SKRIPSI

PENGEMBANGAN METODE PEMBOBOTAN BERBASIS JARINGAN SARAF GRAF PADA REGRESI TERBOBOTI GEOGRAFIS DAN IMPLIKASINYA TERHADAP INFERENSI

DEVELOPMENT OF GRAPH NEURAL NETWORK-BASED WEIGHTING METHOD FOR GEOGRAPHICALLY WEIGHTED REGRESSION AND ITS IMPLICATIONS FOR INFERENCE

Kurniawan Chandra Wijaya
22/497908/PA/21466

PROGRAM SARJANA PROGRAM STUDI STATISTIKA
DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS GADJAH MADA
YOGYAKARTA

2025

SKRIPSI

PENGEMBANGAN METODE PEMBOBOTAN BERBASIS JARINGAN SARAF GRAF PADA REGRESI TERBOBOTI GEOGRAFIS DAN IMPLIKASINYA TERHADAP INFERENSI

DEVELOPMENT OF GRAPH NEURAL NETWORK-BASED WEIGHTING METHOD FOR GEOGRAPHICALLY WEIGHTED REGRESSION AND ITS IMPLICATIONS FOR INFERENCE

Diajukan untuk memenuhi salah satu syarat memperoleh derajat
Sarjana Program Sarjana Program Studi Statistika

Kurniawan Chandra Wijaya
22/497908/PA/21466

PROGRAM SARJANA PROGRAM STUDI STATISTIKA
DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS GADJAH MADA
YOGYAKARTA

2025

HALAMAN PENGESAHAN

SKRIPSI

PENGEMBANGAN METODE PEMBOBOTAN BERBASIS JARINGAN SARAF GRAF PADA REGRESI TERBOBOTI GEOGRAFIS DAN IMPLIKASINYA TERHADAP INFERENSI

Telah dipersiapkan dan disusun oleh

Kurniawan Chandra Wijaya
22/497908/PA/21466

Telah dipertahankan di depan Tim Penguji
pada tanggal tanggal bulan tahun

Susunan Tim Penguji

Prof. Dr. Abdurakhman, S.Si., M.Si. Pembimbing Utama	Ketua Penguji Ketua Tim Penguji
	Penguji 1 Penguji
	Penguji 2 Penguji

PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama	:	Kurniawan Chandra Wijaya
NIM	:	22/497908/PA/21466
Tahun terdaftar	:	2025
Program Studi	:	Program Sarjana Program Studi Statistika
Fakultas	:	Fakultas Matematika dan Ilmu Pengetahuan Alam

menyatakan bahwa dalam dokumen ilmiah skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan sumbernya secara lengkap dalam daftar pustaka.

Dengan demikian saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi dan apabila dokumen ilmiah skripsi ini di kemudian hari terbukti merupakan plagiasi dari hasil karya penulis lain dan/atau dengan sengaja mengajukan karya atau pendapat yang merupakan hasil karya penulis lain, maka penulis bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Yogyakarta, tanggal bulan tahun

Materai Rp10.000,00

Kurniawan Chandra Wijaya

22/497908/PA/21466

Karya sederhana ini penulis persembahkan untuk ibu,

bapak, dan diri penulis.

Man jadda wajada.

Barang siapa bersungguh-sungguh, maka ia akan berhasil.

PRAKATA

Bismillahirrahmanirrahim. Puji syukur kepada Allah SWT atas limpahan berkah dan rahmat yang diberikan sehingga penulis dapat menyelesaikan tugas akhir yang berjudul ... sebagai salah satu syarat mendapakan gelar sarjana di program studi statisika Universias Gadjah Mada. Tugas akhir ini merupakan salah satu hasil dari proses belajar yang penulis jalani selama empat tahun berkuliah di universitas tercinta.

Penulis sadar bahwa dalam proses penyusunan tugas akhir ini tidak akan berjalan tanpa adanya dukungan dan bimbingan dari berbagai pihak. Maka dari itu, penulis ingin mengucapkan terima kasih kepada:

1. Bapak Prof. Dr.Eng Kuwat Triyana, M.Si., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Gadjah Mada.
2. Bapak Dr. Nanang Susyanto, M.Sc., selaku Ketua Departemen Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Gadjah Mada.
3. Bapak Prof. Dr.rer.nat Dedi Rosadi, M.Sc., selaku Ketua Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Gadjah Mada.

Seperti manusia pada umumnya, penulis menyadari bahwa tugas akhir ini masih jauh dari kata sempurna. Oleh karena itu, semua kritik dan saran dari pembaca menjadi masukan yang sangat berharga bagi penulis. Semoga tugas akhir ini dapat bermenfaat untuk pembaca. Terima kasih.

Yogyakarta, Juni 2024

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

INTISARI

PENGEMBANGAN METODE PEMBOBOTAN BERBASIS JARINGAN SARAF GRAF PADA REGRESI TERBOBOTI GEOGRAFIS DAN IMPLIKASINYA TERHADAP INFERENSI

Oleh

Kurniawan Chandra Wijaya
22/497908/PA/21466

TULIS ABSTRAK,

Kata Kunci: kata kunccci.

ABSTRACT

DEVELOPMENT OF GRAPH NEURAL NETWORK-BASED WEIGHTING METHOD FOR GEOGRAPHICALLY WEIGHTED REGRESSION AND ITS IMPLICATIONS FOR INFERENCE

By

Kurniawan Chandra Wijaya
22/497908/PA/21466

TULIS ABSTRAK ENGGRES,

Keyword: keywordsssss

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perekonomian Indonesia senantiasa menghadapi dinamika yang kompleks seiring dengan variasi spasial dan temporal yang terjadi di setiap wilayah. Fenomena inflasi, misalnya, tidak hanya dipengaruhi oleh faktor makroekonomi secara nasional, tetapi juga oleh karakteristik lokal seperti struktur industri, tingkat pendidikan, dan infrastruktur wilayah. Penelitian **?** menunjukkan bahwa model *Space-Time Autoregressive* (STAR) dan *Generalized STAR* (GSTAR) mampu menangkap keterkaitan spasial antarwilayah dalam peramalan indeks harga konsumen (IHK) di beberapa kota besar di Sumatra. Namun, kedua model tersebut masih bersifat linear dan global sehingga belum sepenuhnya mampu menggambarkan ketidakstasioneran hubungan ekonomi antarwilayah dan antarwaktu.

Dalam literatur ekonometrika spasial, ketidakstasioneran spasial (*spatial heterogeneity*) dan ketergantungan spasial (*spatial dependence*) menjadi dua aspek penting yang perlu diperhatikan (**?**). Untuk mengakomodasi hal tersebut, **?** mengembangkan model *Geographically Weighted Regression* (GWR) yang memungkinkan koefisien regresi bervariasi antar lokasi. Model ini kemudian diperluas oleh **?** menjadi *Geographically and Temporally Weighted Regression* (GTWR) dengan menambahkan dimensi waktu, sehingga mampu menangkap dinamika spasial sekaligus temporal dalam satu kerangka analisis. Walaupun interpretatif, model GTWR masih bergantung pada asumsi bentuk fungsi kernel dan pemilihan *bandwidth* yang bersifat subjektif serta tidak mampu merepresentasikan hubungan nonlinear yang kompleks antara jarak spasial dan bobot pembobotan.

Perkembangan terkini di bidang *geospatial artificial intelligence* (GeoAI) menghadirkan pendekatan *machine learning* yang dapat mempelajari pola nonlinear secara adaptif. Salah satunya adalah *Geographically Weighted Artificial Neural*

Network (GWANN) yang dikembangkan oleh **?** serta *Spatial and Attribute Neural Network Weighted Regression* (SANNWR) oleh **?**. Keduanya berupaya memanfaatkan jaringan saraf tiruan untuk mengestimasi fungsi pembobot spasial yang kompleks. Lebih lanjut, **?** memperkenalkan model *Geographically Neural Network Weighted Regression* (GNNWR) dan *Geographically and Temporally Neural Network Weighted Regression* (GTNNWR), yang memadukan kerangka *varying coefficient* dengan kemampuan pembelajaran representasi nonlinear dari *neural networks*. Model tersebut menunjukkan peningkatan akurasi yang signifikan dalam memetakan hubungan spasial-temporal pada data lingkungan dan sosial.

Di sisi lain, **?** mengusulkan *Spatial Regression Graph Convolutional Neural Networks* (SRGCNN) yang mengintegrasikan *graph convolutional neural network* (GCN) ke dalam paradigma regresi spasial. Pendekatan ini memperlakukan data spasial sebagai graf, dengan simpul merepresentasikan lokasi dan sisi (*edges*) merepresentasikan keterhubungan spasial antarunit observasi. Dengan mekanisme *message passing* yang mengagregasi informasi dari tetangga terdekat, SRGCNN tidak hanya mampu menangani struktur data non-Euclidean, tetapi juga mendukung pembelajaran semi-terawasi (*semi-supervised learning*) yang memungkinkan model belajar dari data yang sebagian tidak teramati. Pendekatan ini membuka peluang besar untuk memperkuat model regresi spasial konvensional melalui pembobotan berbasis graf.

Berdasarkan perkembangan tersebut, penelitian ini berupaya mengintegrasikan prinsip pembobotan adaptif berbasis GNN sebagaimana pada GNNWR dan SRGCNN ke dalam kerangka regresi spasial-temporal seperti GTWR. Model yang diusulkan adalah *Graph Neural Network–Geographically and Temporally Varying Coefficient* (GNN-GTVC) serta *Graph Neural Network–Geographically and Temporally Weighted Regression* (GNN-GTWR) . Kedua model ini dirancang untuk menangkap dinamika spasial-temporal dengan lebih fleksibel melalui pembelajaran berbasis graf yang mengoptimasi matriks bobot. Dalam konteks empiris, model ini akan diterapkan pada analisis spasial-temporal inflasi antarprovinsi di Indonesia periode 2024–2025, guna memahami disparitas inflasi antarwilayah dan dinamika faktor-faktor ekonomi yang memengaruhinya. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi metodologis dalam bidang ekonometrika spasial sekaligus implikasi kebijakan dalam pengendalian inflasi regional di Indonesia.

1.2 Pembatasan Masalah

Penelitian ini berfokus pada pengembangan dan penerapan model GNN-GTWR dan GNN-GTVC untuk menganalisis dinamika spasial-temporal inflasi antarprovinsi di Indonesia. Ruang lingkup penelitian dibatasi pada level provinsi sebagai unit analisis dengan periode pengamatan Januari 2024 hingga Agustus 2025. Variabel yang digunakan meliputi indikator makroekonomi utama seperti inflasi, upah minimum, tingkat pengangguran, dan indeks pembangunan manusia (IPM) yang bersumber dari Badan Pusat Statistik (BPS) dan Bank Indonesia (BI). Penelitian ini tidak mencakup analisis mikroekonomi individu maupun agregasi lintas negara, serta tidak membahas secara mendalam mekanisme kausalitas antarvariabel.

1.3 Tujuan dan Manfaat Penelitian

Tujuan utama penelitian ini adalah mengembangkan model pembelajaran spasial-temporal berbasis graf untuk menganalisis variasi inflasi antarwilayah di Indonesia. Secara khusus, penelitian ini bertujuan untuk:

1. mengonstruksi model *Graph Neural Network–Geographically and Temporally Weighted Regression* (GNN-GTWR) dan *Graph Neural Network–Geographically and Temporally Varying Coefficient* (GNN-GTVC) sebagai perluasan dari model GTWR dan VCM yang mampu menangkap ketidakstasioneran spasial dan temporal secara simultan melalui pembobotan adaptif berbasis GNN;

2. menerapkan model yang dikembangkan pada studi kasus inflasi antarprovinsi di Indonesia periode 2024–2025 untuk mengidentifikasi pola spasial-temporal dan faktor-faktor ekonomi yang berpengaruh; dan

3. mengevaluasi kinerja model terhadap model konvensional (GTWR dan GWR) dalam hal ketepatan estimasi, stabilitas, dan interpretabilitas koefisien lokal.

Secara praktis, hasil penelitian ini diharapkan dapat memberikan kontribusi bagi lembaga perencana kebijakan, khususnya Bank Indonesia dan Kementerian Keuangan, dalam memahami dinamika inflasi regional dan merancang kebijakan pengendalian harga yang adaptif terhadap kondisi spasial-temporal.

1.4 Tinjauan Pustaka

Penelitian mengenai regresi spasial dan spasial-temporal telah berkembang pesat dalam dua dekade terakhir. Model klasik seperti GWR (**?**) dan GTWR (**?**) menjadi dasar utama dalam analisis hubungan spasial yang tidak stasioner. Namun, keterbatasannya pada asumsi linearitas dan bentuk kernel mendorong munculnya model berbasis kecerdasan buatan seperti GWANN (**?**), SANNWR (**?**), dan GNNWR (**?**). Di sisi lain, pendekatan *graph-based learning* seperti SRGCNN (**?**) dan GSTRGCN (**?**) memperlihatkan potensi besar dalam memodelkan struktur dependensi spasial-temporal yang kompleks melalui pembelajaran representasi graf. Berangkat dari temuan tersebut, penelitian ini berupaya menggabungkan kekuatan metodologis antara regresi terboboti spasial-temporal dan pembelajaran berbasis graf guna mengembangkan model yang lebih adaptif, akurat, dan interpretatif.

1.5 Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini meliputi studi literatur dan studi kasus. Studi literatur dilakukan untuk mengkaji teori dan penelitian terdahulu terkait regresi spasial-temporal, jaringan saraf graf, serta pendekatan pembelajaran semi-terawasi yang relevan dengan pengembangan model GNN-GTWR dan GNN-GTVC. Studi kasus dilakukan dengan menggunakan data sekunder dari BPS dan BI, yang mencakup 38 provinsi di Indonesia untuk periode Januari 2024 hingga Agustus 2025. Analisis dan komputasi dilakukan menggunakan bahasa pemrograman Python dengan pustaka *PyTorch Geometric* untuk implementasi jaringan saraf graf, serta perangkat lunak statistik seperti R atau GeoDa untuk validasi spasial.

1.6 Sistematika Penulisan

Pada penyusunan skripsi ini, penulis mengacu pada sistematika penulisan sebagai berikut.

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang, pembatasan masalah, tujuan dan manfaat penelitian, tinjauan pustaka, metodologi penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini membahas dasar-dasar teori yang digunakan, termasuk teori regresi spasial-temporal, jaringan saraf graf, dan pembelajaran semi-terawasi.

BAB III PENGEMBANGAN PEMBOBOTAN BERBASIS JARINGAN SARAF GRAF UNTUK REGRESI SPASIAL-TEMPORAL

Bab ini menjelaskan rancangan model GNN-GTVC dan GNN-GTWR, formulasi matematis, serta prosedur estimasi dan validasi model.

BAB IV STUDI KASUS

Bab ini berisi hasil implementasi model pada data inflasi antarprovinsi di Indonesia, analisis spasial-temporal, serta perbandingan kinerja model.

BAB V PENUTUP

Bab ini menyajikan kesimpulan yang diperoleh dari hasil penelitian dan memberikan saran untuk penelitian selanjutnya.

BAB II

LANDASAN TEORI

Pada bab ini akan dibahas mengenai konsep, teori, dan referensi yang menjadi dasar dalam melakukan analisis suatu metode statistik. Konsep dasar yang dibahas pada bab ini adalah aljabar matriks, teori ukuran dan probabilitas, konvergensi dan laju pertumbuhan, teori probabilitas asimtotik, statistika inferensial, analisis regresi linear, analisis regresi spasial, jaringan saraf tiruan, dan jaringan saraf graf.

2.1 Aljabar Matriks

Aljabar linear merupakan salah satu fondasi matematis terpenting dalam statistika, ekonometrika, serta pembelajaran mesin. Banyak metode statistika tradisional maupun modern dapat direpresentasikan dalam aljabar matriks, seperti jaringan saraf graf atau *graph neural networks* (GNN). Oleh karena itu, pemahaman yang kuat mengenai struktur vektor dan matriks serta sifat-sifat aljabarnya diperlukan sebelum membahas metode regresi terboboti geografis.

2.1.1 Ruang Vektor dan Matriks

Pembahasan aljabar linear umumnya berangkat dari konsep lapangan atau *field* sebagai struktur aljabar dasar tempat bilangan berlaku. Dari lapangan, dibangun ruang vektor, kemudian matriks sebagai representasi transformasi linear, dan tensor sebagai generalisasi multidimensi.

Definisi 2.1.1 (**Lapangan, ?**) *Misalkan K adalah subhimpunan dari bilangan kompleks $\mathbb{C}$, K disebut sebagai lapangan apabila memenuhi kondisi berikut.*

(a) Apabila x, y adalah elemen dari K, maka $x + y$ dan xy juga merupakan elemen dari K.

(b) Apabila $x \in K$, maka $-x$ juga merupakan elemen dari K. Lebih lanjut, jika

$x \neq 0$, maka x^{-1} merupakan elemen dari K.

(c) 0 dan 1 merupakan elemen dari K.

Berdasarkan definisi di atas, dapat diperhatikan bahwa $\mathbb{R}$ dan $\mathbb{C}$ merupakan lapangan.

Contoh 2.1.2 Apabila dinotasikan $\mathbb{Q}$ sebagai himpunan bilangan rasional, yaitu

$$\mathbb{Q} = \left\{ \frac{m}{n} : m, n \in \mathbb{Z},\, n \neq 0 \right\}, \tag{2.1.1}$$

maka dapat dinyatakan bahwa $\mathbb{Q}$ membentuk suatu lapangan. Misalkan $\frac{a}{b}, \frac{c}{d} \in \mathbb{Q}$ dengan $a, b, c, d \in \mathbb{Z}$ dan $b, d \neq 0$. Jelas bahwa

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \in \mathbb{Q}, \tag{2.1.2}$$

sehingga sifat tertutup terhadap penjumlahan, sebagai salah satu aksioma lapangan, terpenuhi. Lebih lanjut, untuk setiap $\frac{a}{b} \in \mathbb{Q}$ berlaku pula bahwa $-\frac{a}{b} \in \mathbb{Q}$. Selain itu, jika $a \neq 0$, maka invers perkalian $\frac{a}{b}$ diberikan oleh

$$\left(\frac{a}{b}\right)^{-1} = \frac{b}{a} \in \mathbb{Q}. \tag{2.1.3}$$

Dengan demikian, keberadaan invers aditif dan invers perkalian juga terjamin. Akhirnya, apabila dipilih $\frac{a}{b} = \frac{0}{1}$ dan $\frac{a}{b} = \frac{1}{1}$, diperoleh bahwa elemen identitas penjumlahan 0 dan identitas perkalian 1 termasuk dalam $\mathbb{Q}$. Di sisi lain, dapat dengan mudah diperiksa bahwa $\mathbb{Z}$ bukan merupakan lapangan, karena apabila $n \in \mathbb{Z}$ dan $n \neq 0$, jelas bahwa $n^{-1} \notin \mathbb{Z}$, kecuali untuk kasus $n = 1$ atau $n = -1$.

Dalam ruang vektor dan matriks, dikenal juga istilah sublapangan. Apabila K dan L keduanya merupakan lapangan, serta dimisalkan bahwa $K \subseteq L$, maka K dapat disebut sebagai sublapangan atau *subfield* dari L. Elemen-elemen dari lapangan disebut sebagai bilangan atau skalar (**?**).

Definisi 2.1.3 (**Ruang Vektor, ?**) *Suatu ruang vektor V atas suatu lapangan K*

adalah sebuah himpunan objek yang dapat dijumlahkan dan dikalikan dengan elemen-elemen dari K, sedemikian rupa sehingga hasil penjumlahan dua elemen dari V merupakan elemen V kembali, dan hasil perkalian sebuah elemen V dengan sebuah elemen dari K juga merupakan elemen dari V. Selain itu, sifat-sifat berikut dipenuhi:

(a) $\forall u, v, w \in V$ *berlaku* $(u + v) + w = u + (v + w)$.

(b) $\exists\, 0 \in V$ *sedemikian sehingga* $\forall u \in V$ *berlaku* $u + 0 = u$.

(c) $\forall u \in V$, $\exists - u \in V$ *sedemikian sehingga* $u + (-u) = 0$.

(d) $\forall u, v \in V$ *berlaku* $u + v = v + u$.

(e) $\forall c \in K$ *dan* $\forall u, v \in V$ *berlaku* $c\,(u + v) = cu + cv$.

(f) $\forall a, b \in K$ *dan* $\forall v \in V$ *berlaku* $(a + b)v = av + bv$.

(g) $\forall a, b \in K$ *dan* $\forall v \in V$ *berlaku* $(ab)v = a(bv)$.

(h) $\forall u \in V$ *berlaku* $1 \cdot u = u$, *dengan* 1 *adalah elemen identitas pada* K.

Dalam mendefinisikan ruang vektor, lapangan tempat ruang vektor berada harus didefinisikan dengan spesifik, misalnya $\mathbb{C}^n$ merupakan ruang vektor atas $\mathbb{C}$, tetapi $\mathbb{R}^n$ bukan ruang vektor atas $\mathbb{C}$. $\mathbb{R}^n$ merupakan ruang vektor atas $\mathbb{R}$.

Definisi 2.1.4 (**Ruang Matriks, (?)**) *Misalkan K merupakan lapangan dan $m, n \geq 1 \in \mathbb{Z}$. Matriks adalah suatu larik atau* array *dari bilangan-bilangan dalam K yang dinotasikan sebagai*

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Notasi matriks pada definisi di atas dapat dipersingkat dengan menulis (a_{ij}), $i = 1, 2, \ldots, m$ dan $j = 1, 2, \ldots, n$. Matriks tersebut merupakan matriks $m \times n$ yang berarti bahwa matriks tersebut memiliki m-baris dan n-kolom. Sebagai contoh, kolom pertama dari matriks tersebut adalah

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}$$

dan baris pertamanya adalah $(a_{11}, a_{12}, \ldots, a_{1n})$. Nilai a_{ij} disebut sebagai komponen dari matriks.

Dua buah matriks dikatakan sama apabila keduanya memiliki ukuran yang sama dan elemen-elemen pada posisi yang bersesuaian juga sama. Oleh karena itu, jika $\mathbf{A} = (a_{ij})$ dan $\mathbf{B} = (b_{ij})$, maka $\mathbf{A} = \mathbf{B}$ apabila $a_{ij} = b_{ij}$, untuk seiap i, j.

Definisi 2.1.5 (Kebebasan Linear, (?)) *Misalkan V ruang vektor atas lapangan K dan misalkan pula $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ elemen-elemen dari V, maka $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ dikatakan bebas linear jika dan hanya jika kapan pun $a_1, a_2, \ldots, a_n$ merupakan bilangan sehingga*

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = \mathbf{0}, \tag{2.1.4}$$

maka $a_i = 0, \forall i = 1, 2, \ldots, n$.

Vektor $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ dikatakan bergantung linear terhadap K apabila terdapat elemen-elemen $a_1, a_2, \ldots, a_n$ di K yang semuanya tidak sama dengan nol, sedemikian sehingga Persamaan (2.4) terpenuhi.

Contoh 2.1.6 Misalkan $V = K^n$ dan vektor $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ didefinisikan sebagai berikut.

$$\begin{aligned} \mathbf{v}_1 &= (1, 0, \ldots, 0) \\ &\vdots \\ \mathbf{v}_n &= (0, 0, \ldots, 1). \end{aligned}$$

Vektor-vektor $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ dikatakan bebas linear. Misalkan $a_1, a_2, \ldots, a_n$ merupakan bilangan-bilangan sedemikian sehingga Persamaan (2.4) terpenuhi. Sebab

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = (a_1, a_2, \ldots, a_n), \tag{2.1.5}$$

dapat disimpulkan bahwa $a_i = 0, \forall i = 1, 2, \ldots, n$.

Definisi 2.1.7 (***Span* atau Jangkauan Linear, (?)**) *Misalkan V ruang vektor atas lapangan K dan $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\} \subseteq V$.* Span *dari S, ditulis* $\operatorname{span}(S)$, *adalah himpunan semua kombinasi linear dari elemen-elemen S, yaitu*

$$\operatorname{span}(S) = \left\{ \sum_{i=1}^{m} a_i\mathbf{v}_i : a_i \in K \right\}. \tag{2.1.6}$$

Dengan kata lain, $\operatorname{span}(S)$ *adalah subruang terkecil dari V yang memuat S. Di sisi lain,* span *dari himpunan kosong* $\{\}$ *didefinisikan sebagai* $\{0\}$.

Contoh 2.1.8 Pada $\mathbb{R}^2$, ambil $S = \{(1,0),(0,1)\}$, maka

$$\operatorname{span}(S) = \{a(1,0) + b(0,1) : a, b \in \mathbb{R}\} = \mathbb{R}^2. \tag{2.1.7}$$

Hal ini berarti dua vektor standar membentang atau menjangkau (*spanning*) seluruh bidang $\mathbb{R}^2$. Pada $\mathbb{R}^3$, ambil $S = \{(1,0,0),(0,1,0)\}$, maka

$$\operatorname{span}(S) = \{a(1,0,0) + b(0,1,0) : a, b \in \mathbb{R}\} = \{(x,y,0) : x, y \in \mathbb{R}\}, \tag{2.1.8}$$

yaitu bidang xy dalam $\mathbb{R}^3$.

Definisi 2.1.9 (**Basis, (?)**) *Suatu basis V adalah himpunan vektor-vektor di V yang bebas linear dan menjangkau (*spanning*) V.*

Contoh 2.1.10 Perhatikan vektor-vektor $\mathbf{v}_1 = (1, 2, -4)$ dan $\mathbf{v}_2 = (7, -5, 6)$ di dalam $\mathbb{R}^3$. Jelas bahwa $\mathbf{v}_1$ dan $\mathbf{v}_2$ bebas linear, sebab tidak terdapat skalar $c \in \mathbb{R}$ sehingga $\mathbf{v}_2 = c\mathbf{v}_1$. Namun, himpunan $\{\mathbf{v}_1, \mathbf{v}_2\}$ hanya terdiri dari dua vektor, sehingga $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ paling jauh merupakan subruang berdimensi 2, yaitu sebuah bidang melalui titik asal di $\mathbb{R}^3$. Sebab $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\} \subsetneq \mathbb{R}^3$, maka himpunan ini tidak dapat dijadikan basis dari $\mathbb{R}^3$.

Teorema 2.1.11 (**Karakterisasi Basis, (?)**) *Sebuah* list *dari vektor-vektor* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ *merupakan basis dari* V *jika dan hanya jika* $\forall \mathbf{v} \in V$ *dapat dituliskan secara unik dalam bentuk*

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n, \tag{2.1.9}$$

dengan $a_1, a_2, \ldots, a_n \in \mathbb{R} \cup \mathbb{C}$.

Bukti. Pertama andaikan bahwa $\mathbf{v}_1, \ldots, \mathbf{v}_n$ adalah basis dari V. Sebab $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ menjangkau V, maka untuk setiap $\mathbf{v} \in V$ terdapat skalar $a_1, \ldots, a_n \in F$ sehingga Persamaan (2.9) terpenuhi. Untuk menunjukkan bahwa representasi tersebut unik, misalkan terdapat skalar $c_1, \ldots, c_n \in F$ dengan

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n. \tag{2.1.10}$$

Dengan mengurangkan Persamaan (2.10) dari Persamaan (2.9) akan diperoleh

$$0 = (a_1 - c_1)\mathbf{v}_1 + \cdots + (a_n - c_n)\mathbf{v}_n. \tag{2.1.11}$$

Sebab $\mathbf{v}_1, \ldots, \mathbf{v}_n$ bebas linear, maka setiap koefisien harus nol, yaitu $a_k - c_k = 0$ untuk $k = 1, \ldots, n$. Dengan demikian $a_1 = c_1, \ldots, a_n = c_n$, sehingga representasi (2.9) bersifat unik. Hal ini menyelesaikan pembuktian arah pertama.

Sebaliknya, andaikan setiap $\mathbf{v} \in V$ dapat dituliskan secara unik dalam bentuk

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n, \tag{2.1.12}$$

maka jelas $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ menjangkau V. Untuk membuktikan bahwa $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ bebas linear, andaikan terdapat $a_1, \ldots, a_n \in F$ sehingga

$$0 = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n. \tag{2.1.13}$$

Sebab representasi (2.12) bersifat unik, khususnya untuk $\mathbf{v} = 0$, maka harus berlaku

$a_1 = \cdots = a_n = 0$. Jadi $\mathbf{v}_1, \ldots, \mathbf{v}_n$ bebas linear. Dengan demikian $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ adalah basis V. ■

Definisi 2.1.12 (**Ruang Vektor Berdimensi-Hingga (?)**) *Suatu ruang vektor V atas medan F dikatakan* berdimensi-hingga *apabila terdapat suatu basis $\{v_1, \ldots, v_n\}$ dari V yang terdiri atas sejumlah hingga (n buah) vektor. Dengan kata lain, V berdimensi-hingga jika V dapat dijangkau oleh suatu himpunan vektor berukuran hingga.*

Definisi 2.1.13 (**Dimensi (?)**) *Dimensi dari suatu ruang vektor berdimensi-hingga V adalah banyaknya elemen dalam setiap basis V. Sebab semua basis dari V memiliki panjang yang sama, maka bilangan ini terdefinisi dengan baik. Dimensi V dilambangkan dengan* $\dim V$.

Contoh 2.1.14 Jika $U = \{(x, x, y) \in \mathbb{R}^3 : x, y \in \mathbb{R}\}$, maka $\dim U = 2$ karena $(1, 1, 0), (0, 0, 1)$ merupakan basis dari U.

2.1.2 Sifat-Sifat Matriks

Seluruh bagian ini bekerja di atas suatu lapangan K (umumnya $K = \mathbb{R}$ atau $K = \mathbb{C}$). Untuk $m, n \geq 1$, himpunan seluruh matriks $m \times n$ berelemen di K dilambangkan dengan $\mathrm{Mat}_{m\times n}(K)$. Dengan penjumlahan dan perkalian skalar yang didefinisikan komponen-demi-komponen, $\mathrm{Mat}_{m\times n}(K)$ membentuk ruang vektor atas lapangan K (**?**).

Definisi 2.1.15 (**Kesamaan, Penjumlahan, dan Perkalian dengan Skalar**) *Untuk $\mathbf{A} = (a_{ij})$, $\mathbf{B} = (b_{ij}) \in \mathrm{Mat}_{m\times n}(K)$ dan $c \in K$, berlaku beberapa hal berikut.*

(a) $\mathbf{A} = \mathbf{B}$ *jika dan hanya jika $a_{ij} = b_{ij}$ untuk semua i, j;*

(b) $(\mathbf{A} + \mathbf{B})_{ij} = a_{ij} + b_{ij}$ *(didefinisikan hanya bila ukuran sama); dan*

(c) $(c\,\mathbf{A})_{ij} = c\,a_{ij}$.

Definisi 2.1.16 (**Perkalian Matriks, (?)**) *Jika* $\mathbf{A} \in \mathrm{Mat}_{m\times n}(K)$ *dan* $\mathbf{B} \in \mathrm{Mat}_{n\times p}(K)$, *maka hasil kali* $\mathbf{AB} \in \mathrm{Mat}_{m\times p}(K)$ *didefinisikan oleh*

$$(\mathbf{AB})_{ik} = \sum_{j=1}^{n} a_{ij}\, b_{jk} \qquad (1 \le i \le m,\ 1 \le k \le p). \tag{2.1.14}$$

Contoh 2.1.17 Perkalian matriks berukuran 3×2 dan 2×4:

$$\underbrace{\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}}_{3\times 2} \underbrace{\begin{pmatrix} 6 & 5 & 4 & 3 \\ 2 & 1 & 0 & -1 \end{pmatrix}}_{2\times 4} = \underbrace{\begin{pmatrix} 10 & 7 & 4 & 1 \\ 26 & 19 & 12 & 5 \\ 42 & 31 & 20 & 9 \end{pmatrix}}_{3\times 4}. \tag{2.1.15}$$

Proposisi 2.1.18 (Sifat-Sifat Elementer Perkalian Matriks, (?)) *Untuk ukuran yang sesuai, berlaku:*

(a) sifat asosiatif, yaitu $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$*;*

(b) sifat distributif, yaitu $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$ *dan* $(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$*;*

(c) sifat identitas, yaitu $\mathbf{I}_m\mathbf{A} = \mathbf{A}$ *dan* $\mathbf{AI}_n = \mathbf{A}$ *untuk* $\mathbf{A} \in \mathrm{Mat}_{m\times n}(K)$*; dan*

(d) dapat terjadi $\mathbf{AB} \neq \mathbf{BA}$.

Bukti. Di bawah ini adalah pembuktian untuk keempat sifat elementer perkalian matriks.

(a) Untuk $(\mathbf{AB})\mathbf{C}$, berlaku

$$\begin{aligned} \big((\mathbf{AB})\mathbf{C}\big)_{ik} &= \sum_{\ell=1}^{p} (\mathbf{AB})_{i\ell}\, c_{\ell k} \\ &= \sum_{\ell=1}^{p} \Big(\sum_{j=1}^{n} a_{ij} b_{j\ell}\Big) c_{\ell k} \\ &= \sum_{j=1}^{n} \sum_{\ell=1}^{p} a_{ij} b_{j\ell} c_{\ell k}. \end{aligned} \tag{2.1.16}$$

Di sisi lain, untuk $\mathbf{A}(\mathbf{BC})$ berlaku

$$
\begin{aligned}
\big(\mathbf{A}(\mathbf{BC})\big)_{ik} &= \sum_{j=1}^{n} a_{ij} \, (\mathbf{BC})_{jk} \\
&= \sum_{j=1}^{n} a_{ij} \Big(\sum_{\ell=1}^{p} b_{j\ell} c_{\ell k} \Big) \\
&= \sum_{j=1}^{n} \sum_{\ell=1}^{p} a_{ij} b_{j\ell} c_{\ell k}. \qquad (2.1.17)
\end{aligned}
$$

Keduanya merupakan matriks yang sama.

(b) Untuk $\mathbf{A}(\mathbf{B}+\mathbf{C})$, berlaku

$$
\begin{aligned}
\big(\mathbf{A}(\mathbf{B}+\mathbf{C})\big)_{ik} &= \sum_{j=1}^{n} a_{ij}(b_{jk} + c_{jk}) \\
&= \sum_{j=1}^{n} a_{ij} b_{jk} + \sum_{j=1}^{n} a_{ij} c_{jk} \\
&= (\mathbf{AB})_{ik} + (\mathbf{AC})_{ik}. \qquad (2.1.18)
\end{aligned}
$$

Kasus $(\mathbf{A}+\mathbf{B})\mathbf{C}$ serupa.

(c) Untuk $\mathbf{I}_m$ merupakan matriks identitas berukuran $m \times m$, berlaku

$$
\big(\mathbf{I}_m \mathbf{A}\big)_{ik} = \sum_{j=1}^{m} (\mathbf{I}_m)_{ij} a_{jk} = \sum_{j=1}^{m} \delta_{ij} a_{jk} = a_{ik}, \qquad (2.1.19)
$$

dan

$$
\big(\mathbf{A}\mathbf{I}_n\big)_{ik} = \sum_{j=1}^{n} a_{ij} (\mathbf{I}_n)_{jk} = \sum_{j=1}^{n} a_{ij} \delta_{jk} = a_{ik}, \qquad (2.1.20)
$$

dengan

$$
\delta_{ij} = \begin{cases} 1, & \text{jika } i = j, \\ 0, & \text{jika } i \neq j. \end{cases} \qquad (2.1.21)
$$

Dapat disimpulkan bahwa $\mathbf{IA} = \mathbf{A}$.

(d) Sifat ini akan dibuktikan dengan *counterexample*. Pertimbangkan perkalian matriks berikut.

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \tag{2.1.22}$$

Apabila urutan dibalik, akan memberikan matriks $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, sehingga kontradiksi apabila $\mathbf{AB} = \mathbf{BA}$ untuk semua $\mathbf{A}$ dan $\mathbf{B}$.

■

Definisi 2.1.19 (Perkalian Hadamard, (?)) *Jika diberikan* $\mathbf{A} = (a_{ij}) \in \mathrm{Mat}_{m\times n}$ *dan* $\mathbf{B} = (b_{ij}) \in \mathrm{Mat}_{m\times n}$*, maka perkalian Hadamard atau perkalian Schur dari* $\mathbf{A}$ *dan* $\mathbf{B}$ *merupakan perkalian elemen per elemen dari matriks tersebut atau dapat dituliskan sebagai berikut.*

$$\mathbf{A} \odot \mathbf{B} = (a_{ij}b_{ij}) \in \mathrm{Mat}_{m\times n}. \tag{2.1.23}$$

Notasi perkalian Hadamard $\odot$ dalam beberapa literatur dapat dituliskan juga dengan $\circ$.

Contoh 2.1.20 Misalkan

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}.$$

Maka hasil perkalian Hadamard adalah

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} 1\cdot 7 & 2\cdot 8 & 3\cdot 9 \\ 4\cdot 10 & 5\cdot 11 & 6\cdot 12 \end{bmatrix} = \begin{bmatrix} 7 & 16 & 27 \\ 40 & 55 & 72 \end{bmatrix}.$$

Definisi 2.1.21 (*Transpose*) *Jika* $\mathbf{A} \in \mathrm{Mat}_{m\times n}(K)$*, maka* transpose *dari* $\mathbf{A}$*, dilambangkan sebagai* $\mathbf{A}^\top$*, dengan* $\mathbf{A}^\top \in \mathrm{Mat}_{n\times m}(K)$ *didefinisikan oleh* $(\mathbf{A}^\top)_{ji} = a_{ij}$*.*

Contoh 2.1.22 Salah satu contoh *transpose* dari matriks adalah

$$\mathbf{A} = \begin{pmatrix} -5 & 2 & 4 \\ 3 & 6 & -2 \end{pmatrix} \iff \mathbf{A}^\top = \begin{pmatrix} -5 & 3 \\ 2 & 6 \\ 4 & -2 \end{pmatrix}, \tag{2.1.24}$$

sedangkan untuk vektor adalah

$$\mathbf{a} = \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix} \iff \mathbf{a}^\top = (2, -3, 1). \tag{2.1.25}$$

Notasi untuk *transpose* pada vektor $\mathbf{a}$ dapat juga dinotasikan dengan $\mathbf{a}'$. Di sisi lain, untuk skalar di lapangan K atau $c \in K$, $c^\top = c$.

Teorema 2.1.23 (Sifat-Sifat *Transpose*) *Untuk* $\mathbf{A}, \mathbf{B}$ *berukuran sesuai dan* $c \in K$, *berlaku:*

(a) $(\mathbf{A}^\top)^\top = \mathbf{A}$*;*

(b) $(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$*;*

(c) $(c\,\mathbf{A})^\top = c\,\mathbf{A}^\top$*; dan*

(d) $(\mathbf{AB})^\top = \mathbf{B}^\top\mathbf{A}^\top$.

Bukti. Berikut ini adalah bukti dari keempat sifat *transpose* tersebut.

(a) Elemen-elemen dari $(\mathbf{A}^\top)^\top$ akan sama dengan elemen-elemen dari $\mathbf{A}$, karena

$$\left((\mathbf{A}^\top)^\top\right)_{ij} = a_{ij} = (\mathbf{A})_{ij}. \tag{2.1.26}$$

(b) Elemen-elemen dari $(\mathbf{A} + \mathbf{B})^\top$ adalah $\left((\mathbf{A} + \mathbf{B})^\top\right)_{ij} = a_{ji} + b_{ji}$ yang merupakan elemen-elemen dari $\mathbf{A}^\top + \mathbf{B}^\top$.

(c) Elemen-elemen dari $(c\mathbf{A})^\top$ adalah ca_{ji} yang merupakan elemen-elemen dari $c\mathbf{A}^\top$.

(d) Dengan menuliskan $c_{ik} = (\mathbf{AB})_{ik} = \sum_j a_{ij}b_{jk}$ akan didapatkan

$$\begin{aligned}\left((\mathbf{AB})^\top\right)_{ki} &= c_{ik} \\ &= \sum_j a_{ij}b_{jk} \\ &= \sum_j (\mathbf{B}^\top)_{kj}(\mathbf{A}^\top)_{ji} \\ &= \left(\mathbf{B}^\top\mathbf{A}^\top\right)_{ki}. \end{aligned} \tag{2.1.27}$$

■

Definisi 2.1.24 (**Matriks Simetris, Diagonal, dan Identitas**) *Berikut ini adalah beberapa definisi terkait matriks simetris, diagonal, dan identitas.*

(a) $\mathbf{A} \in \mathrm{Mat}_{n\times n}(K)$ *disebut* simetris *jika* $\mathbf{A}^\top = \mathbf{A}$.

(b) $\mathbf{A}$ *disebut* diagonal *jika* $a_{ij} = 0$ *untuk* $i \neq j$.

(c) $\mathbf{I}_n$ *adalah matriks identitas berukuran* n*, dengan elemen diagonal* 1 *dan selainnya* 0.

Contoh 2.1.25 Matriks $\mathbf{A} = \begin{pmatrix} 3 & -2 & 4 \\ -2 & 10 & -7 \\ 4 & -7 & 9 \end{pmatrix}$ adalah matriks simetris karena $\mathbf{A}^\top = \mathbf{A}$.

Definisi 2.1.26 (**Matriks Idempoten, (?)**) *Suatu matriks persegi* $\mathbf{A}$ *dikatakan idempoten jika dan hanya jika*

$$\mathbf{AA} = \mathbf{A}^2 = \mathbf{A}. \tag{2.1.28}$$

Contoh 2.1.27 Matriks proyeksi merupakan contoh klasik dari matriks idempoten. Misalkan

$$\mathbf{P} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Dapat diverifikasi bahwa $\mathbf{P}$ adalah matriks idempoten karena

$$\mathbf{P}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \mathbf{P}.$$

Contoh lain adalah matriks proyeksi ortogonal ke subruang yang direntang oleh vektor $\mathbf{u} = (1, 1)^\top$, yaitu

$$\mathbf{P}_u = \frac{\mathbf{u}\mathbf{u}^\top}{\mathbf{u}^\top\mathbf{u}} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Dapat diperiksa bahwa $\mathbf{P}_u^2 = \mathbf{P}_u$, sehingga $\mathbf{P}_u$ juga idempoten.

Definisi 2.1.28 (Jejak atau *Trace*) *Untuk* $\mathbf{A} = (a_{ij}) \in \mathrm{Mat}_{n\times n}(K)$*, jejak atau* trace *didefinisikan sebagai*

$$\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}. \tag{2.1.29}$$

Contoh 2.1.29 Misalkan

$$\mathbf{A} = \begin{pmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \\ 2 & 6 & 5 \end{pmatrix}.$$

Jejak dari matriks $\mathbf{A}$ adalah jumlah elemen-elemen pada diagonal utamanya, yaitu

$$\mathrm{tr}(\mathbf{A}) = a_{11} + a_{22} + a_{33} = 3 + 5 + 5 = 13.$$

Perhatikan bahwa untuk matriks identitas $\mathbf{I}_n$ berukuran $n \times n$, berlaku $\mathrm{tr}(\mathbf{I}_n) = n$ karena setiap elemen diagonal bernilai 1.

Teorema 2.1.30 (Sifat-Sifat *Trace*, (?)) *Untuk ukuran yang sesuai, berlaku:*

(a) $\mathrm{tr}(\alpha\mathbf{A} + \beta\mathbf{B}) = \alpha\ \mathrm{tr}(\mathbf{A}) + \beta\ \mathrm{tr}(\mathbf{B})$*;*

(b) $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$*, lebih umum* $\mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{BCA})$*; dan*

(c) jika $\mathbf{B}$ *invertibel, maka* $\mathrm{tr}(\mathbf{B}^{-1}\mathbf{AB}) = \mathrm{tr}(\mathbf{A})$.

Bukti. Berikut ini adalah bukti-bukti untuk ketiga sifat di atas.

(a)

$$\begin{aligned}\mathrm{tr}(\alpha\mathbf{A}+\beta\mathbf{B}) &= \sum_i(\alpha a_{ii}+\beta b_{ii})\\ &= \alpha\sum_i a_{ii}+\beta\sum_i b_{ii}\\ &= \alpha\ \mathrm{tr}(\mathbf{A})+\beta\ \mathrm{tr}(\mathbf{B}).\end{aligned}\tag{2.1.30}$$

(b) Untuk dua faktor,

$$\begin{aligned}\mathrm{tr}(\mathbf{AB}) &= \sum_i(\mathbf{AB})_{ii}\\ &= \sum_i\sum_j a_{ij}b_{ji}\\ &= \sum_j\sum_i b_{ji}a_{ij}\\ &= \sum_j(\mathbf{BA})_{jj}\\ &= \mathrm{tr}(\mathbf{BA}).\end{aligned}\tag{2.1.31}$$

Untuk tiga faktor,

$$\begin{aligned}\mathrm{tr}(\mathbf{ABC}) &= \sum_i(\mathbf{ABC})_{ii}\\ &= \sum_i\sum_j\sum_k a_{ij}b_{jk}c_{ki}\\ &= \sum_j\sum_k\sum_i b_{jk}c_{ki}a_{ij}\\ &= \mathrm{tr}(\mathbf{BCA}).\end{aligned}\tag{2.1.32}$$

(c) Dengan sifat (b), dapat diketahui bahwa

$$\begin{aligned}\mathrm{tr}(\mathbf{B}^{-1}\mathbf{AB}) &= \mathrm{tr}(\mathbf{ABB}^{-1}) \\ &= \mathrm{tr}(\mathbf{AI}) \\ &= \mathrm{tr}(\mathbf{A}). \end{aligned} \tag{2.1.33}$$

■

Definisi 2.1.31 (Matriks Semi-Definit Positif) *Misalkan* $\mathbf{A}$ *merupakan suatu matriks persegi berdimensi* m *dan* $\mathbf{x}$ *merupakan vektor dengan* m *elemen.* $\mathbf{A}$ *dikatakan semi-definit positif jika dan hanya jika untuk semua vektor* $\mathbf{x}$ *berlaku*

$$\mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0. \tag{2.1.34}$$

Matriks $\mathbf{A}$ *diaktakan definit positif jika dan hanya jika untuk* $\mathbf{x}$ *tak kosong berlaku*

$$\mathbf{x}^\top \mathbf{A}\mathbf{x} > 0. \tag{2.1.35}$$

Contoh 2.1.32 Pertimbangkan matriks

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Untuk sembarang vektor $\mathbf{x} = (x_1, x_2)^\top$, berlaku

$$\begin{aligned}\mathbf{x}^\top \mathbf{A}\mathbf{x} &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= 2x_1^2 + 2x_1x_2 + 2x_2^2 \\ &= x_1^2 + (x_1 + x_2)^2 + x_2^2 \geq 0. \end{aligned} \tag{2.1.36}$$

Lebih lanjut, jika $\mathbf{x} \neq \mathbf{0}$, maka $\mathbf{x}^\top \mathbf{A}\mathbf{x} > 0$. Oleh karena itu, $\mathbf{A}$ adalah matriks definit positif. Sebagai contoh lain, matriks identitas $\mathbf{I}_n$ selalu definit positif karena

$\mathbf{x}^\top \mathbf{I}_n \mathbf{x} = \|\mathbf{x}\|^2 > 0$ untuk setiap $\mathbf{x} \neq \mathbf{0}$.

Definisi 2.1.33 (**Matriks Definit Negatif**) *Misalkan matriks* $\mathbf{A}$ *merupakan matriks persegi berukuran* m*, maka* $\mathbf{A}$ *dikatakan definit negatif jika dan hanya jika* $-\mathbf{A}$ *definit positif. Hal yang sama juga berlaku untuk semi-definit negatif.*

Contoh 2.1.34 Pertimbangkan matriks

$$\mathbf{B} = \begin{pmatrix} -3 & 1 \\ 1 & -2 \end{pmatrix}.$$

Matriks $-\mathbf{B}$ adalah

$$-\mathbf{B} = \begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}.$$

Untuk memeriksa apakah $-\mathbf{B}$ definit positif, dapat dihitung nilai eigennya atau diperiksa minor utamanya. Secara umum, suatu matriks simetris $\mathbf{A}$ adalah definit positif jika dan hanya jika semua nilai eigennya positif, karena untuk setiap $\mathbf{x} \neq \mathbf{0}$ berlaku $\mathbf{x}^\top \mathbf{A}\mathbf{x} = \sum_i \lambda_i (\mathbf{u}_i^\top \mathbf{x})^2 > 0$ dengan λ_i nilai eigen dan $\mathbf{u}_i$ vektor eigen ortonormal. Alternatifnya, berdasarkan kriteria Sylvester, matriks simetris definit positif jika dan hanya jika semua minor utama utamanya bernilai positif.

Kriteria Sylvester menyatakan bahwa untuk matriks simetris $\mathbf{A}$ berukuran $n \times n$, matriks tersebut definit positif jika dan hanya jika semua minor utama utama (*leading principal minors*) bernilai positif. Minor utama utama ke-k adalah determinan dari submatriks pojok kiri atas berukuran $k \times k$, yaitu

$$\Delta_k = \det \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix}, \quad k = 1, 2, \ldots, n. \tag{2.1.37}$$

Dengan demikian, syarat definit positif adalah $\Delta_1 > 0$, $\Delta_2 > 0$, …, $\Delta_n > 0$.

Minor utama pertama matriks $-\mathbf{B}$ adalah $\Delta_1 = 3 > 0$ dan minor utama kedua adalah $\Delta_2 = \det(-\mathbf{B}) = 3 \cdot 2 - (-1)(-1) = 5 > 0$. Sebab semua minor utama positif, maka $-\mathbf{B}$ definit positif. Oleh karena itu, $\mathbf{B}$ adalah matriks definit negatif.

Beberapa literatur menggunakan notasi $\succeq$ untuk menunjukkan matriks semidefinit positif, seperti $\mathbf{A} \succeq 0$ berarti $\mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0$, $\forall \mathbf{x}$.

2.1.3 Invers Matriks

Sesuai kerangka **?**, konsep invers dipahami melalui operator linear dan representasinya dalam suatu basis. Sebuah matriks bujur sangkar merepresentasikan operator linear $T : K^n \to K^n$, dan invers matriks berkorespondensi dengan invers operator linear T^{-1} ketika T bijektif.

Definisi 2.1.35 (**Invers Matriks, (?)**) *Misalkan* $\mathbf{A} \in K^{n\times n}$. *Matriks* $\mathbf{A}$ *disebut* invertibel *(atau* non-singular*) apabila terdapat matriks* $\mathbf{B} \in K^{n\times n}$ *sedemikian sehingga*

$$\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A} = \mathbf{I}_n, \tag{2.1.38}$$

dengan $\mathbf{I}_n$ *matriks identitas berukuran* $n \times n$. *Matriks* $\mathbf{B}$ *disebut invers dari* $\mathbf{A}$ *dan dinotasikan* $\mathbf{A}^{-1}$.

Contoh 2.1.36 Pertimbangkan matriks

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix}.$$

Untuk menunjukkan bahwa $\mathbf{A}$ invertibel, perlu dicari matriks $\mathbf{B}$ sehingga $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A} = \mathbf{I}_2$. Dengan menggunakan rumus invers matriks 2×2, diperoleh

$$\mathbf{B} = \mathbf{A}^{-1} = \frac{1}{2 \cdot 3 - 1 \cdot 5} \begin{pmatrix} 3 & -1 \\ -5 & 2 \end{pmatrix} = \begin{pmatrix} 3 & -1 \\ -5 & 2 \end{pmatrix}.$$

Dapat diverifikasi bahwa

$$\mathbf{AB} = \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ -5 & 2 \end{pmatrix} = \begin{pmatrix} 6-5 & -2+2 \\ 15-15 & -5+6 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}_2.$$

Dengan demikian, $\mathbf{A}$ adalah matriks invertibel dengan invers $\mathbf{A}^{-1} = \begin{pmatrix} 3 & -1 \\ -5 & 2 \end{pmatrix}$.

