Frontier Topics in Empirical Economics: Week 2
Non-parametric Method

Zibin Huang [1]

[1] College of Business, Shanghai University of Finance and Economics

September 18, 2024

Non-parametric Method: Introduction

- Common Parametric Models
  Linear Model: $y = X\beta + e$, $e \sim N(0, \sigma^2)$;
  Probit/Logit Model: $P(y|X) = G(X\beta)$ where G is a nonlinear function
- Explicit Parametric Structure for Distribution
- Common Estimator
  OLS, MLE, Nonlinear LS, Efficient GMM etc.
- Key Properties of the Estimator
  Consistency, BLUE, Asymptotic Efficiency etc.

Non-parametric Method: Introduction

- Common Parametric Models
  Linear Model: $y \quad X^{\urcorner} \quad e, e \quad N\ 0; \ ^{2}\ ;$
  Probit/Logit Model: $P\ y¶X \quad G\ X$ where G is a nonlinear function
- Explicit Parametric Structure for Distribution
- Common Estimator
  OLS, MLE, Nonlinear LS, Efficient GMM etc.
- Key Properties of the Estimator
  Consistency, BLUE, Asymptotic Efficiency etc.

Non-parametric Method: Introduction

- Common Parametric Models
  Linear Model: $y \quad X^{\urcorner} \quad e, e \quad N \; 0; \; ^{2} \; ;$
  Probit/Logit Model: $P \; y¶X \quad G \; X$ where G is a nonlinear function
- Explicit Parametric Structure for Distribution
- Common Estimator
  OLS, MLE, Nonlinear LS, Efficient GMM etc.
- Key Properties of the Estimator
  Consistency, BLUE, Asymptotic Efficiency etc.

Non-parametric Method: Introduction

- Common Parametric Models
  Linear Model: $y \quad X^{\urcorner} \quad e, e \quad N\ 0;\ {}^{2}\ ;$
  Probit/Logit Model: $P\ \ y¶X \quad G\ \ X$ where G is a nonlinear function
- Explicit Parametric Structure for Distribution
- Common Estimator
  OLS, MLE, Nonlinear LS, Efficient GMM etc.
- Key Properties of the Estimator
  Consistency, BLUE, Asymptotic Efficiency etc.

Non-parametric Method: Introduction

- Common Parametric Models
  Linear Model: $y \quad X \quad e, e \quad N\ 0; \ ^{2} \ ;$
  Probit/Logit Model: $P\ y¶X \quad G\ X$ where G is a nonlinear function
- Explicit Parametric Structure for Distribution
- Common Estimator
  OLS, MLE, Nonlinear LS, Efficient GMM etc.
- Key Properties of the Estimator
  Consistency, BLUE, Asymptotic Efficiency etc.

Non-parametric Method: Introduction

- In linear model, we have to assume that CEF is linear
- Why linear? Simple? Why not $y \quad x^3 \quad lnx \quad e$?
- What if linear specification is wrong?
- Everything collapses. No data can save.
- It becomes only a linear approximation
- For example, if true model is Logit, but not linear regression
- Functional form can be wrong

Non-parametric Method: Introduction

- In linear model, we have to assume that CEF is linear
- Why linear? Simple? Why not $y \quad x^3 \quad lnx \quad e$?
- What if linear specification is wrong?
- Everything collapses. No data can save.
- It becomes only a linear approximation
- For example, if true model is Logit, but not linear regression
- Functional form can be wrong

Non-parametric Method: Introduction

- In linear model, we have to assume that CEF is linear
- Why linear? Simple? Why not $y \quad x^3 \quad lnx \quad e$?
- What if linear specification is wrong?
- Everything collapses. No data can save.
- It becomes only a linear approximation
- For example, if true model is Logit, but not linear regression
- Functional form can be wrong

Non-parametric Method: Introduction

- In linear model, we have to assume that CEF is linear
- Why linear? Simple? Why not $y \quad x^3 \quad lnx \quad e$?
- What if linear specification is wrong?
- Everything collapses. No data can save.
- It becomes only a linear approximation
- For example, if true model is Logit, but not linear regression
- Functional form can be wrong

Non-parametric Method: Introduction

- In linear model, we have to assume that CEF is linear
- Why linear? Simple? Why not $y \quad x^3 \quad lnx \quad e$?
- What if linear specification is wrong?
- Everything collapses. No data can save.
- It becomes only a linear approximation
- For example, if true model is Logit, but not linear regression
- Functional form can be wrong

Non-parametric Method: Introduction

- In linear model, we have to assume that CEF is linear
- Why linear? Simple? Why not $y \quad x^3 \quad lnx \quad e$?
- What if linear specification is wrong?
- Everything collapses. No data can save.
- It becomes only a linear approximation
- For example, if true model is Logit, but not linear regression
- Functional form can be wrong

Non-parametric Method: Introduction

- In linear model, we have to assume that CEF is linear
- Why linear? Simple? Why not $y \quad x^3 \quad lnx \quad e$?
- What if linear specification is wrong?
- Everything collapses. No data can save.
- It becomes only a linear approximation
- For example, if true model is Logit, but not linear regression
- Functional form can be wrong

Non-parametric Method: Introduction

- In linear model, we have to assume that CEF is linear
- Why linear? Simple? Why not $y \quad x^3 \quad lnx \quad e$?
- What if linear specification is wrong?
- Everything collapses. No data can save.
- It becomes only a linear approximation
- For example, if true model is Logit, but not linear regression
- Functional form can be wrong

Non-parametric Method: Introduction

- Parametric statistics are based on assumptions about the distribution of population from which the sample was taken
- Non-parametric statistics are NOT based on functional form assumptions
- The data can be collected from a sample that does not follow a specific distribution

Non-parametric Method: Introduction

- Parametric statistics are based on assumptions about the distribution of population from which the sample was taken
- Non-parametric statistics are NOT based on functional form assumptions
- The data can be collected from a sample that does not follow a specific distribution

Non-parametric Method: Introduction

- Parametric statistics are based on assumptions about the distribution of population from which the sample was taken
- Non-parametric statistics are NOT based on functional form assumptions
- The data can be collected from a sample that does not follow a specific distribution

Non-parametric Method: Introduction

- Parametric statistics are based on assumptions about the distribution of population from which the sample was taken
- Non-parametric statistics are NOT based on functional form assumptions
- The data can be collected from a sample that does not follow a specific distribution

Non-parametric Method: Introduction

- Potential Outcome Framework is intrinsically non-parametric
- If we can directly get estimations of $E[y|x=1]$ and $E[y|x=0]$
- We can estimate the ATE/ATT in a more general way without regression
- There are many other statistical modeling methods
- Non-parametric, semi-parametric to estimate CEF directly
- To understand tools beyond linear regression

Non-parametric Method: Introduction

- Potential Outcome Framework is intrinsically non-parametric
- If we can directly get estimations of $E[y|x=1]$ and $E[y|x=0]$
- We can estimate the ATE/ATT in a more general way without regression
- There are many other statistical modeling methods
- Non-parametric, semi-parametric to estimate CEF directly
- To understand tools beyond linear regression

Non-parametric Method: Introduction

- Potential Outcome Framework is intrinsically non-parametric
- If we can directly get estimations of $E[y|x=1]$ and $E[y|x=0]$
- We can estimate the ATE/ATT in a more general way without regression
- There are many other statistical modeling methods
- Non-parametric, semi-parametric to estimate CEF directly
- To understand tools beyond linear regression

Non-parametric Method: Introduction

- Potential Outcome Framework is intrinsically non-parametric
- If we can directly get estimations of $E \;\; y\P x \quad 1$ and $E \;\; y\P x \quad 0$
- We can estimate the ATE/ATT in a more general way without regression
- There are many other statistical modeling methods
- Non-parametric, semi-parametric to estimate CEF directly
- To understand tools beyond linear regression

Non-parametric Method: Introduction

- Potential Outcome Framework is intrinsically non-parametric
- If we can directly get estimations of $E[y|x=1]$ and $E[y|x=0]$
- We can estimate the ATE/ATT in a more general way without regression
- There are many other statistical modeling methods
- Non-parametric, semi-parametric to estimate CEF directly
- To understand tools beyond linear regression

Non-parametric Method: Introduction

- Potential Outcome Framework is intrinsically non-parametric
- If we can directly get estimations of $E[y|x = 1]$ and $E[y|x = 0]$
- We can estimate the ATE/ATT in a more general way without regression
- There are many other statistical modeling methods
- Non-parametric, semi-parametric to estimate CEF directly
- To understand tools beyond linear regression

Non-parametric Method: Introduction

- Potential Outcome Framework is intrinsically non-parametric
- If we can directly get estimations of $E[y|x=1]$ and $E[y|x=0]$
- We can estimate the ATE/ATT in a more general way without regression
- There are many other statistical modeling methods
- Non-parametric, semi-parametric to estimate CEF directly
- To understand tools beyond linear regression

Non-parametric Method: Introduction

- Let's forget about the model functional form
- Give up the "parametric" model like linear regression
- Do not assume that CEF is linear
- Go back to the original question to estimate $E[y_i|x_i]$ without imposing any functional form assumption

Non-parametric Method: Introduction

- Let's forget about the model functional form
- Give up the "parametric" model like linear regression
- Do not assume that CEF is linear
- Go back to the original question to estimate $E[y_i|x_i]$ without imposing any functional form assumption

Non-parametric Method: Introduction

- Let's forget about the model functional form
- Give up the "parametric" model like linear regression
- Do not assume that CEF is linear
- Go back to the original question to estimate $E\left(y_i|x_i\right)$ without imposing any functional form assumption

Non-parametric Method: Introduction

- Let's forget about the model functional form
- Give up the "parametric" model like linear regression
- Do not assume that CEF is linear
- Go back to the original question to estimate $E[y_i|x_i]$ without imposing any functional form assumption

Non-parametric Method: Introduction

- Let's forget about the model functional form
- Give up the "parametric" model like linear regression
- Do not assume that CEF is linear
- Go back to the original question to estimate $E[y_i | x_i]$ without imposing any functional form assumption

Non-parametric Method: Introduction

- Notation: x_i, y_i denotes random variable; X_i, Y_i denotes realizations; x, y denotes random variables or some value of the random variables
- Realizations are given (sample), they are NOT random in our context
  $\text{D}\, x <_i^n X_i dx \quad <_i^n X_i \,\text{D}\, x dx$

Non-parametric Method: Introduction

- Notation: $x_i; y_i$ denotes random variable; $X_i; Y_i$ denotes realizations; $x; y$ denotes random variables or some value of the random variables
- Realizations are given (sample), they are NOT random in our context
  $D\, x <_i^n X_i dx \qquad <_i^n X_i\, D\, xdx$

Non-parametric Method: Introduction

- Notation: $x_i; y_i$ denotes random variable; $X_i; Y_i$ denotes realizations; $x; y$ denotes random variables or some value of the random variables
- Realizations are given (sample), they are NOT random in our context
  $\mathrm{D}\, x <_i^n X_i dx \quad <_i^n X_i\, \mathrm{D}\, xdx$

Non-parametric Method: Kernel Regression

- Let's consider the first non-parametric method: Kernel regression
- It is super intuitive and interesting
- Instead of assuming $E\ y_i|x_i \quad x_i'$, we consider this CEF point by point
- That is, estimate $E\ y_i|x_i$ for each possible point of $x_i \quad X$

Non-parametric Method: Kernel Regression

- Let's consider the first non-parametric method: Kernel regression
- It is super intuitive and interesting
- Instead of assuming $E[y_i|x_i] = x_i'\beta$, we consider this CEF point by point
- That is, estimate $E[y_i|x_i]$ for each possible point of $x_i = x$

Non-parametric Method: Kernel Regression

- Let's consider the first non-parametric method: Kernel regression
- It is super intuitive and interesting
- Instead of assuming $E \; y_i | x_i \quad x_i^{\top}$, we consider this CEF point by point
- That is, estimate $E \; y_i | x_i$ for each possible point of $x_i \quad x$

Non-parametric Method: Kernel Regression

- Let's consider the first non-parametric method: Kernel regression
- It is super intuitive and interesting
- Instead of assuming $E\ y_i|x_i \quad x_i'$, we consider this CEF point by point
- That is, estimate $E\ y_i|x_i$ for each possible point of $x_i \quad x$

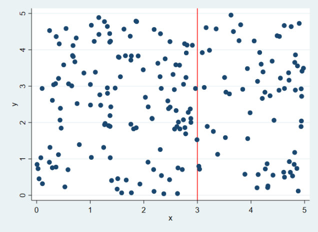
Non-parametric Method: Kernel Regression

- Let's consider the first non-parametric method: Kernel regression
- It is super intuitive and interesting
- Instead of assuming $E \; y_i | x_i \quad x_i'$, we consider this CEF point by point
- That is, estimate $E \; y_i | x_i$ for each possible point of $x_i \quad x$

Non-parametric Method: Kernel Regression

Step 1: Estimating a cumulative density

- Consider estimating a cumulative density function (CDF)

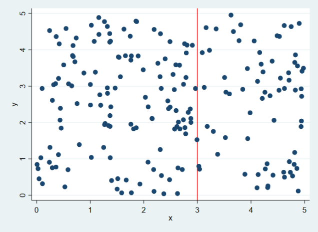


- What is the CDF at x 3? $\hat{F}$ x 3 ?
- Go back to kindergarten!

