Foundations of Statistical Inference

Julien Berestycki

Department of Statistics
University of Oxford

MT 2016

Lecture 5 - Completenes, Lehmann-Scheffé Theorem, Method of Moments

Previously

Definition (Complete Sufficient Statistics)

Let $T(X_1, \ldots, X_n)$ be a sufficient statistic for θ. The statistic T is complete if, whenever $h(T)$ is a function of T for which $\mathbb{E}[h(T)] = 0$ for all θ, then $h(T) \equiv 0$ almost everywhere.

Previously

Definition (Complete Sufficient Statistics)

Let $T(X_1, \ldots, X_n)$ be a sufficient statistic for θ. The statistic T is complete if, whenever $h(T)$ is a function of T for which $\mathbb{E}[h(T)] = 0$ for all θ, then $h(T) \equiv 0$ almost everywhere.

recipe for finding a MVUE

Let T be a complete sufficient statistic and θ an unbiased estimator. Then

- either $\hat{\theta}_T = E[\hat{\theta}|T]$
- or find g such that $\hat{\theta}_T = g(T)$ is unbiased

produce the unique MVUE.

Complete Sufficiency in EFs

Lemma (6)

If the rv X has a distribution belonging to a k-parameter exponential family, then under the usual regularity conditions, the statistic

$$\left(\sum_{i=1}^{n} B_1(X_i), \sum_{i=1}^{n} B_2(X_i), \ldots, \sum_{i=1}^{n} B_k(X_i)\right)$$

which we already know is minimal sufficient, is complete.

Complete Sufficiency in EFs

Lemma (6)

If the rv X has a distribution belonging to a k-parameter exponential family, then under the usual regularity conditions, the statistic

$$\left(\sum_{i=1}^{n} B_1(X_i), \sum_{i=1}^{n} B_2(X_i), \ldots, \sum_{i=1}^{n} B_k(X_i)\right)$$

which we already know is minimal sufficient, is complete.

Comment We showed before that MLEs for exponential families were functions of this same statistic.

Complete Sufficiency in EFs

Lemma (6)

If the rv X has a distribution belonging to a k-parameter exponential family, then under the usual regularity conditions, the statistic

$$\left(\sum_{i=1}^{n} B_1(X_i), \sum_{i=1}^{n} B_2(X_i), \ldots, \sum_{i=1}^{n} B_k(X_i)\right)$$

which we already know is minimal sufficient, is complete.

Comment We showed before that MLEs for exponential families were functions of this same statistic.

Therefore, for a member of an exponential family if the MLE is unbiased (note : not all MLEs are unbiased), then by Lemma 5, the MLE will be MVUE. If there is an unbiased estimator that attains the CRLB then, by Corollary 3, the MLE will attain the CRLB.

Summary

CRLB (How good can you get?)

1. $\hat{\theta}$ unbiased estimator of θ. $V(\hat{\theta}) \geq I_\theta^{-1}$
2. $\exists \hat{\theta}$ attains CRLB $\Leftrightarrow \frac{\partial \ell}{\partial \theta} = I_\theta(\hat{\theta} - \theta) \Rightarrow X \in$ expo. family.
3. If $\tilde{\theta}$ attains CRLB and $\hat{\theta}$ MLE, then $\tilde{\theta} = \hat{\theta}$

Summary

CRLB (How good can you get?)

1. $\hat{\theta}$ unbiased estimator of θ. $V(\hat{\theta}) \geq I_\theta^{-1}$
2. $\exists \hat{\theta}$ attains CRLB $\Leftrightarrow \frac{\partial \ell}{\partial \theta} = I_\theta(\hat{\theta} - \theta) \Rightarrow X \in$ expo. family.
3. If $\tilde{\theta}$ attains CRLB and $\hat{\theta}$ MLE, then $\tilde{\theta} = \hat{\theta}$

Rao-Blackwell (How to be better?)

