Math 3070, Applied Statistics

Section 1

October 23, 2019

Lecture Outline, 10/23

Section 6.2

- Maximum Likelihood Estimators

Likelihood Functions

Given a random sample $X_1 = x_1, \ldots, X_n = x_n$. **Likelihood function** L is a function which varies with parameter(s) of interest λ.

$$L(\lambda) = \prod_{i=1}^{n} f_{X_i}(x_i, \lambda)$$

Here $f_{X_i}(x_i, \lambda)$ is the density function of X_i and λ is a parameter. Note: this product formula assumes uses that the X_i of the sample are independent.

Maximum Likelihood Estimator (MLE)

The **Maximum Likelihood Estimator** (MLE) is the value of λ which maximizes the Likelihood Function L.

In practice, one often maximizes the **log likelihood function** ℓ instead since they are maximized at the same values.

$$\ell(\lambda) = \log[L(\lambda)] = \sum_{i=1}^{n} \log[f(x_i, \lambda)]$$

Note: The log function preserves maximums of functions. The maximums of $f(x)$ and $\log f(x)$ occur at the same place.

Log Tranforms and MLEs

Note: The log function preserves maximums of postive functions. The maximums of $f(x)$ and $\log f(x)$ occur at the same place.
Proof:

$$\begin{aligned}
\frac{d}{dx}\log[f(x)] &= \frac{f'(x)}{f(x)} \\
\frac{d^2}{dx^2}\log[f(x)] &= \frac{d}{dx}\frac{f'(x)}{f(x)} \\
&= \frac{f''(x)}{f(x)} - \frac{f'(x)}{f^2(x)}
\end{aligned}$$

Suppose $f(x)$ is maximized at c

$$f'(c) = \frac{f'(c)}{f(c)} = \frac{0}{f(c)} = 0$$

$$f''(c) = \frac{f''(c)}{f(c)} - \frac{f'(c)}{f^2(c)} = \frac{f''(c)}{f(c)} - \frac{0}{f^2(c)} > 0$$

Note: PDFs are nonnegative and their maxima should be positive.

Maximum Likelihood Estimation

Suppose a Poisson process has unknown rate λ. We observe the process for 2 hours, and the number of events which occur is $X = 40$. What value of λ gives the largest probability $P(X = 40)$?

X is a Poisson random variable with mean 2λ, so

$$P(X = 40) = \frac{e^{-2\lambda}(2\lambda)^{40}}{40!}$$

The function $L(\lambda) = \frac{e^{-2\lambda}(2\lambda)^{40}}{40!}$ here is called the *likelihood function*. To maximize $L(\lambda)$, we take the derivative:

$$L'(\lambda) = \frac{1}{40!}(e^{-2\lambda} \cdot 40(2\lambda)^{39} \cdot 2 - 2e^{-2\lambda}(2\lambda)^{40})$$

Setting this equal to zero gives $40 - 2\lambda = 0$, so $\lambda = 20$. This is called the *maximum likelihood estimate* of λ.

Maximum Likelihood Estimation

In the previous problem, the likelihood function was given by

$$L(\lambda) = \frac{e^{-2\lambda}(2\lambda)^{40}}{40!}$$

The *log-likelihood function* is

$$\ell(\lambda) = \ln(L(\lambda)) = -2\lambda + 40\ln(2\lambda) - \ln(40!)$$

The value of λ which maximizes $\ell(\lambda)$ will also maximize $L(\lambda)$. The log-likelihood function is often easier to differentiate than the likelihood function. In this case,

$$\ell'(\lambda) = -2 + \frac{40}{\lambda}$$

Setting this equal to zero gives $\lambda = 20$ as before.

Maximum Likelihood Estimation

In general, given a random variable X with pmf or pdf $f(x;\theta)$ depending on a parameter θ, the *maximum likelihood estimator* (mle) is the value $\hat{\theta}$ of θ that maximizes $f(X;\theta)$.

More generally, given a random sample $X_1, \ldots, X_n$, the mle is the value $\hat{\theta}$ that maximizes the joint pmf or joint pdf of $X_1, \ldots, X_n$: $f(X_1, \ldots, X_n; \theta) = f(X_1;\theta)f(X_2;\theta)\cdots f(X_n;\theta)$.

Under fairly general conditions, the maximum likelihood estimator $\hat{\theta}$ satisfies some desirable statistical properties:

- $\hat{\theta}$ exists and is unique.
- $\hat{\theta}$ is *asymptotically unbiased*: for large sample sizes, it is practically an unbiased estimator.
- $\hat{\theta}$ is *asymptotically efficient*: for large sample sizes, it approximately achieves the smallest possible variance for an unbiased estimator.
- $\hat{\theta}$ is *asymptotically normal*: for large sample sizes, it has approximately a normal distribution.

Invariance Principle for MLE

Let $h(\theta)$ be a function of θ. If $\hat{\theta}$ is the mle for θ, then $h(\hat{\theta})$ is the mle for $h(\theta)$.

Example: Given a random sample $X_1, \ldots, X_n$ from an exponential distribution with rate λ, we found that the mle for λ was

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} X_i}$$

Therefore, the mle for $\mu = 1/\lambda$ is

$$\hat{\mu} = 1/\hat{\lambda} = \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X}$$

Thus, the mle for μ is simply the sample mean. It can be shown that this in fact is the minimum variance unbiased estimator for μ.

Invariance Principle for MLE

Example: Given an observation of an exponential random variable X with unknown rate λ, we found that the mle for λ was

$$\hat{\lambda} = \frac{1}{X}$$

Therefore, the mle for $\mu = 1/\lambda$ is

$$\hat{\mu} = 1/\hat{\lambda} = \frac{1}{1/X} = X$$

Thus, the mle for μ is simply the observed value X. It can be shown that this in fact is the minimum variance unbiased estimator for μ.