BAB II

TEORI PENUNJANG

2.1. VARIABEL RANDOM

Definisi 2.1.1 :

Variabel random X adalah suatu fungsi berharga real dengan domain Ω (yaitu masing-masing $\omega \in \Omega$, $X \in \mathbb{R} = \{y : -\infty < y < +\infty\}$).

Definisi 2.1.2 :

X disebut variabel random kontinu (absolut) satu dimensi jika d.f F(x) dapat dinyatakan sebagai:

$$F(x) = \int_{-\infty}^{x} f(y)dy, \qquad (-\infty < x < +\infty)$$

di mana $f(y)$ adalah fungsi densitas probabilitas (pdf) yang didefinisikan sebagai :

i) $f(y) \geq 0 \;\; \forall \; y$

ii) $\int_{-\infty}^{\infty} f(y)dy = 1$

Contoh 2.1.

Misalkan diberikan suatu fungsi

$$g(x) = \begin{cases} 3(1-x)^2, & 0<x<1 \\ 0 & , x \text{ yang lain} \end{cases}$$

Akan dibuktikan bahwa g(x) merupakan fungsi densitas.

i). Untuk 0<x<1 dan x yang lainnya, jelas bahwa

$g(x) \geq 0$

ii). $\int_{-\infty}^{\infty} g(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{1} 3(1-x)^2 dx + \int_{1}^{\infty} 0dx$

$= 3x - 3x^2 + x^3 |_0^1$

$= 1$ (syarat (ii) terpenuhi)

Dengan demikian g(x) merupakan fungsi densitas.

Definisi 2.1.3 :

Misalkan X dan Y adalah variabel random kontinu berserikat. Fungsi $f(x,y)$ dikatakan fungsi densitas bersama jika :

i) $f(x,y) \geq 0, \forall x,y$

ii). $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dxdy = 1$

iii). $P[(x,y) \in A] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)dxdy \quad \forall$ daerah A di bidang xy.

Contoh 2.2.

Diberikan fungsi $f(x,y) = \begin{cases} \frac{2}{5}(2x+3y) & , 0<x<1 \\ & 0<y<1 \\ 0 & , \text{yang lain} \end{cases}$

Akan ditunjukkan bahwa $f(x,y)$ merupakan fungsi densitas bersama.

i). Untuk $0 \leq x \leq 1$ dan $0 \leq y \leq 1$ jelas bahwa $f(x,y) \geq 0$ sehingga syarat i) terpenuhi.

ii). $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = \int_0^1\int_0^1 \frac{2}{5}(2x+3y)dxdy$

$$= \int_0^1 \left(\frac{2}{5}x^2 + \frac{6}{5}xy\right)\Big|_0^1 dy$$

$$= \int_0^1 \left(\frac{2}{5} + \frac{6}{5}y\right)dy$$

$$= \left(\frac{2}{5}y + \frac{3}{5}y\right)\Big|_0^1$$

= 1 (syarat (ii) terpenuhi)

iii). Karena A adalah daerah $\{(x,y)|0\le x\le 1, 0\le y\le 1\}$

maka $P[(x,y)\in A] = P[0\le x\le 1, 0\le y\le 1]$

$$= \int_0^1\int_0^1 \frac{2}{5}(2x+3y)dxdy$$

$$= 1$$

Jadi $f(x,y)$ merupakan fungsi densitas gabungan.

Definisi 2.1.4 :

Misalkan X dan Y variabel random kontinu bersama dengan fungsi densitas bersama $f(x,y)$. Fungsi densitas bersyarat dari Y untuk X=x dinyatakan dengan $f(y|x)$ didefinisikan sebagai :

$$f(y|x) = \frac{f(x,y)}{f(x)}$$

Definisi 2.1.5 :

Misalkan X dan Y variabel random kontinu bersama dengan fungsi densitas bersama $f(x,y)$. Maka fungsi densitas marginal dari X dan Y masing-masing didefinisikan sebagai :

i). $f(x) = \int_{-\infty}^{\infty} f(x,y)dy$ dan $f(y) = \int_{-\infty}^{\infty} f(x,y)dx$

ii). $\int_{-\infty}^{\infty} f(x)dx = 1$ dan $\int_{-\infty}^{\infty} f(y)dy = 1$

Contoh :2.3 (pada contoh 2.2)

$$f(x) = \int_0^1 \frac{2}{5}(2x+3y)dy$$

$$= \frac{4}{5}xy + \frac{3}{5}y^2\Big|_0^1$$

$= \frac{4}{5}x + \frac{3}{5}$, untuk $0 \le x \le 1$ dan $f(x) = 0$ untuk x yang lain.