Non-parametric Method: Kernel Regression

Step 1: Estimating a cumulative density

- Consider estimating a cumulative density function (CDF)

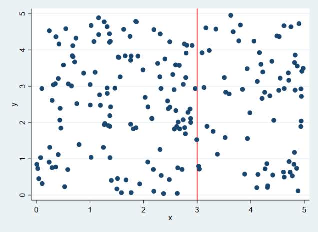


- What is the CDF at x 3? $\hat{F}$ x 3 ?
- Go back to kindergarten!

Non-parametric Method: Kernel Regression

Step 1: Estimating a cumulative density

- Consider estimating a cumulative density function (CDF)

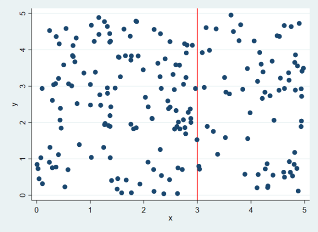


- What is the CDF at x 3? $\hat{F}$ x 3 ?
- Go back to kindergarten!

Non-parametric Method: Kernel Regression

Step 1: Estimating a cumulative density

- Consider estimating a cumulative density function (CDF)

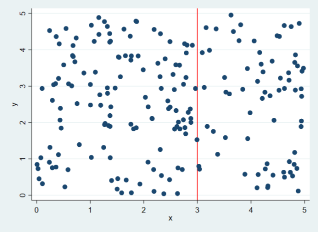


- What is the CDF at x 3? $\hat{F}$ x 3 ?
- Go back to kindergarten!

Non-parametric Method: Kernel Regression

Step 1: Estimating a cumulative density

- Consider estimating a cumulative density function (CDF)



- What is the CDF at x 3? $\hat{F}$ x 3 ?
- Go back to kindergarten!

Non-parametric Method: Kernel Regression

- Just count how many points lie on the left to the red line:

$$\hat{F}(x = 3) = \frac{1}{n}\sum \mathbf{1}(X_i \le 3)$$

- In general, we have an estimation of $F(x)$ as:

$$F(x) = P(X \le x) \quad \hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \le x)$$

- The proportion of points (realizations) that are smaller than x

Non-parametric Method: Kernel Regression

- Just count how many points lie on the left to the red line:

$$\hat{F}\ x \quad 3 \qquad \frac{1}{n} = \mathbf{1}\ X_i \ \&\ 3$$

- In general, we have an estimation of $F\ x$ as:

$$F\ x \qquad P\ X \ \&\ x \qquad \hat{F}\ x \qquad \frac{1}{n} \overset{n}{\underset{i\ 1}{=}} \mathbf{1}\ X_i \ \&\ x$$

- The proportion of points (realizations) that are smaller than x

Non-parametric Method: Kernel Regression

- Just count how many points lie on the left to the red line:

$$\hat{F}\ x \quad 3 \quad \frac{1}{n} = \mathbf{1}\ X_i\ \&\ 3$$

- In general, we have an estimation of $F\ x$ as:

$$F\ x \quad P\ X\ \&\ x \quad \hat{F}\ x \quad \frac{1}{n} \overset{n}{\underset{i\ \ 1}{=}} \mathbf{1}\ X_i\ \&\ x$$

- The proportion of points (realizations) that are smaller than x

Non-parametric Method: Kernel Regression

- Just count how many points lie on the left to the red line:

$$\hat{F} \;\, x \quad 3 \qquad \frac{1}{n} = \mathbf{1} \;\, X_i \;\&\; 3$$

- In general, we have an estimation of $F \;\, x$ as:

$$F \;\, x \qquad P \;\, X \;\&\; x \qquad \hat{F} \;\, x \qquad \frac{1}{n} \overset{n}{\underset{i \;\; 1}{=}} \mathbf{1} \;\, X_i \;\&\; x$$

- The proportion of points (realizations) that are smaller than x

Non-parametric Method: Kernel Regression

Step 2: Estimating a probability density

- Consider estimating a probability density function (PDF)
- PDF represents a marginal increase in CDF at some point (derivative)

$$f(x) = \frac{dF(x)}{dx} = \lim_{h \to 0} \frac{F(x+h) - F(x-h)}{2h}$$

$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h}$$

- Changes of $F(x)$ in a very small interval (with length $2h$)
- h is called "bandwidth"

Non-parametric Method: Kernel Regression

Step 2: Estimating a probability density

- Consider estimating a probability density function (PDF)
- PDF represents a marginal increase in CDF at some point (derivative)

$$f(x) = \frac{dF(x)}{dx} = \lim_{h \to 0} \frac{F(x+h) - F(x-h)}{2h}$$

$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h}$$

- Changes of $F(x)$ in a very small interval (with length $2h$)
- h is called "bandwidth"

Non-parametric Method: Kernel Regression

Step 2: Estimating a probability density

- Consider estimating a probability density function (PDF)
- PDF represents a marginal increase in CDF at some point (derivative)

$$f(x) = \frac{dF(x)}{dx} = \lim_{h \to 0} \frac{F(x+h) - F(x-h)}{2h}$$

$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h}$$

- Changes of $F(x)$ in a very small interval (with length $2h$)
- h is called "bandwidth"

Non-parametric Method: Kernel Regression

Step 2: Estimating a probability density

- Consider estimating a probability density function (PDF)
- PDF represents a marginal increase in CDF at some point (derivative)

$$f(x) = \frac{dF(x)}{dx} = \lim_{h \to 0} \frac{F(x+h) - F(x-h)}{2h}$$

$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h}$$

- Changes of $F(x)$ in a very small interval (with length $2h$)
- h is called "bandwidth"

Non-parametric Method: Kernel Regression

Step 2: Estimating a probability density

- Consider estimating a probability density function (PDF)
- PDF represents a marginal increase in CDF at some point (derivative)

$$f(x) = \frac{dF(x)}{dx} = \lim_{h \to 0} \frac{F(x+h) - F(x-h)}{2h}$$

$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h}$$

- Changes of $F(x)$ in a very small interval (with length $2h$)
- h is called "bandwidth"

Non-parametric Method: Kernel Regression

Step 2: Estimating a probability density

- Consider estimating a probability density function (PDF)
- PDF represents a marginal increase in CDF at some point (derivative)

$$f(x) = \frac{dF(x)}{dx} = \lim_{h \to 0} \frac{F(x+h) - F(x-h)}{2h}$$

$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h}$$

- Changes of $F(x)$ in a very small interval (with length $2h$)
- h is called "bandwidth"

Non-parametric Method: Kernel Regression

Step 2: Estimating a probability density

- Consider estimating a probability density function (PDF)
- PDF represents a marginal increase in CDF at some point (derivative)

$$f(x) = \frac{dF(x)}{dx} = \lim_{h \to 0} \frac{F(x+h) - F(x-h)}{2h}$$

$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h}$$

- Changes of $F(x)$ in a very small interval (with length $2h$)
- h is called "bandwidth"

Non-parametric Method: Kernel Regression

- Then we can write the probability density $f\ x$ at some value x as:

$$\hat{f}\ x \qquad \frac{1}{2h}\,\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\ X_i \,\&\, x \quad h \qquad \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\ X_i \,\&\, x \quad h$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2h}\mathbf{1}\ x \quad h \,\&\, X_i \,\&\, x \quad h$$

- How to interpret this?
- We count the number of obs within a small interval around x, dividing by the length and the total number of obs
- $\sum_{i=1}^{n}$ $<$ $\frac{1}{2h}\mathbf{1}\ x \quad h \,\&\, X_i \,\&\, x \quad h$ is the number of obs per unit length
- When n is large, we can choose very small h

Non-parametric Method: Kernel Regression

- Then we can write the probability density $f\ \ x$ at some value x as:

$$\hat{f}\ x \qquad \frac{1}{2h}\ \frac{1}{n} \overset{n}{\underset{i\ \ 1}{=}} \mathbf{1}\ X_i \,\&\, x \quad h \qquad \frac{1}{n} \overset{n}{\underset{i\ \ 1}{=}} \mathbf{1}\ X_i \,\&\, x \quad h$$

$$\frac{1}{n} \overset{n}{\underset{i\ \ 1}{=}} \frac{1}{2h}\mathbf{1}\ x \quad h \,\&\, X_i \,\&\, x \quad h$$

- How to interpret this?
- We count the number of obs within a small interval around x, dividing by the length and the total number of obs
- $\overset{n}{\underset{i\ \ 1}{<}} \frac{1}{2h}\mathbf{1}\ x \quad h \,\&\, X_i \,\&\, x \quad h$ is the number of obs per unit length
- When n is large, we can choose very small h

Non-parametric Method: Kernel Regression

- Then we can write the probability density $f \;\; x$ at some value x as:

$$\hat{f} \;\; x \qquad \frac{1}{2h}\,\frac{1}{n} \overset{n}{\underset{i \;\; 1}{=}} \mathbf{1} \;\; X_i \;\&\; x \quad h \qquad \frac{1}{n} \overset{n}{\underset{i \;\; 1}{=}} \mathbf{1} \;\; X_i \;\&\; x \quad h$$

$$\frac{1}{n} \overset{n}{\underset{i \;\; 1}{=}} \frac{1}{2h}\mathbf{1} \;\; x \quad h \;\&\; X_i \;\&\; x \quad h$$

- How to interpret this?
- We count the number of obs within a small interval around x, dividing by the length and the total number of obs
- $\overset{n}{\underset{i \;\; 1}{<}} \frac{1}{2h}\mathbf{1} \;\; x \quad h \;\&\; X_i \;\&\; x \quad h$ is the number of obs per unit length
- When n is large, we can choose very small h

Non-parametric Method: Kernel Regression

- Then we can write the probability density $f \; x$ at some value x as:

$$\hat{f} \; x \qquad \frac{1}{2h} \frac{1}{n} \overset{n}{\underset{i \;\; 1}{=}} \mathbf{1} \; X_i \; \& \; x \quad h \qquad \frac{1}{n} \overset{n}{\underset{i \;\; 1}{=}} \mathbf{1} \; X_i \; \& \; x \quad h$$

$$\frac{1}{n} \overset{n}{\underset{i \;\; 1}{=}} \frac{1}{2h} \mathbf{1} \; x \quad h \; \& \; X_i \; \& \; x \quad h$$

- How to interpret this?
- We count the number of obs within a small interval around x, dividing by the length and the total number of obs
- $\overset{n}{\underset{i \;\; 1}{<}} \frac{1}{2h} \mathbf{1} \; x \quad h \; \& \; X_i \; \& \; x \quad h$ is the number of obs per unit length
- When n is large, we can choose very small h

Non-parametric Method: Kernel Regression

- Then we can write the probability density $f\ \ x$ at some value x as:

$$\hat{f}\ \ x \quad \frac{1}{2h}\ \frac{1}{n}\ \overset{n}{\underset{i\ \ 1}{=}}\ \mathbf{1}\ X_i\ \&\ x \quad h \quad \frac{1}{n}\ \overset{n}{\underset{i\ \ 1}{=}}\ \mathbf{1}\ X_i\ \&\ x \quad h$$

$$\frac{1}{n}\ \overset{n}{\underset{i\ \ 1}{=}}\ \frac{1}{2h}\mathbf{1}\ x \quad h\ \&\ X_i\ \&\ x \quad h$$

- How to interpret this?
- We count the number of obs within a small interval around x, dividing by the length and the total number of obs
- $\overset{n}{\underset{i\ \ 1}{<}}\ \frac{1}{2h}\mathbf{1}\ x \quad h\ \&\ X_i\ \&\ x \quad h$ is the number of obs per unit length
- When n is large, we can choose very small h

