Biostatistics 602 - Statistical Inference
Lecture 10
Maximum Likelihood Estimator

Hyun Min Kang

February 12th, 2013

Last Lecture

1. What is a point estimator, and a point estimate?

Last Lecture

1. What is a point estimator, and a point estimate?
2. What is a method of moment estimator?

Last Lecture

1. What is a point estimator, and a point estimate?
2. What is a method of moment estimator?
3. What are advantages and disadvantages of method of moment estimator?

Last Lecture

1. What is a point estimator, and a point estimate?
2. What is a method of moment estimator?
3. What are advantages and disadvantages of method of moment estimator?
4. What is a maximum likelihood estimator (MLE)?

Last Lecture

1. What is a point estimator, and a point estimate?
2. What is a method of moment estimator?
3. What are advantages and disadvantages of method of moment estimator?
4. What is a maximum likelihood estimator (MLE)?
5. How can you find an MLE?

Recap - Method of Moment Estimator

- Point Estimation - Estimate θ or $\tau(\theta)$.
- Method of Moment

$$m_1 = \frac{1}{n}\sum X_i = E\mathbf{X} = \mu_1$$
$$m_2 = \frac{1}{n}\sum X_i^2 = E\mathbf{X}^2 = \mu_2$$
$$\vdots$$
$$m_k = \frac{1}{n}\sum X_i^k = E\mathbf{X}^k = \mu_k$$

Recap - Example of Method of Moment Estimator

$$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$$

$$\hat{\mu} = \overline{X}$$

$$\hat{\mu}^2 + \hat{\sigma}^2 = E\mathbf{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$\hat{\sigma}^2 = \sum (X_i - \overline{X})^2 / n$$

Recap - Example of Method of Moment Estimator

$$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$$
$$\hat{\mu} = \overline{X}$$
$$\hat{\mu}^2 + \hat{\sigma}^2 = E\mathbf{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$
$$\hat{\sigma}^2 = \sum (X_i - \overline{X})^2 / n$$

- Easy to implement

Recap - Example of Method of Moment Estimator

$$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$$
$$\hat{\mu} = \overline{X}$$
$$\hat{\mu}^2 + \hat{\sigma}^2 = E\mathbf{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$
$$\hat{\sigma}^2 = \sum (X_i - \overline{X})^2 / n$$

- Easy to implement
- Easy to understand

Recap - Example of Method of Moment Estimator

$$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$$

$$\hat{\mu} = \overline{X}$$

$$\hat{\mu}^2 + \hat{\sigma}^2 = E\mathbf{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$\hat{\sigma}^2 = \sum (X_i - \overline{X})^2 / n$$

- Easy to implement
- Easy to understand
- Estimators can be improved; use as initial value to get other estimators

Recap - Example of Method of Moment Estimator

$$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$$
$$\hat{\mu} = \overline{X}$$
$$\hat{\mu}^2 + \hat{\sigma}^2 = E\mathbf{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$
$$\hat{\sigma}^2 = \sum (X_i - \overline{X})^2 / n$$

- Easy to implement
- Easy to understand
- Estimators can be improved; use as initial value to get other estimators
- No guarantee that the estimator will fall into the range of valid parameter space.

Recap - Likelihood Function

Definition

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} f_X(x|\theta)$. The join distribution of $\mathbf{X} = (X_1, \cdots, X_n)$ is

$$f_{\mathbf{X}}(\mathbf{x}|\theta) = \prod_{i=1}^{n} f_X(x_i|\theta)$$

Recap - Likelihood Function

Definition

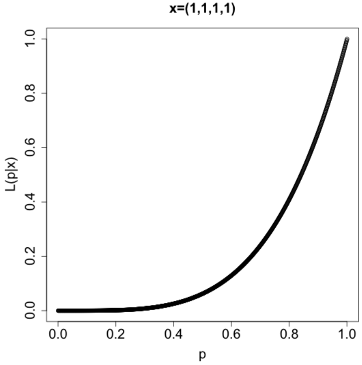
$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} f_X(x|\theta)$. The join distribution of $\mathbf{X} = (X_1, \cdots, X_n)$ is

$$f_{\mathbf{X}}(\mathbf{x}|\theta) = \prod_{i=1}^{n} f_X(x_i|\theta)$$

Given that $\mathbf{X} = \mathbf{x}$ is observed, the function of θ defined by $L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$ is called the likelihood function.

