

STA 332/MATH 343: Homework 4 – Solutions

Duke University, Fall 2026

Due: October 3rd, 11:59 PM on Gradescope

UMVUE, Completeness, and Exponential Families

1. Review: give brief but formal definitions of the following concepts for a statistical model $\{f_\theta : \theta \in \Theta\}$:

- (a) Ancillary statistic
- (b) Sufficient statistic
- (c) Point estimator of $g(\theta)$;
- (d) Unbiased estimator of $g(\theta)$;
- (e) Uniformly minimum-variance unbiased estimator of $g(\theta)$

- (a) A statistic $A = A(\underline{X})$ is ancillary if its distribution does not depend on θ .
- (b) A statistic $T = T(\underline{X})$ is sufficient if the conditional distribution of the full sample $\underline{X}$ given T does not depend on θ .
- (c) A point estimator of $g(\theta)$ is a statistic $U = U(\underline{X})$ used to estimate $g(\theta)$.
- (d) It is unbiased if

$$\mathbb{E}_\theta[U] = g(\theta) \quad \text{for every } \theta \in \Theta.$$

- (e) An estimator U^* is a UMVUE of $g(\theta)$ if it is unbiased and, for every other unbiased estimator U of $g(\theta)$,

$$\text{Var}_\theta(U^*) \leq \text{Var}_\theta(U) \quad \text{for every } \theta \in \Theta.$$

2. Properties of conditional expectations

- (a) Let U be a discrete random variable. Show that $\mathbb{E}[f(U) \mid U] = f(U)$.
- (b) Let X and Y be independent continuous random variables. Show that $\mathbb{E}[X \mid Y] = \mathbb{E}[X]$.
- (c) Let $X_1, X_2 \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, 1)$, so that $U = (X_1 + X_2)/2$ is an unbiased estimator for θ . Let $S = X_2$ and compute $\mathbb{E}_\theta[U \mid S]$. Is it a statistic? What can we conclude about whether S is sufficient?

- (a) For any u with $\mathbb{P}(U = u) > 0$, the conditional probability $\mathbb{P}(U = v \mid U = u)$ is one when $v = u$ and zero otherwise. Thus

$$\mathbb{E}[f(U) \mid U = u] = \sum_v f(v) \mathbb{P}(U = v \mid U = u) = f(u).$$

Substituting U for u gives $\mathbb{E}[f(U) \mid U] = f(U)$ almost surely.

- (b) Independence gives $f_{X,Y}(x, y) = f_X(x)f_Y(y)$. For any y with $f_Y(y) > 0$, we therefore have $f_{X|Y}(x | y) = f_X(x)$, and

$$\mathbb{E}[X | Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x | y) dx = \int_{-\infty}^{\infty} x f_X(x) dx = \mathbb{E}[X].$$

Substituting Y for y gives $\mathbb{E}[X | Y] = \mathbb{E}[X]$ almost surely.

- (c) By linearity and independence,

$$\mathbb{E}_{\theta}[U | S] = \frac{1}{2}(\mathbb{E}_{\theta}[X_1 | X_2] + \mathbb{E}_{\theta}[X_2 | X_2]) = \frac{\theta + X_2}{2}.$$

This depends on the unknown θ , so it is not a statistic. Therefore $S = X_2$ is not sufficient: sufficiency would make this conditional expectation a function of S that does not depend on θ .

3. A factory inspects $n \geq 2$ independently produced components. Let $X_i = 1$ if component i passes inspection and $X_i = 0$ otherwise. Suppose $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Ber}(p)$, where the pass probability $0 < p < 1$ is unknown.

- (a) Show that $\text{Ber}(p)$ is an exponential family, and use this to show $T = \sum_{i=1}^n X_i$ is sufficient and complete.
- (b) Show that $V = (T, X_1)$ is not complete by finding a nonzero function of V whose expectation is zero for every p .
- (c) Find the UMVUE of p^2 by conditioning $X_1 X_2$ on T . Make sure to consider the cases $T < 2$ and $T \geq 2$ separately.

- (a) For $x \in \{0, 1\}$,

$$f_p(x) = p^x(1-p)^{1-x} = \exp\left\{x \log \frac{p}{1-p} + \log(1-p)\right\}.$$

Thus $R(x) = x$, $b(p) = \log(p/(1-p))$, and $c(p) = -\log(1-p)$. The support is fixed and $b(p)$ ranges over $\mathbb{R}$, so $T = \sum_i X_i$ is sufficient and complete by the exponential-family theorem.

