Applied Statistics

Sofia Triantafillou

sof.triantafillou@gmail.com

University of Crete
Department of Mathematics and Applied Mathematics

Lecture Summary

- Statistical Inference (Chapter 7.1).
- Frequentist Inference (Chapter 7.1).
- Maximum Likelihood Estimation (Chapter 7.5).
- Properties of MLE estimators (Chapter 7.6).

Statistical Inference

In our last lecture, we saw several distributions with probability functions: $f(x|\theta)$, where θ for parameters, taking values in some parameter space Ω.

Examples

- The height of a student is approximately normal with mean θ and some known variance.
- The number of people that have a disease out of a group of N people follows the Binomial (N, θ) distribution.
- The lifetime of an electronic component follows an exponential distribution with rate θ.

In practice, we do not know θ.

Statistical Inference

What can we *infer* about θ given the observed data?
Assuming that we observe random variables $X_1, \ldots, X_n$ following some distribution with parameter θ, what conclusions can we draw about parameter θ?

Statistical Inference Tasks

- Prediction.
- Estimation.
- Decision problems (e.g., Hypothesis testing).
- Experimental Design.

Formalizing Statistical Inference

Statistical Model

- An identification of random variables of interest
- A specification of a distribution or a family of possible joint distributions for these variables.

Statistical Inference

A procedure that produces a probabilistic statement for some or all parts of a statistical model.

Formalizing Statistical Inference

Statistic

Suppose that the observable random variables of interest are $X_1, \ldots, X_n$. Let r be an arbitrary real-valued function of n real variables. Then the random variable $T = r(X_1, \ldots, X_n)$ is a statistic.

Examples:

- The mean of $X_1, \ldots, X_n$: $T = \frac{1}{n} \sum_i X_i$.
- The maximum of $X_1, \ldots, X_n$: $T = max\{X_1, \ldots, X_n\}$.
- A constant, e.g., T=3.
- Absolute difference of the mean from 175:
  $T = |\frac{1}{n} \sum_i X_i - 175|$.

Estimation

Estimate (predict) the unknown parameter θ. E.g. We estimated the prevalence of the disease as $\hat{\theta}$.

One of the most common tasks in statistical inference. Two schools:

- Bayesian inference: Treat θ as a random variable.
- Frequentist inference: Treat θ as a number.

In this course we will focus on frequentist approaches.

Likelihood

Likelihood function: $f(\mathbf{x}|\theta), \mathcal{L}(\mathbf{x};\theta)$

The likelihood function (often simply called the likelihood) describes the joint probability of the observed data $x_1, \ldots, x_n$ as a function of the parameters θ of the chosen statistical model.

- Assume you test 10 people for a disease, and you get the following result (0:negative, 1:positive)
- $x_1 = 0, x_2 = 0, x3 = 0, x_4 = 1, x_5 = 1, x_6 = 0, x_7 = 0, x_8 = 1, x_9 = 0, x_{10} = 0$
- Let X_i be the outcome of the i-th patient:
- $X_i \sim Bernoulli(\theta)$.
- $f(x_1, \ldots, x_{10}|\theta) =$, for $\theta = 0.2$, $\theta = 0.8$.

Maximum Likelihood Estimation

Maximum Likelihood Estimator

For any given observations $\mathbf{x}$ we pick the $\theta \in \Omega$ that maximizes $f(\mathbf{x}|\theta)$.

Maximum Likelihood Estimate

For given data $\mathbf{X} = \mathbf{x}$, the maximum likelihood estimate (MLE) will be a function of θ.

- Estimator $\hat{\Theta}(\mathbf{X})$ is a function mapping the random sample $\mathbf{X}$ to the parameter space.
- Estimate $\hat{\theta}$ is a value of the estimator for a particular sample.
- Sometimes $\hat{\theta}$ is to denote both estimator and estimate.

Log likelihood

For numerical reasons (e.g., avoid multiplying numbers in [0,1]), it is often easier to maximize the log likelihood $\mathcal{LL}(\theta) = log(f(x|\theta))$.

Log properties.

Logarithm is monotonic, so $argmax_x f(x) = argmax_\theta log(f(x))$.

$$log(ab) = log(a) + log(b)$$

$$log(\prod_i f(x)) = \sum_i log(f(x))$$

$$log(a^n) = nlog(a)$$

MLE estimation

- Assume you test 10 people for a disease, and you get the following result (0:negative, 1:positive)
- $x_1 = 0, x_2 = 0, x3 = 0, x_4 = 1, x_5 = 1, x_6 = 0, x_7 = 0, x_8 = 1, x_9 = 0, x_{10} = 0$
- Let X_i be the outcome of the i-th patient:
- $X_i \sim Bernoulli(\theta)$.
- Find the MLE for θ.

Maximizing functions

How do we maximize a function?

A point $(x, f(x))$ for which:

$$\frac{df}{dx} = 0, \qquad \frac{d^2 f}{dx^2} < 0$$

Corresponds to a maximum.

- May not be unique.
- For curved exponential families, the log likelihood is concave.
  - Unique maximum!

Properties of Estimators

Theorem (Invariance)

If $\hat{\theta}$ is the MLE of θ and $g(\theta)$ is a function of θ then $g(\hat{\theta})$ is the MLE of $g(\theta)$.

Example

If $\hat{p}$ is the MLE for $\hat{p}$ in the $Binomial(n, p)$, then the MLE for the odds $\frac{p}{1-p}$ is $\frac{\hat{p}}{1-\hat{p}}$.

Properties of Estimators

Definition (Consistency)

A sequence of estimators that converges in probability to the unknown value of the parameter being estimated, as $n \to \infty$ is called a consistent sequence of estimators.

Consistency of MLEs

Under some conditions (typically satisfied in practical problems) the sequence of M.L.E.'s is a consistent sequence of estimators.

Maximum Likelihood Estimation

Example: M.L.E. of normal with known variance

- Assume you observe the heights of n students.
- Let X_i be the height of the i-th student you picked.
- $X_i \sim Norm(\theta, 9)$ (we assume we know the variance).
- We get data $x_1, \ldots, x_n$.
- Find the MLE of θ (as a function of the data).