Recap - Example Likelihood Function

- $X_1, X_2, X_3, X_4 \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p),\ 0 < p < 1.$
- $\mathbf{x} = (1, 1, 1, 1)^T$
- Intuitively, it is more likely that p is larger than smaller.
- $L(p|\mathbf{x}) = f(\mathbf{x}|p) = \prod_{i=1}^{4} p^{x_i}(1-p)^{1-x_i} = p^4.$



How do we find MLE?

If the function is differentiable with respect to θ,

How do we find MLE?

If the function is differentiable with respect to θ,

1. Find candidates that makes first order derivative to be zero

How do we find MLE?

If the function is differentiable with respect to θ,

1. Find candidates that makes first order derivative to be zero
2. Check second-order derivative to check local maximum.

How do we find MLE?

If the function is differentiable with respect to θ,

1. Find candidates that makes first order derivative to be zero
2. Check second-order derivative to check local maximum.
    - For one-dimensional parameter, negative second order derivative implies local maximum.

How do we find MLE?

If the function is differentiable with respect to θ,

1. Find candidates that makes first order derivative to be zero
2. Check second-order derivative to check local maximum.
    - For one-dimensional parameter, negative second order derivative implies local maximum.
    - For two-dimensional parameter, suppose $L(\theta_1, \theta_2)$ is the likelihood function. Then we need to show

How do we find MLE?

If the function is differentiable with respect to θ,

1. Find candidates that makes first order derivative to be zero
2. Check second-order derivative to check local maximum.
   - For one-dimensional parameter, negative second order derivative implies local maximum.
   - For two-dimensional parameter, suppose $L(\theta_1, \theta_2)$ is the likelihood function. Then we need to show
     (a) $\partial^2 L(\theta_1, \theta_2)^2 / \partial\theta_1^2 < 0$ or $\partial^2 L(\theta_1, \theta_2)^2 / \partial\theta_2^2 < 0$.

How do we find MLE?

If the function is differentiable with respect to θ,

1. Find candidates that makes first order derivative to be zero
2. Check second-order derivative to check local maximum.
   - For one-dimensional parameter, negative second order derivative implies local maximum.
   - For two-dimensional parameter, suppose $L(\theta_1, \theta_2)$ is the likelihood function. Then we need to show
     (a) $\partial^2 L(\theta_1, \theta_2)^2 / \partial\theta_1^2 < 0$ or $\partial^2 L(\theta_1, \theta_2)^2 / \partial\theta_2^2 < 0$.
     (b) Determinant of second-order derivative is positive

How do we find MLE?

If the function is differentiable with respect to θ,

1. Find candidates that makes first order derivative to be zero
2. Check second-order derivative to check local maximum.
   - For one-dimensional parameter, negative second order derivative implies local maximum.
   - For two-dimensional parameter, suppose $L(\theta_1, \theta_2)$ is the likelihood function. Then we need to show
     (a) $\partial^2 L(\theta_1, \theta_2)^2 / \partial\theta_1^2 < 0$ or $\partial^2 L(\theta_1, \theta_2)^2 / \partial\theta_2^2 < 0$.
     (b) Determinant of second-order derivative is positive
   - Check boundary points to see whether boundary gives global maximum.

How do we find MLE?

If the function is differentiable with respect to θ,

1. Find candidates that makes first order derivative to be zero
2. Check second-order derivative to check local maximum.
   - For one-dimensional parameter, negative second order derivative implies local maximum.
   - For two-dimensional parameter, suppose $L(\theta_1, \theta_2)$ is the likelihood function. Then we need to show
     (a) $\partial^2 L(\theta_1, \theta_2)^2 / \partial\theta_1^2 < 0$ or $\partial^2 L(\theta_1, \theta_2)^2 / \partial\theta_2^2 < 0$.
     (b) Determinant of second-order derivative is positive
   - Check boundary points to see whether boundary gives global maximum.

If the function is NOT differentiable with respect to θ.

- Use numerical methods

How do we find MLE?

If the function is differentiable with respect to θ,

1. Find candidates that makes first order derivative to be zero
2. Check second-order derivative to check local maximum.
   - For one-dimensional parameter, negative second order derivative implies local maximum.
   - For two-dimensional parameter, suppose $L(\theta_1, \theta_2)$ is the likelihood function. Then we need to show
     (a) $\partial^2 L(\theta_1, \theta_2)^2 / \partial\theta_1^2 < 0$ or $\partial^2 L(\theta_1, \theta_2)^2 / \partial\theta_2^2 < 0$.
     (b) Determinant of second-order derivative is positive
   - Check boundary points to see whether boundary gives global maximum.

If the function is NOT differentiable with respect to θ.

- Use numerical methods
- Or perform direct maximization, using inequalities, or properties of the function.

How do we find MLE?