1. $\hat{\theta}_T = E[\hat{\theta}|T]$ where T sufficient for θ and $\hat{\theta}$ unbiased, then $\hat{\theta}_T = f(T)$ and $V(\hat{\theta}_T) \leq V(\hat{\theta})$.
2. If $\exists$ MVUE and T is min. suff. then $\exists h(T)$ which is MVUE

Summary

CRLB (How good can you get?)

1. $\hat{\theta}$ unbiased estimator of θ. $V(\hat{\theta}) \geq I_\theta^{-1}$
2. $\exists \hat{\theta}$ attains CRLB $\Leftrightarrow \frac{\partial \ell}{\partial \theta} = I_\theta(\hat{\theta} - \theta) \Rightarrow X \in$ expo. family.
3. If $\tilde{\theta}$ attains CRLB and $\hat{\theta}$ MLE, then $\tilde{\theta} = \hat{\theta}$

Rao-Blackwell (How to be better?)

1. $\hat{\theta}_T = E[\hat{\theta}|T]$ where T sufficient for θ and $\hat{\theta}$ unbiased, then $\hat{\theta}_T = f(T)$ and $V(\hat{\theta}_T) \leq V(\hat{\theta})$.
2. If $\exists$ MVUE and T is min. suff. then $\exists h(T)$ which is MVUE

Lehmann-Scheffé (How to be the best?)

1. T is complete min. suff. and $h(T)$ unbiased, then it is MVUE
2. If T is complete suff. and $\hat{\theta}$ is unbiased, then $\hat{\theta}_T$ is MVUE.

Quiz time!

- If a MVUE $\hat{\theta}$ exists, does it always attains the CRLB?

Quiz time!

- If a MVUE $\hat{\theta}$ exists, does it always attains the CRLB? See next example
- In what kind of situation is it the case that $\nexists$ an MVUE?

Quiz time!

- If a MVUE $\hat{\theta}$ exists, does it always attains the CRLB? See next example
- In what kind of situation is it the case that $\nexists$ an MVUE? Make example with $\hat{\theta}_!$ and $\hat{\theta}_{'}2$
- If the MLE is an unbiased estimator and solves $\partial \ell / \partial \theta = 0$, is it always the MVUE (under reg conditions)? Does it attains CRLB?

Quiz time!

- If a MVUE $\hat{\theta}$ exists, does it always attains the CRLB? See next example
- In what kind of situation is it the case that $\nexists$ an MVUE? Make example with $\hat{\theta}_{!}$ and $\hat{\theta}_{,}2$
- If the MLE is an unbiased estimator and solves $\partial \ell / \partial \theta = 0$, is it always the MVUE (under reg conditions)? Does it attains CRLB?
- Can a MLE have a lower variance than CRLB? Higher variance?

Quiz time!

- If a MVUE $\hat{\theta}$ exists, does it always attains the CRLB? See next example
- In what kind of situation is it the case that $\nexists$ an MVUE? Make example with $\hat{\theta}_!$ and $\hat{\theta}{`}2$
- If the MLE is an unbiased estimator and solves $\partial \ell / \partial \theta = 0$, is it always the MVUE (under reg conditions)? Does it attains CRLB?
- Can a MLE have a lower variance than CRLB? Higher variance? See next example
- In Rao-Blackwell, if $\hat{\theta}_T = \hat{\theta}$ for any sufficient T, is it the MVUE?

Quiz time!

- If a MVUE $\hat{\theta}$ exists, does it always attains the CRLB? See next example
- In what kind of situation is it the case that $\nexists$ an MVUE? Make example with $\hat{\theta}_{!}$ and $\hat{\theta}_{\cdot}2$
- If the MLE is an unbiased estimator and solves $\partial \ell / \partial \theta = 0$, is it always the MVUE (under reg conditions)? Does it attains CRLB?
- Can a MLE have a lower variance than CRLB? Higher variance? See next example
- In Rao-Blackwell, if $\hat{\theta}_T = \hat{\theta}$ for any sufficient T, is it the MVUE? Try to compare to Lehmann-Scheffé

Example 13

Let $X_1, \ldots, X_n$ be a random sample from $\mathrm{N}(\mu, \sigma^2)$.

