BAB II
DASAR TEORI

Sebelum pembahasan tentang metode algoritma EM untuk mengestimasi nilai parameter pada data tidak lengkap disajikan, berikut akan diuraikan terlebih dahulu teori-teori yang mendasari metode tersebut.

2.1 Barisan dan Kekonvergenan

Algoritma EM berjalan secara iteratif sehingga menghasilkan suatu barisan, di mana iterasi berhenti jika kekonvergenan barisan tersebut sudah diperoleh. Untuk itu akan dibahas definisi dan teorema tentang barisan dan kekonvergenan yang ditulis oleh Robert **G. Bartle** dan **Donald R. Sherbert**, 1982.

Definisi 2.1.1

Barisan dari bilangan riil (atau barisan di dalam R) adalah suatu fungsi pada kumpulan bilangan asli N, dengan range yang termuat di dalam kumpulan bilangan riil R.

Dengan kata lain, barisan bilangan riil R ditentukan oleh pemetaan bilangan asli N. Bilangan riil yang terdapat di dalam barisan disebut sebagai elemen dari barisan atau nilai barisan. Secara umum barisan dinotasikan dengan

$$X = \{x_n\} \text{ dengan } x_n \in R \text{ dan } n \in N \qquad (2.1.1)$$

Definsisi 2.1.2

Barisan $\{x_n\}$ terbatas ke atas, $\exists\, u \in R \Rightarrow \{x_n\} \leq u$ dan

terbatas ke bawah, $\exists\, p \in R \Rightarrow \{x_n\} \geq p$

Definsisi 2.1.3

Misalkan $S \subseteq R$, maka

- Untuk S terbatas ke atas, suatu batas atas u dikatakan sebagai supremum (batas atas terkecil) dari S, jika $u \leq w$; $\forall w$ batas atas S.
- Untuk S terbatas ke bawah, suatu batas bawah p dikatakan sebagai infimum (batas bawah terbesar) dari S, jika $p \geq q$; $\forall q$ batas bawah S.

Definisi supremum, dapat dinyatakan dengan suatu bilangan $u \in R$ di mana $S \subseteq R$ dan memenuhi 2 kondisi berikut

1) $s \leq u, \forall s \in S$

2) $s \leq w \Rightarrow u \leq w, \forall s \in S$

Kondisi 1) di atas menyatakan bahwa, u merupakan batas atas dari S dan kondisi 2) menunjukkan bahwa u sebagai supremum, karena lebih kecil dari semua batas atas dari kumpulan S

Lemma 2.1.1

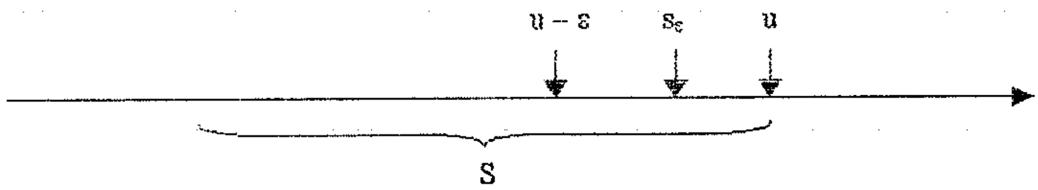
Suatu batas atas u dari kumpulan tak kosong $S \subseteq R$ merupakan supremum S, $\forall \varepsilon > 0 \Leftrightarrow \exists s_\varepsilon \in S$ dengan $u - \varepsilon < s_\varepsilon$

Bukti,

Andaikan u memenuhi kondisi yang ditetapkan, akan ditunjukkan bahwa $u = \sup S$. Misalkan s adalah sebarang batas atas dari S dan $w \neq u$. Jika diasumsikan bahwa $w < u$ dan mengambil $\varepsilon = u - w > 0$, hal ini mengimplikasikan bahwa $\exists s_\varepsilon \in S$, $\ni w = u - \varepsilon < s_\varepsilon$. Akan tetapi hal ini kontradiksi dengan pernyataan w sebagai batas

atas dari S. Dengan hipotesis w < u yang menyatakan kontardiksi, dapat disimpulkan bahwa u < w, sehingga dapat ditunjukkan bahwa u = Sup S.

