

**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 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 mendapatkan gelar sarjana di program studi statistika Universitas 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 bermanfaat untuk pembaca. Terima kasih.

Yogyakarta, Juni 2024

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL</b>	<b>i</b>
<b>HALAMAN PENGESAHAN</b>	<b>ii</b>
<b>HALAMAN PERNYATAAN BEBAS PLAGIASI</b>	<b>iii</b>
<b>HALAMAN PERSEMBAHAN</b>	<b>iv</b>
<b>HALAMAN MOTTO</b>	<b>v</b>
<b>PRAKATA</b>	<b>vi</b>
<b>DAFTAR ISI</b>	<b>vii</b>
<b>DAFTAR TABEL</b>	<b>xi</b>
<b>DAFTAR GAMBAR</b>	<b>xii</b>
<b>INTISARI</b>	<b>xiii</b>
<b>ABSTRACT</b>	<b>xiv</b>
<b>I PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang	1
1.2 Pembatasan Masalah	3
1.3 Tujuan dan Manfaat Penelitian	3
1.4 Tinjauan Pustaka	4
1.5 Metodologi Penelitian	4
1.6 Sistematika Penulisan	5
<b>II LANDASAN TEORI</b>	<b>6</b>
2.1 Aljabar Matriks	6
2.1.1 Ruang Vektor dan Matriks	6
2.1.2 Sifat-Sifat Matriks	13
2.1.3 Invers Matriks	22
2.1.4 Determinan dan <i>Rank</i>	31
2.1.5 Nilai Eigen dan Vektor Eigen	40
2.1.6 Perkalian Kronecker dan Operasi Vektorisasi	43
2.1.7 Norma Vektor dan Matriks	45
2.1.8 Diferensial Vektor dan Matriks	48
2.2 Teori Ukuran dan Probabilitas	57
2.2.1 Ruang Ukur dan Ukuran Probabilitas	57
2.2.2 Variabel Acak sebagai Fungsi Terukur	60
2.2.3 Distribusi dan Ekspektasi	62
2.2.4 Momen, Variansi, dan Kovariansi	71



2.2.5	Probabilitas Bersyarat dan Independensi . . . . .	73
2.2.6	Hukum Probabilitas Total dan Teorema Bayes . . . . .	76
2.2.7	Hukum Ekspektasi Total dan Variansi Total . . . . .	79
2.3	Konvergensi dan Laju Pertumbuhan . . . . .	82
2.3.1	Barisan Variabel Acak . . . . .	83
2.3.2	Konvergensi Barisan Variabel Acak . . . . .	84
2.3.3	Hubungan Antarmode Konvergensi . . . . .	86
2.4	Teori Probabilitas Asimtotik . . . . .	90
2.4.1	Hukum Bilangan Besar . . . . .	90
2.4.2	Konsistensi Estimator . . . . .	91
2.4.3	Fungsi Karakteristik . . . . .	91
2.4.4	Teorema Pemetaan Kontinu dan Teorema Slutsky . . . . .	95
2.4.5	Notasi <i>Big-O</i> dan <i>Little-o</i> (Deterministik) . . . . .	101
2.4.6	Notasi <i>Big-O</i> dan <i>Little-o</i> dalam Probabilitas . . . . .	104
2.4.7	Teorema Limit Pusat . . . . .	111
2.4.8	Metode Delta . . . . .	118
2.5	Analisis Regresi Linear . . . . .	120
2.5.1	Regresi Linear Biasa atau <i>Ordinary Least Squares</i> (OLS) . . . . .	120
2.5.2	Teori Asimtotik untuk OLS . . . . .	126
2.5.3	Regresi Terboboti . . . . .	131
2.5.4	Teori Asimtotik untuk GLS/WLS . . . . .	135
2.6	Pemodelan Polinomial Lokal . . . . .	137
2.6.1	Asumsi Model Polinomial Lokal . . . . .	139
2.6.2	Estimasi Model Polinomial Lokal . . . . .	140
2.7	Analisis Regresi Spasial . . . . .	142
2.7.1	Regresi dengan Dependensi Spasial . . . . .	143
2.7.2	Regresi dengan Keragaman Spasial . . . . .	145
2.8	Jaringan Saraf Tiruan . . . . .	165
2.8.1	Model Dasar dan Notasi . . . . .	165
2.8.2	Fungsi Aktivasi . . . . .	166
2.8.3	Fungsi Kerugian dan Kriteria Pembelajaran . . . . .	169
2.8.4	Pembelajaran dengan Propagasi Mundur . . . . .	170
2.8.5	Optimisasi Parameter . . . . .	172
2.8.6	Jaringan Saraf Tiruan sebagai Aproksimasi Universal . . . . .	178
2.9	Jaringan Saraf Graf . . . . .	181
2.9.1	Dasar Graf dan Laplacian . . . . .	181

2.9.2	Kerangka Penyampaian Pesan pada Jaringan Saraf . . . . .	183
2.9.3	Arsitektur Umum dalam GNN . . . . .	184
2.9.4	Sifat Simetri dan Invariansi Permutasi . . . . .	186
2.9.5	Daya Aproksimasi Jaringan Saraf Graf . . . . .	186
<b>III KERANGKA METODOLOGIS DAN ANALISIS ASIMTOTIK GRAPH ATTENTION-BASED GEOGRAPHICALLY WEIGHTED REGRESSION . . . . .</b>		<b>188</b>
3.1	Keterbatasan Pembobotan Spasial Berbasis <i>Kernel</i> Konvensional . .	188
3.1.1	Prinsip Pembobotan Spasial dalam GWR . . . . .	188
3.1.2	Bias Induktif pada <i>Kernel</i> Tetap . . . . .	189
3.2	Pembelajaran <i>Kernel</i> Spasial dengan Jaringan Saraf Graf . . . . .	191
3.2.1	Representasi Graf untuk Data Spasial . . . . .	191
3.2.2	Fleksibilitas Metrik Jarak dalam Konstruksi Representasi . .	193
3.2.3	Fungsi Skor Spasial dan Jaringan Saraf Graf . . . . .	196
3.2.4	Konstruksi <i>Kernel</i> Spasial Berbasis GNN . . . . .	197
3.3	Integrasi <i>Kernel</i> Terestimasi ke dalam Kerangka GWR . . . . .	202
3.3.1	Definisi Model <i>Graph Attention-based</i> GWR (GA-GWR) .	202
3.3.2	<i>Kernel</i> Terestimasi sebagai <i>Nuisance Parameter</i> . . . . .	203
3.3.3	Permasalahan Endogenitas dan Ketergantungan Data . . . .	203
3.3.4	Skema <i>Cross-Fitting</i> untuk Inferensi Valid . . . . .	206
3.4	Analisis Asimtotik Koefisien Lokal pada GA-GWR . . . . .	212
3.4.1	Asumsi-Asumsi Regularitas . . . . .	212
3.4.2	Analisis Bias Asimtotik Koefisien Lokal . . . . .	216
3.4.3	Analisis Bias Asimtotik Variansi Galat . . . . .	220
3.4.4	Konsistensi Koefisien Lokal . . . . .	225
3.4.5	Konsistensi Variansi Koefisien Lokal . . . . .	228
3.4.6	Distribusi Asimtotik Koefisien Lokal . . . . .	233
3.4.7	Distribusi Asimtotik Uji Statistik . . . . .	238
3.5	Analisis Komputasional GA-GWR . . . . .	242
3.5.1	Graf Komputasional dan Diferensiabilitas . . . . .	243
3.5.2	Analisis <i>Vanishing</i> dan <i>Exploding Gradient</i> . . . . .	246
3.5.3	Konstanta Lipschitz dan Stabilitas <i>Pipeline</i> . . . . .	249
3.5.4	Algoritma Pelatihan dan Kompleksitas Komputasional . . .	251
3.5.5	Diagnostik Model dan Validasi Inferensi . . . . .	254
<b>IV STUDI KASUS . . . . .</b>		<b>259</b>
4.1	Catatan Penting . . . . .	259

4.2	Pembahasan . . . . .	259
4.2.1	Pembahasan 1 . . . . .	259
4.2.2	Pembahasan 2 . . . . .	259
<b>V</b>	<b>PENUTUP . . . . .</b>	<b>261</b>
5.1	Kesimpulan . . . . .	261
5.2	Saran . . . . .	261
	<b>DAFTAR PUSTAKA . . . . .</b>	<b>262</b>
<b>A</b>	<b>Data . . . . .</b>	<b>266</b>
<b>B</b>	<b>Syntax R . . . . .</b>	<b>267</b>

## **DAFTAR TABEL**

## DAFTAR GAMBAR

2.1	Ilustrasi transformasi linear $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dengan $f(x, y) = (2x, y)$ .	32
2.2	Ilustrasi bentuk bilinear anti-simetri $\omega(v_1, v_2) = x_1y_2 - x_2y_1$ pada $\mathbb{R}^2$ .	36
2.3	Perbandingan luas parallelogram sebelum dan sesudah transformasi $A$ .	38
2.4	Ilustrasi vektor eigen dan nilai eigen untuk transformasi $T(x, y) = (2x, 3y)$ .	41
2.5	Diagram hubungan antarmode konvergensi.	87
4.1	SKRIPSI TU DIKERJAIN	259

## **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 kunci.

## **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 Amri (2020) 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 (Anselin, 1988). Untuk mengakomodasi hal tersebut, Fotheringham, Brunson, dan Charlton (2002) mengembangkan model *Geographically Weighted Regression* (GWR) yang memungkinkan koefisien regresi bervariasi antar lokasi. Model ini kemudian diperluas oleh Huang, Wu, dan Barry (2010) 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 nonlinier yang kompleks antara jarak spasial dan bobot pembobotan.

Perkembangan terkini di bidang *geospatial artificial intelligence* (GeoAI) menghadirkan pendekatan *machine learning* yang dapat mempelajari pola nonline-



ar secara adaptif. Salah satunya adalah *Geographically Weighted Artificial Neural Network* (GWANN) yang dikembangkan oleh Hagenauer dan Helbich (2022) serta *Spatial and Attribute Neural Network Weighted Regression* (SANNWR) oleh Ni, Wu, dan Zhang (2022). Keduanya berupaya memanfaatkan jaringan saraf tiruan untuk mengestimasi fungsi pembobot spasial yang kompleks. Lebih lanjut, Yin, Huang, Du, Wu, dkk. (2024) 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, Zhu, Liu, Yao, dan Fischer (2022) 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 antar-unit 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 antarpro-

vinsi 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 (Fotheringham dkk., 2002) dan GTWR (Huang dkk., 2010) 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 (Hagenauer dan Helbich, 2022), SANNWR (Ni dkk., 2022), dan GNNWR (Yin dkk., 2024). Di sisi lain, pendekatan *graph-based learning* seperti SR-GCNN (Zhu dkk., 2022) 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, Lang 1987)** Misalkan  $\mathcal{F}$  adalah subhimpunan dari bilangan kompleks  $\mathbb{C}$ ,  $\mathcal{F}$  disebut sebagai lapangan apabila memenuhi kondisi berikut.

- (a) Apabila  $x, y$  adalah elemen dari  $\mathcal{F}$ , maka  $x + y$  dan  $xy$  juga merupakan elemen dari  $\mathcal{F}$ .
- (b) Apabila  $x \in \mathcal{F}$ , maka  $-x$  juga merupakan elemen dari  $\mathcal{F}$ . Lebih lanjut, jika

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

(c) 0 dan 1 merupakan elemen dari  $\mathcal{F}$ .

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

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

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

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

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

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

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  $\mathcal{F}$  dan  $\mathcal{G}$  keduanya merupakan lapangan, serta dimisalkan bahwa  $\mathcal{F} \subseteq \mathcal{G}$ , maka  $\mathcal{F}$  dapat disebut sebagai sublapangan atau *subfield* dari  $\mathcal{G}$ . Elemen-elemen dari lapangan disebut sebagai bilangan atau skalar (Lang, 1987).

Definisi terkait lapangan atau *field* juga dirumuskan oleh Billingsley (1995)

yang digunakan dalam konteks teori ukuran dan probabilitas.

**Definisi 2.1.3 (Lapangan, Billingsley 1995)** Suatu keluarga  $\mathcal{F}$  dari kejadian-kejadian dalam ruang sampel  $\Omega$  disebut sebagai lapangan (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 ( $\emptyset$ ) dan himpunan dari seluruh ruang sampel ( $\Omega$ ) selalu merupakan elemen dari lapangan  $\mathcal{F}$ . Selain itu, lapangan juga tertutup terhadap operasi irisan, yaitu jika  $A, B \in \mathcal{F}$ , maka  $A \cap B \in \mathcal{F}$ .

**Contoh 2.1.4** Diberikan ruang sampel atau semesta  $\Omega = \{1, 2, 3\}$  dan himpunan  $\mathcal{F}_1, \mathcal{F}_2$  berikut.

$$\mathcal{F}_1 = \{\emptyset, \Omega, \{1\}, \{2\}, \{2, 3\}\},$$

$$\mathcal{F}_2 = \{\emptyset, \Omega, \{2\}, \{1, 3\}\}.$$

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

**Definisi 2.1.5 (Ruang vektor, Lang 1987)** Suatu ruang vektor  $\mathcal{V}$  atas suatu lapangan  $\mathcal{F}$  adalah sebuah himpunan objek yang dapat dijumlahkan dan dikalikan dengan elemen-elemen dari  $\mathcal{F}$ , sedemikian rupa sehingga hasil penjumlahan dua elemen dari  $\mathcal{V}$  merupakan elemen  $\mathcal{V}$  kembali, dan hasil perkalian sebuah elemen  $\mathcal{V}$  dengan sebuah elemen dari  $\mathcal{F}$  juga merupakan elemen dari  $\mathcal{V}$ . Selain itu, sifat-sifat berikut dipenuhi:

- (a)  $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{V}$  berlaku  $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ .
- (b)  $\exists \mathbf{0} \in \mathcal{V}$  sedemikian sehingga  $\forall \mathbf{u} \in \mathcal{V}$  berlaku  $\mathbf{u} + \mathbf{0} = \mathbf{u}$ .

- (c)  $\forall \mathbf{u} \in \mathcal{V}, \exists -\mathbf{u} \in \mathcal{V}$  sedemikian sehingga  $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$ .
- (d)  $\forall \mathbf{u}, \mathbf{v} \in \mathcal{V}$  berlaku  $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ .
- (e)  $\forall c \in \mathcal{F}$  dan  $\forall \mathbf{u}, \mathbf{v} \in \mathcal{V}$  berlaku  $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$ .
- (f)  $\forall a, b \in \mathcal{F}$  dan  $\forall \mathbf{v} \in \mathcal{V}$  berlaku  $(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$ .
- (g)  $\forall a, b \in \mathcal{F}$  dan  $\forall \mathbf{v} \in \mathcal{V}$  berlaku  $(ab)\mathbf{v} = a(b\mathbf{v})$ .
- (h)  $\forall \mathbf{u} \in \mathcal{V}$  berlaku  $1 \cdot \mathbf{u} = \mathbf{u}$ , dengan 1 adalah elemen identitas pada  $\mathcal{F}$ .

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 2.1.6 (Matriks, Lang 1987)** Misalkan  $\mathcal{F}$  merupakan lapangan dan  $m, n \geq 1 \in \mathbb{Z}$ . Matriks adalah suatu larik atau array dari bilangan-bilangan dalam  $\mathcal{F}$  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  $(a_{ij})$ ,  $i = 1, 2, \dots, m$  dan  $j = 1, 2, \dots, n$ . 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  $(a_{11}, a_{12}, \dots, a_{1n})$ . 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 setiap  $i, j$ .

**Definisi 2.1.7 (Kebebasan linear, Lang 1987)** Misalkan  $\mathcal{V}$  ruang vektor atas lapangan  $\mathcal{F}$  dan misalkan pula  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  elemen-elemen dari  $\mathcal{V}$ , maka  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  dikatakan bebas linear jika dan hanya jika  $a_1, a_2, \dots, a_n$  merupakan bilangan sehingga

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n = \mathbf{0}, \quad (2.1.4)$$

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

Vektor  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  dikatakan bergantung linear terhadap  $\mathcal{F}$  apabila terdapat elemen-elemen  $a_1, a_2, \dots, a_n$  di  $\mathcal{F}$  yang semuanya tidak sama dengan nol, sedemikian sehingga Persamaan (2.1.4) terpenuhi.

**Contoh 2.1.8** Misalkan  $\mathcal{V} = \mathcal{F}^n$  dan vektor  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  didefinisikan sebagai berikut.

$$\mathbf{v}_1 = (1, 0, \dots, 0)$$

$$\vdots$$

$$\mathbf{v}_n = (0, 0, \dots, 1).$$

Vektor-vektor  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  dikatakan bebas linear. Misalkan  $a_1, a_2, \dots, a_n$  merupakan bilangan-bilangan sedemikian sehingga Persamaan (2.1.4) terpenuhi. Sebab

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n = (a_1, a_2, \dots, a_n), \quad (2.1.5)$$

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

**Definisi 2.1.9 (Span atau jangkauan linear, Axler 2024)** Misalkan  $\mathcal{V}$  ruang vektor atas lapangan  $\mathcal{F}$  dan  $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\} \subseteq \mathcal{V}$ . Span dari  $S$ , ditulis  $\text{span}(S)$ ,

adalah himpunan semua kombinasi linear dari elemen-elemen  $S$ , yaitu

$$\text{span}(S) = \left\{ \sum_{i=1}^m a_i \mathbf{v}_i : a_i \in \mathcal{F} \right\}. \quad (2.1.6)$$

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

**Contoh 2.1.10** Pada  $\mathbb{R}^2$ , ambil  $S = \{(1, 0), (0, 1)\}$ , maka

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

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

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

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

**Definisi 2.1.11 (Basis, Axler 2024)** Suatu basis  $\mathcal{V}$  adalah himpunan vektor-vektor di  $\mathcal{V}$  yang bebas linear dan menjangkau (*spanning*)  $\mathcal{V}$ .

**Contoh 2.1.12** 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  $\text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$  paling jauh merupakan subruang berdimensi 2, yaitu sebuah bidang melalui titik asal di  $\mathbb{R}^3$ . Sebab  $\text{span}\{\mathbf{v}_1, \mathbf{v}_2\} \subsetneq \mathbb{R}^3$ , maka himpunan ini tidak dapat dijadikan basis dari  $\mathbb{R}^3$ .

**Teorema 2.1.13 (Axler, 2024)** Sebuah list dari vektor-vektor  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  merupakan basis dari  $\mathcal{V}$  jika dan hanya jika  $\forall \mathbf{v} \in \mathcal{V}$  dapat dituliskan secara unik dalam bentuk

$$\mathbf{v} = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \dots + a_n \mathbf{v}_n, \quad (2.1.9)$$

dengan  $a_1, a_2, \dots, a_n \in \mathcal{F}$ , dengan  $\mathcal{F}$  dapat berupa  $\mathbb{R}$  ataupun  $\mathbb{C}$ .

**Bukti.** ( $\Rightarrow$ ) Pertama andaikan bahwa  $\mathbf{v}_1, \dots, \mathbf{v}_n$  adalah basis dari  $\mathcal{V}$ . Sebab  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  menjangkau  $\mathcal{V}$ , maka untuk setiap  $\mathbf{v} \in \mathcal{V}$  terdapat skalar  $a_1, \dots, a_n \in \mathcal{F}$  sehingga Persamaan (2.1.9) terpenuhi. Untuk menunjukkan bahwa representasi tersebut unik, misalkan terdapat skalar  $c_1, \dots, c_n \in \mathcal{F}$  dengan

$$\mathbf{v} = c_1 \mathbf{v}_1 + \dots + c_n \mathbf{v}_n. \quad (2.1.10)$$

Dengan mengurangkan Persamaan (2.1.10) dari Persamaan (2.1.9) akan diperoleh

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

Sebab  $\mathbf{v}_1, \dots, \mathbf{v}_n$  bebas linear, maka setiap koefisien harus nol, yaitu  $a_k - c_k = 0$  untuk  $k = 1, \dots, n$ . Dengan demikian  $a_1 = c_1, \dots, a_n = c_n$ , sehingga representasi (2.1.9) bersifat unik. Hal ini menyelesaikan pembuktian arah pertama.

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

$$\mathbf{v} = a_1 \mathbf{v}_1 + \dots + a_n \mathbf{v}_n, \quad (2.1.12)$$

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

$$0 = a_1 \mathbf{v}_1 + \dots + a_n \mathbf{v}_n. \quad (2.1.13)$$

Sebab representasi (2.1.12) bersifat unik, khususnya untuk  $\mathbf{v} = 0$ , maka harus berlaku  $a_1 = \dots = a_n = 0$ . Jadi  $\mathbf{v}_1, \dots, \mathbf{v}_n$  bebas linear. Dengan demikian  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  adalah basis  $\mathcal{V}$ . ■

**Definisi 2.1.14 (Ruang vektor berdimensi-hingga, Axler 2024)** Suatu ruang vektor  $\mathcal{V}$  atas lapangan  $\mathcal{F}$  dikatakan berdimensi-hingga apabila terdapat suatu basis  $\{v_1, \dots, v_n\}$  dari  $\mathcal{V}$  yang terdiri atas sejumlah hingga ( $n$  buah) vektor. Dengan kata lain,  $\mathcal{V}$  berdimensi-hingga jika  $\mathcal{V}$  dapat dijangkau oleh suatu himpunan vektor berukuran hingga.

**Definisi 2.1.15 (Dimensi, Axler 2024)** *Dimensi dari suatu ruang vektor berdimensi-hingga  $\mathcal{V}$  adalah banyaknya elemen dalam setiap basis  $\mathcal{V}$ . Sebab semua basis dari  $\mathcal{V}$  memiliki panjang yang sama, maka bilangan ini terdefinisi dengan baik. Dimensi  $\mathcal{V}$  dilambangkan dengan  $\dim \mathcal{V}$ .*

**Contoh 2.1.16** 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  $\mathcal{F}$  (umumnya  $\mathcal{F} = \mathbb{R}$  atau  $\mathcal{F} = \mathbb{C}$ ). Untuk  $m, n \geq 1$ , himpunan seluruh matriks  $m \times n$  berelemen di  $\mathcal{F}$  dilambangkan dengan  $\text{Mat}_{m \times n}(\mathcal{F})$ . Dengan penjumlahan dan perkalian skalar yang didefinisikan komponen-demi-komponen,  $\text{Mat}_{m \times n}(\mathcal{F})$  membentuk ruang vektor atas lapangan  $\mathcal{F}$  (Axler, 2024).

**Definisi 2.1.17** Untuk  $\mathbf{A} = (a_{ij}), \mathbf{B} = (b_{ij}) \in \text{Mat}_{m \times n}(\mathcal{F})$  dan  $c \in \mathcal{F}$ , 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).
- (c)  $(c\mathbf{A})_{ij} = c a_{ij}$ .

**Definisi 2.1.18 (Perkalian matriks, Axler 2024)** Jika  $\mathbf{A} \in \text{Mat}_{m \times n}(\mathcal{F})$  dan  $\mathbf{B} \in \text{Mat}_{n \times p}(\mathcal{F})$ , maka hasil kali  $\mathbf{AB} \in \text{Mat}_{m \times p}(\mathcal{F})$  didefinisikan oleh

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

**Contoh 2.1.19** Perkalian matriks berukuran  $3 \times 2$  dan  $2 \times 4$ :

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

**Proposisi 2.1.20 (Axler, 2024)** 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  $\mathbf{A} \mathbf{I}_n = \mathbf{A}$  untuk  $\mathbf{A} \in \text{Mat}_{m \times n}(\mathcal{F})$ ; 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

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

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

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

Keduanya merupakan matriks yang sama.

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

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

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

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

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

dan

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

dengan

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

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

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

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \quad (2.1.22)$$

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 2.1.21 (Perkalian Hadamard, Horn dan Johnson 2012)** Jika diberikan

$\mathbf{A} = (a_{ij}) \in \text{Mat}_{m \times n}(\mathcal{F})$  dan  $\mathbf{B} = (b_{ij}) \in \text{Mat}_{m \times n}(\mathcal{F})$ , 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.

$$\mathbf{A} \odot \mathbf{B} = (a_{ij}b_{ij}) \in \text{Mat}_{m \times n}(\mathcal{F}). \quad (2.1.23)$$

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

**Contoh 2.1.22** 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 2.1.23 (Transpose)** Jika  $\mathbf{A} \in \text{Mat}_{m \times n}(\mathcal{F})$ , maka transpose dari  $\mathbf{A}$ , dilambangkan sebagai  $\mathbf{A}^\top$ , dengan  $\mathbf{A}^\top \in \text{Mat}_{n \times m}(\mathcal{F})$  didefinisikan oleh  $(\mathbf{A}^\top)_{ji} = a_{ij}$ .

**Contoh 2.1.24** Salah satu contoh *transpose* dari matriks adalah

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

sedangkan untuk vektor adalah

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

Notasi untuk *transpose* pada vektor  $\mathbf{a}$  dapat juga dinotasikan dengan  $\mathbf{a}'$ . Di sisi lain,

untuk skalar di lapangan  $\mathcal{F}$  atau  $c \in \mathcal{F}$ ,  $c^\top = c$ .

**Teorema 2.1.25** Untuk  $\mathbf{A}, \mathbf{B}$  berukuran sesuai dan  $c \in \mathcal{F}$ , 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

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

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

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

■

**Definisi 2.1.26 (Matriks simetris, diagonal, dan identitas)** Berikut ini adalah beberapa definisi terkait matriks simetris, diagonal, dan identitas.



- (a)  $\mathbf{A} \in \text{Mat}_{n \times n}(\mathcal{F})$  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 2.1.27** 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 2.1.28 (Matriks ortogonal, Dhrymes 2013)** Misalkan  $\mathbf{a}$  dan  $\mathbf{b}$  adalah vektor-vektor di  $\mathbb{R}^n$ . Vektor-vektor tersebut dikatakan ortogonal jika dan hanya jika

$$\mathbf{a}^\top \mathbf{b} = 0. \quad (2.1.28)$$

Sebuah matriks persegi  $\mathbf{A}$  dikatakan ortogonal jika dan hanya jika

$$\mathbf{A}^\top \mathbf{A} = \mathbf{A} \mathbf{A}^\top = \mathbf{I}_n. \quad (2.1.29)$$

**Contoh 2.1.29** Misalkan

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

Dapat diverifikasi bahwa  $\mathbf{A}$  adalah matriks ortogonal karena

$$\mathbf{A}^\top \mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}_2.$$

**Definisi 2.1.30 (Matriks idempoten, Dhrymes 2013)** Suatu matriks persegi  $\mathbf{A}$  dikatakan idempoten jika dan hanya jika

$$\mathbf{A} \mathbf{A} = \mathbf{A}^2 = \mathbf{A}. \quad (2.1.30)$$

**Contoh 2.1.31** 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 2.1.32 (Jejak atau *trace*, Dhrymes 2013)** Untuk suatu matriks persegi  $\mathbf{A}$  berukuran  $n \times n$  atau  $\mathbf{A} = (a_{ij}) \in \text{Mat}_{n \times n}(\mathcal{F})$ , jejak atau *trace* didefinisikan sebagai

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}. \quad (2.1.31)$$

**Contoh 2.1.33** 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

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

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

**Teorema 2.1.34 (Sifat-Sifat *Trace*, Axler 2024)** Untuk ukuran yang sesuai, berla-

ku:

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

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

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

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

(a)

$$\begin{aligned} \operatorname{tr}(\alpha \mathbf{A} + \beta \mathbf{B}) &= \sum_i (\alpha a_{ii} + \beta b_{ii}) \\ &= \alpha \sum_i a_{ii} + \beta \sum_i b_{ii} \\ &= \alpha \operatorname{tr}(\mathbf{A}) + \beta \operatorname{tr}(\mathbf{B}). \end{aligned} \tag{2.1.32}$$

(b) Untuk dua faktor,

$$\begin{aligned} \operatorname{tr}(\mathbf{AB}) &= \sum_i (\mathbf{AB})_{ii} \\ &= \sum_i \sum_j a_{ij} b_{ji} \\ &= \sum_j \sum_i b_{ji} a_{ij} \\ &= \sum_j (\mathbf{BA})_{jj} \\ &= \operatorname{tr}(\mathbf{BA}). \end{aligned} \tag{2.1.33}$$

Untuk tiga faktor,

$$\begin{aligned}
 \text{tr}(\mathbf{ABC}) &= \sum_i (\mathbf{ABC})_{ii} \\
 &= \sum_i \sum_j \sum_k a_{ij} b_{jk} c_{ki} \\
 &= \sum_j \sum_k \sum_i b_{jk} c_{ki} a_{ij} \\
 &= \text{tr}(\mathbf{BCA}).
 \end{aligned} \tag{2.1.34}$$

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

$$\begin{aligned}
 \text{tr}(\mathbf{B}^{-1}\mathbf{AB}) &= \text{tr}(\mathbf{ABB}^{-1}) \\
 &= \text{tr}(\mathbf{AI}) \\
 &= \text{tr}(\mathbf{A}).
 \end{aligned} \tag{2.1.35}$$

■

**Definisi 2.1.35 (Matriks semi-definit positif, Horn dan Johnson 2012)** Suatu matriks simetris  $\mathbf{A} \in \text{Mat}_{n \times n}(\mathbb{R})$  dikatakan semi-definit positif jika untuk setiap vektor tak nol  $\mathbf{x} \in \mathbb{R}^n$  berlaku

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} \geq 0. \tag{2.1.36}$$

Matriks  $\mathbf{A}$  dikatakan definit positif jika dan hanya jika untuk setiap vektor tak nol  $\mathbf{x} \in \mathbb{R}^n$  berlaku

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0. \tag{2.1.37}$$

**Contoh 2.1.36** Pertimbangkan matriks

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

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

$$\begin{aligned}\mathbf{x}^\top \mathbf{A} \mathbf{x} &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= 2x_1^2 + 2x_1x_2 + 2x_2^2 \\ &= x_1^2 + (x_1 + x_2)^2 + x_2^2 \geq 0.\end{aligned}\tag{2.1.38}$$

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  $\mathbf{x} \neq \mathbf{0}$ .

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 Axler, konsep invers dipahami melalui operator linear dan representasinya dalam suatu basis. Sebuah matriks bujur sangkar merepresentasikan operator linear  $T : \mathcal{F}^n \rightarrow \mathcal{F}^n$ , dan invers matriks berkorespondensi dengan invers operator linear  $T^{-1}$  ketika  $T$  bijektif.

**Definisi 2.1.37 (Invers matriks, Axler 2024)** Misalkan  $\mathbf{A} \in \text{Mat}_{n \times n}(\mathcal{F})$ . Matriks  $\mathbf{A}$  disebut invertibel (atau non-singular) apabila terdapat matriks  $\mathbf{B} \in \text{Mat}_{n \times n}(\mathcal{F})$  sedemikian sehingga

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n,\tag{2.1.39}$$

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

**Contoh 2.1.38** Pertimbangkan matriks

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

Untuk menunjukkan bahwa  $\mathbf{A}$  invertibel, perlu dicari matriks  $\mathbf{B}$  sehingga  $\mathbf{AB} = \mathbf{BA} = \mathbf{I}_2$ . Dengan menggunakan rumus invers matriks  $2 \times 2$ , 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 2.1.39** *Jika  $\mathbf{A}$  invertibel, maka  $\mathbf{A}^{-1}$  tunggal (unik).*

**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

$$\mathbf{B} = \mathbf{BI}_n = \mathbf{B}(\mathbf{AC}) = (\mathbf{BA})\mathbf{C} = \mathbf{I}_n\mathbf{C} = \mathbf{C}. \quad (2.1.40)$$

Oleh karena itu, invers matriks adalah unik. ■

**Proposisi 2.1.40** *Misalkan  $\mathbf{A}, \mathbf{B} \in \text{Mat}_{n \times n}(\mathcal{F})$  invertibel dan  $\alpha \in \mathcal{F}$ . Berlaku:*

- (a)  $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$ ;
- (b)  $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$ ;
- (c)  $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ ; dan
- (d) jika  $\mathbf{A} = \text{diag}(\mathbf{B}_1, \dots, \mathbf{B}_k)$  dengan setiap  $\mathbf{B}_i$  invertibel (ukuran mungkin berbeda), maka

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

**Bukti.** Berikut adalah bukti untuk keempat sifat invers matriks.

- (a) Berdasarkan definisi invers,  $\mathbf{A}^{-1}$  adalah matriks yang memenuhi  $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$ . Oleh karena itu,  $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$ .

(b) Akan ditunjukkan bahwa  $(\mathbf{A}^{-1})^\top$  adalah invers dari  $\mathbf{A}^\top$ . Perhatikan bahwa

$$\begin{aligned}\mathbf{A}^\top (\mathbf{A}^{-1})^\top &= (\mathbf{A}^{-1} \mathbf{A})^\top \quad (\text{sifat transpose perkalian}) \\ &= \mathbf{I}_n^\top \\ &= \mathbf{I}_n.\end{aligned}\tag{2.1.41}$$

Demikian pula,

$$\begin{aligned}(\mathbf{A}^{-1})^\top \mathbf{A}^\top &= (\mathbf{A} \mathbf{A}^{-1})^\top \\ &= \mathbf{I}_n^\top \\ &= \mathbf{I}_n.\end{aligned}\tag{2.1.42}$$

Sebab  $(\mathbf{A}^{-1})^\top$  memenuhi  $\mathbf{A}^\top (\mathbf{A}^{-1})^\top = (\mathbf{A}^{-1})^\top \mathbf{A}^\top = \mathbf{I}_n$ , berdasarkan definisi invers,  $(\mathbf{A}^{-1})^\top = (\mathbf{A}^\top)^{-1}$ .

(c) Akan ditunjukkan bahwa  $\mathbf{B}^{-1} \mathbf{A}^{-1}$  adalah invers dari  $\mathbf{A} \mathbf{B}$ . Perhatikan bahwa

$$\begin{aligned}(\mathbf{A} \mathbf{B})(\mathbf{B}^{-1} \mathbf{A}^{-1}) &= \mathbf{A}(\mathbf{B} \mathbf{B}^{-1}) \mathbf{A}^{-1} \quad (\text{sifat asosiatif}) \\ &= \mathbf{A} \mathbf{I}_n \mathbf{A}^{-1} \\ &= \mathbf{A} \mathbf{A}^{-1} \\ &= \mathbf{I}_n.\end{aligned}\tag{2.1.43}$$

Demikian pula,

$$\begin{aligned}(\mathbf{B}^{-1} \mathbf{A}^{-1})(\mathbf{A} \mathbf{B}) &= \mathbf{B}^{-1}(\mathbf{A}^{-1} \mathbf{A}) \mathbf{B} \\ &= \mathbf{B}^{-1} \mathbf{I}_n \mathbf{B} \\ &= \mathbf{B}^{-1} \mathbf{B} \\ &= \mathbf{I}_n.\end{aligned}\tag{2.1.44}$$

Sebab  $\mathbf{B}^{-1} \mathbf{A}^{-1}$  memenuhi  $(\mathbf{A} \mathbf{B})(\mathbf{B}^{-1} \mathbf{A}^{-1}) = (\mathbf{B}^{-1} \mathbf{A}^{-1})(\mathbf{A} \mathbf{B}) = \mathbf{I}_n$ , berdasarkan definisi invers,  $\mathbf{B}^{-1} \mathbf{A}^{-1} = (\mathbf{A} \mathbf{B})^{-1}$ .

(d) Misalkan  $\mathbf{A} = \text{diag}(\mathbf{B}_1, \dots, \mathbf{B}_k)$  adalah matriks blok diagonal dengan setiap  $\mathbf{B}_i$  berukuran  $n_i \times n_i$  dan invertibel. Akan ditunjukkan bahwa  $\mathbf{A}^{-1} = \text{diag}(\mathbf{B}_1^{-1}, \dots, \mathbf{B}_k^{-1})$ .

Perhatikan struktur perkalian matriks blok diagonal:

$$\begin{aligned} \mathbf{A} \cdot \text{diag}(\mathbf{B}_1^{-1}, \dots, \mathbf{B}_k^{-1}) &= \text{diag}(\mathbf{B}_1, \dots, \mathbf{B}_k) \cdot \text{diag}(\mathbf{B}_1^{-1}, \dots, \mathbf{B}_k^{-1}) \\ &= \text{diag}(\mathbf{B}_1 \mathbf{B}_1^{-1}, \dots, \mathbf{B}_k \mathbf{B}_k^{-1}) \\ &= \text{diag}(\mathbf{I}_{n_1}, \dots, \mathbf{I}_{n_k}) \\ &= \mathbf{I}_n, \end{aligned} \tag{2.1.45}$$

dengan  $n = n_1 + n_2 + \dots + n_k$ . Kesamaan kedua berlaku karena perkalian matriks blok diagonal bersifat *block-wise*, yaitu blok ke- $i$  dari hasil kali adalah  $\mathbf{B}_i \mathbf{B}_i^{-1}$ . Analog,  $\text{diag}(\mathbf{B}_1^{-1}, \dots, \mathbf{B}_k^{-1}) \cdot \mathbf{A} = \mathbf{I}_n$ . Dengan demikian,  $\mathbf{A}^{-1} = \text{diag}(\mathbf{B}_1^{-1}, \dots, \mathbf{B}_k^{-1})$ . ■

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  $\mathbf{E}_k \cdots \mathbf{E}_1$ , yaitu matriks yang diperoleh dari matriks identitas dengan melakukan satu operasi baris elementer) sehingga  $\mathbf{E}_k \cdots \mathbf{E}_1 \mathbf{A} = \mathbf{I}$ . Dengan demikian,

$$\mathbf{A}^{-1} = \mathbf{E}_k \cdots \mathbf{E}_1. \tag{2.1.46}$$

Secara praktis, dapat dilakukan reduksi baris teraugmentasi, yaitu reduksi baris pada matriks yang diperoleh dengan menggabungkan  $\mathbf{A}$  dan  $\mathbf{I}$  secara horizontal, sehingga

$$[\mathbf{A} \mid \mathbf{I}] \xrightarrow{\text{operasi baris}} [\mathbf{I} \mid \mathbf{A}^{-1}].$$

**Contoh 2.1.41** Ambil

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



Dengan melakukan eliminasi Gauss–Jordan pada  $[A \mid I]$  diperoleh

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

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

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

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

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

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

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

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

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 2.1.42 (Kofaktor dan matriks adjoin, Dhrymes, 2013)** Misalkan  $\mathbf{A} = (a_{ij}) \in \text{Mat}_{n \times n}(\mathcal{F})$ . 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

$$C_{ij} = (-1)^{i+j} M_{ij}. \quad (2.1.47)$$

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

$$\text{adj}(\mathbf{A}) = \mathbf{C}^\top. \quad (2.1.48)$$

**Teorema 2.1.43** Jika  $\mathbf{A} \in \mathcal{F}^{n \times n}$  dengan  $\det(\mathbf{A}) \neq 0$ , maka

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \text{adj}(\mathbf{A}). \quad (2.1.49)$$

**Bukti.** Akan ditunjukkan bahwa  $\mathbf{A} \cdot \text{adj}(\mathbf{A}) = \det(\mathbf{A}) \cdot \mathbf{I}_n$ . Elemen  $(i, j)$  dari

matriks  $\mathbf{A} \cdot \text{adj}(\mathbf{A})$  adalah

$$(\mathbf{A} \cdot \text{adj}(\mathbf{A}))_{ij} = \sum_{k=1}^n a_{ik} \cdot (\text{adj}(\mathbf{A}))_{kj} = \sum_{k=1}^n a_{ik} C_{jk}, \quad (2.1.50)$$

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

Teorema ekspansi Laplace menyatakan bahwa determinan matriks dapat dihitung dengan ekspansi kofaktor sepanjang baris ke- $i$ :

$$\det(\mathbf{A}) = \sum_{k=1}^n a_{ik} C_{ik}. \quad (2.1.51)$$

Oleh karena itu,  $(\mathbf{A} \cdot \text{adj}(\mathbf{A}))_{ii} = \det(\mathbf{A})$ .