If the function is differentiable with respect to θ,

1. Find candidates that makes first order derivative to be zero
2. Check second-order derivative to check local maximum.
   - For one-dimensional parameter, negative second order derivative implies local maximum.
   - For two-dimensional parameter, suppose $L(\theta_1, \theta_2)$ is the likelihood function. Then we need to show
     (a) $\partial^2 L(\theta_1, \theta_2)^2 / \partial\theta_1^2 < 0$ or $\partial^2 L(\theta_1, \theta_2)^2 / \partial\theta_2^2 < 0$.
     (b) Determinant of second-order derivative is positive
   - Check boundary points to see whether boundary gives global maximum.

If the function is NOT differentiable with respect to θ.

- Use numerical methods
- Or perform direct maximization, using inequalities, or properties of the function.

Example of MLE : Uniform Distribution

Problem

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}(0, \theta)$, where $X_i \in [0, \theta]$ and $\theta > 0$.

Example of MLE : Uniform Distribution

Problem

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}(0, \theta)$, where $X_i \in [0, \theta]$ and $\theta > 0$.

Solution

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\theta} I(0 \leq x_i \leq \theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I(0 \leq x_i \leq \theta)$$

Example of MLE : Uniform Distribution

Problem

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}(0, \theta)$, where $X_i \in [0, \theta]$ and $\theta > 0$.

Solution

$$\begin{aligned} L(\theta|\mathbf{x}) &= \prod_{i=1}^{n} \frac{1}{\theta} I(0 \leq x_i \leq \theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I(0 \leq x_i \leq \theta) \\ &= \frac{1}{\theta^n} I(0 \leq x_1 \leq \theta \wedge \cdots \wedge 0 \leq x_n \leq \theta) \end{aligned}$$

Example of MLE : Uniform Distribution

Problem

$X_1, \cdots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}(0, \theta)$, where $X_i \in [0, \theta]$ and $\theta > 0$.

Solution

$$\begin{aligned} L(\theta|\mathbf{x}) &= \prod_{i=1}^{n} \frac{1}{\theta} I(0 \leq x_i \leq \theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I(0 \leq x_i \leq \theta) \\ &= \frac{1}{\theta^n} I(0 \leq x_1 \leq \theta \wedge \cdots \wedge 0 \leq x_n \leq \theta) \\ &= \frac{1}{\theta^n} I(x_{(n)} \leq \theta) I(x_{(1)} \geq 0) \end{aligned}$$

Example of MLE : Uniform Distribution

Problem

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}(0, \theta)$, where $X_i \in [0, \theta]$ and $\theta > 0$.

Solution

$$\begin{aligned} L(\theta|\mathbf{x}) &= \prod_{i=1}^{n} \frac{1}{\theta} I(0 \leq x_i \leq \theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I(0 \leq x_i \leq \theta) \\ &= \frac{1}{\theta^n} I(0 \leq x_1 \leq \theta \wedge \cdots \wedge 0 \leq x_n \leq \theta) \\ &= \frac{1}{\theta^n} I(x_{(n)} \leq \theta) I(x_{(1)} \geq 0) \end{aligned}$$

We need to maximize $1/\theta^n$ subject to constraint that $0 \leq x_{(n)} \leq \theta$.

Example of MLE : Uniform Distribution

Problem

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Uniform}(0, \theta)$, where $X_i \in [0, \theta]$ and $\theta > 0$.

Solution

$$\begin{aligned} L(\theta|\mathbf{x}) &= \prod_{i=1}^{n} \frac{1}{\theta} I(0 \leq x_i \leq \theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I(0 \leq x_i \leq \theta) \\ &= \frac{1}{\theta^n} I(0 \leq x_1 \leq \theta \wedge \cdots \wedge 0 \leq x_n \leq \theta) \\ &= \frac{1}{\theta^n} I(x_{(n)} \leq \theta) I(x_{(1)} \geq 0) \end{aligned}$$

We need to maximize $1/\theta^n$ subject to constraint that $0 \leq x_{(n)} \leq \theta$. Because $1/\theta^n$ decreases in θ, the MLE is $\hat{\theta}(\mathbf{X}) = X_{(n)}$.

Example of MLE : Normal Distribution

Problem

Suppose n pairs of data $(X_1, Y_1), \cdots, (X_n, Y_n)$ where X_i is generated from an unknown distribution, and Y_i are generated conditionally on X_i.

$$Y_i|X_i \sim \mathcal{N}(\alpha + \beta X_i, \sigma^2)$$

Find the MLE of $(\alpha, \beta, \sigma^2)$.

