MTH211A: Theory of Statistics

Problem set 3

Part I: Completeness and Ancillarity

1. Show that the following families are not complete:
   (a) Poisson distribution with parameter θ, where $\theta \in \{0, 1\}$.
   (b) The family of distribution of $(\bar{X}, S_n^2)$, when $\{X_1, \ldots, X_n\}$ is a random sample from normal distribution with mean θ and θ^2.
   (c) Uniform distribution supported on $(-\theta, \theta)$ where $\theta > 0$.

   [Hint: To show a distribution is not complete, you need to find a function g such that $E_\theta(g(T)) = 0$, but $P(g(T) = 0) < 1$.]

2. Find a complete-sufficient statistic for θ (or $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$) when $X_1, \ldots, X_n$ are random samples from each of the following distributions:
   (a) X has the pdf $f_X(x) = 2x\theta^{-1}$, with $0 < x < \theta$ and $\theta > 0$.
   (b) X has the pdf $f_X(x) = e^{(x-\theta)} \exp[-\exp\{-(x-\theta)\}]$, with $x \in \mathbb{R}$ and $\theta \in \mathbb{R}$.
   (c) The distribution of $X_{(1)}$ when X is distributed as location exponential family (see PS2) with parameter θ.
   (d) Inverse Gaussian distribution with parameters (θ_1, θ_2) (see PS2).

3. For a `Normal`(μ, σ^2) population with σ^2 known, show that $\bar{X}$ and S_n^2 are independent, where S_n^2 is the sample variance.

4. The random variable X takes 3 values $0, 1, 2$ according to one of the following distributions:

	$P(X=0)$	$P(X=1)$	$P(X=0)$	Range of p
Distribution 1	p	$3p$	$(1-4p)$	$(0, 0.25)$
Distribution 2	p	p^2	$(1-p-p^2)$	$(0, 0.50)$

   In each case determine if the family of distributions is complete.

5. Let $X_1, \ldots, X_n$ be n random samples from some distribution with mean μ and finite variance σ^2.
   (a) Show that both $T_1 = (X_1 - X_2)^2/2$ and $T_2 = S_n^{\star 2}$ are unbiased for σ^2.
   (b) Show that both X_1 and $\bar{X}_n$ are unbiased estimators of μ. Is the statistic $T(\mathbf{X}) = E(X_1 \mid \bar{X}_n = \bar{x})$ unbiased for μ?

6. Let Y_1 and Y_2 be two independent unbiased estimators of θ. Assume that variance of Y_1 is twice the variance of Y_2. Find the constants k_1 and k_2 so that $k_1 Y_1 + k_2 Y_2$ is an unbiased estimator of θ with smallest possible variance for such a linear combinations.

7. Consider a random sample $X_1, \ldots, X_n$ which belongs to a location-scale family of the form

$$X_i = \theta_1 + \theta_2 W_i, \quad i = 1, \ldots, n,$$

where W_is are iid with the common pdf (or pmf) f_W, free of (θ_1, θ_2).

   (a) What property should a statistic $S(\cdot)$ would satisfy so that it becomes ancillary for both θ_1 and θ_2?

   (b) Using part (a), show that for a random sample of size n from $\texttt{Normal}(\mu, \sigma^2)$, the distribution of $S_i(\mathbf{X}) = (X_i - \bar{X})/S_n$ is independent of $\bar{X}_n$.

8. Let $X_1, \ldots, X_n$ be a random sample from $\texttt{Exponential}(\theta)$ distribution. Show that the statistics

$$T_1(\mathbf{X}) = \frac{X_1 + X_2}{\sum_{i=1}^n X_i} \qquad \text{and} \qquad T_2(\mathbf{X}) = \sum_{i=1}^n X_i$$

are independent.

9. Show that $Y = |X|$ is a complete sufficient statistic for $\theta > 0$, where X is a random sample $(n = 1)$ from a $\texttt{Uniform}(-\theta, \theta)$ distribution, $\theta > 0$. Further, show that Y is independent of $Z = \text{sign}(X)$.