Sifat ekspansi kofaktor asing menyatakan bahwa jumlah  $\sum_{k=1}^n a_{ik} C_{jk}$  dengan  $i \neq j$  sama dengan determinan matriks yang memiliki dua baris identik (baris ke- $i$  dan baris ke- $j$ ), sehingga bernilai nol:

$$\sum_{k=1}^n a_{ik} C_{jk} = 0 \quad \text{untuk } i \neq j. \quad (2.1.52)$$

Dengan menggabungkan kedua kasus, diperoleh

$$(\mathbf{A} \cdot \text{adj}(\mathbf{A}))_{ij} = \delta_{ij} \det(\mathbf{A}), \quad (2.1.53)$$

dengan  $\delta_{ij}$  adalah delta Kronecker, yaitu fungsi yang bernilai 1 jika  $i = j$  dan 0 jika  $i \neq j$ . Ini berarti  $\mathbf{A} \cdot \text{adj}(\mathbf{A}) = \det(\mathbf{A}) \cdot \mathbf{I}_n$ . Sebab  $\det(\mathbf{A}) \neq 0$ , dapat dibagi kedua ruas dengan  $\det(\mathbf{A})$  sehingga

$$\mathbf{A} \cdot \frac{1}{\det(\mathbf{A})} \text{adj}(\mathbf{A}) = \mathbf{I}_n. \quad (2.1.54)$$

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

**Contoh 2.1.44** 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

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

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

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

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

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

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

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

Dengan demikian, hasil perhitungan invers terverifikasi benar.

**Contoh 2.1.45** Untuk  $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$  dengan  $ad - bc \neq 0$ , matriks adjoin adalah

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

sehingga

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \quad (2.1.55)$$

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,  $\mathcal{V}$  menyatakan ruang vektor berdimensi hingga atas suatu lapangan  $\mathcal{F}$  dan  $\mathcal{L}(\mathcal{V}, \mathcal{W})$  himpunan semua pemetaan linear  $\mathcal{V} \rightarrow \mathcal{W}$ .

**Definisi 2.1.46 (Pemetaan linear, Axler 2024)** Misalkan  $\mathcal{V}, \mathcal{W}$  ruang vektor atas suatu lapangan  $\mathcal{F}$ . Suatu pemetaan  $T : \mathcal{V} \rightarrow \mathcal{W}$  disebut pemetaan linear jika untuk semua  $\mathbf{u}, \mathbf{v} \in \mathcal{V}$  dan  $\lambda \in \mathcal{F}$  berlaku

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}), \quad T(\lambda \mathbf{v}) = \lambda T(\mathbf{v}). \quad (2.1.56)$$

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

**Contoh 2.1.47** Berikut ini adalah beberapa contoh pemetaan linear.

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

$$T(\mathbf{x}, \mathbf{y}) = (\mathbf{x} + 2\mathbf{y}, 3\mathbf{x} - \mathbf{y}) \quad (2.1.57)$$

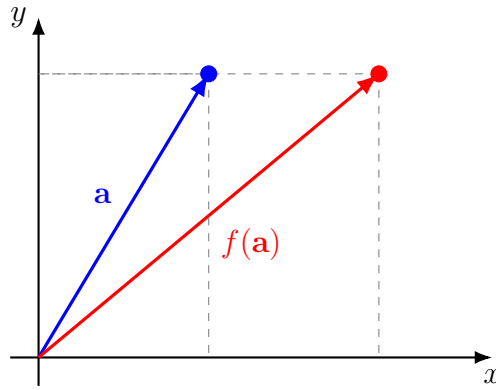
adalah elemen  $\mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$ .

(b) Jika  $\mathcal{V} = \mathbb{R}^3$  dan  $\mathcal{W} = \mathbb{R}^2$ , maka setiap  $T \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$  dapat direpresenta-

sikan oleh matriks  $2 \times 3$ , misalnya

$$T(x_1, x_2, x_3) = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \quad (2.1.58)$$

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  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  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  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  dengan  $f(x, y) = (2x, y)$ .

**Definisi 2.1.48** (*Null space, range, dan rank, Axler 2024*) Untuk  $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ , null space dan range didefinisikan sebagai

$$\text{null } T := \{\mathbf{v} \in \mathcal{V} : T\mathbf{v} = \mathbf{0}\}, \quad \text{range } T := \{T\mathbf{v} : \mathbf{v} \in \mathcal{V}\} \subseteq \mathcal{W}.$$

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

**Contoh 2.1.49** Misalkan

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

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

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

sehingga

$$\text{null } T = \text{span}\{(5, -4, 1)^\top\} \quad \text{dan} \quad \dim(\text{null } T) = 1. \quad (2.1.59)$$

(b) Kolom pertama dan kedua bebas linear, maka

$$\text{range } T = \text{span}\{(1, 0)^\top, (2, 1)^\top\} \quad (2.1.60)$$

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

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

**Teorema 2.1.50 (Teorema Rank–Nulitas, Axler 2024)** *Jika  $\mathcal{V}$  berdimensi-hingga dan  $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ , maka*

$$\dim \mathcal{V} = \dim(\text{null } T) + \dim(\text{range } T). \quad (2.1.61)$$

**Bukti.** Misalkan  $\dim(\text{null } T) = k$  dan pilih basis  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$  untuk  $\text{null } T$ . Sebab  $\text{null } T$  adalah subruang dari  $\mathcal{V}$ , basis ini dapat diperluas menjadi basis  $\mathcal{V}$ , misalkan  $\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{v}_1, \dots, \mathbf{v}_m\}$  dengan  $k + m = \dim \mathcal{V}$ .

Akan ditunjukkan bahwa  $\{T\mathbf{v}_1, T\mathbf{v}_2, \dots, T\mathbf{v}_m\}$  adalah basis untuk  $\text{range } T$ . Ambil sembarang  $\mathbf{w} \in \text{range } T$ . Maka terdapat  $\mathbf{v} \in \mathcal{V}$  sehingga  $T\mathbf{v} = \mathbf{w}$ . Sebab



$\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{v}_1, \dots, \mathbf{v}_m\}$  adalah basis  $\mathcal{V}$ , maka

$$\mathbf{v} = a_1\mathbf{u}_1 + \dots + a_k\mathbf{u}_k + b_1\mathbf{v}_1 + \dots + b_m\mathbf{v}_m \quad (2.1.62)$$

untuk suatu skalar  $a_1, \dots, a_k, b_1, \dots, b_m$ . Dengan menerapkan  $T$  pada kedua ruas, diperoleh

$$\begin{aligned} \mathbf{w} = T\mathbf{v} &= a_1T\mathbf{u}_1 + \dots + a_kT\mathbf{u}_k + b_1T\mathbf{v}_1 + \dots + b_mT\mathbf{v}_m \\ &= a_1 \cdot \mathbf{0} + \dots + a_k \cdot \mathbf{0} + b_1T\mathbf{v}_1 + \dots + b_mT\mathbf{v}_m \\ &= b_1T\mathbf{v}_1 + \dots + b_mT\mathbf{v}_m, \end{aligned} \quad (2.1.63)$$

sebab  $\mathbf{u}_1, \dots, \mathbf{u}_k \in \text{null } T$  sehingga  $T\mathbf{u}_i = \mathbf{0}$ . Jadi,  $\mathbf{w}$  dapat ditulis sebagai kombinasi linear dari  $\{T\mathbf{v}_1, \dots, T\mathbf{v}_m\}$ .

Andaikan terdapat skalar  $c_1, \dots, c_m$  sehingga

$$c_1T\mathbf{v}_1 + c_2T\mathbf{v}_2 + \dots + c_mT\mathbf{v}_m = \mathbf{0}. \quad (2.1.64)$$

Sebab  $T$  linear, hal ini ekuivalen dengan

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_m\mathbf{v}_m) = \mathbf{0}, \quad (2.1.65)$$

yang berarti  $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_m\mathbf{v}_m \in \text{null } T$ . Sebab  $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$  adalah basis untuk  $\text{null } T$ , terdapat skalar  $d_1, \dots, d_k$  sehingga

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_m\mathbf{v}_m = d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + \dots + d_k\mathbf{u}_k. \quad (2.1.66)$$

Dengan menyusun ulang, diperoleh

$$(-d_1)\mathbf{u}_1 + \dots + (-d_k)\mathbf{u}_k + c_1\mathbf{v}_1 + \dots + c_m\mathbf{v}_m = \mathbf{0}. \quad (2.1.67)$$

Sebab  $\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{v}_1, \dots, \mathbf{v}_m\}$  adalah basis  $\mathcal{V}$  (bebas linear), semua koefisien ha-

rus nol:

$$d_1 = \cdots = d_k = 0 \quad \text{dan} \quad c_1 = \cdots = c_m = 0. \quad (2.1.68)$$

Dengan demikian,  $\{T\mathbf{v}_1, \dots, T\mathbf{v}_m\}$  bebas linear.

Sebab  $\{T\mathbf{v}_1, \dots, T\mathbf{v}_m\}$  menjangkau  $\text{range } T$  dan bebas linear, maka himpunan ini adalah basis untuk  $\text{range } T$ . Oleh karena itu,

$$\dim(\text{range } T) = m. \quad (2.1.69)$$

Dengan demikian,

$$\dim V = k + m = \dim(\text{null } T) + \dim(\text{range } T). \quad (2.1.70)$$

■

**Contoh 2.1.51** Berdasarkan Contoh 2.1.49, diketahui bahwa  $\dim(\text{null } T) = 1$  dan  $\dim(\text{range } T) = 2$ . Konsisten dengan Rank–Nulitas,  $1 + 2 = 3 = \dim \mathbb{R}^3$ .

**Definisi 2.1.52 (Bentuk multilinear anti-simetri, Axler 2024)** Misalkan  $\mathcal{V}$  ruang vektor atas lapangan  $\mathcal{F}$ . Suatu fungsi  $\omega : \mathcal{V}^n \rightarrow \mathcal{F}$  disebut bentuk  $n$ -linear jika linear pada setiap argumennya. Bentuk  $n$ -linear  $\omega$  disebut anti-simetri atau alternating jika

$$\omega(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n) = -\omega(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_i, \dots, \mathbf{v}_n) \quad (2.1.71)$$

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

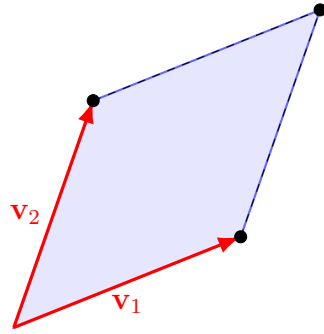
**Contoh 2.1.53** Pada  $\mathcal{V} = \mathbb{R}^2$ , fungsi

$$\omega((x_1, y_1), (x_2, y_2)) = x_1 y_2 - x_2 y_1 \quad (2.1.72)$$

adalah bentuk bilinear anti-simetri, karena

$$\omega((x_2, y_2), (x_1, y_1)) = x_2y_1 - x_1y_2 = -\omega((x_1, y_1), (x_2, y_2)). \quad (2.1.73)$$

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



**Gambar 2.2** Ilustrasi bentuk bilinear anti-simetri  $\omega(v_1, v_2) = x_1y_2 - x_2y_1$  pada  $\mathbb{R}^2$ .

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

$$\begin{aligned} |\mathbf{v}_1||\mathbf{v}_2| \sin(\theta) &= \sqrt{|\mathbf{v}_1|^2|\mathbf{v}_2|^2 - (x_1x_2 + y_1y_2)^2} \\ &= \sqrt{x_1^2x_2^2 + x_1^2y_2^2 + y_1^2x_2^2 + y_1^2y_2^2 - (x_1^2x_2^2 + 2x_1x_2y_1y_2 + y_1^2y_2^2)} \\ &= \sqrt{x_1^2y_2^2 + y_1^2x_2^2 - 2x_1x_2y_1y_2} \\ &= \sqrt{(x_1y_2 - x_2y_1)^2} \\ &= |x_1y_2 - x_2y_1|. \end{aligned} \quad (2.1.74)$$

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

Contoh 2.1.53 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 2.1.54 (Determinan, Axler 2024)** Misalkan  $\mathcal{V}$  berdimensi  $n \geq 1$ . Ruang semua bentuk  $n$ -linear beranti-simetri pada  $\mathcal{V}$  memiliki dimensi 1. Untuk setiap  $T \in \mathcal{L}(\mathcal{V})$ , determinan  $T$ , ditulis  $\det T$ , adalah satu-satunya skalar di  $\mathcal{F}$  dengan sifat

$$\omega(T\mathbf{v}_1, \dots, T\mathbf{v}_n) = (\det T) \omega(\mathbf{v}_1, \dots, \mathbf{v}_n), \quad (2.1.75)$$

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

**Contoh 2.1.55** Ambil  $\mathcal{V} = \mathbb{R}^2$  dan operator linear  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  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

$$\omega(\mathbf{v}_1, \mathbf{v}_2) = \det \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = 3. \quad (2.1.76)$$

Sekarang apabila dilihat bayangan kedua vektor tersebut,

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

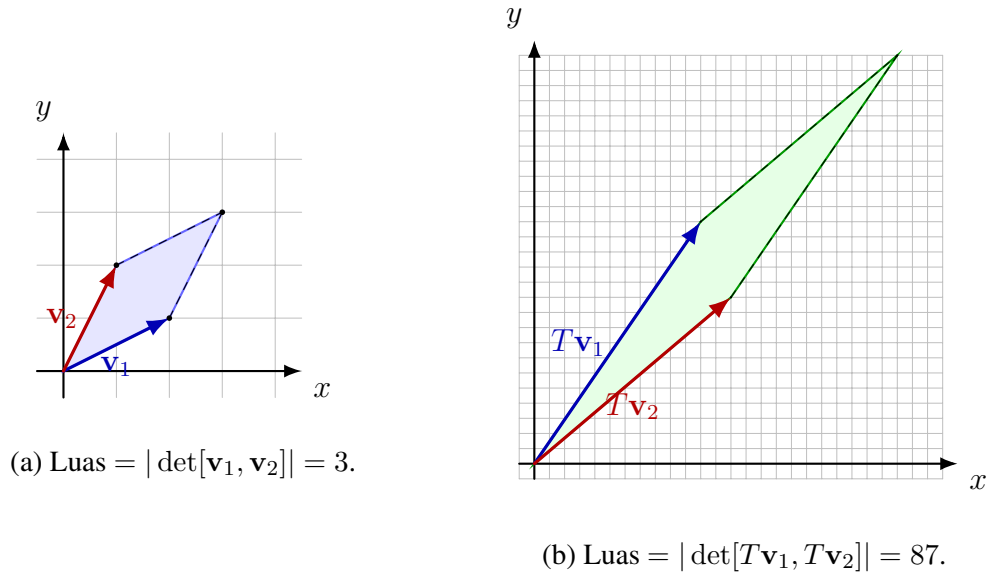
sehingga

$$\omega(T\mathbf{v}_1, T\mathbf{v}_2) = \det \begin{pmatrix} 11 & 13 \\ 16 & 11 \end{pmatrix} = 11 \cdot 11 - 13 \cdot 16 = -87. \quad (2.1.77)$$

Dengan demikian, sesuai definisi determinan operator,

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

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



**Gambar 2.3** Perbandingan luas parallelogram sebelum dan sesudah transformasi  $A$ .

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

**Teorema 2.1.56 (Axler, 2024)** Untuk  $S, T \in \mathcal{L}(\mathcal{V})$  dan skalar  $c \in \mathcal{F}$  berlaku:

- (a)  $\det \mathbf{I}_{\mathcal{V}} = 1$ ;
- (b)  $\det(ST) = \det S \cdot \det T$ ; dan
- (c)  $T$  invertibel jika dan hanya jika  $\det T \neq 0$ .

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

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

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

$$\omega(\mathbf{I}_{\mathcal{V}}\mathbf{v}_1, \dots, \mathbf{I}_{\mathcal{V}}\mathbf{v}_n) = \omega(\mathbf{v}_1, \dots, \mathbf{v}_n). \quad (2.1.78)$$

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

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

$$\begin{aligned} \omega((ST)\mathbf{v}_1, \dots, (ST)\mathbf{v}_n) &= \omega(S(T\mathbf{v}_1), \dots, S(T\mathbf{v}_n)) \\ &= (\det S) \omega(T\mathbf{v}_1, \dots, T\mathbf{v}_n) \\ &= (\det S)(\det T) \omega(\mathbf{v}_1, \dots, \mathbf{v}_n). \end{aligned} \quad (2.1.79)$$

Di sisi lain, berdasarkan definisi determinan untuk  $ST$ :

$$\omega((ST)\mathbf{v}_1, \dots, (ST)\mathbf{v}_n) = (\det(ST)) \omega(\mathbf{v}_1, \dots, \mathbf{v}_n). \quad (2.1.80)$$

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

- (c)  $(\Rightarrow)$  Andaikan  $T$  invertibel, maka terdapat  $T^{-1}$  sehingga  $TT^{-1} = \mathbf{I}_{\mathcal{V}}$ . Dengan sifat (a) dan (b):

$$1 = \det \mathbf{I}_{\mathcal{V}} = \det(TT^{-1}) = \det T \cdot \det T^{-1}. \quad (2.1.81)$$

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  $\text{null } T \neq \{\mathbf{0}\}$  atau  $\text{range } T \neq \mathcal{V}$ . 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  $\omega$ :

$$\omega(T\mathbf{v}_1, \dots, T\mathbf{v}_n) = 0. \quad (2.1.82)$$

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:

$$0 = \omega(T\mathbf{v}_1, \dots, T\mathbf{v}_n) = (\det T) \omega(\mathbf{v}_1, \dots, \mathbf{v}_n), \quad (2.1.83)$$

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 nilai eigen dan vektor eigen yang memberikan wawasan tentang bagaimana suatu operator linear bertindak pada vektor tertentu tanpa mengubah arahnya.

**Definisi 2.1.57 (Vektor eigen dan nilai eigen, (Axler 2024, Lang, 1987))** Misalkan  $T : \mathcal{V} \rightarrow \mathcal{V}$  adalah operator linear pada ruang vektor  $\mathcal{V}$  atas  $\mathbb{C}$ .

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

$$T(\mathbf{v}) = \lambda\mathbf{v}. \quad (2.1.84)$$

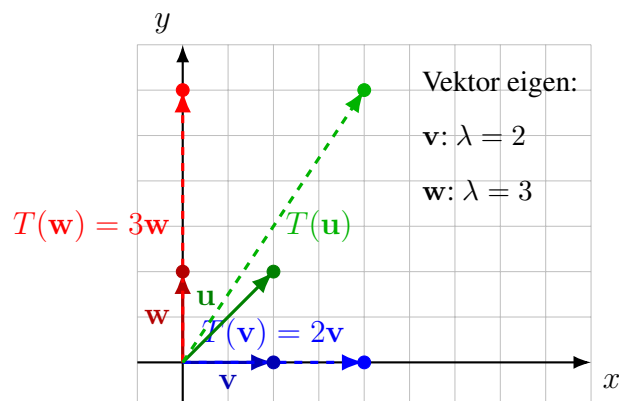
(b) Skalar  $\lambda$  yang memenuhi persamaan di atas disebut nilai eigen dari  $T$ .

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

**Contoh 2.1.58** Pertimbangkan  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  dengan

$$T(x, y) = (2x, 3y). \quad (2.1.85)$$

- (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  $T(x, y) = (2x, 3y)$ .

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  $\mathbf{u}$ .

**Contoh 2.1.59** Berikut akan dihitung nilai eigen dan vektor eigen dari matriks

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



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

$$\begin{aligned}\det \begin{pmatrix} 4 - \lambda & 1 \\ 2 & 3 - \lambda \end{pmatrix} &= 0 \\ (4 - \lambda)(3 - \lambda) - (1)(2) &= 0 \\ 12 - 4\lambda - 3\lambda + \lambda^2 - 2 &= 0 \\ \lambda^2 - 7\lambda + 10 &= 0 \\ (\lambda - 5)(\lambda - 2) &= 0.\end{aligned}\tag{2.1.86}$$

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 dan Operasi Vektorisasi

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 2.1.60 (Perkalian Kronecker, Dhrymes 2013)** Untuk  $\mathbf{A} \in \text{Mat}_{m \times n}(\mathcal{F})$  dan  $\mathbf{B} \in \text{Mat}_{p \times q}(\mathcal{F})$ , perkalian Kronecker dari matriks tersebut dituliskan sebagai  $\mathbf{A} \otimes \mathbf{B}$ , didefinisikan sebagai

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix} \in \text{Mat}_{mp \times nq}(\mathcal{F}). \quad (2.1.87)$$

**Contoh 2.1.61** Apabila diambil

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

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

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} 1 \cdot \mathbf{B} & 2 \cdot \mathbf{B} \\ 3 \cdot \mathbf{B} & 4 \cdot \mathbf{B} \end{pmatrix} = \begin{pmatrix} 0 & 5 & 0 & 10 \\ 6 & 7 & 12 & 14 \\ 0 & 15 & 0 & 20 \\ 18 & 21 & 24 & 28 \end{pmatrix}. \quad (2.1.88)$$

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

Operasi vektorisasi (vec) adalah teknik yang digunakan untuk mengubah matriks menjadi vektor dengan cara menumpuk kolom-kolom matriks secara vertikal. Operasi ini sangat berguna dalam berbagai aplikasi matematika dan statistik, terutama dalam konteks aljabar linear dan analisis data multivariat. Vektorisasi memungkinkan kita untuk merepresentasikan data matriks dalam bentuk vektor, yang seringkali lebih mudah untuk dianalisis dan diolah menggunakan teknik-teknik statistik dan komputasi.

**Definisi 2.1.62 (Vektorisasi, Dhrymes 2013)** Untuk matriks berukuran  $m \times n$  yaitu  $\mathbf{A} = (a_{ij}) \in \text{Mat}_{m \times n}(\mathcal{F})$ , operasi vektorisasi, ditulis sebagai  $\text{vec}(\mathbf{A})$ , didefinisikan sebagai vektor kolom berukuran  $mn \times 1$  yang diperoleh dengan menumpuk kolom-kolom matriks  $\mathbf{A}$  secara vertikal dengan  $n$  elemen pertama berasal dari kolom pertama matriks  $A$ , yaitu  $a_{.1}$ , diikuti oleh  $n$  elemen berikutnya dari kolom kedua matriks  $A$ , yaitu  $a_{.2}$ , dan seterusnya, sehingga

$$\text{vec}(\mathbf{A}) = (a_{.1}^\top, a_{.2}^\top, \dots, a_{.n}^\top)^\top, \quad (2.1.89)$$

Sebaliknya, vektorisasi baris, ditulis sebagai  $\text{rvec}(\mathbf{A})$ , didefinisikan sebagai vektor kolom berukuran  $mn \times 1$  yang diperoleh dengan menumpuk baris-baris matriks  $\mathbf{A}$  secara vertikal dengan  $m$  elemen pertama berasal dari baris pertama matriks  $A$ , yaitu  $a_{1.}$ , diikuti oleh  $m$  elemen berikutnya dari baris kedua matriks  $A$ , yaitu  $a_{2.}$ , dan seterusnya, sehingga

$$\text{rvec}(\mathbf{A}) = (a_{1.}, a_{2.}, \dots, a_{m.}). \quad (2.1.90)$$

**Contoh 2.1.63** Misalkan

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

Vektorisasi dari matriks  $\mathbf{A}$  adalah

$$\text{vec}(\mathbf{A}) = \begin{pmatrix} 1 \\ 4 \\ 2 \\ 5 \\ 3 \\ 6 \end{pmatrix} \in \mathbb{R}^{6 \times 1}. \quad (2.1.91)$$

Sedangkan vektorisasi barisnya adalah

$$\text{rvec}(\mathbf{A}) = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{pmatrix} \in \mathbb{R}^{6 \times 1}. \quad (2.1.92)$$

### 2.1.7 Norma Vektor dan Matriks

Jarak Euclidean, yaitu jarak yang didefinisikan dengan  $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2$ , adalah salah satu konsep dasar dalam aljabar linear dan analisis vektor. Jarak ini mengukur seberapa jauh dua vektor berada satu sama lain dalam ruang Euclidean, yang merupakan ruang vektor berdimensi  $n$  dengan norma Euclidean. Norma dapat didefinisikan sebagai generalisasi dari konsep panjang atau ukuran vektor dalam ruang vektor.

**Definisi 2.1.64 (Norma, Horn dan Johnson 2012)** Misalkan  $\mathcal{V}$  adalah ruang vektor atas lapangan  $\mathcal{F}$  (baik  $\mathbb{R}$  maupun  $\mathbb{C}$ ). Fungsi  $\|\cdot\| : \mathcal{V} \rightarrow \mathbb{R}$  disebut norma pada  $\mathcal{V}$  jika memenuhi sifat-sifat berikut untuk setiap  $\mathbf{u}, \mathbf{v} \in \mathcal{V}$  dan skalar  $c \in \mathcal{F}$ :

- (a)  $\|\mathbf{v}\| \geq 0$ ;
- (b)  $\|\mathbf{v}\| = 0$  jika dan hanya jika  $\mathbf{v} = \mathbf{0}$ ;

(c)  $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$ ; dan

(d)  $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$  (ketaksamaan segitiga).

Konsep norma dapat diperluas ke matriks, yang memungkinkan kita untuk mengukur "ukuran" atau "panjang" dari suatu matriks. Norma matriks sering digunakan dalam analisis numerik, optimisasi, dan teori kontrol untuk menilai kestabilan sistem dan konvergensi algoritma.

Dalam konteks norma Euclidean, jarak antara dua vektor  $\mathbf{x}$  dan  $\mathbf{y}$  dalam ruang Euclidean didefinisikan sebagai norma dari selisih kedua vektor tersebut, yaitu

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (2.1.93)$$

Norma tersebut sering disebut juga norma  $L_2$  atau norma dua dan merupakan ukuran jarak yang paling umum digunakan dalam berbagai aplikasi matematika dan ilmu komputer.

Di sisi lain, terdapat pula perkalian dalam atau (*inner product*) yang merupakan operasi biner pada ruang vektor yang menghasilkan skalar. Perkalian dalam ini memungkinkan kita untuk mengukur sudut antara dua vektor dan menentukan apakah dua vektor ortogonal (tegak lurus) satu sama lain.

**Definisi 2.1.65 (Perkalian dalam, Horn dan Johnson 2012)** Misalkan  $\mathcal{V}$  adalah ruang vektor atas lapangan  $\mathcal{F}$  (baik  $\mathbb{R}$  maupun  $\mathbb{C}$ ). Fungsi  $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{F}$  disebut perkalian dalam pada  $\mathcal{V}$  jika memenuhi sifat-sifat berikut untuk setiap  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{V}$  dan skalar  $c \in \mathcal{F}$ :

(a)  $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$  dan  $\langle \mathbf{v}, \mathbf{v} \rangle = 0$  jika dan hanya jika  $\mathbf{v} = \mathbf{0}$ ;

(b)  $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$ ;

(c)  $\langle c\mathbf{v}, \mathbf{w} \rangle = c\langle \mathbf{v}, \mathbf{w} \rangle$ ; dan

(d)  $\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}$  (simetri Hermitian).

Beberapa contoh norma yang umum digunakan pada vektor dan matriks adalah sebagai berikut:

- (a) **Norma  $L_1$  (Norma Manhattan)**: Untuk vektor  $\mathbf{x} \in \mathbb{R}^n$ , norma  $L_1$  didefinisikan sebagai

$$\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|. \quad (2.1.94)$$

- (b) **Norma  $L_2$  (Norma Euclidean)**: Untuk vektor  $\mathbf{x} \in \mathbb{R}^n$ , norma  $L_2$  didefinisikan sebagai

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}. \quad (2.1.95)$$

- (c) **Norma  $L_p$** : Untuk vektor  $\mathbf{x} \in \mathbb{R}^n$  dan  $p \geq 1$ , norma  $L_p$  didefinisikan sebagai

$$\|\mathbf{x}\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}. \quad (2.1.96)$$

- (d) **Norma Tak Terbatas ( $L_\infty$ )**: Untuk vektor  $\mathbf{x} \in \mathbb{R}^n$ , norma tak terbatas didefinisikan sebagai

$$\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|. \quad (2.1.97)$$

- (e) **Norma Frobenius**: Untuk matriks  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , norma Frobenius didefinisikan sebagai

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}. \quad (2.1.98)$$

Definisi dari norma dan perkalian dalam tidak membutuhkan syarat bahwa ruang vektor harus berdimensi hingga (Horn dan Johnson, 2012). Berikut adalah empat contoh norma pada ruang vektor  $\mathcal{V}[a, b]$  dari semua fungsi kontinu atau kompleks pada interval real tertutup  $[a, b]$ .

- (a) Untuk  $f \in \mathcal{V}[a, b]$ , didefinisikan  $\|f\|_2 = \left( \int_a^b |f(x)|^2 dx \right)^{1/2}$ .

- (b) Untuk  $f \in \mathcal{V}[a, b]$ , didefinisikan  $\|f\|_1 = \int_a^b |f(x)| dx$ .

- (c) Untuk  $f \in \mathcal{V}[a, b]$ , didefinisikan  $\|f\|_p = \left( \int_a^b |f(x)|^p dx \right)^{1/p}$ , dengan  $1 \leq p < \infty$ .
- (d) Untuk  $f \in \mathcal{V}[a, b]$ , didefinisikan  $\|f\|_\infty = \sup_{x \in [a, b]} |f(x)| = \max_{x \in [a, b]} |f(x)|$ , karena  $f$  kontinu pada interval tertutup  $[a, b]$  sehingga supremum tercapai.

### 2.1.8 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 2.1.66 (Dhrymes, 2013)** *Misalkan*

$$\mathbf{y} = \psi(\mathbf{x}), \quad (2.1.99)$$

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

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \left[ \frac{\partial y_i}{\partial x_j} \right], \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n \quad (2.1.100)$$

merupakan notasi dari matriks derivatif parsial orde pertama (atau matriks Jacobian) dari transformasi  $\mathbf{x}$  ke  $\mathbf{y}$  sedemikian sehingga baris ke- $i$  berisi turunan parsial dari  $y_i$  terhadap elemen dari  $\mathbf{x}$ , yaitu

$$\frac{\partial y_i}{\partial x_1}, \frac{\partial y_i}{\partial x_2}, \dots, \frac{\partial y_i}{\partial x_n}. \quad (2.1.101)$$

**Proposisi 2.1.67** *Jika*

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad (2.1.102)$$

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

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{A}. \quad (2.1.103)$$

**Bukti.** Matriks Jacobian  $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$  didefinisikan sebagai matriks dengan elemen  $(i, j)$  dari

matriks tersebut adalah  $\frac{\partial y_i}{\partial x_j}$ . Perhatikan bahwa

$$y_i = \sum_{k=1}^n a_{ik} x_k. \quad (2.1.104)$$

Untuk setiap  $i$  dan  $j$ , turunan parsialnya adalah

$$\frac{\partial y_i}{\partial x_j} = \frac{\partial}{\partial x_j} \sum_{k=1}^n a_{ik} x_k = \sum_{k=1}^n a_{ik} \frac{\partial x_k}{\partial x_j}. \quad (2.1.105)$$

Sebab  $\frac{\partial x_k}{\partial x_j} = \delta_{kj}$  (delta Kronecker), maka

$$\frac{\partial y_i}{\partial x_j} = \sum_{k=1}^n a_{ik} \delta_{kj} = a_{ij}. \quad (2.1.106)$$

Oleh karena itu, elemen  $(i, j)$  dari matriks Jacobian adalah  $a_{ij}$ , yang berarti

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{A}. \quad (2.1.107)$$

■

**Proposisi 2.1.68** *Jika*

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad (2.1.108)$$

*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 berukuran  $r$ , maka*

$$\frac{\partial \mathbf{y}}{\partial \boldsymbol{\alpha}} = (\mathbf{x}^\top \otimes \mathbf{I}_m) \frac{\partial \text{vec}(\mathbf{A})}{\partial \boldsymbol{\alpha}} + \mathbf{A} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\alpha}}. \quad (2.1.109)$$

**Bukti.** Diferensiasikan kedua ruas persamaan  $\mathbf{y} = \mathbf{A}\mathbf{x}$  terhadap parameter skalar  $\alpha_j$ , yaitu

$$\frac{\partial \mathbf{y}}{\partial \alpha_j} = \frac{\partial \mathbf{A}}{\partial \alpha_j} \mathbf{x} + \mathbf{A} \frac{\partial \mathbf{x}}{\partial \alpha_j}. \quad (2.1.110)$$



Fokus pada suku pertama  $\frac{\partial \mathbf{A}}{\partial \alpha_j} \mathbf{x}$ . Elemen ke- $i$  dari vektor ini adalah

$$\left( \frac{\partial \mathbf{A}}{\partial \alpha_j} \mathbf{x} \right)_i = \sum_{k=1}^n \frac{\partial a_{ik}}{\partial \alpha_j} x_k. \quad (2.1.111)$$

Dengan menggunakan identitas vektorisasi  $\text{vec}(\mathbf{A}\mathbf{x}) = (\mathbf{x}^\top \otimes \mathbf{I}_m) \text{vec}(\mathbf{A})$ , didapatkan

$$\text{vec} \left( \frac{\partial \mathbf{A}}{\partial \alpha_j} \mathbf{x} \right) = (\mathbf{x}^\top \otimes \mathbf{I}_m) \text{vec} \left( \frac{\partial \mathbf{A}}{\partial \alpha_j} \right). \quad (2.1.112)$$

Apabila disusun kolom-kolom  $\text{vec} \left( \frac{\partial \mathbf{A}}{\partial \alpha_j} \right)$  untuk  $j = 1, \dots, r$  sebagai matriks  $mn \times r$ :

$$\frac{\partial \text{vec}(\mathbf{A})}{\partial \boldsymbol{\alpha}} = \begin{bmatrix} \text{vec} \left( \frac{\partial \mathbf{A}}{\partial \alpha_1} \right) & \cdots & \text{vec} \left( \frac{\partial \mathbf{A}}{\partial \alpha_r} \right) \end{bmatrix}, \quad (2.1.113)$$

akan diperoleh

$$(\mathbf{x}^\top \otimes \mathbf{I}_m) \frac{\partial \text{vec}(\mathbf{A})}{\partial \boldsymbol{\alpha}} = \begin{bmatrix} \frac{\partial \mathbf{A}}{\partial \alpha_1} \mathbf{x} & \cdots & \frac{\partial \mathbf{A}}{\partial \alpha_r} \mathbf{x} \end{bmatrix} = \frac{\partial \mathbf{A}}{\partial \boldsymbol{\alpha}} \mathbf{x}. \quad (2.1.114)$$

Untuk suku kedua, elemen ke- $i$  dari  $\mathbf{A} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\alpha}}$  adalah

$$\left( \mathbf{A} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\alpha}} \right)_i = \sum_{k=1}^n a_{ik} \frac{\partial x_k}{\partial \boldsymbol{\alpha}}. \quad (2.1.115)$$

Dengan menggabungkan kedua suku, diperoleh

$$\frac{\partial \mathbf{y}}{\partial \boldsymbol{\alpha}} = (\mathbf{x}^\top \otimes \mathbf{I}_m) \frac{\partial \text{vec}(\mathbf{A})}{\partial \boldsymbol{\alpha}} + \mathbf{A} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\alpha}}. \quad (2.1.116)$$

■

**Proposisi 2.1.69** *Jika*

$$\mathbf{y} = \mathbf{z}^\top \mathbf{A} \mathbf{x}, \quad (2.1.117)$$

*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

$$\frac{\partial \mathbf{y}}{\partial \mathbf{z}} = \mathbf{x}^\top \mathbf{A}^\top, \quad \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{z}^\top \mathbf{A}. \quad (2.1.118)$$

**Bukti.** Elemen  $(i, j)$  dari matriks Jacobian  $\frac{\partial \mathbf{y}}{\partial \mathbf{z}}$  adalah  $\frac{\partial y}{\partial z_i}$ . Perhatikan bahwa  $y$  adalah skalar, sehingga

$$y = \mathbf{z}^\top \mathbf{A} \mathbf{x} = \sum_{i=1}^m z_i (\mathbf{A} \mathbf{x})_i = \sum_{i=1}^m z_i \sum_{k=1}^n a_{ik} x_k. \quad (2.1.119)$$

Turunan parsialnya terhadap  $z_i$  adalah

$$\frac{\partial y}{\partial z_i} = \sum_{k=1}^n a_{ik} x_k = (\mathbf{A} \mathbf{x})_i. \quad (2.1.120)$$

Dalam bentuk vektor (sebagai baris),

$$\frac{\partial \mathbf{y}}{\partial \mathbf{z}} = \mathbf{x}^\top \mathbf{A}^\top. \quad (2.1.121)$$

Untuk turunan terhadap  $\mathbf{x}$ , elemen  $(j, 1)$  dari matriks Jacobian  $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$  adalah  $\frac{\partial y}{\partial x_j}$ .

$$\frac{\partial y}{\partial x_j} = \frac{\partial}{\partial x_j} \sum_{i=1}^m z_i \sum_{k=1}^n a_{ik} x_k = \sum_{i=1}^m z_i a_{ij} = (\mathbf{z}^\top \mathbf{A})_j. \quad (2.1.122)$$

Dalam bentuk vektor (sebagai baris),

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{z}^\top \mathbf{A}. \quad (2.1.123)$$

■

**Proposisi 2.1.70** Jika

$$\mathbf{y} = \mathbf{x}^\top \mathbf{A} \mathbf{x}, \quad (2.1.124)$$

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

maka

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{x}^\top (\mathbf{A} + \mathbf{A}^\top). \quad (2.1.125)$$

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

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = 2\mathbf{x}^\top \mathbf{A}. \quad (2.1.126)$$

**Bukti.** Elemen  $(j, 1)$  dari matriks Jacobian  $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$  adalah  $\frac{\partial y}{\partial x_j}$ . Perhatikan bahwa

$$y = \mathbf{x}^\top \mathbf{A} \mathbf{x} = \sum_{i=1}^n \sum_{k=1}^n x_i a_{ik} x_k. \quad (2.1.127)$$

