*Math 408 - Mathematical Statistics*

# Lecture 23a. Fundamental Concepts of Modern Statistical Inference

March 29, 2013

# Agenda

- Statistical Models
- Point Estimates
- Confidence Intervals
- Hypothesis Testing
- Summary

# Statistical Inference

Statistical inference, or "learning", is the process of using data to infer the distribution that generated the data. The basic statistical inference problem is the following:

## Basic Problem

*We observe $X_1, \ldots, X_n \sim \pi$. We want to infer (or estimate, or learn) $\pi$ or some features of $\pi$ such as its mean.*

## Definition

A **statistical model** is a set of distributions or a set of densities $\mathcal{F}$.

- A **parametric model** is a set $\mathcal{F}$ that can be parameterized by a finite number of parameters.
- A **nonparametric model** is a set $\mathcal{F}$ that cannot be parameterized by a finite set of parameters.

# Examples of Statistical Models

Examples:

- If we assume that the data come from a normal distribution, then the model is
$$\mathcal{F} = \left\{ \pi(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad \mu, \sigma^2 \in \mathbb{R} \right\}$$
This is a two-parameter model. In $\pi(x|\mu,\sigma^2)$, $x$ is a value of the random variable, whereas $\mu$ and $\sigma^2$ are parameters.
- A nonparametric model:
$$\mathcal{F}_{\mathrm{all}} = \{\text{all PDFs}\}$$

We will focus on parametric models.
In general, a parametric model takes the form

$$\boxed{\mathcal{F} = \{\pi(x|\theta), \ \ \theta \in \Theta\}}$$

where $\theta$ is an unknown parameter and $\Theta$ is the parameter space.

Remark: $\theta$ can be a vector, for instance, $\theta = (\mu, \sigma^2)$

# Point Estimation

Given a parametric model, $\mathcal{F} = \{\pi(x|\theta), \ \ \theta \in \Theta\}$, the problem of inference is then to estimate (to learn) the parameter $\theta$ from the data.

Almost all problems in statistical inference can be identified as being one of three types: **point estimates**, **confidence intervals**, and **hypothesis testing**.

- Point Estimation refers to providing a single "best guess."
  Suppose $X_1, \ldots, X_n \sim \pi(x|\theta)$, where $\pi(x|\theta) \in \mathcal{F}$.
  A point estimator $\hat{\theta}_n$ of a parameter $\theta$ is some function of $X_1, \ldots, X_n$:

  $$\hat{\theta}_n = f(X_1, \ldots, X_n)$$

  Remember: $\theta$ is fixed but unknown, $\hat{\theta}_n$ is random since depends on $X_1, \ldots, X_n$. We say that $\hat{\theta}_n$ is unbiased if

  $$\mathbb{E}[\hat{\theta}_n] = \theta$$

# Confidence Intervals and Hypothesis Testing

- A $100(1-\alpha)\%$ Confidence Interval for a parameter $\theta$ is a random interval $I_n = (a, b)$ where $a = a(X_1, \ldots, X_n)$ and $b = b(X_1, \ldots, X_n)$ such that

$$\mathbb{P}(\theta \in I_n) = 1 - \alpha$$

  In words: $(a, b)$ traps $\theta$ with probability $1-\alpha$.
  $(1-\alpha)$ is called coverage of the confidence interval.
  In practice, $\alpha = 0.05$ is often used.

- In Hypothesis Testing, we start with some default theory, called a null hypothesis, and we ask if the data provide sufficient evidence to reject the theory. If not, we accept the null hypothesis.

  Example:

  $X_1, \ldots, X_n \sim \mathrm{Bernoulli}(p)$: $n$ independent coin flips.
  We want to test if the coin is fair $\Rightarrow$ the null hypothesis $H_0 : p = 1/2$
  The alternative hypothesis is then: $H_1 : p \neq 1/2$
  It seems reasonable to reject $H_0$ if

$$\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{2} \right| \quad \text{is large}$$

# Summary

- A parametric model is a set $\mathcal{F}$ that can be parameterized by a finite number of parameters.
  - General parametric model:

$$\mathcal{F} = \{\pi(x|\theta), \quad \theta \in \Theta\}$$

- A nonparametric model is a set $\mathcal{F}$ that cannot be parameterized by a finite set of parameters.
- Almost all problems in statistical inference can be identified as being one of three types:
  - Point Estimates
  - Confidence Intervals
  - Hypothesis Testing