Example of MLE : Normal Distribution

Problem

Suppose n pairs of data $(X_1, Y_1), \cdots, (X_n, Y_n)$ where X_i is generated from an unknown distribution, and Y_i are generated conditionally on X_i.

$$Y_i | X_i \sim \mathcal{N}(\alpha + \beta X_i, \sigma^2)$$

Find the MLE of $(\alpha, \beta, \sigma^2)$.

Solution

The joint distribution of $(X_1, Y_1), \cdots, (X_n, Y_n)$ is

Example of MLE : Normal Distribution

Problem

Suppose n pairs of data $(X_1, Y_1), \cdots, (X_n, Y_n)$ where X_i is generated from an unknown distribution, and Y_i are generated conditionally on X_i.

$$Y_i | X_i \sim \mathcal{N}(\alpha + \beta X_i, \sigma^2)$$

Find the MLE of $(\alpha, \beta, \sigma^2)$.

Solution

The joint distribution of $(X_1, Y_1), \cdots, (X_n, Y_n)$ is

$$f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) \prod_{i=1}^{n} f_{\mathbf{Y}}(y_i | x_i)$$

Example of MLE : Normal Distribution

Problem

Suppose n pairs of data $(X_1, Y_1), \cdots, (X_n, Y_n)$ where X_i is generated from an unknown distribution, and Y_i are generated conditionally on X_i.

$$Y_i|X_i \sim \mathcal{N}(\alpha + \beta X_i, \sigma^2)$$

Find the MLE of $(\alpha, \beta, \sigma^2)$.

Solution

The joint distribution of $(X_1, Y_1), \cdots, (X_n, Y_n)$ is

$$\begin{aligned} f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) &= f_{\mathbf{X}}(\mathbf{x}) \prod_{i=1}^{n} f_{\mathbf{Y}}(y_i|x_i) \\ &= f_{\mathbf{X}}(\mathbf{x}) \prod_{i=1}^{n} \frac{1}{2\pi\sigma^2} \exp\left[-\frac{(y_i - \alpha - \beta x_i)^2}{2\sigma^2}\right] \end{aligned}$$

Solution : Normal Distribution (cont'd)

The likelihood function is

$$L(\alpha, \beta, \sigma^2 | \mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})(2\pi\sigma^2)^{-n/2} \exp\left[-\frac{\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2}{2\sigma^2}\right]$$

Solution : Normal Distribution (cont'd)

The likelihood function is

$$L(\alpha, \beta, \sigma^2|\mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})(2\pi\sigma^2)^{-n/2} \exp\left[-\frac{\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2}{2\sigma^2}\right]$$

The log-likelihood function can be simplied as

Solution : Normal Distribution (cont'd)

The likelihood function is

$$L(\alpha, \beta, \sigma^2 | \mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})(2\pi\sigma^2)^{-n/2} \exp\left[-\frac{\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2}{2\sigma^2}\right]$$

The log-likelihood function can be simplied as

$$l(\alpha, \beta, \sigma^2) = C - \frac{n}{2}\log(2\pi\sigma^2) - \frac{\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2}{2\sigma^2}$$

Solution : Normal Distribution (cont'd)

The likelihood function is

$$L(\alpha, \beta, \sigma^2 | \mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})(2\pi\sigma^2)^{-n/2} \exp\left[-\frac{\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2}{2\sigma^2}\right]$$

The log-likelihood function can be simplied as

$$l(\alpha, \beta, \sigma^2) = C - \frac{n}{2}\log(2\pi\sigma^2) - \frac{\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2}{2\sigma^2}$$

$$\frac{\partial l}{\partial \alpha} = \frac{2\sum_{i=1}^n (y_i - \alpha - \beta x_i)}{2\sigma^2} = \frac{n\bar{y} - n\alpha - n\beta\bar{x}}{\sigma^2} = 0$$

Solution : Normal Distribution (cont'd)

The likelihood function is

$$L(\alpha, \beta, \sigma^2 | \mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})(2\pi\sigma^2)^{-n/2} \exp\left[-\frac{\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2}{2\sigma^2}\right]$$

The log-likelihood function can be simplied as

$$l(\alpha, \beta, \sigma^2) = C - \frac{n}{2}\log(2\pi\sigma^2) - \frac{\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2}{2\sigma^2}$$

$$\frac{\partial l}{\partial \alpha} = \frac{2\sum_{i=1}^n (y_i - \alpha - \beta x_i)}{2\sigma^2} = \frac{n\overline{y} - n\alpha - n\beta\overline{x}}{\sigma^2} = 0$$