Teorema 2.1.37 (Keunikan Invers) *Jika $\mathbf{A}$ invertibel, maka $\mathbf{A}^{-1}$ tunggal.*

Bukti. Jika $\mathbf{B}$ dan $\mathbf{C}$ keduanya memenuhi $\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n$ dan $\mathbf{AC} = \mathbf{CA} = \mathbf{I}_n$, maka

$$\mathbf{B} = \mathbf{BI}_n = \mathbf{B}(\mathbf{AC}) = (\mathbf{BA})\mathbf{C} = \mathbf{I}_n\mathbf{C} = \mathbf{C}. \tag{2.1.39}$$

■

Proposisi 2.1.38 (Sifat-Sifat Invers, (?)) *Misalkan $\mathbf{A}, \mathbf{B} \in K^{n\times n}$ invertibel dan $\alpha \in K$. Berlaku:*

(a) $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$*;*

(b) $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$*; pada $K = \mathbb{C}$ juga berlaku* $(\mathbf{A}^*)^{-1} = (\mathbf{A}^{-1})^*$*;*

(c) $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$*;*

(d) Jika $\mathbf{A} = \operatorname{diag}(\mathbf{B}_1, \ldots, \mathbf{B}_k)$ dengan setiap $\mathbf{B}_i$ invertibel (ukuran mungkin berbeda), maka

$$\mathbf{A}^{-1} = \operatorname{diag}(\mathbf{B}_1^{-1}, \ldots, \mathbf{B}_k^{-1}).$$

Salah satu cara atau algoritma perhitungan invers matriks adalah dengan invers Gauss-Jordan. Jika $\mathbf{A}$ invertibel, maka terdapat barisan operasi baris elementer (matriks elementer $\mathbf{E}_k \cdots \mathbf{E}_1$) sehingga $\mathbf{E}_k \cdots \mathbf{E}_1\, \mathbf{A} = \mathbf{I}$. Dengan demikian,

$$\mathbf{A}^{-1} = \mathbf{E}_k \cdots \mathbf{E}_1. \tag{2.1.40}$$

Secara praktis, dapat dilakukan lakukan reduksi baris teraugmentasi:

$$\left[\,\mathbf{A}\;\middle|\;\mathbf{I}\,\right] \xrightarrow{\text{operasi baris}} \left[\,\mathbf{I}\;\middle|\;\mathbf{A}^{-1}\,\right].$$

Contoh 2.1.39 Ambil

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

Dengan melakukan eliminasi Gauss–Jordan pada $[\,\mathbf{A} \mid \mathbf{I}\,]$ diperoleh

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 0 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array}\right] \xrightarrow{R_1 \leftarrow R_1 - \frac{1}{2}R_2} \left[\begin{array}{ccc|ccc} \frac{3}{2} & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array}\right]$$

$$\xrightarrow{R_2 \leftarrow R_2 - R_3} \left[\begin{array}{ccc|ccc} \frac{3}{2} & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 1 & 1 & -1 & 0 & 1 & -1 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array}\right] \xrightarrow{R_3 \leftarrow R_3 - R_2} \left[\begin{array}{ccc|ccc} \frac{3}{2} & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 1 & 1 & -1 & 0 & 1 & -1 \\ -1 & 0 & 3 & 0 & -1 & 2 \end{array}\right]$$

$$\xrightarrow{R_1 \leftarrow R_1 + \frac{1}{2}R_3} \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 0 & 1 & -1 \\ -1 & 0 & 3 & 0 & -1 & 2 \end{array}\right] \xrightarrow{R_2 \leftarrow R_2 - R_1} \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & -1 & 1 \\ 0 & 1 & -2 & -1 & 2 & -2 \\ -1 & 0 & 3 & 0 & -1 & 2 \end{array}\right]$$

$$\xrightarrow{R_3 \leftarrow R_3 + R_1} \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & -1 & 1 \\ 0 & 1 & -2 & -1 & 2 & -2 \\ 0 & 0 & 4 & 1 & -2 & 3 \end{array}\right] \xrightarrow{R_3 \leftarrow \frac{1}{4}R_3} \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & -1 & 1 \\ 0 & 1 & -2 & -1 & 2 & -2 \\ 0 & 0 & 1 & \frac{1}{4} & -\frac{1}{2} & \frac{3}{4} \end{array}\right]$$

$$\xrightarrow{R_1 \leftarrow R_1 - R_3,\ R_2 \leftarrow R_2 + 2R_3} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{3}{4} & -\frac{1}{2} & \frac{1}{4} \\ 0 & 1 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{4} & -\frac{1}{2} & \frac{3}{4} \end{array}\right].$$

Maka

$$\mathbf{A}^{-1} = \begin{pmatrix} \frac{3}{4} & -\frac{1}{2} & \frac{1}{4} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ \frac{1}{4} & -\frac{1}{2} & \frac{3}{4} \end{pmatrix}.$$

Definisi 2.1.40 (Kofaktor dan Matriks Adjoin, (?)) *Misalkan* $\mathbf{A} = (a_{ij}) \in K^{n \times n}$. Minor *dari elemen* a_{ij}*, dinotasikan* M_{ij}*, adalah determinan submatriks berukuran* $(n-1) \times (n-1)$ *yang diperoleh dengan menghapus baris ke-*i *dan kolom ke-*j *dari* $\mathbf{A}$. Kofaktor *dari* a_{ij} *didefinisikan sebagai*

$$C_{ij} = (-1)^{i+j} M_{ij}. \tag{2.1.41}$$

Matriks kofaktor adalah $\mathbf{C} = (C_{ij})$. Matriks adjoin *(atau* adjugate*) dari* $\mathbf{A}$*, dinotasikan* $\operatorname{adj}(\mathbf{A})$*, adalah transpose dari matriks kofaktor:*

$$\operatorname{adj}(\mathbf{A}) = \mathbf{C}^\top. \tag{2.1.42}$$

Teorema 2.1.41 (Invers Matriks dengan Adjoin) *Jika* $\mathbf{A} \in K^{n \times n}$ *dengan* $\det(\mathbf{A}) \neq 0$*, maka*

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \operatorname{adj}(\mathbf{A}). \tag{2.1.43}$$

Bukti. Akan ditunjukkan bahwa $\mathbf{A} \cdot \operatorname{adj}(\mathbf{A}) = \det(\mathbf{A}) \cdot \mathbf{I}_n$. Berdasarkan definisi, elemen (i, j) dari matriks $\mathbf{A} \cdot \operatorname{adj}(\mathbf{A})$ adalah

$$(\mathbf{A} \cdot \operatorname{adj}(\mathbf{A}))_{ij} = \sum_{k=1}^{n} a_{ik} \cdot (\operatorname{adj}(\mathbf{A}))_{kj} = \sum_{k=1}^{n} a_{ik} C_{jk}, \tag{2.1.44}$$

dengan C_{jk} adalah kofaktor dari elemen a_{jk} pada matriks $\mathbf{A}$.

Kasus 1: Jika $i = j$, maka

$$\sum_{k=1}^{n} a_{ik} C_{ik} \tag{2.1.45}$$

merupakan ekspansi kofaktor dari $\det(\mathbf{A})$ sepanjang baris ke-i. Berdasarkan rumus ekspansi Laplace, berlaku

$$\det(\mathbf{A}) = \sum_{k=1}^{n} a_{ik} C_{ik}, \tag{2.1.46}$$

sehingga $(\mathbf{A} \cdot \operatorname{adj}(\mathbf{A}))_{ii} = \det(\mathbf{A})$.

Kasus 2: Jika $i \neq j$, maka

$$\sum_{k=1}^{n} a_{ik} C_{jk} \tag{2.1.47}$$

dapat diinterpretasikan sebagai ekspansi kofaktor dari suatu matriks $\mathbf{A}'$ yang diperoleh dengan mengganti baris ke-j dari $\mathbf{A}$ dengan baris ke-i dari $\mathbf{A}$. Sebab matriks $\mathbf{A}'$ memiliki dua baris yang identik (baris ke-i dan baris ke-j), maka $\det(\mathbf{A}') = 0$. Oleh karena itu, $(\mathbf{A} \cdot \operatorname{adj}(\mathbf{A}))_{ij} = 0$ untuk $i \neq j$.

Dari kedua kasus di atas, diperoleh

$$\mathbf{A} \cdot \operatorname{adj}(\mathbf{A}) = \det(\mathbf{A}) \cdot \mathbf{I}_n. \tag{2.1.48}$$

Sebab $\det(\mathbf{A}) \neq 0$, dapat dibagi kedua ruas dengan $\det(\mathbf{A})$ sehingga

$$\mathbf{A} \cdot \frac{1}{\det(\mathbf{A})} \operatorname{adj}(\mathbf{A}) = \mathbf{I}_n. \tag{2.1.49}$$

Dengan demikian, $\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \operatorname{adj}(\mathbf{A})$. ■

Contoh 2.1.42 Untuk matriks $\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 5 & 6 & 0 \end{pmatrix}$, akan dihitung inversnya menggunakan rumus adjoin. Pertama, dihitung kofaktor-kofaktor dari setiap elemen matriks:

$$C_{11} = (-1)^{1+1} \det \begin{pmatrix} 1 & 4 \\ 6 & 0 \end{pmatrix} = 1 \cdot (1 \cdot 0 - 4 \cdot 6) = -24,$$

$$C_{12} = (-1)^{1+2} \det \begin{pmatrix} 0 & 4 \\ 5 & 0 \end{pmatrix} = -1 \cdot (0 \cdot 0 - 4 \cdot 5) = 20,$$

$$C_{13} = (-1)^{1+3} \det \begin{pmatrix} 0 & 1 \\ 5 & 6 \end{pmatrix} = 1 \cdot (0 \cdot 6 - 1 \cdot 5) = -5,$$

$$C_{21} = (-1)^{2+1} \det \begin{pmatrix} 2 & 3 \\ 6 & 0 \end{pmatrix} = -1 \cdot (2 \cdot 0 - 3 \cdot 6) = 18,$$

$$C_{22} = (-1)^{2+2} \det \begin{pmatrix} 1 & 3 \\ 5 & 0 \end{pmatrix} = 1 \cdot (1 \cdot 0 - 3 \cdot 5) = -15,$$

$$C_{23} = (-1)^{2+3} \det \begin{pmatrix} 1 & 2 \\ 5 & 6 \end{pmatrix} = -1 \cdot (1 \cdot 6 - 2 \cdot 5) = 4,$$

$$C_{31} = (-1)^{3+1} \det \begin{pmatrix} 2 & 3 \\ 1 & 4 \end{pmatrix} = 1 \cdot (2 \cdot 4 - 3 \cdot 1) = 5,$$

$$C_{32} = (-1)^{3+2} \det \begin{pmatrix} 1 & 3 \\ 0 & 4 \end{pmatrix} = -1 \cdot (1 \cdot 4 - 3 \cdot 0) = -4,$$

$$C_{33} = (-1)^{3+3} \det \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} = 1 \cdot (1 \cdot 1 - 2 \cdot 0) = 1.$$

Matriks kofaktor adalah

$$\mathbf{C} = \begin{pmatrix} -24 & 20 & -5 \\ 18 & -15 & 4 \\ 5 & -4 & 1 \end{pmatrix},$$

sehingga matriks adjoin (transpose dari matriks kofaktor) adalah

$$\mathrm{adj}(\mathbf{A}) = \mathbf{C}^\top = \begin{pmatrix} -24 & 18 & 5 \\ 20 & -15 & -4 \\ -5 & 4 & 1 \end{pmatrix}.$$

Selanjutnya, determinan $\mathbf{A}$ dihitung menggunakan ekspansi kofaktor sepanjang baris pertama:

$$\det(\mathbf{A}) = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} = 1\cdot(-24) + 2\cdot 20 + 3\cdot(-5) = -24 + 40 - 15 = 1.$$

Sebab $\det(\mathbf{A}) = 1 \neq 0$, maka $\mathbf{A}$ invertibel dan

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})}\mathrm{adj}(\mathbf{A}) = \frac{1}{1}\begin{pmatrix} -24 & 18 & 5 \\ 20 & -15 & -4 \\ -5 & 4 & 1 \end{pmatrix} = \begin{pmatrix} -24 & 18 & 5 \\ 20 & -15 & -4 \\ -5 & 4 & 1 \end{pmatrix}.$$

Untuk memverifikasi hasil, dapat diperiksa bahwa $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_3$:

$$\begin{aligned}
\mathbf{A}\mathbf{A}^{-1} &= \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 5 & 6 & 0 \end{pmatrix}\begin{pmatrix} -24 & 18 & 5 \\ 20 & -15 & -4 \\ -5 & 4 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1(-24) + 2(20) + 3(-5) & 1(18) + 2(-15) + 3(4) & 1(5) + 2(-4) + 3(1) \\ 0(-24) + 1(20) + 4(-5) & 0(18) + 1(-15) + 4(4) & 0(5) + 1(-4) + 4(1) \\ 5(-24) + 6(20) + 0(-5) & 5(18) + 6(-15) + 0(4) & 5(5) + 6(-4) + 0(1) \end{pmatrix} \\
&= \begin{pmatrix} -24 + 40 - 15 & 18 - 30 + 12 & 5 - 8 + 3 \\ 0 + 20 - 20 & 0 - 15 + 16 & 0 - 4 + 4 \\ -120 + 120 + 0 & 90 - 90 + 0 & 25 - 24 + 0 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathbf{I}_3.
\end{aligned}$$

Dengan demikian, hasil perhitungan invers terverifikasi benar.

Contoh 2.1.43 (Invers 2×2 dengan Rumus Klasik) Untuk $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ dengan $ad - bc \neq 0$, matriks adjoin adalah

$$\mathrm{adj}(\mathbf{A}) = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

sehingga

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{2.1.50}$$

Rumus ini konsisten dengan Gauss–Jordan dan merupakan kasus khusus dari rumus adjoin ketika $n = 2$.

2.1.4 Determinan dan *Rank*

Pada seluruh bagian ini, V menyatakan ruang vektor berdimensi hingga atas suatu lapangan K dan $\mathcal{L}(V, W)$ himpunan semua pemetaan linear $V \to W$.

Definisi 2.1.44 (Pemetaan Linear, (?)) *Misalkan V, W ruang vektor atas suatu lapangan K. Suatu pemetaan $T : V \to W$ disebut* pemetaan linear *jika untuk semua $\mathbf{u}, \mathbf{v} \in V$ dan $\lambda \in K$ berlaku*

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}), \qquad T(\lambda\mathbf{v}) = \lambda T(\mathbf{v}). \tag{2.1.51}$$

Notasi $\mathcal{L}(V, W)$ menyatakan himpunan semua pemetaan linear dari V ke W. Jika $V = W$, maka $\mathcal{L}(V) = \mathcal{L}(V, V)$ adalah himpunan semua operator linear pada V.

Contoh 2.1.45 Berikut ini adalah beberapa contoh pemetaan linear.

(a) Jika $V = W = \mathbb{R}^2$, maka pemetaan

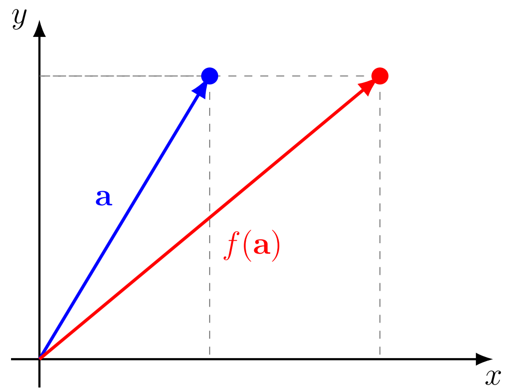
$$T(\mathbf{x}, \mathbf{y}) = (\mathbf{x} + 2\mathbf{y},\, 3\mathbf{x} - \mathbf{y}) \tag{2.1.52}$$

adalah elemen $\mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$.

(b) Jika $V = \mathbb{R}^3$ dan $W = \mathbb{R}^2$, maka setiap $T \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$ dapat direpresentasikan oleh matriks 2×3, misalnya

$$T(x_1, x_2, x_3) = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \tag{2.1.53}$$

Transformasi linear dapat divisualisasikan sebagai pemetaan yang membawa suatu vektor ke vektor lain dengan cara yang teratur (tanpa membengkokkan ruang). Sebagai contoh, Gambar 2.1 berikut memperlihatkan fungsi linear $f : \mathbb{R}^2 \to \mathbb{R}^2$ dengan $f(x, y) = (2x, y)$ yang meregangkan komponen x dari setiap vektor dengan faktor 2. Vektor $\mathbf{a}$ (biru) dipetakan ke $f(\mathbf{a})$ (merah).



Gambar 2.1 Ilustrasi transformasi linear $f : \mathbb{R}^2 \to \mathbb{R}^2$ dengan $f(x, y) = (2x, y)$.
(Sumber: Dokumen penulis)

Definisi 2.1.46 (***Null Space*, *Range*, dan *Rank*, (?)**) *Untuk $T \in \mathcal{L}(V, W)$,* null space *dan* range *didefinisikan sebagai*

$$\operatorname{null} T := \{\mathbf{v} \in V : T\mathbf{v} = 0\}, \qquad \operatorname{range} T := \{T\mathbf{v} : \mathbf{v} \in V\} \subseteq W.$$

Jika V berdimensi-hingga, rank T *didefinisikan oleh* $\operatorname{rank} T := \dim(\operatorname{range} T)$. *Dalam notasi matriks, untuk $\mathbf{A} \in K^{m \times n}$ yang merepresentasikan T relatif terhadap basis yang dipilih,* $\operatorname{rank}(\mathbf{A}) := \operatorname{rank}(T) = \dim(\operatorname{range} T)$.

Contoh 2.1.47 Misalkan

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \end{pmatrix} \in \mathbb{R}^{2\times 3}, \quad T : \mathbb{R}^3 \to \mathbb{R}^2, \text{ dan } \quad T(x) = \mathbf{A}\mathbf{x}.$$

(a) Sistem homogen $\mathbf{A}\mathbf{x} = 0$ memberi

$$x_2 = -4t, \quad x_1 = 5t, \quad x_3 = t, \qquad (t \in \mathbb{R}),$$

sehingga

$$\operatorname{null} T = \operatorname{span}\{(5, -4, 1)^\top\} \quad \text{dan} \quad \dim(\operatorname{null} T) = 1. \tag{2.1.54}$$

(b) Kolom pertama dan kedua bebas linear, maka

$$\operatorname{range} T = \operatorname{span}\{(1, 0)^\top, (2, 1)^\top\} \tag{2.1.55}$$

sehingga $\dim(\operatorname{range} T) = 2$.

Contoh sebelumnya menyebutkan tentang Teorema Rank-Nulitas yang merupakan teorema fundamental dalam pemetaan linear. Lebih lanjut, secara formal, teorema tersebut dirumuskan sebagai berikut.

Teorema 2.1.48 (Teorema Fundamental Pemetaan Linear atau Rank–Nulitas, (?)) *Jika V berdimensi-hingga dan $T \in \mathcal{L}(V, W)$, maka*

$$\dim V \;=\; \dim(\operatorname{null} T) \;+\; \dim(\operatorname{range} T). \tag{2.1.56}$$

Contoh 2.1.49 Berdasarkan Contoh 2.1.47, diketahui bahwa $\dim(\operatorname{null} T) = 1$ dan $\dim(\operatorname{range} T) = 2$. Konsisten dengan Rank–Nulitas, $1 + 2 = 3 = \dim \mathbb{R}^3$.

Definisi 2.1.50 (Bentuk Multilinear Anti-Simetri, (?)) *Misalkan V ruang vektor atas lapangan K. Suatu fungsi $\omega : V^n \to K$ disebut* bentuk n-linear *jika linear pada setiap argumennya. Bentuk n-linear ω disebut* anti-simetri *atau* alternating *jika*

$$\omega(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n) = -\,\omega(\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_n) \tag{2.1.57}$$

untuk setiap pertukaran dua argumen $\mathbf{v}_i, \mathbf{v}_j \in V$*. Konsekuensinya, jika ada dua argumen sama, maka* $\omega(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_n) = 0$.

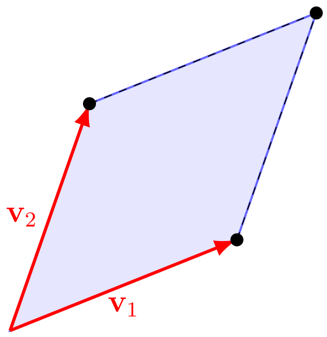
Contoh 2.1.51 Pada $V = \mathbb{R}^2$, fungsi

$$\omega\big((x_1, y_1), (x_2, y_2)\big) = x_1 y_2 - x_2 y_1 \tag{2.1.58}$$

adalah bentuk bilinear anti-simetri, karena

$$\omega\big((x_2, y_2), (x_1, y_1)\big) = x_2 y_1 - x_1 y_2 = -\omega\big((x_1, y_1), (x_2, y_2)\big). \tag{2.1.59}$$

Nilainya sama dengan luas terorientasi jajaran genjang yang dibentuk oleh kedua vektor.



Gambar 2.2 Ilustrasi bentuk bilinear anti-simetri $\omega(v_1, v_2) = x_1 y_2 - x_2 y_1$ pada $\mathbb{R}^2$. (Sumber: Dokumen penulis)

Luas dari jajar genjang pada Gambar 2.2 adalah

$$|\mathbf{v}_1||\mathbf{v}_2| \cdot \sin(\theta),$$

dengan $\cos(\theta) = \frac{\mathbf{v}_1\mathbf{v}_2}{|\mathbf{v}_1||\mathbf{v}_2|}$. Perhatikan bahwa $\sin(\theta) = \sqrt{1-\cos^2(\theta)}$, maka

$$\begin{aligned}
|\mathbf{v}_1||\mathbf{v}_2|\sin(\theta) &= \sqrt{|\mathbf{v}_1|^2|\mathbf{v}_2|^2 - (x_1x_2 + y_1y_2)^2} \\
&= \sqrt{x_1^2x_2^2 + x_1^2y_2^2 + y_1^2x_2^2 + y_1^2y_2^2 - (x_1^2x_2^2 + 2x_1x_2y_1y_2 + y_1^2y_2^2)} \\
&= \sqrt{x_1^2y_2^2 + y_1^2x_2^2 - 2x_1x_2y_1y_2} \\
&= \sqrt{(x_1y_2 - x_2y_1)^2} \\
&= |x_1y_2 - x_2y_1|. \qquad (2.1.60)
\end{aligned}$$

Persamaan terakhir menunjukkan bahwa luas jajar genjang pada Gambar 2.2 adalah $|x_1y_2 - x_2y_1| = |\omega(v_1, v_2)|$.

Contoh 2.1.51 memberikan intuisi penting tentang determinan. Fungsi bilinear anti-simetri $\omega(v_1, v_2) = x_1y_2 - x_2y_1$ tidak lain adalah determinannya pada $\mathbb{R}^2$. Secara geometri, determinan mengukur luas terorientasi dari jajar genjang yang dibentang oleh dua vektor. Nilai absolut $|\det|$ memberikan ukuran luas (atau volume pada dimensi lebih tinggi), sedangkan tanda determinan membedakan orientasi. Jika $\mathbf{v}_1$ ke $\mathbf{v}_2$ berputar berlawanan arah jarum jam, maka $\det(\cdot) > 0$ dan sebaliknya jika searah jarum jam, maka $\det(\cdot) < 0$.

Definisi 2.1.52 (Determinan, (?)) *Misalkan V berdimensi $n \geq 1$. Ruang semua bentuk n-linear beranti-simetri pada V memiliki dimensi 1. Untuk setiap $T \in \mathcal{L}(V)$,* determinan *T, ditulis $\det T$, adalah satu-satunya skalar di K dengan sifat*

$$\omega(T\mathbf{v}_1, \dots, T\mathbf{v}_n) = (\det T)\,\omega(\mathbf{v}_1, \dots, \mathbf{v}_n), \qquad (2.1.61)$$

untuk setiap bentuk n-linear beranti-simetri ω pada V. Jika $\mathbf{A}$ adalah matriks $n \times n$ yang merepresentasikan T pada suatu basis V, maka determinan matriks *didefinisikan oleh $\det(\mathbf{A}) := \det(T)$.*

Contoh 2.1.53 Ambil $V = \mathbb{R}^2$ dan operator linear $T : \mathbb{R}^2 \to \mathbb{R}^2$ yang direpresentasikan oleh matriks

$$\mathbf{A} = \begin{pmatrix} 3 & 5 \\ 7 & 2 \end{pmatrix}.$$

Untuk dua vektor $\mathbf{v}_1 = (2,1)^\top$ dan $\mathbf{v}_2 = (1,2)^\top$, bentuk bilinear anti-simetri standar di $\mathbb{R}^2$ adalah

$$\omega(\mathbf{v}_1, \mathbf{v}_2) = \det \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = 3. \tag{2.1.62}$$

Sekarang apabila dilihat bayangan kedua vektor tersebut,

$$T\mathbf{v}_1 = \mathbf{A}\mathbf{v}_1 = (11, 16)^\top, \qquad T\mathbf{v}_2 = \mathbf{A}\mathbf{v}_2 = (13, 11)^\top.$$

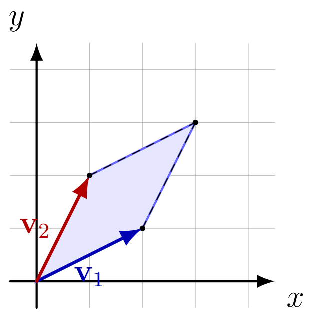
sehingga

$$\omega(T\mathbf{v}_1, T\mathbf{v}_2) = \det \begin{pmatrix} 11 & 13 \\ 16 & 11 \end{pmatrix} = 11 \cdot 11 - 13 \cdot 16 = -87. \tag{2.1.63}$$

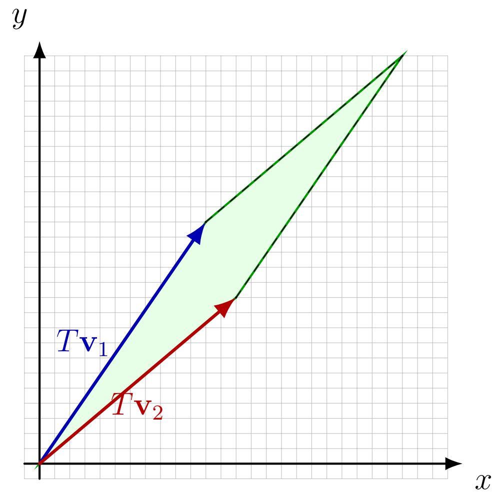
Dengan demikian, sesuai definisi determinan operator,

$$\omega(T\mathbf{v}_1, T\mathbf{v}_2) = (\det T)\, \omega(\mathbf{v}_1, \mathbf{v}_2) \qquad \Longrightarrow \qquad \det T = -87/3 = -29.$$

Sebab $\det(\mathbf{A}) := \det(T)$, diperoleh $\det(\mathbf{A}) = -29$, hasil ini konsisten dengan rumus determinan matriks 2×2 yang sudah dikenal, yaitu $\det \mathbf{A} = 6 - 35 = -29$.



(a) Luas $= |\det[\mathbf{v}_1, \mathbf{v}_2]| = 3$.



(b) Luas $= |\det[T\mathbf{v}_1, T\mathbf{v}_2]| = 87$.

Gambar 2.3 Perbandingan luas parallelogram sebelum dan sesudah transformasi A. (Sumber: Dokumen penulis)

Berdasarkan ilustrasi di atas, terlihat bahwa $\det(\mathbf{A})$ memiliki makna geometris sebagai skala luas jajar genjang hasil transformasi oleh $\mathbf{A}$.

Teorema 2.1.54 (**Sifat Dasar Determinan, (?)**) *Untuk* $S, T \in \mathcal{L}(V)$ *dan skalar* $c \in K$ *berlaku:*

(a) $\det \mathbf{I}_V = 1$*;*

(b) $\det(ST) = \det S \cdot \det T$*; dan*

(c) T *invertibel jika dan hanya jika* $\det T \neq 0$.

Jika $\mathbf{A}, \mathbf{B} \in K^{n \times n}$ *merepresentasikan* S, T*, maka* $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$ *dan* $\mathbf{A}$ *invertibel jika dan hanya jika* $\det(\mathbf{A}) \neq 0$.

Bukti. Berikut ini adalah bukti untuk ketiga sifat dasar determinan.

(a) Operator identitas $\mathbf{I}_V$ memetakan setiap vektor ke dirinya sendiri, sehingga untuk setiap bentuk n-linear beranti-simetri ω berlaku

$$\omega(\mathbf{I}_V\mathbf{v}_1, \ldots, \mathbf{I}_V\mathbf{v}_n) = \omega(\mathbf{v}_1, \ldots, \mathbf{v}_n). \tag{2.1.64}$$

Berdasarkan definisi determinan, $(\det \mathbf{I}_V)\,\omega(\mathbf{v}_1, \dots, \mathbf{v}_n) = \omega(\mathbf{v}_1, \dots, \mathbf{v}_n)$, sehingga $\det \mathbf{I}_V = 1$.

(b) Untuk $S, T \in \mathcal{L}(V)$ dan bentuk n-linear beranti-simetri ω, berlaku

$$\begin{aligned} \omega((ST)\mathbf{v}_1, \dots, (ST)\mathbf{v}_n) &= \omega(S(T\mathbf{v}_1), \dots, S(T\mathbf{v}_n)) \\ &= (\det S)\,\omega(T\mathbf{v}_1, \dots, T\mathbf{v}_n) \\ &= (\det S)(\det T)\,\omega(\mathbf{v}_1, \dots, \mathbf{v}_n). \end{aligned} \tag{2.1.65}$$

Di sisi lain, berdasarkan definisi determinan untuk ST:

$$\omega((ST)\mathbf{v}_1, \dots, (ST)\mathbf{v}_n) = (\det(ST))\,\omega(\mathbf{v}_1, \dots, \mathbf{v}_n). \tag{2.1.66}$$

Dengan membandingkan kedua persamaan, diperoleh $\det(ST) = \det S \cdot \det T$.

(c) $(\Rightarrow)$ Andaikan T invertibel. Maka terdapat T^{-1} sehingga $TT^{-1} = \mathbf{I}_V$. Dengan sifat (a) dan (b):

$$1 = \det \mathbf{I}_V = \det(TT^{-1}) = \det T \cdot \det T^{-1}. \tag{2.1.67}$$

Sebab $\det T \cdot \det T^{-1} = 1$, maka $\det T \neq 0$.

$(\Leftarrow)$ Andaikan $\det T \neq 0$. Misalkan T tidak invertibel, maka T tidak bijektif sehingga $\operatorname{null} T \neq \{\mathbf{0}\}$ atau $\operatorname{range} T \neq V$. Dalam kedua kasus, terdapat vektor-vektor $\mathbf{v}_1, \dots, \mathbf{v}_n$ yang bebas linear sedemikian sehingga $T\mathbf{v}_1, \dots, T\mathbf{v}_n$ bergantung linear. Sebab vektor-vektor hasil transformasi bergantung linear, maka untuk setiap bentuk n-linear beranti-simetri ω:

$$\omega(T\mathbf{v}_1, \dots, T\mathbf{v}_n) = 0. \tag{2.1.68}$$

Namun, sebab $\mathbf{v}_1, \dots, \mathbf{v}_n$ bebas linear, berlaku $\omega(\mathbf{v}_1, \dots, \mathbf{v}_n) \neq 0$. Dari definisi determinan:

$$0 = \omega(T\mathbf{v}_1, \ldots, T\mathbf{v}_n) = (\det T)\, \omega(\mathbf{v}_1, \ldots, \mathbf{v}_n), \tag{2.1.69}$$

sehingga $\det T = 0$, yang berkontradiksi dengan asumsi $\det T \neq 0$. Jadi, T haruslah invertibel.

■

2.1.5 Nilai Eigen dan Vektor Eigen

Dalam aljabar linear, salah satu konsep penting adalah *eigenvalues* dan *eigenvectors* yang memberikan wawasan tentang bagaimana suatu operator linear bertindak pada vektor tertentu tanpa mengubah arahnya.

Definisi 2.1.55 (**Vektor Eigen dan Nilai Eigen, (?, ?)**) *Misalkan $T : V \to V$ adalah operator linear pada ruang vektor V atas $\mathbb{C}$.*

(a) Suatu vektor $\mathbf{v} \in V$, $\mathbf{v} \neq \mathbf{0}$, disebut sebagai vektor eigen dari T jika terdapat skalar $\lambda \in \mathbb{C}$ sehingga

$$T(\mathbf{v}) = \lambda\mathbf{v}. \tag{2.1.70}$$

(b) Skalar λ yang memenuhi persamaan di atas disebut nilai eigen dari T.

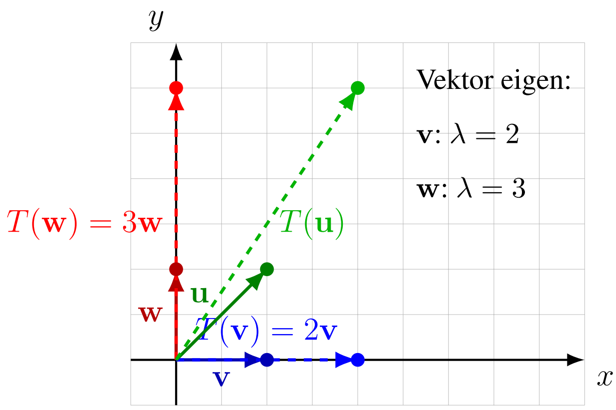
Secara intuitif, vektor eigen adalah vektor yang tidak berubah arah setelah dikenakan transformasi T, melainkan hanya mengalami skala sebesar λ.

Contoh 2.1.56 Pertimbangkan $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ dengan

$$T(x, y) = (2x, 3y). \tag{2.1.71}$$

(a) Untuk vektor $\mathbf{v} = (1, 0)$, berlaku $T(1, 0) = (2, 0) = 2(1, 0)$, sehingga $\mathbf{v}$ adalah vektor eigen dengan nilai $\lambda = 2$.

(b) Untuk vektor $\mathbf{w} = (0, 1)$, berlaku $T(0, 1) = (0, 3) = 3(0, 1)$, sehingga $\mathbf{w}$ adalah vektor eigen dengan nilai $\lambda = 3$.



Gambar 2.4 Ilustrasi vektor eigen dan nilai eigen untuk transformasi $T(x, y) = (2x, 3y)$. Vektor eigen v dan w tetap searah setelah transformasi (hanya mengalami penskalaan), sedangkan vektor u yang bukan vektor eigen berubah arah setelah transformasi. (Sumber: Dokumen penulis)

Gambar 2.4 mengilustrasikan bahwa vektor eigen $\mathbf{v}$ dan $\mathbf{w}$ tetap searah setelah transformasi T, hanya mengalami penskalaan sebesar nilai eigennya masing-masing. Sebaliknya, vektor $\mathbf{u} = (1, 1)$ yang bukan vektor eigen berubah menjadi $T(\mathbf{u}) = (2, 3)$ yang tidak searah dengan $\mathbf{u}$.

Teorema 2.1.57 (Keberadaan Nilai Eigen pada Ruang Vektor Kompleks, (?)) *Jika V adalah ruang vektor berdimensi-hingga atas $\mathbb{C}$ dan $T \colon V \to V$ operator linear, maka T memiliki setidaknya satu nilai eigen.*

Bukti. Misalkan $\dim V = n$. Pilih basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ dari V. Dalam basis ini, T direpresentasikan oleh suatu matriks $\mathbf{A} \in \mathrm{Mat}_n(\mathbb{C})$. Selanjutnya, definisikan polinomial karakteristik

$$p_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}_n). \tag{2.1.72}$$

Polinomial ini berderajat n dalam variabel λ, dengan koefisien dalam $\mathbb{C}$. Teorema dasar aljabar menyatakan bahwa setiap polinomial kompleks nonkonstan memiliki akar. Sebab $\deg p_{\mathbf{A}}(\lambda) = n \geq 1$, maka terdapat $\lambda_0 \in \mathbb{C}$ sehingga

$$p_{\mathbf{A}}(\lambda_0) = \det(\mathbf{A} - \lambda_0 \mathbf{I}_n) = 0. \tag{2.1.73}$$

Jika $\det(\mathbf{A} - \lambda_0 \mathbf{I}_n) = 0$, maka matriks $\mathbf{A} - \lambda_0 \mathbf{I}_n$ tidak invertibel. Hal ini berarti terdapat $\mathbf{v} \neq \mathbf{0}$ dalam $\mathbb{C}^n$ dengan

$$(\mathbf{A} - \lambda_0 \mathbf{I}_n)\mathbf{v} = \mathbf{0}. \tag{2.1.74}$$

Persamaan di atas ekuivalen dengan $\mathbf{A}\mathbf{v} = \lambda_0\mathbf{v}$. Jadi, $\mathbf{v}$ adalah vektor eigen tak-nol dari $\mathbf{A}$, dan λ_0 adalah nilainya. Sebab $\mathbf{A}$ merepresentasikan operator T, maka kesimpulan $\mathbf{A}\mathbf{v} = \lambda_0\mathbf{v}$ identik dengan

$$T(\mathbf{v}) = \lambda_0\mathbf{v}, \quad \mathbf{v} \neq \mathbf{0}. \tag{2.1.75}$$

Dengan demikian, T memiliki setidaknya satu nilai eigen di $\mathbb{C}$. ■

Contoh 2.1.58 Berikut akan dihitung nilai eigen dan vektor eigen dari matriks

$$\mathbf{A} = \begin{pmatrix} 4 & 1 \\ 2 & 3 \end{pmatrix}.$$

Nilai eigen λ diperoleh dari persamaan karakteristik $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$:

$$\begin{aligned} \det \begin{pmatrix} 4-\lambda & 1 \\ 2 & 3-\lambda \end{pmatrix} &= 0 \\ (4-\lambda)(3-\lambda) - (1)(2) &= 0 \\ 12 - 4\lambda - 3\lambda + \lambda^2 - 2 &= 0 \\ \lambda^2 - 7\lambda + 10 &= 0 \\ (\lambda - 5)(\lambda - 2) &= 0. \end{aligned} \tag{2.1.76}$$

Dengan demikian, diperoleh nilai eigen $\lambda_1 = 5$ dan $\lambda_2 = 2$.

Vektor eigen $\mathbf{v}_1$ memenuhi $(\mathbf{A} - 5\mathbf{I})\mathbf{v}_1 = \mathbf{0}$:

$$\begin{pmatrix} 4-5 & 1 \\ 2 & 3-5 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Dari baris pertama didapatkan $-v_1 + v_2 = 0$, sehingga $v_2 = v_1$. Dengan memilih $v_1 = 1$, diperoleh vektor eigen

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Vektor eigen $\mathbf{v}_2$ memenuhi $(\mathbf{A} - 2\mathbf{I})\mathbf{v}_2 = \mathbf{0}$:

$$\begin{pmatrix} 4-2 & 1 \\ 2 & 3-2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Dari baris pertama didapatkan $2v_1 + v_2 = 0$, sehingga $v_2 = -2v_1$. Dengan memilih $v_1 = 1$, diperoleh vektor eigen

$$\mathbf{v}_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}.$$

2.1.6 Perkalian Kronecker

Perkalian Kronecker merupakan operasi matriks yang menghasilkan matriks blok dengan ukuran yang lebih besar dari kedua matriks operand-nya. Berbeda dengan perkalian matriks biasa yang mensyaratkan kesesuaian dimensi, perkalian Kronecker dapat dilakukan pada dua matriks dengan ukuran sembarang. Operasi ini memiliki peran penting dalam berbagai aplikasi, termasuk pemrosesan sinyal, analisis sistem linear, dan representasi tensor. Dalam konteks pembelajaran mesin dan jaringan saraf graf, perkalian Kronecker sering digunakan untuk mengkonstruksi matriks bobot yang merepresentasikan interaksi antar fitur atau simpul secara efisien.

Definisi 2.1.59 (Perkalian Kronecker, (?)) *Untuk* $\mathbf{A} \in \mathrm{Mat}_{m\times n}$ *dan* $\mathbf{B} \in \mathrm{Mat}_{p\times q}$*, perkalian Kronecker dari matriks tersebut dituliskan sebagai* $\mathbf{A} \otimes \mathbf{B}$*, didefinisikan* *sebagai*

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix} \in \mathrm{Mat}_{mp \times nq}. \tag{2.1.77}$$

Contoh 2.1.60 Apabila diambil

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \in \mathbb{R}^{2\times 2}, \qquad \mathbf{B} = \begin{pmatrix} 0 & 5 \\ 6 & 7 \end{pmatrix} \in \mathbb{R}^{2\times 2}.$$

Perkalian Kronecker $\mathbf{A} \otimes \mathbf{B}$ adalah

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} 1 \cdot \mathbf{B} & 2 \cdot \mathbf{B} \\ 3 \cdot \mathbf{B} & 4 \cdot \mathbf{B} \end{pmatrix} = \begin{pmatrix} 0 & 5 & 0 & 10 \\ 6 & 7 & 12 & 14 \\ 0 & 15 & 0 & 20 \\ 18 & 21 & 24 & 28 \end{pmatrix}. \tag{2.1.78}$$

Hasilnya adalah matriks berukuran 4×4 karena $\mathbf{A}$ berukuran 2×2 dan $\mathbf{B}$ juga 2×2, sehingga ukurannya $2 \cdot 2 \times 2 \cdot 2 = 4 \times 4$.

2.1.7 Diferensial Vektor dan Matriks

Kalkulus matriks adalah perpanjangan dari kalkulus diferensial pada ruang Euclidean ke fungsi yang melibatkan vektor dan matriks sebagai argumen maupun hasil.

Definisi 2.1.61 (Konvensi Turunan pada Vektor dan Matriks, (?)) *Misalkan*

$$\mathbf{y} = \psi(\mathbf{x}), \tag{2.1.79}$$

dengan $\mathbf{y}$ *dan* $\mathbf{x}$ *secara berurutan merupakan vektor kolom berukuran* m *dan* n*. Notasi*

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \left[\frac{\partial y_i}{\partial x_j}\right], \quad i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, n \tag{2.1.80}$$

merupakan notasi dari matriks derivatif parsial orde pertama (atau matriks Jacobi-

an) dari transformasi $\mathbf{x}$ *ke* $\mathbf{y}$ *sedemikian sehingga baris ke-*i *berisi turunan parsial dari* y_i *terhadap elemen dari* x*, yaitu*

$$\frac{\partial y_i}{\partial x_1}, \frac{\partial y_i}{\partial x_2}, \ldots, \frac{\partial y_i}{\partial x_n}. \tag{2.1.81}$$

Proposisi 2.1.62 *Jika*

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{2.1.82}$$

dengan $\mathbf{A}$ *merupakan matriks berukuran* $m \times n$ *yang tidak bergantung pada* $\mathbf{x}$*, maka berlaku*

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{A}. \tag{2.1.83}$$

Proposisi 2.1.63 *Jika*

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{2.1.84}$$

dengan $\mathbf{y}$ *berukuran* $m \times 1$*,* $\mathbf{A}$ *berukuran* $m \times n$*,* $\mathbf{x}$ *berukuran* $n \times 1$*, serta* $\mathbf{A}$ *dan* $\mathbf{x}$ *bergantung pada vektor* $\boldsymbol{\alpha}$ *berukuran* r*, maka*

$$\frac{\partial \mathbf{y}}{\partial \boldsymbol{\alpha}} = \left(\mathbf{x}^\top \otimes \mathbf{I}_m\right) \frac{\partial \mathbf{A}}{\partial \boldsymbol{\alpha}} + \mathbf{A}\frac{\partial \mathbf{x}}{\partial \boldsymbol{\alpha}}. \tag{2.1.85}$$

Proposisi 2.1.64 *Jika*

$$\mathbf{y} = \mathbf{z}^\top \mathbf{A}\mathbf{x}, \tag{2.1.86}$$

dengan $\mathbf{z}$ *berukuran* $m \times 1$*,* $\mathbf{A}$ *berukuran* $m \times n$*,* $\mathbf{x}$ *berukuran* $n \times 1$*, serta* $\mathbf{A}$ *independen dari* $\mathbf{z}$ *dan* $\mathbf{x}$*, maka*

$$\frac{\partial \mathbf{y}}{\partial \mathbf{z}} = \mathbf{x}^\top \mathbf{A}^\top, \quad \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{z}^\top \mathbf{A}. \tag{2.1.87}$$

Proposisi 2.1.65 *Jika*

$$\mathbf{y} = \mathbf{x}^\top \mathbf{A}\mathbf{x}, \tag{2.1.88}$$

dengan $\mathbf{x}$ *berukuran* $n \times 1$ *serta* $\mathbf{A}$ *berukuran* $n \times n$ *dan independen dengan* $\mathbf{x}$*, maka*

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{x}^\top (\mathbf{A} + \mathbf{A}^\top). \tag{2.1.89}$$

Jika $\mathbf{A}$ *adalah matriks simetri, maka*

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = 2\mathbf{x}^\top \mathbf{A}. \tag{2.1.90}$$

Proposisi 2.1.66 *Misalkan matriks* $\mathbf{A}$ *nonsingular yang berukuran* $m \times m$ *dan bergantung dengan parameter skalar* α, *maka*

$$\frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{A}^{-1}. \tag{2.1.91}$$

Teorema 2.1.67 (Aturan Jejak, (?)) *Untuk matriks dengan ukuran yang sesuai berlaku:*

$$\frac{\partial}{\partial \mathbf{X}} \operatorname{tr}(\mathbf{AX}) = \mathbf{A}^\top, \qquad \frac{\partial}{\partial \mathbf{X}} \operatorname{tr}(\mathbf{X}^\top \mathbf{AX}) = (\mathbf{A} + \mathbf{A}^\top) X.$$

Bukti. Berikut ini adalah bukti untuk kedua aturan jejak tersebut.

(a) Untuk $\operatorname{tr}(\mathbf{AX})$, perhatikan bahwa

$$\operatorname{tr}(\mathbf{AX}) = \sum_{i=1}^{m} (\mathbf{AX})_{ii} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_{ji}. \tag{2.1.92}$$

Turunan parsial terhadap x_{kl} adalah

$$\frac{\partial}{\partial x_{kl}} \operatorname{tr}(\mathbf{AX}) = \frac{\partial}{\partial x_{kl}} \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_{ji} = a_{lk}, \tag{2.1.93}$$

karena satu-satunya suku yang mengandung x_{kl} adalah ketika $j = k$ dan $i = l$. Dengan demikian,

$$\left(\frac{\partial}{\partial \mathbf{X}} \operatorname{tr}(\mathbf{AX}) \right)_{kl} = a_{lk} = (\mathbf{A}^\top)_{kl}, \tag{2.1.94}$$

sehingga $\frac{\partial}{\partial \mathbf{X}} \operatorname{tr}(\mathbf{AX}) = \mathbf{A}^\top$.

(b) Untuk $\mathrm{tr}(\mathbf{X}^\top \mathbf{A}\mathbf{X})$, perhatikan bahwa

$$\mathrm{tr}(\mathbf{X}^\top \mathbf{A}\mathbf{X}) = \sum_{i,j,k} x_{ji} a_{jk} x_{ki}. \tag{2.1.95}$$

Turunan parsial terhadap x_{pq} adalah

$$\begin{aligned}
\frac{\partial}{\partial x_{pq}} \mathrm{tr}(\mathbf{X}^\top \mathbf{A}\mathbf{X}) &= \frac{\partial}{\partial x_{pq}} \sum_{i,j,k} x_{ji} a_{jk} x_{ki} \\
&= \sum_k a_{pk} x_{kq} + \sum_j x_{jq} a_{jp} \\
&= (\mathbf{A}\mathbf{X})_{pq} + (\mathbf{X}^\top \mathbf{A}^\top)^\top_{qp} \\
&= (\mathbf{A}\mathbf{X})_{pq} + (\mathbf{A}^\top \mathbf{X})_{pq}.
\end{aligned} \tag{2.1.96}$$

Oleh karena itu,

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{tr}(\mathbf{X}^\top \mathbf{A}\mathbf{X}) = \mathbf{A}\mathbf{X} + \mathbf{A}^\top \mathbf{X} = (\mathbf{A} + \mathbf{A}^\top)\mathbf{X}. \tag{2.1.97}$$

■

Contoh 2.1.68 Dalam banyak permasalahan estimasi, seringkali ingin dicari vektor parameter $\boldsymbol{\beta}$ yang meminimumkan suatu fungsi kuadratik. Sebagai ilustrasi, misalkan terdapat n buah observasi yang dikumpulkan dalam vektor $\mathbf{y} \in \mathbb{R}^n$ dan matriks $\mathbf{X} \in \mathbb{R}^{n\times p}$ yang berisi nilai-nilai pengukuran. Tujuannya adalah mencari vektor $\boldsymbol{\beta} \in \mathbb{R}^p$ yang meminimumkan jumlah kuadrat selisih antara $\mathbf{y}$ dan $\mathbf{X}\boldsymbol{\beta}$, yaitu

$$S(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \tag{2.1.98}$$

Fungsi $S(\boldsymbol{\beta})$ ini mengukur seberapa jauh vektor $\mathbf{X}\boldsymbol{\beta}$ dari vektor $\mathbf{y}$. Dengan mengekspansi perkalian dalam, diperoleh

$$\begin{aligned}S(\boldsymbol{\beta}) &= (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\\ &= \mathbf{y}^\top\mathbf{y}-\mathbf{y}^\top\mathbf{X}\boldsymbol{\beta}-(\mathbf{X}\boldsymbol{\beta})^\top\mathbf{y}+(\mathbf{X}\boldsymbol{\beta})^\top\mathbf{X}\boldsymbol{\beta}\\ &= \mathbf{y}^\top\mathbf{y}-\mathbf{y}^\top\mathbf{X}\boldsymbol{\beta}-\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{y}+\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}. \end{aligned} \tag{2.1.99}$$

Perhatikan bahwa $\mathbf{y}^\top\mathbf{X}\boldsymbol{\beta}$ adalah skalar, sehingga $\mathbf{y}^\top\mathbf{X}\boldsymbol{\beta} = (\mathbf{y}^\top\mathbf{X}\boldsymbol{\beta})^\top = \boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{y}$. Dengan demikian,

$$S(\boldsymbol{\beta}) = \mathbf{y}^\top\mathbf{y}-2\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{y}+\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}. \tag{2.1.100}$$

Untuk meminimumkan $S(\boldsymbol{\beta})$, dihitung turunan parsialnya terhadap $\boldsymbol{\beta}$ dan disamakan dengan nol. Menggunakan proposisi-proposisi diferensial matriks:

- Suku $\mathbf{y}^\top\mathbf{y}$ adalah konstanta (tidak bergantung pada $\boldsymbol{\beta}$), sehingga turunannya adalah $\mathbf{0}$.
- Suku $-2\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{y}$ berbentuk $\mathbf{c}^\top\boldsymbol{\beta}$ dengan $\mathbf{c} = -2\mathbf{X}^\top\mathbf{y}$. Turunan dari $\mathbf{c}^\top\boldsymbol{\beta}$ terhadap $\boldsymbol{\beta}$ adalah $\mathbf{c}^\top = -2\mathbf{y}^\top\mathbf{X}$.
- Suku $\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta}$ berbentuk $\boldsymbol{\beta}^\top\mathbf{A}\boldsymbol{\beta}$ dengan $\mathbf{A} = \mathbf{X}^\top\mathbf{X}$ yang simetris. Berdasarkan proposisi sebelumnya, turunannya adalah $2\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}$.

Dengan demikian, gradien dari $S(\boldsymbol{\beta})$ adalah

$$\frac{\partial S(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = -2\mathbf{y}^\top\mathbf{X}+2\boldsymbol{\beta}^\top\mathbf{X}^\top\mathbf{X}. \tag{2.1.101}$$

Untuk menemukan titik stasioner, disamakan gradien dengan nol:

$$\begin{aligned}-2\mathbf{y}^\top\mathbf{X}+2\widehat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathbf{X} &= \mathbf{0}^\top\\ \widehat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathbf{X} &= \mathbf{y}^\top\mathbf{X}. \end{aligned} \tag{2.1.102}$$

Dengan mengambil *transpose* kedua ruas, diperoleh

$$\mathbf{X}^\top \mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{y}. \tag{2.1.103}$$

Jika matriks $\mathbf{X}^\top \mathbf{X}$ bersifat *nonsingular*, maka solusinya adalah

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y}. \tag{2.1.104}$$

Untuk memastikan bahwa $\widehat{\boldsymbol{\beta}}$ adalah titik minimum (bukan maksimum atau titik pelana), diperiksa turunan kedua (Hessian) dari $S(\boldsymbol{\beta})$:

$$\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = 2\mathbf{X}^\top \mathbf{X}. \tag{2.1.105}$$

Matriks $\mathbf{X}^\top \mathbf{X}$ selalu bersifat semi-definit positif karena untuk sembarang vektor $\mathbf{v} \in \mathbb{R}^p$,

$$\mathbf{v}^\top \mathbf{X}^\top \mathbf{X}\mathbf{v} = (\mathbf{X}\mathbf{v})^\top (\mathbf{X}\mathbf{v}) = \|\mathbf{X}\mathbf{v}\|^2 \geq 0. \tag{2.1.106}$$

Jika $\mathbf{X}$ memiliki *rank* penuh (yaitu $\operatorname{rank}(\mathbf{X}) = p$), maka $\mathbf{X}^\top \mathbf{X}$ bersifat definit positif, sehingga $\widehat{\boldsymbol{\beta}}$ adalah peminimum global yang unik.

2.2 Teori Ukuran dan Probabilitas

Teori probabilitas modern dibangun di atas fondasi teori ukuran. Probabilitas dipandang sebagai suatu ukuran terstandarisasi pada kelas himpunan tertentu. Pendekatan ini memungkinkan formulasi yang utuh terhadap variabel acak umum, integrasi ekspektasi, serta analisis limit yang menjadi dasar inferensi statistik asimtotik (**??**).

Definisi 2.2.1 (***Field***) *Suatu keluarga $\mathcal{F}$ dari kejadian-kejadian dalam ruang sampel Ω disebut sebagai* field *apabila*

(a) $\Omega \in \mathcal{F}$;

(b) jika $A \in \mathcal{F}$, maka $A^c \in \mathcal{F}$; dan

(c) jika $A, B \in \mathcal{F}$, maka $A \cup B \in \mathcal{F}$.

Konsekuensi dari definisi di atas adalah bahwa himpunan kosong ($\varnothing$) dan himpunan dari seluruh ruang sampel (Ω) selalu merupakan elemen dari field $\mathcal{F}$. Selain itu, field juga tertutup terhadap operasi irisan, yaitu jika $A, B \in \mathcal{F}$, maka $A \cap B \in \mathcal{F}$.

Contoh 2.2.2 Diberikan ruang sampel atau semesta $\Omega = \{1, 2, 3\}$ dan himpunan $\mathcal{F}_1$, $\mathcal{F}_2$ berikut.

$$\begin{aligned}
\mathcal{F}_1 &= \{\varnothing, \Omega, \{1\}, \{2\}, \{2, 3\}\}, \\
\mathcal{F}_2 &= \{\varnothing, \Omega, \{2\}, \{1, 3\}\}.
\end{aligned}$$

Dapat diperiksa bahwa $\mathcal{F}_1$ bukan merupakan field karena meskipun $\{2\} \in \mathcal{F}_1$, tetapi komplemennya $\{1, 3\} \notin \mathcal{F}_1$. Sebaliknya, $\mathcal{F}_2$ adalah field karena memenuhi ketiga aksioma di atas.

2.2.1 Ruang Ukur dan Ukuran Probabilitas

Definisi 2.2.3 (**σ-Aljabar**, **?**) *Misalkan Ω adalah suatu himpunan tak kosong. Suatu koleksi $\mathcal{F} \subseteq 2^{\Omega}$ disebut sebagai σ-aljabar apabila memenuhi:*

(a) $\Omega \in \mathcal{F}$;

(b) jika $A \in \mathcal{F}$, maka $A^c \in \mathcal{F}$; dan

(c) jika $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$, maka $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Contoh 2.2.4 Misalkan $\Omega = \{1, 2, 3\}$ adalah ruang sampel dengan tiga elemen. Berikut adalah beberapa contoh σ-aljabar pada Ω.

(a) σ-aljabar trivial: $\mathcal{F}_1 = \{\varnothing, \Omega\} = \{\varnothing, \{1, 2, 3\}\}$. Koleksi ini merupakan σ-aljabar terkecil pada Ω.

(b) σ-aljabar power set: $\mathcal{F}_2 = 2^{\Omega} = \{\varnothing, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$. Koleksi ini merupakan σ-aljabar terbesar pada Ω yang memuat semua subhimpunan.

(c) σ-aljabar yang dibangkitkan oleh $\{1\}$: $\mathcal{F}_3 = \{\varnothing, \{1\}, \{2,3\}, \{1,2,3\}\}$. Dapat diperiksa bahwa $\mathcal{F}_3$ memenuhi ketiga aksioma σ-aljabar, yaitu $\Omega \in \mathcal{F}_3$, tertutup terhadap komplemen (misalnya $\{1\}^c = \{2,3\} \in \mathcal{F}_3$), dan tertutup terhadap gabungan terhitung.

Di sisi lain, koleksi $\mathcal{G} = \{\varnothing, \{1\}, \{2\}, \Omega\}$ bukan merupakan σ-aljabar karena $\{1\} \cup \{2\} = \{1,2\} \notin \mathcal{G}$, sehingga tidak tertutup terhadap gabungan.

Definisi 2.2.5 (**Himpunan Borel**) *Misalkan* $\Omega = \mathbb{R}$. Aljabar Borel *pada* $\mathbb{R}$*, dilambangkan dengan* $\mathcal{B}(\mathbb{R})$*, adalah* σ*-aljabar terkecil yang memuat semua interval terbuka di* $\mathbb{R}$*. Himpunan-himpunan dalam* $\mathcal{B}(\mathbb{R})$ *disebut sebagai* himpunan Borel.

Contoh 2.2.6 Beberapa himpunan yang termasuk dalam $\mathcal{B}(\mathbb{R})$ adalah:

- Interval terbuka: (a, b) untuk sembarang $a, b \in \mathbb{R}$ dengan $a < b$.
- Interval tertutup: $[a, b]$ untuk sembarang $a, b \in \mathbb{R}$ dengan $a < b$.
- Himpunan bilangan rasional: $\mathbb{Q} \subseteq \mathbb{R}$.
- Himpunan bilangan irasional: $\mathbb{R} \setminus \mathbb{Q}$.

