Biostatistics Preparatory Course: Methods and Computing

Lecture 9

Maximum Likelihood & the Bootstrap

Overview: Maximum Likelihood Estimation

Consider estimating a parameter θ given a sample of data, $\{X_1, \ldots, X_n\}$

What is maximum likelihood estimation?

- A statistical method that estimates θ as the value that maximizes the likelihood of obtaining the observed data
- That is, the maximum likelihood estimator (MLE) provides the greatest amount of agreement between the selected model and the data

Overview: Maximum Likelihood Estimation

What is the likelihood function?

- In math - $\mathcal{L}(\theta) = f(x_1, \ldots, x_n|\theta)$ where $f(\cdot)$ denotes the joint density of the data
- In words - the function that dictates the probability (relative frequency) of observing the data as a function of θ

The definition of the MLE is:

$$\hat{\theta}_{MLE} = \arg\max_\theta \mathcal{L}(\theta)$$

Simple Setting

We will focus on the setting of *iid* observations, that is,

$$\{X_1, \ldots, X_n\} \quad \text{is a } \textit{simple random sample}$$

- The likelihood then simplifies to

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

- In practice, we typically maximize the log of the likelihood:

$$\ell(\theta) = \log\{\mathcal{L}(\theta)\} = \sum_{i=1}^{n} \log\{f(x_i|\theta)\}$$

  since taking the derivative of a sum is typically easier than a product and the likelihood can be very small for large n (a computational issue)

Why is maximum likelihood estimation so popular?

- Provides a unified framework for estimation
- Under mild regularity conditions, MLEs are:
  1. consistent $\rightarrow$ converge to the true value in probability as $n \rightarrow \infty$, i.e.

  $$\lim_{n\rightarrow\infty} P(|\hat{\theta} - \theta| \leq \epsilon) = 1 \quad \forall\epsilon > 0$$

  2. asymptotically normal $\rightarrow$ $\sqrt{n}(\hat{\theta} - \theta) \sim N(0, \sigma^2)$ for large n
  3. asymptotically efficient $\rightarrow$ achieve the lowest variance for large n
  4. invariant $\rightarrow$ if $\hat{\theta}$ is the MLE for θ then $g(\hat{\theta})$ is the MLE for $g(\theta)$
- Many algorithms exist for maximum likelihood estimation

Steps to find the MLE

1. Write out the likelihood

$$\mathcal{L}(\theta) = f(x_1, \ldots, x_n|\theta)$$

2. Simplify the log likelihood

$$\ell(\theta) = \log\{\mathcal{L}(\theta)\}$$

3. Take the derivative of $\ell(\theta)$ with respect to the parameter of interest, θ
4. Set $= 0$
5. Solve for θ (this is your $\hat{\theta}_{MLE}$)
6. Check that $\hat{\theta}_{MLE}$ is a maximum $\left(\frac{\partial^2}{\partial\theta^2}\ell(\theta) < 0\right)$

MLE Exercises

MLE Exercises

1. Suppose we have an iid sample $\{X_1, \ldots, X_{100}\}$ with $X_i \sim Ber(p)$. Find the MLE for p. Recall that the density for a Bernoulli random variable can be written as:

$$p^{X_i}(1-p)^{1-X_i}$$

MLE Exercises

1. Suppose we have an iid sample $\{X_1, \ldots, X_{100}\}$ with $X_i \sim Ber(p)$. Find the MLE for p. Recall that the density for a Bernoulli random variable can be written as:

$$p^{X_i}(1-p)^{1-X_i}$$

2. Suppose we have an iid sample $\{X_1, \ldots, X_n\}$ with $X_i \sim N(\mu, \sigma^2)$ Find the MLE for μ. Recall that the density for a normal random variable can be written as:

$$\frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{1}{2\sigma^2}(X_i - \mu)^2)$$

MLE Exercises in R

We are going to use R to derive the MLE in more complex cases.

