Topic 8: Maximum Likelihood Estimation

Maximum Likelihood

Method for estimating a parameter in a statistical model where the distribution of Y is specified.

Examples

- Normal distribution
- Bernoulli distribution
- Poisson distribution

Normal distribution

$Y \sim N(\mu, 1)$ with $-\infty < \mu < \infty$. The density at y is

$$f(y) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(y-\mu)^2\right] \text{ where } exp(y) = e^y$$

Assume that we want to

- estimate the unknown parameter μ,
- based on a sample of n independent observations (draws/realizations from the $N(\mu, 1)$ distribution.

For a sample $Y_1, \ldots, Y_n$ the density is

$$f(y_1, \ldots, y_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2\right]$$

Likelihood function

The likelihood function is the density evaluated at the data $Y_1, \ldots, Y_n$, viewed as a function of the parameter μ.

The log likelihood function is more useful

$$L_n(\mu) = -\frac{1}{2}\sum_{i=1}^{n}(Y_i - \mu)^2 - n\log(\sqrt{2\pi})$$

The log likelihood function depends on

- the sample size n
- the parameter μ
- and the data.

Maximum Likelihood Estimator

The MLE is the parameter value $\hat{\mu}$ that maximized $L(\mu)$. Differentiate $L(\mu)$ with respect to μ

$$L_n'(\mu) = \frac{dL_n(\mu)}{d\mu} = \sum_{i=1}^{n}(Y_i - \mu)$$

set $L_n'(\mu) = 0$ and solve for $\hat{\mu}$:

$$\hat{\mu} = \bar{Y}$$

Since

$$L_n''(\mu) = \frac{d^2 L_n(\mu)}{d\mu^2} = -n$$

$\bar{Y}$ is the maximum and not the minimum.

Bernoulli distribution $Y = 0, 1$

The Bernoulli distribution is a Binomial$(1, p)$ distribution, where $0 < p < 1$ and $P(Y = y) = p^y(1-p)^{1-y}$. The probability that $Y_i = y_i$ for $i = 1, \ldots, n$ is

$$\prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$$

which leads to log likelihood function

$$\begin{aligned} L_n(p) &= \sum_{i=1}^{n} [Y_i \log p + (1 - Y_i) \log(1-p)] \\ &= S \log(p) + (n - S) \log(1-p) \end{aligned}$$

where $S = Y_1 + \ldots + Y_n$.

Bernoulli distribution

Find the MLE for p.

Poisson distribution

$Y = 0, 1, 2, 3, \ldots$

$$P(Y_i = y_i) = \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}$$

Joint probability for a sample of n independent observations

$$\prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} = e^{-n\lambda}\lambda^{y_1+\ldots+y_n} \prod_{i=1}^{n} \frac{1}{y_i!}$$

Log likelihood function, with $S = Y_1 + \ldots + Y_n$

$$L_n(\lambda) = -n\lambda + \lambda S - \sum_{i=1}^{n} \log Y_i!$$

Find the MLE for λ

Properties of MLE

In the three examples, the small sample distribution of the MLE is known (up to parameters):

- Normal: $\hat{\mu} \sim N(\mu_0, 1/n)$
- Bernoulli: $S \sim Binomial(n, p_0)$; $\hat{p} = S/n$ is $1/n$ times a Binomial.
- Poisson: $S \sim Poisson(n\lambda_0)$; $\hat{\lambda} = S/n$ is $1/n$ times a Poisson.

Here, μ_0, p_0 and λ_0 denote the true population parameters.

In general, the exact distribution of an MLE is unknown, and we work with the approximate distribution for large n.

Properties of MLE

Theorem

Suppose $Y_1, \ldots, Y_n$ are IID with probability distribution governed by the parameter θ. Let θ_0 be the true value of θ. Under regularity conditions, the MLE for θ is asymptotically normal. The asymptotic mean of the MLE is θ_0, the asymptotic variance $(nI_{\theta_0})^{-1}$.

Discussion

- $I_\theta = -E[L_1''(\theta)]$ is the *Fisher Information*, the negative of the expected value of the second derivative of the log likelihood function, for a sample of size 1.
- Not knowing θ_0, we cannot know I_{θ_0}. The asymptotic variance v_n can be computed in two ways:
  i) $(nI_{\hat{\theta}})^{-1}$
  ii) $[-L_n''(\hat{\theta})]^{-1}$
- The theorem says that $(\hat{\theta} - \theta_0)/\sqrt{(v_n)}$ is nearly $N(0,1)$ when the sample is large (and the assumption of IID sampling from $f(\theta_0)$ is correct).

Applications

State the asymptotic distributions of the following MLEs:

- $\hat{\mu}$ when sampling from an IID Normal
- $\hat{p}$ when sampling from an IID Bernoulli
- $\hat{\lambda}$ when sampling from an IID Poisson

Examples

i) In a random sample of 1000 Swiss employees, 78% stated that they received a pay increase for the current year.
   - What is the MLE for the share of employees with a pay raise?
   - Use the "observed information" approach to construct a 95% confidence interval.

ii) Consider the annual number of off-piste skiers/snowboarders involved in avalanche accidents in Switzerland between 1997 and 2007. The total number of fatalities between 1997 and 2001 was 31; the total number between 2002 and 2007 was 47.

   Assume that the annual number of fatalities is Poisson distributed. Use the ML approach to test whether the mean number increased in the 2002-2007 period relative to the earlier 1997-2001 period.

More than one parameter

When the parameter vector θ is p-dimensional, then

- $L'(\theta)$ is a p-vector with $j-$th component $\partial L/\partial\theta_j$.
- $L''(\theta)$ is a $p \times p$-matrix with $ij-$th component

$$\frac{\partial^2 L}{\partial\theta_i\partial\theta_j} = \frac{\partial^2 L}{\partial\theta_j\partial\theta_i}$$

In this case, $I_\theta = -E[L_1''(\theta)]$ is a (symmetric) $p \times p$-matrix. The diagonal elements of $(nI_{\theta_0})^{-1}$ give the asymptotic variances for the components of $\hat{\theta}$, the off-diagonal elements the covariances.

Inference in the $N(\mu, \sigma^2)$ model

n IID observations from $N(\mu, \sigma^2)$

There are now two parameters. Find

- the log-likelihood function
- the MLEs for μ and σ^2
- the asymptotic distribution for $(\hat{\mu}, \hat{\sigma}^2)$.

Outlook

The multivariate extensions are essential when we look at ML estimation of regression models

Ordinarily, the MLE is biased. Small sample distributions may be far from normal. A decent sample size n is needed.

Often, MLEs are not available in closed form. In this case, numerical optimization is needed, which is implemented in statistical software for standard models