

Lecture 12: Unbiased estimation and UMVUE

Lecturer: Yiqiao Zhong[Report an issue](#)*Last update: October 19, 2026*

These notes support STAT 709 at the University of Wisconsin–Madison. Use them alongside the cited textbooks; corrections and questions are welcome.

Priority guide

High Core concepts and calculations to master; directly examined.

Mid Supporting concepts useful in other lectures; in scope for examinations.

Low Enrichment supporting understanding; will not be examined.

Further reading Optional self-study, references, extensions, and historical notes.

A star (*) marks optional material; priority labels give the scope of each block. Lecture-specific exemptions and the optional status of Part VI override the general exam guide.

Exam scope for this lecture. Section 4 and its linked Exercises 3, 4 and 6 are for understanding and will not be directly tested in exams. Other high- and mid-priority material remains in examination scope.

High priority *Overview*

Conditioning removes variation that carries no information about the parameter. Completeness then makes the remaining unbiased function unique. We develop this two-step route to uniformly minimum variance unbiased estimation, work through Normal, Uniform, Poisson and exponential examples, and distinguish optimality within unbiased rules from optimality among all estimators.

This begins Part IV. The part overview connects these lectures and includes a final self-check.

Lecture 11 judged estimators as $n \rightarrow \infty$; this lecture returns to a fixed n .

Recap from earlier lectures

From Lectures 5–7: conditioning, sufficiency and risk

Reminder only: this block introduces no new exam requirements.

- **Sufficiency.** Given a sufficient statistic, the conditional data law has a version independent of the unknown parameter. Used in Section 1.
- **Factorization and exponential families.** Factorization separates a parameter-dependent function of a statistic from a parameter-free factor. Used in Section 2.
- **Conditional means.** The tower rule preserves expectations, and conditional averaging cannot increase mean squared error. Used in Section 1.
- **Bias and MSE.** Unbiasedness is one constraint; squared-error risk also includes variance. Used in Section 1.

Learning goals

- Compare the MSE of an estimator with that of its conditional average.

- Explain why sufficiency is needed for a parameter-independent averaging rule.
- Use completeness and Lehmann–Scheffé to identify a UMVUE.
- Work through the Section 4 examples for understanding, respecting their exam exemption.

High priority *Introduction*

In previous lectures, we showed that the bias–variance tradeoff is a general phenomenon in statistics. An optimal estimator may not exist or may depend on a case-by-case analysis. However, if we only consider unbiased estimators, there is a systematic way of finding optimal estimators.

1 Motivations

High priority *Motivations*

A well-studied question in estimation theory is the construction of optimal (or near-optimal) estimators. After all, there are many ways to find estimators, but some are better than others. Consider the following simple setting. Suppose that we have data $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{E}(\mu)$ where $\mathcal{E}(\mu)$ is the exponential distribution with parameter μ , i.e., the pdf is

$$f_\mu(x) = \frac{1}{\mu} e^{-x/\mu} \mathbf{1}\{x \geq 0\}.$$

In order to estimate the unknown parameter $\mu > 0$ (assume $n > 2$), a natural estimator is the sample mean $\bar{X}$ since $\mu = \mathbb{E}[X_i]$. But imagine that a portion of the data— X_1, X_2 —is stored in one data center or server, and the rest is stored in another data center. Each data center may estimate μ with

$$\hat{\mu}_1 = \frac{1}{2}(X_1 + X_2), \quad \hat{\mu}_2 = \frac{1}{n-2} \sum_{i>2} X_i.$$

We can combine the two estimators from the two data centers:

$$\hat{\mu}_3 = \frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2).$$

Which one should we choose? The principle of sufficiency says that we should choose an estimator that “sufficiently” summarizes the data. We can check that $\hat{\mu}_1$ and $\hat{\mu}_2$ are not individually sufficient; $\hat{\mu}_3$ is generally not sufficient either (except when $n = 4$, when it equals $\bar{X}$), but $(\hat{\mu}_1, \hat{\mu}_2)$ is a sufficient statistic.

Conditioning reduces variance. If $\mathbb{E}Y^2 < \infty$ and Z is any random variable, then

$$\text{Var}(\mathbb{E}[Y \mid Z]) \leq \text{Var}(Y). \tag{1}$$

This follows from the variance decomposition discussed in Section 4.2 of Lecture 5.

Instructor’s Note *Conditioning and the unknown parameter*

Unbiasedness is unnecessary. Think of conditioning as averaging away some of the estimator’s noise. This averaging keeps the estimator’s mean, and hence its bias, unchanged. Its variance goes down, so its mean squared error goes down too, even if we started with a biased estimator.