Sebaliknya, andaikan u = sup S dan $\varepsilon > 0$. Maka $u - \varepsilon < u$, sehingga $u - \varepsilon$ bukan suatu batas atas dari S, karena beberapa elemen $s_\varepsilon \in S$ melebihi $u - \varepsilon$ mengakibatkan $u - \varepsilon < s_\varepsilon$



gambar 1. u = sup S

Sedangkan definisi infimum, dapat dinyatakan dengan suatu bilangan $p \in R$ di mana $S \subseteq R$ dan memenuhi 2 kondisi berikut

1) $s \geq p \;, \forall s \in S$

2) $s \geq q \Rightarrow p \geq q, \forall s \in S$

Kondisi 1) di atas menyatakan bahwa p merupakan batas bawah dari S dan kondisi 2) menunjukkan bahwa p sebagai infimum, karena lebih besar dari semua batas bawah dari kumpulan S

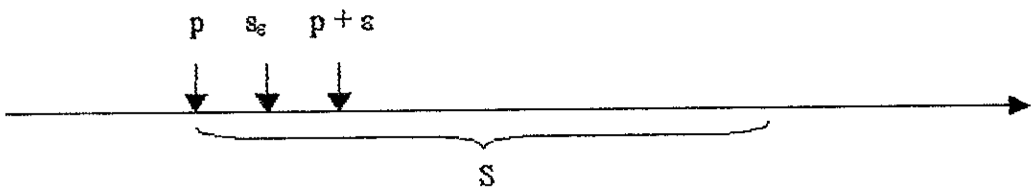
Lemma 2.1.2

Suatu batas bawah p dari kumpulan tak kosong $S \subseteq R$ merupakan infimum S, $\forall \varepsilon > 0 \Leftrightarrow \exists\, s_\varepsilon \in S$ dengan $p + \varepsilon > s_\varepsilon$

Bukti,

Andaikan p memenuhi kondisi yang ditetapkan, akan ditunjukkan bahwa p = inf S. Misalkan ε adalah sebarang batas bawah dari X dengan $q \neq p$. Jika diasumsikan bahwa $q > p$ dan mengambil $\varepsilon = q - p > 0$, hal ini mengimplikasikan bahwa $\exists\, s_\varepsilon \in S, \ni$ $q = p + \varepsilon > s_\varepsilon$. Akan tetapi hal ini kontradiksi dengan pernyataan q sebagai batas bawah dari S. Dengan hipotesis $q > p$ yang menyatakan kontardiksi, dapat disimpulkan bahwa $p > q$, sehingga dapat ditunjukkan bahwa $p = \inf S$.

Sebaliknya, andaikan $p = \inf S$ dan $\varepsilon > 0$. Maka $p+\varepsilon > p$ sehingga $p+\varepsilon$ bukan suatu batas bawah dari S, karena ada beberapa elemen $s_\varepsilon \in S$ yang kurang dari $p+\varepsilon$ mengakibatkan $p+\varepsilon > s_\varepsilon$



gambar 2. $p = \inf S$

Definisi 2.1.4

Misalkan $X = \{x_n\}$ suatu barisan dari bilangan riil. Maka bilangan riil x disebut sebagai batas dari X, $\forall V$ dari $x \Rightarrow \exists\, K(V) \in N, \ni \forall\, n \geq K(V)$, x_n termasuk dalam V. Jika x adalah batas dari barisan X, dapat dikatakan bahwa X mempunyai batas atau konvergen ke x.

Notasi dari K(V) digunakan untuk memberikan kesan bahwa pemilihan K bergantung pada pemilihan persekitaran V. Hal ini menjelaskan bahwa persekitaran V yang "kecil" akan selalu membutuhkan suatu "nilai yang besar" dari K(V) yang menjamin $x_n \in V$, $\forall\, n \geq K(V)$. Untuk barisan bilangan riil $X = \{x_n\}$ yang mempunyai limit $x \in R$, biasa dinotasikan dengan

$$\lim_{n\to\infty} \{x_n\} = x \qquad (2.1.2)$$

Teorema 2.1.1

Misalkan $X = \{x_n\}$ adalah barisan dari bilangan riil dan $x \in R$. Maka pernyataan-pernyataan berikut adalah ekuivalen

a) X konvergen ke x

b) $\forall V_\varepsilon$ dari x, $\exists K(\varepsilon) \in N, \ni \forall n \geq K(\varepsilon)$, dikatakan x_n termasuk dalam V.

