BAB II

LANDASAN TEORI

Pada bab ini akan dibahas teori – teori yang mendukung metode *upper level set scan statistics*, antara lain peubah acak, distribusi gamma, fungsi gamma, fungsi likelihood, dan uji rasio likelihood.

2.1 PEUBAH ACAK

Misalkan dilakukan percobaan acak dengan ruang sampel $\mathscr{C}$. Sebuah ruang sampel $\mathscr{C}$ mungkin menggambarkan elemen dari $\mathscr{C}$ yang bukan angka. Misalnya dapat kita lihat dalam percobaan acak pelemparan sebuah koin, ruang sampel yang berkaitan dengan percobaan tersebut adalah $\mathscr{C}$ = { c | c adalah muka atau c adalah belakang }. Misalkan X adalah sebuah fungsi sedemikian sehingga X(c) = 0 jika c adalah muka dan X(c) = 1 jika c adalah belakang. Fungsi X disebut peubah acak.

Definisi 1

Misalkan sebuah percobaan acak dengan ruang sample $\mathscr{C}$. suatu fungsi X yang memetakan setiap elemen c $\in \mathscr{C}$ ke satu dan hanya satu bilangan riil X(c) = x, disebut peubah acak. Domain dari X adalah $\mathscr{C}$ dan range dari X adalah $\mathscr{A}$ = { x| $X(c) = x$, c $\in \mathscr{C}$ } .

C $\subset \mathscr{C}$ terjadi jika dan hanya jika A $\subset \mathscr{A}$ terjadi, dimana **C** = {c| c $\in \mathscr{C}$, X(c) $\in$ A}, sehingga probabilitas dari kejadian A sama dengan probabilitas bahwa hasil suatu percobaan berada di C atau maka Pr{$X \in$ A } = P(C). Ada dua jenis peubah acak, yaitu peubah acak diskrit dan peubah acak kontinu. Peubah acak diskrit dan peubah acak kontinu dibedakan berdasarkan jenis ruang sampelnya.

Definisi 2

Misalkan X adalah peubah acak dengan ruang hasil $\mathscr{A}$, $\mathscr{A} \subset \mathbb{R}$, dan f adalah suatu fungsi pada bilangan riil. Jika :

1. $f(x) \geq 0,\ \forall x \in \mathscr{A}$
2. $\int_{\mathscr{A}} f(x)dx = 1$
3. A $\subset \mathscr{A}$ berlaku P{A} = Pr{x $\in$ A} = $\int_{A} f(x)dx$

dipenuhi, maka X disebut peubah acak kontinu dan f(x) disebut p.d.f (*Probability Density Function*) atau fungsi kepadatan dari X.

Misalkan peubah acak X memiliki probabilitas P(A), A $\subset \mathbb{R}$. Ambil suatu bilangan riil x dan anggap himpunan A adalah himpunan tak terbatas dari $-\infty$ sampai x, titik x termasuk dalam himpunan tersebut. Untuk setiap himpunan A berlaku P(A) = Pr{X $\in$ A} = Pr{X ≤ x}. F(x) = Pr{X ≤ x}, fungsi F(x)

disebut fungsi distribusi dari peubah acak X. Karena F(x) = Pr{X ≤ x}, maka dengan menggunakan p.d.f $f(x)$, fungsi distribusi dari X dapat dinyatakan dengan $F(x) = \int_{w \le x} f(w)dw$ untuk peubah acak kontinu.

2.2 DISTRIBUSI GAMMA

Suatu peubah acak X dikatakan berdistribusi Gamma, jika peubah acak tersebut memiliki p.d.f sebagai berikut:

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(k)\beta^k} x^{k-1} e^{-x/\beta}, & 0 < x < \infty \\ 0, & lainnnya \end{cases}$$

dengan parameter k dan β, kedua parameter bernilai positif.