Example 13

Let $X_1, \ldots, X_n$ be a random sample from $\mathrm{N}(\mu, \sigma^2)$.

Then we know the MLEs are $\widehat{\mu} = \bar{X}$, $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2$.

Example 13

Let $X_1, \ldots, X_n$ be a random sample from $\mathrm{N}(\mu, \sigma^2)$.

Then we know the MLEs are $\widehat{\mu} = \bar{X}$, $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2$.

$\widehat{\mu}$ is unbiased, but $\widehat{\sigma^2}$ is biased.

Example 13

Let $X_1, \dots, X_n$ be a random sample from $N(\mu, \sigma^2)$.

Then we know the MLEs are $\widehat{\mu} = \bar{X}$, $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2$.

$\widehat{\mu}$ is unbiased, but $\widehat{\sigma^2}$ is biased.

Exercise The minimal sufficient complete statistic is $\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$.

Example 13

Let $X_1, \ldots, X_n$ be a random sample from $\mathrm{N}(\mu, \sigma^2)$.

Then we know the MLEs are $\widehat{\mu} = \bar{X}$, $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2$.

$\widehat{\mu}$ is unbiased, but $\widehat{\sigma^2}$ is biased.

Exercise The minimal sufficient complete statistic is $\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$.

So $\widehat{\mu}$ is MVUE and attains the CRLB with variance σ^2/n.

Example 13

Let $X_1, \ldots, X_n$ be a random sample from $\mathrm{N}(\mu, \sigma^2)$.

Then we know the MLEs are $\widehat{\mu} = \bar{X}$, $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2$.

$\widehat{\mu}$ is unbiased, but $\widehat{\sigma^2}$ is biased.

Exercise The minimal sufficient complete statistic is $\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$.

So $\widehat{\mu}$ is MVUE and attains the CRLB with variance σ^2/n.

The sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ is unbiased and is a function of the minimal sufficient complete statistic so is MVUE with variance $2\sigma^4/(n-1)$ which is larger than the CRLB of $2\sigma^4/n$.

Method of moments

Generate estimators by equating observed statistics with their expected values

$$\mathbb{E}\{t(X_1, \ldots, X_n)\} = t(x_1, \ldots, x_n).$$

Method of moments

Generate estimators by equating observed statistics with their expected values

$$\mathbb{E}\{t(X_1, \ldots, X_n)\} = t(x_1, \ldots, x_n).$$

Comment Simple, easy to use. Often have good properties, but not guaranteed. Can provide good starting point for an iterative method for finding MLEs.

Method of moments

Generate estimators by equating observed statistics with their expected values

$$\mathbb{E}\{t(X_1, \ldots, X_n)\} = t(x_1, \ldots, x_n).$$

Comment Simple, easy to use. Often have good properties, but not guaranteed. Can provide good starting point for an iterative method for finding MLEs.

Example 14 Uniform iid sample $X = (X_1, X_2, ..., X_n)$ on $(0, \theta)$.

$$L(\theta; x) = f(x; \theta) = \theta^{-n},\ 0 < x_1, \ldots, x_n < \theta.$$

Method of moments

Generate estimators by equating observed statistics with their expected values

$$\mathbb{E}\{t(X_1,\dots,X_n)\} = t(x_1,\dots,x_n).$$

Comment Simple, easy to use. Often have good properties, but not guaranteed. Can provide good starting point for an iterative method for finding MLEs.

Example 14 Uniform iid sample $X = (X_1, X_2, ..., X_n)$ on $(0,\theta)$.

$$L(\theta;x) = f(x;\theta) = \theta^{-n},\ 0 < x_1,\dots,x_n < \theta.$$

Let $X_{(n)} = \max_i X_i$. In this case the moment relation

$$\mathbb{E}\left[X_{(n)}\right] = x_{(n)}$$

leads to an unbiased sufficient statistic, $\widehat{\theta} = \frac{n+1}{n} X_{(n)}$.