c) $\forall \varepsilon > 0, \exists K(\varepsilon) \in N, \ni \forall n \geq K(\varepsilon)$ maka $x - \varepsilon < x_n < x + \varepsilon$.

d) $\forall \varepsilon > 0, \exists K(\varepsilon) \in N, \ni \forall n \geq K(\varepsilon)$ maka $| x_n - x | < \varepsilon$

Penjelasannya,

a) $\Rightarrow$ b) Jika X konvergen ke x sesuai definisi 2.1.4 maka $V_\varepsilon = (x - \varepsilon, x + \varepsilon)$, merupakan suatu persekitaran dari x, $\exists K(\varepsilon) = K(V_\varepsilon), \ni \forall n \geq K(\varepsilon)$, dikatakan x_n termasuk dalam V.

b) $\Rightarrow$ c) $x_n \in V_\varepsilon \Rightarrow x - \varepsilon < x_n < x + \varepsilon$.

c) $\Rightarrow$ d) $x - \varepsilon < x_n < x + \varepsilon \Rightarrow | x_n - x | < \varepsilon$.

d) $\Rightarrow$ a) Andaikan dimiliki pilihan yang berubah-ubah untuk $\varepsilon > 0$, dapat ditunjukkan bahwa $X = \{x_n\}$ konvergen ke x sesuai dengan definisi 2.1.4, dimisalkan V persekitaran yang berubah-ubah dari x. Dengan definisi persekitaran dari suatu titik $\exists \varepsilon > 0, \ni V \supseteq V_\varepsilon = (x-\varepsilon, x+\varepsilon)$. Untuk $n \geq K(\varepsilon)$ maka $| x_n - x | < \varepsilon$; hal ini menunjukkan bahwa untuk $n \geq K(\varepsilon)$ akan diperoleh $x - \varepsilon < x_n < x + \varepsilon$. Oleh sebab itu, $n \geq K(\varepsilon) \Rightarrow x_n \in V$, sehingga dapat diambil K(V) sesuai definisi 2.1.4, yang akan sama dengan $K(\varepsilon)$.

Contoh 2.1.1

Dipunyai barisan $\{\frac{1}{n}\}$ akan konvergen ke 0

Penyelesaian,

Diberikan $\varepsilon > 0$ maka $1/\varepsilon > 0$, jika $K(\varepsilon)$ adalah bilangan asli dimana $K(\varepsilon) > 1/\varepsilon$ maka $\forall\ n \in N, \ni n \geq K(\varepsilon)$, akan diperoleh $n > 1/\varepsilon$ dan $1/n < \varepsilon$. Untuk itu jika $n \geq K(\varepsilon)$, maka

$$\left| \frac{1}{n} - 0 \right| = \frac{1}{n} < \varepsilon$$

Sehingga barisan $\{1/n\}$ akan konvergen ke 0, tidak masalah berapapun nilai ε asalkan $\varepsilon > 0$ untuk mendapatkan sebuah bilangan asli $K(\varepsilon) > 1/\varepsilon$.

2.2 Distribusi Bivariat, Marjinal dan Ekspektasi Bersyarat

Dalam berbagai keadaan seringkali diharuskan menghubungkan dua atau lebih variabel random secara serentak. Berikut akan disajikan distribusi probabilitas bersama dua variabel untuk memperoleh distribusi marjinal dan bersyarat; ekspektasi bersyarat; serta kovarian dan korelasi yang diambil dari tulisan **Robert G. D. Stell** dan **James H. Torrie**, 1995 serta **William G. Hines** dan **Douglas Montgomery**, 1990.

Definisi 2.2.1

Fungsi probabilita untuk dua variabel (bivariat) random adalah,

1. Kasus diskrit, untuk setiap hasil $[x_i, y_j]$ dari $[X, Y]$ digabungkan jumlahnya $p(x_i, y_j) = P(X = x_i \text{ dan } Y = y_j)$, di mana

   a) $p(x_i, y_j) \geq 0, \forall\ i,j$ (2.2.1)

   b) $\sum_i \sum_j p(x_i, y_j) = 1$ (2.2.2)

   c) $P(X{=}a, Y{=}b) = p(b).p(a \mid b) = p(a).p(b \mid a)$, (2.2.3)

   untuk kejadian saling bebas $p(a \mid b) = p(a)$ dan $p(b \mid a) = p(b)$, maka

$$P(X=a, Y=b) = p(a).p(b) = p(b).p(a) \tag{2.2.4}$$

2. Kasus kontinu, jika [X, Y] adalah sebuah nilai random kontinu dalam ruang range R, fungsi densitas bersamanya mempunyai sifat

   a) $f(x,y) \geq 0, \forall(x,y) \in R$ (2.2.5)

   b) $\iint\limits_{R} f(x,y)dxdy = 1$ (2.2.6)

   c) $P(a \leq X \leq b, c \leq Y \leq d) = \int_{c}^{d}\int_{a}^{b} f(x,y)dxdy$ (2.2.7)

