Inference 2, 2023, lecture 2

Rolf Larsson

November 6, 2023

Today

Chap. 3. Inference Principles:

3.1 Likelihood function

3.2
- Score function
- Fisher information

Likelihood function

Example 1: Poisson distribution

- Suppose that $\mathbf{X} = (X_1, X_2, X_3)$ where each X_i is Poisson with parameter λ.
- The X_i are independent.
- We observe $x_1 = 1, x_2 = 2, x_3 = 3$.
- The probability to obtain these observations is (why?)

$$P(X_1 = 1, X_2 = 2, X_3 = 3) = \frac{\lambda^6}{12} \exp(-3\lambda).$$

Likelihood function

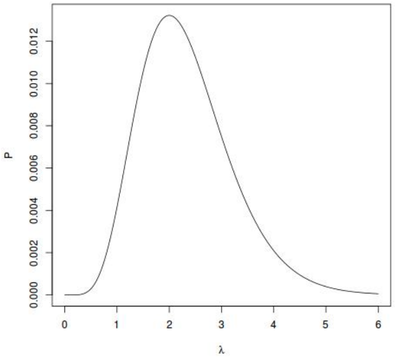

Figure: Probability of Poisson sample.

Likelihood function

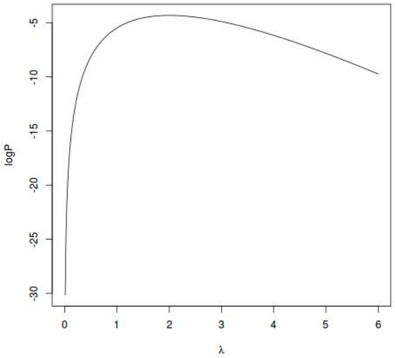

Figure: Log probability of Poisson sample.

Likelihood function

Definition (3.1)

For a fixed observation $\mathbf{x}$ of a random variable $\mathbf{X}$ with probability function $p(\cdot;\theta)$, the **likelihood function** $L(\cdot;\mathbf{x}) : \Theta \to \mathcal{R}_+$ is defined by

$$L(\theta;\mathbf{x}) = p(\mathbf{x};\theta).$$

If $\mathbf{X} = (X_1, ..., X_n)$ where the X_i are independent, then

$$L(\theta;\mathbf{x}) = P(X_1 = x_1, ..., X_n = x_n) = \prod_{i=1}^{n} \mathrm{P}_{i,\theta}(x_i)$$

in the discrete case and

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

in the continuous case.

Likelihood function

- Suppose $\mathbf{X} = (X_1, ..., X_n)$, where the X_i are independent.
- We observe $\mathbf{x} = (x_1, ..., x_n)$.
- Write down the likelihood function if
  1. X_i is Exponential with intensity β.
  2. X_i is Uniform on the interval $[0, \theta]$.

Likelihood function

Definition

The **maximum likelihood principle** says that the likelihood should be large(st) at the parameter which fits best to the data.

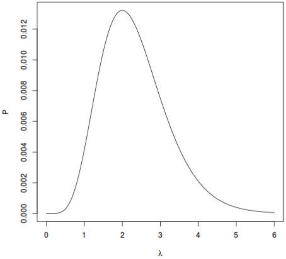

Figure: Likelihood function for a Poisson sample 1,2,3.

Likelihood function

Example 1: Poisson distribution

- Suppose that $\mathbf{X} = (X_1, X_2, X_3)$ where each X_i is Poisson with parameter λ.
- The X_i are independent.
- We observe $x_1 = 1, x_2 = 2, x_3 = 3$.
- Calculate the maximum likelihood estimate (MLE) of λ.

Likelihood function

Lemma (3.1)

Define the log likelihood $l(\theta; \mathbf{X}) = \log\{L(\theta; \mathbf{X})\}$.
Let $\theta = \theta_0$ *be the true (underlying) parameter value. Then,*

$$\mathrm{E}_{\theta_0}\{l(\theta; \mathbf{X})\} \leq \mathrm{E}_{\theta_0}\{l(\theta_0; \mathbf{X})\}$$

for all $\theta_0, \theta \in \Theta$.

Score function

Suppose the following regularity conditions are satisfied (2-3 are stated only in the scalar case):

1. The distributions $\{\mathrm{P} : \theta \in \Theta\}$ have common support, so that the set
$$\mathcal{A} = \{\mathbf{x} : p(\mathbf{x}; \theta) > 0\}$$
is independent of θ.
2. The parameter space Θ is an open interval (finite or infinite).
3. For any $\mathbf{x} \in \mathcal{A}$ and all $\theta \in \Theta$, the derivative $\partial p(\mathbf{x}; \theta)/\partial\theta$ exists and is finite.

