

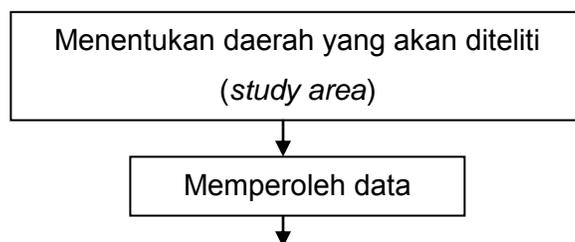
BAB III

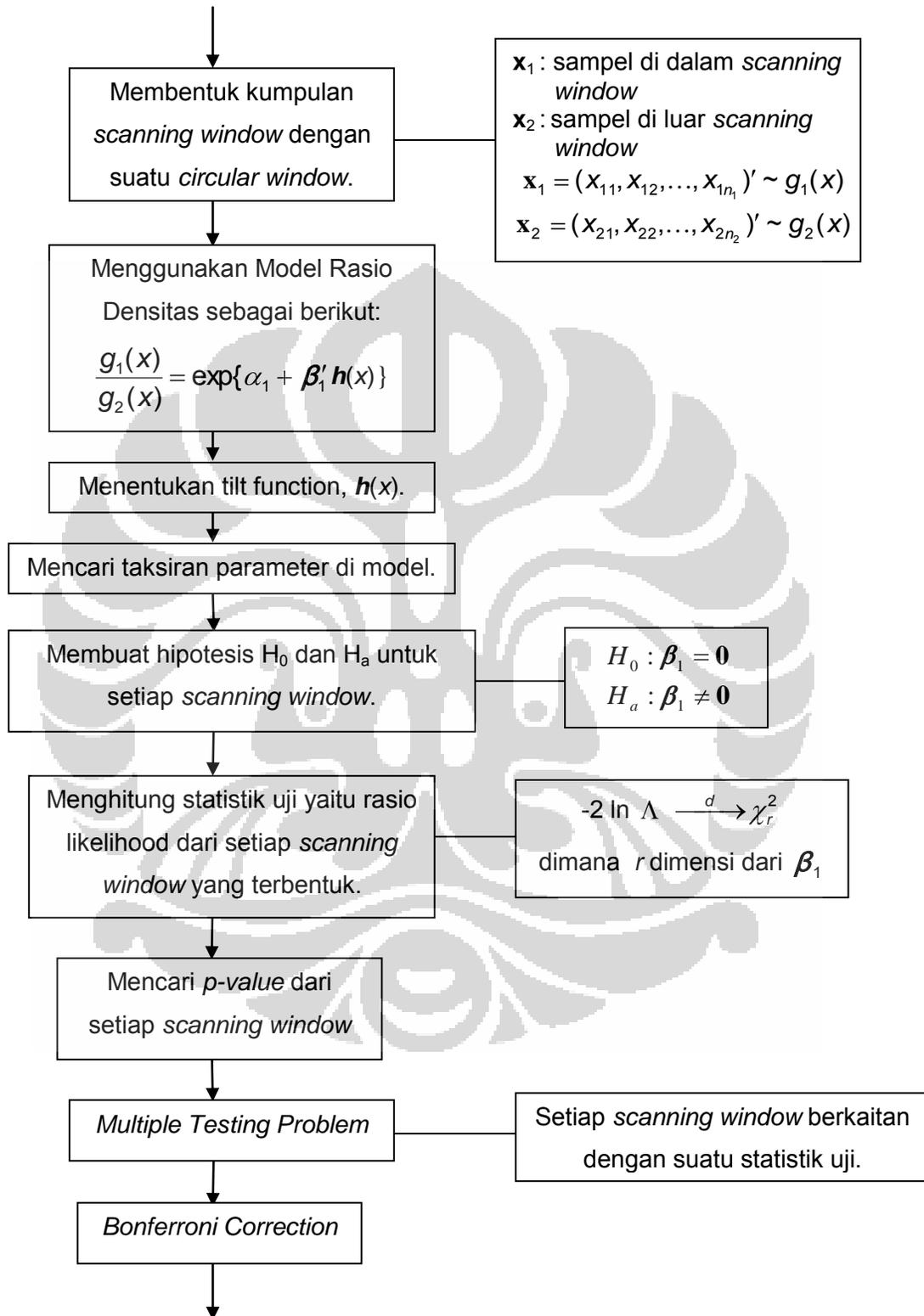
PENDETEKSIAN CLUSTER DENGAN METODE SEMI-PARAMETRIK

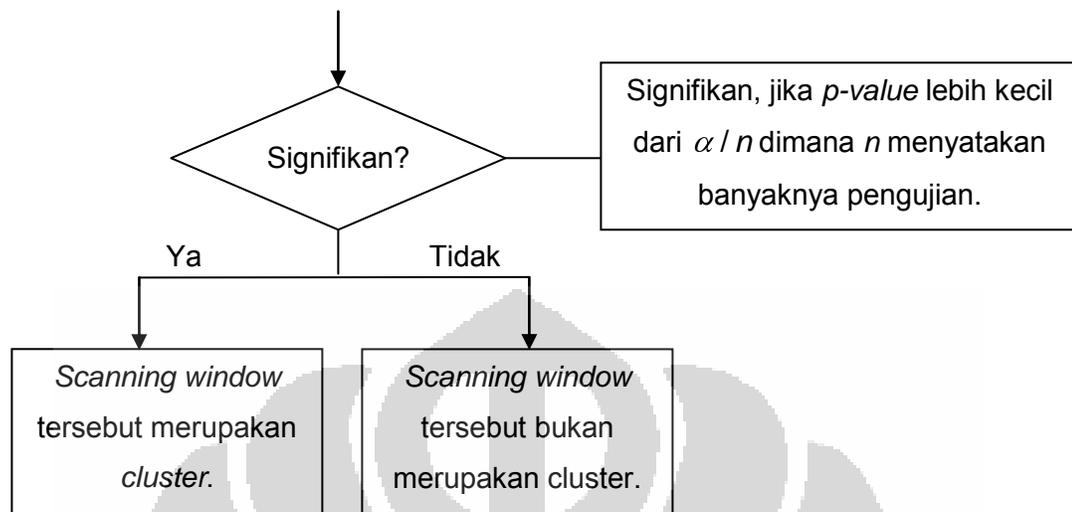
Metode pendeteksian cluster semi-parametrik merupakan suatu metode untuk mendeteksi cluster jika *underlying distribution* dari populasi dimana data diambil tidak diketahui. Metode semi-parametrik menggunakan model rasio densitas semi-parametrik dan prosedur Kulldorf untuk mendeteksi potensial cluster dari suatu kejadian.

