Maximum Likelihood Estimators: Examples

Mathematics 47: Lecture 19

Dan Sloughter

Furman University

April 5, 2006

Example

Example

- Suppose X_1, X_2, ..., X_n is a random sample from $N(\mu, \sigma^2)$.

Example

- Suppose X_1, X_2, ..., X_n is a random sample from $N(\mu, \sigma^2)$.
- Then

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}.$$

Example

- Suppose $X_1, X_2, \ldots, X_n$ is a random sample from $N(\mu, \sigma^2)$.
- Then

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}.$$

- And so

$$\log(L(\mu, \sigma^2)) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Example (cont'd)

Example (cont'd)

- Hence

$$\frac{\partial}{\partial \mu} \log(L(\mu, \sigma^2)) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu)$$

and

$$\frac{\partial}{\partial \sigma^2} \log(L(\mu, \sigma^2)) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2.$$

Example (cont'd)

- Hence

$$\frac{\partial}{\partial \mu} \log(L(\mu, \sigma^2)) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu)$$

and

$$\frac{\partial}{\partial \sigma^2} \log(L(\mu, \sigma^2)) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2.$$

- Hence

$$\frac{\partial}{\partial \mu} \log(L(\mu, \sigma^2)) = 0$$
$$\frac{\partial}{\partial \sigma^2} \log(L(\mu, \sigma^2)) = 0$$

when

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0$$

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = 0.$$

Example (cont'd)

Example (cont'd)

- From the first, we have

$$\sum_{i=1}^{n} x_i - n\mu = 0,$$

and so

$$\mu = \frac{1}{n}\sum_{i=1}^{n} = \bar{x}.$$

Example (cont'd)

- From the first, we have

$$\sum_{i=1}^{n} x_i - n\mu = 0,$$

and so

$$\mu = \frac{1}{n}\sum_{i=1}^{n} = \bar{x}.$$

- From the second, we now have

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0,$$

and so

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Example (cont'd)

Example (cont'd)

- It may be shown that these values yield an absolute minimum for $L(\mu, \sigma^2)$.

Example (cont'd)

- It may be shown that these values yield an absolute minimum for $L(\mu, \sigma^2)$.
- And so

$$\hat{\mu} = \bar{X}$$
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

are the maximum likelihood estimators for μ and σ^2.

Example

Example

- Suppose X_1, X_2, ..., X_n is a random sample from a uniform distribution on $(0, \theta)$.

Example

- Suppose X_1, X_2, ..., X_n is a random sample from a uniform distribution on $(0, \theta)$.
- Then

$$L(\theta) = \begin{cases} \frac{1}{\theta^n}, & \text{if } \theta > x_{(n)}, \\ 0, & \text{otherwise.} \end{cases}$$

Example

- Suppose X_1, X_2, ..., X_n is a random sample from a uniform distribution on $(0, \theta)$.
- Then

$$L(\theta) = \begin{cases} \dfrac{1}{\theta^n}, & \text{if } \theta > x_{(n)}, \\ 0, & \text{otherwise.} \end{cases}$$

- Note: L is a decreasing function on $(x_{(n)}, \infty)$.

Example

- Suppose X_1, X_2, ..., X_n is a random sample from a uniform distribution on $(0, \theta)$.
- Then

$$L(\theta) = \begin{cases} \dfrac{1}{\theta^n}, & \text{if } \theta > x_{(n)}, \\ 0, & \text{otherwise.} \end{cases}$$

- Note: L is a decreasing function on $(x_{(n)}, \infty)$.
- Hence L is maximized when $\theta = x_{(n)}$.

Example

- Suppose X_1, X_2, ..., X_n is a random sample from a uniform distribution on $(0, \theta)$.
- Then
$$L(\theta) = \begin{cases} \frac{1}{\theta^n}, & \text{if } \theta > x_{(n)}, \\ 0, & \text{otherwise.} \end{cases}$$
- Note: L is a decreasing function on $(x_{(n)}, \infty)$.
- Hence L is maximized when $\theta = x_{(n)}$.
- That is, the maximum likelihood estimator for θ is $\hat{\theta} = X_{(n)}$.

Example

Example

- Suppose X is a discrete random variable with range $\{1, 2, \ldots, k\}$.

Example

- Suppose X is a discrete random variable with range $\{1, 2, \ldots, k\}$.
- Suppose X_1, X_2, ..., X_n is a random sample from the distribution of X.

Example

- Suppose X is a discrete random variable with range $\{1, 2, \ldots, k\}$.
- Suppose X_1, X_2, ..., X_n is a random sample from the distribution of X.
- For $i = 1, 2, \ldots, k$, let $p_i = P(X = i)$.

Example

- Suppose X is a discrete random variable with range $\{1, 2, \ldots, k\}$.
- Suppose X_1, X_2, ..., X_n is a random sample from the distribution of X.
- For $i = 1, 2, \ldots, k$, let $p_i = P(X = i)$.
- Let

$$Y_i = \text{ number of outcomes equal to } i.$$

Example

- Suppose X is a discrete random variable with range $\{1, 2, \ldots, k\}$.
- Suppose X_1, X_2, ..., X_n is a random sample from the distribution of X.
- For $i = 1, 2, \ldots, k$, let $p_i = P(X = i)$.
- Let

$$Y_i = \text{ number of outcomes equal to } i.$$

- Then, for $0 \leq p_i \leq 1$, $i = 1, 2, \ldots, n$,

$$L(p_1, p_2, \ldots, p_{k-1}) = p_1^{y_1} p_2^{y_2} \cdots p_{k-1}^{y_{k-1}} (1 - p_1 - p_2 - \cdots - p_{k-1})^{y_k}.$$