Method of moments

The distribution of $X_{(n)} = \max_i X_i$ is obtained from the CDF

$$P(X_{(n)} \leq y) = \left(\frac{y}{\theta}\right)^n$$

(the probability all iid X_i fall in $(0, y)$)

Method of moments

The distribution of $X_{(n)} = \max_i X_i$ is obtained from the CDF

$$P(X_{(n)} \leq y) = \left(\frac{y}{\theta}\right)^n$$

(the probability all iid X_i fall in $(0, y)$) so

$$f_{X_{(n)}}(y;\theta) = \frac{ny^{n-1}}{\theta^n},\ 0 < y < \theta$$

Method of moments

The distribution of $X_{(n)} = \max_i X_i$ is obtained from the CDF

$$P(X_{(n)} \leq y) = \left(\frac{y}{\theta}\right)^n$$

(the probability all iid X_i fall in $(0, y)$) so

$$f_{X_{(n)}}(y; \theta) = \frac{ny^{n-1}}{\theta^n}, \; 0 < y < \theta$$

and

$$\mathbb{E}\left[X_{(n)}\right] = \frac{n}{n+1}\theta,$$

Method of moments

The distribution of $X_{(n)} = \max_i X_i$ is obtained from the CDF

$$P(X_{(n)} \le y) = \left(\frac{y}{\theta}\right)^n$$

(the probability all iid X_i fall in $(0, y)$) so

$$f_{X_{(n)}}(y; \theta) = \frac{ny^{n-1}}{\theta^n}, \; 0 < y < \theta$$

and

$$\mathbb{E}\left[X_{(n)}\right] = \frac{n}{n+1}\theta, \text{ so } \widehat{\theta} = \frac{n+1}{n} X_{(n)}$$

is unbiased.

Now check $\widehat{\theta}$ is sufficient.

Now check $\widehat{\theta}$ is sufficient.

The distribution of $X \mid X_{(n)} = y$ is

$$f(x|X_{(n)} = y; \theta) = \frac{f(x; \theta)}{f_{X_{(n)}}(y; \theta)}$$

Now check $\widehat{\theta}$ is sufficient.

The distribution of $X \mid X_{(n)} = y$ is

$$f(x|X_{(n)} = y; \theta) = \frac{f(x;\theta)}{f_{X_{(n)}}(y;\theta)} = \frac{1}{ny^{n-1}}$$

which does not depend on θ.

Now check $\widehat{\theta}$ is sufficient.

The distribution of $X \mid X_{(n)} = y$ is

$$f(x|X_{(n)} = y; \theta) = \frac{f(x; \theta)}{f_{X_{(n)}}(y; \theta)} = \frac{1}{ny^{n-1}}$$

which does not depend on θ.

$\widehat{\theta}$ is minimal sufficient since

$$\frac{L(\theta; x)}{L(\theta; y)} = \frac{\theta^{-n} I[x_{(n)} < \theta]}{\theta^{-n} I[y_{(n)} < \theta]}$$

does not depend on θ if $x_{(n)} = y_{(n)}$ i.e. $\widehat{\theta}$ forms a Lehmann-Scheffé partition.

Finally, we can show that $X_{(n)}$ is complete.

Finally, we can show that $X_{(n)}$ is complete.

If $\mathbb{E}[h(X_{(n)}] = 0$ for all $\theta > 0$ then

$$\int_0^\theta h(y) \frac{ny^{n-1}}{\theta^n} dy = 0$$

Finally, we can show that $X_{(n)}$ is complete.

If $\mathbb{E}[h(X_{(n)})] = 0$ for all $\theta > 0$ then

$$\int_0^\theta h(y)\frac{ny^{n-1}}{\theta^n}dy = 0$$

for all $\theta > 0$ and hence

$$\int_0^\theta h(y)y^{n-1}dy = 0$$

so that

$$\int_0^\theta h^-(y)y^{n-1}dy = \int_0^\theta h^+(y)y^{n-1}dy$$

for all θ (where $h^=$ and h^+ are negative/positive parts).