Sufficiency makes the averaging rule usable. Write the model parameter as θ (μ in the example above). When we form

$$\mathbb{E}_\theta[T_2(X) \mid T_1(X)] = g_\theta(T_1(X)),$$

the subscript θ matters: the expectation is taken under the probability measure P_θ , which depends on the unknown parameter. Without sufficiency, conditioning can bring that unknown θ into the resulting rule. Then $g_\theta(T_1(X))$ is not a statistic we can calculate from the data alone. Sufficiency removes θ from the conditional law of X given $T_1(X)$, so the averaging produces one function $g(T_1(X))$ that we can actually use. Intuitively, we can average over the variation left after observing T_1 without first knowing the parameter we want to estimate.

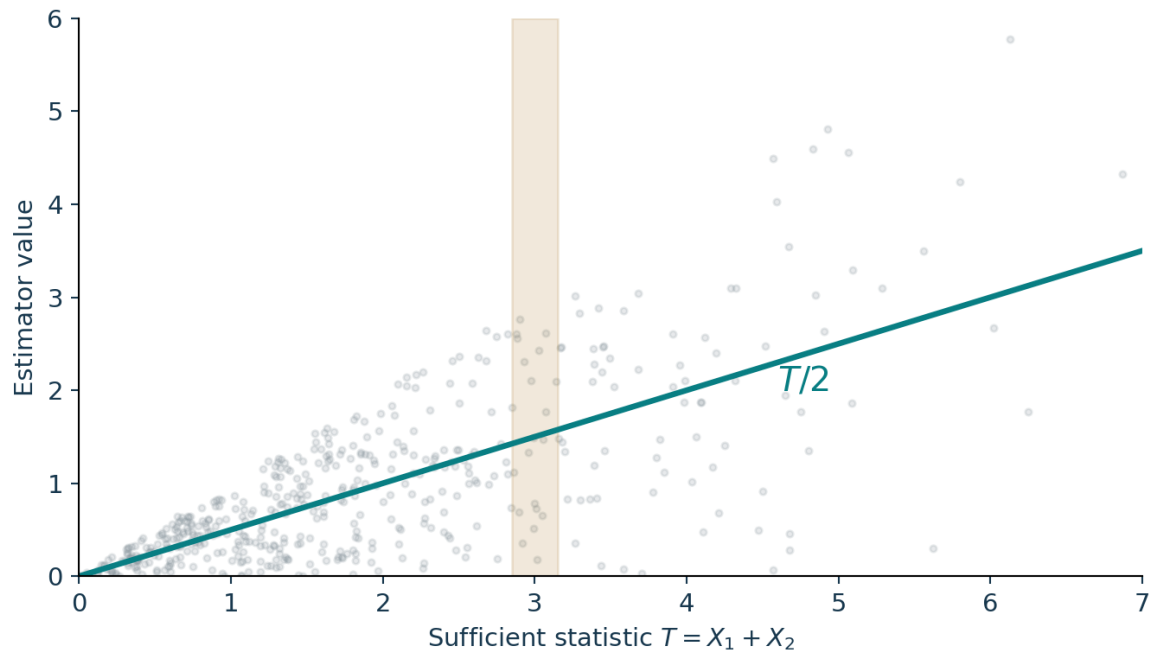
Why the MSE also decreases. Fix a model distribution P_μ , suppose $\mathbb{E}_\mu[T_2(X)^2] < \infty$, and set $g_\mu(T_1) = \mathbb{E}_\mu[T_2(X) \mid T_1(X)]$. For any statistic T_1 , the exact identity is

$$\mathbb{E}_\mu[(T_2 - \mu)^2] = \mathbb{E}_\mu[(g_\mu(T_1) - \mu)^2] + \mathbb{E}_\mu[\text{Var}_\mu(T_2 \mid T_1)]. \quad (2)$$

Consequently, $\mathbb{E}_\mu[(g_\mu(T_1) - \mu)^2] \leq \mathbb{E}_\mu[(T_2 - \mu)^2]$. **Unbiasedness is not required.** The tower rule gives $\mathbb{E}_\mu g_\mu(T_1) = \mathbb{E}_\mu T_2$, so conditioning preserves the bias while reducing variance. If T_2 is unbiased, the conditional version is unbiased too. The reduction is strict at a given μ exactly when $\mathbb{E}_\mu \text{Var}_\mu(T_2 \mid T_1) > 0$; equality holds exactly when $T_2 = g_\mu(T_1)$ almost surely under P_μ . See Exercise 7 for the complete decomposition proof.

What sufficiency contributes. **Sufficiency is not needed for the inequality at a fixed model distribution.** Its role is to make the conditional expectation a usable estimator: a sufficient T_1 allows one common function g , independent of the unknown μ , with $g(T_1) = \mathbb{E}_\mu[T_2 \mid T_1]$ for every μ . Without sufficiency, the function g_μ may depend on the parameter we are trying to estimate. For example, if $X \sim \mathcal{E}(\mu)$, $T_2 = X$ and $T_1 \equiv 0$, then $g_\mu(T_1) = \mu$. Its MSE is zero, but we cannot compute it from the data without already knowing μ . A nonsufficient statistic can still work for a particular T_2 ; sufficiency supplies the general guarantee. This is the **Rao–Blackwell theorem**, which does not require completeness. See Theorem 2.5 of [1] and Shao’s Stat 609 notes, slides 13–15.

