Math 3081 (Probability and Statistics)

Lecture #15 of 27 $\sim$ July 29, 2021

Properties of Estimators: Bias and Efficiency

- Biased and Unbiased Estimators
- Efficiency of Estimators
- The Cramèr-Rao Bound

This material represents §3.1.2-3.1.3 from the course notes.

Recall

Last lecture, we discussed how to perform maximum likelihood estimates to construct estimators for unknown parameters.

- In general, we might hope that the estimators that we can cook up using maximum likelihood estimation would be "the best" possible estimators of our unknown parameters.
- In general, there are many possible estimators for any given parameter, and it is not always clear which one we should use.
- Our present goal is to discuss various properties of estimators that capture various aspects of this desire to find "the best" possible estimator.

Motivation

Recall the German tank problem from last time: we sample the uniform distribution on $[0, \theta]$ to obtain the values $x_1, x_2, \ldots, x_n$.

- The maximum likelihood estimator is $\hat{\theta} = \max(x_1, x_2, \ldots, x_n)$.
- However... does it seem entirely reasonable that "best estimate" $\hat{\theta}$ for the number of enemy tanks is simply the largest number observed?

Motivation

Recall the German tank problem from last time: we sample the uniform distribution on $[0, \theta]$ to obtain the values $x_1, x_2, \dots, x_n$.

- The maximum likelihood estimator is $\hat{\theta} = \max(x_1, x_2, \dots, x_n)$.
- However... does it seem entirely reasonable that "best estimate" $\hat{\theta}$ for the number of enemy tanks is simply the largest number observed?
- Perhaps not! After all, $\hat{\theta}$ is always the lowest feasible number of tanks that is consistent with the observed data.
- That means there is a reasonably good chance that the actual number of tanks is larger than $\hat{\theta}$, since it is not especially likely we would actually see the one with the largest number.
- This suggests that the maximum likelihood estimate should, in general, tend to underestimate the actual correct value of θ.

Bias, I

Intuitively, we might prefer to search for an estimator that tends to have the smallest systematic error.

- The most basic possible requirement is to ask that the estimator not have any "bias", on average, away from the expected value of the parameter.
- To study this idea more precisely, we will shift our emphasis and begin viewing estimators as random variables on the space of possible input data.

Bias, II

When we view the estimator as a random variable, we can phrase this requirement for a lack of bias in terms of expected value:

Definition

An estimator $\hat{\theta}(x_1, x_2, \ldots, x_n)$ for a set of observations $x_1, \ldots, x_n$ drawn by randomly sampling a random variable X with probability density function $f_X(x; \theta)$ is <u>unbiased</u> if $E(\hat{\theta}) = \theta$ for all θ.

More verbosely, if we fix θ and then average over all possible samples $x_1, \ldots, x_n$ of X for a fixed value of θ, then $\hat{\theta}$ is unbiased when the expected value of the estimator $\hat{\theta}$ is equal to the true value of the parameter θ.

Bias, III

Example: Show that the maximum likelihood estimate $\hat{\mu} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$ for sampling the normal distribution with mean μ and fixed standard deviation σ is unbiased.

Bias, III

<u>Example</u>: Show that the maximum likelihood estimate $\hat{\mu} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$ for sampling the normal distribution with mean μ and fixed standard deviation σ is unbiased.

- Note that by properties of expected value, we have $E(\hat{\mu}) = \dfrac{1}{n}[E(x_1) + E(x_2) + \cdots + E(x_n)]$.
- Furthermore, we have $E(x_i) = \mu$ because each x_i is sampled randomly from a distribution with mean μ.
- Thus, we have $E(\hat{\mu}) = \dfrac{1}{n}[n\mu] = \mu$, and so μ is unbiased.

Bias, IV

We can actually generalize the previous example quite substantially, as follows:

- Suppose X is any random variable with finite mean μ, and $x_1, \ldots, x_n$ is an independent random sample of X.
- Then the estimator $\hat{\mu} = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$ for any choice of constants a_i such that $\sum_i a_i = 1$ will be unbiased.
- This follows by the same argument we gave on the previous slide (i.e., by using the additivity and linearity of expected value).

Bias, V

<u>Example</u>: Show that the maximum likelihood estimate for the variance $\hat{\sigma}^2 = \frac{1}{n}(x_1^2 + x_2^2 + \cdots + x_n^2) - \left[\frac{x_1 + x_2 + \cdots + x_n}{n}\right]^2$ from sampling the normal distribution with unknown mean μ and standard deviation σ is biased.

Bias, V

Example: Show that the maximum likelihood estimate for the variance $\hat{\sigma}^2 = \frac{1}{n}(x_1^2 + x_2^2 + \cdots + x_n^2) - \left[\frac{x_1 + x_2 + \cdots + x_n}{n}\right]^2$ from sampling the normal distribution with unknown mean μ and standard deviation σ is biased.

- This example is a bit complicated and requires using a number of properties of expected value.
- Recall that $\mu = E(x_i)$ and $\sigma^2 = E(x_i^2) - E(x_i)^2$, so $E(x_i^2) = \sigma^2 + \mu^2$.
- Then, because variance is additive for independent random variables, we have $\mathrm{var}(x_1+x_2+\cdots+x_n) = \mathrm{var}(x_1)+\mathrm{var}(x_2)+\cdots+\mathrm{var}(x_n) = n\sigma^2$.
- Since $\mathrm{var}(S) = E(S^2) - E(S)^2$, applying this for $S = x_1 + x_2 + \cdots + x_n$ yields $E(S^2) = \mathrm{var}(S) + E(S)^2 = n\sigma^2 + n^2\mu^2$.