Begitu pula untuk $f(y) = \int_0^1 \frac{2}{5}(2x+3y)dx$

$$= \frac{2}{5}x^2 + \frac{6}{5}xy\Big|_0^1$$

$$= \frac{2}{5} + \frac{6}{5}y$$

Selanjutnya :

$$\int_0^1 \left(\frac{4}{5}x + \frac{3}{5}\right)dx = \frac{2}{5}x^2 + \frac{3}{5}x\Big|_0^1$$

$$= \frac{2}{5} + \frac{3}{5} = 1$$

$$\int_0^1 \left(\frac{2}{5} + \frac{6}{5}y\right)dy = \frac{2}{5}y + \frac{3}{5}y^2\Big|_0^1$$

$$= \frac{2}{5} + \frac{3}{5} = 1$$

Jadi $f(x)$ dan $f(y)$ merupakan fungsi densitas marginal.

2.2 SIMETRIS DAN KONTINUITAS

Definisi 2.2.1 :

Suatu fungsi densitas dikatakan fungsi simetri di sekitar μ jika nilai dari $f(x)$ pada $\mu+x$ sama dengan nilai $f(x)$ pada $\mu-x$ untuk setiap x.

Dengan kata lain $f(\mu+x) = f(\mu-x)$.

Contoh 2.4.

Fungsi Kernel Quartic $K(U) = \frac{15}{16}(1-U^2)^2$ merupakan fungsi yang simetri terhadap 0.

Akan ditunjukkan bahwa $K(U)=K(-U)$

Misalkan diambil $U=\frac{1}{2}$, maka :

$$K(\frac{1}{2}) = K(-\frac{1}{2})$$

$$\frac{15}{16}(1-(\frac{1}{2})^2)^2 = \frac{15}{16}(1-(-\frac{1}{2})^2)^2$$

$$\frac{15}{16}(\frac{3}{4})^2 = \frac{15}{16}(\frac{3}{4})^2$$

Jadi Kernel Quartic K(U) adalah simetri di 0

Definisi 2.2.2 :

Suatu fungsi $f(x)$ dikatakan kontinu di $x=x_o$ jika :

i). $f(x_o)$ terdefinisi

ii). $\lim_{x \to x_o} f(x_o)$ ada

iii). $\lim_{x \to x_o} f(x_o) = f(x_o)$

Contoh 2.5.

Andaikan diambil $f(x)$ = fungsi Kernel Quartic K(U).

Fungsi ini kontinu di $x=\frac{1}{2}$

Harus ditunjukkan bahwa ketiga syarat dipenuhi.

i). $K(\frac{1}{2}) = \frac{15}{16}(1-(\frac{1}{2})^2)^2 = \frac{15}{16}\cdot\frac{9}{16}$

($K(\frac{1}{2})$ terdefinisi)

ii). $\lim\limits_{x\to x_o} \frac{15}{16}(1-(\frac{1}{2})^2)^2 = \frac{15}{16}\cdot\frac{9}{16}$ (ada)

iii). Dari (i) dan (ii) jelas bahwa $\lim\limits_{x\to x_o} f(x_o) = f(x_o)$. Jadi Kernel Quartic K(U) merupakan fungsi yang kontinu di $x=\frac{1}{2}$.

2.3 EKSPEKTASI DAN VARIANSI

Definisi 2.3.1 :

Jika X adalah variabel random maka ekspektasi dari X, ditulis sebagai E[x] didefinisikan sebagai :

$$E[x] = \int_{-\infty}^{\infty} x f(x)dx$$, dengan $f(x)$ merupakan fungsi densitas.

Definisi 2.3.2 :

Jika F(y|x) adalah kontinu (absolut) maka mean (rata-rata) bersyarat dari Y dengan diberikan x adalah :

$$E[Y|X=x] = \int_{-\infty}^{\infty} y f(y|x)dy$$

Definisi 2.3.3 :

Misalkan X variabel random dan $\mu = E[x]$. Varian dari X dinyatakan Var[x] didefinisikan sebagai :

$$Var[x] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx, \quad \text{dengan fungsi densitas } f(x)$$

$$= E[x-\mu]$$

2.4. SIFAT-SIFAT ESTIMATOR

Definisi 2.4.1 (Tak Bias)

$t_n(x_1, \ldots, x_n)$ adalah estimator tak bias (unbias) untuk ϑ jika $E(t_n) = \vartheta$

Kriteria ini menyatakan bahwa rata-rata semua harga $t_n(x_1, \ldots, x_n)$ yang mungkin akan sama dengan ϑ.