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$$

Solution : Normal Distribution (cont'd)

$$\frac{\partial l}{\partial \beta} = \frac{2\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)x_i}{2\sigma^2} = \frac{\sum_{i=1}^{n} x_i y_i - n\alpha\bar{x} - \beta\sum_{i=1}^{n} x_i^2}{\sigma^2} = 0$$

$$\sum_{i=1}^{n} x_i y_i - n\bar{x}(\bar{y} - \beta\bar{x}) - \beta\sum_{i=1}^{n} x_i^2 = 0$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2}\frac{2\pi}{2\pi\sigma} + \frac{\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2}{2(\sigma^2)^2} = 0$$

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)^2$$

Solution : Normal Distribution (cont'd)

$$\frac{\partial l}{\partial \beta} = \frac{2\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)x_i}{2\sigma^2} = \frac{\sum_{i=1}^{n} x_i y_i - n\alpha\bar{x} - \beta\sum_{i=1}^{n} x_i^2}{\sigma^2} = 0$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2}\frac{2\pi}{2\pi\sigma} + \frac{\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2}{2(\sigma^2)^2} = 0$$

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)^2$$

Solution : Normal Distribution (cont'd)

$$\frac{\partial l}{\partial \beta} = \frac{2\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)x_i}{2\sigma^2} = \frac{\sum_{i=1}^{n} x_i y_i - n\alpha\overline{x} - \beta\sum_{i=1}^{n} x_i^2}{\sigma^2} = 0$$

$$\sum_{i=1}^{n} x_i y_i - n\overline{x}(\overline{y} - \beta\overline{x}) - \beta\sum_{i=1}^{n} x_i^2 = 0$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}$$

Solution : Normal Distribution (cont'd)

$$\frac{\partial l}{\partial \beta} = \frac{2\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)x_i}{2\sigma^2} = \frac{\sum_{i=1}^{n} x_i y_i - n\alpha\overline{x} - \beta\sum_{i=1}^{n} x_i^2}{\sigma^2} = 0$$

$$\sum_{i=1}^{n} x_i y_i - n\overline{x}(\overline{y} - \beta\overline{x}) - \beta\sum_{i=1}^{n} x_i^2 = 0$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2}\frac{2\pi}{2\pi\sigma} + \frac{\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2}{2(\sigma^2)^2} = 0$$

Solution : Normal Distribution (cont'd)

$$\frac{\partial l}{\partial \beta} = \frac{2\sum_{i=1}^n (y_i - \alpha - \beta x_i)x_i}{2\sigma^2} = \frac{\sum_{i=1}^n x_i y_i - n\alpha\overline{x} - \beta\sum_{i=1}^n x_i^2}{\sigma^2} = 0$$

$$\sum_{i=1}^n x_i y_i - n\overline{x}(\overline{y} - \beta\overline{x}) - \beta\sum_{i=1}^n x_i^2 = 0$$

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i - n\overline{xy}}{\sum_{i=1}^n x_i^2 - n\overline{x}^2}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2}\frac{2\pi}{2\pi\sigma} + \frac{\sum_{i=1}^n (y_i - \alpha - \beta x_i)^2}{2(\sigma^2)^2} = 0$$

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2$$

Putting Things Together

Therefore, the MLE of $(\alpha, \beta, \sigma^2)$ is

Putting Things Together

Therefore, the MLE of $(\alpha, \beta, \sigma^2)$ is

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}$$

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)^2$$

Example : Normal Distribution with Known Variance

Problem

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$ where $\mu \geq 0$. Find MLE of μ.

Example : Normal Distribution with Known Variance

Problem

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$ where $\mu \geq 0$. Find MLE of μ.

Solution

$$L(\mu|\mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x_i - \mu)^2}{2}\right] = (2\pi)^{-n/2} \exp\left[-\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2}\right]$$

Example : Normal Distribution with Known Variance

Problem

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$ where $\mu \geq 0$. Find MLE of μ.

Solution

$$L(\mu|\mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x_i-\mu)^2}{2}\right] = (2\pi)^{-n/2} \exp\left[-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2}\right]$$

$$l(\mu|\mathbf{x}) = \log L(\mu, \mathbf{x}) = C - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2}$$

Example : Normal Distribution with Known Variance

Problem

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$ where $\mu \geq 0$. Find MLE of μ.

Solution

$$L(\mu|\mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x_i-\mu)^2}{2}\right] = (2\pi)^{-n/2} \exp\left[-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2}\right]$$

$$l(\mu|\mathbf{x}) = \log L(\mu, \mathbf{x}) = C - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2}$$

$$\frac{\partial l}{\partial \mu} = \frac{2\sum_{i=1}^{n}(x_i-\mu)}{2} = 0, \qquad \frac{\partial^2 l}{\partial \mu^2} < 0$$

Example : Normal Distribution with Known Variance

Problem

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$ where $\mu \geq 0$. Find MLE of μ.