Finally, we can show that $X_{(n)}$ is complete.

If $\mathbb{E}[h(X_{(n)})] = 0$ for all $\theta > 0$ then

$$\int_0^\theta h(y) \frac{ny^{n-1}}{\theta^n} dy = 0$$

for all $\theta > 0$ and hence

$$\int_0^\theta h(y) y^{n-1} dy = 0$$

so that

$$\int_0^\theta h^-(y) y^{n-1} dy = \int_0^\theta h^+(y) y^{n-1} dy$$

for all θ (where $h^=$ and h^+ are negative/positive parts).We conclude that $h^- \equiv h^+$.

Since $\widehat{\theta}$ is not only sufficient but also complete and unbiased, it is MVUE.

Finally, we can show that $X_{(n)}$ is complete.

If $\mathbb{E}[h(X_{(n)})] = 0$ for all $\theta > 0$ then

$$\int_0^\theta h(y)\frac{ny^{n-1}}{\theta^n}dy = 0$$

for all $\theta > 0$ and hence

$$\int_0^\theta h(y)y^{n-1}dy = 0$$

so that

$$\int_0^\theta h^-(y)y^{n-1}dy = \int_0^\theta h^+(y)y^{n-1}dy$$

for all θ (where $h^=$ and h^+ are negative/positive parts).We conclude that $h^- \equiv h^+$.

Since $\widehat{\theta}$ is not only sufficient but also complete and unbiased, it is MVUE.

We can't use the CRLB as the problem is non-regular.

Exercise $\ell(\theta; x) = -n \log(\theta)$ so

$$\mathbb{E}\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right] = n^2/\theta^2$$

and the CRLB would be $\text{Var}(\widehat{\theta}) \geq \theta^2/n^2$.

Exercise $\ell(\theta; x) = -n \log(\theta)$ so

$$\mathbb{E}\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right] = n^2/\theta^2$$

and the CRLB would be $\mathrm{Var}(\widehat{\theta}) \geq \theta^2/n^2$.

However, you can check (using $f_{X_{(n)}}(y; \theta)$ above) that

$$\mathrm{Var}(\widehat{\theta}) = \frac{\theta^2}{n(n+1)}$$

which is smaller than θ^2/n^2 for any $n \geq 1$.

Exercise $\ell(\theta; x) = -n \log(\theta)$ so

$$\mathbb{E}\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right] = n^2/\theta^2$$

and the CRLB would be $\text{Var}(\widehat{\theta}) \geq \theta^2/n^2$.

However, you can check (using $f_{X_{(n)}}(y; \theta)$ above) that

$$\text{Var}(\widehat{\theta}) = \frac{\theta^2}{n(n+1)}$$

which is smaller than θ^2/n^2 for any $n \geq 1$.

This is not a contradiction, as $f(x; \theta)$ doesn't satisfy the regularity conditions (limits of x depend on θ).

We can find the MLE in this example

$$\widetilde{\theta} = X_{(n)}$$

We can find the MLE in this example

$$\widetilde{\theta} = X_{(n)}$$

but this estimator is

1. biased and the Uniform density does not satisfy the regularity condition needed for CRLB so it does not apply.

We can find the MLE in this example

$$\widetilde{\theta} = X_{(n)}$$

but this estimator is

1. biased and the Uniform density does not satisfy the regularity condition needed for CRLB so it does not apply.
2. does not satisfy $\partial \ell / \partial \theta = 0$, so even if the CRLB did apply we cannot make the link between the MLE and the lower bound.

Exercise

Let $(X_1, \ldots, X_n)$ be i.i.d r.v.'s from $U[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$ with $\theta \in \mathbb{R}$. Show that $T = (T_1, T_2) = (\min X_i, \max X_i)$ is sufficient but **not** complete.