Non-parametric Method: Kernel Regression

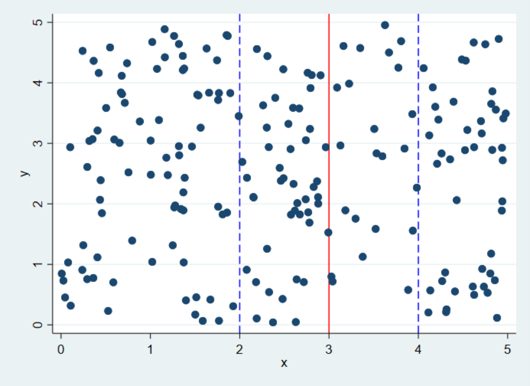
- Then we can write the probability density $f\ \ x$ at some value x as:

$$\hat{f}\ \ x \qquad \frac{1}{2h}\ \frac{1}{n} \overset{n}{\underset{i\ \ 1}{=}} \mathbf{1}\ \ X_i \,\&\, x \quad h \qquad \frac{1}{n} \overset{n}{\underset{i\ \ 1}{=}} \mathbf{1}\ \ X_i \,\&\, x \quad h$$

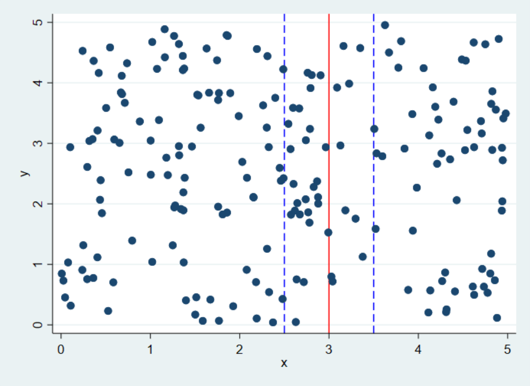
$$\frac{1}{n} \overset{n}{\underset{i\ \ 1}{=}} \frac{1}{2h}\mathbf{1}\ \ x \quad h \,\&\, X_i \,\&\, x \quad h$$

- How to interpret this?
- We count the number of obs within a small interval around x, dividing by the length and the total number of obs
- $\overset{n}{\underset{i\ \ 1}{<}} \frac{1}{2h}\mathbf{1}\ \ x \quad h \,\&\, X_i \,\&\, x \quad h$ is the number of obs per unit length
- When n is large, we can choose very small h

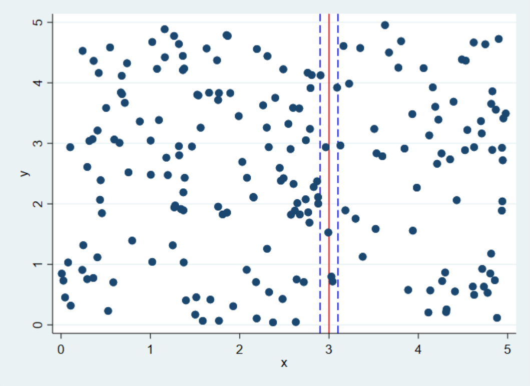
Non-parametric Method: Kernel Regression



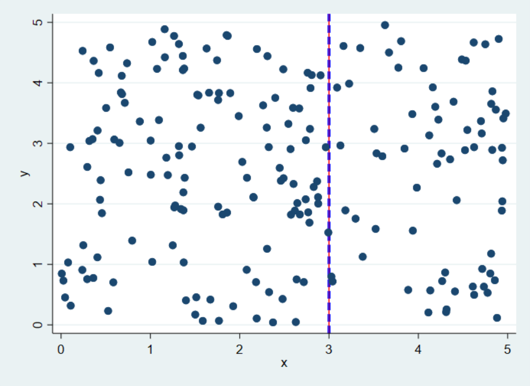
Non-parametric Method: Kernel Regression



Non-parametric Method: Kernel Regression



Non-parametric Method: Kernel Regression



Non-parametric Method: Kernel Regression

- Define $k(v) = \frac{1}{2}\mathbf{1}(\|v\| \leq 1)$. Then we have:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}k\left(\frac{X_i - x}{h}\right)$$

- We call $k(v)$ a uniform kernel function
- This $\hat{f}(x)$ is a kernel estimator of the PDF (uniform kernel)
- Kernel is weight!
- There can be other kinds of kernel functions, when we assign different weights to different observations

Non-parametric Method: Kernel Regression

- Define $k\ v\quad \frac{1}{2}\mathbf{1}\ ¶v¶ \ \&\ 1\ .$ Then we have:

$$\hat{f}\ x \quad \frac{1}{n} \overset{n}{\underset{i\ \ 1}{=}} \frac{1}{h} k\ \frac{X_i \quad x}{h}$$

- We call $k\ v$ a uniform kernel function
- This $\hat{f}\ x$ is a kernel estimator of the PDF (uniform kernel)
- Kernel is weight!
- There can be other kinds of kernel functions, when we assign different weights to different observations

Non-parametric Method: Kernel Regression

- Define $k(v) = \frac{1}{2}\mathbf{1}(|v| \le 1)$. Then we have:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} k\left(\frac{X_i - x}{h}\right)$$

- We call $k(v)$ a uniform kernel function
- This $\hat{f}(x)$ is a kernel estimator of the PDF (uniform kernel)
- Kernel is weight!
- There can be other kinds of kernel functions, when we assign different weights to different observations

Non-parametric Method: Kernel Regression

- Define $k(v) = \frac{1}{2}\mathbf{1}(|v| \le 1)$. Then we have:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}k\left(\frac{X_i - x}{h}\right)$$

- We call $k(v)$ a uniform kernel function
- This $\hat{f}(x)$ is a kernel estimator of the PDF (uniform kernel)
- Kernel is weight!
- There can be other kinds of kernel functions, when we assign different weights to different observations

Non-parametric Method: Kernel Regression

- Define $k(v) = \frac{1}{2}\mathbf{1}(|v| \le 1)$. Then we have:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}k\left(\frac{X_i - x}{h}\right)$$

- We call $k(v)$ a uniform kernel function
- This $\hat{f}(x)$ is a kernel estimator of the PDF (uniform kernel)
- Kernel is weight!
- There can be other kinds of kernel functions, when we assign different weights to different observations

Non-parametric Method: Kernel Regression

- Define $k\ \ v \quad \frac{1}{2}\mathbf{1}\ \P v \P\ \&\ 1\ .$ Then we have:

$$\hat{f}\ \ x \qquad \frac{1}{n} \overset{n}{\underset{i\ \ 1}{=}} \frac{1}{h} k\ \frac{X_i \quad x}{h}$$

- We call $k\ \ v$ a uniform kernel function
- This $\hat{f}\ \ x$ is a kernel estimator of the PDF (uniform kernel)
- Kernel is weight!
- There can be other kinds of kernel functions, when we assign different weights to different observations

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
  - $k\ v$ is integrated to 1
  - $k\ v$ is symmetric with $k\ v\ \ k\ \ v$
- The weights sum to one; The weights are symmetric
- Triangular Kernel: $k\ v\ \ \ 1\ \ \P v \P\ \mathbf{1}\ \P v \P\ \&\ 1$
- Epanechnikov Kernel: $k\ v\ \ \ \frac{3}{4}\ 1\ \ v^2\ \mathbf{1}\ \P v \P\ \&\ 1$
- Gaussian Kernel: $k\ v\ \ \ \frac{1}{2} e^{\frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
  - $k(v)$ is integrated to 1
  - $k(v)$ is symmetric with $k(v) = k(-v)$
- The weights sum to one; The weights are symmetric
- Triangular Kernel: $k(v) = (1 - |v|)\mathbf{1}(|v| \leq 1)$
- Epanechnikov Kernel: $k(v) = \frac{3}{4}(1 - v^2)\mathbf{1}(|v| \leq 1)$
- Gaussian Kernel: $k(v) = \frac{1}{2} e^{\frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
  - $k\ v$ is integrated to 1
  - $k\ v$ is symmetric with $k\ v$ $\quad k \quad v$
- The weights sum to one; The weights are symmetric
- Triangular Kernel: $k\ v \quad 1 \quad ¶v¶$ **1** $¶v¶$ & 1
- Epanechnikov Kernel: $k\ v \quad \frac{3}{4}\ 1 \quad v^2$ **1** $¶v¶$ & 1
- Gaussian Kernel: $k\ v \quad \frac{1}{2} e^{\frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
  - $k\ \ v$ is integrated to 1
  - $k\ \ v$ is symmetric with $k\ \ v \quad k \quad v$
- The weights sum to one; The weights are symmetric
- Triangular Kernel: $k\ \ v \quad 1 \quad$ ¶v¶ **1** ¶v¶ & 1
- Epanechnikov Kernel: $k\ \ v \quad \frac{3}{4}\ 1 \quad v^2$ **1** ¶v¶ & 1
- Gaussian Kernel: $k\ \ v \quad \frac{1}{2} e^{\frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
  - $k(v)$ is integrated to 1
  - $k(v)$ is symmetric with $k(v) = k(-v)$
- The weights sum to one; The weights are symmetric
- Triangular Kernel: $k(v) = (1 - |v|)\mathbf{1}\{|v| \leq 1\}$
- Epanechnikov Kernel: $k(v) = \frac{3}{4}(1 - v^2)\mathbf{1}\{|v| \leq 1\}$
- Gaussian Kernel: $k(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
  - $k\ v$ is integrated to 1
  - $k\ v$ is symmetric with $k\ v\quad k\quad v$
- The weights sum to one; The weights are symmetric
- Triangular Kernel: $k\ v\quad 1\quad ¶v¶\ \mathbf{1}\ ¶v¶ \& 1$
- Epanechnikov Kernel: $k\ v\quad \frac{3}{4}\ 1\quad v^2\ \mathbf{1}\ ¶v¶ \& 1$
- Gaussian Kernel: $k\ v\quad \frac{1}{2} e^{\frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
  - $k\ \ v$ is integrated to 1
  - $k\ \ v$ is symmetric with $k\ \ v \quad k \quad v$
- The weights sum to one; The weights are symmetric
- Triangular Kernel: $k\ \ v \quad 1 \quad ¶v¶\ \mathbf{1}\ ¶v¶ \ \&\ 1$
- Epanechnikov Kernel: $k\ \ v \quad \frac{3}{4}\ 1 \quad v^2\ \mathbf{1}\ ¶v¶ \ \&\ 1$
- Gaussian Kernel: $k\ \ v \quad \frac{1}{2} e^{\ \frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
  - $k\ \ v$ is integrated to 1
  - $k\ \ v$ is symmetric with $k\ \ v\quad k\quad v$
- The weights sum to one; The weights are symmetric
- Triangular Kernel: $k\ \ v\quad 1\quad$ ¶v¶ **1** ¶v¶ & 1
- Epanechnikov Kernel: $k\ \ v\quad \frac{3}{4}\ 1\quad v^2$ **1** ¶v¶ & 1
- Gaussian Kernel: $k\ \ v\quad \frac{1}{2} e^{\ \frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

- A function can be used as a kernel if
  - $k\ v$ is integrated to 1
  - $k\ v$ is symmetric with $k\ v \quad k \quad v$
- The weights sum to one; The weights are symmetric
- Triangular Kernel: $k\ v \quad 1 \quad \P v \P\ \mathbf{1}\ \P v \P \ \& \ 1$
- Epanechnikov Kernel: $k\ v \quad \frac{3}{4}\ 1 \quad v^2\ \mathbf{1}\ \P v \P \ \& \ 1$
- Gaussian Kernel: $k\ v \quad \frac{1}{2} e^{\ \frac{v^2}{2}}$
- Usually, Epanechnikov Kernel and Triangular Kernel are preferred

Non-parametric Method: Kernel Regression

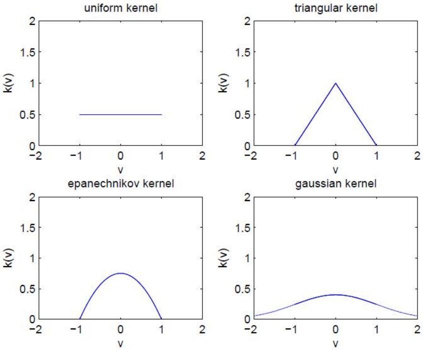


Figure 1: Various Kernels

Non-parametric Method: Kernel Regression

- For multivariate case, let $v = (v_1, v_2, \cdots, v_q)$.
- Define product kernel: $K(v) = (k(v_1) k(v_2) \cdots, k(v_q))$.
- The estimator becomes:

$$\hat{f}(x) = \frac{1}{n h_1 h_2 \cdots h_q} \sum_i K\left(\frac{X_i - x}{h}\right)$$