- (b) The function

$$a(V) = X_1 - \frac{T}{n}$$

has expectation $p - p = 0$ for every p . It is not almost surely zero: the event $X_1 = 1, X_2 = \dots = X_n = 0$ has positive probability, and there $a(V) = 1 - 1/n \neq 0$. Thus V is not complete.

- (c) Independence gives $\mathbb{E}_p[X_1 X_2] = p^2$. Given $T = t \geq 2$,

$$\begin{aligned} \mathbb{E}_p[X_1 X_2 | T = t] &= \mathbb{P}_p(X_1 = 1, X_2 = 1 | \sum_{i=1}^n X_i = t) \\ &= \frac{\mathbb{P}_p(X_1 = 1, X_2 = 1, \sum_{i=3}^n X_i = t-2)}{\mathbb{P}_p(T = t)} \\ &= \frac{p^2 \binom{n-2}{t-2} p^{t-2} (1-p)^{n-t}}{\binom{n}{t} p^t (1-p)^{n-t}} = \binom{n-2}{t-2} / \binom{n}{t} \end{aligned}$$

$$\begin{aligned}
&= \frac{(n-2)! / ((n-t)!(t-2)!)}{n! / ((n-t)!t!)} = \frac{(n-2)!t!}{n!(t-2)!} \\
&= \frac{t(t-1)}{n(n-1)}
\end{aligned}$$

We could have also gotten to the second equality in the third line by noting that each placement of t successes among n positions is equally likely, and of those $\binom{n-2}{t-2}$ of them put two successes in the first two slots. When $t < 2$, we have $\mathbb{E}_p[X_1 X_2 \mid T = t] = \mathbb{P}_p(X_1 = 1, X_2 = 1 \mid \sum_{i=1}^n X_i = t) = 0$. Therefore,

$$\mathbb{E}_p[X_1 X_2 \mid T] = \begin{cases} \frac{T(T-1)}{n(n-1)}, & T \geq 2, \\ 0, & T < 2 \end{cases}$$

is the UMVUE for p^2 . It is an unbiased function of the complete sufficient statistic T .

4. A workshop cuts n square panels with dimensions $X_1 \times X_1, X_2 \times X_2, \dots, X_n \times X_n$, where $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Unif}(0, \theta)$, and $\theta > 0$ is the largest permitted side length. We want to estimate θ^2 , the area of the largest permitted panel. Let

$$U = 3X_1^2 \quad \text{and} \quad M = X_{(n)}.$$

- (a) Show that U is unbiased for θ^2 .
 (b) Compute the UMVUE for θ^2 by conditioning U on M . You may use the following conditional description:

$$(X_1 \mid M = m) = \begin{cases} m, & \text{with prob. } \frac{1}{n}, \\ \text{Unif}(0, m), & \text{with prob. } 1 - \frac{1}{n}. \end{cases}$$

- (a) Since

$$\mathbb{E}_\theta[X_1^2] = \frac{1}{\theta} \int_0^\theta x^2 dx = \frac{\theta^2}{3},$$

$U = 3X_1^2$ is unbiased for θ^2 .

- (b) Using the given conditional description,

$$\mathbb{E}_\theta[X_1^2 \mid M = m] = \frac{1}{n}m^2 + \frac{n-1}{n} \frac{m^2}{3} = \frac{n+2}{3n}m^2.$$

Therefore,

$$U^* = \mathbb{E}_\theta[3X_1^2 \mid M] = \frac{n+2}{n}M^2.$$

The tower property gives $\mathbb{E}_\theta[U^*] = \mathbb{E}_\theta[U] = \theta^2$. Since M is complete and sufficient, as established in Lecture 10, Lehmann–Scheffé shows that U^* is the UMVUE of θ^2 .

