UNIVERSITAS INDONESIA

ESTIMASI PARAMETER MODEL REGRESI LOGISTIK SPASIAL MENGGUNAKAN METODE VARIASIONAL

SKRIPSI

HAKIIM NUR RIZKA
1706047391

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
PROGRAM SARJANA STATISTIKA
DEPOK
JANUARI 2021

UNIVERSITAS INDONESIA

ESTIMASI PARAMETER MODEL REGRESI LOGISTIK SPASIAL MENGGUNAKAN METODE VARIASIONAL

SKRIPSI

Diajukan sebagai salah satu syarat memperoleh gelar sarjana

HAKIIM NUR RIZKA
1706047391

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
PROGRAM SARJANA STATISTIKA
DEPOK
JANUARI 2021

HALAMAN PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya saya sendiri,
dan semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.

Nama	**: Hakiim Nur Rizka**
NPM	**: 1706047391**
Tanda Tangan	**: ____________________**
Tanggal	**: 26 Januari 2021**

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Hakiim Nur Rizka

NPM : 1706047391

Program Studi : Statistika

Judul Skripsi : Estimasi Parameter Model Regresi Logistik Spasial Menggunakan Metode Variasional

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Indonesia.

Type text here

DEWAN PENGUJI

Pembimbing : Dr. Dra. Yekti Widyaningsih, M.Si (....................)

Penguji : Sarini Abdullah S.Si., M.Stats., Ph.D (....................)

Penguji : Dra. Rianti Setiadi, M.Si. (....................)

Ditetapkan di : Depok

Tanggal: 26 Januari 2021

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmatNya penulis dapat menyelesaikan naskah skripsi ini sebagai salah satu syarat mendapatkan gelar sarjana. Shalawat serta salam tidak lupa penulis ucapkan kepada teladan Nabi Muhammad SAW. Mengingat skripsi ini tidak lepas dari kontribusi berbagai pihak, pada kesempatan ini penulis ingin menghaturkan ucapan terima kasih. Ucapan terima kasih penulis haturkan kepada :

[1] Ibu Dr. Dra. Yekti Widyaningsih, M.Si yang telah bersedia menjadi pembimbing penulis dalam penelitian ini serta meluangkan banyak waktunya untuk memberikan kritik dan saran selama penulis menyusun skripsi ini.
[2] Ibu Sarini Abdullah S.Si., M.Stats., Ph.D selaku Ketua Program Studi Sarjana Statistika dan tim penguji yang telah memberikan pengarahan kepada penulis selama masa perkuliahan.
[3] Ibu Dra. Rianti Setiadi, M.Si selaku tim penguji sidang skripsi.
[4] Ibu Dr. Titin Siswantining, DEA selaku pembimbing akademik yang telah banyak membagikan ilmu kepada penulis selama masa perkuliahan.
[5] Ibu Dr. Dian Lestari, DEA selaku ketua Departemen Matematika Universitas Indonesia.
[6] Seluruh bapak dan ibu dosen di Departemen Matematika Universitas Indonesia yang telah memberikan ilmu kepada penulis selama masa perkuliahan.
[7] Orang tua kandung penulis atas setiap doa, dukungan, dan motivasi untuk penulis.
[8] Kakak kandung penulis atas dukungan dan motivasi untuk penulis.
[9] Yudha selaku teman penulis yang memberikan saran terkait sumber untuk studi literatur pada penelitian ini.
[10] Habib selaku teman penulis yang memberikan bantuan dalam melakukan uji plagiarisme untuk naskah skripsi ini.

Tidak lupa kepada semua pihak yang penulis tidak dapat sebutkan satu per satu.

Dengan penuh kesadaran dan kerendahan hati, penulis memahami masih terdapat kekurangan dalam skripsi ini. Maka dari itu, kritik dan saran akan sangat penulis apresiasi untuk menyempurnakan naskah skripsi ini. Akhir kata, penulis berharap penelitian pada skripsi ini dapat membawa manfaat untuk peneliti mendatang.

Depok, 26 Januari 2021

Penulis

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Indonesia, saya yang bertanda tangan di bawah ini:

Nama	: Hakiim Nur Rizka
NPM	: 1706047391
Program Studi	: Statistika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam
Jenis karya	: Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Indonesia **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul :

Estimasi Parameter Model Regresi Logistik Spasial Menggunakan Metode Variasional

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Indonesia berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Depok

Pada tanggal : 26 Januari 2021

Yang menyatakan

(Hakiim Nur Rizka)

ABSTRAK

Nama : Hakiim Nur Rizka
Program Studi : Sarjana Statistika
Judul : Estimasi Parameter Model Regresi Logistik Spasial Menggunakan Metode Variasional
Pembimbing : Dr. Dra. Yekti Widyaningsih, M.Si

Model regresi logistik spasial membangun persamaan dengan jenis variabel dependen adalah biner serta mempertimbangkan dependensi spasial pada data. Proses estimasi parameter pada model ini memerlukan algoritma EM. Namun, bentuk ekspektasi dari *complete log-likelihood* pada *E-step* tidak tersedia dalam *closed-form*. Dalam menangani permasalahan ini, metode terbaru yang sedang diteliti oleh Cecille Hardouin memanfaatkan pendekatan deterministik dikenal dengan metode variasional. Metode variasional merupakan metode aproksimasi distribusi yang memanfaatkan suatu batas bawah fungsi distribusi yang akan diaproksimasi lalu mengoptimalkan batas bawah ini. Metode variasional untuk estimasi parameter model regresi logistik spasial bekerja dengan mencari suatu batas bawah dari *complete log-likelihood* lalu memaksimumkan fungsi ini terhadap parameter model. Dalam studi literatur, didapatkan bahwa metode variasional memiliki akurasi lebih baik daripada algoritma EM dengan aproksimasi Laplace ketika dependensi spasial pada data relatif besar.

Kata Kunci :
Metode Variasional, Model Regresi Logistik Spasial, Algoritma EM

ABSTRACT

Name	:	Hakiim Nur Rizka
Study Program	:	Bachelor of Statistics
Title	:	Parameter Estimation of Spatial Logistic Regression Model using Variational Method
Counsellor	:	Dr. Dra. Yekti Widyaningsih, M.Si

The spatial logistic regression model builds equations in which the dependent variable is binary and considers the spatial dependency on the data. Estimation procedure of the parameters in this model require EM algorithm. However, the expected form of the complete log-likelihood on the E-step is not available in closed-form. In order to deal with this problem, a recent method being researched by Cecille Hardouin utilizes a deterministic approach known as the variational method. The variational method is a distribution approximation method that utilizes a lower bound of the distribution function to be approximated and then optimizes this lower bound. The variational method for estimating the parameters of the spatial logistic regression model works by finding a lower limit of the complete log-likelihood and then maximizing this function to the model parameters. In the literature study, it was found that the variational method has better accuracy than the EM algorithm with Laplace's approximation when the spatial dependence on the data is relatively large.

Key Words :
Variational Methods, Spatial Logistic Regression Model, EM Algorithm

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1. 1. Latar Belakang

Analisis regresi merupakan metode yang digunakan dalam memodelkan hubungan antara satu variabel dependen dengan satu atau lebih variabel independen. Dalam analisis regresi, dibangun suatu persamaan yang dikenal sebagai persamaan regresi. Secara prinsip, persamaan regresi dibentuk oleh variabel dependen dan independen yang berjenis numerik kontinu. Namun, terdapat kasus dimana variabel yang menjadi perhatian berjenis kategorik. Regresi dengan variabel independen berupa kategorik masih termasuk kedalam regresi linear, hanya saja terdapat perlakuan khusus terhadap variabel independen kategorik. Untuk menangani kasus dimana variabel dependen berjenis kategorik, terdapat suatu jenis model regresi yang dikenal sebagai regresi logistik.

Prinsip dasar regresi logistik mirip dengan regresi linear yaitu menghubungkan variabel dependen dengan independen melalui suatu persamaan. Fungsi logistik digunakan dalam membangun hubungan kedua jenis variabel tersebut.

Regresi spasial merupakan salah satu model regresi yang dikembangkan untuk menangani data dengan dependensi spasial. Dependensi spasial pada data dapat diidentifikasi ketika terdapat efek pada suatu objek/wilayah tertentu yang mempengaruhi objek/wilayah didekatnya. Efek ini diwakili melalui matriks kovariansi spasial yang dapat dikarakterisasi melalui suatu fungsi kovariansi.

Dalam pemodelan spasial dikenal juga model *geographically weighted regression*(GWR) yang merupakan model regresi dengan mempertimbangkan adanya dependensi spasial pada data. Perbedaan utama model GWR dengan regresi spasial adalah regresi spasial berlaku global untuk setiap sub wilayah sedangkan GWR membangun persamaan-persamaan regresi untuk masing-masing sub wilayah.

Regresi logistik spasial merupakan model yang dikembangkan untuk mengatasi permasalahan ketika jenis variabel dependen adalah biner serta data memiliki dependensi spasial. Bentuk persamaan regresi logistik spasial memiliki prinsip dan asumsi yang mirip dengan regresi logistik biasa. Dalam membangun model regresi logistik spasial, masalah pada estimasi parameter menjadi perhatian utama. Hal ini mengingat estimasi parameter menggunakan metode likelihood maksimum tidak memiliki bentuk *closed-form*. Beberapa pendekatan dilakukan untuk menangani masalah ini, salah satunya adalah algoritma *expectation-maximization*(EM).

Algoritma EM merupakan metode aproksimasi yang bekerja secara iteratif untuk mendapatkan estimasi parameter dengan menggunakan fungsi *complete likelihood*. *Complete likelihood* adalah fungsi likelihood yang mengasumsikan adanya nilai-nilai tak teramati dari data. Oleh algoritma EM, parameter dalam model diestimasi dengan dua langkah : (1) menghitung ekspektasi dari fungsi *complete likelihood* dan (2) memaksimumkan ekspektasi tersebut terhadap setiap parameter dalam model. Kedua langkah ini dikerjakan secara iteratif hingga mencapai taraf konvergensi yang ditetapkan. Di akhir prosedur iteratif, didapatkan bentuk ekspektasi *complete likelihood* yang telah dimaksimumkan terhadap setiap parameter dalam model. Hal ini setara dengan prosedur estimasi metode likelihood maksimum. Nilai parameter pada fungsi ekspektasi *complete likelihood* yang didapatkan di iterasi terakhir tersebut, digunakan sebagai estimasi parameter dalam model.

Dalam penerapan algoritma EM pada estimasi parameter model regresi logistik spasial, diperlukan aproksimasi Laplace pada bentuk ekspektasi *complete likelihood*. Pada penelitian ini dibahas mengenai alternatif dari penggunaan aproksimasi Laplace tersebut yang dikenal dengan nama metode variasional.

Metode variasional atau dikenal juga sebagai bayes variasional merupakan metode aproksimasi yang digunakan dalam mengaproksimasi fungsi distribusi. Metode variasional melakukan aproksimasi distribusi pada permasalahan asli melalui pendefinisian suatu keluarga distribusi. Setelah mendefinisikan keluarga distribusi, dipilih suatu distribusi dari keluarga distribusi tersebut yang dikenal sebagai distribusi variasional. Selanjutnya distribusi variasional dimaksimumkan terhadap parameternya secara iteratif. Distribusi variasional yang didapatkan diakhir prosedur iteratif tersebut digunakan sebagai wakil dari distribusi pada permasalahan asli.

Pada penelitian ini akan ditegaskan mengenai penerapan metode variasional pada estimasi parameter regresi logistik spasial.

1. 2. Rumusan Masalah

Dalam penelitian ini, diajukan rumusan masalah sebagai berikut :

1) Bagaimana prosedur untuk mendapatkan estimasi parameter model regresi logistik spasial menggunakan metode variasional?
2) Bagaimana sifat estimasi parameter menggunakan metode variasional pada model regresi logistik spasial?

1. 3. Tujuan Penelitian

Berdasarkan rumusan masalah yang diajukan, maka tujuan penelitian ini adalah :

1) Menganalisis prosedur dalam mendapatkan estimasi parameter model regresi logistik spasial menggunakan metode variasional
2) Menganalisis sifat dari estimasi parameter menggunakan metode variasional pada model regresi logistik spasial

1. 4. Batasan Masalah

Untuk membatasi permasalahan dalam penelitian ini, berikut adalah batasan yang dilakukan :

1) Model regresi logistik spasial dibangun menggunakan *hierarchical framework*.
2) Untuk mengetahui sifat dari estimasi, dilakukan studi literatur terkait.

1. 5. Metodologi Penelitian

Dalam menyelesaikan penelitian ini, digunakan metodologi penelitian berupa studi literatur. Studi literatur memanfaatkan sumber tertulis serta video. Sumber-sumber tertulis berupa buku, makalah, jurnal, dan website yang dapat dipercaya. Sumber lain berupa video merupakan video pembelajaran atau perkuliahan yang disampaikan oleh orang-orang dengan latar belakang akademis yang relevan dan dapat dipercaya.

BAB II

LANDASAN TEORI

Bab ini berisi penjelasan mengenai kalkulus diferensial, fungsional, teorema bayes, proses stokastik, distribusi normal, distribusi binomial, metode likelihood maksimum, regresi logistik, dan regresi spasial.

2. 1. Turunan Parsial

Misal terdapat suatu fungsi f dengan dua variabel x dan y, turunan parsial fungsi f terhadap x dapat dinyatakan sebagai berikut :

$$f_x(x,y) = \lim_{h\to 0} \frac{f(x+h,y) - f(x,y)}{h} = \frac{\partial f}{\partial x}$$

Sedangkan turunan parsial f terhadap y adalah :

$$f_y(x,y) = \lim_{h\to 0} \frac{f(x,y+h) - f(x,y)}{h} = \frac{\partial f}{\partial y}$$

Dari definisi diatas, turunan parsial dapat diperumum untuk fungsi dengan variabel lebih dari dua.

2. 2. Gradien

Misalkan untuk suatu fungsi f dengan $\boldsymbol{p} = (x,y) = [x \;\; y]$. Gradien yang dimaksudkan merujuk pada vektor $(f_x(\boldsymbol{p}), f_y(\boldsymbol{p}))$ yang dinotasikan dengan $\nabla f(\boldsymbol{p})$. Untuk kasus umum, misalkan $\boldsymbol{p}$ adalah suatu vektor ukuran n dengan fungsi $f : \mathbf{R}^n \rightarrow \mathbf{R}$, gradien fungsi f adalah :

$$\nabla f(\mathbf{p}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{p}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{p}) \end{bmatrix}$$

2. 3. Fungsi Lagrangian

Lagrangian biasa digunakan untuk mendapatkan nilai maksimum dan minimum suatu fungsi relative terhadap suatu fungsi batas yang biasa dinamakan *constraint*.

Sebagai ilustrasi, misalkan $f\ (x, y)$ suatu fungsi dengan turunan terhadap *x* dan *y* kontinu. Fungsi f ingin dimaksimumkan dengan syarat *constraint* $h\ (x, y) = c$ dimana c adalah suatu nilai konstanta. Dalam kasus ini dapat dibangun fungsi Lagrange sebagai berikut

$$L(x, y, \lambda) = f(x, y) - \lambda[h(x, y) - c]$$

Dimana λ dikenal sebagai pengali Lagrange. Permasalahan tersebut dapat diselesaikan dengan menyelesaikan persamaan berikut

$$\nabla_{x,y,\lambda} L(x, y, \lambda) = \mathbf{0}$$

Dari ilustrasi tersebut, fungsi lagrange dapat dibentuk untuk memaksimumkan fungsi dengan lebih dari satu variabel dan *constraint* lebih dari satu.

2. 4. Fungsional

Fungsional dapat dianggap sebagai suatu fungsi yang memetakan fungsi ke bilangan (Keng, 2017). Salah satu contohnya adalah integral.

$$F[f(x)] = \int f(x)\, dx$$

Dari pengertian fungsional secara intuitif tersebut, diperkenalkan notasi Lagrangian.

2. 5. Teorema Bayes

Misalkan terdapat 2 himpunan kejadian A dan B dengan P(A|B) menyatakan distribusi A bersyarat B. Dalam definisi distribusi bersyarat

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B \cap A^c) + P(B \cap A)}$$

Didapatkan aturan bayes

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B|A^c) \times P(A^c) + P(B|A) \times P(A)}$$

Dalam inferensi bayes misalkan terdapat nilai-nilai teramati *y* dengan himpunan parameter Θ, distribusi parameter dengan syarat nilai teramati (biasa dikenal dengan distribusi posterior) dapat dituliskan sebagai berikut

$$p(\Theta|y) = \frac{p(y|\Theta) \times p(\Theta)}{p(y)}$$

Dimana $p(y|\Theta)$ menyatakan likelihood untuk *y* pada model, $p(\Theta)$ menyatakan distribusi prior himpunan parameter Θ , dan $p(y)$ adalah likelihood marjinal dari *y*. Likelihood marjinal *y*

$$p(y) = \int p(y|\Theta) \times p(\Theta) \, \mathrm{d}\Theta$$

menyatakan suatu konstanta sedemikian bentuk $p(y|\Theta)$ dipastikan merupakan suatu distribusi yang sesuai dan hasil integrasinya sama dengan 1.

2. 6. Proses Stokastik

Merujuk pada buku Ross (2010), proses stokastik *[X(t), t ∈T]* merupakan suatu himpunan dari variabel acak sedemikian sehingga untuk setiap $t \in T$, *X(t)* merupakan variabel acak. Istilah yang digunakan untuk *X(t)* adalah *state* pada indeks *t*. Indeks dapat berupa suatu ruang atau waktu. Contoh untuk hal ini adalah *X(s)* menotasikan banyaknya angka kelahiran di wilayah yang direpresentasikan dengan *s*. Salah satu jenis proses stokastik adalah rantai Markov.

Rantai Markov memiliki sifat khusus yaitu *memoriless* yaitu sifat yang mengatakan bahwa *state* selanjutnya hanya bergantung dari *state* masa kini. Sifat lain yang menjadi perhatian adalah sifat *ergodic*. Untuk menjelaskan sifat ini berikut adalah ulasannya

Misalkan P_{ij}^n menotasikan peluang bahwa proses di *state i* akan berada di *state j* setelah *n* transisi. Suatu *state i* memiliki periode *d* jika $P_{ii}^n = 0$ jika *n* tidak dapat dibagi oleh *d* dan *d* bilangan bulat terbesar yang memenuhi kondisi ini. *State* dikatakan *aperiodic* jika memiliki periode 1. *State j accesible* dari *state i* yaitu jika terdapat $p_{ij}^{n} > 0$ untuk suatu *n*. Jika dua *state i dan j* saling *accesible*, maka dua *state* dikatakan *communicate*. Selanjutnya misalkan f_i menotasikan peluang bahwa untuk suatu proses pada *state i* suatu saat akan kembali ke *state i*. *State i* dikatakan *recurrent* jika $f_i = 1$ dan *transient* jika $f_i = 0$. Jika suatu *state recurrent*, maka *state* tersebut dikatakan *recurrent* positif jika, ketika proses dimulai dari *state i*, ekspektasi waktu hingga proses kembali ke *state i* adalah berhingga.

Dari ulasan diatas, jika suatu rantai Markov memiliki *state* yang *recurrent* positif dan *aperiodic* maka rantai Markov tersebut dinamakan ergodik.