Since there are many ways to find sufficient statistics, it is natural to seek a sufficient statistic that is minimal in a certain sense; in the motivating example, both $(\hat{\mu}_1, \hat{\mu}_2)$ and $\bar{X}$ are sufficient statistics, but $\bar{X}$ is simpler because it is a function of $(\hat{\mu}_1, \hat{\mu}_2)$. Completeness is a separate property: it will ensure uniqueness of unbiased functions of a sufficient statistic.



As a separate illustration, let X_1, X_2 be independent exponential observations with unknown mean θ . Their sum T is sufficient, and symmetry gives $\mathbb{E}_\theta[X_1 | T] = T/2$. Both estimate θ without bias, but their variances are θ^2 and $\theta^2/2$. The synthetic scatter uses $\theta = 1$ and 500 pairs; it illustrates within-group variation rather than proving the identity.

Conditioning and variance. For square-integrable Y , $\text{Var}(Y) = \text{Var}(\mathbb{E}[Y | T]) + \mathbb{E} \text{Var}(Y | T)$. Conditioning removes the second nonnegative term while preserving the mean.

Mid priority *Practice: Weighting two independent server means*

Exercise prompt. For the two-server exponential example, minimize the variance of $a\hat{\mu}_1 + (1-a)\hat{\mu}_2$ over real a .

See Exercise 1 in Section 7 for the full statement and complete solution.

2 Complete statistics

Mid priority *Complete statistics*

Recall that a statistic $T(X)$ based on data $X = (X_1, \dots, X_n)$ drawn from $P \in \mathcal{P}$ is sufficient iff it retains relevant information for inferring the unknown distribution P , or in other words the conditional distribution $X|T(X)$ does not depend on the choice of P . However, there may be redundant information in $T(X)$; e.g., $T(X) = X$ is trivially sufficient but not practically useful.

We introduce completeness as a criterion for good summary statistics¹.

Definition 1 (Complete Statistics). $T(X)$ is complete for the family $\mathcal{P}$ if and only if for all Borel functions f with $\mathbb{E}_P |f(T)| < \infty$ for every $P \in \mathcal{P}$, $\mathbb{E}_P[f(T)] = 0$, $\forall P \in \mathcal{P}$ implies that $f(T) = 0$ a.s. P .

¹A related notion is *minimal sufficient*: a sufficient statistic that is, up to model-null sets, a measurable function of every sufficient statistic. Under the usual dominated-model conditions, it is known that a complete and sufficient statistic is minimal sufficient, whereas a minimal sufficient statistic is not guaranteed to be complete and sufficient.

From this definition, if $T(X)$ is complete and ψ, ψ^{-1} are one-one measurable functions, then $\psi(T(X))$ is also complete. It is easy to find complete statistics in an exponential family.

Proposition 1. Let $\mathcal{P} = \{P_\eta : \eta \in \Xi\}$ be an exponential family of full rank (contains an open set) with p.d.f. given by

$$f_\eta(x) = \exp\{\eta^T T(x) - \xi(\eta)\} h(x).$$

Then $T(X)$ is sufficient and complete for $\eta \in \Xi$.

Below are examples taken from [1, Examples 2.15–2.17].

Example 1. Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$. The joint p.d.f. of $X_1, \dots, X_n$ is

$$(2\pi)^{-n/2} \exp\{\eta_1 T_1 + \eta_2 T_2 - n\xi(\eta)\},$$

where $T_1 = \sum_{i=1}^n X_i$, $T_2 = -\sum_{i=1}^n X_i^2$, and $\eta = (\eta_1, \eta_2) = (\frac{\mu}{\sigma^2}, \frac{1}{2\sigma^2})$. Thus, $T = (T_1, T_2)$ is complete and sufficient for η . Also, $(\bar{X}, S^2)$ is complete and sufficient for $\theta = (\mu, \sigma^2)$.

Example 2. Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} P_\theta = U(0, \theta)$, $\theta > 0$. Check that $X_{(n)}$ is complete and sufficient for $\theta > 0$.

Example 3. Let $\mathcal{P}$ be the family of all distributions on $\mathbb{R}$ having a Lebesgue p.d.f., and let the sample be i.i.d. Then $T(X) = (X_{(1)}, \dots, X_{(n)})$ is complete for $\mathcal{P}$.