Sementara itu estimator yang tidak unbias disebut estimator bias. Jika $t_n(x_1, \ldots x_n)$ adalah estimator bias untuk ϑ dan bias dari estimator yang dinyatakan sebagai $B(t_n)$ maka :

$$E(t_n) - \vartheta = B(t_n)$$

Sehingga Bias estimator $B(t_n) = E(t_n) - \vartheta$

$$= E(t_n - \vartheta)$$

Definisi 2.4.2 :

$t_n(x_1, \ldots, x_n)$ dikatakan estimator tak bias asymtotis untuk ϑ jika $\lim_{n \to \infty} B(t_n) = 0$

Definisi 2.4.3 :(Konsisten)

Jika $X_1,\ldots,X_n$ variabel randomdengan d.f $F(x_1,\ldots,x_n|\vartheta)$ dengan ϑ tak diketahui maka barisan $t_1, t_2,\ldots,t_n$ dinamakan barisan estimator konsisten untuk ϑ jika :

$$t_n \xrightarrow{P} \vartheta$$

artinya $\lim\limits_{n\to\infty} P(|t_n-\vartheta|<\varepsilon) = 1$

Contoh 2.6.

Misalkan $X_1, X_2, \ldots,X_n$ iid dengan mean μ dan varian σ^2. Misalkan juga $\bar{X} = \dfrac{X_1+X_2+\ldots+X_n}{n}$

$$S^2 = \frac{\Sigma(X_i-\bar{X})^2}{n-1}$$

$$U^2 = \frac{\Sigma(X_i-\bar{X})^2}{n}$$

maka :

$$E[\bar{X}] = E\left[\frac{\Sigma X_i}{n}\right] = \frac{1}{n}nE[X] = \mu$$

$$E[S_n^2] = E\left[\frac{\Sigma(X_i-\bar{X})^2}{n-1}\right] = \frac{1}{n-1}\left[\Sigma(X_i-\bar{X})^2\right]$$

$$= \frac{1}{n-1}\, n(\mu^2-n\bar{X})-n(\frac{\sigma^2}{n}+\mu^2)$$

$$= \sigma^2$$

Jadi $\bar{X}$ dan S^2 merupakan estimator tak bias untuk μ dan σ^2. Sedangkan U_n merupakan estimator yang tidak tak bias karena :

$$E[U_n] = \frac{1}{n} E\left[\Sigma(X_i - \overline{X})^2 \right]$$

$$= \frac{1}{n} (n-1)E[S^2]$$

$$= \frac{n-1}{n}\sigma^2$$

Definisi 2.4.4 :(Mean Square Error atau MSE)

$t_n(X_1,\ldots,X_n)$ adalah estimator untuk ϑ yang didasarkan pada random variabel $X_1,X_2,\ldots,X_n$ maka MSE didefinisikan sebagai :

$$MSE[t_n] = E[(t_n-\vartheta)^2]$$

Atau dapat ditulis dalam bentuk lain sebagai:

$$MSE[t_n] = E[(t_n - E(t_n)+E(t_n)-\vartheta)^2]$$

$$= E[(t_n - E(t_n))+(E(t_n)-\vartheta))^2]$$

$$= E[(t_n - E(t_n))^2+2(t_n-E(t_n))(E(t_n)-\vartheta)+ (E(t_n)-\vartheta)^2]$$

$$= E[(t_n - E(t_n)]^2+(E(t_n)-\vartheta))^2$$

$$= Var(t_n) + (Bias)^2$$

Jadi $MSE[t_n]$ sama dengan variansi ditambah bias kuadrat. Jika t_n adalah estimator yang tak bias maka MSE sama dengan variansinya.

Definisi 2.4.5 :

$t_n(X_1,X_2,\ldots,X_n)$ adalah estimator untuk ϑ dikatakan konsisten dalam mean square (kuadrat rata-rata) jika

$$E[t_n-\vartheta]^2 \xrightarrow{p} 0 \text{ untuk } n \to \infty$$

Teorema 2.4.6 :

Jika $E[t_n-\vartheta]^2 \xrightarrow{p} 0$ untuk setiap ϑ maka $t_n \xrightarrow{p} \vartheta$ untuk setiap ϑ.