Exercise

Let $(X_1, \ldots, X_n)$ be i.i.d r.v.'s from $U[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$ with $\theta \in \mathbb{R}$. Show that $T = (T_1, T_2) = (\min X_i, \max X_i)$ is sufficient but **not** complete.
Hint : choose $h(t_1, t_2) = t_2 - t_1 - \frac{n-1}{n+1}$

Example 15

A sample from $N(\mu, \sigma^2)$. The moment relations

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = n\mu, \qquad \mathbb{E}\left[\sum_{i=1}^{n} X_i^2\right] = n(\mu^2 + \sigma^2)$$

lead to the estimating equations

$$\widehat{\mu} = \bar{X}, \qquad \widehat{\mu}^2 + \widehat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} X_i^2.$$

Example 15

A sample from $N(\mu, \sigma^2)$. The moment relations

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = n\mu, \qquad \mathbb{E}\left[\sum_{i=1}^{n} X_i^2\right] = n(\mu^2 + \sigma^2)$$

lead to the estimating equations

$$\widehat{\mu} = \bar{X}, \qquad \widehat{\mu}^2 + \widehat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} X_i^2.$$

We have seen that these are just the equations for the MLE's in this 2-dimensional exponential linear family, and solve to give the MLE

$$\widehat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Example 16

A sample from Pois(λ). We have two possible moment relations

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \lambda, \qquad \Rightarrow \widehat{\lambda}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \lambda, \qquad \Rightarrow \widehat{\lambda}_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Example 16

A sample from Pois(λ). We have two possible moment relations

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \lambda, \qquad \Rightarrow \widehat{\lambda}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \lambda, \qquad \Rightarrow \widehat{\lambda}_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

For a sample of size n we have
$A(\lambda) = \log\lambda$, $B(x) = \sum_{i=1}^{n} x_i$, $C(x) = -\log(\prod_{i=1}^{n} x_i!)$, $D(\lambda) = -n\lambda$

Example 16

A sample from Pois(λ). We have two possible moment relations

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \lambda, \qquad \Rightarrow \widehat{\lambda}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2\right] = \lambda, \qquad \Rightarrow \widehat{\lambda}_2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2$$

For a sample of size n we have
$A(\lambda) = \log\lambda$, $B(x) = \sum_{i=1}^n x_i$, $C(x) = -\log(\prod_{i=1}^n x_i!)$, $D(\lambda) = -n\lambda$

By Lemma 6 we have that $B(x) = \sum_{i=1}^n x_i$ is a sufficient complete statistic. Thus by the Lehmann-Scheffé Theorem we have that arithmetic mean $\widehat{\lambda}_1$ is MVUE. $\widehat{\lambda}_2$ is not a function of $B(x)$.

Example 16

A sample from Pois(λ). We have two possible moment relations

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \lambda, \qquad \Rightarrow \widehat{\lambda}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \lambda, \qquad \Rightarrow \widehat{\lambda}_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

For a sample of size n we have
$A(\lambda) = \log\lambda$, $B(x) = \sum_{i=1}^{n} x_i$, $C(x) = -\log(\prod_{i=1}^{n} x_i!)$, $D(\lambda) = -n\lambda$

By Lemma 6 we have that $B(x) = \sum_{i=1}^{n} x_i$ is a sufficient complete statistic. Thus by the Lehmann-Scheffé Theorem we have that arithmetic mean $\widehat{\lambda}_1$ is MVUE. $\widehat{\lambda}_2$ is not a function of $B(x)$.

Exercise Show that $\widehat{\lambda}_1$ attains the CRLB.

Method of moments asymptotics

Suppose the dimension of θ is $d = 1$ and $\bar{H} = n^{-1} \sum_{i=1}^{n} h(x_i)$ where $\mathbb{E}(\bar{H}) = k(\theta)$ and $\widehat{\theta}$ is the solution to $\bar{H} = k(\theta)$.

