

THE UNIVERSITY OF NEW SOUTH WALES

DEPARTMENT OF STATISTICS

Additional Exercises for MATH5905, Statistical Inference

Part two: Data reduction. Sufficient statistics. Classical estimation

Sufficiency

1. Use the factorization criterion to find a sufficient statistic for the parameter when X_1, X_2, \dots, X_n are independent random variables each with distribution
 - a) $N(\mu, 1)$,
 - b) $N(0, \sigma^2)$,
 - c) Uniform $(\theta, \theta + 1)$,
 - d) Poisson (λ) .

Check your answer in (d) using the definition of sufficiency.

2. Let X_1, X_2, X_3 be a sample of size 3 from the Bernoulli (p) distribution. Consider the 2 statistics $S = X_1 + X_2 + X_3$ and $T = X_1 X_2 + X_3$. Show that S is sufficient for p and T is not.
3. A random variable $X = (X_1, X_2)$ has the following distribution (with $1 < \theta < 3$):

(x_1, x_2)	(0,0)	(0,1)	(1,0)	(1,1)
$P(X_1 = x_1, X_2 = x_2)$	$\frac{1}{12}(12 - 7\theta + \theta^2)$	$\frac{\theta}{12}(4 - \theta)$	$\frac{\theta}{12}(4 - \theta)$	$\frac{\theta}{12}(\theta - 1)$

Check whether $X_1 + X_2$ or $X_1 X_2$ is sufficient for θ .

4. If X_1, X_2, \dots, X_n are independent Bernoulli (p) random variables, prove that X_1 is not sufficient for p .
5. Given that θ is an integer and that X_1 and X_2 are independent random variables which are Uniformly distributed on the integers $1, 2, \dots, \theta$, prove that $X_1 + X_2$ is not sufficient for θ .
6. Suppose X_1, X_2, \dots, X_n are independent discrete random variables each with probability function $f(x; \theta), \theta$ unknown. Prove that $(X_1, X_2, \dots, X_{n-1})$ is not sufficient for θ .
7. Find a minimal sufficient statistic for the parameter when X_1, X_2, \dots, X_n are independent random variables each with distribution
 - a) Poisson (λ) ,
 - b) $N(0, \sigma^2)$,
 - c) Gamma (α) , (With a density $f(x, \alpha) = \frac{1}{\Gamma(\alpha)} \exp(-x)x^{\alpha-1}, x > 0$. (Here the Gamma function is defined as $\Gamma(\alpha) = \int_0^\infty e^{-x}x^{\alpha-1}dx$ and has the property $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$. In particular, for a natural number $n : \Gamma(n + 1) = n!$ holds).
 - d) Uniform $(0, \theta)$.
 - e) Uniform $(\theta, \theta + 1)$.
 - f) Uniform (θ_1, θ_2) .

8. If X_1, X_2, \dots, X_n are i.i.d. random variables with densities $f(x; \theta)$ given below, find a sufficient statistics for θ .
- $f(x; \theta) = \theta x^{\theta-1} I_{(0,1)}(x), \theta \in (0, \infty)$.
 - $f(x; \theta) = \frac{1}{6\theta^4} x^3 e^{-x/\theta} I_{(0,\infty)}(x), \theta \in (0, \infty)$.
9. Show that the minimal sufficient statistic T_n for the parameter σ of the Scale-Cauchy family $f(x, \sigma) = \frac{\sigma}{\pi(x^2 + \sigma^2)}$ has dimension n and is equal to $T_n = (X_{(1)}^2, X_{(2)}^2, \dots, X_{(n)}^2)$ where $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ is the variation sequence.
10. Let X_1, X_2, \dots, X_n be i.i.d. observations from a scale parameter family $\{F_\sigma(X)\}, \sigma > 0$ with $F_\sigma(x) = F(x/\sigma), \sigma > 0 (F(\cdot))$ - a given continuous cumulative distribution function.) Show that any statistic that depends on the sample through the $n - 1$ values $X_1/X_n, X_2/X_n, \dots, X_{n-1}/X_n$ is an ancillary statistic.

