Keener, Theoretical Statistics
Chapter 4: Unbiased Estimation

Junhyeok Kil

17 Sep 2025

Covariance of Random Variables

Definition

The covariance between two random variables X and Y is:

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)] = EXY - (EX)(EY)$$

Covariance of Random Variables

Definition

The covariance between two random variables X and Y is:

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)] = EXY - (EX)(EY)$$

Special Case

If either mean, EX or EY equals 0:

$$\text{Cov}(X, Y) = EXY$$

The Covariance Inequality

Let $\sigma_X = \sqrt{\text{Var}(X)}$ and $\sigma_Y = \sqrt{\text{Var}(Y)}$.

The expectation of any squared random variable is non-negative:

$$E\left[\left(\frac{X - EX}{\sigma_X} \pm \frac{Y - EY}{\sigma_Y}\right)^2\right] \geq 0$$

The Covariance Inequality

Let $\sigma_X = \sqrt{\text{Var}(X)}$ and $\sigma_Y = \sqrt{\text{Var}(Y)}$.
The expectation of any squared random variable is non-negative:

$$E\left[\left(\frac{X - EX}{\sigma_X} \pm \frac{Y - EY}{\sigma_Y}\right)^2\right] \geq 0$$

Expanding this expression leads to:

$$E\left[\frac{(X - EX)^2}{\sigma_X^2} \pm 2\frac{(X - EX)(Y - EY)}{\sigma_X \sigma_Y} + \frac{(Y - EY)^2}{\sigma_Y^2}\right] \geq 0$$

$$\frac{\text{Var}(X)}{\text{Var}(X)} \pm 2\frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} + \frac{\text{Var}(Y)}{\text{Var}(Y)} \geq 0$$

$$2 \pm 2\frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \geq 0$$

$$1 \pm \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \geq 0$$

The Covariance Inequality

Expanding this expression leads to:

$$E\left[\frac{(X-EX)^2}{\sigma_X^2} \pm 2\frac{(X-EX)(Y-EY)}{\sigma_X\sigma_Y} + \frac{(Y-EY)^2}{\sigma_Y^2}\right] \geq 0$$

$$\frac{\text{Var}(X)}{\text{Var}(X)} \pm 2\frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y} + \frac{\text{Var}(Y)}{\text{Var}(Y)} \geq 0$$

$$2 \pm 2\frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y} \geq 0$$

$$1 \pm \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y} \geq 0$$

The Resulting Bound

This directly gives us the covariance inequality:

$$|\text{Cov}(X,Y)| \leq \sigma_X\sigma_Y \quad \text{or} \quad \text{Cov}^2(X,Y) \leq \text{Var}(X)\text{Var}(Y) \qquad (1)$$

A General Variance Bound

- Let δ be an unbiased estimator of $g(\theta)$.
- Let ψ be an arbitrary random variable.

A General Variance Bound

- Let δ be an unbiased estimator of $g(\theta)$.
- Let ψ be an arbitrary random variable.

Using the inequality $\mathrm{Cov}^2(\delta, \psi) \leq \mathrm{Var}(\delta)\mathrm{Var}(\psi)$, we can rearrange it to form a lower bound on the variance of our estimator δ.

The bound is:

$$\mathrm{Var}_\theta(\delta) \geq \frac{\mathrm{Cov}_\theta^2(\delta, \psi)}{\mathrm{Var}_\theta(\psi)} \tag{2}$$

The Challenge with the Bound

So we have:

$$\mathrm{Var}_\theta(\delta) \geq \frac{\mathrm{Cov}_\theta^2(\delta, \psi)}{\mathrm{Var}_\theta(\psi)}$$

The problem is that the RHS still depends on the estimator δ itself through the covariance term.

To find a truly useful bound, we need to choose the random variable ψ cleverly. The goal is to choose a ψ such that $\mathrm{Cov}_\theta(\delta, \psi)$ is the same for **all** estimators δ that are unbiased for $g(\theta)$.

The Setup

Assumptions

- Let $P = \{P_\theta : \theta \in \Omega\}$ be a dominated family of distributions with densities p_θ, where $\Omega \subset \mathbb{R}$.
- Let $\delta(X)$ be any unbiased estimator of $g(\theta)$, meaning $E_\theta \delta(X) = g(\theta)$.

