Lecture 4: Parameter Estimation

May 10, 2012
GENOME 560, Spring 2012

Su-In Lee, CSE & GS
suinlee@uw.edu

Review: Probability Distributions

Discrete:

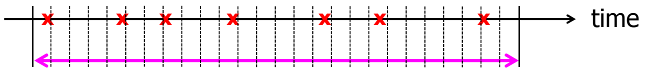
- Binomial distribution
- Hypergeometric distribution
- Poisson distribution

Continuous:

- Uniform distribution
- Exponential distribution
- Gamma distribution
- Normal distribution

Relationship among Binomial, Poisson, and Exponential Distributions

- Let's revisit the e-mail example
  - Assume that e-mails come at a constant rate

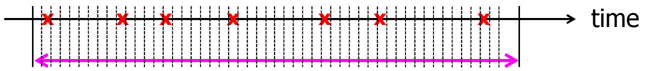


- What is the probability of getting *k* e-mails in a certain time interval?
  - Let's discretize the time interval into subintervals
  - In each subinterval, there are only two possible outcomes – getting an e-mail with prob *p* or not getting an e-mail with prob (1-*p*)
  - We can calculate the probability based on Binomial distribution

$$P\{X=k\}=\binom{n}{k}p^k(1-p)^{n-k}$$

Relationship among Binomial, Poisson, and Exponential Distributions

- Let's revisit the e-mail example
  - Assume that e-mails come at a constant rate

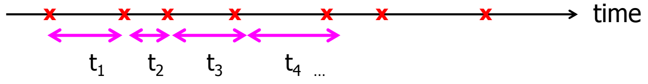


- Let's now make the sub-interval as small as possible
  - We are essentially doing n→∞ and p→0
  - Then, the Poisson distribution becomes equivalent to Binomial distribution

$$P\{X=k\}=\frac{1}{k!}\lambda^k e^{-\lambda}$$

Relationship among Binomial, Poisson, and Exponential Distributions

- Let’s revisit the e-mail example
  - Assume that e-mails come at a constant rate



- Intervals between adjacent events follow the exponential distribution

$$P(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Goals

- Basic concepts on parameter estimation
- Maximum likelihood estimation (MLE)
- Bayesian inference
- Confidence interval

What Is Parameter?

- *Variables* vs. *Parameters*
- According to Bard & Yonathan (1974)[*]
  - … Usually a probabilistic model is designed to explain the relationships that exist among quantities which can be measured independently in an experiment; these are the variables of the model. To formulate these relationships, however, one frequently introduces "constants" which stand for inherent properties of nature. These are the parameters.
- We often denote by θ

[*] Bard, Yonathan (1974). *Nonlinear Parameter Estimation.* New York

---

Examples

- Binomial distribution (coin tossing)
  - X: number of Heads after *n* coin tosses

$$P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}$$

variable parameter $\theta = p$

- Poisson distribution
  - X: number of e-mails within a week

$$P\{X = k\} = e^{-\lambda} \frac{\lambda^k}{k!}$$

variable parameter $\theta = \lambda$

Parameter Estimation

- **Estimator:** Statistic whose calculated value is used to estimate a parameter, θ
- **Estimate:** A particular realization of an estimator, θ
- Types of estimators:
  - **Point estimate:** single number that can be regarded as the most plausible value of θ
  - **Interval estimate:** a range of numbers, called a confidence interval indicating, can be regarded as likely containing the true value of θ

Methods of Point Estimation

- Maximum likelihood estimation (MLE)
- Bayesian inference

Properties of Good Estimators

- What are characteristics of good estimators?
- How well they *explain* the world?
- Say that you flip a coin
  - Let's say that a rv. *X* represents the outcome
  - *p* = probability of getting Head

- If you a flip a coin many times, maybe we can figure out.
  - Realization of the random variable
  - Observation data D = {HHTHHTHTHTHTHTH …}

samples (or instances)

Introduction to Likelihood

- **Before** an experiment is performed, the outcome is unknown. Probability function allows us to predict the probability of any outcome based on **known** parameters:

  P(Data | θ)

- For example, say that we know that probability of getting a Head in a coin toss is *p* = 0.6
  - Then, we can calculate the probability P(Data | θ) for ANY data

$$D_1 = \{HTHHHHTHHHT\} \qquad P(D\,|\,\theta) = p^7(1-p)^3$$

$$D_2 = \{HTH\} \qquad P(D\,|\,\theta) = p^2(1-p)$$

$$D_3 = \{TTTH\} \qquad P(D\,|\,\theta) = p^3(1-p)$$

:

  - If *p* were a different value, the above probabilities would have been different…

Introduction to Likelihood

- **After** an experiment is performed, the outcome is known. Now we talk about the likelihood that a parameter would generate the observed data:

  $$L(\theta : D) = P(\text{Data} \mid \theta)$$

- Estimation proceeds by finding the value of θ that makes the observed data most *likely*
  - *Maximum Likelihood Estimate (MLE)* $\hat{\theta}$

The Coin Example

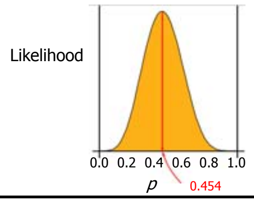
- Let's toss a coin *n* times with probability *p* of heads

- Probability of outcome D = {HHTHTTTTHTTH} is

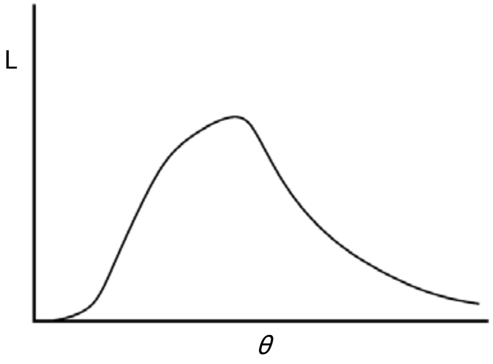
  $$pp(1-p)p(1-p)(1-p)(1-p)(1-p)p(1-p)(1-p)p$$

- The likelihood is then $L = P(D \mid p) = p^5(1-p)^6$

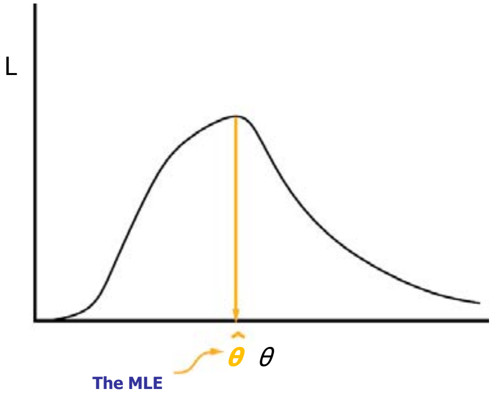
- Plotting *L* against *p* to find its maximum



