Keener, Theoretical Statistics
Chapter 4: Unbiased Estimation

Junhyeok Kil

August 20, 2025

Unbiased Estimators and U-Estimability

Definition: Unbiased Estimator

An estimator $\delta(X)$ is called **unbiased** for a function of the parameter, $g(\theta)$, if its expected value equals the function itself.

$$E_\theta[\delta(X)] = g(\theta), \quad \forall \theta \in \Omega$$

Unbiased Estimators and U-Estimability

Definition: Unbiased Estimator

An estimator $\delta(X)$ is called **unbiased** for a function of the parameter, $g(\theta)$, if its expected value equals the function itself.

$$E_\theta[\delta(X)] = g(\theta), \quad \forall \theta \in \Omega$$

Definition: U-Estimable

If an unbiased estimator exists for $g(\theta)$, then $g(\theta)$ is said to be **U-estimable**.

Example 4.1: Uniform Distribution

Let X have a uniform distribution on the interval $(0, \theta)$, denoted $X \sim U(0, \theta)$.

An estimator $\delta(X)$ is unbiased for $g(\theta)$ if:

$$E_\theta[\delta(X)] = \int_0^\theta \delta(x) \frac{1}{\theta} \, dx = g(\theta)$$

Example 4.1: Uniform Distribution

Let X have a uniform distribution on the interval $(0, \theta)$, denoted $X \sim U(0, \theta)$.

An estimator $\delta(X)$ is unbiased for $g(\theta)$ if:

$$E_\theta[\delta(X)] = \int_0^\theta \delta(x) \frac{1}{\theta} \, dx = g(\theta)$$

Rearranging the equation, we get:

$$\int_0^\theta \delta(x) \, dx = \theta g(\theta), \quad \forall \theta > 0$$

Example 4.1: Uniform Distribution

Let X have a uniform distribution on the interval $(0, \theta)$, denoted $X \sim U(0, \theta)$.

An estimator $\delta(X)$ is unbiased for $g(\theta)$ if:

$$E_\theta[\delta(X)] = \int_0^\theta \delta(x) \frac{1}{\theta}\, dx = g(\theta)$$

Rearranging the equation, we get:

$$\int_0^\theta \delta(x)\, dx = \theta g(\theta), \quad \forall \theta > 0$$

If we assume $g'(\theta)$ exists, we can differentiate both sides with respect to θ using the Fundamental Theorem of Calculus:

$$\delta(\theta) = \frac{d}{d\theta}[\theta g(\theta)] = g(\theta) + \theta g'(\theta)$$

Replacing θ with the random variable X, we find the estimator:

$$\delta(X) = g(X) + X g'(X)$$

Example 4.1: A Specific Case

Continuing with $X \sim U(0, \theta)$, let's find the unbiased estimator for $g(\theta) = \theta$.

- We have $g(\theta) = \theta$.
- The derivative is $g'(\theta) = 1$.

Example 4.1: A Specific Case

Continuing with $X \sim U(0, \theta)$, let's find the unbiased estimator for $g(\theta) = \theta$.

- We have $g(\theta) = \theta$.
- The derivative is $g'(\theta) = 1$.

Using the formula derived on the previous slide:

$$\delta(X) = g(X) + Xg'(X)$$

Example 4.1: A Specific Case

Continuing with $X \sim U(0, \theta)$, let's find the unbiased estimator for $g(\theta) = \theta$.

- We have $g(\theta) = \theta$.
- The derivative is $g'(\theta) = 1$.

Using the formula derived on the previous slide:

$$\delta(X) = g(X) + Xg'(X)$$

Substituting our specific function and its derivative:

$$\delta(X) = X + X(1) = 2X$$

Thus, for a $U(0, \theta)$ distribution, $\delta(X) = 2X$ is an unbiased estimator for θ.

Example 4.2: Binomial Distribution

Let $X \sim \text{Binomial}(n, \theta)$, and consider estimating $g(\theta) = \sin(\theta)$.

For an estimator $\delta(X)$ to be unbiased, the following must hold for all $\theta \in (0, 1)$:

$$E_\theta[\delta(X)] = \sum_{k=0}^{n} \delta(k) \binom{n}{k} \theta^k (1-\theta)^{n-k} = \sin(\theta)$$

Example 4.2: Binomial Distribution

Let $X \sim \text{Binomial}(n, \theta)$, and consider estimating $g(\theta) = \sin(\theta)$.

For an estimator $\delta(X)$ to be unbiased, the following must hold for all $\theta \in (0, 1)$:

$$E_\theta[\delta(X)] = \sum_{k=0}^{n} \delta(k) \binom{n}{k} \theta^k (1-\theta)^{n-k} = \sin(\theta)$$

The Problem

The left-hand side of the equation is a **polynomial** in θ of degree at most n.

However, the right-hand side, $\sin(\theta)$, is a transcendental function and **cannot** be expressed as a finite-degree polynomial.

Example 4.2: Binomial Distribution

Let $X \sim \text{Binomial}(n, \theta)$, and consider estimating $g(\theta) = \sin(\theta)$.

For an estimator $\delta(X)$ to be unbiased, the following must hold for all $\theta \in (0, 1)$:

$$E_\theta[\delta(X)] = \sum_{k=0}^{n} \delta(k) \binom{n}{k} \theta^k (1-\theta)^{n-k} = \sin(\theta)$$

The Problem

The left-hand side of the equation is a **polynomial** in θ of degree at most n.

However, the right-hand side, $\sin(\theta)$, is a transcendental function and **cannot** be expressed as a finite-degree polynomial.
Conclusion: No such estimator $\delta(X)$ can exist. Therefore, $g(\theta) = \sin(\theta)$ is **not U-estimable** for the binomial distribution.

