BAB III

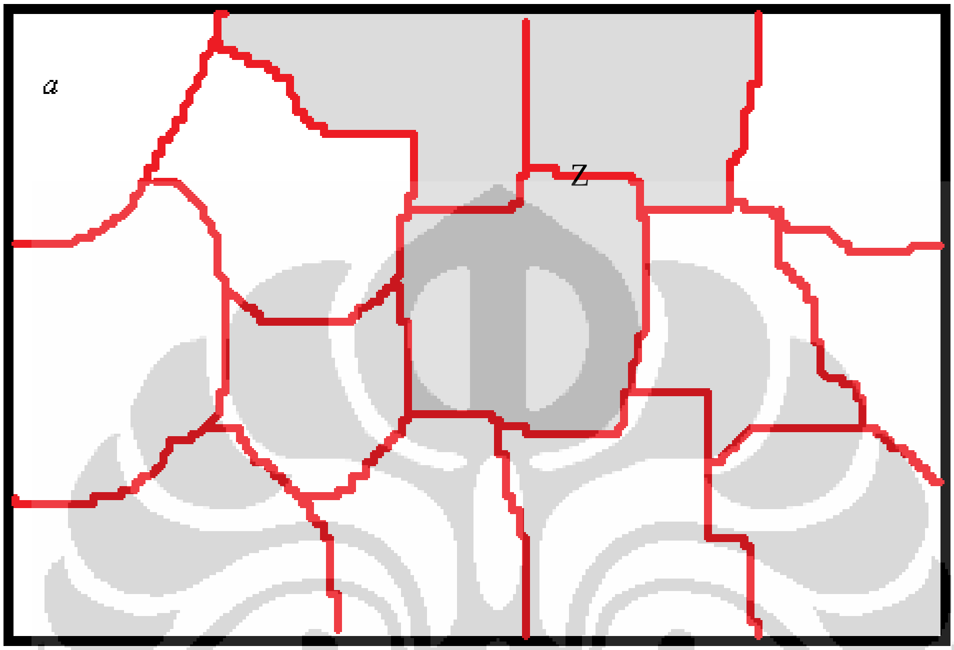
UPPER LEVEL SET SCAN STATISTICS

Bab ini akan membahas mengenai metode *upper level set scan statistics*. Selain itu, akan dibahas juga hal-hal yang berkaitan dengan metode *upper level set scan statistics*.

Berikut ini adalah istilah-istilah yang akan digunakan dalam metode *upper level set scan statistics*:

1. *Hotspot* adalah daerah yang memiliki intensitas tertinggi pada suatu lokasi tertentu
2. *Study area* adalah keseluruhan daerah yang akan diteliti, dinotasikan dengan R.
3. *Study area* dipartisi menjadi beberapa wilayah tertentu (dapat berdasarkan negara, propinsi, kode pos, dan sebagainya). Wilayah-wilayah ini disebut sel dan dinotasikan dengan a.
4. Zona adalah kumpulan sel-sel yang terhubung di dalam *study area* R dan dinotasikan dengan Z.
5. *Scanning window* adalah kumpulan daerah yang memiliki potensi untuk menjadi *most likely cluster*, dapat terdiri dari sel-sel maupun zona, dan dinotasikan dengan **Z**.
6. Level adalah suatu *rate* atau nilai yang digunakan untuk menentukan anggota-anggota *scanning window*, dinotasikan dengan g.

Agar lebih memahami istilah di atas, perhatikan gambar



Gambar 3.1 *Study area* (R), sel (*a*), dan zona (Z)

