# BAB II

# MATERI PENUNJANG

Analisis distribusi probabilitas gelombang laut memerlukan beberapa materi yang dapat menunjang dalam pengambilan keputusan untuk mencari distribusi yang paling representatif terhadap data observasi. Beberapa materi penunjang yang akan dibahas pada bab ini adalah fungsi distribusi probabilitas, model regresi linier, metode estimasi maksimum likelihood dan uji kebaikan suai distribusi.

## 2.1. Fungsi Distribusi Probabilitas

Dalam ilmu pengetahuan alam dan rekayasa, terdapat fenomena acak yang berhubungan dengan hasil numerik. Dalam segala hal, suatu hasil atau peristiwa dapat dinyatakan dengan nilai disebut dengan variabel random. Variabel random dapat dipandang sebagai suatu fungsi yang memetakan peristiwa-peristiwa dalam ruang sampel ke suatu bilangan riel. Terdapat 2 variabel random, yaitu variabel random diskrit dan kontinu. Variabel random diskrit adalah jika suatu ruang sampel dari variabel random X memuat titik-titik yang banyaknya berhingga Sedangkan variabel random kontinu adalah jika ruang sampel dari variabel random X berupa interval atau kumpulan dari interval-interval.

Nilai dari suatu variabel random menyatakan suatu peristiwa, maka nilai variabel random dapat berupa besaran numerik hanya dalam probabilitas yang bersangkutan atau ukuran probabilitas. Aturan untuk menyatakan ukuran probabilitas yang berkaitan dengan semua harga suatu variabel random adalah distribusi probabilitas.

Distribusi probabilitas untuk suatu variabel random kontinu X dapat dinyatakan dalam fungsi densitas probabilitas, sehingga bila $f_X(x)$ adalah fungsi densitas probabilitas dari X, maka probabilitas dari X adalah:

$$P(a \leq x \leq b) = \int_a^b f_X(x)\,dx \tag{1}$$

Fungsi $f_X(x)$ sebagai fungsi densitas probabilitas variabel random kontinu X yang didefinisikan di atas himpunan semua bilangan riil R memenuhi sifat-sifat berikut ini, yaitu :

a. $f_X(x) \geq 0$, untuk semua $X \in R$

b. $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

c. $P(a \leq x \leq b) = \int_a^b f_X(x)\,dx$

Fungsi distribusi probabilitas variabel random kontinu X dapat dinyatakan dengan fungsi distribusi kumulatif yang didefinisikan di atas himpunan semua bilangan riil R, dapat dinyatakan dalam :

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt \tag{2}$$

Sumbu ordinat dari lengkung probabilitas kumulatif menyatakan probabilitas harga suatu variabel random kontinu disamai atau dilampaui, yang dapat ditulis dalam bentuk $1 - P(X \leq x)$ atau dinyatakan dengan $P(X \geq x)$.

Jika $F_X(x)$ memiliki turunan pertama, maka dari persamaan (2) dapat diperoleh :

$$f_X(x) = \frac{d F_X(x)}{dx} \tag{3}$$

Fungsi distribusi kumulatif mempunyai sifat-sifat sebagai berikut :

1. $\lim_{x \to \infty} F(x) = 1$
2. $\lim_{x \to -\infty} F(x) = 0$
3. Tidak turun, yaitu jika $b \geq a$, maka $F(b) \geq F(a)$
4. Kontinu dari kanan, yaitu untuk seluruh x dan $\delta > 0$ , $\lim_{\delta \to 0} [F(x+\delta) - F(x)] = 0$

**2.2. Model Regresi Linier**

Regresi linier sederhana adalah suatu model untuk menentukan hubungan antara sebuah variabel bebas tunggal X dan sebuah variabel tidak bebas Y. Pola hubungan antar peubah Y dan X yang bersifat linier dan sederhana dapat dimodelkan dengan persamaan :

$$X^{(m)} = A\,Y^{(m)} + B + \varepsilon_i \qquad , m = 1,2,\ldots N \tag{4}$$

dimana : $X^{(m)}$ adalah variabel tidak bebas pada pengamatan ke-m , $Y^{(m)}$ adalah variabel bebas pada pengamatan ke-m, A dan B adalah parameter yang belum diketahui nilainya atau koefisien regresi, $\varepsilon_i$ adalah residu pada pengamatan ke-i dengan asumsi bahwa $\varepsilon_i$ berdistribusi bebas normal dengan rata-rata 0 dan varian $\sigma^2$ .