Proof. Let $\mathcal{P}_0$ be exponential family

$$f(x) = C(\theta_1, \dots, \theta_n) \exp(-x^{2n} + \theta_1 x + \theta_2 x^2 + \dots + \theta_n x^n),$$

where $C(\theta_1, \dots, \theta_n)$ is the normalizing constant. $\mathcal{P}_0 \subset \mathcal{P}$, and $(\sum X_i, \dots, \sum X_i^n)$ is complete for $\mathcal{P}_0$. Since every density in $\mathcal{P}_0$ is strictly positive on $\mathbb{R}$, a null event for its i.i.d. sample law is Lebesgue-null on $\mathbb{R}^n$ and hence null for every i.i.d. sample law with a Lebesgue density, $(\sum X_i, \dots, \sum X_i^n)$ is complete for $\mathcal{P}$. There exists one-one mapping $(\sum X_i, \dots, \sum X_i^n) \leftrightarrow (X_{(1)}, \dots, X_{(n)})$. Therefore, $(X_{(1)}, \dots, X_{(n)})$ is complete for $\mathcal{P}$. $\square$

Instructor's Note No invisible unbiased correction

Completeness says that a function of the summary cannot secretly add a nonzero correction whose average is zero under every possible parameter. Once a complete sufficient summary has produced an unbiased answer, there is no second unbiased answer hidden inside that same summary.

Mid priority Practice: Completeness for one Bernoulli variable

Exercise prompt. Let $X \sim \text{Bernoulli}(p)$, $0 < p < 1$. Prove that X is complete.

See Exercise 2 in Section 7 for the full statement and complete solution.

3 UMVUE

High priority *UMVUE*

Recall that we say $T(X)$ is unbiased for an unknown parameter ν if and only if

$$\mathbb{E}[T(X)] = \nu, \quad \forall P \in \mathcal{P}.$$

Definition 2 (UMVUE). *Suppose that $T(X)$ is unbiased for $\nu(P)$ for every $P \in \mathcal{P}$ and has a finite second moment under every such P . We say $T(X)$ is a uniformly minimum variance unbiased estimator (**UMVUE**) of ν if and only if*

$$\text{Var}(T(X)) \leq \text{Var}(U(X)), \quad \forall P \in \mathcal{P}, \quad \forall \text{ unbiased } U(X).$$

It is clear that UMVUE implies optimality among all unbiased estimators in terms of mean squared error (MSE). For general loss functions instead of the quadratic loss, “minimum variance” is replaced by “minimum risk”.

Remark 1. *UMVUE does not always exist a priori. For fixed $P \in \mathcal{P}$, we can solve*

$$\min_{\text{unbiased } T} \text{Var}(T(X)),$$

but the minimizer is not guaranteed to be uniform for every $P \in \mathcal{P}$.

The following Lehmann–Scheffé theorem [1][Thm. 3.1] states that given a sufficient and complete statistic, often we can find a UMVUE.

Theorem 1 (Lehmann–Scheffé). *Suppose that there exists a sufficient and complete statistic $T(X)$ for $\mathcal{P}$. If ν admits an unbiased estimator with finite second moment under every $P \in \mathcal{P}$, then there exists a unique² UMVUE, which is of the form $h(T)$ where h is a Borel function.*

Proof. This is a consequence of the Rao–Blackwell Theorem and completeness, which we prove below. If $U(X)$ is unbiased for ν , then by (1),

$$\text{Var}(\mathbb{E}[U|T]) \leq \text{Var}(U).$$

Note that $\mathbb{E}[U|T]$ is a function of T , i.e. $\mathbb{E}[U|T] = h(T)$. Sufficiency permits a version of h that does not depend on the unknown parameter. For two unbiased starting estimators, their two conditional functions have difference with mean zero for every P ; completeness makes that difference zero almost surely. Thus $h(T)$ does not depend on the choice of U , up to these null sets, so $h(T)$ has minimum variance among all unbiased estimators U , which implies that $h(T)$ is UMVUE.

If $\tilde{T}$ is another UMVUE, then the variance of $\mathbb{E}[\tilde{T}|T]$ must be no larger than that of $h(T)$, thus equal. Since MSE is based on a strictly convex loss, $\tilde{T}$ must be a function of T (by [1, Theorem 2.5]). Therefore, $\tilde{T} = h(T)$ a.s. $\forall P \in \mathcal{P}$. $\square$

The UMVUE route. Sufficiency makes Rao–Blackwell conditioning independent of the unknown parameter. Completeness then makes unbiased functions of that statistic unique, giving the Lehmann–Scheffé route to a UMVUE.

4 Three ways to find UMVUE

²Unique up to P -null sets for every $P \in \mathcal{P}$.

Mid priority Three ways to find UMVUE

How to study this section. It is good to go through the examples and proofs at least once to understand how the methods work. This section will not be directly tested in exams. The mid-priority blocks guide your study; Method 3 remains low priority.

We will introduce three ways of finding UMVUEs. Examples are taken from [1][Section 3.1].

