Biostatistics 602 - Statistical Inference
Lecture 09
Likelihood and Point Estimation

Hyun Min Kang

February 7th, 2013

Likelihood Function

Definition

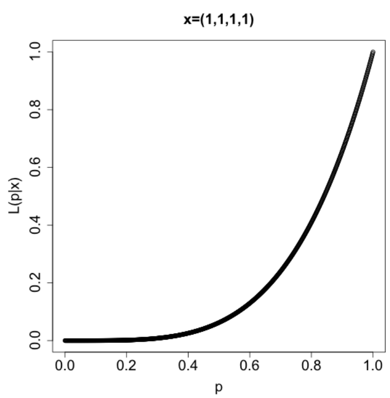
$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} f_X(x|\theta)$. The join distribution of $\mathbf{X} = (X_1, \cdots, X_n)$ is

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

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.

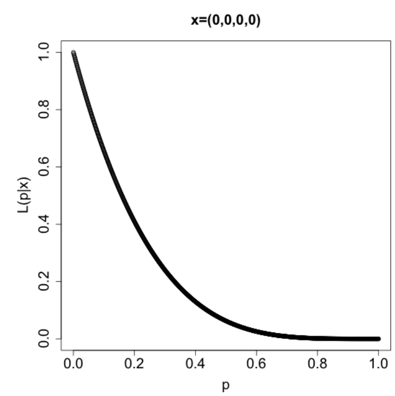
Examples of Likelihood Function - 1/3

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



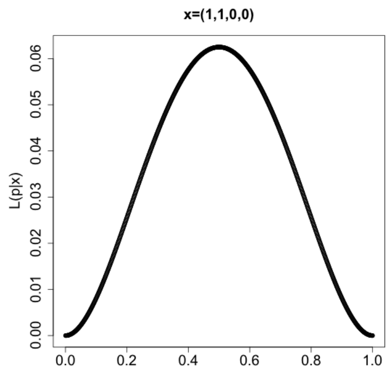
Examples of Likelihood Function - 2/3

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



Examples of Likelihood Function - 3/3

- $X_1, X_2, X_3, X_4 \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p),\ 0 < p < 1.$
- $\mathbf{x} = (1, 1, 0, 0)^T$
- Intuitively, it is more likely that p is somewhere in the middle than in the extremes.
- $L(p|\mathbf{x}) = f(\mathbf{x}|p) = \prod_{i=1}^{4} p^{x_i}(1-p)^{1-x_i} = p^2(1-p)^2.$



Point Estimation : Ingredients

- Data: $\mathbf{x} = (x_1, \cdots, x_n)$ - realizations of random variables $(X_1, \cdots, X_n)$.
- $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} f_X(x|\theta)$.
- Assume a model $\mathcal{P} = \{f_X(x|\theta) : \theta \in \Omega \subset \mathbb{R}^p\}$ where the functional form of $f_X(x|\theta)$ is known, but θ is unknown.
- Task is to use data $\mathbf{x}$ to make inference on θ

Point Estimation

Definition

If we use a function of sample $w(X_1, \cdots, X_n)$ as a "guess" of $\tau(\theta)$, where $\tau(\theta)$ is a function of true parameter θ. Then $w(\mathbf{X}) = w(X_1, \cdots, X_n)$ is called a *point estimator* of $\tau(\theta)$. The realization of the estimation, $w(\mathbf{x}) = w(x_1, \cdots, x_n)$ is called the *estimate* of $\tau(\theta)$.

Example

- $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, 1)$, where $\theta \in \Omega \in \mathbb{R}$.
- Suppose $n = 6$, and $(x_1, \cdots, x_6) = (2.0, 2.1, 2.9, 2.6, 1.2, 1.8)$.
- Define $w_1(X_1, \cdots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} = 2.1$.
- Define $w_2(X_1, \cdots, X_n) = X_{(1)} = 1.2$.

Method of Moments

A method to equate sample moments to population moments and solve equations.

Sample moments	Population moments
$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$	$\mu'_1 = E[X\|\theta] = \mu'_1(\theta)$
$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$	$\mu'_2 = E[X\|\theta] = \mu'_2(\theta)$
$m_3 = \frac{1}{n}\sum_{i=1}^{n} X_i^3$	$\mu'_3 = E[X\|\theta] = \mu'_3(\theta)$
$\vdots$	$\vdots$

Point estimator of $T(\theta)$ is obtained by solving equations like this.

$$\begin{aligned} m_1 &= \mu'_1(\theta) \\ m_2 &= \mu'_2(\theta) \\ &\vdots \\ m_k &= \mu'_k(\theta) \end{aligned}$$

Examples of method of moments estimator

Problem

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

Solution

$$\begin{aligned} \mu'_1 &= E\mathbf{X} = \mu = \overline{X} \\ \mu'_2 &= E\mathbf{X}^2 = [E\mathbf{X}]^2 + \mathrm{Var}(\mathbf{X}) = \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 \end{aligned}$$

$$\begin{cases} \hat{\mu} = \overline{X} \\ \hat{\mu}^2 + \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 \end{cases}$$

Solving the two equations above, $\hat{\mu} = \overline{X}$, $\hat{\sigma^2} = \sum_{i=1}^n (X_i - \overline{X})^2/n$.