Part II: UMVUE

10. Let X follows some distribution with p.m.f. as follows:

$$f_X(-1;\theta) = \theta, \quad f_X(x;\theta) = \theta^x(1-\theta)^2, \quad x = 0, 1, \ldots, \text{ and } f_X(x;\theta) = 0 \text{ otherwise}, \ 0 < \theta < 1.$$

Show that T is an unbiased estimator of 0 (based on a sample X) iff $T(x) = -xT(-1)$ for each $x = 0, 1, \ldots$.

11. Show that if T_j is the UMVUE of $g_j(\theta)$ for $j = 1, \ldots, k$, then $\sum_{j=1}^k c_j T_j$ is the UMVUE of $\sum_{j=1}^k c_j g_j(\theta)$ for real constants $c_1, \ldots, c_k$.

12. Let $X_1, \ldots, X_n$ be i.i.d. samples from the following distributions. In each case, find the UMVUE of θ.

   (a) $\texttt{Bernoulli}(p)$, where (i) $\theta = p(1-p)$, (ii) $\theta = P(X_1 + \ldots + X_5 = k)$, and k is a positive integer less than or equal to 5, (iii) $\theta = p + (1-p)e^2$.

   (b) $\texttt{Uniform}(0, \alpha)$ and $\theta = \alpha^k$ where k is an integer bigger than $-n$.

   (c) $\texttt{Location} - \texttt{scale Exponential}(\mu, \sigma)$ with p.d.f.

$$f_{\mathbf{X}}(x;\mu,\sigma) = \begin{cases} \sigma^{-1}\exp\{-\sigma^{-1}(x-\mu)\} & x > \mu, \\ 0 & \text{otherwise,} \end{cases}$$

   and (i) $\theta = \mu$ when σ is known, (ii) $\theta = \sigma$ when μ is known, (iii) $\theta = \mu$ when σ is unknown, (iv) $\theta = \sigma$ when μ is unknown.

   [Hint: For the part (iii) and (iv), use the fact that $\mathbf{T}(\mathbf{X}) = (X_{(1)}, \sum_{i=1}^n X_i)$ is complete-sufficient for (μ, σ)]

   (d) $\texttt{Normal}(\mu, \sigma^2)$, (i) $\theta = \exp\{2\mu\}$ when σ^2 is known, (ii) $\theta = \mu^2$, (iii) $\theta = \sigma^p$, where $p > 0$.

   (e) $\texttt{Negative Binomial}(r, p)$ distribution with p.m.f.

$$f_X(x) = \begin{cases} \binom{x+r-1}{r-1} p^r (1-p)^x, & x = 0, 1, \ldots, \\ 0 & \text{otherwise,} \end{cases}$$

   and $\theta = p^{-2}$ when r is known.

13. Let $X_1, \ldots, X_n$ be a random sample from $\texttt{Normal}(\theta, \theta^2)$ distribution. Show that $\bar{X}_n$ can not be the UMVUE of θ.

14. Let $X_1, \ldots, X_n$ be independent samples from a Gaussian polynomial regression model such that $X_i \sim N(\alpha t_i + \beta t_i^2, 1)$ where $t_i,\ i = 1, \ldots, n$ are known constants. Find UMVUE of α and β.

15. Let $Y_i,\ i = 1, \ldots, n$ be the order statistics of a random sample of size $n = 2m + 1$ from $\texttt{Uniform}(0, \theta)$ distribution. Show that, $T = 2Y_m$ is an unbiased estimator of θ. Find the UMVUE of θ by Rao-Blackwellizing T.

16. Let $W_1, \ldots, W_k$ be unbiased estimators of a parameter θ with known variances $\text{var}(W_i) = \sigma_i^2,\ i = 1, \ldots, k$. Find the best unbiased estimator of θ of the form $\sum_{i=1}^k a_i W_i$.

17. Suppose that when the radius of a circle is measured, a random error is made, which is modeled as $N(0, \sigma^2)$. If n repeated independent measurements are made, then find an unbiased estimator of area of the circle. Is it the UMVUE?

18. For a random sample of size n from the $\texttt{Poisson}(\lambda)$ distribution, find $E\left(S_n^{\star 2} \mid \bar{X}_n\right)$.