Keterangan

: batas study area, R

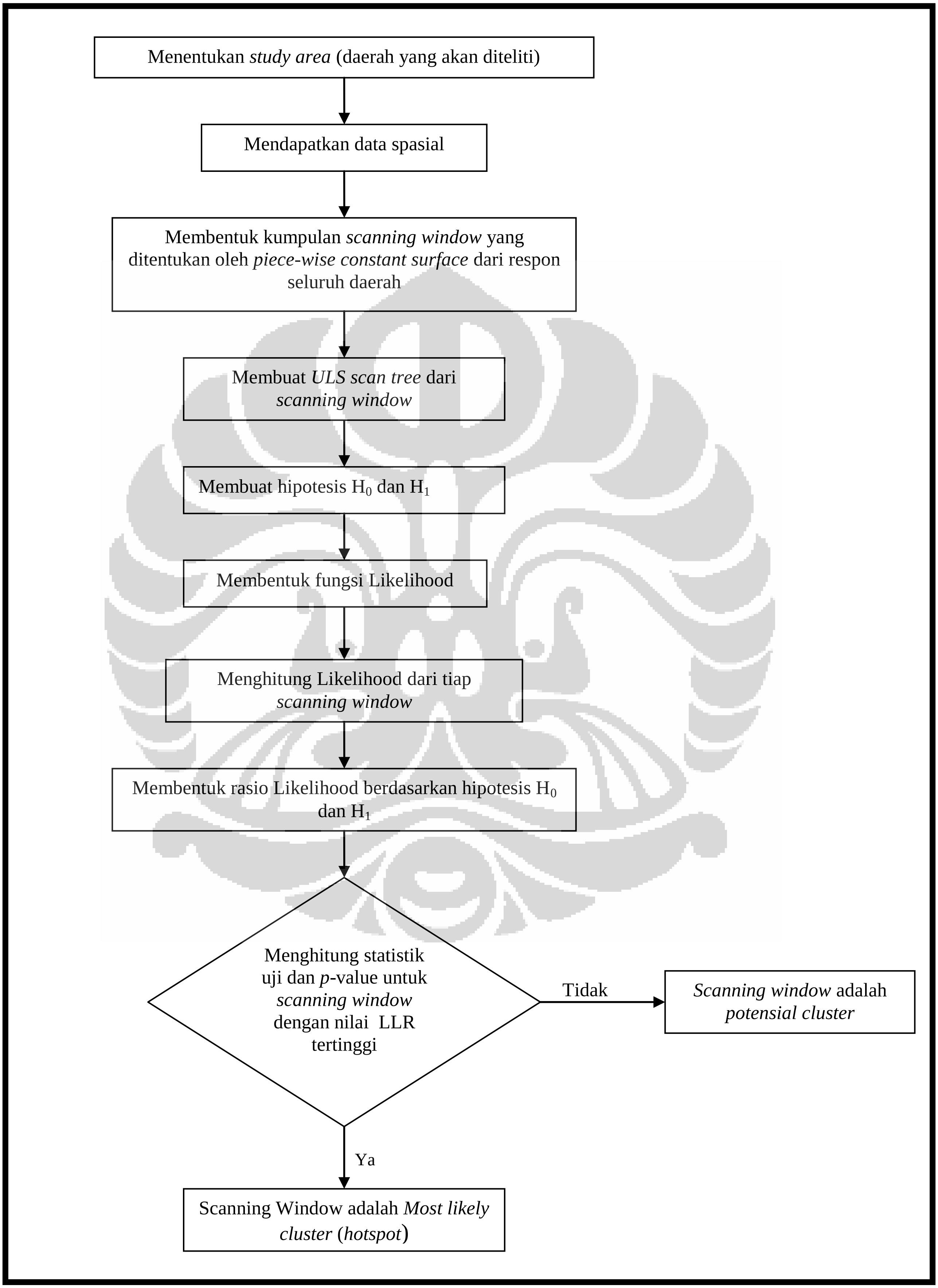
: batas sel, *a*

: zona, Z

3.1 METODE UPPER LEVEL SET SCAN STATISTICS

Upper level set scan statistics adalah suatu metode yang dapat digunakan untuk mendeteksi pengelompokan daerah yang memiliki intensitas yang paling tinggi dari suatu kejadian (*hotspot*) dan mengevaluasi signifikansinya secara statistik.

Langkah - langkah atau cara kerja metode *upper level set scan statistics* dapat dilihat pada diagram alur berikut



Gambar 3.2 Diagram Alur metode *upper level set scan statistics*

3.2 DATA SPASIAL

Data spasial adalah suatu hasil pengukuran yang memuat informasi mengenai lokasi dari pengukuran. Dalam penelitian ini digunakan data spasial yang terdiri dari intensitas pengukuran dan informasi lokasi.

3.3 *SCANNING WINDOW*

Calon-calon kelompok daerah hotspot didapat dengan membentuk kumpulan *scanning window*, yang ditentukan oleh suatu *rate* atau tingkat kejadian dari tiap sel.

Terdapat dua tahap dalam menentukan *scanning window*, yaitu

1. **menghitung rate**

   Dari tiap sel a, hitung rate kejadian, G_a,

$$G_a = \frac{X_a}{A_a}$$

dengan X_a adalah banyaknya kejadian pada sel a yang diasumsikan mengikuti suatu distribusi tertentu dan A_a adalah ukuran populasi pada sel a yang diasumsikan diketahui dan *fixed*.

Rate kejadian, G_a, mendefinisikan suatu *piece-wise constant surface* pada *study area*. Tiap level g menentukan suatu *upper level set*

$$U_g = \{a; G_a \geq g\}$$

Di mana U_g merupakan sel – sel yang memiliki *rate* – *rate* lebih tinggi atau sama dengan dari level g, $g \in G = \{G_a : a \in R\}$.

2. menentukan sel atau zona yang menjadi *scanning window* berdasarkan level g yang telah ditentukan dengan suatu *upper level set*.

Dalam menentukan *scanning window* dengan suatu *upper level set* terdapat dua tahap, yaitu

- pertama, tentukan level g
- kedua, setiap sel a yang memiliki rate $G_a \geq g$, merupakan bagian dari suatu *scanning window*. Jika sel-sel tersebut bertetangga (*adjacent*) maka sel-sel dapat terhubung membentuk zona.

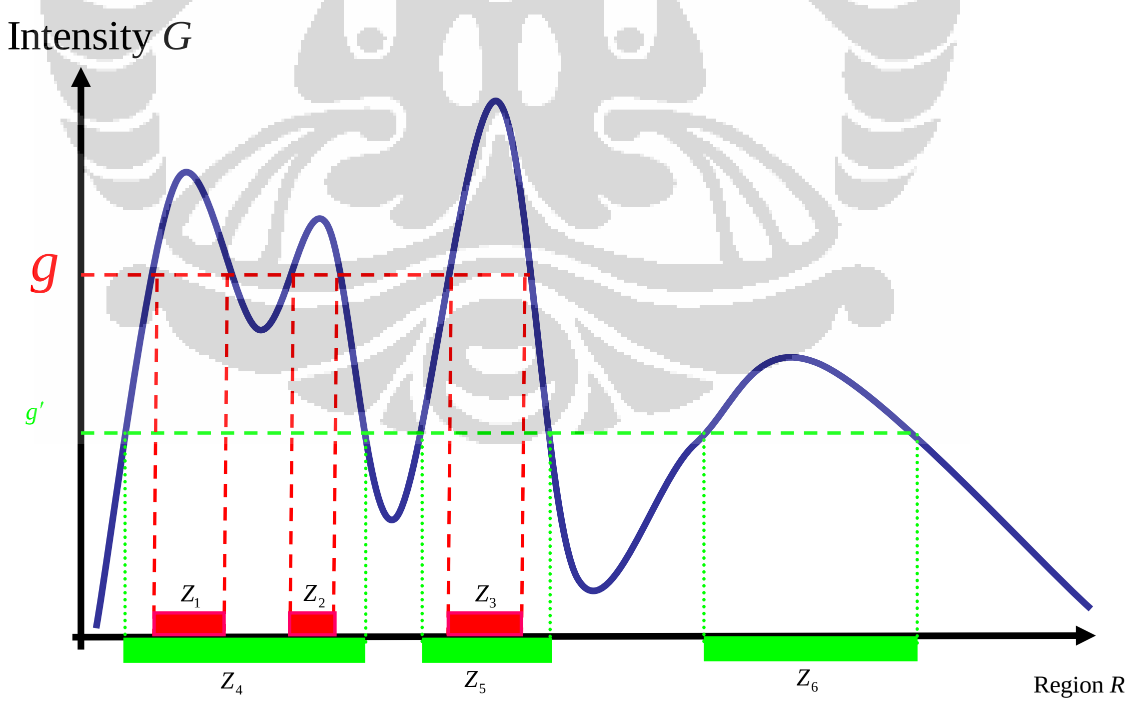
Setelah didapat *scanning window*, dilakukan penghitungan likelihood dari tiap anggota *scanning window* tersebut. Untuk mengetahui apakah kumpulan daerah dalam *scanning window* tersebut signifikan secara statistik, dapat diketahui dengan menghitung nilai signifikansi atau *p-value*.

Penurunan level g

Dalam metode *upper level set scan statistics* kita dapat menurunkan level g menjadi g'. Untuk setiap sel yang memiliki rate $G_a \geq g'$, akan terbentuk suatu *scanning window* baru. Lalu hitung likelihood dari tiap anggota *scanning window* tersebut, kemudian untuk mengetahui apakah kumpulan daerah dalam *scanning window* tersebut signifikan secara statistik, hitung nilai signifikansinya.