4.1 Method 1

Mid priority Method 1

Find sufficient and complete statistic $T(X)$, then find $h(T)$, such that

$$\mathbb{E}[h(T)] = \nu, \quad \forall P \in \mathcal{P}.$$

Example 4 (Example 3.1). Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Unif}([0, \theta])$, $\theta > 0$. Let $\nu = g(\theta)$ with $g \in C^1((0, \infty))$. We know that $X_{(n)} = \max_{1 \leq i \leq n} X_i$ is complete, with p.d.f. given by

$$f_\theta(x) = n \frac{1}{\theta^n} x^{n-1} I_{(0, \theta)}(x).$$

Our goal is to find h , such that

$$\theta^n g(\theta) = n \int_0^\theta h(x) x^{n-1} dx.$$

Taking the derivative on both sides, we get

$$n\theta^{n-1}g(\theta) + \theta^n g'(\theta) = nh(\theta)\theta^{n-1}.$$

Rearranging gives $h(x) = g(x) + xg'(x)/n$. For unbiasedness, also require $\lim_{x \downarrow 0} x^n g(x) = 0$ and integrability of $h(x)x^{n-1}$ on $(0, \theta)$; for UMVUE require its square integrability. The boundary term is essential: differentiating the integral equation alone can lose a nonzero integration constant. In particular, if $\nu = \theta$, then

$$h(X_{(n)}) = (1 + \frac{1}{n})X_{(n)}.$$

Example 5 (Example 3.2). Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Poisson}(\theta)$, $\theta > 0$. The total $S = \sum X_i \sim \text{Poisson}(n\theta)$ is complete sufficient because its natural parameter ranges over $\mathbb{R}$. The identities $\mathbb{E}S = n\theta$ and $\mathbb{E}S(S-1) = n^2\theta^2$ show that S/n and $S(S-1)/n^2$ are UMVUEs of θ and θ^2 , respectively.

Example 6 (Example 3.6). Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} P \in \mathcal{P}$, where $\mathcal{P}$ is the family of all distributions with a Lebesgue p.d.f. satisfying the moment conditions specified below, and $n \geq 2$. Let $T = (X_{(1)}, \dots, X_{(n)})$ be a complete and sufficient statistic. Note that $\varphi(X_1, \dots, X_n)$ is a function of T if and only if for any permutation π ,

$$\varphi(X_1, \dots, X_n) = \varphi(X_{\pi(1)}, \dots, X_{\pi(n)}).$$

We use T to construct UMVUEs on the relevant moment classes: finite second moment for the mean and squared mean, and finite fourth moment for the variance rule. These assumptions ensure finite estimator variance. Writing $S^2 = (n-1)^{-1} \sum_i (X_i - \bar{X})^2$, the squared-mean rule has the cancellation

$$\frac{1}{n} \sum_i X_i^2 - S^2 = \frac{2}{n(n-1)} \sum_{i < j} X_i X_j.$$

Each product has finite second moment because independence gives $\mathbb{E}(X_i X_j)^2 = (\mathbb{E}X_1^2)^2 < \infty$. Thus a fourth moment is unnecessary for this rule, although it is sufficient for the variance estimator S^2 .

- (1) $\nu = \mathbb{E}[X_i]$. Then $\bar{X}$.
- (2) $\nu = \text{Var}(X_i)$. Then S^2 .
- (3) $\nu = (\mathbb{E}[X_i])^2$. Then $\frac{1}{n} \sum_{i=1}^n X_i^2 - S^2$. Note that generally $\mathbb{E}[\bar{X}^2] > (\mathbb{E}[\bar{X}])^2 = (\mathbb{E}[X_i])^2$.
- (4) $\nu = \mathbb{P}(X_i \leq t)$. Then $F_n(t) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, t]}(X_i)$.

Mid priority Practice: Estimating a Poisson zero probability

Exercise prompt. For i.i.d. Poisson data, estimate $\nu = e^{-\theta}$ by conditioning $I\{X_1 = 0\}$ on $S = \sum X_i$. See Exercise 3 in Section 7 for the full statement and complete solution.

4.2 Method 2

Mid priority Method 2

Find a sufficient and complete $T(X)$ and an unbiased $U(X)$, then apply conditioning (as in the proof at Sec 3). Note that any unbiased $U(X)$ (as long as its second moment exists) is okay. We have the freedom to choose one so that $\mathbb{E}[U|T]$ is easy to simplify.

Example 7 (Binomial). Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Bin}(k, \theta)$, where k is known, $\theta \in (0, 1)$ is unknown. Need to calculate

$$\mathbb{E}[I(X_i = 1)|T] = \mathbb{P}(X_i = 1|T),$$

where $T = \sum_{i=1}^n X_i$. For $S = T = s$, the nk Bernoulli trials have s successes uniformly placed among their positions. Therefore

$$\mathbb{P}(X_1 = 1 | S = s) = \frac{\binom{k}{1} \binom{(n-1)k}{s-1}}{\binom{nk}{s}}, \quad 0 \leq s \leq nk,$$

with impossible binomial coefficients interpreted as zero. This is the UMVUE of $k\theta(1-\theta)^{k-1}$.

