Lecture 3 - Estimators, Minimum Variance Unbiased Estimators and the Cramér-Rao Lower Bound.
Definition 5 A point estimate for $\theta$ is a statistic of the data.

$$\hat{\theta} = \hat{\theta}(x) = t(x_1, \ldots, x_n).$$
Definition 5 A point estimate for $\theta$ is a statistic of the data.

$$\hat{\theta} = \hat{\theta}(x) = t(x_1, \ldots, x_n).$$

Definition 6 An interval estimate is a set valued function $C(X) \subseteq \Theta$ such that $\theta \in C(X)$ with a specified probability.
Estimators

Definition 5 A point estimate for $\theta$ is a statistic of the data.

$$\hat{\theta} = \hat{\theta}(x) = t(x_1, \ldots, x_n).$$

Definition 6 An interval estimate is a set valued function $C(X) \subseteq \Theta$ such that $\theta \in C(X)$ with a specified probability.

Definition 7: Maximum likelihood estimation
If $L(\theta)$ is differentiable and there is a unique maximum in the interior of $\theta \in \Theta$, then the MLE $\hat{\theta}$ is the solution of

$$\frac{\partial}{\partial \theta} L(\theta; x) = 0 \text{ or } \frac{\partial}{\partial \theta} \ell(\theta) = 0,$$
Lemma 2 : MLEs and exponential families

Consider a $k$-dimensional exponential family in canonical form

$$L(\theta; x) = \exp \left\{ \sum_{j=1}^{k} \phi_j \left( \sum_{i=1}^{n} B_j(x_i) \right) + nD(\phi) + \sum_{i=1}^{n} C(x_i) \right\}.$$
Consider a $k$-dimensional exponential family in canonical form

$$L(\theta; x) = \exp \left\{ \sum_{j=1}^{k} \phi_j \left( \sum_{i=1}^{n} B_j(x_i) \right) + nD(\phi) + \sum_{i=1}^{n} C(x_i) \right\}.$$ 

Let $T_j(X) = \sum_{i=1}^{n} B_j(X_i)$, $j = 1, \ldots, k$. 

The MLEs of $\phi_1, \ldots, \phi_k$ are the solution of

$$t_j = \mathbb{E}_X(T_j)$$

for $j = 1, \ldots, k$. [If the family is not in canonical form, there is a similar slightly more complicated matrix equation]
Lemma 2 : MLEs and exponential families

Consider a $k$-dimensional exponential family in canonical form

$$L(\theta; x) = \exp \left\{ \sum_{j=1}^{k} \phi_j \left( \sum_{i=1}^{n} B_j(x_i) \right) + nD(\phi) + \sum_{i=1}^{n} C(x_i) \right\}.$$

Let $T_j(X) = \sum_{i=1}^{n} B_j(X_i)$, $j = 1, \ldots, k$. If the realized data are $X = x$, then the statistics evaluated on the data are $T_j(x) = t_j$. 

Jonathan Marchini (University of Oxford)
Lemma 2: MLEs and exponential families

Consider a $k$-dimensional exponential family in canonical form

$$L(\theta; x) = \exp \left\{ \sum_{j=1}^{k} \phi_j \left( \sum_{i=1}^{n} B_j(x_i) \right) + nD(\phi) + \sum_{i=1}^{n} C(x_i) \right\}.$$ 

Let $T_j(X) = \sum_{i=1}^{n} B_j(X_i)$, $j = 1, \ldots, k$. If the realized data are $X = x$, then the statistics evaluated on the data are $T_j(x) = t_j$. The MLEs of $\phi_1, \ldots, \phi_k$ are the solution of

$$t_j = \mathbb{E}_X(T_j), \ j = 1, \ldots, k.$$
Lemma 2 : MLEs and exponential families

Consider a $k$-dimensional exponential family in canonical form

$$L(\theta; x) = \exp \left\{ \sum_{j=1}^{k} \phi_j \left( \sum_{i=1}^{n} B_j(x_i) \right) + nD(\phi) + \sum_{i=1}^{n} C(x_i) \right\}.$$

Let $T_j(X) = \sum_{i=1}^{n} B_j(X_i)$, $j = 1, \ldots, k$. If the realized data are $X = x$, then the statistics evaluated on the data are $T_j(x) = t_j$.