2. 7. Distribusi Normal

Distribusi normal merupakan salah satu keluarga distribusi kontinu yang dijelaskan melalui suatu persamaan yang dikenal dengan nama persamaan normal. Persamaan normal dibangun melalui bentuk integral :

$$I = \int_{-\infty}^{\infty} \exp\left[-\frac{y^2}{2}\right] dy$$

Untuk mendapatkan nilai $I > 0$ digunakan I^2 dengan menuliskannya sebagai:

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left[-\frac{x^2 + y^2}{2}\right] dxdy$$

$x = r\cos\theta$ dan $y = r\sin\theta$

$$I^2 = \int_{0}^{2\pi}\int_{0}^{\infty} \exp\left[-\frac{r^2}{2}\right] r\,drd\theta = \int_{0}^{2\pi} d\theta = 2\pi$$

Didapatkan $I = \sqrt{2\pi}$ serta bentuk persamaan

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{y^2}{2}\right] dy = 1$$

Yang kemudian diperumum untuk $y = \frac{x-a}{b}$ menghasilkan bentuk integral :

$$\int_{-\infty}^{\infty} \frac{1}{b\sqrt{2\pi}} \exp\left[-\frac{(x-a)^2}{2\,b^2}\right] dx$$

Fungsi pembangkit momen untuk distribusi ini :

$$M(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{b\sqrt{2\pi}} \exp\left[-\frac{(x-a)^2}{2\,b^2}\right] dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{b\sqrt{2\pi}} \exp\left[-\frac{-2b^2tx+(x-a)^2}{2\,b^2}\right] dx$$

$$= exp\left[-\frac{-b^4t^2-2b^2at}{2\,b^2}\right] \int_{-\infty}^{\infty} \frac{1}{b\sqrt{2\pi}} \exp\left[-\frac{-2b^2tx+(x-a)^2+b^4t^2+2b^2at}{2\,b^2}\right] dx$$

$$= exp\left[at+\frac{b^2t^2}{2}\right] \int_{-\infty}^{\infty} \frac{1}{b\sqrt{2\pi}} \exp\left[-\frac{(x-a-b^2t)^2}{2\,b^2}\right] dx$$

$$= exp\left[at+\frac{b^2t^2}{2}\right]$$

Momen pertama dan kedua :

$$M'(t) = [a+b^2t]exp\left[at+\frac{b^2t^2}{2}\right]$$

$$M''(t) = a[a+b^2t]exp\left[at+\frac{b^2t^2}{2}\right] + b^2exp\left[at+\frac{b^2t^2}{2}\right] + b^2t[a+b^2t]exp\left[at+\frac{b^2t^2}{2}\right]$$

$$M''(t) = [a+b^2t]^2\, exp\left[at+\frac{b^2t^2}{2}\right] + b^2exp\left[at+\frac{b^2t^2}{2}\right]$$

Akhirnya didapatkan :

Mean : $\mu = M'(0) = a$

Variansi : $\sigma^2 = M''(0) - \mu^2 = b^2$

Dari penjabaran diatas, didapatkan bentuk fungsi distribusi normal secara umum :

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

2. 8. Distribusi Normal Multivariat

Distribusi ini merupakan distribusi bersama n variabel acak, pembentukanya adalah sebagai berikut :

Misalkan terdapat **A** matriks simetris positif definit ukuran n x n, vektor konstan **μ**'= [μ_1 , μ_2 , ... , μ_n] , dan **x'** = [x_1 , x_2 , ... , x_n]. Fungsi nonnegatif x_1, x_2, ..., x_n dengan bentuk sebagai berikut :

$$f(x_1, x_2, \dots, x_n) = C \exp\left[-\frac{(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{\mu})}{2}\right], -\infty < x_i < \infty, i = 1,2, \dots n$$

Untuk suatu konstan C yang sesuai dapat ditunjukkan

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n)\, dx_1 \dots dx_n = 1$$

Misalkan **t'** = [t_1 , t_2 , ... , t_n] dengan t_1 , t_2 , ... , t_n adalah sebarang bilangan riil. Integral diatas kita tuliskan menjadi

$$C \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[\boldsymbol{t}'\boldsymbol{x} - \frac{(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{\mu})}{2}\right] dx_1 \dots dx_n$$

Transformasi x_1 , x_2 , ... , x_n ke y_1 , y_2 , ... , y_n melalui **x-μ = y.** Bentuk integral menjadi

$$C \exp\left(\boldsymbol{t}'\boldsymbol{\mu}\right) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[\boldsymbol{t}'\boldsymbol{y} - \frac{\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}}{2}\right] dx_1 \dots dx_n$$

Karena **A** matriks simetris definit positif, dapat ditemukan suatu matrik ortogonal **L** sedemikian **L'AL** = diag[a_1, a_2, ..., a_n]. a_1, a_2, ..., a_n merupkan n bilangan karakteristik dari matriks **A**. Selanjutnya transformasi y_1 , y_2 , ... , y_n ke z_1 , z_2 , ... , z_n melalui **y = Lz**. Karena **L'** ortogonal(**L'L=I**) maka nilai mutlak jacobian transformasi tersebut adalah 1.

$$C \exp\left(\boldsymbol{t}'\boldsymbol{\mu}\right) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[\boldsymbol{t}'\boldsymbol{L}\boldsymbol{z} - \frac{\boldsymbol{z}\boldsymbol{L}'\boldsymbol{A}\boldsymbol{L}\boldsymbol{z}}{2}\right] dx_1 \dots dx_n$$

Serta untuk memudahkan notasi $\boldsymbol{t}'\boldsymbol{L} = \boldsymbol{w}'$

$$\exp[\,\boldsymbol{t}'\boldsymbol{L}\boldsymbol{z}] = \exp[\,\boldsymbol{w}'\boldsymbol{z}] = \exp\left(\sum_{i=1}^{n} w_i z_i\right)$$

$$\exp\left[\frac{\boldsymbol{z}(\boldsymbol{L}'\boldsymbol{A}\boldsymbol{L})\boldsymbol{z}}{2}\right] = \exp\left[\frac{\sum_{i=1}^{n} a_i {z_i}^2}{2}\right]$$

Bentuk integral menjadi

$$C \exp(\boldsymbol{w}'\boldsymbol{L}'\boldsymbol{\mu}) \prod_{i=1}^{n} \int_{-\infty}^{\infty} \exp\left[w_i z_i - \frac{a_i {z_i}^2}{2}\right] dz_1$$

$$= C \exp(\boldsymbol{w}'\boldsymbol{L}'\boldsymbol{\mu}) \prod_{i=1}^{n} \sqrt{\frac{2\pi}{a_i}} \int_{-\infty}^{\infty} \frac{\exp\left[w_i z_i - \frac{a_i {z_i}^2}{2}\right]}{\sqrt{2\pi/a_i}} dz_1$$

$$= C \exp(\boldsymbol{w}'\boldsymbol{L}'\boldsymbol{\mu}) \prod_{i=1}^{n} \sqrt{\frac{2\pi}{a_i}} \int_{-\infty}^{\infty} \exp(w_i z_i) \times \frac{\exp\left[-\frac{a_i {z_i}^2}{2}\right]}{\sqrt{2\pi/a_i}} dz_1$$

$$= C \exp(\boldsymbol{w}'\boldsymbol{L}'\boldsymbol{\mu}) \prod_{i=1}^{n} \sqrt{\frac{2\pi}{a_i}} \exp\left(\frac{{w_i}^2}{2\, a_i}\right)$$

$$= C \exp(\boldsymbol{w}'\boldsymbol{L}'\boldsymbol{\mu}) \sqrt{\frac{2\pi}{a_1 a_2 \dots a_n}} \exp\left(\sum_{i=1}^{n} \frac{{w_i}^2}{2\, a_i}\right)$$

$$= C \exp(\boldsymbol{w}'\boldsymbol{L}'\boldsymbol{\mu}) \sqrt{(2\pi)^n |\boldsymbol{L}'\, \boldsymbol{A}^{-1}\boldsymbol{L}|} \exp\left(\frac{\boldsymbol{w}'\boldsymbol{L}'\boldsymbol{A}^{-1}\boldsymbol{L}\boldsymbol{w}}{2}\right)$$

$$= C \exp(\boldsymbol{t}'\boldsymbol{\mu}) \sqrt{(2\pi)^n |\, \boldsymbol{A}^{-1}|} \exp\left(\frac{\boldsymbol{t}'\boldsymbol{A}^{-1}\boldsymbol{t}}{2}\right)$$

Dengan mengatur nilai $t_1 = t_2 = \dots = t_n = 0$ dan menyatakan ke persamaan

$$C \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[-\frac{(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{\mu})}{2}\right] dx_1 \dots dx_n = 1$$

$$C \exp(\boldsymbol{t}'\boldsymbol{\mu}) \sqrt{(2\pi)^n |\, \boldsymbol{A}^{-1}|} \exp\left(\frac{\boldsymbol{t}'\boldsymbol{A}^{-1}\boldsymbol{t}}{2}\right) = 1$$

$$C \sqrt{(2\pi)^n |\, \boldsymbol{A}^{-1}|} = 1$$

Akhirnya didapatkan bentuk yang merepresentasikan fungsi distribusi bersama x_1, x_2, ..., x_n

$$f(x_1, x_2, \dots, x_n) = \frac{1}{\sqrt{(2\pi)^n |\,\boldsymbol{A}^{-1}|}} \exp\left[-\frac{(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{\mu})}{2}\right],$$

$$-\infty < x_i < \infty, \qquad i = 1,2, \dots n$$

2. 9. Distribusi Binomial

Percobaan Bernoulli merupakan dasar dari pembentukan distribusi binomial. Percobaan Bernoulli merupakan yang himpunan kejadian yang mungkin terjadi adalah tepat dua kejadian yang saling bebas dan saling lepas. Biasanya dua kejadian ini dinamai "sukses"dan "gagal". Pendefinisian menggunakan variabel acak dapat dituliskan sebagai berikut : misalkan X(.) menotasikan variabel acak untuk percobaan Bernoulli, maka hal ini dapat dituliskan dengan $X(sukses) = 1$ dan $X\,(gagal) = 0$.

Untuk fungsi densitas untuk X tersebut adalah :

$$f(x) = p^x(1-p)^{1-x}, x = 0,1$$

Dimana $0 \le p \le 1$ menyatakan probabilitas "sukses" yang didefinisikan untuk percobaan tersebut. Terkait momen untuk X diatas dapat dituliskan sebagai berikut :

$$E(X) = \sum_{x=0}^{1} x\, p^x(1-p)^{1-x} = p$$

$$E(X^2) = \sum_{x=0}^{1} x^2\, p^x(1-p)^{1-x} = p$$

Maka didapatkan mean dan variansi untuk X yaitu :

$$\mu = E(X) = p$$

$$\sigma^2 = E(X^2) - E(X)^2 = p - p^2 = p(1-p)$$

Dengan notasi sama, X, didefinisikan sebagai variabel acak yang menyatakan banyaknya "sukses". Maka ketika percobaan Bernoulli diulang sebanyak n, maka nilai X = 0, 1, 2, ..., n. Dari hal tersebut terbentuk konsep distribusi dari dari keluarga distribusi diskrit yang dikenal sebagai distribusi binomial. Karena percobaan saling bebas dengan peluang “sukses” dan “gagal” adalah p dan $(1-p)$, maka untuk percobaan berulang n kali $p^x(1-p)^{n-x}$. Didapatkan bentuk distribusi binomial adalah sebagai berikut :

$$f(x) = \binom{n}{x} p^x(1-p)^{n-x}, \;\; x = 0,1,\dots,n$$

$$= 0, \text{ lainnya}$$

2. 10. Metode Likelihood Maksimum

Misalkan dimiliki $X_1,\dots,X_n$ sampel acak suatu distribusi yang diketahui memiliki fungsi densitas $f(x;\theta),\; \theta \in \Omega$ dimana Ω menotasikan ruang parameter. Dapat dibangun fungsi likelihood dari informasi tersebut dengan membentuk fungsi densitas gabungan dari sampel acak sebagai berikut :

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

$$L(\theta;x_1,x_2,\dots,x_n) = f(x_1;\theta)f(x_2;\theta)\dots f(x_n;\theta), \qquad \theta \in \Omega$$

Mencari nilai maksimum fungsi likelihood L dilakukan dengan mencari solusi persamaan turunan pertama. Proses sebelum mencari nilai maksimum dapat melalui bentuk logaritma dari fungsi likelihood. Logaritma dari fungsi likelihood disebut fungsi log-likelihood. Proses tersebut dapat dituliskan sebagai berikut :

$$\ell(\theta;x_1,x_2,\dots,x_n) = \log L(\theta;x_1,x_2,\dots,x_n)$$

$$\frac{\partial \ell(\theta;x_1,x_2,\dots,x_n)}{\partial \theta} = 0$$

Solusi dari persamaan terakhir diatas digunakan sebagai taksiran untuk parameter θ yang dapat dinotasikan dengan $\hat{\theta}$.

2. 11. Regresi Logistik

Regresi logistik merupakan model yang mengakomodasi pemodelan untuk variabel dependen berjenis biner. Secara umum, persamaan regresi logistik dapat dituliskan sebagai berikut :

Misalkan terdapat k variabel independen, dan dibangun fungsi

$$E(y) = \pi = \frac{\exp\,(\beta_0 + \beta_1 x_1 + .. + \beta_k x_k)}{1 + \exp\,(\beta_0 + \beta_1 x_1 + .. + \beta_k x_k)}$$

Dimana

$y = 1$, jika kejadian A terjadi

$y = 0$, jika kejadian B terjadi

Dari persamaan diatas, terlihat bahwa bentuk model tidak linear. Untuk membangun model linear dari persamaan diatas, digunakan bentuk logaritma dari rasio π terhadap 1- π. Sehingga didapatkan :

$$\pi = \frac{\exp\,(\beta_0 + \beta_1 x_1 + .. + \beta_k x_k)}{1 + \exp\,(\beta_0 + \beta_1 x_1 + .. + \beta_k x_k)}$$

$$\frac{\pi}{1-\pi} = \left[\frac{\exp(\beta_0 + \beta_1 x_1 + .. + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + .. + \beta_k x_k)}\right] / \left[\frac{1}{1 + \exp(\beta_0 + \beta_1 x_1 + .. + \beta_k x_k)}\right]$$

$$= \exp(\beta_0 + \beta_1 x_1 + .. + \beta_k x_k)$$

$$\log \frac{\pi}{1-\pi} = \log \exp(\beta_0 + \beta_1 x_1 + .. + \beta_k x_k)$$

$$\log \frac{\pi}{1-\pi} = \pi * = \beta_0 + \beta_1 x_1 + .. + \beta_k x_k$$

Bentuk π^* dikenal dengan nama log-odds. Jika diperhatikan bentuk log-odds dapat diartikan sebagai logaritma rasio “kejadian A” terjadi terhadap “kejadian B” terjadi. Menggunakan definisi log-odds diatas, model linear untuk regresi logistik dididapatkan.

$$\pi^* = \beta_0 + \beta_1 x_1 + .. + \beta_k x_k$$

Dalam praktiknya, prosedur estimasi parameter regresi logistik banyak dilakukan menggunakan metode likelihood maksimum. Selanjutnya dibahas mengenai fungsi likelihood untuk regresi logistik.

Misalkan terdapat n nilai teramati, maka untuk setiap nilai teramati dapat dinyatakan dalam persamaan regresi logistik menjadi :

$$E(y_i) = \pi_i = \frac{\exp\,(\boldsymbol{x'}_i\boldsymbol{\beta})}{1 + \exp\,(\boldsymbol{x'}_i\boldsymbol{\beta})}$$

Dimana $\boldsymbol{x'_i}$ = [1, x_{i1}, x_{i2}, ..., x_{ik}], dan $\boldsymbol{\beta}' = [\beta_0, \beta_1, \beta_2, \ldots, \beta_k]$.

Selanjutnya nilai-nilai teramati diasumsikan berasal dari distribusi Bernoulli, sehingga didapatkan fungsi densitas :

$$f_i(y_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i}, \qquad i = 1, 2, .., n$$

Sehingga dapat dibentuk fungsi likelihood dan log-likelihood

$$L(\boldsymbol{\beta}; y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n} f_i(y_i) = \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i}$$

$$\log L(\boldsymbol{\beta}; y_1, y_2, \ldots, y_n) = \log \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i}$$

$$= \sum_{i=1}^{n} y_i \log \pi_i + \sum_{i=1}^{n} (1-y_i) \log(1-\pi_i)$$

$$= \sum_{i=1}^{n} y_i \log \pi_i + \sum_{i=1}^{n} \log(1-\pi_i) - \sum_{i=1}^{n} y_i \log(1-\pi_i)$$

$$= \sum_{i=1}^{n} y_i \log \frac{\pi_i}{1-\pi_i} + \sum_{i=1}^{n} \log(1-\pi_i)$$

Selanjutnya bentuk akhir diatas diganti dengan notasi matriks menjadi :

$$\log L(\boldsymbol{\beta}; y_1, y_2, \ldots, y_n) = \sum_{i=1}^{n} y_i \log \exp\,(\boldsymbol{x'_i\beta}) + \sum_{i=1}^{n} \log[1+\exp(\boldsymbol{x'_i\beta})]^{-1}$$

$$\log L(\boldsymbol{\beta}; y_1, y_2, \ldots, y_n) = \sum_{i=1}^{n} y_i\, \boldsymbol{x'_i\beta} + \sum_{i=1}^{n} \log[1+\exp(\boldsymbol{x'_i\beta})]^{-1}$$

$$\log L(\boldsymbol{\beta}; y_1, y_2, \ldots, y_n) = \sum_{i=1}^{n} y_i\, \boldsymbol{x'_i\beta} - \sum_{i=1}^{n} \log[1+\exp(\boldsymbol{x'_i\beta})]$$

Untuk bentuk log-likelihood dengan adanya perulangan nilai teramati :

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} y_i\, \boldsymbol{x'_i\beta} - \sum_{i=1}^{n} n_i \log[1+\exp(\boldsymbol{x'_i\beta})]$$

Prosedur untuk mendapatkan estimasi $\hat{\boldsymbol{\beta}}$ dapat dilanjutkan dengan metode numerik. Salah satu metode yang dapat digunakan adalah *least-square* dengan pembobotan berulang iteratif melalui metode Newton-Raphson. Penerapan metode Newton-Raphson untuk estimasi parameter regresi logistik terdapat pada lampiran 1. Pada penelitian ini, digunakan model logistik pada kerangka kerja *hierarchical*. Penggunaan langkah kerja *hierarchical* dipilih karena dianggap lebih representatif terhadap keacakan data. Hal juga ini berkaitan penerapan teorema Bayes dalam model regresi logistik.

2. 12. Regresi Spasial

Model dependen spasial merupakan salah satu metode statistika yang menangani data dengan informasi lokasi. Data dengan informasi lokasi menjadi perhatian karena didalam informasi lokasi sering ditemukan adanya pola. Munculnya pola dalam data dengan informasi lokasi ini menjadi dasar pemodelan dependen spasial. Salah satu hasil pengembangan dengan adanya model dependen spasial adalah regresi spasial.

Regresi spasial, layaknya analisis regresi, membangun model melalui persamaan regresi. Perbedaan terbesar didalam model regresi spasial adalah adanya komponen lokasi pada data. Terdapat dua bentuk umum dari regresi spasial yang dikenal dengan nama model lag spasial dan model error spasial.

Bentuk regresi lag spasial memiliki komponen variabel dependen yang disaring secara spasial. Model lag spasial dapat memiliki variabel independen yang tersaring spasial. Komponen ini ditransformasi oleh matriks bobot spasial serta koefisien lag spasial. Oleh Anselin(1988) bentuknya dapat dituliskan sebagai berikut :

$$\boldsymbol{y} = \boldsymbol{X\beta} + \rho\boldsymbol{W}\boldsymbol{y} + \boldsymbol{\varepsilon}$$

Dimana $\boldsymbol{W}$ menotasikan matriks dependensi spasial dan ρ adalah koefisen lag spasial. Bentuk diatas dapat diubah ke bentuk persamaan dengan variabel dependen berada di bagian kiri. Prosesnya sebagai berikut :

$$\boldsymbol{y} = \boldsymbol{X\beta} + \rho\boldsymbol{W}\boldsymbol{y} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{y} - \rho \boldsymbol{W}\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$(\boldsymbol{I} - \rho \boldsymbol{W})\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{y} = (\boldsymbol{I} - \rho \boldsymbol{W})^{-1}\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{I} - \rho \boldsymbol{W})^{-1}\boldsymbol{\varepsilon}$$

Bentuk persamaan terakhir diatas menjadi dasar untuk digunakan dalam mengestimasi parameter-parameter dalam persamaan tersebut.

Selanjutnya untuk bentuk regresi spasial error, komponen yang disaring spasial adalah bagian error. Oleh Anselin (1988) bentuk persamaannya dituliskan sebagai berikut :

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$$

$$\boldsymbol{u} = \lambda \boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon}$$

Dengan λ menotasikan koefisien autokorelasi spasial.

Pada penelitian ini, fokus yang dipakai adalah model spasial lag. Hal ini merujuk pada Anselin (2002) dimana model yang menyangkut nilai-nilai tak teramati hanya dapat digunakan pada model lag spasial. Namun dalam pembangunan model logistik spasial pada bab 4 akan digunakan notasi khusus. Hal ini dikarenakan model regresi logistik yang digunakan berupa kerangka model *hierarchical*.

BAB III

METODE *VARIATIONAL INFERENCE*

Bab ini membahas dasar dari metode *variational inference* atau singkatnya metode variasional. Selain itu dibahas juga mengenai salah satu jenis keluarga variasional yaitu *mean-field* variasional. Terakhir, dibahas mengenai penelitian-penelitian terkait metode variasional untuk memberikan ilustrasi mengenai teori estimasi metode variasional.

3. 1. Dasar Metodologi *Variational Inference*

Metode telah dikenal sebagai salah satu metode aproksimasi eksak yang banyak digunakan di beberapa bidang. Penerapan metode variasional seperti pada bidang statistika (Rustagi, 1976), mekanika kuantum (Sakurai, 1985), statisitika mekanik (Parisi, 1988). Secara prinsip, penerapan metode variasional pada bidang-bidang tersebut adalah dengan mengubah permasalahan asli yang kompleks ke bentuk permaslahan yang lebih sederhana. Sebagai ilustrasi, gambar 3.1 merupakan

Oleh metode variasional, pemisahan permasalahan kompleks dilakukan melalui ekspansi dari permasalahan asli dengan menggunakan suatu parameter tambahan yang dikenal sebagai parameter variasonal. Parameter variasional ini digunakan sebagai parameter baru dalam fungsi objektif. Istilah variasi dalam metode variasional berasal dari kalkulus variasi.

3. 1. 1. Kalkulus Variasi

Kalkulus variasi merupakan salah satu cabang dalam analisis matematika yang membahas mengenai permasalahan maksimum dan minimum dalam analisis fungsional (Rustagi, 1976). Selanjutnya dibahas mengenai beberapa definisi dan teorema yang membangun dasar dari kalkulus variasi.

Teknik variasional pada analisis fungsional dapat dikatakan memiliki peran yang sama dengan teori maksimum dan minimum pada kalkulus diferensial. Teknik variasional bermula ketika muncul permasalahan optimasi fungsional. Permasalahan ini dapat diilustrasikan sebagai berikut :

Misalkan terdapat suatu fungsi *f(x)* bernilai real yang terdefinisi pada [x_1, x_2]. Diketahui juga *f(x)* memiliki turunan terhadap *x*, *f'(x)*, yang kontinu. Misalkan *L[x, f(x), f'(x)]*(yang dikenal sebagai Lagrangian) adalah suatu fungsi kontinu dan *differentiable* hingga orde kedua. Selanjutnya didefinisikan suatu fungsional *I[f(x)]* sebagai berikut :

$$I[f(x)] = \int_{x_1}^{x_2} L[x, f(x), f'(x)]dx \qquad (3.1)$$

Kalkulus variasi memiliki tujuan untuk mengoptimalkan fungsional pada persamaan (3.1) pada kelas $\mathcal{A}$ yang merupakan himpunan fungsi *f(x)* yang kontinu dan terturunkan pada [x_1, x_2]. Kelas $\mathcal{A}$ ini dikenal dengan nama *admissible class* dari fungsi.

Dalam permasalahan variasional, ekstremum global menjadi perhatian utama. Namun, dalam banyak kasus, ekstremum global cukup sulit dipenuhi. Sebagai ilustrasi ;

Misalkan dicari suatu maksimum global dari fungsional *I[f(x)]*, maka perlu dicari suatu $f_0(x) \in \mathcal{A}$ sedemikian sehingga $I[f_0(x)] \geq I[f(x)]$ untuk setiap $f(x) \in \mathcal{A}$. (Definisi 1)

Selain menunjukkan bahwa terdapat $f_0(x)$ yang memenuhi definisi 1, perlu juga ditunjukkan bahwa $f_0(x)$ unik. Karena hal ini, ekstremum lokal perlu dipertimbangkan.