Metode untuk mengestimasi parameter A dan B pada persamaan regresi linier menggunakan metode kuadrat terkecil. Metode ini dilakukan dengan meminimumkan jumlah kuadrat residu $\sum_{i=1}^{N}\varepsilon_i^2$ terhadap A dan B. Untuk lebih mempermudah maka selanjutnya fungsi jumlah kuadrat residu dinyatakan dalam :

$$S = \sum_{i=1}^{N}\varepsilon_i^2 \tag{5}$$

$$= \sum_{i=1}^{N}(X^{(m)} - B - AY^{(m)})^2 \tag{6}$$

Jika diturunkan terhadap A dan B di dapat :

$$\frac{\partial S}{\partial B} = \sum_{i=1}^{N} 2(X^{(m)} - B - AY^{(m)})(-1) = 0 \tag{7}$$

$$\frac{\partial S}{\partial A} = \sum_{i=1}^{N} 2(X^{(m)} - B - AY^{(m)})(-Y^{(m)}) = 0 \tag{8}$$

Sehingga didapat nilai estimasi untuk A dan B yang dinyatakan dalam bentuk:

$$\hat{A} = \frac{N\sum_{i=1}^{N}X_i^{(m)}Y_i^{(m)} - \sum_{i=1}^{N}X_i^{(m)}\sum_{i=1}^{N}Y_i^{(m)}}{N\sum_{i=1}^{N}(Y_i^{(m)})^2 - \left(\sum_{i=1}^{N}Y_i^{(m)}\right)^2} \tag{9}$$

$$\hat{B} = \overline{X}^{(m)} - \hat{A}\overline{Y}^{(m)} \tag{10}$$

### 2.3. Metode Estimasi Maksimum Likelihood

Metode estimasi maksimum likelihood merupakan salah satu cara untuk mengestimasi parameter yang tidak diketahui. Prosedur estimasi maksimum likelihood menguji apakah estimasi maksimum yang tidak diketahui dari fungsi likelihood suatu sampel nilainya sudah memaksimumkan fungsi likelihood.

Misalkan $X_1, X_2, \ldots, X_N$ adalah variabel random dari populasi dengan distribusi fungsi $f(x, \theta)$, dengan $\theta$ adalah parameter yang tidak diketahui. Maka fungsi likelihood sampel tersebut adalah :

$$L(x_1, x_2, \ldots x_N; \theta) = f(x_1, \theta).\, f(x_2, \theta) \ldots f(x_N, \theta) \qquad (11)$$

Bila fungsi likelihood terdiferensialkan ke-$\theta$, maka estimasi maksimum likelihood dapat diperoleh melalui persamaan berikut :

$$(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k) \rightarrow \frac{\partial L(x_1, x_2, \ldots, x_N; \theta)}{\partial \theta_i} = 0 \qquad (12)$$

Dalam banyak kasus, penggunaan diferensiasi akan lebih mudah bekerja pada logaritma natural dari $L(x_1, x_2, \ldots, x_N; \theta)$, yaitu :

$$K(x_1, x_2, \ldots, x_N; \theta) = \ln L(x_1, x_2, \ldots, x_N; \theta)$$

Jika terdapat dua parameter yang tidak diketahui dalam fungsi likelihood, maka langkah-langkah untuk menentukan estimasi maksimum likelihood dari $\theta_1$ dan $\theta_2$ dapat dituliskan sebagai berikut :

1. Menentukan fungsi likelihood: $L(x_1, x_2, \ldots, x_N; \theta_i) = f(x_1; \theta_i)\, f(x_2; \theta_i) \ldots f(x_N; \theta_i)$
2. Membentuk logaritma natural likelihood $K(x_1, x_2, \ldots, x_N; \theta_i) = \mathrm{Ln}\, L(x_1, x_2, \ldots, x_N; \theta_i)$
3. Membentuk persamaan likelihood dan menyelesaikan $\dfrac{\partial K(x_1, x_2, \ldots, x_N, \theta_i)}{\partial \theta_i} = 0$
4. Didapat estimasi maksimum likelihood dari $\theta_1$ yaitu $\hat{\theta}_1$ dan $\theta_2$ yaitu $\hat{\theta}_2$
5. Membuktikan bahwa $\hat{\theta}_i$ benar-benar memaksimumkan $L(x_1, x_2, \ldots x_N; \theta_i)$

   dengan $\left.\dfrac{\partial^2 L}{\partial \hat{\theta}_i^2}\right|_{\theta_i} = \hat{\theta}_i < 0$ dengan i =1,2