Turunan parsial terhadap  $x_j$  adalah

$$\begin{aligned} \frac{\partial y}{\partial x_j} &= \frac{\partial}{\partial x_j} \sum_{i=1}^n \sum_{k=1}^n x_i a_{ik} x_k \\ &= \sum_{i=1}^n a_{ij} x_i + \sum_{k=1}^n a_{jk} x_k \\ &= \sum_{i=1}^n a_{ji} x_i + \sum_{k=1}^n a_{jk} x_k \\ &= (\mathbf{A}^\top \mathbf{x})_j + (\mathbf{A} \mathbf{x})_j. \end{aligned} \quad (2.1.128)$$

Dalam bentuk vektor (sebagai baris), ini adalah

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{x}^\top \mathbf{A}^\top + \mathbf{x}^\top \mathbf{A} = \mathbf{x}^\top (\mathbf{A} + \mathbf{A}^\top). \quad (2.1.129)$$

Jika  $\mathbf{A}$  simetri, maka  $\mathbf{A}^\top = \mathbf{A}$ , sehingga

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{x}^\top \mathbf{A} + \mathbf{x}^\top \mathbf{A} = 2\mathbf{x}^\top \mathbf{A}. \quad (2.1.130)$$

■

**Proposisi 2.1.71** Misalkan matriks  $\mathbf{A}$  nonsingular yang berukuran  $m \times m$  dan ber-

gantung dengan parameter skalar  $\alpha$ , maka

$$\frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{A}^{-1}. \quad (2.1.131)$$

**Bukti.** Dari identitas  $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_m$ , diferensiasikan kedua ruas terhadap parameter  $\alpha$ :

$$\frac{\partial}{\partial \alpha}(\mathbf{A}\mathbf{A}^{-1}) = \frac{\partial \mathbf{I}_m}{\partial \alpha} = \mathbf{0}. \quad (2.1.132)$$

Dengan aturan hasil kali untuk matriks, diperoleh

$$\frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{A}^{-1} + \mathbf{A} \frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = \mathbf{0}. \quad (2.1.133)$$

Analisis per elemen  $(i, j)$  dari identitas di atas:

$$\sum_k \frac{\partial a_{ik}}{\partial \alpha} (\mathbf{A}^{-1})_{kj} + \sum_k a_{ik} \frac{\partial (\mathbf{A}^{-1})_{kj}}{\partial \alpha} = 0. \quad (2.1.134)$$

Pindahkan suku kedua ke ruas kanan:

$$\mathbf{A} \frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = -\frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{A}^{-1}. \quad (2.1.135)$$

Kalikan kedua ruas dari kiri dengan  $\mathbf{A}^{-1}$ :

$$\frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{A}^{-1}. \quad (2.1.136)$$

Verifikasi per elemen  $(i, j)$ :

$$\frac{\partial (\mathbf{A}^{-1})_{ij}}{\partial \alpha} = -\sum_k \sum_\ell (\mathbf{A}^{-1})_{ik} \frac{\partial a_{k\ell}}{\partial \alpha} (\mathbf{A}^{-1})_{\ell j}. \quad (2.1.137)$$

■

**Teorema 2.1.72 (Dhrymes, 2013)** Untuk matriks dengan ukuran yang sesuai berlaku:

$$\frac{\partial}{\partial \mathbf{X}} \text{tr}(\mathbf{A}\mathbf{X}) = \mathbf{A}^\top, \quad \frac{\partial}{\partial \mathbf{X}} \text{tr}(\mathbf{X}^\top \mathbf{A}\mathbf{X}) = (\mathbf{A} + \mathbf{A}^\top) \mathbf{X}.$$

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

(a) Untuk  $\text{tr}(\mathbf{A}\mathbf{X})$ , perhatikan bahwa

$$\text{tr}(\mathbf{A}\mathbf{X}) = \sum_{i=1}^m (\mathbf{A}\mathbf{X})_{ii} = \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_{ji}. \quad (2.1.138)$$

Turunan parsial terhadap  $x_{kl}$  adalah

$$\frac{\partial}{\partial x_{kl}} \text{tr}(\mathbf{A}\mathbf{X}) = \frac{\partial}{\partial x_{kl}} \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_{ji} = a_{lk}, \quad (2.1.139)$$

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

$$\left( \frac{\partial}{\partial \mathbf{X}} \text{tr}(\mathbf{A}\mathbf{X}) \right)_{kl} = a_{lk} = (\mathbf{A}^\top)_{kl}, \quad (2.1.140)$$

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

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

$$\text{tr}(\mathbf{X}^\top \mathbf{A}\mathbf{X}) = \sum_{i,j,k} x_{ji} a_{jk} x_{ki}. \quad (2.1.141)$$

Turunan parsial terhadap  $x_{pq}$  adalah

$$\begin{aligned} \frac{\partial}{\partial x_{pq}} \text{tr}(\mathbf{X}^\top \mathbf{A}\mathbf{X}) &= \frac{\partial}{\partial x_{pq}} \sum_{i,j,k} x_{ji} a_{jk} x_{ki} \\ &= \sum_k a_{pk} x_{kq} + \sum_j x_{jq} a_{jp} \\ &= (\mathbf{A}\mathbf{X})_{pq} + (\mathbf{X}^\top \mathbf{A}^\top)_{qp}^\top \\ &= (\mathbf{A}\mathbf{X})_{pq} + (\mathbf{A}^\top \mathbf{X})_{pq}. \end{aligned} \quad (2.1.142)$$

Oleh karena itu,

$$\frac{\partial}{\partial \mathbf{X}} \text{tr}(\mathbf{X}^\top \mathbf{A}\mathbf{X}) = \mathbf{A}\mathbf{X} + \mathbf{A}^\top \mathbf{X} = (\mathbf{A} + \mathbf{A}^\top) \mathbf{X}. \quad (2.1.143)$$



**Contoh 2.1.73** Dalam banyak permasalahan estimasi, seringkali ingin dicari vektor parameter  $\beta$  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  $\beta \in \mathbb{R}^p$  yang meminimumkan jumlah kuadrat selisih antara  $\mathbf{y}$  dan  $\mathbf{X}\beta$ , atau kuadrat dari norma  $L_2$  dari vektor residual  $\mathbf{y} - \mathbf{X}\beta$ . Fungsi yang akan diminimumkan adalah

$$S(\beta) = \|\mathbf{y} - \mathbf{X}\beta\|_2^2 = (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta). \quad (2.1.144)$$

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

$$\begin{aligned} S(\beta) &= (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta) \\ &= \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{X}\beta - (\mathbf{X}\beta)^\top \mathbf{y} + (\mathbf{X}\beta)^\top \mathbf{X}\beta \\ &= \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{X}\beta - \beta^\top \mathbf{X}^\top \mathbf{y} + \beta^\top \mathbf{X}^\top \mathbf{X}\beta. \end{aligned} \quad (2.1.145)$$

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

$$S(\beta) = \mathbf{y}^\top \mathbf{y} - 2\beta^\top \mathbf{X}^\top \mathbf{y} + \beta^\top \mathbf{X}^\top \mathbf{X}\beta. \quad (2.1.146)$$

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

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

- Suku  $\beta^\top \mathbf{X}^\top \mathbf{X} \beta$  berbentuk  $\beta^\top \mathbf{A} \beta$  dengan  $\mathbf{A} = \mathbf{X}^\top \mathbf{X}$  yang simetris. Berdasarkan proposisi sebelumnya, turunannya adalah  $2\beta^\top \mathbf{X}^\top \mathbf{X}$ .

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

$$\frac{\partial S(\beta)}{\partial \beta} = -2\mathbf{y}^\top \mathbf{X} + 2\beta^\top \mathbf{X}^\top \mathbf{X}. \quad (2.1.147)$$

Untuk menemukan titik stasioner, disamakan gradien dengan nol:

$$\begin{aligned} -2\mathbf{y}^\top \mathbf{X} + 2\hat{\beta}^\top \mathbf{X}^\top \mathbf{X} &= \mathbf{0}^\top \\ \hat{\beta}^\top \mathbf{X}^\top \mathbf{X} &= \mathbf{y}^\top \mathbf{X}. \end{aligned} \quad (2.1.148)$$

Dengan mengambil *transpose* kedua ruas, diperoleh

$$\mathbf{X}^\top \mathbf{X} \hat{\beta} = \mathbf{X}^\top \mathbf{y}. \quad (2.1.149)$$

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

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (2.1.150)$$

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

$$\frac{\partial^2 S(\beta)}{\partial \beta \partial \beta^\top} = 2\mathbf{X}^\top \mathbf{X}. \quad (2.1.151)$$

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

$$\mathbf{v}^\top \mathbf{X}^\top \mathbf{X} \mathbf{v} = (\mathbf{X} \mathbf{v})^\top (\mathbf{X} \mathbf{v}) = \|\mathbf{X} \mathbf{v}\|^2 \geq 0. \quad (2.1.152)$$

Jika  $\mathbf{X}$  memiliki *rank* penuh (yaitu  $\text{rank}(\mathbf{X}) = p$ ), maka  $\mathbf{X}^\top \mathbf{X}$  bersifat definit positif, sehingga  $\hat{\beta}$  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.

### 2.2.1 Ruang Ukur dan Ukuran Probabilitas

**Definisi 2.2.1** (Aljabar- $\sigma$ , Ash 1972) Misalkan  $\Omega$  adalah suatu himpunan tak kosong. Suatu koleksi  $\mathcal{F} \subseteq 2^\Omega$  disebut sebagai aljabar- $\sigma$  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 2.2.2** Misalkan  $\Omega = \{1, 2, 3\}$  adalah ruang sampel dengan tiga elemen. Berikut adalah beberapa contoh aljabar- $\sigma$  pada  $\Omega$ .

- (a) Aljabar- $\sigma$  trivial:  $\mathcal{F}_1 = \{\emptyset, \Omega\} = \{\emptyset, \{1, 2, 3\}\}$ . Koleksi ini merupakan aljabar- $\sigma$  terkecil pada  $\Omega$ .
- (b) Aljabar- $\sigma$  power set:  $\mathcal{F}_2 = 2^\Omega = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$ . Koleksi ini merupakan aljabar- $\sigma$  terbesar pada  $\Omega$  yang memuat semua subhimpunan.
- (c) Aljabar- $\sigma$  yang dibangkitkan oleh  $\{1\}$ :  $\mathcal{F}_3 = \{\emptyset, \{1\}, \{2, 3\}, \{1, 2, 3\}\}$ . Dapat diperiksa bahwa  $\mathcal{F}_3$  memenuhi ketiga aksioma aljabar- $\sigma$ , yaitu  $\Omega \in \mathcal{F}_3$ , tertutup terhadap komplement (misalnya  $\{1\}^c = \{2, 3\} \in \mathcal{F}_3$ ), dan tertutup terhadap gabungan terhitung.

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



Kita tahu bahwa aljabar- $\sigma$  terbesar dari suatu himpunan  $\Omega$  adalah power set  $2^\Omega$ , yang memuat semua subhimpunan dari  $\Omega$ . Jika  $\mathcal{G}$  adalah suatu kelas himpunan, maka aljabar- $\sigma$  terkecil yang memuat  $\mathcal{G}$  dilambangkan dengan  $\sigma(\mathcal{G})$  dan disebut sebagai aljabar- $\sigma$  yang *dibangkitkan* oleh  $\mathcal{G}$ .

**Definisi 2.2.3 (Himpunan Borel, Ash 1972)** Misalkan  $\Omega = \mathbb{R}$ . Aljabar Borel pada  $\mathbb{R}$ , dilambangkan dengan  $\mathcal{B}(\mathbb{R})$ , adalah aljabar- $\sigma$  terkecil yang memuat semua interval terbuka  $(a, b]$ ,  $a, b \in \mathbb{R}$ . Himpunan-himpunan dalam  $\mathcal{B}(\mathbb{R})$  disebut sebagai himpunan Borel.

**Contoh 2.2.4** 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 2.2.5 (Ukuran)** Diberikan ruang sampel  $\Omega$  dan aljabar- $\sigma$   $\mathcal{F}$  relatif terhadap  $\Omega$ . Suatu fungsi  $\mu : \mathcal{F} \rightarrow \mathbb{R}^+$  disebut sebagai ukuran apabila:

- (a)  $\mu(F) \geq 0$  untuk semua  $F \in \mathcal{F}$ ;
- (b)  $\mu(\emptyset) = 0$ ; dan
- (c) untuk setiap barisan himpunan saling lepas ( $F_i \cap F_j = \emptyset$  untuk  $i \neq j$ ), maka  $\{F_n\}_{n=1}^\infty \subseteq \mathcal{F}$  berlaku.

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

**Definisi 2.2.6 (Ukuran probabilitas, Billingsley 1995)** Suatu fungsi  $\Pr : \mathcal{F} \rightarrow [0, 1]$  disebut sebagai ukuran probabilitas apabila:

- (a)  $\Pr(\Omega) = 1$  dan  $\Pr(\emptyset) = 0$ ;
- (b)  $0 \leq \Pr(A) \leq 1$  untuk semua  $A \in \mathcal{F}$ ; dan
- (c) untuk setiap barisan himpunan saling lepas  $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$  berlaku

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

Triple  $(\Omega, \mathcal{F}, \Pr)$  disebut sebagai *ruang probabilitas*. Penyokong atau *support* dari ukuran probabilitas  $\Pr$  adalah himpunan terkecil  $S \in \mathcal{F}$  sehingga  $\Pr(S) = 1$ .

**Contoh 2.2.7** Misalkan  $\Omega = \{1, 2, 3, 4, 5, 6\}$  adalah ruang sampel pelemparan sebuah dadu seimbang, dan  $\mathcal{F} = 2^{\Omega}$  adalah aljabar- $\sigma$  yang memuat semua subhimpunan dari  $\Omega$ . Didefinisikan fungsi  $\Pr : \mathcal{F} \rightarrow [0, 1]$  dengan

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

untuk setiap  $A \in \mathcal{F}$ , dengan  $|A|$  menyatakan banyaknya elemen dalam  $A$ . Akan diperiksa bahwa  $\Pr$  memenuhi aksioma ukuran probabilitas.

- (a)  $\Pr(\Omega) = \Pr(\{1, 2, 3, 4, 5, 6\}) = \frac{6}{6} = 1$  dan  $\Pr(\emptyset) = \frac{0}{6} = 0$ .
- (b) Untuk setiap  $A \in \mathcal{F}$ , berlaku  $0 \leq |A| \leq 6$ , sehingga  $0 \leq \Pr(A) = \frac{|A|}{6} \leq 1$ .
- (c) Untuk himpunan saling lepas  $A_1, A_2, \dots \subseteq \mathcal{F}$ , berlaku

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

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

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

**Definisi 2.2.8 (Ukuran Lebesgue-Stieltjes, Ash 1972)** Suatu ukuran Lebesgue-Stieltjes pada  $\mathbb{R}$  adalah ukuran  $\mu$  pada  $\mathcal{B}(\mathbb{R})$  sedemikian sehingga  $\mu(I) < \infty$  untuk setiap interval terbatas  $I \subset \mathbb{R}$ .

**Definisi 2.2.9 (Fungsi distribusi, Ash 1972)** Suatu fungsi distribusi pada  $\mathbb{R}$  adalah pemetaan  $F : \mathbb{R} \rightarrow \mathbb{R}$  yang increasing (jika  $a < b$ , maka  $F(a) \leq F(b)$ ) dan right-continuous ( $\lim_{x \rightarrow x^+} F(x) = F(x_0)$ ).

Apabila dipilih  $\mu(a, b] = F(b) - F(a) = \int_a^b f(x) dx$  untuk suatu fungsi  $f : \mathbb{R} \rightarrow \mathbb{R}, f \geq 0$  yang terintegral (Riemann), maka  $F$  adalah fungsi distribusi. Sebagai contoh, apabila dipilih  $f(x) = 1, \forall x$  dan  $F(x) = x$ , maka  $\mu(a, b] = b - a$ . Himpunan fungsi  $\mu$  yang diperoleh dengan cara ini disebut sebagai *ukuran Lebesgue* pada  $\mathcal{B}(\mathbb{R})$  (Ash, 1972).

**Definisi 2.2.10 (Ukuran Lebesgue)** Ukuran Lebesgue pada  $\mathcal{B}(\mathbb{R})$  adalah ukuran Lebesgue-Stieltjes yang didefinisikan oleh  $\mu(a, b] = b - a$  untuk setiap interval terbatas  $(a, b] \subset \mathbb{R}$ .

## 2.2.2 Variabel Acak sebagai Fungsi Terukur

**Definisi 2.2.11 (Variabel acak)** Diberikan ruang probabilitas  $(\Omega, \mathcal{F}, \Pr)$  dan ruang terukur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Suatu fungsi  $X : \Omega \rightarrow \mathbb{R}$  disebut sebagai variabel acak apabila  $X$  terukur, yaitu untuk setiap  $B \in \mathcal{B}(\mathbb{R})$  berlaku

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

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

**Teorema 2.2.12** Misalkan triple  $(\Omega, \mathcal{F}, \Pr)$  adalah ruang probabilitas dan  $X : \Omega \rightarrow \mathbb{R}$  adalah variabel acak, maka  $X$  menginduksi ruang probabilitas  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \Pr_X)$  pada garis real, dengan  $\Pr_X$  didefinisikan oleh

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

Ukuran  $\Pr_X$  disebut sebagai distribusi 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,

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

- (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

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

- (c)  **$\sigma$ -aditivitas:** Misalkan  $\{B_n\}_{n=1}^{\infty}$  adalah barisan himpunan saling lepas dalam  $\mathcal{B}(\mathbb{R})$ , yaitu  $B_i \cap B_j = \emptyset$  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  $X(\omega) \in B_i \cap B_j = \emptyset$ , yang merupakan kontradiksi. Jadi,  $X^{-1}(B_i) \cap X^{-1}(B_j) = \emptyset$  untuk  $i \neq j$ .

Selanjutnya, perhatikan bahwa

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

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

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

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

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

**Definisi 2.2.14 (Fungsi kepadatan probabilitas)** *Misalkan  $X$  adalah variabel acak dengan distribusi  $\Pr_X$ . Jika terdapat fungsi terintegralkan  $f_X : \mathbb{R} \rightarrow [0, \infty)$  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 2.2.15** Berikut adalah beberapa distribusi probabilitas yang sering digunakan dalam statistika dan pembelajaran mesin.

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

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

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  $X \sim \text{Bernoulli}(p)$ , jika

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

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

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

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

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  $X \sim \text{Exp}(\lambda)$ , jika memiliki fungsi kepadatan probabilitas

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

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

**Definisi 2.2.16 (Fungsi sederhana, Dudley 2002)** Suatu fungsi  $s : \Omega \rightarrow \mathbb{R}$  disebut fungsi sederhana apabila  $s$  hanya mengambil sejumlah hingga nilai berbeda. Fungsi sederhana dapat dituliskan dalam bentuk

$$s(\omega) = \sum_{k=1}^n c_k \mathbf{1}_{A_k}(\omega), \quad (2.2.17)$$

dengan  $c_1, \dots, c_n \in \mathbb{R}$  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 2.2.17** Misalkan  $\Omega = [0, 3]$  dan didefinisikan fungsi  $s : \Omega \rightarrow \mathbb{R}$  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  $0, 1, \dots, n$ .

**Definisi 2.2.18 (Dudley, 2002)** 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

$$\int_{\Omega} s \, d\mu := \sum_{k=1}^m c_k \mu(A_k) \in [0, \infty]. \quad (2.2.18)$$

**Definisi 2.2.19 (Dudley, 2002)** Untuk fungsi terukur non-negatif  $f : \Omega \rightarrow [0, \infty]$ , integral Lebesgue didefinisikan sebagai

$$\int_{\Omega} f \, d\mu := \sup \left\{ \int_{\Omega} s \, d\mu : s \text{ sederhana}, 0 \leq s \leq f \right\}. \quad (2.2.19)$$

Integral ini selalu terdefinisi (meskipun mungkin bernilai  $\infty$ ).

**Definisi 2.2.20 (Dudley, 2002)** Untuk fungsi terukur  $f : \Omega \rightarrow \mathbb{R}$ , definisikan bagian positif  $f^+ := \max(f, 0)$  dan bagian negatif  $f^- := -\min(f, 0) = \max(-f, 0)$ , sehingga  $f = f^+ - f^-$ . Integral Lebesgue dari  $f$  didefinisikan sebagai

$$\int_{\Omega} f \, d\mu := \int_{\Omega} f^+ \, d\mu - \int_{\Omega} f^- \, d\mu, \quad (2.2.20)$$

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  $\int_{\Omega} |f| \, d\mu < \infty$ .

Apabila  $\mu$  adalah ukuran Lebesgue  $\lambda$  pada  $\mathcal{B}(\mathbb{R})$ , maka  $d\lambda(x)$  dapat ditulis menjadi  $dx$ .

**Contoh 2.2.21** Misalkan  $f : [0, 1] \rightarrow \mathbb{R}$  dengan  $f(x) = x^2$ . Integral Lebesgue dari  $f$  terhadap ukuran Lebesgue  $\lambda$  adalah

$$\int_{[0,1]} x^2 \, d\lambda(x) = \int_0^1 x^2 \, dx = \frac{1}{3}. \quad (2.2.21)$$

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 2.2.22 (Ekspektasi Lebesgue, Dudley 2002)** Misalkan  $(\Omega, \mathcal{F}, \Pr)$  adalah ruang probabilitas dan  $X : \Omega \rightarrow \mathbb{R}$  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

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\Pr(\omega). \quad (2.2.22)$$

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

$$\mathbb{E}[X] = \int_{\mathbb{R}} x d\Pr_X(x). \quad (2.2.23)$$

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

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx. \quad (2.2.24)$$

Untuk kasus diskrit, ekspektasi dapat dihitung menggunakan fungsi massa probabilitas  $p_X$  sebagai berikut:

$$\mathbb{E}[X] = \sum_{x \in \text{support}(X)} x \cdot p_X(x). \quad (2.2.25)$$

**Contoh 2.2.23** Berikut adalah contoh perhitungan ekspektasi untuk beberapa distribusi.

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

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

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

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

Ekspektasi distribusi normal adalah parameter lokasinya  $\mu$ , yang menyatakan pusat simetri dari distribusi.

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

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

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

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

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

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

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

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

$$\mathbb{E}[X] \leq \mathbb{E}[Y]. \quad (2.2.31)$$

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

$$\mathbb{E}[c] = c. \quad (2.2.32)$$

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

$$\mathbb{E}[X] \geq 0. \quad (2.2.33)$$

(e) **Ketaksamaan segitiga:** Berlaku

$$|\mathbb{E}[X]| \leq \mathbb{E}[|X|]. \quad (2.2.34)$$

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

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]. \quad (2.2.35)$$

**Bukti.** Berikut adalah pembuktian untuk setiap sifat ekspektasi.

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

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

(b) **Monotonisitas:** Jika  $X \leq Y$  hampir pasti, maka  $Y - X \geq 0$  hampir pasti.

Berdasarkan sifat integral Lebesgue untuk fungsi non-negatif, berlaku

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

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

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

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

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

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

$$-\mathbb{E}[|X|] \leq \mathbb{E}[X] \leq \mathbb{E}[|X|], \quad (2.2.40)$$

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

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

■

**Teorema 2.2.25 (Hukum Statistikawan Tak Sadar (*Law of the Unconscious Statistician*, LOTUS), Ross 1972)** Misalkan  $X$  adalah variabel acak dengan fungsi distribusi  $F_X$  dan  $g : \mathbb{R} \rightarrow \mathbb{R}$  adalah fungsi terukur. Maka ekspektasi dari  $g(X)$

dapat dihitung tanpa perlu mengetahui distribusi  $g(X)$  secara eksplisit, yaitu

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

Secara khusus:

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

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

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

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

**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

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

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

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

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

**Contoh 2.2.26** 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 2.2.27 (Momen, Billingsley 1995)** 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 rata-rata atau nilai harapan dari  $X$ . Selain itu, momen pusat ke- $k$  didefinisikan sebagai

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

**Definisi 2.2.28 (Variansi, Billingsley 1995)** Untuk variabel acak  $X$  dengan  $\mathbb{E}[|X|^2] < \infty$ , variansi didefinisikan sebagai momen pusat kedua, yaitu

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]. \quad (2.2.51)$$

Variansi dapat juga dihitung menggunakan rumus alternatif

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

**Contoh 2.2.29** 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 (rata-rata) dari  $X$  adalah

$$\begin{aligned} \mathbb{E}[X] &= \sum_x x \cdot \Pr(X = x) \\ &= 1 \cdot 0,3 + 2 \cdot 0,5 + 3 \cdot 0,2 = 0,3 + 1,0 + 0,6 = 1,9. \end{aligned} \quad (2.2.53)$$

Momen kedua dari  $X$  adalah

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_x x^2 \cdot \Pr(X = x) \\ &= 1^2 \cdot 0,3 + 2^2 \cdot 0,5 + 3^2 \cdot 0,2 = 0,3 + 2,0 + 1,8 = 4,1. \end{aligned} \quad (2.2.54)$$

Variansi dari  $X$  dapat dihitung dengan rumus alternatif, yaitu

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 4,1 - (1,9)^2 = 4,1 - 3,61 = 0,49. \quad (2.2.55)$$

**Definisi 2.2.30 (Kovariansi, Billingsley 1995)** Untuk dua variabel acak  $X$  dan  $Y$  dengan momen orde dua hingga, kovariansi didefinisikan sebagai

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]. \quad (2.2.56)$$

*Rumus alternatif untuk kovariansi adalah*

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

**Contoh 2.2.31** Misalkan  $(X, Y)$  adalah variabel acak gabungan dengan distribusi bersama

$$\begin{aligned} \Pr(X = 0, Y = 0) &= 0,2, & \Pr(X = 0, Y = 1) &= 0,1, \\ \Pr(X = 1, Y = 0) &= 0,3, & \Pr(X = 1, Y = 1) &= 0,4. \end{aligned} \quad (2.2.58)$$

Rataan dari  $X$  dan  $Y$  adalah

$$\mathbb{E}[X] = 0 \cdot (0,2 + 0,1) + 1 \cdot (0,3 + 0,4) = 0,7, \quad (2.2.59)$$

$$\mathbb{E}[Y] = 0 \cdot (0,2 + 0,3) + 1 \cdot (0,1 + 0,4) = 0,5. \quad (2.2.60)$$

Nilai harapan dari  $XY$  adalah

$$\begin{aligned} \mathbb{E}[XY] &= \sum_{x,y} xy \cdot \Pr(X = x, Y = y) \\ &= 0 \cdot 0 \cdot 0,2 + 0 \cdot 1 \cdot 0,1 + 1 \cdot 0 \cdot 0,3 + 1 \cdot 1 \cdot 0,4 = 0,4. \end{aligned} \quad (2.2.61)$$

Kovariansi dari  $X$  dan  $Y$  adalah

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0,4 - 0,7 \cdot 0,5 = 0,4 - 0,35 = 0,05. \quad (2.2.62)$$

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 2.2.32 (Probabilitas bersyarat, Ash 1972)** 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  $\Pr(A | B)$ , didefinisikan sebagai

$$\Pr(A | B) = \frac{\Pr(A \cap B)}{\Pr(B)}. \quad (2.2.63)$$

**Contoh 2.2.33** 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

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

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

Probabilitas bersyarat dari  $A$  diberikan  $B$  adalah

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

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

**Definisi 2.2.34 (Independensi dua kejadian, Billingsley 1995)** Dua kejadian  $A$  dan  $B$  dikatakan independen apabila

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B). \quad (2.2.67)$$

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

**Contoh 2.2.35** 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 sera-

gam, yaitu  $\Pr(\omega) = \frac{1}{4}$  untuk setiap  $\omega \in \Omega$ , sehingga

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

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

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

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 2.2.36 (Independensi variabel acak, Billingsley 1995)** Barisan variabel acak  $\{X_i\}_{i \in I}$  dikatakan saling independen apabila untuk setiap himpunan Borel  $A_i \subseteq \mathbb{R}$  dan setiap himpunan indeks hingga  $\{i_1, i_2, \dots, i_n\} \subseteq I$  berlaku

$$\Pr(X_{i_1} \in A_{i_1}, X_{i_2} \in A_{i_2}, \dots, X_{i_n} \in A_{i_n}) = \prod_{j=1}^n \Pr(X_{i_j} \in A_{i_j}). \quad (2.2.71)$$

Secara khusus, dua variabel acak  $X$  dan  $Y$  dikatakan independen apabila barisan  $\{X, Y\}$  saling independen.

Akibat dari definisi di atas adalah bahwa jika  $X$  dan  $Y$  independen, maka distribusi bersama mereka sama dengan hasil kali distribusi marginal, yaitu

$$\Pr_{X,Y}(A, B) = \Pr_X(A) \cdot \Pr_Y(B) \quad (2.2.72)$$

untuk setiap himpunan Borel  $A, B \subseteq \mathbb{R}$ . Hal ini juga mengimplikasikan bahwa fungsi distribusi kumulatif bersama dapat ditulis sebagai

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y). \quad (2.2.73)$$

**Contoh 2.2.37** Misalkan  $X$  dan  $Y$  adalah variabel acak independen dengan  $X \sim$

Bernoulli(0,5) dan  $Y \sim \text{Bernoulli}(0,5)$ . Maka

$$\Pr(X = 1, Y = 1) = \Pr(X = 1) \cdot \Pr(Y = 1) = 0,5 \cdot 0,5 = 0,25, \quad (2.2.74)$$

$$\Pr(X = 0, Y = 1) = \Pr(X = 0) \cdot \Pr(Y = 1) = 0,5 \cdot 0,5 = 0,25. \quad (2.2.75)$$

Distribusi bersama sepenuhnya ditentukan oleh hasil kali distribusi marginal.

**Definisi 2.2.38 (Independensi bersyarat, Billingsley 1995)** Dua kejadian  $A$  dan  $B$  dikatakan independen bersyarat diberikan kejadian  $C$  (dengan  $\Pr(C) > 0$ ) apabila

$$\Pr(A \cap B \mid C) = \Pr(A \mid C) \cdot \Pr(B \mid C). \quad (2.2.76)$$

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

$$\Pr(X \in A, Y \in B \mid Z = z) = \Pr(X \in A \mid Z = z) \cdot \Pr(Y \in B \mid Z = z). \quad (2.2.77)$$

### 2.2.6 Hukum Probabilitas Total dan Teorema Bayes

**Teorema 2.2.39 (Hukum probabilitas total, Ash 1972)** Misalkan barisan kejadian  $\{B_1, B_2, \dots, B_n\}$  adalah partisi dari ruang sampel  $\Omega$ , yaitu  $B_i \cap B_j = \emptyset$  untuk  $i \neq j$  dan  $\bigcup_{i=1}^n B_i = \Omega$ , dengan  $\Pr(B_i) > 0$  untuk semua  $i$ . Maka untuk setiap kejadian  $A$  berlaku

$$\Pr(A) = \sum_{i=1}^n \Pr(A \cap B_i) = \sum_{i=1}^n \Pr(A \mid B_i) \cdot \Pr(B_i). \quad (2.2.78)$$

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

$$\Pr(A) = \int_{-\infty}^{\infty} \Pr(A \mid Y = y) \cdot f_Y(y) \, dy. \quad (2.2.79)$$

**Bukti.** Sebab  $\{B_1, B_2, \dots, B_n\}$  adalah partisi dari  $\Omega$ , maka

$$A = A \cap \Omega = A \cap \left( \bigcup_{i=1}^n B_i \right) = \bigcup_{i=1}^n (A \cap B_i). \quad (2.2.80)$$

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

$$\Pr(A) = \sum_{i=1}^n \Pr(A \cap B_i) = \sum_{i=1}^n \Pr(A \mid B_i) \cdot \Pr(B_i). \quad (2.2.81)$$

■

**Contoh 2.2.40** 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:

$$\begin{aligned} \Pr(\text{Cacat}) &= \Pr(\text{Cacat} \mid \text{Mesin 1}) \cdot \Pr(\text{Mesin 1}) \\ &\quad + \Pr(\text{Cacat} \mid \text{Mesin 2}) \cdot \Pr(\text{Mesin 2}) \\ &\quad + \Pr(\text{Cacat} \mid \text{Mesin 3}) \cdot \Pr(\text{Mesin 3}) \\ &= 0,02 \cdot 0,50 + 0,03 \cdot 0,30 + 0,05 \cdot 0,20 \\ &= 0,01 + 0,009 + 0,01 = 0,029. \end{aligned} \quad (2.2.82)$$

Jadi, probabilitas suatu komponen adalah cacat adalah 2,9%.

**Teorema 2.2.41 (Teorema Bayes, Ross 1972)** Misalkan  $\{B_1, B_2, \dots, B_n\}$  adalah partisi dari ruang sampel  $\Omega$  dengan  $\Pr(B_i) > 0$  untuk semua  $i$ , dan misalkan  $A$  adalah kejadian dengan  $\Pr(A) > 0$ . Maka untuk setiap  $j = 1, 2, \dots, n$  berlaku

$$\Pr(B_j \mid A) = \frac{\Pr(A \mid B_j) \cdot \Pr(B_j)}{\sum_{i=1}^n \Pr(A \mid B_i) \cdot \Pr(B_i)}. \quad (2.2.83)$$

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

$$\Pr(B | A) = \frac{\Pr(A | B) \cdot \Pr(B)}{\Pr(A)}. \quad (2.2.84)$$

**Bukti.** Berdasarkan definisi probabilitas bersyarat:

$$\Pr(B_j | A) = \frac{\Pr(A \cap B_j)}{\Pr(A)} = \frac{\Pr(A | B_j) \cdot \Pr(B_j)}{\Pr(A)}. \quad (2.2.85)$$

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

**Contoh 2.2.42** 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:

$$\begin{aligned} \Pr(\text{Mesin 3} | \text{Cacat}) &= \frac{\Pr(\text{Cacat} | \text{Mesin 3}) \cdot \Pr(\text{Mesin 3})}{\Pr(\text{Cacat})} \\ &= \frac{0,05 \cdot 0,20}{0,029} = \frac{0,01}{0,029} \approx 0,345. \end{aligned} \quad (2.2.86)$$

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.

Dalam konteks teorema Bayes, terdapat beberapa terminologi penting:

- (a)  $\Pr(B)$  disebut probabilitas *prior*, yaitu keyakinan awal tentang  $B$  sebelum mengamati data  $A$ .
- (b)  $\Pr(A | B)$  disebut *likelihood*, yaitu probabilitas mengamati data  $A$  diberikan  $B$  benar.
- (c)  $\Pr(B | A)$  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:  $\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$ .

### 2.2.7 Hukum Ekspektasi Total dan Variansi Total

**Teorema 2.2.43 (Hukum ekspektasi total (*Law of Total Expectation*), Ross 1972)**

Misalkan  $X$  dan  $Y$  adalah variabel acak pada ruang probabilitas yang sama dengan  $\mathbb{E}[|X|] < \infty$ . Maka

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X | Y]]. \quad (2.2.87)$$

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

$$\mathbb{E}[X] = \sum_j \mathbb{E}[X | Y = y_j] \cdot \Pr(Y = y_j). \quad (2.2.88)$$

Untuk kasus kontinu dengan  $Y$  memiliki kepadatan  $f_Y$ :

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X | Y = y] \cdot f_Y(y) dy. \quad (2.2.89)$$

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

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X | Y]] &= \sum_j \mathbb{E}[X | Y = y_j] \cdot \Pr(Y = y_j) \\ &= \sum_j \left( \sum_i x_i \cdot \Pr(X = x_i | Y = y_j) \right) \cdot \Pr(Y = y_j) \\ &= \sum_i \sum_j x_i \cdot \Pr(X = x_i | Y = y_j) \cdot \Pr(Y = y_j) \\ &= \sum_i \sum_j x_i \cdot \Pr(X = x_i, Y = y_j) \\ &= \sum_i x_i \cdot \Pr(X = x_i) = \mathbb{E}[X]. \end{aligned} \quad (2.2.90)$$

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

menggunakan definisi ekspektasi bersyarat untuk variabel acak kontinu:

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

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

$$\begin{aligned}
 \mathbb{E} [\mathbb{E}[X | Y]] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f_{X,Y}(x, y) dx dy \\
 &= \int_{-\infty}^{\infty} x \left( \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \right) dx \\
 &= \int_{-\infty}^{\infty} x \cdot f_X(x) dx = \mathbb{E}[X], \tag{2.2.92}
 \end{aligned}$$

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 2.2.44** 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:

$$\begin{aligned}
 \mathbb{E}[S] &= \mathbb{E} [\mathbb{E}[S | N]] = \mathbb{E} \left[ \mathbb{E} \left[ \sum_{i=1}^N X_i | N \right] \right] \\
 &= \mathbb{E}[N \cdot \mu] = \mu \cdot \mathbb{E}[N] = \mu\lambda. \tag{2.2.93}
 \end{aligned}$$

**Teorema 2.2.45 (Ross, 1972)** Misalkan  $X$  dan  $Y$  adalah variabel acak dengan  $\text{Var}(X) < \infty$ . Maka

$$\text{Var}(X) = \mathbb{E} [\text{Var}(X | Y)] + \text{Var} (\mathbb{E}[X | Y]). \tag{2.2.94}$$

*Rumus ini dikenal juga sebagai dekomposisi variansi, dengan interpretasi:*

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

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

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

$$\text{Var}(X | Y) = \mathbb{E}[X^2 | Y] - (\mathbb{E}[X | Y])^2. \quad (2.2.95)$$

Dengan mengambil ekspektasi kedua ruas:

$$\mathbb{E}[\text{Var}(X | Y)] = \mathbb{E}[\mathbb{E}[X^2 | Y]] - \mathbb{E}[(\mathbb{E}[X | Y])^2] = \mathbb{E}[X^2] - \mathbb{E}[(\mathbb{E}[X | Y])^2]. \quad (2.2.96)$$

Di sisi lain:

$$\begin{aligned} \text{Var}(\mathbb{E}[X | Y]) &= \mathbb{E}[(\mathbb{E}[X | Y])^2] - (\mathbb{E}[\mathbb{E}[X | Y]])^2 \\ &= \mathbb{E}[(\mathbb{E}[X | Y])^2] - (\mathbb{E}[X])^2. \end{aligned} \quad (2.2.97)$$

Dengan menjumlahkan kedua hasil:

$$\begin{aligned} \mathbb{E}[\text{Var}(X | Y)] + \text{Var}(\mathbb{E}[X | Y]) &= \mathbb{E}[X^2] - \mathbb{E}[(\mathbb{E}[X | Y])^2] \\ &\quad + \mathbb{E}[(\mathbb{E}[X | Y])^2] - (\mathbb{E}[X])^2 \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \text{Var}(X). \end{aligned} \quad (2.2.98)$$

■

**Contoh 2.2.46** Melanjutkan contoh sebelumnya, variansi total penjualan harian  $S = \sum_{i=1}^N X_i$  dapat dihitung dengan hukum variansi total. Misalkan  $\text{Var}(X_i) = \sigma^2$ .