Kemudian lakukan langkah penurunan level g kembali. Saat penurunan level g, terdapat tiga kemungkinan, yaitu terjadi penggabungan zona, perluasan zona, dan pembentukan zona baru. Gambar 3.3 merupakan ilustrasi untuk perubahan zona dengan penurunan level g:

Pada suatu level g, yang merupakan scanning window **Z_1** adalah Z_1, Z_2, dan Z_3. Lalu nilai g diturunkan menjadi g', didapat suatu scanning window baru **Z_2** yaitu penggabungan zona Z_1 dan Z_2 menjadi Z_4, perluasan Z_3 menjadi Z_5, dan terbentuknya zona baru Z_6.



Gambar 3.3 Grafik perubahan zona karena penurunan level g

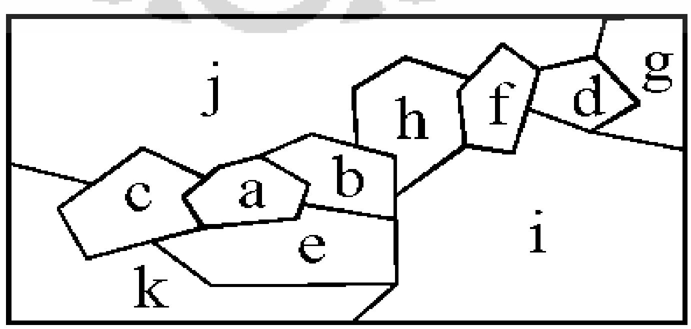
Kumpulan dari *scanning window* juga dapat dilihat sebagat struktur data yang membentuk *tree* (*ULS scan tree*). Seluruh sel dari kumpulan *scanning window* merupakan *node* pada *ULS scan tree*.

Berikut langkah-langkah pembentukan ULS scan *tree*

- Anggap *study area* yang sudah terpartisi sebagai suatu *landform* yang awalnya seluruh permukaan di bawah air , dengan *rate* G_a sebagai tingginya.
- Seiring dengan menurunnya 'level air' (*rate* g), maka akan terlihat beberapa *landform*. Pada tiap 'level air', sel-sel akan muncul atau tidak muncul. Akan muncul pulau (zona atau sel) baru, penggabungan dua atau lebih pulau (zona), atau perluasan pulau (zona).

Ilustrasi dari pembentukan *ULS scan tree* adalah sebagai berikut:

Gambar 3.4a adalah suatu *study area* yang sudah terpartisi menjadi 10 bagian (sel), masing-masing sel dilabelkan dengan a, b, c, …, k.

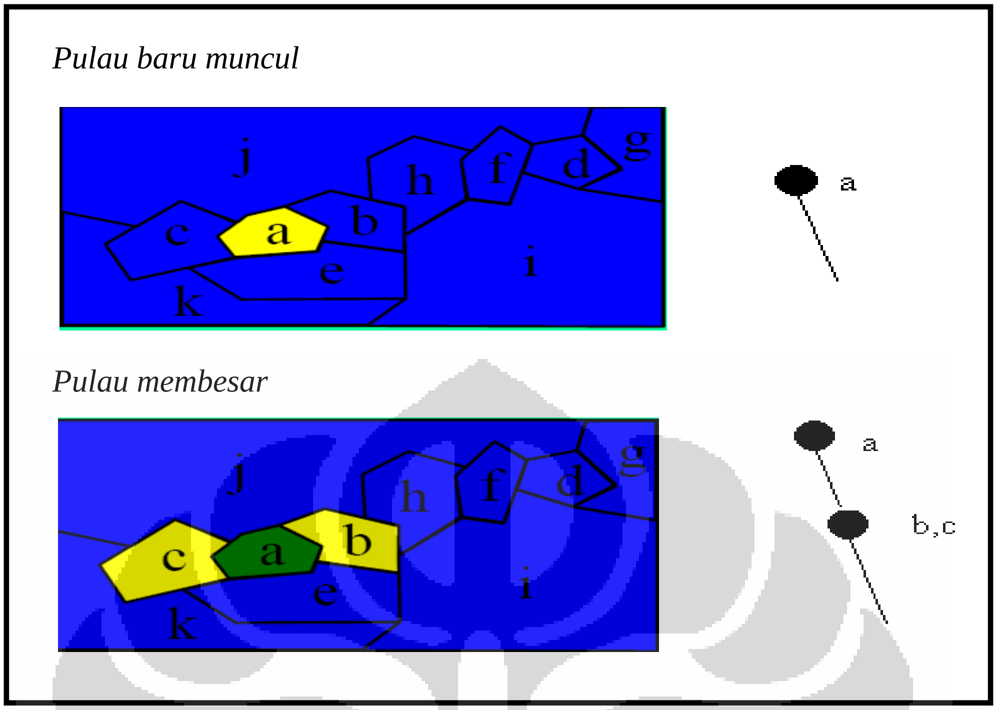


Gambar 3.4a *Study area* yang telah terpartisi

Pada tiap sel akan dihitung *rate*-nya. Selanjutnya anggap *study area* yang sudah terpartisi tersebut sebagai suatu *landform* (permukaan) yang awalnya seluruh permukaan berada di bawah air dengan *rate* sebagai tingginya dan g sebagai 'level air' yang semakin surut (menurun). Pada pembentukan *ULS scan tree* ingin dilihat bagaimana proses konektivitas sel-sel seiring dengan menurunnya 'level air'.