Method of moments asymptotics

Suppose the dimension of θ is $d = 1$ and $\bar{H} = n^{-1} \sum_{i=1}^{n} h(x_i)$ where $\mathbb{E}(\bar{H}) = k(\theta)$ and $\widehat{\theta}$ is the solution to $\bar{H} = k(\theta)$.

Theorem (Asymptotic normality of moment estimators)

As $n \to \infty$, $\widehat{\theta}$ is asymptotically normally distributed with mean θ and variance $n^{-1}\sigma_h^2/(k'(\theta))^2$, where

$$\sigma_h^2 = \int h(x)^2 f(x;\theta)dx - k(\theta)^2$$

Method of moments asymptotics

Suppose the dimension of θ is $d = 1$ and $\bar{H} = n^{-1}\sum_{i=1}^{n} h(x_i)$ where $\mathbb{E}(\bar{H}) = k(\theta)$ and $\widehat{\theta}$ is the solution to $\bar{H} = k(\theta)$.

Theorem (Asymptotic normality of moment estimators)

As $n \to \infty$, $\widehat{\theta}$ is asymptotically normally distributed with mean θ and variance $n^{-1}\sigma_h^2/(k'(\theta))^2$, where

$$\sigma_h^2 = \int h(x)^2 f(x;\theta)dx - k(\theta)^2$$

Proof $\mathbb{E}(\bar{H}) = k(\theta)$, so by the Central Limit Theorem,

$$\frac{(\bar{H} - k(\theta))\sqrt{n}}{\sigma_h} \xrightarrow{\mathcal{D}} N(0,1)$$

Method of moments asymptotics

Suppose the dimension of θ is $d = 1$ and $\bar{H} = n^{-1} \sum_{i=1}^{n} h(x_i)$ where $\mathbb{E}(\bar{H}) = k(\theta)$ and $\widehat{\theta}$ is the solution to $\bar{H} = k(\theta)$.

Theorem (Asymptotic normality of moment estimators)

As $n \to \infty$, $\widehat{\theta}$ is asymptotically normally distributed with mean θ and variance $n^{-1}\sigma_h^2/(k'(\theta))^2$, where

$$\sigma_h^2 = \int h(x)^2 f(x;\theta)dx - k(\theta)^2$$

Proof $\mathbb{E}(\bar{H}) = k(\theta)$, so by the Central Limit Theorem,

$$\frac{(\bar{H} - k(\theta))\sqrt{n}}{\sigma_h} \xrightarrow{\mathcal{D}} N(0,1)$$

Approximately

$$\bar{H} = k(\widehat{\theta}) = k(\theta) + (\widehat{\theta} - \theta)k'(\theta)$$

Method of moments asymptotics

Rearrange to get

$$\frac{\sqrt{n}(\widehat{\theta} - \theta)k'(\theta)}{\sigma_h} = \frac{(\bar{H} - k(\theta))\sqrt{n}}{\sigma_h} \xrightarrow{\mathcal{D}} N(0, 1)$$

Method of moments asymptotics

Rearrange to get

$$\frac{\sqrt{n}(\widehat{\theta}-\theta)k'(\theta)}{\sigma_h} = \frac{(\bar{H}-k(\theta))\sqrt{n}}{\sigma_h} \xrightarrow{\mathcal{D}} N(0,1)$$

so

$$\widehat{\theta} \approx N\left(\theta, \frac{\sigma_h^2}{n(k'(\theta))^2}\right)$$

as $n \to \infty$.

Example 17 - Exponential distribution

$$f(x;\theta) = \theta \exp(-\theta x), \ x > 0$$

Example 17 - Exponential distribution

$$f(x;\theta) = \theta \exp(-\theta x), \; x > 0$$

$$\mathbb{E}[\bar{X}] = \theta^{-1}, \; \widehat{\theta} = \bar{X}^{-1}$$

Example 17 - Exponential distribution

$$f(x;\theta) = \theta \exp(-\theta x),\ x > 0$$

$$\mathbb{E}[\bar{X}] = \theta^{-1},\ \widehat{\theta} = \bar{X}^{-1}$$

That is

$$k(\theta) = \theta^{-1},\ \bar{H} = \bar{X},\ h(X) = X,\ \mathbb{E}\big[h(X)\big] = \theta^{-1}$$