The Setup

Assumptions

- Let $P = \{P_\theta : \theta \in \Omega\}$ be a dominated family of distributions with densities p_θ, where $\Omega \subset \mathbb{R}$.
- Let $\delta(X)$ be any unbiased estimator of $g(\theta)$, meaning $E_\theta \delta(X) = g(\theta)$.

Consider the quantity $E_{\theta+\Delta}[\delta] - E_\theta[\delta]$ which has the same value for **any** unbiased estimator δ:

$$E_{\theta+\Delta}[\delta] - E_\theta[\delta] = g(\theta + \Delta) - g(\theta)$$

Our goal is to express this difference as a covariance under P_θ.

The Likelihood Ratio

To write $E_{\theta+\Delta}[\delta]$ as an expectation under P_θ, we introduce a likelihood ratio.

Assumption

Suppose that the "support" of $p_{\theta+\Delta}$ is contained within the support of p_θ. Formally:

$$\text{If } p_\theta(x) = 0, \text{ then } p_{\theta+\Delta}(x) = 0.$$

The Likelihood Ratio

To write $E_{\theta+\Delta}[\delta]$ as an expectation under P_θ, we introduce a likelihood ratio.

Assumption

Suppose that the "support" of $p_{\theta+\Delta}$ is contained within the support of p_θ. Formally:

$$\text{If } p_\theta(x) = 0, \text{ then } p_{\theta+\Delta}(x) = 0.$$

Definition: Likelihood Ratio

We define the likelihood ratio $L(x)$ as:

$$L(x) = \begin{cases} \frac{p_{\theta+\Delta}(x)}{p_\theta(x)} & \text{if } p_\theta(x) > 0 \\ 1 & \text{otherwise} \end{cases}$$

Changing the Expectation

By the definition, $L(x)p_\theta(x) = p_{\theta+\Delta}(x)$ almost everywhere. So we can rewrite expectations.

General Formula

For any function h that is integrable under $P_{\theta+\Delta}$:

$$\begin{aligned} E_{\theta+\Delta}[h(X)] &= \int h(x)p_{\theta+\Delta}(x)\,d\mu \\ &= \int h(x)L(x)p_\theta(x)\,d\mu = E_\theta[L(X)h(X)] \end{aligned}$$

Changing the Expectation

By the definition, $L(x)p_\theta(x) = p_{\theta+\Delta}(x)$ almost everywhere. So we can rewrite expectations.

General Formula

For any function h that is integrable under $P_{\theta+\Delta}$:

$$E_{\theta+\Delta}[h(X)] = \int h(x)p_{\theta+\Delta}(x)\,d\mu$$
$$= \int h(x)L(x)p_\theta(x)\,d\mu = E_\theta[L(X)h(X)]$$

Applying the Formula

- Taking $h = 1$, we get $E_{\theta+\Delta}[1] = 1 = E_\theta[L(X)]$.
- Taking $h = \delta$, we get $E_{\theta+\Delta}[\delta(X)] = E_\theta[L(X)\delta(X)]$.

Finding the Covariance

Let's define our specific random variable ψ.

Definition of ψ

Let $\psi(X) = L(X) - 1$. Since $E_\theta[L(X)] = 1$,

$$E_\theta[\psi(X)] = E_\theta[L(X) - 1] = 1 - 1 = 0.$$

Finding the Covariance

Let's define our specific random variable ψ.

Definition of ψ

Let $\psi(X) = L(X) - 1$. Since $E_\theta[L(X)] = 1$,

$E_\theta[\psi(X)] = E_\theta[L(X) - 1] = 1 - 1 = 0.$

Since $E_\theta[\psi] = 0$,

$$\begin{aligned}
\mathrm{Cov}_\theta(\delta, \psi) &= E_\theta(\delta\psi) - (E_\theta\delta)(E_\theta\psi) \\
&= E_\theta(\delta\psi) \\
&= E_\theta[\delta(L-1)] = E_\theta[L\delta] - E_\theta[\delta] \\
&= E_{\theta+\Delta}[\delta] - E_\theta[\delta] \\
&= g(\theta + \Delta) - g(\theta)
\end{aligned}$$

This is now independent of the specific estimator δ.