- Define $h = (h_1, h_2, \cdots, h_q)$
- $K\left(\frac{X_i - x}{h}\right)$ is the weighted sum of points within the q-dimension hypercube
- $h_1 h_2 \cdots h_q$ is the volumn of this q-dimension hypercube

Non-parametric Method: Kernel Regression

- For multivariate case, let $v \quad v_1; v_2; \quad ; v_q$.
- Define product kernel: $K \; v \quad k \; v_1 \; k \; v_2 \quad ; k \; v_q$.
- The estimator becomes:

$$\hat{f} \; x \quad \frac{1}{nh_1 h_2 \quad h_q} \sum_i K \; \frac{X_i \quad x}{h}$$

- Define $h \quad h_1; h_2; \quad ; h_q$
- $K \; \frac{X_i \quad x}{h}$ is the weighted sum of points within the q-dimension hypercube
- $h_1 h_2 \quad h_q$ is the volumn of this q-dimension hypercube

Non-parametric Method: Kernel Regression

- For multivariate case, let $v \quad v_1; v_2; \quad ; v_q$.
- Define product kernel: $K \; v \quad k \; v_1 \; k \; v_2 \quad ; k \; v_q$.
- The estimator becomes:

$$\hat{f} \; x \quad \frac{1}{nh_1h_2 \quad h_q} \underset{i}{=} K \; \frac{X_i \quad x}{h}$$

- Define $h \quad h_1; h_2; \quad ; h_q$
- $K \; \frac{X_i \quad x}{h}$ is the weighted sum of points within the q-dimension hypercube
- $h_1h_2 \quad h_q$ is the volumn of this q-dimension hypercube

Non-parametric Method: Kernel Regression

- For multivariate case, let $v \quad v_1; v_2; \quad ; v_q \;.$
- Define product kernel: $K \; v \quad k \; v_1 \; k \; v_2 \quad ; k \; v_q \;.$
- The estimator becomes:

$$\hat{f} \; x \quad \frac{1}{nh_1 h_2 \quad h_q} \underset{i}{=} K \; \frac{X_i \quad x}{h}$$

- Define $h \quad h_1; h_2; \quad ; h_q$
- $K \; \frac{X_i \quad x}{h}$ is the weighted sum of points within the q-dimension hypercube
- $h_1 h_2 \quad h_q$ is the volumn of this q-dimension hypercube

Non-parametric Method: Kernel Regression

- For multivariate case, let $v \quad v_1; v_2; \quad ; v_q$.
- Define product kernel: $K\ v \quad k\ v_1\ k\ v_2 \quad ; k\ v_q$.
- The estimator becomes:

$$\hat{f}\ x \quad \frac{1}{nh_1 h_2 \quad h_q} \underset{i}{=} K\ \frac{X_i \quad x}{h}$$

- Define $h \quad h_1; h_2; \quad ; h_q$
- $K\ \frac{X_i \quad x}{h}$ is the weighted sum of points within the q-dimension hypercube
- $h_1 h_2 \quad h_q$ is the volumn of this q-dimension hypercube

Non-parametric Method: Kernel Regression

- For multivariate case, let $v \quad v_1; v_2; \quad ; v_q \quad .$
- Define product kernel: $K \; v \quad k \; v_1 \; k \; v_2 \quad ; k \; v_q \quad .$
- The estimator becomes:

$$\hat{f} \; x \qquad \frac{1}{n h_1 h_2 \quad h_q} \underset{i}{=} K \; \frac{X_i \quad x}{h}$$

- Define $h \quad h_1; h_2; \quad ; h_q$
- $K \; \frac{X_i \quad x}{h}$ is the weighted sum of points within the q-dimension hypercube
- $h_1 h_2 \quad h_q$ is the volumn of this q-dimension hypercube

Non-parametric Method: Kernel Regression

- For multivariate case, let $v \quad v_1; v_2; \quad ; v_q$.
- Define product kernel: $K \; v \quad k \; v_1 \; k \; v_2 \quad ; k \; v_q$.
- The estimator becomes:

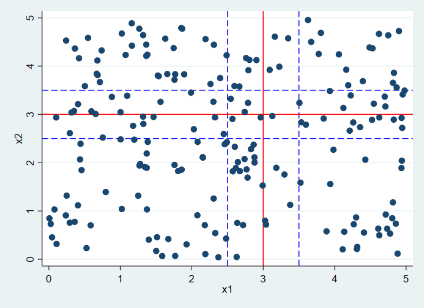
$$\hat{f} \; x \qquad \frac{1}{nh_1 h_2 \quad h_q} \underset{i}{=} K \; \frac{X_i \quad x}{h}$$

- Define $h \quad h_1; h_2; \quad ; h_q$
- $K \; \frac{X_i \quad x}{h}$ is the weighted sum of points within the q-dimension hypercube
- $h_1 h_2 \quad h_q$ is the volumn of this q-dimension hypercube

Non-parametric Method: Kernel Regression

In two dimension case, we have

- $K \ \frac{X_i \quad x}{h}$ is the weighted sum of points within the rectangular
- $h_1 h_2 \quad h_q$ is the area of this rectangular



Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Finally, let's see how to estimate a CEF using kernel method
- Not like linear regression, we estimate the CEF point by point
- Assume that we have CEF:

$$Y = g(X) + u$$
$$E(Y|X) = g(X)$$

- u has a conditional variance $Var(u|X) = \sigma^2(x)$

Step 3: Estimating a CEF

- Finally, let's see how to estimate a CEF using kernel method
- Not like linear regression, we estimate the CEF point by point
- Assume that we have CEF:

$$Y \quad g \; X \quad u$$

$$E \; Y|X \quad g \; X$$

- u has a conditional variance $Var \; u|X \quad {}^2 \; x$

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Finally, let's see how to estimate a CEF using kernel method
- Not like linear regression, we estimate the CEF point by point
- Assume that we have CEF:

$$Y \quad g \; X \quad u$$
$$E \; Y¶X \quad g \; X$$

- u has a conditional variance $Var \; u¶X \quad {}^{2} \; x$

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Finally, let's see how to estimate a CEF using kernel method
- Not like linear regression, we estimate the CEF point by point
- Assume that we have CEF:

$$Y \quad g\ X \quad u$$

$$E\ Y¶X \quad g\ X$$

- u has a conditional variance $Var\ \ u¶X \quad {}^2\ x$

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Finally, let's see how to estimate a CEF using kernel method
- Not like linear regression, we estimate the CEF point by point
- Assume that we have CEF:

$$Y \quad g \; X \quad u$$
$$E \; Y|X \quad g \; X$$

- u has a conditional variance $Var \; u|X \quad {}^2 \; x$

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Finally, let's see how to estimate a CEF using kernel method
- Not like linear regression, we estimate the CEF point by point
- Assume that we have CEF:

$$Y \quad g \; X \quad u$$
$$E \; Y¶X \quad g \; X$$

- u has a conditional variance $Var \; u¶X \quad {}^{2} \; x$

Step 3: Estimating a CEF

- Based on the CDF and PDF we've got, we have Nadaraya-Watson Estimator (N-W) for CEF as follows:

$$\hat{g}(x) = \sum_{i=1}^{n} Y_i K_h(X_i - x); \quad where \quad K_h(X_i - x) = \frac{K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)}$$

- Intuition: The conditional Expectation of Y given X=x is estimated as a weighted average of observed Y_i closely around x (within the range of bandwidth h).
- Weights are determined by the kernel function

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Based on the CDF and PDF we've got, we have Nadaraya-Watson Estimator (N-W) for CEF as follows:

$$\hat{g}(x) = \sum_{i=1}^{n} Y_i K_h(X_i - x); \quad where \quad K_h(X_i - x) = \frac{K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)}$$

- Intuition: The conditional Expectation of Y given X=x is estimated as a weighted average of observed Y_i closely around x (within the range of bandwidth h).
- Weights are determined by the kernel function

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Based on the CDF and PDF we've got, we have Nadaraya-Watson Estimator (N-W) for CEF as follows:

$$\hat{g}(x) = \sum_{i=1}^{n} Y_i K_h(X_i - x); \quad where \quad K_h(X_i - x) = \frac{K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)}$$

- Intuition: The conditional Expectation of Y given X=x is estimated as a weighted average of observed Y_i closely around x (within the range of bandwidth h).
- Weights are determined by the kernel function

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Based on the CDF and PDF we've got, we have Nadaraya-Watson Estimator (N-W) for CEF as follows:

$$\hat{g}(x) = \sum_{i=1}^{n} Y_i K_h(X_i - x); \quad where \quad K_h(X_i - x) = \frac{K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)}$$

- Intuition: The conditional Expectation of Y given X=x is estimated as a weighted average of observed Y_i closely around x (within the range of bandwidth h).
- Weights are determined by the kernel function

Non-parametric Method: Kernel Regression

Step 3: Estimating a CEF

- Based on the CDF and PDF we've got, we have Nadaraya-Watson Estimator (N-W) for CEF as follows:

$$\hat{g}(x) = \sum_{i=1}^{n} Y_i K_h(X_i - x); \quad \textit{where} \quad K_h(X_i - x) = \frac{K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)}$$

- Intuition: The conditional Expectation of Y given X=x is estimated as a weighted average of observed Y_i closely around x (within the range of bandwidth h).
- Weights are determined by the kernel function

Non-parametric Method: Kernel Regression

Homework:

- 1. Derive NW Estimator from the kernel estimator of CDF and PDF. This can be a little bit hard. You can refer to Notes from Carol (or Hansen's book) for help.
- 2. What is NW Estimator, if we use the uniform kernel?

Non-parametric Method: Kernel Regression

Homework:

- 1. Derive NW Estimator from the kernel estimator of CDF and PDF. This can be a little bit hard. You can refer to Notes from Carol (or Hansen's book) for help.
- 2. What is NW Estimator, if we use the uniform kernel?

Non-parametric Method: Kernel Regression

Homework:

- 1. Derive NW Estimator from the kernel estimator of CDF and PDF. This can be a little bit hard. You can refer to Notes from Carol (or Hansen's book) for help.
- 2. What is NW Estimator, if we use the uniform kernel?

Non-parametric Method: Kernel Regression

Homework:

- 1. Derive NW Estimator from the kernel estimator of CDF and PDF. This can be a little bit hard. You can refer to Notes from Carol (or Hansen's book) for help.
- 2. What is NW Estimator, if we use the uniform kernel?

Non-parametric Method: Kernel Regression

- We have $g(x) = E(Y|X)$ as CEF and $f(x)$ as density for x

Theorem (Asymptotics for N-W Estimator)

Under some regularity conditions, as $n \to \infty$, $h_s \to 0$ $(s = 1, \ldots, q)$, $nh_1 \ldots h_q \to \infty$ *and* $nh_1 \ldots h_q \sum_{s=1}^{q} h_s^6 \to 0$, *we have:*

$$\sqrt{nh_1 \ldots h_q}\left(\hat{g}(x) - g(x) - \sum_{s=1}^{q} h_s^2 B_s(x)\right) \xrightarrow{d} N\left(0, \frac{\sigma^2(x)}{f(x)}\left(\int k(v)^2 dv\right)^q\right)$$

$$\text{where} \quad B_s(x) = \frac{\int v^2 k(v)\, dv}{2f(x)}\left[2\frac{\partial f(x)}{\partial x_s}\frac{\partial g(x)}{\partial x_s} + f(x)\frac{\partial^2 g(x)}{\partial x_s^2}\right]$$

Non-parametric Method: Kernel Regression

- We have $g(x) = E(Y|X = x)$ as CEF and $f(x)$ as density for x

Theorem (Asymptotics for N-W Estimator)

Under some regularity conditions, as $n \to \infty$, $h_s \to 0$ $(s = 1, \ldots, q)$, $nh_1 \ldots h_q \to \infty$ and $nh_1 \ldots h_q \sum_{s=1}^{q} h_s^6 \to 0$, we have:

$$\sqrt{nh_1 \ldots h_q}\left(\hat{g}(x) - g(x) - \sum_{s=1}^{q} h_s^2 B_s(x)\right) \xrightarrow{d} N\left(0, \frac{\sigma^2(x)}{f(x)}\left(\int k(v)^2 dv\right)^q\right)$$

$$\text{where} \quad B_s(x) = \frac{\int v^2 k(v)\, dv}{2f(x)}\left(2\frac{\partial f(x)}{\partial x_s}\frac{\partial g(x)}{\partial x_s} + f(x)\frac{\partial^2 g(x)}{\partial x_s^2}\right)$$

Non-parametric Method: Kernel Regression

Asymptotic Bias $\sum_{s=1}^{q} h_s^2 \frac{\int v^2 k(v)\, dv}{2f(x)} \left(2 \frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s} + f(x) \frac{\partial^2 g(x)}{\partial x_s^2} \right)$