Distribusi dua variabel yang telah didefinisikan, kadang disebut dengan distribusi probabilitas bersama (dalam kasus kontinu disebut densitas bersama), variabel yang muncul sebagai distribusi X atau Y sendiri disebut distribusi marjinal.

Definisi 2.2.2

1. Untuk kasus diskrit, distribusi marjinal dari X dan Y adalah

$$p(x_i) = \sum_j p(x_i, y_j),\ i = 1, 2, \ldots \tag{2.2.8}$$

$$p(y_j) = \sum_i p(x_i, y_j),\ j = 1, 2, \ldots \tag{2.2.9}$$

2. Untuk kasus kontinu, distribusi marjinal dari X dan Y adalah

$$f(x) = \int_{-\infty}^{\infty} f(x,y)dy \tag{2.2.10}$$

$$f(y) = \int_{-\infty}^{\infty} f(x,y)dx \tag{2.2.11}$$

Ekspektasi dan varian dapat ditentukan melalui distribusi marjinalnya, untuk kasus diskrit

$$E(X) = \mu_x = \sum_i x_i p_1(x_i) = \sum_i \sum_j x_i p(x_i, y_j) \tag{2.2.12}$$

$$V(X) = \sigma_x^2 = \sum_i (x_i - \mu_x)^2 p(x_i)$$

$$= \sum_i \sum_j (x_i - \mu_x)^2 p(x_i, y_j)$$

$$= \sum_i x_i^2 p(x_i) - \mu_x^2$$

$$= \sum_i \sum_j x_i^2 p(x_i, y_j) - \mu_x^2 \tag{2.2.13}$$

dan

$$E(Y) = \mu_y = \sum_j y_j p(y_j) = \sum_i \sum_j y_j p(x_i, y_j) \tag{2.2.14}$$

$$V(Y) = \sigma_y^2 = \sum_j (y_j - \mu_y)^2 p(y_j)$$

$$= \sum_i \sum_j (y_j - \mu_y)^2 p(x_i, y_j)$$

$$= \sum_j y_j^2 p(y_j) - \mu_y^2$$

$$= \sum_i \sum_j y_j^2 p(x_i, y_j) - \mu_y^2 \tag{2.2.15}$$

Sedangkan pada kasus kontinu,

$$E(X) = \mu_x = \int_{-\infty}^{\infty} x.f(x)dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x.f(x,y)dydx \tag{2.2.16}$$

$$V(X) = \sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x)dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x,y)dydx$$

$$= \int_{-\infty}^{\infty} x^2 f(x) - \mu_x^2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 p(x,y)dydx - \mu_x^2 \quad (2.2.17)$$

dan

$$E(Y) = \mu_y = \int_{-\infty}^{\infty} y.f(y)dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y.f(x,y)dxdy \quad (2.2.18)$$

$$V(Y) = \sigma_y{}^2 = \int_{-\infty}^{\infty} (y-\mu_y)^2 f(y)dy$$

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

$$= \int_{-\infty}^{\infty} y^2 f(y) - \mu_y^2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2 f(x,y)dxdy - \mu_y^2 \quad (2.2.19)$$

Selain distribusi bersama dan marjinal di atas, bisa juga dicari distribusi bersyarat yaitu mendapatkan distribusi dari satu variabel diberikan dari variabel lainnya.

Definisi 2.2.3

Untuk kasus diskrit, distribusi probabilita bersyaratnya adalah

$$p_{x|y}(x) = \frac{p(x,y)}{p(y)} \quad (2.2.20)$$

$$p_{y|x}(y) = \frac{p(x,y)}{p(x)} \quad (2.2.21)$$

dengan $p(x) > 0$ dan $p(y) > 0$

Untuk kasus kontinu, densitas bersyaratnya adalah

$$f_{x|y}(x) = \frac{f(x,y)}{f(y)} \qquad (2.2.22)$$

$$f_{y|x}(y) = \frac{f(x,y)}{f(x)} \qquad (2.2.23)$$

dengan $f(x) > 0$ dan $f(y) > 0$

Bila diperhatikan terdapat ekspektasi bersyarat, yaitu $E(X|Y = y)$ untuk setiap nilai y, di mana setiap nilai $E(X|Y = y)$ bergantung pada nilai y yang perubahannya dipengaruhi oleh fungsi probabilitasnya.