Example 17 - Exponential distribution

$$f(x;\theta) = \theta \exp(-\theta x),\ x > 0$$

$$\mathbb{E}[\bar{X}] = \theta^{-1},\ \widehat{\theta} = \bar{X}^{-1}$$

That is

$$k(\theta) = \theta^{-1},\ \bar{H} = \bar{X},\ h(X) = X,\ \mathbb{E}\big[h(X)\big] = \theta^{-1}$$

$$\text{Var}\big[h(X)\big] = \sigma_h^2 = \theta^{-2}, k'(\theta) = -\theta^{-2}$$

Now

$$\widehat{\theta} \approx N\left(\theta, \frac{\sigma_h^2}{n(k'(\theta))^2}\right)$$

Example 17 - Exponential distribution

$$f(x;\theta) = \theta \exp(-\theta x),\ x > 0$$

$$\mathbb{E}[\bar{X}] = \theta^{-1},\ \widehat{\theta} = \bar{X}^{-1}$$

That is

$$k(\theta) = \theta^{-1},\ \bar{H} = \bar{X},\ h(X) = X,\ \mathbb{E}\big[h(X)\big] = \theta^{-1}$$

$$\mathrm{Var}\big[h(X)\big] = \sigma_h^2 = \theta^{-2}, k'(\theta) = -\theta^{-2}$$

Now

$$\widehat{\theta} \approx N\left(\theta, \frac{\sigma_h^2}{n(k'(\theta))^2}\right)$$

so

$$\widehat{\theta} \approx N(\theta, n^{-1}\theta^2)$$

Exponential families - Method of moments and MLEs

MLE are moment estimates because we are solving

$$\mathbb{E}\Big[\sum_{i=1}^{n} t(X_i)\Big] = \sum_{i=1}^{n} t(x_i)$$

for $\widehat{\theta}$.

Exponential families - Method of moments and MLEs

MLE are moment estimates because we are solving

$$\mathbb{E}\Big[\sum_{i=1}^n t(X_i)\Big] = \sum_{i=1}^n t(x_i)$$

for $\widehat{\theta}$. We showed earlier that

$$\mathbb{E}\Big[t(X)\Big] = -D'(\theta)$$

so

$$\mathbb{E}\Big[\sum_{i=1}^n t(X_i)\Big] = -nD'(\theta)$$

Exponential families - Method of moments and MLEs

MLE are moment estimates because we are solving

$$\mathbb{E}\Big[\sum_{i=1}^{n} t(X_i)\Big] = \sum_{i=1}^{n} t(x_i)$$

for $\widehat{\theta}$. We showed earlier that

$$\mathbb{E}\Big[t(X)\Big] = -D'(\theta)$$

so

$$\mathbb{E}\Big[\sum_{i=1}^{n} t(X_i)\Big] = -nD'(\theta)$$

Now

$$k(\theta) = -D'(\theta),\ k'(\theta) = -D''(\theta).$$

Exponential families - Method of moments and MLEs

MLE are moment estimates because we are solving

$$\mathbb{E}\Big[\sum_{i=1}^{n} t(X_i)\Big] = \sum_{i=1}^{n} t(x_i)$$

for $\widehat{\theta}$. We showed earlier that

$$\mathbb{E}\Big[t(X)\Big] = -D'(\theta)$$

so

$$\mathbb{E}\Big[\sum_{i=1}^{n} t(X_i)\Big] = -nD'(\theta)$$

Now

$$k(\theta) = -D'(\theta),\ k'(\theta) = -D''(\theta).$$

$$\mathrm{Var}(\widehat{\theta}) \approx \frac{\mathrm{Var}(t(X))}{nD''(\theta)^2} = \frac{-D''(\theta)}{nD''(\theta)^2} = I_\theta^{-1}$$