Maximum Likelihood Estimation

Miaoyan Wang

Department of Statistics
UW Madison

General Distribution: ML Estimation

- In a general setting, let $Y_1, \ldots, Y_n$ be iid with probability density function $f(y; \boldsymbol{\theta})$.
- With $\boldsymbol{y} = (y_1, \ldots, y_n)'$, the likelihood function for $\boldsymbol{\theta}$ is

$$\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}) = \prod_{i=1}^{n} f(y_i; \boldsymbol{\theta}).$$

- Find the value of $\boldsymbol{\theta}$ that maximizes $\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y})$.
- Equivalently, find the value of $\boldsymbol{\theta}$ that maximizes the log-likelihood

$$\ell(\boldsymbol{\theta}; \boldsymbol{y}) = \log \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}) = \log \prod_{i=1}^{n} f(y_i; \boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(y_i; \boldsymbol{\theta}).$$

- Intuition:
  Find the parameter value of $\boldsymbol{\theta}$ that most likely produced the data.

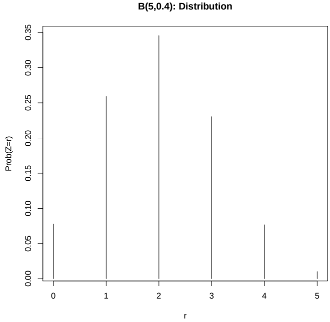
Binomial Distribution: Probability

- Suppose $Y \sim B(n, \pi)$ with probability density function

$$P(Y = y) = \frac{n!}{y!(n-y)!}\pi^y(1-\pi)^{n-y},$$

where $y = 0, 1, \ldots, n$.

- For example, $n = 5$ and $\pi = 0.4$. Plot $P(Y = y)$ versus y:



Binomial Distribution: Statistics

- Suppose there are $n = 5$ trials and the observed number of successes is $y = 2$.
- Q: How to estimate π?
- Consider the probability mass function evaluated at $y = 2$:

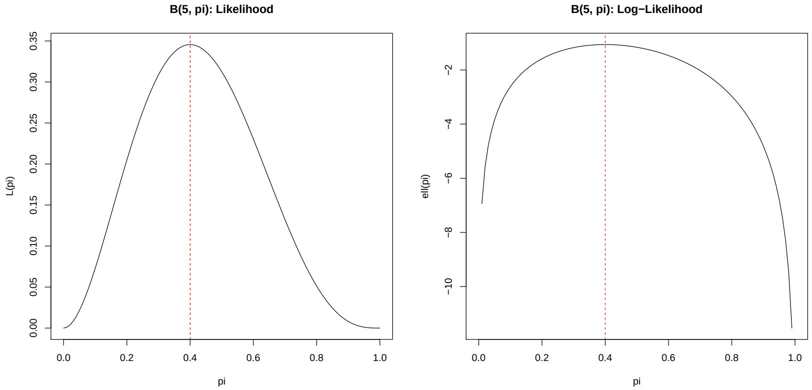
$$P(Y = 2) = \frac{5!}{2!3!}\pi^2(1-\pi)^3.$$

- Thus, we have

π	0.2	0.4	0.6	0.8
$P(Y = 2)$	0.2048	0.3456	0.2304	0.0512

Likelihood Function

- Plot $P(Y = 2)$ versus $\pi = 0.01, 0.02, \ldots, 0.98, 0.99$:



- Q: What value of π makes the given data most likely?

Likelihood Function

- That is, find the value of π that maximizes

$$\frac{5!}{2!3!}\pi^2(1-\pi)^3$$

- Given n and y, the function

$$\mathcal{L}(\pi) = \frac{n!}{y!(n-y)!}\pi^y(1-\pi)^{n-y}$$

  is the **likelihood function** of the unknown parameter π.
- Further, the **log-likelihood function** of π is:

$$\ell(\pi) = y\ln(\pi) + (n-y)\ln(1-\pi) + \ln\left\{\frac{n!}{y!(n-y)!}\right\}.$$

- The **maximum likelihood estimate (MLE)** of π is:

$$\hat{\pi} = \frac{y}{n} = \frac{2}{5}.$$

  (Proof in class)

[MLE] The MLE for a parameter θ is the statistics $\hat{\theta} = T(y)$ whose value for the given data y satisfies the condition

$$L(\hat{\theta}|y) = \sup_{\theta \in \Theta} L(\theta|y),$$

where $L(\theta|y)$ is the likelihood function for θ.
Properties (STAT 609/709):

- MLEs are invariant; i.e., $MLE(g(\theta)) = g(MLE(\theta)) = g(\hat{\theta})$.
- MLEs are asymptotically normal and asymptotically unbiased.