Goal: Minimizing Variance

When evaluating estimators, we often use a **loss function**. A common choice is the squared error loss:

$$L(\theta, d) = (d - g(\theta))^2$$

Goal: Minimizing Variance

When evaluating estimators, we often use a **loss function**. A common choice is the squared error loss:

$$L(\theta, d) = (d - g(\theta))^2$$

The **risk** of an estimator is its expected loss. For an unbiased estimator $\delta(X)$, the risk simplifies to be the variance:

$$\begin{aligned} R(\theta, \delta) &= E_\theta\left[(\delta(X) - g(\theta))^2\right] \\ &= E_\theta\left[(\delta(X) - E_\theta[\delta(X)])^2\right] \\ &= \mathrm{Var}_\theta(\delta(X)) \end{aligned}$$

Goal: Minimizing Variance

When evaluating estimators, we often use a **loss function**. A common choice is the squared error loss:

$$L(\theta, d) = (d - g(\theta))^2$$

The **risk** of an estimator is its expected loss. For an unbiased estimator $\delta(X)$, the risk simplifies to be the variance:

$$\begin{aligned} R(\theta, \delta) &= E_\theta\left[(\delta(X) - g(\theta))^2\right] \\ &= E_\theta\left[(\delta(X) - E_\theta[\delta(X)])^2\right] \\ &= \mathrm{Var}_\theta(\delta(X)) \end{aligned}$$

The Objective

For the class of unbiased estimators, the best estimator under squared error loss is the one with the smallest variance.

The UMVU Estimator

Definition: Uniformly Minimum Variance Unbiased (UMVU)

An unbiased estimator $\delta(X)$ is called **UMVU** if its variance is less than or equal to the variance of any other unbiased estimator $\delta^*(X)$ for all possible values of the parameter θ.

$$\mathrm{Var}_\theta(\delta) \leq \mathrm{Var}_\theta(\delta^*), \quad \forall \theta \in \Omega$$

The UMVU Estimator

Definition: Uniformly Minimum Variance Unbiased (UMVU)

An unbiased estimator $\delta(X)$ is called **UMVU** if its variance is less than or equal to the variance of any other unbiased estimator $\delta^*(X)$ for all possible values of the parameter θ.

$$\mathrm{Var}_\theta(\delta) \leq \mathrm{Var}_\theta(\delta^*), \quad \forall\theta \in \Omega$$

A Key Question

In a general setting, there is no guarantee that a UMVU estimator will exist. When can we find one?

The Lehmann-Scheffé Theorem

Theorem 4.4

Suppose $g(\theta)$ is U-estimable (i.e., at least one unbiased estimator exists) and T is a **complete sufficient statistic** for the family of distributions.

Then there exists an essentially unique unbiased estimator based on T that is the UMVU estimator for $g(\theta)$.

Proof Outline (Part 1: Construction)

The proof relies on the Rao-Blackwell Theorem.

1. **Start with any unbiased estimator.**
   Let $\delta(X)$ be any unbiased estimator for $g(\theta)$, so $E_\theta[\delta(X)] = g(\theta)$.

Proof Outline (Part 1: Construction)

The proof relies on the Rao-Blackwell Theorem.

1. **Start with any unbiased estimator.**
   Let $\delta(X)$ be any unbiased estimator for $g(\theta)$, so $E_\theta[\delta(X)] = g(\theta)$.
2. **Condition on the sufficient statistic T.**
   Define a new estimator, $\eta(T)$, by "smoothing" $\delta(X)$:

$$\eta(T) = E[\delta(X)|T]$$

Proof Outline (Part 1: Construction)

The proof relies on the Rao-Blackwell Theorem.

1. **Start with any unbiased estimator.**
   Let $\delta(X)$ be any unbiased estimator for $g(\theta)$, so $E_\theta[\delta(X)] = g(\theta)$.
2. **Condition on the sufficient statistic T.**
   Define a new estimator, $\eta(T)$, by "smoothing" $\delta(X)$:
   $$\eta(T) = E[\delta(X)|T]$$
3. **Show the new estimator is unbiased.**
   Using the law of total expectation:
   $$\begin{aligned} E_\theta[\eta(T)] &= E_\theta[E[\delta(X)|T]] \\ &= E_\theta[\delta(X)] = g(\theta) \end{aligned}$$
   So, $\eta(T)$ is also an unbiased estimator for $g(\theta)$.

Proof Outline (Part 2: Uniqueness)

Here, the **completeness** of T is crucial.

- Suppose $\eta^*(T)$ is another unbiased estimator based on T.
- Then we have $E_\theta[\eta(T)] = g(\theta)$ and $E_\theta[\eta^*(T)] = g(\theta)$.
- This implies:

$$E_\theta[\eta(T) - \eta^*(T)] = 0, \quad \forall \theta \in \Omega$$

Proof Outline (Part 2: Uniqueness)

Here, the **completeness** of T is crucial.

- Suppose $\eta^*(T)$ is another unbiased estimator based on T.
- Then we have $E_\theta[\eta(T)] = g(\theta)$ and $E_\theta[\eta^*(T)] = g(\theta)$.
- This implies:

$$E_\theta[\eta(T) - \eta^*(T)] = 0, \quad \forall \theta \in \Omega$$

- By the definition of a complete statistic, if the expectation of a function of T is zero for all θ, the function itself must be zero (almost everywhere).

$$\eta(T) - \eta^*(T) = 0 \quad \text{(a.e.)}$$

Proof Outline (Part 2: Uniqueness)

Here, the **completeness** of T is crucial.

- Suppose $\eta^*(T)$ is another unbiased estimator based on T.
- Then we have $E_\theta[\eta(T)] = g(\theta)$ and $E_\theta[\eta^*(T)] = g(\theta)$.
- This implies:

$$E_\theta[\eta(T) - \eta^*(T)] = 0, \quad \forall\theta \in \Omega$$

- By the definition of a complete statistic, if the expectation of a function of T is zero for all θ, the function itself must be zero (almost everywhere).

$$\eta(T) - \eta^*(T) = 0 \quad \text{(a.e.)}$$

- **Conclusion:** The estimator $\eta(T)$ is **essentially unique**. Any unbiased estimator that is a function of T must be equal to $\eta(T)$.