Definisi 2.2.4

Untuk kasus diskrit didefinikan ekspektasi bersyaratnya,

$$E(X|Y = y_j) = \sum_i x_i p_{x|y_j}(x_i) \qquad (2.2.24)$$

$$E(Y|X = x_i) = \sum_j y_j p_{y|x_i}(y_j) \qquad (2.2.25)$$

Untuk kasus kontinu didefinisikan ekspektasi bersyaratnya,

$$E(X|Y = y) = \int_{-\infty}^{\infty} x.f_{x|y}(x)dx \qquad (2.2.26)$$

$$E(Y|X = x) = \int_{-\infty}^{\infty} y.f_{y|x}(y)dy \qquad (2.2.27)$$

Diketahui bahwa $E(X) = \mu_x$ dan $V(X) = \sigma_x^2$ adalah rata-rata dan varian dari X. Nilai E(X) dan V(X) dapat ditentukan dari distribusi marjinal X, dengan pengertian yang sama μ_y dan σ_y^2 adalah rata-rata dan varian dari Y. Sedangkan gambaran yang digunakan untuk melihat tingkat asosiasi antara X dan Y adalah kovarian dan koefisien korelasi.

Definisi 2.2.10

Jika [X,Y] variabel random berdimensi dua, kovarian yang dinotasikan dengan σ_{xy}^2 adalah

$$\mathrm{cov}(X,Y) = \sigma_{xy}^2 = E[(X - E(X))(Y - E(Y))] \tag{2.2.28}$$

Dan koefisien korelasi yang dinotasikan dengan ρ, adalah

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{\sigma_{xy}^2}{\sigma_x . \sigma_y} \tag{2.2.29}$$

nilai koefisien korelasi $-1 \le \rho \le 1$

2.3 Distribusi Normal

Algoritma EM pada penulisan tugas akhir diasumsikan berdistribusi normal atau Gaussian, maka berikut akan diuraikan distribusi probabilitas variabel random kontinu berdistribusi normal, sesuai dengan tulisan **William G. Hines** dan **Douglas C. Montgomery**, 1990.

Sebuah variabel random X, disebut mempunyai distribusi normal dengan rata–rata $\mu(-\infty<\mu<\infty)$ dan varian $\sigma^2>0$, fungsi densitas distribusi normal didefinisikan dengan,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2\right], -\infty < x < \infty \tag{2.3.1}$$

Rata-rata dari distribusi normal dapat dicari, yaitu

$$E(X) = \int_{-\infty}^{\infty} \frac{x}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2\right]dx\,, \tag{2.3.2}$$

jika dimisalkan $z = (x - \mu)/\sigma$, akan diperoleh

$$E(X) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} (\mu + \sigma z) \exp\left(-\frac{z^2}{2}\right)dz$$

$$= \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)dz + \sigma \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} z.\exp\left(-\frac{z^2}{2}\right)dz\,, \tag{2.3.3}$$

karena integral bagian pertama adalah sebuah densitas normal standar, maka nilai integral pertama adalah satu. Bagian integral kedua mempunyai nilai nol, yaitu:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} z.\exp\left(-\frac{z^2}{2}\right)dz = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)\Bigg|_{-\infty}^{\infty} = 0\,, \tag{2.3.4}$$

sehingga,

$$E(X) = \mu[1] + \sigma[0] = \mu \tag{2.3.5}$$

Untuk memperoleh varian harus dihitung

$$V(X) = E[(X - \mu)^2]$$

$$= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2\right]dx \tag{2.3.6}$$

dengan memisalkan $z = (x - \mu)/\sigma$, diperoleh

$$V(X) = \int_{-\infty}^{\infty} \sigma^2 z^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)dz$$

$$= \sigma^2 \left[\int_{-\infty}^{\infty} \frac{z^2}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \right]$$

$$= \sigma^2 \left[\frac{-z.\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}} + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \right]$$

$$= \sigma^2 [\ 0 + 1\ ] = \sigma^2 \qquad (2.3.7)$$

Distribusi normal, $X \sim N(\mu, \sigma^2)$ fungsi pembangkit momennya dinotasikan oleh,

$$M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2) \qquad (2.3.8)$$

dengan sifat :