Asymptotic Variance $\frac{1}{nh_1 \ldots h_q} \frac{\sigma^2(x)}{f(x)} \left(\int k(v)^2 dv \right)^q$

- (1) h_s Bias ; Variance
  we have trade-off in choosing kernel bandwidth.
- (2) q Variance exponentially
  We call this "Curse of Dimensionality".
- (3) Kernel more concentrated Bias $\int v^2 k(v)\, dv$; Variance $\int k(v)^2 dv$
- (4) Slope Effect and Curvature Effect on bias: $\frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s}$; $\frac{\partial^2 g(x)}{\partial x_s^2}$
- (5) $f(x)$ Bias ; Variance (more observations)

Non-parametric Method: Kernel Regression

Asymptotic Bias $\sum_{s=1}^{q} h_s^2 \frac{\int v^2 k(v)\,dv}{2f(x)} \left[2\frac{\partial f(x)}{\partial x_s}\frac{\partial g(x)}{\partial x_s} + f(x)\frac{\partial^2 g(x)}{\partial x_s^2} \right]$

Asymptotic Variance $\frac{1}{nh_1 \ldots h_q} \frac{\sigma^2(x)}{f(x)} \left[\int k(v)^2 dv \right]^q$

- (1) h_s Bias ; Variance
  we have trade-off in choosing kernel bandwidth.
- (2) q Variance exponentially
  We call this "Curse of Dimensionality".
- (3) Kernel more concentrated Bias $\int v^2 k(v)\,dv$; Variance $\int k(v)^2 dv$
- (4) Slope Effect and Curvature Effect on bias: $\frac{\partial f(x)}{\partial x_s}\frac{\partial g(x)}{\partial x_s}$; $\frac{\partial^2 g(x)}{\partial x_s^2}$
- (5) $f(x)$ Bias ; Variance (more observations)

Non-parametric Method: Kernel Regression

Asymptotic Bias $\sum_{s=1}^{q} h_s^2 \frac{\int v^2 k(v)\,dv}{2f(x)} \left(2\frac{\partial f(x)}{\partial x_s}\frac{\partial g(x)}{\partial x_s} + f(x)\frac{\partial^2 g(x)}{\partial x_s^2} \right)$

Asymptotic Variance $\frac{1}{nh_1 \ldots h_q}\frac{\sigma^2(x)}{f(x)}\left[\int k(v)^2 dv\right]^q$

- (1) h_s *Bias* ; *Variance*
  we have trade-off in choosing kernel bandwidth.
- (2) q *Variance* *exponentially*
  We call this "Curse of Dimensionality".
- (3) Kernel more concentrated *Bias* $\int v^2 k(v)\,dv$; *Variance* $\int k(v)^2 dv$
- (4) Slope Effect and Curvature Effect on bias: $\frac{\partial f(x)}{\partial x_s}\frac{\partial g(x)}{\partial x_s}; \frac{\partial^2 g(x)}{\partial x_s^2}$
- (5) $f(x)$ *Bias* ; *Variance* (more observations)

Non-parametric Method: Kernel Regression

Asymptotic Bias $\sum_{s=1}^{q} h_s^2 \frac{\int v^2 k(v)\,dv}{2f(x)} \left(2\frac{\partial f(x)}{\partial x_s}\frac{\partial g(x)}{\partial x_s} + f(x)\frac{\partial^2 g(x)}{\partial x_s^2} \right)$

Asymptotic Variance $\frac{1}{nh_1 \ldots h_q} \frac{{}^2(x)}{f(x)} \left(\int k(v)^2 dv \right)^q$

- (1) h_s *Bias* ; *Variance*
  we have trade-off in choosing kernel bandwidth.
- (2) q *Variance* *exponentially*
  We call this "Curse of Dimensionality".
- (3) Kernel more concentrated *Bias* $\int v^2 k(v)\,dv$; *Variance* $\int k(v)^2 dv$
- (4) Slope Effect and Curvature Effect on bias: $\frac{\partial f(x)}{\partial x_s}\frac{\partial g(x)}{\partial x_s}$; $\frac{\partial^2 g(x)}{\partial x_s^2}$
- (5) $f(x)$ *Bias* ; *Variance* (more observations)

Non-parametric Method: Kernel Regression

Asymptotic Bias $\sum_{s=1}^{q} h_s^2 \frac{\int v^2 k(v)\,dv}{2f(x)} \left(2 \frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s} + f(x) \frac{\partial^2 g(x)}{\partial x_s^2}\right)$

Asymptotic Variance $\frac{1}{nh_1 \ldots h_q} \frac{\sigma^2(x)}{f(x)} \left[\int k(v)^2 dv\right]^q$

- (1) h_s Bias ; Variance
  we have trade-off in choosing kernel bandwidth.
- (2) q Variance exponentially
  We call this "Curse of Dimensionality".
- (3) Kernel more concentrated Bias $\int v^2 k(v)\,dv$; Variance $\int k(v)^2 dv$
- (4) Slope Effect and Curvature Effect on bias: $\frac{\partial f(x)}{\partial x_s} \frac{\partial g(x)}{\partial x_s}, \frac{\partial^2 g(x)}{\partial x_s^2}$
- (5) $f(x)$ Bias ; Variance (more observations)

Non-parametric Method: Kernel Regression

Asymptotic Bias $\overset{q}{\underset{s\ \ 1}{<}}\ h_s^2 \frac{\mathrm{D}\, v^2 k\ \ v\ \ dv}{2f\ \ x}\ \ 2\frac{@f\ \ x}{@x_s}\frac{@g\ \ x}{@x_s}\ \ \ f\ \ x\ \ \frac{@^2 g\ \ x}{@x_s^2}$

Asymptotic Variance $\frac{1}{nh_1 ::: h_q}\ \frac{{}^2\ \ x}{f\ \ x}\ \ \mathrm{D}\, k\ \ v\ \ {}^2 dv\ \ {}^q$

- (1) h_s *Bias* ; *Variance*
  we have trade-off in choosing kernel bandwidth.
- (2) q *Variance* *exponentially*
  We call this "Curse of Dimensionality".
- (3) Kernel more concentrated *Bias* $\mathrm{D}\, v^2 k\ \ v\ \ dv$; *Variance* $\mathrm{D}\, k\ \ v\ \ {}^2 dv$
- (4) Slope Effect and Curvature Effect on bias: $\frac{@f\ \ x}{@x_s}\frac{@g\ \ x}{@x_s};\ \frac{@^2 g\ \ x}{@x_s^2}$
- (5) $f\ \ x$ *Bias* ; *Variance* (more observations)

Non-parametric Method: Kernel Regression

Asymptotic Bias $\overset{q}{\underset{s\ \ 1}{<}} h_s^2 \frac{D\, v^2 k\ \ v\ \ dv}{2f\ \ x}\ \ 2\frac{@f\ \ x}{@x_s}\frac{@g\ \ x}{@x_s}\ \ \ f\ \ x\ \ \frac{@^2 g\ \ x}{@x_s^2}$

Asymptotic Variance $\frac{1}{nh_1 ::: h_q}\frac{{}^2\ \ x}{f\ \ x}\ \ D\, k\ \ v\ \ {}^2 dv\ \ {}^q$

- (1) h_s $Bias$ $; Variance$
  we have trade-off in choosing kernel bandwidth.
- (2) q $Variance$ $exponentially$
  We call this "Curse of Dimensionality".
- (3) Kernel more concentrated $Bias$ $D\, v^2 k\ \ v\ \ dv$ $; Variance$ $D\, k\ \ v\ \ {}^2 dv$
- (4) Slope Effect and Curvature Effect on bias: $\frac{@f\ \ x}{@x_s}\frac{@g\ \ x}{@x_s}; \frac{@^2 g\ \ x}{@x_s^2}$
- (5) $f\ \ x$ $Bias$ $; Variance$ (more observations)

Non-parametric Method: Local Polynomial

- Another widely used kernel-based method is local polynomial
- In linear regression, we use a global linear function to fit data
- In local polynomial, we use piece-wise polynomial (linear) function to fit data interval by interval

Non-parametric Method: Local Polynomial

- Another widely used kernel-based method is local polynomial
- In linear regression, we use a global linear function to fit data
- In local polynomial, we use piece-wise polynomial (linear) function to fit data interval by interval

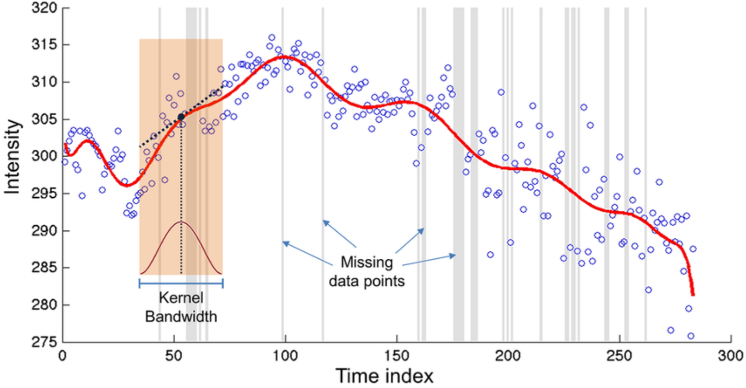
Non-parametric Method: Local Polynomial

- Another widely used kernel-based method is local polynomial
- In linear regression, we use a global linear function to fit data
- In local polynomial, we use piece-wise polynomial (linear) function to fit data interval by interval

Non-parametric Method: Local Polynomial

- Another widely used kernel-based method is local polynomial
- In linear regression, we use a global linear function to fit data
- In local polynomial, we use piece-wise polynomial (linear) function to fit data interval by interval

Non-parametric Method: Local Polynomial



For some X x, we fit g x by choosing samples very close to x. Then we fit a polynomial for these observations. (Here, linear)

Non-parametric Method: Local Polynomial

- For $g(x)$, we solve the following optimization problem at each point x:

$$\min_{b_0, b_1, \cdots, b_p} \sum_{i=1}^{n} k\left(\frac{X_i - x}{h}\right)\left(Y_i - b_0 - b_1(X_i - x) - b_2(X_i - x)^2 - \cdots - b_p(X_i - x)^p\right)^2$$

- When $p = 1$, we call it local linear regression
- When $p = 2$, we call it local quadratic regression

Non-parametric Method: Local Polynomial

- For $g(x)$, we solve the following optimization problem at each point x:

$$\min_{b_0, b_1, \ldots, b_p} \sum_{i=1}^{n} k\left(\frac{X_i - x}{h}\right)\left(Y_i - b_0 - b_1(X_i - x) - b_2(X_i - x)^2 - \cdots - b_p(X_i - x)^p\right)^2$$

- When $p = 1$, we call it local linear regression
- When $p = 2$, we call it local quadratic regression

Non-parametric Method: Local Polynomial

- For $g(x)$, we solve the following optimization problem at each point x:

$$\min_{b_0, b_1, \ldots, b_p} \sum_{i=1}^{n} k\left(\frac{X_i - x}{h}\right)\left(Y_i - b_0 - b_1 (X_i - x) - b_2 (X_i - x)^2 - \cdots - b_p (X_i - x)^p\right)^2$$

- When $p = 1$, we call it local linear regression
- When $p = 2$, we call it local quadratic regression

Non-parametric Method: Local Polynomial

- For $g \quad x$, we solve the following optimization problem at each point x:

$$\min_{b_0;b_1;\quad;b_p} \overset{n}{\underset{i \quad 1}{=}} k \; \frac{X_i \quad x}{h} \quad Y_i \quad b_0 \quad b_1 \; X_i \quad x \quad b_2 \; X_i \quad x \;^{2} \qquad b_p \; X_i \quad x \;^{p \;\; 2}$$

- When p 1, we call it local linear regression
- When p 2, we call it local quadratic regression

Non-parametric Method: Series-based Methods

- Both kernel and local polynomial regressions are Kernel-based methods
- There are three disadvantages of this method:
  - Computational burden is large (point by point estimation)
  - Hard to include information or restriction over functional form
  - Requirement of large sample
- Series-based methods alleviate these problems

Non-parametric Method: Series-based Methods

- Both kernel and local polynomial regressions are Kernel-based methods
- There are three disadvantages of this method:
  - Computational burden is large (point by point estimation)
  - Hard to include information or restriction over functional form
  - Requirement of large sample
- Series-based methods alleviate these problems

Non-parametric Method: Series-based Methods

- Both kernel and local polynomial regressions are Kernel-based methods
- There are three disadvantages of this method:
  - Computational burden is large (point by point estimation)
  - Hard to include information or restriction over functional form
  - Requirement of large sample
- Series-based methods alleviate these problems