Bias, VI

Example: Show that the maximum likelihood estimate for the variance $\hat{\sigma}^2 = \frac{1}{n}(x_1^2 + x_2^2 + \cdots + x_n^2) - \left[\frac{x_1 + x_2 + \cdots + x_n}{n}\right]^2$ from sampling the normal distribution with unknown mean μ and standard deviation σ is biased.

- Note $E(x_i^2) = \sigma^2 + \mu^2$, $E[(x_1 + \cdots + x_n)^2] = n\sigma^2 + n^2\mu^2$.
- Then by properties of expected value, we have $E(\hat{\sigma}^2) = \frac{1}{n}E(x_1^2 + x_2^2 + \cdots + x_n^2) - \frac{1}{n^2}E(x_1 + x_2 + \cdots + x_n)^2$ $= \frac{1}{n} \cdot n(\sigma^2 + \mu^2) - \frac{1}{n^2}(n\sigma^2 + n^2\mu^2) = \frac{n-1}{n}\sigma^2$.
- The expected value is not equal to σ^2 because of the factor of $\frac{n-1}{n}$, so this estimator is biased.

Bias, VII

In this last example, we can construct an unbiased estimator of σ^2 by scaling $\hat{\sigma}^2$ by $\dfrac{n}{n-1}$:

Definition

The estimator $S^2 = \dfrac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n-1}$, *where* $\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$ *is the sample average, is called the* *sample variance*. *Its square root* S *is called the* *sample standard deviation*.

- Despite the fact that $E(S) \neq \sigma$ (although this itself is not easy to show), S is quite commonly used as an estimator for σ because the estimate of σ^2 by S^2 is unbiased.

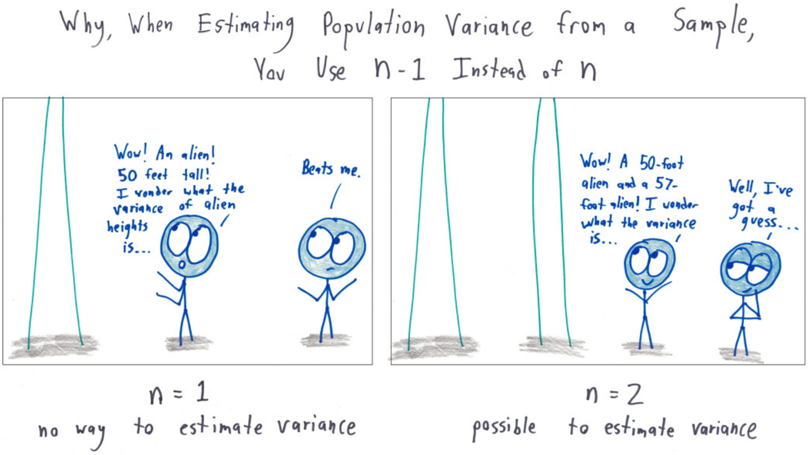
Bias, VIII

The use of $n - 1$ in place of n in the denominator of the sample variance is known as "Bessel's correction".

- Roughly speaking, the correction is required because measuring the variance of the sample relative to the sample mean (rather than relative to the unknown true mean μ) will always lower the estimated variance.
- Thus, as we calculated, a correction is needed to unbias the estimate.

Bias, IX

Here is some intuitive motivation for why the factor $\frac{1}{n-1}$ appears:



Drawing credit: Ben Orlin, `mathwithbaddrawings.com` (2017).

Bias, X

Example: Show that the maximum likelihood estimator $\hat{\theta} = \max(x_1, x_2, \ldots, x_n)$ from sampling the uniform distribution on $[0, \theta]$ is biased.

Bias, X

Example: Show that the maximum likelihood estimator $\hat{\theta} = \max(x_1, x_2, \ldots, x_n)$ from sampling the uniform distribution on $[0, \theta]$ is biased.

- Here, we need to compute the expected value of $\hat{\theta}$, which requires us to find the underlying probability distribution.
- Observe that, for any $0 \leq x \leq \theta$, we have $P(\hat{\theta} \leq x) = (x/\theta)^n$ because $\hat{\theta} \leq k$ occurs precisely when all of the values $x_1, x_2, \ldots, x_n$ lie in the interval $[0, x]$, which occurs with probability $(x/\theta)^n$.
- This means that the cumulative distribution function for $\hat{\theta}$ is $g_{\hat{\theta}}(x) = (x/\theta)^n$ for $0 \leq x \leq \theta$, and so its probability distribution function is the derivative $g'_{\hat{\theta}}(x) = nx^{n-1}/\theta^n$.

Bias, XI

Example: Show that the maximum likelihood estimator $\hat{\theta} = \max(x_1, x_2, \ldots, x_n)$ from sampling the uniform distribution on $[0, \theta]$ is biased.

- We have the pdf $p_{\hat{\theta}}(x) = nx^{n-1}/\theta^n$ for $0 \leq x \leq \theta$.

Bias, XI

Example: Show that the maximum likelihood estimator $\hat{\theta} = \max(x_1, x_2, \ldots, x_n)$ from sampling the uniform distribution on $[0, \theta]$ is biased.

- We have the pdf $p_{\hat{\theta}}(x) = nx^{n-1}/\theta^n$ for $0 \le x \le \theta$.
- Now we may compute $E(\hat{\theta}) = \int_0^\theta x p_{\hat{\theta}}(x)\,dx = \int_0^\theta nx^n/\theta^n\,dx = \dfrac{n}{n+1}\theta$.
- Since this is not equal to θ, we see that $\hat{\theta}$ is biased, as claimed.