Bab ini akan membahas metode pendeteksian cluster semi-parametrik yang meliputi proses dan cara kerja dari metode semi-parametrik untuk mendeteksi cluster. Oleh karena itu, bab ini akan membahas hal-hal yang berkaitan dengan proses dan cara kerja dari metode semi-parametrik untuk mendeteksi cluster, yaitu data, pembentukan *scanning window*, model rasio densitas, pemilihan *tilt function*, penaksiran parameter, hipotesis, rasio likelihood, uji signifikansi, aturan keputusan, dan kesimpulan.

Berikut ini merupakan bagan cara kerja metode semi-parametrik untuk mendeteksi cluster:







3.1 DATA

Data yang digunakan pada metode semi-parametrik secara umum terdiri dari data mengenai kejadian dan data mengenai lokasi kejadian. Data mengenai kejadian misalnya data mengenai banyaknya individu yang terjangkiti demam berdarah. Data mengenai lokasi kejadian yaitu data mengenai lokasi geografi dari setiap kejadian.

3.2 PEMBENTUKAN SCANNING WINDOW

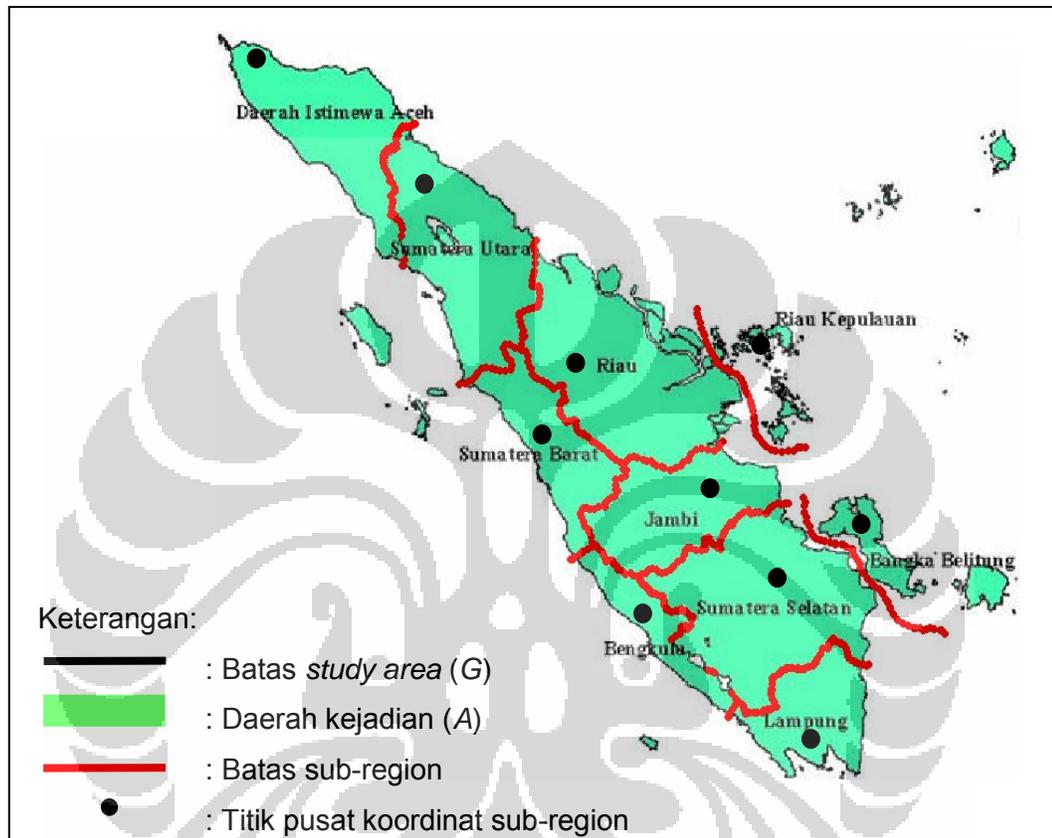
Scanning window merupakan suatu daerah yang potensial untuk menjadi cluster. *Scanning window* yang digunakan dalam tugas akhir ini dihasilkan dari suatu *circular window*. Pembentukan *scanning window* pada

metode semi-parametrik sama dengan pembentukan *scanning window* pada metode Kulldorf.

Berikut ini adalah istilah-istilah yang digunakan dalam pembentukan *scanning window*.

1. *Study area* adalah keseluruhan daerah yang akan diteliti, dinotasikan dengan G .
2. *Sub-region* adalah *study area* yang dipartisi ke dalam beberapa wilayah tertentu, dinotasikan dengan SR_i , $i = 1, 2, \dots, k$ dengan k menyatakan banyaknya *sub-region* dalam *study area*. Contoh: negara bagian, propinsi, kabupaten atau kotamadya, atau wilayah berdasarkan kode pos.
3. Setiap *sub-region* diwakili oleh satu titik pusat koordinat dari *sub-region* tersebut, dinotasikan dengan s_i .
4. Daerah kejadian adalah suatu daerah dalam *study area* yang di dalam daerah tersebut terdapat minimal satu kejadian yang akan diteliti, dinotasikan dengan A , $A \subseteq G$.
5. Titik kejadian adalah titik pusat koordinat dari *sub-region*, yang di dalam *sub-region* tersebut terdapat minimal satu kejadian.
6. *Scanning window* adalah suatu daerah yang potensial untuk menjadi cluster, dinotasikan dengan Z dan $Z \subseteq G$. *Scanning window* yang berbentuk lingkaran disebut *circular window*.

Berikut ini adalah gambar untuk membantu penjelasan dari istilah-istilah di atas:



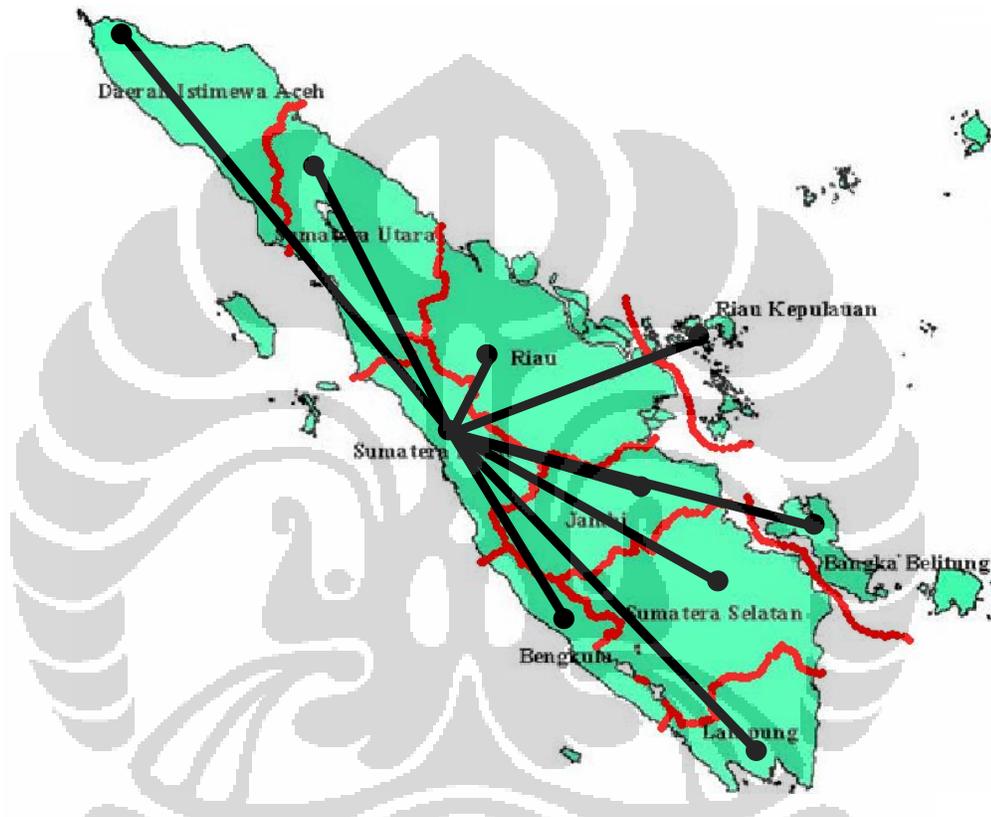
Gambar 3.1 *Study area*

Berikut ini akan dijelaskan algoritma dalam pembentukan *scanning window* yang dihasilkan dari suatu *circular window*.

1. Menentukan jarak.
 - a. Memilih satu *sub-region* secara sembarang yang diwakili oleh titik pusat koordinat *sub-region* tersebut. Hitung jarak *sub-region* yang dipilih ke *sub-region* lain. Jarak yang dihitung adalah jarak Euclid

antara titik pusat koordinat *sub-region* yang dipilih dengan *sub-region* lainnya. Urutkan jarak tersebut dari jarak terdekat sampai jarak terjauh.

Contoh:

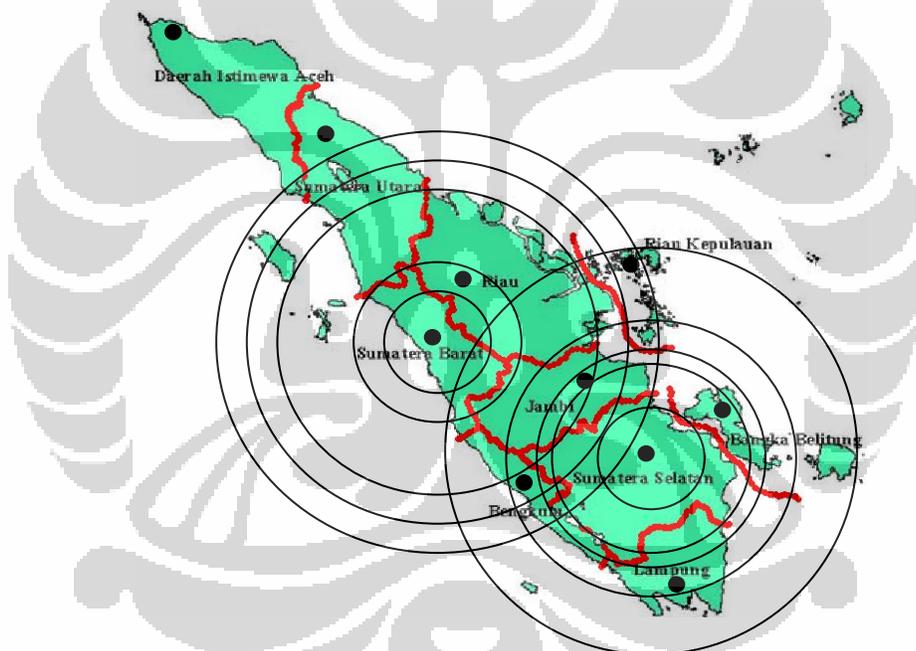


Gambar 3. 2 Menghitung jarak suatu *sub-region* ke *sub-region* lainnya

- b. Ulangi langkah (a) untuk setiap *sub-region*.
2. Membentuk *circular window*.
 - a. Memilih satu *sub-region* secara sembarang yang diwakili oleh titik pusat koordinat *sub-region* tersebut. Buat suatu lingkaran yang berpusat di titik pusat koordinat *sub-region* tersebut dan secara kontinu jari-jari lingkaran tersebut semakin diperbesar. Tidak ada aturan tertentu untuk membatasi ukuran *scanning window*, tetapi

dalam hal ini ukuran *scanning window* dibatasi tidak melebihi 50% dari jumlah populasi dalam *study area* ataupun tidak melebihi 50% dari luas *study area*. Untuk setiap titik pusat koordinat yang masuk ke dalam lingkaran, catat sampel data kejadian, x_1 yang menyatakan sampel di dalam *scanning window* dan x_2 yang menyatakan sampel di luar *scanning window*.

Contoh:



Gambar 3. 3 Pembentukan *scanning window*

- b. Ulangi langkah (a) untuk setiap titik pusat koordinat *sub-region*.
- c. Untuk setiap *scanning window*, hitung nilai rasio likelihood setelah memilih *tilt function*. Pembahasan lebih lanjut mengenai pemilihan *tilt function* dapat dilihat pada subbab 3.4 dan mengenai rasio likelihood dapat dilihat pada subbab 3.7.

3.3 MODEL RASIO DENSITAS

Metode semi-parametrik yang digunakan dalam tugas akhir ini didasarkan pada model rasio densitas. Model rasio densitas digunakan untuk menguji kesamaan distribusi jika diberikan dua atau lebih populasi. *Scanning window* yang telah dibahas pada subbab sebelumnya memisahkan *study area* menjadi dua bagian, yaitu daerah di dalam dan di luar *scanning window*. Oleh karena itu, model rasio densitas yang digunakan pada tugas akhir ini bertujuan untuk menguji kesamaan distribusi jika diberikan dua populasi, yaitu populasi di dalam dan di luar *scanning window*. Pengujian kesamaan distribusi ini berguna untuk menentukan apakah *scanning window* tersebut merupakan suatu cluster atau tidak.