The MLEs of $\phi_1, \ldots, \phi_k$ are the solution of

$$t_j = \mathbb{E}_X(T_j), \ j = 1, \ldots, k.$$

i.e. set the expected values of the sufficient statistics equal to their realised values and solve for $\phi_j$. [If the family is not in canonical form, there is a similar slightly more complicated matrix equation]
Proof

\[ \ell = \log L = \text{const} + \sum_{j=1}^{k} \phi_j t_j + nD(\phi) \]
Proof

\[ \ell = \log L = \text{const} + \sum_{j=1}^{k} \phi_j t_j + nD(\phi) \]

\[ \Rightarrow \quad \frac{\partial}{\partial \phi_j} \ell = t_j + n \frac{\partial}{\partial \phi_j} D(\phi) \]
Proof

\[ \ell = \log L = \text{const} + \sum_{j=1}^{k} \phi_j t_j + nD(\phi) \]

\[ \Rightarrow \quad \frac{\partial}{\partial \phi_j} \ell = t_j + n \frac{\partial}{\partial \phi_j} D(\phi) \]

However, since \( \mathbb{E}_X[B_i(X)] = -\frac{\partial}{\partial \phi_i} D(\phi) \) and \( T_j(X) = \sum_{i=1}^{n} B_j(X_i) \) we know that

\[ \mathbb{E}_X[T_j] = -n \frac{\partial}{\partial \phi_j} D(\phi), \]
Proof

\[ \ell = \log L = \text{const} + \sum_{j=1}^{k} \phi_j t_j + nD(\phi) \]

\[ \Rightarrow \quad \frac{\partial}{\partial \phi_j} \ell = t_j + n \frac{\partial}{\partial \phi_j} D(\phi) \]

However, since \( \mathbb{E}_X[B_i(X)] = -\frac{\partial}{\partial \phi_i} D(\phi) \) and \( T_j(X) = \sum_{i=1}^{n} B_j(X_i) \) we know that

\[ \mathbb{E}_X[T_j] = -n \frac{\partial}{\partial \phi_j} D(\phi), \text{ so} \]

\[ \frac{\partial}{\partial \phi_j} \ell = t_j - \mathbb{E}_X(T_j) = 0 \]

is equivalent to \( t_j = \mathbb{E}_X(T_j) \).
$T_n = T(X_1, \ldots, X_n)$ is a statistic.

Definition 8 $T_n$ is **unbiased** for a function $g(\theta)$ if

$$
\mathbb{E}_X(T_n) = \int_X t_n(x)f(x; \theta)dx = g(\theta), \text{ for all } \theta \in \Theta.
$$
$T_n = T(X_1, \ldots, X_n)$ is a statistic.

**Definition 8** $T_n$ is unbiased for a function $g(\theta)$ if

$$E_X(T_n) = \int_X t_n(x)f(x; \theta)dx = g(\theta), \quad \text{for all } \theta \in \Theta.$$ 

**Definition 9** The bias of an estimator $T_n$ is $\text{bias}(T_n) = E_X[T_n - g(\theta)]$.
Bias, Variance, Mean Squared Error

\[ T_n = T(X_1, \ldots, X_n) \text{ is a statistic.} \]

**Definition 8** \( T_n \) is unbiased for a function \( g(\theta) \) if

\[
\mathbb{E}_X(T_n) = \int_X t_n(x)f(x; \theta)dx = g(\theta), \quad \text{for all } \theta \in \Theta.
\]

**Definition 9** The bias of an estimator \( T_n \) is \( \text{bias}(T_n) = \mathbb{E}_X [T_n - g(\theta)] \)

**Definition 10** \( T_n \) is a consistent estimator if

\[
\forall \epsilon > 0, \ P(|T_n - \theta| > \epsilon) \rightarrow 0 \text{ as } n \rightarrow \infty.
\]
Bias, Variance, Mean Squared Error

\( T_n = T(X_1, \ldots, X_n) \) is a statistic.