Non-parametric Method: Series-based Methods

- Both kernel and local polynomial regressions are Kernel-based methods
- There are three disadvantages of this method:
  - Computational burden is large (point by point estimation)
  - Hard to include information or restriction over functional form
  - Requirement of large sample
- Series-based methods alleviate these problems

Non-parametric Method: Series-based Methods

- Both kernel and local polynomial regressions are Kernel-based methods
- There are three disadvantages of this method:
  - Computational burden is large (point by point estimation)
  - Hard to include information or restriction over functional form
  - Requirement of large sample
- Series-based methods alleviate these problems

Non-parametric Method: Series-based Methods

- Both kernel and local polynomial regressions are Kernel-based methods
- There are three disadvantages of this method:
  - Computational burden is large (point by point estimation)
  - Hard to include information or restriction over functional form
  - Requirement of large sample
- Series-based methods alleviate these problems

Non-parametric Method: Series-based Methods

- Both kernel and local polynomial regressions are Kernel-based methods
- There are three disadvantages of this method:
  - Computational burden is large (point by point estimation)
  - Hard to include information or restriction over functional form
  - Requirement of large sample
- Series-based methods alleviate these problems

Non-parametric Method: Series-based Methods

- As usual, we have a CEF model:

$$Y = g(X) + u$$
$$g(X) = E(Y|X)$$

- We expand the CEF by Taylor Series at zero:

$$g(X) = \sum_{k=0}^{\infty} \frac{g^{(k)}(0)}{k!} X^k$$

Non-parametric Method: Series-based Methods

- As usual, we have a CEF model:

$$Y = g(X) + u$$
$$g(X) = E(Y|X)$$

- We expand the CEF by Taylor Series at zero:

$$g(X) = \sum_{k=0}^{\infty} \frac{g^{(k)}(0)}{k!} X^k$$

Non-parametric Method: Series-based Methods

- As usual, we have a CEF model:

$$Y \quad g \; X \quad u$$
$$g \; X \quad E \; Y|X$$

- We expand the CEF by Taylor Series at zero:

$$g \; X \quad \overset{\infty}{\underset{k \;\; 0}{=}} \frac{g^{\;k} \;\; 0}{k!} X^k$$

Non-parametric Method: Series-based Methods

- This infinite series can be approximated by a K-order global polynomial:

$$g(X) = \sum_{k=0}^{K} \beta_k p_k(X)$$

$$p_0(x) = 1; p_1(x) = x; p_2(x) = x^2; \ldots; p_K(x) = x^K$$

- We can use OLS to estimate this polynomial
- The vector of $[p_0; p_1; p_2; \ldots; p_K]'$ is called "basis"
- This is "global" polynomial, in contrast to "local" polynomial

Non-parametric Method: Series-based Methods

- This infinite series can be approximated by a K-order global polynomial:

$$g\ X \qquad \overset{K}{\underset{k\quad 0}{=}} \qquad {}_k p_k\ X$$

$$p_0\ x \qquad 1; p_1\ x \qquad x; p_2\ x \qquad x^2; :::; p_K\ x \qquad x^K$$

- We can use OLS to estimate this polynomial
- The vector of r$p_0; p_1; p_2; :::; p_K$x is called "basis"
- This is "global" polynomial, in contrast to "local" polynomial

Non-parametric Method: Series-based Methods

- This infinite series can be approximated by a K-order global polynomial:

$$g\ X \quad \overset{K}{\underset{k\ \ 0}{=}} \quad {}_k p_k\ X$$

$$p_0\ x \quad 1; p_1\ x \quad x; p_2\ x \quad x^2; ::: ; p_K\ x \quad x^K$$

- We can use OLS to estimate this polynomial
- The vector of r$p_0; p_1; p_2; :::; p_K$x is called "basis"
- This is "global" polynomial, in contrast to "local" polynomial

Non-parametric Method: Series-based Methods

- This infinite series can be approximated by a K-order global polynomial:

$$g(X) = \sum_{k=0}^{K} \beta_k p_k(X)$$

$$p_0(x) = 1; p_1(x) = x; p_2(x) = x^2; \ldots; p_K(x) = x^K$$

- We can use OLS to estimate this polynomial
- The vector of $[p_0, p_1, p_2, \ldots, p_K]$ is called "basis"
- This is "global" polynomial, in contrast to "local" polynomial

Non-parametric Method: Series-based Methods

- This infinite series can be approximated by a K-order global polynomial:

$$g \;\; X \quad \overset{K}{\underset{k \;\; 0}{=}} \quad {}_k p_k \;\; X$$

$$p_0 \;\; x \qquad 1; p_1 \;\; x \qquad x; p_2 \;\; x \qquad x^2; \ldots; p_K \;\; x \qquad x^K$$

- We can use OLS to estimate this polynomial
- The vector of r$p_0; p_1; p_2; \ldots; p_K$x is called "basis"
- This is "global" polynomial, in contrast to "local" polynomial

Non-parametric Method: Series-based Methods

- Polynomial is the simplest choice of basis
- In multivariate case (2 variables), it becomes: $r1; x_1; x_2; x_1x_2; x_1^2; x_2^2; x_1x_2^2; x_1^2x_2; x_1^2x_2^2 ... x$
- Polynomial series has several problems
- It is very sensitive to outliers
- The biggest problem for polynomial series is Runge's phenomenon

Non-parametric Method: Series-based Methods

- Polynomial is the simplest choice of basis
- In multivariate case (2 variables), it becomes: r1; $x_1; x_2; x_1x_2; x_1^2; x_2^2; x_1x_2^2; x_1^2x_2; x_1^2x_2^2 ::: x$
- Polynomial series has several problems
- It is very sensitive to outliers
- The biggest problem for polynomial series is Runge's phenomenon

Non-parametric Method: Series-based Methods

- Polynomial is the simplest choice of basis
- In multivariate case (2 variables), it becomes:
  r1; x_1; x_2; x_1x_2; x_1^2; x_2^2; $x_1x_2^2$; $x_1^2x_2$; $x_1^2x_2^2$:::x
- Polynomial series has several problems
- It is very sensitive to outliers
- The biggest problem for polynomial series is Runge's phenomenon

Non-parametric Method: Series-based Methods

- Polynomial is the simplest choice of basis
- In multivariate case (2 variables), it becomes: r1; x_1; x_2; x_1x_2; x_1^2; x_2^2; $x_1x_2^2$; $x_1^2x_2$; $x_1^2x_2^2$:::x
- Polynomial series has several problems
- It is very sensitive to outliers
- The biggest problem for polynomial series is Runge's phenomenon

Non-parametric Method: Series-based Methods

- Polynomial is the simplest choice of basis
- In multivariate case (2 variables), it becomes:
  r1; $x_1; x_2; x_1x_2; x_1^2; x_2^2; x_1x_2^2; x_1^2x_2; x_1^2x_2^2 ...$x
- Polynomial series has several problems
- It is very sensitive to outliers
- The biggest problem for polynomial series is Runge's phenomenon

Non-parametric Method: Series-based Methods

- Polynomial is the simplest choice of basis
- In multivariate case (2 variables), it becomes:
  r1; $x_1; x_2; x_1x_2; x_1^2; x_2^2; x_1x_2^2; x_1^2x_2; x_1^2x_2^2 ...$x
- Polynomial series has several problems
- It is very sensitive to outliers
- The biggest problem for polynomial series is Runge's phenomenon

Non-parametric Method: Series-based Methods

- Runge's phenomenon
- Red: original true function; Blue: fifth-order poly; Green: ninth-order poly

- Since the power polynomials are forced to vary somewhere
- It may be pushed to the boundary
- The boundary part is approximated very poorly

Non-parametric Method: Series-based Methods

- Runge's phenomenon
- Red: original true function; Blue: fifth-order poly; Green: ninth-order poly
- Since the power polynomials are forced to vary somewhere
- It may be pushed to the boundary
- The boundary part is approximated very poorly

Non-parametric Method: Series-based Methods

- Runge's phenomenon
- Red: original true function; Blue: fifth-order poly; Green: ninth-order poly

- Since the power polynomials are forced to vary somewhere
- It may be pushed to the boundary
- The boundary part is approximated very poorly

Non-parametric Method: Series-based Methods

- Runge's phenomenon
- Red: original true function; Blue: fth-order poly; Green: ninth-order poly

- Since the power polynomials are forced to vary somewhere
- It may be pushed to the boundary
- The boundary part is approximated very poorly

Non-parametric Method: Series-based Methods

- Runge's phenomenon
- Red: original true function; Blue: fth-order poly; Green: ninth-order poly

- Since the power polynomials are forced to vary somewhere
- It may be pushed to the boundary
- The boundary part is approximated very poorly

Non-parametric Method: Series-based Methods

- Runge's phenomenon
- Red: original true function; Blue: fth-order poly; Green: ninth-order poly

- Since the power polynomials are forced to vary somewhere
- It may be pushed to the boundary
- The boundary part is approximated very poorly

Non-parametric Method: Series-based Methods

- How to choose the optimal order?
- We will discuss this problem in details in the next lecture when considering model selection and machine learning
- But in general, high order polynomial behaves very bad
- Some other basis are better

Non-parametric Method: Series-based Methods

- How to choose the optimal order?
- We will discuss this problem in details in the next lecture when considering model selection and machine learning
- But in general, high order polynomial behaves very bad
- Some other basis are better

Non-parametric Method: Series-based Methods

- How to choose the optimal order?
- We will discuss this problem in details in the next lecture when considering model selection and machine learning
- But in general, high order polynomial behaves very bad
- Some other basis are better

Non-parametric Method: Series-based Methods

- How to choose the optimal order?
- We will discuss this problem in details in the next lecture when considering model selection and machine learning
- But in general, high order polynomial behaves very bad
- Some other basis are better

Non-parametric Method: Series-based Methods

- How to choose the optimal order?
- We will discuss this problem in details in the next lecture when considering model selection and machine learning
- But in general, high order polynomial behaves very bad
- Some other basis are better

Non-parametric Method: Series-based Methods

- Fourier basis, derived by Fourier expansion

- Excellent for approximating periodic functions
- Better than poly, but still not good at boundary/jumping point (Gibbs' phenomenon)

Non-parametric Method: Series-based Methods

- Fourier basis, derived by Fourier expansion

- Excellent for approximating periodic functions
- Better than poly, but still not good at boundary/jumping point (Gibbs' phenomenon)

Non-parametric Method: Series-based Methods

- Fourier basis, derived by Fourier expansion

- Excellent for approximating periodic functions
- Better than poly, but still not good at boundary/jumping point (Gibbs' phenomenon)

Non-parametric Method: Series-based Methods

- Fourier basis, derived by Fourier expansion

- Excellent for approximating periodic functions
- Better than poly, but still not good at boundary/jumping point (Gibbs' phenomenon)

Non-parametric Method: Series-based Methods

- Better than poly, but still not good at boundary/jumping point (Gibbs' phenomenon)
- Let's see an approximation of Fourier series to the square wave

Non-parametric Method: Series-based Methods

- Better than poly, but still not good at boundary/jumping point (Gibbs' phenomenon)
- Let's see an approximation of Fourier series to the square wave

Non-parametric Method: Series-based Methods

- Better than poly, but still not good at boundary/jumping point (Gibbs' phenomenon)
- Let's see an approximation of Fourier series to the square wave

Non-parametric Method: Series-based Methods

- There are more basis
- Such as Spline basis and Wavelet basis
- They are complicated, rarely seen in Applied works
- But Carol claims that Spline basis is in general a better choice
- If interested, you can read her notes

Non-parametric Method: Series-based Methods

- There are more basis
- Such as Spline basis and Wavelet basis
- They are complicated, rarely seen in Applied works
- But Carol claims that Spline basis is in general a better choice
- If interested, you can read her notes

Non-parametric Method: Series-based Methods

- There are more basis
- Such as Spline basis and Wavelet basis
- They are complicated, rarely seen in Applied works
- But Carol claims that Spline basis is in general a better choice
- If interested, you can read her notes

Non-parametric Method: Series-based Methods

- There are more basis
- Such as Spline basis and Wavelet basis
- They are complicated, rarely seen in Applied works
- But Carol claims that Spline basis is in general a better choice
- If interested, you can read her notes

Non-parametric Method: Series-based Methods

- There are more basis
- Such as Spline basis and Wavelet basis
- They are complicated, rarely seen in Applied works
- But Carol claims that Spline basis is in general a better choice
- If interested, you can read her notes