Maka

$$\text{Var}(S \mid N) = N \cdot \sigma^2, \quad (2.2.99)$$

$$\mathbb{E}[S \mid N] = N \cdot \mu. \quad (2.2.100)$$

Dengan hukum variansi total:

$$\begin{aligned} \text{Var}(S) &= \mathbb{E}[\text{Var}(S \mid N)] + \text{Var}(\mathbb{E}[S \mid N]) \\ &= \mathbb{E}[N \cdot \sigma^2] + \text{Var}(N \cdot \mu) \\ &= \sigma^2 \cdot \mathbb{E}[N] + \mu^2 \cdot \text{Var}(N) \\ &= \sigma^2 \lambda + \mu^2 \lambda = \lambda(\sigma^2 + \mu^2). \end{aligned} \quad (2.2.101)$$

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

### 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 Billingsley (1995), Dudley (2002), dan van der Vaart (1998). 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, Ross 1972)** *Misalkan  $X$  adalah variabel acak non-negatif dengan ekspektasi  $\mathbb{E}[X] < \infty$ . Maka untuk setiap  $a > 0$  berlaku*

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

**Bukti.** Dengan definisi ekspektasi:

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

Karena  $X$  non-negatif, maka

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

Oleh karena itu,

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

■

**Teorema 2.3.2 (Ketaksamaan Chebyshev, Ross 1972)** Misalkan  $X$  adalah variabel acak dengan ekspektasi  $\mathbb{E}[X]$  dan variansi  $\text{Var}(X) < \infty$ . Maka untuk setiap  $\varepsilon > 0$  berlaku

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

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

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

■

### 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  $n$  (Billingsley, 1995).

### 2.3.2 Konvergensi Barisan Variabel Acak

Pemahaman mengenai berbagai mode konvergensi barisan variabel acak sangat penting dalam statistika asimtotik.

**Definisi 2.3.3 (Pishro-Nik, 2014)** Barisan variabel acak  $\{X_n\}$  dikatakan konvergen hampir pasti (almost surely) ke  $X$ , ditulis  $X_n \xrightarrow{a.s.} X$ , apabila

$$\Pr(\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1. \quad (2.3.7)$$

**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

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

Untuk  $\omega = 1$ , berlaku  $X_n(1) = 1$  untuk semua  $n$ , sehingga  $\lim_{n \rightarrow \infty} X_n(1) = 1$ . Definisikan  $X(\omega) = 0$  untuk  $\omega \in [0, 1)$  dan  $X(1) = 1$ . Maka

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

sehingga  $X_n \rightarrow X$  hampir pasti.

**Definisi 2.3.5 (Pishro-Nik, 2014)** Barisan  $\{X_n\}$  dikatakan konvergen dalam probabilitas ke  $X$ , ditulis  $X_n \xrightarrow{p} X$ , apabila untuk setiap  $\varepsilon > 0$  berlaku

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

$X_n \xrightarrow{p} X$  dapat dinotasikan juga sebagai  $\text{plim}_{n \rightarrow \infty} X_n = X$ .

**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

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

ketika  $n \rightarrow \infty$ . Oleh karena itu,  $X_n \xrightarrow{p} 0$ . Namun, perhatikan bahwa  $X_n$  tidak konvergen hampir pasti ke 0, karena untuk setiap  $\omega$  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 (Pishro-Nik, 2014)** Barisan  $\{X_n\}$  dikatakan konvergen dalam distribusi ke  $X$ , ditulis  $X_n \Rightarrow X$  atau  $X_n \xrightarrow{d} X$ , apabila

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \quad (2.3.9)$$

untuk setiap titik  $x$  di mana  $F_X$  kontinu. Alternatifnya, konvergensi dalam distribusi dapat didefinisikan sebagai

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)] \quad (2.3.10)$$

untuk setiap fungsi kontinu terbatas  $f$ .

**Contoh 2.3.8** Misalkan  $X_n \sim \mathcal{N}(0, 1 + 1/n)$  untuk  $n = 1, 2, \dots$ . 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  $n \rightarrow \infty$ , variansi  $1 + 1/n \rightarrow 1$ , sehingga

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

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 (Pishro-Nik, 2014)** Barisan variabel acak  $\{X_n\}$  dikatakan konvergen dalam rata-rata kuadrat (mean square) ke  $X$ , ditulis  $X_n \xrightarrow{L_2} X$  atau  $X_n \xrightarrow{m.s.} X$ , apabila

$$\lim_{n \rightarrow \infty} \mathbb{E}[(X_n - X)^2] = 0. \quad (2.3.11)$$

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

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

ketika  $n \rightarrow \infty$ . Oleh karena itu,  $X_n \xrightarrow{L_2} 0$ .

**Contoh 2.3.11** Misalkan  $X_n \sim \mathcal{N}(0, 1/n)$  untuk  $n = 1, 2, \dots$ . Akan ditunjukkan bahwa  $X_n \xrightarrow{L_2} 0$ . Perhatikan bahwa

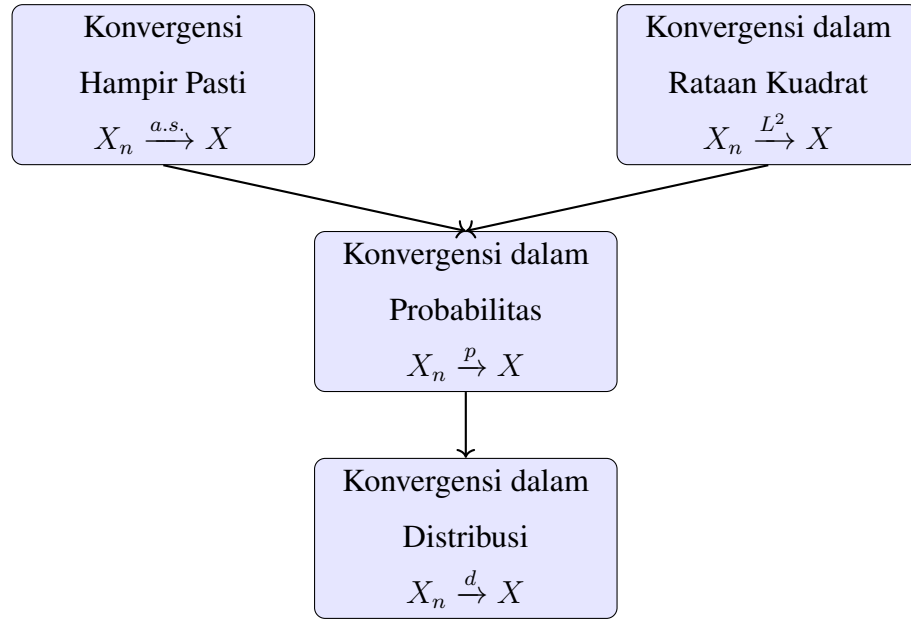
$$\mathbb{E}[(X_n - 0)^2] = \mathbb{E}[X_n^2] = \text{Var}(X_n) + (\mathbb{E}[X_n])^2 = \frac{1}{n} + 0 = \frac{1}{n} \rightarrow 0$$

ketika  $n \rightarrow \infty$ . 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  $\Pr(\lim_{n \rightarrow \infty} X_n = X) = 1$ . Definisikan himpunan

$$A = \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\}. \quad (2.3.12)$$

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,

$$A \subseteq \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} B_n(\varepsilon)^c. \quad (2.3.13)$$

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

$$\Pr\left(\limsup_{n \rightarrow \infty} B_n(\varepsilon)\right) \leq \Pr(A^c) = 0. \quad (2.3.14)$$

Berdasarkan lemma Borel–Cantelli, jika  $\Pr(\limsup_n B_n) = 0$ , maka  $\Pr(B_n) \rightarrow 0$ .

Oleh karena itu,

$$\lim_{n \rightarrow \infty} \Pr(|X_n - X| > \varepsilon) = 0, \quad (2.3.15)$$

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

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

$$\Pr(|X_n - X| > \varepsilon) = \Pr((X_n - X)^2 > \varepsilon^2) \leq \frac{\mathbb{E}[(X_n - X)^2]}{\varepsilon^2}. \quad (2.3.16)$$

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

$$\lim_{n \rightarrow \infty} \Pr(|X_n - X| > \varepsilon) \leq \lim_{n \rightarrow \infty} \frac{\mathbb{E}[(X_n - X)^2]}{\varepsilon^2} = 0. \quad (2.3.17)$$

Sebab probabilitas selalu non-negatif, diperoleh  $\Pr(|X_n - X| > \varepsilon) \rightarrow 0$ , 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  $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ . 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

$$\begin{aligned} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| &= |\mathbb{E}[f(X_n) - f(X)]| \\ &\leq \mathbb{E}[|f(X_n) - f(X)|]. \end{aligned} \quad (2.3.18)$$

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

$$\begin{aligned}\mathbb{E}[|f(X_n) - f(X)|] &= \mathbb{E}[|f(X_n) - f(X)| \cdot \mathbf{1}_{\{|X_n - X| \leq \eta\}}] \\ &\quad + \mathbb{E}[|f(X_n) - f(X)| \cdot \mathbf{1}_{\{|X_n - X| > \eta\}}].\end{aligned}\quad (2.3.19)$$

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

$$\mathbb{E}[|f(X_n) - f(X)| \cdot \mathbf{1}_{\{|X_n - X| \leq \eta\}}] < \delta. \quad (2.3.20)$$

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

$$\mathbb{E}[|f(X_n) - f(X)| \cdot \mathbf{1}_{\{|X_n - X| > \eta\}}] \leq 2M \cdot \Pr(|X_n - X| > \eta). \quad (2.3.21)$$

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

$$\limsup_{n \rightarrow \infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq \delta. \quad (2.3.22)$$

Sebab  $\delta > 0$  sembarang, maka  $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$ , 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  $\mathbb{E}[f(X_n)] \rightarrow f(c)$ .

Untuk setiap  $\varepsilon > 0$ , definisikan fungsi kontinu terbatas

$$f_\varepsilon(x) = \begin{cases} 1, & \text{jika } |x - c| \geq \varepsilon, \\ \frac{|x - c|}{\varepsilon}, & \text{jika } |x - c| < \varepsilon. \end{cases} \quad (2.3.23)$$

Fungsi  $f_\varepsilon$  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

$$\Pr(|X_n - c| \geq \varepsilon) = \mathbb{E}[\mathbf{1}_{\{|X_n - c| \geq \varepsilon\}}] \leq \mathbb{E}[f_\varepsilon(X_n)]. \quad (2.3.24)$$



Sebab  $X_n \xrightarrow{d} c$  dan  $f_\varepsilon$  kontinu terbatas, maka  $\mathbb{E}[f_\varepsilon(X_n)] \rightarrow f_\varepsilon(c) = 0$ . Oleh karena itu,

$$\lim_{n \rightarrow \infty} \Pr(|X_n - c| \geq \varepsilon) = 0, \quad (2.3.25)$$

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

## 2.4 Teori Probabilitas Asimtotik

Teori probabilitas asimtotik mempelajari perilaku limit dari barisan variabel acak 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 van der Vaart (1998) dengan fondasi probabilitas dari Billingsley (1995).

### 2.4.1 Hukum Bilangan Besar

**Teorema 2.4.1 (Weak law of large numbers, Pishro-Nik 2014)** 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

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} \mu. \quad (2.4.1)$$

**Bukti.** Andaikan  $\{X_i\}_{i=1}^\infty$  adalah barisan variabel acak i.i.d. dengan  $\mathbb{E}[X_1] = \mu$  dan  $\text{Var}(X_1) = \sigma^2 < \infty$ . 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

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n} \cdot n\mu = \mu. \quad (2.4.2)$$

Selanjutnya, karena  $X_i$  saling independen, berlaku

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}. \quad (2.4.3)$$

Dengan menggunakan ketaksamaan Chebyshev, untuk setiap  $\varepsilon > 0$  berlaku

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}. \quad (2.4.4)$$

Dengan mengambil limit  $n \rightarrow \infty$ , diperoleh

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \lim_{n \rightarrow \infty} \frac{\sigma^2}{n\varepsilon^2} = 0. \quad (2.4.5)$$

Oleh karena probabilitas selalu non-negatif, berlaku

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0, \quad (2.4.6)$$

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, van der Vaart 1998)** Suatu estimator  $\hat{\theta}_n$  untuk parameter  $\theta$  dikatakan konsisten apabila

$$\hat{\theta}_n \xrightarrow{p} \theta. \quad (2.4.7)$$

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 karakte-

ristik selalu ada untuk setiap variabel acak.

**Definisi 2.4.3 (Fungsi karakteristik, Billingsley 1995)** Misalkan  $X$  adalah variabel acak. Fungsi karakteristik dari  $X$  didefinisikan sebagai

$$\phi_X(t) = \mathbb{E} [e^{itX}] = \int_{-\infty}^{\infty} e^{itx} dF_X(x), \quad t \in \mathbb{R}, \quad (2.4.8)$$

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

$$\phi_X(t) = \exp \left( i\mu t - \frac{\sigma^2 t^2}{2} \right). \quad (2.4.9)$$

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

$$\phi_X(t) = \exp \left( \lambda(e^{it} - 1) \right). \quad (2.4.10)$$

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

$$\phi_X(t) = \frac{\lambda}{\lambda - it}. \quad (2.4.11)$$

**Proposisi 2.4.5** 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)  $\phi_X$  kontinu seragam pada  $\mathbb{R}$ .

(d)  $\overline{\phi_X(t)} = \phi_X(-t)$ , dengan  $\bar{z}$  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** Jika  $\mathbb{E}[|X|^n] < \infty$ , maka fungsi karakteristik  $\phi_X$  terdiferensiasi  $n$  kali dan

$$\phi_X^{(k)}(0) = i^k \mathbb{E}[X^k], \quad k = 1, 2, \dots, n. \quad (2.4.12)$$

Khususnya, jika  $\mathbb{E}[X^2] < \infty$ , maka ekspansi Taylor di sekitar  $t = 0$  memberikan

$$\phi_X(t) = 1 + it\mathbb{E}[X] - \frac{t^2}{2}\mathbb{E}[X^2] + o(t^2). \quad (2.4.13)$$

Teorema berikut merupakan hasil fundamental yang menghubungkan konvergensi fungsi karakteristik dengan konvergensi dalam distribusi.

**Teorema 2.4.7 (Teorema kontinuitas Lévy, Dudley 2002)** Misalkan  $\{X_n\}_{n=1}^{\infty}$  adalah barisan variabel acak dengan fungsi karakteristik  $\phi_{X_n}$ , dan misalkan  $X$  adalah variabel acak dengan fungsi karakteristik  $\phi_X$ . Maka berlaku dua pernyataan berikut.

(a) Jika  $X_n \Rightarrow X$ , maka  $\phi_{X_n}(t) \rightarrow \phi_X(t)$  untuk setiap  $t \in \mathbb{R}$ .

(b) Sebaliknya, jika  $\phi_{X_n}(t) \rightarrow \phi(t)$  untuk setiap  $t \in \mathbb{R}$  dan  $\phi$  kontinu di  $t = 0$ , maka  $\phi$  adalah fungsi karakteristik dari suatu variabel acak  $X$  dan  $X_n \Rightarrow X$ .

**Bukti.**

(a) Asumsikan  $X_n \Rightarrow X$ , artinya untuk setiap fungsi  $f : \mathbb{R} \rightarrow \mathbb{R}$  yang kontinu dan terbatas, berlaku

$$\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]. \quad (2.4.14)$$

Untuk setiap  $t \in \mathbb{R}$  yang tetap, perhatikan bahwa

$$\phi_{X_n}(t) = \mathbb{E}[e^{itX_n}] = \mathbb{E}[\cos(tX_n)] + i\mathbb{E}[\sin(tX_n)]. \quad (2.4.15)$$

Fungsi  $f(x) = \cos(tx)$  adalah kontinu dan terbatas (karena  $|\cos(tx)| \leq 1$ ).

Dengan demikian, konvergensi dalam distribusi menjamin

$$\mathbb{E}[\cos(tX_n)] \rightarrow \mathbb{E}[\cos(tX)]. \quad (2.4.16)$$

Serupa halnya,  $g(x) = \sin(tx)$  adalah kontinu dan terbatas, sehingga

$$\mathbb{E}[\sin(tX_n)] \rightarrow \mathbb{E}[\sin(tX)]. \quad (2.4.17)$$

Menggabungkan kedua hasil, diperoleh

$$\phi_{X_n}(t) \rightarrow \phi_X(t) \quad \text{untuk setiap } t \in \mathbb{R}. \quad (2.4.18)$$

- (b) Asumsikan  $\phi_{X_n}(t) \rightarrow \phi(t)$  untuk setiap  $t \in \mathbb{R}$  dan  $\phi$  kontinu di  $t = 0$ . Akan ditunjukkan bahwa barisan  $\{X_n\}$  adalah *tight*, artinya untuk setiap  $\varepsilon > 0$ , terdapat  $M > 0$  sehingga  $\Pr(|X_n| > M) < \varepsilon$  untuk semua  $n$ .

Perhatikan bahwa untuk setiap  $x \in \mathbb{R}$ ,

$$1 - \cos(x) = 2 \sin^2(x/2) \leq 2(x/2)^2 = \frac{x^2}{2} \quad \text{untuk } |x| \leq \pi. \quad (2.4.19)$$

Dengan demikian, untuk setiap  $\delta > 0$ ,

$$\begin{aligned} \mathbb{E}[1 - \cos(\delta X_n)] &= 1 - \mathbb{E}[\cos(\delta X_n)] = 1 - \operatorname{Re}[\phi_{X_n}(\delta)] \\ &\rightarrow 1 - \operatorname{Re}[\phi(\delta)]. \end{aligned} \quad (2.4.20)$$

Karena  $\phi$  kontinu di  $t = 0$  dengan  $\phi(0) = 1$ , untuk setiap  $\varepsilon > 0$ , terdapat  $\delta > 0$  sehingga  $|1 - \operatorname{Re}[\phi(\delta)]| < \varepsilon$ . Oleh karena itu, untuk  $n$  yang cukup besar,

$$\mathbb{E}[1 - \cos(\delta X_n)] < 2\varepsilon. \quad (2.4.21)$$

Menggunakan pertidaksamaan  $1 - \cos(\delta x) \geq c \mathbf{1}_{\{|x| > M\}}$  untuk suatu konstanta  $c > 0$  dan  $M$  yang dipilih cukup besar, diperoleh bahwa barisan  $\{X_n\}$

adalah tight.

Dengan teorema Helly, setiap subbarisan dari  $\{X_n\}$  memiliki subsubbarisan yang konvergen dalam distribusi ke suatu variabel acak  $X$ . Oleh karena itu,  $X_n \Rightarrow X$ . Lebih lanjut, fungsi karakteristik dari  $X$  haruslah  $\phi$  dari keunikan fungsi karakteristik.

■

**Contoh 2.4.8** Teorema Lévy dapat digunakan untuk membuktikan konvergensi distribusi. Misalkan  $X_n \sim \mathcal{N}(0, 1 + 1/n)$  untuk  $n = 1, 2, \dots$ . Fungsi karakteristik dari  $X_n$  adalah

$$\phi_{X_n}(t) = \exp\left(-\frac{(1 + 1/n)t^2}{2}\right). \quad (2.4.22)$$

Untuk setiap  $t \in \mathbb{R}$ , berlaku

$$\lim_{n \rightarrow \infty} \phi_{X_n}(t) = \exp\left(-\frac{t^2}{2}\right) = \phi_X(t), \quad (2.4.23)$$

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*), Dudley 2002)** Misalkan  $\{X_n\}$  adalah barisan variabel acak dan  $g : \mathbb{R}^k \rightarrow \mathbb{R}^m$  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  $\mathbb{E}[h(X_n)] \rightarrow \mathbb{E}[h(X)]$  untuk setiap fungsi kontinu terbatas  $h$ .

Ambil sembarang fungsi kontinu terbatas  $h : \mathbb{R}^m \rightarrow \mathbb{R}$ . Definisikan  $f = h \circ g : \mathbb{R}^k \rightarrow \mathbb{R}$ . 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

$$\mathbb{E}[h(g(X_n))] = \mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)] = \mathbb{E}[h(g(X))]. \quad (2.4.24)$$

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,  $\Pr(|g(X_n) - g(X)| > \varepsilon') \rightarrow 0$ .

Bagian (c) mengikuti langsung dari sifat fungsi kontinu: jika  $X_n(\omega) \rightarrow X(\omega)$  untuk hampir semua  $\omega$ , dan  $g$  kontinu di  $X(\omega)$ , maka  $g(X_n(\omega)) \rightarrow g(X(\omega))$ .

■

**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

$$X_n^2 \xrightarrow{d} X^2 \sim \chi^2(1). \quad (2.4.25)$$

Demikian pula, jika  $g(x) = e^x$ , maka  $e^{X_n} \xrightarrow{d} e^X$ .

**Teorema 2.4.11 (Teorema Slutsky, van der Vaart 1998)** 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

$$\phi_{(X_n, Y_n)}(t_1, t_2) = \mathbb{E}[e^{i(t_1 X_n + t_2 Y_n)}]. \quad (2.4.26)$$

Perhatikan bahwa

$$\begin{aligned} |\phi_{(X_n, Y_n)}(t_1, t_2) - \phi_X(t_1)e^{it_2 c}| &= |\mathbb{E}[e^{it_1 X_n}(e^{it_2 Y_n} - e^{it_2 c})]| \\ &\leq \mathbb{E}[|e^{it_2 Y_n} - e^{it_2 c}|]. \end{aligned} \quad (2.4.27)$$

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,  $\mathbb{E}[|e^{it_2 Y_n} - e^{it_2 c}|] \rightarrow 0$ .

Selanjutnya, sebab  $X_n \xrightarrow{d} X$ , berdasarkan teorema kontinuitas Lévy, dida-



patkan bahwa  $\phi_{X_n}(t_1) \rightarrow \phi_X(t_1)$ . Dengan demikian,

$$\phi_{(X_n, Y_n)}(t_1, t_2) \rightarrow \phi_X(t_1)e^{it_2c} = \phi_{(X, c)}(t_1, t_2). \quad (2.4.28)$$

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  $\Pr(Y_n = 0) \rightarrow 0$ , 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 rata-rata  $\mu$  dan variansi  $\sigma^2$ , dan  $S_n^2$  adalah variansi sampel. Berdasarkan Teorema Limit Pusat:

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2). \quad (2.4.29)$$

Berdasarkan Hukum Bilangan Besar,  $S_n^2 \xrightarrow{p} \sigma^2$ , sehingga  $S_n \xrightarrow{p} \sigma$ .

Dengan teorema Slutsky:

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \cdot \frac{\sigma}{S_n} \xrightarrow{d} \mathcal{N}(0, 1) \cdot 1 = \mathcal{N}(0, 1). \quad (2.4.30)$$

Hasil ini membenarkan penggunaan statistik  $t$  dalam inferensi asimtotik.

**Teorema 2.4.13** Misalkan  $\{\mathbf{X}_n\}$  adalah barisan vektor acak berdimensi  $p$  dan  $\{\mathbf{A}_n\}$  adalah barisan matriks acak berukuran  $q \times p$ . Jika  $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$  dan  $\mathbf{A}_n \xrightarrow{p} \mathbf{A}$  dengan  $\mathbf{A}$  adalah matriks konstanta, maka:

$$(a) \mathbf{A}_n \mathbf{X}_n \xrightarrow{d} \mathbf{A} \mathbf{X},$$

$$(b) \mathbf{X}_n + \mathbf{b}_n \xrightarrow{d} \mathbf{X} + \mathbf{b} \text{ jika } \mathbf{b}_n \xrightarrow{p} \mathbf{b},$$

$$(c) \mathbf{X}_n^\top \mathbf{A}_n \mathbf{X}_n \xrightarrow{d} \mathbf{X}^\top \mathbf{A} \mathbf{X} \text{ untuk bentuk kuadrat, dan}$$

(d) jika  $\mathbf{A}$  adalah matriks invertibel, maka  $\mathbf{A}_n^{-1} \xrightarrow{p} \mathbf{A}^{-1}$  dan  $\mathbf{A}_n^{-1} \mathbf{X}_n \xrightarrow{d} \mathbf{A}^{-1} \mathbf{X}$ .

**Bukti. Bagian (a):** Perkalian matriks  $g(\mathbf{A}, \mathbf{x}) = \mathbf{A}\mathbf{x}$  adalah fungsi kontinu dari elemen-elemen  $\mathbf{A}$  dan  $\mathbf{x}$ . Akan ditunjukkan bahwa  $(\mathbf{A}_n, \mathbf{X}_n) \xrightarrow{d} (\mathbf{A}, \mathbf{X})$ . Untuk setiap fungsi kontinu terbatas  $h : \mathbb{R}^{qp+p} \rightarrow \mathbb{R}$ , perlu ditunjukkan bahwa  $\mathbb{E}[h(\mathbf{A}_n, \mathbf{X}_n)] \rightarrow \mathbb{E}[h(\mathbf{A}, \mathbf{X})]$ .

Sebab  $\mathbf{A}_n \xrightarrow{p} \mathbf{A}$ , untuk setiap  $\varepsilon > 0$  berlaku  $\Pr(\|\mathbf{A}_n - \mathbf{A}\| > \varepsilon) \rightarrow 0$ .

Definisikan

$$\mathbb{E}[h(\mathbf{A}_n, \mathbf{X}_n)] = \mathbb{E}[h(\mathbf{A}_n, \mathbf{X}_n) \mathbf{1}_{\{\|\mathbf{A}_n - \mathbf{A}\| \leq \varepsilon\}}] + \mathbb{E}[h(\mathbf{A}_n, \mathbf{X}_n) \mathbf{1}_{\{\|\mathbf{A}_n - \mathbf{A}\| > \varepsilon\}}]. \quad (2.4.31)$$

Suku kedua dibatasi oleh  $\|h\|_\infty \cdot \Pr(\|\mathbf{A}_n - \mathbf{A}\| > \varepsilon) \rightarrow 0$ . Untuk suku pertama, pada kejadian  $\{\|\mathbf{A}_n - \mathbf{A}\| \leq \varepsilon\}$ , fungsi  $h(\mathbf{A}_n, \mathbf{X}_n)$  mendekati  $h(\mathbf{A}, \mathbf{X}_n)$  untuk  $\varepsilon$  kecil. Sebab  $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ , maka  $\mathbb{E}[h(\mathbf{A}, \mathbf{X}_n)] \rightarrow \mathbb{E}[h(\mathbf{A}, \mathbf{X})]$ . Dengan demikian,  $(\mathbf{A}_n, \mathbf{X}_n) \xrightarrow{d} (\mathbf{A}, \mathbf{X})$ . Dengan teorema pemetaan kontinu dan fungsi  $g(\mathbf{A}, \mathbf{x}) = \mathbf{A}\mathbf{x}$ , diperoleh  $\mathbf{A}_n \mathbf{X}_n \xrightarrow{d} \mathbf{A}\mathbf{X}$ .

**Bagian (b):** Analog dengan bagian (a), penjumlahan vektor  $g(\mathbf{x}, \mathbf{b}) = \mathbf{x} + \mathbf{b}$  adalah fungsi kontinu.

**Bagian (c):** Bentuk kuadratik  $g(\mathbf{A}, \mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x}$  adalah fungsi kontinu dari elemen-elemen  $\mathbf{A}$  dan  $\mathbf{x}$ . Dengan hasil dari bagian (a),  $(\mathbf{A}_n, \mathbf{X}_n) \xrightarrow{d} (\mathbf{A}, \mathbf{X})$ , sehingga teorema pemetaan kontinu memberikan  $\mathbf{X}_n^\top \mathbf{A}_n \mathbf{X}_n \xrightarrow{d} \mathbf{X}^\top \mathbf{A} \mathbf{X}$ .

**Bagian (d):** Jika  $\mathbf{A}$  invertibel, fungsi invers matriks  $g(\mathbf{A}) = \mathbf{A}^{-1}$  kontinu pada himpunan matriks invertibel. Dengan demikian, berdasarkan teorema pemetaan kontinu,  $\mathbf{A}_n^{-1} \xrightarrow{p} \mathbf{A}^{-1}$ . Selanjutnya, dengan bagian (a), diperoleh  $\mathbf{A}_n^{-1} \xrightarrow{d} \mathbf{A}^{-1} \mathbf{X}$ .

■

**Contoh 2.4.14** Misalkan  $\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma)$  dan  $\hat{\Sigma}_n \xrightarrow{p} \Sigma$  dengan  $\Sigma$  definit positif. Diketahui statistik Wald

$$W_n = n(\hat{\beta}_n - \beta)^\top \hat{\Sigma}_n^{-1} (\hat{\beta}_n - \beta) \quad (2.4.32)$$

memenuhi  $W_n \xrightarrow{d} \chi^2(p)$ , dengan  $p = \dim(\beta)$ . Hal ini karena  $\hat{\Sigma}_n^{-1} \xrightarrow{p} \Sigma^{-1}$  (berdasarkan teorema pemetaan kontinu untuk fungsi invers matriks), dan

$$\sqrt{n}(\hat{\beta}_n - \beta)^\top \hat{\Sigma}_n^{-1} \sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathbf{Z}^\top \Sigma^{-1} \mathbf{Z} \sim \chi^2(p), \quad (2.4.33)$$

dengan  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ .

**Teorema 2.4.15 (Teorema Cramér–Wold, van der Vaart 1998)** *Barisan vektor acak  $\mathbf{X}_n \in \mathbb{R}^p$  konvergen dalam distribusi ke  $\mathbf{X}$  jika dan hanya jika untuk setiap vektor  $\mathbf{t} \in \mathbb{R}^p$  berlaku*

$$\mathbf{t}^\top \mathbf{X}_n \xrightarrow{d} \mathbf{t}^\top \mathbf{X}. \quad (2.4.34)$$

**Bukti.** ( $\Rightarrow$ ) Jika  $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ , maka untuk setiap  $\mathbf{t} \in \mathbb{R}^p$ , fungsi  $g(\mathbf{x}) = \mathbf{t}^\top \mathbf{x}$  adalah linear (kontinu). Berdasarkan teorema pemetaan kontinu,  $\mathbf{t}^\top \mathbf{X}_n \xrightarrow{d} \mathbf{t}^\top \mathbf{X}$ .

( $\Leftarrow$ ) Andaikan untuk setiap  $\mathbf{t} \in \mathbb{R}^p$  berlaku  $\mathbf{t}^\top \mathbf{X}_n \xrightarrow{d} \mathbf{t}^\top \mathbf{X}$ . Akan ditunjukkan  $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$  melalui konvergensi fungsi karakteristik. Fungsi karakteristik dari  $\mathbf{X}_n$  adalah

$$\phi_{\mathbf{X}_n}(\mathbf{t}) = \mathbb{E}[e^{i\mathbf{t}^\top \mathbf{X}_n}]. \quad (2.4.35)$$

Perhatikan bahwa  $e^{i\mathbf{t}^\top \mathbf{X}_n}$  adalah fungsi karakteristik dari variabel acak skalar  $\mathbf{t}^\top \mathbf{X}_n$  dievaluasi di titik 1. Sebab  $\mathbf{t}^\top \mathbf{X}_n \xrightarrow{d} \mathbf{t}^\top \mathbf{X}$ , berdasarkan teorema kontinuitas Lévy, fungsi karakteristik dari  $\mathbf{t}^\top \mathbf{X}_n$  konvergen ke fungsi karakteristik dari  $\mathbf{t}^\top \mathbf{X}$  di setiap titik. Khususnya,

$$\phi_{\mathbf{X}_n}(\mathbf{t}) = \mathbb{E}[e^{i\mathbf{t}^\top \mathbf{X}_n}] \rightarrow \mathbb{E}[e^{i\mathbf{t}^\top \mathbf{X}}] = \phi_{\mathbf{X}}(\mathbf{t}). \quad (2.4.36)$$

Sebab  $\phi_{\mathbf{X}_n}(\mathbf{t}) \rightarrow \phi_{\mathbf{X}}(\mathbf{t})$  untuk setiap  $\mathbf{t} \in \mathbb{R}^p$ , dan  $\phi_{\mathbf{X}}$  kontinu di  $\mathbf{t} = \mathbf{0}$  (sebab merupakan fungsi karakteristik), berdasarkan teorema kontinuitas Lévy multivariat,  $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ . ■

**Contoh 2.4.16** Teorema Cramér–Wold sangat berguna untuk membuktikan Teorema Limit Pusat multivariat. Jika  $\mathbf{X}_1, \mathbf{X}_2, \dots$  adalah vektor acak *i.i.d.* dengan rata-rata

$\mu$  dan matriks kovarians  $\Sigma$ , maka untuk setiap  $\mathbf{t} \in \mathbb{R}^p$ :

$$\mathbf{t}^\top (\sqrt{n}(\bar{\mathbf{X}}_n - \mu)) = \sqrt{n}(\mathbf{t}^\top \bar{\mathbf{X}}_n - \mathbf{t}^\top \mu) \xrightarrow{d} \mathcal{N}(0, \mathbf{t}^\top \Sigma \mathbf{t}), \quad (2.4.37)$$

berdasarkan univariat. Dengan teorema Cramér–Wold,  $\sqrt{n}(\bar{\mathbf{X}}_n - \mu) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma)$ .

#### 2.4.5 Notasi *Big-O* dan *Little-o* (Deterministik)

Notasi asimtotik *Big-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-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

$$a_n = \mathcal{O}(b_n), \quad (2.4.38)$$

jika terdapat konstanta  $C > 0$  dan bilangan bulat  $n_0 \in \mathbb{N}$  sehingga

$$|a_n| \leq Cb_n, \quad \forall n \geq n_0. \quad (2.4.39)$$

Secara ekuivalen,  $a_n = \mathcal{O}(b_n)$  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-O*.

- (a) **Polinomial:** Jika  $a_n = 3n^2 + 5n + 7$ , maka  $a_n = \mathcal{O}(n^2)$ . Untuk  $n \geq 1$ , berlaku

$$|a_n| = 3n^2 + 5n + 7 \leq 3n^2 + 5n^2 + 7n^2 = 15n^2,$$

sehingga dapat dipilih  $C = 15$  dan  $n_0 = 1$ .

- (b) **Logaritma:** Jika  $a_n = \log n$ , maka  $a_n = \mathcal{O}(n^\varepsilon)$  untuk setiap  $\varepsilon > 0$ . Hal ini karena  $\lim_{n \rightarrow \infty} \frac{\log n}{n^\varepsilon} = 0$ , sehingga  $\log n \leq Cn^\varepsilon$  untuk  $n$  cukup besar.

- (c) **Eksponensial:** Jika  $a_n = 2^n$ , maka  $a_n = \mathcal{O}(3^n)$  karena  $2^n \leq 3^n$  untuk semua

$$n \geq 0.$$

- (d) **Rata-rata sampel:** Jika  $\bar{X}_n$  adalah rata-rata sampel dari variabel acak dengan variansi  $\sigma^2$ , maka galat standar  $\sigma/\sqrt{n} = \mathcal{O}(n^{-1/2})$ .

**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

$$a_n = o(b_n), \quad (2.4.40)$$

jika

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0. \quad (2.4.41)$$

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

$$\lim_{n \rightarrow \infty} \frac{n}{n^2} = \lim_{n \rightarrow \infty} \frac{1}{n} = 0.$$

- (b) **Logaritma vs polinomial:** Jika  $a_n = \log n$ , maka  $a_n = o(n^\varepsilon)$  untuk setiap  $\varepsilon > 0$ . Dengan aturan L'Hôpital:

$$\lim_{n \rightarrow \infty} \frac{\log n}{n^\varepsilon} = \lim_{n \rightarrow \infty} \frac{1/n}{\varepsilon n^{\varepsilon-1}} = \lim_{n \rightarrow \infty} \frac{1}{\varepsilon n^\varepsilon} = 0.$$

- (c) **Eksansi Taylor:** Untuk fungsi  $\sin x$  di sekitar  $x = 0$ :

$$\sin x = x - \frac{x^3}{6} + o(x^3) \quad \text{ketika } x \rightarrow 0.$$

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

$$\lim_{n \rightarrow \infty} \frac{1/n^2}{1/n} = \lim_{n \rightarrow \infty} \frac{1}{n} = 0.$$

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** Untuk barisan  $\{a_n\}$  dan  $\{b_n\}$  dengan  $b_n > 0$ , berlaku:

- (a) Jika  $a_n = o(b_n)$ , maka  $a_n = \mathcal{O}(b_n)$ , tetapi tidak sebaliknya.
- (b)  $a_n = o(b_n)$  jika dan hanya jika  $a_n = \mathcal{O}(b_n)$  dan  $\limsup_{n \rightarrow \infty} |a_n/b_n| = 0$ .

**Bukti.**

- (a) Andaikan  $a_n = o(b_n)$ . Maka  $\lim_{n \rightarrow \infty} a_n/b_n = 0$ . 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  $a_n = \mathcal{O}(b_n)$  dengan  $C = 1$ .

Sebaliknya tidak berlaku:  $a_n = 1$  dan  $b_n = 1$  memberikan  $a_n = \mathcal{O}(b_n)$  (dengan  $C = 1$ ), tetapi  $\lim_{n \rightarrow \infty} a_n/b_n = 1 \neq 0$ , sehingga  $a_n \neq o(b_n)$ .

- (b) Langsung dari definisi:  $a_n = o(b_n)$  ekuivalen dengan  $\lim_{n \rightarrow \infty} |a_n/b_n| = 0$ , yang mengimplikasikan barisan  $\{|a_n/b_n|\}$  terbatas (sehingga  $a_n = \mathcal{O}(b_n)$ ) dan  $\limsup_{n \rightarrow \infty} |a_n/b_n| = 0$ .

■

**Proposisi 2.4.22** Untuk barisan  $\{a_n\}$ ,  $\{b_n\}$ ,  $\{c_n\}$  dengan suku positif yang sesuai, berlaku:

- (a) **Transitivitas:** Jika  $a_n = \mathcal{O}(b_n)$  dan  $b_n = \mathcal{O}(c_n)$ , maka  $a_n = \mathcal{O}(c_n)$ .
- (b) **Penjumlahan:**  $\mathcal{O}(a_n) + \mathcal{O}(b_n) = \mathcal{O}(\max\{a_n, b_n\})$ .
- (c) **Perkalian:**  $\mathcal{O}(a_n) \cdot \mathcal{O}(b_n) = \mathcal{O}(a_n b_n)$ .
- (d) **Penyerapan:**  $o(a_n) + \mathcal{O}(a_n) = \mathcal{O}(a_n)$  dan  $o(a_n) \cdot \mathcal{O}(b_n) = o(a_n b_n)$ .

### 2.4.6 Notasi Big- $\mathcal{O}$ dan Little- $o$ dalam Probabilitas

Notasi  $\mathcal{O}_p$  dan  $o_p$  merupakan perluasan dari notasi deterministik  $\mathcal{O}(\cdot)$  dan  $o(\cdot)$  ke dalam konteks stokastik. Notasi ini sangat penting dalam perumusan representasi asimtotik estimator dan digunakan secara luas dalam statistika asimtotik (van der Vaart, 1998).