**Definition 8** \( T_n \) is unbiased for a function \( g(\theta) \) if

\[
\mathbb{E}_X(T_n) = \int \chi t_n(x)f(x; \theta)dx = g(\theta), \quad \text{for all } \theta \in \Theta.
\]

**Definition 9** The bias of an estimator \( T_n \) is \( \text{bias}(T_n) = \mathbb{E}_X [T_n - g(\theta)] \)

**Definition 10** \( T_n \) is a consistent estimator if

\[
\forall \epsilon > 0, P(|T_n - \theta| > \epsilon) \rightarrow 0 \text{ as } n \rightarrow \infty.
\]

**Definition 11** The Mean Squared Error (MSE) of \( T_n \) is

\[
\text{MSE}(T_n) = \mathbb{E}_X [T_n - g(\theta)]^2 = V_X(T_n) + [\text{bias}(T_n)]^2
\]
\( T_n = T(X_1, \ldots, X_n) \) is a statistic.

**Definition 8** \( T_n \) is unbiased for a function \( g(\theta) \) if

\[
\mathbb{E}_X(T_n) = \int X^n(x)f(x;\theta)dx = g(\theta), \text{ for all } \theta \in \Theta.
\]

**Definition 9** The bias of an estimator \( T_n \) is \( \text{bias}(T_n) = \mathbb{E}_X[T_n - g(\theta)] \)

**Definition 10** \( T_n \) is a consistent estimator if

\[
\forall \epsilon > 0, P(|T_n - \theta| > \epsilon) \to 0 \text{ as } n \to \infty.
\]

**Definition 11** The Mean Squared Error (MSE) of \( T_n \) is

\[
\text{MSE}(T_n) = \mathbb{E}_X[T_n - g(\theta)]^2 = V_X(T_n) + [\text{bias}(T_n)]^2
\]

**Example 10** \( N(\mu, \sigma^2) \). \( \hat{\mu} = \bar{X} \) and \( S^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \) are unbiased estimates of \( \mu \) and \( \sigma^2 \).
Minimum Variance Unbiased Estimators (MVUE)

If we want to find a good estimator then one obvious strategy is to try to find estimators that minimise MSE. This is often difficult.
Minimum Variance Unbiased Estimators (MVUE)

- If we want to find a good estimator then one obvious strategy is to try to find estimators that minimise MSE. This is often difficult.

- For example, if we choose the estimator \( \hat{\theta} = \theta_0 \) then this has MSE=0 when \( \theta = \theta_0 \), so no other estimator can be uniformly best unless it has zero MSE everywhere.
Minimum Variance Unbiased Estimators (MVUE)

- If we want to find a good estimator then one obvious strategy is to try to find estimators that minimise MSE. This is often difficult.
- For example, if we choose the estimator $\hat{\theta} = \theta_0$ then this has MSE=0 when $\theta = \theta_0$, so no other estimator can be uniformly best unless it has zero MSE everywhere.
- If we restrict attention to unbiased estimators then the situation becomes more tractable. In this case, MSE reduces to the variance of the estimator and we can focus on minimising the variance of estimators. That is, we search for minimum variance unbiased estimators (MVUE).
Theorem 2 : Cramér-Rao inequality (and bound).

If \( \hat{\theta} \) is an unbiased estimator of \( \theta \), then subject to certain regularity conditions on \( f(x; \theta) \), we have

\[
\text{Var}(\hat{\theta}) \geq I_\theta^{-1}.
\]

where \( I_\theta \), the Fisher information, is given by

\[
I_\theta = -\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \ell(\theta) \right]
\]
Theorem 2 : Cramér-Rao inequality (and bound).

If \( \hat{\theta} \) is an unbiased estimator of \( \theta \), then subject to certain regularity conditions on \( f(x; \theta) \), we have

\[
\text{Var}(\hat{\theta}) \geq I^{-1}_\theta.
\]

where \( I_\theta \), the Fisher information, is given by

\[
I_\theta = -\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \ell(\theta) \right].
\]

**Comment** This bound tells us the minimum possible variance. If an estimator achieves the bound then it is MVUE. There is no guarantee that the bound will be attainable. In many cases it is attainable asymptotically.
Theorem 2: Cramér-Rao inequality (and bound).

