Parameter Estimation
Fitting Probability Distributions
Maximum Likelihood

MIT 18.443

Dr. Kempthorne

Spring 2015

Outline

1. Maximum Likelihood Estimation
   - Framework/Definitions

Likelihood Definition

General Model

- Data Model : $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ vector-valued random variable with joint density given by
  $$f(x_1, \ldots, x_n \mid \theta)$$
- Data Realization: $\mathbf{X} = \mathbf{x} = (x_1, \ldots, x_n)$
- **Likelihood of** θ (given $\mathbf{x}$):
  $$lik(\theta) = f(x_1, \ldots, x_n \mid \theta)$$
  Note:
  - $(x_1, \ldots, x_n)$ treated as constants – realization of $\mathbf{X} = \mathbf{x}$
  - $lik(\theta) = lik(\theta \mid \mathbf{x})$ a function of θ given $\mathbf{x}$

Case A: One-Sample Model

- $X_1, X_2, \ldots, X_n$ i.i.d. r.v.'s. with density function $f(x \mid \theta)$

$$f(x_1, \ldots, x_n \mid \theta) = f(x_1 \mid \theta) \times \cdots \times f(x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

$$\Longrightarrow \quad lik(\theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

Likelihood Definition

Case B: Time-Series Model

- $X_1, X_2, \ldots, X_n$ are observations of a time series $\{X_t, t = 1, 2, \ldots\}$
- Joint density of $X = (X_1, X_2, \ldots, X_n)$ is given by:

$$\begin{aligned} f(x_1, \ldots, x_n \mid \theta) &= f(x_1 \mid \theta) \times f(x_2 \mid \theta, x_1) \times f(x_3 \mid \theta, x_1, x_2) \times \cdots \\ &\quad \times f(x_n \mid \theta, x_1, x_2, \ldots, x_{n-1}) \\ &= \prod_{i=1}^{n} f(x_i \mid \theta, \{x_j; j < i\}) \\ \Longrightarrow \quad lik(\theta) &= \prod_{i=1}^{n} f(x_i \mid \theta, \{x_j; j < i\}) \end{aligned}$$

Maximum Likelihood Estimate (MLE)

- Model/Likelihood Assumptions
  - Data Model : $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ vector-valued random variable with joint density given by
    $$f(x_1, \ldots, x_n \mid \theta)$$
  - Data Realization: $\mathbf{X} = \mathbf{x} = (x_1, \ldots, x_n)$
  - **Likelihood of** θ (given $\mathbf{x}$):
    $$lik(\theta) = f(x_1, \ldots, x_n \mid \theta)$$
- The **Maximum Likelihood Estimate** maximizes $lik(\theta)$
  $$lik(\hat{\theta}_{MLE}) = \max_{\theta} lik(\theta).$$
  - $\hat{\theta}_{MLE} = \hat{\theta}_{MLE}(\mathbf{x})$ (depends on $\mathbf{x}$!)
  - $\hat{\theta}_{MLE}$ is parameter for which realization $\mathbf{x}$ is "most likely"
- $\hat{\theta}_{MLE}$ maximizes the **log likelihood**
  $$\ell(\theta) = \log lik(\theta) = \sum_{i=1}^{n} \log[f(x_i \mid \theta)], \quad \text{(Case A)}$$

Specifying the MLE

Example 8.5.A: Poisson Distribution

- $X_1, \ldots, X_n$ i.i.d. $Poisson(\lambda)$
- $f(x \mid \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$
- $\ell(\lambda) = \sum_{i=1}^{n} [x_i \ln(\lambda) - \lambda - \ln(x!)]$

MLE $\hat{\lambda}_{MLE}$

- $\ell(\hat{\lambda}_{MLE})$ maximizes $\ell(\lambda)$
- $\hat{\lambda}_{MLE}$ solves: $$\frac{d\ell(\lambda)}{d\lambda} = 0$$
- $\hat{\lambda}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}$.

Properties of $\hat{\lambda}_{MLE}$

- $\hat{\lambda}_{MLE}$ same as $\hat{\lambda}_{MOM}$
- Sampling distribution of $\hat{\lambda}_{MLE}$ known.