Namun, terdapat himpunan-himpunan tertentu yang tidak termasuk dalam $\mathcal{B}(\mathbb{R})$, seperti himpunan Vitali yang dibangun menggunakan aksioma pilihan. Himpunan ini tidak dapat dibentuk melalui operasi hitung pada interval terbuka, sehingga tidak termasuk dalam aljabar Borel.

Definisi 2.2.7 (**Ukuran**) *Diberikan ruang sampel* Ω *dan* σ*-aljabar* $\mathcal{F}$ *relatif terhadap* Ω*. Suatu fungsi* $\mu : \mathcal{F} \to \mathbb{R}^+$ *disebut sebagai* ukuran *apabila:*

(a) $\mu(F) \geq 0$ *untuk semua* $F \in \mathcal{F}$*;*

(b) $\mu(\varnothing) = 0$*; dan*

(c) untuk setiap barisan himpunan saling lepas ($F_i \cap F_j = \varnothing$ *untuk* $i \neq j$*), maka* $\{F_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ *berlaku.*

Triple $(\Omega, \mathcal{F}, \mu)$ disebut sebagai *ruang ukur*.

Definisi 2.2.8 (**Ukuran Probabilitas**, (**?**)) *Suatu fungsi* $\Pr : \mathcal{F} \to [0, 1]$ *disebut sebagai ukuran probabilitas apabila:*

(a) $\Pr(\Omega) = 1$*;*

(b) $\Pr(A) \geq 0$ *untuk semua* $A \in \mathcal{F}$*; dan*

(c) untuk setiap barisan himpunan saling lepas $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ *berlaku*

$$\Pr\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \Pr(A_n). \tag{2.2.1}$$

Triple $(\Omega, \mathcal{F}, \Pr)$ disebut sebagai *ruang probabilitas*.

Contoh 2.2.9 Misalkan $\Omega = \{1, 2, 3, 4, 5, 6\}$ adalah ruang sampel pelemparan sebuah dadu seimbang, dan $\mathcal{F} = 2^{\Omega}$ adalah σ-aljabar yang memuat semua subhimpunan dari Ω. Didefinisikan fungsi $\Pr : \mathcal{F} \to [0, 1]$ dengan

$$\Pr(A) = \frac{|A|}{6} \tag{2.2.2}$$

untuk setiap $A \in \mathcal{F}$, dengan $|A|$ menyatakan banyaknya elemen dalam A.

Akan diperiksa bahwa $\Pr$ memenuhi aksioma ukuran probabilitas.

(a) $\Pr(\Omega) = \Pr(\{1, 2, 3, 4, 5, 6\}) = \frac{6}{6} = 1$.

(b) Untuk setiap $A \in \mathcal{F}$, berlaku $|A| \geq 0$ sehingga $\Pr(A) = \frac{|A|}{6} \geq 0$.

(c) Untuk himpunan saling lepas $A_1, A_2, \ldots \subseteq \mathcal{F}$, berlaku

$$\Pr\left(\bigcup_{n=1}^{\infty} A_n\right) = \frac{|\bigcup_{n=1}^{\infty} A_n|}{6} = \frac{\sum_{n=1}^{\infty} |A_n|}{6} = \sum_{n=1}^{\infty} \frac{|A_n|}{6} = \sum_{n=1}^{\infty} \Pr(A_n). \tag{2.2.3}$$

Dengan demikian, $\Pr$ adalah ukuran probabilitas pada $(\Omega, \mathcal{F})$. Sebagai contoh perhitungan, probabilitas memperoleh angka genap adalah

$$\Pr(\{2, 4, 6\}) = \frac{3}{6} = \frac{1}{2}. \tag{2.2.4}$$

2.2.2 Variabel Acak sebagai Fungsi Terukur

Definisi 2.2.10 (Variabel Acak, (?)) *Diberikan ruang probabilitas* $(\Omega, \mathcal{F}, \Pr)$ *dan ruang terukur* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. *Suatu fungsi* $X : \Omega \to \mathbb{R}$ *disebut sebagai variabel acak apabila* X *terukur, yaitu untuk setiap* $B \in \mathcal{B}(\mathbb{R})$ *berlaku*

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}. \tag{2.2.5}$$

Pendefinisian ini memungkinkan penggunaan teori integrasi Lebesgue untuk mendefinisikan ekspektasi dan momen variabel acak secara umum.

Teorema 2.2.11 (Variabel Acak Menginduksi Ruang Probabilitas) *Misalkan triple* $(\Omega, \mathcal{F}, \Pr)$ *adalah ruang probabilitas dan* $X : \Omega \to \mathbb{R}$ *adalah variabel acak. Maka* X *menginduksi ruang probabilitas* $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \Pr_X)$ *pada garis real, dengan* $\Pr_X$ *didefinisikan oleh*

$$\Pr_X(B) = \Pr(X^{-1}(B)) = \Pr(\{\omega \in \Omega : X(\omega) \in B\}), \quad \forall B \in \mathcal{B}(\mathbb{R}). \tag{2.2.6}$$

Ukuran $\Pr_X$ *disebut sebagai distribusi atau hukum dari variabel acak* X.

Bukti. Akan ditunjukkan bahwa $\Pr_X$ memenuhi ketiga aksioma ukuran probabilitas.

(a) **Normalisasi:** Perhatikan bahwa $X^{-1}(\mathbb{R}) = \{\omega \in \Omega : X(\omega) \in \mathbb{R}\} = \Omega$, karena X memetakan setiap $\omega \in \Omega$ ke suatu bilangan real. Oleh karena itu,

$$\Pr_X(\mathbb{R}) = \Pr(X^{-1}(\mathbb{R})) = \Pr(\Omega) = 1. \tag{2.2.7}$$

(b) **Non-negativitas:** Untuk setiap $B \in \mathcal{B}(\mathbb{R})$, karena X adalah variabel acak (fungsi terukur), maka $X^{-1}(B) \in \mathcal{F}$. Sebab $\Pr$ adalah ukuran probabilitas pada $(\Omega, \mathcal{F})$, berlaku

$$\Pr_X(B) = \Pr(X^{-1}(B)) \geq 0. \tag{2.2.8}$$

(c) σ**-aditivitas:** Misalkan $\{B_n\}_{n=1}^{\infty}$ adalah barisan himpunan saling lepas dalam $\mathcal{B}(\mathbb{R})$, yaitu $B_i \cap B_j = \varnothing$ untuk $i \neq j$. Akan ditunjukkan bahwa $\{X^{-1}(B_n)\}_{n=1}^{\infty}$ juga saling lepas dalam $\mathcal{F}$. Andaikan terdapat $\omega \in X^{-1}(B_i) \cap X^{-1}(B_j)$ untuk $i \neq j$. Maka $X(\omega) \in B_i$ dan $X(\omega) \in B_j$, sehingga $X(\omega) \in B_i \cap B_j = \varnothing$, yang merupakan kontradiksi. Jadi, $X^{-1}(B_i) \cap X^{-1}(B_j) = \varnothing$ untuk $i \neq j$.

Selanjutnya, perhatikan bahwa

$$X^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right) = \bigcup_{n=1}^{\infty} X^{-1}(B_n). \tag{2.2.9}$$

Sebab $\Pr$ adalah ukuran probabilitas yang memenuhi σ-aditivitas, diperoleh

$$\begin{aligned}
\Pr_X\left(\bigcup_{n=1}^{\infty} B_n\right) &= \Pr\left(X^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right)\right) \\
&= \Pr\left(\bigcup_{n=1}^{\infty} X^{-1}(B_n)\right) \\
&= \sum_{n=1}^{\infty} \Pr(X^{-1}(B_n)) \\
&= \sum_{n=1}^{\infty} \Pr_X(B_n). \qquad (2.2.10)
\end{aligned}$$

Dengan demikian, $\Pr_X$ memenuhi ketiga aksioma ukuran probabilitas, sehingga $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \Pr_X)$ adalah ruang probabilitas yang diinduksi oleh variabel acak X. ■

2.2.3 Distribusi dan Ekspektasi

Definisi 2.2.12 (**Distribusi Variabel Acak**, (**?**)) *Distribusi dari variabel acak* X *adalah ukuran probabilitas* $\Pr_X$ *pada* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *yang didefinisikan oleh*

$$\Pr_X(B) = \Pr(X \in B), \qquad B \in \mathcal{B}(\mathbb{R}). \tag{2.2.11}$$

Definisi 2.2.13 (**Fungsi Kepadatan Probabilitas**, (**?**)) *Misalkan* X *adalah variabel acak dengan distribusi* $\Pr_X$*. Jika terdapat fungsi terintegralkan* $f_X : \mathbb{R} \to [0, \infty)$ *sedemikian sehingga untuk setiap himpunan Borel* $B \in \mathcal{B}(\mathbb{R})$ *berlaku*

$$\Pr_X(B) = \Pr(X \in B) = \int_B f_X(x)\, dx, \tag{2.2.12}$$

maka f_X *disebut sebagai* fungsi kepadatan probabilitas *(fkp) atau* probability density function *(pdf) dari* X*. Fungsi kepadatan probabilitas memenuhi sifat:*

(a) $f_X(x) \geq 0$ *untuk semua* $x \in \mathbb{R}$*, dan*

(b) $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$.

Variabel acak yang memiliki fungsi kepadatan probabilitas disebut variabel acak kontinu. *Di sisi lain, untuk variabel acak diskrit, digunakan* fungsi massa probabilitas *(fmp) yang didefinisikan sebagai* $p_X(x) = \Pr(X = x)$.

Perlu dicatat bahwa distribusi dan kepadatan adalah konsep yang berbeda. Distribusi $\Pr_X$ adalah ukuran probabilitas yang memetakan himpunan ke probabilitas, sedangkan kepadatan f_X adalah fungsi yang menggambarkan "kepadatan" probabilitas pada setiap titik. Tidak semua distribusi memiliki fungsi kepadatan; sebagai contoh, distribusi diskrit seperti Bernoulli tidak memiliki kepadatan terhadap ukuran Lebesgue, melainkan memiliki fungsi massa probabilitas.

Contoh 2.2.14 Berikut adalah beberapa distribusi probabilitas yang sering digunakan dalam statistika dan pembelajaran mesin.

(a) **Distribusi Normal (Gaussian).** Variabel acak X berdistribusi normal dengan parameter μ (rataan) dan σ^2 (variansi), ditulis $X \sim \mathcal{N}(\mu, \sigma^2)$, jika memiliki fungsi kepadatan probabilitas

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}. \tag{2.2.13}$$

Distribusi normal standar diperoleh ketika $\mu = 0$ dan $\sigma^2 = 1$.

(b) **Distribusi Bernoulli.** Variabel acak X berdistribusi Bernoulli dengan parameter $p \in [0, 1]$, ditulis $X \sim$ Bernoulli(p), jika

$$\Pr(X = k) = p^k(1-p)^{1-k}, \quad k \in \{0, 1\}. \tag{2.2.14}$$

Distribusi ini memodelkan percobaan dengan dua kemungkinan hasil (sukses atau gagal).

(c) **Distribusi Poisson.** Variabel acak X berdistribusi Poisson dengan parameter $\lambda > 0$, ditulis $X \sim$ Poisson(λ), jika

$$\Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \in \{0, 1, 2, \ldots\}. \tag{2.2.15}$$

Distribusi ini memodelkan banyaknya kejadian langka dalam interval waktu atau ruang tertentu.

(d) **Distribusi Eksponensial.** Variabel acak X berdistribusi eksponensial dengan parameter $\lambda > 0$, ditulis $X \sim$ Exp(λ), jika memiliki fungsi kepadatan probabilitas

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0. \tag{2.2.16}$$

Distribusi ini sering digunakan untuk memodelkan waktu tunggu antara kejadian berurutan.

Definisi 2.2.15 (Ukuran Lebesgue, (?)) *Ukuran Lebesgue pada* $\mathbb{R}$ *adalah ukuran* $\lambda : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ *yang memenuhi:*

(a) untuk setiap interval $[a, b]$ *berlaku* $\lambda([a, b]) = b - a$*;*

(b) *λ bersifat σ-aditif, yaitu untuk barisan himpunan saling lepas $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{B}(\mathbb{R})$ berlaku*

$$\lambda\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \lambda(A_n); \tag{2.2.17}$$

(c) *λ bersifat invarian terhadap translasi, yaitu $\lambda(A + x) = \lambda(A)$ untuk setiap $A \in \mathcal{B}(\mathbb{R})$ dan $x \in \mathbb{R}$.*

Ukuran Lebesgue merupakan perluasan natural dari konsep panjang interval ke himpunan-himpunan Borel yang lebih umum.

Contoh 2.2.16 Beberapa contoh perhitungan ukuran Lebesgue adalah sebagai berikut.

(a) Untuk interval $[2, 5]$, berlaku $\lambda([2, 5]) = 5 - 2 = 3$.

(b) Untuk himpunan singleton $\{a\}$, berlaku $\lambda(\{a\}) = 0$ karena $\{a\} = \bigcap_{n=1}^{\infty}[a - 1/n, a + 1/n]$ dan $\lambda([a - 1/n, a + 1/n]) = 2/n \to 0$.

(c) Untuk himpunan bilangan rasional $\mathbb{Q}$, berlaku $\lambda(\mathbb{Q}) = 0$ karena $\mathbb{Q}$ terhitung dan merupakan gabungan terhitung dari singleton.

Definisi 2.2.17 (**Fungsi Sederhana**, (**?**)) *Suatu fungsi $s : \Omega \to \mathbb{R}$ disebut fungsi sederhana apabila s hanya mengambil sejumlah hingga nilai berbeda. Fungsi sederhana dapat dituliskan dalam bentuk*

$$s(\omega) = \sum_{k=1}^{m} c_k \mathbf{1}_{A_k}(\omega), \tag{2.2.18}$$

dengan $c_1, \ldots, c_m \in \mathbb{R}$ adalah nilai-nilai yang berbeda, $A_k = s^{-1}(\{c_k\})$ adalah himpunan terukur, dan $\mathbf{1}_{A_k}$ adalah fungsi indikator dari A_k.

Contoh 2.2.18 Misalkan $\Omega = [0, 3]$ dan didefinisikan fungsi $s : \Omega \to \mathbb{R}$ sebagai berikut:

$$s(\omega) = \begin{cases} 2, & \text{jika } \omega \in [0,1), \\ 5, & \text{jika } \omega \in [1,2), \\ -1, & \text{jika } \omega \in [2,3]. \end{cases}$$

Fungsi s adalah fungsi sederhana karena hanya mengambil tiga nilai berbeda, yaitu $c_1 = 2, c_2 = 5$, dan $c_3 = -1$. Himpunan-himpunan terukur yang bersesuaian adalah $A_1 = [0,1)$, $A_2 = [1,2)$, dan $A_3 = [2,3]$. Dengan demikian, s dapat dituliskan sebagai

$$s(\omega) = 2 \cdot \mathbf{1}_{[0,1)}(\omega) + 5 \cdot \mathbf{1}_{[1,2)}(\omega) + (-1) \cdot \mathbf{1}_{[2,3]}(\omega).$$

Sebagai contoh lain, fungsi tangga atau *step function* $s(\omega) = \lfloor \omega \rfloor$ pada interval $[0,n]$ dengan $n \in \mathbb{Z}^+$ adalah fungsi sederhana yang mengambil nilai-nilai $0, 1, \ldots, n$.

Definisi 2.2.19 (**Integral Lebesgue untuk Fungsi Sederhana**, (**?**)) *Untuk fungsi sederhana non-negatif $s = \sum_{k=1}^{m} c_k \mathbf{1}_{A_k}$ pada ruang ukur $(\Omega, \mathcal{F}, \mu)$, integral Lebesgue didefinisikan sebagai*

$$\int_\Omega s \, d\mu = \sum_{k=1}^{m} c_k \, \mu(A_k). \tag{2.2.19}$$

Definisi 2.2.20 (**Integral Lebesgue untuk Fungsi Terukur Non-Negatif**, (**?**)) *Untuk fungsi terukur non-negatif $f : \Omega \to [0, \infty]$, integral Lebesgue didefinisikan sebagai*

$$\int_\Omega f \, d\mu = \sup \left\{ \int_\Omega s \, d\mu : s \text{ sederhana}, 0 \leq s \leq f \right\}. \tag{2.2.20}$$

Integral ini selalu terdefinisi (meskipun mungkin bernilai ∞).

Definisi 2.2.21 (**Integral Lebesgue untuk Fungsi Terukur Umum**, (**?**)) *Untuk fungsi terukur $f : \Omega \to \mathbb{R}$, definisikan bagian positif $f^+ = \max(f, 0)$ dan bagian negatif $f^- = \max(-f, 0)$, sehingga $f = f^+ - f^-$. Integral Lebesgue dari f*

didefinisikan sebagai

$$\int_{\Omega} f \, d\mu = \int_{\Omega} f^{+} \, d\mu - \int_{\Omega} f^{-} \, d\mu, \tag{2.2.21}$$

dengan syarat setidaknya salah satu dari $\int_{\Omega} f^{+} \, d\mu$ *atau* $\int_{\Omega} f^{-} \, d\mu$ *bernilai hingga. Fungsi* f *dikatakan* terintegralkan Lebesgue *jika keduanya bernilai hingga, yaitu* $\int_{\Omega} |f| \, d\mu < \infty$.

Contoh 2.2.22 Misalkan $f : [0, 1] \to \mathbb{R}$ dengan $f(x) = x^2$. Integral Lebesgue dari f terhadap ukuran Lebesgue λ adalah

$$\int_{[0,1]} x^2 \, d\lambda(x) = \int_{0}^{1} x^2 \, dx = \frac{1}{3}. \tag{2.2.22}$$

Dalam kasus ini, integral Lebesgue sama dengan integral Riemann karena f kontinu pada interval tertutup.

Dengan menggunakan konsep integral Lebesgue yang telah didefinisikan, ekspektasi variabel acak dapat dirumuskan secara formal sebagai berikut.

Definisi 2.2.23 (**Ekspektasi Lebesgue**, (**?**)) *Misalkan* $(\Omega, \mathcal{F}, \Pr)$ *adalah ruang probabilitas dan* $X : \Omega \to \mathbb{R}$ *adalah variabel acak. Jika* X *terintegralkan Lebesgue (yaitu* $\int_{\Omega} |X(\omega)| \, d\Pr(\omega) < \infty$*), maka ekspektasi* X *didefinisikan sebagai integral Lebesgue terhadap ukuran probabilitas* $\Pr$*, yaitu*

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) \, d\Pr(\omega). \tag{2.2.23}$$

Secara ekuivalen, dengan menggunakan distribusi $\Pr_X$ *yang diinduksi oleh* X *pada* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$*, ekspektasi dapat dituliskan sebagai*

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \, d\Pr_X(x). \tag{2.2.24}$$

Jika X memiliki fungsi kepadatan probabilitas f_X, maka ekspektasi diberikan oleh

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx. \tag{2.2.25}$$

Contoh 2.2.24 Berikut adalah contoh perhitungan ekspektasi untuk beberapa distribusi.

(a) **Ekspektasi Distribusi Bernoulli.** Jika $X \sim \text{Bernoulli}(p)$, maka

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=0}^{1} k \cdot P(X = k) \\ &= 0 \cdot (1 - p) + 1 \cdot p = p. \end{aligned} \tag{2.2.26}$$

Ekspektasi menyatakan proporsi keberhasilan dalam jangka panjang. Jika $p = 0.3$, maka rata-rata 30% percobaan akan sukses.

(b) **Ekspektasi Distribusi Normal.** Jika $X \sim \mathcal{N}(\mu, \sigma^2)$, maka

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \mu. \tag{2.2.27}$$

Ekspektasi distribusi normal adalah parameter lokasinya μ, yang menyatakan pusat simetri dari distribusi.

(c) **Ekspektasi Distribusi Poisson.** Jika $X \sim \text{Poisson}(\lambda)$, maka

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!} \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda. \end{aligned} \tag{2.2.28}$$

Ekspektasi menyatakan rata-rata banyaknya kejadian dalam jangka panjang. Jika rata-rata terjadi $\lambda = 5$ kecelakaan per hari, maka ekspektasi banyaknya kecelakaan adalah 5.

(d) **Ekspektasi Distribusi Eksponensial.** Jika $X \sim \text{Exp}(\lambda)$, maka

$$\mathbb{E}[X] = \int_0^\infty x \cdot \lambda e^{-\lambda x} dx = \frac{1}{\lambda}. \tag{2.2.29}$$

Ekspektasi menyatakan rata-rata waktu tunggu antara kejadian. Jika $\lambda = 2$ kejadian per jam, maka rata-rata waktu tunggu adalah $1/2$ jam atau 30 menit.

Teorema 2.2.25 (**Sifat-Sifat Ekspektasi, (?)**) *Misalkan X dan Y adalah variabel acak pada ruang probabilitas $(\Omega, \mathcal{F}, \Pr)$ dengan ekspektasi terdefinisi. Berlaku sifat-sifat berikut.*

(a) ***Linearitas:*** *Untuk konstanta $a, b \in \mathbb{R}$, berlaku*

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]. \tag{2.2.30}$$

(b) ***Monotonisitas:*** *Jika $X \leq Y$ hampir pasti, maka*

$$\mathbb{E}[X] \leq \mathbb{E}[Y]. \tag{2.2.31}$$

(c) ***Ekspektasi konstanta:*** *Untuk konstanta $c \in \mathbb{R}$, berlaku*

$$\mathbb{E}[c] = c. \tag{2.2.32}$$

(d) ***Ketaknegatifan:*** *Jika $X \geq 0$ hampir pasti, maka*

$$\mathbb{E}[X] \geq 0. \tag{2.2.33}$$

(e) ***Ketaksamaan segitiga:*** *Berlaku*

$$|\mathbb{E}[X]| \leq \mathbb{E}[|X|]. \tag{2.2.34}$$

(f) ***Independensi:*** *Jika X dan Y independen, maka*

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]. \tag{2.2.35}$$

Bukti. Berikut adalah pembuktian untuk setiap sifat ekspektasi.

(a) **Linearitas:** Dengan menggunakan sifat integral Lebesgue, berlaku

$$\begin{aligned}\mathbb{E}[aX + bY] &= \int_\Omega (aX(\omega) + bY(\omega))\, d\mathrm{Pr}(\omega) \\ &= a \int_\Omega X(\omega)\, d\mathrm{Pr}(\omega) + b \int_\Omega Y(\omega)\, d\mathrm{Pr}(\omega) \\ &= a\mathbb{E}[X] + b\mathbb{E}[Y]. \end{aligned} \tag{2.2.36}$$

(b) **Monotonisitas:** Jika $X \leq Y$ hampir pasti, maka $Y - X \geq 0$ hampir pasti. Berdasarkan sifat integral Lebesgue untuk fungsi non-negatif, berlaku

$$\mathbb{E}[Y - X] = \mathbb{E}[Y] - \mathbb{E}[X] \geq 0, \tag{2.2.37}$$

sehingga $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

(c) **Ekspektasi konstanta:** Untuk konstanta c, variabel acak $X(\omega) = c$ untuk semua $\omega \in \Omega$, sehingga

$$\mathbb{E}[c] = \int_\Omega c\, d\mathrm{Pr}(\omega) = c \cdot \mathrm{Pr}(\Omega) = c \cdot 1 = c. \tag{2.2.38}$$

(d) **Ketaknegatifan:** Jika $X \geq 0$ hampir pasti, maka berdasarkan definisi integral Lebesgue untuk fungsi non-negatif, berlaku

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, d\mathrm{Pr}(\omega) \geq 0. \tag{2.2.39}$$

(e) **Ketaksamaan segitiga:** Perhatikan bahwa $-|X| \leq X \leq |X|$. Dengan sifat monotonisitas, diperoleh

$$-\mathbb{E}[|X|] \leq \mathbb{E}[X] \leq \mathbb{E}[|X|], \tag{2.2.40}$$

yang ekuivalen dengan $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.

(f) **Independensi:** Jika X dan Y independen, maka distribusi bersama sama dengan hasil kali distribusi marginal, yaitu $\Pr_{X,Y} = \Pr_X \times \Pr_Y$. Dengan teorema Fubini, berlaku

$$\begin{aligned}\mathbb{E}[XY] &= \int_{\mathbb{R}^2} xy \, d\Pr_{X,Y}(x,y) \\ &= \int_{\mathbb{R}} x \, d\Pr_X(x) \cdot \int_{\mathbb{R}} y \, d\Pr_Y(y) \\ &= \mathbb{E}[X] \cdot \mathbb{E}[Y]. \end{aligned} \tag{2.2.41}$$

■

Teorema 2.2.26 (Hukum Statistikawan Tak Sadar (*Law of the Unconscious Statistician*, LOTUS), (?)) *Misalkan X adalah variabel acak dengan fungsi distribusi F_X dan $g : \mathbb{R} \to \mathbb{R}$ adalah fungsi terukur. Maka ekspektasi dari $g(X)$ dapat dihitung tanpa perlu mengetahui distribusi $g(X)$ secara eksplisit, yaitu*

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \, dF_X(x). \tag{2.2.42}$$

Secara khusus:

(a) Jika X adalah variabel acak diskrit dengan fungsi massa probabilitas $p_X(x)$, maka

$$\mathbb{E}[g(X)] = \sum_x g(x) \, p_X(x). \tag{2.2.43}$$

(b) Jika X adalah variabel acak kontinu dengan fungsi kepadatan probabilitas $f_X(x)$, maka

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \, f_X(x) \, dx. \tag{2.2.44}$$

Bukti. Akan dibuktikan untuk kasus variabel acak kontinu dengan fungsi kepadatan probabilitas $f_X(x)$. Misalkan $Y = g(X)$ adalah variabel acak baru. Berdasarkan definisi ekspektasi, seharusnya perlu dicari fungsi kepadatan probabilitas $f_Y(y)$ dari Y terlebih dahulu, kemudian menghitung $\mathbb{E}[Y] = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy$.

Namun, dengan menggunakan teori ukuran, ekspektasi dari $g(X)$ dapat dihitung langsung sebagai integral Lebesgue terhadap ukuran probabilitas yang diinduksi oleh X. Secara formal, untuk setiap fungsi terukur g, berlaku

$$\mathbb{E}[g(X)] = \int_{\Omega} g(X(\omega))\, d\Pr(\omega) = \int_{\mathbb{R}} g(x)\, d\Pr_X(x), \tag{2.2.45}$$

dengan $\Pr_X$ adalah distribusi yang diinduksi oleh X. Sebab X memiliki fungsi kepadatan probabilitas $f_X(x)$, maka $d\Pr_X(x) = f_X(x)\, dx$, sehingga

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx. \tag{2.2.46}$$

Hasil ini diperoleh langsung dari teorema perubahan variabel untuk integral Lebesgue tanpa perlu menghitung distribusi $g(X)$ secara eksplisit. ■

Contoh 2.2.27 Misalkan $X \sim \text{Exp}(1)$ dengan fungsi kepadatan probabilitas $f_X(x) = e^{-x}$ untuk $x \geq 0$. Akan dihitung $\mathbb{E}[X^2]$.

Dengan menggunakan LOTUS, tidak perlu dicari distribusi dari $Y = X^2$ terlebih dahulu. Langsung dapat dihitung

$$\begin{aligned} \mathbb{E}[X^2] &= \int_0^{\infty} x^2 \cdot e^{-x}\, dx \\ &= \Gamma(3) = 2! = 2. \end{aligned} \tag{2.2.47}$$

Hasil ini diperoleh dengan menggunakan integral Gamma $\Gamma(n) = \int_0^{\infty} x^{n-1} e^{-x}\, dx = (n-1)!$ untuk n bilangan bulat positif.

Sebagai contoh lain, jika $X \sim \mathcal{N}(0, 1)$ dan ingin dihitung $\mathbb{E}[e^X]$, maka dengan LOTUS:

$$\begin{aligned}\mathbb{E}[e^X] &= \int_{-\infty}^{\infty} e^x \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{x-x^2/2}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x^2-2x)/2}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-1)^2/2+1/2}\, dx \\ &= e^{1/2} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-1)^2/2}\, dx \\ &= e^{1/2} = \sqrt{e}. \end{aligned} \tag{2.2.48}$$

2.2.4 Momen, Variansi, dan Kovariansi

Definisi 2.2.28 (**Momen**, (**?**)) *Untuk variabel acak X dan bilangan bulat positif k, momen ke-k dari X didefinisikan sebagai*

$$\mu_k = \mathbb{E}[X^k], \tag{2.2.49}$$

dengan syarat $\mathbb{E}[|X|^k] < \infty$. Momen pertama $\mu_1 = \mathbb{E}[X]$ disebut sebagai rataan atau nilai harapan dari X. Selain itu, momen pusat ke-k didefinisikan sebagai

$$\mu'_k = \mathbb{E}\big[(X - \mathbb{E}[X])^k\big]. \tag{2.2.50}$$

Definisi 2.2.29 (**Variansi**, (**?**)) *Untuk variabel acak X dengan $\mathbb{E}[|X|^2] < \infty$, variansi didefinisikan sebagai momen pusat kedua, yaitu*

$$\mathrm{Var}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big]. \tag{2.2.51}$$

Variansi dapat juga dihitung menggunakan rumus alternatif

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2. \tag{2.2.52}$$

Contoh 2.2.30 Misalkan X adalah variabel acak diskrit dengan distribusi

$$\Pr(X = 1) = 0.3, \quad \Pr(X = 2) = 0.5, \quad \Pr(X = 3) = 0.2.$$

Momen pertama (rataan) dari X adalah

$$\begin{aligned}\mathbb{E}[X] &= \sum_x x \cdot \Pr(X = x) \\ &= 1 \cdot 0.3 + 2 \cdot 0.5 + 3 \cdot 0.2 = 0.3 + 1.0 + 0.6 = 1.9. \end{aligned} \tag{2.2.53}$$

Momen kedua dari X adalah

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_x x^2 \cdot \Pr(X = x) \\ &= 1^2 \cdot 0.3 + 2^2 \cdot 0.5 + 3^2 \cdot 0.2 = 0.3 + 2.0 + 1.8 = 4.1. \end{aligned} \tag{2.2.54}$$

Variansi dari X dapat dihitung dengan rumus alternatif, yaitu

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 4.1 - (1.9)^2 = 4.1 - 3.61 = 0.49. \tag{2.2.55}$$

Definisi 2.2.31 (Kovariansi, (?)) *Untuk dua variabel acak X dan Y dengan momen orde dua hingga, kovariansi didefinisikan sebagai*

$$\mathrm{Cov}(X, Y) = \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big]. \tag{2.2.56}$$

Rumus alternatif untuk kovariansi adalah

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \tag{2.2.57}$$

Contoh 2.2.32 Misalkan (X, Y) adalah variabel acak gabungan dengan distribusi bersama

$$\begin{aligned}
&\Pr(X=0,Y=0)=0.2, \quad \Pr(X=0,Y=1)=0.1, \\
&\Pr(X=1,Y=0)=0.3, \quad \Pr(X=1,Y=1)=0.4. \qquad (2.2.58)
\end{aligned}$$

Rataan dari X dan Y adalah

$$\mathbb{E}[X]=0\cdot(0.2+0.1)+1\cdot(0.3+0.4)=0.7, \qquad (2.2.59)$$

$$\mathbb{E}[Y]=0\cdot(0.2+0.3)+1\cdot(0.1+0.4)=0.5. \qquad (2.2.60)$$

Nilai harapan dari XY adalah

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_{x,y} xy\cdot\Pr(X=x,Y=y) \\
&= 0\cdot 0\cdot 0.2+0\cdot 1\cdot 0.1+1\cdot 0\cdot 0.3+1\cdot 1\cdot 0.4=0.4. \qquad (2.2.61)
\end{aligned}$$

Kovariansi dari X dan Y adalah

$$\mathrm{Cov}(X,Y)=\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]=0.4-0.7\cdot 0.5=0.4-0.35=0.05. \quad (2.2.62)$$

Nilai kovariansi positif menunjukkan bahwa X dan Y cenderung bergerak searah.

2.2.5 Probabilitas Bersyarat dan Independensi

Probabilitas bersyarat merupakan konsep fundamental yang menggambarkan bagaimana informasi baru mengubah keyakinan kita terhadap suatu kejadian. Konsep ini menjadi dasar bagi inferensi statistik dan pembelajaran mesin.

Definisi 2.2.33 (Probabilitas Bersyarat, (?)) *Misalkan A dan B adalah dua kejadian pada ruang probabilitas $(\Omega,\mathcal{F},\Pr)$ dengan $\Pr(B)>0$. Probabilitas bersyarat dari A diberikan B, dinotasikan $\Pr(A\mid B)$, didefinisikan sebagai*

$$\Pr(A\mid B)=\frac{\Pr(A\cap B)}{\Pr(B)}. \qquad (2.2.63)$$

Contoh 2.2.34 Misalkan sebuah dadu seimbang dilempar. Definisikan kejadian $A = \{\text{hasil genap}\} = \{2, 4, 6\}$ dan $B = \{\text{hasil lebih dari 3}\} = \{4, 5, 6\}$. Maka

$$\Pr(A) = \frac{3}{6} = \frac{1}{2}, \quad \Pr(B) = \frac{3}{6} = \frac{1}{2}, \tag{2.2.64}$$

$$\Pr(A \cap B) = \Pr(\{4, 6\}) = \frac{2}{6} = \frac{1}{3}. \tag{2.2.65}$$

Probabilitas bersyarat dari A diberikan B adalah

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{1/3}{1/2} = \frac{2}{3}. \tag{2.2.66}$$

Hasil ini menunjukkan bahwa jika diketahui hasil dadu lebih dari 3, maka probabilitas hasil genap meningkat dari $1/2$ menjadi $2/3$.

Definisi 2.2.35 (Independensi Dua Kejadian, (?)) *Dua kejadian A dan B dikatakan independen apabila*

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B). \tag{2.2.67}$$

Secara ekuivalen, jika $\Pr(B) > 0$, maka A dan B independen jika dan hanya jika $\Pr(A \mid B) = \Pr(A)$.

Contoh 2.2.36 Pertimbangkan pelemparan dua koin seimbang secara independen. Definisikan $A = \{\text{koin pertama muncul kepala}\}$ dan $B = \{\text{koin kedua muncul kepala}\}$. Ruang sampel adalah $\Omega = \{KK, KE, EK, EE\}$ dengan probabilitas seragam. Maka

$$\Pr(A) = \Pr(\{KK, KE\}) = \frac{2}{4} = \frac{1}{2}, \tag{2.2.68}$$

$$\Pr(B) = \Pr(\{KK, EK\}) = \frac{2}{4} = \frac{1}{2}, \tag{2.2.69}$$

$$\Pr(A \cap B) = \Pr(\{KK\}) = \frac{1}{4}. \tag{2.2.70}$$

Sebab $\Pr(A \cap B) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \Pr(A) \cdot \Pr(B)$, maka A dan B independen.

Definisi 2.2.37 (Independensi Variabel Acak, (?)) *Variabel acak X dan Y dikatakan independen apabila untuk setiap himpunan Borel $A, B \subseteq \mathbb{R}$ berlaku*

$$\Pr(X \in A,\, Y \in B) = \Pr(X \in A) \Pr(Y \in B). \tag{2.2.71}$$

Secara ekuivalen, X dan Y independen jika dan hanya jika fungsi distribusi kumulatif bersama memenuhi

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y), \quad \forall x, y \in \mathbb{R}. \tag{2.2.72}$$

Contoh 2.2.38 Misalkan X dan Y adalah variabel acak independen dengan $X \sim$ Bernoulli(0.5) dan $Y \sim$ Bernoulli(0.5). Maka

$$\Pr(X = 1, Y = 1) = \Pr(X = 1) \cdot \Pr(Y = 1) = 0.5 \cdot 0.5 = 0.25, \tag{2.2.73}$$
$$\Pr(X = 0, Y = 1) = \Pr(X = 0) \cdot \Pr(Y = 1) = 0.5 \cdot 0.5 = 0.25. \tag{2.2.74}$$

Distribusi bersama sepenuhnya ditentukan oleh hasil kali distribusi marginal.

Definisi 2.2.39 (Independensi Bersyarat, (?)) *Dua kejadian A dan B dikatakan independen bersyarat diberikan kejadian C (dengan $\Pr(C) > 0$) apabila*

$$\Pr(A \cap B \mid C) = \Pr(A \mid C) \cdot \Pr(B \mid C). \tag{2.2.75}$$

Untuk variabel acak, X dan Y independen bersyarat diberikan Z apabila untuk setiap nilai z dengan $\Pr(Z = z) > 0$ berlaku

$$\Pr(X \in A, Y \in B \mid Z = z) = \Pr(X \in A \mid Z = z) \cdot \Pr(Y \in B \mid Z = z). \tag{2.2.76}$$

2.2.6 Teorema Bayes dan Hukum Probabilitas Total

Teorema 2.2.40 (Hukum Probabilitas Total, (?)) *Misalkan himpunan $\{B_1, B_2, \ldots, B_n\}$ adalah partisi dari ruang sampel Ω, yaitu $B_i \cap B_j = \varnothing$ untuk $i \neq j$ dan $\bigcup_{i=1}^{n} B_i =$*

Ω, *dengan* $\Pr(B_i) > 0$ *untuk semua* i. *Maka untuk setiap kejadian* A *berlaku*

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A \mid B_i) \cdot \Pr(B_i). \tag{2.2.77}$$

Untuk kasus kontinu dengan variabel acak Y*, hukum ini dapat ditulis sebagai*

$$\Pr(A) = \int_{-\infty}^{\infty} \Pr(A \mid Y = y) \cdot f_Y(y)\, dy. \tag{2.2.78}$$

Bukti. Sebab $\{B_1, B_2, \ldots, B_n\}$ adalah partisi dari Ω, maka

$$A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{n} B_i \right) = \bigcup_{i=1}^{n} (A \cap B_i). \tag{2.2.79}$$

Sebab kejadian $A \cap B_i$ saling lepas untuk i berbeda, maka berdasarkan σ-aditivitas:

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A \cap B_i) = \sum_{i=1}^{n} \Pr(A \mid B_i) \cdot \Pr(B_i). \tag{2.2.80}$$

■

Contoh 2.2.41 Suatu pabrik memiliki tiga mesin yang memproduksi komponen elektronik. Mesin 1 memproduksi 50% dari total produksi dengan tingkat cacat 2%, mesin 2 memproduksi 30% dengan tingkat cacat 3%, dan mesin 3 memproduksi 20% dengan tingkat cacat 5%. Probabilitas suatu komponen yang dipilih secara acak adalah cacat dapat dihitung dengan hukum probabilitas total:

$$\begin{aligned} \Pr(\text{Cacat}) &= \Pr(\text{Cacat} \mid \text{Mesin } 1) \cdot \Pr(\text{Mesin } 1) \\ &\quad + \Pr(\text{Cacat} \mid \text{Mesin } 2) \cdot \Pr(\text{Mesin } 2) \\ &\quad + \Pr(\text{Cacat} \mid \text{Mesin } 3) \cdot \Pr(\text{Mesin } 3) \\ &= 0.02 \cdot 0.50 + 0.03 \cdot 0.30 + 0.05 \cdot 0.20 \\ &= 0.01 + 0.009 + 0.01 = 0.029. \end{aligned} \tag{2.2.81}$$

Jadi, probabilitas suatu komponen adalah cacat adalah 2.9%.

Teorema 2.2.42 (**Teorema Bayes**, (**?**)) *Misalkan* $\{B_1, B_2, \ldots, B_n\}$ *adalah partisi dari ruang sampel* Ω *dengan* $\Pr(B_i) > 0$ *untuk semua* i*, dan misalkan* A *adalah kejadian dengan* $\Pr(A) > 0$*. Maka untuk setiap* $j = 1, 2, \ldots, n$ *berlaku*

$$\Pr(B_j \mid A) = \frac{\Pr(A \mid B_j) \cdot \Pr(B_j)}{\sum_{i=1}^{n} \Pr(A \mid B_i) \cdot \Pr(B_i)}. \tag{2.2.82}$$

Dalam bentuk yang lebih sederhana untuk dua kejadian A *dan* B*:*

$$\Pr(B \mid A) = \frac{\Pr(A \mid B) \cdot \Pr(B)}{\Pr(A)}. \tag{2.2.83}$$

Bukti. Berdasarkan definisi probabilitas bersyarat:

$$\Pr(B_j \mid A) = \frac{\Pr(A \cap B_j)}{\Pr(A)} = \frac{\Pr(A \mid B_j) \cdot \Pr(B_j)}{\Pr(A)}. \tag{2.2.84}$$

Dengan mensubstitusi $\Pr(A)$ menggunakan hukum probabilitas total, diperoleh hasil yang diinginkan. ∎

Contoh 2.2.43 Melanjutkan contoh pabrik sebelumnya, misalkan suatu komponen yang dipilih secara acak ternyata cacat. Probabilitas komponen tersebut berasal dari mesin 3 dapat dihitung dengan teorema Bayes:

$$\begin{aligned}\Pr(\text{Mesin 3} \mid \text{Cacat}) &= \frac{\Pr(\text{Cacat} \mid \text{Mesin 3}) \cdot \Pr(\text{Mesin 3})}{\Pr(\text{Cacat})} \\ &= \frac{0.05 \cdot 0.20}{0.029} = \frac{0.01}{0.029} \approx 0.345. \end{aligned} \tag{2.2.85}$$

Meskipun mesin 3 hanya memproduksi 20% dari total produksi, probabilitas komponen cacat berasal dari mesin 3 adalah sekitar 34.5% karena tingkat kecacatannya yang lebih tinggi.

Definisi 2.2.44 (**Terminologi Bayesian**) *Dalam konteks teorema Bayes, terdapat beberapa terminologi penting:*

(a) $\Pr(B)$ *disebut probabilitas* prior*, yaitu keyakinan awal tentang* B *sebelum mengamati data* A*.*

(b) $\Pr(A \mid B)$ *disebut* likelihood*, yaitu probabilitas mengamati data A diberikan B benar.*

(c) $\Pr(B \mid A)$ *disebut probabilitas* posterior*, yaitu keyakinan tentang B setelah mengamati data A.*

(d) $\Pr(A)$ *disebut* marginal likelihood *atau* evidence*, yang berfungsi sebagai konstanta normalisasi.*

Teorema Bayes dapat diringkas sebagai: Posterior $\propto$ Likelihood $\times$ Prior.

2.2.7 Hukum Ekspektasi Total dan Variansi Total

Teorema 2.2.45 (**Hukum Ekspektasi Total (Law of Total Expectation)**, (**?**)) *Misalkan X dan Y adalah variabel acak pada ruang probabilitas yang sama dengan $\mathbb{E}[|X|] < \infty$. Maka*

$$\mathbb{E}[X] = \mathbb{E}\left[\mathbb{E}[X \mid Y]\right]. \tag{2.2.86}$$

Untuk kasus diskrit dengan Y mengambil nilai $y_1, y_2, \ldots$:

$$\mathbb{E}[X] = \sum_j \mathbb{E}[X \mid Y = y_j] \cdot \Pr(Y = y_j). \tag{2.2.87}$$

Untuk kasus kontinu dengan Y memiliki kepadatan f_Y:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X \mid Y = y] \cdot f_Y(y)\, dy. \tag{2.2.88}$$

Bukti. Untuk kasus diskrit, dengan menggunakan definisi ekspektasi bersyarat:

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}[X \mid Y]\right] &= \sum_j \mathbb{E}[X \mid Y = y_j] \cdot \Pr(Y = y_j) \\
&= \sum_j \left(\sum_i x_i \cdot \Pr(X = x_i \mid Y = y_j) \right) \cdot \Pr(Y = y_j) \\
&= \sum_i \sum_j x_i \cdot \Pr(X = x_i \mid Y = y_j) \cdot \Pr(Y = y_j) \\
&= \sum_i \sum_j x_i \cdot \Pr(X = x_i, Y = y_j) \\
&= \sum_i x_i \cdot \Pr(X = x_i) = \mathbb{E}[X].
\end{aligned} \tag{2.2.89}$$

Untuk kasus kontinu, misalkan Y memiliki fungsi kepadatan probabilitas $f_Y(y)$ dan ekspektasi bersyarat $\mathbb{E}[X \mid Y = y]$ terdefinisi dengan baik. Dengan menggunakan definisi ekspektasi bersyarat untuk variabel acak kontinu:

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}[X \mid Y]\right] &= \int_{-\infty}^{\infty} \mathbb{E}[X \mid Y = y] \cdot f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x \cdot f_{X|Y}(x \mid y)\, dx \right) \cdot f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x \mid y) \cdot f_Y(y)\, dx\, dy.
\end{aligned} \tag{2.2.90}$$

Dengan menggunakan hubungan $f_{X,Y}(x, y) = f_{X|Y}(x \mid y) \cdot f_Y(y)$, diperoleh

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}[X \mid Y]\right] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f_{X,Y}(x, y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy \right) dx \\
&= \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx = \mathbb{E}[X],
\end{aligned} \tag{2.2.91}$$

dengan $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy$ adalah fungsi kepadatan marginal dari X. Pertukaran urutan integrasi dijamin oleh teorema Fubini dengan syarat $\mathbb{E}[|X|] < \infty$. ■

Contoh 2.2.46 Misalkan N adalah variabel acak diskrit yang menyatakan jumlah pelanggan yang datang ke suatu toko dalam sehari, dengan $N \sim \text{Poisson}(\lambda)$. Setiap pelanggan membeli sejumlah barang X_i yang berdistribusi identik dan independen dengan $\mathbb{E}[X_i] = \mu$. Total penjualan harian adalah $S = \sum_{i=1}^{N} X_i$. Ekspektasi total penjualan dapat dihitung dengan hukum ekspektasi total:

$$\begin{aligned}\mathbb{E}[S] = \mathbb{E}\left[\mathbb{E}[S \mid N]\right] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{N} X_i \mid N\right]\right] \\ &= \mathbb{E}[N \cdot \mu] = \mu \cdot \mathbb{E}[N] = \mu\lambda. \end{aligned} \tag{2.2.92}$$

Teorema 2.2.47 (**Hukum Variansi Total (Law of Total Variance)**, (**?**)) *Misalkan X dan Y adalah variabel acak dengan* $\text{Var}(X) < \infty$. *Maka*

$$\text{Var}(X) = \mathbb{E}\left[\text{Var}(X \mid Y)\right] + \text{Var}\left(\mathbb{E}[X \mid Y]\right). \tag{2.2.93}$$

Rumus ini dikenal juga sebagai dekomposisi variansi, dengan interpretasi:

(a) $\mathbb{E}\left[\text{Var}(X \mid Y)\right]$ *adalah rata-rata variansi dalam kelompok (*within-group variance*).*

(b) $\text{Var}\left(\mathbb{E}[X \mid Y]\right)$ *adalah variansi antar kelompok (*between-group variance*).*

Bukti. Dengan menggunakan rumus $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$, diperoleh

$$\text{Var}(X \mid Y) = \mathbb{E}[X^2 \mid Y] - (\mathbb{E}[X \mid Y])^2. \tag{2.2.94}$$

Dengan mengambil ekspektasi kedua ruas:

$$\mathbb{E}\left[\text{Var}(X \mid Y)\right] = \mathbb{E}\left[\mathbb{E}[X^2 \mid Y]\right] - \mathbb{E}\left[(\mathbb{E}[X \mid Y])^2\right] = \mathbb{E}[X^2] - \mathbb{E}\left[(\mathbb{E}[X \mid Y])^2\right]. \tag{2.2.95}$$

Di sisi lain:

$$\begin{aligned}\text{Var}\left(\mathbb{E}[X \mid Y]\right) &= \mathbb{E}\left[(\mathbb{E}[X \mid Y])^2\right] - \left(\mathbb{E}\left[\mathbb{E}[X \mid Y]\right]\right)^2 \\ &= \mathbb{E}\left[(\mathbb{E}[X \mid Y])^2\right] - (\mathbb{E}[X])^2. \end{aligned} \tag{2.2.96}$$

Dengan menjumlahkan kedua hasil:

$$\begin{aligned}\mathbb{E}\left[\operatorname{Var}(X \mid Y)\right] + \operatorname{Var}\left(\mathbb{E}[X \mid Y]\right) &= \mathbb{E}[X^2] - \mathbb{E}\left[(\mathbb{E}[X \mid Y])^2\right] \\ &\quad + \mathbb{E}\left[(\mathbb{E}[X \mid Y])^2\right] - (\mathbb{E}[X])^2 \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \operatorname{Var}(X). \end{aligned} \tag{2.2.97}$$

■

Contoh 2.2.48 Melanjutkan contoh sebelumnya, variansi total penjualan harian $S = \sum_{i=1}^{N} X_i$ dapat dihitung dengan hukum variansi total. Misalkan $\operatorname{Var}(X_i) = \sigma^2$. Maka

$$\operatorname{Var}(S \mid N) = N \cdot \sigma^2, \tag{2.2.98}$$

$$\mathbb{E}[S \mid N] = N \cdot \mu. \tag{2.2.99}$$

Dengan hukum variansi total:

$$\begin{aligned}\operatorname{Var}(S) &= \mathbb{E}\left[\operatorname{Var}(S \mid N)\right] + \operatorname{Var}\left(\mathbb{E}[S \mid N]\right) \\ &= \mathbb{E}[N \cdot \sigma^2] + \operatorname{Var}(N \cdot \mu) \\ &= \sigma^2 \cdot \mathbb{E}[N] + \mu^2 \cdot \operatorname{Var}(N) \\ &= \sigma^2\lambda + \mu^2\lambda = \lambda(\sigma^2 + \mu^2). \end{aligned} \tag{2.2.100}$$

Hasil ini menunjukkan bahwa variansi total penjualan terdiri dari dua komponen: variansi akibat fluktuasi pembelian per pelanggan dan variansi akibat fluktuasi jumlah pelanggan.

Ketergantungan antar variabel acak merupakan aspek fundamental dalam pemodelan statistik, khususnya dalam analisis spasial dan data bertautan, ketika asumsi independensi sering kali tidak terpenuhi.

2.3 Konvergensi dan Laju Pertumbuhan

Analisis asimtotik dalam statistika memerlukan pemahaman yang jelas mengenai perilaku barisan variabel acak ketika ukuran sampel meningkat. Konsep konvergensi dan notasi laju pertumbuhan menyediakan bahasa formal untuk menyatakan konsistensi, distribusi limit, serta ukuran galat dari suatu estimator. Pembahasan dalam bagian ini mengikuti kerangka teori probabilitas klasik dan statistik asimtotik sebagaimana dikembangkan oleh **?**, **?**, dan **?**. Namun, sebelum memasuki konvergensi dan laju pertumbuhan, terdapat dua ketaksamaan penting yang sering digunakan dalam pembuktian teorema limit, yaitu ketaksamaan Markov dan Chebyshev.

Teorema 2.3.1 (**Ketaksamaan Markov**, (**?**)) *Misalkan X adalah variabel acak non-negatif dengan ekspektasi $\mathbb{E}[X] < \infty$. Maka untuk setiap $a > 0$ berlaku*

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}. \tag{2.3.1}$$

Bukti. Dengan definisi ekspektasi:

$$\mathbb{E}[X] = \int_0^\infty \Pr(X \geq t)\, dt. \tag{2.3.2}$$

Karena X non-negatif, maka

$$\mathbb{E}[X] \geq \int_a^\infty \Pr(X \geq t)\, dt \geq \int_a^\infty \Pr(X \geq a)\, dt = \Pr(X \geq a) \cdot (\infty - a). \tag{2.3.3}$$

Oleh karena itu,

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}. \tag{2.3.4}$$

■

Teorema 2.3.2 (**Ketaksamaan Chebyshev**, (**?**)) *Misalkan X adalah variabel acak dengan ekspektasi $\mathbb{E}[X]$ dan variansi $\mathrm{Var}(X) < \infty$. Maka untuk setiap $\varepsilon > 0$ berlaku*

$$\Pr(|X - \mathbb{E}[X]| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}. \tag{2.3.5}$$

Bukti. Definisikan variabel acak non-negatif $Y = (X - \mathbb{E}[X])^2$. Dengan menggunakan ketaksamaan Markov pada Y, diperoleh

$$\Pr(|X - \mathbb{E}[X]| \geq \varepsilon) = \Pr(Y \geq \varepsilon^2) \leq \frac{\mathbb{E}[Y]}{\varepsilon^2} = \frac{\mathrm{Var}(X)}{\varepsilon^2}. \tag{2.3.6}$$

■

2.3.1 Barisan Variabel Acak

Suatu *barisan variabel acak* adalah koleksi variabel acak $\{X_n\}_{n\geq 1}$ yang didefinisikan pada ruang probabilitas yang sama $(\Omega, \mathcal{F}, \Pr)$. Dalam konteks statistika, barisan ini biasanya merepresentasikan estimator yang bergantung pada ukuran sampel n (**?**).

2.3.2 Konvergensi Barisan Variabel Acak

Pemahaman mengenai berbagai mode konvergensi barisan variabel acak sangat penting dalam statistika asimtotik. Tiga mode konvergensi yang paling umum adalah konvergensi hampir pasti, konvergensi dalam probabilitas, dan konvergensi dalam distribusi.

Definisi 2.3.3 (**Konvergensi Hampir Pasti**, (**?**)) *Barisan variabel acak $\{X_n\}$ dikatakan konvergen hampir pasti (*almost surely*) ke X, ditulis $X_n \xrightarrow{a.s.} X$, apabila*

$$\Pr\big(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\big) = 1. \tag{2.3.7}$$

Contoh 2.3.4 Misalkan $\Omega = [0, 1]$ dengan ukuran probabilitas Lebesgue $\Pr$ dan definisikan barisan variabel acak $X_n(\omega) = \omega^n$ untuk $\omega \in [0, 1]$. Untuk setiap $\omega \in [0, 1)$, berlaku

$$\lim_{n\to\infty} X_n(\omega) = \lim_{n\to\infty} \omega^n = 0.$$

Untuk $\omega = 1$, berlaku $X_n(1) = 1$ untuk semua n, sehingga $\lim_{n\to\infty} X_n(1) = 1$.

Definisikan $X(\omega) = 0$ untuk $\omega \in [0, 1)$ dan $X(1) = 1$. Maka

$$\Pr\big(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\big) = \Pr([0, 1]) = 1,$$

sehingga $X_n \to X$ hampir pasti.

Definisi 2.3.5 (**Konvergensi dalam Probabilitas**, (**?**)) *Barisan $\{X_n\}$ dikatakan konvergen dalam probabilitas ke X, ditulis $X_n \xrightarrow{p} X$, apabila untuk setiap $\varepsilon > 0$ berlaku*

$$\lim_{n\to\infty} \Pr(|X_n - X| > \varepsilon) = 0. \tag{2.3.8}$$

$X_n \xrightarrow{p} X$ dapat dinotasikan juga sebagai $\operatorname*{plim}_{n\to\infty} X_n = X$.