Cramer-Rao Bound. UMVUE

11. Calculate the Cramer-Rao lower bound for the variance of an unbiased estimator of θ and find a statistic with variance equal to the bound when X_1, X_2, \dots, X_n are independent random variables each with distribution
- $f(x, \theta) = \frac{1}{\theta} e^{-x/\theta}, x > 0,$
 - Bernoulli (θ),
 - $N(\theta, 1),$
 - $N(0, \theta).$
- e) Prove that no unbiased estimator of θ has variance equal to the bound when the distribution is $N(0, \theta^2).$
12. If X_1, X_2, \dots, X_n are independent Poisson (λ) random variables, find the *umvue* of $e^{-2\lambda}$. Check that the estimator has mean $e^{-2\lambda}$ and compare the variance of the estimator with the Cramer-Rao lower bound for the variance of an unbiased estimator of $e^{-2\lambda}$.
13. Suppose random variables X and Y have joint density

$$f_{X,Y}(x, y) = 8xy, 0 < y < x, 0 < x < 1.$$

For this pair of random variables, verify directly the lemma which states that if $a(x) = \mathbf{E}(Y|X = x)$, then $\mathbf{E}a(X) = \mathbf{E}(Y)$ and $\text{Var}\{a(X)\} \leq \text{Var}(Y)$.

14. Find the *umvue* of θ^2 when X_1, X_2, \dots, X_n are independent Bernoulli (θ) random variables. Check that your estimator does have mean θ^2 .
15. Find the *umvue* of θ^2 when X_1, X_2, \dots, X_n are independent random variables each with density

$$f(x; \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, x > 0; \theta > 0.$$

Hint: consider \bar{X}^2 .

16. Suppose X_1, X_2, \dots, X_n are independent Uniform $(0, \theta)$ random variables.
- Find the *umvue* of θ^2 and calculate its variance.
 - Find the *umvue* of $\frac{1}{\theta}$.

17. Suppose X_1, X_2, \dots, X_n are independent random variables, each with density

$$f(x; \theta) = \theta e^{-\theta x}, x > 0, \theta > 0. \text{ Let } T = \sum_{i=1}^n X_i.$$

- Prove (*) that the density of T is given by $f(t; \theta) = \frac{1}{\Gamma(n)} \theta^n t^{n-1} e^{-\theta t}, t > 0$.
- Prove that the indicator function of the event $\{X_1 > k\}$ is an unbiased estimator of $e^{-k\theta}$, where k is a known constant.
- If $T = \sum_{i=1}^n X_i$, take for granted (or try to prove using a) (*)) that the conditional density of X_1 given $T = t$ is

$$f(x_1|t) = \frac{(n-1)}{t} \left(1 - \frac{x_1}{t}\right)^{n-2}, 0 < x_1 < t < \infty.$$

Then find the *umvue* of $e^{-k\theta}$.

18. The random variable X takes values 0,1,2,3 with probabilities

$\frac{P(X=0)}{2\theta^2}$	$\frac{P(X=1)}{\theta - 2\theta^3}$	$\frac{P(X=2)}{\theta^2}$	$\frac{P(X=3)}{1 + 2\theta^3 - 3\theta^2 - \theta}$
----------------------------	-------------------------------------	---------------------------	---

The range of the parameter θ is $\theta \in \Theta = (0, \frac{1}{5})$. Is the family of distributions $\{P_\theta(X)\}, \theta \in \Theta$, complete? Give reasons for your answer.

19. Is the following statistic complete:

- $T = \bar{X}$ when the random variables $\mathbf{X} = (X_1, X_2, \dots, X_n)$ are i.i.d. $N(0, \theta)$.
- $T = \sum_{i=1}^n X_i$ when the random variables $\mathbf{X} = (X_1, X_2, \dots, X_n)$ are i.i.d. Bernoulli (θ).
- $T = \sum_{i=1}^n X_i$ when the random variables $\mathbf{X} = (X_1, X_2, \dots, X_n)$ are i.i.d. Poisson (θ).
- $T = X_{(n)}$ when the random variables $\mathbf{X} = (X_1, X_2, \dots, X_n)$ are i.i.d. Uniform $(0, \theta)$.

Answers:

- a) bound: $\frac{\theta^2}{n}$, UMVUE: \bar{X} ;
- b) bound: $\frac{\theta(1-\theta)}{n}$, UMVUE: \bar{X} ;
- c) bound: $\frac{1}{n}$, UMVUE: \bar{X} ;
- d) bound: $\frac{2\theta^2}{n}$, UMVUE: $\frac{\sum_{i=1}^n X_i^2}{n}$;
- e) bound: $\frac{\theta^2}{2n}$, the score is $-\frac{n}{\theta} + \frac{\sum_{i=1}^n X_i^2}{\theta^3}$ and can not be written as $K(\theta, n)(T - \theta)$.
- $(1 - \frac{2}{n})^{\sum_{i=1}^n X_i}$, CR bound: $\frac{4\lambda e^{-4\lambda}}{n}$ and is smaller than the variance of the UMVUE.
- $Var(a(X)) = \frac{8}{675}, Var(Y) = \frac{11}{225}$.
- $\frac{T(T-1)}{n(n-1)}, T = \sum_{i=1}^n X_i$.
- $\frac{T^2}{n(n+1)}, T = \sum_{i=1}^n X_i$.
- a) $\frac{(n+2)T^2}{n}, T = X_{(n)}$.
- UMVUE: $\{\frac{T-k}{T}\}^{n-1} I_{(k, \infty)}(T), T = \sum_{i=1}^n X_i$.
- Not complete.
- a) Not complete; b) Complete; c) Complete; d) Complete.

MLE and their properties. Asymptotic properties of estimators

20. A sample of size n_1 is to be drawn from a normal population with mean μ_1 and variance σ_1^2 . A second sample of size n_2 is to be drawn from a normal population with mean μ_2 and variance σ_2^2 . What is the MLE of $\theta = \mu_1 - \mu_2$? If we assume that the total sample size $n = n_1 + n_2$ is fixed, how should the n observations be divided between the two populations in order to minimize the variance of the MLE?

21. Let X_1, X_2, \dots, X_n be a sample from the density $f(x; \theta) = \theta x^{-2} I_{[\theta, \infty)}(x)$ where $\theta > 0$.

- a) Find the MLE of θ .
- b) Is $Y = X_{(1)}$ a sufficient statistic?
22. Let X_1, X_2, \dots, X_n be a sample from the density function $f(x; \theta) = \theta x^{\theta-1} I_{(0,1)}(x)$ where $\theta > 0$.
- a) Find the MLE of $\tau(\theta) = \frac{\theta}{1+\theta}$.
- b) State the asymptotic distribution of the MLE of $\tau(\theta)$ in a).
- c) Find a sufficient statistic, and check completeness. Is $\sum_{i=1}^n X_i$ a sufficient statistic?
- d) Is there a function of θ for which there exists an unbiased estimator whose variance coincides with the Cramer-Rao lower Bound? What is the Cramer-Rao lower bound?
23. Let X_1, X_2, \dots, X_n be a sample from normal distribution $N(\mu, \sigma^2)$ where μ is known and σ^2 is the parameter to be estimated.
- a) Find the MLE and state its asymptotic distribution.
- b) Assume now that σ is to be estimated. Find the MLE and state its asymptotic distribution.
24. Consider n i.i.d. observations from a Poisson (λ) distribution.
- a) Suppose the parameter of interest is $\tau(\lambda) = \frac{1}{\lambda}$.
- i) What is the MLE of $\tau(\lambda)$?
- ii) What is its variance?
- iii) What is its asymptotic variance?
- b) Assume that the parameter of interest is $\tau(\lambda) = \sqrt{\lambda}$.
- i) State the asymptotic distribution of the MLE of $\sqrt{\lambda}$. In particular, show that the asymptotic variance does not depend on λ (we say in that case that $\sqrt{\lambda}$ is a *variance stabilising transformation*).
- ii) For a given small value of $\alpha \in (0, 1)$ and using the result in i), how would you construct a confidence interval for λ that asymptotically has a level $1 - \alpha$.
- Answers:**
20. MLE: $\bar{X}_{n_1} - \bar{Y}_{n_2}; n_1 = \frac{\sigma_1}{\sigma_1 + \sigma_2} n$
21. $\hat{\theta}_{mle} = X_{(1)}$, sufficient.
22. a) $\frac{n}{n - \sum_{i=1}^n \log X_i}$; b) $N(0, \frac{\theta^2}{(1+\theta)^4})$
- c) $\sum_{i=1}^n \log X_i$ is sufficient and complete; $\sum_{i=1}^n X_i$ is **not** sufficient.
- d) $\tau(\theta) = \frac{1}{\theta}$ is such a function. The attainable bound in this case is $\frac{1}{n\theta^2}$.
23. a) MLE of σ^2 is $\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$. The asymptotic distribution: $N(0, 2\sigma^4)$.
- b) MLE is $\sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2}$. The asymptotic distribution is $N(0, \frac{1}{2}\sigma^2)$.
24. a) Variance of MLE is infinite but **asymptotic** variance of MLE is finite and equals $\frac{1}{\lambda^3}$.
- b) Asymptotic variance is 0.25.