1) $M_{ax}(t) = M_x(at)$ (2.3.9)

2) $M_{a+bx}(t) = \exp(at).\ M_x(bt)$ (2.3.10)

Teorema 2.3.1

Jika $X \sim N(\mu, \sigma^2)$, maka $\frac{X-\mu}{\sigma} \sim N(0,1)$

Bukti,

$$M_{X-\mu}(t) = \exp(-\mu t).M_X(t)$$

$$= \exp(-\mu t).\exp(\mu t + \frac{1}{2}\sigma^2 t^2)$$

$$= \exp(\frac{1}{2}\sigma^2 t^2) \qquad (2.3.11)$$

$$M_{\frac{x-\mu}{\sigma}}(t) = M_{x-\mu}\left(\frac{t}{\sigma}\right)$$

$$= \exp\left(\frac{1}{2}\sigma^2\left(\frac{t}{\sigma}\right)^2\right)$$

$$= \exp\left(\frac{1}{2}t^2\right) \Rightarrow N(0,1) \qquad (2.3.12)$$

Didefinisikan fungsi distribusi F, dengan persamaan

$$F(x) = P(X \leq x)$$

$$= \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) du \qquad (2.3.13)$$

Hal ini tidak dimungkinkan untuk menghitung integral persamaan 2.3.13, tanpa menggunakan metode secara numerik, dan seringkali penilaian dilakukan dengan menyelesaikan persamaan masing-masing pasangan (μ, σ^2). Meskipun transformasi sederhana dari variabel $z = (x - \mu)/\sigma$, menyebabkan penilaian bebas dari μ dan σ. Yaitu,

$$F(x) = P(X \leq x)$$

$$= P\left(Z \leq \frac{x-\mu}{\sigma}\right)$$

$$= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$$

$$= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \varphi(z) dz$$

$$= \Phi\left(\frac{x-\mu}{\sigma}\right), \tag{2.3.15}$$

dan densitasnya adalah

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right), \ -\infty < z < \infty \tag{2.3.16}$$

Dengan Z ~ N(0, 1) atau mempunyai distribusi normal standar. Fungsi distribusi yang bersesuaian adalah Φ, di mana

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2}{2}\right) du \ ,$$

dan hasil fungsi ini biasanya disajikan dalam tabel normal standar

2.4 Pertidaksamaan Jensen

Salah satu konsep yang mendasari algoritma EM yaitu, untuk setiap iterasi nilai ekspektasinya akan meningkat. Untuk itu akan diuraikan pertidaksamaan Jensen sesuai tulisan **Robert J Boik**, 2002 dan **Roger Koenker**, 2004 yang digunakan untuk melihat peningkatan nilai setiap iterasi dari algoritma EM.

Definisi 2.4.1

- *Convex Set.* Andaikan S merupakan subset ruang berdimensi n. Subset S disebut *convex* jika

$$x \in S \text{ and } y \in S \Rightarrow \alpha x + (1-\alpha)y \in S, \ \forall \ \alpha \in (0; 1) \tag{2.4.1}$$

- Fungsi *Convex.* Suatu fungsi g(x) didefinisikan pada *convex set* S, merupakan fungsi *convex* jika

$$g[\alpha x + (1-\alpha)y] \le \alpha g(x) + (1-\alpha)g(y),\ \forall\ \alpha \in (0;1) \qquad (2.4.2)$$

Teorema 2.4.1

Jika x dan g(x) adalah variabel random yang dapat diintegralkan dan g(.) fungsi *convex*, maka $g[E(x)] \le E[g(x)]$

Bukti,

Convex dari g mengimplikasikan bahwa untuk setiap ξ terdapat suatu garis L yang melewati titik $(\xi, g(\xi))$ sedemikian sehingga grafik g di atas garis. Contoh

$$g(x) \ge g(\xi) + \lambda(x-\xi) \qquad (2.4.3)$$

misalkan, $\xi = E(X)$ maka untuk setiap x

$$g(x) \ge g(E(X)) + \lambda(x - E(X)) \qquad (2.4.4)$$

sebagai catatan bahwa λ bergantung pada ξ bukan pada x. sekarang dimisalkan x = X akan diperoleh

$$g(X) \ge g(E(X)) + \lambda(X - E(X)) \qquad (2.4.5)$$

Dari persamaan 2.4.5 terlihat bahwa $g(X) \ge g(E(X))$ sedemikian sehingga $E(g(X)) \ge g(E(X))$.