Contoh 2.3.6 Misalkan X_n adalah variabel acak dengan distribusi

$$\Pr(X_n = 0) = 1 - \frac{1}{n}, \quad \Pr(X_n = n) = \frac{1}{n}.$$

Akan ditunjukkan bahwa $X_n \xrightarrow{p} 0$. Untuk setiap $\varepsilon > 0$, berlaku

$$\Pr(|X_n - 0| > \varepsilon) = \Pr(X_n > \varepsilon) = \Pr(X_n = n) = \frac{1}{n} \to 0$$

ketika $n \to \infty$. Oleh karena itu, $X_n \xrightarrow{p} 0$. Namun, perhatikan bahwa X_n tidak konvergen hampir pasti ke 0, karena untuk setiap ω terdapat tak hingga banyak n sedemikian sehingga $X_n(\omega) = n$ (dengan probabilitas positif).

Konvergensi dalam probabilitas merupakan konsep utama dalam pembahasan konsistensi estimator, sedangkan konvergensi hampir pasti sering digunakan sebagai alat teknis dalam pembuktian teorema limit.

Definisi 2.3.7 (**Konvergensi dalam Distribusi**, (**?**)) *Barisan $\{X_n\}$ dikatakan konvergen dalam distribusi ke X, ditulis $X_n \Rightarrow X$ atau $X_n \xrightarrow{d} X$, apabila*

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)] \tag{2.3.9}$$

untuk setiap fungsi kontinu terbatas f.

Contoh 2.3.8 Misalkan $X_n \sim \mathcal{N}(0, 1+1/n)$ untuk $n = 1, 2, \ldots$. Akan ditunjukkan bahwa $X_n \Rightarrow X$ dengan $X \sim \mathcal{N}(0, 1)$. Untuk setiap fungsi kontinu terbatas f, berlaku

$$\mathbb{E}[f(X_n)] = \int_{-\infty}^{\infty} f(x) \cdot \frac{1}{\sqrt{2\pi(1+1/n)}} \exp\left(-\frac{x^2}{2(1+1/n)}\right) dx.$$

Ketika $n \to \infty$, variansi $1 + 1/n \to 1$, sehingga

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \int_{-\infty}^{\infty} f(x) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = \mathbb{E}[f(X)].$$

Oleh karena itu, $X_n \Rightarrow X \sim \mathcal{N}(0, 1)$.

Konvergensi dalam distribusi memungkinkan pendekatan distribusi *sampling* yang kompleks dengan distribusi limit yang lebih sederhana, dan merupakan dasar utama Teorema Limit Pusat.

Definisi 2.3.9 (**Konvergensi dalam Rataan Kuadrat**, (**?**)) *Barisan variabel acak $\{X_n\}$ dikatakan konvergen dalam rataan kuadrat (*mean square*) ke X, ditulis $X_n \xrightarrow{L^2} X$ atau $X_n \xrightarrow{m.s.} X$, apabila*

$$\lim_{n\to\infty} \mathbb{E}\left[(X_n - X)^2\right] = 0. \tag{2.3.10}$$

Konvergensi ini juga dikenal sebagai konvergensi dalam L^2 atau konvergensi kuadrat rata-rata.

Contoh 2.3.10 Misalkan X_n adalah variabel acak dengan distribusi

$$\Pr(X_n = 0) = 1 - \frac{1}{n^2}, \quad \Pr(X_n = 1) = \frac{1}{n^2}.$$

Akan ditunjukkan bahwa $X_n \xrightarrow{L^2} 0$. Perhatikan bahwa

$$\mathbb{E}\left[(X_n - 0)^2\right] = \mathbb{E}[X_n^2] = 0^2 \cdot \left(1 - \frac{1}{n^2}\right) + 1^2 \cdot \frac{1}{n^2} = \frac{1}{n^2} \to 0$$

ketika $n \to \infty$. Oleh karena itu, $X_n \xrightarrow{L^2} 0$.

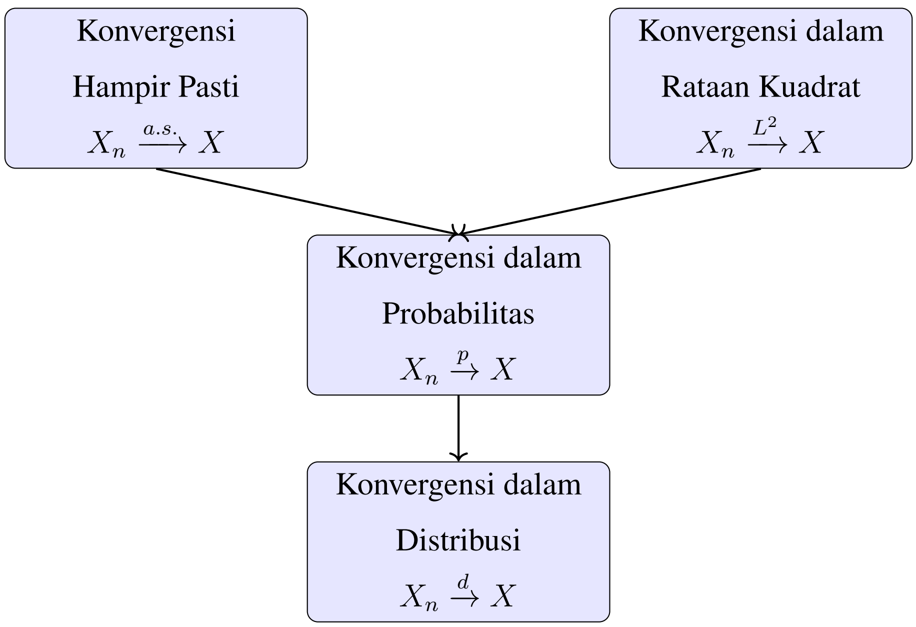
Contoh 2.3.11 Misalkan $X_n \sim \mathcal{N}(0, 1/n)$ untuk $n = 1, 2, \ldots$. Akan ditunjukkan bahwa $X_n \xrightarrow{L^2} 0$. Perhatikan bahwa

$$\mathbb{E}\left[(X_n - 0)^2\right] = \mathbb{E}[X_n^2] = \mathrm{Var}(X_n) + (\mathbb{E}[X_n])^2 = \frac{1}{n} + 0 = \frac{1}{n} \to 0$$

ketika $n \to \infty$. Oleh karena itu, $X_n \xrightarrow{L^2} 0$.

2.3.3 Hubungan Antarmode Konvergensi

Berbagai mode konvergensi memiliki hubungan hierarkis yang penting dalam teori probabilitas dan statistika asimtotik. Hubungan ini dapat divisualisasikan dalam diagram berikut.



Gambar 2.5 Diagram hubungan antarmode konvergensi

Teorema 2.3.12 *Jika $X_n \xrightarrow{a.s.} X$, maka $X_n \xrightarrow{p} X$.*

Bukti. Andaikan $X_n \xrightarrow{a.s.} X$, yaitu $\Pr\left(\lim_{n\to\infty} X_n = X\right) = 1$. Definisikan himpunan

$$A = \left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}. \tag{2.3.11}$$

Berdasarkan hipotesis, $\Pr(A) = 1$. Untuk setiap $\omega \in A$ dan setiap $\varepsilon > 0$, terdapat $N(\omega, \varepsilon)$ sehingga $|X_n(\omega) - X(\omega)| < \varepsilon$ untuk semua $n \geq N(\omega, \varepsilon)$.

Definisikan himpunan $B_n(\varepsilon) = \{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}$. Untuk setiap $\omega \in A$, terdapat N sehingga $\omega \notin B_n(\varepsilon)$ untuk $n \geq N$. Dengan kata lain,

$$A \subseteq \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} B_n(\varepsilon)^c. \tag{2.3.12}$$

Hal ini berarti $\limsup_{n\to\infty} B_n(\varepsilon) \subseteq A^c$, sehingga

$$\Pr\left(\limsup_{n\to\infty} B_n(\varepsilon)\right) \leq \Pr(A^c) = 0. \tag{2.3.13}$$

Berdasarkan lemma Borel–Cantelli, jika $\Pr(\limsup_n B_n) = 0$, maka $\Pr(B_n) \to 0$. Oleh karena itu,

$$\lim_{n\to\infty} \Pr(|X_n - X| > \varepsilon) = 0, \tag{2.3.14}$$

yang berarti $X_n \xrightarrow{p} X$. ■

Teorema 2.3.13 *Jika* $X_n \xrightarrow{L^2} X$*, maka* $X_n \xrightarrow{p} X$.

Bukti. Andaikan $X_n \xrightarrow{L^2} X$, yaitu $\mathbb{E}[(X_n - X)^2] \to 0$. Akan ditunjukkan bahwa untuk setiap $\varepsilon > 0$, berlaku $\Pr(|X_n - X| > \varepsilon) \to 0$.

Dengan menggunakan ketaksamaan Markov pada variabel acak non-negatif $(X_n - X)^2$, diperoleh

$$\Pr(|X_n - X| > \varepsilon) = \Pr\left((X_n - X)^2 > \varepsilon^2\right) \leq \frac{\mathbb{E}[(X_n - X)^2]}{\varepsilon^2}. \tag{2.3.15}$$

Sebab $\mathbb{E}[(X_n - X)^2] \to 0$ ketika $n \to \infty$, maka

$$\lim_{n\to\infty} \Pr(|X_n - X| > \varepsilon) \leq \lim_{n\to\infty} \frac{\mathbb{E}[(X_n - X)^2]}{\varepsilon^2} = 0. \tag{2.3.16}$$

Sebab probabilitas selalu non-negatif, diperoleh $\Pr(|X_n - X| > \varepsilon) \to 0$, yang berarti $X_n \xrightarrow{p} X$. ■

Teorema 2.3.14 *Jika $X_n \xrightarrow{p} X$, maka $X_n \xrightarrow{d} X$.*

Bukti. Andaikan $X_n \xrightarrow{p} X$. Akan ditunjukkan bahwa untuk setiap fungsi kontinu terbatas f, berlaku $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$.

Misalkan f adalah fungsi kontinu terbatas dengan $|f(x)| \leq M$ untuk semua x. Untuk setiap $\varepsilon > 0$, karena f kontinu terbatas pada $\mathbb{R}$, maka f kontinu seragam pada setiap himpunan kompak. Oleh karena itu, untuk setiap $\delta > 0$, terdapat $\eta > 0$ sehingga $|x - y| < \eta$ mengimplikasikan $|f(x) - f(y)| < \delta$.

Perhatikan bahwa

$$\begin{aligned} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| &= |\mathbb{E}[f(X_n) - f(X)]| \\ &\leq \mathbb{E}[|f(X_n) - f(X)|]. \end{aligned} \tag{2.3.17}$$

Dekomposisi ekspektasi berdasarkan kejadian $\{|X_n - X| \leq \eta\}$ dan $\{|X_n - X| > \eta\}$ memberikan

$$\begin{aligned} \mathbb{E}[|f(X_n) - f(X)|] &= \mathbb{E}[|f(X_n) - f(X)| \cdot \mathbf{1}_{\{|X_n - X| \leq \eta\}}] \\ &\quad + \mathbb{E}[|f(X_n) - f(X)| \cdot \mathbf{1}_{\{|X_n - X| > \eta\}}]. \end{aligned} \tag{2.3.18}$$

Untuk suku pertama, karena $|X_n - X| \leq \eta$ mengimplikasikan $|f(X_n) - f(X)| < \delta$, maka

$$\mathbb{E}[|f(X_n) - f(X)| \cdot \mathbf{1}_{\{|X_n - X| \leq \eta\}}] < \delta. \tag{2.3.19}$$

Untuk suku kedua, karena $|f(x)| \leq M$, maka $|f(X_n) - f(X)| \leq 2M$, sehingga

$$\mathbb{E}[|f(X_n) - f(X)| \cdot \mathbf{1}_{\{|X_n - X| > \eta\}}] \leq 2M \cdot \Pr(|X_n - X| > \eta). \tag{2.3.20}$$

Sebab $X_n \xrightarrow{p} X$, maka $\Pr(|X_n - X| > \eta) \to 0$ ketika $n \to \infty$. Oleh karena itu,

$$\limsup_{n \to \infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq \delta. \tag{2.3.21}$$

Sebab $\delta > 0$ sembarang, maka $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$, yang berarti $X_n \xrightarrow{d} X$. ■

Teorema 2.3.15 *Jika $X_n \xrightarrow{d} c$ dengan c adalah konstanta, maka $X_n \xrightarrow{p} c$.*

Bukti. Andaikan $X_n \xrightarrow{d} c$. Berdasarkan definisi konvergensi dalam distribusi, untuk setiap fungsi kontinu terbatas f, berlaku $\mathbb{E}[f(X_n)] \to f(c)$.

Untuk setiap $\varepsilon > 0$, definisikan fungsi kontinu terbatas

$$f_\varepsilon(x) = \begin{cases} 1, & \text{jika } |x-c| \geq \varepsilon, \\ \frac{|x-c|}{\varepsilon}, & \text{jika } |x-c| < \varepsilon. \end{cases} \tag{2.3.22}$$

Fungsi f_ε kontinu dan memenuhi $0 \leq f_\varepsilon(x) \leq 1$ untuk semua x, serta $f_\varepsilon(c) = 0$.

Perhatikan bahwa $\mathbf{1}_{\{|X_n - c| \geq \varepsilon\}} \leq f_\varepsilon(X_n)$, sehingga

$$\Pr(|X_n - c| \geq \varepsilon) = \mathbb{E}[\mathbf{1}_{\{|X_n-c|\geq\varepsilon\}}] \leq \mathbb{E}[f_\varepsilon(X_n)]. \tag{2.3.23}$$

Sebab $X_n \xrightarrow{d} c$ dan f_ε kontinu terbatas, maka $\mathbb{E}[f_\varepsilon(X_n)] \to f_\varepsilon(c) = 0$. Oleh karena itu,

$$\lim_{n\to\infty} \Pr(|X_n - c| \geq \varepsilon) = 0, \tag{2.3.24}$$

yang berarti $X_n \xrightarrow{p} c$. ■

Contoh 2.3.16 Misalkan $\Omega = [0,1]$ dengan ukuran probabilitas Lebesgue. Definisikan barisan variabel acak sebagai berikut: untuk setiap $n \geq 1$, tulis $n = 2^k + j$ dengan $0 \leq j < 2^k$, dan definisikan

$$X_n(\omega) = \mathbf{1}_{[j/2^k,(j+1)/2^k]}(\omega).$$

Barisan ini disebut barisan *typewriter*. Untuk setiap $\varepsilon \in (0,1)$,

$$\Pr(|X_n| > \varepsilon) = \Pr(X_n = 1) = \frac{1}{2^k} \to 0$$

ketika $n \to \infty$ (sebab $k \to \infty$). Jadi, $X_n \xrightarrow{p} 0$.

Namun, untuk setiap $\omega \in [0,1]$, terdapat tak hingga banyak n sehingga $X_n(\omega) = 1$ (karena setiap titik ω akan tercakup oleh interval $[j/2^k, (j+1)/2^k]$ untuk tak hingga banyak nilai k). Oleh karena itu, $\limsup_{n\to\infty} X_n(\omega) = 1 \neq 0$ untuk semua ω, sehingga X_n tidak konvergen hampir pasti ke 0.

Contoh 2.3.17 Misalkan $X \sim \mathcal{N}(0,1)$ dan definisikan $X_n = -X$ untuk semua n. Maka $X_n \sim \mathcal{N}(0,1) = X$ untuk semua n, sehingga $X_n \xrightarrow{d} X$ secara trivial.

Namun, untuk $\varepsilon = 1$,

$$\Pr(|X_n - X| > 1) = \Pr(|-X - X| > 1) = \Pr(|2X| > 1) = \Pr(|X| > 0.5) \approx 0.617 \neq 0.$$

Oleh karena itu, X_n tidak konvergen dalam probabilitas ke X.

2.4 Teori Probabilitas Asimtotik

Teori probabilitas asimtotik mempelajari perilaku limit dari barisan variabel random dan estimator ketika ukuran sampel menuju tak hingga. Dalam statistika, hasil-hasil asimtotik digunakan sebagai pendekatan terhadap distribusi *sampling* yang umumnya sulit diperoleh secara eksak. Pembahasan dalam bagian ini mengikuti kerangka statistik asimtotik sebagaimana dikembangkan oleh **?** dengan fondasi probabilitas dari **?**.

2.4.1 Hukum Bilangan Besar

Teorema 2.4.1 (***Weak Law of Large Numbers***, (**?**)) *Misalkan $\{X_i\}_{i=1}^{\infty}$ adalah barisan variabel acak i.i.d. dengan $\mathbb{E}[X_1] = \mu$ dan $\mathbb{E}[|X_1|] < \infty$. Maka*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu. \tag{2.4.1}$$

Bukti. Andaikan $\{X_i\}_{i=1}^{\infty}$ adalah barisan variabel acak i.i.d. dengan $\mathbb{E}[X_1] = \mu$ dan $\operatorname{Var}(X_1) = \sigma^2 < \infty$. Definisikan rata-rata sampel $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Akan ditunjukkan bahwa $\bar{X}_n \xrightarrow{p} \mu$.

Pertama, perhatikan bahwa

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\cdot n\mu = \mu. \tag{2.4.2}$$

Selanjutnya, karena X_i saling independen, berlaku

$$\mathrm{Var}(\bar{X}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}. \tag{2.4.3}$$

Dengan menggunakan ketaksamaan Chebyshev, untuk setiap $\varepsilon > 0$ berlaku

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\mathrm{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}. \tag{2.4.4}$$

Dengan mengambil limit $n \to \infty$, diperoleh

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon^2} = 0. \tag{2.4.5}$$

Oleh karena probabilitas selalu non-negatif, berlaku

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0, \tag{2.4.6}$$

yang ekuivalen dengan $\bar{X}_n \xrightarrow{p} \mu$. Dengan demikian, *Weak Law of Large Numbers* terbukti. ■

Hukum Bilangan Besar memberikan dasar probabilistik bagi konsistensi estimator berbasis rata-rata sampel.

2.4.2 Konsistensi Estimator

Definisi 2.4.2 (Konsistensi, (?)) *Suatu estimator $\hat{\theta}_n$ untuk parameter θ dikatakan konsisten apabila*

$$\hat{\theta}_n \xrightarrow{p} \theta. \tag{2.4.7}$$

Konsistensi sering diperoleh sebagai konsekuensi langsung dari Hukum Bilangan

Besar, khususnya untuk estimator yang dapat diekspresikan sebagai fungsi rata-rata sampel.

2.4.3 Fungsi Karakteristik

Fungsi karakteristik merupakan alat fundamental dalam teori probabilitas yang memungkinkan karakterisasi lengkap distribusi suatu variabel acak. Berbeda dengan fungsi pembangkit momen yang mungkin tidak terdefinisi, fungsi karakteristik selalu ada untuk setiap variabel acak.

Definisi 2.4.3 (**Fungsi Karakteristik**, (**?**)) *Misalkan X adalah variabel acak. Fungsi karakteristik dari X didefinisikan sebagai*

$$\phi_X(t) = \mathbb{E}\left[e^{itX}\right] = \int_{-\infty}^{\infty} e^{itx}\, dF_X(x), \quad t \in \mathbb{R}, \tag{2.4.8}$$

dengan $i = \sqrt{-1}$ adalah unit imajiner dan F_X adalah fungsi distribusi kumulatif dari X.

Contoh 2.4.4 Berikut adalah fungsi karakteristik untuk beberapa distribusi umum.

(a) **Distribusi Normal.** Jika $X \sim \mathcal{N}(\mu, \sigma^2)$, maka

$$\phi_X(t) = \exp\left(i\mu t - \frac{\sigma^2 t^2}{2}\right). \tag{2.4.9}$$

Khususnya, untuk distribusi normal standar $X \sim \mathcal{N}(0, 1)$, berlaku $\phi_X(t) = e^{-t^2/2}$.

(b) **Distribusi Poisson.** Jika $X \sim \text{Poisson}(\lambda)$, maka

$$\phi_X(t) = \exp\left(\lambda(e^{it} - 1)\right). \tag{2.4.10}$$

(c) **Distribusi Eksponensial.** Jika $X \sim \text{Exp}(\lambda)$, maka

$$\phi_X(t) = \frac{\lambda}{\lambda - it}. \tag{2.4.11}$$

Proposisi 2.4.5 (**Sifat-Sifat Fungsi Karakteristik**, (**?**)) *Fungsi karakteristik $\phi_X(t)$ memenuhi sifat-sifat berikut.*

(a) $\phi_X(0) = 1$.

(b) $|\phi_X(t)| \leq 1$ *untuk semua* $t \in \mathbb{R}$.

(c) ϕ_X *kontinu seragam pada* $\mathbb{R}$.

(d) $\overline{\phi_X(t)} = \phi_X(-t)$, *dengan* $\overline{z}$ *menyatakan konjugat kompleks dari* z.

(e) Jika $Y = aX + b$ *untuk konstanta* $a, b \in \mathbb{R}$, *maka* $\phi_Y(t) = e^{ibt}\phi_X(at)$.

(f) Jika X *dan* Y *independen, maka* $\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t)$.

Proposisi 2.4.6 (**Hubungan dengan Momen**, (**?**)) *Jika* $\mathbb{E}[|X|^n] < \infty$, *maka fungsi karakteristik* ϕ_X *terdiferensiasi* n *kali dan*

$$\phi_X^{(k)}(0) = i^k\mathbb{E}[X^k], \quad k = 1, 2, \ldots, n. \tag{2.4.12}$$

Khususnya, jika $\mathbb{E}[X^2] < \infty$, *maka ekspansi Taylor di sekitar* $t = 0$ *memberikan*

$$\phi_X(t) = 1 + it\mathbb{E}[X] - \frac{t^2}{2}\mathbb{E}[X^2] + o(t^2). \tag{2.4.13}$$

Teorema berikut merupakan hasil fundamental yang menghubungkan konvergensi fungsi karakteristik dengan konvergensi dalam distribusi.

Teorema 2.4.7 (**Teorema Kontinuitas Lévy**, (**?**)) *Misalkan* $\{X_n\}_{n=1}^{\infty}$ *adalah barisan variabel acak dengan fungsi karakteristik* ϕ_{X_n}, *dan misalkan* X *adalah variabel acak dengan fungsi karakteristik* ϕ_X. *Maka berlaku dua pernyataan berikut.*

(a) Jika $X_n \Rightarrow X$, *maka* $\phi_{X_n}(t) \to \phi_X(t)$ *untuk setiap* $t \in \mathbb{R}$.

(b) Sebaliknya, jika $\phi_{X_n}(t) \to \phi(t)$ *untuk setiap* $t \in \mathbb{R}$ *dan* ϕ *kontinu di* $t = 0$, *maka* ϕ *adalah fungsi karakteristik dari suatu variabel acak* X *dan* $X_n \Rightarrow X$.

Contoh 2.4.8 Teorema Lévy dapat digunakan untuk membuktikan konvergensi distribusi. Misalkan $X_n \sim \mathcal{N}(0, 1+1/n)$ untuk $n = 1, 2, \ldots$. Fungsi karakteristik dari X_n adalah

$$\phi_{X_n}(t) = \exp\left(-\frac{(1+1/n)t^2}{2}\right). \tag{2.4.14}$$

Untuk setiap $t \in \mathbb{R}$, berlaku

$$\lim_{n\to\infty} \phi_{X_n}(t) = \exp\left(-\frac{t^2}{2}\right) = \phi_X(t), \tag{2.4.15}$$

dengan $\phi_X(t)$ adalah fungsi karakteristik dari $X \sim \mathcal{N}(0, 1)$. Sebab limit ini kontinu di $t = 0$, berdasarkan Teorema Lévy, diperoleh $X_n \Rightarrow X \sim \mathcal{N}(0, 1)$.

Teorema Lévy sangat penting dalam pembuktian Teorema Limit Pusat, di mana konvergensi dalam distribusi dibuktikan melalui konvergensi fungsi karakteristik ke fungsi karakteristik distribusi normal.

2.4.4 Teorema Pemetaan Kontinu dan Teorema Slutsky

Teorema pemetaan kontinu (*Continuous Mapping Theorem*) dan teorema Slutsky merupakan alat fundamental dalam statistika asimtotik yang memungkinkan manipulasi limit variabel acak melalui fungsi kontinu dan operasi aritmetika.

Teorema 2.4.9 **(Teorema Pemetaan Kontinu (*Continuous Mapping Theorem*), (?))** *Misalkan $\{X_n\}$ adalah barisan variabel acak dan $g : \mathbb{R}^k \to \mathbb{R}^m$ adalah fungsi yang kontinu hampir di mana-mana relatif terhadap distribusi limit. Maka berlaku implikasi berikut.*

(a) Jika $X_n \xrightarrow{d} X$, maka $g(X_n) \xrightarrow{d} g(X)$.

(b) Jika $X_n \xrightarrow{p} X$, maka $g(X_n) \xrightarrow{p} g(X)$.

(c) Jika $X_n \xrightarrow{a.s.} X$, maka $g(X_n) \xrightarrow{a.s.} g(X)$.

Bukti. Akan dibuktikan bagian (a) untuk kasus konvergensi dalam distribusi. Misalkan $X_n \xrightarrow{d} X$ dan g kontinu pada himpunan C dengan $\Pr(X \in C) = 1$. Akan ditunjukkan bahwa $g(X_n) \xrightarrow{d} g(X)$.

Berdasarkan teorema Portmanteau, $X_n \xrightarrow{d} X$ ekuivalen dengan $\mathbb{E}[h(X_n)] \to \mathbb{E}[h(X)]$ untuk setiap fungsi kontinu terbatas h.

Ambil sembarang fungsi kontinu terbatas $h : \mathbb{R}^m \to \mathbb{R}$. Definisikan $f = h \circ g : \mathbb{R}^k \to \mathbb{R}$. Sebab g kontinu pada C dan h kontinu pada $\mathbb{R}^m$, maka f kontinu pada C.

Untuk setiap $\varepsilon > 0$, himpunan diskontinuitas D_f dari f memenuhi $D_f \subseteq D_g$ (himpunan diskontinuitas g). Sebab $\Pr(X \in D_g) = 0$, maka $\Pr(X \in D_f) = 0$.

Dengan teorema Portmanteau yang diperluas, sebab f kontinu hampir di mana-mana relatif terhadap distribusi X, berlaku

$$\mathbb{E}[h(g(X_n))] = \mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)] = \mathbb{E}[h(g(X))]. \tag{2.4.16}$$

Sebab h adalah fungsi kontinu terbatas sembarang, berdasarkan teorema Portmanteau, diperoleh $g(X_n) \xrightarrow{d} g(X)$.

Untuk bagian (b), jika $X_n \xrightarrow{p} X$, maka untuk setiap $\varepsilon > 0$ dan $\delta > 0$, terdapat N sehingga $\Pr(|X_n - X| > \delta) < \varepsilon$ untuk $n \geq N$. Sebab g kontinu pada C, maka g kontinu seragam pada setiap himpunan kompak. Dengan demikian, untuk setiap $\varepsilon' > 0$, terdapat $\delta > 0$ sehingga $|x - y| < \delta$ mengimplikasikan $|g(x) - g(y)| < \varepsilon'$ pada himpunan kompak yang memuat support distribusi. Oleh karena itu, $\Pr(|g(X_n) - g(X)| > \varepsilon') \to 0$.

Bagian (c) mengikuti langsung dari sifat fungsi kontinu: jika $X_n(\omega) \to X(\omega)$ untuk hampir semua ω, dan g kontinu di $X(\omega)$, maka $g(X_n(\omega)) \to g(X(\omega))$.
■

Contoh 2.4.10 Misalkan $X_n \xrightarrow{d} X$ dengan $X \sim \mathcal{N}(0, 1)$. Dengan teorema pemetaan kontinu dan fungsi $g(x) = x^2$ yang kontinu, diperoleh

$$X_n^2 \xrightarrow{d} X^2 \sim \chi^2(1). \tag{2.4.17}$$

Demikian pula, jika $g(x) = e^x$, maka $e^{X_n} \xrightarrow{d} e^X$.

Teorema 2.4.11 (**Teorema Slutsky**, (**?**)) *Misalkan* $\{X_n\}$ *dan* $\{Y_n\}$ *adalah barisan variabel acak. Jika* $X_n \xrightarrow{d} X$ *dan* $Y_n \xrightarrow{p} c$ *dengan* c *adalah konstanta, maka berlaku:*

(a) $X_n + Y_n \xrightarrow{d} X + c$,

(b) $X_n \cdot Y_n \xrightarrow{d} c \cdot X$, *dan*

(c) $X_n/Y_n \xrightarrow{d} X/c$ *jika* $c \neq 0$.

Bukti. Akan dibuktikan ketiga bagian secara berurutan.

Bagian (a): Akan ditunjukkan bahwa $(X_n, Y_n) \xrightarrow{d} (X, c)$, kemudian menerapkan teorema pemetaan kontinu dengan $g(x, y) = x + y$.

Untuk menunjukkan $(X_n, Y_n) \xrightarrow{d} (X, c)$, digunakan fungsi karakteristik. Fungsi karakteristik bersama dari (X_n, Y_n) adalah

$$\phi_{(X_n,Y_n)}(t_1, t_2) = \mathbb{E}[e^{i(t_1 X_n + t_2 Y_n)}]. \tag{2.4.18}$$

Perhatikan bahwa

$$\begin{aligned} \left|\phi_{(X_n,Y_n)}(t_1, t_2) - \phi_X(t_1)e^{it_2 c}\right| &= \left|\mathbb{E}[e^{it_1 X_n}(e^{it_2 Y_n} - e^{it_2 c})]\right| \\ &\leq \mathbb{E}\left[|e^{it_2 Y_n} - e^{it_2 c}|\right]. \end{aligned} \tag{2.4.19}$$

Sebab $Y_n \xrightarrow{p} c$, maka $e^{it_2 Y_n} \xrightarrow{p} e^{it_2 c}$ (berdasarkan teorema pemetaan kontinu untuk konvergensi dalam probabilitas). Sebab $|e^{it_2 Y_n} - e^{it_2 c}| \leq 2$ terbatas, berdasarkan teorema konvergensi terdominasi, $\mathbb{E}[|e^{it_2 Y_n} - e^{it_2 c}|] \to 0$.

Selanjutnya, sebab $X_n \xrightarrow{d} X$, berdasarkan teorema kontinuitas Lévy, $\phi_{X_n}(t_1) \to \phi_X(t_1)$. Dengan demikian,

$$\phi_{(X_n,Y_n)}(t_1, t_2) \to \phi_X(t_1)e^{it_2 c} = \phi_{(X,c)}(t_1, t_2). \tag{2.4.20}$$

Berdasarkan teorema Lévy, $(X_n, Y_n) \xrightarrow{d} (X, c)$.

Dengan teorema pemetaan kontinu dan $g(x, y) = x + y$ yang kontinu, diperoleh $X_n + Y_n \xrightarrow{d} X + c$.

Bagian (b): Dengan hasil $(X_n, Y_n) \xrightarrow{d} (X, c)$ dan fungsi $g(x, y) = xy$ yang kontinu, teorema pemetaan kontinu memberikan $X_n \cdot Y_n \xrightarrow{d} X \cdot c = cX$.

Bagian (c): Jika $c \neq 0$, fungsi $g(x, y) = x/y$ kontinu pada $\{(x, y) : y \neq 0\}$. Sebab $Y_n \xrightarrow{p} c \neq 0$, maka $\Pr(Y_n = 0) \to 0$, sehingga (X, c) terkonsentrasi pada himpunan di mana g kontinu. Dengan teorema pemetaan kontinu, $X_n/Y_n \xrightarrow{d} X/c$. ■

Contoh 2.4.12 Misalkan $\bar{X}_n$ adalah rata-rata sampel dari variabel acak i.i.d. dengan rataan μ dan variansi σ^2, dan S_n^2 adalah variansi sampel. Berdasarkan Teorema Limit Pusat:

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2). \tag{2.4.21}$$

Berdasarkan Hukum Bilangan Besar, $S_n^2 \xrightarrow{p} \sigma^2$, sehingga $S_n \xrightarrow{p} \sigma$.

Dengan teorema Slutsky:

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \cdot \frac{\sigma}{S_n} \xrightarrow{d} \mathcal{N}(0, 1) \cdot 1 = \mathcal{N}(0, 1). \tag{2.4.22}$$

Hasil ini membenarkan penggunaan statistik t dalam inferensi asimtotik.

Teorema 2.4.13 (**Teorema Slutsky untuk Vektor dan Matriks**, (**?**)) *Misalkan $\{\boldsymbol{X}_n\}$ adalah barisan vektor acak berdimensi p dan $\{\boldsymbol{A}_n\}$ adalah barisan matriks acak berukuran $q \times p$. Jika $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$ dan $\boldsymbol{A}_n \xrightarrow{p} \boldsymbol{A}$ dengan $\boldsymbol{A}$ adalah matriks konstanta, maka:*

(a) $\boldsymbol{A}_n\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{A}\boldsymbol{X}$,

(b) $\boldsymbol{X}_n + \boldsymbol{b}_n \xrightarrow{d} \boldsymbol{X} + \boldsymbol{b}$ *jika* $\boldsymbol{b}_n \xrightarrow{p} \boldsymbol{b}$,

(c) $\boldsymbol{X}_n^\top \boldsymbol{A}_n \boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}^\top \boldsymbol{A}\boldsymbol{X}$ *untuk bentuk kuadratik, dan*

(d) jika $\boldsymbol{A}$ adalah matriks invertibel, maka $\boldsymbol{A}_n^{-1} \xrightarrow{p} \boldsymbol{A}^{-1}$ dan $\boldsymbol{A}_n^{-1}\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{A}^{-1}\boldsymbol{X}$.

Bukti. Bagian (a): Perkalian matriks $g(\boldsymbol{A}, \boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ adalah fungsi kontinu dari elemen-elemen $\boldsymbol{A}$ dan $\boldsymbol{x}$. Akan ditunjukkan bahwa $(\boldsymbol{A}_n, \boldsymbol{X}_n) \xrightarrow{d} (\boldsymbol{A}, \boldsymbol{X})$.

Untuk setiap fungsi kontinu terbatas $h : \mathbb{R}^{qp+p} \to \mathbb{R}$, perlu ditunjukkan bahwa $\mathbb{E}[h(\boldsymbol{A}_n, \boldsymbol{X}_n)] \to \mathbb{E}[h(\boldsymbol{A}, \boldsymbol{X})]$.

Sebab $\boldsymbol{A}_n \xrightarrow{p} \boldsymbol{A}$, untuk setiap $\varepsilon > 0$ berlaku $\Pr(\|\boldsymbol{A}_n - \boldsymbol{A}\| > \varepsilon) \to 0$. Definisikan

$$\mathbb{E}[h(\boldsymbol{A}_n, \boldsymbol{X}_n)] = \mathbb{E}[h(\boldsymbol{A}_n, \boldsymbol{X}_n)\mathbf{1}_{\{\|\boldsymbol{A}_n - \boldsymbol{A}\| \le \varepsilon\}}] + \mathbb{E}[h(\boldsymbol{A}_n, \boldsymbol{X}_n)\mathbf{1}_{\{\|\boldsymbol{A}_n - \boldsymbol{A}\| > \varepsilon\}}]. \tag{2.4.23}$$

Suku kedua dibatasi oleh $\|h\|_\infty \cdot \Pr(\|\boldsymbol{A}_n - \boldsymbol{A}\| > \varepsilon) \to 0$. Untuk suku pertama, pada kejadian $\{\|\boldsymbol{A}_n - \boldsymbol{A}\| \le \varepsilon\}$, fungsi $h(\boldsymbol{A}_n, \boldsymbol{X}_n)$ mendekati $h(\boldsymbol{A}, \boldsymbol{X}_n)$ untuk ε kecil. Sebab $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$, maka $\mathbb{E}[h(\boldsymbol{A}, \boldsymbol{X}_n)] \to \mathbb{E}[h(\boldsymbol{A}, \boldsymbol{X})]$.

Dengan demikian, $(\boldsymbol{A}_n, \boldsymbol{X}_n) \xrightarrow{d} (\boldsymbol{A}, \boldsymbol{X})$. Dengan teorema pemetaan kontinu dan fungsi $g(\boldsymbol{A}, \boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, diperoleh $\boldsymbol{A}_n\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{A}\boldsymbol{X}$.

Bagian (b): Analog dengan bagian (a), penjumlahan vektor $g(\boldsymbol{x}, \boldsymbol{b}) = \boldsymbol{x} + \boldsymbol{b}$ adalah fungsi kontinu.

Bagian (c): Bentuk kuadratik $g(\boldsymbol{A}, \boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ adalah fungsi kontinu dari elemen-elemen $\boldsymbol{A}$ dan $\boldsymbol{x}$. Dengan hasil dari bagian (a), $(\boldsymbol{A}_n, \boldsymbol{X}_n) \xrightarrow{d} (\boldsymbol{A}, \boldsymbol{X})$, sehingga teorema pemetaan kontinu memberikan $\boldsymbol{X}_n^\top \boldsymbol{A}_n \boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}^\top \boldsymbol{A}\boldsymbol{X}$.

Bagian (d): Jika $\boldsymbol{A}$ invertibel, fungsi invers matriks $g(\boldsymbol{A}) = \boldsymbol{A}^{-1}$ kontinu pada himpunan matriks invertibel. Dengan demikian, berdasarkan teorema pemetaan kontinu, $\boldsymbol{A}_n^{-1} \xrightarrow{p} \boldsymbol{A}^{-1}$. Selanjutnya, dengan bagian (a), diperoleh $\boldsymbol{A}_n^{-1} \xrightarrow{d} \boldsymbol{A}^{-1}\boldsymbol{X}$. ■

Contoh 2.4.14 Misalkan $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma)$ dan $\hat{\Sigma}_n \xrightarrow{p} \boldsymbol{\Sigma}$ dengan $\boldsymbol{\Sigma}$ definit positif. Maka statistik Wald

$$W_n = n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})^\top \hat{\boldsymbol{\Sigma}}_n^{-1} (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \tag{2.4.24}$$

memenuhi $W_n \xrightarrow{d} \chi^2(p)$, dengan $p = \dim(\boldsymbol{\beta})$.

Hal ini karena $\hat{\boldsymbol{\Sigma}}_n^{-1} \xrightarrow{p} \boldsymbol{\Sigma}^{-1}$ (berdasarkan teorema pemetaan kontinu untuk fungsi invers matriks), dan

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})^\top \hat{\boldsymbol{\Sigma}}_n^{-1} \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} \boldsymbol{Z}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{Z} \sim \chi^2(p), \tag{2.4.25}$$

dengan $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$.

Teorema 2.4.15 (**Teorema Cramér–Wold**, (**?**)) *Barisan vektor acak $\boldsymbol{X}_n \in \mathbb{R}^p$ konvergen dalam distribusi ke $\boldsymbol{X}$ jika dan hanya jika untuk setiap vektor $\boldsymbol{t} \in \mathbb{R}^p$ berlaku*

$$\boldsymbol{t}^\top \boldsymbol{X}_n \xrightarrow{d} \boldsymbol{t}^\top \boldsymbol{X}. \tag{2.4.26}$$

Bukti. ($\Rightarrow$) Jika $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$, maka untuk setiap $\boldsymbol{t} \in \mathbb{R}^p$, fungsi $g(\boldsymbol{x}) = \boldsymbol{t}^\top \boldsymbol{x}$ adalah linear (kontinu). Berdasarkan teorema pemetaan kontinu, $\boldsymbol{t}^\top \boldsymbol{X}_n \xrightarrow{d} \boldsymbol{t}^\top \boldsymbol{X}$.

($\Leftarrow$) Andaikan untuk setiap $\boldsymbol{t} \in \mathbb{R}^p$ berlaku $\boldsymbol{t}^\top \boldsymbol{X}_n \xrightarrow{d} \boldsymbol{t}^\top \boldsymbol{X}$. Akan ditunjukkan $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$ melalui konvergensi fungsi karakteristik.

Fungsi karakteristik dari $\boldsymbol{X}_n$ adalah

$$\phi_{\boldsymbol{X}_n}(\boldsymbol{t}) = \mathbb{E}[e^{i\boldsymbol{t}^\top \boldsymbol{X}_n}]. \tag{2.4.27}$$

Perhatikan bahwa $e^{i\boldsymbol{t}^\top \boldsymbol{X}_n}$ adalah fungsi karakteristik dari variabel acak skalar $\boldsymbol{t}^\top \boldsymbol{X}_n$ dievaluasi di titik 1. Sebab $\boldsymbol{t}^\top \boldsymbol{X}_n \xrightarrow{d} \boldsymbol{t}^\top \boldsymbol{X}$, berdasarkan teorema kontinuitas Lévy, fungsi karakteristik dari $\boldsymbol{t}^\top \boldsymbol{X}_n$ konvergen ke fungsi karakteristik dari $\boldsymbol{t}^\top \boldsymbol{X}$ di setiap titik. Khususnya,

$$\phi_{\boldsymbol{X}_n}(\boldsymbol{t}) = \mathbb{E}[e^{i \cdot 1 \cdot (\boldsymbol{t}^\top \boldsymbol{X}_n)}] \to \mathbb{E}[e^{i \cdot 1 \cdot (\boldsymbol{t}^\top \boldsymbol{X})}] = \phi_{\boldsymbol{X}}(\boldsymbol{t}). \tag{2.4.28}$$

Sebab $\phi_{\boldsymbol{X}_n}(\boldsymbol{t}) \to \phi_{\boldsymbol{X}}(\boldsymbol{t})$ untuk setiap $\boldsymbol{t} \in \mathbb{R}^p$, dan $\phi_{\boldsymbol{X}}$ kontinu di $\boldsymbol{t} = \boldsymbol{0}$ (sebab merupakan fungsi karakteristik), berdasarkan teorema kontinuitas Lévy multivariat, $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$. ■

Contoh 2.4.16 Teorema Cramér–Wold sangat berguna untuk membuktikan Teorema Limit Pusat multivariat. Jika $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ adalah vektor acak i.i.d. dengan rataan

$\boldsymbol{\mu}$ dan matriks kovarians $\boldsymbol{\Sigma}$, maka untuk setiap $\boldsymbol{t} \in \mathbb{R}^p$:

$$\boldsymbol{t}^\top \left(\sqrt{n}(\bar{\boldsymbol{X}}_n - \boldsymbol{\mu})\right) = \sqrt{n}(\boldsymbol{t}^\top \bar{\boldsymbol{X}}_n - \boldsymbol{t}^\top \boldsymbol{\mu}) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{t}^\top \boldsymbol{\Sigma} \boldsymbol{t}), \tag{2.4.29}$$

berdasarkan TLP univariat. Dengan teorema Cramér–Wold, $\sqrt{n}(\bar{\boldsymbol{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$.

2.4.5 Notasi Big-$\mathcal{O}$ dan Little-o (Deterministik)

Notasi asimtotik Big-$\mathcal{O}$ dan Little-o merupakan alat fundamental dalam analisis matematika untuk menyatakan laju pertumbuhan dan tingkat ketakterabaian suatu besaran deterministik.

Definisi 2.4.17 (**Notasi Big-$\mathcal{O}$**) *Misalkan $\{a_n\}$ dan $\{b_n\}$ adalah barisan bilangan real dengan $b_n > 0$ untuk semua n cukup besar. Barisan $\{a_n\}$ dikatakan* berorde paling banyak b_n*, ditulis*

$$a_n = O(b_n), \tag{2.4.30}$$

jika terdapat konstanta $C > 0$ dan bilangan bulat $n_0 \in \mathbb{N}$ sehingga

$$|a_n| \le C b_n, \quad \forall n \ge n_0. \tag{2.4.31}$$

Secara ekuivalen, $a_n = O(b_n)$ jika dan hanya jika barisan $\{a_n/b_n\}$ terbatas untuk n cukup besar.

Contoh 2.4.18 Berikut adalah beberapa contoh penggunaan notasi Big-$\mathcal{O}$.

(a) **Polinomial:** Jika $a_n = 3n^2 + 5n + 7$, maka $a_n = O(n^2)$. Untuk $n \ge 1$, berlaku

$$|a_n| = 3n^2 + 5n + 7 \le 3n^2 + 5n^2 + 7n^2 = 15n^2,$$

sehingga dapat dipilih $C = 15$ dan $n_0 = 1$.

(b) **Logaritma:** Jika $a_n = \log n$, maka $a_n = O(n^\varepsilon)$ untuk setiap $\varepsilon > 0$. Hal ini karena $\lim_{n\to\infty} \frac{\log n}{n^\varepsilon} = 0$, sehingga $\log n \le C n^\varepsilon$ untuk n cukup besar.

(c) **Eksponensial:** Jika $a_n = 2^n$, maka $a_n = O(3^n)$ karena $2^n \le 3^n$ untuk semua

$n \geq 0$.

(d) **Rata-rata sampel:** Jika $\bar{X}_n$ adalah rata-rata sampel dari variabel acak dengan variansi σ^2, maka galat standar $\sigma/\sqrt{n} = O(n^{-1/2})$.

Definisi 2.4.19 (**Notasi Little-**o) *Misalkan* $\{a_n\}$ *dan* $\{b_n\}$ *adalah barisan bilangan real dengan* $b_n > 0$ *untuk semua* n *cukup besar. Barisan* $\{a_n\}$ *dikatakan* berorde lebih kecil dari b_n, *ditulis*

$$a_n = o(b_n), \tag{2.4.32}$$

jika

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 0. \tag{2.4.33}$$

Secara ekuivalen, $a_n = o(b_n)$ *jika dan hanya jika untuk setiap* $\varepsilon > 0$*, terdapat* $n_0 \in \mathbb{N}$ *sehingga* $|a_n| < \varepsilon b_n$ *untuk semua* $n \geq n_0$*.*

Contoh 2.4.20 Berikut adalah beberapa contoh penggunaan notasi Little-o.

(a) **Polinomial:** Jika $a_n = n$, maka $a_n = o(n^2)$ karena

$$\lim_{n\to\infty} \frac{n}{n^2} = \lim_{n\to\infty} \frac{1}{n} = 0.$$

(b) **Logaritma vs polinomial:** Jika $a_n = \log n$, maka $a_n = o(n^\varepsilon)$ untuk setiap $\varepsilon > 0$. Dengan aturan L'Hôpital:

$$\lim_{n\to\infty} \frac{\log n}{n^\varepsilon} = \lim_{n\to\infty} \frac{1/n}{\varepsilon n^{\varepsilon-1}} = \lim_{n\to\infty} \frac{1}{\varepsilon n^\varepsilon} = 0.$$

(c) **Ekspansi Taylor:** Untuk fungsi $\sin x$ di sekitar $x = 0$:

$$\sin x = x - \frac{x^3}{6} + o(x^3) \quad \text{ketika } x \to 0.$$

Suku $o(x^3)$ menunjukkan bahwa galat sisa tumbuh lebih lambat dari x^3.

(d) **Perhitungan numerik:** Misalkan $a_n = 1/n^2$ dan $b_n = 1/n$. Maka $a_n =$

$o(b_n)$ karena

$$\lim_{n\to\infty} \frac{1/n^2}{1/n} = \lim_{n\to\infty} \frac{1}{n} = 0.$$

Secara numerik, untuk $n = 100$: $a_{100} = 0.0001$ dan $b_{100} = 0.01$, sehingga $a_{100}/b_{100} = 0.01$.

Teorema 2.4.21 (Hubungan antara O dan o) *Untuk barisan $\{a_n\}$ dan $\{b_n\}$ dengan $b_n > 0$, berlaku:*

(a) Jika $a_n = o(b_n)$, maka $a_n = O(b_n)$, tetapi tidak sebaliknya.

(b) $a_n = o(b_n)$ jika dan hanya jika $a_n = O(b_n)$ dan $\limsup_{n\to\infty} |a_n/b_n| = 0$.

Bukti.

(a) Andaikan $a_n = o(b_n)$. Maka $\lim_{n\to\infty} a_n/b_n = 0$. Dengan definisi limit, untuk $\varepsilon = 1$, terdapat n_0 sehingga $|a_n/b_n| < 1$ untuk $n \geq n_0$. Dengan demikian, $|a_n| < b_n$ untuk $n \geq n_0$, sehingga $a_n = O(b_n)$ dengan $C = 1$.

Sebaliknya tidak berlaku: $a_n = 1$ dan $b_n = 1$ memberikan $a_n = O(b_n)$ (dengan $C = 1$), tetapi $\lim_{n\to\infty} a_n/b_n = 1 \neq 0$, sehingga $a_n \neq o(b_n)$.

(b) Langsung dari definisi: $a_n = o(b_n)$ ekuivalen dengan $\lim_{n\to\infty} |a_n/b_n| = 0$, yang mengimplikasikan barisan $\{|a_n/b_n|\}$ terbatas (sehingga $a_n = O(b_n)$) dan $\limsup_{n\to\infty} |a_n/b_n| = 0$.

■

Proposisi 2.4.22 (Sifat-Sifat Notasi Asimtotik) *Untuk barisan $\{a_n\}$, $\{b_n\}$, $\{c_n\}$ dengan suku positif yang sesuai, berlaku:*

*(a) **Transitivitas:** Jika $a_n = O(b_n)$ dan $b_n = O(c_n)$, maka $a_n = O(c_n)$.*

*(b) **Penjumlahan:** $O(a_n) + O(b_n) = O(\max\{a_n, b_n\})$.*

*(c) **Perkalian:** $O(a_n) \cdot O(b_n) = O(a_n b_n)$.*

*(d) **Penyerapan:** $o(a_n) + O(a_n) = O(a_n)$ dan $o(a_n) \cdot O(b_n) = o(a_n b_n)$.*

2.4.6 Notasi Big-$\mathcal{O}$ dan Little-o dalam Probabilitas

Notasi $\mathcal{O}_p$ dan o_p merupakan perluasan dari notasi deterministik $O(\cdot)$ dan $o(\cdot)$ ke dalam konteks stokastik. Notasi ini sangat penting dalam perumusan representasi asimtotik estimator dan digunakan secara luas dalam statistika asimtotik (**?**).

Definisi 2.4.23 (**Notasi $\mathcal{O}_p$ (Terbatas dalam Probabilitas)**, (**?**)) *Barisan variabel acak $\{X_n\}$ dikatakan* terbatas dalam probabilitas *dengan orde a_n, ditulis*

$$X_n = \mathcal{O}_p(a_n), \tag{2.4.34}$$

apabila untuk setiap $\varepsilon > 0$ terdapat $M > 0$ dan $N \in \mathbb{N}$ sehingga

$$\Pr(|X_n| > Ma_n) < \varepsilon, \quad \forall n \geq N. \tag{2.4.35}$$

Secara ekuivalen, $X_n = \mathcal{O}_p(a_n)$ jika dan hanya jika barisan $\{X_n/a_n\}$ terbatas dalam probabilitas, yaitu

$$\sup_{n \geq N} \Pr\left(\left|\frac{X_n}{a_n}\right| > M\right) < \varepsilon. \tag{2.4.36}$$

Contoh 2.4.24 Misalkan $X_1, X_2, \ldots$ adalah barisan variabel acak i.i.d. dengan $\mathbb{E}[X_i] = \mu$ dan $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Definisikan rata-rata sampel $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Akan ditunjukkan bahwa $\bar{X}_n - \mu = \mathcal{O}_p(n^{-1/2})$.

Berdasarkan Teorema Limit Pusat, berlaku

$$\sqrt{n}(\bar{X}_n - \mu) \Rightarrow \mathcal{N}(0, \sigma^2).$$

Sebab $\sqrt{n}(\bar{X}_n - \mu)$ konvergen dalam distribusi ke variabel acak normal, maka barisan ini terbatas dalam probabilitas (*bounded in probability*). Dengan demikian, untuk setiap $\varepsilon > 0$, terdapat $M > 0$ sehingga

$$\sup_n \Pr\left(|\sqrt{n}(\bar{X}_n - \mu)| > M\right) < \varepsilon,$$

yang ekuivalen dengan

$$\sup_n \Pr\left(|\bar{X}_n - \mu| > M \cdot n^{-1/2}\right) < \varepsilon.$$

Oleh karena itu, $\bar{X}_n - \mu = \mathcal{O}_p(n^{-1/2})$.

Contoh 2.4.25 (**Perhitungan Numerik**) Misalkan $X_n \sim \mathcal{N}(0, 1/n)$ untuk $n = 1, 2, \ldots$. Akan ditunjukkan bahwa $X_n = \mathcal{O}_p(n^{-1/2})$.

Perhatikan bahwa $\sqrt{n}X_n \sim \mathcal{N}(0, 1)$. Untuk $\varepsilon = 0.05$ dan $M = 1.96$, berlaku

$$\Pr(|\sqrt{n}X_n| > 1.96) = 2(1 - \Phi(1.96)) \approx 0.05.$$

Dengan demikian,

$$\Pr(|X_n| > 1.96 \cdot n^{-1/2}) = 0.05 < \varepsilon,$$

sehingga $X_n = \mathcal{O}_p(n^{-1/2})$ dengan $M = 1.96$.

Definisi 2.4.26 (**Notasi o_p (Konvergen ke Nol dalam Probabilitas)**, (**?**)) *Barisan variabel acak $\{X_n\}$ dikatakan* konvergen ke nol dalam probabilitas *dengan orde a_n, ditulis*

$$X_n = o_p(a_n), \tag{2.4.37}$$

apabila

$$\frac{X_n}{a_n} \xrightarrow{p} 0, \tag{2.4.38}$$

yaitu untuk setiap $\varepsilon > 0$ dan $\delta > 0$, terdapat $N \in \mathbb{N}$ sehingga

$$\Pr\left(\left|\frac{X_n}{a_n}\right| > \varepsilon\right) < \delta, \quad \forall n \geq N. \tag{2.4.39}$$

Contoh 2.4.27 Misalkan $X_1, X_2, \ldots$ adalah barisan variabel acak i.i.d. dengan $\mathbb{E}[X_i] = \mu$ dan $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Definisikan rata-rata sampel $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Akan ditunjukkan bahwa $\bar{X}_n - \mu = o_p(1)$.

Berdasarkan Hukum Bilangan Besar, berlaku

$$\bar{X}_n \xrightarrow{p} \mu,$$

yang berarti $\bar{X}_n - \mu \xrightarrow{p} 0$. Dengan definisi notasi o_p, hal ini ekuivalen dengan

$$\frac{\bar{X}_n - \mu}{1} \xrightarrow{p} 0,$$

sehingga $\bar{X}_n - \mu = o_p(1)$.

Contoh 2.4.28 (**Perhitungan Numerik**) Misalkan $X_n \sim \mathcal{N}(0, 1/n^2)$ untuk $n = 1, 2, \ldots$. Akan ditunjukkan bahwa $X_n = o_p(n^{-1/2})$.

Perhatikan bahwa

$$\frac{X_n}{n^{-1/2}} = n^{1/2} X_n \sim \mathcal{N}\left(0, \frac{n}{n^2}\right) = \mathcal{N}\left(0, \frac{1}{n}\right).$$

Untuk setiap $\varepsilon > 0$, dengan ketaksamaan Chebyshev diperoleh

$$\Pr\left(\left|n^{1/2} X_n\right| > \varepsilon\right) \le \frac{\mathrm{Var}(n^{1/2} X_n)}{\varepsilon^2} = \frac{1/n}{\varepsilon^2} = \frac{1}{n\varepsilon^2} \to 0$$

ketika $n \to \infty$. Dengan demikian, $n^{1/2} X_n \xrightarrow{p} 0$, sehingga $X_n = o_p(n^{-1/2})$.

