BAB II

TEORI PENUNJANG

2.1. Definisi-definisi

2.1.1. Sampel Kontinyu

Definisi : Suatu ruang sampel S yang mempunyai anggota semua titik dalam suatu interval atau dalam suatu penggabungan interval-interval pada suatu garis nyata dinamakan sampel kontinyu

Contoh : S = { x:$0 \leq x \leq 10$ }

S = { x:$0 \leq x \leq 1$ atau $2 \leq x \leq 3$ }

dinamakan sampel kontinyu. Ruang sampel kontinyu mempunyai jumlah anggota tak hingga.

2.1.2. Kontinuitas

Definisi : Suatu fungsi f(x) dikatakan kontinyu di x = x_0 jika :

i. $f(x_0)$ terdefinisi

ii. $\lim\limits_{x \to x_0} f(x)$ ada

iii. $\lim\limits_{x \to x_0} f(x) = f(x_0)$.

Contoh : Ambil $f(x) = \frac{3}{4}(1 - x^2)$ kontinyu di $x = \frac{1}{2}$. Selanjutnya akan diperlihatkan f(x) memenuhi ketiga syarat tersebut.

i. $f(1/2)=\frac{3}{4}(1-(1/2)^2)=\frac{3}{4}\times\frac{3}{4}=\frac{9}{16}$ (terdefinisi)

ii. $\lim\limits_{x\to\frac{1}{2}} \frac{3}{4}(1-x^2)=\frac{9}{16}$ (ada)

iii. Dari pers.(i) dan pers.(ii) terlihat dengan jelas bahwa

$$\lim_{x\to\frac{1}{2}} \frac{3}{4}(1-x^2)=\frac{9}{16}=f(1/2).$$

maka jelaslah bahwa fungsi $f(x)=\frac{3}{4}(1-x^2)$ kontinyu di $x=\frac{1}{2}$.

2.1.3. Peubah Acak

Peubah acak adalah suatu peubah yang nilainya dari bilangan yang ditentukan melalui suatu hasil percobaan.

2. 1. 4. Fungsi Densitas

Fungsi distribusi dari peubah acak kontinyu X adalah fungsi $F(x)=P\{X\le x\}$ didefinisikan untuk setiap x dari -∞ sampai ∞.

Definisi : Turunan dari F(x) yaitu $f(x)=\frac{dF(x)}{dx}$ disebut fungsi densitas dari peubah acak kontinyu X.

2.1.5. Fungsi Densitas Gabungan

Bila X dan Y dua peubah acak kontinyu, distribusi peluang terjadinya serentak X dan Y dinyatakan fungsi $f(x,y)$ untuk setiap pasangan nilai (x,y) dalam rentang peubah acak kontinyu X dan Y.

Definisi : Fungsi $f(x,y)$ disebut fungsi densitas gabungan peubah acak kontinyu X dan Y bila

1. $f(x,y) \geq 0$ untuk semua (x,y)

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

3. $P[(X,Y) \in A] = \iint_A f(x,y)dxdy$ untuk setiap daerah A dibidang xy.

Contoh : Suatu perusahaan coklat mengirim berkotak-kotak coklat dengan campuran krem, tofe serta kacang berlapis coklat cerah dan pekat. Bila kotak dipilih secara serentak, serta X dan Y masing-masing menyatakan proporsi yang krem berlapis coklat cerah dan pekat dan misalkan fungsi densitas gabungannya adalah

$f(x,y) = \frac{2}{5}(2x+3y)$, untuk $0 \leq x \leq 1$,$0 \leq y \leq 1$

$= 0$, untuk x,y yang lain.

Akan ditunjukkan bahwa ketiga syarat diatas terpenuhi.

i. Untuk $0 \leq x \leq 1$ dan $0 \leq y \leq 1$ terlihat dengan jelas bahwa $f(x,y) = \frac{2}{5}(2x+3y) \geq 0$ dan $f(x,y) = 0$ untuk x,y yang lain, sehingga syarat 1 terpenuhi

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

$$= \int_0^1 \frac{2x^2}{5} + \frac{6xy}{5}\Bigg|_{x=0}^{x=1} dy$$

$$= \int_0^1 \left(\frac{2}{5} + \frac{6y}{5}\right) dy = \frac{2y}{5} + \frac{3y^2}{5}\Bigg|_{y=0}^{y=1}$$

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

terlihat syarat 2 terpenuhi.

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

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

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

(seperti bukti no.ii)

terlihat bahwa $f(x,y)$ memenuhi ketiga syarat tersebut sehingga dapat kita katakan bahwa $f(x,y)$ merupakan fungsi densitas gabungan.