Like with the sample variance of the normal distribution from earlier, we can rescale the maximum likelihood estimate to obtain an unbiased estimator of θ, namely, $\hat{\theta} = \dfrac{n+1}{n}\max(x_1, x_2, \ldots, x_n)$.

Bias, XII

Example: Show that the estimator $\hat{\theta} = \frac{2}{n}(x_1 + x_2 + \cdots + x_n)$ from sampling the uniform distribution on $[0, \theta]$ is unbiased.

Bias, XII

Example: Show that the estimator $\hat{\theta} = \frac{2}{n}(x_1 + x_2 + \cdots + x_n)$ from sampling the uniform distribution on $[0, \theta]$ is unbiased.

- Since each x_i is drawn from the uniform distribution on $[0, \theta]$, we have $E(x_i) = \int_0^\theta x \cdot \frac{1}{\theta} dx = \frac{\theta}{2}$.
- Then, by properties of expected value, we have $E(\hat{\theta}) = \frac{2}{n}(E[x_1] + E[x_2] + \cdots + E[x_n]) = \frac{2}{n} \cdot n \cdot \frac{\theta}{2} = \theta$.
- Therefore, $\hat{\theta}$ is unbiased, as claimed.

Efficiency, I

As we have already remarked, for any given parameter estimation problem, there are many different possible choices for estimators.

- One desirable quality for an estimator is that it be unbiased. However, this requirement alone does not impose a substantial condition, since (as we have seen) there can exist several different unbiased estimators for a given parameter.
- In the last two examples, we constructed two different unbiased estimators for the parameter θ, given a random sample $x_1, x_2, \ldots, x_n$ from the uniform distribution on $[0, \theta]$: namely, $\hat{\theta}_1 = \dfrac{n+1}{n} \max(x_1, x_2, \ldots, x_n)$ and $\hat{\theta}_2 = \dfrac{2}{n}(x_1 + x_2 + \cdots + x_n)$.

Efficiency, II

- Likewise, it is also not hard to see that, given a random sample x_1, x_2 from the normal distribution with mean θ and standard deviation σ, the estimators $\hat{\theta}_1 = \frac{1}{2}(x_1 + x_2)$ and $\hat{\theta}_2 = \frac{1}{3}(x_1 + 2x_2)$ are also both unbiased.
- More generally, any estimator of the form $ax_1 + (1-a)x_2$ will be an unbiased estimator of the mean.

We would now like to know if there is a meaningful way to say one of these unbiased estimators is better than the other.

Efficiency, III

In the abstract, it seems reasonable to say that an estimator with a smaller variance is better than one with a larger variance, since a smaller variance would indicate that the value of the estimator stays closer to the "true" parameter value more often.

We formalize this as follows:

Definition

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators for the parameter θ, we say that $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ if $\mathrm{var}(\hat{\theta}_1) < \mathrm{var}(\hat{\theta}_2)$.

Efficiency, IV

Example: Suppose that a random sample x, y is taken from the normal distribution with mean θ and standard deviation σ.

1. Find the variance of the estimator $\hat{\theta}_1 = \frac{1}{2}(x + y)$.
2. Find the variance of the estimator $\hat{\theta}_2 = \frac{1}{3}(x + 2y)$.
3. Show that $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of θ.
4. Which of $\hat{\theta}_1, \hat{\theta}_2$ is a more efficient estimator of θ?
5. More generally, for $\hat{\theta}_a = ax + (1 - a)y$, which value of a produces the most efficient estimator?

Efficiency, IV

Example: Suppose that a random sample x, y is taken from the normal distribution with mean θ and standard deviation σ.

1. Find the variance of the estimator $\hat{\theta}_1 = \frac{1}{2}(x+y)$.
2. Find the variance of the estimator $\hat{\theta}_2 = \frac{1}{3}(x+2y)$.
3. Show that $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of θ.
4. Which of $\hat{\theta}_1, \hat{\theta}_2$ is a more efficient estimator of θ?
5. More generally, for $\hat{\theta}_a = ax + (1-a)y$, which value of a produces the most efficient estimator?

- To compute the variances and check for unbiasedness, we will use properties of expected value and the additivity of variance for independent variables.

Efficiency, V

Example: Suppose that a random sample x, y is taken from the normal distribution with mean θ and standard deviation σ.

1. Find the variance of the estimator $\hat{\theta}_1 = \frac{1}{2}(x + y)$.

Efficiency, V

Example: Suppose that a random sample x, y is taken from the normal distribution with mean θ and standard deviation σ.

1. Find the variance of the estimator $\hat{\theta}_1 = \frac{1}{2}(x + y)$.

- Note that because x and y are independent, their variances are additive, and $\mathrm{var}(x) = \mathrm{var}(y) = \sigma^2$.
- Then, we have

$$\begin{aligned}
\mathrm{var}(\hat{\theta}_1) &= \mathrm{var}(\tfrac{1}{2}x + \tfrac{1}{2}y) \\
&= \mathrm{var}(\tfrac{1}{2}x) + \mathrm{var}(\tfrac{1}{2}y) \\
&= \tfrac{1}{4}\mathrm{var}(x) + \tfrac{1}{4}\mathrm{var}(y) \\
&= \tfrac{1}{4}\sigma^2 + \tfrac{1}{4}\sigma^2 = \tfrac{1}{2}\sigma^2.
\end{aligned}$$

- Thus, $\mathrm{var}(\hat{\theta}_2) = (1/2)\sigma^2$.

Efficiency, VI

Example: Suppose that a random sample x, y is taken from the normal distribution with mean θ and standard deviation σ.

2. Find the variance of the estimator $\hat{\theta}_2 = \frac{1}{3}(x + 2y)$.

Efficiency, VI

Example: Suppose that a random sample x, y is taken from the normal distribution with mean θ and standard deviation σ.