Solution

$$
\begin{aligned}
L(\mu|\mathbf{x}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x_i-\mu)^2}{2}\right] = (2\pi)^{-n/2} \exp\left[-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2}\right] \\
l(\mu|\mathbf{x}) &= \log L(\mu, \mathbf{x}) = C - \frac{\sum_{i=1}^n (x_i-\mu)^2}{2} \\
\frac{\partial l}{\partial \mu} &= \frac{2\sum_{i=1}^n (x_i-\mu)}{2} = 0, \qquad \frac{\partial^2 l}{\partial \mu^2} < 0 \\
\hat{\mu} &= \sum_{i=1}^{n} x_i/n = \overline{x}
\end{aligned}
$$

Are we done?

The MLE parameter must be within the parameter space

We need to check whether $\hat{\mu}$ is within the parameter space $[0, \infty)$.

The MLE parameter must be within the parameter space

We need to check whether $\hat{\mu}$ is within the parameter space $[0, \infty)$.

- If $\overline{x} \geq 0$, $\hat{\mu} = \overline{x}$ falls into the parameter space.

The MLE parameter must be within the parameter space

We need to check whether $\hat{\mu}$ is within the parameter space $[0, \infty)$.

- If $\overline{x} \geq 0$, $\hat{\mu} = \overline{x}$ falls into the parameter space.
- If $\overline{x} < 0$, $\hat{\mu} = \overline{x}$ does NOT fall into the parameter space.

The MLE parameter must be within the parameter space

We need to check whether $\hat{\mu}$ is within the parameter space $[0, \infty)$.

- If $\overline{x} \geq 0$, $\hat{\mu} = \overline{x}$ falls into the parameter space.
- If $\overline{x} < 0$, $\hat{\mu} = \overline{x}$ does NOT fall into the parameter space.

When $\overline{x} < 0$

The MLE parameter must be within the parameter space

We need to check whether $\hat{\mu}$ is within the parameter space $[0, \infty)$.

- If $\overline{x} \geq 0$, $\hat{\mu} = \overline{x}$ falls into the parameter space.
- If $\overline{x} < 0$, $\hat{\mu} = \overline{x}$ does NOT fall into the parameter space.

When $\overline{x} < 0$

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^{n} (x_i - \mu) = n(\overline{x} - \mu) < 0$$

The MLE parameter must be within the parameter space

We need to check whether $\hat{\mu}$ is within the parameter space $[0, \infty)$.

- If $\overline{x} \geq 0$, $\hat{\mu} = \overline{x}$ falls into the parameter space.
- If $\overline{x} < 0$, $\hat{\mu} = \overline{x}$ does NOT fall into the parameter space.

When $\overline{x} < 0$

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^{n}(x_i - \mu) = n(\overline{x} - \mu) < 0$$

for $\mu \geq 0$. Therefore, $l(\mu|\mathbf{x})$ is a decreasing function of μ. So $\hat{\mu} = 0$ when $\overline{x} < 0$.

The MLE parameter must be within the parameter space

We need to check whether $\hat{\mu}$ is within the parameter space $[0, \infty)$.

- If $\overline{x} \geq 0$, $\hat{\mu} = \overline{x}$ falls into the parameter space.
- If $\overline{x} < 0$, $\hat{\mu} = \overline{x}$ does NOT fall into the parameter space.

When $\overline{x} < 0$

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^{n} (x_i - \mu) = n(\overline{x} - \mu) < 0$$

for $\mu \geq 0$. Therefore, $l(\mu|\mathbf{x})$ is a decreasing function of μ. So $\hat{\mu} = 0$ when $\overline{x} < 0$. Therefore, MLE is

$$\hat{\mu}(\mathbf{X}) = \max(\overline{X}, 0)$$

Invariance Property of MLE

Question

If $\hat{\theta}$ is the MLE of θ, what is the MLE of $\tau(\theta)$?

Invariance Property of MLE

Question

If $\hat{\theta}$ is the MLE of θ, what is the MLE of $\tau(\theta)$?

Example

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$ where $0 < p < 1$.

1. What is the MLE of p?

Invariance Property of MLE

Question

If $\hat{\theta}$ is the MLE of θ, what is the MLE of $\tau(\theta)$?

Example

$X_1, \cdots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$ where $0 < p < 1$.