**Definisi 2.4.23 (Notasi  $\mathcal{O}_p$ , van der Vaart 1998)** Barisan variabel acak  $\{X_n\}$  dikatakan terbatas dalam probabilitas dengan orde  $a_n$ , ditulis

$$X_n = \mathcal{O}_p(a_n), \quad (2.4.42)$$

apabila untuk setiap  $\varepsilon > 0$  terdapat  $M > 0$  dan  $N \in \mathbb{N}$  sehingga

$$\Pr(|X_n| > Ma_n) < \varepsilon, \quad \forall n \geq N. \quad (2.4.43)$$

Secara ekuivalen,  $X_n = \mathcal{O}_p(a_n)$  jika dan hanya jika barisan  $\{X_n/a_n\}$  terbatas dalam probabilitas, yaitu

$$\sup_{n \geq N} \Pr\left(\left|\frac{X_n}{a_n}\right| > M\right) < \varepsilon. \quad (2.4.44)$$

**Contoh 2.4.24** Misalkan  $X_1, X_2, \dots$  adalah barisan variabel acak *i.i.d.* dengan  $\mathbb{E}[X_i] = \mu$  dan  $\text{Var}(X_i) = \sigma^2 < \infty$ . 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

$$\sup_n \Pr(|\sqrt{n}(\bar{X}_n - \mu)| > M) < \varepsilon,$$

yang ekuivalen dengan

$$\sup_n \Pr(|\bar{X}_n - \mu| > M \cdot n^{-1/2}) < \varepsilon.$$

Oleh karena itu,  $\bar{X}_n - \mu = \mathcal{O}_p(n^{-1/2})$ .

**Contoh 2.4.25** Misalkan  $X_n \sim \mathcal{N}(0, 1/n)$  untuk  $n = 1, 2, \dots$ . 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  $o_p$ , van der Vaart 1998)** Barisan variabel acak  $\{X_n\}$  dikatakan konvergen ke nol dalam probabilitas dengan orde  $a_n$ , ditulis

$$X_n = o_p(a_n), \quad (2.4.45)$$

apabila

$$\frac{X_n}{a_n} \xrightarrow{p} 0, \quad (2.4.46)$$

yaitu untuk setiap  $\varepsilon > 0$  dan  $\delta > 0$ , terdapat  $N \in \mathbb{N}$  sehingga

$$\Pr\left(\left|\frac{X_n}{a_n}\right| > \varepsilon\right) < \delta, \quad \forall n \geq N. \quad (2.4.47)$$

**Contoh 2.4.27** Misalkan  $X_1, X_2, \dots$  adalah barisan variabel acak *i.i.d.* dengan  $\mathbb{E}[X_i] = \mu$  dan  $\text{Var}(X_i) = \sigma^2 < \infty$ . 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  $\bar{X}_n - \mu = o_p(1)$ .

**Teorema 2.4.28** 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)  $\mathcal{O}_p(a_n) + \mathcal{O}_p(b_n) = \mathcal{O}_p(\max\{a_n, b_n\})$ .
- (b)  $\mathcal{O}_p(a_n) \cdot \mathcal{O}_p(b_n) = \mathcal{O}_p(a_n b_n)$ .
- (c)  $o_p(a_n) = \mathcal{O}_p(a_n)$ , tetapi tidak sebaliknya.
- (d)  $o_p(a_n) + \mathcal{O}_p(a_n) = \mathcal{O}_p(a_n)$ .
- (e)  $o_p(a_n) \cdot \mathcal{O}_p(b_n) = o_p(a_n b_n)$ .
- (f)  $X_n \xrightarrow{p} c$  jika dan hanya jika  $X_n = c + o_p(1)$ .
- (g) 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

$$\begin{aligned}\Pr(|X_n + Y_n| > Mc_n) &\leq \Pr(|X_n| + |Y_n| > (M_1 + M_2)c_n) \\ &\leq \Pr(|X_n| > M_1a_n) + \Pr(|Y_n| > M_2b_n) \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon.\end{aligned}\tag{2.4.48}$$

- (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

$$\Pr(|X_n| > M_1a_n) < \varepsilon/2, \quad \Pr(|Y_n| > M_2b_n) < \varepsilon/2.$$

Perhatikan bahwa

$$\begin{aligned}\Pr(|X_n Y_n| > M_1 M_2 a_n b_n) &\leq \Pr(|X_n| > M_1 a_n) + \Pr(|Y_n| > M_2 b_n) \\ &< \varepsilon.\end{aligned}\tag{2.4.49}$$

- (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 2.4.29** Misalkan  $\{X_n\}$  adalah barisan variabel acak. Berlaku pernyataan-pernyataan berikut.

- (a) Jika  $\mathbb{E}[|X_n|] = \mathcal{O}(a_n)$  (dalam arti deterministik), maka  $X_n = \mathcal{O}_p(a_n)$ .
- (b) Jika  $\mathbb{E}[X_n] = \mu_n$  dan  $\text{Var}(X_n) = \mathcal{O}(a_n^2)$ , maka  $X_n - \mu_n = \mathcal{O}_p(a_n)$ .
- (c) 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) Andaikan  $\mathbb{E}[|X_n|] \leq Ca_n$  untuk suatu konstanta  $C > 0$  dan semua  $n$  cukup besar. Dengan ketaksamaan Markov, untuk setiap  $M > 0$ :

$$\Pr(|X_n| > Ma_n) \leq \frac{\mathbb{E}[|X_n|]}{Ma_n} \leq \frac{Ca_n}{Ma_n} = \frac{C}{M}. \quad (2.4.50)$$

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) Andaikan  $\text{Var}(X_n) \leq Ca_n^2$  untuk suatu konstanta  $C > 0$ . Dengan ketaksamaan Chebyshev, untuk setiap  $M > 0$ :

$$\Pr(|X_n - \mu_n| > Ma_n) \leq \frac{\text{Var}(X_n)}{M^2a_n^2} \leq \frac{Ca_n^2}{M^2a_n^2} = \frac{C}{M^2}. \quad (2.4.51)$$

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) Andaikan  $\mathbb{E}[X_n^2] \leq Ca_n^2$  untuk suatu konstanta  $C > 0$ . Dengan ketaksamaan Markov pada  $X_n^2$ :

$$\Pr(|X_n| > Ma_n) = \Pr(X_n^2 > M^2a_n^2) \leq \frac{\mathbb{E}[X_n^2]}{M^2a_n^2} \leq \frac{Ca_n^2}{M^2a_n^2} = \frac{C}{M^2}. \quad (2.4.52)$$

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 2.4.30** Misalkan  $X_1, X_2, \dots$  adalah barisan variabel acak *i.i.d.* dengan rata-rata  $\mathbb{E}[X_i] = \mu$  dan  $\text{Var}(X_i) = \sigma^2 < \infty$ . Definisikan rata-rata sampel  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ .

$$\mathbb{E}[\bar{X}_n] = \mu, \quad (2.4.53)$$

$$\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n} = \mathcal{O}(n^{-1}) = \mathcal{O}((n^{-1/2})^2). \quad (2.4.54)$$

Berdasarkan Teorema Dekomposisi Momen bagian (b), diperoleh

$$\bar{X}_n - \mu = \mathcal{O}_p(n^{-1/2}).$$

**Teorema 2.4.31** Misalkan  $\{X_n\}$  adalah barisan variabel acak. Berlaku pernyataan-pernyataan berikut.

- (a) Jika  $\mathbb{E}[|X_n|] = o(a_n)$  (dalam arti deterministik), maka  $X_n = o_p(a_n)$ .
- (b) Jika  $\mathbb{E}[X_n] = o(a_n)$  dan  $\text{Var}(X_n) = o(a_n^2)$ , maka  $X_n = o_p(a_n)$ .
- (c) 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) Andaikan  $\mathbb{E}[|X_n|]/a_n \rightarrow 0$  ketika  $n \rightarrow \infty$ . Dengan ketaksamaan Markov, untuk setiap  $\varepsilon > 0$ :

$$\Pr\left(\left|\frac{X_n}{a_n}\right| > \varepsilon\right) \leq \frac{\mathbb{E}[|X_n|/a_n]}{\varepsilon} = \frac{\mathbb{E}[|X_n|]}{\varepsilon a_n} \rightarrow 0. \quad (2.4.55)$$

Dengan demikian,  $X_n/a_n \xrightarrow{p} 0$ , sehingga  $X_n = o_p(a_n)$ .

- (b) Andaikan  $\mathbb{E}[X_n] = o(a_n)$  dan  $\text{Var}(X_n) = o(a_n^2)$ . Maka  $\mathbb{E}[X_n]/a_n \rightarrow 0$  dan

$\text{Var}(X_n)/a_n^2 \rightarrow 0$ . Dengan ketaksamaan Chebyshev:

$$\Pr \left( \left| \frac{X_n - \mathbb{E}[X_n]}{a_n} \right| > \varepsilon \right) \leq \frac{\text{Var}(X_n)}{\varepsilon^2 a_n^2} \rightarrow 0. \quad (2.4.56)$$

Sehingga  $(X_n - \mathbb{E}[X_n])/a_n \xrightarrow{p} 0$ . Sebab  $\mathbb{E}[X_n]/a_n \rightarrow 0$ , 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) Andaikan  $\mathbb{E}[X_n^2]/a_n^2 \rightarrow 0$ . Dengan ketaksamaan Markov:

$$\Pr \left( \left| \frac{X_n}{a_n} \right| > \varepsilon \right) = \Pr \left( \frac{X_n^2}{a_n^2} > \varepsilon^2 \right) \leq \frac{\mathbb{E}[X_n^2]}{\varepsilon^2 a_n^2} \rightarrow 0. \quad (2.4.57)$$

Dengan demikian,  $X_n/a_n \xrightarrow{p} 0$ , sehingga  $X_n = o_p(a_n)$ .

■

**Contoh 2.4.32** Misalkan  $X_n \sim \mathcal{N}(0, 1/n^3)$  untuk  $n = 1, 2, \dots$ . Akan ditunjukkan bahwa  $X_n = o_p(n^{-1})$ .

$$\mathbb{E}[X_n^2] = \text{Var}(X_n) = \frac{1}{n^3}. \quad (2.4.58)$$

Perhatikan bahwa

$$\frac{\mathbb{E}[X_n^2]}{(n^{-1})^2} = \frac{1/n^3}{1/n^2} = \frac{1}{n} \rightarrow 0. \quad (2.4.59)$$

Berdasarkan Teorema Dekomposisi  $o_p$  bagian (c), diperoleh  $X_n = o_p(n^{-1})$ .

$$\frac{X_n}{n^{-1}} = nX_n \sim \mathcal{N} \left( 0, \frac{n^2}{n^3} \right) = \mathcal{N} \left( 0, \frac{1}{n} \right). \quad (2.4.60)$$

Untuk setiap  $\varepsilon > 0$ :

$$\Pr(|nX_n| > \varepsilon) = 2 \left( 1 - \Phi \left( \frac{\varepsilon}{\sqrt{1/n}} \right) \right) = 2 (1 - \Phi(\varepsilon\sqrt{n})) \rightarrow 0 \quad (2.4.61)$$

ketika  $n \rightarrow \infty$ . Dengan demikian,  $nX_n \xrightarrow{p} 0$ , sehingga  $X_n = o_p(n^{-1})$ .

#### 2.4.7 Teorema Limit Pusat

**Teorema 2.4.33 (Teorema limit pusat, van der Vaart 1998)** Misalkan  $\{X_i\}_{i=1}^{\infty}$  adalah barisan variabel acak i.i.d. dengan  $\mathbb{E}[X_1] = \mu$  dan  $\text{Var}(X_1) = \sigma^2 < \infty$ , maka

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n X_i - \mu \right) \Rightarrow \mathcal{N}(0, \sigma^2). \quad (2.4.62)$$

**Bukti.** Pembuktian dilakukan dengan menggunakan fungsi karakteristik. Definisikan  $Y_i = X_i - \mu$  sehingga  $\mathbb{E}[Y_i] = 0$  dan  $\text{Var}(Y_i) = \sigma^2$ . Selanjutnya, definisikan

$$S_n = \sum_{i=1}^n Y_i = \sum_{i=1}^n (X_i - \mu). \quad (2.4.63)$$

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

$$\phi_Y(t) = 1 + it\mathbb{E}[Y_i] + \frac{(it)^2}{2}\mathbb{E}[Y_i^2] + o(t^2) = 1 - \frac{t^2\sigma^2}{2} + o(t^2). \quad (2.4.64)$$

Sebab  $Y_1, Y_2, \dots, Y_n$  saling independen dan berdistribusi identik, fungsi karakteristik dari  $S_n/\sqrt{n\sigma^2}$  adalah

$$\begin{aligned} \phi_{S_n/\sqrt{n\sigma^2}}(t) &= \mathbb{E} \left[ \exp \left( \frac{itS_n}{\sqrt{n\sigma^2}} \right) \right] \\ &= \mathbb{E} \left[ \exp \left( \sum_{i=1}^n \frac{itY_i}{\sqrt{n\sigma^2}} \right) \right] \\ &= \prod_{i=1}^n \mathbb{E} \left[ \exp \left( \frac{itY_i}{\sqrt{n\sigma^2}} \right) \right] \\ &= \left[ \phi_Y \left( \frac{t}{\sqrt{n\sigma^2}} \right) \right]^n. \end{aligned} \quad (2.4.65)$$

Dengan substitusi ekspansi Taylor, diperoleh

$$\begin{aligned}\phi_{S_n/\sqrt{n\sigma^2}}(t) &= \left[1 - \frac{1}{2} \cdot \frac{t^2}{n\sigma^2} \cdot \sigma^2 + o\left(\frac{1}{n}\right)\right]^n \\ &= \left[1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right]^n.\end{aligned}\quad (2.4.66)$$

Dengan mengambil limit  $n \rightarrow \infty$  dan menggunakan fakta tentang limit pada fungsi eksponensial bahwa  $\lim_{n \rightarrow \infty} \left(1 + \frac{a}{n}\right)^n = e^a$ , diperoleh

$$\lim_{n \rightarrow \infty} \phi_{S_n/\sqrt{n\sigma^2}}(t) = \exp\left(-\frac{t^2}{2}\right), \quad (2.4.67)$$

yang merupakan fungsi karakteristik dari distribusi  $\mathcal{N}(0, 1)$ .

Berdasarkan teorema kontinuitas Lévy, konvergensi fungsi karakteristik mengimplikasikan konvergensi dalam distribusi. Oleh karena itu,

$$\frac{S_n}{\sqrt{n\sigma^2}} = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \Rightarrow \mathcal{N}(0, 1). \quad (2.4.68)$$

Dengan mengalikan kedua ruas dengan  $\sigma$ , diperoleh

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n X_i - \mu \right) \Rightarrow \mathcal{N}(0, \sigma^2). \quad (2.4.69)$$

■

Teorema Limit Pusat atau *Central Limit Theorem* (CLT) merupakan hasil fundamental yang memungkinkan pendekatan distribusi normal terhadap rata-rata sampel, dan menjadi dasar utama inferensi statistik asimtotik.

**Teorema 2.4.34 (Teorema Lindeberg–Lévy, Billingsley 1995)** Misalkan  $\{X_i\}_{i=1}^\infty$  adalah barisan variabel acak i.i.d. dengan  $\mathbb{E}[X_i] = \mu$  dan  $\text{Var}(X_i) = \sigma^2 < \infty$ . Maka

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (2.4.70)$$

**Bukti.** Teorema Lindeberg–Lévy merupakan kasus khusus dari Teorema Limit Pu-

sat klasik yang telah dibuktikan sebelumnya. Definisikan  $Y_i = X_i - \mu$  sehingga  $\mathbb{E}[Y_i] = 0$  dan  $\text{Var}(Y_i) = \sigma^2$ . Dengan menggunakan fungsi karakteristik, ekspansi Taylor dari  $\phi_Y(t)$  di sekitar  $t = 0$  memberikan

$$\phi_Y(t) = 1 - \frac{t^2 \sigma^2}{2} + o(t^2). \quad (2.4.71)$$

Fungsi karakteristik dari  $\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n Y_i$  adalah

$$\left[ \phi_Y \left( \frac{t}{\sigma\sqrt{n}} \right) \right]^n = \left[ 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \right]^n \rightarrow e^{-t^2/2} \quad (2.4.72)$$

ketika  $n \rightarrow \infty$ . Berdasarkan teorema kontinuitas Lévy, konvergensi ini mengimplikasikan konvergensi dalam distribusi ke  $\mathcal{N}(0, 1)$ . ■

**Teorema 2.4.35 (Teorema Lindeberg–Feller, van der Vaart 1998)** *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*

$$s_n^2 = \sum_{i=1}^n \text{Var}(X_{n,i}) = \sum_{i=1}^n \sigma_{n,i}^2. \quad (2.4.73)$$

*Jika kondisi Lindeberg terpenuhi, yaitu untuk setiap  $\varepsilon > 0$ ,*

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^n \mathbb{E} \left[ X_{n,i}^2 \cdot \mathbf{1}_{\{|X_{n,i}| > \varepsilon s_n\}} \right] = 0, \quad (2.4.74)$$

*maka*

$$\frac{1}{s_n} \sum_{i=1}^n X_{n,i} \xrightarrow{d} \mathcal{N}(0, 1). \quad (2.4.75)$$

**Bukti.** Untuk membuktikan konvergensi dalam distribusi, akan menunjukkan bahwa fungsi karakteristik dari  $\frac{1}{s_n} \sum_{i=1}^n X_{n,i}$  konvergen ke  $e^{-t^2/2}$ . Fungsi karakteristik dari  $\frac{1}{s_n} \sum_{i=1}^n X_{n,i}$  adalah

$$\phi_n(t) = \mathbb{E} \left[ \exp \left( it \frac{1}{s_n} \sum_{j=1}^n X_{n,j} \right) \right]. \quad (2.4.76)$$



Sebab variabel-variabel  $X_{n,j}$  independen,

$$\phi_n(t) = \prod_{j=1}^n \mathbb{E} \left[ \exp \left( it \frac{X_{n,j}}{s_n} \right) \right] = \prod_{j=1}^n \phi_{n,j} \left( \frac{t}{s_n} \right), \quad (2.4.77)$$

dengan  $\phi_{n,j}$  adalah fungsi karakteristik dari  $X_{n,j}$ . Ekspansi Taylor dari  $\phi_{n,j}(u)$  di sekitar  $u = 0$  memberikan

$$\phi_{n,j}(u) = 1 + iu\mathbb{E}[X_{n,j}] - \frac{u^2}{2}\mathbb{E}[X_{n,j}^2] + R_{n,j}(u), \quad (2.4.78)$$

dengan  $\mathbb{E}[X_{n,j}] = 0$ , sehingga

$$\phi_{n,j}(u) = 1 - \frac{u^2 \sigma_{n,j}^2}{2} + R_{n,j}(u), \quad (2.4.79)$$

dengan  $R_{n,j}(u)$  adalah suku sisanya yang memenuhi  $|R_{n,j}(u)| \leq C|u|^3\mathbb{E}[|X_{n,j}|^3]$  untuk suatu konstanta  $C > 0$ .

Untuk setiap  $\varepsilon > 0$ , definisikan

$$X_{n,j}^{(\varepsilon)} := X_{n,j} \mathbf{1}_{\{|X_{n,j}| \leq \varepsilon s_n\}}, \quad (2.4.80)$$

$$X_{n,j}^{(\varepsilon,c)} := X_{n,j} \mathbf{1}_{\{|X_{n,j}| > \varepsilon s_n\}}, \quad (2.4.81)$$

sehingga  $X_{n,j} = X_{n,j}^{(\varepsilon)} + X_{n,j}^{(\varepsilon,c)}$ . Melalui substitusi  $u = t/s_n$  dalam ekspansi Taylor, didapatkan

$$\phi_{n,j} \left( \frac{t}{s_n} \right) = 1 - \frac{t^2 \sigma_{n,j}^2}{2s_n^2} + R_{n,j} \left( \frac{t}{s_n} \right). \quad (2.4.82)$$

Logaritma dari fungsi karakteristik  $\phi_n(t)$  tersebut adalah

$$\begin{aligned} \log \phi_n(t) &= \sum_{j=1}^n \log \phi_{n,j} \left( \frac{t}{s_n} \right) \\ &= \sum_{j=1}^n \log \left( 1 - \frac{t^2 \sigma_{n,j}^2}{2s_n^2} + R_{n,j} \left( \frac{t}{s_n} \right) \right). \end{aligned} \quad (2.4.83)$$

Dengan menggunakan aproksimasi  $\log(1+x) \approx x - \frac{x^2}{2}$  untuk  $|x|$  kecil, didapatkan

$$\begin{aligned}
 \log \phi_n(t) &= \sum_{j=1}^n \left( -\frac{t^2 \sigma_{n,j}^2}{2s_n^2} + R_{n,j} \left( \frac{t}{s_n} \right) \right) + O \left( \sum_{j=1}^n \left| \frac{t^2 \sigma_{n,j}^2}{2s_n^2} \right|^2 \right) \\
 &= -\frac{t^2}{2s_n^2} \sum_{j=1}^n \sigma_{n,j}^2 + \sum_{j=1}^n R_{n,j} \left( \frac{t}{s_n} \right) + O \left( \frac{t^4}{s_n^4} \sum_{j=1}^n \sigma_{n,j}^4 \right) \\
 &= -\frac{t^2}{2} + \sum_{j=1}^n R_{n,j} \left( \frac{t}{s_n} \right) + O \left( \frac{t^4}{s_n^4} \sum_{j=1}^n \sigma_{n,j}^4 \right). \tag{2.4.84}
 \end{aligned}$$

Untuk melengkapi bukti, perlu juga ditunjukkan bahwa

$$\sum_{j=1}^n R_{n,j} \left( \frac{t}{s_n} \right) \rightarrow 0 \quad \text{saat } n \rightarrow \infty. \tag{2.4.85}$$

Suku sisa  $R_{n,j}(u)$  dapat diperkirakan sebagai berikut. Dengan kondisi Lindeberg, untuk setiap  $\varepsilon > 0$ ,

$$\begin{aligned}
 \left| \sum_{j=1}^n R_{n,j} \left( \frac{t}{s_n} \right) \right| &\leq \sum_{j=1}^n C \left| \frac{t}{s_n} \right|^3 \mathbb{E}[|X_{n,j}|^3] \\
 &= C \frac{|t|^3}{s_n^3} \sum_{j=1}^n \mathbb{E}[|X_{n,j}|^3]. \tag{2.4.86}
 \end{aligned}$$

Momen ketiga dapat didekomposisi menggunakan fungsi indikator Lindeberg:

$$\mathbb{E}[|X_{n,j}|^3] = \mathbb{E}[|X_{n,j}|^3 \mathbf{1}_{\{|X_{n,j}| \leq \varepsilon s_n\}}] + \mathbb{E}[|X_{n,j}|^3 \mathbf{1}_{\{|X_{n,j}| > \varepsilon s_n\}}]. \tag{2.4.87}$$

Suku pertama dibatasi oleh  $(\varepsilon s_n)^3$  dikalikan dengan probabilitas, yang terbatas. Suku kedua dapat dikendalikan melalui kondisi Lindeberg, yaitu untuk setiap  $\varepsilon > 0$ ,

$$\sum_{j=1}^n \mathbb{E}[|X_{n,j}|^3 \mathbf{1}_{\{|X_{n,j}| > \varepsilon s_n\}}] \leq \frac{1}{\varepsilon s_n} \sum_{j=1}^n \mathbb{E}[X_{n,j}^2 \mathbf{1}_{\{|X_{n,j}| > \varepsilon s_n\}}] \rightarrow 0 \tag{2.4.88}$$

berdasarkan kondisi Lindeberg.

Asumsi yang masuk akal adalah  $s_n \rightarrow \infty$  dan  $\max_j \sigma_{n,j}/s_n \rightarrow 0$  (sehingga

tidak ada satu variabel yang mendominasi). Dengan ini,

$$\frac{|t|^3}{s_n^3} \sum_{j=1}^n \mathbb{E}[|X_{n,j}|^3] \rightarrow 0. \quad (2.4.89)$$

Oleh karena itu,  $\log \phi_n(t) \rightarrow -\frac{t^2}{2}$ , sehingga  $\phi_n(t) \rightarrow e^{-t^2/2}$ . Berdasarkan teorema kontinuitas Lévy,

$$\frac{1}{s_n} \sum_{i=1}^n X_{n,i} \xrightarrow{d} \mathcal{N}(0, 1). \quad (2.4.90)$$

■

**Contoh 2.4.36** Misalkan  $X_{n,i} = Y_i/\sqrt{n}$  untuk  $i = 1, \dots, n$ , dengan  $Y_i$  adalah variabel acak *i.i.d.* dengan  $\mathbb{E}[Y_i] = 0$  dan  $\text{Var}(Y_i) = \sigma^2 < \infty$ . Maka  $\mathbb{E}[X_{n,i}] = 0$  dan  $\text{Var}(X_{n,i}) = \sigma^2/n$ , sehingga  $s_n^2 = \sigma^2$ .

Kondisi Lindeberg dapat diperiksa sebagai berikut. Untuk setiap  $\varepsilon > 0$ ,

$$\begin{aligned} \frac{1}{s_n^2} \sum_{i=1}^n \mathbb{E}[X_{n,i}^2 \cdot \mathbf{1}_{\{|X_{n,i}| > \varepsilon s_n\}}] &= \frac{1}{\sigma^2} \sum_{i=1}^n \mathbb{E}\left[\frac{Y_i^2}{n} \cdot \mathbf{1}_{\{|Y_i| > \varepsilon \sigma \sqrt{n}\}}\right] \\ &= \frac{1}{\sigma^2} \mathbb{E}[Y_1^2 \cdot \mathbf{1}_{\{|Y_1| > \varepsilon \sigma \sqrt{n}\}}]. \end{aligned} \quad (2.4.91)$$

Sebab  $\mathbb{E}[Y_1^2] = \sigma^2 < \infty$ , maka  $\mathbb{E}[Y_1^2 \cdot \mathbf{1}_{\{|Y_1| > \varepsilon \sigma \sqrt{n}\}}] \rightarrow 0$  ketika  $n \rightarrow \infty$  berdasarkan teorema konvergensi terdominasi. Dengan demikian, kondisi Lindeberg terpenuhi dan Teorema Lindeberg–Feller berlaku.

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 2.4.37** Misalkan  $\{X_{n,i} : i = 1, \dots, n\}$  adalah *triangular array* variabel

acak dengan  $\mathbb{E}[X_{n,i}] = 0$  dan varians total

$$\sum_{i=1}^n \text{Var}(X_{n,i}) \rightarrow \sigma^2. \quad (2.4.92)$$

Jika kondisi Lindeberg terpenuhi, maka

$$\sum_{i=1}^n X_{n,i} \Rightarrow \mathcal{N}(0, \sigma^2). \quad (2.4.93)$$

**Bukti.** Definisikan  $s_n^2 = \sum_{i=1}^n \text{Var}(X_{n,i}) \rightarrow \sigma^2$  dan  $Y_{n,i} = X_{n,i}/s_n$  sehingga  $\sum_{i=1}^n \text{Var}(Y_{n,i}) = 1$ . Akan ditunjukkan bahwa  $\sum_{i=1}^n Y_{n,i} \Rightarrow \mathcal{N}(0, 1)$ .

Dengan teorema Lindeberg–Feller, jika kondisi Lindeberg terpenuhi untuk  $\{Y_{n,i}\}$ , maka  $\sum_{i=1}^n Y_{n,i} \Rightarrow \mathcal{N}(0, 1)$ . Kondisi Lindeberg untuk  $\{X_{n,i}\}$  menyatakan bahwa untuk setiap  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^n \mathbb{E}[X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}| > \varepsilon s_n\}}] = 0. \quad (2.4.94)$$

Untuk  $Y_{n,i} = X_{n,i}/s_n$ , kondisi Lindeberg menjadi: untuk setiap  $\varepsilon > 0$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{E}[Y_{n,i}^2 \mathbf{1}_{\{|Y_{n,i}| > \varepsilon\}}] &= \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{E}\left[\frac{X_{n,i}^2}{s_n^2} \mathbf{1}_{\left\{\frac{|X_{n,i}|}{s_n} > \varepsilon\right\}}\right] \\ &= \lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^n \mathbb{E}[X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}| > \varepsilon s_n\}}] = 0. \end{aligned} \quad (2.4.95)$$

Dengan demikian, kondisi Lindeberg untuk  $\{X_{n,i}\}$  juga terpenuhi untuk  $\{Y_{n,i}\}$ .

Berdasarkan teorema Lindeberg–Feller,

$$\sum_{i=1}^n Y_{n,i} \Rightarrow \mathcal{N}(0, 1). \quad (2.4.96)$$

Dengan transformasi balik,  $\sum_{i=1}^n Y_{n,i} = \frac{1}{s_n} \sum_{i=1}^n X_{n,i}$ , sehingga

$$\frac{1}{s_n} \sum_{i=1}^n X_{n,i} \Rightarrow \mathcal{N}(0, 1). \quad (2.4.97)$$

Mengalikan kedua ruas dengan  $s_n$ , dan menggunakan fakta bahwa  $s_n^2 \rightarrow \sigma^2$ , diperoleh

$$\sum_{i=1}^n X_{n,i} \Rightarrow \mathcal{N}(0, \sigma^2), \quad (2.4.98)$$

karena jika  $Z_n \Rightarrow \mathcal{N}(0, 1)$  dan  $a_n \rightarrow a > 0$ , maka  $a_n Z_n \Rightarrow \mathcal{N}(0, a^2)$ . ■

Hasil ini sangat relevan untuk analisis asimtotik estimator lokal, apabila kontribusi masing-masing observasi bergantung pada  $n$ .

### 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  $\theta$  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 2.4.38 (Metode  $\Delta$ , van der Vaart 1998)** *Misalkan*

$$\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow \mathcal{N}(0, \Sigma), \quad (2.4.99)$$

*dan  $g$  adalah fungsi terdiferensialkan di  $\theta$ , maka*

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \Rightarrow \mathcal{N}(0, g'(\theta)\Sigma g'(\theta)^\top). \quad (2.4.100)$$

**Bukti.** Dengan menggunakan ekspansi Taylor orde pertama dari  $g$  di sekitar  $\theta$ , diperoleh

$$g(\hat{\theta}_n) = g(\theta) + g'(\theta)(\hat{\theta}_n - \theta) + o(|\hat{\theta}_n - \theta|). \quad (2.4.101)$$

Sebab  $\hat{\theta}_n \xrightarrow{p} \theta$ , suku sisa  $o(|\hat{\theta}_n - \theta|)$  memenuhi

$$\frac{o(|\hat{\theta}_n - \theta|)}{|\hat{\theta}_n - \theta|} \xrightarrow{p} 0. \quad (2.4.102)$$

Dengan mengalikan kedua ruas ekspansi Taylor dengan  $\sqrt{n}$ , diperoleh

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) = g'(\theta) \cdot \sqrt{n}(\hat{\theta}_n - \theta) + \sqrt{n} \cdot o(|\hat{\theta}_n - \theta|). \quad (2.4.103)$$

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,

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) = g'(\theta) \cdot \sqrt{n}(\hat{\theta}_n - \theta) + o_p(1). \quad (2.4.104)$$

Berdasarkan teorema Slutsky, sebab  $\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow \mathcal{N}(0, \Sigma)$  dan  $o_p(1) \xrightarrow{p} 0$ , maka

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \Rightarrow g'(\theta) \cdot \mathcal{N}(0, \Sigma) = \mathcal{N}(0, g'(\theta)\Sigma g'(\theta)^\top). \quad (2.4.105)$$

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 2.4.39** Misalkan  $X_1, X_2, \dots, X_n$  adalah sampel acak *i.i.d.* dari distribusi dengan rata-rata  $\mu > 0$  dan variansi  $\sigma^2$ . 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

$$\sqrt{n}(\log(\bar{X}_n) - \log(\mu)) \Rightarrow \mathcal{N}\left(0, \frac{\sigma^2}{\mu^2}\right). \quad (2.4.106)$$

Dengan demikian,  $\log(\bar{X}_n)$  adalah asimtotik normal dengan rata-rata  $\log(\mu)$  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  $X$ ). (Montgomery, Peck, dan Vining, 2012)

### 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, Montgomery dkk. 2012)** *Model regresi linear sederhana merupakan model dengan satu variabel penjelas  $X$  yang mempunyai hubungan garis lurus terhadap satu respons  $Y$ . Model ini dirumuskan dengan*

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x} + \boldsymbol{\varepsilon}, \quad (2.5.1)$$

*dengan intersep  $\beta_0$  dan koefisien  $\beta_1$  merupakan konstanta yang tidak diketahui, serta  $\boldsymbol{\varepsilon}$  adalah vektor galat acak.  $\mathbf{y}$  dan  $\mathbf{x}$  berturut-turut adalah vektor acak  $Y$  dan  $X$ . Parameter  $\beta_0$  dan  $\beta_1$  biasa disebut sebagai koefisien-koefisien regresi.*

**Definisi 2.5.2 (Regresi Linear Berganda, Montgomery dkk. 2012)** *Regresi linear berganda merupakan model regresi yang memiliki lebih dari satu variabel independen. Model ini merupakan generalisasi dari regresi linear sederhana. Vektor respons  $\mathbf{y}$  mungkin memiliki hubungan dengan  $p$ -buah variabel penjelas, yaitu matriks acak  $\mathbf{X}$  yang dibentuk atas  $p$ -buah vektor acak  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ , sehingga*

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \dots + \beta_p \mathbf{x}_p + \boldsymbol{\varepsilon}. \quad (2.5.2)$$

Model di atas dapat dituliskan dalam notasi matriks dan vektor, yaitu

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \quad (2.5.3)$$

atau

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (2.5.4)$$

Model OLS mengasumsikan bahwa galat memiliki rata-rata nol dan variansi konstan  $\sigma^2$  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  $\varepsilon \stackrel{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma^2)$ , untuk suatu  $\sigma^2$  yang bersifat konstan.

Model regresi linear memandang vektor acak  $\mathbf{y}$  sebagai fungsi dari matriks acak  $\mathbf{X}$ . Oleh karena itu, rata-rata dari distribusi  $\mathbf{y}$  dapat dituliskan sebagai

$$\mathbb{E}[\mathbf{y} \mid \mathbf{X}] = \mathbb{E}[\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}] = \mathbf{X}\boldsymbol{\beta} \quad (2.5.5)$$

dan variansinya

$$\text{Var}(\mathbf{y} \mid \mathbf{X}) = \text{Var}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n, \quad (2.5.6)$$

karena  $\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$  dan  $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n$ . Oleh karena itu, rata-rata dari  $\mathbf{y}$  adalah fungsi linear dari  $\mathbf{X}$  meskipun variansinya tidak bergantung pada nilai  $\mathbf{X}$ .

Model regresi linear klasik didasarkan pada sejumlah asumsi yang dikenal sebagai asumsi Gauss–Markov. Asumsi-asumsi ini diperlukan agar penduga OLS memiliki sifat optimal sebagai *best linear unbiased estimator* (BLUE). Asumsi-asumsi tersebut adalah sebagai berikut (Wooldridge, 2020).

**Asumsi OLS.1 (Linear dalam parameter)** Diberikan sampel data  $\{(\mathbf{y}_i, \mathbf{x}_i) : i =$



$1, \dots, n\}$ , model populasi regresi linear dinyatakan sebagai

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.5.7)$$

dengan  $\mathbf{y}$  adalah vektor respons,  $\mathbf{X}$  adalah matriks desain yang terdiri dari variabel penjelas,  $\boldsymbol{\beta}$  adalah vektor parameter yang tidak diketahui, dan  $\boldsymbol{\varepsilon}$  adalah vektor galat acak.

**Asumsi OLS.2 (Pengambilan sampel acak dan independen)** Sampel data  $\{(\mathbf{y}_i, \mathbf{x}_i) : i = 1, \dots, n\}$  diambil secara acak dan independen dari populasi.

**Asumsi OLS.3 (Rank penuh dari matriks desain)** Matriks desain  $\mathbf{X}$  memiliki rank penuh, yaitu  $\text{rank}(\mathbf{X}) = p$ , sehingga matriks  $\mathbf{X}^\top \mathbf{X}$  nonsingular. Dengan kata lain, tidak ada kolinearitas sempurna di antara variabel penjelas.

**Asumsi OLS.4 (Eksogenitas)** Nilai tengah galat bersyarat pada  $\mathbf{X}$  adalah nol, yaitu

$$\mathbb{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \mathbf{0}. \quad (2.5.8)$$

Asumsi eksogenitas memberikan akibat sebagai berikut.

**Akibat 2.5.3** Di bawah asumsi OLS.1 hingga OLS.4, didapatkan

$$\mathbb{E}[\mathbf{X}^\top \boldsymbol{\varepsilon}] = \mathbf{0}. \quad (2.5.9)$$

**Bukti.** Dengan menggunakan hukum iterasi ekspektasi, diperoleh

$$\mathbb{E}[\mathbf{X}^\top \boldsymbol{\varepsilon}] = \mathbb{E}[\mathbb{E}[\mathbf{X}^\top \boldsymbol{\varepsilon} \mid \mathbf{X}]] = \mathbb{E}[\mathbf{X}^\top \mathbb{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}]] = \mathbb{E}[\mathbf{X}^\top \mathbf{0}] = \mathbf{0}. \quad (2.5.10)$$

■

**Asumsi OLS.5 (Homoskedastisitas)** Ragam galat bersyarat pada  $\mathbf{X}$  adalah konstan, yaitu

$$\text{Var}(\boldsymbol{\varepsilon} \mid \mathbf{X}) = \sigma^2 \mathbf{I}_n, \quad (2.5.11)$$

dengan  $\sigma^2$  adalah konstanta positif.

Asumsi OLS.1 hingga OLS.5 dikenal sebagai asumsi Gauss–Markov.

**Asumsi OLS.6 (Normalitas galat)** Galat populasi  $\varepsilon$  bersifat independen dan terdistribusi normal, yaitu

$$\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n). \quad (2.5.12)$$

Asumsi OLS.6 tidak termasuk dalam asumsi Gauss–Markov, tetapi sering digunakan untuk tujuan inferensi statistik, seperti pengujian hipotesis dan pembuatan interval kepercayaan. Asumsi OLS.1 hingga OLS.6 membentuk dasar dari model regresi linear klasik atau *classical linear model* (CLM).

Estimasi parameter dalam regresi linear dilakukan dengan metode kuadrat terkecil atau *ordinary least squares* (OLS). Prinsip dasarnya adalah mencari nilai  $\beta$  yang meminimumkan jumlah kuadrat galat

$$S(\beta) = (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta). \quad (2.5.13)$$

Perhatikan bahwa fungsi kuadrat terkecil dari persamaan di atas dapat dituliskan sebagai

$$S(\beta) = (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta) = \mathbf{y}^\top \mathbf{y} - 2\beta^\top \mathbf{X}^\top \mathbf{y} + \beta^\top \mathbf{X}^\top \mathbf{X}\beta. \quad (2.5.14)$$

Dengan diferensial matriks, gradien terhadap  $\beta$  adalah

$$\nabla_\beta S(\beta) = -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X}\beta. \quad (2.5.15)$$

Syarat orde pertama meminimumkan  $S$  adalah  $\nabla_\beta S(\beta) = \mathbf{0}$ , sehingga diperoleh

$$\mathbf{X}^\top \mathbf{X}\hat{\beta} = \mathbf{X}^\top \mathbf{y}, \quad (2.5.16)$$

yang disebut persamaan normal. Selanjutnya, matriks Hessian atau orde kedua ada-

lah

$$\nabla_{\beta}^2 S(\beta) = 2 \mathbf{X}^\top \mathbf{X} \succeq \mathbf{0}, \quad (2.5.17)$$

sehingga  $S$  konveks dalam  $\beta$ . Jika  $\mathbf{X}^\top \mathbf{X}$  non-singular atau setara  $\text{rank}(\mathbf{X}) = p$ , maka Hessian definit positif dan solusi orde pertama unik serta memberikan minimum global.