Method of moments estimator - Binomial

Problem

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Binomial}(k, p)$. Find an estimator for k, p.

Solution

$$f_X(x|k,p) = \binom{k}{x} p^x (1-p)^{k-x} \qquad x \in \{0, 1, \cdots, k\}$$

Equating first two sample moments,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n X_i &= \overline{x} = \mu'_1 = E\mathbf{X} = kp \\ \frac{1}{n}\sum_{i=1}^n X_i^2 &= \mu'_2 = E[\mathbf{X}^2] = (E\mathbf{X})^2 + \mathrm{Var}(\mathbf{X}) = k^2p^2 + kp(1-p) \end{aligned}$$

Method of moments estimator - Binomial (cont'd)

The method of moments estimators are

$$\begin{aligned} \hat{k} &= \frac{\overline{X}^2}{\overline{X} - \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2} \\ \hat{p} &= \frac{\overline{X}}{\hat{k}} \end{aligned}$$

These are not the best estimators. It is possible to get negative estimates of k and p.

Examples of MoM estimator - Negative Binomial

Problem

$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Negative Binomial}(r, p)$. Find estimator for (r, p).

Solution

$$\begin{aligned} m_1 &= \frac{1}{n}\sum_{i=1}^n X_i = E\mathbf{X} = \frac{r(1-p)}{p} \\ m_2 &= \frac{1}{n}\sum_{i=1}^n X_i^2 = E\mathbf{X}^2 = \left(\frac{r(1-p)}{p}\right)^2 + \frac{r(1-p)}{p^2} \\ \hat{p} &= \frac{m_1}{m_2 - m_1^2} = \frac{\overline{X}}{\frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}^2} \\ \hat{r} &= \frac{m_1\hat{p}}{1-\hat{p}} = \frac{\overline{X}\hat{p}}{1-\hat{p}} \end{aligned}$$

Satterthwaite Approximation

Problem

Let $Y_1, \cdots, Y_k$ are independently (but not identically) distributed random variables from $\chi^2_{r_1}, \cdots, \chi^2_{r_k}$, respectively. We know that the distribution $\sum_{i=1}^n Y_i$ is also chi-squared with degrees of freedom equal to $\sum_{i=1}^k r_i$.

However, the distribution of $\sum_{i=1}^k a_i Y_i$, where a_is are known constants with $\sum_{i=1}^n a_i r_i = 1$, in general, the distribution is hard to obtain.

It is often reasonable to assume that the distribution of $\sum_{i=1}^k a_i Y_i$ follows $\frac{1}{\nu}\chi^2_\nu$ approximately. Find a moment-based estimator of ν.

A Naive Solution

To match the first moment, let $X \sim \chi^2_\nu/\nu$. Then $E(X) = 1$, and $\mathrm{Var}(X) = 2/\nu$.

$$E\left(\sum_{i=1}^k a_i Y_i\right) = \sum_{i=1}^k a_i E Y_i = \sum_{i=1}^k a_i r_i = 1 = E(X)$$

To match the second moments,

$$E\left(\sum_{i=1}^k a_i Y_i\right)^2 = E\left(X^2\right) = \frac{2}{\nu} + 1$$

Therefore, the method of moment estimator of ν is

$$\hat{\nu} = \frac{2}{(\sum_{i=1}^k a_i Y_i)^2 - 1}$$

Note that ν can be negative, which is not desirable.

An alternative Solution

To match the second moments,

$$\begin{aligned}
E\left(\sum_{i=1}^k a_i Y_i\right)^2 &= \mathrm{Var}\left(\sum_{i=1}^k a_i Y_i\right) + \left[E(\sum_{i=1}^k a_i Y_i)\right]^2 \\
&= \left[E(\sum_{i=1}^k a_i Y_i)\right]^2 \left[\frac{\mathrm{Var}(\sum_{i=1}^k a_i Y_i)}{\left[E(\sum_{i=1}^k a_i Y_i)\right]^2} + 1\right] \\
&= \left[\frac{\mathrm{Var}(\sum_{i=1}^k a_i Y_i)}{\left[E(\sum_{i=1}^k a_i Y_i)\right]^2} + 1\right] = \frac{2}{\nu} + 1 \\
\nu &= \frac{2\left[E(\sum_{i=1}^k a_i Y_i)\right]^2}{\mathrm{Var}(\sum_{i=1}^k a_i Y_i)}
\end{aligned}$$

Alternative Solution (cont'd)

To match the second moments, Finally, use the fact that $Y_1, \cdots, Y_k$ are independent chi-squared random variables.

$$\begin{aligned}
\mathrm{Var}(\sum_{i=1}^n a_i Y_i) &= \sum_{i=1}^k a_i^2 \mathrm{Var}(Y_i) \\
&= 2\sum_{i=1}^n \frac{a_i^2 (EY_i)^2}{r_i}
\end{aligned}$$