The regularity conditions imply that

$$\frac{\partial}{\partial\theta} \int_{\mathcal{A}} f(\mathbf{x}; \theta) d\mathbf{x} = \int_{\mathcal{A}} \frac{\partial}{\partial\theta} f(\mathbf{x}; \theta) d\mathbf{x}$$

in the continuous case, and similarly in the discrete case.

Score function

Do the regularity conditions hold for

1. the exponential distribution?
2. the uniform distribution?

Score function

Definition (3.2)

Suppose regularity conditions 1-3 are satisfied. For every $\mathbf{x} \in \mathcal{A}$, we define the **score function** as the derivative of the log-likelihood function, i.e.

$$V(\theta; \mathbf{x}) = l'(\theta; \mathbf{x}) = \frac{\partial}{\partial \theta} \log\{L(\theta; \mathbf{x})\}.$$

Theorem (3.1)

Suppose regularity conditions 1-3 are satisfied. We have

$$\mathrm{E}_\theta\{V(\theta; \mathbf{X})\} = 0$$

for all $\theta \in \Theta$.

Score function

- Suppose $\mathbf{X} = (X_1, ..., X_n)$, where the X_i are independent.
- We observe $\mathbf{x} = (x_1, ..., x_n)$.
- Calculate the score function and verify that it has expectation 0 when the X_i have the following distributions:
  1. Exponential with intensity β.
  2. $N(\mu, \sigma^2)$ with known σ^2.

Fisher information

Definition (3.3)

Suppose regularity conditions 1-3 are satisfied.
The **Fisher information** is defined by

$$I_{\mathbf{X}}(\theta) = \mathrm{Var}_\theta\{V(\theta; \mathbf{X})\}.$$

Fisher information

Regularity condition 4: For all $\mathbf{x} \in \mathcal{A}$ and all $\theta \in \Theta$ the log-likelihood is twice differentiable and for all $\theta \in \Theta$,

$$\frac{\partial^2}{\partial\theta^2}\int_{\mathcal{A}} f(\mathbf{x};\theta)d\mathbf{x} = \int_{\mathcal{A}} \frac{\partial^2}{\partial\theta^2} f(\mathbf{x};\theta)d\mathbf{x}$$

in the continuous case, and similarly in the discrete case.

The score function is the derivative of $\log L$ w.r.t. θ.
If (minus) this derivative changes quickly w.r.t. θ, then the "information" contained in the sample is high.

Theorem (3.1)

Suppose regularity conditions 1-4 are satisfied. Then

$$I_{\mathbf{X}}(\theta) = -\mathrm{E}_\theta\{l''(\theta;\mathbf{X})\} = -\mathrm{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log\{p(\mathbf{X};\theta)\}\right].$$

Fisher information

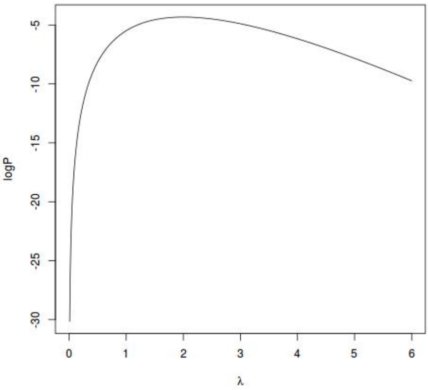

Figure: Log likelihood function for a Poisson sample 1,2,3.

Fisher information

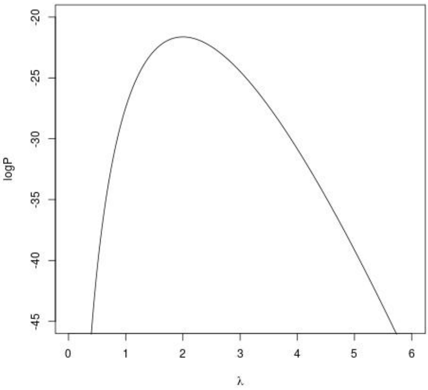

Figure: Log likelihood function for a Poisson sample 1,2,3,1,2,3,1,2,3,1,2,3,1,2,3.

Fisher information

- Suppose $\mathbf{X} = (X_1, ..., X_n)$, where the X_i are independent.
- We observe $\mathbf{x} = (x_1, ..., x_n)$.
- Calculate the Fisher information and verify theorem 3.1 by direct calculation when the X_i have the following distributions:
  1. Exponential with intensity β.
  2. $N(\mu, \sigma^2)$ with known σ^2.

News of today

- Likelihood
- MLE
- Score function (derivative of log likelihood)
- Fisher information
    - the variance of the score
    - minus the expected second derivative of the log likelihood