Proof Outline (Part 3: Minimum Variance)

The minimum variance property comes directly from the Rao-Blackwell Theorem.

- Let δ^* be *any* competing unbiased estimator.
- Let $\eta^*(T) = E[\delta^*|T]$. We already know $\eta^*(T)$ is unbiased.
- From the uniqueness step, we must have $\eta^*(T) = \eta(T)$.

Proof Outline (Part 3: Minimum Variance)

The minimum variance property comes directly from the Rao-Blackwell Theorem.

- Let δ^* be *any* competing unbiased estimator.
- Let $\eta^*(T) = E[\delta^*|T]$. We already know $\eta^*(T)$ is unbiased.
- From the uniqueness step, we must have $\eta^*(T) = \eta(T)$.
- The Rao-Blackwell Theorem states that for squared error loss, conditioning reduces risk (variance):

$$\mathrm{Var}_\theta(\delta^*) \geq \mathrm{Var}_\theta(E[\delta^*|T])$$

Proof Outline (Part 3: Minimum Variance)

The minimum variance property comes directly from the Rao-Blackwell Theorem.

- Let δ^* be *any* competing unbiased estimator.
- Let $\eta^*(T) = E[\delta^*|T]$. We already know $\eta^*(T)$ is unbiased.
- From the uniqueness step, we must have $\eta^*(T) = \eta(T)$.
- The Rao-Blackwell Theorem states that for squared error loss, conditioning reduces risk (variance):

$$\mathrm{Var}_\theta(\delta^*) \geq \mathrm{Var}_\theta(E[\delta^*|T])$$

- Substituting our terms:

$$\mathrm{Var}_\theta(\delta^*) \geq \mathrm{Var}_\theta(\eta^*(T)) = \mathrm{Var}_\theta(\eta(T))$$

Proof Outline (Part 3: Minimum Variance)

The minimum variance property comes directly from the Rao-Blackwell Theorem.

- Let δ^* be *any* competing unbiased estimator.
- Let $\eta^*(T) = E[\delta^* | T]$. We already know $\eta^*(T)$ is unbiased.
- From the uniqueness step, we must have $\eta^*(T) = \eta(T)$.
- The Rao-Blackwell Theorem states that for squared error loss, conditioning reduces risk (variance):

$$\mathrm{Var}_\theta(\delta^*) \geq \mathrm{Var}_\theta(E[\delta^* | T])$$

- Substituting our terms:

$$\mathrm{Var}_\theta(\delta^*) \geq \mathrm{Var}_\theta(\eta^*(T)) = \mathrm{Var}_\theta(\eta(T))$$

- **Conclusion:** The variance of $\eta(T)$ is less than or equal to the variance of any other unbiased estimator δ^*.

Therefore, $\eta(T)$ is the **UMVU estimator**.

A Practical Strategy

From the Lehmann-Scheffé Theorem, we get a powerful strategy:

1. Find a **complete sufficient statistic**, T.

A Practical Strategy

From the Lehmann-Scheffé Theorem, we get a powerful strategy:

1. Find a **complete sufficient statistic**, T.
2. Find a function of T, let's call it $\eta(T)$, that is an **unbiased estimator** for $g(\theta)$.

$$E_\theta[\eta(T)] = g(\theta)$$

A Practical Strategy

From the Lehmann-Scheffé Theorem, we get a powerful strategy:

1. Find a **complete sufficient statistic**, T.
2. Find a function of T, let's call it $\eta(T)$, that is an **unbiased estimator** for $g(\theta)$.

$$E_\theta[\eta(T)] = g(\theta)$$

3. The uniqueness part of the theorem guarantees that this estimator $\eta(T)$ **must be the UMVU estimator**.

Instead of searching through all possible estimators, we only need to solve the unbiasedness equation for a function of T.

Example: Uniform Distribution

The Setup

Let $X_1, \ldots, X_n$ be i.i.d. random variables from a Uniform$(0, \theta)$ distribution.

Example: Uniform Distribution

The Setup

Let $X_1, \ldots, X_n$ be i.i.d. random variables from a Uniform$(0, \theta)$ distribution.

- From previous examples, we know that $T = \max\{X_1, \ldots, X_n\}$ is a **complete and sufficient** statistic for θ.

Example: Uniform Distribution

The Setup

Let $X_1, \ldots, X_n$ be i.i.d. random variables from a Uniform$(0, \theta)$ distribution.

- From previous examples, we know that $T = \max\{X_1, \ldots, X_n\}$ is a **complete and sufficient** statistic for θ.
- The probability density function (PDF) of T is:

$$f_T(t) = \frac{nt^{n-1}}{\theta^n}, \quad \text{for } 0 < t < \theta$$

Example: Uniform Distribution

The Setup

Let $X_1, \ldots, X_n$ be i.i.d. random variables from a Uniform$(0, \theta)$ distribution.

- From previous examples, we know that $T = \max\{X_1, \ldots, X_n\}$ is a **complete and sufficient** statistic for θ.
- The probability density function (PDF) of T is:

$$f_T(t) = \frac{nt^{n-1}}{\theta^n}, \quad \text{for } 0 < t < \theta$$

Our goal is to find an estimator $\eta(T)$ such that $E_\theta[\eta(T)] = g(\theta)$.