5. A maintenance team records independent repair durations $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Gamma}(\alpha, 1/\theta)$, where $n \geq 2$, the shape $\alpha > 0$ is known, and the scale $\theta > 0$ is unknown. Write $S = \sum_i X_i$ and $M = X_{(n)}$. Recall from Lecture 5 and HW2 that $X_i = \theta Z_i$ with $Z_i \stackrel{\text{i.i.d.}}{\sim} \text{Gamma}(\alpha, 1)$, and

$$f_\theta(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)} \mathbb{1}_{\{x>0\}} \exp\left\{-\frac{x}{\theta} - \alpha \log \theta\right\}.$$

- (a) Use this representation to establish that S is complete and sufficient, and hence minimal sufficient.
- (b) Show that (M, S) is not complete. *Hint: consider the function $a(M, S) = M/S$. Show its expectation is some c which does not depend on θ .*
- (c) Find the UMVUE of θ . You may use $S \sim \text{Gamma}(n\alpha, 1/\theta)$.
- (d) Now suppose $n\alpha > 2$. Compute $\mathbb{E}_\theta[S^{-1}]$ and $\mathbb{E}_\theta[S^{-2}]$, and find the UMVUE of the rate $1/\theta$. Is the reciprocal of the UMVUE from part (c) also unbiased for $1/\theta$?

- (a) Here $R(x) = x$ and $b(\theta) = -1/\theta$ ranges over $(-\infty, 0)$. The support is fixed, so S is complete and sufficient by the exponential-family theorem, and therefore minimal sufficient.

- (b) We have

$$\frac{M}{S} = \frac{\max_i Z_i}{\sum_i Z_i},$$

so this ratio has a parameter-independent distribution. It is nonconstant for $n \geq 2$: the joint Gamma density is positive on $(0, \infty)^n$, and the ratio varies on open sets. Thus $M/S - c$ is a nonzero mean-zero function of (M, S) , proving noncompleteness.

- (c) Write $a = n\alpha$. Since $\mathbb{E}_\theta[S] = a\theta$, S/a is unbiased for θ . It is a function of the complete sufficient statistic S , so it is the UMVUE by Lehmann–Scheffé.
- (d) Recognizing Gamma kernels gives, for $a > r$,

$$\mathbb{E}_\theta[S^{-r}] = \frac{1}{\Gamma(a)\theta^a} \int_0^\infty s^{a-r-1} e^{-s/\theta} ds = \frac{\Gamma(a-r)}{\Gamma(a)\theta^r}.$$

In particular,

$$\mathbb{E}_\theta[S^{-1}] = \frac{1}{(a-1)\theta}, \quad \mathbb{E}_\theta[S^{-2}] = \frac{1}{(a-1)(a-2)\theta^2}.$$

Thus $(a-1)/S$ is unbiased for $1/\theta$. Since it is a function of the complete sufficient statistic S , it is the UMVUE by Lehmann–Scheffé. In contrast, a/S , the reciprocal of S/a , has expectation $a/((a-1)\theta)$ and is biased upward.

6. A sensor gives independent signed readings $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, \theta^2)$ of a signal, where $\theta > 0$ is the mean response and the standard deviation is also θ . Assume $n \geq 2$. Define $T_1 = \sum_i X_i$ and $T_2 = \sum_i X_i^2$.

- (a) Show that $\mathcal{N}(\theta, \theta^2)$ is not a one-parameter exponential family.
- (b) Show that θ is a scale parameter and that (T_1, T_2) is minimal sufficient.
- (c) Compute $\mathbb{E}_\theta[T_1]$, $\mathbb{E}_\theta[T_1^2]$, and $\mathbb{E}_\theta[T_2]$.
- (d) Use your answers to construct a nonzero function of (T_1, T_2) with expectation zero for every θ . What does this say about completeness?

- (a) Expanding the density gives

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{x}{\theta} - \frac{x^2}{2\theta^2} - \log \theta - \frac{1}{2}\right\}.$$

The two terms involving both x and θ cannot be combined into one product $b(\theta)R(x)$.

- (b) If $Z_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(1, 1)$, then $\theta Z_i \sim \mathcal{N}(\theta, \theta^2)$, so θ is a scale parameter. The joint density is

$$f_{\theta}(\underline{x}) = (2\pi)^{-n/2} \theta^{-n} \exp \left\{ \frac{T_1(\underline{x})}{\theta} - \frac{T_2(\underline{x})}{2\theta^2} - \frac{n}{2} \right\}.$$

Factorization makes (T_1, T_2) sufficient. For two sample points $\underline{x}, \underline{y}$, write $\Delta_j = T_j(\underline{x}) - T_j(\underline{y})$. Their likelihood ratio is

$$\frac{f_{\theta}(\underline{x})}{f_{\theta}(\underline{y})} = \exp \left\{ \frac{\Delta_1}{\theta} - \frac{\Delta_2}{2\theta^2} \right\}.$$

This is independent of $\theta > 0$ if and only if the polynomial $\Delta_1 u - \Delta_2 u^2/2$ is constant for $u = 1/\theta > 0$, which holds if and only if $\Delta_1 = \Delta_2 = 0$. The likelihood-ratio criterion therefore shows that (T_1, T_2) is minimal sufficient.