k merupakan *shape parameter* dan β merupakan *scale parameter*. Mean dari peubah acak X yang berdistribusi gamma dengan parameter k dan β adalah

$$
\begin{aligned}
E[X] = \mu[X] &= \int_0^\infty x\, f(x;k,\beta)dx \\
&= \int_0^\infty x \frac{1}{\Gamma(k)\beta^k} x^{k-1} e^{-x/\beta} dx \\
&= \int_0^\infty \frac{1}{\Gamma(k)\beta^k} x^k e^{-x/\beta} dx \\
&= \frac{1}{\Gamma(k)\beta^k} \int_0^\infty x^k e^{-x/\beta} dx \\
&= \frac{1}{\Gamma(k)\beta^k} \int_0^\infty (y\beta)^k e^{-y} \beta dy \\
&= \frac{1}{\Gamma(k)\beta^k} \beta^{k+1} \int_0^\infty y^k e^{-y} dy \\
&= \frac{1}{\Gamma(k)\beta^k} \beta^{k+1} \Gamma(k+1) \\
&= \frac{1}{\Gamma(k)\beta^k} \beta^{k+1} k\Gamma(k) \\
&= k\beta
\end{aligned}
$$

Misalkan

$$y = \frac{x}{\beta}$$

$$dy = \frac{1}{\beta} dx$$

Variansi dari peubah acak X yang berdistribusi gamma dengan parameter k dan β adalah

$$
\begin{aligned}
\mathrm{Var}[X] = \sigma^2[X] = E[X^2] - [E[X]]^2 \\
&= \int_0^\infty x^2 \, f(x;k,\beta)dx - [k\beta]^2 \\
&= \int_0^\infty x^2 \frac{1}{\Gamma(k)\beta^k} x^{k-1} e^{-x/\beta} dx - k^2\beta^2 \\
&= \int_0^\infty \frac{1}{\Gamma(k)\beta^k} x^{k+1} e^{-x/\beta} dx - k^2\beta^2 \\
&= \frac{1}{\Gamma(k)\beta^k} \int_0^\infty x^{k+1} e^{-x/\beta} dx - k^2\beta^2 \\
&= \frac{1}{\Gamma(k)\beta^k} \int_0^\infty (y\beta)^{k+1} e^{-y} \beta dy - k^2\beta^2 \\
&= \frac{1}{\Gamma(k)\beta^k} \beta^{k+2} \int_0^\infty y^{k+1} e^{-y} dy - k^2\beta^2 \\
&= \frac{1}{\Gamma(k)\beta^k} \beta^{k+2} \Gamma(k+2) - k^2\beta^2 \\
&= \frac{1}{\Gamma(k)\beta^k} \beta^{k+2} (k+1)\Gamma(k+1) - k^2\beta^2 \\
&= \frac{1}{\Gamma(k)\beta^k} \beta^{k+2} (k+1)k\Gamma(k) - k^2\beta^2 \\
&= (k+1)k\beta^2 - k^2\beta^2 \\
&= k^2\beta^2 + k\beta^2 - k^2\beta^2 \\
&= k\beta^2
\end{aligned}
$$

2.3 FUNGSI GAMMA DAN FUNGSI DIGAMMA

Fungsi Gamma

Fungsi gamma yang dinyatakan oleh $\Gamma(\alpha)$ didefinisikan oleh

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

untuk $\alpha > 0$.

Sebuah rumus rekursif untuk fungsi gamma adalah

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$$

Khusus untuk α adalah sebuah bilangan bulat positif, maka

$$\Gamma(\alpha+1) = \alpha! \qquad \alpha = 1, 2, 3, \ldots$$

Fungsi Digamma

Fungsi digamma yang dinyatakan dengan $\psi(\alpha)$ merupakan turunan ln fungsi gamma.

$$\psi(\alpha) = \frac{d \ln \Gamma(\alpha)}{d\alpha} = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

2.4 FUNGSI LIKELIHOOD

Fungsi likelihood merupakan joint p.d.f (p.d.f bersama) dari beberapa peubah acak.

Definisi 3

Misalkan $X_1, X_2, ..., X_n$ merupakan peubah acak yang memiliki p.d.f $f(x;\theta), \theta \in \Omega$ dengan θ merupakan suatu parameter yang tidak diketahui. Maka fungsi likelihood adalah sebagai berikut

$$L(\theta; x_1, x_2, ..., x_n) = f(x_1;\theta) f(x_2;\theta)... f(x_n;\theta)$$

Akan dicari taksiran θ yang dapat memaksimumkan fungsi likelihood dengan metode maximum likelihood.

2.5 MAXIMUM LIKELIHOOD ESTIMATOR

Maximum likelihood estimator adalah metode yang digunakan untuk menaksir suatu parameter yang tidak diketahui dalam suatu fungsi probabilitas.