Example: Derivation of the Estimator

We set up the expectation equation:

$$E_\theta[\eta(T)] = \int_0^\theta \eta(t) f_T(t)\, dt = g(\theta)$$

Example: Derivation of the Estimator

We set up the expectation equation:

$$E_\theta[\eta(T)] = \int_0^\theta \eta(t) f_T(t)\, dt = g(\theta)$$

Substituting the PDF of T:

$$\int_0^\theta \eta(t) \frac{nt^{n-1}}{\theta^n}\, dt = g(\theta)$$

Example: Derivation of the Estimator

We set up the expectation equation:

$$E_\theta[\eta(T)] = \int_0^\theta \eta(t) f_T(t)\, dt = g(\theta)$$

Substituting the PDF of T:

$$\int_0^\theta \eta(t) \frac{nt^{n-1}}{\theta^n}\, dt = g(\theta)$$

Rearranging the terms gives:

$$\int_0^\theta t^{n-1} \eta(t)\, dt = \frac{\theta^n g(\theta)}{n}$$

Example: Derivation of the Estimator

We set up the expectation equation:

$$E_\theta[\eta(T)] = \int_0^\theta \eta(t) f_T(t)\, dt = g(\theta)$$

Substituting the PDF of T:

$$\int_0^\theta \eta(t) \frac{nt^{n-1}}{\theta^n}\, dt = g(\theta)$$

Rearranging the terms gives:

$$\int_0^\theta t^{n-1}\eta(t)\, dt = \frac{\theta^n g(\theta)}{n}$$

Assuming differentiability, we differentiate both sides with respect to θ using the Fundamental Theorem of Calculus:

$$\theta^{n-1}\eta(\theta) = \frac{1}{n}\frac{d}{d\theta}\left[\theta^n g(\theta)\right]$$

Example: The UMVU Estimator

Solving for $\eta(\theta)$ from the previous step:

$$\eta(\theta) = \frac{1}{n\theta^{n-1}} \frac{d}{d\theta} \left[\theta^n g(\theta)\right]$$

Example: The UMVU Estimator

Solving for $\eta(\theta)$ from the previous step:

$$\eta(\theta) = \frac{1}{n\theta^{n-1}} \frac{d}{d\theta} \left[\theta^n g(\theta)\right]$$

Using the product rule for differentiation:

$$\begin{aligned}
\eta(\theta) &= \frac{1}{n\theta^{n-1}} \left[n\theta^{n-1} g(\theta) + \theta^n g'(\theta)\right] \\
&= g(\theta) + \frac{\theta}{n} g'(\theta)
\end{aligned}$$

Example: The UMVU Estimator

Solving for $\eta(\theta)$ from the previous step:

$$\eta(\theta) = \frac{1}{n\theta^{n-1}} \frac{d}{d\theta} \left[\theta^n g(\theta)\right]$$

Using the product rule for differentiation:

$$\begin{aligned} \eta(\theta) &= \frac{1}{n\theta^{n-1}} \left[n\theta^{n-1} g(\theta) + \theta^n g'(\theta)\right] \\ &= g(\theta) + \frac{\theta}{n} g'(\theta) \end{aligned}$$

The UMVU estimator for $g(\theta)$ is $\eta(T) = g(T) + \frac{T}{n} g'(T)$.

Example: The UMVU Estimator

Solving for $\eta(\theta)$ from the previous step:

$$\eta(\theta) = \frac{1}{n\theta^{n-1}} \frac{d}{d\theta} \left[\theta^n g(\theta)\right]$$

Using the product rule for differentiation:

$$\begin{aligned} \eta(\theta) &= \frac{1}{n\theta^{n-1}} \left[n\theta^{n-1} g(\theta) + \theta^n g'(\theta)\right] \\ &= g(\theta) + \frac{\theta}{n} g'(\theta) \end{aligned}$$

The UMVU estimator for $g(\theta)$ is $\eta(T) = g(T) + \frac{T}{n} g'(T)$.

Note on Completeness

If we let $g(\theta) = c$ (a constant), the formula gives $\eta(T) = c + \frac{T}{n}(0) = c$. Since $E[c] = c$, this shows that the only unbiased estimator of a constant is the constant itself, which is a property of complete statistics.

Case Study: Estimating θ

Let's find the UMVU estimator for $g(\theta) = \theta$ for i.i.d. $X_i \sim U(0, \theta)$.

From the previous derivation, the general form of the estimator is:

$$\eta(t) = g(t) + \frac{t}{n} g'(t)$$

Case Study: Estimating θ

Let's find the UMVU estimator for $g(\theta) = \theta$ for i.i.d. $X_i \sim U(0, \theta)$.

From the previous derivation, the general form of the estimator is:

$$\eta(t) = g(t) + \frac{t}{n} g'(t)$$

For our case, $g(t) = t$ and $g'(t) = 1$. Substituting these in:

$$\eta(t) = t + \frac{t}{n}(1) = t\left(1 + \frac{1}{n}\right) = \frac{n+1}{n} t$$

Case Study: Estimating θ

Let's find the UMVU estimator for $g(\theta) = \theta$ for i.i.d. $X_i \sim U(0, \theta)$.

From the previous derivation, the general form of the estimator is:

$$\eta(t) = g(t) + \frac{t}{n} g'(t)$$

For our case, $g(t) = t$ and $g'(t) = 1$. Substituting these in:

$$\eta(t) = t + \frac{t}{n}(1) = t\left(1 + \frac{1}{n}\right) = \frac{n+1}{n} t$$

The UMVU estimator for θ is $\eta(T) = \frac{n+1}{n} T$, where $T = \max\{X_i\}$.

Comparing Two Unbiased Estimators

We have two unbiased estimators for θ:

1. The UMVU estimator: $\eta(T) = \frac{n+1}{n} T$
2. Another simple estimator: $\delta = 2\bar{X}$

Comparing Two Unbiased Estimators

We have two unbiased estimators for θ:

1. The UMVU estimator: $\eta(T) = \frac{n+1}{n} T$
2. Another simple estimator: $\delta = 2\bar{X}$

Theoretical Guarantee

The Lehmann-Scheffé theorem guarantees that the UMVU estimator has the smallest variance.

$$\mathrm{Var}_\theta(\eta(T)) \leq \mathrm{Var}_\theta(\delta)$$

Comparing Two Unbiased Estimators

We have two unbiased estimators for θ:

1. The UMVU estimator: $\eta(T) = \frac{n+1}{n} T$
2. Another simple estimator: $\delta = 2\bar{X}$

Theoretical Guarantee

The Lehmann-Scheffé theorem guarantees that the UMVU estimator has the smallest variance.

$$\mathrm{Var}_\theta(\eta(T)) \leq \mathrm{Var}_\theta(\delta)$$

Let's verify this by calculating the variances explicitly.