2. Find the variance of the estimator $\hat{\theta}_2 = \frac{1}{3}(x + 2y)$.

- In the same way, we have

$$\begin{aligned} \mathrm{var}(\hat{\theta}_2) &= \mathrm{var}(\tfrac{1}{3}x + \tfrac{2}{3}y) \\ &= \mathrm{var}(\tfrac{1}{3}x) + \mathrm{var}(\tfrac{2}{3}y) \\ &= \tfrac{1}{9}\mathrm{var}(x) + \tfrac{4}{9}\mathrm{var}(y) \\ &= \tfrac{1}{9}\sigma^2 + \tfrac{4}{9}\sigma^2 = \tfrac{5}{9}\sigma^2 \end{aligned}$$

- Thus, $\mathrm{var}(\hat{\theta}_2) = (5/9)\sigma^2$.

Efficiency, VII

<u>Example</u>: Suppose that a random sample x, y is taken from the normal distribution with mean θ and standard deviation σ.

3. Show that $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of θ.

Efficiency, VII

Example: Suppose that a random sample x, y is taken from the normal distribution with mean θ and standard deviation σ.

3. Show that $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of θ.

- We have $E(\hat{\theta}_1) = E(\frac{1}{2}(x+y)) = \frac{1}{2}E(x) + \frac{1}{2}E(y) = \theta$.
- Likewise, $E(\hat{\theta}_1) = E(\frac{2}{3}(x+2y)) = \frac{1}{3}E(x) + \frac{2}{3}E(y) = \theta$.
- Thus, both estimators are unbiased.

4. Which of $\hat{\theta}_1, \hat{\theta}_2$ is a more efficient estimator of θ?

Efficiency, VII

Example: Suppose that a random sample x, y is taken from the normal distribution with mean θ and standard deviation σ.

3. Show that $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of θ.
   - We have $E(\hat{\theta}_1) = E(\frac{1}{2}(x+y)) = \frac{1}{2}E(x) + \frac{1}{2}E(y) = \theta$.
   - Likewise, $E(\hat{\theta}_1) = E(\frac{2}{3}(x+2y)) = \frac{1}{3}E(x) + \frac{2}{3}E(y) = \theta$.
   - Thus, both estimators are unbiased.
4. Which of $\hat{\theta}_1, \hat{\theta}_2$ is a more efficient estimator of θ?
   - Since $\mathrm{var}(\hat{\theta}_1) = (1/2)\sigma^2$ while $\mathrm{var}(\hat{\theta}_2) = (5/9)\sigma^2$, we see $\hat{\theta}_1$ is more efficient since $1/2 < 5/9$.

Efficiency, VIII

Example: Suppose that a random sample x, y is taken from the normal distribution with mean θ and standard deviation σ.

5. More generally, for $\hat{\theta}_a = ax + (1-a)y$, which value of a produces the most efficient estimator?

Efficiency, VIII

Example: Suppose that a random sample x, y is taken from the normal distribution with mean θ and standard deviation σ.

5. More generally, for $\hat{\theta}_a = ax + (1-a)y$, which value of a produces the most efficient estimator?

- In the same way as before, we can compute

$$\begin{aligned}\operatorname{var}(\hat{\theta}_a) &= \operatorname{var}(ax + (1-a)y) \\ &= a^2\operatorname{var}(x) + (1-a)^2\operatorname{var}(y) \\ &= [a^2 + (1-a)^2]\sigma^2 = (2a^2 - 2a + 1)\sigma^2.\end{aligned}$$

- By calculus (the derivative is $4a - 2$ which is zero for $a = 1/2$) or by completing the square $(2a^2 - 2a + 1 = 2(a - 1/2)^2 + 1/2)$ we can see that the minimum of the quadratic occurs when $a = 1/2$.
- Thus, in fact, $\hat{\theta}_1$ is the most efficient estimator of this form.

Efficiency, IX

<u>Example</u>: Suppose that a random sample x, y is taken from the normal distribution with mean θ and standard deviation σ.

5. More generally, for $\hat{\theta}_a = ax + (1-a)y$, which value of a produces the most efficient estimator? (Answer: $a = 1/2$.)

- Intuitively, this last calculation should make sense, because if we put more weight on one observation, its variation will tend to dominate the calculation.
- In the extreme situation of taking $\hat{\theta}_3 = x_2$ (which corresponds to $a = 0$), for example, we see that the variance is simply σ^2, which is much larger than the variance arising from the average.
- This is quite reasonable, since the average $\frac{1}{2}(x_1 + x_2)$ uses a bigger sample and thus captures more information than just using a single observation.

Efficiency, X

Example: A random sample $x_1, x_2, \ldots, x_n$ is taken from the uniform distribution on $[0, \theta]$.

1. Find the variance of the unbiased estimator $\hat{\theta}_1 = \dfrac{n+1}{n} \max(x_1, \ldots, x_n)$.
2. Find the variance of the unbiased estimator $\hat{\theta}_2 = \dfrac{2}{n}(x_1 + x_2 + \cdots + x_n)$.
3. Which estimator is a more efficient estimator for θ?

Efficiency, XI

Example: A random sample $x_1, x_2, \ldots, x_n$ is taken from the uniform distribution on $[0, \theta]$.

1. Find the variance of $\hat{\theta}_1 = \frac{n+1}{n} \max(x_1, \ldots, x_n)$.

Efficiency, XI

<u>Example</u>: A random sample $x_1, x_2, \dots, x_n$ is taken from the uniform distribution on $[0, \theta]$.