Terkait ekstremum lokal, berikut definisi untuk dua jenis minimum lokal :

Misalkan jarak antara dua fungsi *p(x)* dan *q(x)* adalah

$$d_0(p,q) = \sup_{t \in [x_1, x_2]} |p(x) - q(x)|$$

$f_o(x)$ adalah minimum lokal kuat (*strong local minimum*) untuk fungsional *I[f(x)]* jika $I[f_0(x)] \leq I[f(x)]$ untuk setiap $f(x) \in \mathcal{A}$ sedemikian sehingga terdapat $\delta > 0, d_0(p,q) < \delta$.

Sedangkan untuk minimum lokal lemah (*weak local minimum*), digunakan definisi jarak

$$d_1(p,q) = \sup_{t \in [x_1, x_2]} |p(x) - q(x)| + \sup_{t \in [x_1, x_2]} |p'(x) - q'(x)|$$

$f_o(x)$ merupakan minimum lokal lemah untuk fungsional *I[f(x)]* jika $I[f_0(x)] \leq I[f(x)]$ untuk setiap $f(x) \in \mathcal{A}$ sedemikian sehingga terdapat $\delta > 0, d_1(p,q) < \delta$.

Secara umum, pada permasalahan variasional, hanya terdapat satu jenis ekstremum lokal dari yang telah disebutkan diatas (Rustagi, 1976).

3. 1. 2. Euler-Lagrange Equation

Selanjutnya dibahas mengenai kondisi yang diperlukan untuk suatu ekstremum integral (3.1) dibawah kondisi batas :

$$f(x_1) = y_1 \text{ dan } f(x_2) = y_2 \tag{3.2}$$

Pertama dimisalkan $f^0(x)$ adalah adalah ekstremum untuk (3.1) dan memenuhi kondisi batas (3.2). Diperkenalkan juga fungsi $\eta(x)$ yang memenuhi

$$\eta(x_1) = \eta(x_2) = 0 \tag{3.3}$$

Selanjutnya didefinisikan fungsi baru

$$f(x) = f^0(x) + \alpha\, \eta(x), x_1 \leq x \leq x_2 \tag{3.4}$$

dimana fungsi ini juga diasumsikan memenuhi kondisi batas (3.2) serta *differentiable* terhadap x hingga orde kedua. Permasalahan yang menjadi perhatian utama adalah menentukan fungsi $f(x)$ tertentu yang merupakan ekstremum untuk integral (3.1). Pertama dengan memperhatikan persamaan (3.4), dapat didefinisikan integral (3.1) sebagai suatu fungsi dari α yang dapat dituliskan sebagai berikut

$$I(\alpha) = \int_{x_1}^{x_2} L[x, f(x), f'(x)]dx \tag{3.5}$$

Bentuk (3.5) diatas dapat dibuktikan dengan mempertimbangkan persamaan (3.4) dan bentuk integral terhadap variabel x. dengan menyelesaikan integral tentu pada (3.5), variabel α menjadi satu-satunya variabel didalam fungsi I tersebut. Oleh karena itu untuk mendapatkan ekstremum (3.5) dapat digunakan persamaan turunan pertama dan mencari solusi untuk persamaan tersebut. Namun, solusi untuk persamaan ini telah jelas jika melihat Kembali persamaan (3.4). Sehingga dapat disimpulkan persamaan berikut berlaku :

$$\left.\frac{dI(\alpha)}{d\alpha}\right|_{\alpha=0} = 0 \tag{3.6}$$

pernyataan (3.6) merupakan fakta yang didapatkan bahwa solusi untuk persamaan diferensial untuk mencari ekstremum I adalah ketika $\alpha = 0$. Hal ini karena dengan merujuk pada persamaan (3.4), dapat dilihat bahwa ketika $\alpha = 0$ maka $f(x) = f^0(x)$ dan $f^0(x)$ telah dimisalkan sebagai ekstremum. Selanjutnya persamaan (3.6) diterapkan untuk fungsi pada (3.5) dan didapatkan

$$\frac{d}{d\alpha}\int_{x_1}^{x_2} L[x, f(x), f'(x)]dx\Bigg|_{\alpha=0} = 0$$

$$\int_{x_1}^{x_2} \frac{\partial}{\partial\alpha} L[x, f(x), f'(x)]\Big|_{\alpha=0} dx = 0 \tag{3.7}$$

dan untuk mendapatkan bentuk turunan dari fungsi didalam integral, dilakukan dengan menggunakan aturan rantai

$$\int_{x_1}^{x_2} \left[\frac{\partial L[x, f(x), f'(x)]}{\partial f(x)} \frac{\partial f(x)}{\partial\alpha}\right]\Bigg|_{\alpha=0} + \left[\frac{\partial L[x, f(x), f'(x)]}{\partial f'(x)} \frac{\partial f'(x)}{\partial\alpha}\right]\Bigg|_{\alpha=0} dx = 0 \tag{3.8}$$

Selanjutnya digunakan bentuk turunan parsial fungsi f dan f' terhadap α berdasarkan persamaan (3.4)

$$f'(x) = \frac{\partial f(x)}{\partial x} = f^{0\prime}(x) + \alpha\,\eta'(x) \tag{3.9}$$

$$\frac{\partial f(x)}{\partial\alpha} = \eta(x) \tag{3.10}$$

$$\frac{\partial f'(x)}{\partial\alpha} = \eta'(x) \tag{3.11}$$

dengan menerapkan persamaan (3.10) dan (3.11) kedalam (3.8), didapatkan

$$\int_{x_1}^{x_2} \left[\frac{\partial L[x, f(x), f'(x)]}{\partial f(x)} \eta(x)\right]\Bigg|_{\alpha=0} + \int_{x_1}^{x_2} \left[\frac{\partial L[x, f(x), f'(x)]}{\partial f'(x)} \eta'(x)\right]\Bigg|_{\alpha=0} dx = 0 \tag{3.12}$$

bagian kanan pada fungsi didalam integral (3.12)

$$\int_{x_1}^{x_2} \frac{\partial L[x, f(x), f'(x)]}{\partial f'(x)} \eta'(x) dx \tag{3.13}$$

Dengan integral parsial, bentuk (3.13) dapat dituliskan menjadi

$$\int_{x_1}^{x_2} \frac{\partial L[x, f(x), f'(x)]}{\partial f'(x)} \eta'(x) dx = -\int_{x_1}^{x_2} \left[\int \eta'(x)\ dx\right] \frac{d}{dx} \frac{\partial L[x, f(x), f'(x)]}{\partial f'(x)}\ dx$$

$$+ \frac{\partial L[x, f(x), f'(x)]}{\partial f'(x)} \int_{x_1}^{x_2} \eta'(x) dx \tag{3.14}$$

$$= \frac{\partial L[x, f(x), f'(x)]}{\partial f'(x)} [\eta(x)]_{x_2}^{x_1}$$

$$- \int_{x_1}^{x_2} \eta(x) \frac{d}{dx} \frac{\partial L[x, f(x), f'(x)]}{\partial f'(x)}\ dx \tag{3.15}$$

menerapkan persamaan (3.3) pada (3.15) maka didapatkan bentuk akhir persamaan

$$\int_{x_1}^{x_2} \frac{\partial L[x, f(x), f'(x)]}{\partial f'(x)} \eta'(x) dx = -\int_{x_1}^{x_2} \eta(x) \frac{d}{dx} \frac{\partial L[x, f(x), f'(x)]}{\partial f'(x)}\ dx \tag{3.16}$$

kemudian (3.16) diterapkan ke persamaan (3.12)

$$\int_{x_1}^{x_2} \left[\frac{\partial L[x, f(x), f'(x)]}{\partial f(x)} \eta(x)\right]\Bigg|_{\alpha=0}$$

$$- \left[\frac{d}{dx} \frac{\partial L[x, f(x), f'(x)]}{\partial f'(x)} \eta(x)\right]\Bigg|_{\alpha=0} dx = 0 \tag{3.17}$$

$$\int_{x_1}^{x_2} \eta(x) \left[\frac{\partial L[x, f(x), f'(x)]}{\partial f(x)} - \frac{d}{dx} \frac{\partial L[x, f(x), f'(x)]}{\partial f'(x)}\right]\Bigg|_{\alpha=0} dx = 0 \tag{3.18}$$

selanjutnya digunakan (3.4) sehingga dapat dituliskan bahwa ketika $\alpha = 0$ maka $f(x) = f^0(x)$

$$\int_{x_1}^{x_2} \eta(x) \left[\frac{\partial L[x, f(x), f'(x)]}{\partial f^0(x)} - \frac{d}{dx} \frac{\partial L[x, f(x), f'(x)]}{\partial f^{0\prime}(x)} \right] dx = 0 \qquad (3.19)$$

karena $\eta(x)$ suatu fungsi yang sembarang, maka satu-satunya cara yang memastikan persamaan (3.19) terpenuhi adalah

$$\frac{\partial L[x, f(x), f'(x)]}{\partial f^0(x)} - \frac{d}{dx} \frac{\partial L[x, f(x), f'(x)]}{\partial f^{0\prime}(x)} = 0 \qquad (3.20)$$

Persamaan (3.20) menyatakan persamaan yang dikenal dengan nama persamaan Euler-Lagrange. Dari penjabaran diatas, dapat disimpulkan bahwa jika suatu fungsi $f^0(x)$ adalah ekstremum dari $I[f(x)]$ pada (3.1), maka $f^0(x)$ harus memenuhi persamaan Euler-Lagrange (3.20).

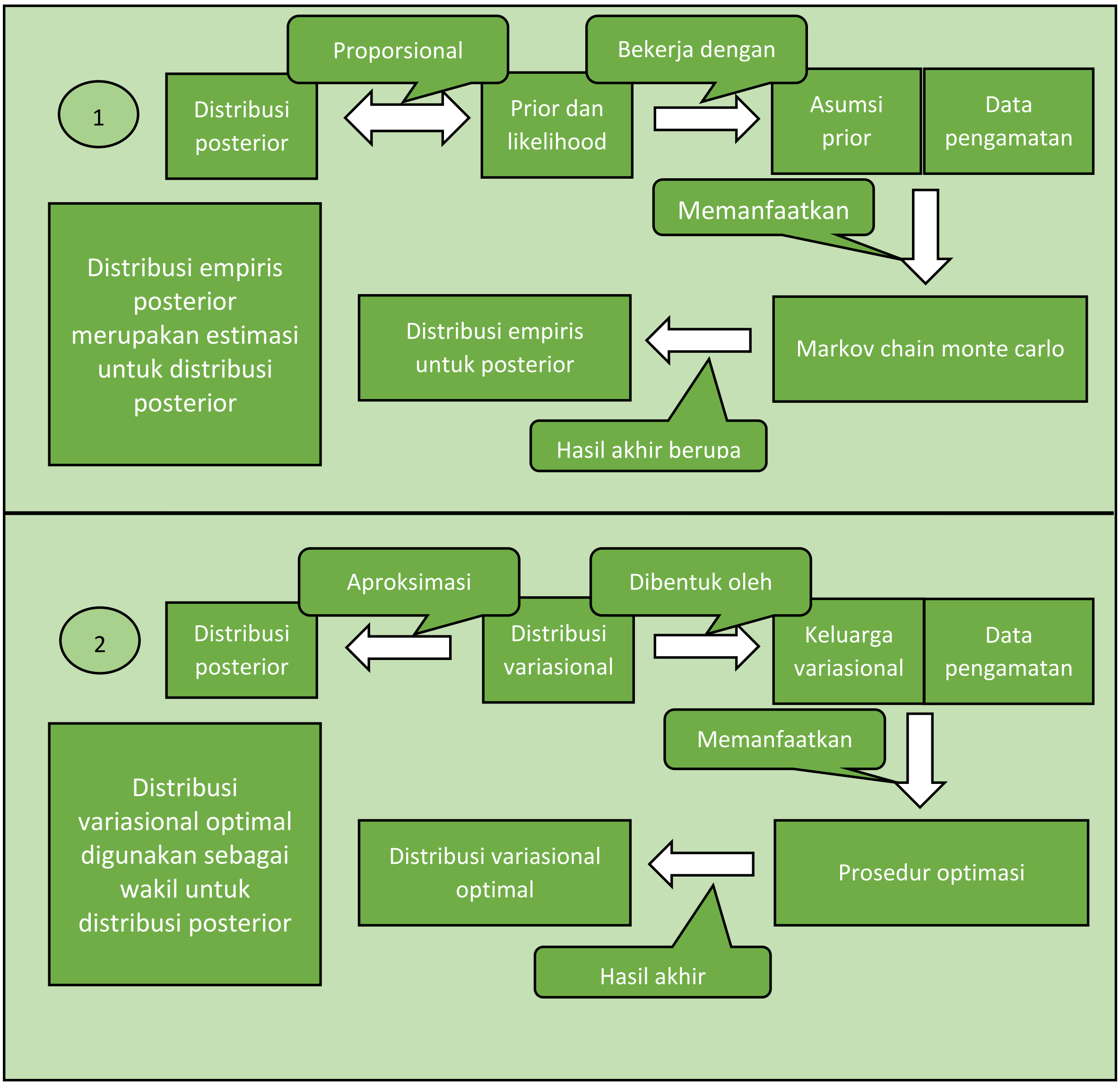
3. 2. Metode *Variational Inference*

Menggunakan dasar-dasar pada kalkulus variasi, metode *Variational Inference* diajukan sebagai alternatif dalam menangani permasalahan estimasi fungsi densitas distribusi. Permasalahan estimasi distribusi densitas ini sangat erat hubungannya dengan statistika Bayesian.

Untuk membahas keterkaitan metode variasional dengan statistika Bayesian, misalkan terdapat nilai-nilai teramati ***X*** dan tak teramati ***Z*** dengan distribusi bersama keduanya *p(**x**,**z**)*. Dengan mendefinisikan adanya nilai-nilai tak teramati ini membantu dalam menentukan distribusi pada populasi (Blei dkk, 2018). Model Bayesian mengambil nilai-nilai tak teramati dari distribusi prior *p(**z**)* dan menghubungkannya dengan nilai-nilai teramati melalui *likelihood p(**x**|**z**).* Prosedur selanjutnya adalah mengkonstruksikan bentuk distribusi posterior *p(**z**|**x**)*. Metode aproksimasi diperlukan dalam konstruksi posterior. Markov Chain Monte Carlo (MCMC) merupakan metode yang paling sering digunakan dalam membangun distribusi posterior (Hastings, 1970).

Metode MCMC diawali dengan membuat rantai Markov ergodic pada nilai-nilai tak teramati ***z*** yang distribusi stasionernya adalah distribusi posterior *p(**z**|**x**)*. Selanjutnya dilakukan pengambilan sampel dari rantai yang telah dibangun untuk mengumpulkan sampel dari distribusi posterior. Aproksimasi dari posterior merupakan distribusi empiris yang dibangun dari sampel yang didapatkan tersebut. Walau banyak digunakan, metode MCMC tidak lepas dari kekurangan.

Masalah pada penggunaan metode MCMC ditemui ketika berhadapan dengan data yang besar dan model yang kompleks. Pada keadaan tersebut, perhatian utama adalah mengenai waktu komputasi. Metode variasional memberikan solusi untuk masalah tersebut. Sebagai tambahan untuk memberikan ilustrasi pada perbedaan kerja pada metode Bayes dengan MCMC dan metode variasional, dapat diperhatikan pada gambar 3.1.



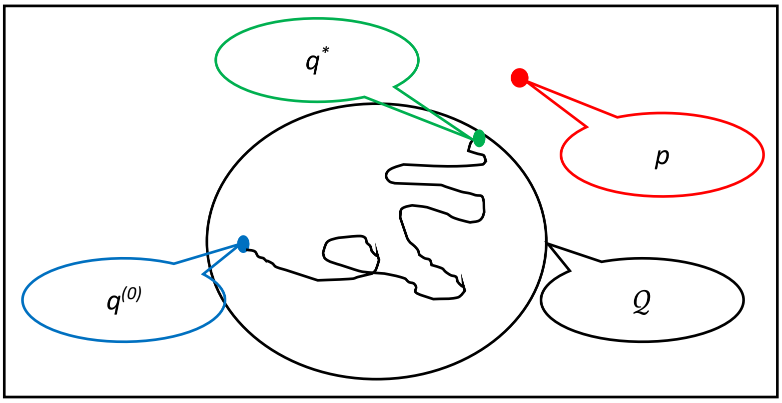
Gambar 3.1 Bagan Alur Kerja Metode Bayes (1) dan Metode Variasional (2)

Prinsip utama metode variasional adalah melalui proses optimasi pada aproksimasi distribusi densitas nilai-nilai tak teramati bersyarat data teramati. Blei (2017) mengatakan bahwa metode variasional secara prinsip mengubah teknik inferensi menjadi teknik optimasi. Metode variasional dikenal juga sebagai Bayes variasional.

Misalkan terdapat data teramati $\boldsymbol{X}$ dengan distribusi posterior $p(\boldsymbol{z}|\boldsymbol{x})$ dimana $\boldsymbol{Z}$ mewakili nilai-nilai tak teramati. Oleh aturan Bayes distribusi posterior ini dapat dituliskan sebagai :

$$p(\boldsymbol{z}|\boldsymbol{x}) = \frac{p(\boldsymbol{z},\boldsymbol{x})}{p(\boldsymbol{x})} = \frac{p(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z})}{\int p(\boldsymbol{z},\boldsymbol{x})d\boldsymbol{z}}$$

Dalam metode variasional, $p(\boldsymbol{z}|\boldsymbol{x})$ diestimasi oleh suatu distribusi $q(\boldsymbol{z}|\boldsymbol{\nu})$ yang dikenal sebagai distrbusi variasional. Parameter $\boldsymbol{\nu}$ pada distribusi variasional dikenal sebagai parameter variasional. Pemilihan distribusi $q(\boldsymbol{z}|\ \boldsymbol{\nu})$ terbatas pada keluarga distribusi $\mathcal{Q}$ yang merupakan himpunan distribusi nilai-nilai tak teramati.



Gambar 3.2 Ilustrasi Proses Kerja Metode Variasional

Setelah distribusi variasional $q(\boldsymbol{z}|\boldsymbol{\nu})$ didefinisikan, selanjutnya dilakukan prosedur optimasi berdasarkan divergensi Kullback-Leibler (KL) sedemikian sehingga di akhir optimasi didapatkan distribusi yang meminimumkan divergensi KL. Distribusi variasional yang didapatkan di akhir prosedur optimasi digunakan sebagai wakil dari distribusi bersyarat yang sebenarnya. Gambar 3.2 memberikan ilustrasi proses kerja metode variasional. Diinginkan aproksimasi fungsi distribusi p melalui metode variasional. $\mathcal{Q}$ menotasikan keluarga variasional, $q^{(0)}$ menotasikan estimasi awal untuk distribusi variasional, dan q^* menotasikan distribusi variasional yang telah dioptimalkan. Divergensi KL digambarkan sebagai kedekatan antara p dengan q pada gambar 3.2. Proses yang dilakukan dari $q^{(0)}$ hingga q^* merupakan proses iterasi untuk meminimumkan divergensi KL.

3. 2. 1. Permasalahan Inferensi Aproksimasi

Misalkan $\boldsymbol{x}_i$ menotasikan nilai-nilai teramati ke-i dan $\boldsymbol{z}_i$ merepresentasikan nilai-nilai tak teramati ke-i yang berkorespondensi, sedemikian sehingga dapat dibentuk densitas gabungan dari keduanya $p(\boldsymbol{x},\boldsymbol{z})$. Masalah inferensi, terpusat pada penghitungan densitas nilai-nilai tak teramati bersyarat nilai-nilai teramati $p(\boldsymbol{z}|\boldsymbol{x})$. Densitas bersyarat ini dapat ditulis kembali menggunakan aturan Bayes

$$p(\boldsymbol{z}|\boldsymbol{x}) = \frac{p(\boldsymbol{z},\boldsymbol{x})}{p(\boldsymbol{x})} \tag{3.21}$$

dengan marjinalitas fungsi densitas, pembagi pada (3.21) dapat dituliskan sebagai

$$p(\boldsymbol{x}) = \int p(\boldsymbol{z},\boldsymbol{x})d\boldsymbol{z} \tag{3.22}$$

sehingga distribusi bersyarat pada (3.21) dapat dituliskan menjadi

$$p(\boldsymbol{z}|\boldsymbol{x}) = \frac{p(\boldsymbol{z},\boldsymbol{x})}{\int p(\boldsymbol{z},\boldsymbol{x})d\boldsymbol{z}} \tag{3.23}$$

Kebanyakan dari model akan menemui setidaknya satu dari masalah berikut : (1) bentuk integral (3.22) tidak tersedia dalam *closed-form*; (2) waktu komputasi yang membesar secara eksponensial seiring dengan banyaknya pengamatan (Blei dkk, 2018).

3. 2. 2. *Evidence of Lower Bound* (ELBO)

Pada *Variational Inference*, ditentukan suatu keluarga distribusi $\mathcal{Q}$ atas nilai-nilai tak teramati. Untuk setiap distribusi $q(\boldsymbol{z}) \in \mathcal{Q}$ adalah kandidat untuk aproksimasi distribusi bersyarat $p(\boldsymbol{z}|\boldsymbol{x})$ pada (3.21). Setelah ditentukan keluarga distribusi $\mathcal{Q}$, dipilih suatu distribusi aproksimasi awal $q^0(\boldsymbol{z}) \in \mathcal{Q}$. Prosedur selanjutnya adalah mengoptimasi distribusi awal ini.

Pemilihan distribusi aproksimasi terbaik $q^*(\boldsymbol{z})$ adalah dengan mencari distribusi $q(\boldsymbol{z}) \in \mathcal{Q}$ yang memiliki divergensi Kullback-Leibler (KL) paling kecil. Pernyataan matematis untuk hal ini dapat dituliskan sebagai berikut :

$$q^*(\boldsymbol{z}) = \underset{q(\boldsymbol{z}) \in \mathcal{Q}}{\arg\min}\, KL\big(q(\boldsymbol{z}) \parallel p(\boldsymbol{z}|\boldsymbol{x})\big)$$

prosedur optimasi selesai dengan mendapatkan distribusi $q^*(\boldsymbol{z})$ yang merupakan aproksimasi terbaik relative pada keluarga Q.