Variance of the UMVU Estimator $\eta(T)$

We use $\text{Var}(\eta(T)) = E[\eta(T)^2] - (E[\eta(T)])^2$. Since $E[\eta(T)] = \theta$, we need $E[\eta(T)^2]$.

First, we find $E[T^2]$:

$$E_\theta[T^2] = \int_0^\theta t^2 \left(\frac{nt^{n-1}}{\theta^n} \right) dt = \frac{n}{\theta^n} \int_0^\theta t^{n+1} dt = \frac{n}{n+2} \theta^2$$

Variance of the UMVU Estimator $\eta(T)$

We use $\text{Var}(\eta(T)) = E[\eta(T)^2] - (E[\eta(T)])^2$. Since $E[\eta(T)] = \theta$, we need $E[\eta(T)^2]$.

First, we find $E[T^2]$:

$$E_\theta[T^2] = \int_0^\theta t^2 \left(\frac{nt^{n-1}}{\theta^n} \right) dt = \frac{n}{\theta^n} \int_0^\theta t^{n+1} dt = \frac{n}{n+2}\theta^2$$

Now we can find $E[\eta(T)^2]$:

$$E_\theta[\eta(T)^2] = E_\theta\left[\left(\frac{n+1}{n} T \right)^2 \right] = \left(\frac{n+1}{n} \right)^2 E_\theta[T^2] = \frac{(n+1)^2}{n(n+2)}\theta^2$$

Variance of the UMVU Estimator $\eta(T)$

We use $\mathrm{Var}(\eta(T)) = E[\eta(T)^2] - (E[\eta(T)])^2$. Since $E[\eta(T)] = \theta$, we need $E[\eta(T)^2]$.

First, we find $E[T^2]$:

$$E_\theta[T^2] = \int_0^\theta t^2 \left(\frac{nt^{n-1}}{\theta^n}\right) dt = \frac{n}{\theta^n}\int_0^\theta t^{n+1} dt = \frac{n}{n+2}\theta^2$$

Now we can find $E[\eta(T)^2]$:

$$E_\theta[\eta(T)^2] = E_\theta\left[\left(\frac{n+1}{n}T\right)^2\right] = \left(\frac{n+1}{n}\right)^2 E_\theta[T^2] = \frac{(n+1)^2}{n(n+2)}\theta^2$$

Finally, the variance is:

$$\begin{aligned}
\mathrm{Var}_\theta(\eta(T)) &= \frac{(n+1)^2}{n(n+2)}\theta^2 - \theta^2 \\
&= \left(\frac{n^2+2n+1-(n^2+2n)}{n(n+2)}\right)\theta^2 = \frac{\theta^2}{n(n+2)}
\end{aligned}$$

Variance of the Competing Estimator δ

Now let's find the variance of $\delta = 2\bar{X}$.

Recall that for a single $X_i \sim U(0, \theta)$:

$$E[X_i] = \frac{\theta}{2}, \quad E[X_i^2] = \frac{\theta^2}{3},$$

$$\text{Var}(X_i) = E[X_i^2] - (E[X_i])^2 = \frac{\theta^2}{3} - \left(\frac{\theta}{2}\right)^2 = \frac{\theta^2}{12}.$$

Variance of the Competing Estimator δ

Now let's find the variance of $\delta = 2\bar{X}$.

Recall that for a single $X_i \sim U(0, \theta)$:

$$E[X_i] = \frac{\theta}{2}, \quad E[X_i^2] = \frac{\theta^2}{3},$$

$$\mathrm{Var}(X_i) = E[X_i^2] - (E[X_i])^2 = \frac{\theta^2}{3} - \left(\frac{\theta}{2}\right)^2 = \frac{\theta^2}{12}.$$

$$\begin{aligned}\mathrm{Var}(\bar{X}) &= \mathrm{Var}\left(\frac{1}{n}\sum X_i\right) \\ &= \frac{1}{n^2}\sum \mathrm{Var}(X_i) = \frac{n \cdot \mathrm{Var}(X_i)}{n^2} \\ &= \frac{\mathrm{Var}(X_i)}{n} = \frac{\theta^2}{12n}\end{aligned}$$

Variance of the Competing Estimator δ

Now let's find the variance of $\delta = 2\bar{X}$.

Recall that for a single $X_i \sim U(0, \theta)$:

$$E[X_i] = \frac{\theta}{2}, \quad E[X_i^2] = \frac{\theta^2}{3},$$

$$\text{Var}(X_i) = E[X_i^2] - (E[X_i])^2 = \frac{\theta^2}{3} - \left(\frac{\theta}{2}\right)^2 = \frac{\theta^2}{12}.$$

$$\begin{aligned}\text{Var}(\bar{X}) &= \text{Var}\left(\frac{1}{n}\sum X_i\right) \\ &= \frac{1}{n^2}\sum \text{Var}(X_i) = \frac{n \cdot \text{Var}(X_i)}{n^2} \\ &= \frac{\text{Var}(X_i)}{n} = \frac{\theta^2}{12n}\end{aligned}$$

Therefore, the variance of $\delta = 2\bar{X}$ is:

$$\text{Var}_\theta(\delta) = \text{Var}(2\bar{X}) = 4 \cdot \text{Var}(\bar{X}) = 4 \cdot \frac{\theta^2}{12n} = \frac{\theta^2}{3n}$$

The Final Comparison

Let's compare the two variances by taking their ratio.

- $\mathrm{Var}_\theta(\eta(T)) = \frac{\theta^2}{n(n+2)}$
- $\mathrm{Var}_\theta(\delta) = \frac{\theta^2}{3n}$

The Final Comparison

Let's compare the two variances by taking their ratio.