6. Membuktikan bahwa $B^2$ -AC < 0 dan A + C < 0 relatif maksimum pada $(\hat{\theta}_1, \hat{\theta}_2)$

dengan $A = \frac{\partial^2 L}{\partial \theta_1^2}(\hat{\theta}_1, \hat{\theta}_2)$ , $B = \frac{\partial^2 L}{\partial \theta_1 \partial \theta_2}(\hat{\theta}_1, \hat{\theta}_2)$ dan $C = \frac{\partial^2 L}{\partial \theta_2^2}(\hat{\theta}_1, \hat{\theta}_2)$

## 2.4. Uji Kebaikan Suai Distribusi

Uji hipotesis adalah suatu prosedur membuat keputusan tentang kebenaran atau kesalahan hipotesis, yang bertujuan memperoleh suatu keputusan menerima hipotesis berdasarkan analisis sampel tertentu. Salah satu analisis sampel untuk menguji keabsahan dari distribusi teoritik yang telah diasumsikan, yang dapat dibenarkan atau disangkal secara statistik dengan uji kebaikan suai (*goodness of fit*) adalah uji Kolmogorov – Smirnov.

Jika terdapat $x_1, x_2, \ldots, x_N$ sebagai nilai-nilai dari data sampel yang telah disusun secara meningkat, N adalah besarnya sampel dan k merupakan banyak observasi yang sama atau kurang dari x, maka dari data sampel yang terurut dapat dibentuk suatu fungsi distribusi kumulatif observasi, $S_N(x)$ sebagai berikut :

$$S_N(x) = \begin{cases} 0 & , x < x_1 \\ \frac{k}{N} & , x_k \leq x < x_{k+1} \\ 1 & , x \geq x_N \end{cases} \tag{14}$$

Kemudian dari nilai $S_N(x)$ dibandingkan dengan $F^*(x)$ sebagai fungsi distribusi kumulatif teoritik untuk melihat apakah ada alasan yang kuat untuk menyatakan bahwa $F^*(x)$ adalah fungsi distribusi populasi yang sebenarnya.

Terdapat rumusan hipotesis 2 arah, yang dinyatakan dalam :

$H_0 : F(x) = F^*(x)\ ,\ -\infty < x < \infty$

$H_1 : F(x) \neq F^*(x)$, untuk suatu x

dengan F(x) merupakan fungsi distribusi yang tidak diketahui.

Pada uji Kolmogorov–Smirnov, $D_N$ merupakan selisih maksimum antara $F^*(x)$ dan $S_N(x)$ untuk seluruh rentang X yang merupakan pengukur perbedaan antara model teoritik dan data observasi. Maka statistik uji 2 arah Kolmogorov-Smirnov dapat dinyatakan dalam bentuk persamaan:

$$D_N = \underset{x}{\text{Supremum}}\ |\ F^*(x) - S_N(x)\ | \qquad (15)$$

Untuk suatu taraf nyata $\alpha$ tertentu, pengujian Kolmogorov-Smirnov membandingkan selisih maksimum pengamatan dalam persamaan (15) dengan nilai kritis $D_N^\alpha$, yang didefinisikan dengan :

$$P\ (D_N \leq D_N^\alpha) = 1 - \alpha \qquad (16)$$

Nilai-nilai kritis $D_N^\alpha$ untuk taraf nyata $\alpha$ yang berbeda-beda disajikan dalam tabel uji Kolmogorov-Smirnov pada lampiran 6 untuk berbagai nilai N.

Aturan pengambilan keputusan yaitu terima $H_0$ jika $D_N$ yang diamati kurang dari nilai kritis $D_N^\alpha$, maka distribusi yang diusulkan dapat diterima pada taraf nyata $\alpha$ yang ditentukan, jika tidak, distribusi yang diajukan akan ditolak.