The Hammersley-Chapman-Robbins Inequality

We now substitute our results into the general variance bound:

$$\mathsf{Var}_\theta(\delta) \geq \frac{\mathsf{Cov}_\theta^2(\delta, \psi)}{\mathsf{Var}_\theta(\psi)}$$

The Hammersley-Chapman-Robbins Inequality

We now substitute our results into the general variance bound:

$$\mathrm{Var}_\theta(\delta) \geq \frac{\mathrm{Cov}_\theta^2(\delta, \psi)}{\mathrm{Var}_\theta(\psi)}$$

- The numerator is $[g(\theta + \Delta) - g(\theta)]^2$.
- The denominator is $\mathrm{Var}_\theta(\psi) = E_\theta(\psi^2) - (E_\theta\psi)^2 = E_\theta(\psi^2)$ since $E_\theta(\psi) = 0$.

The Hammersley-Chapman-Robbins Inequality

We now substitute our results into the general variance bound:

$$\mathrm{Var}_\theta(\delta) \geq \frac{\mathrm{Cov}_\theta^2(\delta, \psi)}{\mathrm{Var}_\theta(\psi)}$$

- The numerator is $[g(\theta + \Delta) - g(\theta)]^2$.
- The denominator is $\mathrm{Var}_\theta(\psi) = E_\theta(\psi^2) - (E_\theta\psi)^2 = E_\theta(\psi^2)$ since $E_\theta(\psi) = 0$.

The HCR Inequality

Putting it all together, we get the bound:

$$\mathrm{Var}_\theta(\delta) \geq \frac{[g(\theta + \Delta) - g(\theta)]^2}{E_\theta\left[\left(\frac{p_{\theta+\Delta}(X)}{p_\theta(X)} - 1\right)^2\right]} \tag{3}$$

This is the Hammersley-Chapman-Robbins inequality.

The Limit of the HCR Bound

We can rewrite the Hammersley-Chapman-Robbins inequality's lower bound as:

$$\frac{\left(\frac{g(\theta+\Delta)-g(\theta)}{\Delta}\right)^2}{E_\theta\left[\left(\frac{p_{\theta+\Delta}(X)-p_\theta(X)}{\Delta}\right)^2 \frac{1}{p_\theta(X)^2}\right]}$$

The Limit of the HCR Bound

We can rewrite the Hammersley-Chapman-Robbins inequality's lower bound as:

$$\frac{\left(\frac{g(\theta+\Delta)-g(\theta)}{\Delta}\right)^2}{E_\theta\left[\left(\frac{p_{\theta+\Delta}(X)-p_\theta(X)}{\Delta}\right)^2 \frac{1}{p_\theta(X)^2}\right]}$$

Taking the Limit

Under suitable regularity conditions, the dominated convergence theorem can be used to show that as $\Delta \to 0$, this bound converges to:

$$\frac{[g'(\theta)]^2}{E_\theta\left[\left(\frac{\partial p_\theta(X)/\partial\theta}{p_\theta(X)}\right)^2\right]}$$

Fisher Information

The Score Function

The term inside the square is the partial derivative of the log-likelihood:

$$\frac{\partial \log p_\theta(X)}{\partial \theta} = \frac{1}{p_\theta(X)} \frac{\partial p_\theta(X)}{\partial \theta}$$

This is often called the **score function**.

Fisher Information

The Score Function

The term inside the square is the partial derivative of the log-likelihood:

$$\frac{\partial \log p_\theta(X)}{\partial \theta} = \frac{1}{p_\theta(X)} \frac{\partial p_\theta(X)}{\partial \theta}$$

This is often called the **score function**.

Fisher Information, $I(\theta)$

The expected squared score is called the **Fisher Information**, denoted $I(\theta)$:

$$I(\theta) = E_\theta \left[\left(\frac{\partial \log p_\theta(X)}{\partial \theta} \right)^2 \right] \tag{4}$$

An Alternative Formula for $I(\theta)$

- With enough regularity, we can interchange integration and differentiation.
- We start with the identity $\int p_\theta(x) d\mu(x) = 1$.

An Alternative Formula for $I(\theta)$

- With enough regularity, we can interchange integration and differentiation.
- We start with the identity $\int p_\theta(x) d\mu(x) = 1$.

Differentiating Under the Integral

$$
\begin{aligned}
0 = \frac{\partial}{\partial \theta} 1 = \frac{\partial}{\partial \theta} \int p_\theta(x) d\mu(x) &= \int \frac{\partial}{\partial \theta} p_\theta(x) d\mu(x) \\
&= \int \frac{\partial \log p_\theta(x)}{\partial \theta} p_\theta(x) d\mu(x) \\
&= E_\theta \left[\frac{\partial \log p_\theta(X)}{\partial \theta} \right]
\end{aligned}
$$

The expected value of the score function is zero.