Example: MLE for Gaussian Distribution

- Suppose $Y_1, Y_2, \ldots, Y_n \sim$ iid $N(\mu, \sigma^2)$.
- Goal: estimate μ and σ^2
- Given the data $y_1, y_2, \ldots, y_n$, the likelihood function of μ, σ^2 is

$$\mathcal{L}(\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right\}.$$

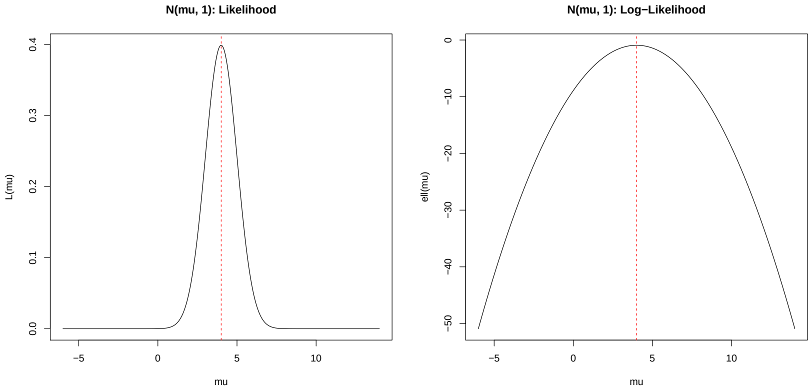
- The log-likelihood function of μ, σ^2 is

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2.$$

- The maximum likelihood estimate (MLE) for μ, σ^2 are:

$$\begin{aligned} \hat{\mu} &= \bar{y} \\ \tilde{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2. \end{aligned}$$

Gaussian Distribution: ML Estimation



Point Estimation

A good estimate $\hat{\theta}$ should

- Be unbiased: $\mathbb{E}(\hat{\theta}) = \theta$
- Have small sampling variance: small $\mathrm{Var}(\hat{\theta})$
- Be efficient: its mean squared error (MSE) is minimum among all competitors.

$$\mathrm{MSE}(\hat{\theta}) \equiv \mathbb{E}(\hat{\theta} - \theta)^2 = \mathrm{Bias}^2(\hat{\theta}) + \mathrm{Var}(\hat{\theta}),$$

  where $\mathrm{Bias}(\theta) = \mathbb{E}(\hat{\theta}) - \theta$.

- Be consistent:

$$\hat{\theta} = \hat{\theta}(n) \to \theta \quad \text{in probability, as the sample size} \quad n \to \infty.$$

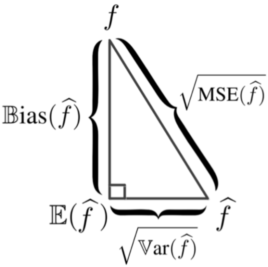
Proof of MSE decomposition

Bias-Variance Decomposition

$$\mathsf{MSE}(\hat{\theta}) \equiv \mathbb{E}(\hat{\theta} - \theta)^2 = \mathsf{Bias}^2(\hat{\theta}) + \mathsf{Var}(\hat{\theta}).$$

Proof

$$\begin{aligned}\mathsf{MSE}(\hat{\theta}) \equiv \mathbb{E}(\hat{\theta} - \theta)^2 &= \mathbb{E}(\hat{\theta} - \mathbb{E}\hat{\theta} + \mathbb{E}\hat{\theta} - \theta)^2 \\ &= \underbrace{\mathbb{E}(\hat{\theta} - \mathbb{E}\hat{\theta})^2}_{=\mathsf{Var}(\hat{\theta})} + \underbrace{(\mathbb{E}\hat{\theta} - \theta)^2}_{\mathsf{Bias}^2} + \underbrace{2(\mathbb{E}\hat{\theta} - \mathbb{E}\hat{\theta})(\mathbb{E}\hat{\theta} - \theta)}_{=0}\end{aligned}$$



Comparison

Method of Moment:

- Pros: easy to compute, consistent
- Cons: not necessarily the most efficient estimate; sometimes outside the valid range; may not be unique.

Maximum likelihood estimator:

- Pros: **asymptotically** unbiased, consistent, normally distributed, and efficient
- Cons: can be highly biased for small samples; sometimes, MLE has no closed-form.