- (c) Since $\mathbb{E}_{\theta}[X_i] = \theta$ and $\text{Var}_{\theta}(X_i) = \theta^2$,

$$\mathbb{E}_{\theta}[T_1] = n\theta, \quad \mathbb{E}_{\theta}[T_1^2] = n\theta^2 + n^2\theta^2 = n(n+1)\theta^2, \quad \mathbb{E}_{\theta}[T_2] = 2n\theta^2.$$

- (d) Take

$$a(T_1, T_2) = \frac{T_1^2}{n(n+1)} - \frac{T_2}{2n}.$$

Part (c) gives $\mathbb{E}_{\theta}[a(T_1, T_2)] = 0$ for every θ . For $n \geq 2$, this polynomial is nonzero when $x_1 = 1$ and $x_2 = \dots = x_n = 0$. By continuity, it remains nonzero on a neighborhood of that point. The joint normal density is positive everywhere, so $a(T_1, T_2)$ is not almost surely zero. Thus (T_1, T_2) is not complete.

7. Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Beta}(\alpha, \beta)$, with density

$$f_{\alpha, \beta}(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \mathbb{1}_{\{0 < x < 1\}}.$$

- (a) Suppose $\alpha > 0$ is known and $\beta > 0$ is unknown. Write the density as a one-parameter exponential family. Identify a complete sufficient statistic.
 (b) Now suppose $\beta > 0$ is known and $\alpha > 0$ is unknown. Repeat part (a).

- (a) With α fixed, write

$$f_{\beta}(x) = x^{\alpha-1} \mathbb{1}_{\{0 < x < 1\}} \exp \left\{ (\beta-1) \log(1-x) + \log \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right\}.$$

Here $R(x) = \log(1-x)$, $b(\beta) = \beta-1$, and $c(\beta) = \log \Gamma(\alpha) + \log \Gamma(\beta) - \log \Gamma(\alpha + \beta)$. The support is fixed and the natural-parameter range is $(-1, \infty)$. Thus $\sum_i \log(1-X_i)$ is complete and sufficient, and hence minimal sufficient.

- (b) With β fixed, write

$$f_{\alpha}(x) = (1-x)^{\beta-1} \mathbb{1}_{\{0 < x < 1\}} \exp \left\{ (\alpha-1) \log x + \log \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right\}.$$

Now $R(x) = \log x$, $b(\alpha) = \alpha-1$, and the natural-parameter range is again $(-1, \infty)$. Thus $\sum_i \log X_i$ is complete and sufficient, and hence minimal sufficient.

8. Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, where μ is known and $\sigma^2 > 0$ is unknown. In HW3 you found the MLE $\hat{\sigma}^2 = n^{-1} \sum_i (X_i - \mu)^2$.

- (a) Write this model as a one-parameter exponential family in σ^2 . Identify a complete sufficient statistic.
 (b) Show that the MLE above is also the UMVUE of σ^2 .

- (a) The density is

$$f_{\sigma^2}(x) = \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 - \frac{1}{2} \log(2\pi\sigma^2) \right\}.$$

Since μ is known, $R(x) = (x - \mu)^2$ is a statistic. The natural parameter $-1/(2\sigma^2)$ ranges over $(-\infty, 0)$ and the support is $\mathbb{R}$. Hence $Q = \sum_i (X_i - \mu)^2$ is complete and sufficient, and therefore minimal sufficient.

- (b) We have $\mathbb{E}[Q] = \sum_i \mathbb{E}[(X_i - \mu)^2] = n\sigma^2$, so Q/n is unbiased for σ^2 . It is a function of the complete sufficient statistic Q , and hence is the UMVUE by Lehmann–Scheffé.

9. Let $X_1, \dots, X_n$ be i.i.d. with density

$$f_\theta(x) = \frac{1}{\theta} (1+x)^{-(1+\theta)/\theta} \mathbb{1}_{\{x>0\}}, \quad \theta > 0.$$

- (a) Determine the distribution of $Y_i = \log(1 + X_i)$.
 (b) Write the original density $f_\theta(x)$ as a one-parameter exponential family. Identify a complete sufficient statistic T for θ , and conclude that it is minimal sufficient.
 (c) Determine the distribution of T and find the UMVUE of θ .
 (d) Find the MLE of θ and compare it with the UMVUE you found.