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (2.5.18)$$

Jika  $\mathbf{X}^\top \mathbf{X}$  singular, himpunan minimum tidak tunggal. Salah satu solusi adalah solusi *norma-minimum*  $\hat{\beta} = \mathbf{X}^+ \mathbf{y}$  dengan  $\mathbf{X}^+$  merupakan *Moore–Penrose pseudo-inverse*.

**Teorema 2.5.4 (Teorema Gauss–Markov, Wooldridge 2020)** *Di bawah asumsi OLS.1 hingga OLS.5, penduga OLS  $\hat{\beta}$  bersifat tak bias, yaitu  $\mathbb{E}[\hat{\beta}] = \beta$ , dan memiliki kovariansi*

$$\text{Var}(\hat{\beta}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}. \quad (2.5.19)$$

Selain itu,  $\hat{\beta}$  merupakan best linear unbiased estimator (*BLUE*), artinya  $\hat{\beta}$  memiliki variansi terkecil di antara semua penduga linear tak bias.

**Bukti.** Penduga OLS dapat dituliskan sebagai

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (2.5.20)$$

Substitusi  $\mathbf{y} = \mathbf{X}\beta + \epsilon$  menghasilkan

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\mathbf{X}\beta + \epsilon) = \beta + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \epsilon, \quad (2.5.21)$$

sehingga rata-rata dari  $\hat{\beta}$  adalah

$$\mathbb{E}[\hat{\beta}] = \beta + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbb{E}[\epsilon] = \beta, \quad (2.5.22)$$

yang menunjukkan bahwa  $\hat{\beta}$  tak bias. Kovariansnya diperoleh dari

$$\text{Var}(\hat{\beta}) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \text{Var}(\epsilon) \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}. \quad (2.5.23)$$

Untuk sifat BLUE, dipertimbangkan penduga linear tak bias umum  $\tilde{\beta} = \mathbf{C}\mathbf{y}$  dengan  $\mathbf{C}\mathbf{X} = \mathbf{I}$ . Variansi dari  $\tilde{\beta}$  adalah

$$\text{Var}(\tilde{\beta}) = \mathbf{C}\sigma^2\mathbf{I}\mathbf{C}^\top = \sigma^2\mathbf{C}\mathbf{C}^\top. \quad (2.5.24)$$

Melalui dekomposisi  $\mathbf{C} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top + \mathbf{D}$  dengan  $\mathbf{D}\mathbf{X} = \mathbf{0}$ , diperoleh

$$\text{Var}(\tilde{\beta}) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1} + \sigma^2\mathbf{D}\mathbf{D}^\top \succeq \text{Var}(\hat{\beta}), \quad (2.5.25)$$

yang membuktikan bahwa  $\hat{\beta}$  memiliki variansi minimum di antara semua penduga linear tak bias. ■

**Teorema 2.5.5** *Jika asumsi normalitas galat terpenuhi, maka penduga kemungkinan maksimum atau maximum likelihood estimation (MLE) untuk  $\beta$  identik dengan penduga kuadrat terkecil (OLS), yaitu*

$$\hat{\beta}_{MLE} = \hat{\beta}_{OLS} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}. \quad (2.5.26)$$

*Penduga MLE untuk varians galat adalah*

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\beta})^\top(\mathbf{y} - \mathbf{X}\hat{\beta}). \quad (2.5.27)$$

**Bukti.** Dengan asumsi  $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ , fungsi *likelihood* untuk  $\beta$  dan  $\sigma^2$  adalah

$$L(\beta, \sigma^2 \mid \mathbf{y}) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)^\top(\mathbf{y} - \mathbf{X}\beta)\right), \quad (2.5.28)$$

yang memiliki *log-likelihood*

$$\ell(\beta, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)^\top(\mathbf{y} - \mathbf{X}\beta). \quad (2.5.29)$$

Maksimisasi terhadap  $\beta$  dengan turunan pertama didapatkan

$$\begin{aligned}
 \frac{\partial \ell(\beta, \sigma^2)}{\partial \beta} &= \frac{2\mathbf{X}^\top}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta) \\
 &= \frac{1}{\sigma^2}\mathbf{X}^\top \mathbf{y} - \frac{1}{\sigma^2}\mathbf{X}^\top \mathbf{X}\beta \\
 \mathbf{0} &= \mathbf{X}^\top \mathbf{y} - \mathbf{X}^\top \mathbf{X}\hat{\beta} \\
 \mathbf{X}^\top \mathbf{X}\hat{\beta} &= \mathbf{X}^\top \mathbf{y} \\
 \hat{\beta} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}.
 \end{aligned} \tag{2.5.30}$$

Perhatikan bahwa turunan kedua terhadap  $\beta$  menghasilkan  $-\frac{1}{\sigma^2}(\mathbf{X}^\top \mathbf{X})$ . Sebab  $\mathbf{X}^\top \mathbf{X} \succeq 0$ , maka  $\frac{\partial^2 \ell}{\partial \beta \partial \beta^\top} \preceq 0$  yang menjadikan  $\hat{\beta}$  memaksimumkan fungsi log-likelihood. Oleh karena itu

$$\hat{\beta}_{\text{MLE}} = \hat{\beta}_{\text{OLS}}. \tag{2.5.31}$$

Selanjutnya, turunan pertama  $\ell$  terhadap  $\sigma^2$  akan menghasilkan

$$\begin{aligned}
 \frac{\partial \ell(\beta, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta) \\
 0 &= -n\hat{\sigma}^2 + (\mathbf{y} - \mathbf{X}\hat{\beta})^\top (\mathbf{y} - \mathbf{X}\hat{\beta}) \\
 \hat{\sigma}^2 &= \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\beta})^\top (\mathbf{y} - \mathbf{X}\hat{\beta}).
 \end{aligned} \tag{2.5.32}$$

■

### 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 (Wooldridge, 2020)** *Di bawah asumsi OLS.1 hingga OLS.4, penduga OLS  $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$  bersifat konsisten, yaitu*

$$\hat{\beta} \xrightarrow{p} \beta, \tag{2.5.33}$$

dengan kondisi tambahan berupa  $\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^\top \mathbf{X} = \mathbf{Q}$  untuk suatu matriks  $\mathbf{Q}$  yang definit positif.

**Bukti.** Penduga OLS dapat dituliskan sebagai

$$\begin{aligned} \hat{\beta} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\mathbf{X}\beta + \varepsilon) \\ &= \beta + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \varepsilon \\ &= \beta + \left( \frac{1}{n} \mathbf{X}^\top \mathbf{X} \right)^{-1} \left( \frac{1}{n} \mathbf{X}^\top \varepsilon \right). \end{aligned} \quad (2.5.34)$$

Berdasarkan kondisi tambahan,  $\frac{1}{n} \mathbf{X}^\top \mathbf{X} \xrightarrow{p} \mathbf{Q}$  dengan  $\mathbf{Q}$  definit positif. Dengan teorema pemetaan kontinu,  $\left( \frac{1}{n} \mathbf{X}^\top \mathbf{X} \right)^{-1} \xrightarrow{p} \mathbf{Q}^{-1}$ . Selanjutnya, perlu ditunjukkan bahwa  $\frac{1}{n} \mathbf{X}^\top \varepsilon \xrightarrow{p} \mathbf{0}$ . Perhatikan bahwa

$$\frac{1}{n} \mathbf{X}^\top \varepsilon = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \varepsilon_i. \quad (2.5.35)$$

Sebab  $\mathbb{E}[\mathbf{x}_i \varepsilon_i] = \mathbb{E}[\mathbb{E}[\mathbf{x}_i \varepsilon_i \mid \mathbf{X}]] = \mathbb{E}[\mathbf{x}_i \mathbb{E}[\varepsilon_i \mid \mathbf{X}]] = \mathbf{0}$ , Hukum Bilangan Besar menjamin bahwa  $\frac{1}{n} \mathbf{X}^\top \varepsilon \xrightarrow{p} \mathbf{0}$ .

Dengan teorema Slutsky, diperoleh

$$\hat{\beta} - \beta = \left( \frac{1}{n} \mathbf{X}^\top \mathbf{X} \right)^{-1} \left( \frac{1}{n} \mathbf{X}^\top \varepsilon \right) \xrightarrow{p} \mathbf{Q}^{-1} \cdot \mathbf{0} = \mathbf{0}, \quad (2.5.36)$$

sehingga  $\hat{\beta} \xrightarrow{p} \beta$ . ■

**Teorema 2.5.7** Di bawah kondisi-kondisi Teorema konsistensi OLS dan dengan tambahan asumsi OLS.5, penduga variansi

$$s^2 = \frac{1}{n - p - 1} (\mathbf{y} - \mathbf{X}\hat{\beta})^\top (\mathbf{y} - \mathbf{X}\hat{\beta}) = \frac{1}{n - p - 1} \sum_{i=1}^n \hat{\varepsilon}_i^2 \quad (2.5.37)$$

bersifat konsisten, yaitu  $s^2 \xrightarrow{p} \sigma^2$ .

**Bukti.** Definisikan residual  $\hat{\varepsilon} = \mathbf{y} - \mathbf{X}\hat{\beta} = \mathbf{M}\mathbf{y}$  dengan  $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$

adalah matriks proyeksi idempoten, maka

$$\hat{\epsilon} = \mathbf{M}(\mathbf{X}\beta + \epsilon) = \mathbf{M}\epsilon, \quad (2.5.38)$$

sebab  $\mathbf{MX} = \mathbf{0}$ . Oleh karena itu,

$$\hat{\epsilon}^\top \hat{\epsilon} = \epsilon^\top \mathbf{M}^\top \mathbf{M} \epsilon = \epsilon^\top \mathbf{M} \epsilon. \quad (2.5.39)$$

Dengan menggunakan sifat jejak dan ekspektasi, dapat ditunjukkan bahwa

$$\begin{aligned} \mathbb{E}[\hat{\epsilon}^\top \hat{\epsilon}] &= \mathbb{E}[\text{tr}(\epsilon^\top \mathbf{M} \epsilon)] \\ &= \mathbb{E}[\text{tr}(\mathbf{M} \epsilon \epsilon^\top)] \\ &= \text{tr}(\mathbf{M} \mathbb{E}[\epsilon \epsilon^\top]) \\ &= \sigma^2 \text{tr}(\mathbf{M}) \end{aligned} \quad (2.5.40)$$

$$= \sigma^2(n - p - 1). \quad (2.5.41)$$

Dengan Hukum Bilangan Besar,  $\frac{1}{n} \hat{\epsilon}^\top \hat{\epsilon} \xrightarrow{p} \sigma^2$ . Sebab  $\frac{n}{n-p-1} \rightarrow 1$ , maka  $s^2 \xrightarrow{p} \sigma^2$ . ■

**Teorema 2.5.8** Misalkan kondisi-kondisi konsistensi terpenuhi dan tambahan kondisi Lindeberg–Feller terpenuhi, yaitu

$$\frac{1}{\sqrt{n}} \mathbf{X}^\top \epsilon \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}), \quad (2.5.42)$$

maka penduga OLS memiliki distribusi asimtotik

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}^{-1}). \quad (2.5.43)$$

**Bukti.** Dari pembuktian konsistensi, dapat dituliskan

$$\begin{aligned} \sqrt{n}(\hat{\beta} - \beta) &= \sqrt{n} \left( \frac{1}{n} \mathbf{X}^\top \mathbf{X} \right)^{-1} \left( \frac{1}{n} \mathbf{X}^\top \epsilon \right) \\ &= \left( \frac{1}{n} \mathbf{X}^\top \mathbf{X} \right)^{-1} \left( \frac{1}{\sqrt{n}} \mathbf{X}^\top \epsilon \right). \end{aligned} \quad (2.5.44)$$

Berdasarkan kondisi Lindeberg–Feller,  $\frac{1}{\sqrt{n}}\mathbf{X}^\top \boldsymbol{\varepsilon} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q})$ . Dengan teorema Slutsky, sebab  $\left(\frac{1}{n}\mathbf{X}^\top \mathbf{X}\right)^{-1} \xrightarrow{p} \mathbf{Q}^{-1}$ , maka

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathbf{Q}^{-1} \cdot \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}) = \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}^{-1}). \quad (2.5.45)$$

■

**Akibat 2.5.9** Dengan penduga variansi konsisten  $s^2 \xrightarrow{p} \sigma^2$ , berlaku

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \sigma^2 \left(\frac{1}{n}\mathbf{X}^\top \mathbf{X}\right)^{-1}\right), \quad (2.5.46)$$

dan penduga matriks kovarians asimtotik yang konsisten adalah

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = s^2(\mathbf{X}^\top \mathbf{X})^{-1}. \quad (2.5.47)$$

Di bawah ini akan dipaparkan hasil-hasil mengenai distribusi eksak dan asimtotik dari uji- $t$  dan uji- $F$  untuk model regresi linear.

**Teorema 2.5.10** Untuk menguji hipotesis  $H_0 : \beta_k = \beta_k^0$  lawan  $H_1 : \beta_k \neq \beta_k^0$ , statistik uji- $t$  didefinisikan sebagai

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{\text{se}(\hat{\beta}_k)}, \quad (2.5.48)$$

dengan  $\text{se}(\hat{\beta}_k) = s\sqrt{[(\mathbf{X}^\top \mathbf{X})^{-1}]_{kk}}$  adalah galat standar dari  $\hat{\beta}_k$ .

(a) Distribusi eksak di bawah asumsi normalitas galat  $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$  dan  $H_0$  benar adalah  $t_k \sim t_{n-p-1}$ .

(b) Distribusi asimtotik di bawah kondisi regularitas asimtotik dan  $H_0$  benar adalah  $t_k \xrightarrow{d} \mathcal{N}(0, 1)$ .

**Bukti.**

(a) Di bawah normalitas,  $\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1})$ , sehingga  $\hat{\beta}_k \sim \mathcal{N}(\beta_k, \sigma^2[(\mathbf{X}^\top \mathbf{X})^{-1}]_{kk})$ .



Oleh karena itu,

$$Z_k = \frac{\hat{\beta}_k - \beta_k}{\sigma \sqrt{[(\mathbf{X}^\top \mathbf{X})^{-1}]_{kk}}} \sim \mathcal{N}(0, 1). \quad (2.5.49)$$

Selain itu,  $(n - p - 1)s^2/\sigma^2 \sim \chi_{n-p-1}^2$  dan independen dari  $\hat{\beta}$ . Dengan demikian,

$$t_k = \frac{Z_k}{\sqrt{(n - p - 1)s^2/\sigma^2/(n - p - 1)}} = \frac{Z_k}{\sqrt{s^2/\sigma^2}} \sim t_{n-p-1}. \quad (2.5.50)$$

(b) Dari normalitas asimtotik,  $\sqrt{n}(\hat{\beta}_k - \beta_k) \xrightarrow{d} \mathcal{N}(0, \sigma^2[\mathbf{Q}^{-1}]_{kk})$ . Dengan  $s^2 \xrightarrow{p} \sigma^2$  dan teorema Slutsky,

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{s \sqrt{[(\mathbf{X}^\top \mathbf{X})^{-1}]_{kk}}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (2.5.51)$$

■

Pada praktiknya, perbedaan penggunaan distribusi eksak dan asimtotik uji- $t$  biasanya kecil untuk ukuran sampel yang cukup besar, sehingga uji- $t$  asimtotik sering digunakan sebagai pendekatan yang baik.

**Teorema 2.5.11** Untuk menguji hipotesis linear umum  $H_0 : \mathbf{R}\beta = \mathbf{r}$  dengan  $\mathbf{R}$  matriks berukuran  $q \times (p+1)$  dan  $\text{rank}(\mathbf{R}) = q$ , statistik uji- $F$  didefinisikan sebagai

$$F = \frac{(\mathbf{R}\hat{\beta} - \mathbf{r})^\top [\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r})}{q \cdot s^2}. \quad (2.5.52)$$

(a) Distribusi eksak di bawah asumsi normalitas galat dan  $H_0$  benar adalah

$$F \sim F_{q, n-p-1}.$$

(b) Distribusi asimtotik di bawah kondisi regularitas asimtotik dan  $H_0$  benar adalah  $qF \xrightarrow{d} \chi_q^2$ .

**Bukti.**

(a) Distribusi eksak di bawah  $H_0$  dan normalitas,  $\mathbf{R}\hat{\beta} \sim \mathcal{N}(\mathbf{r}, \sigma^2 \mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top)$ .

Oleh karena itu,

$$\frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^\top [\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})}{\sigma^2} \sim \chi_q^2. \quad (2.5.53)$$

Sebab  $(n - p - 1)s^2/\sigma^2 \sim \chi_{n-p-1}^2$  dan independen dari pembilang, maka

$$F = \frac{\chi_q^2/q}{\chi_{n-p-1}^2/(n-p-1)} \sim F_{q,n-p-1}. \quad (2.5.54)$$

(b) Distribusi asimtotik di bawah kondisi regularitas asimtotik dan  $H_0$  benar adalah

$${}_qF = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^\top [\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})}{s^2} \xrightarrow{d} \chi_q^2. \quad (2.5.55)$$

■

Uji  $F$  eksak dan asimtotik umumnya memberikan hasil yang serupa untuk ukuran sampel yang besar, sehingga uji  $F$  asimtotik sering digunakan sebagai pendekatan yang baik dalam praktik. Uji ini berguna untuk menguji hipotesis yang melibatkan beberapa parameter secara bersamaan, misalnya dengan menggunakan matriks restriksi  $\mathbf{R}$  berupa matriks

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}, \quad (2.5.56)$$

sehingga  $\mathbf{R}\boldsymbol{\beta} = \mathbf{0}$  menguji apakah semua koefisien regresi selain intercept sama dengan nol.

### 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 karena asumsi OLS.5 tidak terpenuhi.

**Definisi 2.5.12** (*Generalized least squares, Greene 2002*) *Pertimbangkan model regresi linear*

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathbb{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \mathbf{0}, \quad (2.5.57)$$

*dengan struktur variansi-kovariansi galat*

$$\text{Var}(\boldsymbol{\varepsilon} \mid \mathbf{X}) = \sigma^2 \boldsymbol{\Omega}, \quad (2.5.58)$$

*dengan  $\boldsymbol{\Omega}$  adalah matriks simetri definit-positif berukuran  $n \times n$ . Model yang memperkenalkan struktur variansi-kovariansi galat disebut sebagai model kuadrat terkecil tergeneralisasi atau generalized least squares (GLS).*

Penduga GLS diperoleh dengan meminimumkan jumlah kuadrat terbobot

$$S_{\mathbf{W}}(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad (2.5.59)$$

dengan  $\mathbf{W} = \boldsymbol{\Omega}^{-1}$ . Diferensiasi terhadap  $\boldsymbol{\beta}$  dan disamakan dengan nol memberikan

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} S_{\mathbf{W}}(\boldsymbol{\beta}) &= -2\mathbf{W}\mathbf{X}^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (\text{karena } \mathbf{W} \text{ simetris}) \\ \mathbf{0} &= -2\mathbf{X}^\top \mathbf{W}\mathbf{y} - 2\mathbf{X}^\top \mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}} \\ \mathbf{X}^\top \mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}^\top \mathbf{W}\mathbf{y}, \end{aligned} \quad (2.5.60)$$

sehingga pada asumsi  $\mathbf{X}^\top \mathbf{W}\mathbf{X}$  nonsingular,

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}^\top \mathbf{W}\mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}\mathbf{y}. \quad (2.5.61)$$

Penduga ini tak bias jika  $\mathbf{W}$  tidak acak atau independen dari  $\boldsymbol{\varepsilon}$  dan  $\mathbb{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \mathbf{0}$ , sebab  $\mathbb{E}[\hat{\boldsymbol{\beta}}_{\text{GLS}} \mid \mathbf{X}] = \boldsymbol{\beta}$ .

Kasus umum dari GLS yang sering dipakai adalah kuadrat terkecil terboboti

atau *weighted least squares* (WLS), yaitu  $\Omega = \text{diag}(\omega_1, \dots, \omega_n)$  yang merupakan heteroskedastisitas diagonal dengan bobot  $\mathbf{W} = \Omega^{-1} = \text{diag}(w_1, \dots, w_n)$ ,  $w_i = 1/\omega_i$ . Bobot  $\mathbf{W}$  dianggap nonstokastik atau *fixed*. Model GLS sesuai apabila galat memiliki korelasi serial, yaitu  $\Omega$  bukan matriks diagonal. Di sisi lain, model WLS sesuai apabila galat tidak berkorelasi tetapi memiliki variansi yang berbeda-beda.

**Teorema 2.5.13 (Teorema Aitken)** *Misalkan model regresi linear*

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathbb{E}[\boldsymbol{\varepsilon} \mid \mathbf{X}] = 0, \quad \text{Var}(\boldsymbol{\varepsilon} \mid \mathbf{X}) = \sigma^2 \Omega,$$

*dengan  $\Omega$  simetris definit-positif, maka penduga*

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Omega^{-1} \mathbf{y} \quad (2.5.62)$$

*adalah best linear unbiased estimator (BLUE) dari  $\boldsymbol{\beta}$ .*

**Bukti.** Pertama-tama perhatikan bahwa setiap penduga linear bagi  $\boldsymbol{\beta}$  dapat ditulis dalam bentuk  $\tilde{\boldsymbol{\beta}} = \mathbf{A}\mathbf{y}$  dengan  $\mathbf{A} \in \mathbb{R}^{p \times n}$ . Agar tak bias, berlaku

$$\mathbb{E}[\tilde{\boldsymbol{\beta}} \mid \mathbf{X}] = \mathbf{A}\mathbb{E}[\mathbf{y} \mid \mathbf{X}] = \mathbf{A}\mathbf{X}\boldsymbol{\beta}. \quad (2.5.63)$$

Supaya sama dengan  $\boldsymbol{\beta}$  untuk semua  $\boldsymbol{\beta}$ , syarat yang harus dipenuhi adalah  $\mathbf{A}\mathbf{X} = \mathbf{I}_p$ . Selanjutnya, kovariansi dari  $\tilde{\boldsymbol{\beta}}$  diberikan oleh

$$\text{Var}(\tilde{\boldsymbol{\beta}} \mid \mathbf{X}) = \mathbf{A} \sigma^2 \Omega \mathbf{A}^\top. \quad (2.5.64)$$

Di antara semua matriks  $\mathbf{A}$  yang memenuhi  $\mathbf{A}\mathbf{X} = \mathbf{I}_p$ , pilih

$$\mathbf{A}_0 = (\mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Omega^{-1}. \quad (2.5.65)$$

Jelas bahwa  $\mathbf{A}_0 \mathbf{X} = \mathbf{I}_p$ , sehingga  $\mathbf{A}_0$  menghasilkan penduga tak bias.

Sekarang, ambil sembarang penduga tak bias lain  $\mathbf{A}$ . Matriks  $\mathbf{A}$  tersebut dapat ditulis sebagai  $\mathbf{A} = \mathbf{A}_0 + \mathbf{D}$  dengan  $\mathbf{D}\mathbf{X} = \mathbf{0}$ , sebab  $\mathbf{A}_0 \mathbf{X} = \mathbf{I}_p$  dan syarat

$\mathbf{AX} = \mathbf{I}_p$  harus tetap dipenuhi. Dengan demikian,

$$\tilde{\boldsymbol{\beta}} = \mathbf{A}\mathbf{y} = \mathbf{A}_0\mathbf{y} + \mathbf{D}\mathbf{y}. \quad (2.5.66)$$

Variansi dari penduga ini adalah

$$\text{Var}(\tilde{\boldsymbol{\beta}} \mid \mathbf{X}) = \text{Var}(\mathbf{A}_0\mathbf{y} \mid \mathbf{X}) + \sigma^2 \mathbf{D}\boldsymbol{\Omega}\mathbf{D}^\top. \quad (2.5.67)$$

Bagian pertama adalah kovarians dari  $\hat{\boldsymbol{\beta}}_{\text{GLS}}$ , sedangkan bagian kedua selalu semi definit positif karena  $\boldsymbol{\Omega}$  definit-positif. Hal ini berarti tambahan  $\mathbf{D}\mathbf{y}$  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,  $\hat{\boldsymbol{\beta}}_{\text{GLS}}$  meminimalkan kovarians di antara semua penduga linear tak bias, sehingga merupakan BLUE. ■

**Teorema 2.5.14** *Jika  $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{\Omega})$ , maka penduga maksimum likelihood (MLE) untuk  $\boldsymbol{\beta}$  adalah*

$$\hat{\boldsymbol{\beta}}_{\text{MLE}} = \hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{y}, \quad (2.5.68)$$

*dan MLE untuk  $\sigma^2$  (dengan  $\boldsymbol{\Omega}$  diketahui) adalah*

$$\hat{\sigma}_{\text{MLE}}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \quad (2.5.69)$$

**Bukti.** Di bawah normalitas galat, fungsi *likelihood* untuk  $\boldsymbol{\beta}$  dan  $\sigma^2$  adalah

$$L(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{y}) = (2\pi\sigma^2)^{-n/2} |\boldsymbol{\Omega}|^{-1/2} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right).$$

Hal ini berarti fungsi log-*likelihood* untuk  $\boldsymbol{\beta}$  dan  $\sigma^2$  adalah

$$\ell(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log |\boldsymbol{\Omega}| - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Maksimisasi  $\ell$  untuk  $\beta$  dengan turunan pertama menghasilkan

$$\begin{aligned}\frac{\partial \ell(\beta, \sigma^2 | \mathbf{y})}{\partial \beta} &= \frac{2}{2\sigma^2} \mathbf{X}^\top \Omega^{-1} (\mathbf{y} - \mathbf{X}\beta) \\ 0 &= \frac{1}{\sigma^2} \mathbf{X}^\top \Omega^{-1} \mathbf{y} - \frac{1}{\sigma^2} \mathbf{X}^\top \Omega^{-1} \mathbf{X} \hat{\beta} \\ \frac{1}{\sigma^2} \mathbf{X}^\top \Omega^{-1} \mathbf{X} \hat{\beta} &= \frac{1}{\sigma^2} \mathbf{X}^\top \Omega^{-1} \mathbf{y} \\ \hat{\beta} &= (\mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Omega^{-1} \mathbf{y},\end{aligned}\tag{2.5.70}$$

sehingga  $\hat{\beta}_{\text{MLE}} = \hat{\beta}_{\text{GLS}}$ . Selanjutnya, turunan terhadap  $\sigma^2$  memberi  $\partial \ell / \partial \sigma^2 = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} (\mathbf{y} - \mathbf{X}\hat{\beta})^\top \Omega^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta}) = 0$ , yang menghasilkan rumus  $\hat{\sigma}_{\text{MLE}}^2$  di atas. ■

Dalam analisis nyata,  $\Omega$  biasanya jarang diketahui. *Feasible GLS* (FGLS) memperkirakan  $\Omega$ , misalnya dari residu OLS atau model varian parametrik, lalu menyusun  $\widehat{\mathbf{W}} = \widehat{\Omega}^{-1}$  dan menghitung  $\hat{\beta}_{\text{FGLS}} = (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{y}$ . Di bawah kondisi reguler,  $\hat{\beta}_{\text{FGLS}}$  konsisten dan asimtotik efisien. Namun, untuk sampel kecil, sifat *finite-sample* bergantung pada kualitas spesifikasi  $\widehat{\Omega}$ .

#### 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 2.5.15** Misalkan model regresi linear dengan heteroskedastisitas

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \quad \mathbb{E}[\varepsilon | \mathbf{X}] = \mathbf{0}, \quad \text{Var}(\varepsilon | \mathbf{X}) = \sigma^2 \Omega,$$

dengan  $\Omega$  diketahui dan definit positif. Jika kondisi-kondisi berikut terpenuhi:

- (i)  $\frac{1}{n} \mathbf{X}^\top \Omega^{-1} \mathbf{X} \xrightarrow{p} \mathbf{Q}_\Omega$  dengan  $\mathbf{Q}_\Omega$  matriks definit positif,
- (ii)  $\frac{1}{n} \mathbf{X}^\top \Omega^{-1} \varepsilon \xrightarrow{p} \mathbf{0}$ ,

maka penduga GLS  $\hat{\beta}_{GLS} = (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{y}$  bersifat konsisten, yaitu

$$\hat{\beta}_{GLS} \xrightarrow{p} \beta. \quad (2.5.71)$$

**Bukti.** Penduga GLS dapat dituliskan sebagai

$$\begin{aligned} \hat{\beta}_{GLS} &= (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{y} \\ &= (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} (\mathbf{X}\beta + \boldsymbol{\varepsilon}) \\ &= \beta + (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} \\ &= \beta + \left( \frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X} \right)^{-1} \left( \frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} \right). \end{aligned} \quad (2.5.72)$$

Berdasarkan kondisi (i),  $\frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X} \xrightarrow{p} \mathbf{Q}_\Omega$  dengan  $\mathbf{Q}_\Omega$  definit positif. Dengan teorema pemetaan kontinu,  $\left( \frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X} \right)^{-1} \xrightarrow{p} \mathbf{Q}_\Omega^{-1}$ . Berdasarkan kondisi (ii),  $\frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} \xrightarrow{p} \mathbf{0}$ . Dengan teorema Slutsky, diperoleh

$$\hat{\beta}_{GLS} - \beta = \left( \frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X} \right)^{-1} \left( \frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} \right) \xrightarrow{p} \mathbf{Q}_\Omega^{-1} \cdot \mathbf{0} = \mathbf{0}, \quad (2.5.73)$$

sehingga  $\hat{\beta}_{GLS} \xrightarrow{p} \beta$ . ■

Berdasarkan teorema di atas, penduga GLS konsisten jika dan hanya jika matriks bobot terspesifikasi dengan benar dan kondisi-kondisi konsistensi

**Teorema 2.5.16** *Misalkan kondisi-kondisi konsistensi terpenuhi dan tambahan kondisi berikut dipenuhi:*

$$(i) \quad \frac{1}{\sqrt{n}} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}_\Omega) \text{ (kondisi Lindeberg–Feller terpenuhi).}$$

*Maka penduga GLS memiliki distribusi asimtotik*

$$\sqrt{n}(\hat{\beta}_{GLS} - \beta) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}_\Omega^{-1}). \quad (2.5.74)$$

**Bukti.** Dari pembuktian konsistensi, dapat dituliskan

$$\begin{aligned}\sqrt{n}(\hat{\beta}_{\text{GLS}} - \beta) &= \sqrt{n} \left( \frac{1}{n} \mathbf{X}^\top \Omega^{-1} \mathbf{X} \right)^{-1} \left( \frac{1}{n} \mathbf{X}^\top \Omega^{-1} \epsilon \right) \\ &= \left( \frac{1}{n} \mathbf{X}^\top \Omega^{-1} \mathbf{X} \right)^{-1} \left( \frac{1}{\sqrt{n}} \mathbf{X}^\top \Omega^{-1} \epsilon \right).\end{aligned}\quad (2.5.75)$$

Berdasarkan kondisi (i),  $\frac{1}{\sqrt{n}} \mathbf{X}^\top \Omega^{-1} \epsilon \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}_\Omega)$ . Dengan teorema Slutsky, sebab  $\left( \frac{1}{n} \mathbf{X}^\top \Omega^{-1} \mathbf{X} \right)^{-1} \xrightarrow{p} \mathbf{Q}_\Omega^{-1}$ , maka

$$\sqrt{n}(\hat{\beta}_{\text{GLS}} - \beta) \xrightarrow{d} \mathbf{Q}_\Omega^{-1} \cdot \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}_\Omega) = \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}_\Omega^{-1}). \quad (2.5.76)$$

■

## 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 Fan dan Gijbels (1996).

**Definisi 2.6.1 (Model regresi nonparametrik)** *Pertimbangkan model regresi nonparametrik*

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (2.6.1)$$

dengan  $m(\cdot)$  adalah fungsi regresi yang tidak diketahui,  $X_i$  adalah variabel prediktor,  $Y_i$  adalah variabel respons, dan  $\varepsilon_i$  adalah galat acak dengan  $\mathbb{E}[\varepsilon_i | X_i] = 0$  dan  $\text{Var}(\varepsilon_i | X_i) = \sigma^2(X_i)$ .

**Definisi 2.6.2 (Aproksimasi polinomial lokal, Fan dan Gijbels 1996)** *Untuk mengestimasi  $m(x_0)$  pada titik target  $x_0$ , fungsi regresi  $m(X)$  diaproksimasi secara lokal*



menggunakan ekspansi Taylor orde- $p$  di sekitar  $x_0$ :

$$m(X) \approx \sum_{j=0}^p \frac{m^{(j)}(x_0)}{j!} (X - x_0)^j, \quad (2.6.2)$$

dengan mendefinisikan  $\beta_j = \frac{m^{(j)}(x_0)}{j!}$  untuk  $j = 0, 1, \dots, p$ , aproksimasi lokal dapat dituliskan sebagai

$$m(X) \approx \sum_{j=0}^p \beta_j (X - x_0)^j. \quad (2.6.3)$$

$X$  adalah ketetanggaan dari  $x_0$  dengan aproksimasi ini berlaku dengan baik.

Estimator polinomial lokal  $\hat{m}(x_0)$  diperoleh dengan menyelesaikan masalah optimisasi berikut:

$$\min_{\beta_0, \beta_1, \dots, \beta_p} \sum_{i=1}^n K_h(X_i - x_0) \left[ Y_i - \sum_{j=0}^p \beta_j (X_i - x_0)^j \right]^2, \quad (2.6.4)$$

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  $\hat{m}(x_0) = \hat{\beta}_0(x_0)$ , dengan  $\hat{\beta}_0(x_0)$  adalah solusi optimal dari masalah di atas.

**Contoh 2.6.3** Dua kasus khusus yang paling umum digunakan adalah sebagai berikut.

- (a) **Estimator Nadaraya–Watson** ( $p = 0$ ): Dengan  $p = 0$ , diperoleh estimator rata-rata lokal

$$\hat{m}(x_0) = \frac{\sum_{i=1}^n K_h(X_i - x_0) Y_i}{\sum_{i=1}^n K_h(X_i - x_0)}. \quad (2.6.5)$$

- (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

Pertimbangkan model regresi nonparametrik

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (2.6.6)$$

dengan  $m(\cdot)$  adalah fungsi regresi yang tidak diketahui,  $X_i$  adalah variabel prediktor,  $Y_i$  adalah variabel respons, dan  $\varepsilon_i$  adalah galat acak dengan  $\mathbb{E}[\varepsilon_i | X_i] = 0$  dan  $\text{Var}(\varepsilon_i | X_i) = \sigma^2(X_i)$ . Untuk memperoleh sifat-sifat asimtotik estimator polinomial lokal, diperlukan asumsi-asumsi regularitas berikut.

**Asumsi LPM.1** Untuk mengaproksimasi  $m(X)$  di sekitar titik  $x_0$ , fungsi regresi  $m(X)$  diasumsikan memiliki turunan kontinu hingga orde  $(p + 1)$  di sekitar  $x_0$ .

**Asumsi LPM.2** Galat  $\varepsilon_i$  memenuhi  $\mathbb{E}[\varepsilon_i | X_i] = 0$  dan  $\text{Var}(\varepsilon_i | X_i) = \sigma^2(X_i)$  dengan  $\sigma^2(x)$  kontinu di  $x_0$ .

**Asumsi LPM.3** Fungsi kernel  $K(\cdot)$  adalah yang bekerja sebagai mekanisme pembobot dengan ketentuan sebagai berikut.

- (a)  $K(u) \geq 0$  untuk semua  $u \in \mathbb{R}$ .
- (b)  $\int_{\mathbb{R}} K(u) \, du = 1$ .
- (c)  $\int_{\mathbb{R}} uK(u) \, du = 0$  atau  $K(u) = K(-u)$  untuk semua  $u \in \mathbb{R}$ .

**Asumsi LPM.4** Seiring  $n \rightarrow \infty$ , berlaku  $h \rightarrow 0$  dan  $nh \rightarrow \infty$ .

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 Estimasi Model Polinomial Lokal

**Definisi 2.6.4** *Regresi tertimbang lokal (locally weighted least squares atau LWLS) model polinomial lokal pada titik  $x_0$  menyelesaikan masalah optimisasi berikut:*

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n K_h(X_i - x_0) [Y_i - \mathbf{z}_i^\top \boldsymbol{\beta}]^2, \quad (2.6.7)$$

dengan  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$  dan  $\mathbf{z}_i = (1, X_i - x_0, \dots, (X_i - x_0)^p)^\top$ . Solusi optimal  $\hat{\boldsymbol{\beta}}(x_0)$  memberikan estimator polinomial lokal  $\hat{m}(x_0) = \hat{\beta}_0(x_0)$ .

**Teorema 2.6.5** *Definisikan matriks desain lokal  $\mathbf{X}_{x_0}$  dan matriks bobot  $\mathbf{W}_{x_0}$  sebagai*

$$\mathbf{X}_{x_0} = \begin{pmatrix} 1 & (X_1 - x_0) & (X_1 - x_0)^2 & \cdots & (X_1 - x_0)^p \\ 1 & (X_2 - x_0) & (X_2 - x_0)^2 & \cdots & (X_2 - x_0)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (X_n - x_0) & (X_n - x_0)^2 & \cdots & (X_n - x_0)^p \end{pmatrix} \in \mathbb{R}^{n \times (p+1)}, \quad (2.6.8)$$

dan

$$\mathbf{W}_{x_0} = \text{diag}(K_h(X_1 - x_0), K_h(X_2 - x_0), \dots, K_h(X_n - x_0)) \in \mathbb{R}^{n \times n}. \quad (2.6.9)$$

Maka estimator polinomial lokal diberikan oleh

$$\hat{\boldsymbol{\beta}}(x_0) = (\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{y}, \quad (2.6.10)$$

dengan  $\mathbf{y} = (Y_1, Y_2, \dots, Y_n)^\top$ . Khususnya, estimator untuk  $m(x_0)$  adalah

$$\hat{m}(x_0) = \mathbf{e}_1^\top \hat{\boldsymbol{\beta}}(x_0) = \mathbf{e}_1^\top (\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{y}, \quad (2.6.11)$$

dengan  $\mathbf{e}_1 = (1, 0, \dots, 0)^\top \in \mathbb{R}^{p+1}$ .