If \( \hat{\theta} \) is an unbiased estimator of \( \theta \), then subject to certain regularity conditions on \( f(x; \theta) \), we have

\[
\text{Var}(\hat{\theta}) \geq I^{-1}_\theta.
\]

where \( I_\theta \), the Fisher information, is given by

\[
I_\theta = -\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \ell(\theta) \right]
\]

**Comment** This bound tells us the minimum possible variance. If an estimator achieves the bound then it is MVUE. There is no guarantee that the bound will be attainable. In many cases it is attainable asymptotically. Intuitively, the more ‘information’ we have about \( \theta \), the larger \( I_\theta \) will be and lowest possible variance of the estimator will be smaller.
We will not be concerned with the details of the required regularity conditions.
We will not be concerned with the details of the required regularity conditions.

The main reason they are needed is to ensure that it is ok to interchange integration and differentiation during parts of the proof.
We will not be concerned with the details of the required regularity conditions.

The main reason they are needed is to ensure that it is ok to interchange integration and differentiation during parts of the proof.

One condition that is often easy to check is that the range of the rv $X$ must not depend on $\theta$. So for example, the result can not be applied when working with the uniform distribution $U[0, \theta]$ and we wish to estimate $\theta$. 
In order to prove the CRLB we will need to use a few results.

**Proposition 1**: Variance-Covariance inequality
Let $U$ and $V$ be scalar rv. Then

$$\text{cov}(U, V)^2 \leq \text{var}(U)\text{var}(V)$$

with equality if and only if $U = aV + b$ for constants and $a \neq 0$. 

Jonathan Marchini (University of Oxford)
The Fisher Information $I_\theta$, which is used in the Cramér-Rao lower bound, can be expressed in two different forms.

Lemma 3 Under regularity conditions

\[ I_\theta = -E\left[ \frac{\partial^2}{\partial \theta^2} \ell(\theta) \right] = E\left[ \left( \frac{\partial \ell}{\partial \theta} \right)^2 \right] = \text{Var}\left[ S(X; \theta) \right], \]

where the score function $s(x; \theta)$ is defined as

\[ s(x; \theta) = \frac{\partial}{\partial \theta} \ell(\theta) = f'(x; \theta) f(x; \theta). \]
The Fisher Information $I_\theta$, which is used in the Cramér-Rao lower bound, can be expressed in two different forms.

Lemma 3 Under regularity conditions

$$I_\theta = -\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \ell(\theta) \right] = \mathbb{E} \left[ \left( \frac{\partial \ell}{\partial \theta} \right)^2 \right] = \text{Var}[S(X; \theta)],$$

where the score function $s(x; \theta)$ is defined as

$$s(x; \theta) = \frac{\partial}{\partial \theta} \ell(\theta) = \frac{f'(x; \theta)}{f(x; \theta)}$$
The Fisher Information $I_\theta$, which is used in the Cramér-Rao lower bound, can be expressed in two different forms.

**Lemma 3** Under regularity conditions

$$I_\theta = -\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \ell(\theta) \right] = \mathbb{E} \left[ \left( \frac{\partial \ell}{\partial \theta} \right)^2 \right] = \text{Var}[S(X; \theta)],$$

where the **score function** $s(x; \theta)$ is defined as

$$s(x; \theta) = \frac{\partial}{\partial \theta} \ell(\theta) = \frac{f'(x; \theta)}{f(x; \theta)}$$
Lemma 3 - Proof

We need to prove $-\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \ell(\theta) \right] = \mathbb{E} \left[ \left( \frac{\partial \ell}{\partial \theta} \right)^2 \right]$. 

The second term has expectation zero because $\mathbb{E} \left[ \frac{1}{L} \frac{\partial^2 L}{\partial \theta^2} \right] = \int \frac{1}{L} \frac{\partial^2 L}{\partial \theta^2} L dx = \int \frac{\partial^2 L}{\partial \theta^2} dx = \frac{\partial^2}{\partial \theta^2} \int L dx = 0$. 

The alternative form $I_\theta = \text{Var} \left[ S(X; \theta) \right]$ follows from $\mathbb{E} \left[ \frac{\partial \ell}{\partial \theta} \right] = 0$. 