Setelah pembentukan *scanning window* diperoleh sampel sebagai berikut:

$$\mathbf{x}_1 = (x_{11}, x_{12}, \dots, x_{1n_1})' \sim g_1(x)$$

$$\mathbf{x}_2 = (x_{21}, x_{22}, \dots, x_{2n_2})' \sim g_2(x)$$

dimana

\mathbf{x}_1 : sampel di dalam *scanning window*

n_1 : ukuran sampel di dalam *scanning window*

\mathbf{x}_2 : sampel di luar *scanning window*

n_2 : ukuran sampel di luar *scanning window*

$g_1(x)$: p.d.f dari \mathbf{x}_1

$g_2(x)$: p.d.f dari \mathbf{x}_2

Dengan memilih sampel di luar *scanning window* sebagai sampel acuan dan $g_2(x)$ sebagai densitas acuan, maka model rasio densitas mempunyai bentuk sebagai berikut:

$$\frac{g_1(x)}{g_2(x)} = \exp\{\alpha_1 + \beta_1' h(x)\} \quad (3.3.1)$$

dimana $h(x)$ adalah fungsi dari x yang dapat berupa skalar, seperti x , x^2 , atau $\log x$, atau berupa vektor, seperti $(x, x^2)'$ atau $(x, \log x)'$, dan lainnya. Untuk selanjutnya $h(x)$ disebut sebagai *tilt function* yang akan dipilih berdasarkan data yang diperoleh. α_1 berupa skalar, tetapi β_1 dapat berupa skalar atau vektor tergantung pada $h(x)$.

3.4 PEMILIHAN TILT FUNCTION

Dikarenakan tidak diketahui informasi apapun mengenai *underlying distribution*, pemilihan *tilt function* didasarkan pada plot histogram dari data yang diperoleh. Berdasarkan pada plot dan *tilt function* yang diperoleh dari beberapa distribusi yang dijabarkan pada Lampiran 3, diperoleh pola pemilihan *tilt function* sebagai berikut:

1. Jika plot data cenderung simetri atau cenderung *skewed* tetapi menuju simetri, *tilt function* yang dipilih yaitu $h(x) = (x, x^2)$.
2. Jika plot data cenderung *skewed* tetapi tidak menuju simetri, *tilt function* yang dipilih yaitu $h(x) = (x, \ln x)$, $h(x) = (x, 1/x)$.

3. Jika plot data asimptotik, *tilt function* yang dipilih yaitu $h(x) = (x, \ln x)$.

Selanjutnya, setelah memilih *tilt function* dapat diperoleh taksiran dari β_1 . Jika nilai β_1 yang bersesuaian dengan suku dari $h(x)$ tertentu relatif kecil, maka suku tersebut dapat dihilangkan.

3.5 HIPOTESIS

Metode semi-parametrik yang digunakan dalam tugas akhir ini bertujuan untuk mendeteksi ada atau tidak adanya cluster dari suatu daerah yang diteliti.

Dari persamaan (3.3.1), jika $\beta_1 = \mathbf{0}$ maka $\alpha_1 = 0$, sehingga hipotesis yang digunakan untuk mendeteksi ada atau tidaknya cluster sebagai berikut:

Hipotesis nol:

$$H_0 : \beta_1 = \mathbf{0}$$

Hipotesis nol tersebut menyatakan bahwa densitas di dalam *scanning window* sama dengan densitas di luar *scanning window*. Hal ini berarti *scanning window* tersebut bukan merupakan cluster.

Hipotesis alternatif:

$$H_0 : \beta_1 \neq 0$$

Hipotesis alternatif tersebut menyatakan bahwa densitas di dalam *scanning window* tidak sama dengan densitas di luar *scanning window*. Hal ini berarti *scanning window* tersebut merupakan cluster.

PENAKSIRAN PARAMETER PADA MODEL

Setelah pembentukkan *scanning window* diperoleh dua sampel yaitu $\mathbf{x}_1 = (x_{11}, x_{12}, \dots, x_{1n_1})$ yang saling bebas dengan densitas $g_1(x) = \exp\{\alpha_1 + \beta_1' h(x)\} g_2(x)$ dan $\mathbf{x}_2 = (x_{21}, x_{22}, \dots, x_{2n_2})$ yang saling bebas dengan densitas $g_2(x)$.

Misalkan $\mathbf{t} = (t_1, t_2, \dots, t_n)' = (\mathbf{x}'_1, \mathbf{x}'_2)'$ menyatakan data gabungan dari kedua sampel. Fungsi likelihood yaitu

$$L(\alpha_1, \beta_1, g_2) = g_1(x_{11}) \cdot g_1(x_{12}) \cdot \dots \cdot g_1(x_{1n_1}) \cdot g_2(x_{21}) \cdot g_2(x_{22}) \cdot \dots \cdot g_2(x_{2n_2})$$

$$L(\alpha_1, \beta_1, g_2) = [\exp\{\alpha_1 + \beta_1' h(x_{11})\} \cdot g_2(x_{11})] \cdot [\exp\{\alpha_1 + \beta_1' h(x_{12})\} \cdot g_2(x_{12})] \cdot \dots \cdot [\exp\{\alpha_1 + \beta_1' h(x_{1n_1})\} \cdot g_2(x_{1n_1})] \cdot g_2(x_{21}) \cdot g_2(x_{22}) \cdot \dots \cdot g_2(x_{2n_2})$$

$$L(\alpha_1, \beta_1, g_2) = g_2(x_{11}) \cdot g_2(x_{12}) \cdot g_2(x_{13}) \cdot \dots \cdot g_2(x_{1n_1}) \cdot g_2(x_{21}) \cdot g_2(x_{22}) \cdot \dots \cdot g_2(x_{2n_2}) \cdot \exp\{\alpha_1 + \beta_1' h(x_{11})\} \cdot$$

$$\exp\{\alpha_1 + \beta_1' h(x_{12})\} \cdot \exp\{\alpha_1 + \beta_1' h(x_{13})\} \cdot \dots \cdot$$

$$\exp\{\alpha_1 + \beta_1' h(x_{1n_1})\}$$

$$L(\alpha_1, \beta_1, g_2) = g_2(x_{11}) \cdot g_2(x_{12}) \cdot \dots \cdot g_2(x_{1n_1}) \cdot g_2(x_{21}) \cdot g_2(x_{22}) \cdot \dots \cdot g_2(x_{2n_2}) \cdot \prod_{j=1}^{n_1} \exp\{\alpha_1 + \beta_1' h(x_{1j})\}$$

Definisikan $p_i = g_2(t_i)$, $i = 1, 2, \dots, n$ dimana $n = n_1 + n_2$, maka

$$L(\alpha_1, \beta_1, g_2) = p_1 \cdot p_2 \cdot \dots \cdot p_{n_1} \cdot p_{n_1+1} \cdot p_{n_1+2} \cdot \dots \cdot p_{n_1+n_2} \prod_{j=1}^{n_1} \exp\{\alpha_1 + \beta_1' h(x_{1j})\}$$

sehingga diperoleh fungsi likelihood sebagai berikut:

$$L(\alpha_1, \beta_1, p_i) = \prod_{i=1}^n p_i \prod_{j=1}^{n_1} \exp\{\alpha_1 + \beta_1' h(x_{1j})\} \quad (3.5.1)$$

Dengan menggunakan prosedur yang dibahas oleh Fokianos (2001), maka prosedur dalam penaksir parameter pada model sebagai berikut:

1. Menyatakan setiap p_i dalam suku α_1 dan β_1 .
2. Mensubstitusi kembali p_i ke fungsi likelihood untuk memperoleh fungsi likelihood yang hanya dalam α_1 dan β_1 .
3. Mencari taksiran dari α_1 dan β_1 dengan menggunakan fungsi likelihood yang hanya dalam α_1 dan β_1 dengan metode *Maximum Likelihood Estimator (MLE)*.