1. What is the MLE of p?
2. What is the MLE of odds, defined by $\eta = p/(1-p)$?

MLE of p

$$L(p|\mathbf{x}) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

MLE of p

$$
\begin{aligned}
L(p|\mathbf{x}) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i} \\
l(p|\mathbf{x}) &= \log p \sum_{i=1}^{n} x_i + \log(1-p)(n - \sum_{i=1}^{n} x_i)
\end{aligned}
$$

MLE of p

$$
\begin{aligned}
L(p|\mathbf{x}) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i} \\
l(p|\mathbf{x}) &= \log p \sum_{i=1}^{n} x_i + \log(1-p)(n - \sum_{i=1}^{n} x_i) \\
\frac{\partial l}{\partial p} &= \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p} = 0
\end{aligned}
$$

MLE of p

$$
\begin{aligned}
L(p|\mathbf{x}) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i} \\
l(p|\mathbf{x}) &= \log p \sum_{i=1}^{n} x_i + \log(1-p)(n - \sum_{i=1}^{n} x_i) \\
\frac{\partial l}{\partial p} &= \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p} = 0 \\
\hat{p} &= \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}
\end{aligned}
$$

MLE of $\eta = \frac{p}{1-p}$

- $\eta = p/(1-p) = \tau(p)$
- $p = \eta/(1+\eta) = \tau^{-1}(\eta)$

MLE of $\eta = \frac{p}{1-p}$

- $\eta = p/(1-p) = \tau(p)$
- $p = \eta/(1+\eta) = \tau^{-1}(\eta)$

$$L^*(\eta|\mathbf{x}) = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

MLE of $\eta = \frac{p}{1-p}$

- $\eta = p/(1-p) = \tau(p)$
- $p = \eta/(1+\eta) = \tau^{-1}(\eta)$

$$\begin{aligned} L^*(\eta|\mathbf{x}) &= p^{\sum x_i}(1-p)^{n-\sum x_i} \\ &= \frac{p}{1-p}^{\sum x_i}(1-p)^n = \frac{\eta^{\sum x_i}}{(1+\eta)^n} \end{aligned}$$

MLE of $\eta = \frac{p}{1-p}$

- $\eta = p/(1-p) = \tau(p)$
- $p = \eta/(1+\eta) = \tau^{-1}(\eta)$

$$
\begin{aligned}
L^{*}(\eta|\mathbf{x}) &= p^{\sum x_i}(1-p)^{n-\sum x_i} \\
&= \frac{p}{1-p}^{\sum x_i}(1-p)^n = \frac{\eta^{\sum x_i}}{(1+\eta)^n} \\
l^{*}(\eta|\mathbf{x}) &= \sum_{i=1}^{n} x_i \log \eta - n \log(1+\eta)
\end{aligned}
$$

MLE of $\eta = \frac{p}{1-p}$

- $\eta = p/(1-p) = \tau(p)$
- $p = \eta/(1+\eta) = \tau^{-1}(\eta)$

$$\begin{aligned}
L^*(\eta|\mathbf{x}) &= p^{\sum x_i}(1-p)^{n-\sum x_i} \\
&= \frac{p}{1-p}^{\sum x_i}(1-p)^n = \frac{\eta^{\sum x_i}}{(1+\eta)^n} \\
l^*(\eta|\mathbf{x}) &= \sum_{i=1}^{n} x_i \log \eta - n \log(1+\eta) \\
\frac{\partial l^*}{\partial \eta} &= \frac{\sum_{i=1}^n x_i}{\eta} - \frac{n}{1+\eta} = 0
\end{aligned}$$

MLE of $\eta = \frac{p}{1-p}$

- $\eta = p/(1-p) = \tau(p)$
- $p = \eta/(1+\eta) = \tau^{-1}(\eta)$

$$
\begin{aligned}
L^*(\eta|\mathbf{x}) &= p^{\sum x_i}(1-p)^{n-\sum x_i} \\
&= \frac{p}{1-p}^{\sum x_i}(1-p)^n = \frac{\eta^{\sum x_i}}{(1+\eta)^n} \\
l^*(\eta|\mathbf{x}) &= \sum_{i=1}^{n} x_i \log\eta - n\log(1+\eta) \\
\frac{\partial l^*}{\partial \eta} &= \frac{\sum_{i=1}^n x_i}{\eta} - \frac{n}{1+\eta} = 0 \\
\hat{\eta} &= \frac{\sum_{i=1}^n x_i/n}{1-\sum_{i=1}^n x_i/n} = \tau(\hat{p})
\end{aligned}
$$

Another way to get MLE of $\eta = \frac{p}{1-p}$

$$L^*(\eta|\mathbf{x}) = \frac{\eta^{\sum x_i}}{(1+\eta)^n}$$

Another way to get MLE of $\eta = \frac{p}{1-p}$

$$L^*(\eta|\mathbf{x}) = \frac{\eta^{\sum x_i}}{(1+\eta)^n}$$

- From MLE of $\hat{p}$, we know $L^*(\eta|\mathbf{x})$ is maximized when $p = \eta/(1+\eta) = \hat{p}$.