Substituting this expression for the variance and removing expectations, we obtain Satterthwaite's estimator

$$\hat{\nu} = \frac{\sum_{i=1}^n a_i Y_i}{\sum_{i=1}^n \frac{a_i^2}{r_i} Y_i^2}$$

Maximum Likelihood Estimator

Definition

- For a given sample point $\mathbf{x} = (x_1, \cdots, x_n)$,
- let $\hat{\theta}(\mathbf{x})$ be the value such that
- $L(\theta|\mathbf{x})$ attains its maximum.
- More formally, $L(\hat{\theta}(\mathbf{x})|\mathbf{x}) \geq L(\theta|\mathbf{x})\ \forall \theta \in \Omega$ where $\hat{\theta}(\mathbf{x}) \in \Omega$.
- $\hat{\theta}(\mathbf{x})$ is called the *maximum likelihood estimate* of θ based on data $\mathbf{x}$,
- and $\hat{\theta}(\mathbf{X})$ is the *maximum likelihood estimator (MLE)* of θ.

Example of MLE - Exponential Distribution

Problem

Let $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Exponential}(\beta)$. Find MLE of β.

Solution

$$\begin{aligned} L(\beta|\mathbf{x}) &= f_{\mathbf{X}}(\mathbf{x}|\theta) = \prod_{i=1}^n f_X(x_i|\theta) \\ &= \prod_{i=1}^n \left[\frac{1}{\beta} e^{-x_i/\beta}\right] = \frac{1}{\beta^n} \exp\left(-\sum_{i=1}^n \frac{x_i}{\beta}\right) \end{aligned}$$

where $\beta > 0$.

Use the derivative to find potential MLE

To maximize the likelihood function $L(\beta|\mathbf{x})$ is equivalent to maximize the log-likelihood function

$$\begin{aligned} l(\beta|\mathbf{x}) &= \log L(\beta|\mathbf{x}) = \log\left[\frac{1}{\beta^n}\exp\left(-\sum_{i=1}^n \frac{x_i}{\beta}\right)\right] \\ &= -\frac{\sum_{i=1}^n x_i}{\beta} - n\log\beta \\ \frac{\partial l}{\partial \beta} &= \frac{\sum_{i=1}^n x_i}{\beta^2} - \frac{n}{\beta} = 0 \\ \sum_{i=1}^n x_i &= n\beta \\ \hat{\beta} &= \frac{\sum_{i=1}^n x_i}{n} = \overline{x} \end{aligned}$$

Use the double derivative to confirm local maximum

$$\begin{aligned} \left.\frac{\partial^2 l}{\partial \beta^2}\right|_{\beta=\overline{x}} &= \left.-2\frac{\sum_{i=1}^n x_i}{\beta^3} + \frac{n}{\beta^2}\right|_{\beta=\overline{x}} \\ &= \left.\frac{1}{\beta^2}\left(-\frac{2\sum_{i=1}^n x_i}{\beta} + n\right)\right|_{\beta=\overline{x}} \\ &= \frac{1}{\overline{x}^2}\left(-\frac{2n\overline{x}}{\overline{x}} + n\right) \\ &= \frac{1}{\overline{x}^2}(-n) < 0 \end{aligned}$$

Therefore, we can conclude that $\hat{\beta}(\mathbf{X}) = \overline{X}$ is unique local maximum on the interval

Check boundary and confirm global maximum

$\beta \in (0, \infty)$. If $\beta \to \infty$

$$\begin{aligned} l(\beta|\mathbf{x}) &= -\frac{\sum_{i=1}^n x_i}{\beta} - n \log \beta \to -\infty \\ L(\beta|\mathbf{x}) &\to 0 \end{aligned}$$

If $\beta \to 0$, use $\log(x) = \lim_{\beta \to 0} \frac{1}{\beta}(x^\beta - 1)$

$$\begin{aligned} l(\beta|\mathbf{x}) &= -\frac{\sum_{i=1}^n x_i}{\beta} - n \log \beta \\ &= -\frac{\sum_{i=1}^n x_i}{\beta} - n\left(\frac{1}{\beta}\beta^\beta - 1\right) \\ &= -\frac{\sum_{i=1}^n x_i - n(\beta^\beta - 1)}{\beta} \to -\infty \\ L(\beta|\mathbf{x}) &\to 0 \end{aligned}$$

Putting Things Together

1. $\frac{\partial l}{\partial \beta} = 0$ at $\hat{\beta} = \overline{x}$
2. $\frac{\partial^2 l}{\partial \beta^2} < 0$ at $\hat{\beta} = \overline{x}$
3. $L(\beta|\mathbf{x}) \to 0$ (lowest bound) when β approaches the boundary

Therefore $l(\beta|\mathbf{x})$ and $L(\beta|\mathbf{x})$ attains the global maximum when $\hat{\beta} = \overline{x}$
$\hat{\beta}(\mathbf{X}) = \overline{X}$ is the MLE of β.

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 directly maximization, using inequalities, or properties of the function.

Summary

Today

- Likelihood Function
- Point Estimator
- Method of Moments Estimator
- Maximum Likelihood Estimator

Next Lecture

- Maximum Likelihood Estimator