Example

- Suppose X is a discrete random variable with range $\{1, 2, \ldots, k\}$.
- Suppose X_1, X_2, ..., X_n is a random sample from the distribution of X.
- For $i = 1, 2, \ldots, k$, let $p_i = P(X = i)$.
- Let

$$Y_i = \text{ number of outcomes equal to } i.$$

- Then, for $0 \leq p_i \leq 1$, $i = 1, 2, \ldots, n$,

$$L(p_1, p_2, \ldots, p_{k-1}) = p_1^{y_1} p_2^{y_2} \cdots p_{k-1}^{y_{k-1}} (1 - p_1 - p_2 - \cdots - p_{k-1})^{y_k}.$$

- And so

$$\log(L(p_1, p_2, \ldots, p_{k-1})) = y_1 \log(p_1) + y_2 \log(p_2) + \cdots + y_{k-1} \log(p_{k-1}) + y_k \log(1 - p_1 - p_2 - \cdots - p_{k-1}).$$

Example (cont'd)

Example (cont'd)

- Hence, for $i = 1, 2, \ldots, k-1$,

$$\frac{\partial}{\partial p_i} \log(L(p_1, p_2, \ldots, p_{k-1})) = \frac{y_i}{p_i} - \frac{y_k}{p_k}.$$

Example (cont'd)

- Hence, for $i = 1, 2, \ldots, k-1$,

$$\frac{\partial}{\partial p_i} \log(L(p_1, p_2, \ldots, p_{k-1})) = \frac{y_i}{p_i} - \frac{y_k}{p_k}.$$

- Thus $\frac{\partial}{\partial p_i} \log(L(p_1, p_2, \ldots, p_{k-1})) = 0$, for $i = 1, 2, \ldots, k-1$, when

$$\frac{y_i}{p_i} = \frac{y_k}{p_k}.$$

Example (cont'd)

- Hence, for $i = 1, 2, \ldots, k-1$,

$$\frac{\partial}{\partial p_i} \log(L(p_1, p_2, \ldots, p_{k-1})) = \frac{y_i}{p_i} - \frac{y_k}{p_k}.$$

- Thus $\frac{\partial}{\partial p_i} \log(L(p_1, p_2, \ldots, p_{k-1})) = 0$, for $i = 1, 2, \ldots, k-1$, when

$$\frac{y_i}{p_i} = \frac{y_k}{p_k}.$$

- Note: this equation in fact holds for $i = 1, 2, \ldots, k$.

Example (cont'd)

- Hence, for $i = 1, 2, \ldots, k-1$,

$$\frac{\partial}{\partial p_i} \log(L(p_1, p_2, \ldots, p_{k-1})) = \frac{y_i}{p_i} - \frac{y_k}{p_k}.$$

- Thus $\frac{\partial}{\partial p_i} \log(L(p_1, p_2, \ldots, p_{k-1})) = 0$, for $i = 1, 2, \ldots, k-1$, when

$$\frac{y_i}{p_i} = \frac{y_k}{p_k}.$$

- Note: this equation in fact holds for $i = 1, 2, \ldots, k$.
- Hence

$$\sum_{i=1}^{k} y_i = \frac{y_k}{p_k} \sum_{i=1}^{k} p_i.$$

Example (cont'd)

- Hence, for $i = 1, 2, \ldots, k-1$,

$$\frac{\partial}{\partial p_i} \log(L(p_1, p_2, \ldots, p_{k-1})) = \frac{y_i}{p_i} - \frac{y_k}{p_k}.$$

- Thus $\frac{\partial}{\partial p_i} \log(L(p_1, p_2, \ldots, p_{k-1})) = 0$, for $i = 1, 2, \ldots, k-1$, when

$$\frac{y_i}{p_i} = \frac{y_k}{p_k}.$$

- Note: this equation in fact holds for $i = 1, 2, \ldots, k$.
- Hence

$$\sum_{i=1}^{k} y_i = \frac{y_k}{p_k} \sum_{i=1}^{k} p_i.$$

- It follows that $n = \frac{y_k}{p_k}$.

Example

Example

- Thus, for $i = 1, 2, \ldots, n$,

$$\frac{y_i}{p_i} = n.$$

Example

- Thus, for $i = 1, 2, \ldots, n$,
$$\frac{y_i}{p_i} = n.$$
- Hence, for $i = 1, 2, \ldots, k$, the maximum likelihood estimator for p_i is
$$\hat{p}_i = \frac{y_i}{n}.$$

Example

- Thus, for $i = 1, 2, \ldots, n$,

$$\frac{y_i}{p_i} = n.$$

- Hence, for $i = 1, 2, \ldots, k$, the maximum likelihood estimator for p_i is

$$\hat{p}_i = \frac{y_i}{n}.$$

- In particular, if X_1, X_2, ..., X_n is a random sample from a Bernoulli distribution with probability of success p, the maximum likelihood estimator for p is

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Sufficient statistics

- Recall: If T is a sufficient statistic for a parameter θ, then $L(\theta) = g(t, \theta)$ for some function g.

Sufficient statistics

- Recall: If T is a sufficient statistic for a parameter θ, then $L(\theta) = g(t, \theta)$ for some function g.
- Hence the value of θ which maximizes L must depend on the sample only through t.

Sufficient statistics

- Recall: If T is a sufficient statistic for a parameter θ, then $L(\theta) = g(t, \theta)$ for some function g.
- Hence the value of θ which maximizes L must depend on the sample only through t.
- That is, the maximum likelihood estimator $\hat{\theta}$ for θ must be a function of T.