Divergensi KL memiliki definisi (2) :

Misalkan terdapat dua distribusi $q(z)$ dan $p(z|x)$. Ukuran divergensi KL dari *p(z)* ke *q(z)* dapat dihitung menggunakan

$$\begin{aligned} D_{KL}\big(q(z) \parallel p(z|x)\big) &= E_z\left[\log \frac{q(z)}{p(z|x)}\right] \\ &= E_z[\log q(z)] - E_z[\log p(z|x)] \qquad (3.24) \end{aligned}$$

dengan semua bentuk ekspektasi dihitung terhadap $q(z)$.

Divergensi KL dikenal sebagai ukuran bertambah atau berkurangnya informasi ketika satu distribusi dipilih dibandingkan distribusi oposisi. Dalam konteks statistika Bayesian, divergensi KL pada persamaan (3.24) dapat diartikan sebagai ukuran informasi yang hilang jika distribusi *q(z)* digunakan untuk mengaproksimasi posterior *p(z|x)*. Divergensi KL juga memiliki sifat bahwa nilainya selalu non negatif ($D_{KL} \geq 0$) (Kullback dan Leibler, 1951).

Ketika dilakukan ekspansi pada persamaan (3.24) terlihat ada ketergantungan divergensi KL pada distribusi *p(x)*, sebagai berikut

$$\begin{aligned} D_{KL}\big(q(z) \parallel p(z|x)\big) &= E_z[\log q(z)] - E_z[\log p(z|x)] \\ &= E_z[\log q(z)] - E_z[\log p(x,z)] \\ &\qquad + E_z[\log p(x)] \\ &= E_z[\log q(z)] - E_z[\log p(x,z)] + \log p(x) \qquad (3.25) \end{aligned}$$

karena adanya ketergantungan ini, penghitungan divergensi KL (D_{KL}) tidak dapat dilakukan. Untuk mengatasi masalah ini, digunakan suatu bentuk batas bawah yang dikenal sebagai *Evidence of Lower Bound* (ELBO).

Untuk membangun ELBO, diselidiki bentuk log *p(x)* pada (3.25). Pertama dimisalkan *p(x,z)* adalah distribusi bersama *X* dan *Z* dengan *Z* kontinu. Bentuk log *p(x)* pada (3.25) diekspansi menggunakan marjinalitas dengan *Z* menghasilkan

$$
\begin{aligned}
\log p(x) \quad &= \log \int p(x,z)dz \\
&= \log \int \frac{p(x,z)}{q(z)} q(z)dz \\
&= \log E_z \left[\frac{p(x,z)}{q(z)}\right] \qquad (3.26)
\end{aligned}
$$

Selanjutnya dibahas mengenai pertaksamaan Jensen yang diperlukan untuk membangun ELBO. Teorema Jensen untuk fungsi *concave* mengatakan bahwa untuk sembarang fungsi *concave f(x)*, berlaku pertaksamaan berikut

$$E[f(x)] \leq f[E(x)] \qquad (3.27)$$

Bukti untuk pertaksamaan (3.27) adalah : misalkan *f differentiable*. Fungsi *f concave* jika untuk sembarang titik x_1 dan x_2 berlaku

$$f(x_1) \ \leq f(x_2) + (x_1 - x_2)f'(x_2) \qquad (3.28)$$

Selanjutnya misalkan $x_1 = X$ dan $x_2 = E(X)$, maka pertaksamaan (3.28) dapat dituliskan sebagai

$$f(X) \ \leq f[E(X)] + [X - E(X)]\, f'[E(X)] \qquad (3.29)$$

Karena pertaksamaan (3.29) berlaku untuk semua *X*, maka dapat diterapkan ekspektasi untuk kedua sisi pertaksamaan (3.29) sehingga didapatkan

$$
\begin{aligned}
E[f(X)] \quad &\leq E\{f[E(X)] + [X - E(X)]\, f'[E(X)]\} \\
&\leq E\{f[E(X)]\} + E\{[X - E(X)]\, f'[E(X)]\} \\
&\leq E\{f[E(X)]\} + E\{[X - E(X)]\, \}f'[E(X)] \\
&\leq E\{f[E(X)]\} \\
&\leq f[E(X)] \qquad (3.30)
\end{aligned}
$$

Menggunakan pertaksamaan Jensen (3.30), batas bawah untuk log *p(x)* dapat diturunkan. Menerapkan pertaksamaan (3.30) pada persamaan (3.26) didapatkan pertaksamaan

$$
\begin{aligned}
\log E_z \left[\frac{p(x,z)}{q(z)}\right] \quad &\geq \ E_z \left[\log \frac{p(x,z)}{q(z)}\right] \\
&\geq \ E_z[\log p(x,z)] - \ E_z[\log q(z)] \qquad (3.31)
\end{aligned}
$$

akhirnya didapatkan pada ruas kanan pertaksamaan (3.31) bentuk *Evidence of Lower Bound* (ELBO) atau dikenal juga sebagai *variational free energy*

$$\begin{aligned} ELBO &= E_z\left[\log \frac{p(x,z)}{q(z)}\right] \\ &= E_z[\log p(x,z)] - E_z[\log q(z)] \end{aligned} \tag{3.32}$$

ELBO ini memiliki hubungan dengan divergensi KL yang dapat dituliskan sebagai berikut

$$ELBO = -D_{KL}\big(q(z) \parallel p(z|x)\big) + \log p(x) \tag{3.33}$$

Memaksimumkan ELBO ekivalen dengan meminimumkan divergensi KL.

Eksplorasi mengenai kriteria pemilihan model berdasarkan batas variasional telah banyak dilakukan. Seperti pada model *mixture* (Ueda dan Ghahramani, 2002; McGrory dan Titterington 2007) dan model berbasis *cross-validation* untuk prediksi log densitas. Walaupun pemilihan model menggunakan batas ini cukup baik hasilnya secara praktek, teori mengenai ini belum memberikan pembenaran.

Sebagai catatan, dapat diperhatikan bahwa bentuk bagian pertama ELBO pada (3.32) merupakan ekspektasi dari *complete likelihood*. *Complete likelihood* merupakan fungsi likelihood yang dihitung dengan mempertimbangkan adanya nilai-nilai tak teramati. *Complete likelihood* merupakan bentuk likelihood yang dibahas Dempster (1977). *Complete likelihood* dapat dioptimasi menggunakan algoritma *expectation-maximization* (EM) sedemikian sehingga didapatkan estimasi parameter yang memiliki kualitas sebaik estimasi metode likelihood maksimum (Dempster, 1977).

Prinsip dasar algoritma EM adalah bekerja secara iterative antara menghitung ekspektasi *complete likelihood* berdasarkan posterior $p(z|x)$ (dikenal dengan *E-Step*) lalu mengoptimasi ekspektasi *complete likelihood* ini berdasarkan setiap parameter dalam model (dikenal dengan *M-Step*). Asumsi yang digunakan pada algoritma EM adalah bahwa ekspektasi dibawah distribusi bersyarat $p(z|x)$ dapat dihitung. Hal ini berbeda dengan metode variasional yang bekerja dibawah asumsi bahwa ekspektasi *complete likelihood* pada model distribusi bersyarat $p(z|x)$ tidak dapat dihitung. Selanjutnya dibahas mengenai salah satu jenis keluarga variasional yang sekaligus memberikan contoh penerapan ELBO. Keluarga variasional yang dibahas selanjutnya dikenal dengan nama *variational mean-field* atau *Mean-fiel* variasional.

3. 2. 3. *Mean-field* Variasional

Evidence of lower bound (ELBO) pada persamaan (3.32) digunakan pada prosedur optimasi untuk divergensi Kullback-Leibler (KL). Pertama didefinisikan suatu keluarga variasional Q untuk digunakan pada prosedur optimasi. Tingkat kompleksitas pada keluarga distribusi yang dipilih akan menentukan tingkat kesulitan prosedur optimasi (Blei dkk, 2018). Pembahasan selanjutnya difokuskan pada *mean-field* variasional yang dapat dikelompokkan sebagai bentuk keluarga variasional.

Mean-field adalah metode yang mempelajari model kompleks melalui perilaku distribusi yang lebih sederhana yang mengaproksimasi distribusi sebenarnya pada model. Estimasi model kompleks dilakukan dengan mempartisi distribusi asli dimana masing-masing partisi saling independen. Berikut adalah ilustrasi terkait penerapan metode *mean-field*.

Misalkan $p(\boldsymbol{\theta}|\mathrm{x})$ adalah distribusi yang akan diestimasi oleh $q(\boldsymbol{\theta})$. $\boldsymbol{\theta}=(\theta_1, \theta_2, \dots, \theta_N)^{\mathrm{T}}$ menyatakan vektor parameter. Oleh *mean-field*, $q(\boldsymbol{\theta})$ dipartisi menjadi N distribusi (fungsi distribusi partisi ke-i dinotasikan $q_i(\theta_i)$) yang lebih sederhana dan saling independen sedemikian sehingga :

$$p(\boldsymbol{\theta}|x) \approx q(\boldsymbol{\theta}) = q(\theta_1, \theta_2, \dots, \theta_n) \tag{3.34}$$

$$q(\boldsymbol{\theta}) = \prod_{i=1}^{N} q_i(\theta_i) \tag{3.35}$$

Transisi dari (3.34) ke (3.35) didapatkan dari asumsi bahwa setiap partisinya independent. Selain itu, diasumsikan untuk setiap partisi hanya memiliki satu variabel.

Selanjutnya diturunkan bentuk fungsional dari $q_j(\theta_j)$. Pertama bentuk (3.35) diterapkan pada ELBO (3.32) dengan penyesuaian notasi. Variabel pada partisi ke-i dinotasikan dengan θ_i. $\mathcal{L}[q(\boldsymbol{\theta})]$ menotasikan ELBO

$$\begin{aligned} \mathcal{L}[q(\boldsymbol{\theta})] &= -E_{\boldsymbol{\theta}}[\log q(\boldsymbol{\theta})] + E_{\boldsymbol{\theta}}[\log p(x, \boldsymbol{\theta})] \\ \mathcal{L}[q_1, q_2, \dots, q_n] &= \int \dots \int q(\boldsymbol{\theta}) \log \frac{p(\boldsymbol{\theta}, x)}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \end{aligned} \tag{3.36}$$

$$= \int \dots \int \prod_{i=1}^{N} q_i(\theta_i) \log \frac{p(\boldsymbol{\theta}, x)}{\prod_{j=1}^{N} q_j(\theta_j)} d\theta_1 \dots d\theta_N$$

$$= \int \dots \int \prod_{i=1}^{N} q_i(\theta_i) \, [\log p(\boldsymbol{\theta}, x)$$

$$- \log \prod_{j=1}^{N} q_j(\theta_j)] d\theta_1 \dots d\theta_N \quad (3.37)$$

Bentuk (3.37) merupakan hasil penerapan persamaan (3.35) pada persamaan (3.36). Selanjutnya dilakukan pengaturan untuk mengeluarkan factor $q_k(\theta_k)$ dari setiap bagian didalam integral (3.37)

$$\mathcal{L}[q_1, q_2, \dots, q_n] = \int \dots \int q_k(\theta_k) \prod_{i \neq k} q_i(\theta_i) \, [\log p(\boldsymbol{\theta}, x)$$

$$- \log \prod_{j=1}^{N} q_j(\theta_j)] d\theta_1 \dots d\theta_N$$

$$= \int \dots \int q_k(\theta_k) \prod_{i \neq k} q_i(\theta_i) \log p(\boldsymbol{\theta}, x) \, d\theta_1 \dots d\theta_N$$

$$- \int \dots \int q_k(\theta_k) \prod_{i \neq k} q_i(\theta_i) \log \prod_{j=1}^{N} q_j(\theta_j) \; d\theta_1 \dots d\theta_N$$

$$= \int \dots \int q_k(\theta_k) \prod_{i \neq k} q_i(\theta_i) \log p(\boldsymbol{\theta}, x) \, d\theta_1 \dots d\theta_N$$

$$- \int \dots \int q_k(\theta_k) \prod_{i \neq k} q_i(\theta_i) \sum_{j=1}^{N} \log q_j(\theta_j) \, d\theta_1 \dots d\theta_N \quad (3.38)$$

Untuk menyederhanakan notasi, selanjutnya digunakan definisi

$$E_{m|m \neq k}[\log p(\boldsymbol{\theta}, x)]$$

$$= \int \prod_{i \neq k}^{N} q_i(\theta_i) \log p(\boldsymbol{\theta}, x) d\theta_1 \dots d\theta_{k-1} d\theta_{k+1} \dots d\theta_N \quad (3.39)$$

Menerapkan notasi baru dari (3.39) ke (3.38)

$$\mathcal{L}[q_1, q_2, \dots, q_n] = \int q_k(\theta_k) E_{m|m \neq k}[\log p(\boldsymbol{\theta}, x)] d\boldsymbol{\theta}$$

$$-\int \dots \int q_k(\theta_k) \prod_{i \neq k} q_i(\theta_i) \sum_{j=1}^{N} \log q_j(\theta_j)\, d\theta_1 \dots d\theta_N \quad (3.40)$$

Selanjutnya dilakukan pengaturan pada bagian kedua sisi kanan persamaan pada (3.40)

$$\begin{aligned}
\mathcal{L}[q_1, q_2, \dots, q_n] &= \int q_k(\theta_k) E_{m|m\neq k}[\log p(\boldsymbol{\theta}, x)] d\boldsymbol{\theta} \\
&\quad - \int \dots \int q_k(\theta_k) \prod_{i \neq k} q_i(\theta_i) \log q_k(\theta_k)\, d\theta_1 \dots d\theta_N \\
&\quad - \int \dots \int q_k(\theta_k) \prod_{i \neq k} q_i(\theta_i) \sum_{j \neq k} \log q_j(\theta_j)\, d\theta_1 \dots d\theta_N \\
&= \int q_k(\theta_k) E_{m|m\neq k}[\log p(\boldsymbol{\theta}, x)] d\boldsymbol{\theta} \\
&\quad - \int q_k(\theta_k) \log q_k(\theta_k) \int \prod_{i \neq k} q_i(\theta_i)\, d\theta_1 \dots d\theta_N \\
&\quad - \int q_k(\theta_k) \int \prod_{i \neq k} q_i(\theta_i) \sum_{j \neq k} \log q_j(\theta_j)\, d\theta_1 \dots d\theta_N \quad (3.41) \\
&= \int q_k(\theta_k) E_{m|m\neq k}[\log p(\boldsymbol{\theta}, x)] d\boldsymbol{\theta} \\
&\quad - \int q_k(\theta_k) \log q_k(\theta_k)\, d\theta_k \\
&\quad - \int \prod_{i \neq k} q_i(\theta_i) \sum_{j \neq k} \log q_j(\theta_j)\, d\theta_1 \dots d\theta_{k-1} d\theta_{k+1} \dots d\theta_N \quad (3.42)
\end{aligned}$$

Perubahan yang terjadi dari (3.41) ke (3.42) adalah penyelesaian integral. Selanjutnya didefinisikan

$$\begin{aligned}
G(q_1, \dots, q_{k-1}, q_{k+1}, \dots, q_n) &= \int \prod_{i \neq k} q_i(\theta_i) \sum_{j \neq k} \log q_j(\theta_j) \\
&\quad d\theta_1 \dots d\theta_{k-1} d\theta_{k+1} \dots d\theta_N \quad (3.43)
\end{aligned}$$

Sehingga persamaan (3.42) dapat dituliskan menjadi

$$\begin{aligned}
\mathcal{L}[q_1, q_2, \dots, q_n] &= \int q_k(\theta_k) \{E_{m|m\neq k}[\log p(\boldsymbol{\theta}, x)] - \log q_k(\theta_k)\} d\boldsymbol{\theta} \\
&\quad - G(q_1, \dots, q_{k-1}, q_{k+1}, \dots, q_n) \quad (3.44)
\end{aligned}$$

Persamaan (3.44) menunjukkan bahwa fungsional $\mathcal{L}[q_1, \dots, q_n]$ dapat dituliskan dalam bentuk yang mengandung $q_k(\theta_k)$ dan $E_{m|m\neq k}[\log p(\boldsymbol{\theta}, x)]$ serta bentuk fungsi-fungsi q_i lainnya. Selanjutnya dituliskan bentuk lagrangian berdasarkan persamaan (3.44)

$$L[q_1, \dots, q_n] = \mathcal{L}[q_1, \dots, q_n] - \sum_{i=1}^{N} \lambda_i \int q_i(\theta_i) d\theta_i \tag{3.45}$$

Lagrangian pada (3.45) dibentuk menggunakan *constraints* bahwa untuk setiap $q_i(\theta_i)$ menyatakan fungsi densitas distribusi. Selanjutnya fungsi lagrangian (3.45) diturunkan terhadap $q_k(\theta_k)$

$$\begin{aligned} \frac{\delta L[q_1, \dots, q_n]}{\delta q_k(\theta_k)} &= \frac{\partial}{\partial q_k} \{q_k(\theta_k)\{E_{m|m\neq k}[\log p(\boldsymbol{\theta}, x)] \\ &\qquad - \log q_k(\theta_k)\} - \lambda_k q_k(\theta_k)\} \\ &= E_{m|m\neq k}[\log p(\boldsymbol{\theta}, x)] - \log q_k(\theta_k) - 1 - \lambda_k \end{aligned} \tag{3.46}$$

Kemudian dibangun persamaan turunan pada (3.46) menggunakan persamaan Euler-Lagrange

$$\frac{\delta L[q_1, \dots, q_n]}{\delta q_k(\theta)} = 0$$

$$E_{m|m\neq k}[\log p(\boldsymbol{\theta}, x)] - \log q_k(\theta_k) - 1 - \lambda_k = 0$$

$$\log q_k(\theta_k) = E_{m|m\neq k}[\log p(\boldsymbol{\theta}, x)] - 1 - \lambda_k$$

$$q_k(\theta_k) = \frac{\exp\left\{E_{m|m\neq k}[\log p(\boldsymbol{\theta}, x)]\right\}}{Z_k} \tag{3.47}$$

Dimana Z_k menyatakan konstanta normalisasi untuk distribusi $q_k(\theta_k)$.

Dari penjabaran diatas, didapatkan hasil akhir berupa bentuk fungsional dari penerapan *mean-field* pada divergensi KL. Pada prakteknya, Ketika memasukkan $E_{m|m\neq k}[\log p(\boldsymbol{\theta}, x)]$ yang sering terjadi adalah bentuk pada persamaan (3.47) akan bersesuaian dengan suatu jenis distribusi (Keng, 2017).

Sebagai implikasi dari penjabaran diatas, oleh Keng (2017) dituliskan prosedur iterative dalam algoritma optimasi sebagai berikut :

1) Mulai dengan sembarang nilai untuk setiap parameter dalam fungsi $q_k(\theta_k)$

2) Untuk setiap q_k, gunakan persamaan (3.47) untuk meminimumkan divergensi KL dengan memperbaharui $q_k(\theta_k)$ (dengan menganggap partisi lainnya konstan)
3) Ulangi proses hingga mencapai kriteria konvergensi yang diinginkan

Prosedur diatas juga disebutkan dalam Blei dkk (2018) dan disebut sebagai *coordinate ascent variational inference* (CAVI).

3. 2. 4. *Coordinate Ascent* pada *Mean-field* Variasional

Telah diberikan prosedur iteratif *coordinate ascent variational inference* (CAVI) yang digunakan untuk meminimumkan divergensi KL melalui proses memaksimumkan ELBO. Prosedur iteratif ini memaksimumkan ELBO hingga suatu titik maksimum local (Blei dkk, 2018). Bentuk algoritma secara formal dapat dituliskan sebagai berikut :

Algoritma 1 : *Coordinate ascent variational inference*

Masukan : Model *p(x,z)*; dataset *X*

Keluaran : Fungsi densitas variasional

$$q(\boldsymbol{\theta}) = \prod_{k=1}^{N} q_k(\theta_k)$$

Inisialisasi : Faktor-faktor variasional $q_k(\theta_k)$

$while\ ELBO\ has\ not\ converged\ do$

$|\ for\ j\ \in \{1,2,\dots,N\}\ do$

$|\ |\ define\ \exp\{E_{m|m\neq k}[\log p(\boldsymbol{\theta}, x)]\}$

$|\ |\ set\ q_k(\theta_k) = \frac{\exp\{E_{m|m\neq k}[\log p(\boldsymbol{\theta}, x)]\}}{Z_k}$

$|\ end$

$|\ compute\ ELBO(q) = \mathbb{E}[\log p(x, \boldsymbol{\theta})] - \mathbb{E}[\log q(\boldsymbol{\theta})]$

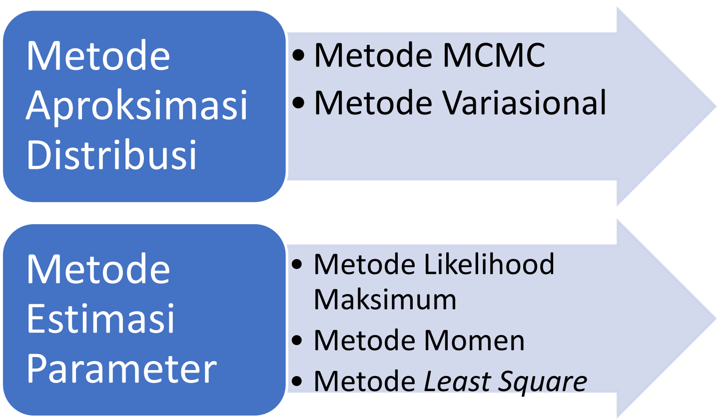
end

$return\ q(\boldsymbol{\theta})$

Prosedur iteratif CAVI pada algoritma 1 akan memuncak hingga suatu titik maksimum lokal. Beberapa aplikasi dari CAVI yang pernah dilakukan seperti pada *variational message passing* (VMP) oleh Winn dan Bishop (2005) dan VMP pada model non konjugasi oleh Knowles dan Minka (2011).

Hubungan CAVI juga cukup erat dengan metode inferensi sampling Gibbs (Geman dan Geman, 1984; Gelfan dan Smith, 2000). Pada Teknik sampling Gibbs, fokus yang menjadi perhatian adalah bentuk realisasi dari nilai-nilai tak teramati serta secara iteratif melakukan pengambilan sampel untuk setiap distribusi bersyarat dari variabel. Sedangkan pada CAVI, fokusnya adalah menggunakan bentuk (3.47) dan secara iteratif memperbarui setiap factor variasional pada fungsi objektif (Blei dkk, 2018).