19. Consider a random sample of size n from the $\texttt{Binomial}(m, p)$ distribution, where m is known. Find the UMVUE of $p(1-p)^{m-1}$.

20. Consider the $\texttt{Normal}(\mu, \sigma^2)$ distribution.

    (a) Among the estimators S_n^2 and $S_n^{\star 2}$ of σ^2, which one has a lower MSE?

    (b) Among the estimators S_n^2 and $S_n^{\star 2}$ of σ^2, which is the UMVUE? Why?

    (c) Find the value of $c > 0$ for which the class of estimators $S_c(\mathbf{X}) = c \sum_{i=1}^n (X_i - \bar{X})^2$ has the lowest MSE.

Part III: CRLB

21. Let $X_1, \ldots, X_n$ are i.i.d. samples from the following distribution. Find the Cramér Rao lower bound of variance of the estimators $T(\mathbf{X})$:

    (a) Gamma(α, β) and $T(\mathbf{X}) = \bar{X}_n$, when α is known.

    (b) Normal(μ, σ^2) and (i) $T(\mathbf{X}) = n^{-1} \sum_{i=1}^n |X_i - \bar{X}_n|$ when μ is known, (ii) $T(\mathbf{X}) = \bar{X}_n^2 - \sigma^2/n$ when σ is known.

22. Show that for a scale family with scale parameter θ the Fisher information is of the form c/θ^2. Further, show that the Fisher information of $\xi = \log(\theta)$ is free of θ.

23. Let $X_1, \ldots, X_n$ be an i.i.d. sample from the following distributions. In each case, find the Fisher information matrix, $\mathbf{I}_n(\boldsymbol{\theta})$, for the parameter $\boldsymbol{\theta}$:

    (a) $\texttt{Gamma}(\alpha, \beta)$ and $\boldsymbol{\theta} = (\alpha, \beta)$.

    (b) $\texttt{Normal}(\mu, \sigma^2)$ and $\boldsymbol{\theta} = (\alpha, \beta)$.
    [Note: The Fisher information matrix for a vector valued parameter $\boldsymbol{\theta}$ is defined as

$$\mathbf{I}_n(\boldsymbol{\theta}) = E\left[\left(\frac{\partial}{\partial \boldsymbol{\theta}} \log f_{\mathbf{X}}(\mathbf{X}; \boldsymbol{\theta})\right)\left(\frac{\partial}{\partial \boldsymbol{\theta}} \log f_{\mathbf{X}}(\mathbf{X}; \boldsymbol{\theta})\right)^\top\right] = -E\left[\left(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \log f_{\mathbf{X}}(\mathbf{X}; \boldsymbol{\theta})\right)\right]]$$

    (c) From the above definition of $\mathbf{I}_n(\boldsymbol{\theta})$ show that, when $X_1, \ldots, X_n$ are i.i.d., then $\mathbf{I}_n(\boldsymbol{\theta}) = n\mathbf{I}_1(\boldsymbol{\theta})$, where $\mathbf{I}_1(\boldsymbol{\theta})$ is the Fisher information matrix for one sample.

24. Let $X_1, \ldots, X_n$ be a random sample from $\texttt{Gamma}(\alpha, \beta)$ distribution with α known, and $T = \sum_{i=1}^{n} X_i$. Show that

$$E\left(X_{(1)} \mid T\right) = T \frac{E\left(X_{(1)}\right)}{E\left(T\right)}.$$

25. Let $X_1, \ldots, X_n$ be a random sample from $\texttt{Normal}(0, \theta)$ distribution.

    (a) Find UMVUE of $\sqrt{\theta}$.

    (b) Is the UMVUE an efficient estimator of $\sqrt{\theta}$?

26. Let $X \sim \texttt{Beta}(\alpha, 2)$ distribution, show that

$$E(\log X) = \frac{\Gamma'(\alpha+2)}{\Gamma(\alpha+2)} - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}, \quad \text{where } \Gamma'(u) = \frac{\partial}{\partial u}\Gamma(u).$$