An Alternative Formula for $I(\theta)$

Differentiating Under the Integral

$$
\begin{aligned}
0 = \frac{\partial}{\partial\theta}1 = \frac{\partial}{\partial\theta}\int p_\theta(x)d\mu(x) &= \int \frac{\partial}{\partial\theta}p_\theta(x)d\mu(x) \\
&= \int \frac{\partial \log p_\theta(x)}{\partial\theta}p_\theta(x)d\mu(x) \\
&= E_\theta\left[\frac{\partial \log p_\theta(X)}{\partial\theta}\right]
\end{aligned}
$$

The expected value of the score function is zero.

$I(\theta)$ as a Variance

Since $E[Y] = 0 \implies \mathrm{Var}(Y) = E[Y^2]$, it follows that:

$$
I(\theta) = \mathrm{Var}_\theta\left(\frac{\partial \log p_\theta(X)}{\partial\theta}\right) \tag{5}
$$

Another Formula for Calculation

If we can pass a second derivative inside the integral $\int p_\theta d\mu = 1$:

$$\int \frac{\partial^2 p_\theta(x)}{\partial \theta^2} d\mu(x) = E_\theta \left[\frac{\partial^2 p_\theta(X)/\partial \theta^2}{p_\theta(X)} \right] = 0$$

Another Formula for Calculation

If we can pass a second derivative inside the integral $\int p_\theta d\mu = 1$:

$$\int \frac{\partial^2 p_\theta(x)}{\partial \theta^2} d\mu(x) = E_\theta \left[\frac{\partial^2 p_\theta(X)/\partial \theta^2}{p_\theta(X)} \right] = 0$$

Now, consider the second derivative of the log-likelihood (using the quotient rule):

$$\begin{aligned}
\frac{\partial^2 \log p_\theta(X)}{\partial \theta^2} &= \frac{\partial}{\partial \theta} \left(\frac{\partial p_\theta(X)/\partial \theta}{p_\theta(X)} \right) = \frac{\partial}{\partial \theta} \frac{p'_\theta(X)}{p_\theta(X)} \\
&= \frac{p''_\theta(X) p_\theta(X) - (p'_\theta(X))^2}{(p_\theta(X))^2} \\
&= \frac{\partial^2 p_\theta(X)/\partial \theta^2}{p_\theta(X)} - \left(\frac{\partial p_\theta(X)/\partial \theta}{p_\theta(X)} \right)^2 \\
&= \frac{\partial^2 p_\theta(X)/\partial \theta^2}{p_\theta(X)} - \left(\frac{\partial \log p_\theta(X)}{\partial \theta} \right)^2
\end{aligned}$$

Another Formula for Calculation

The Second Derivative Formula

Taking the expectation and rearranging:

$$E_\theta\left[\frac{\partial^2 \log p_\theta(X)}{\partial \theta^2}\right] = E_\theta\left[\frac{\partial^2 p_\theta(X)/\partial \theta^2}{p_\theta(X)}\right] - E_\theta\left[\left(\frac{\partial \log p_\theta(X)}{\partial \theta}\right)^2\right]$$

$$E_\theta\left[\frac{\partial^2 \log p_\theta(X)}{\partial \theta^2}\right] = 0 - I(\theta)$$

$$I(\theta) = -E_\theta\left[\frac{\partial^2 \log p_\theta(X)}{\partial \theta^2}\right] \tag{6}$$

This formula is often more convenient for calculations.

Deriving the Covariance Term

- We can derive a lower bound based on Fisher information in a similar way to the HCR inequality.
- Let δ be an estimator with mean $g(\theta) = E_\theta(\delta)$.
- We make the clever choice for ψ to be the score function:

$$\psi = \frac{\partial \log p_\theta(X)}{\partial \theta}$$

Deriving the Covariance Term

- We can derive a lower bound based on Fisher information in a similar way to the HCR inequality.
- Let δ be an estimator with mean $g(\theta) = E_\theta(\delta)$.
- We make the clever choice for ψ to be the score function:

$$\psi = \frac{\partial \log p_\theta(X)}{\partial \theta}$$

Differentiating Under the Integral Sign

With sufficient regularity, we can find $g'(\theta)$:

$$\begin{aligned} g'(\theta) = \frac{\partial}{\partial \theta} \int \delta(x) p_\theta(x)\, d\mu &= \int \delta(x) \frac{\partial}{\partial \theta} p_\theta(x)\, d\mu \\ &= \int \delta(x) \left(\frac{\partial \log p_\theta(x)}{\partial \theta} \right) p_\theta(x)\, d\mu \\ &= E_\theta[\delta(X)\psi(X)] \end{aligned}$$