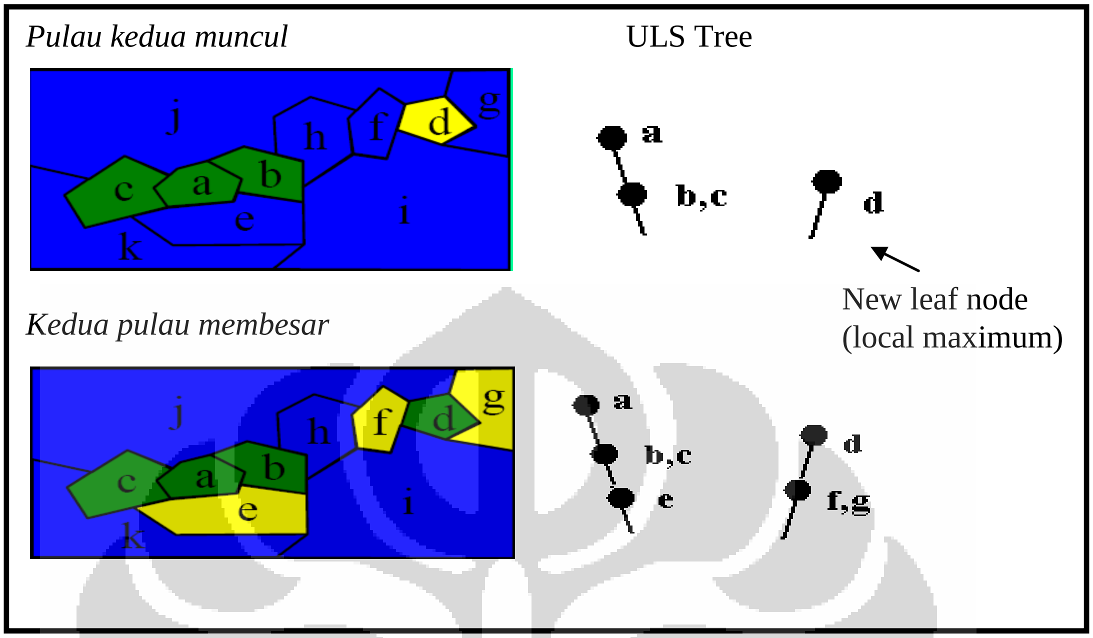
Seiring dengan turunnya 'level air', akan tampak beberapa *landform* yang muncul. Misalnya pada tiap 'level air', sel diberi warna yaitu hijau untuk sel muncul pertama kali pada level g_1, warna kuning untuk sel yang muncul kemudian, dan warna biru untuk sel yang belum muncul.

Gambar 3.4b menunjukkan bahwa sel yang pertama kali muncul adalah sel a yang memiliki *rate* tertinggi. Kemudian disusul sel b dan sel c secara bersamaan, sel b dan c memiliki *rate* yang sama. Untuk *ULS scan tree*, buat satu *node* untuk tiap sel yang baru muncul. Pada ilustrasi ini, buat satu *node* untuk sel a, namakan a, kemudian satu *node* untuk sel b dan sel c, namakan b,c, yang terletak agak di bawah *node* a. Oleh karena sel a, sel b, dan sel c ber-*adjacent* maka *node* a dan *node* b,c dihubungkan oleh suatu *edge*. Pada gambar tersebut sel a, sel b, dan sel c terhubung membentuk zona yang semakin membesar atau seperti pulau yang melebar.



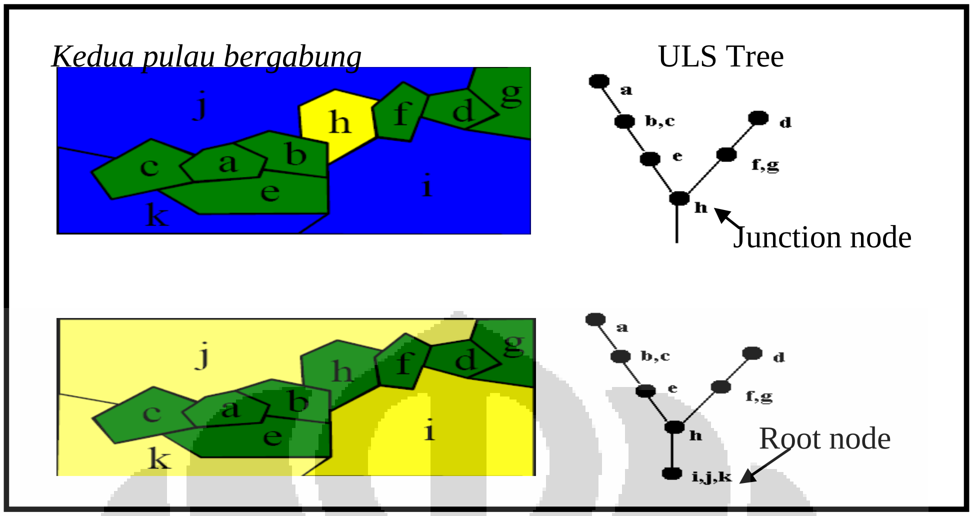
Gambar 3.4b Proses pembentukan *ULS scan tree* (langkah 1 dan 2)

Pada Gambar 3.4c, seiring dengan turunnya ‘level air’, sel yang muncul selanjutnya adalah sel d lalu buat *node* untuk sel d, *node* d. Karena sel d tidak ber-*adjacent* dengan sel a, sel b maupun sel c maka *node* d terletak tidak di bawah *node* a dan *node* b,c melainkan di samping *node* d. Munculnya sel d seolah-olah terbentuknya suatu pulau baru. Kemudian disusul dengan munculnya sel e yang ber-*adjacent* dengan sel a, sel b, dan sel c juga munculnya sel f dan sel g yang ber-*adjacent* dengan sel d.



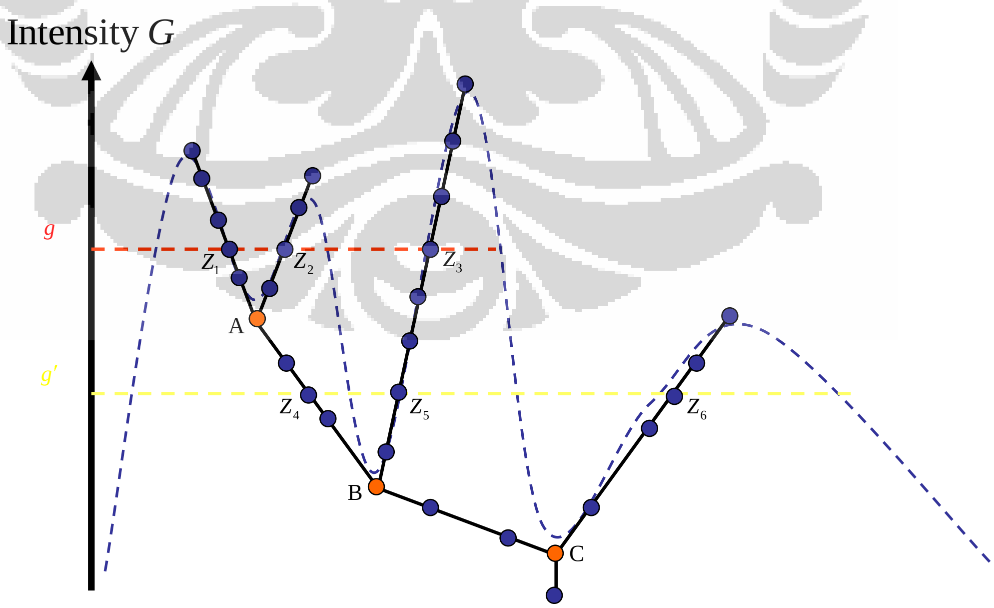
Gambar 3.4c Proses pembentukan *ULS scan tree* (langkah 3 dan 4)

Gambar 3.4d menunjukkan proses pembentukan *ULS scan tree* yang masih berlanjut dengan munculnya sel h. Sel h menghubungkan zona yang mengandung sel a dengan zona yang mengandung sel d sehingga dengan munculnya sel h, kedua zona tersebut bergabung menjadi satu zona atau seperti bergabungnya dua pulau. Kemudian disusul dengan munculnya sel i, sel j, dan sel k.