1. Find the variance of $\hat{\theta}_1 = \frac{n+1}{n} \max(x_1, \dots, x_n)$.

- To compute the variance of $\hat{\theta}_1$, we use the pdf of $\max(x_1, \dots, x_n)$, which is $g(x) = nx^{n-1}/\theta^n$ for $0 \le x \le \theta$.
- Then $E[\max(x_1, \dots, x_n)^2] = \int_0^\theta x^2 \cdot nx^{n-1}/\theta^n \, dx = \frac{n}{n+2}\theta^2$.
- Also, $E[\max(x_1, \dots, x_n)] = \int_0^\theta x \cdot nx^{n-1}/\theta^n \, dx = \frac{n}{n+1}\theta$, so $\mathrm{var}[\max(x_1, \dots, x_n)] = \frac{n}{n+2}\theta^2 - \left[\frac{n}{n+1}\theta\right]^2 = \frac{n}{(n+2)(n+1)^2}\theta^2$.
- Therefore,
$$\mathrm{var}(\hat{\theta}_1) = \left[\frac{n+1}{n}\right]^2 \mathrm{var}[\max(x_1, \dots, x_n)] = \frac{1}{n(n+2)}\theta^2.$$

Efficiency, XII

Example: A random sample $x_1, x_2, \ldots, x_n$ is taken from the uniform distribution on $[0, \theta]$.

2. Find the variance of $\hat{\theta}_2 = \dfrac{2}{n}(x_1 + x_2 + \cdots + x_n)$.

Efficiency, XII

Example: A random sample $x_1, x_2, \dots, x_n$ is taken from the uniform distribution on $[0, \theta]$.

2. Find the variance of $\hat{\theta}_2 = \frac{2}{n}(x_1 + x_2 + \cdots + x_n)$.

- For $\hat{\theta}_2$, since the x_i are independent, their variances are additive.
- We have $\mathrm{var}(x_i) = \int_0^\theta (x - \theta/2)^2 \cdot \frac{1}{\theta}\, dx = \frac{\theta^2}{12}$.
- Thus,

  $\mathrm{var}(\hat{\theta}_2) = \mathrm{var}(\frac{2}{n}x_1) + \cdots + \mathrm{var}(\frac{2}{n}x_n) = n \cdot \frac{4}{n^2} \cdot \frac{\theta^2}{12} = \frac{1}{3n}\theta^2$.

Efficiency, XIII

Example: A random sample $x_1, x_2, \ldots, x_n$ is taken from the uniform distribution on $[0, \theta]$.

3. Which estimator is a more efficient estimator for θ?

Efficiency, XIII

Example: A random sample $x_1, x_2, \ldots, x_n$ is taken from the uniform distribution on $[0, \theta]$.

3. Which estimator is a more efficient estimator for θ?

- We just calculated $\mathrm{var}(\hat{\theta}_1) = \dfrac{1}{n(n+2)}\theta^2$ and $\mathrm{var}(\hat{\theta}_2) = \dfrac{1}{3n}\theta^2$.
- For $n = 1$ these variances are the same (this is unsurprising because when $n = 1$ the estimators themselves are the same!).
- For $n > 1$ we see that the variance of $\hat{\theta}_1$ is smaller since $\dfrac{1}{n+2} < \dfrac{1}{3}$, so $\hat{\theta}_1$ is more efficient.

Efficiency, XIV

Example: Suppose x and y are respectively drawn from two independent normal distributions X and Y with the same unknown mean $E(X) = E(Y) = \theta$ but different known variances $\mathrm{var}(X) = \sigma^2$ and $\mathrm{var}(Y) = 2\sigma^2$.

1. Show that for any parameter $0 \le a \le 1$ the estimator $\hat{\theta}_a = ax + (1-a)y$ is an unbiased estimator of θ.
2. Find the value of a yielding the most efficient estimator of θ.

Efficiency, XV

Example: Suppose x and y are respectively drawn from two independent normal distributions X and Y with the same unknown mean $E(X) = E(Y) = \theta$ but different known variances $\mathrm{var}(X) = \sigma^2$ and $\mathrm{var}(Y) = 2\sigma^2$.

1. Show that for any parameter $0 \le a \le 1$ the estimator $\hat{\theta}_a = ax + (1-a)y$ is an unbiased estimator of θ.

Efficiency, XV

Example: Suppose x and y are respectively drawn from two independent normal distributions X and Y with the same unknown mean $E(X) = E(Y) = \theta$ but different known variances $\mathrm{var}(X) = \sigma^2$ and $\mathrm{var}(Y) = 2\sigma^2$.

1. Show that for any parameter $0 \leq a \leq 1$ the estimator $\hat{\theta}_a = ax + (1-a)y$ is an unbiased estimator of θ.

- By the linearity of expected value, we have $E(\hat{\theta}_a) = aE(x) + (1-a)E(y) = a\theta + (1-a)\theta = \theta$. Thus, $\hat{\theta}_a$ is unbiased for each value of a.
- Note that this is essentially the same calculation we have made several times before, and has nothing to do with the standard deviations of the distributions given.

Efficiency, XVI

Example: Suppose x and y are respectively drawn from two independent normal distributions X and Y with the same unknown mean $E(X) = E(Y) = \theta$ but different known variances $\mathrm{var}(X) = \sigma^2$ and $\mathrm{var}(Y) = 2\sigma^2$.

2. Find the value of a yielding the most efficient estimator of θ.

Efficiency, XVI

<u>Example</u>: Suppose x and y are respectively drawn from two independent normal distributions X and Y with the same unknown mean $E(X) = E(Y) = \theta$ but different known variances $\mathrm{var}(X) = \sigma^2$ and $\mathrm{var}(Y) = 2\sigma^2$.