Deriving the Covariance Term

Differentiating Under the Integral Sign

With sufficient regularity, we can find $g'(\theta)$:

$$\begin{aligned} g'(\theta) = \frac{\partial}{\partial\theta}\int \delta(x)p_\theta(x)\,d\mu &= \int \delta(x)\frac{\partial}{\partial\theta}p_\theta(x)\,d\mu \\ &= \int \delta(x)\left(\frac{\partial \log p_\theta(x)}{\partial\theta}\right)p_\theta(x)\,d\mu \\ &= E_\theta[\delta(X)\psi(X)] \end{aligned}$$

The Covariance

Since we know $E_\theta(\psi) = 0$, the expression $E_\theta(\delta\psi)$ is precisely the covariance:

$$\mathrm{Cov}_\theta(\delta,\psi) = E_\theta[\delta\psi] - (E_\theta\delta)(E_\theta\psi) = g'(\theta)$$

Theorem 4.9: The Cramér-Rao Bound

Theorem Statement

Let $P = \{P_\theta : \theta \in \Omega\}$ be a dominated family with Ω an open set in $\mathbb{R}$ and densities p_θ differentiable with respect to θ. If for all $\theta \in \Omega$:

- $E_\theta(\psi) = 0$
- $E_\theta(\delta^2) < \infty$
- The regularity condition $g'(\theta) = E_\theta(\delta\psi)$ holds,

then the variance of the estimator δ is bounded by:

$$\mathrm{Var}_\theta(\delta) \geq \frac{[g'(\theta)]^2}{I(\theta)}$$

This result is called the **Cramér-Rao**, or **information**, bound.

The Trouble with the Regularity Condition

A Potential Problem

The regularity condition from the theorem,

$$g'(\theta) = E_\theta(\delta\psi)$$

is troublesome because it involves the estimator δ itself.

The Trouble with the Regularity Condition

A Potential Problem

The regularity condition from the theorem,

$$g'(\theta) = E_\theta(\delta\psi)$$

is troublesome because it involves the estimator δ itself.

Implication

This leaves open the possibility that some estimators, which do not satisfy this condition, may have a variance *below* the stated bound.

The Trouble with the Regularity Condition

A Potential Problem

The regularity condition from the theorem,

$$g'(\theta) = E_\theta(\delta\psi)$$

is troublesome because it involves the estimator δ itself.

Implication

This leaves open the possibility that some estimators, which do not satisfy this condition, may have a variance *below* the stated bound.

How This is Addressed

- **Weak Conditions:** Woodroofe and Simons (1983) show the bound holds for any estimator δ for *almost all* θ.
- **Restrictive Conditions:** Other authors impose stricter conditions on the model P, but show the bound holds for *all* δ at *all* $\theta \in \Omega$.

Fisher Information Under Reparameterization

- Let h be a one-to-one function from Ξ to Ω, with $\theta = h(\xi)$.
- The new family is $\tilde{P} = \{Q_\xi : \xi \in \Xi\}$ with densities $q_\xi = p_{h(\xi)}$.

Fisher Information Under Reparameterization

- Let h be a one-to-one function from Ξ to Ω, with $\theta = h(\xi)$.
- The new family is $\tilde{P} = \{Q_\xi : \xi \in \Xi\}$ with densities $q_\xi = p_{h(\xi)}$.