Gambar 3.4c Proses pembentukan *ULS scan tree* (langkah 5 dan 6)

Pada proses pembentukan *ULS scan tree* di atas, dapat disimpulkan bahwa untuk tiap sel yang baru muncul terdapat tiga kemungkinan, yaitu membentuk suatu pulau baru, terdapat pulau yang semakin membesar, atau bergabung nya dua atau lebih pulau.



Gambar 3.5 *ULS scan tree* dari gambar 3.3

Gambar 3.5 di atas adalah *tree* (*ULS scan tree*) dari gambar 3.3

3.4 RESPON BERDISTRIBUSI GAMMA

Metode *upper level set (ULS) scan statistics* dapat digunakan untuk mendeteksi *hotspot* tidak hanya untuk data dengan variable random diskrit akan tetapi dapat juga untuk data yang kontinu. Dalam *ULS scan statistics*, karena pendeteksian *hotspot* bergantung pada ukuran sel (A_a) sehingga mean dan variansi dari model akan dinyatakan dalam term A_a. Diasumsikan banyak atau besarnya kejadian antar sel (X_a) adalah saling bebas dan ukuran sel (A_a) diketahui dan *fixed*. Distribusi yang digunakan dalam metode *ULS scan statistics* adalah distribusi yang memenuhi asumsi sebagai berikut:

1. Mean proporsional terhadap A_a.
2. Variabilitas menurun seiring dengan besarnya A_a.

Distribusi Gamma merupakan distribusi kontinu yang memenuhi asumsi di atas, yang selanjutnya akan digunakan dalam penelitian ini. Sehingga penelitian ini merupakan pendeteksian hotspot menggunakan metode *ULS scan statistics* dengan model respon Gamma.

Dalam metode *upper level set scan statistics* ini, $X_a \sim \text{Gamma}(k_a, \beta_a)$, k_a dan β_a dapat bervariasi dari sel ke sel. Untuk memenuhi asumsi di atas maka kita gunakan parameter indeks *k* yang proporsional terhadap ukuran A_a dari sel:

$$k_a = \frac{A_a}{c}$$

Dimana *c* merupakan parameter yang tidak diketahui yang nilainya sama untuk semua a.

Sehingga mean:

$$
\begin{aligned}
E[X_a] = \mu[X_a] &= k_a\beta_a \\
&= \frac{A_a}{c}\beta_a \\
&= \frac{\beta_a A_a}{c}
\end{aligned}
$$

Dan koefisien variasinya adalah

$$
\begin{aligned}
\mathrm{CV}^2[X_a] &= \left(\frac{\sigma[X_a]}{\mu[X_a]}\right)^2 \\
&= \left(\frac{\sqrt{k_a\beta_a^2}}{k_a\beta_a}\right)^2 \\
&= \frac{k_a\beta_a^2}{k_a^2\beta_a^2} \\
&= \frac{1}{k_a} \\
&= \frac{1}{A_a/c} \\
&= \frac{c}{A_a}
\end{aligned}
$$

3.5 HIPOTESIS

Metode *upper level set scan statistic*s bertujuan untuk mendeteksi apakah suatu kelompok daerah **Z** tertentu memiliki intensitas (*rate*) lebih tinggi dari daerah lainnya secara signifikan. Hipotesis null H_0 pada penelitian ini adalah tidak terdapat daerah yang memiliki intensitas lebih tinggi dari daerah lainnya, dengan kata lain intensitas sama untuk semua daerah **Z** (tidak terdapat *hotspot*). Hipotesis alternatif H_1 menyatakan bahwa intensitas pada daerah **Z** lebih tinggi dari daerah lainnya secara signifikan.

H_0 : tidak terdapat *hotspot* (β_a sama untuk semua a)

H_1 : terdapat *hotspot* dan nilai parameter $\beta_0, \beta_1 > 0$ sedemikian sehingga

$$\beta_a = \begin{cases} \beta_1 & \text{untuk gabungan } a \text{ di dalam } \mathbf{Z} \\ \beta_0 & \text{untuk gabungan } a \text{ di luar } \mathbf{Z} \end{cases} \quad \text{dan } \beta_1 > \beta_0$$

3.6 FUNGSI LIKELIHOOD

Berikut ini akan dijelaskan mengenai fungsi likelihood untuk model di atas. Fungsi probabilitas f(x) untuk distribusi gamma

$$f(x_a) = \begin{cases} \dfrac{1}{\Gamma(\frac{A_a}{c})\beta_1^{\frac{A_a}{c}}} x_a^{\frac{A_a}{c}-1} e^{-x_a/\beta_1}, & a \in Z \\ \dfrac{1}{\Gamma(\frac{A_a}{c})\beta_0^{\frac{A_a}{c}}} x_a^{\frac{A_a}{c}-1} e^{-x_a/\beta_0}, & a \notin Z \end{cases} \qquad (3.6.1)$$