2.1.6. Fungsi Densitas Marginal

Bila diketahui fungsi densitas gabungan $f(x,y)$ dari peubah acak kontinyu X dan Y maka distribusi peluang $g(x)$ dari X sendiri dapat diperoleh dengan pengintegralan $f(x,y)$ terhadap Y, begitu juga distribusi peluang $h(y)$ dapat diperoleh dengan pengintegralan $f(x,y)$ terhadap X. Selanjutnya kita sebut $g(x)$ dan $h(y)$ masing-masing sebuah fungsi densitas marginal dari X dan Y.

Definisi : Fungsi Densitas Marginal dari X sendiri dan Y sendiri didefinisikan sebagai

$$g(x)=\int_{-\infty}^{\infty} f(x,y)dy \text{ dan } h(y)=\int_{-\infty}^{\infty} f(x,y)dx$$

Contoh : Dari contoh 2.1.2. cari $g(x)$ dan $h(y)$nya.

Penyelesaian : Menurut definisi $g(x)=\int_{-\infty}^{\infty} f(x,y)dy$

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

$$=\frac{4xy}{5}+\frac{6y^2}{10}\Big|_{y=0}^{y=1}=\frac{4x+3}{5}$$

untuk $0 \le x \le 1$ dan $g(x)=0$ untuk yang lain, begitu pula

untuk $h(x)=\int_0^1 \frac{2}{5}(2x+3y)dx=\frac{2x^2}{5}+\frac{6xy}{5}\Big|_{x=0}^{x=1}=\frac{2}{5}(1+3y)$

2.1.7. Simetris

Definisi : Variabel random X dikatakan simetris terhadap suatu harga α jika $P(x \ge \alpha + x) = P(x \le \alpha - x)$ untuk setiap x,

Pada dasarnya, jika X adalah variabel random kontinyu, maka X adalah simetris dengan pusat α jika dan hanya jika $F(\alpha - x) = F(\alpha + x)$ untuk $\forall x \in R$. Jika α=0, secara singkat kita katakan bahwa X adalah simetris (atau F adalah simetris).

Contoh : Fungsi $f(x)=\frac{3}{4}(1-x^2)$ adalah simetris terhadap 0.

Akan ditunjukkan bahwa $f(x)=f(-x)$

$$f(1/2)=f(-1/2)$$

$$\frac{3}{4}(1-(1/2)^2)=\frac{3}{4}(1-(-1/2)^2)$$

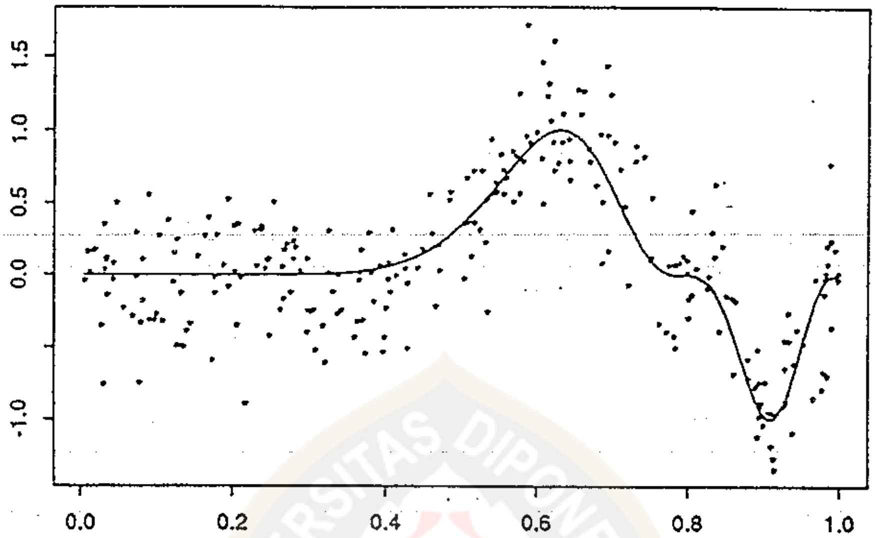
$$\frac{3}{4}\times\frac{3}{4}=\frac{3}{4}\times\frac{3}{4}$$

$$\frac{9}{16}=\frac{9}{16}$$

terlihat bahwa fungsi tersebut simetris terhadap 0.

2.2. Estimasi Densitas Kernel

Penghalusan dari data $\{(X_i,Y_i)\}_{i=1}^{n}$ meliputi perkiraan dari kurva tanggapan m dalam hubungan regresi seperti terlihat dalam persamaan(1.1.1) yaitu $Y_i=m(X_i)+\varepsilon_i$ dengan i = 1, 2, 3, . . . , n. Dari data yang diperoleh pada umumnya harga-harga dari (X_i,Y_i) tidak menunjukkan suatu penyimpangan yang sangat berarti walaupun kadang hal ini terjadi. Dengan kata lain data tersebut terlihat lebih terpusat pada suatu kelompk-kelompok tertentu (data mengelompok) seperti terlihat pada gambar (2.1). Perkiraan yang sesuai untuk kurva regresi m(x) oleh karenanya akan merupakan suatu harga yang mendekati pusat dari kelompok-kelompok itu sendiri.



