BAB II

TEORI DASAR

2.1. Fungsi densitas

Dalam menarik kesimpulan dari suatu data observasi, teori probabilitas memegang peranan yang sangat penting. Dengan diketahui nilai probabilitas, kita dapat memperoleh informasi dari data tersebut. Dalam hal ini fungsi densitas berguna dalam perhitungan nilai probabilitas. Selanjutnya akan dijelaskan mengenai definisi fungsi densitas.

Definisi 2.1.1

Suatu fungsi densitas f(x) merupakan fungsi densitas atau fungsi probabilitas dari suatu variabel random kontinu x jika :

i. $f(x) \geq 0$ untuk semua x

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

iii. $\int_{a}^{b} f(x)\, dx = P(a \leq x \leq b)$

dimana a dan b merupakan dua buah nilai yang berlainan dari x dan a<b.

Definisi 2.1.2

Suatu fungsi densitas dikatakan simetri di sekitar μ jika untuk setiapx barlaku $f(\mu+x) = f(\mu-x)$.

2.2. Estimasi fungsi densitas

Di muka telah dijelaskan bahwa fungsi densitas sangat berguna untuk mendapatkan informasi dari suatu data. Akan tetapi, fungsi densitas ini biasanya tidak bisa diketahui secara langsung dari data yang diperoleh. Oleh karena itu dari data yang ada perlu dilakukan suatu pendekatan terhadap fungsi densitas yang belum diketahui. Terdapat dua macam pendekatan yaitu :

1. Pendekatan parametrik

   Pendekatan ini dilakukan jika fungsi densitas f mempunyai parameter-parameter yang belum diketahui. Oleh karena itu dilakukan estimasi terhadap parameter-parameter tersebut. Dengan kata lain pendekatan ini identik dengan estimasi parameter.

2. Pendekatan non parametrik

   Pendekatan ini tidak membatasi kemungkinan bentuk dari fungsi densitas dengan asumsi f termasuk dalam keluarga fungsi densitas yang telah diketahui. Hal pokok dalam pendekatan ini adalah mengestimasi kurva kerapatannya. Pendekatan ini dilakukan ketika tidak diketahui informasi yang tepat tentang bentuk dan kelas fungsi densitas yang sebenarnya

2.3. Estimator Unbias

Jika suatu parameter diestimasi kemudian diperoleh suatu hasil yang sedemikian rupa sehingga estimator dan yang diestimasi adalah sama maka statistik itu dinamakan *estimator yang unbias*. Secara tepat, definisi unbias suatu estimator berkaitan dengan nilai harapan dari suatu variabel random X.

Definisi 2.3.1

Misalkan $\hat{\theta}$ adalah estimator titik untuk parameter θ, maka $\hat{\theta}$ sebuah estimator unbias dari θ jika

$$E[\hat{\theta}] = \theta$$

Contoh :

Misalkan X adalah variabel random dengan rata-rata μ dan varian σ^2. Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel random dari X. Akan ditunjukkan bahwa $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ adalah estimator yang unbias untuk μ.

$$E(\overline{X}) = E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right]$$

$$= \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} E(X_i)$$

Karena $E(X_i) = \mu$ untuk semua $i=1,2,\ldots,n$ maka

$$E(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

Jadi rata-rata sampel $\overline{X}$ adalah estimator yang unbias terhadap rata-rata populasi μ.

Rata-rata error kuadrat suatu estimator $\hat{\theta}$ didefinisikan sebagai :

$$\begin{aligned} MSE(\hat{\theta}) &= E(\hat{\theta} - \theta)^2 \\ &= E[\hat{\theta} - E(\hat{\theta})]^2 + [E(\hat{\theta}) - \theta]^2 \\ &= Var(\hat{\theta}) + [Bias(\hat{\theta})]^2 \end{aligned}$$

Rata-rata error kuadrat merupakan ukuran keakuratan suatu estimator. Jika $\hat{\theta}$ adalah sebuah estimator yang unbias pada θ maka $MSE(\hat{\theta}) = Var(\hat{\theta})$.

Dalam sekelompok estimator yang unbias, esimator yamg baik mempunyai varian yang kecil. Akan tetapi, kadang-kadang estimator yang bias lebih disukai daripada estimator yang unbias. Hal ini disebabkan karena estimator yang bias tersebut mempunyai rata-rata error kuadrat yang kecil.

2.4. Estimator Maximum Likelihood

Salah satu metode yang yang digunakan untuk memperoleh suatu estimator tunggal adalah metode *maximum likelihood.* Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel random dari suatu populasi dengan densitas $f(X,\theta)$ dimana θ adalah suatu parameter yang tidak diketahui.

Fungsi likelihood didefinisikan sebagai :

$$L(\theta) = \prod_{i=1}^{n} f(X_i,\theta)$$

Estimator maximum likelihood adalah nilai θ yang memaksimumkan fungsi likelihood $L(\theta)$ atau memaksimalkan probabilitas kejadian hasil sampel. Jika $\hat{\theta}$ memaksimalkan $L(\theta)$ maka $\hat{\theta}$ juga memaksimalkan $\log L(\theta)$, yaitu logaritma likelihood, sehingga dengan mendeferensialkan logaritma ini akan diperoleh estimator maximum likelihood. Secara matematis persamaan likelihood :

$$\frac{\partial \text{LogL}(\hat{\theta})}{\partial \hat{\theta}} = 0$$

akan menghasilkan $\hat{\theta}_i$ sebagai estimator maximum likelihood.

Contoh :

$X_1, X_2, \ldots, X_n$ variabel random independen dari populasi berdistribusi poisson dengan parameter θ. Tentukan estimator maximum likelihood untuk θ.