- (a) For $y > 0$, the inverse transformation is $x = e^y - 1$, with derivative e^y . Thus

$$f_{Y_i}(y) = f_\theta(e^y - 1)e^y = \frac{1}{\theta} e^{-y/\theta} \mathbb{1}_{\{y>0\}}.$$

Therefore $Y_i \sim \text{Exp}(1/\theta)$, with rate $1/\theta$.

- (b) Write

$$f_\theta(x) = \frac{\mathbb{1}_{\{x>0\}}}{1+x} \exp \left\{ -\frac{1}{\theta} \log(1+x) - \log \theta \right\}.$$

Here $R(x) = \log(1+x)$, $b(\theta) = -1/\theta \in (-\infty, 0)$, and $c(\theta) = \log \theta$. The support is fixed, so $T = \sum_i \log(1 + X_i)$ is complete and sufficient, and hence minimal sufficient.

- (c) Gamma additivity gives $T \sim \text{Gamma}(n, 1/\theta)$. Consequently $\mathbb{E}_\theta[T/n] = \theta$. Since T/n is a function of the complete sufficient statistic T , it is the UMVUE of θ by Lehmann–Scheffé.
 (d) The log likelihood is $\ell(\theta) = -n \log \theta - T/\theta - \sum_i \log(1 + X_i)$. Its derivative is $(T - n\theta)/\theta^2$, positive for $\theta < T/n$ and negative for $\theta > T/n$. Since $T > 0$ almost surely, the unique MLE is T/n , which equals the UMVUE.

10. Let $X_1, \dots, X_n$ be i.i.d. with density

$$f_\theta(x) = \frac{1}{x \log(\theta)} \mathbb{1}_{\{1 < x < \theta\}}, \quad \theta > 1,$$

and let $M = X_{(n)}$.

- (a) Prove that M is sufficient for θ .
- (b) Find the cdf and density of M .
- (c) Prove directly that M is complete, and use this to conclude that M is minimal sufficient. *Hint: if $\mathbb{E}_\theta[a(M)] = 0$ for every $\theta > 1$, multiply by $[\log \theta]^n$ and differentiate the resulting integral with respect to its upper limit.*
- (d) Find the UMVUE of $\log \theta$. *Hint: show $\log X_i \sim \text{Unif}(0, \log \theta)$ and use the expectation calculation from Example 3, Lecture 10 without reproving it.*

- (a) The joint density is

$$f_\theta(\underline{x}) = \left(\prod_i \frac{1}{x_i} \mathbb{1}_{\{x_i > 1\}} \right) [\log(\theta)]^{-n} \mathbb{1}_{\{M < \theta\}}.$$

Factorization gives sufficiency of M .

- (b) For $1 < m < \theta$,

$$F_M(m) = \left(\frac{\log m}{\log(\theta)} \right)^n, \quad f_M(m) = \frac{n(\log m)^{n-1}}{m[\log(\theta)]^n}.$$

The cdf is zero for $m \leq 1$ and one for $m \geq \theta$; the density is zero outside $(1, \theta)$.

- (c) Suppose $a(M)$ is integrable with mean zero for every θ . Then

$$\int_1^\theta a(m) \frac{(\log m)^{n-1}}{m} dm = 0 \quad \text{for all } \theta > 1.$$

Writing $u = \theta$, differentiation gives $a(u)(\log u)^{n-1}/u = 0$ for almost every $u > 1$. The weight is positive there, so $a(u) = 0$ almost everywhere on $(1, \infty)$. Every distribution of M has a density on this interval, hence $a(M) = 0$ almost surely for every θ . Thus M is complete and sufficient, and therefore minimal sufficient.

- (d) Write $\eta = \log(\theta) > 0$. For $0 < y < \eta$, $\mathbb{P}_\theta(\log X_i \leq y) = \mathbb{P}_\theta(X_i \leq e^y) = \log(e^y)/\log \theta = y/\eta$. Thus $\log X_i \sim \text{Unif}(0, \eta)$, whose sample maximum is $\log M$. By the uniform-maximum expectation calculation, $\mathbb{E}_\theta[\log M] = n\eta/(n+1)$. Hence

$$\hat{\eta} = \frac{n+1}{n} \log M$$

is unbiased for η . It is a function of the complete sufficient statistic M , so it is the UMVUE of $\log \theta$.