- $\mathrm{Var}_\theta(\eta(T)) = \frac{\theta^2}{n(n+2)}$
- $\mathrm{Var}_\theta(\delta) = \frac{\theta^2}{3n}$

The ratio of the variances is:

$$\frac{\mathrm{Var}_\theta(\eta(T))}{\mathrm{Var}_\theta(\delta)} = \frac{\theta^2/(n(n+2))}{\theta^2/(3n)} = \frac{3n}{n(n+2)} = \frac{3}{n+2}$$

The Final Comparison

Let's compare the two variances by taking their ratio.

- $\text{Var}_\theta(\eta(T)) = \frac{\theta^2}{n(n+2)}$
- $\text{Var}_\theta(\delta) = \frac{\theta^2}{3n}$

The ratio of the variances is:

$$\frac{\text{Var}_\theta(\eta(T))}{\text{Var}_\theta(\delta)} = \frac{\theta^2/(n(n+2))}{\theta^2/(3n)} = \frac{3n}{n(n+2)} = \frac{3}{n+2}$$

Conclusion

The ratio is always less than 1 for $n > 1$. As $n \to \infty$, the ratio tends to zero. This shows that the UMVU estimator $\eta(T)$ is significantly more accurate (has lower variance) than δ, especially for large sample sizes.

An Alternative Strategy

The proof of the Lehmann-Scheffé theorem suggests another powerful method for finding UMVU estimators.

1. Find **any** unbiased estimator for $g(\theta)$, let's call it $\delta(X)$. It doesn't have to be a good one.

An Alternative Strategy

The proof of the Lehmann-Scheffé theorem suggests another powerful method for finding UMVU estimators.

1. Find **any** unbiased estimator for $g(\theta)$, let's call it $\delta(X)$. It doesn't have to be a good one.
2. Find a **complete sufficient statistic**, T.

An Alternative Strategy

The proof of the Lehmann-Scheffé theorem suggests another powerful method for finding UMVU estimators.

1. Find **any** unbiased estimator for $g(\theta)$, let's call it $\delta(X)$. It doesn't have to be a good one.
2. Find a **complete sufficient statistic**, T.
3. Compute the conditional expectation of your simple estimator given the sufficient statistic:

$$\eta(T) = E[\delta(X)|T]$$

An Alternative Strategy

The proof of the Lehmann-Scheffé theorem suggests another powerful method for finding UMVU estimators.

1. Find **any** unbiased estimator for $g(\theta)$, let's call it $\delta(X)$. It doesn't have to be a good one.
2. Find a **complete sufficient statistic**, T.
3. Compute the conditional expectation of your simple estimator given the sufficient statistic:

$$\eta(T) = E[\delta(X)|T]$$

4. The resulting estimator, $\eta(T)$, is guaranteed to be the **UMVU estimator** for $g(\theta)$.

Example: Bernoulli Trials

The Setup

Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(θ) variables, where $P_\theta(X_i = 1) = \theta$.

Example: Bernoulli Trials

The Setup

Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(θ) variables, where $P_\theta(X_i = 1) = \theta$.

- The joint probability mass function belongs to an exponential family.
- The sum $T = \sum_{i=1}^{n} X_i$ is a **complete sufficient statistic**.
- We know that $T \sim \text{Binomial}(n, \theta)$.

Example: Bernoulli Trials

The Setup

Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(θ) variables, where $P_\theta(X_i = 1) = \theta$.

- The joint probability mass function belongs to an exponential family.
- The sum $T = \sum_{i=1}^n X_i$ is a **complete sufficient statistic**.
- We know that $T \sim \text{Binomial}(n, \theta)$.

Our Goal

Find the UMVU estimator for $g(\theta) = \theta^2$.

Example: Applying the Strategy

Step 1: Find any unbiased estimator for θ^2.

- Consider the estimator $\delta = X_1 X_2$.
- Is it unbiased? Let's check its expectation:

$$E_\theta[\delta] = E_\theta[X_1 X_2]$$

Example: Applying the Strategy

Step 1: Find any unbiased estimator for θ^2.

- Consider the estimator $\delta = X_1 X_2$.
- Is it unbiased? Let's check its expectation:

$$E_\theta[\delta] = E_\theta[X_1 X_2]$$

$$= E_\theta[X_1] E_\theta[X_2] \quad \text{(by independence)}$$
$$= \theta \cdot \theta = \theta^2$$

- Yes, $\delta = X_1 X_2$ is an unbiased estimator.

Example: Applying the Strategy

Step 1: Find any unbiased estimator for θ^2.

- Consider the estimator $\delta = X_1 X_2$.
- Is it unbiased? Let's check its expectation:

$$E_\theta[\delta] = E_\theta[X_1 X_2]$$

$$= E_\theta[X_1]E_\theta[X_2] \quad \text{(by independence)}$$
$$= \theta \cdot \theta = \theta^2$$

- Yes, $\delta = X_1 X_2$ is an unbiased estimator.

Step 2: Condition on the sufficient statistic T.

- The UMVU estimator is $\eta(T) = E[\delta | T] = E[X_1 X_2 | T]$.
- This is equivalent to finding $P(X_1 = 1, X_2 = 1 | T = t)$.

Example: The Calculation

We need to compute
$P(X_1 = 1, X_2 = 1 | T = t) = \frac{P(X_1=1, X_2=1, T=t)}{P(T=t)}$.

Numerator: The event $\{X_1 = 1, X_2 = 1, T = t\}$ means that X_1 and X_2 are 1, and the sum of the remaining $n - 2$ variables is $t - 2$.

$$
\begin{aligned}
& P(X_1 = 1, X_2 = 1, \sum_{i=3}^{n} X_i = t - 2) \\
& = P(X_1 = 1)P(X_2 = 1)P(\sum_{i=3}^{n} X_i = t - 2) \\
& = \theta \cdot \theta \cdot \binom{n-2}{t-2} \theta^{t-2}(1-\theta)^{(n-2)-(t-2)} \\
& = \binom{n-2}{t-2} \theta^{t}(1-\theta)^{n-t}
\end{aligned}
$$

Example: The Calculation

We need to compute
$P(X_1 = 1, X_2 = 1 | T = t) = \frac{P(X_1=1, X_2=1, T=t)}{P(T=t)}$.