Sebagai contoh numerik, untuk $n = 100$ dan $\varepsilon = 0.1$:

$$\Pr(|n^{1/2} X_n| > 0.1) \le \frac{1}{100 \cdot 0.01} = 1.$$

Batas ini tidak informatif, tetapi dengan menggunakan sifat distribusi normal:

$$\Pr(|n^{1/2} X_n| > 0.1) = 2\left(1 - \Phi\left(\frac{0.1}{\sqrt{1/100}}\right)\right) = 2(1 - \Phi(1)) \approx 0.317.$$

Untuk $n = 10000$:

$$\Pr(|n^{1/2} X_n| > 0.1) = 2\left(1 - \Phi\left(\frac{0.1}{\sqrt{1/10000}}\right)\right) = 2(1 - \Phi(10)) \approx 0.$$

Teorema 2.4.29 (Sifat-Sifat Notasi $\mathcal{O}_p$ dan o_p, (?)) *Misalkan $\{X_n\}$ dan $\{Y_n\}$ adalah barisan variabel acak, serta $\{a_n\}$ dan $\{b_n\}$ adalah barisan bilangan positif. Berlaku sifat-sifat berikut.*

(a) ***Penjumlahan:*** $\mathcal{O}_p(a_n) + \mathcal{O}_p(b_n) = \mathcal{O}_p(\max\{a_n, b_n\})$.

(b) ***Perkalian:*** $\mathcal{O}_p(a_n) \cdot \mathcal{O}_p(b_n) = \mathcal{O}_p(a_n b_n)$.

(c) ***Relasi o_p dan $\mathcal{O}_p$:*** $o_p(a_n) = \mathcal{O}_p(a_n)$, *tetapi tidak sebaliknya.*

(d) ***Penyerapan:*** $o_p(a_n) + \mathcal{O}_p(a_n) = \mathcal{O}_p(a_n)$.

(e) ***Perkalian dengan o_p:*** $o_p(a_n) \cdot \mathcal{O}_p(b_n) = o_p(a_n b_n)$.

(f) ***Konvergensi dalam probabilitas:*** $X_n \xrightarrow{p} c$ *jika dan hanya jika* $X_n = c + o_p(1)$.

(g) ***Konvergensi dalam distribusi:*** *Jika* $X_n \Rightarrow X$ *dengan* X *variabel acak, maka* $X_n = \mathcal{O}_p(1)$.

Bukti. Berikut adalah bukti untuk beberapa sifat utama.

(a) **Penjumlahan:** Misalkan $X_n = \mathcal{O}_p(a_n)$ dan $Y_n = \mathcal{O}_p(b_n)$. Untuk setiap $\varepsilon > 0$, terdapat $M_1, M_2 > 0$ sehingga

$$\Pr(|X_n| > M_1 a_n) < \varepsilon/2, \quad \Pr(|Y_n| > M_2 b_n) < \varepsilon/2.$$

Dengan ketaksamaan segitiga, $|X_n + Y_n| \leq |X_n| + |Y_n|$. Definisikan $c_n = \max\{a_n, b_n\}$ dan $M = M_1 + M_2$. Maka

$$\begin{aligned} \Pr(|X_n + Y_n| > Mc_n) &\leq \Pr(|X_n| + |Y_n| > (M_1 + M_2)c_n) \\ &\leq \Pr(|X_n| > M_1 a_n) + \Pr(|Y_n| > M_2 b_n) \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned} \tag{2.4.40}$$

(b) **Perkalian:** Misalkan $X_n = \mathcal{O}_p(a_n)$ dan $Y_n = \mathcal{O}_p(b_n)$. Untuk setiap $\varepsilon > 0$, terdapat $M_1, M_2 > 0$ sehingga

$$\Pr(|X_n| > M_1 a_n) < \varepsilon/2, \quad \Pr(|Y_n| > M_2 b_n) < \varepsilon/2.$$

Perhatikan bahwa

$$\begin{aligned}\Pr(|X_n Y_n| > M_1 M_2 a_n b_n) &\leq \Pr(|X_n| > M_1 a_n) + \Pr(|Y_n| > M_2 b_n) \\ &< \varepsilon. \end{aligned} \tag{2.4.41}$$

(c) **Relasi o_p dan $\mathcal{O}_p$:** Jika $X_n = o_p(a_n)$, maka $X_n/a_n \xrightarrow{p} 0$. Untuk setiap $\varepsilon > 0$ dan $M > 0$, terdapat N sehingga untuk $n \geq N$:

$$\Pr(|X_n/a_n| > M) < \varepsilon,$$

yang berarti $X_n = \mathcal{O}_p(a_n)$. Sebaliknya tidak berlaku: $X_n \sim \mathcal{N}(0,1)$ memenuhi $X_n = \mathcal{O}_p(1)$ tetapi bukan $o_p(1)$.

(f) **Konvergensi dalam probabilitas:** Jika $X_n \xrightarrow{p} c$, maka $X_n - c \xrightarrow{p} 0$, yang ekuivalen dengan $(X_n - c)/1 \xrightarrow{p} 0$, sehingga $X_n - c = o_p(1)$, atau $X_n = c + o_p(1)$. Sebaliknya juga berlaku.

(g) **Konvergensi dalam distribusi:** Jika $X_n \Rightarrow X$, maka untuk setiap $\varepsilon > 0$, terdapat M sehingga $\Pr(|X| > M) < \varepsilon/2$. Untuk n cukup besar, $\Pr(|X_n| > M) < \varepsilon$ karena konvergensi dalam distribusi. Dengan demikian, $X_n = \mathcal{O}_p(1)$.

■

Teorema 2.4.30 (Dekomposisi $\mathcal{O}_p$ Berdasarkan Momen, (?)) *Misalkan $\{X_n\}$ adalah barisan variabel acak. Berlaku pernyataan-pernyataan berikut.*

(a) ***Dekomposisi ekspektasi:*** *Jika $\mathbb{E}[|X_n|] = \mathcal{O}(a_n)$ (dalam arti deterministik), maka $X_n = \mathcal{O}_p(a_n)$.*

(b) ***Dekomposisi variansi:*** *Jika $\mathbb{E}[X_n] = \mu_n$ dan $\mathrm{Var}(X_n) = \mathcal{O}(a_n^2)$, maka $X_n - \mu_n = \mathcal{O}_p(a_n)$.*

(c) ***Dekomposisi momen kedua:*** *Jika* $\mathbb{E}[X_n^2] = \mathcal{O}(a_n^2)$*, maka* $X_n = \mathcal{O}_p(a_n)$*.*

Bukti. Berikut adalah bukti untuk ketiga pernyataan.

(a) **Dekomposisi ekspektasi:** Andaikan $\mathbb{E}[|X_n|] \leq Ca_n$ untuk suatu konstanta $C > 0$ dan semua n cukup besar. Dengan ketaksamaan Markov, untuk setiap $M > 0$:

$$\Pr(|X_n| > Ma_n) \leq \frac{\mathbb{E}[|X_n|]}{Ma_n} \leq \frac{Ca_n}{Ma_n} = \frac{C}{M}. \tag{2.4.42}$$

Untuk setiap $\varepsilon > 0$, pilih $M = C/\varepsilon$, maka $\Pr(|X_n| > Ma_n) < \varepsilon$. Dengan demikian, $X_n = \mathcal{O}_p(a_n)$.

(b) **Dekomposisi variansi:** Andaikan $\mathrm{Var}(X_n) \leq Ca_n^2$ untuk suatu konstanta $C > 0$. Dengan ketaksamaan Chebyshev, untuk setiap $M > 0$:

$$\Pr(|X_n - \mu_n| > Ma_n) \leq \frac{\mathrm{Var}(X_n)}{M^2 a_n^2} \leq \frac{Ca_n^2}{M^2 a_n^2} = \frac{C}{M^2}. \tag{2.4.43}$$

Untuk setiap $\varepsilon > 0$, pilih $M = \sqrt{C/\varepsilon}$, maka $\Pr(|X_n - \mu_n| > Ma_n) < \varepsilon$. Dengan demikian, $X_n - \mu_n = \mathcal{O}_p(a_n)$.

(c) **Dekomposisi momen kedua:** Andaikan $\mathbb{E}[X_n^2] \leq Ca_n^2$ untuk suatu konstanta $C > 0$. Dengan ketaksamaan Markov pada X_n^2:

$$\Pr(|X_n| > Ma_n) = \Pr(X_n^2 > M^2 a_n^2) \leq \frac{\mathbb{E}[X_n^2]}{M^2 a_n^2} \leq \frac{Ca_n^2}{M^2 a_n^2} = \frac{C}{M^2}. \tag{2.4.44}$$

Untuk setiap $\varepsilon > 0$, pilih $M = \sqrt{C/\varepsilon}$, maka $\Pr(|X_n| > Ma_n) < \varepsilon$. Dengan demikian, $X_n = \mathcal{O}_p(a_n)$.

■

Contoh 2.4.31 (**Aplikasi Dekomposisi Momen**) Misalkan $X_1, X_2, \ldots$ adalah barisan variabel acak i.i.d. dengan $\mathbb{E}[X_i] = \mu$ dan $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Definisikan rata-rata sampel $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Menggunakan dekomposisi variansi:

$$\mathbb{E}[\bar{X}_n] = \mu, \tag{2.4.45}$$

$$\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n} = \mathcal{O}(n^{-1}) = \mathcal{O}\left((n^{-1/2})^2\right). \tag{2.4.46}$$

Berdasarkan Teorema Dekomposisi Momen bagian (b), diperoleh

$$\bar{X}_n - \mu = \mathcal{O}_p(n^{-1/2}).$$

Perhitungan numerik: Untuk $n = 100$, $\mu = 0$, dan $\sigma^2 = 1$:

$$\mathrm{Var}(\bar{X}_{100}) = \frac{1}{100} = 0.01, \tag{2.4.47}$$

$$\Pr(|\bar{X}_{100}| > 0.2) \leq \frac{0.01}{0.04} = 0.25. \tag{2.4.48}$$

Dengan distribusi eksak $\bar{X}_{100} \sim \mathcal{N}(0, 0.01)$:

$$\Pr(|\bar{X}_{100}| > 0.2) = 2(1 - \Phi(2)) \approx 0.0455.$$

Ini konsisten dengan $\bar{X}_n - \mu = \mathcal{O}_p(n^{-1/2})$ karena $0.2 = 2 \cdot 0.1 = 2 \cdot n^{-1/2}$.

Teorema 2.4.32 (Dekomposisi o_p Berdasarkan Momen, (?)) *Misalkan $\{X_n\}$ adalah barisan variabel acak. Berlaku pernyataan-pernyataan berikut.*

(a) ***Dekomposisi ekspektasi:*** *Jika $\mathbb{E}[|X_n|] = o(a_n)$ (dalam arti deterministik), maka $X_n = o_p(a_n)$.*

(b) ***Dekomposisi variansi:*** *Jika $\mathbb{E}[X_n] = o(a_n)$ dan $\mathrm{Var}(X_n) = o(a_n^2)$, maka $X_n = o_p(a_n)$.*

(c) ***Dekomposisi momen kedua:*** *Jika $\mathbb{E}[X_n^2] = o(a_n^2)$, maka $X_n = o_p(a_n)$.*

Bukti. Berikut adalah bukti untuk ketiga pernyataan.

(a) **Dekomposisi ekspektasi:** Andaikan $\mathbb{E}[|X_n|]/a_n \to 0$ ketika $n \to \infty$. Dengan ketaksamaan Markov, untuk setiap $\varepsilon > 0$:

$$\Pr\left(\left|\frac{X_n}{a_n}\right| > \varepsilon\right) \leq \frac{\mathbb{E}[|X_n|/a_n]}{\varepsilon} = \frac{\mathbb{E}[|X_n|]}{\varepsilon a_n} \to 0. \tag{2.4.49}$$

Dengan demikian, $X_n/a_n \xrightarrow{p} 0$, sehingga $X_n = o_p(a_n)$.

(b) **Dekomposisi variansi:** Andaikan $\mathbb{E}[X_n] = o(a_n)$ dan $\mathrm{Var}(X_n) = o(a_n^2)$. Maka $\mathbb{E}[X_n]/a_n \to 0$ dan $\mathrm{Var}(X_n)/a_n^2 \to 0$. Dengan ketaksamaan Chebyshev:

$$\Pr\left(\left|\frac{X_n - \mathbb{E}[X_n]}{a_n}\right| > \varepsilon\right) \leq \frac{\mathrm{Var}(X_n)}{\varepsilon^2 a_n^2} \to 0. \tag{2.4.50}$$

Sehingga $(X_n - \mathbb{E}[X_n])/a_n \xrightarrow{p} 0$. Sebab $\mathbb{E}[X_n]/a_n \to 0$, maka

$$\frac{X_n}{a_n} = \frac{X_n - \mathbb{E}[X_n]}{a_n} + \frac{\mathbb{E}[X_n]}{a_n} \xrightarrow{p} 0.$$

Dengan demikian, $X_n = o_p(a_n)$.

(c) **Dekomposisi momen kedua:** Andaikan $\mathbb{E}[X_n^2]/a_n^2 \to 0$. Dengan ketaksamaan Markov:

$$\Pr\left(\left|\frac{X_n}{a_n}\right| > \varepsilon\right) = \Pr\left(\frac{X_n^2}{a_n^2} > \varepsilon^2\right) \leq \frac{\mathbb{E}[X_n^2]}{\varepsilon^2 a_n^2} \to 0. \tag{2.4.51}$$

Dengan demikian, $X_n/a_n \xrightarrow{p} 0$, sehingga $X_n = o_p(a_n)$.

■

Contoh 2.4.33 (**Aplikasi Dekomposisi** o_p) Misalkan $X_n \sim \mathcal{N}(0, 1/n^3)$ untuk $n = 1, 2, \ldots$. Akan ditunjukkan bahwa $X_n = o_p(n^{-1})$.

Menggunakan dekomposisi momen kedua:

$$\mathbb{E}[X_n^2] = \mathrm{Var}(X_n) = \frac{1}{n^3}. \tag{2.4.52}$$

Perhatikan bahwa

$$\frac{\mathbb{E}[X_n^2]}{(n^{-1})^2} = \frac{1/n^3}{1/n^2} = \frac{1}{n} \to 0. \tag{2.4.53}$$

Berdasarkan Teorema Dekomposisi o_p bagian (c), diperoleh $X_n = o_p(n^{-1})$.

Verifikasi langsung:

$$\frac{X_n}{n^{-1}} = nX_n \sim \mathcal{N}\left(0, \frac{n^2}{n^3}\right) = \mathcal{N}\left(0, \frac{1}{n}\right). \tag{2.4.54}$$

Untuk setiap $\varepsilon > 0$:

$$\Pr(|nX_n| > \varepsilon) = 2\left(1 - \Phi\left(\frac{\varepsilon}{\sqrt{1/n}}\right)\right) = 2\left(1 - \Phi(\varepsilon\sqrt{n})\right) \to 0 \tag{2.4.55}$$

ketika $n \to \infty$. Dengan demikian, $nX_n \xrightarrow{p} 0$, sehingga $X_n = o_p(n^{-1})$.

2.4.7 Teorema Limit Pusat Klasik

Teorema 2.4.34 (Teorema Limit Pusat, (?)) *Misalkan $\{X_i\}_{i=1}^{\infty}$ adalah barisan variabel acak i.i.d. dengan $\mathbb{E}[X_1] = \mu$ dan $\mathrm{Var}(X_1) = \sigma^2 < \infty$. Maka*

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \Rightarrow \mathcal{N}(0, \sigma^2). \tag{2.4.56}$$

Bukti. Pembuktian dilakukan dengan menggunakan fungsi karakteristik. Definisikan $Y_i = X_i - \mu$ sehingga $\mathbb{E}[Y_i] = 0$ dan $\mathrm{Var}(Y_i) = \sigma^2$. Selanjutnya, definisikan

$$S_n = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n}(X_i - \mu). \tag{2.4.57}$$

Akan ditunjukkan bahwa $S_n/\sqrt{n\sigma^2} \Rightarrow \mathcal{N}(0, 1)$.

Fungsi karakteristik dari Y_i adalah $\phi_Y(t) = \mathbb{E}[e^{itY_i}]$. Dengan ekspansi Taylor di sekitar $t = 0$, diperoleh

$$\phi_Y(t) = 1 + it\mathbb{E}[Y_i] + \frac{(it)^2}{2}\mathbb{E}[Y_i^2] + o(t^2) = 1 - \frac{t^2\sigma^2}{2} + o(t^2). \tag{2.4.58}$$

Sebab $Y_1, Y_2, \ldots, Y_n$ saling independen dan berdistribusi identik, fungsi karakteristik dari $S_n/\sqrt{n\sigma^2}$ adalah

$$\begin{aligned}
\phi_{S_n/\sqrt{n\sigma^2}}(t) &= \mathbb{E}\left[\exp\left(\frac{itS_n}{\sqrt{n\sigma^2}}\right)\right] \\
&= \mathbb{E}\left[\exp\left(\sum_{i=1}^{n}\frac{itY_i}{\sqrt{n\sigma^2}}\right)\right] \\
&= \prod_{i=1}^{n}\mathbb{E}\left[\exp\left(\frac{itY_i}{\sqrt{n\sigma^2}}\right)\right] \\
&= \left[\phi_Y\left(\frac{t}{\sqrt{n\sigma^2}}\right)\right]^n.
\end{aligned} \tag{2.4.59}$$

Dengan substitusi ekspansi Taylor, diperoleh

$$\begin{aligned}
\phi_{S_n/\sqrt{n\sigma^2}}(t) &= \left[1 - \frac{1}{2}\cdot\frac{t^2}{n\sigma^2}\cdot\sigma^2 + o\left(\frac{1}{n}\right)\right]^n \\
&= \left[1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right]^n.
\end{aligned} \tag{2.4.60}$$

Dengan mengambil limit $n \to \infty$ dan menggunakan fakta bahwa $\lim_{n\to\infty}\left(1+\frac{a}{n}\right)^n = e^a$, diperoleh

$$\lim_{n\to\infty}\phi_{S_n/\sqrt{n\sigma^2}}(t) = \exp\left(-\frac{t^2}{2}\right), \tag{2.4.61}$$

yang merupakan fungsi karakteristik dari distribusi $\mathcal{N}(0,1)$.

Berdasarkan teorema kontinuitas Lévy, konvergensi fungsi karakteristik mengimplikasikan konvergensi dalam distribusi. Oleh karena itu,

$$\frac{S_n}{\sqrt{n\sigma^2}} = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) \Rightarrow \mathcal{N}(0,1). \tag{2.4.62}$$

Dengan mengalikan kedua ruas dengan σ, diperoleh

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}X_i - \mu\right) \Rightarrow \mathcal{N}(0,\sigma^2). \tag{2.4.63}$$

■

Teorema Limit Pusat merupakan hasil fundamental yang memungkinkan pendekatan distribusi normal terhadap rata-rata sampel, dan menjadi dasar utama inferensi statistik asimtotik.

Dalam banyak aplikasi statistik, khususnya estimator berbobot dan estimator lokal, observasi tidak lagi identik terdistribusi. Hal ini memotivasi penggunaan CLT untuk *triangular arrays*. *Triangular arrays* adalah kumpulan variabel acak yang disusun dalam bentuk segitiga, dengan setiap baris dapat memiliki distribusi yang berbeda.

Teorema 2.4.35 (**CLT untuk *Triangular Arrays***, (**?**)) *Misalkan* $\{X_{n,i} : i = 1, \dots, n\}$ *adalah triangular array variabel acak dengan* $\mathbb{E}[X_{n,i}] = 0$ *dan varians total*

$$\sum_{i=1}^{n} \mathrm{Var}(X_{n,i}) \to \sigma^2. \tag{2.4.64}$$

Jika kondisi Lindeberg terpenuhi, maka

$$\sum_{i=1}^{n} X_{n,i} \Rightarrow \mathcal{N}(0, \sigma^2). \tag{2.4.65}$$

Hasil ini sangat relevan untuk analisis asimtotik estimator lokal, di mana kontribusi masing-masing observasi bergantung pada n.

Teorema 2.4.36 (**Teorema Lindeberg–Lévy**, (**?**)) *Misalkan* $\{X_i\}_{i=1}^{\infty}$ *adalah barisan variabel acak i.i.d. dengan* $\mathbb{E}[X_i] = \mu$ *dan* $\mathrm{Var}(X_i) = \sigma^2 < \infty$. *Maka*

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{2.4.66}$$

Bukti. Teorema Lindeberg–Lévy merupakan kasus khusus dari Teorema Limit Pusat klasik yang telah dibuktikan sebelumnya. Definisikan $Y_i = X_i - \mu$ sehingga $\mathbb{E}[Y_i] = 0$ dan $\mathrm{Var}(Y_i) = \sigma^2$. Dengan menggunakan fungsi karakteristik, ekspansi Taylor dari $\phi_Y(t)$ di sekitar $t = 0$ memberikan

$$\phi_Y(t) = 1 - \frac{t^2\sigma^2}{2} + o(t^2). \tag{2.4.67}$$

Fungsi karakteristik dari $\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n} Y_i$ adalah

$$\left[\phi_Y\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n = \left[1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right]^n \to e^{-t^2/2} \tag{2.4.68}$$

ketika $n \to \infty$. Berdasarkan teorema kontinuitas Lévy, konvergensi ini mengimplikasikan konvergensi dalam distribusi ke $\mathcal{N}(0,1)$. ■

Teorema 2.4.37 (**Teorema Lindeberg–Feller**, (**?**)) *Misalkan* $\{X_{n,i} : i = 1, \dots, n\}$ *adalah* triangular array *variabel acak independen (tidak harus identik) dengan* $\mathbb{E}[X_{n,i}] = 0$ *untuk semua* n *dan* i*. Definisikan*

$$s_n^2 = \sum_{i=1}^{n} \mathrm{Var}(X_{n,i}) = \sum_{i=1}^{n} \sigma_{n,i}^2. \tag{2.4.69}$$

Jika kondisi Lindeberg terpenuhi, yaitu untuk setiap $\varepsilon > 0$*,*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} \mathbb{E}\left[X_{n,i}^2 \cdot \mathbf{1}_{\{|X_{n,i}| > \varepsilon s_n\}}\right] = 0, \tag{2.4.70}$$

maka

$$\frac{1}{s_n} \sum_{i=1}^{n} X_{n,i} \xrightarrow{d} \mathcal{N}(0,1). \tag{2.4.71}$$

Bukti. Pembuktian dilakukan dengan menggunakan fungsi karakteristik. Definisikan $S_n = \sum_{i=1}^{n} X_{n,i}$ dan $Z_n = S_n/s_n$. Fungsi karakteristik dari Z_n adalah

$$\phi_{Z_n}(t) = \prod_{i=1}^{n} \phi_{X_{n,i}}\left(\frac{t}{s_n}\right). \tag{2.4.72}$$

Untuk setiap i, dengan ekspansi Taylor fungsi karakteristik diperoleh

$$\phi_{X_{n,i}}\left(\frac{t}{s_n}\right) = 1 + i\frac{t}{s_n}\mathbb{E}[X_{n,i}] - \frac{t^2}{2s_n^2}\mathbb{E}[X_{n,i}^2] + R_{n,i}(t), \tag{2.4.73}$$

dengan $\mathbb{E}[X_{n,i}] = 0$ dan sisa $R_{n,i}(t)$ memenuhi $|R_{n,i}(t)| \le \frac{|t|^3}{6s_n^3}\mathbb{E}[|X_{n,i}|^3]$.

Sehingga,

$$\phi_{X_{n,i}}\left(\frac{t}{s_n}\right) = 1 - \frac{t^2\sigma_{n,i}^2}{2s_n^2} + R_{n,i}(t). \tag{2.4.74}$$

Kondisi Lindeberg menjamin bahwa kontribusi suku-suku dengan $|X_{n,i}|$ besar dapat diabaikan. Secara lebih rinci, kondisi ini memastikan bahwa

$$\max_{1\le i\le n} \frac{\sigma_{n,i}^2}{s_n^2} \to 0 \tag{2.4.75}$$

ketika $n \to \infty$, yang berarti tidak ada variabel tunggal yang mendominasi variansi total.

Dengan menggunakan fakta bahwa $\log(1+x) \approx x$ untuk $|x|$ kecil dan

$$\log \phi_{Z_n}(t) = \sum_{i=1}^{n} \log \phi_{X_{n,i}}\left(\frac{t}{s_n}\right), \tag{2.4.76}$$

diperoleh

$$\begin{aligned}
\log \phi_{Z_n}(t) &= \sum_{i=1}^{n} \log\left(1 - \frac{t^2\sigma_{n,i}^2}{2s_n^2} + R_{n,i}(t)\right) \\
&\approx \sum_{i=1}^{n}\left(-\frac{t^2\sigma_{n,i}^2}{2s_n^2}\right) + \sum_{i=1}^{n} R_{n,i}(t) \\
&= -\frac{t^2}{2s_n^2}\sum_{i=1}^{n}\sigma_{n,i}^2 + \sum_{i=1}^{n} R_{n,i}(t) \\
&= -\frac{t^2}{2} + \sum_{i=1}^{n} R_{n,i}(t).
\end{aligned} \tag{2.4.77}$$

Kondisi Lindeberg menjamin bahwa $\sum_{i=1}^{n} R_{n,i}(t) \to 0$ ketika $n \to \infty$. Oleh karena itu,

$$\lim_{n\to\infty} \phi_{Z_n}(t) = e^{-t^2/2}, \tag{2.4.78}$$

yang merupakan fungsi karakteristik dari $\mathcal{N}(0,1)$. Berdasarkan teorema kontinuitas Lévy, $Z_n \xrightarrow{d} \mathcal{N}(0,1)$. ■

Contoh 2.4.38 Misalkan $X_{n,i} = Y_i/\sqrt{n}$ untuk $i = 1,\ldots,n$, dengan Y_i adalah variabel acak i.i.d. dengan $\mathbb{E}[Y_i] = 0$ dan $\mathrm{Var}(Y_i) = \sigma^2 < \infty$. Maka $\mathbb{E}[X_{n,i}] = 0$ dan $\mathrm{Var}(X_{n,i}) = \sigma^2/n$, sehingga $s_n^2 = \sigma^2$.

Kondisi Lindeberg dapat diperiksa sebagai berikut. Untuk setiap $\varepsilon > 0$,

$$\begin{aligned} \frac{1}{s_n^2} \sum_{i=1}^{n} \mathbb{E}\left[X_{n,i}^2 \cdot \mathbf{1}_{\{|X_{n,i}| > \varepsilon s_n\}}\right] &= \frac{1}{\sigma^2} \sum_{i=1}^{n} \mathbb{E}\left[\frac{Y_i^2}{n} \cdot \mathbf{1}_{\{|Y_i| > \varepsilon\sigma\sqrt{n}\}}\right] \\ &= \frac{1}{\sigma^2} \mathbb{E}\left[Y_1^2 \cdot \mathbf{1}_{\{|Y_1| > \varepsilon\sigma\sqrt{n}\}}\right]. \end{aligned} \tag{2.4.79}$$

Sebab $\mathbb{E}[Y_1^2] = \sigma^2 < \infty$, maka $\mathbb{E}[Y_1^2 \cdot \mathbf{1}_{\{|Y_1| > \varepsilon\sigma\sqrt{n}\}}] \to 0$ ketika $n \to \infty$ berdasarkan teorema konvergensi terdominasi. Dengan demikian, kondisi Lindeberg terpenuhi dan Teorema Lindeberg–Feller berlaku.

2.4.8 Metode Delta

Dalam banyak aplikasi statistik, sering kali diperlukan distribusi asimtotik dari suatu fungsi transformasi estimator, bukan hanya distribusi estimator itu sendiri. Sebagai contoh, jika $\hat{\theta}_n$ adalah estimator untuk parameter θ dan diperlukan inferensi terhadap $g(\theta)$ untuk suatu fungsi g, maka perlu diketahui distribusi asimtotik dari $g(\hat{\theta}_n)$. Metode Delta menyediakan alat yang elegan untuk menurunkan distribusi limit dari transformasi halus suatu estimator yang asimtotik normal. Metode ini memanfaatkan ekspansi Taylor orde pertama untuk mengaproksimasi fungsi nonlinear secara lokal sebagai fungsi linear, sehingga sifat normalitas asimtotik dapat dipertahankan melalui transformasi.

Teorema 2.4.39 (**Metode** Δ, (**?**)) *Misalkan*

$$\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow \mathcal{N}(0, \Sigma), \tag{2.4.80}$$

dan g adalah fungsi terdiferensialkan di θ. Maka

$$\sqrt{n}\big(g(\hat{\theta}_n) - g(\theta)\big) \Rightarrow \mathcal{N}\big(0, g'(\theta)\Sigma g'(\theta)^\top\big). \tag{2.4.81}$$

Bukti. Dengan menggunakan ekspansi Taylor orde pertama dari g di sekitar θ, diperoleh

$$g(\hat{\theta}_n) = g(\theta) + g'(\theta)(\hat{\theta}_n - \theta) + o(|\hat{\theta}_n - \theta|). \tag{2.4.82}$$

Sebab $\hat{\theta}_n \xrightarrow{p} \theta$, suku sisa $o(|\hat{\theta}_n - \theta|)$ memenuhi

$$\frac{o(|\hat{\theta}_n - \theta|)}{|\hat{\theta}_n - \theta|} \xrightarrow{p} 0. \tag{2.4.83}$$

Dengan mengalikan kedua ruas ekspansi Taylor dengan $\sqrt{n}$, diperoleh

$$\sqrt{n}\big(g(\hat{\theta}_n) - g(\theta)\big) = g'(\theta) \cdot \sqrt{n}(\hat{\theta}_n - \theta) + \sqrt{n} \cdot o(|\hat{\theta}_n - \theta|). \tag{2.4.84}$$

Perhatikan bahwa $\sqrt{n} \cdot o(|\hat{\theta}_n - \theta|) = o_p(1)$ karena $\sqrt{n}(\hat{\theta}_n - \theta) = \mathcal{O}_p(1)$ berdasarkan konvergensi dalam distribusi ke normal. Dengan demikian,

$$\sqrt{n}\big(g(\hat{\theta}_n) - g(\theta)\big) = g'(\theta) \cdot \sqrt{n}(\hat{\theta}_n - \theta) + o_p(1). \tag{2.4.85}$$

Berdasarkan teorema Slutsky, sebab $\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow \mathcal{N}(0, \Sigma)$ dan $o_p(1) \xrightarrow{p} 0$, maka

$$\sqrt{n}\big(g(\hat{\theta}_n) - g(\theta)\big) \Rightarrow g'(\theta) \cdot \mathcal{N}(0, \Sigma) = \mathcal{N}\big(0, g'(\theta)\Sigma g'(\theta)^\top\big). \tag{2.4.86}$$

Kesamaan terakhir mengikuti dari sifat transformasi linear variabel acak normal: jika $Z \sim \mathcal{N}(0, \Sigma)$, maka $AZ \sim \mathcal{N}(0, A\Sigma A^\top)$ untuk matriks $A = g'(\theta)$. ■

Contoh 2.4.40 Misalkan $X_1, X_2, \ldots, X_n$ adalah sampel acak i.i.d. dari distribusi dengan rataan $\mu > 0$ dan variansi σ^2. Akan dicari distribusi asimtotik dari $\log(\bar{X}_n)$.

Berdasarkan Teorema Limit Pusat, $\sqrt{n}(\bar{X}_n - \mu) \Rightarrow \mathcal{N}(0, \sigma^2)$. Dengan $g(x) = \log(x)$ dan $g'(x) = 1/x$, metode Delta memberikan

$$\sqrt{n}\big(\log(\bar{X}_n) - \log(\mu)\big) \Rightarrow \mathcal{N}\left(0, \frac{\sigma^2}{\mu^2}\right). \tag{2.4.87}$$

Dengan demikian, $\log(\bar{X}_n)$ adalah asimtotik normal dengan rataan $\log(\mu)$ dan variansi asimtotik $\sigma^2/(n\mu^2)$.

2.5 Analisis Regresi Linear

Analisis regresi merupakan suatu teknik statistik untuk menginvestigasi dan memodelkan hubungan antarvariabel. Aplikasi dari analisis regresi sangat luas dan digunakan dalam hampir semua bidang, seperti teknik atau *engineering*, ilmu fisika dan kimia, ekonomi, manajemen, ilmu biologi, dan sosiologi. Dalam analisis regresi, variabel respons atau variabel dependen (dinotasikan dengan variabel acak Y) dimodelkan dengan fungsi dalam variabel penjelas atau variabel independen (dinotasikan dengan variabel acak X). (**?**)

2.5.1 Regresi Linear Biasa atau *Ordinary Least Squares* (OLS)

Analisis regresi linear memodelkan hubungan antarvariabel dengan fungsi linear. Apabila terdapat satu variabel prediktor maka model yang terbentuk disebut sebagai regresi linear sederhana, sedangkan apabila terdapat lebih dari satu variabel prediktor maka model yang terbentuk disebut sebagai regresi linear berganda.

Definisi 2.5.1 (**Regresi Linear Sederhana, (?)**) *Model regresi linear sederhana merupakan model dengan satu variabel penjelas X yang mempunyai hubungan garis lurus terhadap satu respons Y. Model ini dirumuskan dengan*

$$\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{x} + \boldsymbol{\varepsilon}, \tag{2.5.1}$$

dengan intersep β_0 dan koefisien β_1 merupakan konstanta yang tidak diketahui, serta ε adalah vektor galat acak. $\boldsymbol{y}$ dan $\boldsymbol{x}$ berturut-turut adalah vektor acak Y dan X. Parameter β_0 dan β_1 biasa disebut sebagai koefisien-koefisien regresi.

Definisi 2.5.2 (**Regresi Linear Berganda, (?)**) *Regresi linear berganda merupakan model regresi yang memiliki lebih dari satu variabel independen. Model ini merupakan generalisasi dari regresi linear sederhana. Vektor respons $\boldsymbol{y}$ mungkin memiliki hubungan dengan p-buah variabel penjelas, yaitu matriks acak $\boldsymbol{X}$ yang dibentuk atas p-buah vektor acak $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p$, sehingga*

$$\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{x}_1 + \beta_2 \boldsymbol{x}_2 + \cdots + \beta_p \boldsymbol{x}_p + \boldsymbol{\varepsilon}. \tag{2.5.2}$$

Model di atas dapat dituliskan dalam notasi matriks dan vektor, yaitu

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \tag{2.5.3}$$

atau

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{2.5.4}$$

Model OLS mengasumsikan bahwa galat memiliki rataan nol dan variansi konstan σ^2 yang tidak diketahui nilainya. Lebih lanjut, model ini juga mengasumsikan bahwa komponen galat tidak memiliki autokorelasi yang berarti bahwa nilai dari suatu galat tidak bergantung pada galat lainnya. Dengan kata lain $\varepsilon \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma^2)$, untuk suatu σ^2 yang bersifat konstan.

Model regresi linear memandang vektor acak $\boldsymbol{y}$ sebagai fungsi dari matriks acak $\boldsymbol{X}$. Oleh karena itu, rataan dari distribusi $\boldsymbol{y}$ dapat dituliskan sebagai

$$\mathbb{E}[\boldsymbol{y} \mid \boldsymbol{X}] = \mathbb{E}[\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}] = \boldsymbol{X}\boldsymbol{\beta} \tag{2.5.5}$$

dan variansinya

$$\operatorname{Var}(\boldsymbol{y} \mid \boldsymbol{X}) = \operatorname{Var}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n, \tag{2.5.6}$$

karena $\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$ dan $\operatorname{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n$. Oleh karena itu, rataan dari $\boldsymbol{y}$ adalah fungsi linear dari $\boldsymbol{X}$ meskipun variansinya tidak bergantung pada nilai $\boldsymbol{X}$.

Definisi 2.5.3 (**Asumsi Regresi Linear Klasik, (???)**) *Model regresi linear klasik didasarkan pada sejumlah asumsi yang dikenal sebagai asumsi Gauss–Markov. Asumsi-asumsi ini diperlukan agar penduga kuadrat terkecil memiliki sifat optimal sebagai* best linear unbiased estimator *(BLUE). Asumsi-asumsi tersebut adalah sebagai berikut.*

(a) Model bersifat linear dalam parameter, yaitu $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

(b) Nilai tengah galat adalah nol bersyarat pada $\boldsymbol{X}$*, yaitu* $\mathbb{E}[\boldsymbol{\varepsilon} \mid \boldsymbol{X}] = \mathbf{0}$.

(c) *Matriks desain* $\boldsymbol{X}$ *memiliki rank penuh sehingga* $\boldsymbol{X}^\top\boldsymbol{X}$ *nonsingular.*

(d) *Galat memiliki ragam konstan,* $\mathrm{Var}(\varepsilon_i \mid \boldsymbol{X}) = \sigma^2$ *untuk semua* i*.*

(e) *Komponen galat tidak berkorelasi,* $\mathrm{Cov}(\varepsilon_i, \varepsilon_j \mid \boldsymbol{X}) = 0$ *untuk* $i \neq j$*.*

(f) *Untuk keperluan inferensi, diasumsikan* $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_n)$.

Estimasi parameter dalam regresi linear dilakukan dengan metode kuadrat terkecil atau *ordinary least squares* (OLS). Prinsip dasarnya adalah mencari nilai $\boldsymbol{\beta}$ yang meminimumkan jumlah kuadrat galat

$$S(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}). \tag{2.5.7}$$

Perhatikan bahwa fungsi kuadrat terkecil dari persamaan di atas dapat dituliskan sebagai

$$S(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = \boldsymbol{y}^\top\boldsymbol{y} - 2\,\boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{y} + \boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}. \tag{2.5.8}$$

Dengan diferensial matriks, gradien terhadap $\boldsymbol{\beta}$ adalah

$$\nabla_{\boldsymbol{\beta}} S(\boldsymbol{\beta}) \;=\; -2\,\boldsymbol{X}^\top\boldsymbol{y} \;+\; 2\,\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}. \tag{2.5.9}$$

Syarat orde pertama meminimumkan S adalah $\nabla_{\boldsymbol{\beta}} S(\boldsymbol{\beta}) = \mathbf{0}$, sehingga diperoleh

$$\boldsymbol{X}^\top\boldsymbol{X}\,\hat{\boldsymbol{\beta}} \;=\; \boldsymbol{X}^\top\boldsymbol{y}, \tag{2.5.10}$$

yang disebut persamaan normal. Selanjutnya, matriks Hessian atau orde kedua adalah

$$\nabla^2_{\boldsymbol{\beta}} S(\boldsymbol{\beta}) \;=\; 2\,\boldsymbol{X}^\top\boldsymbol{X} \;\succeq\; \mathbf{0}, \tag{2.5.11}$$

sehingga S konveks dalam $\boldsymbol{\beta}$. Jika $\boldsymbol{X}^\top\boldsymbol{X}$ non-singular atau setara $\mathrm{rank}(\boldsymbol{X}) = p$, maka Hessian definit positif dan solusi orde pertama unik serta memberikan peminimum global.

$$\hat{\boldsymbol{\beta}} \;=\; (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}. \tag{2.5.12}$$

Jika $\boldsymbol{X}^\top \boldsymbol{X}$ singular, himpunan peminimum tidak tunggal. Salah satu solusi adalah solusi *norma-minimum* $\hat{\boldsymbol{\beta}} = \boldsymbol{X}^+ \boldsymbol{y}$ dengan $\boldsymbol{X}^+$ merupakan *Moore–Penrose pseudoinverse*.

Teorema 2.5.4 (**Teorema Gauss–Markov, (??)**) *Di bawah asumsi regresi linear klasik, penduga OLS $\hat{\boldsymbol{\beta}}$ bersifat tak bias, yaitu $\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$, dan memiliki kovariansi*

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}. \tag{2.5.13}$$

Selain itu, $\hat{\boldsymbol{\beta}}$ merupakan best linear unbiased estimator *(BLUE), artinya $\hat{\boldsymbol{\beta}}$ memiliki variansi terkecil di antara semua penduga linear tak bias.*

Bukti. Penduga OLS dapat dituliskan sebagai

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}. \tag{2.5.14}$$

Substitusi $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ menghasilkan

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\varepsilon}, \tag{2.5.15}$$

sehingga rataan dari $\hat{\boldsymbol{\beta}}$ adalah

$$\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbb{E}[\boldsymbol{\varepsilon}] = \boldsymbol{\beta}, \tag{2.5.16}$$

yang menunjukkan bahwa $\hat{\boldsymbol{\beta}}$ tak bias. Kovariansnya diperoleh dari

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathrm{Var}(\boldsymbol{\varepsilon})\, \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}. \tag{2.5.17}$$

Untuk sifat BLUE, dipertimbangkan penduga linear tak bias umum $\tilde{\boldsymbol{\beta}} = \mathbf{C}\boldsymbol{y}$ dengan $\mathbf{C}\boldsymbol{X} = \mathbf{I}$. Variansi dari $\tilde{\boldsymbol{\beta}}$ adalah

$$\mathrm{Var}(\tilde{\boldsymbol{\beta}}) = \mathbf{C}\sigma^2 \mathbf{I} \mathbf{C}^\top = \sigma^2 \mathbf{C}\mathbf{C}^\top. \tag{2.5.18}$$

Melalui dekomposisi $\mathbf{C} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top + \mathbf{D}$ dengan $\mathbf{D}\boldsymbol{X} = \mathbf{0}$, diperoleh

$$\mathrm{Var}(\tilde{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1} + \sigma^2\mathbf{D}\mathbf{D}^\top \succeq \mathrm{Var}(\hat{\boldsymbol{\beta}}), \tag{2.5.19}$$

yang membuktikan bahwa $\hat{\boldsymbol{\beta}}$ memiliki variansi minimum di antara semua penduga linear tak bias. ■

Teorema 2.5.5 (Kepadanan OLS dan MLE, (?)) *Jika asumsi normalitas galat terpenuhi, maka penduga kemungkinan maksimum (PKM) atau* maximum likelihood estimation *(MLE) untuk $\boldsymbol{\beta}$ identik dengan penduga kuadrat terkecil (OLS), yaitu*

$$\hat{\boldsymbol{\beta}}_{MLE} = \hat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y}. \tag{2.5.20}$$

Penduga MLE untuk varians galat adalah

$$\hat{\sigma}^2 = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}). \tag{2.5.21}$$

Bukti. Dengan asumsi $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, fungsi *likelihood* untuk $\boldsymbol{\beta}$ dan σ^2 adalah

$$L(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}) = (2\pi\sigma^2)^{-n/2} \exp\left(-\tfrac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right), \tag{2.5.22}$$

yang memiliki log-*likelihood*

$$\ell(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}). \tag{2.5.23}$$

Maksimisasi terhadap $\boldsymbol{\beta}$ dengan turunan pertama didapatkan

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} &= \frac{2\boldsymbol{X}^\top}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\
&= \frac{1}{\sigma^2}\boldsymbol{X}^\top \boldsymbol{y} - \frac{1}{\sigma^2}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} \\
\mathbf{0} &= \boldsymbol{X}^\top \boldsymbol{y} - \boldsymbol{X}^\top \boldsymbol{X}\widehat{\boldsymbol{\beta}} \\
\boldsymbol{X}^\top \boldsymbol{X}\widehat{\boldsymbol{\beta}} &= \boldsymbol{X}^\top \boldsymbol{y} \\
\widehat{\boldsymbol{\beta}} &= \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y}.
\end{aligned} \tag{2.5.24}$$

Perhatikan bahwa turunan kedua terhadap $\boldsymbol{\beta}$ menghasilkan $-\frac{1}{\sigma^2}\left(\boldsymbol{X}^\top\boldsymbol{X}\right)$. Sebab $\boldsymbol{X}^\top\boldsymbol{X} \succeq 0$, maka $\frac{\partial^2 \ell}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} \preceq 0$ yang menjadikan $\widehat{\boldsymbol{\beta}}$ memaksimumkan fungsi log-*likelihood*. Oleh karena itu

$$\hat{\boldsymbol{\beta}}_{\text{MLE}} = \hat{\boldsymbol{\beta}}_{\text{OLS}}. \tag{2.5.25}$$

Selanjutnya, turunan pertama ℓ terhadap σ^2 akan menghasilkan

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\
0 &= -n\widehat{\sigma}^2 + (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\
\widehat{\sigma}^2 &= \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})^\top(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}). \qquad (2.5.26)
\end{aligned}$$

■

2.5.2 Teori Asimtotik untuk OLS

Teori asimtotik mempelajari perilaku penduga OLS ketika ukuran sampel n menuju tak hingga. Hasil-hasil asimtotik sangat penting karena memberikan justifikasi teoritis untuk inferensi statistik pada sampel besar, bahkan ketika asumsi normalitas galat tidak terpenuhi.

Teorema 2.5.6 (**Konsistensi Penduga OLS, (?)**) *Misalkan model regresi linear* $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ *memenuhi kondisi-kondisi berikut.*

(i) $\mathbb{E}[\boldsymbol{\varepsilon} \mid \boldsymbol{X}] = \boldsymbol{0}$ *(eksogenitas ketat).*

(ii) $\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X} \xrightarrow{p} \boldsymbol{Q}$ *dengan* $\boldsymbol{Q}$ *matriks definit positif.*

Maka penduga OLS $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$ *bersifat konsisten, yaitu*

$$\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}. \tag{2.5.27}$$

Bukti. Penduga OLS dapat dituliskan sebagai

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y} \\ &= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{\varepsilon} \\ &= \boldsymbol{\beta} + \left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{\varepsilon}\right). \end{aligned} \tag{2.5.28}$$

Berdasarkan kondisi (ii), $\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X} \xrightarrow{p} \boldsymbol{Q}$ dengan $\boldsymbol{Q}$ definit positif. Dengan teorema pemetaan kontinu, $\left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \xrightarrow{p} \boldsymbol{Q}^{-1}$.

Selanjutnya, perlu ditunjukkan bahwa $\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{\varepsilon} \xrightarrow{p} \boldsymbol{0}$. Perhatikan bahwa

$$\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \varepsilon_i. \tag{2.5.29}$$

Sebab $\mathbb{E}[\boldsymbol{x}_i \varepsilon_i] = \mathbb{E}[\mathbb{E}[\boldsymbol{x}_i \varepsilon_i \mid \boldsymbol{X}]] = \mathbb{E}[\boldsymbol{x}_i \mathbb{E}[\varepsilon_i \mid \boldsymbol{X}]] = \boldsymbol{0}$ (dari kondisi (i)) dan dengan asumsi kondisi regularitas tambahan (seperti $\mathbb{E}[\|\boldsymbol{x}_i\|^2 \varepsilon_i^2] < \infty$), Hukum Bilangan Besar menjamin bahwa $\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{\varepsilon} \xrightarrow{p} \boldsymbol{0}$.

Dengan teorema Slutsky, diperoleh

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{\varepsilon}\right) \xrightarrow{p} \boldsymbol{Q}^{-1} \cdot \boldsymbol{0} = \boldsymbol{0}, \tag{2.5.30}$$

sehingga $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$. ■

Teorema 2.5.7 (Konsistensi Penduga Variansi, (?)) *Di bawah kondisi-kondisi Teorema konsistensi OLS dan dengan tambahan asumsi $\mathbb{E}[\varepsilon_i^2 \mid \boldsymbol{X}] = \sigma^2$ (homoskedastisitas), penduga variansi*

$$s^2 = \frac{1}{n-p-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = \frac{1}{n-p-1}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \tag{2.5.31}$$

bersifat konsisten, yaitu $s^2 \xrightarrow{p} \sigma^2$.

Bukti. Definisikan residual $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{M}\boldsymbol{y}$ dengan $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$

adalah matriks proyeksi idempoten. Maka

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{M}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{M}\boldsymbol{\varepsilon}, \tag{2.5.32}$$

sebab $\boldsymbol{M}\boldsymbol{X} = \boldsymbol{0}$. Oleh karena itu,

$$\hat{\boldsymbol{\varepsilon}}^\top \hat{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}^\top \boldsymbol{M}^\top \boldsymbol{M}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^\top \boldsymbol{M}\boldsymbol{\varepsilon}. \tag{2.5.33}$$

Dengan menggunakan sifat jejak dan ekspektasi, dapat ditunjukkan bahwa

$$\mathbb{E}[\hat{\boldsymbol{\varepsilon}}^\top \hat{\boldsymbol{\varepsilon}}] = \mathbb{E}[\mathrm{tr}(\boldsymbol{\varepsilon}^\top \boldsymbol{M}\boldsymbol{\varepsilon})] = \mathbb{E}[\mathrm{tr}(\boldsymbol{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top)] = \mathrm{tr}(\boldsymbol{M}\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top]) = \sigma^2\mathrm{tr}(\boldsymbol{M}) = \sigma^2(n-p-1). \tag{2.5.34}$$

Dengan Hukum Bilangan Besar, $\frac{1}{n}\hat{\boldsymbol{\varepsilon}}^\top \hat{\boldsymbol{\varepsilon}} \xrightarrow{p} \sigma^2$. Sebab $\frac{n}{n-p-1} \to 1$, maka $s^2 \xrightarrow{p} \sigma^2$. ■

Teorema 2.5.8 (**Normalitas Asimtotik Penduga OLS, (?)**) *Misalkan kondisi-kondisi konsistensi terpenuhi dan tambahan kondisi berikut dipenuhi.*

(i) $\mathbb{E}[\varepsilon_i^2 \mid \boldsymbol{X}] = \sigma^2$ *(homoskedastisitas).*

(ii) $\frac{1}{\sqrt{n}}\boldsymbol{X}^\top \boldsymbol{\varepsilon} \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{Q})$ *(kondisi Lindeberg–Feller terpenuhi).*

Maka penduga OLS memiliki distribusi asimtotik

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{Q}^{-1}). \tag{2.5.35}$$

Bukti. Dari pembuktian konsistensi, dapat dituliskan

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \sqrt{n}\left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{\varepsilon}\right) \\ &= \left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1}\left(\frac{1}{\sqrt{n}}\boldsymbol{X}^\top \boldsymbol{\varepsilon}\right). \end{aligned} \tag{2.5.36}$$

Berdasarkan kondisi (ii), $\frac{1}{\sqrt{n}}\boldsymbol{X}^\top \boldsymbol{\varepsilon} \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{Q})$. Dengan teorema Slutsky, sebab $\left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \xrightarrow{p} \boldsymbol{Q}^{-1}$, maka

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \boldsymbol{Q}^{-1} \cdot \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{Q}) = \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{Q}^{-1}). \tag{2.5.37}$$

■

Akibat 2.5.9 (Distribusi Asimtotik dengan Variansi Diestimasi) *Dengan penduga variansi konsisten $s^2 \xrightarrow{p} \sigma^2$, berlaku*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \sigma^2 \left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1}\right), \tag{2.5.38}$$

dan penduga matriks kovarians asimtotik yang konsisten adalah

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = s^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1}. \tag{2.5.39}$$

Teorema 2.5.10 (Distribusi Eksak dan Asimtotik Uji-t, (??)) *Untuk menguji hipotesis $H_0 : \beta_k = \beta_k^0$ terhadap $H_1 : \beta_k \neq \beta_k^0$, statistik uji-$t$ didefinisikan sebagai*

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{\mathrm{se}(\hat{\beta}_k)}, \tag{2.5.40}$$

dengan $\mathrm{se}(\hat{\beta}_k) = s\sqrt{[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}]_{kk}}$ *adalah galat standar dari $\hat{\beta}_k$.*

(a) ***Distribusi eksak:*** *Di bawah asumsi normalitas galat $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$ dan H_0 benar, $t_k \sim t_{n-p-1}$.*

(b) ***Distribusi asimtotik:*** *Di bawah kondisi regularitas asimtotik dan H_0 benar, $t_k \xrightarrow{d} \mathcal{N}(0, 1)$.*

Bukti.

(a) **Distribusi eksak:** Di bawah normalitas, $\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1})$, sehingga $\hat{\beta}_k \sim \mathcal{N}(\beta_k, \sigma^2[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}]_{kk})$. Oleh karena itu,

$$Z_k = \frac{\hat{\beta}_k - \beta_k}{\sigma\sqrt{[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}]_{kk}}} \sim \mathcal{N}(0, 1). \tag{2.5.41}$$

Selain itu, $(n - p - 1)s^2/\sigma^2 \sim \chi^2_{n-p-1}$ dan independen dari $\hat{\boldsymbol{\beta}}$. Dengan demikian,

$$t_k = \frac{Z_k}{\sqrt{(n-p-1)s^2/\sigma^2/(n-p-1)}} = \frac{Z_k}{\sqrt{s^2/\sigma^2}} \sim t_{n-p-1}. \tag{2.5.42}$$

(b) **Distribusi asimtotik:** Dari normalitas asimtotik, $\sqrt{n}(\hat{\beta}_k - \beta_k) \xrightarrow{d} \mathcal{N}(0, \sigma^2[\boldsymbol{Q}^{-1}]_{kk})$. Dengan $s^2 \xrightarrow{p} \sigma^2$ dan teorema Slutsky,

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{s\sqrt{[(\boldsymbol{X}^\top \boldsymbol{X})^{-1}]_{kk}}} \xrightarrow{d} \mathcal{N}(0,1). \tag{2.5.43}$$

■

Teorema 2.5.11 (Distribusi Eksak dan Asimtotik Uji-F, (??)) *Untuk menguji hipotesis linear umum $H_0 : \boldsymbol{R\beta} = \boldsymbol{r}$ dengan $\boldsymbol{R}$ matriks berukuran $q \times (p+1)$ dan $\operatorname{rank}(\boldsymbol{R}) = q$, statistik uji-$F$ didefinisikan sebagai*

$$F = \frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top [\boldsymbol{R}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{R}^\top]^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})}{q \cdot s^2}. \tag{2.5.44}$$

*(a) **Distribusi eksak:** Di bawah asumsi normalitas galat dan H_0 benar, $F \sim F_{q,n-p-1}$.*

*(b) **Distribusi asimtotik:** Di bawah kondisi regularitas asimtotik dan H_0 benar, $qF \xrightarrow{d} \chi^2_q$.*

Bukti.

