January 05, 2014.

$\Rightarrow$ Theoretical statistics

Compare different statistical methods.

$\Rightarrow$ Buy the textbook: Essential of statistical Inference
chapter 1 to chapter 8

$\Rightarrow$ Lab: 3:30 PM to 4:50 PM

One week one student

$\Rightarrow$ Lab question $\rightarrow$ term-test $\rightarrow$ Final exam

plus presentation skill

$\Rightarrow$ two term test (subjective) + (attendance)

Consider only the higher mark

Final time: April

$\Rightarrow$ what is statistical inference?

$\Rightarrow$ observations on n units

$x_1, x_2, \cdots, x_n$ (x_i may be vector)

$\Rightarrow$ population Distribution

regard $x_1, x_2, \cdots, x_n$ as a realization

of random variables $x_1, \cdots, x_n$.

$\Rightarrow$ For simplicity we assume that

$$x_1, x_2, \cdots x_n \overset{iid}{\sim} f(x)$$

$f(x)$ is called a population distribution

Non iid example: spatial data (spatial correlation)
time series.

$\Rightarrow$ Furthermore we assume $f(x)$ is of known

analytic form, but involves unknow parameters

Example $\quad f(x, \theta) = \dfrac{1}{\sqrt{2\pi}\,\sigma} \; e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$

parametric statistics.

$\Rightarrow$ parameter space

$\theta \in \Theta$

Θ: all possible values of θ

Example: $\quad x_1, \cdots, x_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$$\sigma \in [0, +\infty)$$

$\Rightarrow$ statistical inference:

We want to learn some aspects of

unknow θ from the observations $x_1, \cdots x_n$

(from sample to population)

probability:

θ is known $\longrightarrow$ $f(x_1, \cdots, x_n \mid \theta)$

Statistics :

θ is unknow $\longleftarrow$ $x_1, \cdots, x_n$

$\Rightarrow$ Different types of inference :

(1) point estimation $\hat{\theta}(x_1, \cdots, x_n)$

$\hat{\theta}$ is a single number to capture θ

(2) interval estimation

$\theta \in (L(x_1, \cdots, x_n), M(x_1, \cdots, x_n))$

The true parameter is within this interval

(3) Hypothesis testing

$\theta = \theta_0$ Vs $\theta \neq \theta_0$

$\theta \in \theta_0$ Vs $\theta \in \theta_1$

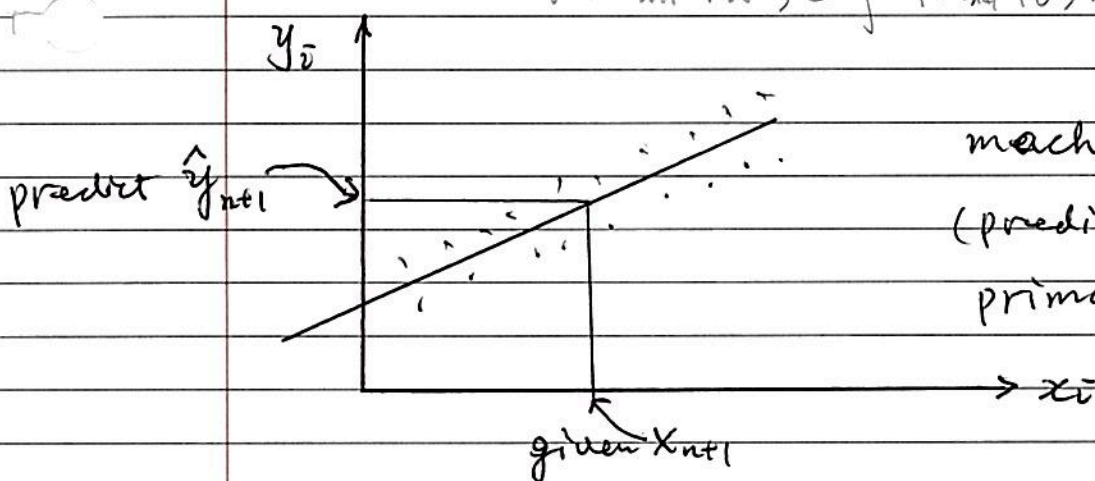$\theta < \theta_0$ Vs $\theta > \theta_0$

(4) Predictive inference $x_1, x_2, \cdots, x_n \xrightarrow{\text{predict}} x_{n+1}$

Example : $f(y_0 \mid x_i, \hat{\theta})$

from $(x_1, y_1), \cdots, (x_n, y_n)$

We want to predict y_{n+1} given x_{n+1}

$$f(x_{n+1} \mid X) \overset{?}{=} \int f(x_{n+1} \mid \theta) \pi(\theta \mid X) \, dx$$

machine learning
(prediction is the
primary goal)

$\Rightarrow$ how to do : standard paradigms for

stat inference :

- Bayesian inference

  prior distribution $\theta \sim \Pi(\theta)$

  data model $D (\theta \sim f(d \mid \theta))$

  posterior : $f(\theta \mid D = d) = \dfrac{\pi(\theta) f(d \mid \theta)}{\int \pi(\theta) f(d \mid \theta) \, d\theta}$

- In 1920, fisher devises a system of

  inference approaches that we indep of π

  repeated sampling principle :

  $\hat{\theta}(x_1, \cdots, x_n)$ has a sampling distribution