Misalkan X adalah peubah acak yang mempunyai bentuk fungsi probabilitas tertentu tetapi fungsi probabilitasnya bergantung kepada suatu parameter θ yang tidak diketahui. Fungsi probabilitas X dapat ditulis sebagai $f(x;\theta), \theta \in \Omega$, dengan Ω adalah ruang parameter.

Dalam metode maksimum likelihood, misalkan terdapat peubah acak $X_1, X_2, ..., X_n$ dengan p.d.f $f(x;\theta), \theta \in \Omega$ dan fungsi likelihood L(θ ; $x_1, x_2, ..., x_n$), akan dicari suatu fungsi dari $x_1, x_2, ..., x_n$, misal s($x_1, x_2, ..., x_n$), sedemikian sehingga jika θ diganti oleh s($x_1, x_2, ..., x_n$), maka nilai fungsi likelihood L akan maksimum dan s($x_1, x_2, ..., x_n$) disebut sebagai maksimum likelihood estimator (MLE) dari θ .

Definisi 4

Misalkan X_1, X_2, …, X_n adalah peubah acak. $L(\theta; x_1, x_2, ..., x_n)$ mencapai maksimum di Ω pada saat $\theta = S(x)$ sehingga

$$\sup_{\theta \in \Omega} L(\theta; x_1, x_2, ..., x_n) = L(S(x)), \quad S(x) \in \Omega$$

Maka statistic $\hat{\theta} = S(x)$ disebut sebagai maksimum likelihood estimator (MLE) dari θ .

2.6 UJI RASIO LIKELIHOOD

Uji rasio likelihood adalah suatu metode yang digunakan untuk menguji hipotesis H_0: $\theta \in \omega$ terhadap hipotesis H_1: $\theta \in \Omega$ dengan Ω adalah ruang parameter keseluruhan dan ω adalah ruang parameter dalam H_0. $\omega \subset \Omega$.

Misalkan X_1, X_2, …, X_n adalah n peubah acak yang saling bebas dengan masing-masing fungsi distribusinya adalah $f_i(x_i, \theta_1, \theta_2, ..., \theta_n), i = 1, 2, ..., n$. Misalkan Ω adalah himpunan yang mengandung seluruh titik parameter $(\theta_1, \theta_2, ..., \theta_n)$, Ω disebut ruang parameter. Misalkan ω adalah subset dari ruang parameter Ω. Akan diuji hipotesis

H_0: $(\theta_1, \theta_2, ..., \theta_m) \in \omega$ terhadap H_1: $(\theta_1, \theta_2, ..., \theta_m) \in \Omega$.

Definisikan fungsi likelihood berikut:

$$L(\omega) = \prod_{i=1}^{n} f_i(x_i; \theta_1, \theta_2, ..., \theta_m), \qquad (\theta_1, \theta_2, ..., \theta_m) \in \omega$$

dan

$$L(\Omega) = \prod_{i=1}^{n} f_i(x_i; \theta_1, \theta_2, ..., \theta_m), \qquad (\theta_1, \theta_2, ..., \theta_m) \in \Omega$$

dengan $L(\omega)$ adalah fungsi likelihood di hipotesis null, dan $L(\Omega)$ adalah fungsi likelihood dalam ruang parameter keseluruhan. Misalkan $L(\hat{\omega})$ adalah nilai maksimum dari $L(\omega)$ di ω dan $L(\hat{\Omega})$ adalah nilai maksimum $L(\Omega)$ di Ω. Rasio $L(\hat{\omega})$ terhadap $L(\hat{\Omega})$ disebut rasio likelihood dan dinotasikan oleh

$$\nabla(x_1, x_2, ..., x_n) = \nabla = \frac{L(\hat{\omega})}{L(\hat{\Omega})}$$

Karena $L(\omega)$ dan $L(\Omega)$ adalah fungsi probabilitas, $\nabla \geq 0$, dan $\omega \subset \Omega$ maka $\nabla \leq 1$. Jika ∇ digunakan sebagai kriteria pengujian maka daerah penolakan H_0 adalah himpunan terdefinisi dalam $0 \leq \nabla \leq \nabla_0$, dengan ∇_0 adalah suatu fungsi positif.