Fungsi likelihoodnya

$$L(Z,\beta_0,\beta_1,c)=\prod_{a\in Z}\frac{1}{\Gamma\left(\frac{A_a}{c}\right)\beta_1^{\frac{A_a}{c}}}x_a^{\frac{A_a}{c}-1}e^{-x_a/\beta_1}\prod_{a\notin Z}\frac{1}{\Gamma\left(\frac{A_a}{c}\right)\beta_0^{\frac{A_a}{c}}}x_a^{\frac{A_a}{c}-1}e^{-x_a/\beta_0}$$

$$=\frac{1}{\beta_1^{\sum_{a\in Z}\frac{A_a}{c}}}e^{-\frac{\sum_{a\in Z}x_a}{\beta_1}}\prod_{a\in Z}\frac{1}{\Gamma\left(\frac{A_a}{c}\right)}x_a^{\frac{A_a}{c}-1}\ \frac{1}{\beta_0^{\sum_{a\notin Z}\frac{A_a}{c}}}e^{-\frac{\sum_{a\notin Z}x_a}{\beta_0}}\prod_{a\notin Z}\frac{1}{\Gamma\left(\frac{A_a}{c}\right)}x_a^{\frac{A_a}{c}-1}$$

$$=\frac{1}{\beta_1^{\sum_{a\in Z}\frac{A_a}{c}}}e^{-\frac{\sum_{a\in Z}x_a}{\beta_1}}\frac{1}{\beta_0^{\sum_{a\notin Z}\frac{A_a}{c}}}e^{-\frac{\sum_{a\notin Z}x_a}{\beta_0}}\prod_{a\in Z}\frac{1}{\Gamma\left(\frac{A_a}{c}\right)}x_a^{\frac{A_a}{c}-1}\prod_{a\notin Z}\frac{1}{\Gamma\left(\frac{A_a}{c}\right)}x_a^{\frac{A_a}{c}-1}$$

$$=\frac{e^{-\frac{\sum_{a\in Z}x_a}{\beta_1}}e^{-\frac{\sum_{a\notin Z}x_a}{\beta_0}}}{\beta_1^{\sum_{a\in Z}\frac{A_a}{c}}\beta_0^{\sum_{a\notin Z}\frac{A_a}{c}}}\prod_{a\in Z}\frac{1}{\Gamma\left(\frac{A_a}{c}\right)}x_a^{\frac{A_a}{c}-1}\prod_{a\notin Z}\frac{1}{\Gamma\left(\frac{A_a}{c}\right)}x_a^{\frac{A_a}{c}-1}$$

$$=\frac{e^{-\left(\frac{\sum_{a\in Z}x_a}{\beta_1}+\frac{\sum_{a\notin Z}x_a}{\beta_0}\right)}}{\beta_1^{\sum_{a\in Z}\frac{A_a}{c}}\beta_0^{\sum_{a\notin Z}\frac{A_a}{c}}}\prod_{a\in R}\frac{1}{\Gamma\left(\frac{A_a}{c}\right)}x_a^{\frac{A_a}{c}-1} \tag{3.6.2}$$

Definisikan :

- Ruang parameter keseluruhan

  $\Omega=\{(Z,\beta_0,\beta_1,c):0<\beta_0\le 1,0<\beta_1\le 1,c,Z\}$

- Ruang parameter yang terdefinisikan dalam kondisi H_0

  $\omega=\{(\beta_0,\beta_1,c):\beta_0=\beta_1,0<\beta_0\le 1,0<\beta_1\le 1\},\omega\in\Omega$

Model full memiliki empat parameter yang tidak diketahui yaitu Z, *c*, β_0, dan β_1 yang harus ditaksir.

3.6.1 Menaksir parameter di bawah kondisi H_0

Fungsi likelihood di bawah kondisi H_0 ($\beta_0 = \beta_1$) didapat dengan mensubtitusi $\beta_0 = \beta_1$ pada persamaan (3.6.2)

$$L(\omega) = L_0(\beta_1, c) = \frac{e^{-\left(\frac{\sum_{a\in Z} x_a}{\beta_1} + \frac{\sum_{a\notin Z} x_a}{\beta_1}\right)}}{\beta_1^{\sum_{a\in Z}\frac{A_a}{c}} \beta_1^{\sum_{a\notin Z}\frac{A_a}{c}}} \prod_{a\in R} \frac{1}{\Gamma\left(\frac{A_a}{c}\right)} x_a^{\frac{A_a}{c}-1}$$

$$= \frac{e^{-\left(\frac{\sum_{a\in R} x_a}{\beta_1}\right)}}{\beta_1^{\sum_{a\in R}\frac{A_a}{c}}} \prod_{a\in R} \frac{1}{\Gamma\left(\frac{A_a}{c}\right)} x_a^{\frac{A_a}{c}-1} \qquad (3.6.3)$$

Di bawah kondisi H_0 terdapat 2 parameter yang akan ditaksir, yaitu β_1 dan c. Untuk mendapatkan taksiran parameter persamaan (3.6.3), digunakan metode Maksimum Likeihood Estimator (MLE).

Untuk mempermudah perhitungan, gunakan $\ln L_0(\beta_1, c)$, karena fungsi likelihood di bawah kondisi H_0, $L_0(\beta_1, c)$, dan $\ln L_0(\beta_1, c)$ akan memaksimumkan untuk nilai β_1 dan c yang sama.

Fungsi ln likelihood di bawah kondisi H_0 adalah

$$\ln L_0(\beta_1, c) = \sum_{a\in R} \ln \frac{1}{\Gamma\left(\frac{A_a}{c}\right)} x_a^{\frac{A_a}{c}-1} - \frac{\sum_{a\in R} x_a}{\beta_1} - \ln \beta_1^{\sum_{a\in R}\frac{A_a}{c}} \tag{3.6.4}$$

1. Menaksir parameter β_1

   Nilai $\ln L_0(\beta_1, c)$ yang maksimum diperoleh dari nilai β_1 yang

merupakan solusi dari $\frac{\partial \ln L_0(\beta_1, c)}{\partial \beta_1} = 0$, sebagai berikut

$$\frac{\partial \ln L_0(\beta_1, c)}{\partial \beta_1} = \frac{\sum_{a\in G} x_a}{\beta_1^2} - \frac{\sum_{a\in G} \frac{A_a}{c}}{\beta_1} = 0 \tag{3.6.5}$$

$$\frac{\sum_{a\in R} x_a}{\beta_1^2} - \frac{\sum_{a\in R} \frac{A_a}{c}}{\beta_1} = 0$$

$$\Leftrightarrow \frac{\sum_{a\in R} x_a}{\beta_1^2} = \frac{\sum_{a\in R} \frac{A_a}{c}}{\beta_1}$$

$$\Leftrightarrow \frac{\sum_{a\in R} x_a}{\beta_1} = \sum_{a\in R} \frac{A_a}{c}$$

$$\Leftrightarrow \hat{\beta}_1 = \frac{\sum_{a\in R} x_a}{\sum_{a\in R} \frac{A_a}{c}}$$

$$\Leftrightarrow \hat{\beta}_1 = \frac{\sum_{a\in R} x_a}{\frac{1}{c}\sum_{a\in R} A_a}$$

$$\Leftrightarrow \hat{\beta}_1 = c\,\frac{\sum_{a\in R} x_a}{\sum_{a\in R} A_a}$$