2. Find the value of a yielding the most efficient estimator of θ.

- Since x and y are independent we have

$$\begin{aligned}\mathrm{var}(\hat{\theta}_a) &= \mathrm{var}[ax] + \mathrm{var}[(1-a)y] \\ &= a^2\mathrm{var}(x) + (1-a)^2\mathrm{var}(y) \\ &= a^2\sigma^2 + (1-a)^2 \cdot 2\sigma^2 = (3a^2 - 4a + 2)\sigma^2.\end{aligned}$$

- Via calculus (the derivative is $6a - 4$ which is zero for $a = 2/3$) or completing the square ($3a^2 - 4a + 2 = 3(a - 2/3)^2 + 2/3$), we see that the minimum variance occurs for $a = 2/3$.
- Thus, $a = 2/3$ yields the most efficient estimator.

Cramèr-Rao, I [FOR FUN ONLY]

The variance of any estimator is always bounded below (since it is by definition nonnegative). So it is quite reasonable to ask whether, for a fixed estimation problem, there might be an optimal unbiased estimator: namely, one of *minimal* variance.

- This question turns out to be quite subtle, because we are not guaranteed that such an estimator necessarily exists.
- For example, it could be the case that the possible variances of unbiased estimators form an open interval of the form (a, ∞) for some $a \geq 0$.
- Then there would be estimators whose variances approach the value a arbitrarily closely, but there is none that actually achieves the lower bound value a.

Cramèr-Rao, II [FOR FUN ONLY]

There is a lower bound on the possible values for the variance of an unbiased estimator:

Theorem (Cramèr-Rao Inequality)

Suppose that $p_X(x;\theta)$ is a probability density function that is differentiable in θ. Also suppose that the support of p, the set of values of x where $p_X(x;\theta) \neq 0$, does not depend on the parameter θ. If $x_1, x_2, \ldots, x_n$ is a random sample drawn from X, $\hat{\theta} = f(x_1, \ldots, f_n)$ is an unbiased estimator of θ, and $\ell = \ln[p_X(x;\theta)]$ denotes the log-pdf of the distribution, then $\mathrm{var}(\hat{\theta}) \geq 1/I(\theta)$ where $I(\theta) = n \cdot E[(\partial\ell/\partial\theta)^2]$.

In the event that p_X is twice-differentiable in θ, it can be shown that $I(\theta)$ can also be calculated as $I(\theta) = -n \cdot E[\partial^2\ell/\partial\theta^2]$.

Cramèr-Rao, III [FOR FUN ONLY]

A few remarks:

- The proof of the Cramèr-Rao inequality is rather technical (although not conceptually difficult), so we will omit the precise details.
- In practice, it is not always so easy to evaluate the lower bound in the Cramèr-Rao inequality.
- Furthermore, there does not always exist an unbiased estimator that actually achieves the Cramèr-Rao bound.
- However, if we are able to find an unbiased estimator whose variance does achieve the Cramèr-Rao bound, then the inequality guarantees that this estimator is the most efficient possible.

Cramèr-Rao, IV [FOR FUN ONLY]

Example: Suppose that a coin with unknown probability θ of landing heads is flipped n times, yielding results $x_1, x_2, \ldots, x_n$ (where we interpret heads as 1 and tails as 0). Let $\hat{\theta} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$.

1. Show that $\hat{\theta}$ is an unbiased estimator of θ.
2. Find the variance of $\hat{\theta}$.
3. Show that $\hat{\theta}$ has the minimum variance of all possible unbiased estimators of θ.

Cramèr-Rao, IV [FOR FUN ONLY]

Example: Suppose that a coin with unknown probability θ of landing heads is flipped n times, yielding results $x_1, x_2, \ldots, x_n$ (where we interpret heads as 1 and tails as 0). Let $\hat{\theta} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$.

1. Show that $\hat{\theta}$ is an unbiased estimator of θ.
2. Find the variance of $\hat{\theta}$.
3. Show that $\hat{\theta}$ has the minimum variance of all possible unbiased estimators of θ.

- We first need to compute the expected value and variance of this estimator. Then we need to evaluate the lower bound in the Cramèr-Rao inequality.
- The claim is that the given estimator actually achieves this lower bound.

Cramèr-Rao, V [FOR FUN ONLY]

Example: Suppose that a coin with unknown probability θ of landing heads is flipped n times, yielding results $x_1, x_2, \ldots, x_n$ (where we interpret heads as 1 and tails as 0). Let $\hat{\theta} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$.

1. Show that $\hat{\theta}$ is an unbiased estimator of θ.

Cramèr-Rao, V [FOR FUN ONLY]

Example: Suppose that a coin with unknown probability θ of landing heads is flipped n times, yielding results $x_1, x_2, \ldots, x_n$ (where we interpret heads as 1 and tails as 0). Let $\hat{\theta} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$.

1. Show that $\hat{\theta}$ is an unbiased estimator of θ.

- Since $x_1 + x_2 + \cdots + x_n$ is binomially distributed with parameters n and θ, its expected value is $n\theta$.
- Then $E(\hat{\theta}) = \frac{1}{n} \cdot n\theta = \theta$, so $\hat{\theta}$ is unbiased.

2. Find the variance of $\hat{\theta}$.

Cramèr-Rao, V [FOR FUN ONLY]

Example: Suppose that a coin with unknown probability θ of landing heads is flipped n times, yielding results $x_1, x_2, \ldots, x_n$ (where we interpret heads as 1 and tails as 0). Let $\hat{\theta} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$.

1. Show that $\hat{\theta}$ is an unbiased estimator of θ.

- Since $x_1 + x_2 + \cdots + x_n$ is binomially distributed with parameters n and θ, its expected value is $n\theta$.
- Then $E(\hat{\theta}) = \frac{1}{n} \cdot n\theta = \theta$, so $\hat{\theta}$ is unbiased.