1. Menyatakan setiap p_i dalam suku α_1 dan β_1 .

Dengan memisalkan α_1 dan β_1 tertentu, fungsi likelihood (3.5.1)

dimaksimumkan dengan memaksimumkan $\prod_{i=1}^n p_i$, dengan dua faktor

kendala,

$$\sum_{i=1}^n p_i = 1 \quad (3.5.2)$$

$$\sum_{i=1}^n p_i [w_1(t_i) - 1] = 0 \quad (3.5.3)$$

dimana $w_1(t) = \exp\{\alpha_1 + \beta_1' h(x_{1j})\}$.

Memaksimumkan $\prod_{i=1}^n p_i$ sama halnya dengan memaksimumkan

$\sum_{i=1}^n \log p_i$. Karena terdapat dua faktor kendala, maka akan digunakan metode

Lagrange dengan menggunakan dua fungsi pengali Lagrange yaitu λ_0 dan

λ_1 . Akan dicari p_i , $i = 1, 2, \dots, n$ dalam suku α_1 dan β_1 , λ_0 dan λ_1 sehingga

fungsi Lagrange diperoleh sebagai berikut:

$$\begin{aligned} H &= \sum_{i=1}^n \log p_i - \lambda_0 \left(\sum_{i=1}^n p_i - 1 \right) - \lambda_1 \sum_{i=1}^n p_i [w_1(t_i) - 1] \\ &= \sum_{i=1}^n \log p_i + \lambda_0 \left(1 - \sum_{i=1}^n p_i \right) - \lambda_1 \sum_{i=1}^n p_i [w_1(t_i) - 1] \end{aligned} \quad (3.5.4)$$

Langkah selanjutnya yang dilakukan adalah mencari λ_0 dan λ_1 melalui

turunan parsial dari fungsi Lagrange terhadap p_i .

$$\frac{\partial H}{\partial p_i} = \frac{1}{p_i} - \lambda_0 - \lambda_1[w_1(t_i) - 1] = 0, \quad i = 1, 2, \dots, n \quad (3.5.5)$$

$$\frac{\partial H}{\partial p_i} = \frac{1 - p_i \lambda_0 - p_i \lambda_1 [w_1(t_i) - 1]}{p_i} = 0, \quad i = 1, 2, \dots, n$$

$$p_i \frac{\partial H}{\partial p_i} = 1 - p_i \lambda_0 - p_i \lambda_1 [w_1(t_i) - 1] = 0, \quad i = 1, 2, \dots, n$$

$$\sum_{i=1}^n p_i \frac{\partial H}{\partial p_i} = \sum_{i=1}^n 1 - p_i \lambda_0 - p_i \lambda_1 [w_1(t_i) - 1] = 0$$

$$\sum_{i=1}^n p_i \frac{\partial H}{\partial p_i} = n - \lambda_0 \sum_{i=1}^n p_i - \lambda_1 \sum_{i=1}^n p_i [w_1(t_i) - 1] = 0$$

Karena $\sum_{i=1}^n p_i = 1$ dan $\sum_{i=1}^n p_i [w_1(t_i) - 1] = 0$ maka diperoleh

$$\sum_{i=1}^n p_i \frac{\partial H}{\partial p_i} = n - \lambda_0(1) - \lambda_1(0) = 0$$

$$\sum_{i=1}^n p_i \frac{\partial H}{\partial p_i} = n - \lambda_0 = 0, \text{ sehingga diperoleh } \lambda_0 = n \text{ dan } \lambda_1 = v_1 n.$$

Dari persamaan (3.5.5) diperoleh

$$\frac{1}{p_i} - \lambda_0 - \lambda_1 [w_1(t_i) - 1] = 0, \quad i = 1, 2, \dots, n$$

Karena $\lambda_0 = n$ dan $\lambda_1 = v_1 n$, maka diperoleh p_i yang hanya dalam suku α_1

dan β_1 ,

$$\frac{1}{p_i} = n + n v_1 [w_1(t_i) - 1], \quad i = 1, 2, \dots, n$$

$$p_i = \frac{1}{n} \frac{1}{1 + v_1[w_1(t_i) - 1]}, \quad i = 1, 2, \dots, n \quad (3.5.6)$$

dengan mensubstitusikan p_i ke faktor kendala (3.5.3) diperoleh

$$\frac{1}{n} \sum_{i=1}^n \frac{w_1(t_i) - 1}{1 + v_1(w_1(t_i) - 1)} = 0 \quad (3.5.7)$$

Lalu substitusikan p_i ke $L(\alpha_1, \beta_1, G)$, sehingga diperoleh fungsi likelihood,

$$L(\alpha_1, \beta_1) = \prod_{i=1}^n \frac{1}{n} \cdot \frac{1}{1 + v_1(w_1(t_i) - 1)} \prod_{j=1}^{n_1} \exp\{\alpha_1 + \beta_1' \mathbf{h}(x_{1j})\} \quad (3.5.8)$$

dan log-likelihood menjadi sebagai berikut:

$$\begin{aligned} \ell &\equiv \log L(\alpha_1, \beta_1) \\ &= \sum_{i=1}^n \log \left(\frac{1}{n} \cdot \frac{1}{1 + v_1(w_1(t_i) - 1)} \right) + \sum_{j=1}^{n_1} \{\alpha_1 + \beta_1' \mathbf{h}(x_{1j})\} \\ &= \sum_{i=1}^n \log \left(\frac{1}{n} \right) + \sum_{i=1}^n \log \left(\frac{1}{1 + v_1(w_1(t_i) - 1)} \right) + \sum_{j=1}^{n_1} \{\alpha_1 + \beta_1' \mathbf{h}(x_{1j})\} \\ &= \sum_{i=1}^n \log n^{-1} + \sum_{i=1}^n \log [1 + v_1(w_1(t_i) - 1)]^{-1} + \sum_{j=1}^{n_1} \{\alpha_1 + \beta_1' \mathbf{h}(x_{1j})\} \\ &= -\sum_{i=1}^n \log n - \sum_{i=1}^n \log [1 + v_1(w_1(t_i) - 1)] + \sum_{j=1}^{n_1} \{\alpha_1 + \beta_1' \mathbf{h}(x_{1j})\} \\ &= -\sum_{i=1}^n \log n - \sum_{i=1}^n \log [1 + v_1(w_1(t_i) - 1)] + \sum_{j=1}^{n_1} \{\alpha_1 + \beta_1' \mathbf{h}(x_{1j})\} \\ &= -n \log n - \sum_{i=1}^n \log [1 + v_1(w_1(t_i) - 1)] + \sum_{j=1}^{n_1} \{\alpha_1 + \beta_1' \mathbf{h}(x_{1j})\} \quad (3.5.9) \end{aligned}$$

Akan tetapi fungsi likelihood yang diperoleh masih mengandung bentuk v_1 . Untuk memperoleh bentuk dari v_1 dicari dengan menggunakan persamaan $\partial \ell / \partial \alpha_1 = 0$ dan menggunakan persamaan (3.5.7).