Pembahasan mengenai *mean-field* variasional yang telah diselesaikan bertujuan untuk memberikan gambaran bagaimana penerapan metode variasional dilakukan. Pada praktiknya, penerapan metode variasional akan memerlukan penyesuaian dengan model yang digunakan serta ukuran data. Untuk menutup pembahasan mengenai metode variasional, gambar 3.3 mengilustrasikan pengelompokan metode berdasarkan jenis target yang diestimasi.



Gambar 3.3 Pengelompokan Metode Berdasarkan Jenis Target yang Diestimasi

3. 3. Teori pada Estimasi Menggunakan Metode Variasional

Untuk mengakhiri bab ini, selanjutnya dibahas mengenai sifat estimasi metode variasional. Jika diperhatikan, prosedur yang dilakukan pada metode variasional dapat disejajarkan sebagai aproksimasi untuk suatu estimasi (Peyrard dkk, 2018). Hal ini dapat dideduksi dengan memperhatikan algoritma 1 yang memaksimumkan ELBO sehingga divergensi KL minimum. Divergensi KL $D_{KL}\left(q(z) \parallel p(z|x)\right)$ menyatakan seberapa besar informasi yang hilang Ketika distribusi $q(z)$ digunakan sebagai wakil untuk $p(z|x)$. Sehingga, proses meminimumkan divergensi KL dapat dianggap sebagai melakukan aproksimasi untuk suatu distribusi $p(z|x)$ menggunakan $q(z)$ sehingga informasi yang hilang minimum (Blei, 2017). Oleh karena itu, estimasi variasional hanyalah sebuah aproksimasi untuk estimasi dari metode likelihood maksimum (Peyrard, 2018).

Dari sudut pandang statistika, sifat dari parameter yang diestimasi metode variasional tidak dapat disamakan dengan sifat estimasi dari metode likelihood maksimum. Sifat estimasi pada metode variasional hanya terbatas untuk model yang diterapkan. Seperti contoh pada model Poisson log-normal (Hall dkk, 2011), model linear bayesian (You dkk, 2014), model blok stokastik (Celisse dkk, 2012; Bickel dkk, 2013).

Hall dkk (2011) menunjukkan adanya konsistensi pada estimasi variasional untuk model Poisson log-normal. Hasil teoritis mengenai sifat estimasi variasional yang dibahas adalah adanya similaritas log-likehood dengan batas bawahnya. Selain itu, dibahas juga mengenai teorema batas nilai estimasi parameter ketika memenuhi syarat tertentu. Serta ditunjukkan bahwa estimasi memiliki normalitas asimtotik.

You dkk (2014) memfokuskan pada posterior variasional untuk model linear Bayesian dengan menggunakan prior normal dan invers gamma. Didapatkan sifat konsisten pada mean parameter untuk posterior variasional *mean-field*. Konsistensi ini terjadi pada kondisi pengaturan standar.

Celisse dkk (2012) dan Bickel dkk (2013) melakukan penelitian pada data *network*. Ditemukan bahwa estimasi parameter oleh variasional *mean-field* untuk model blok stokastik menunjukkan normalitas asimtotik. Selain itu, ditemukan bahwa metode variasional cenderung lebih efisien dalam proses iteratifnya. Celisse dkk menyatakan secara teoritis aproksimasi varisional yang diterapkan pada model dan kondisi yang sama, akan mendekati hasil estimasi dari metode likelihood maksimum.

Selain ketiga contoh penerapan pada model-model diatas, terdapat jurnal yang meringkas capaian penelitian mengenai metode variasional seperti pada Blei dkk (2018) dan Peyrard dkk (2018). Ditambah penerapan pada model regresi logistik oleh Jaakkola dan Jordan (1998) yang menemukan bahwa estimasi parameter oleh metode variasional sangat akurat.

BAB IV
ESTIMASI PARAMETER PADA REGRESI LOGISTIK SPASIAL MENGGUNAKAN METODE VARIASIONAL

Bab ini berisi pembahsan mengenai penerapan metode variasional untuk estimasi parameter pada regresi logistik spasial.

4. 1. Regresi Logistik Spasial

Layaknya regresi logistik biasa, regresi logistik spasial membangun persamaan antara variabel independen dengan variabel dependen melalui fungsi logistik

$$f(y) = \frac{e^y}{1 + e^y} \tag{4.1}$$

Melalui persamaan logistik (4.1) model regresi logistik untuk data yang memiliki dependensi spasial dikembangkan. Pembentukan model selanjutnya akan ditulis menggunakan notasi dari Hardouin (2019) dengan kerangka kerja untuk pemodelan hierarkis. Selain itu, untuk prosedur estimasi parameter tidak sepenuhnya menggunakan metode bayesian. Tidak sepenuhnya menggunakan metode bayesian, maksudnya adalah tidak menerapkan asumsi distribusi tertentu pada parameter dalam model.

4. 1. 1. Model Regresi Logistik Spasial

Misalkan terdapat suatu ruang domain dua dimensi $T \equiv \{\boldsymbol{s}_i\ : i = 1,2,\dots,n\} \subset \boldsymbol{R}^2$ dimana $\boldsymbol{s}_i = (s_{i1}, s_{i2})$. Selanjutnya dimisalkan proses acak terjadi pada T yang dinyatakan dengan $\boldsymbol{Y}^* = (Y^*(\boldsymbol{s}_1), Y^*(\boldsymbol{s}_2), \dots, Y^*(\boldsymbol{s}_n))$. Nilai-nilai realisasi yang mungkin untuk variabel $Y^*(.)$ adalah 0 atau 1. Selanjutnya variabel $Y^*(.)$ dimodelkan dengan distribusi Bernoulli yang nilai meannya bergantung pada *hidden process* $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \dots, Y(\boldsymbol{s}_n))$. Oleh Hardouin (2019) variabel $Y^*(.)$ diasumsikan saling independen bersyarat proses spasial $\boldsymbol{Y}$. Oleh karena itu, distribusi variabel-variabel pada $\boldsymbol{Y}^*$ bersyarat $\boldsymbol{Y}$ untuk setiap $\boldsymbol{s} \in \boldsymbol{T}$ dapat dituliskan sebagai

$$Y^*(\boldsymbol{s}) | Y(\boldsymbol{s}) \sim Bernoulli(p(\boldsymbol{s})) \tag{4.2}$$

dengan

$$p(\boldsymbol{s}) = \frac{\exp(Y(\boldsymbol{s}))}{1 + \exp(Y(\boldsymbol{s}))} \tag{4.3}$$

sehingga didapatkan

$$
\begin{aligned}
p\left(Y^*(\boldsymbol{s}) = z \,\middle|\, Y(\boldsymbol{s})\right) &= p(\boldsymbol{s})^z\left(1 - p(\boldsymbol{s})\right)^{1-z} \\
&= \frac{1}{1 + \exp[-Y(\boldsymbol{s})(2z - 1)]}
\end{aligned}
\tag{4.4}
$$

Hidden process $\boldsymbol{Y}$ dimodelkan dengan

$$
Y(\boldsymbol{s}) = \boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta} + \varepsilon(\boldsymbol{s}) \tag{4.5}
$$

Pada persamaan (4.5) notasi $\boldsymbol{X}(\boldsymbol{s})$ menyatakan matriks untuk q kovariat sedemikian sehingga $\boldsymbol{X}(\boldsymbol{s}) = \left(X_1(\boldsymbol{s}), X_2(\boldsymbol{s}), \dots, X_q(\boldsymbol{s})\right)^T$, $\boldsymbol{\beta}$ menyatakan vektor ukuran q untuk koefisien regresi, dan $\varepsilon(\boldsymbol{s})$ menyatakan variasi spasial yang tidak diamati. Selain itu proses $\boldsymbol{\varepsilon} = (\varepsilon(\boldsymbol{s}_1), \varepsilon(\boldsymbol{s}_2), \dots, \varepsilon(\boldsymbol{s}_n))$ diasumsikan berdistribusi normal multivariat

$$
\boldsymbol{\varepsilon} \sim MN(\boldsymbol{0}, \boldsymbol{\Sigma}) \tag{4.6}
$$

dengan $\boldsymbol{\Sigma}$ adalah matriks kovariansi spasial yang tidak diketahui. Oleh Hardouin (2019) bentuk matriks kovariansi pada (4.6) dapat juga dilihat sebagai fungsi $\boldsymbol{\Sigma} = \sigma^2 Q(\boldsymbol{\theta})$ dimana $Q(\boldsymbol{\theta})$ adalah fungsi kovariansi dengan vektor parameter $\boldsymbol{\theta}$. Prosedur estimasi akan berfokus untuk mengestimasi parameter $\boldsymbol{\beta}$ dan $\boldsymbol{\Sigma}$.

4. 1. 2. Fungsi *complete log-likelihood* Regresi Logistik Spasial

Untuk notasi selanjutnya dimisalkan *[U|V]* sebagai distribusi *U* bersyarat *V*. Sesuai dengan yang telah disebutkan sebelumnya, kerangka kerja yang dipakai adalah hierarkis dengan mempertimbangkan *hidden process*. Prosedur estimasi selanjutnya menggunakan *complete log-likelihood*. Selain itu, parameter pada model tidak diasumsikan memiliki distribusi tertentu.

Untuk membentuk fungsi *complete log-likelihood* L_c, dibentuk distribusi dari variabel teramati dan tak teramati. Dari penjelasan pada subbab 4.1.1, diketahui data *complete* yang menjadi perhatian terdiri atas $\boldsymbol{Y}^*$ yang merupakan nilai teramati dan $\boldsymbol{\varepsilon}$ yang menyatakan hal-hal yang tidak teramati. Oleh karena itu, distribusi untuk data *complete* dapat dinotasikan dengan

$$
[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}, \boldsymbol{\Sigma}] \tag{4.7}
$$

Kemudian menggunakan aturan rantai untuk fungsi distribusi pada (4.7), didapatkan

$$
[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}, \boldsymbol{\Sigma}] = [\boldsymbol{Y}^* \,|\boldsymbol{\varepsilon}, \boldsymbol{\beta}] \times [\boldsymbol{\varepsilon} \,|\boldsymbol{\Sigma}] \tag{4.8}
$$

Selanjutnya menggunakan dekomposisi pada (4.8), fungsi *complete log-likelihood* L_c dibentuk untuk model regresi logistik spasial sebagai berikut

$$\begin{aligned} L_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}] &= \log[\boldsymbol{Y}^* \,|\boldsymbol{\varepsilon}, \boldsymbol{\beta}] \; + \log[\boldsymbol{\varepsilon} \,|\boldsymbol{\Sigma}] \\ &= -\sum_{\boldsymbol{s}\in\boldsymbol{T}} \log\left(1 + e^{Y(\boldsymbol{s})}\right) + \sum_{\boldsymbol{s}\in\boldsymbol{T}} Y(\boldsymbol{s})Y^*(\boldsymbol{s}) \\ &\qquad - \frac{n}{2}\log 2\pi - \frac{1}{2}[\log(\det \boldsymbol{\Sigma}) + \boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{\Sigma}^{-\mathbf{1}}\boldsymbol{\varepsilon}] \end{aligned} \quad (4.9)$$

dimana $Y(\boldsymbol{s})$ sesuai dengan persamaan (4.5).

4. 2. Estimasi Parameter pada Regresi Logistik Spasial Menggunakan Metode Variasional

Complete log-likelihood pada (4.9) digunakan untuk mencari estimasi parameter metode likelihood maksimum. Prosedur yang digunakan untuk menemukan estimasi parameter menggunakan fungsi *complete log-likelihood* adalah algoritma *expectation maximization* (EM). Algoritma EM diajukan pertama kali oleh Dempster (1977). Ide utama algoritma EM adalah prosedur iteratif dimana dalam setiap iterasinya dilakukan penghitungan ekspektasi dari fungsi *complete log-likelihood*, kemudian memaksimumkan ekspektasi dari fungsi *complete log-likelihood* berdasarkan parameter dalam fungsi. Untuk detail mengenai algoritma EM dapat dilihat seperti pada McLachlan dan Krishnan (2008).

Pada kasus spesifik untuk fungsi *complete log-likelihood* (4.9), algoritma EM akan terkendala pada saat menghitung ekspektasi fungsi *complete log-likelihood*

$$E\{L_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}]\big|\boldsymbol{Y}^*, \widehat{\boldsymbol{\beta}}\,, \widehat{\boldsymbol{\Sigma}}\}. \quad (4.10)$$

Mengutip dari Hardouin (2019), fungsi ekspektasi (4.10) tidak tersedia dalam bentuk *closed-form*. Pembahasan untuk solusi masalah ini telah banyak diajukan. McLachlan dan Krishnan (2008) menggunakan integrasi Monte Carlo dan simulasi pada $\boldsymbol{\varepsilon}$. Pemilihan densitas yang "sesuai" juga diajukan sebagai alternatif oleh Robert dan Rosenthal (2001), dimana masalah yang paling dikritik adalah ketika metode ini diaplikasikan ke data berukuran besar, komputasi menjadi lambat. Serta alternatif yang banyak digunakan yaitu aproksimasi Laplace. Alternatif ini bekerja menggunakan ekspansi deret taylor orde kedua pada logaritma dari integral. Untuk aplikasi alternatif menggunakan aproksimasi Laplace dapat dilihat pada Sengupta dan Cressie (2013). Metode yang digunakan pada penelitian ini diturunkan dari aproksimasi pada fungsi logistik (4.1) yang telah diberikan oleh Jaakkola dan Jordan (1998).

Seperti yang telah dijabarkan pada subbab 3.2, secara umum metode variasional mengaproksimasi suatu fungsi distribusi dengan menyisipkan suatu parameter baru yang dikenal sebagai parameter variasional. Oleh Jaakkola dan Jordan (1998) : penerapan metode variasional pada suatu permasalahan, untuk suatu nilai parameter variasional, dapat menghasilkan bentuk solusi yang *closed-form*. Namun, seperti pembahasan pada subbab 3.3, sifat dari hasil estimasi menggunakan metode variasional tidak memiliki teori yang berlaku secara umum. Hal ini karena inferensi variasional bekerja dengan aproksimasi dari distribusi permasalahan awal, atau ekuivalen dengan mengaproksimasi menggunakan aproksimasi fungsi distribusi (Peyrard, 2018). Pembahasan sifat hasil estimasi metode variasional untuk model regresi logistik spasial berdasarkan hasil simulasi dari Hardouin (2019).

4. 2. 1. Fungsi Objektif

Pada kasus spesifik untuk estimasi fungsi *complete log-likelihoodi* pada (4.9), digunakan batas bawah untuk fungsi logistik yang diberikan oleh Jaakola dan Jordan (1998). Untuk mendapatkan batas bawah ini, pertama dipertimbangkan log dari fungsi logistik pada (4.1)

$$\begin{aligned}\log f(y) \quad &= \log \frac{e^y}{1+e^y} \\ &= \log \frac{1}{1+e^{-y}} \\ &= -\log(1+e^{-y}) \end{aligned} \tag{4.11}$$

Selanjutnya dilakukan simetrisasi pada fungsi tersebut. Pendekatan yang dilakukan adalah mencari transformasi variasional yang mempertimbangkan kombinasi bentuk gaussian. Mengikuti bentuk transformasi untuk fungsi (4.11) oleh Jaakkola dan Jordan (1998) didapatkan

$$-\log(1+e^{-y}) \quad = \frac{y}{2} - \log\left[\exp\left(-\frac{y}{2}\right) + \exp\left(\frac{y}{2}\right)\right]$$

$$\log f(y) \qquad = \frac{y}{2} - \log\left[\exp\left(-\frac{y}{2}\right) + \exp\left(\frac{y}{2}\right)\right] \tag{4.12}$$

Ruas kanan suku kedua pada (4.12) didefinisikan sebagai berikut

$$g(y) = \log\left[\exp\left(-\frac{y}{2}\right) + \exp\left(\frac{y}{2}\right)\right] \tag{4.13}$$

$g(y)$ adalah fungsi konveks relatif terhadap variabel y^2 (hal ini dapat ditunjukkan menggunakan aturan turunan kedua). Karena fungsi (4.13) merupakan fungsi konvex, maka dapat digunakan permukaan garis singgung fungsi (4.13) sebagai batas bawah global untuk fungsi tersebut. Ekspansi taylor terhadap variabel y^2 diterapkan pada fungsi (4.13), didapatkan pertaksamaan berikut

$$\begin{aligned} g(y) \quad &\geq g(\tau) + \frac{\partial g(\tau)}{\partial(\tau^2)}(y^2 - \tau^2) \\ &= \log\left[\exp\left(-\frac{\tau}{2}\right) + \exp\left(\frac{\tau}{2}\right)\right] + \frac{\exp(\tau) - 1}{4\tau[\exp(\tau) + 1]}(y^2 - \tau^2) \\ &= \frac{\tau}{2} - \log f(\tau) + \frac{1}{4\tau}\tanh\left(\frac{\tau}{2}\right)(y^2 - \tau^2) \end{aligned} \tag{4.14}$$

Dengan menerapkan 4.14 ke bentuk 4.12, didapatkan batas bawah logaritma fungsi logistik sesuai dengan Hardouin (2019)

$$\log f(y) \quad \geq \frac{y - \tau}{2} + \log f(\tau) - \lambda(\tau)(y^2 - \tau^2) \tag{4.15}$$

Dengan $\lambda(\tau) = \frac{f(\tau) - 1/2}{2\tau} = \frac{1}{4\tau}\tanh\left(\frac{\tau}{2}\right)$. Batas bawah pada pertaksamaan (4.15) eksak ketika $y^2 = \tau^2$ (Hardouin, 2019).

Prosedur estimasi dilanjutkan dengan menerapkan pertaksamaan (4.15) ke bentuk $-\log\left(1 + e^{Y(\boldsymbol{s})}\right) = \log f[-Y(\boldsymbol{s})]$ untuk setiap $\boldsymbol{s} \in T$. Selanjutnya diperkenalkan vektor $\boldsymbol{\tau} = (\tau(\boldsymbol{s}_1), \tau(\boldsymbol{s}_2), \dots, \tau(\boldsymbol{s}_n))$ yang merupakan parameter variasional. Didapatkan pertaksamaan (4.16)

$$\begin{aligned} \log f[-Y(\boldsymbol{s})] \quad &\geq \frac{[-Y(\boldsymbol{s})] - \tau(\boldsymbol{s})}{2} + \log f\big(\tau(\boldsymbol{s})\big) \\ &\qquad -\lambda(\tau(\boldsymbol{s}))([Y(\boldsymbol{s})]^2 - \tau(\boldsymbol{s})^2) \end{aligned} \tag{4.16}$$

Sehingga dapat diperoleh batas bawah untuk bagian awal fungsi *complete log-likelihood*

$$-\sum_{\boldsymbol{s}\in\boldsymbol{T}}\log\left(1+e^{Y(\boldsymbol{s})}\right)+\sum_{\boldsymbol{s}\in\boldsymbol{T}}Y(\boldsymbol{s})Y^*(\boldsymbol{s}) \quad \geq \sum_{\boldsymbol{s}\in\boldsymbol{T}}\frac{[-Y(\boldsymbol{s})]-\tau(\boldsymbol{s})}{2} + \sum_{\boldsymbol{s}\in\boldsymbol{T}}\log f\left(\tau(\boldsymbol{s})\right) - \sum_{\boldsymbol{s}\in\boldsymbol{T}}\lambda(\tau(\boldsymbol{s}))([Y(\boldsymbol{s})]^2-\tau(\boldsymbol{s})^2) + \sum_{\boldsymbol{s}\in\boldsymbol{T}}Y(\boldsymbol{s})Y^*(\boldsymbol{s}) \quad (4.17)$$

Untuk menyederhanakan notasi, batas bawah pada (4.17) selanjutnya akan dilakukan penataan ulang. Batas bawah pada (4.17) dinotasikan dengan S_c. Diterapkan persamaan (4.5) ke (4.17)

$$\begin{aligned} S_c &= \sum_{\boldsymbol{s}\in\boldsymbol{T}}\frac{[-Y(\boldsymbol{s})]-\tau(\boldsymbol{s})}{2}+\sum_{\boldsymbol{s}\in\boldsymbol{T}}\log f\left(\tau(\boldsymbol{s})\right) \\ &\quad - \sum_{\boldsymbol{s}\in\boldsymbol{T}}\lambda(\tau(\boldsymbol{s}))([Y(\boldsymbol{s})]^2-\tau(\boldsymbol{s})^2) + \sum_{\boldsymbol{s}\in\boldsymbol{T}}Y(\boldsymbol{s})Y^*(\boldsymbol{s}) \\ &= \sum_{\boldsymbol{s}\in\boldsymbol{T}}\frac{-[\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}+\varepsilon(\boldsymbol{s})]-\tau(\boldsymbol{s})}{2}+\sum_{\boldsymbol{s}\in\boldsymbol{T}}\log f\left(\tau(\boldsymbol{s})\right) \\ &\quad - \sum_{\boldsymbol{s}\in\boldsymbol{T}}\lambda(\tau(\boldsymbol{s}))([\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}+\varepsilon(\boldsymbol{s})]^2-\tau(\boldsymbol{s})^2) \\ &\quad + \sum_{\boldsymbol{s}\in\boldsymbol{T}}[\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}+\varepsilon(\boldsymbol{s})]Y^*(\boldsymbol{s}) \end{aligned} \quad (4.18)$$

Kemudian setiap komponen yang hanya mengandung komponen parameter variasional τ dikumpulkan menjadi suku pertama persamaan (4.19)

$$\begin{aligned} S_c &= \sum_{\boldsymbol{s}\in\boldsymbol{T}}\left\{\frac{-\tau(\boldsymbol{s})}{2}+\log f\left(\tau(\boldsymbol{s})\right)+\lambda\left(\tau(\boldsymbol{s})\right)\tau(\boldsymbol{s})^2\right\} \\ &\quad + \sum_{\boldsymbol{s}\in\boldsymbol{T}}\left\{\frac{-[\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}+\varepsilon(\boldsymbol{s})]}{2}-\lambda\left(\tau(\boldsymbol{s})\right)([\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}+\varepsilon(\boldsymbol{s})]^2)\right\} \\ &\quad + \sum_{\boldsymbol{s}\in\boldsymbol{T}}\{[\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}+\varepsilon(\boldsymbol{s})]Y^*(\boldsymbol{s})\} \end{aligned} \quad (4.19)$$