Bukti :

Sesuai dengan ketidaksamaan Chebyscev yang menyatakan bahwa untuk setiap $\varepsilon > 0$ berlaku $P(|t_n-\vartheta| \leq \varepsilon) \geq 1 - \dfrac{E[t_n-\vartheta]^2}{\varepsilon^2}$ maka menurut definisi 2.4.3. akan diperoleh :

$$\lim_{n\to\infty} P(|t_n-\vartheta| \leq \varepsilon) \geq \lim_{n\to\infty} \left(1 - \frac{E[t_n-\vartheta]^2}{\varepsilon^2}\right)$$

$$= 1 - \lim_{n\to\infty} \left(\frac{E[t_n-\vartheta]^2}{\varepsilon^2}\right)$$

$$= 1$$

Jadi terbukti bahwa t_n konsisten terhadap ϑ.

2.5. ESTIMASI DENSITAS KERNEL

Salah satu teknik penghalusan dalam regresi nonparametrik adalah teknik kernel dengan barisan bobotnya adalah $\{W_{hi}(x)\}_{i=1}^{n}$ yang ditentukan oleh kernel K dan bandwidth h.

Pada umumnya kernel K didefinisikan sebagai :

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right)$$

dengan h adalah penghalus kernel yang juga dikenal sebagai bandwidth. Rata-rata keseluruhan fungsi kernel dalam suatu observasi ditunjukkan dengan estimator densitas kernel :

$$\hat{f}_h(x) = \frac{1}{n} \Sigma K_h(x-X_i)$$

$$= \frac{1}{nh} \sum K\left(\frac{x-X_i}{h} \right)$$

Sifat-sifat fungsi Kernel :

1. Fungsi kernel adalah simetri di sekitar 0 dan nilai integralnya = 1
2. Karena kernel merupakan fungsi densitas maka estimasi kernel merupakan fungsi densitas pula.

   $$\int K(x)\, dx = 1 \Rightarrow \int \hat{f}_h(x) dx = 1$$

3. Jika K kontinu dan differensiabel (n kali) maka $\hat{f}_h(x)$ juga kontinu differensiabel (n kali)
4. Semua fungsi kernel selalu positif. Ini digunakan untuk memastikan bahwa $\hat{f}_h(\cdot)$ adalah fungsi densitas.

Dikenal 5 macam fungsi kernel sebagai berikut :

1.	Uniform	$\frac{1}{2} I(\lvert U \rvert \le 1)$
2.	Triangle	$(1-\lvert U \rvert)\ I(\lvert U \rvert \le 1)$
3.	Epanechnicov	$\frac{3}{4}(1-U^2)\ I(\lvert U \rvert \le 1)$
4.	Quartic	$\frac{15}{16}(1-U^2)^2\ I(\lvert U \rvert \le 1)$
5.	Triweight	$\frac{35}{32}(1-U^2)^3\ I(\lvert U \rvert \le 1)$

2.6. ALGORITMA DASAR WARPING

Algoritma dasar Warping terdiri dari 3 tahapan,yaitu :

1. Pembentukan bin (Binning the data)
   - Bentuk suatu index subbin $t(x)=j \Leftarrow x \in [(j-1/2)\delta,(j+1/2)\partial)$, $j \in \mathbb{Z}$
   - Hitung berapa n_z yaitu banyaknya observasi dalam

bin z.

- Hitung index j_z untuk bin yang tidak kosong.
- Hitung $Y_{.z}$ sebagai jumlahan Y_i dalam bin z.

2. Pembentukan fungsi bobot (Creating Weight)

   Membentuk fungsi bobot K(l/M) dengan l=1, 2, ..., M-1 dan dengan menganggap bahwa observasi x_i yang terletak dalam bin yang sama mempunyai bobot $W_{Mi}(x)$ yang sama pula.

3. Pembobotan bin (Weighting the bins)

   Menghitung $\hat{m}_M(x)$ untuk bin yang tidak kosong dari perhitungan numerator dan denominator yang telah dihitung pada lanhkag ke-2 yaitu :

$$\hat{m}_M(x) = \frac{\Sigma\ K(l/M)Y_{.t(x)+l}}{\Sigma\ K(l/M)n_{.t(x)+l}}$$

Algoritma dasar Warping ini akan diterapkan dalam kasus regresi nonparametrik dan dengan teknik kernel untuk menentukan taksiran kurva regresi nonparametrik.