Gambar 2.1 : Plot penyebaran data tabel 1 serta kurva regresi sesungguhnya.

Pilihan yang sering dipakai adalah rata-rata dari variabel tanggapan Y mendekati suatu harga X. "Rata-rata lokal" tersebut dapat dipakai sebagai ide dasar dari penghalusan. Untuk lebih jelasnya prosedur ini dapat didefinisikan sebagai

$$\hat{m}_h(x) = n^{-1}\sum_{i=1}^{n} W_{hi}(x) Y_i \qquad \ldots (2.2.1)$$

sedangkan barisan bobot untuk penghalus kernel tersebut didefinisikan sebagai $W_{hi}(x) = \dfrac{K_h(x - X_i)}{\hat{f}_h(x)}$ . . . (2.2.2)

dimana $K_h(x)$ adalah fungsi kernel yang didefinisikan sebagai $K_h(x) = h^{-1}K(x/h)$ dengan sifat simetris terhadap 0, merupakan fungsi kontinyu dan terintegralkan ke 1 dengan faktor skala h (yang lebih dikenal dengan 'Bandwidth'). Dan

$$\hat{f}_h(x) = n^{-1}\sum_{i=1}^{n} K_h(x - X_i) = n^{-1}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \qquad \ldots (2.2.3)$$

sebagai estimator densitas kernel. Terdapat 7(tujuh) bentuk dari fungsi kernel yang telah kita kenal yaitu

Kernel	*K(U)*
Uniform	$\frac{1}{2}\ I(\lvert U\rvert \le 1)$
Triangle	$(1-\lvert U\rvert)\ I(\lvert U\rvert \le 1)$
Epanechnikov	$\frac{3}{4}(1-U^2)\ I(\lvert U\rvert \le 1)$
Quartik	$\frac{15}{16}(1-U^2)^2\ I(\lvert U\rvert \le 1)$
Triweight	$\frac{35}{32}(1-U^2)^3\ I(\lvert U\rvert \le 1)$
Gaussian	$\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}U^2)$
Cosinus	$\frac{\pi}{4}\cos(\frac{\pi}{2}U)\ I(\lvert U\rvert \le 1)$

Selanjutnya fungsi bobot $W_{hi}(x)$ pada persamaan(2.2.2) dikenal dengan nama fungsi bobot Nadaraya-Watson, dan oleh karenanya

$$\hat{m}_h(x) = n^{-1}\sum_{i=1}^{n} W_{hi}(x)Y_i$$

$$= \frac{n^{-1}\sum_{i=1}^{n} K_h(x - X_i)Y_i}{n^{-1}\sum_{i=1}^{n} K_h(x - X_i)} = \frac{\hat{r}_h(x)}{\hat{f}_h(x)} \qquad \ldots (2.2.4)$$

dikenal dengan nama Estimator Nadaraya-Watson.

Selanjutnya untuk mengetahui sifat dari $\hat{m}_h(x)$ terhadap $m(x)$ maka kita pertimbangkan nilai ekspektasi dari elemen $\hat{m}_h(x)$. Bila didefinisikan $r(x) = \int yf(x,y)dy = m(x)f(x)$ maka

$$E[\hat{r}_h(x)] = E[n^{-1}\sum_{i=1}^{n} K_h(x - X_i)Y_i]$$

$$= E[K_h(x - X)Y]$$

$$= \iint yK_h(x-u)f(y|u)f(u)dydu$$

$$= \int K_h(x-u)f(u)(\int yf(y|u)dy)du$$

$$= \int K_h(x-u)f(u)(E[Y|X=u])du$$

$$= \int K_h(x-u)f(u)m(u)du$$

$$= \int K_h(x-u)r(u)du$$

dimisalkan $\frac{x-u}{h} = (-s)$, maka $x + sh = u$ dan $du = hds$ maka persamaan diatas menjadi

$$= h^{-1}\int K(\tfrac{x-u}{h})r(u)du$$

$$= h^{-1}\int K(s)r(x+sh)(hds)$$

$$= \int K(s)[r(x) + shr'(x) + \frac{s^2h^2}{2!}r''(x) + O(h^2)]ds$$

$$= r(x) + \frac{h^2}{2!}r''(x)\mu_2(K) + O(h^2) \qquad ,h \to 0 \qquad \ldots (2.2.5)$$