Jadi, taksiran β_1 yang memaksimumkan $L_0(\beta_1, c)$ dalam term c adalah $\hat{\beta}_1 = c\dfrac{\sum_{a\in R} x_a}{\sum_{a\in R} A_a}$

2. Menaksir parameter c

Nilai $\ln L_0(\beta_1, c)$ yang maksimum diperoleh dari nilai c yang merupakan solusi dari $\dfrac{\partial \ln L_0(\beta_1, c)}{\partial c} = 0$, sebagai berikut

$$\frac{\partial \ln L_0(\beta_1, c)}{\partial\, c} = \frac{1}{c^2}\sum_{a\in R} A_a \ln \beta_1 + \sum_{a\in R}\left(\frac{-A_a \ln x_a}{c^2} + \frac{A_a \psi\left(\frac{A_a}{c}\right)}{c^2}\right) = 0 \qquad (3.6.6)$$

Persamaan (3.6.6) akan menjadi

$$\sum_{a\in R} A_a\left[\ln\left(\frac{A_a}{c}\right) - \psi\left(\frac{A_a}{c}\right)\right] = \left(\sum_{a\in R} A_a\right)\ln\frac{\sum_{a\in R} x_a}{\sum_{a\in R} A_a} - \sum_a A_a \ln\frac{x_a}{A_a} \qquad (3.6.7)$$

dimana $\psi(.)$ adalah fungsi digamma.

Taksiran parameter c pada persamaan (3.6.7) tidak dapat secara langsung dicari. Digunakan metode iterasi Newton-Rapshon untuk mencari solusi dari c pada persamaan tersebut.

3.6.2 Menaksir parameter di bawah kondisi H_1

Fungsi likelihood dalam ruang parameter keseluruhan Ω (di bawah kondisi H_1) dapat ditulis sebagai berilkut:

$$L(\Omega) = L(Z, \beta_0, \beta_1, c) = \begin{cases} \dfrac{e^{-\left(\frac{\sum_{a \in Z} x_a}{\beta_1} + \frac{\sum_{a \notin Z} x_a}{\beta_0}\right)}}{\beta_1^{\sum_{a \in Z} \frac{A_a}{c}} \beta_0^{\sum_{a \notin Z} \frac{A_a}{c}}} \prod_{a \in R} \dfrac{1}{\Gamma\left(\frac{A_a}{c}\right)} x_a^{\frac{A_a}{c} - 1}, \; jika \beta_1 > \beta_0 \\ \dfrac{e^{-\left(\frac{\sum_{a \in R} x_a}{\beta_1}\right)}}{\beta_1^{\sum_{a \in R} \frac{A_a}{c}}} \prod_{a \in R} \dfrac{1}{\Gamma\left(\frac{A_a}{c}\right)} x_a^{\frac{A_a}{c} - 1}, lainnya \end{cases}$$

(3.6.8)

Untuk mendapatkan taksiran parameter persamaan (3.6.8) yang memaksimumkan $L(Z, \beta_0, \beta_1, c)$, digunakan metode Maksimum Likelihod Estimator (MLE).

Dalam persamaan (3.6.7) terdapat dua fungsi, yaitu

1. $$f_1 = \frac{e^{-\left(\frac{\sum_{a \in Z} x_a}{\beta_1} + \frac{\sum_{a \notin Z} x_a}{\beta_0}\right)}}{\beta_1^{\sum_{a \in Z} \frac{A_a}{c}} \beta_0^{\sum_{a \notin Z} \frac{A_a}{c}}} \prod_{a \in R} \frac{1}{\Gamma\left(\frac{A_a}{c}\right)} x_a^{\frac{A_a}{c} - 1} \qquad (3.6.9)$$

2. $$f_2 = -\frac{e^{-\left(\frac{\sum_{a \in R} x_a}{\beta_1}\right)}}{\beta_1^{\sum_{a \in R} \frac{A_a}{c}}} \prod_{a \in R} \frac{1}{\Gamma\left(\frac{A_a}{c}\right)} x_a^{\frac{A_a}{c} - 1} \qquad (3.6.10)$$

Karena persamaan (3.6.8) terdiri dari dua fungsi, sehingga untuk mencari taksiran parameternya dilakukan dengan mencari taksiran parameter untuk persamaan (3.6.9) dan (3.6.10).

Persamaan f_2 pada (3.6.10) sama dengan persamaan (3.6.3), sehingga taksiran parameternya adalah $\hat{\beta}_1 = c\frac{\sum_{a \in R} x_a}{\sum_{a \in R} A_a}$ dan $\hat{c}$ solusi pada persamaan (3.6.7) yang dicari dengan metode iterasi Newton-Rapshon.

Sekarang akan dicari taksiran parameter β_0, β_1, dan *c* dan memaksimumkan persamaan (3.6.9). untuk mempermudah perhitungan, gunakan $\ln f_1$.

$$\ln f_1 = -\frac{\sum_{a \in Z} x_a}{\beta_1} - \frac{\sum_{a \notin Z} x_a}{\beta_0} - \sum_{a \in Z} \frac{A_a}{c} \ln \beta_1 - \sum_{a \notin Z} \frac{A_a}{c} \ln \beta_0 + \sum_{a \in R} \ln \frac{1}{\Gamma\left(\frac{A_a}{c}\right)} x_a^{\frac{A_a}{c}-1} \tag{3.6.11}$$

1. Menaksir parameter β_0

Nilai $\ln f_1$ yang maksimum diperoleh dari nilai β_0 yang merupakan solusi dari $\frac{\partial \ln f_1}{\partial \beta_0} = 0$, sebagai berikut

$$\frac{\partial \ln f_1}{\partial \beta_0} = \frac{\sum_{a \notin Z} x_a}{\beta_0^2} - \frac{\sum_{a \notin Z} \frac{A_a}{c}}{\beta_0} = 0 \tag{3.6.12}$$

$$\frac{\sum_{a \notin Z} x_a}{\beta_0^2} - \frac{\sum_{a \notin Z} \frac{A_a}{c}}{\beta_0} = 0$$

$$\Leftrightarrow \frac{\sum_{a \notin Z} x_a}{\beta_0^2} = \frac{\sum_{a \notin Z} \frac{A_a}{c}}{\beta_0}$$

$$\Leftrightarrow \frac{\sum_{a \notin Z} x_a}{\beta_0} = \sum_{a \notin Z} \frac{A_a}{c}$$

$$\Leftrightarrow \hat{\beta}_0 = \frac{\sum_{a \notin Z} x_a}{\sum_{a \notin Z} \frac{A_a}{c}}$$

$$\Leftrightarrow \hat{\beta}_0 = \frac{\sum_{a \notin Z} x_a}{\frac{1}{c} \sum_{a \notin Z} A_a}$$

$$\Leftrightarrow \hat{\beta}_0 = c \frac{\sum_{a \notin Z} x_a}{\sum_{a \notin Z} A_a}$$