(a) **Distribusi eksak:** Di bawah H_0 dan normalitas, $\boldsymbol{R}\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{r}, \sigma^2 \boldsymbol{R}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{R}^\top)$. Oleh karena itu,

$$\frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top [\boldsymbol{R}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{R}^\top]^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})}{\sigma^2} \sim \chi^2_q. \tag{2.5.45}$$

Sebab $(n-p-1)s^2/\sigma^2 \sim \chi^2_{n-p-1}$ dan independen dari pembilang, maka

$$F = \frac{\chi^2_q/q}{\chi^2_{n-p-1}/(n-p-1)} \sim F_{q,n-p-1}. \tag{2.5.46}$$

(b) **Distribusi asimtotik:** Dengan normalitas asimtotik dan $s^2 \xrightarrow{p} \sigma^2$,

$$qF = \frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top [\boldsymbol{R}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{R}^\top]^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})}{s^2} \xrightarrow{d} \chi^2_q. \tag{2.5.47}$$

■

Teorema 2.5.12 (Uji Wald, (??)) *Untuk menguji hipotesis $H_0 : \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$ dengan restriksi linear, statistik Wald didefinisikan sebagai*

$$W = (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top \left[\boldsymbol{R}\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}})\boldsymbol{R}^\top\right]^{-1} (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}), \tag{2.5.48}$$

dengan $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = s^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$. Di bawah H_0 dan kondisi regularitas asimtotik,

$$W \xrightarrow{d} \chi^2_q. \tag{2.5.49}$$

Bukti. Dari normalitas asimtotik, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{Q}^{-1})$. Dengan transformasi linear,

$$\sqrt{n}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{R}\boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{R}\boldsymbol{Q}^{-1}\boldsymbol{R}^\top). \tag{2.5.50}$$

Di bawah H_0, $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$, sehingga

$$\sqrt{n}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{R}\boldsymbol{Q}^{-1}\boldsymbol{R}^\top). \tag{2.5.51}$$

Statistik Wald dapat dituliskan sebagai

$$W = n(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top \left[\boldsymbol{R}(n\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}))\boldsymbol{R}^\top\right]^{-1} (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}). \tag{2.5.52}$$

Sebab $n\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = s^2 n(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \xrightarrow{p} \sigma^2\boldsymbol{Q}^{-1}$, maka dengan teorema Slutsky dan sifat bentuk kuadratik normal, $W \xrightarrow{d} \chi^2_q$. ■

Teorema 2.5.13 (Uji Rasio Kemungkinan (*Likelihood Ratio Test*), (?)) *Untuk* *menguji $H_0 : \boldsymbol{R\beta} = \boldsymbol{r}$, statistik rasio kemungkinan didefinisikan sebagai*

$$\text{LR} = n\left(\log \hat{\sigma}_R^2 - \log \hat{\sigma}_U^2\right), \tag{2.5.53}$$

dengan $\hat{\sigma}_R^2$ dan $\hat{\sigma}_U^2$ adalah penduga MLE variansi di bawah model terestriksi dan tidak terestriksi. Di bawah H_0 dan kondisi regularitas,

$$\text{LR} \xrightarrow{d} \chi_q^2. \tag{2.5.54}$$

Bukti. Fungsi log-likelihood untuk model regresi normal adalah

$$\ell(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X\beta})^\top(\boldsymbol{y} - \boldsymbol{X\beta}). \tag{2.5.55}$$

Penduga MLE untuk σ^2 pada model tidak terestriksi adalah $\hat{\sigma}_U^2 = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_U)^\top(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_U)$, sedangkan pada model terestriksi adalah $\hat{\sigma}_R^2 = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_R)^\top(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_R)$.

Statistik rasio kemungkinan didefinisikan sebagai $\text{LR} = 2(\ell_U - \ell_R)$, dengan ℓ_U dan ℓ_R adalah nilai log-likelihood maksimum pada model tidak terestriksi dan terestriksi. Substitusi memberikan

$$\begin{aligned}\text{LR} &= 2\left[-\frac{n}{2}\log(\hat{\sigma}_U^2) - \frac{n}{2} + \frac{n}{2}\log(\hat{\sigma}_R^2) + \frac{n}{2}\right] \\ &= n\left(\log \hat{\sigma}_R^2 - \log \hat{\sigma}_U^2\right).\end{aligned} \tag{2.5.56}$$

Di bawah H_0, dengan ekspansi Taylor dari $\log(\hat{\sigma}_R^2/\hat{\sigma}_U^2)$ di sekitar 1 dan menggunakan fakta bahwa $\hat{\sigma}_R^2 - \hat{\sigma}_U^2 = \frac{1}{n}(\boldsymbol{R}\hat{\boldsymbol{\beta}}_U - \boldsymbol{r})^\top[\boldsymbol{R}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{R}^\top]^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}}_U - \boldsymbol{r}) + o_p(n^{-1})$, dapat ditunjukkan bahwa

$$\text{LR} = \frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}}_U - \boldsymbol{r})^\top[\boldsymbol{R}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{R}^\top]^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}}_U - \boldsymbol{r})}{\hat{\sigma}_U^2} + o_p(1). \tag{2.5.57}$$

Sebab $\sqrt{n}(\boldsymbol{R}\hat{\boldsymbol{\beta}}_U - \boldsymbol{r}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{RQ}^{-1}\boldsymbol{R}^\top)$ di bawah H_0 dan $\hat{\sigma}_U^2 \xrightarrow{p} \sigma^2$, maka berdasarkan sifat bentuk kuadratik normal, $\text{LR} \xrightarrow{d} \chi_q^2$. ∎

Teorema 2.5.14 (**Uji Pengali Lagrange (*Lagrange Multiplier Test*), (?)**) *Untuk* *menguji $H_0 : \boldsymbol{R\beta} = \boldsymbol{r}$, statistik pengali Lagrange didefinisikan sebagai*

$$\mathrm{LM} = \frac{1}{\hat{\sigma}_R^2}\hat{\boldsymbol{\varepsilon}}_R^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \hat{\boldsymbol{\varepsilon}}_R, \tag{2.5.58}$$

dengan $\hat{\boldsymbol{\varepsilon}}_R$ adalah residual dari model terestriksi. Di bawah H_0 dan kondisi regularitas,

$$\mathrm{LM} \xrightarrow{d} \chi_q^2. \tag{2.5.59}$$

Proposisi 2.5.15 (**Kesetaraan Asimtotik Uji-Uji Klasik, (?)**) *Di bawah hipotesis nol $H_0 : \boldsymbol{R\beta} = \boldsymbol{r}$ dan kondisi regularitas, ketiga statistik uji—Wald (W), rasio kemungkinan (LR), dan pengali Lagrange (LM)—bersifat setara secara asimtotik dalam arti*

$$W - \mathrm{LR} \xrightarrow{p} 0, \quad \mathrm{LR} - \mathrm{LM} \xrightarrow{p} 0, \quad W - \mathrm{LM} \xrightarrow{p} 0. \tag{2.5.60}$$

Ketiga statistik ini memiliki distribusi limit yang sama, yaitu χ_q^2, sehingga memberikan kesimpulan inferensi yang identik secara asimtotik.

2.5.3 Regresi Terboboti

Regresi terboboti dibentuk apabila model regresi linear klasik melanggar asumsi homoskedastisitas galat. Heteroskedastisitas terjadi apabila komponen galat regresi tidak memiliki variansi yang konstan. Apabila terjadi heteroskedastisitas, maka Teorema Gauss–Markov tidak berlaku sehingga penduga dari model regresi tidak lagi bersifat BLUE.

Definisi 2.5.16 (**Model Regresi Linear dengan Heteroskedastisitas, (???)**) *Pertimbangkan model regresi linear*

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}, \qquad \mathbb{E}[\boldsymbol{\varepsilon} \mid \boldsymbol{X}] = \boldsymbol{0}, \tag{2.5.61}$$

dengan struktur ragam–kovarians galat

$$\mathrm{Var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \sigma^2\boldsymbol{\Omega}, \tag{2.5.62}$$

dengan $\mathbf{\Omega}$ *adalah matriks simetri definit-posistif berukuran* $n \times n$. *Model yang memperkenalkan struktur variansi-kovarians galat disebut sebagai model kuadrat terkecil tergeneralisasi atau* generalized least squares *(GLS).*

Kasus umum dari GLS yang sering dipakai adalah kuadrat terkecil terboboti atau *weighted least squares* (WLS), yaitu $\mathbf{\Omega} = \operatorname{diag}(\omega_1, \ldots, \omega_n)$ yang merupakan heteroskedastisitas diagonal dengan bobot $\mathbf{W} = \mathbf{\Omega}^{-1} = \operatorname{diag}(w_1, \ldots, w_n)$, $w_i = 1/\omega_i$.

Penduga GLS diperoleh dengan meminimumkan jumlah kuadrat terbobot

$$S_{\mathbf{W}}(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^{\top}\mathbf{W}\,(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{2.5.63}$$

dengan $\mathbf{W} = \mathbf{\Omega}^{-1}$. Diferensiasi terhadap $\boldsymbol{\beta}$ dan disamakan dengan nol memberikan

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} S_{\mathbf{W}}(\boldsymbol{\beta}) &= -2\mathbf{W}\boldsymbol{X}^{\top}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \quad \text{(karena } \mathbf{W} \text{ simetris)} \\ \mathbf{0} &= -2\boldsymbol{X}^{\top}\mathbf{W}\boldsymbol{y} - 2\boldsymbol{X}^{\top}\mathbf{W}\boldsymbol{X}\widehat{\boldsymbol{\beta}} \\ \boldsymbol{X}^{\top}\mathbf{W}\boldsymbol{X}\,\widehat{\boldsymbol{\beta}} &= \boldsymbol{X}^{\top}\mathbf{W}\boldsymbol{y}, \end{aligned} \tag{2.5.64}$$

sehingga pada asumsi $\boldsymbol{X}^{\top}\mathbf{W}\boldsymbol{X}$ nonsingular,

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\boldsymbol{X}^{\top}\mathbf{W}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top}\mathbf{W}\boldsymbol{y}. \tag{2.5.65}$$

Penduga ini tak bias jika $\mathbf{W}$ tidak acak atau independen dari $\boldsymbol{\varepsilon}$ dan $\mathbb{E}[\boldsymbol{\varepsilon} \mid \boldsymbol{X}] = \mathbf{0}$, sebab $\mathbb{E}[\hat{\boldsymbol{\beta}}_{\text{GLS}} \mid \boldsymbol{X}] = \boldsymbol{\beta}$.

Teorema 2.5.17 (**Teorema Aitken atau Generalisasi Gauss–Markov, (??)**) *Misalkan model regresi linear*

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \mathbb{E}[\varepsilon \mid \boldsymbol{X}] = 0, \quad \operatorname{Var}(\varepsilon \mid \boldsymbol{X}) = \sigma^2\mathbf{\Omega},$$

dengan $\boldsymbol{\Omega}$ *simetris definit-positif, maka penduga*

$$\widehat{\boldsymbol{\beta}}_{GLS} = (\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{y} \tag{2.5.66}$$

adalah best linear unbiased estimator *(BLUE) dari* β*.*

Bukti. Pertama-tama perhatikan bahwa setiap penduga linear bagi $\boldsymbol{\beta}$ dapat ditulis dalam bentuk $\widetilde{\boldsymbol{\beta}} = \mathbf{A}\boldsymbol{y}$ dengan $\mathbf{A} \in \mathbb{R}^{p\times n}$. Agar tak bias, berlaku

$$\mathbb{E}[\widetilde{\boldsymbol{\beta}} \mid \boldsymbol{X}] = \mathbf{A}\mathbb{E}[\boldsymbol{y} \mid \boldsymbol{X}] = \mathbf{A}\boldsymbol{X}\boldsymbol{\beta}. \tag{2.5.67}$$

Supaya sama dengan $\boldsymbol{\beta}$ untuk semua $\boldsymbol{\beta}$, syarat yang harus dipenuhi adalah $\mathbf{A}\boldsymbol{X} = \mathbf{I}_p$. Selanjutnya, kovarians dari $\widetilde{\boldsymbol{\beta}}$ diberikan oleh

$$\mathrm{Var}(\widetilde{\boldsymbol{\beta}} \mid \boldsymbol{X}) = \mathbf{A}\,\sigma^2\boldsymbol{\Omega}\,\mathbf{A}^\top. \tag{2.5.68}$$

Di antara semua matriks $\mathbf{A}$ yang memenuhi $\mathbf{A}\boldsymbol{X} = \mathbf{I}_p$, pilih

$$\mathbf{A}_0 = (\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Omega}^{-1}. \tag{2.5.69}$$

Jelas bahwa $\mathbf{A}_0\boldsymbol{X} = \mathbf{I}_p$, sehingga $\mathbf{A}_0$ menghasilkan penduga tak bias. Penduga inilah yang dikenal sebagai *generalized least squares*.

Sekarang, ambil sembarang penduga tak bias lain $\mathbf{A}$. Matriks $\mathbf{A}$ tersebut dapat ditulis sebagai $\mathbf{A} = \mathbf{A}_0 + \mathbf{D}$ dengan $\mathbf{D}\boldsymbol{X} = \mathbf{0}$, sebab $\mathbf{A}_0\boldsymbol{X} = \mathbf{I}_p$ dan syarat $\mathbf{A}\boldsymbol{X} = \mathbf{I}_p$ harus tetap dipenuhi. Dengan demikian,

$$\widetilde{\boldsymbol{\beta}} = \mathbf{A}\boldsymbol{y} = \mathbf{A}_0\boldsymbol{y} + \mathbf{D}\boldsymbol{y}. \tag{2.5.70}$$

Variansi dari penduga ini adalah

$$\mathrm{Var}(\widetilde{\boldsymbol{\beta}} \mid \boldsymbol{X}) = \mathrm{Var}(\mathbf{A}_0\boldsymbol{y} \mid \boldsymbol{X}) + \sigma^2\mathbf{D}\boldsymbol{\Omega}\mathbf{D}^\top. \tag{2.5.71}$$

Bagian pertama adalah kovarians dari $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}$, sedangkan bagian kedua selalu semi definit positif karena $\mathbf{\Omega}$ definit-positif. Hal ini berarti tambahan $\mathbf{D}\boldsymbol{y}$ hanya menambah variansi tanpa mengurangi bias. Dengan demikian, kovarians dari sembarang penduga linear tak bias selalu lebih besar atau sama dengan kovarians GLS. Oleh karena itu, $\widehat{\boldsymbol{\beta}}_{\text{GLS}}$ meminimalkan kovarians di antara semua penduga linear tak bias, sehingga merupakan BLUE. ■

Teorema 2.5.18 (**Kepadanan GLS dan MLE di Bawah Normalitas, (??)**) *Jika $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{\Omega})$, maka penduga maksimum likelihood (MLE) untuk $\boldsymbol{\beta}$ adalah*

$$\hat{\boldsymbol{\beta}}_{MLE} = \hat{\boldsymbol{\beta}}_{GLS} = (\boldsymbol{X}^\top\mathbf{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\mathbf{\Omega}^{-1}\boldsymbol{y}, \tag{2.5.72}$$

dan MLE untuk σ^2 (dengan $\mathbf{\Omega}$ diketahui) adalah

$$\hat{\sigma}^2_{MLE} = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top\mathbf{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}). \tag{2.5.73}$$

Bukti. Di bawah normalitas galat, fungsi *likelihood* untuk $\boldsymbol{\beta}$ dan σ^2 adalah

$$L(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}) = (2\pi\sigma^2)^{-n/2}|\mathbf{\Omega}|^{-1/2}\exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top\mathbf{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right).$$

Hal ini berarti fungsi log-*likelihood* untuk $\boldsymbol{\beta}$ dan σ^2 adalah

$$\ell(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log|\mathbf{\Omega}| - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top\mathbf{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}).$$

Maksimisasi ℓ untuk $\boldsymbol{\beta}$ dengan turunan pertama menghasilkan

$$\begin{aligned}
\frac{\partial\ell(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y})}{\partial\boldsymbol{\beta}} &= \frac{2}{2\sigma^2}\boldsymbol{X}^\top\mathbf{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\
0 &= \frac{1}{\sigma^2}\boldsymbol{X}^\top\mathbf{\Omega}^{-1}\boldsymbol{y} - \frac{1}{\sigma^2}\boldsymbol{X}^\top\mathbf{\Omega}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}} \\
\frac{1}{\sigma^2}\boldsymbol{X}^\top\mathbf{\Omega}^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}} &= \frac{1}{\sigma^2}\boldsymbol{X}^\top\mathbf{\Omega}^{-1}\boldsymbol{y} \\
\widehat{\boldsymbol{\beta}} &= \left(\boldsymbol{X}^\top\mathbf{\Omega}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\mathbf{\Omega}^{-1}\boldsymbol{y},
\end{aligned} \tag{2.5.74}$$

sehingga $\hat{\boldsymbol{\beta}}_{\text{MLE}} = \hat{\boldsymbol{\beta}}_{\text{GLS}}$. Selanjutnya, turunan terhadap σ^2 memberi $\partial\ell/\partial\sigma^2 = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top\mathbf{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = 0$, yang menghasilkan rumus $\hat{\sigma}^2_{\text{MLE}}$

di atas. ■

Dalam analisis nyata, $\boldsymbol{\Omega}$ biasanya jarang diketahui. *Feasible GLS* (FGLS) memperkirakan $\boldsymbol{\Omega}$, misalnya dari residu OLS atau model varian parametrik, lalu menyusun $\widehat{\mathbf{W}} = \widehat{\boldsymbol{\Omega}}^{-1}$ dan menghitung $\hat{\boldsymbol{\beta}}_{\text{FGLS}} = (\boldsymbol{X}^\top \widehat{\mathbf{W}} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \widehat{\mathbf{W}} \boldsymbol{y}$. Di bawah kondisi reguler, $\hat{\boldsymbol{\beta}}_{\text{FGLS}}$ konsisten dan asimtotik efisien. Namun, untuk sampel kecil, sifat *finite-sample* bergantung pada kualitas spesifikasi $\hat{\boldsymbol{\Omega}}$.

2.5.4 Teori Asimtotik untuk GLS/WLS

Teori asimtotik untuk penduga GLS/WLS mempelajari perilaku penduga ketika ukuran sampel n menuju tak hingga. Hasil-hasil asimtotik ini penting untuk memberikan justifikasi teoritis bagi inferensi statistik pada sampel besar.

Teorema 2.5.19 (**Konsistensi Penduga GLS, (?)**) *Misalkan model regresi linear dengan heteroskedastisitas*

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \mathbb{E}[\boldsymbol{\varepsilon} \mid \boldsymbol{X}] = \mathbf{0}, \qquad \operatorname{Var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \sigma^2 \boldsymbol{\Omega},$$

dengan $\boldsymbol{\Omega}$ diketahui dan definit positif. Jika kondisi-kondisi berikut terpenuhi:

(i) $\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{X} \xrightarrow{p} \boldsymbol{Q}_\Omega$ *dengan $\boldsymbol{Q}_\Omega$ matriks definit positif,*

(ii) $\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} \xrightarrow{p} \mathbf{0}$*,*

maka penduga GLS $\hat{\boldsymbol{\beta}}_{GLS} = (\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{y}$ bersifat konsisten, yaitu

$$\hat{\boldsymbol{\beta}}_{GLS} \xrightarrow{p} \boldsymbol{\beta}. \tag{2.5.75}$$

Bukti. Penduga GLS dapat dituliskan sebagai

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{\text{GLS}} &= (\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{y} \\
&= (\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\
&= \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} \\
&= \boldsymbol{\beta} + \left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{X}\right)^{-1} \left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}\right).
\end{aligned} \tag{2.5.76}$$

Berdasarkan kondisi (i), $\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X} \xrightarrow{p} \boldsymbol{Q}_\Omega$ dengan $\boldsymbol{Q}_\Omega$ definit positif. Dengan teorema pemetaan kontinu, $\left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1} \xrightarrow{p} \boldsymbol{Q}_\Omega^{-1}$. Berdasarkan kondisi (ii), $\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon} \xrightarrow{p} \mathbf{0}$. Dengan teorema Slutsky, diperoleh

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} - \boldsymbol{\beta} = \left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}\right) \xrightarrow{p} \boldsymbol{Q}_\Omega^{-1}\cdot\mathbf{0} = \mathbf{0}, \tag{2.5.77}$$

sehingga $\hat{\boldsymbol{\beta}}_{\text{GLS}} \xrightarrow{p} \boldsymbol{\beta}$. ■

Teorema 2.5.20 (Konsistensi Penduga Variansi GLS, (?)) *Di bawah kondisi-kondisi konsistensi GLS, penduga variansi*

$$s^2_{GLS} = \frac{1}{n-p-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS})^\top\boldsymbol{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{GLS}) \tag{2.5.78}$$

bersifat konsisten, yaitu $s^2_{GLS} \xrightarrow{p} \sigma^2$.

Bukti. Definisikan residual GLS $\hat{\boldsymbol{\varepsilon}}_{\text{GLS}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\text{GLS}} = \boldsymbol{M}_\Omega\boldsymbol{y}$ dengan $\boldsymbol{M}_\Omega = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}$. Maka

$$\hat{\boldsymbol{\varepsilon}}_{\text{GLS}} = \boldsymbol{M}_\Omega(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{M}_\Omega\boldsymbol{\varepsilon}, \tag{2.5.79}$$

sebab $\boldsymbol{M}_\Omega\boldsymbol{X} = \mathbf{0}$. Oleh karena itu,

$$\hat{\boldsymbol{\varepsilon}}_{\text{GLS}}^\top\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{\varepsilon}}_{\text{GLS}} = \boldsymbol{\varepsilon}^\top\boldsymbol{M}_\Omega^\top\boldsymbol{\Omega}^{-1}\boldsymbol{M}_\Omega\boldsymbol{\varepsilon}. \tag{2.5.80}$$

Dengan menggunakan sifat ekspektasi dan fakta bahwa $\operatorname{tr}(\boldsymbol{M}_\Omega^\top\boldsymbol{\Omega}^{-1}\boldsymbol{M}_\Omega\cdot\sigma^2\boldsymbol{\Omega}) = \sigma^2(n-p-1)$, serta Hukum Bilangan Besar, diperoleh $s^2_{\text{GLS}} \xrightarrow{p} \sigma^2$. ■

Teorema 2.5.21 (Normalitas Asimtotik Penduga GLS, (?)) *Misalkan kondisi-kondisi konsistensi terpenuhi dan tambahan kondisi berikut dipenuhi:*

(i) $\frac{1}{\sqrt{n}}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{Q}_\Omega)$ *(kondisi Lindeberg–Feller terpenuhi).*

Maka penduga GLS memiliki distribusi asimtotik

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{Q}_\Omega^{-1}). \tag{2.5.81}$$

Bukti. Dari pembuktian konsistensi, dapat dituliskan

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\beta}}_{\text{GLS}} - \boldsymbol{\beta}) &= \sqrt{n}\left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}\right)\\ &= \left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1}\left(\frac{1}{\sqrt{n}}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}\right).\end{aligned} \tag{2.5.82}$$

Berdasarkan kondisi (i), $\frac{1}{\sqrt{n}}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{Q}_\Omega)$. Dengan teorema Slutsky, sebab $\left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1} \xrightarrow{p} \boldsymbol{Q}_\Omega^{-1}$, maka

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{\text{GLS}} - \boldsymbol{\beta}) \xrightarrow{d} \boldsymbol{Q}_\Omega^{-1} \cdot \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{Q}_\Omega) = \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{Q}_\Omega^{-1}). \tag{2.5.83}$$

■

Akibat 2.5.22 (Distribusi Asimtotik dengan Variansi Diestimasi) *Dengan penduga variansi konsisten $s_{GLS}^2 \xrightarrow{p} \sigma^2$, berlaku*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \sigma^2\left(\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X}\right)^{-1}\right), \tag{2.5.84}$$

dan penduga matriks kovarians asimtotik yang konsisten adalah

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{GLS}) = s_{GLS}^2(\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}. \tag{2.5.85}$$

Teorema 2.5.23 (Distribusi Uji Statistik untuk GLS/WLS, (?)) *Distribusi asimtotik dari uji-uji statistik untuk model GLS/WLS analog dengan kasus OLS. Secara spesifik:*

(a) ***Uji-t:*** *Statistik uji*

$$t_k = \frac{\hat{\beta}_{k,GLS} - \beta_k^0}{s_{GLS}\sqrt{[(\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}]_{kk}}} \tag{2.5.86}$$

memiliki distribusi t_{n-p-1} secara eksak di bawah normalitas galat, dan $\mathcal{N}(0,1)$

secara asimtotik.

(b) ***Uji-F:*** *Statistik uji untuk hipotesis linear $H_0 : \boldsymbol{R\beta} = \boldsymbol{r}$*

$$F = \frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{r})^\top [\boldsymbol{R}(\boldsymbol{X}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}^\top]^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{r})}{q \cdot s^2_{GLS}} \tag{2.5.87}$$

memiliki distribusi $F_{q,n-p-1}$ secara eksak dan $qF \xrightarrow{d} \chi^2_q$ secara asimtotik.

(c) ***Uji Wald:*** *Statistik Wald*

$$W = (\boldsymbol{R}\hat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{r})^\top \left[\boldsymbol{R}\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_{GLS})\boldsymbol{R}^\top\right]^{-1} (\boldsymbol{R}\hat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{r}) \tag{2.5.88}$$

memiliki distribusi $W \xrightarrow{d} \chi^2_q$ di bawah H_0.

Pembuktian distribusi-distribusi ini mengikuti pola yang serupa dengan kasus OLS (lihat Teorema distribusi uji-t, uji-F, dan uji Wald pada Subbab Teori Asimtotik untuk OLS), dengan mengganti $(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$ menjadi $(\boldsymbol{X}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}$ dan s^2 menjadi s^2_{GLS}.

Proposisi 2.5.24 (**Kesetaraan Asimtotik Uji-Uji Klasik untuk GLS**) *Di bawah hipotesis nol $H_0 : \boldsymbol{R\beta} = \boldsymbol{r}$ dan kondisi regularitas, ketiga statistik uji—Wald (W), rasio kemungkinan (LR), dan pengali Lagrange (LM)—untuk model GLS bersifat setara secara asimtotik. Ketiganya memiliki distribusi limit χ^2_q dan memberikan kesimpulan inferensi yang identik secara asimtotik.*

2.6 Pemodelan Polinomial Lokal

Pemodelan polinomial lokal (*local polynomial modeling*) merupakan teknik nonparametrik yang fleksibel untuk mengestimasi fungsi regresi tanpa mengasumsikan bentuk fungsional tertentu secara global. Pendekatan ini menjadi fondasi penting bagi metode regresi terboboti geografis (GWR) yang akan dibahas pada bagian selanjutnya. Pembahasan dalam bagian ini mengikuti kerangka yang dikembangkan oleh **?** dan **?**.

Definisi 2.6.1 (**Model Regresi Nonparametrik**) *Pertimbangkan model regresi non-*

parametrik

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{2.6.1}$$

dengan $m(\cdot)$ *adalah fungsi regresi yang tidak diketahui,* X_i *adalah variabel prediktor,* Y_i *adalah variabel respons, dan* ε_i *adalah galat acak dengan* $\mathbb{E}[\varepsilon_i \mid X_i] = 0$ *dan* $\mathrm{Var}(\varepsilon_i \mid X_i) = \sigma^2(X_i)$.

Definisi 2.6.2 (**Aproksimasi Polinomial Lokal**) *Untuk mengestimasi* $m(x_0)$ *pada titik target* x_0*, fungsi regresi* $m(x)$ *diaproksimasi secara lokal menggunakan ekspansi Taylor orde-*p *di sekitar* x_0*:*

$$m(x) \approx m(x_0) + m'(x_0)(x - x_0) + \frac{m''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{m^{(p)}(x_0)}{p!}(x - x_0)^p. \tag{2.6.2}$$

Dengan mendefinisikan $\beta_j = \frac{m^{(j)}(x_0)}{j!}$ *untuk* $j = 0, 1, \ldots, p$*, aproksimasi lokal dapat dituliskan sebagai*

$$m(x) \approx \sum_{j=0}^{p} \beta_j (x - x_0)^j. \tag{2.6.3}$$

Definisi 2.6.3 (**Estimator Polinomial Lokal**) *Estimator polinomial lokal* $\hat{m}(x_0)$ *diperoleh dengan menyelesaikan masalah optimisasi berikut:*

$$\min_{\beta_0, \beta_1, \ldots, \beta_p} \sum_{i=1}^{n} K_h(X_i - x_0) \left[Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j \right]^2, \tag{2.6.4}$$

dengan $K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$ *adalah fungsi* kernel *dengan parameter lebar pita (bandwidth)* $h > 0$*. Estimator polinomial lokal pada titik* x_0 *didefinisikan sebagai* $\hat{m}(x_0) = \hat{\beta}_0(x_0)$*, di mana* $\hat{\beta}_0(x_0)$ *adalah solusi optimal dari masalah di atas.*

Contoh 2.6.4 Dua kasus khusus yang paling umum digunakan adalah sebagai berikut.

(a) **Estimator Nadaraya–Watson (**$p = 0$**):** Dengan $p = 0$, diperoleh estimator rata-rata lokal

$$\hat{m}(x_0) = \frac{\sum_{i=1}^{n} K_h(X_i - x_0) Y_i}{\sum_{i=1}^{n} K_h(X_i - x_0)}. \tag{2.6.5}$$

(b) **Estimator Linear Lokal ($p = 1$):** Dengan $p = 1$, model lokal adalah $m(x) \approx \beta_0 + \beta_1(x - x_0)$, dan estimator $\hat{m}(x_0) = \hat{\beta}_0(x_0)$ diperoleh dari regresi linear tertimbang lokal.

2.6.1 Asumsi Model Polinomial Lokal

Untuk memperoleh sifat-sifat asimtotik estimator polinomial lokal, diperlukan asumsi-asumsi regularitas berikut.

(A1) **Kehalusan fungsi regresi:** Fungsi regresi $m(x)$ memiliki turunan kontinu hingga orde $(p + 1)$ di sekitar x_0.

(A2) **Kepadatan prediktor:** Variabel prediktor X memiliki fungsi kepadatan $f(x)$ yang kontinu dan positif di x_0, yaitu $f(x_0) > 0$.

(A3) **Struktur galat:** Galat ε_i memenuhi $\mathbb{E}[\varepsilon_i \mid X_i] = 0$ dan $\mathrm{Var}(\varepsilon_i \mid X_i) = \sigma^2(X_i)$ dengan $\sigma^2(x)$ kontinu di x_0.

(A4) **Sifat *kernel*:** Fungsi *kernel* $K(\cdot)$ adalah fungsi kepadatan simetris dengan momen hingga, yaitu $K(u) \geq 0$, $\int K(u)\,du = 1$, $\int uK(u)\,du = 0$, dan $\mu_2(K) = \int u^2K(u)\,du < \infty$.

(A5) **Kondisi *bandwidth*:** Seiring $n \to \infty$, berlaku $h \to 0$ dan $nh \to \infty$.

Beberapa fungsi *kernel* yang sering digunakan dalam pemodelan polinomial lokal adalah sebagai berikut.

(a) ***Kernel* Gaussian:** $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$.

(b) ***Kernel* Epanechnikov:** $K(u) = \frac{3}{4}(1 - u^2)\mathbf{1}_{|u|\leq 1}$.

(c) ***Kernel* Bi-kuadrat:** $K(u) = \frac{15}{16}(1 - u^2)^2\mathbf{1}_{|u|\leq 1}$.

(d) ***Kernel* Tri-kubik:** $K(u) = \frac{70}{81}(1 - |u|^3)^3\mathbf{1}_{|u|\leq 1}$.

2.6.2 Pendugaan Model Polinomial Lokal

Definisi 2.6.5 (**Regresi Tertimbang Lokal**) *Regresi tertimbang lokal (*locally weighted least squares*, LWLS) pada titik x_0 menyelesaikan masalah optimisasi berikut:*

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} K_h(X_i - x_0) \left[Y_i - \boldsymbol{z}_i^\top \boldsymbol{\beta}\right]^2, \tag{2.6.6}$$

dengan $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^\top$ dan $\boldsymbol{z}_i = (1, X_i - x_0, (X_i - x_0)^2, \ldots, (X_i - x_0)^p)^\top$. Solusi optimal $\hat{\boldsymbol{\beta}}(x_0)$ memberikan estimator polinomial lokal $\hat{m}(x_0) = \hat{\beta}_0(x_0)$.

Teorema 2.6.6 *Definisikan matriks desain lokal $\boldsymbol{X}_{x_0}$ dan matriks bobot $\boldsymbol{W}_{x_0}$ sebagai*

$$\boldsymbol{X}_{x_0} = \begin{pmatrix} 1 & (X_1 - x_0) & (X_1 - x_0)^2 & \cdots & (X_1 - x_0)^p \\ 1 & (X_2 - x_0) & (X_2 - x_0)^2 & \cdots & (X_2 - x_0)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (X_n - x_0) & (X_n - x_0)^2 & \cdots & (X_n - x_0)^p \end{pmatrix} \in \mathbb{R}^{n\times(p+1)}, \tag{2.6.7}$$

dan

$$\boldsymbol{W}_{x_0} = \operatorname{diag}\left(K_h(X_1 - x_0), K_h(X_2 - x_0), \ldots, K_h(X_n - x_0)\right) \in \mathbb{R}^{n\times n}. \tag{2.6.8}$$

Maka estimator polinomial lokal diberikan oleh

$$\hat{\boldsymbol{\beta}}(x_0) = \left(\boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{X}_{x_0}\right)^{-1} \boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{y}, \tag{2.6.9}$$

dengan $\boldsymbol{y} = (Y_1, Y_2, \ldots, Y_n)^\top$. Khususnya, estimator untuk $m(x_0)$ adalah

$$\hat{m}(x_0) = \boldsymbol{e}_1^\top \hat{\boldsymbol{\beta}}(x_0) = \boldsymbol{e}_1^\top \left(\boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{X}_{x_0}\right)^{-1} \boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{y}, \tag{2.6.10}$$

dengan $\boldsymbol{e}_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^{p+1}$.

Bukti. Masalah minimisasi LWLS dapat dituliskan dalam bentuk matriks sebagai

$$\min_{\boldsymbol{\beta}} \left(\boldsymbol{y} - \boldsymbol{X}_{x_0}\boldsymbol{\beta}\right)^\top \boldsymbol{W}_{x_0} \left(\boldsymbol{y} - \boldsymbol{X}_{x_0}\boldsymbol{\beta}\right). \tag{2.6.11}$$

Turunan pertama terhadap $\boldsymbol{\beta}$ adalah

$$\frac{\partial}{\partial \boldsymbol{\beta}} \left(\boldsymbol{y} - \boldsymbol{X}_{x_0}\boldsymbol{\beta}\right)^\top \boldsymbol{W}_{x_0} \left(\boldsymbol{y} - \boldsymbol{X}_{x_0}\boldsymbol{\beta}\right) = -2\boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \left(\boldsymbol{y} - \boldsymbol{X}_{x_0}\boldsymbol{\beta}\right). \tag{2.6.12}$$

Menyamakan dengan nol memberikan persamaan normal tertimbang

$$\boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{X}_{x_0} \hat{\boldsymbol{\beta}}(x_0) = \boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{y}. \tag{2.6.13}$$

Jika $\boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{X}_{x_0}$ nonsingular, solusinya adalah

$$\hat{\boldsymbol{\beta}}(x_0) = \left(\boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{X}_{x_0}\right)^{-1} \boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{y}. \tag{2.6.14}$$

Sebab $\hat{\beta}_0(x_0)$ adalah komponen pertama dari $\hat{\boldsymbol{\beta}}(x_0)$, maka $\hat{m}(x_0) = \boldsymbol{e}_1^\top \hat{\boldsymbol{\beta}}(x_0)$. ■

Teorema 2.6.7 (**Representasi Bobot Ekuivalen**) *Estimator $\hat{m}(x_0)$ dapat dituliskan sebagai kombinasi linear tertimbang dari respons:*

$$\hat{m}(x_0) = \sum_{i=1}^{n} w_i(x_0) Y_i, \tag{2.6.15}$$

dengan bobot efektif $w_i(x_0) = \boldsymbol{e}_1^\top \left(\boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{X}_{x_0}\right)^{-1} \boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{e}_i$, dengan $\boldsymbol{e}_i$ adalah vektor unit ke-i dalam $\mathbb{R}^n$.

Bukti. Dari teorema sebelumnya, estimator polinomial lokal diberikan oleh

$$\hat{m}(x_0) = \boldsymbol{e}_1^\top \left(\boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{X}_{x_0}\right)^{-1} \boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{y}. \tag{2.6.16}$$

Vektor respons $\boldsymbol{y} = (Y_1, Y_2, \ldots, Y_n)^\top$ dapat dituliskan sebagai kombinasi linear dari vektor-vektor unit standar dalam $\mathbb{R}^n$, yaitu

$$\boldsymbol{y} = \sum_{i=1}^{n} Y_i \boldsymbol{e}_i, \tag{2.6.17}$$

dengan $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)^\top$ adalah vektor unit ke-i yang memiliki nilai 1 pada posisi ke-i dan 0 pada posisi lainnya.

Substitusi representasi ini ke dalam rumus estimator memberikan

$$\begin{aligned}
\hat{m}(x_0) &= \boldsymbol{e}_1^\top \left(\boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{X}_{x_0}\right)^{-1} \boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \left(\sum_{i=1}^{n} Y_i \boldsymbol{e}_i\right) \\
&= \sum_{i=1}^{n} Y_i \cdot \boldsymbol{e}_1^\top \left(\boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{X}_{x_0}\right)^{-1} \boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{e}_i, \qquad (2.6.18)
\end{aligned}$$

dengan pertukaran urutan penjumlahan dan perkalian matriks dijamin oleh linearitas operasi matriks.

Dengan mendefinisikan bobot efektif sebagai

$$w_i(x_0) = \boldsymbol{e}_1^\top \left(\boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{X}_{x_0}\right)^{-1} \boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{e}_i, \tag{2.6.19}$$

diperoleh representasi yang diinginkan

$$\hat{m}(x_0) = \sum_{i=1}^{n} w_i(x_0) Y_i. \tag{2.6.20}$$

Perhatikan bahwa $\boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0} \boldsymbol{e}_i$ adalah kolom ke-i dari matriks $\boldsymbol{X}_{x_0}^\top \boldsymbol{W}_{x_0}$, yang sama dengan $K_h(X_i - x_0)\boldsymbol{z}_i$ dengan $\boldsymbol{z}_i = (1, X_i - x_0, \ldots, (X_i - x_0)^p)^\top$. Oleh karena itu, bobot $w_i(x_0)$ bergantung pada jarak X_i dari titik target x_0 melalui fungsi *kernel* $K_h(X_i - x_0)$ dan struktur polinomial lokal. ■

Teorema 2.6.8 (**Bias Estimator Polinomial Lokal**) *Di bawah asumsi regularitas (A1)–(A5), bias bersyarat dari estimator polinomial lokal orde-p adalah*

$$\text{Bias}\left(\hat{m}(x_0) \mid X_1, \ldots, X_n\right) = \frac{m^{(p+1)}(x_0)}{(p+1)!} h^{p+1} \mu_{p+1}(K) + o(h^{p+1}), \tag{2.6.21}$$

dengan $\mu_j(K) = \int u^j K(u)\, du$ *adalah momen ke-*j *dari* kernel K.

Khususnya, untuk estimator linear lokal ($p = 1$) *dengan* kernel *simetris, berlaku*

$$\text{Bias}\,(\hat{m}(x_0)) = \frac{1}{2} m''(x_0) h^2 \mu_2(K) + o(h^2). \tag{2.6.22}$$

Bukti. Untuk estimator linear lokal ($p = 1$), ekspansi Taylor dari $m(X_i)$ di sekitar x_0 adalah

$$m(X_i) = m(x_0) + m'(x_0)(X_i - x_0) + \frac{1}{2} m''(x_0)(X_i - x_0)^2 + O\left((X_i - x_0)^3\right). \tag{2.6.23}$$

Substitusi ke dalam ekspektasi estimator dan menggunakan sifat *kernel* simetris memberikan

$$\mathbb{E}[\hat{m}(x_0)] = m(x_0) + \frac{1}{2} m''(x_0) \mathbb{E}\left[\sum_{i=1}^{n} w_i(x_0)(X_i - x_0)^2\right] + O(h^3). \tag{2.6.24}$$

Dengan analisis asimtotik dan fakta bahwa $\sum_{i=1}^{n} w_i(x_0)(X_i - x_0)^2 \approx h^2 \mu_2(K)$ untuk n besar dan h kecil, diperoleh hasil yang diinginkan. ■

Teorema 2.6.9 (Variansi Estimator Polinomial Lokal) *Di bawah asumsi regularitas (A1)–(A5), variansi bersyarat dari estimator polinomial lokal orde-*p *adalah*

$$\text{Var}\,(\hat{m}(x_0) \mid X_1, \ldots, X_n) = \frac{\sigma^2(x_0) R(K)}{n h f(x_0)} + o\left(\frac{1}{nh}\right), \tag{2.6.25}$$

dengan $R(K) = \int K^2(u)\, du$ *adalah kekasaran (*roughness*) dari* kernel K.

Bukti. Variansi estimator diberikan oleh

$$\text{Var}\,(\hat{m}(x_0) \mid \boldsymbol{X}) = \sum_{i=1}^{n} w_i^2(x_0) \text{Var}(Y_i \mid X_i) = \sum_{i=1}^{n} w_i^2(x_0) \sigma^2(X_i). \tag{2.6.26}$$

Dengan analisis asimtotik dan menggunakan aproksimasi integral Riemann, dapat ditunjukkan bahwa

$$\sum_{i=1}^{n} w_i^2(x_0)\sigma^2(X_i) \approx \frac{\sigma^2(x_0)}{n^2h^2f^2(x_0)} \sum_{i=1}^{n} K^2\left(\frac{X_i - x_0}{h}\right)$$
$$\approx \frac{\sigma^2(x_0)}{nhf(x_0)} \int K^2(u)\,du = \frac{\sigma^2(x_0)R(K)}{nhf(x_0)}. \tag{2.6.27}$$

■

Teorema 2.6.10 (Rataan Kuadrat Galat (*Mean Squared Error*, MSE)) *MSE estimator polinomial lokal orde-p di titik x_0 adalah*

$$\mathrm{MSE}\left(\hat{m}(x_0)\right) = \mathrm{Bias}^2\left(\hat{m}(x_0)\right) + \mathrm{Var}\left(\hat{m}(x_0)\right). \tag{2.6.28}$$

Untuk estimator linear lokal ($p = 1$), MSE asimtotik diberikan oleh

$$\mathrm{MSE}\left(\hat{m}(x_0)\right) = \frac{1}{4}\left(m''(x_0)\right)^2 h^4\mu_2^2(K) + \frac{\sigma^2(x_0)R(K)}{nhf(x_0)} + o\left(h^4 + \frac{1}{nh}\right). \tag{2.6.29}$$

Akibat 2.6.11 (*Bandwidth* Optimal) Bandwidth *optimal yang meminimumkan MSE asimtotik untuk estimator linear lokal diperoleh dengan menyeimbangkan bias dan variansi. Dengan mengambil turunan MSE terhadap h dan menyamakan dengan nol, diperoleh*

$$h_{\mathrm{opt}} = \left(\frac{\sigma^2(x_0)R(K)}{nf(x_0)\left(m''(x_0)\right)^2\mu_2^2(K)}\right)^{1/5} = O\left(n^{-1/5}\right). \tag{2.6.30}$$

Dengan bandwidth *optimal ini, MSE optimal adalah*

$$\mathrm{MSE}_{\mathrm{opt}}\left(\hat{m}(x_0)\right) = O\left(n^{-4/5}\right). \tag{2.6.31}$$

2.6.3 Teori Asimtotik Estimator Polinomial Lokal

Teorema 2.6.12 (**Konsistensi Estimator Polinomial Lokal**) *Di bawah asumsi regularitas (A1)–(A5), estimator polinomial lokal $\hat{m}(x_0)$ bersifat konsisten, yaitu*

$$\hat{m}(x_0) \xrightarrow{p} m(x_0) \quad \textit{ketika } n \to \infty. \tag{2.6.32}$$

Bukti. Dari hasil bias dan variansi, berlaku

$$\operatorname{Bias}(\hat{m}(x_0)) = O(h^{p+1}) \to 0 \quad \text{sebab } h \to 0, \tag{2.6.33}$$

$$\operatorname{Var}(\hat{m}(x_0)) = O\left(\frac{1}{nh}\right) \to 0 \quad \text{sebab } nh \to \infty. \tag{2.6.34}$$

Dengan demikian, MSE($\hat{m}(x_0)$) $\to 0$, yang mengimplikasikan konvergensi dalam probabilitas. ■

Teorema 2.6.13 (**Normalitas Asimtotik Estimator Polinomial Lokal**) *Di bawah asumsi regularitas (A1)–(A5) dan dengan asumsi tambahan bahwa galat memiliki momen keempat terbatas, estimator polinomial lokal memenuhi*

$$\sqrt{nh}\left(\hat{m}(x_0) - m(x_0) - \operatorname{Bias}(\hat{m}(x_0))\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2(x_0)R(K)}{f(x_0)}\right). \tag{2.6.35}$$

Secara ekuivalen, dengan mendefinisikan bias asimtotik $b(x_0) = \frac{m^{(p+1)}(x_0)}{(p+1)!}h^{p+1}\mu_{p+1}(K)$, berlaku

$$\sqrt{nh}\left(\hat{m}(x_0) - m(x_0) - b(x_0)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2(x_0)R(K)}{f(x_0)}\right). \tag{2.6.36}$$

Bukti. Estimator $\hat{m}(x_0)$ dapat dituliskan sebagai

$$\hat{m}(x_0) = \sum_{i=1}^{n} w_i(x_0)Y_i = \sum_{i=1}^{n} w_i(x_0)\left(m(X_i) + \varepsilon_i\right). \tag{2.6.37}$$

Dengan memisahkan komponen deterministik dan stokastik, diperoleh

$$\hat{m}(x_0) - m(x_0) = \underbrace{\sum_{i=1}^{n} w_i(x_0)\left(m(X_i) - m(x_0)\right)}_{\text{bias}} + \underbrace{\sum_{i=1}^{n} w_i(x_0)\varepsilon_i}_{\text{komponen stokastik}} . \tag{2.6.38}$$

Komponen stokastik $\sum_{i=1}^{n} w_i(x_0)\varepsilon_i$ adalah kombinasi linear dari variabel acak independen dengan rataan nol. Dengan Teorema Lindeberg–Feller, dapat ditunjukkan bahwa setelah normalisasi dengan $\sqrt{nh}$, komponen ini konvergen dalam distribusi ke normal.

Secara spesifik, definisikan $Z_n = \sqrt{nh}\sum_{i=1}^{n} w_i(x_0)\varepsilon_i$. Variansi dari Z_n adalah

$$\mathrm{Var}(Z_n) = nh\sum_{i=1}^{n} w_i^2(x_0)\sigma^2(X_i) \to \frac{\sigma^2(x_0)R(K)}{f(x_0)}. \tag{2.6.39}$$

Kondisi Lindeberg dapat diverifikasi dengan menggunakan fakta bahwa bobot $w_i(x_0)$ mengecil secara seragam dan momen keempat galat terbatas. Dengan demikian, Teorema Lindeberg–Feller menjamin

$$Z_n \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2(x_0)R(K)}{f(x_0)}\right). \tag{2.6.40}$$

■

Teorema 2.6.14 (Interval Konfidensi Asimtotik) *Interval konfidensi asimtotik tingkat* $(1-\alpha)\times 100\%$ *untuk* $m(x_0)$ *diberikan oleh*

$$\hat{m}(x_0) \pm z_{\alpha/2}\sqrt{\frac{\hat{\sigma}^2(x_0)R(K)}{nh\hat{f}(x_0)}}, \tag{2.6.41}$$

dengan $z_{\alpha/2}$ *adalah kuantil* $(1-\alpha/2)$ *dari distribusi normal standar,* $\hat{\sigma}^2(x_0)$ *adalah estimator konsisten untuk* $\sigma^2(x_0)$*, dan* $\hat{f}(x_0)$ *adalah estimator kepadatan* kernel *untuk* $f(x_0)$*.*

Interval ini valid secara asimtotik jika bias diabaikan, yang terjadi ketika $nh^{2(p+1)+1} \to 0$ *(*undersmoothing*) atau ketika koreksi bias eksplisit diterapkan.*

Proposisi 2.6.15 (**Pemilihan *Bandwidth* dengan *Cross-Validation***) Bandwidth *optimal dapat dipilih secara adaptif menggunakan* leave-one-out cross-validation *dengan meminimumkan*

$$\mathrm{CV}(h) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{m}_{-i}(X_i; h)\right)^2, \tag{2.6.42}$$

dengan $\hat{m}_{-i}(X_i; h)$ *adalah estimator* leave-one-out *yang dihitung tanpa observasi ke-*i*. Secara asimtotik,* bandwidth *yang dipilih dengan* cross-validation *mendekati* bandwidth *optimal yang meminimumkan* integrated mean squared error *(IMSE).*

2.7 Analisis Regresi Spasial

Analisis regresi spasial merupakan perluasan dari analisis regresi klasik yang secara eksplisit memperhitungkan aspek geografis atau lokasi dalam pemodelan. Terdapat dua fenomena utama dalam data spasial yang perlu diperhatikan, yaitu dependensi spasial (*spatial dependence*) dan keragaman spasial (*spatial heterogeneity*). Kedua konsep ini memiliki karakteristik yang berbeda dan memerlukan pendekatan pemodelan yang berbeda pula (**???**).

Dependensi spasial merujuk pada situasi ketika nilai suatu variabel di suatu lokasi berkorelasi dengan nilai variabel yang sama di lokasi lain yang berdekatan secara geografis. Dengan kata lain, observasi pada lokasi yang berdekatan cenderung memiliki nilai yang serupa (autokorelasi spasial positif) atau berlawanan (autokorelasi spasial negatif). Fenomena ini dapat dianalogikan dengan autokorelasi serial dalam data deret waktu, tetapi terjadi dalam dimensi ruang. Dependensi spasial melanggar asumsi independensi galat dalam regresi linear klasik, yaitu $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) \neq 0$ untuk lokasi i dan j yang berdekatan.

Keragaman spasial atau heterogenitas spasial merujuk pada situasi ketika hubungan struktural antara variabel dependen dan independen bervariasi di berbagai lokasi geografis. Dalam konteks regresi, hal ini berarti koefisien regresi $\boldsymbol{\beta}$ tidak konstan di seluruh wilayah pengamatan, melainkan merupakan fungsi dari lokasi geografis, yaitu $\boldsymbol{\beta} = \boldsymbol{\beta}(s)$ dengan s menyatakan lokasi. Fenomena ini mengindikasikan bahwa proses yang mendasari hubungan antarvariabel berbeda di setiap wilayah.

Tabel 2.1 Perbandingan Dependensi Spasial dan Keragaman Spasial

Aspek	**Dependensi Spasial**	**Keragaman Spasial**
Definisi	Korelasi antarlokasi pada variabel yang sama	Variasi hubungan struktural antarlokasi
Fokus	Hubungan antar-observasi	Hubungan antar-variabel
Manifestasi	$\mathrm{Cov}(Y_i, Y_j) \neq 0$ untuk lokasi berdekatan	$\boldsymbol{\beta}_i \neq \boldsymbol{\beta}_j$ untuk lokasi berbeda
Asumsi yang dilanggar	Independensi galat	Homogenitas parameter
Model penanganan	SAR, SEM, SDM	GWR, GTWR
Contoh	Harga rumah di lokasi berdekatan saling memengaruhi	Pengaruh pendidikan terhadap pendapatan berbeda di kota dan desa

Kedua fenomena ini dapat terjadi secara bersamaan dalam data spasial. Sebagai ilustrasi, pada analisis harga properti, dependensi spasial terjadi ketika harga rumah di suatu lokasi dipengaruhi oleh harga rumah di sekitarnya (efek *spillover*), sedangkan keragaman spasial terjadi ketika faktor-faktor yang memengaruhi harga rumah (seperti luas tanah atau jarak ke pusat kota) memiliki pengaruh yang berbeda di wilayah perkotaan dibandingkan dengan wilayah pinggiran.

2.7.1 Regresi dengan Dependensi Spasial

Dependensi spasial (*spatial dependence*) atau autokorelasi spasial terjadi ketika nilai suatu variabel di suatu lokasi berkorelasi dengan nilai variabel yang sama di lokasi lain yang berdekatan secara geografis. Konsep ini melanggar asumsi independensi dalam regresi linear klasik dan memerlukan perlakuan khusus dalam pemodelan ekonometrika spasial (**??**).

Definisi 2.7.1 (**Dependensi Spasial**) *Misalkan Y_i merupakan nilai variabel di lo-*

kasi i dan Y_j merupakan nilai variabel di lokasi tetangga j. Dependensi spasial terjadi jika

$$\mathrm{Cov}(Y_i, Y_j) \neq 0, \quad \textit{untuk lokasi } i \neq j \textit{ yang berdekatan,} \tag{2.7.1}$$

yang dapat diukur menggunakan statistik Moran's I, yaitu

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \cdot \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}, \tag{2.7.2}$$

dengan w_{ij} adalah bobot spasial antara lokasi i dan j.

Model regresi dengan dependensi spasial dapat diklasifikasikan menjadi beberapa bentuk utama berdasarkan letak dependensi spasial muncul dalam spesifikasi model.

Definisi 2.7.2 (***Spatial Autoregressive Model*** **(SAR)**) *Model SAR atau* spatial lag model *mengasumsikan bahwa variabel dependen di suatu lokasi dipengaruhi oleh nilai variabel dependen di lokasi tetangga, yaitu*

$$y_i = \rho \sum_{j=1}^{n} w_{ij} y_j + \sum_{k=1}^{p} \beta_k x_{ik} + \varepsilon_i, \tag{2.7.3}$$

atau dalam bentuk matriksnya adalah

$$\boldsymbol{y} = \rho \mathbf{W} \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.7.4}$$

dengan ρ adalah parameter autokorelasi spasial dan $\mathbf{W}$ adalah matriks bobot spasial.

Definisi 2.7.3 (***Spatial Error Model*** **(SEM)**) *Model SEM mengasumsikan bahwa dependensi spasial terjadi pada komponen galat, yaitu*

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \quad \boldsymbol{u} = \lambda \mathbf{W} \boldsymbol{u} + \boldsymbol{\varepsilon}, \tag{2.7.5}$$

dengan λ adalah parameter autokorelasi spasial pada galat dan $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.

Definisi 2.7.4 (***Spatial Durbin Model*** **(SDM)**) *Model SDM menggabungkan* lag *spasial dari variabel dependen dan independen, yaitu*

$$\boldsymbol{y} = \rho \mathbf{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{X}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \tag{2.7.6}$$

dengan $\boldsymbol{\delta}$ adalah vektor parameter untuk lag spasial variabel independen.

2.7.2 Regresi dengan Keragaman Spasial

Heterogenitas spasial (*spatial heterogeneity*) terjadi ketika hubungan statistik antara variabel berbeda di berbagai lokasi geografis. Berbeda dengan dependensi spasial yang menangkap korelasi antar lokasi, heterogenitas spasial menangkap variasi dalam struktur hubungan itu sendiri (**?**).

Pemodelan polinomial lokal yang telah dibahas pada bagian sebelumnya menyediakan kerangka teoretis fundamental untuk memahami *Geographically Weighted Regression* (GWR). Dalam pemodelan polinomial lokal, fungsi regresi $m(x)$ diestimasi secara lokal pada titik target x_0 dengan mengaproksimasi fungsi tersebut menggunakan ekspansi Taylor dan memberikan bobot yang lebih besar pada observasi yang dekat dengan titik target. GWR merupakan perluasan natural dari konsep ini ke dalam konteks spasial dua dimensi. Jika dalam pemodelan polinomial lokal standar, kedekatan diukur dalam ruang variabel prediktor satu dimensi, maka dalam GWR, kedekatan diukur dalam ruang geografis dua dimensi yang didefinisikan oleh koordinat (u, v). Dengan demikian, GWR dapat dipandang sebagai kasus khusus dari model koefisien bervariasi (*varying coefficient model*) dengan lokasi geografis sebagai variabel pengindeks variasi koefisien.