Jonathan Marchini (University of Oxford)
Lemma 3 - Proof

We need to prove \(- \mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \ell(\theta) \right] = \mathbb{E} \left[ \left( \frac{\partial \ell}{\partial \theta} \right)^2 \right] \).

\[
\frac{\partial^2 \ell}{\partial \theta^2} = \frac{\partial}{\partial \theta} \left\{ \frac{1}{L} \frac{\partial L}{\partial \theta} \right\} \quad \text{[since \( \frac{\partial \ell}{\partial \theta} = \frac{1}{L} \frac{\partial L}{\partial \theta} \)]}
\]
Lemma 3 - Proof

We need to prove \(-\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \ell(\theta) \right] = \mathbb{E} \left[ \left( \frac{\partial \ell}{\partial \theta} \right)^2 \right]. \)

\[
\frac{\partial^2 \ell}{\partial \theta^2} = \frac{\partial}{\partial \theta} \left\{ \frac{1}{L} \frac{\partial L}{\partial \theta} \right\} \quad \text{since} \quad \frac{\partial \ell}{\partial \theta} = \frac{1}{L} \frac{\partial L}{\partial \theta}
\]

\[
= - \frac{1}{L^2} \left( \frac{\partial L}{\partial \theta} \right)^2 + \frac{1}{L} \frac{\partial^2 L}{\partial \theta^2}
\]
Lemma 3 - Proof

We need to prove $- \mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \ell(\theta) \right] = \mathbb{E} \left[ \left( \frac{\partial \ell}{\partial \theta} \right)^2 \right]$. 

\[
\frac{\partial^2 \ell}{\partial \theta^2} = \frac{\partial}{\partial \theta} \left\{ \frac{1}{L} \frac{\partial L}{\partial \theta} \right\} \quad \text{[since]} \quad \frac{\partial \ell}{\partial \theta} = \frac{1}{L} \frac{\partial L}{\partial \theta} 
\]

\[
= - \frac{1}{L^2} \left( \frac{\partial L}{\partial \theta} \right)^2 + \frac{1}{L} \frac{\partial^2 L}{\partial \theta^2}
\]

\[
= - \left( \frac{\partial \ell}{\partial \theta} \right)^2 + \frac{1}{L} \left( \frac{\partial^2 L}{\partial \theta^2} \right)
\]
Lemma 3 - Proof

We need to prove \(-\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \ell(\theta) \right] = \mathbb{E} \left[ \left( \frac{\partial \ell}{\partial \theta} \right)^2 \right].\)

\[
\frac{\partial^2 \ell}{\partial \theta^2} = \frac{\partial}{\partial \theta} \left\{ \frac{1}{L} \frac{\partial L}{\partial \theta} \right\} \quad \text{[since \(\frac{\partial \ell}{\partial \theta} = \frac{1}{L} \frac{\partial L}{\partial \theta}\)]}
\]

\[
= -\frac{1}{L^2} \left( \frac{\partial L}{\partial \theta} \right)^2 + \frac{1}{L} \frac{\partial^2 L}{\partial \theta^2}
\]

\[
= - \left( \frac{\partial \ell}{\partial \theta} \right)^2 + \frac{1}{L} \left( \frac{\partial^2 L}{\partial \theta^2} \right)
\]

The second term has expectation zero because

\[
\mathbb{E} \left[ \frac{1}{L} \left( \frac{\partial^2 L}{\partial \theta^2} \right) \right] = \int \frac{1}{L} \frac{\partial^2 L}{\partial \theta^2} Ldx
\]
Lemma 3 - Proof

We need to prove $-\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \ell(\theta) \right] = \mathbb{E} \left[ \left( \frac{\partial \ell}{\partial \theta} \right)^2 \right]$.

$$\frac{\partial^2 \ell}{\partial \theta^2} = \frac{\partial}{\partial \theta} \left\{ \frac{1}{L} \frac{\partial L}{\partial \theta} \right\} \quad \text{[since $\frac{\partial \ell}{\partial \theta} = \frac{1}{L} \frac{\partial L}{\partial \theta}$]}
$$

$$= -\frac{1}{L^2} \left( \frac{\partial L}{\partial \theta} \right)^2 + \frac{1}{L} \frac{\partial^2 L}{\partial \theta^2}
$$

$$= -\left( \frac{\partial \ell}{\partial \theta} \right)^2 + \frac{1}{L} \left( \frac{\partial^2 L}{\partial \theta^2} \right)
$$

The second term has expectation zero because

$$\mathbb{E} \left[ \frac{1}{L} \left( \frac{\partial^2 L}{\partial \theta^2} \right) \right] = \int \frac{1}{L} \frac{\partial^2 L}{\partial \theta^2} L \, dx = \int \frac{\partial^2 L}{\partial \theta^2} \, dx$$
Lemma 3 - Proof