2. Find the variance of $\hat{\theta}$.

- Since $x_1 + x_2 + \cdots + x_n$ is binomially distributed with parameters n and θ, its variance is $n\theta(1 - \theta)$.
- Then the variance of $\hat{\theta} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$ is $\mathrm{var}(\hat{\theta}) = \frac{1}{n^2} \cdot n\theta(1 - \theta) = \frac{\theta(1-\theta)}{n}$.

Cramèr-Rao, VI [FOR FUN ONLY]

Example: Suppose that a coin with unknown probability θ of landing heads is flipped n times, yielding results $x_1, x_2, \ldots, x_n$ (where we interpret heads as 1 and tails as 0). Let $\hat{\theta} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$.

3. Show that $\hat{\theta}$ has the minimum variance of all possible unbiased estimators of θ.

Cramèr-Rao, VI [FOR FUN ONLY]

Example: Suppose that a coin with unknown probability θ of landing heads is flipped n times, yielding results $x_1, x_2, \dots, x_n$ (where we interpret heads as 1 and tails as 0). Let $\hat{\theta} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$.

3. Show that $\hat{\theta}$ has the minimum variance of all possible unbiased estimators of θ.

- We compute the Cramèr-Rao bound: if $\ell = \ln[p_X(x;\theta)]$ is the log-pdf of the distribution, then $\mathrm{var}(\hat{\theta}) \geq 1/I(\theta)$ where $I(\theta) = n \cdot E[(\partial\ell/\partial\theta)^2]$.
- Here, the likelihood function can be written as $L(x;\theta) = \theta^x(1-\theta)^{1-x}$ (it is θ if $x = 1$ and $1-\theta$ if $x = 0$), so that $\ell = x\ln\theta + (1-x)\ln(1-\theta)$.
- Differentiating twice yields $\dfrac{\partial^2\ell}{\partial\theta^2} = -\dfrac{x}{\theta^2} + \dfrac{1-x}{(1-\theta)^2}$.

Cramèr-Rao, VII [FOR FUN ONLY]

Example: Suppose that a coin with unknown probability θ of landing heads is flipped n times, yielding results $x_1, x_2, \ldots, x_n$ (where we interpret heads as 1 and tails as 0). Let $\hat{\theta} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$.

3. Show that $\hat{\theta}$ has the minimum variance of all possible unbiased estimators of θ.

- So, since $E(x) = \theta$, the expected value is

$$
\begin{aligned}
E[\frac{\partial^2 \ell}{\partial \theta^2}] &= \frac{E(x)}{\theta^2} + \frac{E(1-x)}{(1-\theta)^2} \\
&= -\frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2} \\
&= -\frac{1}{\theta(1-\theta)}.
\end{aligned}
$$

Cramèr-Rao, VIII [FOR FUN ONLY]

Example: Suppose that a coin with unknown probability θ of landing heads is flipped n times, yielding results $x_1, x_2, \ldots, x_n$ (where we interpret heads as 1 and tails as 0). Let $\hat{\theta} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$.

3. Show that $\hat{\theta}$ has the minimum variance of all possible unbiased estimators of θ.

- Using the calculation on the previous slide shows that the Cramèr-Rao bound is $\mathrm{var}(\hat{\theta}) \geq \frac{\theta(1-\theta)}{n}$.
- But we calculated before that for our unbiased estimator $\hat{\theta} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$ we do in fact have $\mathrm{var}(\hat{\theta}) = \frac{\theta(1-\theta)}{n}$.
- Therefore, our estimator $\hat{\theta}$ achieves the Cramèr-Rao bound, so it has the minimum variance of all possible unbiased estimators of θ, as claimed.

So, in fact, the obvious estimator is actually the best possible!

Cramèr-Rao, IX [FOR FUN ONLY]

Example: Show that the maximum-likelihoood estimator $\hat{\theta}_\mu = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$ is the most efficient possible unbiased estimator of the mean of a normal distribution with unknown mean $\theta = \mu$ and known standard deviation σ.

Cramèr-Rao, IX [FOR FUN ONLY]

Example: Show that the maximum-likelihoood estimator $\hat{\theta}_\mu = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$ is the most efficient possible unbiased estimator of the mean of a normal distribution with unknown mean $\theta = \mu$ and known standard deviation σ.

- We will show that this estimator achieves the Cramèr-Rao bound.
- For this, we first compute the log-pdf $\ell = \ln(\sqrt{2\pi}) - \frac{1}{2}\ln(\sigma) - \dfrac{(x-\theta)^2}{2\sigma^2}$.
- Differentiating yields $\dfrac{\partial \ell}{\partial \theta} = -\dfrac{x-\theta}{\sigma^2}$ and then $\dfrac{\partial^2 \ell}{\partial \theta^2} = \dfrac{1}{\sigma^2}$. Since this is constant we simply see $E[\dfrac{\partial^2 \ell}{\partial \theta^2}] = \dfrac{1}{\sigma^2}$.
- Then, the Cramèr-Rao bound dictates that $\mathrm{var}(\hat{\theta}) \geq \sigma^2/n$ for any estimator $\hat{\theta}$.

Cramèr-Rao, X [FOR FUN ONLY]

Example: Show that the maximum-likelihoood estimator $\hat{\theta}_\mu = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$ is the most efficient possible unbiased estimator of the mean of a normal distribution with unknown mean $\theta = \mu$ and known standard deviation σ.