Suku kedua persamaan (4.20) adalah bentuk fungsi yang hanya mengandung τ dan $\boldsymbol{\beta}$. Suku ketiga persamaan (4.20) adalah bentuk fungsi yang mengandung τ, $\boldsymbol{\beta}$, dan ε.

$$\begin{aligned} S_c \;=\; & \sum_{s\in T}\left\{\frac{-\tau(\boldsymbol{s})}{2}+\log f\big(\tau(\boldsymbol{s})\big)+\lambda\big(\tau(\boldsymbol{s})\big)\tau(\boldsymbol{s})^2\right\} \\ & +\sum_{s\in T}\left\{\frac{-\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}}{2}-\lambda\big(\tau(\boldsymbol{s})\big)[\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}]^2+Y^*(\boldsymbol{s})\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}\right\} \\ & +\sum_{s\in T}\left\{\frac{-\varepsilon(\boldsymbol{s})}{2}-\lambda\big(\tau(\boldsymbol{s})\big)[2\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}\varepsilon(\boldsymbol{s})+\varepsilon(\boldsymbol{s})^2]+[\varepsilon(\boldsymbol{s})]Y^*(\boldsymbol{s})\right\} \end{aligned} \qquad (4.20)$$

Setelah dilakukan penataan ulang pada persamaan (4.20), didefinisikan fungsi untuk setiap suku

$$T_1(\boldsymbol{\tau}) = \sum_{s\in T}\left\{\frac{-\tau(\boldsymbol{s})}{2}+\log f\big(\tau(\boldsymbol{s})\big)+\lambda\big(\tau(\boldsymbol{s})\big)\tau(\boldsymbol{s})^2\right\} \qquad (4.21)$$

$$T_2(\boldsymbol{\tau},\boldsymbol{\beta}) = \sum_{s\in T}\left\{\frac{-\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}}{2}-\lambda\big(\tau(\boldsymbol{s})\big)[\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}]^2+Y^*(\boldsymbol{s})\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}\right\} \qquad (4.22)$$

$$T_3(\boldsymbol{\tau},\boldsymbol{\beta},\boldsymbol{\Sigma}) = \sum_{s\in T}\left\{\begin{matrix}\frac{-\varepsilon(\boldsymbol{s})}{2}-\lambda\big(\tau(\boldsymbol{s})\big)[2\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}\varepsilon(\boldsymbol{s})+\varepsilon(\boldsymbol{s})^2] \\ +[\varepsilon(\boldsymbol{s})]Y^*(\boldsymbol{s})\end{matrix}\right\} \qquad (4.23)$$

Notasi (4.21), (4.22), dan (4.23) diterapkan pada batas bawah *complete log-likelihood* (4.9) sebagai berikut

$$\begin{aligned} L_c[\boldsymbol{Y}^*,\boldsymbol{\varepsilon}\,|\boldsymbol{\beta},\boldsymbol{\Sigma}] & \geq T_1(\boldsymbol{\tau})+T_2(\boldsymbol{\tau},\boldsymbol{\beta})+T_3(\boldsymbol{\tau},\boldsymbol{\beta},\boldsymbol{\Sigma}) \\ & \quad -\frac{n}{2}\log 2\pi-\frac{1}{2}[\log(\det\boldsymbol{\Sigma})+\boldsymbol{\varepsilon}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}] \\ & = T_1(\boldsymbol{\tau})+T_2(\boldsymbol{\tau},\boldsymbol{\beta})+T_3(\boldsymbol{\tau},\boldsymbol{\beta},\boldsymbol{\Sigma}) \\ & \quad -\frac{1}{2}[\log(\det\boldsymbol{\Sigma})+\boldsymbol{\varepsilon}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}]+c \end{aligned} \qquad (4.24)$$

c menotasikan konstanta $-\frac{n}{2}\log 2\pi$. Batas bawah pada persamaan (4.24) selanjutnya dinotasikan

$$\begin{aligned} \tilde{L}_c[\boldsymbol{Y}^*,\boldsymbol{\varepsilon}\,|\boldsymbol{\beta},\boldsymbol{\Sigma},\boldsymbol{\tau}] & = T_1(\boldsymbol{\tau})+T_2(\boldsymbol{\tau},\boldsymbol{\beta})+T_3(\boldsymbol{\tau},\boldsymbol{\beta},\boldsymbol{\Sigma}) \\ & \quad -\frac{1}{2}[\log(\det\boldsymbol{\Sigma})+\boldsymbol{\varepsilon}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}]+c \end{aligned} \qquad (4.25)$$

Sebagai catatan, batas bawah (4.25) akan eksak ketika $\tau(\boldsymbol{s})^2 = Y(\boldsymbol{s})^2$ untuk setiap $\boldsymbol{s}\in\boldsymbol{T}$.

Prosedur estimasi mempertimbangkan variabel tak teramati sehingga *complete log-likelihood* digunakan, algoritma EM diperlukan untuk memaksimumkan fungsi (4.25). Pada setiap iterasi pada algoritma EM, dihitung ekspektasi *complete log-*

likelihood bersyarat data teramati sesuai dengan persamaan (4.10). Oleh karena itu, selanjutnya dicari distribusi bersyarat data teramati atau lebih tepatnya

$$[\boldsymbol{\varepsilon}|\boldsymbol{Y}^*, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau}] \tag{4.26}$$

Pertama diperkenalkan vektor $\boldsymbol{M} = \left(M(\boldsymbol{s_1}), M(\boldsymbol{s_2}), \ldots, M(\boldsymbol{s_n})\right)^T$ dimana

$$M(\boldsymbol{s}) = Y^*(\boldsymbol{s}) - \frac{1}{2} - 2\lambda\big(\tau(\boldsymbol{s})\big)\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta} \tag{4.27}$$

merupakan bagian dari $T_3(\boldsymbol{\tau}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$ sehingga jika diterapkan pada (4.23) didapatkan notasi bentuk matriks sebagai berikut

$$\begin{aligned} T_3(\boldsymbol{\tau}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) &= \sum_{s \in T} \left\{ \begin{matrix} \dfrac{-\varepsilon(\boldsymbol{s})}{2} - \lambda\big(\tau(\boldsymbol{s})\big)[2\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}\varepsilon(\boldsymbol{s}) + \varepsilon(\boldsymbol{s})^2] \\ +[\varepsilon(\boldsymbol{s})]Y^*(\boldsymbol{s}) \end{matrix} \right\} \\ &= \sum_{s \in T} \left\{ \begin{matrix} -\lambda\big(\tau(\boldsymbol{s})\big)\varepsilon(\boldsymbol{s})^2 \\ +\varepsilon(\boldsymbol{s}) \left[Y^*(\boldsymbol{s}) - \dfrac{1}{2} - \lambda\big(\tau(\boldsymbol{s})\big)[2\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}] \right] \end{matrix} \right\} \\ &= -\boldsymbol{\varepsilon}^T\boldsymbol{\Lambda}(\boldsymbol{\tau})\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^T\boldsymbol{M} \end{aligned} \tag{4.28}$$

Dengan $\boldsymbol{\Lambda}(\boldsymbol{\tau})$ adalah matriks diagonal dengan elemen-elemen diagonal utamanya adalah $\lambda\big(\tau(\boldsymbol{s})\big)$. Bentuk baru (4.28) kemudian diterapkan ke (4.25) sehingga didapatkan

$$\begin{aligned} \tilde{L}_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau}] &= T_1(\boldsymbol{\tau}) + T_2(\boldsymbol{\tau}, \boldsymbol{\beta}) - \boldsymbol{\varepsilon}^T\boldsymbol{\Lambda}(\boldsymbol{\tau})\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^T\boldsymbol{M} \\ &\qquad - \frac{1}{2}[\log(\det\boldsymbol{\Sigma}) + \boldsymbol{\varepsilon}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}] + c \\ &= T_1(\boldsymbol{\tau}) + T_2(\boldsymbol{\tau}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}^T\boldsymbol{M} \\ &\qquad - \frac{1}{2}[\log(\det\boldsymbol{\Sigma}) + 2\boldsymbol{\varepsilon}^T\boldsymbol{\Lambda}(\boldsymbol{\tau})\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}] + c \end{aligned} \tag{4.29}$$

Selanjutnya komponen kuadratik $\boldsymbol{\varepsilon}$ disatukan melalui pendefinisian $\boldsymbol{W}^{-1} = \boldsymbol{\Sigma}^{-1} + 2\boldsymbol{\Lambda}(\boldsymbol{\tau})$, sehingga bentuk persamaan (4.29) menjadi

$$\begin{aligned} \tilde{L}_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau}] &= T_1(\boldsymbol{\tau}) + T_2(\boldsymbol{\tau}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}^T\boldsymbol{M} \\ &\qquad - \frac{1}{2}[\log(\det\boldsymbol{\Sigma}) + \boldsymbol{\varepsilon}^T\boldsymbol{W}^{-1}\boldsymbol{\varepsilon}] + c \end{aligned} \tag{4.30}$$

Untuk mendapatkan distribusi (4.26) digunakan fakta bahwa untuk nilai dalam vektor $\boldsymbol{\tau}$ yang ditetapkan, distribusi tersebut akan proporsional dengan distribusi $[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau}]$ (Hardouin, 2019). Oleh karena itu, ruas kanan persamaan (4.30) dieksponensial.

$$\begin{aligned}\exp(\tilde{L}_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}, \boldsymbol{\tau}]) \quad &= \exp[T_1(\boldsymbol{\tau}) + T_2(\boldsymbol{\tau}, \boldsymbol{\beta})] \times \exp\ (\boldsymbol{\varepsilon}^T \boldsymbol{M}) \\ &\quad \times \exp\left(-\frac{1}{2}\boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{W}^{-1}\boldsymbol{\varepsilon}\right) \times \frac{1}{\sqrt{(\det \boldsymbol{\Sigma})}} \\ &= \exp[T_1(\boldsymbol{\tau}) + T_2(\boldsymbol{\tau}, \boldsymbol{\beta})] \times \frac{1}{\sqrt{(\det \boldsymbol{\Sigma})}} \\ &\quad \times \exp\left(-\frac{1}{2}\boldsymbol{\varepsilon}^T\boldsymbol{W}^{-1}\boldsymbol{\varepsilon} + \ \boldsymbol{\varepsilon}^T\boldsymbol{M}\right) \qquad (4.31)\end{aligned}$$

Bagian konstanta c pada (4.30) diabaikan. Untuk mentransformasi (4.31) ke bentuk yang proporsional dengan distribusi multivariat normal, didefinisikan $\boldsymbol{\mu} = \boldsymbol{W}\boldsymbol{M}$ yang akan diterapkan pada (4.31). Selain itu, bentuk yang bernilai 1 dikalikan pada (4.31) didapatkan

$$\begin{aligned}\exp(\tilde{L}_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}, \boldsymbol{\tau}]) \quad &= \exp[T_1(\boldsymbol{\tau}) + T_2(\boldsymbol{\tau}, \boldsymbol{\beta})] \\ &\quad \times \frac{1}{\sqrt{(\det \boldsymbol{\Sigma})}} \times \exp\left(-\frac{1}{2}\boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{W}^{-1}\boldsymbol{\varepsilon} + \ \boldsymbol{\varepsilon}^T\boldsymbol{M}\right) \\ &\quad \times \exp\left(\frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{W}^{-1}\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{W}^{-1}\boldsymbol{\mu}\right) \\ &= \exp[T_1(\boldsymbol{\tau}) + T_2(\boldsymbol{\tau}, \boldsymbol{\beta}) + \frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{W}^{-1}\boldsymbol{\mu}] \\ &\quad \times \exp\left[-\frac{1}{2}(\boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{W}^{-1}\boldsymbol{\varepsilon} - 2\boldsymbol{\varepsilon}^T\boldsymbol{M} + \boldsymbol{\mu}^T\boldsymbol{W}^{-1}\boldsymbol{\mu})\right] \\ &\quad \times \frac{1}{\sqrt{(\det \boldsymbol{\Sigma})}} \qquad (4.32)\end{aligned}$$

Selanjutnya akan dilakukan pengaturan ulang pada bagian eksponensial kedua pada (4.32). untuk menyederhanakan notasi, bagian ini akan dinotasikan dengan V.

$$\begin{aligned}V \quad &= \exp\left[-\frac{1}{2}(\boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{W}^{-1}\boldsymbol{\varepsilon} - 2\boldsymbol{\varepsilon}^T\boldsymbol{M} + \boldsymbol{\mu}^T\boldsymbol{W}^{-1}\boldsymbol{\mu})\right] \\ &= \exp\left[-\frac{1}{2}(\boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{W}^{-1}\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^T\boldsymbol{M} - \boldsymbol{M}^T\boldsymbol{\varepsilon} + \boldsymbol{\mu}^T\boldsymbol{W}^{-1}\boldsymbol{\mu})\right] \qquad (4.33)\end{aligned}$$

(4.33) dapat dilakukan karena $\boldsymbol{\varepsilon}$ dan $\boldsymbol{M}$ adalah vektor

$$V \quad = \exp\left[-\frac{1}{2}\begin{pmatrix}\boldsymbol{\varepsilon}^T\boldsymbol{W}^{-1}\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^T(\boldsymbol{W}^{-1}\boldsymbol{W})\boldsymbol{M} \\ -\boldsymbol{M}^T(\boldsymbol{W}^{\boldsymbol{T}}(\boldsymbol{W}^{\boldsymbol{T}})^{-\boldsymbol{1}})\boldsymbol{\varepsilon} + \boldsymbol{\mu}^T\boldsymbol{W}^{-1}\boldsymbol{\mu}\end{pmatrix}\right] \qquad (4.34)$$

$$= \exp\left[-\frac{1}{2}(\boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{W}^{-1}\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^T\boldsymbol{W}^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^T\boldsymbol{W}^{-1}\boldsymbol{\varepsilon} + \boldsymbol{\mu}^T\boldsymbol{W}^{-1}\boldsymbol{\mu})\right] \qquad (4.35)$$

Pada transisi dari (4.33) ke (4.34) dilakukan perkalian dengan matriks identitas sehingga tidak ada perubahan nilai. Selanjutnya transisi dari (4.34) ke (4.35) menggunakan definsi dari matriks $\boldsymbol{W}^{-1} = \boldsymbol{\Sigma}^{-1} + 2\boldsymbol{\Lambda}(\boldsymbol{\tau})$ yang mengimplikasikan bahwa matriks $\boldsymbol{W}$ simetris. Jika diperhatikan, (4.35) adalah bentuk kuadratik dari $(\boldsymbol{\varepsilon} - \boldsymbol{\mu})$ sehingga dapat dituliskan menjadi

$$V = \exp\left[-\frac{1}{2}\left((\boldsymbol{\varepsilon} - \boldsymbol{\mu})^T \boldsymbol{W}^{-1}(\boldsymbol{\varepsilon} - \boldsymbol{\mu})\right)\right] \tag{4.36}$$

(4.36) diterapkan kembali ke (4.32)

$$\begin{aligned} \exp(\tilde{L}_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}, \boldsymbol{\tau}]) \quad &= \exp[T_1(\boldsymbol{\tau}) + T_2(\boldsymbol{\tau}, \boldsymbol{\beta}) + \frac{1}{2}\boldsymbol{\mu}^T \boldsymbol{W}^{-1}\boldsymbol{\mu}] \\ &\times \exp\left[-\frac{1}{2}\left((\boldsymbol{\varepsilon} - \boldsymbol{\mu})^{\boldsymbol{T}} \boldsymbol{W}^{-1}(\boldsymbol{\varepsilon} - \boldsymbol{\mu})\right)\right] \\ &\times \frac{1}{\sqrt{(\det \boldsymbol{\Sigma})}} \end{aligned} \tag{4.37}$$

Akhirnya didapatkan (4.37) yang merupakan bentuk distribusi $[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}, \boldsymbol{\tau}]$. Bentuk ini merupakan distribusi multivariat normal dengan vektor mean $\boldsymbol{\mu}$ dan matriks kovariansi $\boldsymbol{W}$ (*MN*$(\boldsymbol{\mu}, \boldsymbol{W})$). Sesuai dengan pernyataan Hardouin (2019) sebelumnya, $[\boldsymbol{\varepsilon}|\boldsymbol{Y}^*, \boldsymbol{\beta}\,, \boldsymbol{\Sigma}, \boldsymbol{\tau}\,]$ proporsional dengan $[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}, \boldsymbol{\tau}]$

$$\begin{aligned} p(\boldsymbol{\varepsilon} \,|\boldsymbol{Y}^*, \boldsymbol{\beta}\,, \boldsymbol{\Sigma}, \boldsymbol{\tau}) \quad &\propto \exp[T_1(\boldsymbol{\tau}) + T_2(\boldsymbol{\tau}, \boldsymbol{\beta}) + \frac{1}{2}\boldsymbol{\mu}^T \boldsymbol{W}^{-1}\boldsymbol{\mu}] \\ &\times \exp\left[-\frac{1}{2}\left((\boldsymbol{\varepsilon} - \boldsymbol{\mu})^T \boldsymbol{W}^{-1}(\boldsymbol{\varepsilon} - \boldsymbol{\mu})\right)\right] \\ &\times \frac{1}{\sqrt{(\det \boldsymbol{\Sigma})}} \end{aligned} \tag{4.38}$$

sehingga perhitungan ekspektasi pada (4.10) didasarkan dengan distribusi *MN*$(\boldsymbol{\mu}, \boldsymbol{W})$.

Karena telah ditemukan bentuk fungsi objektif (4.30) dan distribusi bersyarat yang dibutuhkan (4.38), selanjutnya dibahas mengenai prosedur iteratif EM variasional untuk mendapat estimasi parameter dalam model regresi logistik spasial.

4. 2. 2. Prosedur Iteratif EM Variasional untuk Mendapatkan Estimasi Parameter

Pada pembahasan subbab 3.2.3, didapatkan bentuk fungsi objektif pada *mean-field* variasional. Bentuk fungsi objektif ini ekuivalen dengan bentuk ekspektasi fungsi *complete log-likelihood* (4.10). Sehingga penurunan prosedur iteratif EM variasional serupa dengan kasus *mean-field* variasional.

Secara garis besar, prosedur iteratif EM variasional akan mencari solusi untuk memaksimumkan $\tilde{L}_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}, \boldsymbol{\tau}]$ relatif terhadap $\boldsymbol{\beta}$ dan $\boldsymbol{\Sigma}$ dengan $\boldsymbol{\tau}$ bernilai tetap, sehingga didapatkan $\tilde{L}_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta_{max}}\,, \boldsymbol{\Sigma_{max}}, \boldsymbol{\tau}]$. Sebagai tambahan, langkah memaksimumkan fungsi $\tilde{L}_c$ juga dilakukan terhadap parameter variasional $\boldsymbol{\tau}$ sehingga prosedur iteratif berakhir dengan didapatkannya $\tilde{L}_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta_{max}}\,, \boldsymbol{\Sigma_{max}}, \boldsymbol{\tau_{max}}]$. Oleh Hardouin (2019) diberikan pertaksamaan berikut

$$\begin{aligned} \tilde{L}_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}, \boldsymbol{\tau}] &\leq \tilde{L}_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta_{max}}\,, \boldsymbol{\Sigma_{max}}, \boldsymbol{\tau}] \\ &\leq \tilde{L}_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta_{max}}\,, \boldsymbol{\Sigma_{max}}, \boldsymbol{\tau_{max}}] \end{aligned} \tag{4.39}$$

Berdasarkan pertaksamaan (4.39) tujuan prosedur iteratif adalah mendapatkan

$$\tilde{L}_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta_{max}}\,, \boldsymbol{\Sigma_{max}}, \boldsymbol{\tau_{max}}] \cong L_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta_{max}}\,, \boldsymbol{\Sigma_{max}}]. \tag{4.40}$$

Berdasarkan algoritma EM, setiap iterasi dimulai dengan menghitung ekspektasi dari fungsi *complete log-likelihood*. Oleh karena itu dibentuk fungsi berikut

$$\varphi[\boldsymbol{\beta}\,, \boldsymbol{\Sigma}, \widehat{\boldsymbol{\beta}}^{(l)}\,, \widehat{\boldsymbol{\Sigma}}^{(l)}\,; \boldsymbol{\tau}] \quad = E[\tilde{L}_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}, \boldsymbol{\tau}]|\boldsymbol{Y}^*, \widehat{\boldsymbol{\beta}}^{(l)}\,, \widehat{\boldsymbol{\Sigma}}^{(l)}] \tag{4.41}$$

dengan indeks (l) diatas parameter menandakan parameter tersebut merupakan hasil iterasi ke-l. Fungsi (4.41) menjadi dasar dalam prosedur EM variasional selanjutnya. Berdasarkan (4.10) dan implikasi (4.38), oleh Hardouin (2019) fungsi (4.41) diduga memiliki bentuk sebagai berikut

$$\begin{aligned} \varphi[\boldsymbol{\beta}\,, \boldsymbol{\Sigma}, \widehat{\boldsymbol{\beta}}^{(l)}\,, \widehat{\boldsymbol{\Sigma}}^{(l)}\,; \boldsymbol{\tau}] \quad &= T_1(\boldsymbol{\tau}) + T_2(\boldsymbol{\tau}, \boldsymbol{\beta}) + \widehat{\boldsymbol{\mu}}^{(l)T}\boldsymbol{M} \\ &\quad - \frac{1}{2} tr\left[\left(\widehat{\boldsymbol{W}}^{(l)} + \widehat{\boldsymbol{\mu}}^{(l)}\widehat{\boldsymbol{\mu}}^{(l)T}\right)\boldsymbol{W}^{-1}\right] \\ &\quad - \frac{1}{2}\log(\det \boldsymbol{\Sigma}) \end{aligned} \tag{4.42}$$

dengan $tr[\boldsymbol{A}]$ menotasikan trace untuk matriks $\boldsymbol{A}$ dan $tr\left[\widehat{\boldsymbol{\mu}}^{(l)}\widehat{\boldsymbol{\mu}}^{(l)T}\boldsymbol{W}^{-1}\right] = \widehat{\boldsymbol{\mu}}^{(l)T}\boldsymbol{W}^{-1}\,\widehat{\boldsymbol{\mu}}^{(l)}$.