We need to prove \(-\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \ell(\theta) \right] = \mathbb{E} \left[ \left( \frac{\partial \ell}{\partial \theta} \right)^2 \right] \).

\[
\frac{\partial^2 \ell}{\partial \theta^2} = \frac{\partial}{\partial \theta} \left\{ \frac{1}{L} \frac{\partial L}{\partial \theta} \right\} \quad \text{[since]} \quad \frac{\partial \ell}{\partial \theta} = \frac{1}{L} \frac{\partial L}{\partial \theta}
\]

\[
= -\frac{1}{L^2} \left( \frac{\partial L}{\partial \theta} \right)^2 + \frac{1}{L} \frac{\partial^2 L}{\partial \theta^2}
\]

\[
= -\left( \frac{\partial \ell}{\partial \theta} \right)^2 + \frac{1}{L} \left( \frac{\partial^2 L}{\partial \theta^2} \right)
\]

The second term has expectation zero because

\[
\mathbb{E} \left[ \frac{1}{L} \left( \frac{\partial^2 L}{\partial \theta^2} \right) \right] = \int \frac{1}{L} \frac{\partial^2 L}{\partial \theta^2} L \, dx = \int \frac{\partial^2 L}{\partial \theta^2} \, dx = \frac{\partial^2}{\partial \theta^2} \int L \, dx = 0
\]
Lemma 3 - Proof

We need to prove \(-\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \ell(\theta) \right] = \mathbb{E} \left[ \left( \frac{\partial \ell}{\partial \theta} \right)^2 \right]. \)

\[
\frac{\partial^2 \ell}{\partial \theta^2} = \frac{\partial}{\partial \theta} \left\{ \frac{1}{L} \frac{\partial L}{\partial \theta} \right\} \quad \text{since} \quad \frac{\partial \ell}{\partial \theta} = \frac{1}{L} \frac{\partial L}{\partial \theta}
\]

\[
= -\frac{1}{L^2} \left( \frac{\partial L}{\partial \theta} \right)^2 + \frac{1}{L} \frac{\partial^2 L}{\partial \theta^2}
\]

\[
= -\left( \frac{\partial \ell}{\partial \theta} \right)^2 + \frac{1}{L} \left( \frac{\partial^2 L}{\partial \theta^2} \right)
\]

The second term has expectation zero because

\[
\mathbb{E} \left[ \frac{1}{L} \left( \frac{\partial^2 L}{\partial \theta^2} \right) \right] = \int \frac{1}{L} \frac{\partial^2 L}{\partial \theta^2} L \, dx = \int \frac{\partial^2 L}{\partial \theta^2} \, dx = \frac{\partial^2}{\partial \theta^2} \int L \, dx = 0
\]

The alternative form \(l_\theta = \text{Var}[S(X; \theta)]\) follows from \(\mathbb{E} \left[ \frac{\partial \ell}{\partial \theta} \right] = 0. \)
Proof of the CRLB

We consider only unbiased estimators, so we have

\[ \mathbb{E}(\hat{\theta}) = \int_{x} \hat{\theta}(x)L(\theta; x)dx = \theta \]
Proof of the CRLB

We consider only unbiased estimators, so we have

\[ \mathbb{E}(\hat{\theta}) = \int_{x} \hat{\theta}(x) L(\theta; x) \, dx = \theta \]

Differentiate both sides w.r.t. \( \theta \)

\[ \int_{x} \hat{\theta} \frac{\partial L}{\partial \theta} \, dx = 1 \]
Proof of the CRLB

We consider only unbiased estimators, so we have

\[ \mathbb{E}(\hat{\theta}) = \int_{\chi} \hat{\theta}(x) L(\theta; x) dx = \theta \]

Differentiate both sides w.r.t. \( \theta \)

\[ \int_{\chi} \hat{\theta} \frac{\partial L}{\partial \theta} dx = 1 \]

Now

\[ \frac{\partial L}{\partial \theta} = L \frac{\partial \ell}{\partial \theta} \]
Proof of the CRLB

We consider only unbiased estimators, so we have

$$\mathbb{E}(\hat{\theta}) = \int_{\chi} \hat{\theta}(x)L(\theta; x)dx = \theta$$