- The Cramèr-Rao bound says $\mathrm{var}(\hat{\theta}) \geq \sigma^2/n$.
- For our estimator, since the x_i are all independent and normally distributed with mean θ and standard deviation σ, we have $\mathrm{var}(\hat{\theta}_\mu) = \frac{1}{n^2}[\mathrm{var}(x_1) + \cdots + \mathrm{var}(x_n)] = \frac{1}{n^2} \cdot n\sigma^2 = \sigma^2/n$.
- Thus, the variance of our estimator $\hat{\theta}_\mu$ achieves the Cramèr-Rao bound, meaning that it is the most efficient unbiased estimator possible.

Cramèr-Rao, XI [FOR FUN ONLY]

Example: Show that the maximum-likelihoood estimator $\hat{\theta}_{\mathrm{var}} = \frac{1}{n}\left[(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_n - \mu)^2\right]$ is the most efficient possible estimator of the variance $\theta = \sigma^2$ of a normal distribution with known mean μ.

Cramèr-Rao, XI [FOR FUN ONLY]

Example: Show that the maximum-likelihoood estimator $\hat{\theta}_{\text{var}} = \frac{1}{n}\left[(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_n - \mu)^2\right]$ is the most efficient possible estimator of the variance $\theta = \sigma^2$ of a normal distribution with known mean μ.

- This one is rather lengthy, so we will just sketch the actual calculations (the full details are in the notes).
- As before, we show $\hat{\theta}_{\text{var}}$ achieves the Cramèr-Rao bound.
- Using expected value properties, we can eventually show that $I(\theta) = -E[\partial^2 \ell / \partial \theta^2] = n/(2\theta^2)$, and so the Cramèr-Rao bound says $\text{var}(\hat{\theta}) \geq 2\theta^2/n$.
- For our estimator, by some calculations with the normal distribution and variance properties, we can find $\text{var}(\hat{\theta}_{\text{var}}) = \frac{1}{n^2} \cdot n \cdot 2\theta^2 = \frac{2\theta^2}{n}$.
- Thus, $\hat{\theta}_{\text{var}}$ achieves the Cramèr-Rao bound, as claimed.

Cramèr-Rao, XII [FOR FUN ONLY]

Example: Compare the variance of the unbiased estimator $\hat{\theta}_1 = \frac{n+1}{n} \max(x_1, x_2, \ldots, x_n)$ from sampling the uniform distribution on $[0, \theta]$ to the Cramèr-Rao bound.

Cramèr-Rao, XII [FOR FUN ONLY]

Example: Compare the variance of the unbiased estimator $\hat{\theta}_1 = \dfrac{n+1}{n}\max(x_1, x_2, \ldots, x_n)$ from sampling the uniform distribution on $[0, \theta]$ to the Cramèr-Rao bound.

- We already computed $\mathrm{var}(\hat{\theta}_1) = \frac{1}{n(n+2)}\theta^2$ earlier.
- To compute the bound from Cramèr-Rao, we have $L(\theta) = (1/\theta)^n$ hence $\ell = \ln(L) = -n\ln(\theta)$.
- Then $\partial\ell/\partial\theta = -n/\theta$ so $\partial^2\ell/\partial\theta^2 = n/\theta^2$.
- Since this is constant, $I(\theta) = -n \cdot E[\partial^2\ell/\partial\theta^2] = n^2/\theta^2$.
- Thus, the Cramèr-Rao bound is $\mathrm{var}(\hat{\theta}) \geq \theta^2/n^2$.
- But now notice that $\mathrm{var}(\hat{\theta}_1) < \theta^2/n^2$: this means $\hat{\theta}_1$ actually has a smaller variance than the Cramèr-Rao minimum!

Cramèr-Rao, XIII [FOR FUN ONLY]

Example: Compare the variance of the unbiased estimator $\hat{\theta}_1 = \frac{n+1}{n}\max(x_1, x_2, \ldots, x_n)$ from sampling the uniform distribution on $[0, \theta]$ to the Cramèr-Rao bound.

Cramèr-Rao, XIII [FOR FUN ONLY]

Example: Compare the variance of the unbiased estimator $\hat{\theta}_1 = \dfrac{n+1}{n}\max(x_1, x_2, \ldots, x_n)$ from sampling the uniform distribution on $[0, \theta]$ to the Cramèr-Rao bound.

- We have $\mathrm{var}(\hat{\theta}_1) = \frac{1}{n(n+2)}\theta^2$, but the Cramèr-Rao bound says we should have $\mathrm{var}(\hat{\theta}) \geq \theta^2/n^2$.
- This is not a contradiction, because in fact one of the hypotheses of the Cramèr-Rao theorem is violated.
- Specifically, one hypothesis says that the set of values of x where $p_X(x; \theta) \neq 0$ does not depend on the parameter θ.
- Here, $p_X(x; \theta) \neq 0$ for $x \in [0, \theta]$, and this range clearly does depend on θ. So the theorem does not apply, and we do not have a contradiction.

Cramèr-Rao, XIV [FOR FUN ONLY]

In more general situations, the Cramèr-Rao theorem can be very hard to apply (since we have to compute the expected value of a second derivative), and there is no guarantee that there actually exists an estimator realizing the bound.

- Think of it more as a broad result telling us about the best possible "minimum variance" we might hope to find for our estimation problem.
- If we can find an unbiased estimator whose variance is close to the Cramèr-Rao minimum, we should view this estimator as "good". If the variance is far away from the Cramèr-Rao minimum, that suggests our estimator is probably not so good.

Summary

We discussed biased and unbiased estimators.

We discussed efficiency of estimators.

We stated the Cramèr-Rao bound and used it to show that some of our unbiased estimators were the most efficient possible ones.

Next lecture: Interval estimation and confidence intervals