**Bukti.** Masalah minimisasi LWLS dapat dituliskan dalam bentuk matriks sebagai

$$\min_{\beta} (\mathbf{y} - \mathbf{X}_{x_0}\beta)^\top \mathbf{W}_{x_0} (\mathbf{y} - \mathbf{X}_{x_0}\beta). \quad (2.6.12)$$

Turunan pertama terhadap  $\beta$  adalah

$$\frac{\partial}{\partial \beta} (\mathbf{y} - \mathbf{X}_{x_0}\beta)^\top \mathbf{W}_{x_0} (\mathbf{y} - \mathbf{X}_{x_0}\beta) = -2\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} (\mathbf{y} - \mathbf{X}_{x_0}\beta). \quad (2.6.13)$$

Menyamakan dengan nol memberikan persamaan normal tertimbang

$$\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0} \hat{\beta}(x_0) = \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{y}. \quad (2.6.14)$$

Jika  $\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0}$  nonsingular, solusinya adalah

$$\hat{\beta}(x_0) = (\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{y}. \quad (2.6.15)$$

Sebab  $\hat{\beta}_0(x_0)$  adalah komponen pertama dari  $\hat{\beta}(x_0)$ , maka  $\hat{m}(x_0) = \mathbf{e}_1^\top \hat{\beta}(x_0)$ . ■

**Teorema 2.6.6** *Estimator  $\hat{m}(x_0)$  dapat dituliskan sebagai kombinasi linear tertimbang dari respons:*

$$\hat{m}(x_0) = \sum_{i=1}^n w_i(x_0) Y_i, \quad (2.6.16)$$

dengan bobot efektif  $w_i(x_0) = \mathbf{e}_1^\top (\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{e}_i$ , dengan  $\mathbf{e}_i$  adalah vektor unit ke- $i$  dalam  $\mathbb{R}^n$ .

**Bukti.** Dari teorema sebelumnya, estimator polinomial lokal diberikan oleh

$$\hat{m}(x_0) = \mathbf{e}_1^\top (\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{y}. \quad (2.6.17)$$

Vektor respons  $\mathbf{y} = (Y_1, Y_2, \dots, Y_n)^\top$  dapat dituliskan sebagai kombinasi linear dari vektor-vektor unit standar dalam  $\mathbb{R}^n$ , yaitu

$$\mathbf{y} = \sum_{i=1}^n Y_i \mathbf{e}_i, \quad (2.6.18)$$

dengan  $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)^\top$  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

$$\begin{aligned}\hat{m}(x_0) &= \mathbf{e}_1^\top (\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \left( \sum_{i=1}^n Y_i \mathbf{e}_i \right) \\ &= \sum_{i=1}^n Y_i \cdot \mathbf{e}_1^\top (\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{e}_i,\end{aligned}\quad (2.6.19)$$

dengan pertukaran urutan penjumlahan dan perkalian matriks dijamin oleh linearitas operasi matriks.

Dengan mendefinisikan bobot efektif sebagai

$$w_i(x_0) = \mathbf{e}_1^\top (\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{e}_i, \quad (2.6.20)$$

diperoleh representasi yang diinginkan

$$\hat{m}(x_0) = \sum_{i=1}^n w_i(x_0) Y_i. \quad (2.6.21)$$

Perhatikan bahwa  $\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{e}_i$  adalah kolom ke- $i$  dari matriks  $\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0}$ , yang sama dengan  $K_h(X_i - x_0) \mathbf{z}_i$  dengan  $\mathbf{z}_i = (1, X_i - x_0, \dots, (X_i - x_0)^p)^\top$ . 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. ■

## 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 (Anselin, 1988; Fotheringham dkk.,

2002; LeSage dan Pace, 2009).

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  $\text{Cov}(\varepsilon_i, \varepsilon_j) \neq 0$  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  $\beta$  tidak konstan di seluruh wilayah pengamatan, melainkan merupakan fungsi dari lokasi geografis, yaitu  $\beta = \beta(s)$  dengan  $s$  menyatakan lokasi. Fenomena ini mengindikasikan bahwa proses yang mendasari hubungan antarvariabel berbeda di setiap wilayah.

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 (Anselin, 1988; LeSage dan Pace, 2009).

**Definisi 2.7.1** Misalkan  $Y_i$  merupakan nilai variabel di lokasi  $i$  dan  $Y_j$  merupakan

nilai variabel di lokasi tetangga  $j$ . *Dependensi spasial terjadi jika*

$$\text{Cov}(Y_i, Y_j) \neq 0, \quad \text{untuk lokasi } i \neq j \text{ yang berdekatan}, \quad (2.7.1)$$

yang dapat diukur menggunakan statistik Moran's  $I$ , yaitu

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_{i=1}^n (Y_i - \bar{Y})^2}, \quad (2.7.2)$$

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

$$y_i = \rho \sum_{j=1}^n w_{ij} y_j + \sum_{k=1}^p \beta_k x_{ik} + \varepsilon_i, \quad (2.7.3)$$

atau dalam bentuk matriksnya adalah

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.7.4)$$

dengan  $\rho$  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

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{u}, \quad \mathbf{u} = \lambda \mathbf{W} \mathbf{u} + \boldsymbol{\varepsilon}, \quad (2.7.5)$$

dengan  $\lambda$  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

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \quad (2.7.6)$$

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 (Fotheringham dkk., 2002).

Heterogenitas spasial terjadi ketika parameter regresi  $\boldsymbol{\beta}$  bervariasi menurut lokasi geografis, yaitu

$$\boldsymbol{\beta} = \boldsymbol{\beta}(s_i) = \boldsymbol{\beta}(u_i, v_i), \quad (2.7.7)$$

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.

Pemodelan polinomial lokal yang telah dibahas pada bagian sebelumnya menyediakan kerangka teoretis fundamental untuk memahami 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  $\beta(u, v)$  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.

**Definisi 2.7.5 (*Geographically Weighted Regression (GWR)*, Fotheringham dkk. 2002)** Model GWR untuk observasi ke- $i$  pada lokasi  $(u_i, v_i)$  didefinisikan sebagai

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i)x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (2.7.8)$$

atau dalam bentuk vektor

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta}(u_i, v_i) + \varepsilon_i, \quad (2.7.9)$$

dengan  $y_i$  adalah variabel respons pada lokasi  $i$ ,  $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ip})^\top \in \mathbb{R}^{p+1}$  adalah vektor variabel prediktor,  $\boldsymbol{\beta}(u_i, v_i) = (\beta_0(u_i, v_i), \beta_1(u_i, v_i), \dots, \beta_p(u_i, v_i))^\top$  adalah vektor koefisien lokal,  $(u_i, v_i)$  adalah koordinat geografis lokasi ke- $i$ , dan  $\varepsilon_i$  adalah galat acak.

Dalam bentuk matriks, dapat dituliskan sebagai berikut.

$$\mathbf{y} = (\mathbf{X} \odot \mathcal{B}) \mathbf{1}_{p+1} + \boldsymbol{\varepsilon}, \quad (2.7.10)$$

dengan

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix},$$

$$\mathcal{B} = \begin{pmatrix} \beta_0(u_1, v_1) & \cdots & \beta_p(u_1, v_1) \\ \beta_0(u_2, v_2) & \cdots & \beta_p(u_2, v_2) \\ \vdots & \ddots & \vdots \\ \beta_0(u_n, v_n) & \cdots & \beta_p(u_n, v_n) \end{pmatrix}, \quad \text{dan } \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}. \quad (2.7.11)$$

Operasi  $\odot$  menunjukkan perkalian Hadamard (elemen demi elemen) antara matriks  $\mathbf{X}$  dan  $\mathcal{B}$ .

Model GWR didasarkan pada asumsi-asumsi berikut.

**Asumsi GWR.1** Barisan sampel  $\{(\mathbf{x}_i, (u_i, v_i)) : i = 1, 2, \dots, n\}$  stasioner dan dependen untuk setiap  $i$  dan  $j$  dengan  $\|i - j\|_\infty \leq h$ . Dua vektor acak  $(\mathbf{x}_i, (u_i, v_i))$  dan  $(\mathbf{x}_j, (u_j, v_j))$  bersifat independen ketika  $\|i - j\|_\infty > h$ .

**Asumsi GWR.2** Galat acak  $\varepsilon_i$  memiliki nilai harapan nol, yaitu  $\mathbb{E}[\varepsilon_i \mid \mathbf{X}, (u_i, v_i)] = 0$  untuk semua  $i$ .

**Asumsi GWR.3** Variansi galat acak adalah konstan, yaitu  $\text{Var}(\varepsilon_i \mid \mathbf{X}, (u_i, v_i)) = \sigma^2$  untuk semua  $i$ .

**Asumsi GWR.4** Fungsi koefisien lokal  $\beta_k(u, v)$  untuk  $k = 0, 1, \dots, p$  memiliki turunan parsial kontinu hingga orde kedua di setiap lokasi.

**Asumsi GWR.5** Fungsi kepadatan marginal dari lokasi observasi, yaitu  $f(u, v)$ , memiliki turunan parsial kontinu hingga orde kedua di setiap lokasi.

**Asumsi GWR.6** Ketika  $n \rightarrow \infty$ , dipunyai  $h \rightarrow 0$  dan  $nh^{d-1} \rightarrow \infty$ .

**Asumsi GWR.7** Untuk setiap lokasi  $i$ , matriks  $\mathbf{X}^\top \mathbf{W}_i \mathbf{X}$  adalah nonsingular.

**Teorema 2.7.6** Untuk setiap lokasi  $(u_i, v_i)$ , estimator locally weighted least squares (LWLS) untuk koefisien lokal diperoleh dengan meminimalkan

$$S_i(\boldsymbol{\beta}) = \sum_{j=1}^n w_{ij} (y_j - \mathbf{x}_j^\top \boldsymbol{\beta})^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{W}_i (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad (2.7.12)$$

dengan  $\mathbf{W}_i = \text{diag}(w_{i1}, w_{i2}, \dots, w_{in})$ . Estimator yang dihasilkan adalah

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i \mathbf{y}. \quad (2.7.13)$$

**Bukti.** Untuk lokasi  $(u_i, v_i)$ , tujuannya adalah meminimalkan fungsi objektif kuadrat terkecil tertimbang lokal

$$S_i(\boldsymbol{\beta}) = \sum_{j=1}^n w_{ij} (y_j - \mathbf{x}_j^\top \boldsymbol{\beta})^2. \quad (2.7.14)$$

Dalam notasi matriks, fungsi ini dapat dituliskan sebagai

$$S_i(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{W}_i (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad (2.7.15)$$

dengan  $\mathbf{W}_i = \text{diag}(w_{i1}, w_{i2}, \dots, w_{in})$  adalah matriks diagonal berukuran  $n \times n$  yang berisi bobot spasial. Secara interpretatif, persamaan (2.7.15) menggunakan koefisien  $\boldsymbol{\beta}$  yang sama untuk semua observasi, tetapi bobot  $w_{ij}$  menekankan kontribusi observasi yang lebih dekat ke lokasi target  $(u_i, v_i)$  karena residual dari observasi tersebut memiliki bobot yang lebih besar dalam fungsi objektif.

Ekspansi bentuk kuadratik memberikan

$$\begin{aligned} S_i(\boldsymbol{\beta}) &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{W}_i (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}^\top \mathbf{W}_i \mathbf{y} - \mathbf{y}^\top \mathbf{W}_i \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{W}_i \mathbf{y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{W}_i \mathbf{X}\boldsymbol{\beta}. \end{aligned} \quad (2.7.16)$$

Sebab  $\mathbf{y}^\top \mathbf{W}_i \mathbf{X}\boldsymbol{\beta}$  adalah skalar, maka  $\mathbf{y}^\top \mathbf{W}_i \mathbf{X}\boldsymbol{\beta} = (\mathbf{y}^\top \mathbf{W}_i \mathbf{X}\boldsymbol{\beta})^\top = \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{W}_i^\top \mathbf{y} =$

$\beta^\top \mathbf{X}^\top \mathbf{W}_i \mathbf{y}$  (sebab  $\mathbf{W}_i$  simetris). Dengan demikian,

$$S_i(\beta) = \mathbf{y}^\top \mathbf{W}_i \mathbf{y} - 2\beta^\top \mathbf{X}^\top \mathbf{W}_i \mathbf{y} + \beta^\top \mathbf{X}^\top \mathbf{W}_i \mathbf{X} \beta. \quad (2.7.17)$$

Untuk menemukan minimum, dihitung turunan pertama terhadap  $\beta$  dan disamakan dengan nol. Dengan menggunakan aturan diferensial matriks, turunan dari masing-masing suku adalah sebagai berikut.

- (i) Suku  $\mathbf{y}^\top \mathbf{W}_i \mathbf{y}$  tidak bergantung pada  $\beta$ , sehingga turunannya adalah 0.
- (ii) Suku  $-2\beta^\top \mathbf{X}^\top \mathbf{W}_i \mathbf{y}$  berbentuk  $\mathbf{c}^\top \beta$  dengan  $\mathbf{c} = -2\mathbf{X}^\top \mathbf{W}_i \mathbf{y}$ . Turunannya adalah  $\mathbf{c} = -2\mathbf{X}^\top \mathbf{W}_i \mathbf{y}$ .
- (iii) Suku  $\beta^\top \mathbf{X}^\top \mathbf{W}_i \mathbf{X} \beta$  berbentuk  $\beta^\top \mathbf{A} \beta$  dengan  $\mathbf{A} = \mathbf{X}^\top \mathbf{W}_i \mathbf{X}$  yang simetris (sebab  $\mathbf{W}_i$  simetris). Turunannya adalah  $2\mathbf{A}\beta = 2\mathbf{X}^\top \mathbf{W}_i \mathbf{X} \beta$ .

Dengan demikian, gradien dari  $S_i(\beta)$  terhadap  $\beta$  adalah

$$\frac{\partial S_i(\beta)}{\partial \beta} = -2\mathbf{X}^\top \mathbf{W}_i \mathbf{y} + 2\mathbf{X}^\top \mathbf{W}_i \mathbf{X} \beta. \quad (2.7.18)$$

Syarat orde pertama untuk minimum adalah gradien sama dengan nol:

$$\begin{aligned} -2\mathbf{X}^\top \mathbf{W}_i \mathbf{y} + 2\mathbf{X}^\top \mathbf{W}_i \mathbf{X} \hat{\beta} &= \mathbf{0} \\ \mathbf{X}^\top \mathbf{W}_i \mathbf{X} \hat{\beta} &= \mathbf{X}^\top \mathbf{W}_i \mathbf{y}. \end{aligned} \quad (2.7.19)$$

Persamaan terakhir merupakan persamaan normal tertimbang untuk lokasi  $i$ . Dengan asumsi bahwa matriks  $\mathbf{X}^\top \mathbf{W}_i \mathbf{X}$  adalah nonsingular (invertibel), solusinya adalah

$$\hat{\beta}(u_i, v_i) = (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i \mathbf{y}. \quad (2.7.20)$$

Untuk memverifikasi bahwa solusi ini adalah peminimum, diperiksa turunan kedua (matriks Hessian):

$$\frac{\partial^2 S_i(\beta)}{\partial \beta \partial \beta^\top} = 2\mathbf{X}^\top \mathbf{W}_i \mathbf{X}. \quad (2.7.21)$$

Matriks  $\mathbf{X}^\top \mathbf{W}_i \mathbf{X}$  adalah semi-definit positif sebab untuk sembarang vektor  $\mathbf{v} \in \mathbb{R}^{p+1}$ ,

$$\mathbf{v}^\top \mathbf{X}^\top \mathbf{W}_i \mathbf{X} \mathbf{v} = (\mathbf{X} \mathbf{v})^\top \mathbf{W}_i (\mathbf{X} \mathbf{v}) = \sum_{j=1}^n w_{ij} (\mathbf{x}_j^\top \mathbf{v})^2 \geq 0, \quad (2.7.22)$$

sebab  $w_{ij} \geq 0$  untuk semua  $j$ . Jika  $\mathbf{X}^\top \mathbf{W}_i \mathbf{X}$  nonsingular, maka matriks ini definit positif, sehingga Hessian definit positif dan  $\hat{\beta}(u_i, v_i)$  adalah peminimum global yang unik. ■

**Akibat 2.7.7** Nilai prediksi pada lokasi  $i$  dapat dituliskan sebagai kombinasi linear dari respons, yaitu

$$\hat{y}_i = \mathbf{x}_i^\top \hat{\beta}(u_i, v_i) = \mathbf{r}_i^\top \mathbf{y}, \quad (2.7.23)$$

dengan  $\mathbf{r}_i^\top = \mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i$ . Secara agregat, vektor prediksi  $\hat{\mathbf{y}} = \mathbf{S} \mathbf{y}$ , dengan matriks hat GWR

$$\mathbf{S} = \begin{pmatrix} \mathbf{r}_1^\top \\ \mathbf{r}_2^\top \\ \vdots \\ \mathbf{r}_n^\top \end{pmatrix}. \quad (2.7.24)$$

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.

Fungsi *kernel*  $K_h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  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.

**Asumsi GWR.8** Fungsi kernel  $K_h(\cdot)$  memenuhi kondisi 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 penyokong (support) terbatas yang kompak, yaitu terdapat konstanta  $C > 0$  sehingga  $K_h(d) = 0$  untuk semua  $d > Ch$ .
- (d) Sesuai asumsi ??,  $K_h(\cdot)$  bersifat simeteris, yaitu  $\int_0^\infty dK_h(d)f_D(d) dd = 0$ .

Asumsi normalisasi atau  $\int K_h(d)f_D(d) dd = 1$  tidak diwajibkan dalam konteks GWR karena estimator GWR bersifat invarian terhadap skala bobot.

Beberapa fungsi *kernel* yang umum digunakan adalah sebagai berikut.

- (a) **Kernel Gaussian:**

$$w_{ij} = \exp \left( -\frac{1}{2} \left( \frac{d_{ij}}{h} \right)^2 \right). \quad (2.7.25)$$

- (b) **Kernel Bi-square (kompak):**

$$w_{ij} = \begin{cases} \left( 1 - \left( \frac{d_{ij}}{h} \right)^2 \right)^2 & \text{jika } d_{ij} < h, \\ 0 & \text{jika } d_{ij} \geq h. \end{cases} \quad (2.7.26)$$

- (c) **Kernel Tri-cube (kompak):**

$$w_{ij} = \begin{cases} \left( 1 - \left( \frac{d_{ij}}{h} \right)^3 \right)^3 & \text{jika } d_{ij} < h, \\ 0 & \text{jika } d_{ij} \geq h. \end{cases} \quad (2.7.27)$$

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 2.7.8** 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:**

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}. \quad (2.7.28)$$

(b) **Jarak Manhattan:**

$$d_{ij} = |u_i - u_j| + |v_i - v_j|. \quad (2.7.29)$$

(c) **Jarak Geodesik (Great Circle Distance):**

$$d_{ij} = R \cdot \arccos(\sin(\phi_i) \sin(\phi_j) + \cos(\phi_i) \cos(\phi_j) \cos(\lambda_i - \lambda_j)), \quad (2.7.30)$$

dengan  $R$  adalah jari-jari bumi,  $(\phi_i, \lambda_i)$  dan  $(\phi_j, \lambda_j)$  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* sebagai berikut.

- (a) *Fixed bandwidth*, yaitu parameter  $h$  bernilai konstan untuk semua lokasi. Pendekatan ini sesuai ketika observasi terdistribusi secara relatif homogen di ruang geografis.
- (b) *Adaptive bandwidth*, yaitu 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

$$h_i = d_{i[k]}, \quad (2.7.31)$$

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 2.7.9** Bandwidth optimal  $h^*$  dipilih dengan meminimalkan fungsi leave-one-out cross-validation:

$$CV(h) = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(h))^2, \quad (2.7.32)$$

dengan  $\hat{y}_{\neq i}(h)$  adalah prediksi untuk observasi  $i$  menggunakan model yang di-fit tanpa observasi  $i$ . Secara eksplisit,

$$h^* = \arg \min_h CV(h). \quad (2.7.33)$$

Ketika *bandwidth*  $h$  diestimasi dari data (bukan ditetapkan secara eksogen), terdapat implikasi penting terhadap teori asimtotik, yaitu:

- (a) Estimator  $\hat{h}$  bergantung pada  $\mathbf{y}$ , sehingga matriks bobot  $\mathbf{W}_i(\hat{h})$  menjadi stokastik ( $\mathbf{W}_i(\hat{h})$ ) dan bergantung pada data. Hal ini memperumit derivasi distribusi asimtotik.
- (b) Ketidakpastian dalam  $\hat{h}$  menambah variabilitas pada estimator  $\hat{\beta}$ , 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 2.7.10** Di bawah asumsi GWR.1 sampai GWR.8, estimator GWR memiliki bias yang bergantung pada tingkat kehalusan koefisien dan bandwidth, yaitu

$$\mathbb{E}[\hat{\beta}(u_i, v_i)] = \beta(u_i, v_i) + \mathcal{O}(h^2), \quad (2.7.34)$$

dengan suku bias  $\mathcal{O}(h^2)$  proporsional dengan turunan kedua dari  $\beta(u, v)$  di lokasi  $i$ .



**Bukti.** Berdasarkan asumsi GWR.4, fungsi koefisien  $\beta(u, v)$  memiliki turunan parsial kontinu hingga orde kedua. Didefinisikan vektor perpindahan  $\delta_{ij} = (u_j - u_i, v_j - v_i)^\top$  dengan norma  $d_{ij} = \|\delta_{ij}\|$ . Dengan ekspansi Taylor di sekitar lokasi  $(u_i, v_i)$ , untuk observasi  $j$  dengan lokasi  $(u_j, v_j)$ , diperoleh

$$\beta(u_j, v_j) = \beta(u_i, v_i) + \nabla\beta(u_i, v_i)^\top \delta_{ij} + \frac{1}{2} \delta_{ij}^\top \mathbf{H}_\beta(u_i, v_i) \delta_{ij} + R_{ij}, \quad (2.7.35)$$

dengan  $\nabla\beta$  adalah gradien dan  $\mathbf{H}_\beta$  adalah matriks Hessian dari  $\beta$ .

Suku sisa  $R_{ij} = o(d_{ij}^2)$  karena berdasarkan teorema Taylor dengan sisa Peano, untuk fungsi  $\beta : \mathbb{R}^2 \rightarrow \mathbb{R}^{p+1}$  yang memiliki turunan parsial kontinu hingga orde kedua di sekitar  $(u_i, v_i)$ , suku sisa setelah ekspansi orde kedua memenuhi

$$\lim_{d_{ij} \rightarrow 0} \frac{\|R_{ij}\|}{d_{ij}^2} = 0, \quad (2.7.36)$$

yang merupakan definisi dari  $R_{ij} = o(d_{ij}^2)$ . Secara eksplisit, jika turunan orde ketiga ada dan terbatas, maka  $R_{ij} = \mathcal{O}(d_{ij}^3)$ , yang tentu saja memenuhi  $R_{ij} = o(d_{ij}^2)$ .

Model GWR yang sebenarnya adalah  $y_j = \mathbf{x}_j^\top \beta(u_j, v_j) + \varepsilon_j$ . Substitusi ekspansi Taylor memberikan

$$y_j = \mathbf{x}_j^\top \beta(u_i, v_i) + \mathbf{x}_j^\top \mathbf{r}_{ij} + \varepsilon_j, \quad (2.7.37)$$

dengan  $\mathbf{r}_{ij} = \nabla\beta(u_i, v_i)^\top \delta_{ij} + \frac{1}{2} \delta_{ij}^\top \mathbf{H}_\beta(u_i, v_i) \delta_{ij} + R_{ij}$  adalah suku sisa deterministik yang mengandung turunan pertama dan kedua dari  $\beta$ . Oleh karena itu, nilai harapan dari estimator LWLS dapat dituliskan sebagai

$$\begin{aligned} \mathbb{E}[\widehat{\beta}(u_i, v_i)] &= \mathbb{E}[(\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i \mathbf{y}] \\ &= (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i \mathbb{E}[\mathbf{y}] \\ &= (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i (\mathbf{X} \beta(u_i, v_i) + \mathbf{X} \mathbf{r}_i), \end{aligned} \quad (2.7.38)$$

dengan  $\mathbf{r}_i = (\mathbf{r}_{i1}, \dots, \mathbf{r}_{in})^\top$  adalah vektor suku sisa. Sebab  $(\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i \mathbf{X} =$

I, maka

$$\mathbb{E}[\hat{\beta}(u_i, v_i)] = \beta(u_i, v_i) + (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i \mathbf{X} \mathbf{r}_i. \quad (2.7.39)$$

Dengan menggunakan sifat fungsi *kernel* bahwa bobot  $w_{ij}$  terkonsentrasi pada observasi dengan jarak  $d_{ij} = \mathcal{O}(h)$  dari lokasi  $i$ , suku sisa  $\mathbf{r}_{ij} = \mathcal{O}(h^2)$  karena didominasi oleh suku kuadratik dalam ekspansi Taylor. Oleh karena itu, vektor bias adalah

$$\mathbf{b}_i = (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i \mathbf{X} \mathbf{r}_i = \mathcal{O}(h^2), \quad (2.7.40)$$

yang proporsional dengan turunan kedua dari  $\beta(u, v)$  di lokasi  $i$ . ■

**Teorema 2.7.11 (Fotheringham dkk., 2002)** *Di bawah asumsi GWR.1 sampai GWR.8, matriks kovarians bersyarat dari estimator GWR adalah*

$$\text{Var}(\hat{\beta}(u_i, v_i) \mid \mathbf{X}) = \sigma^2 (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i^2 \mathbf{X} (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1}. \quad (2.7.41)$$

*Dengan normalisasi yang tepat, variansi adalah  $\mathcal{O}((nh^2)^{-1})$ .*

**Bukti.** Estimator GWR dapat dituliskan sebagai

$$\hat{\beta}(u_i, v_i) = (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i \mathbf{y} = \mathbf{C}_i \mathbf{y}, \quad (2.7.42)$$

dengan  $\mathbf{C}_i = (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i$ .

Berdasarkan asumsi GWR.2, galat  $\varepsilon_1, \dots, \varepsilon_n$  bersifat independen dengan  $\mathbb{E}[\varepsilon_j] = 0$  dan  $\text{Var}(\varepsilon_j) = \sigma^2$  untuk semua  $j$ . Dalam notasi matriks,  $\text{Var}(\boldsymbol{\varepsilon} \mid \mathbf{X}) = \sigma^2 \mathbf{I}_n$ . Sebab  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  dengan  $\mathbf{X}$  diperlakukan sebagai tetap (bersyarat), maka

$$\begin{aligned} \text{Var}(\hat{\beta}(u_i, v_i) \mid \mathbf{X}) &= \text{Var}(\mathbf{C}_i \mathbf{y} \mid \mathbf{X}) \\ &= \mathbf{C}_i \text{Var}(\mathbf{y} \mid \mathbf{X}) \mathbf{C}_i^\top \\ &= \mathbf{C}_i (\sigma^2 \mathbf{I}_n) \mathbf{C}_i^\top \\ &= \sigma^2 \mathbf{C}_i \mathbf{C}_i^\top. \end{aligned} \quad (2.7.43)$$

Substitusi  $\mathbf{C}_i = (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i$  memberikan

$$\begin{aligned} \mathbf{C}_i \mathbf{C}_i^\top &= (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i [(\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i]^\top \\ &= (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i \mathbf{W}_i^\top \mathbf{X} (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1}. \end{aligned} \quad (2.7.44)$$

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,  $(\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-\top} = (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1}$  karena matriks ini simetris. Oleh karena itu,

$$\text{Var}(\hat{\beta}(u_i, v_i) \mid \mathbf{X}) = \sigma^2 (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i^2 \mathbf{X} (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1}. \quad (2.7.45)$$

Untuk menunjukkan bahwa variansi adalah  $\mathcal{O}((nh^2)^{-1})$ , 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,  $n_{\text{eff}} = \mathcal{O}(nh^2)$ . Sebab variansi estimator berbanding terbalik dengan jumlah observasi efektif, maka  $\text{Var}(\hat{\beta}(u_i, v_i)) = \mathcal{O}((nh^2)^{-1})$ . ■

**Teorema 2.7.12** *Estimator konsisten untuk  $\sigma^2$  diberikan oleh*

$$\hat{\sigma}^2 = \frac{\mathbf{e}^\top \mathbf{e}}{n - 2\text{tr}(\mathbf{S}) + \text{tr}(\mathbf{S}^\top \mathbf{S})}, \quad (2.7.46)$$

dengan  $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$  adalah vektor residual dan penyebut merepresentasikan derajat bebas efektif residual.

**Bukti.** Vektor nilai prediksi GWR dapat dituliskan sebagai  $\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}$ , dengan  $\mathbf{S}$  adalah matriks topi GWR yang baris ke- $i$ -nya adalah  $\mathbf{s}_i^\top = \mathbf{x}_i^\top (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i$ . Vektor residual adalah

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{S})\mathbf{y}. \quad (2.7.47)$$

Dengan model  $\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}$  dengan  $\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$  dan  $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$ , nilai harapan

dari jumlah kuadrat residual adalah

$$\begin{aligned}
 \mathbb{E}[\mathbf{e}^\top \mathbf{e}] &= \mathbb{E}[\mathbf{y}^\top (\mathbf{I} - \mathbf{S})^\top (\mathbf{I} - \mathbf{S}) \mathbf{y}] \\
 &= \mathbb{E}[\text{tr}((\mathbf{I} - \mathbf{S}) \mathbf{y} \mathbf{y}^\top (\mathbf{I} - \mathbf{S})^\top)] \\
 &= \text{tr}((\mathbf{I} - \mathbf{S}) \mathbb{E}[\mathbf{y} \mathbf{y}^\top] (\mathbf{I} - \mathbf{S})^\top).
 \end{aligned} \tag{2.7.48}$$

Sebab  $\mathbb{E}[\mathbf{y} \mathbf{y}^\top] = \boldsymbol{\mu} \boldsymbol{\mu}^\top + \sigma^2 \mathbf{I}$ , dengan asumsi bahwa model GWR benar sehingga  $(\mathbf{I} - \mathbf{S}) \boldsymbol{\mu} \approx \mathbf{0}$  (mengabaikan bias), diperoleh

$$\begin{aligned}
 \mathbb{E}[\mathbf{e}^\top \mathbf{e}] &\approx \sigma^2 \text{tr}((\mathbf{I} - \mathbf{S})(\mathbf{I} - \mathbf{S})^\top) \\
 &= \sigma^2 \text{tr}(\mathbf{I} - \mathbf{S} - \mathbf{S}^\top + \mathbf{S}^\top \mathbf{S}) \\
 &= \sigma^2 (n - \text{tr}(\mathbf{S}) - \text{tr}(\mathbf{S}^\top) + \text{tr}(\mathbf{S}^\top \mathbf{S})) \\
 &= \sigma^2 (n - 2\text{tr}(\mathbf{S}) + \text{tr}(\mathbf{S}^\top \mathbf{S})),
 \end{aligned} \tag{2.7.49}$$

dengan menggunakan sifat  $\text{tr}(\mathbf{S}^\top) = \text{tr}(\mathbf{S})$ .

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

$$\nu = n - 2\text{tr}(\mathbf{S}) + \text{tr}(\mathbf{S}^\top \mathbf{S}). \tag{2.7.50}$$

Dengan demikian, estimator tak bias untuk  $\sigma^2$  adalah

$$\hat{\sigma}^2 = \frac{\mathbf{e}^\top \mathbf{e}}{n - 2\text{tr}(\mathbf{S}) + \text{tr}(\mathbf{S}^\top \mathbf{S})}, \tag{2.7.51}$$

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  $n \rightarrow \infty$ . ■

Berdasarkan teorema-teorema di atas, diperoleh hasil penting mengenai kondisi *bandwidth* untuk konsistensi estimator GWR, yaitu *trade-off* antara bias dan va-

riansi.

**Akibat 2.7.13** Mean Squared Error (*MSE*) estimator GWR adalah

$$\text{MSE}(\hat{\beta}(u_i, v_i)) = \text{Bias}^2 + \text{Var} = \mathcal{O}(h^4) + \mathcal{O}((nh^2)^{-1}). \quad (2.7.52)$$

Untuk mencapai konsistensi ( $\text{MSE} \rightarrow 0$  ketika  $n \rightarrow \infty$ ), diperlukan:

- (i)  $h \rightarrow 0$  agar bias  $\mathcal{O}(h^2) \rightarrow 0$ , dan
- (ii)  $nh^2 \rightarrow \infty$  agar variansi  $\mathcal{O}((nh^2)^{-1}) \rightarrow 0$ .

Ini menjadi dasar terbentuknya asumsi GWR.6.

Selanjutnya, analisis asimtotik dari estimator GWR dan uji statistik yang relevan dijelaskan di bawah ini.

**Teorema 2.7.14** Di bawah asumsi GWR.1 sampai GWR.8, estimator GWR bersifat konsisten:

$$\hat{\beta}(u_i, v_i) \xrightarrow{p} \beta(u_i, v_i). \quad (2.7.53)$$

**Bukti.** Konsistensi ditunjukkan dengan membuktikan bahwa  $\text{MSE}(\hat{\beta}(u_i, v_i)) \rightarrow 0$  ketika  $n \rightarrow \infty$ . Dari teorema bias dan variansi yang telah dibuktikan sebelumnya, diperoleh

$$\text{MSE}(\hat{\beta}(u_i, v_i)) = \|\text{Bias}\|^2 + \text{tr}(\text{Var}) = \mathcal{O}(h^4) + \mathcal{O}((nh^2)^{-1}). \quad (2.7.54)$$

Di bawah kondisi  $h \rightarrow 0$  dan  $nh^2 \rightarrow \infty$  ketika  $n \rightarrow \infty$ :

- (i) Suku bias  $\mathcal{O}(h^4) \rightarrow 0$  karena  $h \rightarrow 0$ .
- (ii) Suku variansi  $\mathcal{O}((nh^2)^{-1}) \rightarrow 0$  karena  $nh^2 \rightarrow \infty$ .

Oleh karena itu,  $\text{MSE}(\hat{\beta}(u_i, v_i)) \rightarrow 0$ . Dengan ketaksamaan Chebyshev, untuk setiap  $\varepsilon > 0$ ,

$$P\left(\|\hat{\beta}(u_i, v_i) - \beta(u_i, v_i)\| > \varepsilon\right) \leq \frac{\text{MSE}(\hat{\beta}(u_i, v_i))}{\varepsilon^2} \rightarrow 0, \quad (2.7.55)$$

yang menunjukkan konvergensi dalam probabilitas  $\widehat{\beta}(u_i, v_i) \xrightarrow{p} \beta(u_i, v_i)$ . ■

**Teorema 2.7.15** *Di bawah kondisi yang sama dengan teorema konsistensi, estimator variansi  $\hat{\sigma}^2$  bersifat konsisten:*

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2. \quad (2.7.56)$$

**Bukti.** Estimator variansi adalah  $\hat{\sigma}^2 = \mathbf{e}^\top \mathbf{e} / \nu$  dengan  $\nu = n - 2\text{tr}(\mathbf{S}) + \text{tr}(\mathbf{S}^\top \mathbf{S})$ . Vektor residual dapat ditulis sebagai

$$\mathbf{e} = (\mathbf{I} - \mathbf{S})\mathbf{y} = (\mathbf{I} - \mathbf{S})(\boldsymbol{\mu} + \boldsymbol{\varepsilon}), \quad (2.7.57)$$

dengan  $\boldsymbol{\mu} = \mathbb{E}[\mathbf{y}]$ .

Jumlah kuadrat residual dapat didekomposisi sebagai

$$\mathbf{e}^\top \mathbf{e} = \boldsymbol{\varepsilon}^\top (\mathbf{I} - \mathbf{S})^\top (\mathbf{I} - \mathbf{S}) \boldsymbol{\varepsilon} + 2\boldsymbol{\mu}^\top (\mathbf{I} - \mathbf{S})^\top (\mathbf{I} - \mathbf{S}) \boldsymbol{\varepsilon} + \boldsymbol{\mu}^\top (\mathbf{I} - \mathbf{S})^\top (\mathbf{I} - \mathbf{S}) \boldsymbol{\mu}. \quad (2.7.58)$$

Di bawah kondisi konsistensi, estimator GWR konsisten sehingga berlaku  $(\mathbf{I} - \mathbf{S})\boldsymbol{\mu} \xrightarrow{p} \mathbf{0}$ , yang berarti suku ketiga menjadi hilang. Suku kedua memiliki nilai harapan nol. Untuk suku pertama, dengan hukum bilangan besar,

$$\frac{1}{\nu} \boldsymbol{\varepsilon}^\top (\mathbf{I} - \mathbf{S})^\top (\mathbf{I} - \mathbf{S}) \boldsymbol{\varepsilon} \xrightarrow{p} \sigma^2, \quad (2.7.59)$$

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 2.7.16** *Di bawah asumsi GWR.1 sampai GWR.8 dan dengan bandwidth  $h$  diperlakukan sebagai diketahui (atau diestimasi dengan properti oracle), estimator GWR memiliki distribusi asimtotik:*

$$\sqrt{nh^2} \left( \widehat{\beta}(u_i, v_i) - \beta(u_i, v_i) - \mathbf{b}_i \right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_i), \quad (2.7.60)$$

dengan  $\boldsymbol{\Sigma}_i$  adalah matriks kovarians asimtotik yang bergantung pada  $\sigma^2$ , kepadat-

an spasial  $f(u_i, v_i)$ , dan kekasaran kernel.

Secara praktis, dengan mengabaikan bias (melalui undersmoothing atau koreksi bias), berlaku

$$\hat{\beta}(u_i, v_i) \sim \mathcal{N}(\beta(u_i, v_i), \sigma^2(\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i^2 \mathbf{X} (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1}). \quad (2.7.61)$$

**Bukti.** Estimator GWR dapat dituliskan sebagai

$$\hat{\beta}(u_i, v_i) = \beta(u_i, v_i) + \mathbf{b}_i + (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i \boldsymbol{\varepsilon}, \quad (2.7.62)$$

dengan  $\mathbf{b}_i = \mathcal{O}(h^2)$  adalah bias deterministik. Mendefinisikan

$$\mathbf{Z}_i = (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i \boldsymbol{\varepsilon} = \sum_{j=1}^n c_{ij} \varepsilon_j, \quad (2.7.63)$$

dengan  $c_{ij}$  adalah elemen dari matriks  $\mathbf{C}_i = (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i$ . Sebab  $\varepsilon_j$  independen dengan  $\mathbb{E}[\varepsilon_j] = 0$  dan  $\text{Var}(\varepsilon_j) = \sigma^2$ , maka  $\mathbf{Z}_i$  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  $n_{\text{eff}} = \mathcal{O}(nh^2)$  observasi dengan jarak  $\mathcal{O}(h)$  dari lokasi  $i$ . Variansi dari  $\mathbf{Z}_i$  adalah

$$\text{Var}(\mathbf{Z}_i) = \sigma^2 (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i^2 \mathbf{X} (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} = \mathcal{O}((nh^2)^{-1}). \quad (2.7.64)$$

Dengan normalisasi  $\sqrt{nh^2}$ , diperoleh

$$\sqrt{nh^2} \mathbf{Z}_i = \sqrt{nh^2} \sum_{j=1}^n c_{ij} \varepsilon_j. \quad (2.7.65)$$

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  $\mathcal{O}((nh^2)^{-1})$ , yang menuju nol relatif terhadap total variansi.