Pertama, akan dicari $\partial \ell / \partial \alpha_1 = 0$ dan diperoleh sebagai berikut:

$$-v_1 \sum_{i=1}^n \frac{w_1(t_i)}{1 + v_1(w_1(t_i) - 1)} + \sum_{j=1}^{n_1} 1 = 0$$

$$v_1 \sum_{i=1}^n \frac{w_1(t_i)}{1 + v_1(w_1(t_i) - 1)} = n_1 \quad (3.5.10)$$

Dari persamaan (3.5.7) diperoleh

$$\frac{1}{n} \sum_{i=1}^n \frac{w_1(t_i) - 1}{1 + v_1(w_1(t_i) - 1)} = 0$$

$$\sum_{i=1}^n \frac{w_1(t_i) - 1}{1 + v_1(w_1(t_i) - 1)} = 0$$

$$\sum_{i=1}^n \frac{w_1(t_i)}{1 + v_1(w_1(t_i) - 1)} = \sum_{i=1}^n \frac{1}{1 + v_1(w_1(t_i) - 1)} \quad (3.5.11)$$

Substitusikan persamaan (3.5.6) ke persamaan (3.5.2), sehingga diperoleh

$$\sum_{i=1}^n \frac{1}{n} \frac{1}{1 + v_1(w_1(t_i) - 1)} = 1$$

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{1 + v_1(w_1(t_i) - 1)} = 1$$

$$\sum_{i=1}^n \frac{1}{1 + v_1(w_1(t_i) - 1)} = n \quad (3.5.12)$$

Substitusikan persamaan (3.5.12) ke persamaan (3.5.11), maka diperoleh

$$\sum_{i=1}^n \frac{w_1(t_i)}{1 + v_1(w_1(t_i) - 1)} = \sum_{i=1}^n \frac{1}{1 + v_1(w_1(t_i) - 1)} = n \quad (3.5.13)$$

Lalu substitusikan persamaan (3.5.13) ke persamaan (3.5.10) dan diperoleh

$$v_1 n = n_1$$

atau

$$v_1 = \frac{n_1}{n}$$

Substitusikan v_1 yang diperoleh ke persamaan (3.5.6), diperoleh

$$\rho_i = \frac{1}{n} \cdot \frac{1}{1 + \frac{n_1}{n}(w_1(t_i) - 1)}$$

$$\rho_i = \frac{1}{n} \cdot \frac{1}{\frac{n + n_1(w_1(t_i) - 1)}{n}}$$

$$\rho_i = \frac{1}{n} \cdot \frac{n}{n + n_1(w_1(t_i) - 1)}$$

$$\rho_i = \frac{1}{n + n_1(w_1(t_i) - 1)} \cdot \frac{n_2}{n_2}$$

$$\rho_i = \frac{1}{n_2} \cdot \frac{1}{\frac{n + n_1(w_1(t_i) - 1)}{n_2}}$$

$$\rho_i = \frac{1}{n_2} \cdot \frac{1}{\frac{n}{n_2} - \frac{n_1}{n_2} + \frac{n_1}{n_2} w_1(t_i)}$$

$$\rho_i = \frac{1}{n_2} \cdot \frac{1}{\frac{(n_1 + n_2) - n_1}{n_2} + \frac{n_1}{n_2} w_1(t_i)}$$

$$p_i = \frac{1}{n_2} \cdot \frac{1}{1 + \frac{n_1}{n_2} w_1(t_i)}$$

maka diperoleh setiap p_i yang dinyatakan dalam suku α_1 dan β_1 sebagai berikut:

$$p_i = \frac{1}{n_2} \cdot \frac{1}{1 + \rho_1 w_1(t_i)} \quad (3.5.14)$$

dimana $\rho_1 = \frac{n_1}{n_2}$.

2. Mensubstitusi kembali p_i ke fungsi likelihood untuk memperoleh fungsi likelihood yang hanya dalam α_1 dan β_1 .

Substitusikan persamaan (3.5.14) ke persamaan (3.5.1), sehingga diperoleh fungsi likelihood yang hanya dalam suku α_1 dan β_1 sebagai berikut:

$$L(\alpha_1, \beta_1) = \prod_{i=1}^n \frac{1}{n_2} \cdot \frac{1}{1 + \rho_1 w_1(t_i)} \prod_{j=1}^{n_1} \exp\{\alpha_1 + \beta_1' h(x_{1j})\} \quad (3.5.15)$$

maka log-likelihoodnya sebagai berikut:

$$\begin{aligned} \ell(\alpha_1, \beta_1) &= \sum_{i=1}^n \log\left(\frac{1}{n_2} \cdot \frac{1}{1 + \rho_1 w_1(t_i)}\right) + \sum_{j=1}^{n_1} \{\alpha_1 + \beta_1' h(x_{1j})\} \\ &= \sum_{i=1}^n \log\left(\frac{1}{n_2}\right) + \sum_{i=1}^n \log\left(\frac{1}{1 + \rho_1 w_1(t_i)}\right) + \sum_{j=1}^{n_1} \{\alpha_1 + \beta_1' h(x_{1j})\} \\ &= \sum_{i=1}^n \log n_2^{-1} + \sum_{i=1}^n \log\left(\frac{1}{1 + \rho_1 w_1(t_i)}\right) + \sum_{j=1}^{n_1} \{\alpha_1 + \beta_1' h(x_{1j})\} \end{aligned}$$

$$\ell(\alpha_1, \beta_1) = -n \log n_2 - \sum_{i=1}^n \log[1 + \rho_1 w_1(t_i)] + \sum_{j=1}^{n_1} \{\alpha_1 + \beta_1' h(x_{1j})\} \quad (3.5.16)$$