Adanya *closed-form* untuk ekspektasi *complete log-likelihood* (4.42), langkah memaksimumkan *complete log-likelihood* dapat dilanjutkan menggunakan fungsi ini. Langkah memaksimumkan fungsi *complete log-likelihood* dibagi menjadi dua bagian, memaksimumkan fungsi (4.42) berdasarkan parameter dalam model yaitu $\boldsymbol{\beta}$ dan $\boldsymbol{\Sigma}$ serta memaksimumkan fungsi (4.42) berdasarkan parameter variasional $\boldsymbol{\tau}$. Notasi diperjelas dengan menambahkan indeks parameter seperti berikut : $\boldsymbol{M}$ dinotasikan dengan $\boldsymbol{M}(\boldsymbol{\beta}, \boldsymbol{\tau})$, $\boldsymbol{W}$ dinotasikan dengan $\boldsymbol{W}(\boldsymbol{\Sigma}, \boldsymbol{\tau})$, dan $\boldsymbol{\mu}$ dinotasikan dengan $\boldsymbol{\mu}(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau})$. Penotasian indeks untuk iterasi juga dituliskan pada notasi baru tersebut.

Prosedur iterasi dimulai dengan menginisiasi nilai-nilai $\widehat{\boldsymbol{\beta}}^{(0)}$, $\widehat{\boldsymbol{\Sigma}}^{(0)}$, dan $\widehat{\boldsymbol{\tau}}^{(0)}$. Nilai-nilai koefisien dalam $\widehat{\boldsymbol{\beta}}^{(0)}$ dapat diinisiasi dengan estimasi koefisien model regresi logistik biasa dari data. Terkait aturan nilai inisiasi untuk $\widehat{\boldsymbol{\Sigma}}^{(0)}$, Hardouin (2019) menggunakan fungsi kovariansi eksponensial dengan elemennya adalah $\widehat{\boldsymbol{\Sigma}}^{(0)}{}_{ij} = C(\boldsymbol{s}_i\text{-}\boldsymbol{s}_j)$ dimana $C(\boldsymbol{h}) = \sigma^2 exp(|\boldsymbol{h}|/\theta)$. Nilai σ^2 dan θ diinisiasi berdasarkan kesesuaian dengan data. Penggunaan matriks kovariansi lain sangat dimungkinkan namun tidak dibahas dalam penelitian ini. Terakhir, inisiasi untuk nilai-nilai dalam $\widehat{\boldsymbol{\tau}}^{(\mathbf{0})}$ disarankan oleh Hardouin (2019) agar memenuhi $\tau(\boldsymbol{s})^2 = y(\boldsymbol{s})^2$ untuk setiap $\boldsymbol{s} \in T$. Dapat juga menggunakan $\hat{\tau}^{(0)}(\boldsymbol{s}) = [X(\boldsymbol{s})^T\widehat{\boldsymbol{\beta}}^{(\mathbf{0})} + \eta(\boldsymbol{s})] \times (2z - 1)$, dimana $\eta(\boldsymbol{s})$ ~ i.i.d N(0,1).

Berdasarkan semua yang telah dijabarkan diatas, selanjutnya akan dibahas langkah-langkah dalam setiap iterasi untuk mendapatkan estimasi dari parameter model logistik spasial menggunakan metode variasional. Untuk iterasi ke-l , $l = 1, 2, \ldots$ ikuti langkah berikut (Algoritma 2)

1) Hitung : $\widehat{\boldsymbol{W}}_1^{(l-1)} = \boldsymbol{W}\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \widehat{\boldsymbol{\Sigma}}^{(\boldsymbol{l}-\mathbf{1})}\right)$, $\widehat{\boldsymbol{M}}_1^{(l-1)} = \boldsymbol{M}\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \widehat{\boldsymbol{\beta}}^{(\boldsymbol{l}-\mathbf{1})}\right)$, dan $\widehat{\boldsymbol{\mu}}_1^{(l-1)} = \widehat{\boldsymbol{W}}_1^{(l-1)}\widehat{\boldsymbol{M}}_1^{(l-1)}$. Definisi dari tiap-tiap unsur telah dijabarkan sebelumnya.
2) Hitung

   $$\widehat{\boldsymbol{\beta}}^{(\boldsymbol{l})} = \arg\max_{\boldsymbol{\beta}}\left[T_2\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \boldsymbol{\beta}\right) + \left(\widehat{\boldsymbol{\mu}}_1^{(l-1)}\right)^T \boldsymbol{M}\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \boldsymbol{\beta}\right)\right].$$

   Penghitungan ini dapat dilakukan dengan cara berikut. Bentuk fungsi yang akan dimaksimumkan adalah

   $$T(\boldsymbol{\beta}) = \sum_{s\in T} -\lambda\big(\tau(\boldsymbol{s})\big)[\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}]^2 - [\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}][y^*(\boldsymbol{s}) - \frac{1}{2}$$
   $$-2\lambda\big(\tau(\boldsymbol{s})\big)\hat{\mu}(\boldsymbol{s})]$$

   dengan menetapkan nilai-nilai dalam $\boldsymbol{\tau}$, dicari solusi $G(\boldsymbol{\beta}) = \frac{\partial T(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}$ dengan $G(\boldsymbol{\beta}) = \sum_{s\in T}\left\{\boldsymbol{X}(\boldsymbol{s})\begin{bmatrix} -2\lambda\big(\tau(\boldsymbol{s})\big)[\boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}] \\ +y^*(\boldsymbol{s}) - \frac{1}{2} - 2\lambda\big(\tau(\boldsymbol{s})\big)\hat{\mu}(\boldsymbol{s})\end{bmatrix}\right\}$. Namun, untuk ukuran $\boldsymbol{\beta}$

lebih dari dua gunakan metode Newton-Raphson menggunakan prosedur iteratif: $\widehat{\boldsymbol{\beta}}^{(k)} = \widehat{\boldsymbol{\beta}}^{(k-1)} - \left[\frac{\partial G(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]^{-1}_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^{(k-1)}} \partial G\left(\widehat{\boldsymbol{\beta}}^{(k-1)}\right)$, dengan $\frac{\partial G(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{\boldsymbol{s}\in T} -2\lambda\left(\tau(\boldsymbol{s})\right)\boldsymbol{X}(\boldsymbol{s})\boldsymbol{X}(\boldsymbol{s})^{\boldsymbol{T}}$ dan syarat konvergensi ketika $\widehat{\boldsymbol{\beta}}^{(k)} \cong \widehat{\boldsymbol{\beta}}^{(k-1)}$ sehingga gunakan $\widehat{\boldsymbol{\beta}}^{(k)}$ sebagai estimasi untuk $\widehat{\boldsymbol{\beta}}^{(l)}$.

3) Perbarui bagian dari fungsi objektif yang berkaitan dengan hasil langkah kedua, yaitu

$\widehat{\boldsymbol{M}}_2^{(l-1)} = \boldsymbol{M}\left(\widehat{\boldsymbol{\tau}}^{(l-1)}, \widehat{\boldsymbol{\beta}}^{(l)}\right)$ dan $\widehat{\boldsymbol{\mu}}_2^{(l-1)} = \widehat{\boldsymbol{W}}_1^{(l-1)}\widehat{\boldsymbol{M}}_2^{(l-1)}$

4) Hitung

$$\widehat{\boldsymbol{\Sigma}}^{(\boldsymbol{l})} = \arg\max_{\boldsymbol{\Sigma}} \left\{ \begin{matrix} -\frac{1}{2}\, tr\left[\left(\widehat{\boldsymbol{W}}_1^{(l-1)} + \widehat{\boldsymbol{\mu}}_2^{(l-1)}\widehat{\boldsymbol{\mu}}_2^{(l-1)^T}\right)\boldsymbol{W}^{-1}\left(\widehat{\boldsymbol{\tau}}^{(l-1)}, \boldsymbol{\Sigma}\right)\right] \\ -\frac{\mathbf{1}}{\mathbf{2}}\log(\det \boldsymbol{\Sigma}) \end{matrix} \right\}$$

Untuk mendapatkan solusi diatas, $\boldsymbol{\Sigma}$ akan dilihat sebagai fungsi $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{Q}$. Oleh karena itu fungsi objektif untuk mendapatkan estimasi $\boldsymbol{\Sigma}$ berubah menjadi :

$$h(\boldsymbol{Q}, \sigma^2) = -\frac{1}{2\sigma^2} tr\left[\left(\widehat{\boldsymbol{W}}_1^{(l-1)} + \widehat{\boldsymbol{\mu}}_2^{(l-1)}\widehat{\boldsymbol{\mu}}_2^{(l-1)^T}\right)\boldsymbol{Q}^{-\mathbf{1}}\right] - \frac{1}{2}\log[\det(\sigma^2\boldsymbol{Q})]$$

atau dapat disederhanakan menjadi :

$$h(\boldsymbol{Q}, \sigma^2) = \frac{1}{\sigma^2} tr\left[\left(\widehat{\boldsymbol{W}}_1^{(l-1)} + \widehat{\boldsymbol{\mu}}_2^{(l-1)}\widehat{\boldsymbol{\mu}}_2^{(l-1)^T}\right)\boldsymbol{Q}^{-\mathbf{1}}\right] + \log[\det(\boldsymbol{Q})] + n\log\sigma^2$$

dan dimaksimumkan terhadap σ^2 dan θ. Turunan fungsi objektif $h(\boldsymbol{Q}, \sigma^2)$ terhadap σ^2

$$\frac{\partial h}{\partial \sigma^2} = -\frac{1}{(\sigma^2)^2} tr\left[\left(\widehat{\boldsymbol{W}}_1^{(l-1)} + \widehat{\boldsymbol{\mu}}_2^{(l-1)}\widehat{\boldsymbol{\mu}}_2^{(l-1)^T}\right)\boldsymbol{Q}^{-\mathbf{1}}\right] + \frac{n}{\sigma^2}$$

dan didapatkan solusi untuk $\frac{\partial h}{\partial \sigma^2} = 0$ adalah

$$\sigma^2(\boldsymbol{Q}) = \frac{1}{n} tr\left[\left(\widehat{\boldsymbol{W}}_1^{(l-1)} + \widehat{\boldsymbol{\mu}}_2^{(l-1)}\widehat{\boldsymbol{\mu}}_2^{(l-1)^T}\right)\boldsymbol{Q}^{-\mathbf{1}}\right]$$

yang kemudian diterapkan ke fungsi objektif $h(\boldsymbol{Q}, \sigma^2)$ didapatkan

$$h(\boldsymbol{Q}) = n + \log[\det(\boldsymbol{Q})] + n\log\frac{1}{n} tr\left[\left(\widehat{\boldsymbol{W}}_1^{(l-1)} + \widehat{\boldsymbol{\mu}}_2^{(l-1)}\widehat{\boldsymbol{\mu}}_2^{(l-1)^{\boldsymbol{T}}}\right)\boldsymbol{Q}^{-\mathbf{1}}\right]$$

terakhir $h(\boldsymbol{Q})$ dimaksimumkan terhadap $\boldsymbol{Q}$. $\boldsymbol{Q}$ dapat diasumsikan diparameterisasi oleh θ melalui suatu fungsi kovariansi. Misalkan fungsi eksponensial digunakan untuk mengkarakterisasi matriks kovariansi $\mathbf{\Sigma}$, maka elemen-elemennya $\boldsymbol{\Sigma}_{ij} = C(\boldsymbol{s}_i\text{-}\boldsymbol{s}_j)$ dengan $C(\boldsymbol{h}) = \sigma^2\ exp\left(\frac{|\boldsymbol{h}|}{\theta}\right)$. Jika cara ini digunakan, maka proses memaksimumkan $h(\boldsymbol{Q})$ adalah dengan mencari parameter θ.

5) Perbarui bagian dari fungsi objektif yang berkaitan dengan hasil langkah keempat

$$\widehat{\boldsymbol{W}}_2^{(l-1)} = \boldsymbol{W}\left(\hat{\boldsymbol{\tau}}^{(\boldsymbol{l-1})}, \widehat{\mathbf{\Sigma}}^{(\boldsymbol{l})}\right) \text{ dan } \widehat{\boldsymbol{\mu}}_3^{(l-1)} = \widehat{\boldsymbol{W}}_2^{(l-1)} \widehat{\boldsymbol{M}}_2^{(l-1)}$$

6) Perbarui parameter variasional dengan menghitung

$$\hat{\boldsymbol{\tau}}^{(\boldsymbol{l})} = \arg\max_{\boldsymbol{\tau}} \left\{ \begin{array}{c} T_1(\boldsymbol{\tau}) + T_2\left(\boldsymbol{\tau}, \widehat{\boldsymbol{\beta}}^{(\boldsymbol{l})}\right) + \widehat{\boldsymbol{\mu}}_3^{(l-1)^T} \boldsymbol{M}\left(\boldsymbol{\tau}, \widehat{\boldsymbol{\beta}}^{(\boldsymbol{l})}\right) \\ -\frac{1}{2}\, tr\left[\left(\widehat{\boldsymbol{W}}_2^{(l-1)} + \widehat{\boldsymbol{\mu}}_3^{(l-1)} \widehat{\boldsymbol{\mu}}_3^{(l-1)^{\boldsymbol{T}}}\right) \boldsymbol{W}^{-\boldsymbol{1}}\left(\boldsymbol{\tau}, \widehat{\mathbf{\Sigma}}^{(\boldsymbol{l})}\right)\right] \end{array} \right\}$$

Untuk mendapatkan solusi ini, pertama diperkenalkan notasi $\widehat{W}_{ii}$ sebagai elemen diagonal utama ke-*i* dari matriks $\widehat{\boldsymbol{W}}$. Selanjutnya didefinisikan

$\hat{Z}^{(l)} = \left(X(\boldsymbol{s})^T \widehat{\boldsymbol{\beta}}^{(l)}\right)^2 + 2\hat{\mu}(\boldsymbol{s})\left(X(\boldsymbol{s})^T \widehat{\boldsymbol{\beta}}^{(l)}\right) + \widehat{W}_{ii} + \hat{\mu}(\boldsymbol{s})^2$ dan

$A\left(\widehat{\boldsymbol{\beta}}^{(l)}, \widehat{\boldsymbol{\Sigma}}^{(l)}; y\right) = \log f(y) - \frac{y}{2} - \lambda(y)\left(\hat{Z}^{(l)} - y^2\right)$. Kemudian nyatakan

$E\left[\tilde{L}_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon}\,|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau}]\middle|\boldsymbol{Y}^*, \widehat{\boldsymbol{\beta}}^{(l)}, \widehat{\boldsymbol{\Sigma}}^{(l)}\right] = A\left(\widehat{\boldsymbol{\beta}}^{(l)}, \widehat{\boldsymbol{\Sigma}}^{(l)};\ \boldsymbol{\tau}\right) + d$

dengan *d* menyatakan bagian-bagian lain yang tidak mengandung $\boldsymbol{\tau}$. Sehingga untuk mendapatkan pembaruan parameter variasional, hanya diperlukan bagian $A\left(\widehat{\boldsymbol{\beta}}^{(l)}, \widehat{\boldsymbol{\Sigma}}^{(l)};\ \boldsymbol{\tau}\right)$ untuk dimaksimumkan. Selanjutnya dicari solusi persamaan turunan

$$\frac{\partial A\left(\widehat{\boldsymbol{\beta}}^{(l)}, \widehat{\boldsymbol{\Sigma}}^{(l)};\ \boldsymbol{\tau}\right)}{\partial \boldsymbol{\tau}} = \mathbf{0}$$

Caranya dengan melihat bentuk fungsi *A*. Sesuai definisi $\lambda(y) = \frac{f(y)-\frac{1}{2}}{2y}$ dan $\frac{d \log f(y)}{dy} = f(-y)$.

$$\frac{\partial A\left(\widehat{\boldsymbol{\beta}}^{(l)}, \widehat{\boldsymbol{\Sigma}}^{(l)}; y\right)}{\partial y} = -\lambda'(y)\left(\hat{Z}^{(l)} - y^2\right) + 2y\lambda(y) - \frac{1}{2} + f(-y)$$

$$\frac{\partial A\left(\widehat{\boldsymbol{\beta}}^{(l)}, \widehat{\boldsymbol{\Sigma}}^{(l)}; y\right)}{\partial y} = -\lambda'(y)\left(\hat{Z}^{(l)} - y^2\right)$$

selanjutnya akan dicari bentuk $\lambda'(y)$. Dengan definisi $\lambda(y) = \frac{f(y)-1/2}{2y}$, didapatkan bentuk turunannya sebagai berikut

$$\lambda'(y) = \frac{2y\exp(-y) - \exp(-2y) - 1}{4y^2(1+\exp(-y))^2}$$

didapatkan solusi untuk persamaan turunan fungsi $\boldsymbol{A}$ adalah ketika $y^2 = \hat{Z}^{(l)}$ atau $y^2 = 0$. Namun, dengan mempertimbangkan syarat batas bawah (4.15) supaya bernilai eksak, maka dipilih $y^2 = \hat{Z}^{(l)}$. Sehingga didapatkan bentuk persamaan untuk pembaruan parameter variasional sebagai berikut : untuk setiap $\boldsymbol{s} \in T$

$$\hat{\boldsymbol{\tau}}^{(l)}(\boldsymbol{s})^2 = \left(X(\boldsymbol{s})^T\widehat{\boldsymbol{\beta}}^{(l)}\right)^2 + 2\hat{\mu}^{(l)}(\boldsymbol{s})\left(X(\boldsymbol{s})^T\widehat{\boldsymbol{\beta}}^{(l)}\right) + \widehat{W}_{ii} + \hat{\mu}^{(l)}(\boldsymbol{s})^2$$

$$\hat{\boldsymbol{\tau}}^{(l)}(\boldsymbol{s}) = \sqrt{\hat{\boldsymbol{\tau}}^{(l)}(\boldsymbol{s})^2} \times (2Y^*(\boldsymbol{s}) - 1)$$

Dengan selesainya penjabaran mengenai prosedur iteratif EM variasional, selanjutnya akan dibahas mengenai sifat dari hasil estimasi. Pembahasan berdasarkan studi simulasi yang telah dilakukan oleh Hardouin (2019) dan dikaitkan dengan beberapa penelitian terkait estimasi menggunakan metode variasional.

4. 3. Sifat Estimasi Parameter Regresi Logistik Spasial Menggunakan Metode Variasional

Mengulas kembali mengenai teori metode variasional sebelumnya. Karena metode variasional merupakan metode untuk mengaproksimasi suatu bentuk estimasi, teori yang berlaku secara umum dan eksak belum ada (Peyrard, 2018). Sehingga sifat estimasi metode variasional selalu terbatas untuk model yang diterapkan. Namun, penggunaan metode variasional terbukti dari beberapa penelitian mampu menangani permasalahan kompleks dengan mengurainya ke permasalahan yang lebih sederhana (Jordan dkk, 1999). Mengutip Blei (2017) metode variasional mengubah teknik inferensi menjadi teknik aproksimasi. Hal ini mempermudah permasalahan dalam beberapa ranah studi seperti pada statistika bayesian yang terkendala dengan bentuk distribusi posterior.

Selanjutnya dibahas sifat dari estimator EM variasional menggunakan prosedur yang telah dijabarkan sebelumnya. Hardouin (2019) melakukan studi simulasi untuk mempelajari sifat dari estimasi parameter model logistik spasial menggunakan metode variasional.

Secara garis besar simulasi dilakukan untuk membandingkan akurasi estimasi metode EM variasional dengan EM menggunakan aproksimasi laplace. Dua pengaturan berbeda dibuat, satu pengaturan memastikan keadaan dimana dependensi spasial lebih kuat dibanding keadaan yang lain. Untuk memastikan dua keadaan ini, ditetapkan nilai parameter $\theta = 5$ untuk mewakili keadaan data dengan dependensi spasial rendah dan $\theta = 15$ untuk mewakili keadaan data dengan dependensi spasial yang lebih tinggi. Pengaturan parameter $\boldsymbol{\beta}$ sama dengan yang telah dibahas dalam penelitian ini ditambah dengan pengaturan inisiasi parameter $\boldsymbol{\beta}$ yang dipilih menggunakan arahan dari penelitian Aldworth dan Cressie (1999). Nilai-nilai parameter $\boldsymbol{\beta}$ ini adalah : $\beta_0 = 0{,}1, \beta_1 = 0{,}0625$, $\beta_2 = 0{,}0417$. Simulasi dilakukan sebanyak 100 kali. Hasil yang didapatkan adalah metode EM variasional memiliki akurasi lebih baik daripada metode EM dengan aproksimasi laplace. Hal ini terjadi pada data dengan dependensi spasial yang kuat (Hardouin, 2019). Pengukuran akurasi ini didasarkan pada MSE untuk estimasi parameter oleh metode EM variasional yang konsisten rendah untuk semua parameter.

Selain simulasi untuk akurasi, dilakukan juga simulasi bootstrap untuk mengestimasi variansi dari estimator variational. Estimasi variansi ini dibandingkan dengan metode EM menggunakan aproksimasi laplace. Pengaturan parameter yang dilakukan sama seperti simulasi untuk membandingkan akurasi sebelumnya. Hasilnya, walau tidak konsisten dan selisih variansi antara kedua metode sangat kecil, metode EM variasional sedikit lebih unggul dibandingkan metode EM laplace (Hardouin, 2019).

BAB V

KESIMPULAN DAN SARAN

Menutup bagian isi dari penelitian skripsi ini, berikut adalah kesimpulan dan saran yang diajukan oleh penulis.