Numerator: The event $\{X_1 = 1, X_2 = 1, T = t\}$ means that X_1 and X_2 are 1, and the sum of the remaining $n-2$ variables is $t-2$.

$$
\begin{aligned}
& P(X_1 = 1, X_2 = 1, \sum_{i=3}^{n} X_i = t-2) \\
&= P(X_1 = 1)P(X_2 = 1)P(\sum_{i=3}^{n} X_i = t-2) \\
&= \theta \cdot \theta \cdot \binom{n-2}{t-2} \theta^{t-2}(1-\theta)^{(n-2)-(t-2)} \\
&= \binom{n-2}{t-2} \theta^{t}(1-\theta)^{n-t}
\end{aligned}
$$

Denominator: This is just the Binomial PMF.

$$P(T = t) = \binom{n}{t} \theta^{t}(1-\theta)^{n-t}$$

Example: The UMVU Estimator

Now we compute the ratio:

$$\begin{aligned}
\eta(t) &= \frac{\binom{n-2}{t-2}\theta^t(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\
&= \frac{(n-2)!}{(t-2)!(n-t)!} \cdot \frac{t!(n-t)!}{n!} \\
&= \frac{t(t-1)}{n(n-1)}
\end{aligned}$$

Example: The UMVU Estimator

Now we compute the ratio:

$$\begin{aligned}\eta(t) &= \frac{\binom{n-2}{t-2}\theta^t(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\ &= \frac{(n-2)!}{(t-2)!(n-t)!} \cdot \frac{t!(n-t)!}{n!} \\ &= \frac{t(t-1)}{n(n-1)}\end{aligned}$$

The UMVU estimator for $g(\theta) = \theta^2$ is $\eta(T) = \frac{T(T-1)}{n(n-1)}$.

Should We Only Use Unbiased Estimators?

The UMVU framework is powerful, but it rests on a strict premise: that we should only consider unbiased estimators.

A Critical Question

Is a small amount of bias acceptable if it leads to a better estimator overall?

Should We Only Use Unbiased Estimators?

The UMVU framework is powerful, but it rests on a strict premise: that we should only consider unbiased estimators.

A Critical Question

Is a small amount of bias acceptable if it leads to a better estimator overall?

- Estimators with large bias are generally not useful.
- However, an estimator with a small bias might outperform the "best" unbiased estimator.
- This leads to the crucial concept of the **bias-variance tradeoff**.

Uniform Example Revisited: Is UMVU Optimal?

Recall: $X_1, \ldots, X_n \sim U(0, \theta)$ and $T = \max\{X_i\}$.

- The UMVU estimator for θ is $\eta(T) = \frac{n+1}{n} T$.
- This is a multiple of the sufficient statistic T.
- **Question:** Is this the *best* multiple of T?

Uniform Example Revisited: Is UMVU Optimal?

Recall: $X_1, \ldots, X_n \sim U(0, \theta)$ and $T = \max\{X_i\}$.

- The UMVU estimator for θ is $\eta(T) = \frac{n+1}{n} T$.
- This is a multiple of the sufficient statistic T.
- **Question:** Is this the *best* multiple of T?

Let's consider a general estimator of the form $\delta_a = aT$ and find the value of a that minimizes the squared error risk, $R(\theta, \delta_a) = E_\theta[(aT - \theta)^2]$.

Uniform Example: Calculating the Risk

From previous calculations, we know:

$$E_\theta[T] = \frac{n}{n+1}\theta \quad \text{and} \quad E_\theta[T^2] = \frac{n}{n+2}\theta^2$$

Uniform Example: Calculating the Risk

From previous calculations, we know:

$$E_\theta[T] = \frac{n}{n+1}\theta \quad \text{and} \quad E_\theta[T^2] = \frac{n}{n+2}\theta^2$$

The risk of $\delta_a = aT$ is:

$$\begin{aligned}
R(\theta, \delta_a) &= E_\theta[(aT - \theta)^2] \\
&= a^2 E_\theta[T^2] - 2a\theta E_\theta[T] + \theta^2 \\
&= a^2 \left(\frac{n}{n+2}\theta^2\right) - 2a\theta \left(\frac{n}{n+1}\theta\right) + \theta^2 \\
&= \theta^2 \left[\left(\frac{n}{n+2}\right) a^2 - \left(\frac{2n}{n+1}\right) a + 1\right]
\end{aligned}$$

This is a quadratic in a.

Uniform Example: Minimizing the Risk

The quadratic $R(\theta, \delta_a)$ is minimized when $a = \frac{n+2}{n+1}$.

Uniform Example: Minimizing the Risk

The quadratic $R(\theta, \delta_a)$ is minimized when $a = \frac{n+2}{n+1}$.
Let's compare our estimators:

- **UMVU Estimator:** $a_{UMVU} = \frac{n+1}{n}$.
- **Minimum Risk Estimator:** $a_{MR} = \frac{n+1}{n+2}$.

Uniform Example: Minimizing the Risk

The quadratic $R(\theta, \delta_a)$ is minimized when $a = \frac{n+2}{n+1}$.
Let's compare our estimators:

- **UMVU Estimator:** $a_{UMVU} = \frac{n+1}{n}$.
- **Minimum Risk Estimator:** $a_{MR} = \frac{n+1}{n+2}$.

The minimum risk for the biased estimator is $R(\theta, \delta_{a_{MR}}) = \frac{\theta^2}{(n+2)^2}$.
The risk (variance) of the UMVU estimator was
$R(\theta, \eta(T)) = \frac{\theta^2}{n(n+2)}$.

Uniform Example: Minimizing the Risk

The quadratic $R(\theta, \delta_a)$ is minimized when $a = \frac{n+2}{n+1}$.
Let's compare our estimators:

- **UMVU Estimator:** $a_{UMVU} = \frac{n+1}{n}$.
- **Minimum Risk Estimator:** $a_{MR} = \frac{n+1}{n+2}$.