Example 8. (Shao, [1, Example 3.3]) Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{E}(\theta)$ with $n \geq 2$ and take $t \geq 0$. We know that

$$F_\theta(x) = (1 - e^{-x/\theta})I_{(0, \infty)}(x).$$

Let $\nu = 1 - F_\theta(t)$, then

$$T(X) = \mathbb{P}(X_1 > t|\bar{X}),$$

is UMVUE. Note that $\bar{X}$ is complete, and moreover, $\frac{X_1}{\bar{X}}$ is ancillary since $\mathcal{E}(\theta)$ only has a scale parameter. By Basu's Theorem (see Theorem 3), $\bar{X}$ is independent of $\frac{X_1}{\bar{X}}$. So,

$$\mathbb{P}(X_1 > t|\bar{X} = \bar{x}) = \mathbb{P}\left(\frac{X_1}{\bar{X}} > \frac{t}{\bar{x}}|\bar{X} = \bar{x}\right) = \mathbb{P}\left(\frac{X_1}{\bar{X}} > \frac{t}{\bar{x}}\right).$$

It remains to derive the p.d.f. of

$$\frac{X_1}{X_1 + \sum_{i>1} X_i}.$$

Note that X_1 is independent of $\sum_{i>1} X_i$, and $\sum_{i>1} X_i$ has a Gamma distribution. We can use the change-of-variables formula for densities in Lecture 3 to calculate the explicit p.d.f. of this transformation (see Exercise 6). Writing $S = n\bar{X}$, the calculation gives $X_1/S \sim \text{Beta}(1, n-1)$ independent of S , and the UMVUE is $(1 - t/S)_+^{n-1}$. This direct calculation supplies the full route without needing the optional Basu theorem.

Example 9 (Nonexistence of UMVUE). *With one observation $X \sim \text{Bernoulli}(p)$, every estimator has expectation $a(1-p) + bp$, an affine function of p . It cannot equal p^2 for all $p \in (0,1)$. Thus p^2 has no unbiased estimator, and consequently no UMVUE, in this experiment.*

Remark 2. *Let $\mathcal{U}$ be the set of all unbiased estimators,*

$$\text{Var}(T(X)) \leq \text{Var}(U(X)), \quad \forall P \in \mathcal{P}, \quad \forall U \in \mathcal{U}.$$

If $\mathcal{P}$ is small, $\mathcal{U}$ may be large; and conversely, if $\mathcal{P}$ is large, then $\mathcal{U}$ may be small. A larger parameter space does not imply it is harder to find a UMVUE.

Mid priority *Practice: An unbiased rule for Bernoulli variance*

Exercise prompt. Let $S \sim \text{Binomial}(n, p)$ with $n \geq 2$. Find a UMVUE for $p(1-p)$ using S .

See Exercise 4 in Section 7 for the full statement and complete solution.

4.3 Method 3

Low priority *Method 3*

Find UMVUE without knowing the complete statistics.

Theorem 2 (Theorem 3.2 of [1]). *Let*

$$\mathcal{U} = \{U : \mathbb{E}[U(X)] = 0, \text{Var}(U(X)) < \infty, \forall P \in \mathcal{P}\}.$$

T is unbiased for ν with $\mathbb{E}[T(X)^2] < \infty$.

(1) *$T(X)$ is UMVUE if and only if*

$$\mathbb{E}[T(X)U(X)] = 0, \quad \forall U \in \mathcal{U}, \quad \forall P \in \mathcal{P}.$$

(2) *If $\tilde{T}$ is sufficient for $\mathcal{P}$, let*

$$\tilde{\mathcal{U}} = \mathcal{U} \cap \{g(\tilde{T}) : g \text{ Borel}\}.$$

Then, $T = h(\tilde{T})$ is UMVUE if and only if

$$\mathbb{E}[T(X)U(X)] = 0, \quad \forall U \in \tilde{\mathcal{U}}, \quad \forall P \in \mathcal{P}.$$

Proof. For any zero-mean U in the stated class, $T + aU$ is unbiased for every real a . At fixed P ,

$$\text{Var}_P(T + aU) = \text{Var}_P(T) + 2a \mathbb{E}_P(TU) + a^2 \text{Var}_P(U).$$

If T is UMVUE, the derivative at $a = 0$ must vanish, giving $\mathbb{E}_P(TU) = 0$. Conversely, any unbiased competitor V can be written $T + U$ with $\mathbb{E}_P U = 0$ for all P , so orthogonality gives $\text{Var}_P(V) = \text{Var}_P(T) + \text{Var}_P(U) \geq \text{Var}_P(T)$. For (2), first Rao–Blackwellize any competitor on the sufficient statistic and apply the same argument within its functions. $\square$

Instructor's Note *Variation along an unbiased direction*

The original calculus-of-variations picture is useful: move a little in any direction that keeps the estimator unbiased. At a minimum, neither sign of that movement can improve the variance, so the linear term must vanish. Here the feasible class is affine; an arbitrary constrained optimization problem need not have zero directional derivatives in every direction.