Applying the Chain Rule

Fisher information $\tilde{I}$ for the new family $\tilde{P}$ is given by:

$$\begin{aligned}
\tilde{I}(\xi) &= E_\xi \left[\left(\frac{\partial \log q_\xi(X)}{\partial \xi} \right)^2 \right] \\
&= E_\xi \left[\left(\frac{\partial \log p_{h(\xi)}(X)}{\partial \xi} \right)^2 \right] \\
&= E_\theta \left[\left(\frac{\partial \log p_\theta(X)}{\partial \theta} \cdot \frac{d\theta}{d\xi} \right)^2 \right] \\
&= [h'(\xi)]^2 E_\theta \left[\left(\frac{\partial \log p_\theta(X)}{\partial \theta} \right)^2 \right] \\
&= [h'(\xi)]^2 I(\theta)
\end{aligned}$$

Example 4.10: Exponential Families

Canonical Form

Let P be a one-parameter exponential family. The densities p_η are given by:

$$p_\eta(x) = \exp\left[\eta\, T(x) - A(\eta)\right] h(x)$$

Differentiating the log-likelihood gives the score function:

$$\frac{\partial \log p_\eta(X)}{\partial \eta} = T(X) - A'(\eta)$$

Example 4.10: Exponential Families

Canonical Form

Let P be a one-parameter exponential family. The densities p_η are given by:

$$p_\eta(x) = \exp\left[\eta\, T(x) - A(\eta)\right] h(x)$$

Differentiating the log-likelihood gives the score function:

$$\frac{\partial \log p_\eta(X)}{\partial \eta} = T(X) - A'(\eta)$$

Calculating $I(\eta)$

Using $I(\eta) = \mathrm{Var}_\eta(\text{score})$, we find:

$$I(\eta) = \mathrm{Var}_\eta\left(T - A'(\eta)\right) = \mathrm{Var}_\eta(T) = A''(\eta)$$

This also follows from the second derivative formula, as $I(\eta) = -E(\frac{\partial^2 \log p_\eta}{\partial \eta^2}) = -\frac{\partial^2 \log p_\eta}{\partial \eta^2} = A''(\eta)$.

Exponential Families: Reparameterization

Parameterizing by the Mean $\mu = E_\eta[T]$

Recall the Fisher information under the natural parameterization:

$$I(\eta) = A''(\eta) = \text{Var}_\eta(T).$$

Since $\mu = A'(\eta)$, we have

$$\frac{d\mu}{d\eta} = A''(\eta), \quad \Rightarrow \quad \frac{d\eta}{d\mu} = \frac{1}{A''(\eta)}.$$

By the reparameterization rule:

$$I(\mu) = \left(\frac{d\eta}{d\mu}\right)^2 I(\eta).$$

Substituting:

$$I(\mu) = \left(\frac{1}{A''(\eta)}\right)^2 \cdot A''(\eta) = \frac{1}{A''(\eta)} = \frac{1}{\text{Var}_\mu(T)}.$$

Exponential Families: Reparameterization

Sharpness of the Cramér–Rao Bound

Recall the general Cramér–Rao bound:

$$\mathrm{Var}_\mu(\delta) \;\geq\; \frac{[g'(\mu)]^2}{I(\mu)}.$$

If δ is an **unbiased estimator of** μ itself, then

$$E_\mu[\delta] = \mu \quad \Rightarrow \quad g(\mu) = \mu \;\text{ and }\; g'(\mu) = 1.$$

Therefore,

$$\mathrm{Var}_\mu(\delta) \;\geq\; \frac{1}{I(\mu)}.$$

$$\mathrm{Var}_\mu(\delta) \;\geq\; \mathrm{Var}_\mu(T).$$

Since the sufficient statistic T is the UMVUE of μ, its variance exactly equals the lower bound. Therefore, the Cramér–Rao bound is **sharp** in this case.

Example 4.11: Location Families

Definition

- Let ϵ be a random variable with a known density f.
- A location family is the family of distributions for $X = \theta + \epsilon$, where $\theta \in \mathbb{R}$ is the location parameter.

Example 4.11: Location Families

Definition

- Let ϵ be a random variable with a known density f.
- A location family is the family of distributions for $X = \theta + \epsilon$, where $\theta \in \mathbb{R}$ is the location parameter.

Density of a Location Family

By a change of variables, the density $p_\theta(x)$ of X is found to be:

$$p_\theta(x) = f(x - \theta)$$

Fisher Information for Location Families

We calculate $I(\theta)$ using the definition:

$$
\begin{aligned}
I(\theta) &= E_\theta \left[\left(\frac{\partial \log f(X-\theta)}{\partial \theta} \right)^2 \right] \\
&= E_\theta \left[\left(\frac{-f'(X-\theta)}{f(X-\theta)} \right)^2 \right] \quad \text{(by the chain rule)} \\
&= E \left[\left(\frac{f'(\epsilon)}{f(\epsilon)} \right)^2 \right] \quad \text{(since } X-\theta = \epsilon\text{)} \\
&= \int \frac{[f'(x)]^2}{f(x)} dx
\end{aligned}
$$