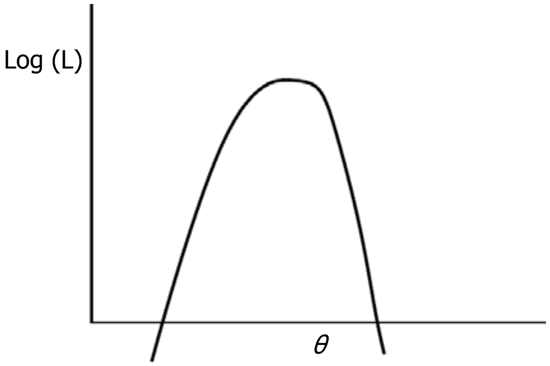
A Likelihood Curve



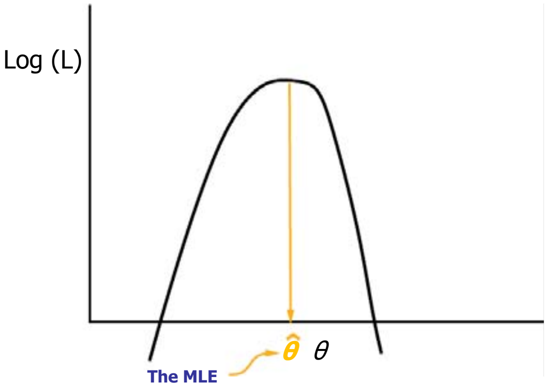
Its Maximum Likelihood Estimate



Better to Plot log (L) than L



Better to Plot log (L) than L



Differentiating to Find the Maximum

- Differentiate the expression for log (*L*) with respect to *p*

$$\log L = \log\left[p^5(1-p)^6\right] = \frac{5}{p} + \frac{6}{1-p}$$

- Equate the derivative to 0

$$\frac{\partial \log L}{\partial p} = \left(\frac{5}{p} - \frac{6}{1-p}\right) = 0$$

$$5 - 11p = 0 \quad \Rightarrow \quad \hat{p} = \frac{5}{11}$$

  - The value of *p* that is at the peak can be found to be $p = 5/11$

Formal Statement of MLE

- Let x[1], x[2], …, x[M] be a sequence of *M* observed values
  - e.g. x[m] = H or x[m] = T in coin tossing
- **Joint probability:**

$$P(D \mid \theta) = P(X = x[1])P(X = x[2]) \cdots P(X = x[M])$$
$$= \prod_{m=1}^{M} P(X = x[m])$$

- **Likelihood** is then:

$$L(\theta : D) = \prod_{m=1}^{M} P(X = x[m])$$

$$\log L(\theta : D) = \sum_{m=1}^{M} \log P(X = x[m])$$

More MLE Example

- Say that we want to estimate the recombination fraction (θ) between locus A and B from 5 heterozygous (AaBb) people. We examined 30 gametes for each and observed 4,3,5,6 and 7 recombinants gametes in the five parents. What is the MLE of the recombination fraction θ?
- Define X = number of recombinant gametes for a single person. Then, the probability that $X = r$ is the following:

$$P\{X=r\}=\binom{n}{r}\theta^{r}(1-\theta)^{n-r}$$

  - where n represents the total number gametes (n = 30)

MLE Example: Specifying Likelihood

- How about all 5 subjects?
- **Probability:**

$$P(X_1=r_1, X_2=r_2,\ldots,X_5=r_5 \mid \theta,n)=P(X_1=r_1 \mid \theta,n)\ \ldots\ P(X_5=r_5 \mid \theta,n)$$

$$=\binom{n}{r_1}\theta^{r_1}(1-\theta)^{n-r_1}\ \cdots\ \binom{n}{r_5}\theta^{r_5}(1-\theta)^{n-r_5}$$

- **Likelihood:**

$$L(\theta : r_1,\ldots,r_5,n)=\prod_{i=1}^{5}\binom{n}{r_i}\theta^{r_i}(1-\theta)^{n-r_i}$$

$$\log L(\theta : r_1,\ldots,r_5,n)=\sum_{i=1}^{5}\left[\log\binom{n}{r_i}+r_i\log\theta+(n-r_i)\log(1-\theta)\right]$$

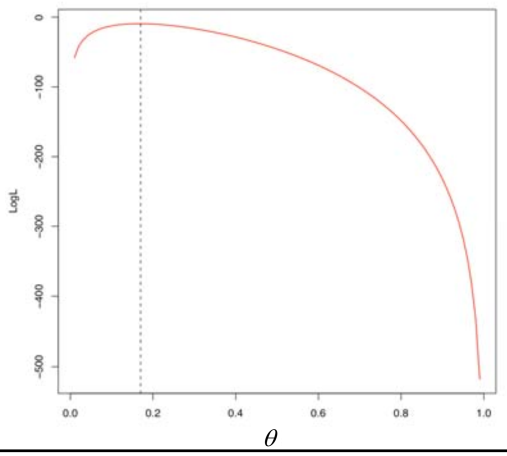
MLE Example: Maximizing the Likelihood

- Again, we want to find p such that log(L) is maximized:

$$\log L(\theta : r_1,\ldots,r_5,n)=\sum_{i=1}^{5}\left[\log\binom{n}{r_i}+r_i\log\theta+(n-r_i)\log(1-\theta)\right]$$