Jadi, taksiran β_0 yang memaksimumkan $L(Z, \beta_0, \beta_1, c)$ dalam term c adalah $\hat{\beta}_0 = c \frac{\sum_{a \notin Z} x_a}{\sum_{a \notin Z} A_a}$

2. Menaksir parameter β_1

Nilai $\ln f_1$ yang maksimum diperoleh dari nilai β_1 yang merupakan solusi dari $\frac{\partial \ln f_1}{\partial \beta_1} = 0$, sebagai berikut

$$\frac{\partial \ln f_1}{\partial \beta_1} = \frac{\sum_{a \in Z} x_a}{\beta_1^2} - \frac{\sum_{a \in Z} \frac{A_a}{c}}{\beta_1} = 0 \tag{3.6.13}$$

$$\frac{\sum_{a \in Z} x_a}{\beta_1^2} - \frac{\sum_{a \in Z} \frac{A_a}{c}}{\beta_1} = 0$$

$$\Leftrightarrow \frac{\sum_{a \in Z} x_a}{\beta_1^2} = \frac{\sum_{a \in Z} \frac{A_a}{c}}{\beta_1}$$

$$\Leftrightarrow \frac{\sum_{a \in Z} x_a}{\beta_1} = \sum_{a \in Z} \frac{A_a}{c}$$

$$\Leftrightarrow \hat{\beta}_1 = \frac{\sum_{a \in Z} x_a}{\sum_{a \in Z} \frac{A_a}{c}}$$

$$\Leftrightarrow \hat{\beta}_1 = \frac{\sum_{a \in Z} x_a}{\frac{1}{c} \sum_{a \in Z} A_a}$$

$$\Leftrightarrow \hat{\beta}_1 = c \frac{\sum_{a \in Z} x_a}{\sum_{a \in Z} A_a}$$

Jadi, taksiran β_1 yang memaksimumkan $L(Z, \beta_0, \beta_1, c)$ dalam term c adalah $\hat{\beta}_1 = c \frac{\sum_{a \in Z} x_a}{\sum_{a \in Z} A_a}$

3. Menaksir parameter c

   Nilai $\ln f_1$ yang maksimum diperoleh dari nilai c yang merupakan solusi dari $\frac{\partial \ln f_1}{\partial c} = 0$, sebagai berikut

$$\frac{\partial \ln f_1}{\partial c} = \frac{1}{c^2}\left(\sum_{a\in Z} A_a \ln \beta_1 + \sum_{a\notin Z} A_a \ln \beta_0\right) + \sum_{a\in R}\left(\frac{-A_a \ln x_a}{c^2} - \frac{-A_a \psi\left(\frac{A_a}{c}\right)}{c^2}\right) = 0$$

(3.6.14)

Persamaan (3.6.13) akan menjadi

$$\sum_a A_a\left[\ln\left(\frac{A_a}{c}\right) - \psi\left(\frac{A_a}{c}\right)\right] = \left(\sum_{a\notin Z} A_a\right)\ln\frac{\sum_{a\notin Z} x_a}{\sum_{a\notin Z} A_a} + \left(\sum_{a\in Z} A_a\right)\ln\frac{\sum_{a\in Z} x_a}{\sum_{a\in Z} A_a} - \sum_a A_a \ln\frac{x_a}{A_a}$$

(3.6.15)

dimana $\psi(.)$ adalah fungsi digamma.

Taksiran parameter c pada persamaan (3.6.15) tidak dapat secara langsung dicari. Digunakan metode iterasi Newton-Rapshon untuk mencari solusi dari c pada persamaan tersebut.

3.7 STATISTIK UJI

Rasio likelihood digunakan untuk menguji hipotesis H_0 terhadap hipotesis alternative H_1. yaitu rasio dari likelihood dalam hipotesis null terhadap likelihood dalam ruang parameter keseluruhan. Berikut ini merupakan rasio likelihoodnya

$$LR = \nabla = \frac{L(\hat{\omega})}{L(\hat{\Omega})}$$

Kemudian rasio likelihood tersebut akan digunakan dalam pengujian. Statistik uji yang digunakan dalam penelitian ini adalah statistik uji Log Rasio Likelihood (LLR), dengan

$$\text{LLR} = -2 \text{ Log } \frac{L(\hat{\omega})}{L(\hat{\Omega})}$$

Penjabaran $\frac{L(\hat{\omega})}{L(\hat{\Omega})}$ terdapat pada lampiran 4. Karena bentuk rumus yang demikian, LLR tidak selalu dapat didekati oleh distribusi Chi kuadrat oleh karena itu untuk mendapatkan nilai *p*-value digunakan metode simulasi Monte Carlo. Monte Carlo adalah suatu metode untuk mendapatkan solusi parameter suatu hipotesis untuk populasi. Metode ini menggunakan barisan random dari bilangan - bilangan untuk membentuk suatu sampel dari populasi, dimana penaksiran parameter dapat diperoleh (F. James, 1980).

3.8 MENGHITUNG *p-VALUE*

Setelah didapatkan *scanning window* atau calon *hotspot* dan dihitung log rasio likelihood (LLR) dari tiap anggota *scanning window*, akan diperiksa apakah *scanning window* tersebut signifikan secara statistik, atau dengan kata lain, *scanning window* tersebut merupakan *hotspot* atau tidak. *p-value* didapat dengan melakukan simulasi pendekatan Monte Carlo.

Langkah – langkah yang dilakukan untuk mendapatkan *p-value* dengan pendekatan Monte Carlo adalah sebagai berikut:

1. Hitung nilai LLR tertinggi dari data, misalkan dinotasikan dengan t_0, yang dihasilkan dari *scanning window* pada *ULS scan tree*.
2. Buat *replicate* data acak yang dibangun di bawah kondisi H_0.
3. Tentukan *scanning window* dan *ULS scan tree* dengan menggunakan data acak tersebut.
4. Dapatkan nilai LLR yang tertinggi dari data acak tersebut (T(x)).
5. Ulangi langkah 2 sampai 4 untuk mendapatkan *m* pengulangan
6. Urutkan *m* + 1 nilai LLR dari yang tertinggi dari nilai yang terkecil sampai yang terbesar.
7. Hitung *p-value*, $p = \frac{\text{banyaknya}(T(x)) \geq t_0)}{m+1}$