Specifying the MLE

Example 8.5.B: Normal Distribution

- $X_1, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$
- $f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$
- $\ell(\mu, \sigma^2) = \sum_{i=1}^{n} \ln[f(x_i \mid \mu, \sigma^2)]$

$$= \sum_{i=1}^{n}[-\tfrac{1}{2}(\ln(2\pi) + \ln(\sigma^2)) - \tfrac{1}{2}\tfrac{(x-\mu)^2}{\sigma^2}]$$

$$= -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

MLE of $\theta = (\mu, \sigma^2)$:

- $\hat{\theta}_{MLE} = (\hat{\mu}_{MLE}, \hat{\sigma}^2_{MLE})$
- $\ell(\hat{\theta}_{MLE})$ maximizes $\ell(\theta) = \ell(\mu, \sigma)$
- $\hat{\theta}_{MLE}$ solves: $\dfrac{\partial \ell(\mu, \sigma^2)}{\partial \mu} = 0$ and $\dfrac{\partial \ell(\mu, \sigma^2)}{\partial \sigma^2} = 0$

MLEs of Normal Distribution Parameters

Normal MLEs

- $\hat{\mu}_{MLE} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- $\hat{\sigma}^2_{MLE} = \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$

Note:

- For any 1-1 function $g(\theta)$, the MLE for $g(\theta)$ is
  $$\widehat{g(\theta)}_{MLE} = g(\hat{\theta}_{MLE})$$
- $\hat{\sigma}_{MLE} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$
- $\hat{\theta}_{MLE} = (\hat{\mu}_{MLE}, \hat{\sigma}^2_{MLE}) = \hat{\theta}_{MOM}$
- Sampling distribution of $\hat{\theta}_{MLE}$ known (joint distribution!)

Specifying the MLE

Example 8.5.C: Gamma Distribution

- $X_1, \ldots, X_n$ i.i.d. $Gamma(\alpha, \lambda)$
- $f(x \mid \alpha, \lambda) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x}$
- $\ell(\alpha, \lambda) = \sum_{i=1}^n \ln[f(x_i \mid \mu, \sigma^2)]$

$$= \sum_{i=1}^n [-\ln(\Gamma(\alpha)) + \alpha \ln(\lambda) + (\alpha - 1)\ln(x_i) - \lambda x_i]$$

$$= -n\ln(\Gamma(\alpha)) + n\alpha \ln(\lambda) + (\alpha - 1)\sum_{i=1}^n \ln(x_i) - \lambda \sum_{i=1}^n x_i$$

MLE of $\theta = (\alpha, \lambda)$:

- $\hat{\theta}_{MLE} = (\hat{\alpha}_{MLE}, \hat{\lambda}_{MLE})$
- $\ell(\hat{\theta}_{MLE})$ maximizes $\ell(\theta) = \ell(\alpha, \lambda)$
- $\hat{\theta}_{MLE}$ solves: $\frac{\partial \ell(\alpha, \lambda)}{\partial \alpha} = 0$ and $\frac{\partial \ell(\mu, \lambda)}{\partial \lambda} = 0$

MLEs of Gamma Distribution Parameters

- $\ell(\alpha, \lambda) = -n\ln(\Gamma(\alpha)) + n\alpha\ln(\lambda) + (\alpha - 1)\sum_{i=1}^{n}\ln(x_i) - \lambda\sum_{i=1}^{n} x_i$
- Partial derivative Equations to solve:

$$0 = \frac{\partial\ell(\alpha,\lambda)}{\partial\alpha} = -n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + n\ln(\lambda) + \sum_{i=1}^{n}\ln(x_i)$$

$$0 = \frac{\partial\ell(\alpha,\lambda)}{\partial\lambda} = \frac{n\alpha}{\lambda} - \sum_{i=1}^{n} x_i$$

- Second equation gives $\hat{\lambda} = \hat{\alpha} \times \frac{n}{\sum_{i=1}^{n} x_i} = \hat{\alpha}/\overline{X}$
- Substitution in first equation gives

$$\begin{aligned} 0 = \frac{\partial\ell(\alpha,\hat{\lambda})}{\partial\alpha} &= -n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + n\ln(\hat{\lambda}) + \sum_{i=1}^{n}\ln(x_i) \\ &= -n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + n\ln(\alpha) - n\ln(\overline{X}) + \sum_{i=1}^{n}\ln(x_i) \end{aligned}$$

$\Longrightarrow 0 = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \ln(\alpha) - \ln(\overline{X}) + \frac{1}{n}\sum_{i=1}^{n}\ln(x_i)$

MLEs of Gamma Distribution Parameters

Note:

- The log likelihood $\ell(\alpha, \lambda)$ for a fixed α is maximized by
$$\hat{\lambda}_\alpha = \alpha/\overline{X}.$$
- Numerical methods required to solve for $\hat{\alpha}_{MLE}$ maximizing the **concentrated likelihood** or **profile likelihood**
$$\ell(\alpha, \hat{\lambda}_\alpha)$$
- The MLE for α does not equal $\hat{\alpha}_{MOM}$
- The sampling distribution of $\hat{\theta}_{MLE} = (\hat{\alpha}_{MLE}, \hat{\lambda}_{MLE})$ is approximated using the **bootstrap** method

Issue:

- How do the sampling distributions of $\hat{\theta}_{MLE}$ and $\hat{\theta}_{MOM}$ compare?
- Better estimates have lower dispersion about true θ.