5. 1. Kesimpulan

Berdasarkan tujuan penelitian yang telah dituliskan di bab pertama skripsi ini, penulis mengulas hasil yang didapatkan dalam penelitian ini sebagai berikut :

1) Berdasarkan penerapan metode estimasi *variational inference* untuk model regresi logistik spasial pada bab 4, telah didapatkan bentuk fungsi objektif pada (4.42). Fungsi (4.42) dioptimasi menggunakan algoritma *variational expectation-maximization* (VEM). Langkah dalam setiap iterasi secara umum berupa penghitungan fungsi (4.42) kemudian langkah memaksimumkan fungsi tersebut berdasarkan parameter dalam model. Detail dari prosedur ini telah dijelaskan pada algoritma 2 pada bab 4.
2) Terkait sifat hasil estimasi parameter menggunakan metode variasional, didapatkan dari studi literatur bahwa teori mengenai sifat estimasi metode variasional tidak dapat diperumum. Hal ini karena secara prinsip, metode variasional adalah pendekatan numerik untuk estimasi fungsi distribusi melalui fungsi distribusi lain. Walaupun demikian, beberapa penelitian mengenai penerapan metode variasional untuk model tertentu telah banyak dilakukan. Di antara penelitian tersebut, beberapa penelitian menemukan konsistensi estimator dan adanya tingkat akurasi yang cukup tinggi pada model yang diterapkan. Berdasarkan studi literatur penerapan metode variasional pada model regresi logistik spasial, didapatkan akurasi estimasi parameter yang relatif lebih baik dan variansi yang lebih kecil daripada metode aproksimasi Laplace untuk data dengan dependensi spasial yang tinggi. Hal ini didasarkan pada studi simulasi yang nilai-nilai parameternya telah dijelaskan pada sub bab 4.3.

5. 2. Saran

Metode *variational inference* telah banyak diteliti sebagai alternatif dalam prosedur estimasi fungsi distribusi dan parameter. Salah satu keuntungan yang banyak ditemukan dalam penerapan metode variasional adalah mengenai waktu komputasi. Hal ini tentu sangat menguntungkan terutama dalam menangani pemodelan dengan data berukuran besar atau pada kasus *big data*. Walaupun begitu, penelitian serta pemakaian metode ini di Indonesia masih relatif rendah. Oleh karena itu, penulis berharap bahwa skripsi ini menjadi pendorong untuk digiatkannya penelitian dan penerapan metode variasional oleh para akademisi dan profesional di bidang terkait di Indonesia.

DAFTAR PUSTAKA

Aldworth, Jeremy dan Cressie, Noel. 1999. "*Sampling designs and prediction methods for Gaussian spatial processes*", dalam buku *Multivariate Analysis, Designs of Experiments, and Survey Sampling* hal 1-54. Markel Dekker, Inc.: New York, NY.

Bickel, P. dkk. 2013. "*Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels*". Dari *The Annals of Statistics vol* 41(4), hal 1922–1943.

Blei, David M. 2017. "*Variational Inference : Foundations and Innovations"*. Video kuliah online https://simons.berkeley.edu/talks/david-blei-2017-5-1

Blei, David M. , Kucukelbir, Alp , dan McAuliffe, Jon D. . 2018. "*Variational Inference : A Review For Statisticians*". New York : Columbia University.

Bolstad, William M. 2007. "*Introduction to Bayesian Statistics second edition*". New Jersey : John Wiley and Sons, inc.

Celisse, A. dkk. 2012. "*Consistency of maximum-likelihood and variational estimators in the stochastic block model*". Dari *Electronic Journal of Statistics vol* 6, hal 1847–1899.

Gelfand, Alan E. 2000."*Gibbs Sampling.*" Dari *Journal of the American Statistical Association vol* 95, no. 452, hal 1300-304. DOI : 10.2307/2669775.

Hall, P., Ormerod, J., dan Wand, M. 2011. "*Theory of Gaussian variational approximation for a Poisson mixed model*". Dari *Statistica Sinica vol* 21, hal 369–389.

Hall, P., Pham, T., Wand, M., dan Wang, S. 2011. "*Asymptotic normality and valid inference for Gaussian variational approximation*". Dari *Annals of Statistics vol* 39(5), hal 2502–2532.

Hardouin, Cécile. 2019. "*A Variational Method for Parameter Estimation in a Logistic Spatial Regression*" dari *Spatial Statistics Journals volume 31* *www.sciencedirect.com/journal/spatial-statistics/vol/31/suppl/C*. DOI: 10.1016/j.spasta.2019.100365. Nanterre : Université Paris Nanterre

Hogg, Robert V. dan Craig, Allen T. . 2004. "*Introduction to Mathematical Statistics fifth edition*". Pearson Education Asia Limited.

Jaakkola, Tommi S. dan Jordan, Michael I. . 1998. "*Bayesian Parameter Estimation Through Variational Methods*". Artikel pada *Statistics and computing vol 10*¸25-37. DOI : 10.1023/A:1008932416310

Jordan, Michael I. dkk. 1999. "*An Introduction to Variational Methods for Graphical Models*". Dari *Springer Journal of Machine Learning vol 37,* 183-233. DOI : 10.1023/A:1007665907178.

Keng, Brian. 2017. "*Variational Bayes and the Mean-Field Approximation*". http://bjlkeng.github.io/posts/variational-bayes-and-the-mean-field-approximation/.

Keng, Brian. 2017. "*The Calculus of Variation*". http://bjlkeng.github.io/posts/the-calculus-of-variations/.

McLachlan, Geoffrey J. dan Krishnan, Thriyambakam. 2008. "*The EM Algorithm and Extensions, Second Edition*". DOI : 10.1002/9780470191613. John Wiley and Sons, inc.

Montgomery, Douglas C., Peck, Elizabeth A., dan Vining, G. Geoffrey. 2012. "*Introduction to Linear Regression Analysis fifth edition*". New Jersey : John Wiley & Sons, inc.

Nisa, Hilwin dkk. 2019. "*Estimation of propensity score using spatial logistic regression*" dari IOP Publishing. DOI : 10.1088/1757-899X/546/5/052048. Malang : Universitas Brawijaya.

Paéz, Antonio dkk. 2010. "*Progress in Spatial Analysis : Methods and Applications*". Berlin : Springer-Verlag Berlin Heidelberg

Parisi, Giorgio. 1988. "*Statistical Field Theory*". Dari *Physics Today vol* 41(12), hal 110. DOI : 10.1063/1.2811677.

Peyrard, Nathalie dkk. 2018. "*Exact or approximate inference in graphical models : why the choice is dictated by the treewidth, and how variable elimination can be exploited*". Artikel akses terbuka pada arxiv : 1506.08544.pdf

Roberts, Gareth O. dan Rosenthal, Jeffrey S. 2001. "*Optimal Scaling for Various Metropolis-Hastings Algorithm*". Artikel akses terbuka pada *Project Euclid : Statistical Science volume 16 number 4(2001),* hal 351-367. The Institute of Mathematical Statistics.

Ross, Sheldon M. . 2010. "*Introduction to Probability Models 10th edition*". Elsevier Inc.

Rustagi, Jagdish S. . 1976. "*Variational Methods in Statistics*". New York : Academic Press, Inc.

Sakurai, J. J. . 1985. "*Modern Quantum Mechanics*". Redwood City, California : Benjamin-Cummings Publishing Co.,Subs. of Addison Wesley Longman.

Sengupta, Aritra dan Cressie, Noel. 2013. "*Hierarchical statistical modeling of big spatial datasets using the exponential family of distributions*" dari *Spatial Statistics Journals volume 31 www.sciencedirect.com/journal/spatial-statistics/vol/31/suppl/C*. DOI : 10.1016/j.spasta.2013.02.002. Ohio : The Ohio State University.

Varberg, Dale , Purcell, Edwin J. , dan Rigdon, Steven E. .2007 . "*Calculus ninth edition*". London : Pearson Education.

You, C., Ormerod, J., dan Muller, S. 2014. "*On variational Bayes estimation and variational information criteria for linear regression models*". *Australian & New Zealand Journal of Statistics vol* 56(1), hal 73–87.

LAMPIRAN 1. Metode Newton-Raphson untuk Estimasi Parameter Regresi Logistik

Misalkan suatu fungsi f terdefinsi pada $[a,b]$, dengan $p^{(0)} \in [a,b]$ sebagai estimasi awal untuk p sedemikian $f'\left(p^{(0)}\right) \neq 0$ dan $|p-p^{(0)}|$ kecil. Dengan polinom taylor pertama untuk f diperluas dengan $p^{(0)}$ dan dievaluasi saat p

$$f(p) = f\left(p^{(0)}\right) + \left(p-p^{(0)}\right)f'\left(p^{(0)}\right) + \frac{(p-p^{(0)})^2}{2}f''(\xi(p))$$

$\xi(p)$ berada diantara p dan $p^{(0)}$. Kemudian karena $f(p)=0$ dan $|p-p^{(0)}|$ kecil menyebabkan kuadratnya lebih kecil lagi persamaannya menjadi

$$0 = f\left(p^{(0)}\right) + \left(p-p^{(0)}\right)f'\left(p^{(0)}\right) + \frac{(p-p^{(0)})^2}{2}f''(\xi(p))$$

$$0 = f\left(p^{(0)}\right) + \left(p-p^{(0)}\right)f'\left(p^{(0)}\right)$$

$$\left(p-p^{(0)}\right)f'\left(p^{(0)}\right) = -f\left(p^{(0)}\right)$$

$$pf'\left(p^{(0)}\right) = -f\left(p^{(0)}\right) + p^{(0)}f'\left(p^{(0)}\right)$$

$$p = -\frac{f\left(p^{(0)}\right)}{f'(p^{(0)})} + p^{(0)} = p^{(1)}$$

Dari hasil diatas, proses diulang dengan $p^{(1)}$ sebagai hasil iterasi pertama untuk estimasi p. Proses iterasi berhenti ketika $p^{(n)}$ mencapai ambang toleransi. Proses ini menghasilkan himpunan yang dapat dinyatakan sebagai berikut

$$\{p^{(n)}\}_{n=0}^{\infty}\ ,\ p^{(n)} = -\frac{f\left(p^{(n-1)}\right)}{f'\left(p^{(n-1)}\right)} + p^{(n-1)}\text{, untuk } n \geq 1$$

Untuk aplikasi metode NR pada regresi logistik, diinginkan nilai parameter yang memaksimumkan fungsi likelihood $L(\boldsymbol{\beta})$. Dalam prosedurnya, dicari nilai parameter sedemikian $\ell'(\boldsymbol{\beta}) = 0$. Untuk model regresi logistik yang mengasumsikan adanya nilai teramati berulang dengan jumlah variabel independen sebanyak K, didapatkan bentuk turunan pertama untuk log-likelihoodnya adalah

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_k} = \frac{\partial}{\partial \beta_k}\left\{\sum_{i=1}^{N} y_i\,\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta} - \sum_{i=1}^{N} n_i \ln[1+\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})]\right\}$$

$$= \frac{\partial}{\partial \beta_k}\sum_{i=1}^{N} y_i\,\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta} - \frac{\partial}{\partial \beta_k}\sum_{i=1}^{N} n_i \ln[1+\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})]$$

(Lanjutan)

$$= \sum_{i=1}^{N} y_i\, x_{ik} - \sum_{i=1}^{N} n_i x_{ik} \frac{\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})}{1+\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})} \quad , \qquad k = 0, 1, \dots, K$$

Menggunakan persamaan setiap turunan parsial diatas dengan nol, dapat ditemukan solusi untuk estimasi likelihood maksimum untuk parameter $\boldsymbol{\beta}$. Metode NR membutuhkan turunan kedua dari log-likelihood

$$\frac{\partial \ell^2(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_k} = \frac{\partial}{\partial \beta_k}\left[\sum_{i=1}^{N} y_i\, x_{ik} - \sum_{i=1}^{N} n_i x_{ik} \frac{\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})}{1+\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})}\right]$$

$$= -\frac{\partial}{\partial \beta_k}\sum_{i=1}^{N} n_i x_{ik} \frac{\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})}{1+\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})}$$

$$= -\frac{\partial}{\partial \beta_k}\sum_{i=1}^{N} n_i x_{ik} \frac{\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})}{1+\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})}$$

$$= -\sum_{i=1}^{N} n_i x_{ik} \frac{\partial}{\partial \beta_k} \frac{\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})}{1+\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})}$$

$$= -\sum_{i=1}^{N} n_i x_{ik} \frac{x_{ik}\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})\ [1+\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})] - \exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta}) x_{ik}\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})}{[1+\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})]^2}$$

$$= -\sum_{i=1}^{N} n_i x_{ik}\left[x_{ik}\frac{\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})}{1+\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})} - x_{ik}\frac{\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})}{[1+\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})]^2}\right]$$

$$= -\sum_{i=1}^{N} n_i x_{ik} x_{ik}\frac{\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})}{1+\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})}\left[1 - \frac{\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})}{1+\exp(\boldsymbol{x'}_{\boldsymbol{i}}\boldsymbol{\beta})}\right]$$

Prosedur NR untuk estimasi vektor parameter $\boldsymbol{\beta}$ dapat dinotasikan menggunakan bentuk matriks sebagai berikut

$$\boldsymbol{\beta}^{(n)} = -\ell'\big(\boldsymbol{\beta}^{(n-1)}\big)\ell''\big(\boldsymbol{\beta}^{(n-1)}\big)^{-1} + \boldsymbol{\beta}^{(n-1)}$$

Dengan $\ell'(\boldsymbol{\beta})$ dan $\ell''(\boldsymbol{\beta})$ menyatakan vektor yang berisi bentuk turunan parsial pertama dan kedua dari log-likelihood terhadap setiap parameter dalam vektor parameter $\boldsymbol{\beta}$.

LAMPIRAN 2. Algoritma *Expectation-Maximization* untuk Estimasi Parameter Regresi Logistik Spasial

Prosedur algoritma *Expectation-Maximization*(EM) untuk estimasi parameter $\boldsymbol{\beta}$ pada model regresi logistik spasial dimulai dengan membentuk likelihood komplit yang mempertimbangkan nilai-nilai tak teramati. Secara matematis, bentuknya dapat dituliskan sebagai berikut

$$\ell_c(\boldsymbol{\beta},\rho) = \boldsymbol{y}'\boldsymbol{X\beta} - \sum_{i=1}^{n} \ln\left[1+\exp\left(\frac{[\boldsymbol{HX\beta}]_i}{\boldsymbol{\Sigma}_{ii}}\right)\right] - \sum_{i=n+1}^{m} \ln\left[1+\exp\left(\frac{[\boldsymbol{HX\beta}]_i}{\boldsymbol{\Sigma}_{ii}}\right)\right]$$

Selanjutnya dengan menggunakan $\boldsymbol{\beta}^{(0)}$ sebagai estimasi awal untuk $\boldsymbol{\beta}$, dihitung nilai

$$H\left(\boldsymbol{\beta},\boldsymbol{\beta}^{(0)}\right) = \boldsymbol{E}\left\{\ell_c(\boldsymbol{\beta},\rho)\middle|\boldsymbol{X},\boldsymbol{\beta}^{(0)}\right\}$$

Dengan mengganti setiap nilai teramati dari variabel dependen oleh y_j, $j = n+1,\dots,m$, menggunakan bentuk ekspektasi (Anderson dan Hardin, 2013)

$$\boldsymbol{E}\left(Y_j\middle|\boldsymbol{x_j},\boldsymbol{\beta}^{(0)}\right) = P\left(Y_j = 1\middle|\boldsymbol{x_j},\boldsymbol{\beta}^{(0)}\right)$$

$$= \frac{1}{1+\exp\left(\frac{-[\boldsymbol{HX\beta}]_{\boldsymbol{i}}}{\boldsymbol{\Sigma}_{\boldsymbol{ii}}}\right)}$$

Prosedur selanjutnya merupakan prosedur iteratif dengan langkah pertama memilih estimasi awal, misal $\widehat{\boldsymbol{\beta}}$**,** kemudian untuk setiap iterasi gunakan persamaan berikut untuk memperbaiki estimasi sebelumnya

$$\boldsymbol{\beta}^{(n)} = \boldsymbol{\beta}^{(n-1)} + (\boldsymbol{X}'\boldsymbol{V}^{(\boldsymbol{n-1})}\boldsymbol{X})^{-\boldsymbol{1}}\boldsymbol{X}'\,(\boldsymbol{y}-\boldsymbol{p}^{(\boldsymbol{n-1})})$$

$\boldsymbol{p}$ menotasikan vektor probabilitas yang diprediksi untuk suatu respon dengan anggotanya dapat dinyatakan sebagai

$$p_i = \frac{1}{1+\exp(-\boldsymbol{x}_{\boldsymbol{i}}'\boldsymbol{\beta})}$$

$\boldsymbol{V}^{(\boldsymbol{n-1})}$ menotasikan matriks diagonal dengan diagonal utamanya adalah ${p_i}^{(n-1)}\left(1-{p_i}^{(n-1)}\right)$. Prosedur iteratif berlanjut hingga estimasi terakhir mencapai batas toleransi kesalahan.

LAMPIRAN 3. Aproksimasi Laplace untuk Fungsi Ekspektasi *Complete Log-likelihood* Model Regresi Logistik Spasial

Lampiran ini berdasarkan *Appendix* pada penelitian Hardouin (2020). Untuk mengaproksimasi bentuk fungsi (4.10) menggunakan metode Laplace, pertama digunakan persamaan (4.9) untuk fungsi *complete log-likelihood* $L_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}]$ dan menotasikan $\boldsymbol{\varepsilon}_m$ sebagai modus fungsi $L_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}]$. Selanjutnya dibangun ekspansi Taylor untuk $L_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}]$ disekitar $\boldsymbol{\varepsilon}_m$ sebagai berikut

$$L_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}] = L_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon}_m \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}] + (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_m)^T \frac{\partial}{\partial \boldsymbol{\varepsilon}} L_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}]\Big|_{\boldsymbol{\varepsilon}=\boldsymbol{\varepsilon}_m}$$

$$+\frac{1}{2}(\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_m)^T \frac{\partial^2}{\partial \boldsymbol{\varepsilon} \partial \boldsymbol{\varepsilon}^T} L_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}]\Big|_{\boldsymbol{\varepsilon}=\boldsymbol{\varepsilon}_m} (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_m) + \cdots$$

Dengan bagian kedua bernilai 0 sehingga didapatkan aproksimasi

$$L_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}] \cong L_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon}_m \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}]$$

$$-\frac{1}{2}(\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_m)^T[-H(\boldsymbol{\varepsilon}_m)](\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_m)$$

$$H(\boldsymbol{\varepsilon}_m) = \frac{\partial^2}{\partial \boldsymbol{\varepsilon} \partial \boldsymbol{\varepsilon}^T} L_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}]\Big|_{\boldsymbol{\varepsilon}=\boldsymbol{\varepsilon}_m}$$

Dideduksi distribusi $[\boldsymbol{\varepsilon} \,|\boldsymbol{Y}^*, \boldsymbol{\beta}\,, \boldsymbol{\Sigma}]$ dapat diaproksimasi proporsional dengan distribusi normal multivariat $(\boldsymbol{\varepsilon}_m, [-H(\boldsymbol{\varepsilon}_m)]^{-1})$, sehingga $E[\boldsymbol{\varepsilon} \,|\boldsymbol{Y}^*, \boldsymbol{\beta}\,, \boldsymbol{\Sigma}] \cong \boldsymbol{\varepsilon}_m$ dan $var[\boldsymbol{\varepsilon} \,|\boldsymbol{Y}^*, \boldsymbol{\beta}\,, \boldsymbol{\Sigma}] = [-H(\boldsymbol{\varepsilon}_m)]^{-1}$ Selanjutnya prosedur yang similar diterapkan untuk mendapatkan ekspansi deret Taylor $\log\left(1 + e^{Y(\boldsymbol{s})}\right)$ disekitar $\boldsymbol{\varepsilon}_m(\boldsymbol{s})$. Dinotasikan $Y_m(\boldsymbol{s}) = \boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta} + \varepsilon_m(\boldsymbol{s})$, didapatkan

$$\log\left(1 + e^{Y(\boldsymbol{s})}\right) = \log\left(1 + e^{Y_m(\boldsymbol{s})}\right) + \left(\varepsilon(\boldsymbol{s}) - \varepsilon_m(\boldsymbol{s})\right)\frac{e^{Y_m(\boldsymbol{s})}}{1 + e^{Y_m(\boldsymbol{s})}}$$

$$+\frac{1}{2}\left(\varepsilon(\boldsymbol{s}) - \varepsilon_m(\boldsymbol{s})\right)^{\mathbf{2}} \frac{e^{Y_m(\boldsymbol{s})}}{\left(1+e^{Y_m(\boldsymbol{s})}\right)^2}$$

Sehingga didapatkan aproksimasi ekspektasi

$$E\left[\log\left(1 + e^{Y(\boldsymbol{s})}\right) |\boldsymbol{Y}^*, \boldsymbol{\beta}\,, \boldsymbol{\Sigma}\right] = \log\left(1 + e^{Y_m(\boldsymbol{s})}\right) - \frac{1}{2}\frac{e^{Y_m(\boldsymbol{s})}}{(1 + e^{Y_m(\boldsymbol{s})})^2}[H(\boldsymbol{\varepsilon}_m)^{-1}]_{ii}$$

Didapatkan aproksimasi untuk bentuk ekspektasi yang ekuivalen dengan persamaan (4.41)

$$\varphi\left[\boldsymbol{\beta}\,, \boldsymbol{\Sigma}, \widehat{\boldsymbol{\beta}}^{(l)}\,, \widehat{\boldsymbol{\Sigma}}^{(l)}\right] = E[L_c[\boldsymbol{Y}^*, \boldsymbol{\varepsilon} \,|\boldsymbol{\beta}\,, \boldsymbol{\Sigma}]|\boldsymbol{Y}^*, \widehat{\boldsymbol{\beta}}^{(l)}\,, \widehat{\boldsymbol{\Sigma}}^{(l)}]$$

(Lanjutan)

$$\cong -\sum_{s\in T}\left\{\log\left(1+e^{Y_m(\boldsymbol{s})}\right)-\frac{1}{2}\frac{e^{Y_m(s)}}{\left(1+e^{Y_m(s)}\right)^2}\left[H(\boldsymbol{\varepsilon}_m)^{-1}\right]_{ii}\right\}$$

$$+\sum_{s\in T} Y_m(\boldsymbol{s})Y^*(\boldsymbol{s})-\frac{1}{2}\log(\det\boldsymbol{\Sigma})$$

$$-\frac{1}{2}\{tr[-\boldsymbol{\Sigma}^{-1}H(\boldsymbol{\varepsilon}_m)^{-1}]+\ \boldsymbol{\varepsilon}_m^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_m\}-\frac{n}{2}\log 2\pi$$