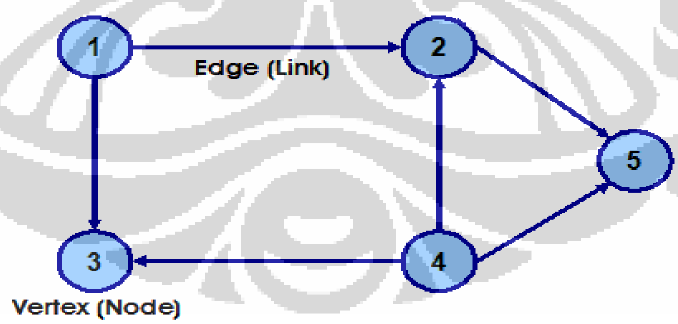
Fungsi ∇ didefinisikan oleh suatu peubah acak $\nabla(X_1, X_2, ..., X_n)$, dan nilai signifikansi dari pengujian diberikan oleh $\alpha = \Pr[\nabla(X_1, X_2, ..., X_n) \leq \nabla_0; H_0]$

2.7 NODE, EDGE, DAN TREE

Graph. Suatu *graph* G = (V, E) adalah suatu himpunan verteks (*node*) V yang dihubungkan oleh *edge*.E,.

Node. *Node* V adalah suatu titik pangkal (*terminal point*) atau titik pertemuan (*intersection point*) pada suatu *graph*.

Edge. *Edge* E adalah sesuatu yang menghubungkan dua *node*. Gambar 2.1 di bawah ini merupakan *graph* yang terdiri dari *node* dan *edge*.

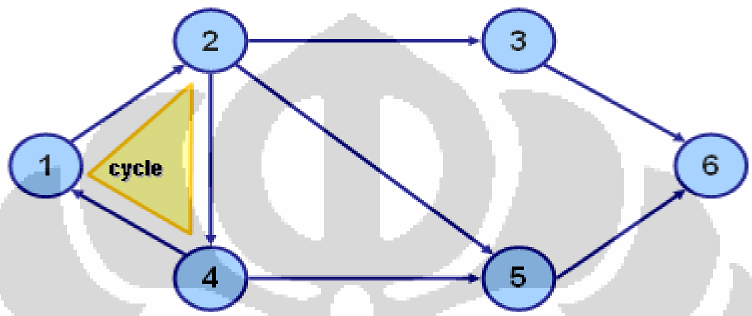


Gambar 2. 1 *Graph*, *Node*, dan *Edge*

Connected graph. *Connected graph* (graph terhubung) adalah *graph* dimana jika setiap pasang *node* yang berbeda terhubung.

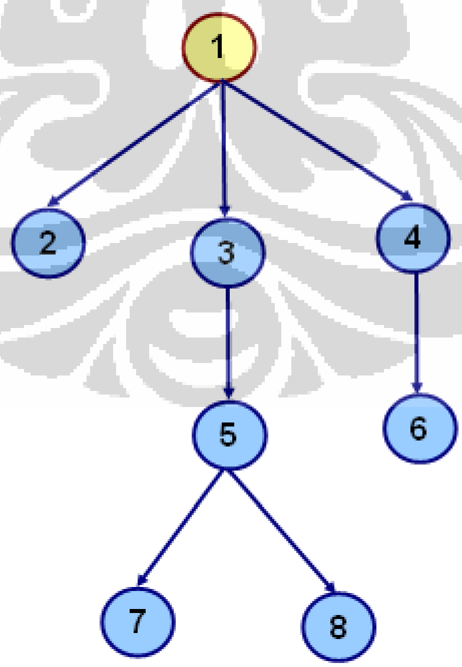
Cycle. *Cycle* adalah *path* pada *graph* yang mulai dan berakhir pada *node* yang sama.

Connected graph dan *Cycle* ditunjukkan oleh gambar 2.2.



Gambar 2. 2 *Connected graph* dan *Cycle*

Tree. *Tree* adalah *graph* terhubung yang tidak mengandung *cycle*.



Gambar 2. 3 *Tree*

Root node merupakan *node* dengan hirarki tertinggi.

Leaf adalah *node* yang tidak memiliki cabang.

Internal node adalah *node* yang bukan merupakan *leaf*.

Pada gambar 2.3 di atas, *node* 1 merupakan *root node*; *node* 2, *node* 6, *node* 7, dan node 8 merupakan *leaf*; dan *node* 1, *node* 3, *node* 4, dan *node* 5 merupakan *internal node*.