Another way to get MLE of $\eta = \frac{p}{1-p}$

$$L^*(\eta|\mathbf{x}) = \frac{\eta^{\sum x_i}}{(1+\eta)^n}$$

- From MLE of $\hat{p}$, we know $L^*(\eta|\mathbf{x})$ is maximized when $p = \eta/(1+\eta) = \hat{p}$.
- Equivalently, $L^*(\eta|\mathbf{x})$ is maximized when $\eta = \hat{p}/(1-\hat{p}) = \tau(\hat{p})$, because τ is a one-to-one function.

Another way to get MLE of $\eta = \frac{p}{1-p}$

$$L^*(\eta|\mathbf{x}) = \frac{\eta^{\sum x_i}}{(1+\eta)^n}$$

- From MLE of $\hat{p}$, we know $L^*(\eta|\mathbf{x})$ is maximized when $p = \eta/(1+\eta) = \hat{p}$.
- Equivalently, $L^*(\eta|\mathbf{x})$ is maximized when $\eta = \hat{p}/(1-\hat{p}) = \tau(\hat{p})$, because τ is a one-to-one function.
- Therefore $\hat{\eta} = \tau(\hat{p})$.

Invariance Property of MLE

Fact

Denote the MLE of θ by $\hat{\theta}$. If $\tau(\theta)$ is an one-to-one function of θ, then MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$.

Invariance Property of MLE

Fact

Denote the MLE of θ by $\hat{\theta}$. If $\tau(\theta)$ is an one-to-one function of θ, then MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$.

Proof

The likelihood function in terms of $\tau(\theta) = \eta$ is

$$L^*(\tau(\theta)|\mathbf{x}) = \prod_{i=1}^{n} f_X(x_i|\theta) = \prod_{i=1}^{n} f(x_i|\tau^{-1}(\eta))$$

Invariance Property of MLE

Fact

Denote the MLE of θ by $\hat{\theta}$. If $\tau(\theta)$ is an one-to-one function of θ, then MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$.

Proof

The likelihood function in terms of $\tau(\theta) = \eta$ is

$$\begin{aligned} L^*(\tau(\theta)|\mathbf{x}) &= \prod_{i=1}^n f_X(x_i|\theta) = \prod_{i=1}^n f(x_i|\tau^{-1}(\eta)) \\ &= L(\tau^{-1}(\eta)|\mathbf{x}) \end{aligned}$$

Invariance Property of MLE

Fact

Denote the MLE of θ by $\hat{\theta}$. If $\tau(\theta)$ is an one-to-one function of θ, then MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$.

Proof

The likelihood function in terms of $\tau(\theta) = \eta$ is

$$\begin{aligned} L^*(\tau(\theta)|\mathbf{x}) &= \prod_{i=1}^{n} f_X(x_i|\theta) = \prod_{i=1}^{n} f(x_i|\tau^{-1}(\eta)) \\ &= L(\tau^{-1}(\eta)|\mathbf{x}) \end{aligned}$$

We know this function is maximized when $\tau^{-1}(\eta) = \hat{\theta}$, or equivalently, when $\eta = \tau(\hat{\theta})$.

Invariance Property of MLE

Fact

Denote the MLE of θ by $\hat{\theta}$. If $\tau(\theta)$ is an one-to-one function of θ, then MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$.

Proof

The likelihood function in terms of $\tau(\theta) = \eta$ is

$$\begin{aligned} L^*(\tau(\theta)|\mathbf{x}) &= \prod_{i=1}^{n} f_X(x_i|\theta) = \prod_{i=1}^{n} f(x_i|\tau^{-1}(\eta)) \\ &= L(\tau^{-1}(\eta)|\mathbf{x}) \end{aligned}$$

We know this function is maximized when $\tau^{-1}(\eta) = \hat{\theta}$, or equivalently, when $\eta = \tau(\hat{\theta})$. Therefore, MLE of $\eta = \tau(\theta)$ is $\tau(\hat{\theta})$.

Summary

Today

- Maximum Likelihood Estimator

Summary

Today

- Maximum Likelihood Estimator

Next Lecture

- Mean Squared Error
- Unbiased Estimator
- Cramer-Rao inequality