

# CH 11: Nonparametric Methods

- Estimation of CDF

Let  $X_1, X_2, \dots, X_n$  iid  $\sim F$ , we need to estimate  $F$ .

Empirical distri.:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I\{X_i < x\}.$$



Home Page

Title Page



Page 1 of 100

Go Back

Full Screen

Close

Quit

- Step function

(1).  $EF_n(x) = F(x),$

$$\text{Var}(F_n(x)) = F(x)(1 - F(x))/n$$

(2). If  $F(\cdot)$  is continuous, then

$$\sup_{x \in R^1} |F_n(x) - F(x)| \longrightarrow 0 \text{ a.s.}$$

Home Page

Title Page

◀ ▶

◀ ▶

Page 2 of 100

Go Back

Full Screen

Close

Quit

(3). If  $0 < F(\cdot) < 1$ , then

$$\frac{\sqrt{n}[F_n(x) - F(x)]}{\sqrt{F(x)(1 - F(x))}} \longrightarrow^d. N(0, 1)$$

(4). If  $F(\cdot)$  is continuous, then

$$\sqrt{n} \sup_{x \in R^1} |F_n(x) - F(x)| \longrightarrow^d. X_0,$$

Home Page

Title Page

◀ ▶

◀ ▶

Page 3 of 100

Go Back

Full Screen

Close

Quit

where  $X_0 \sim Q$  with

$$Q(x) = \sum_{j=-\infty}^{+\infty} (-1)^j \exp\{-2j^2 x^2\}$$

for  $x > 0$ . It can be used to test

$$H_0 : F = F_0, \quad H_1 : F \neq F_0$$

Home Page

Title Page

◀ ▶

◀ ▶

Page 4 of 100

Go Back

Full Screen

Close

Quit

- Quantile estimation

the  $p$ th quantile of  $F$  ( $0 < p < 1$ ):

$$x_p = \inf\{x : F(x) \geq p\}$$

We use the sample order statistic  $X_{([np]+1)}$  to estimate  $x_p$ .

- If  $x_p$  is unique for  $F(x) = p$ , then  $X_{([np]+1)}$  is consistent and

Home Page

Title Page

◀ ▶

◀ ▶

Page 5 of 100

Go Back

Full Screen

Close

Quit

$$\frac{\sqrt{n}(X_{([np]+1)} - x_p)}{\sqrt{p(1-p)}f(x_p)} \longrightarrow^d N(0, 1)$$

for  $f(x_p) > 0$ .

Home Page

Title Page

◀ ▶

◀ ▶

Page 6 of 100

Go Back

Full Screen

Close

Quit

- Density estimation:

$F_n$  estimates  $F$ ,

$F'_n$  estimates  $f = F'$ ?

(1). Kernel estimation

When  $h > 0$  is small, then

$$f(x) \approx \frac{F(x+h) - F(x-h)}{2h}.$$

Home Page

Title Page

◀ ▶

◀ ▶

Page 7 of 100

Go Back

Full Screen

Close

Quit

Take

$$\begin{aligned}\hat{f}_n(x) &= \frac{F_n(x+h) - F_n(x-h)}{2h} \\ &= \frac{1}{nh} \sum_{i=1}^n K_0((X_i - x)/h),\end{aligned}$$

where  $K_0(x) = 0.5 * I\{|x| \leq 1\}$ .

Rosenblatt (1956).

Home Page

Title Page

◀ ▶

◀ ▶

Page 8 of 100

Go Back

Full Screen

Close

Quit



In general

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^n K((X_i - x)/h),$$

$K$  may be taken as a density.

Home Page

Title Page

◀ ▶

◀ ▶

Page 9 of 100

Go Back

Full Screen

Close

Quit

(1). If  $f(x)$  is continuous at  $x$ ,  $K$  is a bounded density and  $\lim_{x \rightarrow \infty} |x|K(x) = 0$ ,  $h \rightarrow 0$ ,  $nh \rightarrow \infty$ , then

$$E \hat{f}_n(x) \longrightarrow f(x)$$

$$f_n(x) \longrightarrow f(x) \text{ a.s.}$$

Home Page

Title Page

◀ ▶

◀ ▶

Page 10 of 100

Go Back

Full Screen

Close

Quit

and

$$\sqrt{nh}(f_n(x) - E\hat{f}_n(x)) \longrightarrow^d N(0, f(x) \int K^2(x)dx)$$

Home Page

Title Page

◀ ▶

◀ ▶

Page 11 of 100

Go Back

Full Screen

Close

Quit

(2). If  $f(x)$  is continuous uniformly,  
 $K$  has bounded variation,  $h \rightarrow 0$ ,  
 $\sum_{n=1}^{\infty} \exp\{-cnh_n^2\} < +\infty$  ( $c > 0$ ), then

$$\sup_x |f_n(x) - f(x)| \longrightarrow 0 \text{ a.s.}$$

Home Page

Title Page

◀ ▶

◀ ▶

Page 12 of 100

Go Back

Full Screen

Close

Quit

- Least neighborhood method:

Given  $a_n \in Z^+$ , let

$$\alpha_n(x) = \min\left\{t : \sum_{i=1}^n I_{(x-t, x+t]}(X_i) \geq a_n\right\}$$

$$f_n^*(x) = \frac{a_n}{2n\alpha_n(x)}.$$

Home Page

Title Page

◀ ▶

◀ ▶

Page 13 of 100

Go Back

Full Screen

Close

Quit

- Test  $H_0 : F = G, \quad H_1 : F \neq G$

Let

$$D_{m,n} = \sup_x |F_m(x) - G_n(x)|$$

then, under  $H_0$ ,

$$\sqrt{\frac{nm}{m+n}} D_{m,n} \xrightarrow{d.} X_0.$$

Home Page

Title Page

◀ ▶

◀ ▶

Page 14 of 100

Go Back

Full Screen

Close

Quit

- Skewness and Kurtosis Test for Normality
- Wilk Test for Normality

$$W = \frac{\sum_{k=1}^{\lfloor n/2 \rfloor} \alpha_k (X_{(n+1-k)} - X_{(k)})}{\sum_{i=1}^n (X_{(i)} - \bar{X})^2}.$$

$\alpha_k$  are some constants.

Home Page

Title Page

◀ ▶

◀ ▶

Page 15 of 100

Go Back

Full Screen

Close

Quit

- $R^2$  and  $D$  Test for Normality
- Count Statistic
- Rank Statistic
- $U$ -Statistic

Home Page

Title Page

◀ ▶

◀ ▶

Page 16 of 100

Go Back

Full Screen

Close

Quit



- Regression function estimation

$$m(x) = E(Y|X = x) = \frac{\int y f(x, y) dy}{f(x)}$$

and

$$\hat{m}_n(x) = \frac{\sum_{i=1}^n Y_i K((X_i - x)/h)}{\sum_{i=1}^n K((X_i - x)/h)}.$$

Home Page

Title Page

◀ ▶

◀ ▶

Page 17 of 100

Go Back

Full Screen

Close

Quit