Secara lebih spesifik, hubungan antara keduanya dapat dilihat dari beberapa aspek berikut.

(a) Pemodelan polinomial lokal mengaproksimasi $m(x) \approx \sum_{j=0}^{p} \beta_j(x_0)(x - x_0)^j$ di sekitar x_0, sedangkan GWR mengaproksimasi $\boldsymbol{\beta}(u, v)$ secara lokal di sekitar setiap lokasi geografis.

(b) Keduanya menggunakan fungsi *kernel* untuk memberikan bobot berdasarkan kedekatan, yaitu dalam polinomial lokal kedekatan di ruang prediktor dan dalam GWR kedekatan di ruang geografis.

(c) Keduanya menggunakan *locally weighted least squares* sebagai metode estimasi parameter lokal.

(d) Parameter *bandwidth* h mengontrol tingkat penghalusan (*smoothing*) dan menentukan trade-off antara bias dan variansi pada kedua pendekatan.

Heterogenitas spasial terjadi ketika parameter regresi $\boldsymbol{\beta}$ bervariasi menurut lokasi geografis, yaitu

$$\boldsymbol{\beta} = \boldsymbol{\beta}(s_i) = \boldsymbol{\beta}(u_i, v_i), \tag{2.7.7}$$

dengan (u_i, v_i) adalah koordinat geografis lokasi ke-i. Hal ini mengindikasikan bahwa hubungan struktural yang sama dapat memiliki kekuatan dan arah yang berbeda di lokasi yang berbeda. Salah satu model regresi yang memperhitungkan heterogenitas spasial adalah *Geographically Weighted Regression* (GWR) atau regresi terboboti geografis yang memungkinkan koefisien regresi bervariasi secara kontinu di ruang geografis.

Definisi 2.7.5 (***Geographically Weighted Regression*** **(GWR)**) *Model GWR untuk observasi ke-i pada lokasi (u_i, v_i) didefinisikan sebagai*

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{2.7.8}$$

atau dalam bentuk vektor

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}(u_i, v_i) + \varepsilon_i, \tag{2.7.9}$$

dengan:

- y_i *adalah variabel respons pada lokasi* i*,*
- $\boldsymbol{x}_i = (1, x_{i1}, \ldots, x_{ip})^\top \in \mathbb{R}^{p+1}$ *adalah vektor variabel prediktor,*

- $\boldsymbol{\beta}(u_i, v_i) = (\beta_0(u_i, v_i), \beta_1(u_i, v_i), \ldots, \beta_p(u_i, v_i))^\top$ *adalah vektor koefisien lokal,*

- (u_i, v_i) *adalah koordinat geografis lokasi ke-i, dan*

- ε_i *adalah galat acak.*

Model GWR didasarkan pada asumsi-asumsi berikut:

(G1) Fungsi koefisien $\beta_k(u, v)$ untuk $k = 0, 1, \ldots, p$ memiliki turunan parsial kontinu hingga orde kedua di setiap lokasi.

(G2) $\mathbb{E}[\varepsilon_i \mid \boldsymbol{X}, (u_i, v_i)] = 0$ untuk semua i.

(G3) $\mathrm{Var}(\varepsilon_i \mid \boldsymbol{X}, (u_i, v_i)) = \sigma^2$ konstan untuk semua i.

(G4) $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ untuk $i \neq j$.

(G5) Lokasi observasi memiliki kepadatan $f(u, v) > 0$ di sekitar setiap titik target.

(G6) Untuk setiap lokasi i, matriks $\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X}$ adalah nonsingular.

Teorema 2.7.6 (Estimator LWLS untuk GWR) *Untuk setiap lokasi (u_i, v_i), estimator* locally weighted least squares *(LWLS) untuk koefisien lokal diperoleh dengan meminimalkan*

$$S_i(\boldsymbol{\beta}) = \sum_{j=1}^{n} w_{ij} \left(y_j - \boldsymbol{x}_j^\top \boldsymbol{\beta}\right)^2 = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \mathbf{W}_i (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{2.7.10}$$

dengan $\mathbf{W}_i = \mathrm{diag}(w_{i1}, w_{i2}, \ldots, w_{in})$. *Estimator yang dihasilkan adalah*

$$\widehat{\boldsymbol{\beta}}(u_i, v_i) = (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{y}. \tag{2.7.11}$$

Bukti. Untuk lokasi (u_i, v_i), tujuannya adalah meminimalkan fungsi objektif kuadrat terkecil tertimbang lokal

$$S_i(\boldsymbol{\beta}) = \sum_{j=1}^{n} w_{ij} \left(y_j - \boldsymbol{x}_j^\top \boldsymbol{\beta}\right)^2. \tag{2.7.12}$$

Dalam notasi matriks, fungsi ini dapat dituliskan sebagai

$$S_i(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \mathbf{W}_i(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{2.7.13}$$

dengan $\mathbf{W}_i = \mathrm{diag}(w_{i1}, w_{i2}, \ldots, w_{in})$ adalah matriks diagonal berukuran $n \times n$ yang berisi bobot spasial.

Ekspansi bentuk kuadratik memberikan

$$\begin{aligned} S_i(\boldsymbol{\beta}) &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top \mathbf{W}_i(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= \boldsymbol{y}^\top \mathbf{W}_i \boldsymbol{y} - \boldsymbol{y}^\top \mathbf{W}_i \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{\beta}^\top \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{y} + \boldsymbol{\beta}^\top \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X}\boldsymbol{\beta}. \end{aligned} \tag{2.7.14}$$

Sebab $\boldsymbol{y}^\top \mathbf{W}_i \boldsymbol{X}\boldsymbol{\beta}$ adalah skalar, maka $\boldsymbol{y}^\top \mathbf{W}_i \boldsymbol{X}\boldsymbol{\beta} = (\boldsymbol{y}^\top \mathbf{W}_i \boldsymbol{X}\boldsymbol{\beta})^\top = \boldsymbol{\beta}^\top \boldsymbol{X}^\top \mathbf{W}_i^\top \boldsymbol{y} = \boldsymbol{\beta}^\top \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{y}$ (sebab $\mathbf{W}_i$ simetris). Dengan demikian,

$$S_i(\boldsymbol{\beta}) = \boldsymbol{y}^\top \mathbf{W}_i \boldsymbol{y} - 2\boldsymbol{\beta}^\top \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{y} + \boldsymbol{\beta}^\top \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X}\boldsymbol{\beta}. \tag{2.7.15}$$

Untuk menemukan minimum, dihitung turunan pertama terhadap $\boldsymbol{\beta}$ dan disamakan dengan nol. Dengan menggunakan aturan diferensial matriks, turunan dari masing-masing suku adalah sebagai berikut.

- Suku $\boldsymbol{y}^\top \mathbf{W}_i \boldsymbol{y}$ tidak bergantung pada $\boldsymbol{\beta}$, sehingga turunannya adalah $\mathbf{0}$.
- Suku $-2\boldsymbol{\beta}^\top \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{y}$ berbentuk $\boldsymbol{c}^\top \boldsymbol{\beta}$ dengan $\boldsymbol{c} = -2\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{y}$. Turunannya adalah $\boldsymbol{c} = -2\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{y}$.
- Suku $\boldsymbol{\beta}^\top \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X}\boldsymbol{\beta}$ berbentuk $\boldsymbol{\beta}^\top \boldsymbol{A}\boldsymbol{\beta}$ dengan $\boldsymbol{A} = \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X}$ yang simetris (sebab $\mathbf{W}_i$ simetris). Turunannya adalah $2\boldsymbol{A}\boldsymbol{\beta} = 2\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X}\boldsymbol{\beta}$.

Dengan demikian, gradien dari $S_i(\boldsymbol{\beta})$ terhadap $\boldsymbol{\beta}$ adalah

$$\frac{\partial S_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{y} + 2\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X}\boldsymbol{\beta}. \tag{2.7.16}$$

Syarat orde pertama untuk minimum adalah gradien sama dengan nol:

$$\begin{aligned}-2\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{y} + 2\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X}\widehat{\boldsymbol{\beta}} &= \mathbf{0} \\ \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X}\widehat{\boldsymbol{\beta}} &= \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{y}. \end{aligned} \tag{2.7.17}$$

Persamaan terakhir merupakan persamaan normal tertimbang untuk lokasi i.

Dengan asumsi bahwa matriks $\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X}$ adalah nonsingular (invertibel), solusinya adalah

$$\widehat{\boldsymbol{\beta}}(u_i, v_i) = (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{y}. \tag{2.7.18}$$

Untuk memverifikasi bahwa solusi ini adalah peminimum (bukan pemaksimum atau titik pelana), diperiksa turunan kedua (matriks Hessian):

$$\frac{\partial^2 S_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = 2\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X}. \tag{2.7.19}$$

Matriks $\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X}$ adalah semi-definit positif sebab untuk sembarang vektor $\boldsymbol{v} \in \mathbb{R}^{p+1}$,

$$\boldsymbol{v}^\top \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X} \boldsymbol{v} = (\boldsymbol{X}\boldsymbol{v})^\top \mathbf{W}_i (\boldsymbol{X}\boldsymbol{v}) = \sum_{j=1}^{n} w_{ij} (\boldsymbol{x}_j^\top \boldsymbol{v})^2 \geq 0, \tag{2.7.20}$$

sebab $w_{ij} \geq 0$ untuk semua j. Jika $\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X}$ nonsingular, maka matriks ini definit positif, sehingga Hessian definit positif dan $\widehat{\boldsymbol{\beta}}(u_i, v_i)$ adalah peminimum global yang unik. ■

Akibat 2.7.7 *Nilai prediksi pada lokasi i dapat dituliskan sebagai kombinasi linear dari respons:*

$$\hat{y}_i = \boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}}(u_i, v_i) = \boldsymbol{r}_i^\top \boldsymbol{y}, \tag{2.7.21}$$

dengan $\boldsymbol{r}_i^\top = \boldsymbol{x}_i^\top (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i$. Secara agregat, vektor prediksi $\hat{\boldsymbol{y}} = \mathbf{S}\boldsymbol{y}$, *dengan matriks* hat *GWR*

$$\mathbf{S} = \begin{pmatrix} \boldsymbol{r}_1^\top \\ \boldsymbol{r}_2^\top \\ \vdots \\ \boldsymbol{r}_n^\top \end{pmatrix}. \tag{2.7.22}$$

Matriks bobot dibangun menggunakan fungsi *kernel* yang mengukur kedekatan spasial antara lokasi. Fungsi *kernel* ini menentukan bobot w_{ij} yang diberikan pada observasi j saat mengestimasi koefisien di lokasi i. Bobot ini biasanya menurun seiring bertambahnya jarak geografis antara lokasi i dan j. Fungsi *kernel* juga melibatkan parameter *bandwidth* h yang mengontrol tingkat penghalusan (*smoothing*) dari estimasi lokal.

Definisi 2.7.8 (Fungsi *Kernel* untuk Pembobotan Spasial) *Fungsi* kernel $K_h : \mathbb{R}^+ \to \mathbb{R}^+$ *memetakan jarak geografis ke bobot non-negatif. Untuk jarak* d_{ij} *antara lokasi* i *dan* j*, bobot didefinisikan sebagai* $w_{ij} = K_h(d_{ij})$*. Syarat umum untuk fungsi* kernel *sesuai dengan kerangka pemodelan polinomial lokal adalah sebagai berikut.*

(a) $K_h(\cdot)$ *adalah fungsi non-negatif, yaitu* $K_h(d) \geq 0$ *untuk semua* $d \geq 0$.

(b) $K_h(\cdot)$ *adalah fungsi menurun, yaitu jika* $d_1 < d_2$*, maka* $K_h(d_1) \geq K_h(d_2)$.

(c) $K_h(\cdot)$ *memiliki dukungan terbatas, yaitu terdapat konstanta* $C > 0$ *sehingga* $K_h(d) = 0$ *untuk semua* $d > Ch$.

(d) $K_h(\cdot)$ *memenuhi kondisi normalisasi, yaitu* $\int_0^\infty K_h(d) f_D(d)\, dd = 1$*, dengan* $f_D(d)$ *adalah fungsi kepadatan jarak antar lokasi.*

Beberapa fungsi *kernel* yang umum digunakan adalah sebagai berikut.

(a) ***Kernel* Gaussian:**

$$w_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right). \tag{2.7.23}$$

(b) ***Kernel* Bi-square (kompak):**

$$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h}\right)^2\right)^2 & \text{jika } d_{ij} < h, \\ 0 & \text{jika } d_{ij} \geq h. \end{cases} \tag{2.7.24}$$

(c) ***Kernel* Tri-cube (kompak):**

$$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h}\right)^3\right)^3 & \text{jika } d_{ij} < h, \\ 0 & \text{jika } d_{ij} \geq h. \end{cases} \tag{2.7.25}$$

Perhitungan jarak geografis d_{ij} antara lokasi i dan j biasanya dilakukan menggunakan beberapa pendekatan, seperti jarak Euclidean, jarak Manhattan, atau jarak geodesik (*great circle distance*). Pilihan metrik jarak ini bergantung pada konteks geografis dan karakteristik data yang dianalisis.

Definisi 2.7.9 (Perhitungan Jarak Geografis) *Misalkan lokasi i memiliki koordinat geografis (u_i, v_i) dan lokasi j memiliki koordinat (u_j, v_j). Jarak geografis d_{ij} antara lokasi i dan j dihitung menggunakan metrik tertentu, seperti:*

(a) ***Jarak Euclidean:***

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}. \tag{2.7.26}$$

(b) ***Jarak Manhattan:***

$$d_{ij} = |u_i - u_j| + |v_i - v_j|. \tag{2.7.27}$$

(c) ***Jarak Geodesik (*****Great Circle Distance*****):***

$$d_{ij} = R \cdot \arccos\left(\sin(\phi_i)\sin(\phi_j) + \cos(\phi_i)\cos(\phi_j)\cos(\lambda_i - \lambda_j)\right), \tag{2.7.28}$$

dengan R adalah jari-jari bumi, (ϕ_i, λ_i) dan (ϕ_j, λ_j) adalah lintang dan bujur lokasi i dan j dalam radian.

Dalam GWR, pemilihan *bandwidth* h sangat penting karena mempengaruhi bias dan variansi dari estimator lokal. *Bandwidth* yang kecil menghasilkan estimasi yang lebih fleksibel namun dengan variansi yang tinggi, sedangkan *bandwidth* yang besar menghasilkan estimasi yang lebih halus namun dengan bias yang lebih besar. Oleh karena itu, pemilihan *bandwidth* yang optimal merupakan langkah krusial dalam penerapan GWR.

Terdapat dua pendekatan utama dalam menentukan *bandwidth*:

(a) ***Fixed bandwidth*:** Parameter h bernilai konstan untuk semua lokasi. Pendekatan ini sesuai ketika observasi terdistribusi secara relatif homogen di ruang geografis.

(b) ***Adaptive bandwidth*:** Parameter h_i bervariasi menurut lokasi, biasanya didefinisikan sehingga setiap lokasi memiliki jumlah tetangga efektif yang sama (misalnya, k tetangga terdekat). Secara formal, h_i dipilih sehingga

$$h_i = d_{i[k]}, \tag{2.7.29}$$

dengan $d_{i[k]}$ adalah jarak ke tetangga terdekat ke-k dari lokasi i. Pendekatan ini lebih sesuai ketika kepadatan observasi bervariasi secara spasial.

Definisi 2.7.10 (***Cross-Validation* untuk *Bandwidth***) Bandwidth *optimal* h^* *dipilih dengan meminimalkan fungsi* leave-one-out cross-validation*:*

$$\mathrm{CV}(h) = \sum_{i=1}^{n} \left(y_i - \hat{y}_{\neq i}(h)\right)^2, \tag{2.7.30}$$

dengan $\hat{y}_{\neq i}(h)$ *adalah prediksi untuk observasi* i *menggunakan model yang di-*fit *tanpa observasi* i*. Secara eksplisit,*

$$h^* = \arg\min_h \mathrm{CV}(h). \tag{2.7.31}$$

Definisi 2.7.11 (**Kriteria AIC untuk Pemilihan *Bandwidth***) *Alternatif lain adalah* *menggunakan Akaike Information Criterion terkoreksi:*

$$\mathrm{AIC}_c(h) = 2n\log(\hat{\sigma}(h)) + n\log(2\pi) + n\left(\frac{n + \mathrm{tr}(\mathbf{S}(h))}{n - 2 - \mathrm{tr}(\mathbf{S}(h))}\right), \qquad (2.7.32)$$

dengan $\hat{\sigma}^2(h) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i(h))^2$ *dan* $\mathrm{tr}(\mathbf{S}(h))$ *adalah derajat bebas efektif model.*

Ketika *bandwidth* h diestimasi dari data (bukan ditetapkan secara eksogen), terdapat implikasi penting terhadap teori asimtotik, yaitu:

(a) Estimator $\hat{h}$ bergantung pada $\boldsymbol{y}$, sehingga matriks bobot $\mathbf{W}_i(\hat{h})$ menjadi stokastik dan bergantung pada data. Hal ini memperumit derivasi distribusi asimtotik.

(b) Ketidakpastian dalam $\hat{h}$ menambah variabilitas pada estimator $\widehat{\boldsymbol{\beta}}$, yang tidak tercermin dalam formula variansi standar.

(c) Dalam praktik, sering diasumsikan bahwa $\hat{h} \xrightarrow{p} h_{\text{opt}}$ dengan laju yang cukup cepat sehingga efek estimasi *bandwidth* dapat diabaikan secara asimtotik (*oracle property*). Hasil asimtotik kemudian diturunkan dengan menganggap h sebagai parameter tetap.

Teorema 2.7.12 *Di bawah asumsi (G1)–(G6), estimator GWR memiliki bias yang bergantung pada tingkat kehalusan koefisien dan* bandwidth*:*

$$\mathbb{E}[\widehat{\boldsymbol{\beta}}(u_i, v_i)] = \boldsymbol{\beta}(u_i, v_i) + O(h^2), \qquad (2.7.33)$$

dengan suku bias $O(h^2)$ *proporsional dengan turunan kedua dari* $\boldsymbol{\beta}(u, v)$ *di lokasi* i*.*

Bukti. Berdasarkan asumsi (G1), fungsi koefisien $\boldsymbol{\beta}(u, v)$ memiliki turunan parsial kontinu hingga orde kedua. Didefinisikan vektor perpindahan $\boldsymbol{\delta}_{ij} = (u_j - u_i, v_j - v_i)^\top$ dengan norma $d_{ij} = \|\boldsymbol{\delta}_{ij}\|$. Dengan ekspansi Taylor di sekitar lokasi (u_i, v_i), untuk observasi j dengan lokasi (u_j, v_j), diperoleh

$$\boldsymbol{\beta}(u_j, v_j) = \boldsymbol{\beta}(u_i, v_i) + \nabla\boldsymbol{\beta}(u_i, v_i)^\top \boldsymbol{\delta}_{ij} + \frac{1}{2}\boldsymbol{\delta}_{ij}^\top \mathbf{H}_{\boldsymbol{\beta}}(u_i, v_i)\boldsymbol{\delta}_{ij} + R_{ij}, \quad (2.7.34)$$

dengan $\nabla\boldsymbol{\beta}$ adalah gradien dan $\mathbf{H}_{\boldsymbol{\beta}}$ adalah matriks Hessian dari $\boldsymbol{\beta}$.

Suku sisa $R_{ij} = o(d_{ij}^2)$ karena berdasarkan teorema Taylor dengan sisa Peano, untuk fungsi $\boldsymbol{\beta} : \mathbb{R}^2 \to \mathbb{R}^{p+1}$ yang memiliki turunan parsial kontinu hingga orde kedua di sekitar (u_i, v_i), suku sisa setelah ekspansi orde kedua memenuhi

$$\lim_{d_{ij}\to 0} \frac{\|R_{ij}\|}{d_{ij}^2} = 0, \quad (2.7.35)$$

yang merupakan definisi dari $R_{ij} = o(d_{ij}^2)$. Secara eksplisit, jika turunan orde ketiga ada dan terbatas, maka $R_{ij} = O(d_{ij}^3)$, yang tentu saja memenuhi $R_{ij} = o(d_{ij}^2)$.

Model GWR yang sebenarnya adalah $y_j = \boldsymbol{x}_j^\top \boldsymbol{\beta}(u_j, v_j) + \varepsilon_j$. Substitusi ekspansi Taylor memberikan

$$y_j = \boldsymbol{x}_j^\top \boldsymbol{\beta}(u_i, v_i) + \boldsymbol{x}_j^\top \boldsymbol{r}_{ij} + \varepsilon_j, \quad (2.7.36)$$

dengan $\boldsymbol{r}_{ij} = \nabla\boldsymbol{\beta}(u_i, v_i)^\top \boldsymbol{\delta}_{ij} + \frac{1}{2}\boldsymbol{\delta}_{ij}^\top \mathbf{H}_{\boldsymbol{\beta}}(u_i, v_i)\boldsymbol{\delta}_{ij} + R_{ij}$ adalah suku sisa deterministik yang mengandung turunan pertama dan kedua dari $\boldsymbol{\beta}$.

Nilai harapan dari estimator LWLS adalah

$$\begin{aligned}
\mathbb{E}[\widehat{\boldsymbol{\beta}}(u_i, v_i)] &= \mathbb{E}[(\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{y}] \\
&= (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i \mathbb{E}[\boldsymbol{y}] \\
&= (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i (\boldsymbol{X}\boldsymbol{\beta}(u_i, v_i) + \boldsymbol{X}\boldsymbol{r}_i), \quad (2.7.37)
\end{aligned}$$

dengan $\boldsymbol{r}_i = (\boldsymbol{r}_{i1}, \ldots, \boldsymbol{r}_{in})^\top$ adalah vektor suku sisa. Sebab $(\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X} = \mathbf{I}$, maka

$$\mathbb{E}[\widehat{\boldsymbol{\beta}}(u_i, v_i)] = \boldsymbol{\beta}(u_i, v_i) + (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X} \boldsymbol{r}_i. \quad (2.7.38)$$

Dengan menggunakan sifat fungsi *kernel* bahwa bobot w_{ij} terkonsentrasi pada observasi dengan jarak $d_{ij} = O(h)$ dari lokasi i, suku sisa $\boldsymbol{r}_{ij} = O(h^2)$ karena didominasi oleh suku kuadratik dalam ekspansi Taylor. Oleh karena itu, vektor bias adalah

$$\boldsymbol{b}_i = (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X} \boldsymbol{r}_i = O(h^2), \tag{2.7.39}$$

yang proporsional dengan turunan kedua dari $\boldsymbol{\beta}(u, v)$ di lokasi i. ■

Teorema 2.7.13 (Variansi Estimator GWR) *Di bawah asumsi (G1)–(G6), matriks kovarians bersyarat dari estimator GWR adalah*

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}(u_i, v_i) \mid \boldsymbol{X}) = \sigma^2 (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i^2 \boldsymbol{X} (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1}. \tag{2.7.40}$$

Dengan normalisasi yang tepat, variansi adalah $O((nh^2)^{-1})$.

Bukti. Estimator GWR dapat dituliskan sebagai

$$\widehat{\boldsymbol{\beta}}(u_i, v_i) = (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{y} = \mathbf{C}_i \boldsymbol{y}, \tag{2.7.41}$$

dengan $\mathbf{C}_i = (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i$.

Berdasarkan asumsi (G3), galat $\varepsilon_1, \ldots, \varepsilon_n$ bersifat independen dengan $\mathbb{E}[\varepsilon_j] = 0$ dan $\mathrm{Var}(\varepsilon_j) = \sigma^2$ untuk semua j. Dalam notasi matriks, $\mathrm{Var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \sigma^2 \mathbf{I}_n$.

Karena $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ dengan $\boldsymbol{X}$ diperlakukan sebagai tetap (bersyarat), maka

$$\begin{aligned} \mathrm{Var}(\widehat{\boldsymbol{\beta}}(u_i, v_i) \mid \boldsymbol{X}) &= \mathrm{Var}(\mathbf{C}_i \boldsymbol{y} \mid \boldsymbol{X}) \\ &= \mathbf{C}_i \mathrm{Var}(\boldsymbol{y} \mid \boldsymbol{X}) \mathbf{C}_i^\top \\ &= \mathbf{C}_i (\sigma^2 \mathbf{I}_n) \mathbf{C}_i^\top \\ &= \sigma^2 \mathbf{C}_i \mathbf{C}_i^\top. \end{aligned} \tag{2.7.42}$$

Substitusi $\mathbf{C}_i = (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i$ memberikan

$$\begin{aligned}\mathbf{C}_i \mathbf{C}_i^\top &= (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i \left[(\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i\right]^\top \\ &= (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i \mathbf{W}_i^\top \boldsymbol{X} (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-\top}. \end{aligned} \tag{2.7.43}$$

Sebab $\mathbf{W}_i$ adalah matriks diagonal simetris, maka $\mathbf{W}_i^\top = \mathbf{W}_i$ dan $\mathbf{W}_i \mathbf{W}_i = \mathbf{W}_i^2$. Demikian pula, $(\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-\top} = (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1}$ karena matriks ini simetris. Oleh karena itu,

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}(u_i, v_i) \mid \boldsymbol{X}) = \sigma^2 (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i^2 \boldsymbol{X} (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1}. \tag{2.7.44}$$

Untuk menunjukkan bahwa variansi adalah $O((nh^2)^{-1})$, perhatikan bahwa dengan *bandwidth* h, jumlah observasi efektif yang berkontribusi signifikan pada estimasi lokal adalah $n_{\text{eff}} \approx n \cdot \pi h^2 / A$, dengan A adalah luas wilayah studi. Dengan asumsi kepadatan observasi yang seragam, $n_{\text{eff}} = O(nh^2)$. Sebab variansi estimator berbanding terbalik dengan jumlah observasi efektif, maka $\mathrm{Var}(\widehat{\boldsymbol{\beta}}(u_i, v_i)) = O((nh^2)^{-1})$. ■

Teorema 2.7.14 (**Estimator Variansi**) *Estimator konsisten untuk σ^2 diberikan oleh*

$$\hat{\sigma}^2 = \frac{\boldsymbol{e}^\top \boldsymbol{e}}{n - 2\mathrm{tr}(\mathbf{S}) + \mathrm{tr}(\mathbf{S}^\top \mathbf{S})}, \tag{2.7.45}$$

dengan $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}}$ adalah vektor residual dan penyebut merepresentasikan derajat bebas efektif residual.

Bukti. Vektor nilai prediksi GWR dapat dituliskan sebagai $\hat{\boldsymbol{y}} = \mathbf{S}\boldsymbol{y}$, dengan S adalah matriks topi GWR yang baris ke-i-nya adalah $\boldsymbol{s}_i^\top = \boldsymbol{x}_i^\top (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i$. Vektor residual adalah

$$\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}} = (\mathbf{I} - \mathbf{S})\boldsymbol{y}. \tag{2.7.46}$$

Dengan model $\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}$ dengan $\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$ dan $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$, nilai harapan dari jumlah kuadrat residual adalah

$$\begin{aligned}\mathbb{E}[\boldsymbol{e}^\top \boldsymbol{e}] &= \mathbb{E}[\boldsymbol{y}^\top(\mathbf{I}-\mathbf{S})^\top(\mathbf{I}-\mathbf{S})\boldsymbol{y}] \\ &= \mathbb{E}[\mathrm{tr}((\mathbf{I}-\mathbf{S})\boldsymbol{y}\boldsymbol{y}^\top(\mathbf{I}-\mathbf{S})^\top)] \\ &= \mathrm{tr}((\mathbf{I}-\mathbf{S})\mathbb{E}[\boldsymbol{y}\boldsymbol{y}^\top](\mathbf{I}-\mathbf{S})^\top). \end{aligned} \tag{2.7.47}$$

Karena $\mathbb{E}[\boldsymbol{y}\boldsymbol{y}^\top] = \boldsymbol{\mu}\boldsymbol{\mu}^\top + \sigma^2\mathbf{I}$, dengan asumsi bahwa model GWR benar sehingga $(\mathbf{I}-\mathbf{S})\boldsymbol{\mu} \approx \mathbf{0}$ (mengabaikan bias), diperoleh

$$\begin{aligned}\mathbb{E}[\boldsymbol{e}^\top \boldsymbol{e}] &\approx \sigma^2\mathrm{tr}((\mathbf{I}-\mathbf{S})(\mathbf{I}-\mathbf{S})^\top) \\ &= \sigma^2\mathrm{tr}(\mathbf{I}-\mathbf{S}-\mathbf{S}^\top+\mathbf{S}^\top\mathbf{S}) \\ &= \sigma^2(n-\mathrm{tr}(\mathbf{S})-\mathrm{tr}(\mathbf{S}^\top)+\mathrm{tr}(\mathbf{S}^\top\mathbf{S})) \\ &= \sigma^2(n-2\mathrm{tr}(\mathbf{S})+\mathrm{tr}(\mathbf{S}^\top\mathbf{S})), \end{aligned} \tag{2.7.48}$$

dengan menggunakan sifat $\mathrm{tr}(\mathbf{S}^\top) = \mathrm{tr}(\mathbf{S})$.

Perlu dicatat bahwa matriks $\mathbf{S}$ dalam GWR umumnya tidak simetris dan tidak idempoten, berbeda dengan matriks *hat* dalam regresi OLS standar. Oleh karena itu, derajat bebas efektif residual bukan $n-p$, melainkan

$$\nu = n - 2\mathrm{tr}(\mathbf{S}) + \mathrm{tr}(\mathbf{S}^\top\mathbf{S}). \tag{2.7.49}$$

Dengan demikian, estimator tak bias untuk σ^2 adalah

$$\hat{\sigma}^2 = \frac{\boldsymbol{e}^\top\boldsymbol{e}}{n-2\mathrm{tr}(\mathbf{S})+\mathrm{tr}(\mathbf{S}^\top\mathbf{S})}, \tag{2.7.50}$$

yang memenuhi $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$. Konsistensi mengikuti dari hukum bilangan besar karena baik pembilang maupun penyebut konvergen ke nilai-nilai yang tepat ketika $n \to \infty$. ■

Berdasarkan teorema-teorema di atas, diperoleh hasil penting mengenai kondisi *bandwidth* untuk konsistensi estimator GWR, yaitu *trade-off* antara bias dan variansi.

Akibat 2.7.15 (**Kondisi Bandwidth untuk Konsistensi**) *Mean Squared Error (MSE) estimator GWR adalah*

$$\mathrm{MSE}(\widehat{\boldsymbol{\beta}}(u_i, v_i)) = \mathrm{Bias}^2 + \mathrm{Var} = O(h^4) + O((nh^2)^{-1}). \tag{2.7.51}$$

Untuk mencapai konsistensi ($\mathrm{MSE} \to 0$ *ketika* $n \to \infty$*), diperlukan:*

(i) $h \to 0$ *agar bias* $O(h^2) \to 0$*, dan*

(ii) $nh^2 \to \infty$ *agar variansi* $O((nh^2)^{-1}) \to 0$.

Selanjutnya, analisis asimtotik dari estimator GWR dan uji statistik yang relevan dijelaskan di bawah ini.

Teorema 2.7.16 (**Konsistensi Estimator GWR**) *Di bawah asumsi (G1)–(G6) dan kondisi* bandwidth $h \to 0$ *serta* $nh^2 \to \infty$ *ketika* $n \to \infty$*, estimator GWR bersifat konsisten:*

$$\widehat{\boldsymbol{\beta}}(u_i, v_i) \xrightarrow{p} \boldsymbol{\beta}(u_i, v_i). \tag{2.7.52}$$

Bukti. Konsistensi ditunjukkan dengan membuktikan bahwa $\mathrm{MSE}(\widehat{\boldsymbol{\beta}}(u_i, v_i)) \to 0$ ketika $n \to \infty$. Dari teorema bias dan variansi yang telah dibuktikan sebelumnya, diperoleh

$$\mathrm{MSE}(\widehat{\boldsymbol{\beta}}(u_i, v_i)) = \|\mathrm{Bias}\|^2 + \mathrm{tr}(\mathrm{Var}) = O(h^4) + O((nh^2)^{-1}). \tag{2.7.53}$$

Di bawah kondisi $h \to 0$ dan $nh^2 \to \infty$ ketika $n \to \infty$:

(i) Suku bias $O(h^4) \to 0$ karena $h \to 0$.

(ii) Suku variansi $O((nh^2)^{-1}) \to 0$ karena $nh^2 \to \infty$.

Oleh karena itu, $\mathrm{MSE}(\widehat{\boldsymbol{\beta}}(u_i, v_i)) \to 0$. Dengan ketaksamaan Chebyshev, untuk setiap $\varepsilon > 0$,

$$P\left(\|\widehat{\boldsymbol{\beta}}(u_i, v_i) - \boldsymbol{\beta}(u_i, v_i)\| > \varepsilon\right) \leq \frac{\mathrm{MSE}(\widehat{\boldsymbol{\beta}}(u_i, v_i))}{\varepsilon^2} \to 0, \tag{2.7.54}$$

yang menunjukkan konvergensi dalam probabilitas $\widehat{\boldsymbol{\beta}}(u_i, v_i) \xrightarrow{p} \boldsymbol{\beta}(u_i, v_i)$. ■

Teorema 2.7.17 (Konsistensi Estimator Variansi GWR) *Di bawah kondisi yang sama dengan teorema konsistensi, estimator variansi $\hat{\sigma}^2$ bersifat konsisten:*

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2. \tag{2.7.55}$$

Bukti. Estimator variansi adalah $\hat{\sigma}^2 = \boldsymbol{e}^\top \boldsymbol{e}/\nu$ dengan $\nu = n - 2\mathrm{tr}(\mathbf{S}) + \mathrm{tr}(\mathbf{S}^\top \mathbf{S})$. Vektor residual dapat ditulis sebagai

$$\boldsymbol{e} = (\mathbf{I} - \mathbf{S})\boldsymbol{y} = (\mathbf{I} - \mathbf{S})(\boldsymbol{\mu} + \boldsymbol{\varepsilon}), \tag{2.7.56}$$

dengan $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{y}]$.

Jumlah kuadrat residual dapat didekomposisi sebagai

$$\boldsymbol{e}^\top \boldsymbol{e} = \boldsymbol{\varepsilon}^\top (\mathbf{I}-\mathbf{S})^\top (\mathbf{I}-\mathbf{S})\boldsymbol{\varepsilon} + 2\boldsymbol{\mu}^\top (\mathbf{I}-\mathbf{S})^\top (\mathbf{I}-\mathbf{S})\boldsymbol{\varepsilon} + \boldsymbol{\mu}^\top (\mathbf{I}-\mathbf{S})^\top (\mathbf{I}-\mathbf{S})\boldsymbol{\mu}. \tag{2.7.57}$$

Di bawah kondisi konsistensi, estimator GWR konsisten sehingga $(\mathbf{I}-\mathbf{S})\boldsymbol{\mu} \xrightarrow{p} \mathbf{0}$, yang berarti suku ketiga menjadi negligible. Suku kedua memiliki nilai harapan nol. Untuk suku pertama, dengan hukum bilangan besar,

$$\frac{1}{\nu} \boldsymbol{\varepsilon}^\top (\mathbf{I} - \mathbf{S})^\top (\mathbf{I} - \mathbf{S})\boldsymbol{\varepsilon} \xrightarrow{p} \sigma^2, \tag{2.7.58}$$

karena $\mathbb{E}[\boldsymbol{\varepsilon}^\top (\mathbf{I} - \mathbf{S})^\top (\mathbf{I} - \mathbf{S})\boldsymbol{\varepsilon}] = \sigma^2 \nu$ dan variansi dari kuantitas ini berkurang seiring bertambahnya n. Oleh karena itu, $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$. ■

Teorema 2.7.18 (Normalitas Asimtotik Estimator GWR) *Di bawah asumsi (G1)–(G6), kondisi* bandwidth *yang sesuai, dan dengan* bandwidth h *diperlakukan sebagai diketahui (atau diestimasi dengan properti oracle), estimator GWR memiliki distribusi asimtotik:*

$$\sqrt{nh^2} \left(\widehat{\boldsymbol{\beta}}(u_i, v_i) - \boldsymbol{\beta}(u_i, v_i) - \boldsymbol{b}_i \right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_i\right), \tag{2.7.59}$$

dengan $\boldsymbol{\Sigma}_i$ adalah matriks kovarians asimtotik yang bergantung pada σ^2, kepadatan spasial $f(u_i, v_i)$, dan kekasaran kernel.

Secara praktis, dengan mengabaikan bias (melalui undersmoothing *atau koreksi bias), berlaku*

$$\widehat{\boldsymbol{\beta}}(u_i, v_i) \dot{\sim} \mathcal{N}\left(\boldsymbol{\beta}(u_i, v_i), \sigma^2(\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i^2 \boldsymbol{X} (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1}\right). \quad (2.7.60)$$

Bukti. Estimator GWR dapat dituliskan sebagai

$$\widehat{\boldsymbol{\beta}}(u_i, v_i) = \boldsymbol{\beta}(u_i, v_i) + \boldsymbol{b}_i + (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{\varepsilon}, \quad (2.7.61)$$

dengan $\boldsymbol{b}_i = O(h^2)$ adalah bias deterministik. Mendefinisikan

$$\boldsymbol{Z}_i = (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{\varepsilon} = \sum_{j=1}^{n} c_{ij} \varepsilon_j, \quad (2.7.62)$$

dengan c_{ij} adalah elemen dari matriks $\mathbf{C}_i = (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i$.

Karena ε_j independen dengan $\mathbb{E}[\varepsilon_j] = 0$ dan $\operatorname{Var}(\varepsilon_j) = \sigma^2$, maka $\boldsymbol{Z}_i$ adalah kombinasi linear dari variabel acak independen. Untuk menerapkan teorema limit pusat (CLT), perlu ditunjukkan bahwa kondisi Lindeberg terpenuhi.

Dengan *bandwidth* h, bobot w_{ij} terkonsentrasi pada $n_{\text{eff}} = O(nh^2)$ observasi dengan jarak $O(h)$ dari lokasi i. Variansi dari $\boldsymbol{Z}_i$ adalah

$$\operatorname{Var}(\boldsymbol{Z}_i) = \sigma^2(\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{W}_i^2 \boldsymbol{X} (\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1} = O((nh^2)^{-1}). \quad (2.7.63)$$

Dengan normalisasi $\sqrt{nh^2}$, diperoleh

$$\sqrt{nh^2} \boldsymbol{Z}_i = \sqrt{nh^2} \sum_{j=1}^{n} c_{ij} \varepsilon_j. \quad (2.7.64)$$

Kondisi Lindeberg terpenuhi karena bobot w_{ij} dari fungsi *kernel* menurun secara halus dan tidak ada observasi tunggal yang mendominasi. Secara spesifik, kontribusi maksimum individual adalah $O((nh^2)^{-1})$, yang menuju nol relatif terhadap total variansi.

Dengan teorema limit pusat Lindeberg-Feller, diperoleh

$$\sqrt{nh^2}\left(\widehat{\boldsymbol{\beta}}(u_i, v_i) - \boldsymbol{\beta}(u_i, v_i) - \boldsymbol{b}_i\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_i\right), \tag{2.7.65}$$

dengan $\boldsymbol{\Sigma}_i = \lim_{n\to\infty} nh^2 \cdot \mathrm{Var}(\boldsymbol{Z}_i)$ adalah matriks kovarians asimtotik. ∎

Ketika *bandwidth* diestimasi menggunakan CV atau AIC:

(a) Jika $\hat{h}$ konsisten dan memiliki laju konvergensi $\hat{h} - h_{\text{opt}} = O_p(n^{-\alpha})$ untuk $\alpha > 0$ yang cukup besar, maka distribusi asimtotik estimator koefisien tidak terpengaruh secara orde pertama (*oracle property*).

(b) Untuk sampel hingga, variabilitas tambahan dari estimasi *bandwidth* dapat menyebabkan interval konfidensi yang sedikit *undercoverage*.

(c) Metode *bootstrap* dapat digunakan untuk memperhitungkan ketidakpastian estimasi *bandwidth*.

Teorema 2.7.19 (Uji Signifikansi Koefisien Lokal) *Untuk menguji* $H_0 : \beta_k(u_i, v_i) = 0$ *terhadap* $H_1 : \beta_k(u_i, v_i) \neq 0$, *statistik uji adalah*

$$t_k(u_i, v_i) = \frac{\hat{\beta}_k(u_i, v_i)}{\mathrm{se}(\hat{\beta}_k(u_i, v_i))}, \tag{2.7.66}$$

dengan

$$\mathrm{se}(\hat{\beta}_k(u_i, v_i)) = \hat{\sigma}\sqrt{\boldsymbol{e}_{k+1}^\top(\boldsymbol{X}^\top\mathbf{W}_i\boldsymbol{X})^{-1}\boldsymbol{X}^\top\mathbf{W}_i^2\boldsymbol{X}(\boldsymbol{X}^\top\mathbf{W}_i\boldsymbol{X})^{-1}\boldsymbol{e}_{k+1}}, \tag{2.7.67}$$

dan $\mathbf{e}_{k+1}$ *adalah vektor unit ke-*$(k+1)$. *Di bawah* H_0, *statistik ini mengikuti distribusi* t *dengan derajat bebas efektif* $\nu = n - 2\mathrm{tr}(\mathbf{S}) + \mathrm{tr}(\mathbf{S}^\top\mathbf{S})$, *atau secara asimtotik* $\mathcal{N}(0, 1)$.

Bukti. Dari teorema normalitas asimtotik, estimator $\hat{\beta}_k(u_i, v_i)$ mengikuti distribusi normal asimtotik. Secara spesifik, komponen ke-$(k+1)$ dari vektor estimator memenuhi

$$\hat{\beta}_k(u_i, v_i) \dot{\sim} \mathcal{N}\left(\beta_k(u_i, v_i), \sigma^2 v_{k,i}\right), \tag{2.7.68}$$

dengan $v_{k,i} = \boldsymbol{e}_{k+1}^\top(\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1}\boldsymbol{X}^\top \mathbf{W}_i^2 \boldsymbol{X}(\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})^{-1}\boldsymbol{e}_{k+1}$ adalah elemen diagonal ke-$(k+1)$ dari matriks kovarians sandwich.

Di bawah hipotesis nol $H_0 : \beta_k(u_i, v_i) = 0$, statistik

$$Z_k = \frac{\hat{\beta}_k(u_i, v_i)}{\sigma\sqrt{v_{k,i}}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{2.7.69}$$

Karena σ^2 tidak diketahui, digunakan estimator konsisten $\hat{\sigma}^2$. Dari teorema konsistensi estimator variansi, $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$. Dengan teorema Slutsky,

$$t_k(u_i, v_i) = \frac{\hat{\beta}_k(u_i, v_i)}{\hat{\sigma}\sqrt{v_{k,i}}} = \frac{Z_k}{\hat{\sigma}/\sigma} \xrightarrow{d} \mathcal{N}(0, 1). \tag{2.7.70}$$

Untuk sampel hingga, aproksimasi distribusi t dengan derajat bebas $\nu = n - 2\mathrm{tr}(\mathbf{S}) + \mathrm{tr}(\mathbf{S}^\top \mathbf{S})$ memberikan hasil yang lebih akurat karena memperhitungkan ketidakpastian dalam estimasi σ^2. ■

Teorema 2.7.20 (Uji Heterogenitas Spasial (Stasioneritas)) *Untuk menguji H_0 : koefisien stasioner (konstan di semua lokasi) terhadap H_1 : koefisien bervariasi secara spasial, digunakan statistik F:*

$$F = \frac{(\mathrm{RSS}_{\mathrm{OLS}} - \mathrm{RSS}_{\mathrm{GWR}})/\nu_1}{\mathrm{RSS}_{\mathrm{GWR}}/\nu_2}, \tag{2.7.71}$$

dengan:

- $\mathrm{RSS}_{\mathrm{OLS}} = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}})^\top(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}})$,
- $\mathrm{RSS}_{\mathrm{GWR}} = (\boldsymbol{y} - \mathbf{S}\boldsymbol{y})^\top(\boldsymbol{y} - \mathbf{S}\boldsymbol{y})$,
- $\nu_1 = \mathrm{tr}(\mathbf{S}) - (p+1)$ *(perbedaan derajat bebas efektif),*
- $\nu_2 = n - 2\mathrm{tr}(\mathbf{S}) + \mathrm{tr}(\mathbf{S}^\top\mathbf{S})$.

Di bawah H_0, statistik ini mengikuti distribusi $F(\nu_1, \nu_2)$ secara aproksimasi.

Bukti. Uji ini membandingkan model terbatas (OLS global) dengan model penuh (GWR lokal). Di bawah H_0 bahwa koefisien konstan di semua lokasi, model GWR mereduksi menjadi model OLS global.

Untuk model OLS, matriks topi adalah $\mathbf{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$ dengan $\mathrm{tr}(\mathbf{H}) = p+1$. Untuk model GWR, matriks topi adalah $\mathbf{S}$ dengan derajat bebas efektif $\mathrm{tr}(\mathbf{S})$.

Perbedaan RSS dapat ditulis sebagai

$$\begin{aligned} \mathrm{RSS}_{\mathrm{OLS}} - \mathrm{RSS}_{\mathrm{GWR}} &= \boldsymbol{y}^\top(\mathbf{I}-\mathbf{H})^\top(\mathbf{I}-\mathbf{H})\boldsymbol{y} - \boldsymbol{y}^\top(\mathbf{I}-\mathbf{S})^\top(\mathbf{I}-\mathbf{S})\boldsymbol{y} \\ &= \boldsymbol{y}^\top \left[(\mathbf{I}-\mathbf{H})^\top(\mathbf{I}-\mathbf{H}) - (\mathbf{I}-\mathbf{S})^\top(\mathbf{I}-\mathbf{S})\right] \boldsymbol{y}. \end{aligned} \tag{2.7.72}$$

Di bawah H_0 dengan galat normal $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, bentuk kuadratik $\boldsymbol{y}^\top\mathbf{A}\boldsymbol{y}/\sigma^2$ mengikuti distribusi χ^2 dengan derajat bebas sama dengan rank matriks $\mathbf{A}$ (jika $\mathbf{A}$ idempoten).

Meskipun $\mathbf{S}$ tidak idempoten, aproksimasi Satterthwaite dapat digunakan untuk menentukan derajat bebas efektif. Derajat bebas untuk pembilang adalah

$$\nu_1 = \mathrm{tr}(\mathbf{S}) - (p+1), \tag{2.7.73}$$

yang merepresentasikan kompleksitas tambahan dari model GWR dibandingkan OLS.

Derajat bebas untuk penyebut adalah derajat bebas efektif residual GWR:

$$\nu_2 = n - 2\mathrm{tr}(\mathbf{S}) + \mathrm{tr}(\mathbf{S}^\top\mathbf{S}). \tag{2.7.74}$$

Dengan menggunakan aproksimasi distribusi F, statistik uji mengikuti $F(\nu_1, \nu_2)$ di bawah H_0. ■

Teorema 2.7.21 (**Uji Variasi Spasial Koefisien Individual**) *Untuk menguji apakah koefisien tertentu β_k bervariasi secara spasial, dapat digunakan statistik berbasis* *variansi:*

$$V_k = \frac{\sum_{i=1}^{n}(\hat{\beta}_k(u_i, v_i) - \bar{\hat{\beta}}_k)^2}{\sum_{i=1}^{n} \mathrm{Var}(\hat{\beta}_k(u_i, v_i))}, \tag{2.7.75}$$

dengan $\bar{\hat{\beta}}_k = \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_k(u_i, v_i)$*. Di bawah* H_0 *bahwa* β_k *konstan,* V_k *mengikuti distribusi* χ^2 *yang dimodifikasi.*

Bukti. Di bawah $H_0 : \beta_k(u_i, v_i) = \beta_k$ untuk semua i (koefisien konstan), estimator lokal $\hat{\beta}_k(u_i, v_i)$ seharusnya tidak bervariasi secara sistematis antar lokasi selain karena variabilitas sampling.

Pembilang $\sum_{i=1}^{n}(\hat{\beta}_k(u_i, v_i) - \bar{\hat{\beta}}_k)^2$ mengukur variasi empiris dari estimator lokal di sekitar rata-ratanya. Di bawah H_0, variasi ini seharusnya konsisten dengan variabilitas sampling yang diharapkan.

Penyebut $\sum_{i=1}^{n} \mathrm{Var}(\hat{\beta}_k(u_i, v_i))$ adalah jumlah variansi teoretis dari estimator lokal. Rasio ini membandingkan variasi yang diamati dengan variasi yang diharapkan.

Di bawah H_0 dan asumsi normalitas, jika estimator lokal independen (yang merupakan aproksimasi untuk lokasi yang berjauhan), maka

$$\frac{\hat{\beta}_k(u_i, v_i) - \beta_k}{\sqrt{\mathrm{Var}(\hat{\beta}_k(u_i, v_i))}} \sim \mathcal{N}(0, 1). \tag{2.7.76}$$

Namun, karena estimator lokal berkorelasi (lokasi yang berdekatan menggunakan observasi yang tumpang tindih), distribusi eksak dari V_k bukanlah χ^2 standar. Distribusi χ^2 yang dimodifikasi dengan derajat bebas efektif $\nu_{\text{eff}} < n - 1$ dapat diturunkan menggunakan aproksimasi momen pertama dan kedua dari V_k. ■

Teorema 2.7.22 (Interval Konfidensi untuk Koefisien Lokal) *Interval konfidensi* $(1 - \alpha) \times 100\%$ *untuk* $\beta_k(u_i, v_i)$ *adalah*

$$\hat{\beta}_k(u_i, v_i) \pm t_{\alpha/2,\nu} \cdot \mathrm{se}(\hat{\beta}_k(u_i, v_i)), \tag{2.7.77}$$

dengan $t_{\alpha/2,\nu}$ *adalah kuantil distribusi* t *dengan derajat bebas* ν*. Untuk sampel* *besar, dapat diganti dengan kuantil normal* $z_{\alpha/2}$.

Bukti. Dari teorema normalitas asimtotik dan konstruksi statistik t pada uji signifikansi koefisien lokal, diperoleh bahwa

$$\frac{\hat{\beta}_k(u_i, v_i) - \beta_k(u_i, v_i)}{\text{se}(\hat{\beta}_k(u_i, v_i))} \sim t_\nu \tag{2.7.78}$$

secara aproksimasi, dengan $\nu = n - 2\text{tr}(\mathbf{S}) + \text{tr}(\mathbf{S}^\top \mathbf{S})$.

Oleh karena itu,

$$P\left(-t_{\alpha/2,\nu} \leq \frac{\hat{\beta}_k(u_i, v_i) - \beta_k(u_i, v_i)}{\text{se}(\hat{\beta}_k(u_i, v_i))} \leq t_{\alpha/2,\nu}\right) = 1 - \alpha. \tag{2.7.79}$$

Dengan manipulasi aljabar,

$$P\left(\hat{\beta}_k(u_i, v_i) - t_{\alpha/2,\nu} \cdot \text{se} \leq \beta_k(u_i, v_i) \leq \hat{\beta}_k(u_i, v_i) + t_{\alpha/2,\nu} \cdot \text{se}\right) = 1 - \alpha, \tag{2.7.80}$$

yang memberikan interval konfidensi $(1 - \alpha) \times 100\%$ untuk $\beta_k(u_i, v_i)$.

Untuk sampel besar ($\nu \to \infty$), distribusi t_ν konvergen ke $\mathcal{N}(0, 1)$, sehingga $t_{\alpha/2,\nu}$ dapat diganti dengan $z_{\alpha/2}$. ■

Teorema 2.7.23 (Diagnostik Model GWR) *Beberapa diagnostik penting untuk evaluasi model GWR meliputi:*

(a) ***Derajat bebas efektif:***

$$\text{df}_{\text{eff}} = \text{tr}(\mathbf{S}), \tag{2.7.81}$$

yang mengukur kompleksitas model.

(b) ***Indeks determinasi lokal:***

$$R_i^2 = 1 - \frac{\sum_{j=1}^n w_{ij}(y_j - \hat{y}_j)^2}{\sum_{j=1}^n w_{ij}(y_j - \bar{y}_w)^2}, \tag{2.7.82}$$

dengan $\bar{y}_w = \frac{\sum_j w_{ij} y_j}{\sum_j w_{ij}}$.

*(c) **Indeks kondisi lokal (multikolinearitas):***

$$\kappa_i = \sqrt{\frac{\lambda_{\max}(\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})}{\lambda_{\min}(\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X})}}, \tag{2.7.83}$$

dengan nilai $\kappa_i > 30$ mengindikasikan multikolinearitas serius.

Bukti.

(a) **Derajat bebas efektif:** Dalam regresi OLS standar, matriks topi $\mathbf{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$ bersifat idempoten ($\mathbf{H}^2 = \mathbf{H}$) dan $\mathrm{tr}(\mathbf{H}) = p + 1$ sama dengan jumlah parameter. Untuk GWR, matriks $\mathbf{S}$ tidak idempoten, namun $\mathrm{tr}(\mathbf{S})$ tetap mengukur kompleksitas model secara analog. Setiap elemen diagonal S_{ii} mengukur pengaruh y_i terhadap prediksinya sendiri $\hat{y}_i$. Jumlah $\mathrm{tr}(\mathbf{S}) = \sum_{i=1}^n S_{ii}$ mengukur total "leverage" dan berfungsi sebagai derajat bebas efektif model.

(b) **Indeks determinasi lokal:** Konstruksi R_i^2 mengikuti prinsip yang sama dengan R^2 global, tetapi menggunakan bobot lokal w_{ij}. Pembilang $\sum_j w_{ij}(y_j - \hat{y}_j)^2$ adalah jumlah kuadrat residual tertimbang (mengukur variasi yang tidak dijelaskan), sedangkan penyebut $\sum_j w_{ij}(y_j - \bar{y}_w)^2$ adalah jumlah kuadrat total tertimbang (variasi total). Rasio ini mengukur proporsi variasi yang tidak dijelaskan, sehingga R_i^2 mengukur proporsi variasi yang dijelaskan oleh model di sekitar lokasi i.

(c) **Indeks kondisi lokal:** Untuk sistem linear $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$, indeks kondisi $\kappa(\mathbf{A}) = \|\mathbf{A}\| \cdot \|\mathbf{A}^{-1}\|$ mengukur sensitivitas solusi terhadap perubahan pada $\boldsymbol{b}$. Untuk matriks simetris definit positif, $\kappa(\mathbf{A}) = \lambda_{\max}/\lambda_{\min}$. Dalam konteks GWR, matriks $\boldsymbol{X}^\top \mathbf{W}_i \boldsymbol{X}$ digunakan untuk mengestimasi koefisien lokal. Indeks kondisi $\kappa_i = \sqrt{\lambda_{\max}/\lambda_{\min}}$ mengukur seberapa sensitif estimator terhadap perubahan kecil pada data. Nilai $\kappa_i > 30$ mengindikasikan bahwa kolom-kolom $\boldsymbol{X}$ (setelah pembobotan) hampir kolinear secara lokal, yang menyebabkan estimator tidak stabil.