- How?
  - Graphically
    - Draw a plot of log(L) (y-axis) against p (x-axis)
  - Calculus
    - Differentiate log (L) with respect to p and equating the derivative to 0
  - Numerically

MLE Example: Finding the MLE of *p*



Goals

- Basic concepts on parameter estimation
- Maximum likelihood estimation (MLE)
- Bayesian inference

World View According to Bayesian's

- The classic philosophy (frequentist) assumes that parameters are ***fixed*** quantities that we want to estimate as precisely as possible
- Bayesian perspective is different: parameters are ***random variables*** with probabilities assigned to particular values of parameters to reflect the degree of evidence for that value

Bayes' Rule

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}$$

- **Discrete** $P(B)=\sum_{i=1}^{n}P(B|A=a_i)P(A=a_i)$
- **Continuous** $P(B)=\int P(B|A)P(A)dA$

Bayesian Estimation

- In order to make probability statements about θ given some observed data, D, we make use of Bayes' rule

$$P(\theta|D)=\frac{P(\theta)P(D|\theta)}{P(D)}=\frac{P(\theta)P(D|\theta)}{\int P(\theta)P(D|\theta)d\theta}$$

Not a function of θ!

$$\boxed{P(\theta|D)\propto P(\theta)P(D|\theta)}$$

Posterior ∝ Prior × Likelihood

- The prior is the probability of the parameter and represents what was thought ***before*** observing the data
- The likelihood is the probability of the data given the parameter and represents the data now available
- The posterior represents what is thought given both prior information and the data just **observed**

Bayesian Estimation

- **Find $\boldsymbol{\theta}$ such that the posterior P(θ|D) is maximized**

$$P(\theta \mid D) = \frac{P(\theta)P(D \mid \theta)}{P(D)} = \frac{P(\theta)P(D \mid \theta)}{\int P(\theta)P(D \mid \theta)d\theta}$$

Not a function of θ!

$$\boxed{P(\theta \mid D) \propto P(\theta)P(D \mid \theta)}$$

Posterior ∝ **Prior** × **Likelihood**

- The **prior** is the probability of the parameter and represents what was thought ***before*** observing the data
- The **likelihood** is the probability of the data given the parameter and represents the data now available
- The **posterior** represents what is thought given both prior information and the data just **observed**

Simple Example

- Say that we want to estimate the recombination fraction (θ) between locus A and B from 5 heterozygous (AaBb) people. We examined 30 gametes for each and observed 4,3,5,6 and 7 recombinants gametes in the five parents. What is the MLE of the recombination fraction θ?
- Let's simplify and ask what the recombination fraction (θ) is for subject # 3, who had 5 observed recombinant gametes.

Specifying The Posterior Density

$$P(\theta \mid D) = P(\theta \mid n = 30, r = 5) = \frac{P(\theta)P(r = 5 \mid \theta, n = 30)}{\int_0^{0.5} P(r = 5 \mid \theta, n = 30)P(\theta)d\theta}$$

- **Prior** $P(\theta) = \text{uniform}[0,0.5] = 0.5$
- **Likelihood** $P(r = 5 \mid \theta, n = 30) = \binom{30}{5}\theta^5(1-\theta)\theta^{30-5}$
- **Normalizing constant** $\int_0^{0.5} P(r = 5 \mid \theta, n = 30)P(\theta)d\theta$

$$= 0.5 \cdot \binom{30}{5}\int_0^{0.5}\theta^5(1-\theta)^{25}d\theta \approx 6531$$

Specifying The Posterior Density

$$P(\theta \mid D) = P(\theta \mid n = 30, r = 5) = \frac{P(\theta)P(r = 5 \mid \theta, n = 30)}{\int_0^{0.5} P(r = 5 \mid \theta, n = 30)P(\theta)d\theta}$$

$$= \frac{0.5 \cdot \binom{30}{5}\theta^5(1-\theta)^{25}}{6531}$$