3. Mencari taksiran dari α_1 dan β_1 dengan metode *Maximum Likelihood Estimator (MLE)*.

Untuk memperoleh taksiran α_1 dan β_1 yang memaksimumkan (3.5.16), dicari turunan parsial dari (3.5.16) untuk masing-masing α_1 dan β_1 lalu menyamakan dengan nol,

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha_1} &= - \sum_{i=1}^n \frac{\rho_1 w_1(t_i)}{1 + \rho_1 w_1(t_i)} + \sum_{i=1}^{n_1} 1 = 0 \\ \rho_1 \sum_{i=1}^n \frac{w_1(t_i)}{1 + \rho_1 w_1(t_i)} &= n_1 \\ \frac{n_1}{n_2} \sum_{i=1}^n \frac{w_1(t_i)}{1 + \rho_1 w_1(t_i)} &= n_1 \\ \sum_{i=1}^n \frac{w_1(t_i)}{1 + \rho_1 w_1(t_i)} &= n_2 \end{aligned} \quad (3.5.17)$$

dan

$$\begin{aligned} \frac{\partial \ell}{\partial \beta_{1k}} &= - \sum_{i=1}^n \frac{\rho_1 h_k(t_i) w_1(t_i)}{1 + \rho_1 w_1(t_i)} + \sum_{j=1}^{n_1} h_k(x_{1j}) = 0, \quad k = 1, \dots, r \\ \rho_1 \sum_{i=1}^n \frac{h_k(t_i) w_1(t_i)}{1 + \rho_1 w_1(t_i)} &= \sum_{j=1}^{n_1} h_k(x_{1j}) \\ \frac{n_1}{n_2} \sum_{i=1}^n \frac{h_k(t_i) w_1(t_i)}{1 + \rho_1 w_1(t_i)} &= \sum_{j=1}^{n_1} h_k(x_{1j}) \end{aligned} \quad (3.5.18)$$

Notasi yang digunakan dalam penaksiran parameter akan disesuaikan dengan notasi yang telah digunakan dalam pembentukan *scanning window* dengan mendefinisikan notasi-notasi yang digunakan sebagai berikut:

N_G : ukuran sampel gabungan pada *study area*.

n_G : banyaknya kejadian pada *study area*.

N_Z : ukuran sampel di dalam *scanning window*.

n_Z : banyaknya kejadian di dalam *scanning window*.

t_i : observasi ke- i dari sample gabungan pada *study area*, $i = 1, \dots, N_G$.

ρ_1 : ukuran sampel relatif, dimana $\rho_1 = \frac{N_Z}{N_G - N_Z}$.

x_{Zj} : observasi ke- j di dalam *scanning window* Z , $j = 1, \dots, N_Z$.

Setelah penyesuaian notasi maka persamaan 3.5.17 dan 3.5.18 menjadi sebagai berikut:

$$\sum_{i=1}^{N_G} \frac{w_1(t_i)}{1 + \frac{N_Z}{N_G - N_Z} w_1(t_i)} = N_G - N_Z \quad (3.5.19)$$

$$\frac{N_Z}{N_G - N_Z} \sum_{i=1}^{N_G} \frac{h_k(t_i) w_1(t_i)}{1 + \frac{N_Z}{N_G - N_Z} w_1(t_i)} = \sum_{j=1}^{N_Z} h_k(x_{Zj}) \quad (3.5.20)$$

Setelah dilakukan pemilihan *tilt function*, maka persamaan (3.5.19) dan (3.5.20) dapat diselesaikan sehingga diperoleh $\hat{\alpha}_1$ dan $\hat{\beta}_1$ dengan menggunakan metode numerik, Newton-Raphson, yang terdapat pada S-plus yaitu fungsi `nlinb`. Metode Newton-Raphson dapat dilihat pada Lampiran 4.

RASIO LIKELIHOOD

Statistik uji yang digunakan dalam tugas akhir ini adalah statistik uji rasio likelihood yaitu $-2 \ln \Lambda \xrightarrow{d} \chi_r^2$, dimana Λ merupakan rasio likelihood dan r merupakan dimensi dari β_1 .

Penentuan rasio likelihood, Λ , sebagai berikut:

Dari persamaan (3.5.15) diperoleh fungsi Likelihood sebagai berikut:

$$L(\alpha_1, \beta_1) = \prod_{i=1}^n \frac{1}{n_2} \cdot \frac{1}{1 + \rho_1 w_1(t_i)} \prod_{j=1}^{n_1} \exp\{\alpha_1 + \beta_1' h(x_{1j})\}$$

$$L(\alpha_1, \beta_1) = \prod_{i=1}^n \frac{1}{n_2} \cdot \frac{1}{1 + \rho_1 \exp\{\alpha_1 + \beta_1' h(t_i)\}} \prod_{j=1}^{n_1} \exp\{\alpha_1 + \beta_1' h(x_{1j})\} \quad (3.7.1)$$

Definisikan:

1. Ruang parameter keseluruhan $\Omega = \{\beta_1 : -\infty < \beta_{1i} < \infty, i = 1, \dots, r\}$.
2. Ruang parameter yang terdefiniskan dalam kondisi H_0 ,
 $\omega = \{\beta_1 : \beta_1 = \mathbf{0}\} \omega \subseteq \Omega$.

Fungsi likelihood di bawah kondisi H_0 ($\beta_1 = \mathbf{0}$) didapat dengan mensubstitusikan $\beta_1 = \mathbf{0}$ dan $\alpha_1 = 0$ ke persamaan (3.7.1) sehingga diperoleh sebagai berikut:

$$L(\omega) = \prod_{i=1}^n \frac{1}{n_2} \cdot \frac{1}{1 + \rho_1 \exp\{0 + \mathbf{0}h(t_i)\}} \prod_{j=1}^{n_1} \exp\{0 + \mathbf{0}h(x_{1j})\}$$

$$L(\omega) = \prod_{i=1}^n \frac{1}{n_2} \cdot \frac{1}{1 + \rho_1(1)} \prod_{j=1}^{n_1} 1$$

$$L(\omega) = \prod_{i=1}^n \frac{1}{n_2} \cdot \frac{1}{1 + \frac{n_1}{n_2}} \quad (1)$$

$$L(\omega) = \prod_{i=1}^n \frac{1}{n_2 + n_1}$$

$$L(\omega) = \prod_{i=1}^n \frac{1}{n} = n^{-n}$$

sehingga diperoleh

$$L(\hat{\omega}) = n^{-n} \quad (3.7.2)$$

Fungsi likelihood dalam ruang parameter keseluruhan Ω sebagai berikut:

$$L(\Omega) = \begin{cases} \prod_{i=1}^n \frac{1}{n_2} \cdot \frac{1}{1 + \rho_1 \exp\{\alpha_1 + \beta_1' \mathbf{h}(t_i)\}} \prod_{j=1}^{n_1} \exp\{\alpha_1 + \beta_1' \mathbf{h}(x_{1j})\} & , \beta_1 \neq 0 \\ n^{-n} & , \beta_1 = 0 \end{cases}$$

dan diperoleh sebagai berikut:

$$L(\hat{\Omega}) = \begin{cases} \prod_{i=1}^n \frac{1}{n_2} \cdot \frac{1}{1 + \rho_1 \exp\{\hat{\alpha}_1 + \hat{\beta}_1' \mathbf{h}(t_i)\}} \prod_{j=1}^{n_1} \exp\{\hat{\alpha}_1 + \hat{\beta}_1' \mathbf{h}(x_{1j})\} & , \beta_1 \neq 0 \\ n^{-n} & , \beta_1 = 0 \end{cases} \quad (3.7.3)$$