2.5 Metode Maksimum *Likelihood*

Metode maksimum *likelihood* merupakan metode yang baik untuk memperoleh sebuah estimator. Berikut akan diuraikan metode maksimum *likelihood* yang ditulis oleh **William W. Hines** dan **Douglas Montgomery**, 1990. Misalkan X variabel random dengan distribusi probabilitas $f(x,\theta)$, di mana parameter θ tidak diketahui. Misalkan $x_1, x_2, \ldots, x_n$ menjadi nilai yang diamati di dalam suatu variabel random yang besarnya n. Maka fungsi *likelihood* sampel tersebut adalah

$$L(\theta \mid X) = \prod_{i=1}^{n} f(x_i, \theta) \tag{2.5.1}$$

Estimator maksimum *likelihood* untuk θ adalah nilai θ yang memaksimumkan fungsi *likelihood* $L(\theta)$. Tujuan utama dari estimator maksimum *likelihood* adalah mencari nilai θ yang memaksimalkan probabilitas kejadian hasil sampel. Untuk memudahkan dalam mencari estimasi parameter, fungsi *likelihood* biasanya ditransformasikan ke dalam bentuk *loglikelihood*, yaitu

$$\lambda(\theta \mid X) = \log\,[L(\theta \mid X)] \tag{2.5.2}$$

Metode maksimum *loglikelihood* dapat digunakan untuk memperkirakan keadaan di mana beberapa parameter tidak diketahui, misalkan $\theta_1, \theta_2, \ldots, \theta_k$. Dalam keadaan seperti ini fungsi *loglikelihood* adalah sebuah fungsi dengan k parameter tidak diketahui. Estimator maksimum *loglikelihood* $\{\hat{\theta}_i\}$ dapat diperoleh dengan menyelesaikan persamaan turunan parsial pertama k, $\partial\lambda(\theta_1, \theta_2, \ldots, \theta_k)/\partial\theta_i = 0$, di mana $i = 1, \ldots, k$.

Sebagai contoh misalkan X berdistribusi normal dengan rata-rata μ dan varian σ^2, di mana μ dan σ^2 tidak diketahui. Untuk menentukan estimator maksimum *likelihood* untuk μ dan σ^2. Maka fungsi *likelihood* untuk sampel random berukuran n adalah

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \tag{2.5.3}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right) \tag{2.5.4}$$

$$\lambda(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \tag{2.5.5}$$

Maka,

$$\frac{\partial\lambda(\mu,\sigma^2)}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \tag{2.5.6}$$

dan

$$\frac{\partial\lambda(\mu,\sigma^2)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \tag{2.5.7}$$

Sehingga penyelesaian MLE adalah,

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}X_i = \overline{X} \tag{2.5.8}$$

dan

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2 \tag{2.5.9}$$

2.6 Teorema Bayesian

Dalam penulisan tugas akhir ini, data yang diolah berupa data tidak lengkap yang memuat beberapa nilai hilang sedangkan nilai parameter yang akan dicari adalah nilai parameter data lengkapnya. Sehingga untuk mencari MLE, parameternya akan bergantung kepada nilai data lengkap. Untuk itu berikut akan diuraikan teorema Bayesian yang ditulis oleh **William W. Hines** dan **Douglas Montgomery**, 1990 serta **Walsh**, 2002.