$\blacksquare$

Teorema 2.7.24 (**Uji Autokorelasi Residual Spasial**) *Untuk menguji apakah residual GWR masih menunjukkan autokorelasi spasial, digunakan statistik Moran's I:*

$$I_{\text{GWR}} = \frac{n}{\sum_i \sum_j w_{ij}^*} \cdot \frac{\sum_i \sum_j w_{ij}^* e_i e_j}{\sum_i e_i^2}, \tag{2.7.84}$$

dengan $e_i = y_i - \hat{y}_i$ adalah residual GWR dan w_{ij}^* adalah bobot ketetanggaan spasial (dapat berbeda dari bobot kernel *GWR). Di bawah H_0 tidak ada autokorelasi spasial residual:*

$$\frac{I_{\text{GWR}} - \mathbb{E}[I]}{\sqrt{\text{Var}(I)}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{2.7.85}$$

Bukti. Statistik Moran's I mengukur korelasi spasial antara nilai-nilai di lokasi yang berdekatan. Dalam notasi matriks, dengan $\mathbf{W}^* = (w_{ij}^*)$ adalah matriks bobot ketetanggaan,

$$I = \frac{n}{S_0} \cdot \frac{\boldsymbol{e}^\top \mathbf{W}^* \boldsymbol{e}}{\boldsymbol{e}^\top \boldsymbol{e}}, \tag{2.7.86}$$

dengan $S_0 = \sum_i \sum_j w_{ij}^*$ adalah jumlah total bobot.

Di bawah H_0 tidak ada autokorelasi spasial, residual e_i bersifat independen secara spasial. Untuk menurunkan distribusi asimtotik, perhatikan bahwa:

(i) Nilai harapan di bawah H_0 (dengan asumsi residual memiliki mean nol) adalah

$$\mathbb{E}[I] = -\frac{1}{n-1}, \tag{2.7.87}$$

yang mendekati nol untuk n besar.

(ii) Variansi I bergantung pada struktur matriks bobot $\mathbf{W}^*$ dan dapat dihitung sebagai

$$\text{Var}(I) = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{S_0^2 (n^2 - 1)} - \mathbb{E}[I]^2, \tag{2.7.88}$$

dengan $S_1 = \frac{1}{2} \sum_i \sum_j (w_{ij}^* + w_{ji}^*)^2$ dan $S_2 = \sum_i (\sum_j w_{ij}^* + \sum_j w_{ji}^*)^2$.

Untuk n besar, dengan kondisi regularitas pada matriks bobot (tidak ada lokasi yang mendominasi), teorema limit pusat dapat diterapkan. Statistik $\boldsymbol{e}^\top \mathbf{W}^* \boldsymbol{e}$ adalah bentuk kuadratik dalam variabel acak, dan dengan normalisasi yang tepat, konvergen ke distribusi normal. Oleh karena itu,

$$Z = \frac{I - \mathbb{E}[I]}{\sqrt{\mathrm{Var}(I)}} \xrightarrow{d} \mathcal{N}(0,1). \tag{2.7.89}$$

Nilai Z yang besar positif mengindikasikan autokorelasi spasial positif (residual yang berdekatan cenderung mirip), sedangkan nilai negatif yang besar mengindikasikan autokorelasi negatif. ■

2.8 Jaringan Saraf Tiruan

Jaringan Saraf Tiruan (JST) atau *artificial neural networks* (ANN) mempelajari pemetaan nonlinier dari ruang masukan ke keluaran melalui komposisi berlapis fungsi-fungsi sederhana (**???**).

2.8.1 Model Dasar dan Notasi

Definisi 2.8.1 (Neuron Tiruan, (?)) *Diberikan $\boldsymbol{x} \in \mathbb{R}^p$, neuron menghasilkan*

$$y \;=\; \phi(\mathbf{w}^\top \boldsymbol{x} + b), \tag{2.8.1}$$

dengan $\mathbf{w} \in \mathbb{R}^p$ adalah bobot, $b \in \mathbb{R}$ adalah bias, dan $\phi : \mathbb{R} \to \mathbb{R}$ adalah fungsi aktivasi nonlinier terdiferensiasi hampir di semua titik.

Definisi 2.8.2 (Jaringan *Feedforward* Multilapis (?)) *Untuk L lapisan tersembunyi, definisikan rekursi*

$$\boldsymbol{z}^{(l)} = \mathbf{W}^{(l)}\boldsymbol{a}^{(l-1)} + \mathbf{b}^{(l)}, \qquad \boldsymbol{a}^{(l)} = \phi^{(l)}\big(\boldsymbol{z}^{(l)}\big), \qquad l = 1, \dots, L, \tag{2.8.2}$$

dengan $\boldsymbol{a}^{(0)} = \boldsymbol{x}$ dan keluaran akhir $\hat{y} = g\left(\boldsymbol{a}^{(L)}\right)$. Parameter jaringan $\Theta = \{\mathbf{W}^{(l)}, \mathbf{b}^{(l)}\}_{l=1}^{L}$ dipelajari dari data.

Contoh 2.8.3 Misalkan $\mathbf{x} = (1,2)^\top$, bobot $\mathbf{w} = (0.5, 0.25)^\top$, bias $b = 1$, dan

$\varphi(z) = \tanh(z)$. Maka keluaran neuron adalah

$$y = \tanh(0.5 \cdot 1 + 0.25 \cdot 2 + 1) = \tanh(2) \approx 0.964.$$

2.8.2 Fungsi Aktivasi

Fungsi aktivasi memperkenalkan non-linearitas pada ANN. Tanpa fungsi aktivasi non-linear, jaringan saraf hanya merepresentasikan transformasi linear berlapis.

Definisi 2.8.4 (**Fungsi Sigmoid**) *Fungsi sigmoid didefinisikan sebagai*

$$\varphi(z) = \frac{1}{1 + e^{-z}}. \tag{2.8.3}$$

Contoh 2.8.5 Jika $z = -10$, maka

$$\varphi(-10) = \frac{1}{1 + e^{10}} \approx \frac{1}{1 + 22026.5} \approx 4.54 \times 10^{-5}.$$

Fungsi sigmoid sangat jenuh untuk $|z|$ besar: gradiennya mendekati nol.

Definisi 2.8.6 (**Fungsi Tanh**) *Fungsi* $\tanh$ *didefinisikan sebagai*

$$\varphi(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}. \tag{2.8.4}$$

Contoh 2.8.7 Jika $z = 0$, maka

$$\varphi(0) = \tanh(0) = 0.$$

Jika $z = 2$, maka

$$\varphi(2) = \frac{e^2 - e^{-2}}{e^2 + e^{-2}} \approx \frac{7.389 - 0.135}{7.389 + 0.135} \approx \frac{7.254}{7.524} \approx 0.964.$$

Jika $z = -2$, maka

$$\varphi(-2) = \frac{e^{-2} - e^2}{e^{-2} + e^2} \approx \frac{0.135 - 7.389}{0.135 + 7.389} \approx \frac{-7.254}{7.524} \approx -0.964.$$

Fungsi $\tanh$ memiliki *range* $(-1, 1)$ dan gradien maksimum di sekitar $z = 0$.

Definisi 2.8.8 **(Fungsi ReLU, (?))** *Fungsi ReLU (*Rectified Linear Unit*) didefinisikan sebagai*

$$\varphi(z) = \max(0, z). \tag{2.8.5}$$

Contoh 2.8.9 Jika $z = -3$, maka $\varphi(z) = 0$. Jika $z = 2.5$, maka $\varphi(z) = 2.5$.

Definisi 2.8.10 **(Fungsi Leaky ReLU)** *Fungsi Leaky ReLU mengatasi masalah* dying ReLU *dengan*

$$\varphi(z) = \begin{cases} z, & z \geq 0, \\ \alpha z, & z < 0, \end{cases} \tag{2.8.6}$$

dengan $\alpha \in (0, 1)$ *kecil, misalnya* 0.01.

Contoh 2.8.11 Jika $z = -3$ dan $\alpha = 0.01$, maka $\varphi(z) = -0.03$. Jika $z = 2$, maka $\varphi(z) = 2$.

Definisi 2.8.12 **(Fungsi GELU)** *Fungsi Gaussian Error Linear Unit (GELU) adalah*

$$\varphi(z) = z \cdot \Phi(z), \tag{2.8.7}$$

dengan $\Phi(z)$ *fungsi distribusi normal standar. Aproksimasi praktisnya adalah sebagai berikut.*

$$\varphi(z) \approx 0.5z \left(1 + \tanh\left[\sqrt{\tfrac{2}{\pi}}(z + 0.0447z^3)\right]\right). \tag{2.8.8}$$

Contoh 2.8.13 Untuk $z = 1$, $\varphi(1) \approx 0.84$. Untuk $z = -1$, $\varphi(-1) \approx -0.16$.

Lema 2.8.14 **(Sifat Fungsi Aktivasi)** *Fungsi aktivasi* $\varphi : \mathbb{R} \to \mathbb{R}$ *yang digunakan pada jaringan saraf tiruan sebaiknya memenuhi:*

(i) φ *non-linear,*

(ii) φ *terdiferensiasi hampir di semua titik,*

(iii) φ *tidak konstan, dan*

(iv) φ *memiliki range memadai, tidak menyebabkan saturasi total.*

Bukti.

(i) *Nonlinearitas.* Jika φ linear, misalnya $\varphi(z) = az + b$, maka suatu multilayer perceptron dengan L lapisan menghasilkan

$$f(\boldsymbol{x}) = \mathbf{W}^{(L)}\varphi\big(\mathbf{W}^{(L-1)}\varphi\big(\dots\varphi(\mathbf{W}^{(1)}\boldsymbol{x})\big)\big) = \mathbf{W}'\boldsymbol{x} + \boldsymbol{b}',$$

dengan $\mathbf{W}' = \prod_{l=1}^{L} a\mathbf{W}^{(l)}$ dan $\boldsymbol{b}'$ kombinasi linear dari b. Komposisi tersebut tetap linear. Akibatnya, jaringan tidak lebih kuat daripada regresi linear biasa dan tidak mampu mengaproksimasi fungsi nonlinear, misalnya XOR.

(ii) *Kediferensialan.* Algoritma *backpropagation* memerlukan turunan $\varphi'(z)$ untuk menghitung gradien fungsi kerugian L. Dengan notasi $\boldsymbol{z}^{(l)} = \mathbf{W}^{(l)}\boldsymbol{a}^{(l-1)} + \boldsymbol{b}^{(l)}$ dan $\boldsymbol{a}^{(l)} = \varphi(\boldsymbol{z}^{(l)})$, bentuk umum gradien adalah

$$\frac{\partial L}{\partial \mathbf{W}^{(l)}} = \delta^{(l)}\,(\boldsymbol{a}^{(l-1)})^{\top}, \qquad \delta^{(l)} = (\mathbf{W}^{(l+1)})^{\top}\delta^{(l+1)} \odot \varphi'(\boldsymbol{z}^{(l)}).$$

Jika $\varphi'(z)$ tidak terdefinisi pada banyak titik, maka gradien tidak dapat dihitung. Oleh karena itu, φ setidaknya harus terdiferensiasi hampir di semua titik. Fungsi seperti ReLU tetap sah karena ketakdiferensialan hanya terjadi pada $z = 0$ (himpunan titik ukuran nol).

(iii) *Nonkonstansi.* Jika $\varphi(z) = c$ konstan, maka $\varphi'(z) = 0$ untuk semua z. Substitusi ke rumus backpropagation menghasilkan $\delta^{(l)} = \mathbf{0}$ untuk semua l, sehingga

$$\frac{\partial L}{\partial \mathbf{W}^{(l)}} = \mathbf{0}\,.$$

Bobot tidak akan pernah diperbarui, sehingga proses pelatihan tidak berjalan.

(iv) *Rentang keluaran yang memadai.* Jika φ cepat jenuh, misalnya sigmoid dengan limit $\varphi(z) \to 0$ untuk $z \to -\infty$ dan $\varphi(z) \to 1$ untuk $z \to +\infty$, maka untuk $|z|$ besar berlaku

$$\varphi'(z) = \varphi(z)(1 - \varphi(z)) \approx 0.$$

Akibatnya, besaran kesalahan terpropagasi $\delta^{(l)}$ mengecil secara eksponensial menuju nol (fenomena *vanishing gradient*). Hal ini dapat menggagalkan pembelajaran pada lapisan-lapisan dalam. Untuk mengatasinya, digunakan fungsi aktivasi dengan rentang yang lebih lebar atau gradien yang tidak menghilang, misalnya ReLU atau Leaky ReLU.

■

2.8.3 Fungsi Kerugian dan Kriteria Pembelajaran

Definisi 2.8.15 (Risiko Empiris, (?)) *Diberikan data terawasi $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ dan model prediksi $\hat{y}_i = f_\Theta(\boldsymbol{x}_i)$, tujuan pembelajaran adalah meminimalkan*

$$L(\Theta) = \frac{1}{n} \sum_{i=1}^{n} \ell(\hat{y}_i, y_i), \tag{2.8.9}$$

dengan fungsi kerugian ℓ dipilih sesuai dengan tugas yang dihadapi.

Contoh 2.8.16 Untuk regresi skalar, sebagai contoh digunakan $\ell(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$. Hal ini berarti jika $\hat{y} = 2.3$ dan $y = 1.5$, maka $\ell = \frac{1}{2}(0.8)^2 = 0.32$.

Contoh 2.8.17 Untuk klasifikasi biner, digunakan probabilitas prediksi $\hat{p} = \sigma(z)$ dan fungsi kerugian

$$\ell(\hat{p}, y) = -\big[y \log \hat{p} + (1 - y) \log(1 - \hat{p})\big], \quad y \in \{0, 1\}. \tag{2.8.10}$$

Fungsi ini merupakan entropi silang antara distribusi target $p(y)$ dan prediksi $\hat{p}$, yaitu

$$\begin{aligned} H(p, \hat{p}) &= -\sum_{y \in \{0,1\}} p(y) \log \hat{p}(y) \\ &= -\big[p(1) \log \hat{p} + p(0) \log(1 - \hat{p})\big]. \end{aligned} \tag{2.8.11}$$

Dengan $p(1) = y$ dan $p(0) = 1 - y$, diperoleh bentuk di atas.

Sebagai contoh, jika $y = 1$ dan $\hat{p} = 0.9$, maka $\ell = -\log 0.9 \approx 0.105$ (kerugian kecil). Jika $y = 0$ dan $\hat{p} = 0.9$, maka $\ell = -\log 0.1 \approx 2.303$ (kerugian besar).

Contoh 2.8.18 Untuk klasifikasi multikelas, dengan logit $\mathbf{z} \in \mathbb{R}^K$, softmax $\hat{p}_k = \exp(z_k) \big/ \sum_{j=1}^{K} \exp(z_j)$ dan

$$\ell(\hat{\mathbf{p}}, y) = -\log \hat{p}_y, \qquad y \in \{1, \ldots, K\}.$$

Fungsi ini juga merupakan entropi silang antara distribusi target $p(y)$ dan prediksi $\hat{\mathbf{p}}$:

$$\begin{aligned} H(p, \hat{\mathbf{p}}) &= -\sum_{k=1}^{K} p(k) \log \hat{p}_k \\ &= -\log \hat{p}_y, \end{aligned} \tag{2.8.12}$$

dengan $p(k) = 1$ jika $k = y$ dan 0 selainnya.

Sebagai contoh, untuk $K = 3$ kelas, jika $\mathbf{z} = (2.0,\ 1.0,\ 0.1)$, maka softmax $\hat{\mathbf{p}} \approx (0.659,\ 0.242,\ 0.099)$. Jika label benar $y = 1$, maka $\ell = -\log 0.659 \approx 0.417$. Jika $y = 3$, maka $\ell = -\log 0.099 \approx 2.313$.

2.8.4 Pembelajaran dengan Propagasi Mundur

Teorema 2.8.19 (**Aturan Rantai pada Jaringan**) *Misalkan keluaran akhir didefinisikan sebagai komposisi $y = f^{(L)} \circ f^{(L-1)} \circ \cdots \circ f^{(1)}(\boldsymbol{x})$. Maka turunan terhadap* *parameter pada lapisan ke-l memenuhi*

$$\frac{\partial y}{\partial \mathbf{W}^{(l)}} = \frac{\partial f^{(L)}}{\partial f^{(L-1)}} \cdots \frac{\partial f^{(l+1)}}{\partial f^{(l)}} \frac{\partial f^{(l)}}{\partial \mathbf{W}^{(l)}}. \tag{2.8.13}$$

Interpretasi teorema ini adalah bahwa gradien dihitung dengan mengalikan turunan setiap lapisan dari keluaran ke lapisan yang diinginkan. Hal ini berarti akan dihitung efek perubahan bobot pada lapisan tersebut terhadap keluaran akhir melalui semua lapisan berikutnya. Apabila efek ini kecil, maka gradien akan mengecil secara eksponensial (fenomena *vanishing gradient*). Apabila efek ini besar, maka gradien akan membesar secara eksponensial (fenomena *exploding gradient*). Secara umum, pembelajaran menginginkan untuk memperbaiki bobot yang memiliki efek signifikan terhadap keluaran, tetapi tidak terlalu besar sehingga menyebabkan ketidakstabilan.

Teorema 2.8.20 (Propagasi Mundur atau *Backpropagation*, (?)) *Dengan notasi $\boldsymbol{z}^{(l)} = \mathbf{W}^{(l)}\boldsymbol{a}^{(l-1)} + \boldsymbol{b}^{(l)}$, $\boldsymbol{a}^{(l)} = \varphi(\boldsymbol{z}^{(l)})$, dan $\ell(\hat{y}, y)$ fungsi kerugian, gradien terhadap bobot $\mathbf{W}^{(l)}$ diberikan oleh*

$$\frac{\partial \ell}{\partial \mathbf{W}^{(l)}} = \delta^{(l)} \, (\boldsymbol{a}^{(l-1)})^\top, \tag{2.8.14}$$

dengan $\delta^{(l)} = (\mathbf{W}^{(l+1)})^\top \delta^{(l+1)} \odot \varphi'(\boldsymbol{z}^{(l)})$. Operator $\odot$ adalah perkalian elemen-per-elemen atau Hadamard product.

Contoh 2.8.21 Misalkan jaringan 2 lapis dengan *input* $\boldsymbol{x} = (x_1, x_2)^\top$, bobot

$$\mathbf{W}^{(1)} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix},$$

bias $\boldsymbol{b}^{(1)} = (0, 0)^\top$, fungsi aktivasi $\varphi(z) = \tanh(z)$, dan lapisan output linear $y = \boldsymbol{w}^{(2)\top}\boldsymbol{a}^{(1)}$ dengan $\boldsymbol{w}^{(2)} = (1, 1)^\top$. Target $y^* = 1$, kerugian $\ell(y, y^*) = \frac{1}{2}(y - y^*)^2$.

(a) *Forward:*

$$\begin{aligned} \boldsymbol{z}^{(1)} &= \mathbf{W}^{(1)}\boldsymbol{x} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ \boldsymbol{a}^{(1)} &= \tanh(\boldsymbol{z}^{(1)}) \\ y &= (1,1) \cdot \boldsymbol{a}^{(1)} \end{aligned}$$

(b) *Backward:*

$$\begin{aligned} \delta^{(2)} &= \frac{\partial \ell}{\partial y} = y - y^* \\ \delta^{(1)} &= \boldsymbol{w}^{(2)} \cdot \delta^{(2)} \odot \varphi'(\boldsymbol{z}^{(1)}) \\ \frac{\partial \ell}{\partial \mathbf{W}^{(1)}} &= \delta^{(1)}(\boldsymbol{x})^\top \end{aligned}$$

Dengan substitusi nilai, gradien dapat dihitung secara eksplisit untuk setiap parameter.

2.8.5 Optimisasi Parameter

Dalam pelatihan jaringan saraf, parameter Θ dioptimalkan untuk meminimalkan fungsi risiko $L(\Theta)$. Beberapa algoritma optimisasi yang akan dibahas adalah *Gradient Descent* (GD), *Stochastic Gradient Descent* (SGD), dan *Adaptive Moment Estimation* (Adam).

Definisi 2.8.22 (***Gradient Descent***) *Diberikan fungsi risiko empiris*

$$L(\Theta) = \tfrac{1}{n} \sum_{i=1}^{n} \ell\big(f_\Theta(\mathbf{x}_i), y_i\big),$$

pembaruan parameter dengan laju belajar $\eta > 0$ adalah

$$\Theta_{t+1} = \Theta_t - \eta \, \nabla_\Theta L(\Theta_t)\,. \tag{2.8.15}$$

Metode ini konvergen ke minimum lokal untuk fungsi non-konveks seperti jaringan *saraf.*

Intuisi dari metode berbasis gradien ini adalah memperbarui parameter model ke arah yang menurunkan fungsi kerugian paling cepat. Pada setiap iterasi, gradien $\nabla_\Theta L(\Theta)$ menunjukkan arah perubahan parameter yang paling efektif untuk mengurangi kerugian. Dengan memilih laju pembelajaran η yang sesuai, proses optimisasi bergerak menuju minimum lokal dari fungsi kerugian. Jika $L(\Theta)$ konveks, metode ini menjamin konvergensi ke minimum global. Namun, pada kasus nonkonveks seperti jaringan saraf, *gradient descent* tetap efektif menemukan solusi yang baik secara empiris. Variasi seperti SGD dan Adam mempercepat konvergensi dan meningkatkan generalisasi dengan memanfaatkan batch acak dan penyesuaian adaptif terhadap gradien.

Contoh 2.8.23 Misalkan fungsi kerugian $L(w) = \frac{1}{2}(w-3)^2$. Gradiennya adalah $\nabla_w L(w) = w - 3$. Dengan laju belajar $\eta = 0.1$ dan inisialisasi $w_0 = 0$, maka iterasi *gradient descent* adalah sebagai berikut.

$$
\begin{aligned}
w_1 &= w_0 - 0.1 \cdot (w_0 - 3) = 0 - 0.1 \cdot (-3) = 0.3 \\
w_2 &= w_1 - 0.1 \cdot (w_1 - 3) = 0.3 - 0.1 \cdot (-2.7) = 0.57 \\
w_3 &= 0.57 - 0.1 \cdot (0.57 - 3) = 0.813 \\
&\vdots
\end{aligned}
$$

Nilai w_t akan konvergen menuju 3.

Definisi 2.8.24 (***Stochastic Gradient Descent*** **(SGD)**) *SGD adalah metode optimisasi yang memperbarui parameter model menggunakan rata-rata gradien dari* batch *kecil acak berukuran m pada setiap iterasi.* Batch *adalah sekumpulan sampel yang digunakan untuk menghitung gradien, serta batch kecil ini diambil secara acak dari* dataset. *Pembaruan parameter dilakukan sebagai berikut.*

$$
\Theta_{t+1} = \Theta_t - \eta \frac{1}{|\mathcal{B}_t|} \sum_{(\mathbf{x}_i, y_i) \in \mathcal{B}_t} \nabla_\Theta \, \ell\big(f_\Theta(\mathbf{x}_i), y_i\big), \tag{2.8.16}
$$

dengan $\mathcal{B}_t$ adalah batch *kecil pada iterasi ke-t, η adalah laju pembelajaran, dan ℓ adalah fungsi kerugian. Metode ini sangat efisien untuk data berukuran besar dan memberikan regularisasi implisit melalui pengacakan* batch.

Contoh 2.8.25 Misalkan fungsi kerugian $L(w) = \frac{1}{2}(w-3)^2$ dan data terdiri dari $n = 1000$ sampel, masing-masing x_i dengan target $y_i = 3$ untuk semua i. Dengan *batch* kecil berukuran $m = 10$, pada setiap iterasi SGD dilakukan langkah berikut.

(a) Pilih acak $m = 10$ sampel dari data, misal indeks $\mathcal{B}_t = \{i_1, \ldots, i_{10}\}$.

(b) Hitung rata-rata gradien pada batch:

$$\nabla_w L_{\mathcal{B}_t}(w) = \frac{1}{m}\sum_{j=1}^{m}(w - y_{i_j}) = w - 3$$

(karena semua $y_{i_j} = 3$).

(c) Perbarui parameter dengan laju belajar η:

$$w_{t+1} = w_t - \eta(w_t - 3)$$

(d) Ulangi langkah (a)–(c) hingga w_t konvergen ke 3.

Dengan batch acak, setiap iterasi hanya menggunakan sebagian kecil data, sehingga proses lebih efisien dan tetap menuju minimum global $w = 3$.

Definisi 2.8.26 (***Adaptive Moment Estimation*** **(Adam)**) *Adam adalah algoritma optimisasi berbasis gradien yang menggabungkan momentum (rata-rata bergerak gradien) dan skala adaptif (rata-rata kuadrat gradien). Adam memperbarui parameter dengan estimasi momen pertama dan kedua yang dikoreksi bias. Dengan gradien $\mathbf{g}_t = \nabla_\Theta L_t(\Theta_t)$, inisialisasi $\mathbf{m}_0 = \mathbf{0}$, $\mathbf{v}_0 = \mathbf{0}$, lalu untuk setiap iterasi*

$t \geq 1$:

$$\mathbf{m}_t = \beta_1 \mathbf{m}_{t-1} + (1-\beta_1)\,\mathbf{g}_t, \tag{2.8.17}$$

$$\mathbf{v}_t = \beta_2 \mathbf{v}_{t-1} + (1-\beta_2)\,(\mathbf{g}_t \odot \mathbf{g}_t), \tag{2.8.18}$$

$$\hat{\mathbf{m}}_t = \frac{\mathbf{m}_t}{1-\beta_1^t}, \quad \hat{\mathbf{v}}_t = \frac{\mathbf{v}_t}{1-\beta_2^t}, \tag{2.8.19}$$

$$\Theta_{t+1} = \Theta_t - \eta\,\frac{\hat{\mathbf{m}}_t}{\sqrt{\hat{\mathbf{v}}_t} + \varepsilon}. \tag{2.8.20}$$

Parameter umum: $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$.

Contoh 2.8.27 Misalkan gradien pada iterasi pertama $\mathbf{g}_1 = (0.1,\ -0.2)$, inisialisasi $\mathbf{m}_0 = (0,\ 0)$, $\mathbf{v}_0 = (0,\ 0)$, $\eta = 0.01$. Maka:

$$\begin{aligned}
\mathbf{m}_1 &= 0.9 \cdot (0,\ 0) + 0.1 \cdot (0.1,\ -0.2) = (0.01,\ -0.02),\\
\mathbf{v}_1 &= 0.999 \cdot (0,\ 0) + 0.001 \cdot (0.01,\ 0.04) = (0.00001,\ 0.00004),\\
\hat{\mathbf{m}}_1 &= (0.01,\ -0.02)/(1-0.9^1) = (0.1,\ -0.2),\\
\hat{\mathbf{v}}_1 &= (0.00001,\ 0.00004)/(1-0.999^1) = (0.01,\ 0.04),\\
\Theta_2 &= \Theta_1 - 0.01 \cdot \frac{(0.1,\ -0.2)}{(\sqrt{0.01},\ \sqrt{0.04}) + 10^{-8}} = \Theta_1 - 0.01 \cdot (1,\ -1).
\end{aligned}$$

Jadi, parameter diperbarui dengan langkah adaptif pada setiap komponen.

2.8.6 Jaringan Saraf Tiruan sebagai Aproksimasi Universal

Salah satu hasil teoretis paling penting tentang jaringan saraf tiruan adalah Teorema Aproksimasi Universal (UAT). Teorema ini menyatakan bahwa jaringan saraf dengan satu lapisan tersembunyi yang cukup besar dapat mengaproksimasi fungsi kontinu pada himpunan kompak dengan presisi setinggi apapun, asalkan fungsi aktivasi memenuhi syarat tertentu.

Teorema 2.8.28 (Teorema Aproksimasi Universal (UAT), (?)) *Jika $\varphi : \mathbb{R} \to \mathbb{R}$ adalah fungsi sigmoid kontinu tak konstan (misal, $\varphi(z) = 1/(1+e^{-z})$ atau $\varphi(z) = \tanh(z)$), maka untuk setiap fungsi kontinu f pada $[0,1]^p$ dan setiap $\varepsilon > 0$, ada* *jaringan satu lapis dengan jumlah neuron cukup banyak sehingga*

$$\sup_{\boldsymbol{x}\in[0,1]^p} |f(\boldsymbol{x}) - g(\boldsymbol{x})| < \varepsilon,$$

dengan $g(\boldsymbol{x}) = \sum_{j=1}^{m} c_j\, \varphi(\boldsymbol{w}_j^\top \boldsymbol{x} + b_j)$.

2.9 Jaringan Saraf Graf

Jaringan Saraf Graf (*Graph Neural Networks*, GNN) merupakan kelas arsitektur pembelajaran representasi yang dirancang khusus untuk data terstruktur dalam bentuk graf. Berbeda dengan jaringan saraf tiruan (JST) yang bekerja pada vektor atau matriks, GNN memanfaatkan informasi topologi melalui graf $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ dengan simpul $\mathcal{V}$ dan sisi $\mathcal{E}$.

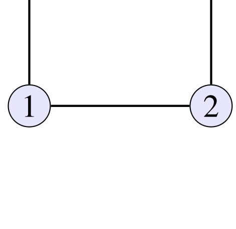
2.9.1 Dasar Graf dan Laplacian

Definisi 2.9.1 (Graf dan Matriks Ketetanggaan) *Graf* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *dengan* $|\mathcal{V}| = n$ *direpresentasikan oleh:*

$$\mathbf{A}_{ij} = \begin{cases} 1 & \textit{jika } (i,j) \in \mathcal{E}, \\ 0 & \textit{lainnya}, \end{cases} \tag{2.9.1}$$

disebut matriks ketetanggaan. Matriks derajat didefinisikan sebagai $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n)$, *dengan* $d_i = \sum_j A_{ij}$.

Contoh 2.9.2 Graf $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ dengan $|\mathcal{V}| = 4$ dan sisi $\mathcal{E} = \{(1,2), (2,3), (3,4), (4,1)\}$ dapat divisualisasikan sebagai berikut.



Matriks ketetanggaan $\mathbf{A}$ untuk graf di atas adalah

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

Dengan kata lain, matriks derajatnya $\mathbf{D} = \mathrm{diag}(2, 2, 2, 2)$.

Definisi 2.9.3 (**Graf Laplacian, (?)**) *Graf Laplacian adalah* $\mathbf{L} = \mathbf{D} - \mathbf{A}$. *Normalisasi simetris diberikan oleh*

$$\mathbf{L}_{sym} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}. \tag{2.9.2}$$

Graf Laplacian merepresentasikan struktur graf dan digunakan dalam berbagai algoritma GNN.

Contoh 2.9.4 Misalkan graf $\mathcal{G}$ dengan 4 simpul dan sisi $\mathcal{E} = \{(1, 2), (2, 3), (3, 4), (4, 1)\}$ (graf siklus). Matriks ketetanggaannya adalah

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

dan matriks derajatnya adalah

$$\mathbf{D} = \mathrm{diag}(2, 2, 2, 2).$$

Graf Laplacian-nya dapat dihitung sebagai

$$\mathbf{L} = \mathbf{D} - \mathbf{A} = \begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ -1 & 0 & -1 & 2 \end{pmatrix}.$$

Sehingga, dengan normalisasi simetris didapatkan

$$\mathbf{L}_{\text{sym}} = \mathbf{I} - \frac{1}{2}\mathbf{A},$$

karena $\mathbf{D}^{-1/2} = \mathrm{diag}(1/\sqrt{2}, 1/\sqrt{2}, 1/\sqrt{2}, 1/\sqrt{2})$.

Teorema 2.9.5 (**Teorema Spektral Laplacian, (?)**) *Matriks Laplacian* $\mathbf{L}$ *adalah simetris dan positif semidefinit, sehingga dapat didekomposisi dengan dekomposisi nilai eigen sebagai berikut.*

$$\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\top}, \tag{2.9.3}$$

dengan $\mathbf{U}$ *ortogonal dan* $\mathbf{\Lambda}$ *diagonal berisi nilai eigen* $\lambda_1, \dots, \lambda_n \geq 0$*. Selalu berlaku* $\lambda_1 = 0$ *dengan vektor eigen* $\mathbf{1}$.

2.9.2 Kerangka Penyampaian Pesan pada Jaringan Saraf

Definisi 2.9.6 (***Message Passing Neural Networks*** **(MPNN), (?)**) *Misalkan graf* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *dengan* $\mathcal{V}$ *adalah himpunan simpul dan* $\mathcal{E}$ *adalah himpunan sisi. Misalkan pula* $\mathcal{N}_u$ *adalah ketetanggaan simpul* $u \in \mathcal{V}$*. Lebih lanjut, misalkan* $\boldsymbol{x}_u$ *adalah fitur simpul* u *dan* $\boldsymbol{e}_{uv}$ *adalah fitur sisi* $(u, v) \in \mathcal{E}$*. Suatu lapisan MPNN didefinisikan sebagai berikut.*

$$\boldsymbol{h}_u = \varphi\left(\boldsymbol{x}_u, \bigoplus_{v \in \mathcal{N}_u} \psi(\boldsymbol{x}_u, \boldsymbol{x}_v, \boldsymbol{e}_{uv}),\right) \tag{2.9.4}$$

dengan φ *dan* ψ *adalah fungsi yang terdiferensiasi (seperti jaringan saraf) dan* $\bigoplus$ *adalah operator agregasi yang bersifat komutatif dan asosiatif (seperti penjumlahan atau rata-rata) atau* permutation invariant.

Kerangka MPNN ini sangat fleksibel dan dapat disesuaikan dengan berbagai jenis data graf. Fungsi ψ dapat dirancang untuk menangkap interaksi spesifik antara simpul dan sisi, sedangkan fungsi φ dapat berupa jaringan saraf yang kompleks untuk pembaruan fitur. Operator agregasi $\bigoplus$ memastikan bahwa model tetap invariant terhadap permutasi tetangga, yang penting dalam konteks graf. Beberapa arsitektur umum GNN, seperti *Graph Convolutional Networks* (GCN) dan *Graph Attention Networks* (GAT), dapat dianggap sebagai kasus khusus dari kerangka MPNN ini.

2.9.3 Arsitektur Umum dalam GNN

Definisi 2.9.7 (**Graph Convolutional Networks (GCN), (?)**) *GCN mendefinisikan operasi propagasi lapisan ke-l sebagai*

$$\boldsymbol{H}^{(l+1)} = \sigma\left(\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-1/2}\boldsymbol{H}^{(l)}\mathbf{W}^{(l)}\right), \tag{2.9.5}$$

dengan:

- $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_n$ *matriks ketetanggaan dengan* self-loop*,*
- $\tilde{\mathbf{D}} = \mathrm{diag}(\sum_j \tilde{A}_{ij})$ *matriks derajat dari* $\tilde{\mathbf{A}}$*,*
- $\boldsymbol{H}^{(l)} \in \mathbb{R}^{n\times d_l}$ *representasi simpul pada lapisan l,*
- $\mathbf{W}^{(l)} \in \mathbb{R}^{d_l\times d_{l+1}}$ *bobot terlatih, dan*
- $\sigma(\cdot)$ *fungsi aktivasi non-linear.*

GCN merupakan kasus khusus dari kerangka MPNN dengan:

- *Fitur sisi diabaikan ($\boldsymbol{e}_{uv}$ tidak digunakan).*
- *Fungsi pesan $\psi(\boldsymbol{x}_u, \boldsymbol{x}_v, \boldsymbol{e}_{uv}) = \boldsymbol{x}_v$ (hanya mengambil fitur tetangga).*
- *Operator agregasi $\bigoplus$ berupa penjumlahan atau rata-rata tertimbang (melalui normalisasi Laplacian).*
- *Fungsi update φ adalah komposisi linear dan aktivasi non-linear: $\varphi(\boldsymbol{x}_u, \mathrm{AGG}) = \sigma(\mathrm{AGG}\cdot\mathbf{W})$.*

Secara eksplisit, propagasi GCN dapat ditulis sebagai

$$\boldsymbol{h}_u^{(l+1)} = \sigma\left(\sum_{v\in\mathcal{N}_u\cup\{u\}} \frac{1}{\sqrt{d_u d_v}}\boldsymbol{h}_v^{(l)}\mathbf{W}^{(l)}\right),$$

dengan d_u *derajat simpul* u *dan* σ *fungsi aktivasi. Matriks normalisasi simetris* $\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-1/2}$ *merepresentasikan agregasi pesan dari tetangga dengan bobot yang sesuai, sehingga GCN adalah MPNN dengan agregasi rata-rata dan update linear.*

Definisi 2.9.8 (Graph Attention Networks (GAT), (?)) *GAT mengganti normalisasi derajat dengan mekanisme perhatian. Koefisien perhatian untuk sisi* (i, j) *adalah*

$$\alpha_{ij} = \frac{\exp\big(\mathrm{LeakyReLU}\big(\boldsymbol{a}^\top[\mathbf{W}\boldsymbol{h}_i \,\|\, \mathbf{W}\boldsymbol{h}_j]\big)\big)}{\sum_{k\in\mathcal{N}(i)}\exp\big(\mathrm{LeakyReLU}\big(\mathbf{a}^\top[\mathbf{W}\boldsymbol{h}_i \,\|\, \mathbf{W}\boldsymbol{h}_k]\big)\big)}, \tag{2.9.6}$$

dengan:

- $\boldsymbol{h}_i \in \mathbb{R}^d$ *vektor fitur simpul* i*,*
- $\mathbf{W}$ *bobot transformasi linear,*
- $\boldsymbol{a} \in \mathbb{R}^{2d'}$ *vektor bobot perhatian, dan*
- $\|$ *operator konkatenasi.*

Pembaruan fitur simpul adalah

$$\mathrm{h}_i' = \sigma\left(\sum_{j\in\mathcal{N}(i)} \alpha_{ij}\mathbf{W}\boldsymbol{h}_j\right). \tag{2.9.7}$$

GAT merupakan kasus khusus dari kerangka MPNN dengan spesifikasi berikut.

- *Fitur sisi* $\boldsymbol{e}_{uv}$ *tidak digunakan (atau dapat diabaikan).*
- *Fungsi pesan* $\psi(\boldsymbol{x}_u, \boldsymbol{x}_v, \boldsymbol{e}_{uv}) = \mathbf{W}\boldsymbol{x}_v$ *(transformasi linear fitur tetangga).*
- *Operator agregasi* $\bigoplus$ *berupa penjumlahan tertimbang, dengan bobot agregasi* α_{uv} *diperoleh dari mekanisme perhatian (*attention*) yang bersifat permutation invariant.*

- *Fungsi update φ adalah komposisi linear dan aktivasi non-linear:* $\varphi(\boldsymbol{x}_u, \mathrm{AGG}) = \sigma(\mathrm{AGG})$.

Secara formal, pembaruan fitur simpul u pada GAT dapat ditulis dalam notasi MPNN sebagai

$$\boldsymbol{h}'_u = \sigma\left(\sum_{v \in \mathcal{N}_u} \alpha_{uv} \psi(\boldsymbol{x}_u, \boldsymbol{x}_v, \boldsymbol{e}_{uv})\right) = \sigma\left(\sum_{v \in \mathcal{N}_u} \alpha_{uv} \mathbf{W} \boldsymbol{x}_v\right), \tag{2.9.8}$$

dengan α_{uv} adalah skor perhatian yang bergantung pada fitur $\boldsymbol{x}_u$ dan $\boldsymbol{x}_v$ melalui fungsi $\mathrm{LeakyReLU}(\mathbf{a}^\top[\mathbf{W}\boldsymbol{x}_u \,\|\, \mathbf{W}\boldsymbol{x}_v])$ *dan normalisasi softmax. Dengan demikian, GAT adalah MPNN dengan agregasi penjumlahan tertimbang dan fungsi pesan linear dengan bobot agregasi ditentukan secara adaptif oleh mekanisme perhatian.*

Definisi 2.9.9 (GraphSAGE, (?)) *GraphSAGE atau* Graph Sample and Aggregation *mendefinisikan pembaruan simpul i sebagai*

$$\boldsymbol{h}_i^{(l+1)} = \sigma\Big(\boldsymbol{W}^{(l)} \cdot \mathrm{AGGREGATE}\left(\{\boldsymbol{h}_i^{(l)}\} \cup \{\boldsymbol{h}_j^{(l)} : j \in \mathcal{N}(i)\}\right)\Big), \tag{2.9.9}$$

dengan $\mathrm{AGGREGATE}$ *fungsi agregasi non-linear yang bisa berupa* mean*,* max-pooling*, atau LSTM.*

2.9.4 Kekuatan Representasi GNN

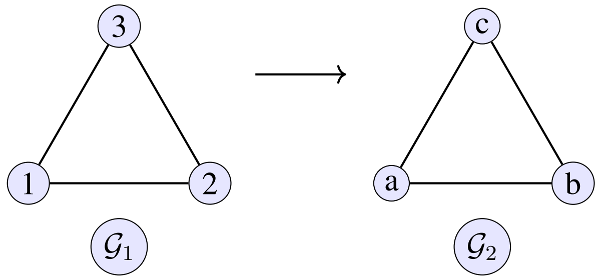
Definisi 2.9.10 (Isomorfisme Graf) *Dua graf $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ dan $\mathcal{G}_2 = (\mathcal{V}_2, \mathcal{E}_2)$ disebut isomorfik jika terdapat bijeksi $\pi : \mathcal{V}_1 \to \mathcal{V}_2$ sehingga*

$$(i, j) \in \mathcal{E}_1 \iff (\pi(i), \pi(j)) \in \mathcal{E}_2. \tag{2.9.10}$$

Sebuah fungsi f dikatakan invarian terhadap isomorfisme graf *jika $f(\mathcal{G}_1) = f(\mathcal{G}_2)$ untuk setiap graf isomorfik $\mathcal{G}_1 \cong \mathcal{G}_2$.*

Contoh 2.9.11 Misalkan $\mathcal{G}_1$ adalah graf dengan simpul $\{1, 2, 3\}$ dan sisi $\{(1, 2), (2, 3), (3, 1)\}$ (graf segitiga). Ambil $\mathcal{G}_2$ dengan simpul $\{a, b, c\}$ dan sisi $\{(a, b), (b, c), (c, a)\}$. Bijeksi π dapat dipilih misal $\pi(1) = a$, $\pi(2) = b$, $\pi(3) = c$. Maka $\mathcal{G}_1$ dan $\mathcal{G}_2$ isomorfik.

Sebagai contoh fungsi invarian, jumlah simpul $|\mathcal{V}|$ dan jumlah sisi $|\mathcal{E}|$ adalah invarian terhadap isomorfisme graf. Demikian juga, spektrum Laplacian graf $\mathbf{L}$ adalah invarian terhadap isomorfisme.



Gambar 2.6 Ilustrasi dua graf isomorfik, yaitu $\mathcal{G}_1$ dan $\mathcal{G}_2$ memiliki struktur yang sama walaupun label simpul berbeda. (Sumber: Dokumen penulis)

Definisi 2.9.12 (**Tes Weisfeiler–Lehman (WL) Satu Dimensi**) *Algoritma WL satu dimensi, juga dikenal sebagai* color refinement*, merupakan prosedur iteratif untuk membedakan graf berdasarkan struktur lokal simpul. Pada setiap iterasi, label simpul $h_i^{(t)}$ diperbarui dengan cara:*

$$h_i^{(t+1)} = \mathrm{HASH}\Big(h_i^{(t)}, \{h_j^{(t)} : j \in \mathcal{N}(i)\}\Big), \tag{2.9.11}$$

dimulai dari label awal $h_i^{(0)}$. Dua graf dikatakan dapat dibedakan oleh tes WL jika multiset label akhirnya berbeda. Dalam konteks pembelajaran mesin, tes WL memberikan kerangka formal untuk mengukur kemampuan model, seperti GNN, dalam membedakan struktur graf yang berbeda melalui proses agregasi informasi dari tetangga.

Fungsi HASH pada algoritma WL berperan sebagai pemetaan yang menggabungkan label simpul saat ini dengan label-label tetangganya menjadi label baru yang unik. Intuisi utamanya adalah memastikan bahwa jika dua simpul memiliki lingkungan yang berbeda, hasil HASH juga berbeda. Dalam implementasi praktis, HASH dapat berupa fungsi injektif seperti pengkodean *string*, *concatenation*, atau fungsi neural yang mampu membedakan multiset masukan secara efektif.

Teorema 2.9.13 (**Keterbatasan GNN Berbasis Agregasi, (?)**) *GNN berbasis agre-*

gasi dengan pembaruan umum

$$h_i^{(l+1)} = \varphi\Big(W^{(l)} \cdot \mathrm{AGGREGATE}\big(\{h_i^{(l)}\} \cup \{h_j^{(l)} : j \in \mathcal{N}(i)\}\big)\Big) \tag{2.9.12}$$

tidak lebih kuat daripada tes WL satu dimensi dalam membedakan graf yang tidak isomorfik. Artinya, kemampuan GNN untuk membedakan struktur graf dibatasi oleh kekuatan agregasi lokal yang serupa dengan tes WL.

Teorema 2.9.14 (Hubungan dengan Tes WL, (?)) *Jika fungsi agregasi* $\mathrm{AGGREGATE}$ *bersifat injektif terhadap multiset, maka GNN memiliki kekuatan representasi setara dengan tes WL satu dimensi. Dengan kata lain, GNN dapat membedakan graf sejauh tes WL mampu membedakannya, asalkan agregasi dan pembaruan fitur dilakukan secara unik untuk setiap lingkungan simpul.*

BAB III

KERANGKA METODOLOGIS DAN ANALISIS ASIMTOTIK REGRESI TERBOBOTI GEOGRAFIS DENGAN KERNEL SPASIAL TERPELAJARI BERBASIS JARINGAN SARAF GRAF

3.1 Keterbatasan Pembobotan Spasial Berbasis Kernel Konvensional

3.1.1 Prinsip Pembobotan Spasial dalam Regresi Terbobot Geografis Klasik

Regresi Terbobot Geografis (Geographically Weighted Regression, GWR) merupakan perluasan lokal dari model regresi linear klasik yang memungkinkan koefisien regresi bervariasi secara spasial (**?**). Misalkan $(u_i, v_i) \in \mathbb{R}^2$ menyatakan koordinat spasial observasi ke-i. Model GWR dituliskan sebagai

$$Y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}(u_i, v_i) + \varepsilon_i, \tag{3.1.1}$$

dengan $\boldsymbol{\beta}(u, v)$ adalah vektor koefisien regresi lokal yang bergantung pada lokasi.

Estimator koefisien lokal pada lokasi target (u_0, v_0) diperoleh melalui metode *locally weighted least squares* (LWLS), yaitu

$$\widehat{\boldsymbol{\beta}}(u_0, v_0) = \left(X^\top W(u_0, v_0) X\right)^{-1} X^\top W(u_0, v_0) Y, \tag{3.1.2}$$

dengan $W(u_0, v_0)$ adalah matriks diagonal bobot spasial.

Definisi 3.1.1 (Kernel Spasial Konvensional) *Fungsi bobot spasial disebut kernel konvensional apabila bobot antarobservasi i dan lokasi target (u_0, v_0) ditentukan oleh*

$$w_i(u_0, v_0) = K\left(\frac{d_i(u_0, v_0)}{h}\right), \tag{3.1.3}$$

dengan $d_i(u_0, v_0)$ adalah jarak Euclidean antara (u_i, v_i) dan (u_0, v_0), $h > 0$ adalah *parameter* bandwidth*, dan $K(\cdot)$ adalah fungsi kernel tetap seperti Gaussian atau bisquare (**?**).*

Pendekatan ini menyiratkan bahwa struktur dependensi spasial sepenuhnya ditentukan oleh jarak geometris dan parameter bandwidth, yang diasumsikan tetap dan diketahui.

3.1.2 Bias Induktif pada Kernel Tetap

Kernel spasial konvensional membawa asumsi struktural yang bersifat *a priori*, terutama isotropi dan monotonisitas pengaruh terhadap jarak Euclidean (**??**). Asumsi ini membatasi kelas fungsi kernel yang dapat digunakan dalam estimasi, sehingga menimbulkan apa yang dikenal sebagai *bias induktif*.

Definisi 3.1.2 (Bias Induktif) *Bias induktif adalah kecenderungan sistematis suatu metode estimasi yang disebabkan oleh pembatasan kelas fungsi atau struktur model yang diasumsikan sebelum data diamati.*

Dalam konteks GWR, bias induktif muncul karena kernel tetap membatasi bobot spasial pada fungsi jarak isotropik tertentu, sehingga tidak mampu merepresentasikan struktur spasial yang bersifat anisotropik, tidak simetris, atau bergantung pada konteks non-geometris (**?**).

Secara matematis, misalkan $\kappa^\star(u_i, u_0)$ menyatakan kernel spasial sejati yang mendasari proses data. Penggunaan kernel konvensional menyiratkan aproksimasi

$$\kappa^\star(u_i, u_0) \approx K\left(\frac{\|u_i - u_0\|}{h}\right), \tag{3.1.4}$$

dengan $K(\cdot)$ dipilih dari kelas fungsi terbatas. Jika $\kappa^\star$ tidak berada dalam kelas tersebut, maka kesalahan aproksimasi tidak dapat dieliminasi meskipun ukuran sampel meningkat.

Keterbatasan ini diperparah oleh penggunaan bandwidth tunggal yang bersifat global, yang mengasumsikan bahwa seluruh hubungan spasial beroperasi pada skala yang sama, padahal banyak proses spasial bersifat multi-skala (**?**).

3.1.3 Implikasi terhadap Inferensi Statistik

Keterbatasan representasi kernel tetap tidak hanya berdampak pada akurasi estimasi koefisien lokal, tetapi juga memiliki implikasi langsung terhadap validitas inferensi statistik.

Inferensi dalam GWR klasik umumnya dilakukan dengan memperlakukan matriks bobot $W(u_0, v_0)$ sebagai kuantitas tetap dan tidak acak. Di bawah asumsi ini, sifat asimtotik koefisien lokal dapat diturunkan menggunakan teori regresi lokal (**?**). Namun, apabila kernel spasial yang digunakan tidak sesuai dengan struktur dependensi spasial yang sesungguhnya, maka estimator koefisien lokal dapat mengalami bias sistematis akibat fenomena *data borrowing* (**?**).

Lebih lanjut, kesalahan spesifikasi kernel dapat menyebabkan distorsi pada estimasi variansi dan distribusi asimtotik koefisien lokal, sehingga uji signifikansi berbasis pendekatan asimtotik menjadi tidak valid (**?**). Kondisi ini menjelaskan mengapa hingga saat ini tidak terdapat konsensus yang mapan mengenai inferensi statistik dalam GWR konvensional (**?**).

Keterbatasan-keterbatasan tersebut memotivasi pengembangan pendekatan pembobotan spasial yang lebih fleksibel dan berbasis data, di mana struktur kernel tidak ditentukan secara *ad hoc*, tetapi dipelajari dari data.

3.2 Pembelajaran Kernel Spasial dengan Jaringan Saraf Graf

3.2.1 Representasi Graf untuk Data Spasial

3.2.2 Pembelajaran Pola Interaksi Spasial melalui Jaringan Saraf Graf

3.2.3 Pendugaan Kernel Spasial Berbasis GNN

3.3 Integrasi Kernel Terestimasi ke dalam Kerangka GWR

3.3.1 Definisi Model *Estimated Kernel*-GWR (EK-GWR)

3.3.2 Kernel Terestimasi sebagai *Nuisance Parameter*

3.3.3 Permasalahan Endogenitas dan Ketergantungan Data

3.3.4 Skema *Cross-Fitting* untuk Inferensi Valid

3.4 Analisis Asimtotik Koefisien Lokal pada EK-GWR

3.4.1 Asumsi-Asumsi Regularitas

3.4.2 Analisis Ketakbiasan dan Bias Asimtotik Koefisien Lokal

3.4.3 Analisis Ketakbiasan dan Bias Asimtotik Variansi Galat

3.4.4 Konsistensi Koefisien Lokal

3.4.5 Konsistensi Variansi Koefisien Lokal

3.4.6 Distribusi Asimtotik Koefisien Lokal

3.4.7 Distribusi Asimtotik Uji Statistik

3.5 Komputasional EK-GWR

3.5.1 Algoritma Pelatihan Jaringan Saraf Graf untuk Estimasi Kernel

3.5.2 Pemilihan *Bandwidth* dan *Hyperparameter*

3.5.3 Diagnostik Model

BAB IV

STUDI KASUS

4.1 Catatan Penting

ISINYA YAA PEMBAHASAN STUDI KASUS KAMUUUU.

4.2 Pembahasan

4.2.1 Pembahasan 1

Dalam membuat tabel, disarankan memakai ”longtable” agar tabel nya bisa dipotong halaman. Biar lebih mudah kamu bisa pakai web https://www.tablesgenerator.com/latex_tables#google_vignette.

Contoh input gambar

Gambar 4.1 SKRIPSI TU DIKERJAIN

4.2.2 Pembahasan 2

Beberapa catatan penulisan yang wajib diperhatikan sebagai berikut.

- Penggunaan kata ’adalah’ dan ’merupakan’

- Penulisan tanda baca seperti titik yang wajib ada di setiap akhir persamaan
- Dalam kalimat '... dihitung menggunakan persamaan berikut.' di akhir kata berikut harus ada 'titik'
- Kamu bisa pakai 'begin{equation}' atau '$$' atau '$$ begin{aligned} ... end{aligned} $$' dalam menuliskan persamaan, tinggal pilih mana yang kamu butuhkan. Contoh

$$\begin{aligned}
(\boldsymbol{A})_{22} &= (\boldsymbol{C})_{22} + \min\{(\boldsymbol{A})_{2(2-1)}, (\boldsymbol{A})_{(2-1)2}, (\boldsymbol{A})_{(2-1)(2-1)}\} \\
&= (\boldsymbol{C})_{22} + \min\{(\boldsymbol{A})_{21}, (\boldsymbol{A})_{12}, (\boldsymbol{A})_{11}\} \\
&= 2 + \min\{3, 5, 2\} \\
&= 2 + 2 = 4.
\end{aligned}$$

$$(\boldsymbol{A})_{11} = (\boldsymbol{C})_{11}. \tag{4.2.1}$$

$$A = (5,\ 6,\ 5,\ 7,\ 6,\ 6,\ 6,\ 6) \qquad B = (7,\ 8,\ 6,\ 10,\ 10,\ 10,\ 8,\ 8)$$

- Pemanggilan persamaan, tabel, dan gambar, WAJIB menggunakan huruf besar di depan. Contohnya adalah '... dapat dihitung menggunakan Persamaan xx', 'Dari Tabel xx', 'Berdasarkan Gambar xx'

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan pembahasan sebelumnya serta analisis yang telah dilakukan, diperoleh kesimpulan sebagai berikut.

1. Kesimpulan 1.
2. Kesimpulan 2.
3. Kesimpulan 3.

5.2 Saran

Saran untuk penelitian selanjutnya

LAMPIRAN A

Data

LAMPIRAN B

Syntax R

```
SYNTAX  R  ALALALALALLALA
```