Differentiate both sides w.r.t. $\theta$

$$\int_{\chi} \hat{\theta} \frac{\partial L}{\partial \theta} dx = 1$$

Now

$$\frac{\partial L}{\partial \theta} = L \frac{\partial \ell}{\partial \theta}$$

so

$$1 = \int_{\chi} \hat{\theta} \frac{\partial \ell}{\partial \theta} L dx = \mathbb{E} \left[ \hat{\theta} \frac{\partial \ell}{\partial \theta} \right]$$
Proof of the CRLB

Now we use the inequality that for two random variables $U, V$
\[
\text{Cov}[U, V]^2 \leq \text{Var}[U]\text{Var}[V]
\]
with $U = \hat{\theta}$, $V = \frac{\partial \ell}{\partial \theta}$.
Proof of the CRLB

Now we use the inequality that for two random variables $U, V$

$$\text{Cov}[U, V]^2 \leq \text{Var}[U] \text{Var}[V]$$

with $U = \hat{\theta}$, $V = \frac{\partial \ell}{\partial \theta}$. We know $\text{Var}[\frac{\partial \ell}{\partial \theta}] = I_\theta$. Must show $\text{Cov}[U, V] = 1$. 
Proof of the CRLB

Now we use the inequality that for two random variables $U, V$

$$\text{Cov}[U, V]^2 \leq \text{Var}[U] \text{Var}[V]$$

with $U = \hat{\theta}$, $V = \frac{\partial \ell}{\partial \theta}$. We know $\text{Var}\left[\frac{\partial \ell}{\partial \theta}\right] = I_\theta$. Must show $\text{Cov}[U, V] = 1$.

$$\text{Cov}[U, V] = \mathbb{E}[UV] - \mathbb{E}[U]\mathbb{E}[V],$$
Proof of the CRLB

Now we use the inequality that for two random variables $U, V$

$$\text{Cov}[U, V]^2 \leq \text{Var}[U]\text{Var}[V]$$

with $U = \hat{\theta}$, $V = \frac{\partial \ell}{\partial \theta}$. We know $\text{Var}[\frac{\partial \ell}{\partial \theta}] = I_\theta$. Must show $\text{Cov}[U, V] = 1$.

$$\text{Cov}[U, V] = \mathbb{E}[UV] - \mathbb{E}[U]\mathbb{E}[V], \quad \mathbb{E}[U] = \theta,$$
Proof of the CRLB

Now we use the inequality that for two random variables $U, V$

$$\text{Cov}[U, V]^2 \leq \text{Var}[U] \text{Var}[V]$$

with $U = \hat{\theta}$, $V = \frac{\partial \ell}{\partial \theta}$. We know $\text{Var}[\frac{\partial \ell}{\partial \theta}] = I_\theta$. Must show $\text{Cov}[U, V] = 1$.

$$\text{Cov}[U, V] = \mathbb{E}[UV] - \mathbb{E}[U] \mathbb{E}[V], \quad \mathbb{E}[U] = \theta, \quad \mathbb{E}\left[\hat{\theta} \frac{\partial \ell}{\partial \theta}\right] = 1$$
Proof of the CRLB

Now we use the inequality that for two random variables $U$, $V$

$$\text{Cov}[U, V]^2 \leq \text{Var}[U]\text{Var}[V]$$

with $U = \hat{\theta}$, $V = \frac{\partial \ell}{\partial \theta}$. We know $\text{Var}[\frac{\partial \ell}{\partial \theta}] = I_\theta$. Must show $\text{Cov}[U, V] = 1$.

$$\text{Cov}[U, V] = \mathbb{E}[UV] - \mathbb{E}[U]\mathbb{E}[V], \quad \mathbb{E}[U] = \theta, \quad \mathbb{E}\left[\hat{\theta} \frac{\partial \ell}{\partial \theta}\right] = 1$$

$$\mathbb{E}[V] = \int_{\chi} \frac{\partial \ell}{\partial \theta} L \, dx = \int_{\chi} \frac{\partial L}{\partial \theta} \, dx$$
Proof of the CRLB