Dari persamaan (3.7.2) dan (3.7.3) diperoleh rasio likelihood sebagai berikut:

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

$$= \begin{cases} \frac{n^{-n}}{\prod_{i=1}^n \frac{1}{n_2} \cdot \frac{1}{1 + \rho_1 \exp\{\hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(t_i)\}} \prod_{j=1}^{n_1} \exp\{\hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(x_{1j})\}} & , \beta_1 \neq 0 \\ 1 & , \beta_1 = 0 \end{cases}$$

sehingga diperoleh

$$\begin{aligned} -2 \ln \Lambda &= -2 \left(\ln n^{-n} - \sum_{i=1}^n \ln \frac{1}{n_2} \cdot \frac{1}{1 + \rho_1 (\exp\{\hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(t_i)\})} - \sum_{j=1}^{n_1} \hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(x_{1j}) \right) \\ &= -2 \left(\ln n^{-n} - \sum_{i=1}^n \ln \frac{1}{n_2} \cdot \frac{1}{1 + \rho_1 (\exp\{\hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(t_i)\})} - \sum_{j=1}^{n_1} \hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(x_{1j}) \right) \\ &= 2n \ln n + 2 \sum_{i=1}^n \ln \frac{1}{n_2} \cdot \frac{1}{1 + \rho_1 (\exp\{\hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(t_i)\})} + 2 \sum_{j=1}^{n_1} \hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(x_{1j}) \\ &= 2n \ln n + 2 \sum_{i=1}^n \ln \frac{1}{n_2} + 2 \sum_{i=1}^n \ln \frac{1}{1 + \rho_1 (\exp\{\hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(t_i)\})} + 2 \sum_{j=1}^{n_1} \hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(x_{1j}) \\ &= 2n \ln n - 2 \sum_{i=1}^n \ln n_2 - 2 \sum_{i=1}^n \ln [1 + \rho_1 (\exp\{\hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(t_i)\})] + 2 \sum_{j=1}^{n_1} \hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(x_{1j}) \\ &= 2n \ln n - 2n \ln n_2 - 2 \sum_{i=1}^n \ln [1 + \rho_1 (\exp\{\hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(t_i)\})] + 2 \sum_{j=1}^{n_1} \hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(x_{1j}) \\ &= 2n \ln \frac{n}{n_2} - 2 \sum_{i=1}^n \ln [1 + \rho_1 (\exp\{\hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(t_i)\})] + 2 \sum_{j=1}^{n_1} \hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(x_{1j}) \\ &= 2n \ln \left(\frac{n_2}{n_2} + \frac{n_1}{n_2} \right) - 2 \sum_{i=1}^n \ln [1 + \rho_1 (\exp\{\hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(t_i)\})] + 2 \sum_{j=1}^{n_1} \hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(x_{1j}) \\ &= 2n \ln \left(1 + \frac{n_1}{n_2} \right) - 2 \sum_{i=1}^n \ln [1 + \rho_1 (\exp\{\hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(t_i)\})] + 2 \sum_{j=1}^{n_1} \hat{\alpha}_1 + \hat{\beta}'_1 \mathbf{h}(x_{1j}) \quad (3.7.4) \end{aligned}$$

Dengan menggunakan teorema 2.5.1 maka di bawah kondisi $H_0 : \beta_1 = \mathbf{0}$, $-2 \ln \Lambda$ secara asimtotik berdistribusi χ^2 dengan derajat bebas r , dimana r merupakan dimensi dari β_1 yang bergantung dari pemilihan *tilt function* $h(x)$.

Setiap *scanning window* yang terbentuk berkaitan dengan suatu statistik uji yang digunakan, dalam tugas akhir ini yaitu rasio likelihood, sehingga terdapat sejumlah besar pengujian dan statistik uji. Hal ini menyebabkan *multiple testing problem*. Oleh karena itu, untuk mengatasinya digunakan *Bonferroni Correction*. Subbab selanjutnya memberikan penjelasan lebih lanjut mengenai uji signifikansi dan *Bonferroni Correction*.

3.6 UJI SIGNIFIKANSI

Dikarenakan pendekatan secara asimtotik dari rasio likelihood diketahui, maka dapat diperoleh *p-value* yang selanjutnya akan digunakan untuk uji signifikansi.

Banyaknya *scanning window* yang terbentuk menghasilkan sejumlah besar pengujian dan statistik uji. Hal ini menyebabkan *multiple testing problem*. Oleh karena itu, untuk mengatasinya digunakan *Bonferroni correction*.

Dengan *Bonferroni correction*, tingkat signifikansi yang digunakan pada setiap pengujian yaitu α / n dimana n menyatakan banyaknya pengujian, sehingga tingkat signifikansi untuk keseluruhan pengujian yaitu α .

3.7 ATURAN KEPUTUSAN

Setelah diperoleh *p-value* dari setiap pengujian dan dilakukan uji signifikansi dengan *Bonferroni correction*, selanjutnya adalah menentukan keputusan apakah hipotesis nol akan ditolak atau diterima. Tentukan tingkat signifikansi tertentu (misalnya 0.05). Berikut ini adalah aturan keputusan yang dibuat untuk mendeteksi cluster dengan metode semi-parametrik untuk setiap pengujian:

1. H_0 akan ditolak, jika $p\text{-value} < 0.05/n$.
2. H_0 akan diterima, jika $p\text{-value} > 0.05/n$.

3.8 KESIMPULAN

Setelah diperoleh keputusan apakah H_0 ditolak atau diterima, yang perlu dilakukan selanjutnya adalah membuat suatu kesimpulan dari hipotesis yang telah dibuat. Berikut ini adalah kesimpulan yang dibuat menurut hasil aturan keputusan yang diperoleh:

1. Jika keputusan H_0 tidak ditolak, maka kesimpulannya adalah *scanning window* tersebut bukan merupakan cluster.
2. Jika keputusan H_0 ditolak, maka kesimpulannya adalah *scanning window* tersebut merupakan cluster.