Dalam metode maksimum *likelihood*, parameter θ merupakan besaran tetap yang tidak diketahui, sedangkan dalam pendekatan Bayesian θ dipandang sebagai variabel random yang mempunyai distribusi khusus yang disebut *prior*. Jadi di dalam pendekatan Bayesian di samping informasi sampel, harus diperhatikan juga tentang informasi θ itu sendiri. Informasi tentang θ dinyatakan dalam bentuk distribusi *prior* $\pi(\theta)$. Misalkan distribusi X dengan $f(x|\theta)$, untuk menunjukkan bahwa distribusi tersebut bergantung pada parameter θ yang tidak diketahui. Bila diambil sampel random X, dengan nilai $x_1, x_2, \ldots, x_n$ maka densitas gabungan atau *likelihood*nya adalah

$$f(x_1, x_2, \ldots, x_n|\theta) = f(x_1|\theta)\, f(x_2|\theta) \ldots f(x_n|\theta) \qquad (2.6.1)$$

Didefinisikan distribusi θ yang baru atau *posterior* sebagai distribusi bersyarat dari θ, memberikan hasil sampel yang memenuhi,

$$\pi(\theta|x_1\, x_2, \ldots, x_n) = \frac{f(x_1, x_2, \ldots, x_n; \theta)}{f(x_1, x_2, \ldots, x_n)}, \qquad (2.6.2)$$

distribusi bersama antara sampel dan θ di dalam pembilang persamaan 2.6.2 adalah hasil *prior* θ dan *likelihood*nya atau

$$f(x_1, x_2, \ldots, x_n; \theta) = f(x_1, x_2, \ldots, x_n|\theta).\pi(\theta) \qquad (2.6.3)$$

Penyebut persamaan 2.6.2 adalah distribusi marjinal yang hanya menormalkan konstanta yang diperoleh dengan

$$f(x_1, x_2, \ldots, x_n) = \begin{cases} \int_{-\infty}^{\infty} f(\theta) f(x_1, x_2, \ldots, x_n | \theta) d\theta & x \text{ kontinu} \\ \sum_{\theta} f(\theta) f(x_1, x_2, \ldots, x_n | \theta) & x \text{ diskrit} \end{cases} \quad (2.6.4)$$

Sebagai konsekuensinya dapat dituliskan *posterior* atau distribusi θ yang baru, yaitu

$$\pi(\theta | x_1, x_2, \ldots, x_n) = \frac{\pi(\theta) f(x_1, x_2, \ldots, x_n | \theta)}{f(x_1, x_2, \ldots, x_n)} \quad (2.6.5)$$

Bila diperhatikan teorema Bayes telah mengubah atau memperbaiki distribusi awal ke distribusi yang baru. Distribusi yang baru merefleksikan tingkat kepercayaan tentang θ yang memberikan keterangan sampel. Selanjutnya distribusi yang baru proporsional untuk menghasilkan distribusi awal dan likelihoodnya, konstanta proporsionalnya menjadi konstanta kenormalan $f(x_1, x_2, \ldots, x_n)$. Kemudian densitas yang baru untuk θ mengungkapkan tingkat kepercayaan tentang θ yang diberikan hasil sampel.

Walsh, 2002 meyebutkan bahwa dalam mencari distribusi *posterior* $\pi(\theta | x)$, melibatkan integral dalam kasus kontinu dan jumlahan dalam kasus diskrit. Cara ini mungkin kurang praktis, ada cara yang singkat sebagai alternatif yaitu menggunakan kesebandingannya,

$$\pi(\theta | x) \propto f(x | \theta).\pi(\theta) \quad (2.6.6)$$

di mana simbol α menyatakan "proporsional dengan".

Keistimewaan lain dari analisa Bayesian adalah dapat dipindahkan atau dihilangkannya efek atau parameter yang mengganggu dengan mengintegralkannya keluar dari distribusi *posterior*, untuk membangun distribusi marjinal *posterior* dari parameter yang diinginkan.

Posterior marjinal mungkin melibatkan beberapa parameter, misalkan parameter tidak diketahui $\theta = (\theta_1, \theta_n)$, misalkan θ_n merupakan parameter yang mengganggu. Maka integral terhadap θ_n akan menghasilkan distribusi marjinal dari θ_1, yaitu

$$\pi(\theta_1 \mid x) = \int_{\theta_n} \pi(\theta_1, \theta_n \mid x) d\theta_n \tag{2.6.7}$$

Untuk mengambil sebuah estiamtor Bayes dari parameter yang tidak diketahui, di mana terdapat beberapa nilai yang mungkin. Sesuai dengan metode maksimum *likelihood* dapat dicari suatu *mode* (nilai maksimal) dari parameternya, yaitu

$$\begin{aligned} \hat{\theta} &= \max_{\theta} \; [p(\theta|x)] \\ &= E\,[(\theta \mid x)] \\ &= \int \theta .\ \pi(\theta \mid x)\ d\theta \end{aligned} \tag{2.6.8}$$