Now we use the inequality that for two random variables $U$, $V$

$$\text{Cov}[U, V]^2 \leq \text{Var}[U] \text{Var}[V]$$

with $U = \hat{\theta}$, $V = \frac{\partial \ell}{\partial \theta}$. We know $\text{Var}[\frac{\partial \ell}{\partial \theta}] = I_\theta$. Must show $\text{Cov}[U, V] = 1$.

$$\text{Cov}[U, V] = \mathbb{E}[UV] - \mathbb{E}[U] \mathbb{E}[V], \quad \mathbb{E}[U] = \theta, \quad \mathbb{E} \left[\hat{\theta} \frac{\partial \ell}{\partial \theta}\right] = 1$$

$$\mathbb{E}[V] = \int_{\chi} \frac{\partial \ell}{\partial \theta} Ldx = \int_{\chi} \frac{\partial L}{\partial \theta} dx = \frac{\partial}{\partial \theta} \left[\int_{\chi} Ldx\right]$$
Proof of the CRLB

Now we use the inequality that for two random variables $U, V$

$$\text{Cov}[U, V]^2 \leq \text{Var}[U]\text{Var}[V]$$

with $U = \hat{\theta}$, $V = \frac{\partial \ell}{\partial \theta}$. We know $\text{Var}[\frac{\partial \ell}{\partial \theta}] = I_\theta$. Must show $\text{Cov}[U, V] = 1$.

$$\text{Cov}[U, V] = \mathbb{E}[UV] - \mathbb{E}[U]\mathbb{E}[V], \quad \mathbb{E}[U] = \theta, \quad \mathbb{E}\left[\hat{\theta}\frac{\partial \ell}{\partial \theta}\right] = 1$$

$$\mathbb{E}[V] = \int_\chi \frac{\partial \ell}{\partial \theta} Ldx = \int_\chi \frac{\partial L}{\partial \theta} dx = \frac{\partial}{\partial \theta} \left[ \int_\chi Ldx \right] = \frac{\partial}{\partial \theta} [1] = 0$$
Proof of the CRLB

Now we use the inequality that for two random variables \( U, V \)
\[
\text{Cov}[U, V]^2 \leq \text{Var}[U] \text{Var}[V]
\]
with \( U = \hat{\theta}, \ V = \frac{\partial \ell}{\partial \theta} \). We know \( \text{Var} \left[ \frac{\partial \ell}{\partial \theta} \right] = I_\theta \). Must show \( \text{Cov}[U, V] = 1 \).

\[
\text{Cov}[U, V] = \mathbb{E}[UV] - \mathbb{E}[U] \mathbb{E}[V], \quad \mathbb{E}[U] = \theta, \quad \mathbb{E} \left[ \hat{\theta} \frac{\partial \ell}{\partial \theta} \right] = 1
\]

\[
\mathbb{E}[V] = \int_{\chi} \frac{\partial \ell}{\partial \theta} L \, dx = \int_{\chi} \frac{\partial L}{\partial \theta} \, dx = \frac{\partial}{\partial \theta} \left[ \int_{\chi} L \, dx \right] = \frac{\partial}{\partial \theta} [1] = 0
\]

\[
\text{Var}[\hat{\theta}] = \text{Var}[U] \geq \frac{\text{Cov}[U, V]^2}{\text{Var}[V]} = \frac{1^2}{I_\theta} = I_\theta^{-1}
\]
Information in a sample of size $n$.

If we have $n$ iid observations then

$$f(x; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$
Information in a sample of size $n$.

If we have $n$ iid observations then

$$f(x; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

and the Fisher information is

$$i_n(\theta) = -\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \ell(\theta) \right] = - \int \sum_{i=1}^{n} \frac{\partial^2}{\partial \theta^2} \log f(x_i; \theta)f(x; \theta) \, dx = ni_1(\theta).$$
Information in a sample of size $n$.

If we have $n$ iid observations then

$$f(x; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

and the Fisher information is

$$i_n(\theta) = -\mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} \ell(\theta) \right] = -\int \sum_{i=1}^{n} \frac{\partial^2}{\partial \theta^2} \log f(x_i; \theta) f(x; \theta) \, dx = n i_1(\theta).$$

That is, $i_1(\theta)$ is calculated from the density as

$$i_1(\theta) = -\int \frac{\partial^2}{\partial \theta^2} \log f(x; \theta) f(x; \theta) \, dx$$