As a consequence, UMVUE satisfies linearity; see [1][Corollary 3.1]. We can also use this theorem to show UMVUE does not exist in some cases.

Example 10. Let $X \sim \text{Unif}(\theta - \frac{1}{2}, \theta + \frac{1}{2})$, $\theta \in \mathbb{R}$. If $\nu = g(\theta)$, where g is not a constant, no UMVUE exists. See proofs in [1][Example 3.8].

5 Further reading

Ancillary statistics

Further reading *Ancillary statistics*

Definition 3. $V(X)$ is called ancillary if its distribution does not depend on the member $P \in \mathcal{P}$, i.e. the distribution remains the same for every $P \in \mathcal{P}$.

Given a family $\mathcal{P}$, an ancillary statistic has no parameter-dependent marginal distribution; it can still be useful jointly with other statistics.

Example 11. If $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$, then $X_i - X_j$ is ancillary.

Example 12. If $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$, then $\frac{X_i}{X_j}$ and $\frac{X_{(i)}}{X_{(j)}}$ are both ancillary.

Theorem 3 (Basu's Theorem). If V is ancillary and T is boundedly complete and sufficient for P . Then $V \perp T$ w.r.t. $\forall P \in \mathcal{P}$.

Example 13. Let $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$. $\bar{X}$ is complete and sufficient, and $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ is ancillary. This implies $\bar{X} \perp S^2$.

Key ideas

- Conditional averaging can reduce squared-error risk without an unbiasedness assumption.
- Sufficiency makes that averaging rule independent of the unknown parameter.
- Completeness turns unbiasedness into uniqueness among functions of a sufficient statistic.
- Section 4 examples and proofs are for understanding and are not directly tested in exams.

6 Detailed reference

High priority *Conditioning and variance*

For square-integrable Y , $\text{Var}(Y) = \text{Var}(\mathbb{E}[Y | T]) + \mathbb{E} \text{Var}(Y | T)$. Conditioning removes the second nonnegative term while preserving the mean. Return to the earlier explanation.

High priority *The UMVUE route*

Sufficiency makes Rao–Blackwell conditioning independent of the unknown parameter. Completeness then makes unbiased functions of that statistic unique, giving the Lehmann–Scheffe route to a UMVUE. Return to the earlier explanation.

Looking ahead

Lecture 13 complements the UMVUE construction with a lower bound on the variance of unbiased estimators. Likelihood sensitivity, measured by Fisher information, determines that bound under regularity. Continue to Lecture 13.

7 Exercises and complete solutions

Mid priority *Weighting two independent server means*

Exercise 1 (Weighting two independent server means).

For the two-server exponential example, minimize the variance of $a\hat{\mu}_1 + (1-a)\hat{\mu}_2$ over real a .

Solution.

Independence gives variance $\mu^2[a^2/2 + (1-a)^2/(n-2)]$. Its derivative vanishes at $a/2 = (1-a)/(n-2)$, hence $a = 2/n$. The second derivative is positive. The resulting rule is $\bar{X}$, with variance μ^2/n . Equal server weights are optimal only at $n = 4$. □

Mid priority *Completeness for one Bernoulli variable*

Exercise 2 (Completeness for one Bernoulli variable).

Let $X \sim \text{Bernoulli}(p)$, $0 < p < 1$. Prove that X is complete.

Solution.

If $\mathbb{E}_p f(X) = (1-p)f(0) + pf(1) = 0$ for every $p \in (0, 1)$, this affine polynomial is identically zero. Its constant coefficient gives $f(0) = 0$ and its slope then gives $f(1) = 0$. Thus $f(X) = 0$ almost surely for every p . □

Mid priority *Estimating a Poisson zero probability*

Exercise 3 (Estimating a Poisson zero probability).

Study practice for Section 4; not directly tested in exams.

For i.i.d. Poisson data, estimate $\nu = e^{-\theta}$ by conditioning $I\{X_1 = 0\}$ on $S = \sum X_i$.

Solution.

For $n \geq 2$, $X_1 | S = s \sim \text{Binomial}(s, 1/n)$, so the conditional expectation is $(1 - 1/n)^S$. Its expectation is $\exp\{n\theta[(1 - 1/n) - 1]\} = e^{-\theta}$. It is bounded and a function of complete sufficient S , hence UMVUE. For $n = 1$ the rule is simply $I\{S = 0\}$. □

Mid priority *An unbiased rule for Bernoulli variance*

Exercise 4 (An unbiased rule for Bernoulli variance).

Study practice for Section 4; not directly tested in exams.

Let $S \sim \text{Binomial}(n, p)$ with $n \geq 2$. Find a UMVUE for $p(1-p)$ using S .