- In the previous two examples, we found a closed-form solution (MLE) for our parameters
- Sometimes, there is no closed-form solution, so we need to use optimization methods to estimate our parameter of interest

The `optim` function

- General-purpose optimization that implements various methods
- It will find the values of some parameters that *minimizes* some function
- You need to specify...
  - The parameters that you want to estimate
  - The function (in our case, the negative log-likelihood; why negative?)
  - The method (I typically use "BFGS")
  - Starting values for your parameters (use random numbers)
  - Other values that you need to pass into your function

MLE Exercises

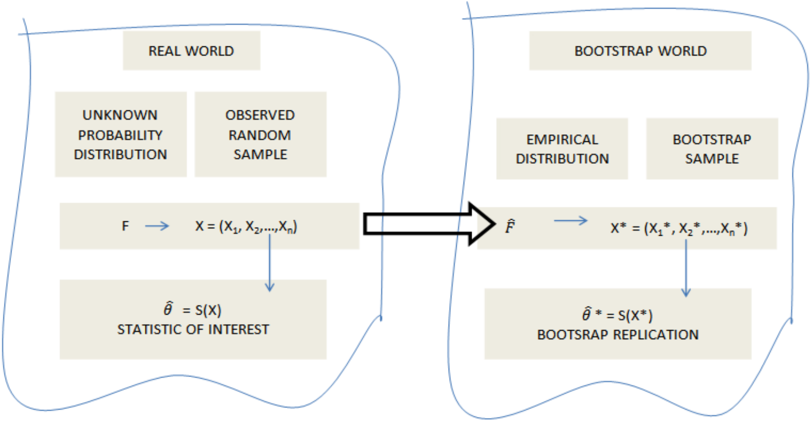
The Bootstrap

- What is the bootstrap?
  - A widely applicable, computer intensive resampling method used to compute standard errors, confidence intervals, and significance tests
- Why is it important?
  - The exact sampling distribution of an estimator can be difficult to obtain
  - Asymptotic expansions are sometimes easier but expressions for standard errors based on large sample theory may not perform well in finite samples

Motivating Analogy

The bootstrap samples should relate to the original sample just as the original sample relates to the population

Overview: The Bootstrap Principle



Without additional information, the sample contains all we know about the underlying distribution so resampling the sample is the best approximation to sampling from the true distribution

The Bootstrap Principle

Suppose $X = \{X_1, \ldots, X_n\}$ is a sample used to estimate some parameter $\theta = T(P)$ of the underlying distribution P. To make inference on θ, we are interested in the properties of our estimator $\hat{\theta} = S(X)$ for θ.

- If we knew P, we could obtain $\{X^b | b = 1, \ldots B\}$ from P and use Monte-Carlo to estimate the sampling distribution of $\hat{\theta}$ (sound familiar?)
- We don't so we do the next best thing and resample from original sample, i.e. *the empirical distribution*, $\hat{P}$
  - We expect the empirical distribution to estimate the underlying distribution well by the *Glivenko-Cantelli Theorem*

Bootstrap procedure

Goal: Find the standard error and confidence intervals for some $\hat{\theta} = S(\mathbf{D})$ where $\mathbf{D}$ encodes our observed data.

- Select B independent bootstrap resamples $\mathbf{D}(b)$, each consisting of N data values drawn with replacement from the data.
- Compute the estimates from each bootstrap resample

$$\hat{\theta}^*(b) = S(\mathbf{D}^*(b)) \quad b = 1, ..., B$$

- Estimate the standard error $se(\hat{\theta})$ by the sample standard deviation of the B replications of $\hat{\theta}^*(b)$
- Estimate the confidence interval by finding the $100(1-\alpha)$ percentile bootstrap CI,

$$(\hat{\theta}_L, \hat{\theta}_U) = (\hat{\theta^*}^{\alpha/2}, \hat{\theta^*}^{1-\alpha/2})$$

Boostrap exercise