Non-parametric Method: Series-based Methods

- There are more basis
- Such as Spline basis and Wavelet basis
- They are complicated, rarely seen in Applied works
- But Carol claims that Spline basis is in general a better choice
- If interested, you can read her notes

Non-parametric Method: Semi-parametric Model

- Non-parametric model is so general that we do not impose any structure
- Totally data driven, no prior information
- Convergence rate is low, variance is high, requirement for data is high
- What if we want to impose some structure, but not the full structure?
- Semi-parametric model

Non-parametric Method: Semi-parametric Model

- Non-parametric model is so general that we do not impose any structure
- Totally data driven, no prior information
- Convergence rate is low, variance is high, requirement for data is high
- What if we want to impose some structure, but not the full structure?
- Semi-parametric model

Non-parametric Method: Semi-parametric Model

- Non-parametric model is so general that we do not impose any structure
- Totally data driven, no prior information
- Convergence rate is low, variance is high, requirement for data is high
- What if we want to impose some structure, but not the full structure?
- Semi-parametric model

Non-parametric Method: Semi-parametric Model

- Non-parametric model is so general that we do not impose any structure
- Totally data driven, no prior information
- Convergence rate is low, variance is high, requirement for data is high
- What if we want to impose some structure, but not the full structure?
- Semi-parametric model

Non-parametric Method: Semi-parametric Model

- Non-parametric model is so general that we do not impose any structure
- Totally data driven, no prior information
- Convergence rate is low, variance is high, requirement for data is high
- What if we want to impose some structure, but not the full structure?
- Semi-parametric model

Non-parametric Method: Semi-parametric Model

- Non-parametric model is so general that we do not impose any structure
- Totally data driven, no prior information
- Convergence rate is low, variance is high, requirement for data is high
- What if we want to impose some structure, but not the full structure?
- Semi-parametric model

Non-parametric Method: Semi-parametric Model

- Partially linear model
- One of the most popular semi-parametric models

$$Y \quad X^\top \quad g \; Z \quad u; \quad E \; u|X;Z \quad 0; \; Var \; u|X;Z \quad ^2$$

- X enters in the model linearly, Z non-parametrically

Non-parametric Method: Semi-parametric Model

- Partially linear model
- One of the most popular semi-parametric models

$$Y \quad X^{\top} \quad g\ Z \quad u; \quad E\ u|X;Z \quad 0; Var\ u|X;Z \quad {}^{2}$$

- X enters in the model linearly, Z non-parametrically

Non-parametric Method: Semi-parametric Model

- Partially linear model
- One of the most popular semi-parametric models

$$Y = X^{\top}\beta + g(Z) + u; \quad E(u|X, Z) = 0; \; Var(u|X, Z) = \sigma^2$$

- X enters in the model linearly, Z non-parametrically

Non-parametric Method: Semi-parametric Model

- Partially linear model
- One of the most popular semi-parametric models

$$Y = X^{\top}\beta + g(Z) + u; \quad E(u|X;Z) = 0; Var(u|X;Z) = \sigma^2$$

- X enters in the model linearly, Z non-parametrically

Non-parametric Method: Semi-parametric Model

- Estimation of β is simple, we follow Robinson (1988)
- In the first step, conditional on Z and then take the subtract:

$$E(Y|Z) = E(X|Z)^{\top}\beta + g(Z)$$

$$Y - E(Y|Z) = (X - E(X|Z))^{\top}\beta + u$$

- $E(Y|Z)$ and $E(X|Z)$ can be estimated using methods introduced previously
- Then we have estimators for $Y - E(Y|Z)$ and $X - E(X|Z)$
- Then we can estimate β using OLS
- Asymptotics of this estimator is complicated

Non-parametric Method: Semi-parametric Model

- Estimation of is simple, we follow Robinson (1988)
- In the rst step, conditional on Z and then take the subtract:

$$E\ Y¶Z \quad E\ X^{\top}¶Z \quad g\ Z$$

$$Y \quad E\ Y¶Z \quad X \quad E\ X¶Z\ ^{\top} \quad u$$

- E Y¶Z and E X¶Z can be estimated using methods introduced previously
- Then we have estimators for Y E Y¶Z and X E X¶Z
- Then we can estimate using OLS
- Asymptotics of this estimator is complicated

Non-parametric Method: Semi-parametric Model

- Estimation of   is simple, we follow Robinson (1988)
- In the rst step, conditional on Z and then take the subtract:

$$E\ Y|Z \qquad E\ X^{\neg}|Z \qquad g\ Z$$

$$Y \quad E\ Y|Z \qquad X \quad E\ X|Z\ ^{\neg} \qquad u$$

- $E\ Y|Z$ and $E\ X|Z$ can be estimated using methods introduced previously
- Then we have estimators for $Y \quad E\ Y|Z$ and $X \quad E\ X|Z$
- Then we can estimate using OLS
- Asymptotics of this estimator is complicated

Non-parametric Method: Semi-parametric Model

- Estimation of is simple, we follow Robinson (1988)
- In the first step, conditional on Z and then take the subtract:

$$E\ Y|Z \quad E\ X^{\neg}|Z \quad g\ Z$$

$$Y \quad E\ Y|Z \quad X \quad E\ X|Z\ ^{\neg} \quad u$$

- $E\ Y|Z$ and $E\ X|Z$ can be estimated using methods introduced previously
- Then we have estimators for $Y \quad E\ Y|Z$ and $X \quad E\ X|Z$
- Then we can estimate using OLS
- Asymptotics of this estimator is complicated

Non-parametric Method: Semi-parametric Model

- Estimation of β is simple, we follow Robinson (1988)
- In the first step, conditional on Z and then take the subtract:

$$E(Y|Z) = E(X|Z)'\beta + g(Z)$$
$$Y - E(Y|Z) = (X - E(X|Z))'\beta + u$$

- $E(Y|Z)$ and $E(X|Z)$ can be estimated using methods introduced previously
- Then we have estimators for $Y - E(Y|Z)$ and $X - E(X|Z)$
- Then we can estimate β using OLS
- Asymptotics of this estimator is complicated

Non-parametric Method: Semi-parametric Model

- Estimation of is simple, we follow Robinson (1988)
- In the rst step, conditional on Z and then take the subtract:

$$E \; Y|Z \qquad E \; X'|Z \qquad g \; Z$$

$$Y \quad E \; Y|Z \qquad X \quad E \; X|Z \; ' \qquad u$$

- $E\;Y|Z$ and $E\;X|Z$ can be estimated using methods introduced previously
- Then we have estimators for $Y \quad E\;Y|Z$ and $X \quad E\;X|Z$
- Then we can estimate using OLS
- Asymptotics of this estimator is complicated

Non-parametric Method: Semi-parametric Model

- Estimation of   is simple, we follow Robinson (1988)
- In the rst step, conditional on Z and then take the subtract:

$$E(Y|Z) \quad E(X|Z)' \quad g(Z)$$

$$Y \quad E(Y|Z) \quad X \quad E(X|Z)' \quad u$$

- $E(Y|Z)$ and $E(X|Z)$ can be estimated using methods introduced previously
- Then we have estimators for $Y \quad E(Y|Z)$ and $X \quad E(X|Z)$
- Then we can estimate using OLS
- Asymptotics of this estimator is complicated

Non-parametric Method: Semi-parametric Model

- In the second step, we subtract $X^\top$ from Y:

$$Y \quad X^\top \quad g \; Z \quad u$$

- g(Z) can be estimated using methods introduced previously

Non-parametric Method: Semi-parametric Model

- In the second step, we subtract $X^{\neg}$ from Y:

$$Y \quad X^{\neg} \quad g\ Z \quad u$$

- g(Z) can be estimated using methods introduced previously

Non-parametric Method: Semi-parametric Model

- In the second step, we subtract $X^{\top}$ from Y:

$$Y \quad X^{\top} \quad g \; Z \quad u$$

- g(Z) can be estimated using methods introduced previously

Non-parametric Method: Semi-parametric Model

- Question: How to estimate the variance of $\hat{Z}$?
- Can we use the variance from the non-parametric regression directly?
- No! Because $Y \quad X^{\top}$ is also estimated
- It contains more uncertainty from the first step
- This is a common mistake in empirical work:
  When you have first stage estimation as known parameter in the second stage, watch out for the std err estimation!

Non-parametric Method: Semi-parametric Model

- Question: How to estimate the variance of $\hat{Z}$?
- Can we use the variance from the non-parametric regression directly?
- No! Because $Y - X^{\top}\hat{\beta}$ is also estimated
- It contains more uncertainty from the first step
- This is a common mistake in empirical work:
  When you have first stage estimation as known parameter in the second stage, watch out for the std err estimation!

Non-parametric Method: Semi-parametric Model

- Question: How to estimate the variance of $\hat{Z}$?
- Can we use the variance from the non-parametric regression directly?
- No! Because $Y - X^{\top}\hat{\beta}$ is also estimated
- It contains more uncertainty from the first step
- This is a common mistake in empirical work:
  When you have first stage estimation as known parameter in the second stage, watch out for the std err estimation!

Non-parametric Method: Semi-parametric Model

- Question: How to estimate the variance of $\hat{Z}$?
- Can we use the variance from the non-parametric regression directly?
- No! Because $Y \quad X$ is also estimated
- It contains more uncertainty from the first step
- This is a common mistake in empirical work:
  When you have first stage estimation as known parameter in the second stage, watch out for the std err estimation!

Non-parametric Method: Semi-parametric Model

- Question: How to estimate the variance of $\hat{Z}$?
- Can we use the variance from the non-parametric regression directly?
- No! Because $Y \quad X^{\urcorner}$ is also estimated
- It contains more uncertainty from the rst step
- This is a common mistake in empirical work:
  When you have rst stage estimation as known parameter in the second stage, watch out for the std err estimation!

Non-parametric Method: Semi-parametric Model

- Question: How to estimate the variance of $\hat{Z}$?
- Can we use the variance from the non-parametric regression directly?
- No! Because Y $X^{\neg}$ is also estimated
- It contains more uncertainty from the rst step
- This is a common mistake in empirical work:
  When you have rst stage estimation as known parameter in the second stage, watch out for the std err estimation!

Non-parametric Method: Semi-parametric Model

- Similarly, how to conduct inference for rst step?
- It is a combination of non-parametric and regression estimations
- No closed-form variance equation is available
- Not possible to directly calculate the standard error

Non-parametric Method: Semi-parametric Model

- Similarly, how to conduct inference for rst step?
- It is a combination of non-parametric and regression estimations
- No closed-form variance equation is available
- Not possible to directly calculate the standard error

Non-parametric Method: Semi-parametric Model

- Similarly, how to conduct inference for rst step?
- It is a combination of non-parametric and regression estimations
- No closed-form variance equation is available
- Not possible to directly calculate the standard error

Non-parametric Method: Semi-parametric Model

- Similarly, how to conduct inference for rst step?
- It is a combination of non-parametric and regression estimations
- No closed-form variance equation is available
- Not possible to directly calculate the standard error

Non-parametric Method: Semi-parametric Model

- Similarly, how to conduct inference for rst step?
- It is a combination of non-parametric and regression estimations
- No closed-form variance equation is available
- Not possible to directly calculate the standard error

Non-parametric Method: Semi-parametric Model

In these two cases, we need bootstrap for inference

Non-parametric Method: Bootstrap

- Bootstrap is a non-parametric method for inference
- It is used when there is no closed-form standard errors
- Instead of deriving the closed-form equation of variance
- We use simulation to estimate it
- Random sampling with replacement

Non-parametric Method: Bootstrap

- Bootstrap is a non-parametric method for inference
- It is used when there is no closed-form standard errors
- Instead of deriving the closed-form equation of variance
- We use simulation to estimate it
- Random sampling with replacement

Non-parametric Method: Bootstrap

- Bootstrap is a non-parametric method for inference
- It is used when there is no closed-form standard errors
- Instead of deriving the closed-form equation of variance
- We use simulation to estimate it
- Random sampling with replacement

Non-parametric Method: Bootstrap

- Bootstrap is a non-parametric method for inference
- It is used when there is no closed-form standard errors
- Instead of deriving the closed-form equation of variance
- We use simulation to estimate it
- Random sampling with replacement

Non-parametric Method: Bootstrap

- Bootstrap is a non-parametric method for inference
- It is used when there is no closed-form standard errors
- Instead of deriving the closed-form equation of variance
- We use simulation to estimate it
- Random sampling with replacement

Non-parametric Method: Bootstrap

- Bootstrap is a non-parametric method for inference
- It is used when there is no closed-form standard errors
- Instead of deriving the closed-form equation of variance
- We use simulation to estimate it
- Random sampling with replacement