Term linier $\int K(s)shr'(x)ds$ = 0 karena diketahui bahwa K adalah simetris terhadap 0. Dari sini $\hat{r}_h(x)$ adalah tidak bias asimtotik untuk $h \to 0$. Selanjutnya untuk menghitung variansi dari $\hat{r}_h(x)$ dimisalkan $s^2(x) = E[Y^2|X = x]$ maka

$$
\begin{aligned}
Var[\hat{r}_h(x)] &= Var[n^{-1}\sum_{i=1}^{n} K_h(x - X_i)Y_i] \\
&= n^{-1}h^{-2}\{E[K^2(\tfrac{x-X}{h})Y^2] - E^2[K(\tfrac{x-X}{h})Y]\} \\
&= n^{-1}h^{-2}\{\iint K^2(\tfrac{x-X}{h})y^2 f(y|u)f(u)dydu \\
&\qquad -[\iint K(\tfrac{x-X}{h})yf(y|u)f(u)dydu]^2\} \\
&= n^{-1}h^{-1}\{\iint K^2(s)y^2 f(y|x+sh)f(x+sh)dyds \\
&\qquad -[\iint K(s)yf(y|x+sh)f(x+sh)dyds]^2\} \\
&= n^{-1}h^{-1}\{\int K^2(s)E[Y^2|X = x+sh]f(x+sh)ds \\
&\qquad -h[\int K(s)E[Y|X = x+sh]f(x+sh)ds]^2\}
\end{aligned}
$$

catatan bahwa $m(x) = E[Y|X = x]$, $r(x) = m(x)f(x)$ dan $s^2(x) = E[Y^2|X = x]$, sekarang dipertimbangkan untuk kasus $h \to 0$ dan $nh \to \infty$:

$$
\begin{aligned}
Var[\hat{r}_h(x)] &= n^{-1}h^{-1}\{\int K^2(s)s^2(x+sh)f(x+sh)ds \\
&\qquad -h[\int K(s)m(x+sh)f(x+sh)ds]^2\}
\end{aligned}
$$

$$= n^{-1}h^{-1}\{\int K^2(s)s^2(x+sh)f(x+sh)ds$$

$$-h[\int K(s)r(x+sh)ds]^2\}$$

$$= n^{-1}h^{-1}\int K^2(s)s^2(x+sh)f(x+sh)ds$$

$$-n^{-1}[\int K(s)r(x+sh)ds]^2$$

$$= n^{-1}h^{-1}[s^2(x)f(x)\int K^2(s)ds + O(h^2)]$$

$$-n^{-1}[r(x)\int K(s)ds + O(h^2)]^2$$

$$= n^{-1}h^{-1}s^2(x)f(x)\int K^2(s)ds + O(n^{-1}h)$$

$$-n^{-1}r^2(x) + O(n^{-1}h^2)$$

$$= n^{-1}h^{-1}s^2(x)f(x)\int K^2(s)ds + O((nh)^{-1})$$

$$= n^{-1}h^{-1}s^2(x)f(x)\|K\|_2^2 + O((nh)^{-1})\ ,\ \ nh \to \infty \qquad ...(2.2.6)$$

Bila kita jumlahkan bias^2 persamaan(2.2.5) dan persamaan(2.2.6) maka kita peroleh formula Rata-rata Kesalahan Kuadrat $MSE[\hat{r}_h(x)]$.

$$MSE[\hat{r}_h(x)] = Var(\hat{r}_h(x)) + [Bias(\hat{r}_h(x))]^2$$

$$= n^{-1}h^{-1}s^2(x)f(x)\|K\|_2^2 + \frac{h^4}{4}(r''(x)\mu_2(K))^2$$

$$+O(h^4) + O((nh)^{-1}) \qquad (h \to 0, nh \to \infty) \qquad ...(2.2.7)$$

dari sini bila kita ambil $h \to 0$ dan $nh \to \infty$, kita peroleh $MSE[\hat{r}_h(x)] \to 0$, sebagai akibatnya adalah $\hat{r}_h(x) \to r(x)$. Begitu juga untuk $\hat{f}_h(x)$nya, dari sini kita peroleh

$$\hat{m}_h(x) = \frac{\hat{r}_h}{\hat{f}_h} \rightarrow \frac{r(x)}{f(x)} = m(x) \qquad (h \rightarrow 0, nh \rightarrow \infty)$$

Dari keterangan diatas kita peroleh suatu konsistensi (ketepatan terhadap kurva m(x)) dari estimasi Nadaraya-Watson $\hat{m}_h(x)$ untuk $h \rightarrow 0$ dan $nh \rightarrow \infty$.