The minimum risk for the biased estimator is $R(\theta, \delta_{a_{MR}}) = \frac{\theta^2}{(n+2)^2}$.
The risk (variance) of the UMVU estimator was
$R(\theta, \eta(T)) = \frac{\theta^2}{n(n+2)}$.
The biased estimator $\frac{n+1}{n+2}T$ has a slightly smaller risk than the UMVU estimator. A small amount of bias was traded for a larger reduction in variance.

The Bias-Variance Decomposition

The risk of any estimator δ under squared error loss can be decomposed:

$$\begin{aligned} R(\theta,\delta) &= E_\theta[(\delta - g(\theta))^2] \\ &= E_\theta[(\delta - E\delta + E\delta - g(\theta))^2] \\ &= E_\theta[(\delta - E\delta)^2] + (E\delta - g(\theta))^2 \\ &= \mathrm{Var}_\theta(\delta) + (\mathrm{Bias}_\theta(\delta))^2 \end{aligned}$$

The Tradeoff

Sometimes, by accepting a small amount of bias, we can decrease the variance by an even larger amount, leading to a lower overall risk. This kind of trade-off between bias and variance arises fairly often in statistics. In nonparametric curve estimations these trade-offs often play a key role (Sec 18.1)

Example: A Truncated Poisson Distribution

The Setup

Suppose X has a Poisson distribution truncated to $\{1, 2, 3, \ldots\}$.
The PMF is:

$$P_\theta(X = x) = \frac{\theta^x e^{-\theta}}{x!(1 - e^{-\theta})}, \quad x = 1, 2, \ldots$$

X is a complete sufficient statistic.

Example: A Truncated Poisson Distribution

The Setup

Suppose X has a Poisson distribution truncated to $\{1, 2, 3, \dots\}$. The PMF is:

$$P_\theta(X = x) = \frac{\theta^x e^{-\theta}}{x!(1 - e^{-\theta})}, \quad x = 1, 2, \dots$$

X is a complete sufficient statistic.

Our Goal

Find the UMVU estimator for $g(\theta) = e^{-\theta}$ (the probability of a zero, which was truncated).

Example: Finding the Unbiased Estimator

If $\delta(X)$ is unbiased for $g(\theta) = e^{-\theta}$, then $E_\theta[\delta(X)] = e^{-\theta}$.

$$\sum_{k=1}^{\infty} \delta(k) \frac{\theta^k e^{-\theta}}{k!(1 - e^{-\theta})} = e^{-\theta}$$

Example: Finding the Unbiased Estimator

If $\delta(X)$ is unbiased for $g(\theta) = e^{-\theta}$, then $E_\theta[\delta(X)] = e^{-\theta}$.

$$\sum_{k=1}^{\infty} \delta(k) \frac{\theta^k e^{-\theta}}{k!(1 - e^{-\theta})} = e^{-\theta}$$

Multiplying both sides by $(1 - e^{-\theta})e^{\theta}$:

$$\sum_{k=1}^{\infty} \frac{\delta(k)}{k!} \theta^k = 1 - e^{-\theta}$$

Example: Finding the Unbiased Estimator

If $\delta(X)$ is unbiased for $g(\theta) = e^{-\theta}$, then $E_\theta[\delta(X)] = e^{-\theta}$.

$$\sum_{k=1}^{\infty} \delta(k) \frac{\theta^k e^{-\theta}}{k!(1 - e^{-\theta})} = e^{-\theta}$$

Multiplying both sides by $(1 - e^{-\theta})e^{\theta}$:

$$\sum_{k=1}^{\infty} \frac{\delta(k)}{k!} \theta^k = 1 - e^{-\theta}$$

We know the power series for $1 - e^{-\theta}$:

$$1 - e^{-\theta} = 1 - \sum_{k=0}^{\infty} \frac{(-\theta)^k}{k!} = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k!} \theta^k$$

Example: Finding the Unbiased Estimator

If $\delta(X)$ is unbiased for $g(\theta) = e^{-\theta}$, then $E_\theta[\delta(X)] = e^{-\theta}$.

$$\sum_{k=1}^{\infty} \delta(k) \frac{\theta^k e^{-\theta}}{k!(1 - e^{-\theta})} = e^{-\theta}$$

Multiplying both sides by $(1 - e^{-\theta})e^{\theta}$:

$$\sum_{k=1}^{\infty} \frac{\delta(k)}{k!} \theta^k = 1 - e^{-\theta}$$

We know the power series for $1 - e^{-\theta}$:

$$1 - e^{-\theta} = 1 - \sum_{k=0}^{\infty} \frac{(-\theta)^k}{k!} = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k!} \theta^k$$

By matching the coefficients of θ^k in the two power series, we must have:

$$\frac{\delta(k)}{k!} = \frac{(-1)^{k+1}}{k!} \implies \delta(k) = (-1)^{k+1}$$

Example: An Absurd Estimator

The unique unbiased estimator (and therefore the UMVU) is:

$$\delta(X) = (-1)^{X+1}$$

Example: An Absurd Estimator

The unique unbiased estimator (and therefore the UMVU) is:

$$\delta(X) = (-1)^{X+1}$$

What does this mean?

- If X is odd (1, 3, 5, ...), the estimate for $e^{-\theta}$ is 1.
- If X is even (2, 4, 6, ...), the estimate for $e^{-\theta}$ is -1.

Example: An Absurd Estimator

The unique unbiased estimator (and therefore the UMVU) is:

$$\delta(X) = (-1)^{X+1}$$

What does this mean?

- If X is odd (1, 3, 5, ...), the estimate for $e^{-\theta}$ is 1.
- If X is even (2, 4, 6, ...), the estimate for $e^{-\theta}$ is -1.

Conclusion

This estimator is absurd as it estimates a probability to be 1 or -1. In this case, the strict requirement of unbiasedness leads to an uninformative result. It would be far better to use a reasonable, but biased, estimator.