Non-parametric Method: Bootstrap

- Step 1: Given full sample with size n, draw R new samples of size n, with replacement. Index each new sample by r
- Step 2: Calculate the simulated variance of $\hat{g}(x)$ by:
  $\hat{V}(x) = \frac{1}{R-1}\sum_{r=1}^{R}\left(\hat{g}_r(x) - \hat{g}(x)\right)^2$
- Step 3: Use $\hat{V}(x)$ to calculate confidence intervals and implement statistical tests
- We call this bootstrapped variance

Non-parametric Method: Bootstrap

- Step 1: Given full sample with size n, draw R new samples of size n, with replacement. Index each new sample by r
- Step 2: Calculate the simulated variance of $g(\hat{x})$ by:
  $\hat{V}(x) = \frac{1}{R-1}\sum_{r=1}^{R}\left(\hat{g}_r(x) - \hat{g}(x)\right)^2$
- Step 3: Use $\hat{V}(x)$ to calculate confidence intervals and implement statistical tests
- We call this bootstrapped variance

Non-parametric Method: Bootstrap

- Step 1: Given full sample with size n, draw R new samples of size n, with replacement. Index each new sample by r.
- Step 2: Calculate the simulated variance of $\hat{g}(x)$ by:

$$\hat{V}(\hat{g}(x)) = \frac{1}{R-1}\sum_{r=1}^{R}\left(\hat{g}_r(x) - \hat{g}(x)\right)^2$$

- Step 3: Use $\hat{V}(\hat{g}(x))$ to calculate confidence intervals and implement statistical tests
- We call this bootstrapped variance

Non-parametric Method: Bootstrap

- Step 1: Given full sample with size n, draw R new samples of size n, with replacement. Index each new sample by r.
- Step 2: Calculate the simulated variance of $\hat{g}(x)$ by:
  $\hat{V}(x) = \frac{1}{R-1}\sum_{r=1}^{R}\left(\hat{g}_r(x) - \hat{g}(x)\right)^2$
- Step 3: Use $\hat{V}(x)$ to calculate confidence intervals and implement statistical tests
- We call this bootstrapped variance

Non-parametric Method: Bootstrap

- Step 1: Given full sample with size n, draw R new samples of size n, with replacement. Index each new sample by r
- Step 2: Calculate the simulated variance of $\hat{g}(x)$ by:
  $\hat{V}(x) = \frac{1}{R-1}\sum_{r=1}^{R}\left(\hat{g}_r(x) - \hat{g}(x)\right)^2$
- Step 3: Use $\hat{V}(x)$ to calculate confidence intervals and implement statistical tests
- We call this bootstrapped variance

Non-parametric Method: Bootstrap

- But using bootstrapped variance to construct confidence interval is a poor choice
- It relies on asymptotic normality, which is not accurate in finite sample
- A better chioce is "percentile interval"
- First, we stack the sample of bootstrap estimates $\hat{\theta}^1, \hat{\theta}^2, \ldots, \hat{\theta}^R$
- We have an empirical distribution of $\hat{\theta}^r$
- The bootstrap $100(1-\alpha)$% confidence interval is then: $\left[q^*_{\alpha/2}, q^*_{1-\alpha/2} \right]$
- q^* is the quantile of this empirical distribution

Non-parametric Method: Bootstrap

- But using bootstrapped variance to construct confidence interval is a poor choice
- It relies on asymptotic normality, which is not accurate in finite sample
- A better chioce is "percentile interval"
- First, we stack the sample of bootstrap estimates $\hat\theta^1, \hat\theta^2, \ldots, \hat\theta^R$
- We have an empirical distribution of $\hat\theta^r$
- The bootstrap $100(1-\alpha)$% confidence interval is then: $[\tilde q_{\alpha/2}, \tilde q_{1-\alpha/2}]$
- $\tilde q$ is the quantile of this empirical distribution

Non-parametric Method: Bootstrap

- But using bootstrapped variance to construct confidence interval is a poor choice
- It relies on asymptotic normality, which is not accurate in finite sample
- A better chioce is "percentile interval"
- First, we stack the sample of bootstrap estimates $\hat{\theta}^1, \hat{\theta}^2, \ldots, \hat{\theta}^R$
- We have an empirical distribution of $\hat{\theta}^r$
- The bootstrap $100(1-\alpha)$% confidence interval is then: $[\tilde{q}_{\alpha/2}, \tilde{q}_{1-\alpha/2}]$
- $\tilde{q}$ is the quantile of this empirical distribution

Non-parametric Method: Bootstrap

- But using bootstrapped variance to construct confidence interval is a poor choice
- It relies on asymptotic normality, which is not accurate in finite sample
- A better chioce is "percentile interval"
- First, we stack the sample of bootstrap estimates $\hat{\theta}^1, \hat{\theta}^2, \ldots, \hat{\theta}^R$
- We have an empirical distribution of $\hat{\theta}^r$
- The bootstrap $100(1-\alpha)$% confidence interval is then: $[\tilde{q}_{\alpha/2}, \tilde{q}_{1-\alpha/2}]$
- $\tilde{q}$ is the quantile of this empirical distribution

Non-parametric Method: Bootstrap

- But using bootstrapped variance to construct con dence interval is a poor choice
- It relies on asymptotic normality, which is not accurate in nite sample
- A better chioce is "percentile interval"
- First, we stack the sample of bootstrap estimates $\hat{}^{1}; \hat{}^{2}; \ldots; \hat{}^{R}$ x
- We have an empirical distribution of $\hat{}^{r}$
- The bootstrap 1001 % con dence interval is then: $\tilde{q}_{©2}; \tilde{q}_{1}$ ©2
- $\tilde{q}$ is the quantile of this empirical distribution

Non-parametric Method: Bootstrap

- But using bootstrapped variance to construct confidence interval is a poor choice
- It relies on asymptotic normality, which is not accurate in finite sample
- A better chioce is "percentile interval"
- First, we stack the sample of bootstrap estimates $\{\hat{\theta}^1, \hat{\theta}^2, \ldots, \hat{\theta}^R\}$
- We have an empirical distribution of $\hat{\theta}^r$
- The bootstrap $100(1-\alpha)$% confidence interval is then: $[\tilde{q}_{\alpha/2}, \tilde{q}_{1-\alpha/2}]$
- $\tilde{q}$ is the quantile of this empirical distribution

Non-parametric Method: Bootstrap

- But using bootstrapped variance to construct confidence interval is a poor choice
- It relies on asymptotic normality, which is not accurate in finite sample
- A better chioce is "percentile interval"
- First, we stack the sample of bootstrap estimates $\{\hat\theta^1, \hat\theta^2, \ldots, \hat\theta^R\}$
- We have an empirical distribution of $\hat\theta^r$
- The bootstrap $100(1-\alpha)\%$ confidence interval is then: $[\tilde q_{\alpha/2}, \tilde q_{1-\alpha/2}]$
- $\tilde q$ is the quantile of this empirical distribution

Non-parametric Method: Bootstrap

- But using bootstrapped variance to construct con dence interval is a poor choice
- It relies on asymptotic normality, which is not accurate in nite sample
- A better chioce is "percentile interval"
- First, we stack the sample of bootstrap estimates $\hat{}^1; \hat{}^2; :::; \hat{}^R$ x
- We have an empirical distribution of $\hat{}^r$
- The bootstrap 1001   % con dence interval is then: $\tilde{q}_{©2}; \tilde{q}_{1 \quad ©2}$
- $\tilde{q}$ is the quantile of this empirical distribution

Non-parametric Method: Application

- Where to apply non-parametric methods?
- Anything related to estimation of CEF
- Potential outcome framework is non-parametric
- Non-parametric inference in complicated models (Bootstrap)
- If you focus on prediction and t, but not the structure behind it
  Predict stock price, machine learning, RDD tting
- We will show these in the following lectures

Non-parametric Method: Application

- Where to apply non-parametric methods?
- Anything related to estimation of CEF
- Potential outcome framework is non-parametric
- Non-parametric inference in complicated models (Bootstrap)
- If you focus on prediction and t, but not the structure behind it
  Predict stock price, machine learning, RDD tting
- We will show these in the following lectures

Non-parametric Method: Application

- Where to apply non-parametric methods?
- Anything related to estimation of CEF
- Potential outcome framework is non-parametric
- Non-parametric inference in complicated models (Bootstrap)
- If you focus on prediction and t, but not the structure behind it
  Predict stock price, machine learning, RDD tting
- We will show these in the following lectures

Non-parametric Method: Application

- Where to apply non-parametric methods?
- Anything related to estimation of CEF
- Potential outcome framework is non-parametric
- Non-parametric inference in complicated models (Bootstrap)
- If you focus on prediction and t, but not the structure behind it
  Predict stock price, machine learning, RDD tting
- We will show these in the following lectures

Non-parametric Method: Application

- Where to apply non-parametric methods?
- Anything related to estimation of CEF
- Potential outcome framework is non-parametric
- Non-parametric inference in complicated models (Bootstrap)
- If you focus on prediction and t, but not the structure behind it
  Predict stock price, machine learning, RDD tting
- We will show these in the following lectures

Non-parametric Method: Application

- Where to apply non-parametric methods?
- Anything related to estimation of CEF
- Potential outcome framework is non-parametric
- Non-parametric inference in complicated models (Bootstrap)
- If you focus on prediction and t, but not the structure behind it
  Predict stock price, machine learning, RDD tting
- We will show these in the following lectures

Non-parametric Method: Application

- Where to apply non-parametric methods?
- Anything related to estimation of CEF
- Potential outcome framework is non-parametric
- Non-parametric inference in complicated models (Bootstrap)
- If you focus on prediction and t, but not the structure behind it
  Predict stock price, machine learning, RDD tting
- We will show these in the following lectures

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
  - Kernel-based methods: N-W estimator, Local polynomial
  - Series-based methods: Polynomial, Fourier, Splines, Wavelet
- They are very useful in causal inference to directly estimate CEF
- However, they have weaknesses: Not always better to make model more exible
  - Hard to incorporate restrictions
  - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
  - Kernel-based methods: N-W estimator, Local polynomial
  - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful in causal inference to directly estimate CEF
- However, they have weaknesses: Not always better to make model more flexible
  - Hard to incorporate restrictions
  - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
  - Kernel-based methods: N-W estimator, Local polynomial
  - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful in causal inference to directly estimate CEF
- However, they have weaknesses: Not always better to make model more exible
  - Hard to incorporate restrictions
  - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
  - Kernel-based methods: N-W estimator, Local polynomial
  - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful in causal inference to directly estimate CEF
- However, they have weaknesses: Not always better to make model more exible
  - Hard to incorporate restrictions
  - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
  - Kernel-based methods: N-W estimator, Local polynomial
  - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful in causal inference to directly estimate CEF
- However, they have weaknesses: Not always better to make model more flexible
  - Hard to incorporate restrictions
  - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
  - Kernel-based methods: N-W estimator, Local polynomial
  - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful in causal inference to directly estimate CEF
- However, they have weaknesses: Not always better to make model more exible
  - Hard to incorporate restrictions
  - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
  - Kernel-based methods: N-W estimator, Local polynomial
  - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful in causal inference to directly estimate CEF
- However, they have weaknesses: Not always better to make model more exible
  - Hard to incorporate restrictions
  - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
  - Kernel-based methods: N-W estimator, Local polynomial
  - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful in causal inference to directly estimate CEF
- However, they have weaknesses: Not always better to make model more exible
  - Hard to incorporate restrictions
  - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
  - Kernel-based methods: N-W estimator, Local polynomial
  - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful in causal inference to directly estimate CEF
- However, they have weaknesses: Not always better to make model more flexible
  - Hard to incorporate restrictions
  - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
  - Kernel-based methods: N-W estimator, Local polynomial
  - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful in causal inference to directly estimate CEF
- However, they have weaknesses: Not always better to make model more flexible
  - Hard to incorporate restrictions
  - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

Final Conclusion

- There are statistical modeling methods other than Linear regression
- Non-parametric methods impose no prior structure, totally data-driven
  - Kernel-based methods: N-W estimator, Local polynomial
  - Series-based methods: Polynomial, Fourier, Spline, Wavelet
- They are very useful in causal inference to directly estimate CEF
- However, they have weaknesses: Not always better to make model more flexible
  - Hard to incorporate restrictions
  - Require large sample size to have accurate estimation
- We will discuss more about it next week
- A semi-parametric model is between non-parametric and parametric

References

Robinson, Peter M. 1988. "Root-N-consistent Semiparametric Regression." *Econometrica* :931–954.