Penyelesaian :

$$X_i \sim P(\theta)$$

$$f(X_i) = \frac{\theta^{x_i}}{x\ !}$$

Fungsi likelihood :

$$L = \prod_{i=1}^{n} f(X_i)$$

$$= \frac{e^{-\theta}.\theta^{x_1}}{x_1!} \cdot \frac{e^{-\theta}.\theta^{x_2}}{x_2!} \ldots \frac{e^{-\theta}.\theta^{x_n}}{x_n!}$$

$$= \frac{e^{-n\theta}.\theta^{\Sigma x_1}}{\prod x_i!}$$

$$\text{Log } L = -n\theta + \Sigma X_i \log \theta - \Sigma \log X_i$$

$$\frac{\partial \log L}{\partial \theta} = -n + \frac{\Sigma X_i}{\theta}$$

Persamaan likelihood

$$\frac{\partial \log L}{\partial \hat{\theta}} = 0 \quad \Leftarrow \quad -n + \frac{\Sigma X_i}{\hat{\theta}} = 0$$

$$\frac{\Sigma X_i}{\hat{\theta}} = n$$

$$\hat{\theta} = \frac{\Sigma X_i}{n}$$

Jadi estimator likelihood untuk θ adalah $\hat{\theta} = \frac{1}{n}\ \Sigma X_i$

2.5. Kriteria Kesalahan Estimasi Densitas

Jika dalam pendekatan parametrik dengan estimator bias menggunakan kriteria rata-rata error kuadrat MSE seperti yang telah dijelaskan di atas, maka dalam pendekatan non parametrik untuk estimasi densitas, perhitungan MSE dapat diperoleh dengan menganggap bahwa f merupakan parameter yang belum diketahui, atau dengan kata lain $\theta = f(x)$ yaitu fungsi densitas itu sendiri. Dengan demikian rata-rata error kuadrat untuk estimasi ini adalah :

$$\begin{aligned} MSE\{\hat{f}(x)\} &= E\{\hat{f}(x) - f(x)\}^2 \\ &= Var\ \{\hat{f}(x)\} + Bias^2\{\hat{f}(x)\} \end{aligned}$$

dimana $Bias[\hat{f}(x)] = E[\hat{f}(x) - f(x)]$

Pada prakteknya, formula MSE di atas sulit dihitung karena fungsi densitas belum diketahui. Terdapat kriteria lain yang lebih mudah yaitu Integrated Squared Error (ISE) dengan rumus :

$$ISE = \int [\hat{f}(x) - f(x)]^2\ dx$$

Mean Integrated Squared Error (MISE), yang merupakan rata-rata dari ISE adalah :

$$MISE = E[ISE]$$
$$= E[\int \{\hat{f}(x) - f(x)\}^2 \, dx]$$

2.6. Estimasi Densitas Kernel

Estimasi densitas dengan kernel adalah suatu metode pendekatan terhadap fungsi densitas yang belum diketahui dengan menggunakan fungsi densitas kernel. Dalam estimasi ini ada suatu parameter yang sangat berpengaruh terhadap keoptimalan hasil yaitu bandwidth h. Di samping itu jenis kernel juga mempengaruhi bentuk estimator densitas kernel.

Ada beberapa kemungkinan yang merupakan pengaruh pemilihan bandwidth terhadap grafik penghalusan fungsi densitas kernel :

- $h \longrightarrow 0$, needleplot, grafik yang acak-acakan
- h kecil, fungsi agak halus
- h besar, fungsi halus
- $h \longrightarrow \infty$ fungsinya datar (terlalu halus)

Secara umum kernel K didefinisikan :

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

Jika didapat sekumpulan observasi $\{X_i\}_{i=1}^{n}$ maka estimator densitas kernel adalah :

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

Keterangan :

$\hat{f}_h(x)$ = estimator densitas kernel

K = Jenis kernel

n = Jumlah observasi (data)

X_i = data ke-i, i=1,2,...n

x = nilai yang didapat dengan pendekatan

Terdapat beberapa jenis fungsi kernel yang dapat dijadikan alteratif pilihan dalam estimasi fungsi densitas yaitu :

Kernel	K(u)
Uniform	$\frac{1}{2} I(\|u\| \le 1)$
Triangle	$(1 - \|u\|) I(\|u\| \le 1)$
Epanechnicov	$\frac{3}{4}(1 - u)^2 I(\|u\| \le 1)$
Quartic	$\frac{15}{16}(1-u^2)^2 I(\|u\| \le 1)$
Triweight	$\frac{35}{32}(1-u^2)^3 I(\|u\| \le 1)$
Gaussian	$\frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}u^2)$
Cosinus	$\frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right) I(\|u\| \le 1)$

Di bawah ini beberapa hal yang perlu diketahui mengenai fungsi densitas kernel :

1. Fungsi kernel simetris di sekitar O dan terintegrasi ke 1.
2. Jika kernel adalah fungsi densitas maka estimasi kernel juga merupakan fungsi densitas.

3. Sifat penghalus kernel diikuti oleh $\hat{f}_h(x)$, jika diferensiabel (n kali) maka $\hat{f}_h(x)$ juga diferensiabel (n kali)
4. Estimasi kernel tidak tergantung dari pemilihan harga awal. Jika bandwidth h dan jenis kernel telah ditentukan maka diperoleh fungsi densitas kernel yang tunggal untuk sekumpulan data yang diberikan.
5. Kernel di atas semua positif untuk memastikan bahwa $\hat{f}_h(\cdot)$ adalah fungsi densitas.