

Estimation

Session 4

Module 1 Probability & Statistical Inference

The Summer Institutes

DEPARTMENT OF BIostatISTICS

SCHOOL OF PUBLIC HEALTH

UNIVERSITY *of* WASHINGTON



Overview

Probability/statistical models depend on parameters

Binomial depends on probability of success π .

Normal depends on mean μ , standard deviation σ .

Parameters are properties of the “population” and are typically unknown.

The process of taking a sample of data to make inferences about these parameters is referred to as **estimation**.

There are a number of different estimation methods ... we will study two estimation methods:

1. **Maximum likelihood (ML)**
2. **Bayes**

Maximum Likelihood

Fisher (1922) invented this general method.

Problem Unknown model parameters θ

Set-up Write the probability of the data \mathbf{Y} in terms of the model parameter: $P(\mathbf{Y} | \theta)$

Solution Estimate θ as the value that makes the data \mathbf{Y} look most likely to occur. This estimate is denoted by $\hat{\theta}$.

 The estimator $\hat{\theta}$ is called the **maximum likelihood estimator (MLE)**.

Maximum Likelihood Estimate

Suppose a man is known to have transmitted allele A1 to his child at a locus that has only two alleles: A1 and A2.

What is the maximum likelihood estimate of the man's genotype?

Solution Let X represent the data (paternal allele in the child) and let θ represent the parameter (man's genotype):

$$X = A1$$

$$\theta = \{A1A1, A1A2, A2A2\}$$

The probability function is based on $P(X | \theta)$

$$P(X = A1 | \theta = A1A1) = 1$$

$$P(X = A1 | \theta = A1A2) = 0.5$$

$$P(X = A1 | \theta = A2A2) = 0$$

Therefore, the MLE is $\hat{\theta} = A1A1$

Maximum Likelihood

Suppose we have a sample of 20 gametes in which the number of recombinants (Z) and nonrecombinants ($N-Z$) for two loci can be counted. *Use these data to estimate the recombination fraction (π) between the two loci.*

Solution The probability of the data can be modeled using a binomial distribution. The **probability distribution function** is:

$$P(Z | \pi) = \binom{20}{Z} \pi^Z (1 - \pi)^{20-Z}$$

where Z is the variable and **π is fixed**.

The **likelihood function** is the same function:

$$L(\pi | Z) = \binom{20}{Z} \pi^Z (1 - \pi)^{20-Z}$$

except now π is the variable and **Z is fixed**.

Maximum Likelihood

$$\text{Recall: } L(\pi | Z) = \binom{20}{Z} \pi^Z (1 - \pi)^{20-Z}$$

- We can use calculus to find the **maximum of the (log) likelihood function**.

$$\begin{aligned}\frac{d \ln L}{d\pi} &= 0 \\ \frac{d}{d\pi} [Z \ln \pi + (20 - Z) \ln(1 - \pi)] &= 0 \\ \frac{Z}{\pi} - \frac{20 - Z}{1 - \pi} &= 0 \\ \hat{\pi} &= \frac{Z}{20}\end{aligned}$$

- Not surprisingly, the likelihood in this example is maximized at the observed proportion, $3/20$.
- Sometimes the MLE has a simple closed form. In more complex problems, numerical optimization is used.
 - Computers can find these maximum values!

Maximum Likelihood

General Notation

$L(\theta)$ = **likelihood** as a function of the parameter θ

$\ell(\theta) = \ln(L(\theta))$, the **log likelihood**

- > Usually more convenient to work with analytically and numerically

$S(\theta) = d\ell(\theta)/d\theta$, the **score**

- > Set to zero and solve for θ to calculate the MLE

$I(\theta) = -d^2\ell(\theta)/d\theta^2$, the **information**

- > Inverse gives variance of $\hat{\theta}$

$\text{Var}(\hat{\theta}) = E[I(\theta)]^{-1}$ (in most cases)

Bayes Estimation

Recall Bayes theorem:
(written in terms of data X and parameter θ)

$$P(\theta | X) = \frac{P(X | \theta)P(\theta)}{\int_{\theta} P(X | \theta)P(\theta)}$$

Notice the change in perspective - **θ is now treated as a random variable** instead of a fixed number.

- > $P(X | \theta)$ is the **likelihood function**, as before.
- > $P(\theta)$ is called the **prior distribution** of θ .
- > $P(\theta | X)$ is called the **posterior distribution** of θ and is used for estimation

Based on $P(\theta | X)$ we can define a number of possible estimators of θ . A commonly used estimate is the maximum a posteriori (MAP) estimate:

$$\hat{\theta}_{MAP} = \max_{\theta} P(\theta | X)$$

We can also use $P(\theta | X)$ to define “credible” intervals for θ .

Comments

- > The Bayesian procedure provides a convenient way of **combining external information** or previous data (through the prior distribution) **with the current data** (through the likelihood) to create a new estimate.
- > As N increases, the data (through the likelihood) **overwhelms** the prior and the Bayes estimator typically converges to the MLE.
- > Controversy arises when $P(\theta)$ is used to incorporate subjective beliefs or opinions.
- > If the prior distribution $P(\theta)$ is simply that θ is uniformly distributed over all possible values, this is called an **uninformative prior**, and the MAP is the same as the MLE.

Bayes Estimation

Suppose a man is known to have transmitted allele A1 to his child at a locus that has only two alleles: A1 and A2.

What is the maximum likelihood estimate of the man's genotype?

Solution Let X represent the data (paternal allele in the child) and let θ represent the parameter (man's genotype):

$$X = A1$$

$$\theta = \{A1A1, A1A2, A2A2\}$$

The probability function is based on $P(X | \theta) \dots$

$$P(X = A1 | \theta = A1A1) = 1$$

$$P(X = A1 | \theta = A1A2) = 0.5$$

$$P(X = A1 | \theta = A2A2) = 0$$

Therefore, the MLE is $\hat{\theta} = A1A1$

Bayes Estimation

Suppose that we know that the frequency of the A1 allele in the general population is only 1%. Assuming Hardy-Weinberg Equilibrium we have

$$P(\theta = A1A1) = 0.01 * 0.01 = 0.0001$$

$$P(\theta = A1A2 \text{ or } A2A1) = 2 * 0.01 * (1-0.01) = 0.0198$$

$$P(\theta = A2A2) = (1-0.01) * (1-0.01) = 0.9801$$

Also

$$P(X) = \sum_{\theta} P(X = A1 | \theta)P(\theta) = 0.01$$

This leads to the **posterior distribution**

$$P(\theta = A1A1 | X = A1) = 0.01$$

$$P(\theta = A1A2 | X = A1) = 0.99$$

$$P(\theta = A2A2 | X = A1) = 0$$

Therefore the Bayesian
MAP estimator is

$$\theta = A1A2$$

Summary

Maximum likelihood is a method of estimating parameters from data

- > ML requires you to write a probability model for the data
- > MLEs may be found analytically or numerically
- > (Inverse of the negative of the) second derivative of the log-likelihood gives variance of estimates

Bayesian procedures allow us to incorporate additional information about the parameters in the form of prior data, external information, or personal beliefs.

Break #1

**Pause the video,
take a break, stretch,
then review relevant exercises
from worksheet.**

Afterwards, continue on!