Fisher Information for Location Families

We calculate $I(\theta)$ using the definition:

$$
\begin{aligned}
I(\theta) &= E_\theta \left[\left(\frac{\partial \log f(X-\theta)}{\partial \theta} \right)^2 \right] \\
&= E_\theta \left[\left(\frac{-f'(X-\theta)}{f(X-\theta)} \right)^2 \right] \quad \text{(by the chain rule)} \\
&= E \left[\left(\frac{f'(\epsilon)}{f(\epsilon)} \right)^2 \right] \quad \text{(since } X-\theta = \epsilon\text{)} \\
&= \int \frac{[f'(x)]^2}{f(x)} dx
\end{aligned}
$$

Key Result

The final expression does not depend on θ. For location families, Fisher information $I(\theta)$ is a constant.

Additivity of Fisher Information

Setup

- Suppose X and Y are **independent** observations.
- X has density $p_\theta(x)$ with information $I_X(\theta)$.
- Y has density $q_\theta(y)$ with information $I_Y(\theta)$.

Additivity of Fisher Information

Setup

- Suppose X and Y are **independent** observations.
- X has density $p_\theta(x)$ with information $I_X(\theta)$.
- Y has density $q_\theta(y)$ with information $I_Y(\theta)$.

Derivation

The joint density is $p_\theta(x)q_\theta(y)$. The joint log-likelihood is:

$$\log\left(p_\theta(X)q_\theta(Y)\right) = \log p_\theta(X) + \log q_\theta(Y)$$

The score function for the joint observation is the sum of the individual scores.

Additivity of Fisher Information

The total information $I_{X,Y}(\theta)$ is the variance of the joint score:

$$I_{X,Y}(\theta) = \mathrm{Var}_\theta \left(\frac{\partial \log p_\theta(X)}{\partial \theta} + \frac{\partial \log q_\theta(Y)}{\partial \theta} \right)$$

Additivity of Fisher Information

The total information $I_{X,Y}(\theta)$ is the variance of the joint score:

$$I_{X,Y}(\theta) = \mathrm{Var}_\theta \left(\frac{\partial \log p_\theta(X)}{\partial \theta} + \frac{\partial \log q_\theta(Y)}{\partial \theta} \right)$$

Because X and Y are independent, their score functions are also independent. For independent random variables, the variance of the sum is the sum of the variances.

Additivity of Fisher Information

The total information $I_{X,Y}(\theta)$ is the variance of the joint score:

$$I_{X,Y}(\theta) = \mathrm{Var}_\theta \left(\frac{\partial \log p_\theta(X)}{\partial \theta} + \frac{\partial \log q_\theta(Y)}{\partial \theta} \right)$$

Because X and Y are independent, their score functions are also independent. For independent random variables, the variance of the sum is the sum of the variances.

Result

$$\begin{aligned} I_{X,Y}(\theta) &= \mathrm{Var}_\theta \left(\frac{\partial \log p_\theta(X)}{\partial \theta} \right) + \mathrm{Var}_\theta \left(\frac{\partial \log q_\theta(Y)}{\partial \theta} \right) \\ &= I_X(\theta) + I_Y(\theta) \end{aligned}$$

Fisher information from independent sources is additive.

Information in a Random Sample

Extending the Additivity Principle

By iterating the result from the previous slide, if we have a random sample of n i.i.d. observations $X_1, \ldots, X_n$, and the Fisher information for a single observation is $I(\theta)$, then the total Fisher information for the entire sample is:

$$I_{\text{sample}}(\theta) = n \cdot I(\theta)$$

The Fisher Information Matrix

When the parameter θ is a vector in $\mathbb{R}^s$, the Fisher Information becomes an $s \times s$ matrix.

Definition

In regular cases, the entry in the i-th row and j-th column of the Fisher Information matrix $I(\theta)$ is given by:

$$\begin{aligned} I(\theta)_{i,j} &= E_\theta \left[\left(\frac{\partial \log p_\theta(X)}{\partial \theta_i} \right) \left(\frac{\partial \log p_\theta(X)}{\partial \theta_j} \right) \right] \\ &= \mathrm{Cov}_\theta \left(\frac{\partial \log p_\theta(X)}{\partial \theta_i}, \frac{\partial \log p_\theta(X)}{\partial \theta_j} \right) \\ &= -E_\theta \left[\frac{\partial^2 \log p_\theta(X)}{\partial \theta_i \partial \theta_j} \right] \end{aligned}$$

The Fisher Information Matrix

When the parameter θ is a vector in $\mathbb{R}^s$, the Fisher Information becomes an $s \times s$ matrix.