Solution.

Since $\mathbb{E}S = np$ and $\mathbb{E}S(S-1) = n(n-1)p^2$, $S/n - S(S-1)/[n(n-1)] = S(n-S)/[n(n-1)]$ has expectation $p - p^2$. The binomial family is full-rank exponential with complete sufficient S , so this bounded unbiased rule is UMVUE. □

Mid priority *Completeness of the Uniform maximum*

Exercise 5 (Completeness of the Uniform maximum).

Prove the sufficiency and completeness of $M = X_{(n)}$ for i.i.d. Uniform(0, θ) data.

Solution.

The joint density is $\theta^{-n} I\{0 < \min X_i, M < \theta\}$, so factorization gives sufficiency. If $\mathbb{E}_\theta f(M) = n\theta^{-n} \int_0^\theta f(x)x^{n-1}dx = 0$ for every $\theta > 0$, the indefinite integral vanishes. Absolute continuity on every compact subinterval implies $f(x)x^{n-1} = 0$ Lebesgue-almost everywhere, hence $f(M) = 0$ almost surely for every θ . □

Mid priority *The exponential share and survival probability*

Exercise 6 (The exponential share and survival probability).

Study practice for Section 4; not directly tested in exams.

Derive the density of $(U, S) = (X_1/(X_1 + \dots + X_n), X_1 + \dots + X_n)$ for exponential data with $n \geq 2$, and the conditional survival formula.

Solution.

Let $V = \sum_{i=2}^n X_i$. Its density is $v^{n-2}e^{-v/\theta}/[\Gamma(n-1)\theta^{n-1}]$, independently of X_1 . Substitute $x_1 = us$, $v = (1-u)s$ with Jacobian s . The joint density factors as

$$(n-1)(1-u)^{n-2}I_{(0,1)}(u) \frac{s^{n-1}e^{-s/\theta}}{\Gamma(n)\theta^n} I_{(0,\infty)}(s).$$

Thus U and S are independent and $U \sim \text{Beta}(1, n-1)$. For $t \geq 0$, integrate its density from t/s to 1 when $s > t$; otherwise the probability is zero. The result is $(1 - t/s)_+^{n-1}$. □

High priority *The exact gain from conditioning*

Exercise 7 (The exact gain from conditioning).

For square-integrable Y and a statistic T , prove $\mathbb{E}(Y - a)^2 = \mathbb{E}(\mathbb{E}[Y | T] - a)^2 + \mathbb{E} \text{Var}(Y | T)$.

Solution.

Let $m(T) = \mathbb{E}[Y | T]$. Square integrability implies that $m(T)$ and $Y - m(T)$ are square-integrable, so their product is integrable. Write

$$Y - a = (Y - m(T)) + (m(T) - a).$$

Since $m(T) - a$ is determined by T and $\mathbb{E}[Y - m(T) | T] = 0$, the cross term vanishes:

$$\mathbb{E}[(Y - m(T))(m(T) - a)] = \mathbb{E}[(m(T) - a) \mathbb{E}[Y - m(T) | T]] = 0.$$

Taking expectations after squaring gives

$$\mathbb{E}(Y - a)^2 = \mathbb{E}(m(T) - a)^2 + \mathbb{E}(Y - m(T))^2 = \mathbb{E}(m(T) - a)^2 + \mathbb{E} \text{Var}(Y | T).$$

The last term is nonnegative. It vanishes if and only if $Y = m(T)$ almost surely. Also, $\mathbb{E} m(T) = \mathbb{E} Y$, so the bias relative to a is unchanged. Neither $\mathbb{E} Y = a$ nor sufficiency of T was used. In an estimation model, apply this identity separately under each P_μ with $Y = T_2(X)$, $T = T_1(X)$ and $a = \mu$; sufficiency is what guarantees a common parameter-independent version of m . □

Low priority *A short proof of Basu independence*

Exercise 8 (A short proof of Basu independence).

If T is boundedly complete and sufficient and V is ancillary, prove their independence, using a parameter-independent conditional distribution given T .

Solution.

For any measurable event B in the range of V , let $h_B(T) = \mathbb{P}(V \in B \mid T)$. Sufficiency gives a common version and ancillarity gives the same number $c_B = \mathbb{P}_P(V \in B)$ for every P . The bounded function $h_B - c_B$ has expectation zero under every P , so bounded completeness gives $h_B(T) = c_B$ almost surely under each P . Multiplying by $I\{T \in A\}$ and taking expectations yields independence for all measurable A, B . □

References

[1] Jun Shao. *Mathematical Statistics*. Springer, 2nd edition, 2003.

What changed

October 19, 2026: Corrected notation, wording and cross-references after the course-wide errata review; refreshed the linked HTML and PDF editions.

October 14, 2026. Added issue-reporting links to the webpage and PDF. The mathematics and priority assignments are unchanged in this update.