Definition

In regular cases, the entry in the i-th row and j-th column of the Fisher Information matrix $I(\theta)$ is given by:

$$\begin{aligned} I(\theta)_{i,j} &= E_\theta \left[\left(\frac{\partial \log p_\theta(X)}{\partial \theta_i} \right) \left(\frac{\partial \log p_\theta(X)}{\partial \theta_j} \right) \right] \\ &= \mathrm{Cov}_\theta \left(\frac{\partial \log p_\theta(X)}{\partial \theta_i}, \frac{\partial \log p_\theta(X)}{\partial \theta_j} \right) \\ &= -E_\theta \left[\frac{\partial^2 \log p_\theta(X)}{\partial \theta_i \partial \theta_j} \right] \end{aligned}$$

Note

The first two lines are equal because the expected value of the score vector is zero, $E_\theta[\nabla_\theta \log p_\theta(X)] = \mathbf{0}$. The third formula requires extra regularity to pass derivatives inside the integral.

The Fisher Information Matrix in Matrix Notation

Using matrix notation, the definitions can be expressed more compactly.

- Let ∇_θ be the gradient operator with respect to θ.
- Let ∇^2_θ be the Hessian matrix of second-order partial derivatives.
- Let prime ($'$) denote the transpose of a vector or matrix.

Matrix Forms of $I(\theta)$

The Fisher Information matrix can be written as:

$$\begin{aligned} I(\theta) &= E_\theta \left[(\nabla_\theta \log p_\theta(X))(\nabla_\theta \log p_\theta(X))' \right] \\ &= \mathrm{Cov}_\theta \left(\nabla_\theta \log p_\theta(X) \right) \\ &= -E_\theta \left[\nabla^2_\theta \log p_\theta(X) \right] \end{aligned}$$

Example 4.12: Exponential Families

s-Parameter Canonical Form

Let P be an s-parameter exponential family with densities:

$$p_\eta(x) = \exp\left[\eta \cdot T(x) - A(\eta)\right] h(x)$$

where $\eta \in \mathbb{R}^s$. The log-likelihood is
$\log p_\eta(X) = \eta \cdot T(X) - A(\eta) + \log h(X)$.

Example 4.12: Exponential Families

s-Parameter Canonical Form

Let P be an s-parameter exponential family with densities:

$$p_\eta(x) = \exp\left[\eta \cdot T(x) - A(\eta)\right] h(x)$$

where $\eta \in \mathbb{R}^s$. The log-likelihood is
$\log p_\eta(X) = \eta \cdot T(X) - A(\eta) + \log h(X)$.

Calculating the Information Matrix

The second-order partial derivative of the log-likelihood is:

$$\frac{\partial^2 \log p_\eta(X)}{\partial \eta_i \partial \eta_j} = -\frac{\partial^2 A(\eta)}{\partial \eta_i \partial \eta_j}$$

Using the formula $I(\eta)_{i,j} = -E_\eta\left[\frac{\partial^2 \log p_\eta(X)}{\partial \eta_i \partial \eta_j}\right]$, we get:

$$I(\eta)_{i,j} = \frac{\partial^2 A(\eta)}{\partial \eta_i \partial \eta_j}$$

Example 4.12: Exponential Families

Calculating the Information Matrix

The second-order partial derivative of the log-likelihood is:

$$\frac{\partial^2 \log p_\eta(X)}{\partial \eta_i \partial \eta_j} = -\frac{\partial^2 A(\eta)}{\partial \eta_i \partial \eta_j}$$

Using the formula $I(\eta)_{i,j} = -E_\eta \left[\frac{\partial^2 \log p_\eta(X)}{\partial \eta_i \partial \eta_j} \right]$, we get:

$$I(\eta)_{i,j} = \frac{\partial^2 A(\eta)}{\partial \eta_i \partial \eta_j}$$

Succinct Form

This means the Fisher Information matrix is simply the Hessian of $A(\eta)$:

$$I(\eta) = \nabla^2 A(\eta)$$