Dengan teorema limit pusat Lindeberg-Feller, diperoleh

$$\sqrt{nh^2} \left( \hat{\beta}(u_i, v_i) - \beta(u_i, v_i) - \mathbf{b}_i \right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma_i), \quad (2.7.66)$$

dengan  $\Sigma_i = \lim_{n \rightarrow \infty} nh^2 \cdot \text{Var}(\mathbf{Z}_i)$  adalah matriks kovarians asimtotik. ■

Ketika *bandwidth* diestimasi menggunakan CV, hal yang akan terjadi adalah sebagai berikut.

- (a) Jika  $\hat{h}$  konsisten dan memiliki laju konvergensi  $\hat{h} - h_{\text{opt}} = \mathcal{O}_p(n^{-\alpha})$  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 2.7.17** Untuk menguji  $H_0 : \beta_k(u_i, v_i) = 0$  lawan  $H_1 : \beta_k(u_i, v_i) \neq 0$ , statistik uji adalah

$$t_k(u_i, v_i) = \frac{\hat{\beta}_k(u_i, v_i)}{\text{se}(\hat{\beta}_k(u_i, v_i))}, \quad (2.7.67)$$

dengan

$$\text{se}(\hat{\beta}_k(u_i, v_i)) = \hat{\sigma} \sqrt{\mathbf{e}_{k+1}^\top (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i^2 \mathbf{X} (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{e}_{k+1}}, \quad (2.7.68)$$

dan  $\mathbf{e}_{k+1}$  adalah vektor unit ke- $(k+1)$ . Di bawah  $H_0$ , statistik ini mengikuti distribusi  $t$  dengan derajat bebas efektif  $\nu = n - 2\text{tr}(\mathbf{S}) + \text{tr}(\mathbf{S}^\top \mathbf{S})$ , 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

$$\hat{\beta}_k(u_i, v_i) \sim \mathcal{N}(\beta_k(u_i, v_i), \sigma^2 v_{k,i}), \quad (2.7.69)$$



dengan  $v_{k,i} = \mathbf{e}_{k+1}^\top (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i^2 \mathbf{X} (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{e}_{k+1}$  adalah elemen diagonal ke- $(k+1)$  dari matriks kovarians sandwich.

Di bawah hipotesis nol  $H_0 : \beta_k(u_i, v_i) = 0$ , statistik

$$Z_k = \frac{\hat{\beta}_k(u_i, v_i)}{\sigma \sqrt{v_{k,i}}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (2.7.70)$$

Sebab  $\sigma^2$  tidak diketahui, digunakan estimator konsisten  $\hat{\sigma}^2$ . Dari teorema konsistensi estimator variansi,  $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ . Dengan teorema Slutsky,

$$t_k(u_i, v_i) = \frac{\hat{\beta}_k(u_i, v_i)}{\hat{\sigma} \sqrt{v_{k,i}}} = \frac{Z_k}{\hat{\sigma}/\sigma} \xrightarrow{d} \mathcal{N}(0, 1). \quad (2.7.71)$$

Untuk sampel hingga, aproksimasi distribusi  $t$  dengan derajat bebas  $\nu = n - 2\text{tr}(\mathbf{S}) + \text{tr}(\mathbf{S}^\top \mathbf{S})$  memberikan hasil yang lebih akurat karena memperhitungkan ketidakpastian dalam estimasi  $\sigma^2$ . ■

**Teorema 2.7.18 (Leung, Mei, dan Zhang, 2000)** Untuk menguji  $H_0$  : koefisien stasioner (konstan di semua lokasi) atau  $\beta_k(u_i, v_i) = \beta_k$  lawan  $H_1$  : koefisien bervariasi secara spasial, digunakan statistik  $F$  berikut.

$$F = \frac{\text{RSS}_{\text{GWR}}/\nu_1}{\text{RSS}_{\text{OLS}}/(n-p-1)}, \quad (2.7.72)$$

dengan  $\text{RSS}_{\text{OLS}} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{OLS}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{OLS}})$ ,  $\text{RSS}_{\text{GWR}} = (\mathbf{y} - \mathbf{S}\mathbf{y})^\top (\mathbf{y} - \mathbf{S}\mathbf{y})$ , dan  $\nu_1 = n - 2\text{tr}(\mathbf{S}) + \text{tr}(\mathbf{S}^\top \mathbf{S})$ . Di bawah  $H_0$ , statistik ini mengikuti distribusi  $F$  secara aproksimasi dengan derajat bebas  $r = \frac{\nu_1^2}{\text{tr}[(\mathbf{I}-\mathbf{S})^\top (\mathbf{I}-\mathbf{S})]^2}$  dan  $n-p-1$ .

**Bukti.** Uji ini membandingkan model terbatas (OLS global) dengan model penuh (GWR lokal). Residual kedua model dapat ditulis sebagai

$$\mathbf{e}_{\text{OLS}} = (\mathbf{I} - \mathbf{H})\mathbf{y}, \quad (2.7.73)$$

$$\mathbf{e}_{\text{GWR}} = (\mathbf{I} - \mathbf{S})\mathbf{y}, \quad (2.7.74)$$

dengan  $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$  adalah matriks topi OLS dan  $\mathbf{S}$  adalah matriks penga-

ruh GWR.

Di bawah  $H_0$  bahwa koefisien konstan, kedua model mempunyai nilai harapan yang sama. Perbedaannya adalah GWR menggunakan estimasi lokal yang lebih fleksibel. Sebab GWR lebih fleksibel, RSS dari GWR akan cenderung lebih kecil dari RSS dari OLS, yaitu  $RSS_{GWR} \leq RSS_{OLS}$ .

Matriks residual OLS mempunyai derajat bebas residual  $\nu_{OLS} = n - p - 1$  karena  $\text{tr}(\mathbf{H}) = p + 1$ . Untuk model GWR, meskipun  $\mathbf{S}$  tidak idempoten, derajat bebas residual efektif dapat didekati melalui Satterthwaite correction sebagai:

$$\nu_{GWR} = n - 2\text{tr}(\mathbf{S}) + \text{tr}(\mathbf{S}^\top \mathbf{S}). \quad (2.7.75)$$

Statistik  $F$  yang digunakan membandingkan mean squared error (MSE) kedua model:

$$F = \frac{MSE_{GWR}}{MSE_{OLS}} = \frac{RSS_{GWR}/\nu_{GWR}}{RSS_{OLS}/(n - p - 1)}, \quad (2.7.76)$$

dimana  $\nu_{GWR} = n - 2\text{tr}(\mathbf{S}) + \text{tr}(\mathbf{S}^\top \mathbf{S}) = \nu_1$  sesuai definisi dalam teorema.

Di bawah  $H_0$  dengan  $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ , bentuk  $RSS_{OLS}/\sigma^2$  dan  $RSS_{GWR}/\sigma^2$  masing-masing mengikuti distribusi chi-square yang dimodifikasi. Rasio dari MSE keduanya akan mengikuti distribusi  $F$  secara aproksimasi. Karena bobot lokal GWR menciptakan korelasi residual, derajat bebas efektif pembilang harus dikalibrasi melalui aproksimasi Satterthwaite, menghasilkan derajat bebas aproksimasi  $r = \frac{\nu_1^2}{\text{tr}[(\mathbf{I} - \mathbf{S})^\top (\mathbf{I} - \mathbf{S})]^2}$  untuk pembilang dan derajat bebas penyebut  $n - p - 1$ .

Di bawah  $H_0$ , statistik  $F$  mengikuti distribusi  $F(r, n - p - 1)$  secara aproksimasi. ■

**Teorema 2.7.19 (Leung dkk., 2000)** Untuk menguji apakah koefisien tertentu  $\beta_k$  bervariasi secara spasial dengan  $H_0 : \beta_k(u_i, v_i) = \beta_k(u_j, v_j), \forall i, j, i \neq j$  (konstan) lawan  $H_1 : \exists i, j$  sehingga  $\beta_k(u_i, v_i) \neq \beta_k(u_j, v_j)$ , dapat digunakan statistik uji variabilitas spasial, yaitu

$$V_k^2 = \frac{1}{n} \sum_{i=1}^n \left( \hat{\beta}_k(u_i, v_i) - \bar{\hat{\beta}}_k \right)^2, \quad (2.7.77)$$

dengan  $\bar{\hat{\beta}}_k = \frac{1}{n} \sum_{i=1}^n \hat{\beta}_k(u_i, v_i)$ . Di bawah  $H_0$ , statistik ini mengikuti distribusi  $F$ .

**Bukti.** Definisikan matriks  $\mathbf{J}$  sebagai matriks  $n \times n$  dengan semua elemen bernilai satu, yaitu

$$\mathbf{J} := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Statistik  $V_k^2$  dapat ditulis ulang sebagai

$$V_k^2 = \frac{1}{n} \widehat{\boldsymbol{\beta}}_k^\top \left( \mathbf{I} - \frac{1}{n} \mathbf{J} \right) \widehat{\boldsymbol{\beta}}_k, \quad (2.7.78)$$

dengan  $\widehat{\boldsymbol{\beta}}_k = [\hat{\beta}_k(u_1, v_1), \hat{\beta}_k(u_2, v_2), \dots, \hat{\beta}_k(u_n, v_n)]^\top$ .

Di bawah  $H_0$ , asumsikan bahwa rata-rata dari koefisien lokal adalah konstan, yaitu  $\mathbb{E}[\hat{\beta}_k(u_i, v_i)] = \mu_k$  untuk semua  $i$ . Dengan demikian, vektor  $\mathbb{E}[\widehat{\boldsymbol{\beta}}_k] = \mu_k \mathbf{1}_n$ , dengan  $\mathbf{1}_n$  adalah vektor satuan berukuran  $n$ . Selanjutnya, perhatikan bahwa  $\mathbf{1}_n^\top (\mathbf{I} - \frac{1}{n} \mathbf{J}) = \mathbf{0}^\top$  dan  $(\mathbf{I} - \frac{1}{n} \mathbf{J}) \mathbf{1}_n = \mathbf{0}$ , sehingga statistik  $V_k^2$  dapat diekspresikan sebagai berikut.

$$V_k^2 = \frac{1}{n} (\widehat{\boldsymbol{\beta}}_k - \mu_k \mathbf{1}_n)^\top \left( \mathbf{I} - \frac{1}{n} \mathbf{J} \right) (\widehat{\boldsymbol{\beta}}_k - \mu_k \mathbf{1}_n). \quad (2.7.79)$$

Definisikan  $\boldsymbol{\delta}_k = \widehat{\boldsymbol{\beta}}_k - \mu_k \mathbf{1}_n = [\delta_{k1}, \delta_{k2}, \dots, \delta_{kn}]^\top$  sebagai deviasi dari nilai harapan, dengan  $\delta_{ki} = \hat{\beta}_k(u_i, v_i) - \mu_k$ . Dari teori GWR, di bawah  $H_0$  dengan asumsi normalitas:

$$\boldsymbol{\delta}_k \sim \mathcal{N}(\mathbf{0}, \sigma_k^2 \boldsymbol{\Sigma}_k), \quad (2.7.80)$$

dimana  $\boldsymbol{\Sigma}_k$  adalah matriks kovarians dari estimator lokal dengan elemen diagonal  $\text{Var}(\hat{\beta}_k(u_i, v_i))$  dan elemen off-diagonal merepresentasikan korelasi spasial antar estimator.

Bentuk kuadrat  $\boldsymbol{\delta}_k^\top (\mathbf{I} - \frac{1}{n} \mathbf{J}) \boldsymbol{\delta}_k / \sigma_k^2$  mengikuti distribusi yang kompleks karena matriks  $(\mathbf{I} - \frac{1}{n} \mathbf{J})$  dan  $\boldsymbol{\Sigma}_k$  keduanya non-idempoten. Namun, dengan menggunakan aproksimasi Satterthwaite untuk mencocokkan derajat bebas efektif, statistik

yang dinormalisasi:

$$F_k = \frac{n \cdot V_k^2 / 1}{\hat{\sigma}_k^2 / (n - 2\text{tr}(\mathbf{S}) + \text{tr}(\mathbf{S}^\top \mathbf{S}))}, \quad (2.7.81)$$

akan mengikuti distribusi  $F$  secara aproksimasi. Di sini, pembilang memiliki derajat bebas  $\nu_1 = 1$  (mengukur variabilitas koefisien spesifik  $\beta_k$ ) dan penyebut memiliki derajat bebas  $\nu_2 = n - 2\text{tr}(\mathbf{S}) + \text{tr}(\mathbf{S}^\top \mathbf{S})$  (derajat bebas residual efektif GWR). ■

## 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 (Bishop, 1995; Goodfellow, Bengio, dan Courville, 2016; Haykin, 2009).

### 2.8.1 Model Dasar dan Notasi

**Definisi 2.8.1 (Neuron tiruan, Haykin 2009)** Diberikan  $\mathbf{x} \in \mathbb{R}^p$ , neuron menghasilkan

$$y = \phi(\mathbf{w}^\top \mathbf{x} + b), \quad (2.8.1)$$

dengan  $\mathbf{w} \in \mathbb{R}^p$  adalah bobot,  $b \in \mathbb{R}$  adalah bias, dan  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  adalah fungsi aktivasi nonlinier terdiferensiasi hampir di semua titik.

**Definisi 2.8.2 (Jaringan *feedforward* multilapis Bishop 1995)** Untuk  $L$  lapisan tersembunyi, definisikan rekursi

$$\mathbf{z}^{(l)} = \mathbf{W}^{(l)} \mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}, \quad \mathbf{a}^{(l)} = \phi^{(l)}(\mathbf{z}^{(l)}), \quad l = 1, \dots, L, \quad (2.8.2)$$

dengan  $\mathbf{a}^{(0)} = \mathbf{x}$  dan keluaran akhir  $\hat{y} = g(\mathbf{a}^{(L)})$ . 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*

$$\varphi(z) = \frac{1}{1 + e^{-z}}. \quad (2.8.3)$$

**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*

$$\varphi(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}. \quad (2.8.4)$$

**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, Nair dan Hinton 2010)** *Fungsi ReLU (Rectified Linear Unit) didefinisikan sebagai*

$$\varphi(z) = \max(0, z). \quad (2.8.5)$$

**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*

$$\varphi(z) = \begin{cases} z, & z \geq 0, \\ \alpha z, & z < 0, \end{cases} \quad (2.8.6)$$

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*

$$\varphi(z) = z \cdot \Phi(z), \quad (2.8.7)$$

dengan  $\Phi(z)$  fungsi distribusi normal standar. Aproksimasi praktisnya adalah sebagai berikut.

$$\varphi(z) \approx 0,5z \left( 1 + \tanh \left[ \sqrt{\frac{2}{\pi}} (z + 0,0447z^3) \right] \right). \quad (2.8.8)$$

**Contoh 2.8.13** Untuk  $z = 1$ ,  $\varphi(1) \approx 0,84$ . Untuk  $z = -1$ ,  $\varphi(-1) \approx -0,16$ .

**Lema 2.8.14** *Fungsi aktivasi  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  yang digunakan pada jaringan saraf*

tiruan sebaiknya memenuhi:

- (i)  $\varphi$  non-linear;
- (ii)  $\varphi$  terdiferensiasi hampir di semua titik,
- (iii)  $\varphi$  tidak konstan, dan
- (iv)  $\varphi$  memiliki range memadai, tidak menyebabkan saturasi total.

**Bukti.**

- (i) Jika  $\varphi$  linear, misalnya  $\varphi(z) = az + b$ , maka suatu multilayer perceptron dengan  $L$  lapisan menghasilkan

$$f(\mathbf{x}) = \mathbf{W}^{(L)}\varphi(\mathbf{W}^{(L-1)}\varphi(\dots\varphi(\mathbf{W}^{(1)}\mathbf{x}))) = \mathbf{W}'\mathbf{x} + \mathbf{b}',$$

dengan  $\mathbf{W}' = \prod_{l=1}^L a\mathbf{W}^{(l)}$  dan  $\mathbf{b}'$  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) Algoritma *backpropagation* memerlukan turunan  $\varphi'(z)$  untuk menghitung gradien fungsi kerugian  $L$ . Dengan notasi  $\mathbf{z}^{(l)} = \mathbf{W}^{(l)}\mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}$  dan  $\mathbf{a}^{(l)} = \varphi(\mathbf{z}^{(l)})$ , bentuk umum gradien adalah

$$\frac{\partial L}{\partial \mathbf{W}^{(l)}} = \delta^{(l)} (\mathbf{a}^{(l-1)})^\top, \quad \delta^{(l)} = (\mathbf{W}^{(l+1)})^\top \delta^{(l+1)} \odot \varphi'(\mathbf{z}^{(l)}).$$

Jika  $\varphi'(z)$  tidak terdefinisi pada banyak titik, maka gradien tidak dapat dihitung. Oleh karena itu,  $\varphi$  setidaknya harus terdiferensiasi hampir di semua titik. Fungsi seperti ReLU tetap sah karena ketakdiferensialan hanya terjadi pada  $z = 0$  (himpunan titik ukuran nol).

- (iii) 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) Jika  $\varphi$  cepat jenuh, misalnya sigmoid dengan limit  $\varphi(z) \rightarrow 0$  untuk  $z \rightarrow -\infty$  dan  $\varphi(z) \rightarrow 1$  untuk  $z \rightarrow +\infty$ , 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, Goodfellow dkk. 2016)** Diberikan data terawasi  $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$  dan model prediksi  $\hat{y}_i = f_{\Theta}(\mathbf{x}_i)$ , tujuan pembelajaran adalah meminimalkan

$$L(\Theta) = \frac{1}{n} \sum_{i=1}^n \ell(\hat{y}_i, y_i), \quad (2.8.9)$$

dengan fungsi kerugian  $\ell$  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

$$\ell(\hat{p}, y) = -[y \log \hat{p} + (1 - y) \log(1 - \hat{p})], \quad y \in \{0, 1\}. \quad (2.8.10)$$

Fungsi ini merupakan entropi silang antara distribusi target  $p(y)$  dan prediksi  $\hat{p}$ ,



yaitu

$$\begin{aligned} H(p, \hat{p}) &= - \sum_{y \in \{0,1\}} p(y) \log \hat{p}(y) \\ &= - [p(1) \log \hat{p} + p(0) \log(1 - \hat{p})]. \end{aligned} \quad (2.8.11)$$

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  $\hat{p}_k = \exp(z_k) / \sum_{j=1}^K \exp(z_j)$  dan

$$\ell(\hat{\mathbf{p}}, y) = -\log \hat{p}_y, \quad y \in \{1, \dots, K\}.$$

Fungsi ini juga merupakan entropi silang antara distribusi target  $p(y)$  dan prediksi  $\hat{\mathbf{p}}$ :

$$\begin{aligned} H(p, \hat{\mathbf{p}}) &= - \sum_{k=1}^K p(k) \log \hat{p}_k \\ &= -\log \hat{p}_y, \end{aligned} \quad (2.8.12)$$

dengan  $p(k) = 1$  jika  $k = y$  dan 0 selainnya.

Sebagai contoh, untuk  $K = 3$  kelas, jika  $\mathbf{z} = (2, 0, 1, 0, 1)$ , 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** Misalkan keluaran akhir didefinisikan sebagai komposisi  $y = f^{(L)} \circ f^{(L-1)} \circ \dots \circ f^{(1)}(\mathbf{x})$ . Maka turunan terhadap parameter pada lapisan ke- $l$  meme-

nuhi

$$\frac{\partial y}{\partial \mathbf{W}^{(l)}} = \frac{\partial f^{(L)}}{\partial f^{(L-1)}} \cdots \frac{\partial f^{(l+1)}}{\partial f^{(l)}} \frac{\partial f^{(l)}}{\partial \mathbf{W}^{(l)}}. \quad (2.8.13)$$

**Bukti.** Teorema ini merupakan implikasi langsung dari aturan rantai dalam kalkulus diferensial. ■

Interpretasi teorema ini adalah 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*, Goodfellow dkk. 2016)** Dengan notasi  $\mathbf{z}^{(l)} = \mathbf{W}^{(l)} \mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}$ ,  $\mathbf{a}^{(l)} = \varphi(\mathbf{z}^{(l)})$ , dan  $\ell(\hat{y}, y)$  fungsi kerugian, gradien terhadap bobot  $\mathbf{W}^{(l)}$  diberikan oleh

$$\frac{\partial \ell}{\partial \mathbf{W}^{(l)}} = \delta^{(l)} (\mathbf{a}^{(l-1)})^\top, \quad (2.8.14)$$

dengan  $\delta^{(l)} = (\mathbf{W}^{(l+1)})^\top \delta^{(l+1)} \odot \varphi'(\mathbf{z}^{(l)})$ . Operator  $\odot$  adalah perkalian elemen-per-elemen atau *Hadamard product*.

**Contoh 2.8.21** Misalkan jaringan 2 lapis dengan *input*  $\mathbf{x} = (x_1, x_2)^\top$ , bobot

$$\mathbf{W}^{(1)} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix},$$

bias  $\mathbf{b}^{(1)} = (0, 0)^\top$ , fungsi aktivasi  $\varphi(z) = \tanh(z)$ , dan lapisan *output* linear  $y = \mathbf{w}^{(2)\top} \mathbf{a}^{(1)}$  dengan  $\mathbf{w}^{(2)} = (1, 1)^\top$ . Target  $y^* = 1$ , kerugian  $\ell(y, y^*) = \frac{1}{2}(y - y^*)^2$ .

(a) *Forward:*

$$\begin{aligned}\mathbf{z}^{(1)} &= \mathbf{W}^{(1)} \mathbf{x} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ \mathbf{a}^{(1)} &= \tanh(\mathbf{z}^{(1)}) \\ y &= (1, 1) \cdot \mathbf{a}^{(1)}\end{aligned}$$

(b) *Backward:*

$$\begin{aligned}\delta^{(2)} &= \frac{\partial \ell}{\partial y} = y - y^* \\ \delta^{(1)} &= \mathbf{w}^{(2)} \cdot \delta^{(2)} \odot \varphi'(\mathbf{z}^{(1)}) \\ \frac{\partial \ell}{\partial \mathbf{W}^{(1)}} &= \delta^{(1)}(\mathbf{x})^\top\end{aligned}$$

Dengan substitusi nilai, gradien dapat dihitung secara eksplisit untuk setiap parameter.

### 2.8.5 Optimisasi Parameter

Dalam pelatihan jaringan saraf, parameter  $\Theta$  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

$$L(\Theta) = \frac{1}{n} \sum_{i=1}^n \ell(f_{\Theta}(\mathbf{x}_i), y_i),$$

pembaruan parameter dengan laju belajar  $\eta > 0$  adalah

$$\Theta_{t+1} = \Theta_t - \eta \nabla_{\Theta} L(\Theta_t). \quad (2.8.15)$$

Metode ini konvergen ke minimum lokal untuk fungsi non-konveks seperti jaringan

saraf.

Algoritma 2.1 merangkum prosedur *gradient descent*.

---

**Algoritma 2.1 Gradient Descent**

---

**Require:** Fungsi kerugian  $L(\Theta)$ , data pelatihan  $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ , laju belajar  $\eta > 0$ , jumlah epoch maksimal  $T_{\max}$ , toleransi konvergensi  $\varepsilon > 0$

**Ensure:** Parameter optimal  $\Theta^*$

```

1: Inisialisasi parameter  $\Theta_0$  secara acak
2:  $t \leftarrow 0$ 
3: while  $t < T_{\max}$  do
4:   Hitung gradien:  $\mathbf{g}_t = \nabla_{\Theta} L(\Theta_t)$ 
5:   Perbarui parameter:  $\Theta_{t+1} \leftarrow \Theta_t - \eta \mathbf{g}_t$ 
6:   if  $\|\mathbf{g}_t\| < \varepsilon$  or  $|L(\Theta_{t+1}) - L(\Theta_t)| < \varepsilon$  then
7:     break
8:   end if
9:    $t \leftarrow t + 1$ 
10: end while
11: return  $\Theta^* \leftarrow \Theta_t$ 

```

---

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  $\eta$  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 non-konveks 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.*

$$\Theta_{t+1} = \Theta_t - \eta \frac{1}{|\mathcal{B}_t|} \sum_{(\mathbf{x}_i, y_i) \in \mathcal{B}_t} \nabla_{\Theta} \ell(f_{\Theta}(\mathbf{x}_i), y_i), \quad (2.8.16)$$

dengan  $\mathcal{B}_t$  adalah batch kecil pada iterasi ke- $t$ ,  $\eta$  adalah laju pembelajaran, dan  $\ell$  adalah fungsi kerugian. Metode ini sangat efisien untuk data berukuran besar dan memberikan regularisasi implisit melalui pengacakan batch.

Algoritma 2.2 merangkum prosedur *stochastic gradient descent*.

---

**Algoritma 2.2 Stochastic Gradient Descent (SGD)**


---

**Require:** Fungsi kerugian  $L(\Theta)$ , data pelatihan  $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ , laju belajar  $\eta > 0$ , ukuran *batch*  $m$ , jumlah epoch maksimal  $T_{\max}$ , toleransi konvergensi  $\varepsilon > 0$

**Ensure:** Parameter optimal  $\Theta^*$

```

1: Inisialisasi parameter  $\Theta_0$  secara acak
2:  $t \leftarrow 0$ 
3: while  $t < T_{\max}$  do
4:   Acak urutan indeks data:  $\pi = \text{permutasi}([1, 2, \dots, n])$ 
5:   for  $i = 0$  to  $n - m$  step  $m$  do
6:     Ambil batch:  $\mathcal{B}_t = \{\pi[i], \pi[i+1], \dots, \pi[i+m-1]\}$ 
7:     Hitung gradien batch:  $\mathbf{g}_t = \frac{1}{m} \sum_{j \in \mathcal{B}_t} \nabla_{\Theta} \ell(f_{\Theta}(\mathbf{x}_j), y_j)$ 
8:     Perbarui parameter:  $\Theta \leftarrow \Theta - \eta \mathbf{g}_t$ 
9:   end for
10:  Hitung nilai kerugian keseluruhan:  $L_t \leftarrow \frac{1}{n} \sum_{i=1}^n \ell(f_{\Theta}(\mathbf{x}_i), y_i)$ 
11:  if  $|L_t - L_{t-1}| < \varepsilon$  then
12:    break
13:  end if
14:   $t \leftarrow t + 1$ 
15: end while
16: return  $\Theta^* \leftarrow \Theta$ 

```

---

**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  $\mathcal{B}_t = \{i_1, \dots, i_{10}\}$ .
- (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  $\eta$ :

$$w_{t+1} = w_t - \eta(w_t - 3)$$

(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  $t \geq 1$ :

$$\mathbf{m}_t = \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1) \mathbf{g}_t, \quad (2.8.17)$$

$$\mathbf{v}_t = \beta_2 \mathbf{v}_{t-1} + (1 - \beta_2) (\mathbf{g}_t \odot \mathbf{g}_t), \quad (2.8.18)$$

$$\hat{\mathbf{m}}_t = \frac{\mathbf{m}_t}{1 - \beta_1^t}, \quad \hat{\mathbf{v}}_t = \frac{\mathbf{v}_t}{1 - \beta_2^t}, \quad (2.8.19)$$

$$\Theta_{t+1} = \Theta_t - \eta \frac{\hat{\mathbf{m}}_t}{\sqrt{\hat{\mathbf{v}}_t} + \varepsilon}. \quad (2.8.20)$$

Parameter umum yang digunakan adalah  $\beta_1 = 0,9$ ,  $\beta_2 = 0,999$ ,  $\varepsilon = 10^{-8}$ .

Di bawah ini adalah Algoritma 2.3 yang merangkum prosedur Adam.

---

**Algoritma 2.3 Adaptive Moment Estimation (Adam)**


---

**Require:** Fungsi kerugian  $L(\Theta)$ , data pelatihan  $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ , laju belajar  $\eta > 0$ , parameter momentum  $\beta_1 = 0.9$ , parameter momentum orde-2  $\beta_2 = 0.999$ , stabilisasi  $\varepsilon = 10^{-8}$ , jumlah iterasi maksimal  $T_{\max}$

**Ensure:** Parameter optimal  $\Theta^*$

- 1: Inisialisasi parameter  $\Theta_0$  secara acak
  - 2: Inisialisasi momen pertama:  $\mathbf{m}_0 \leftarrow \mathbf{0}$
  - 3: Inisialisasi momen kedua:  $\mathbf{v}_0 \leftarrow \mathbf{0}$
  - 4:  $t \leftarrow 0$
  - 5: **while**  $t < T_{\max}$  **do**
  - 6:     Hitung gradien:  $\mathbf{g}_t = \nabla_{\Theta} L(\Theta_t)$
  - 7:     Perbarui momen pertama:  $\mathbf{m}_t \leftarrow \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1) \mathbf{g}_t$
  - 8:     Perbarui momen kedua:  $\mathbf{v}_t \leftarrow \beta_2 \mathbf{v}_{t-1} + (1 - \beta_2) (\mathbf{g}_t \odot \mathbf{g}_t)$
  - 9:     Koreksi bias momen pertama:  $\hat{\mathbf{m}}_t \leftarrow \mathbf{m}_t / (1 - \beta_1^{t+1})$
  - 10:    Koreksi bias momen kedua:  $\hat{\mathbf{v}}_t \leftarrow \mathbf{v}_t / (1 - \beta_2^{t+1})$
  - 11:    Perbarui parameter:  $\Theta_t \leftarrow \Theta_t - \eta \frac{\hat{\mathbf{m}}_t}{\sqrt{\hat{\mathbf{v}}_t} + \varepsilon}$
  - 12:    **if**  $\|\mathbf{g}_t\| < 10^{-6}$  **or**  $t \equiv 0 \pmod{100}$  dan konvergen **then**
  - 13:       **break**
  - 14:    **end if**
  - 15:     $t \leftarrow t + 1$
  - 16: **end while**
  - 17: **return**  $\Theta^* \leftarrow \Theta_t$
- 

**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 (Cybenko, 1989).

Pertanyaan fundamental dalam jaringan saraf adalah apakah kelas fungsi yang dihasilkan oleh suatu arsitektur memiliki daya representasi yang cukup kaya atau tidak. Secara khusus bagian ini akan menyajikan jaringan saraf dengan fungsi aktivasi sigmoid karena aplikasinya secara khusus pada bab selanjutnya. Untuk menjawab pertanyaan tersebut, akan dipelajari sifat-sifat aproksimasi dari fungsi yang berbentuk

$$G(\mathbf{x}) = \sum_{j=1}^N \alpha_j \sigma(\mathbf{w}_j^\top \mathbf{x} + b_j), \quad (2.8.21)$$

dengan  $\sigma(z) = \frac{1}{1+e^{-z}}$  adalah fungsi sigmoid standar,  $\alpha_j \in \mathbb{R}$ ,  $\mathbf{w}_j \in \mathbb{R}^p$ , dan  $b_j \in \mathbb{R}$ .

**Teorema 2.8.28 (Teorema aproksimasi universal (UAT), Cybenko 1989)** *Misalkan  $\sigma : \mathbb{R} \rightarrow \mathbb{R}$  adalah fungsi sigmoid kontinu, yaitu memenuhi*

$$\lim_{t \rightarrow -\infty} \sigma(t) = 0, \quad \lim_{t \rightarrow +\infty} \sigma(t) = 1, \quad (2.8.22)$$

maka kelas fungsi

$$\mathcal{G} = \left\{ \sum_{j=1}^N \alpha_j \sigma(\mathbf{w}_j^\top \mathbf{x} + b_j) : N \in \mathbb{N}, \alpha_j \in \mathbb{R}, \mathbf{w}_j \in \mathbb{R}^p, b_j \in \mathbb{R} \right\} \quad (2.8.23)$$

adalah dense atau padat di ruang kontinu  $C([0, 1]^p)$  terhadap norma supremum, yaitu untuk setiap fungsi kontinu  $f \in C([0, 1]^p)$  dan setiap  $\varepsilon > 0$ , terdapat fungsi  $G \in \mathcal{G}$  sehingga

$$\|f - G\|_\infty = \sup_{\mathbf{x} \in [0, 1]^p} |f(\mathbf{x}) - G(\mathbf{x})| < \varepsilon. \quad (2.8.24)$$

**Bukti.** Bukti dilakukand dengan pendekatan kontradiksi dan didasarkan pada argumen pemisahan linear. Andaikan bahwa kelas fungsi  $\mathcal{G}$  tidak padat di  $C([0, 1]^p)$ . Maka terdapat fungsi kontinu  $f \in C([0, 1]^p)$  yang tidak dapat didekati secara seragam oleh fungsi-fungsi dalam  $\mathcal{G}$ . Secara konseptual, hal ini berarti terdapat suatu alat ukur yang mampu membedakan  $f$  dari semua fungsi dalam  $\mathcal{G}$ .

Secara formal, asumsi tersebut menyiratkan adanya fungsi linear kontinu tak nol  $L : C([0, 1]^p) \rightarrow \mathbb{R}$  sedemikian hingga

$$L(G) = 0 \quad \forall G \in \mathcal{G}, \quad \text{tetapi} \quad L(f) \neq 0. \quad (2.8.25)$$

Fungsi tersebut kontinu pada ruang fungsi kontinu dan dapat direpresentasikan sebagai integral terhadap suatu ukuran bertanda-hingga  $\mu$ . Dengan demikian, fungsi tersebut dapat didefinisikan dalam integral Lebesgue-Stieltjes:

$$L(h) = \int_{[0, 1]^p} h(\mathbf{x}) d\mu(\mathbf{x}) \quad \forall h \in C([0, 1]^p). \quad (2.8.26)$$

Sebab  $L(G) = 0$  untuk semua  $G \in \mathcal{G}$ , maka

$$\int_{[0, 1]^p} \sigma(\mathbf{w}^\top \mathbf{x} + b) d\mu(\mathbf{x}) = 0 \quad \forall \mathbf{w} \in \mathbb{R}^p, b \in \mathbb{R}. \quad (2.8.27)$$

Selanjutnya, dengan memanfaatkan sifat batas dari fungsi sigmoid, untuk

suatu konstanta skala  $\lambda > 0$ , perhatikan fungsi

$$\sigma_\lambda(\mathbf{w}^\top \mathbf{x} + b) = \sigma(\lambda(\mathbf{w}^\top \mathbf{x} + b)). \quad (2.8.28)$$

Ketika  $\lambda \rightarrow +\infty$ , fungsi ini mendekati fungsi indikator

$$\mathbb{I}_{\{\mathbf{x}:\mathbf{w}^\top \mathbf{x}+b \geq 0\}}(\mathbf{x}) = \begin{cases} 1, & \text{jika } \mathbf{w}^\top \mathbf{x} + b \geq 0, \\ 0, & \text{jika } \mathbf{w}^\top \mathbf{x} + b < 0. \end{cases} \quad (2.8.29)$$

Oleh karena itu, dengan menerapkan Teorema Dominasi Konvergensi Lebesgue yang menyatakan bahwa batas integral dapat ditukar dengan integral batas untuk fungsi yang terukur dan terdominasi, diperoleh

$$\int_{[0,1]^p} \mathbb{I}_{\{\mathbf{x}:\mathbf{w}^\top \mathbf{x}+b \geq 0\}}(\mathbf{x}) d\mu(\mathbf{x}) = 0 \quad \forall \mathbf{w} \in \mathbb{R}^p, b \in \mathbb{R}. \quad (2.8.30)$$

Persamaan ini menyiratkan bahwa ukuran bertanda  $\mu$  dari setiap himpunan separasi linear adalah nol. Namun, himpunan separasi linear ini membentuk basis untuk ruang Borel pada  $[0, 1]^p$ . Oleh karena itu, ukuran bertanda  $\mu$  haruslah nol pada semua himpunan Borel di  $[0, 1]^p$ , yang bertentangan dengan asumsi awal bahwa  $L(f) \neq 0$ . Dengan demikian, asumsi bahwa kelas fungsi  $\mathcal{G}$  tidak padat di  $C([0, 1]^p)$  adalah salah, sehingga teorema ini terbukti. ■

Teorema ini menunjukkan bahwa jaringan saraf dengan satu lapisan tersembunyi yang cukup besar dapat mengaproksimasi fungsi kontinu pada himpunan kompak dengan presisi setinggi apapun. Meskipun demikian, teorema ini tidak memberikan petunjuk tentang berapa banyak neuron yang diperlukan untuk mencapai aproksimasi tertentu, atau bagaimana cara melatih jaringan tersebut secara efektif. Selain itu, teorema ini berlaku untuk fungsi kontinu pada himpunan kompak, sehingga tidak mencakup semua jenis fungsi yang mungkin dihadapi dalam praktik.

## 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  $\mathcal{E}$ .

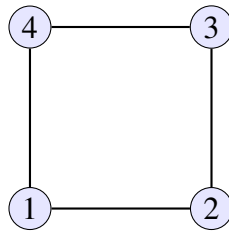
### 2.9.1 Dasar Graf dan Laplacian

**Definisi 2.9.1 (Graf dan matriks ketetanggaan)** Graf  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  dengan  $|\mathcal{V}| = n$  direpresentasikan oleh:

$$\mathbf{A}_{ij} = \begin{cases} 1 & \text{jika } (i, j) \in \mathcal{E}, \\ 0 & \text{lainnya,} \end{cases} \quad (2.9.1)$$

disebut matriks ketetanggaan. Matriks derajat didefinisikan sebagai  $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$ , 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  $\mathbf{D} = \text{diag}(2, 2, 2, 2)$ .

**Definisi 2.9.3 (Graf Laplacian, Chung 1997)** *Graf Laplacian adalah  $\mathbf{L} = \mathbf{D} - \mathbf{A}$ . Normalisasi simetris diberikan oleh*

$$\mathbf{L}_{sym} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}. \quad (2.9.2)$$

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

$$\mathbf{D} = \text{diag}(2, 2, 2, 2).$$

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

$$\mathbf{L}_{sym} = \mathbf{I} - \frac{1}{2} \mathbf{A},$$

karena  $\mathbf{D}^{-1/2} = \text{diag}(1/\sqrt{2}, 1/\sqrt{2}, 1/\sqrt{2}, 1/\sqrt{2})$ .

**Teorema 2.9.5 (Teorema spektral Laplacian, Chung 1997)** *Matriks Laplacian*

$\mathbf{L}$  adalah simetris dan positif semidefinit, sehingga dapat didekomposisi dengan dekomposisi nilai eigen sebagai berikut.

$$\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top, \quad (2.9.3)$$

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)*, Bronstein, Bruna, Cohen, dan Veličković 2021) 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  $\mathbf{x}_u$  adalah fitur simpul  $u$  dan  $\mathbf{e}_{uv}$  adalah fitur sisi  $(u, v) \in \mathcal{E}$ . Suatu lapisan MPNN didefinisikan sebagai berikut.

$$\mathbf{h}_u = \varphi \left( \mathbf{x}_u, \bigoplus_{v \in \mathcal{N}_u} \psi(\mathbf{x}_u, \mathbf{x}_v, \mathbf{e}_{uv}) \right) \quad (2.9.4)$$

dengan  $\varphi$  dan  $\psi$  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  $\psi$  dapat dirancang untuk menangkap interaksi spesifik antara simpul dan sisi, sedangkan fungsi  $\varphi$  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)*, Kipf dan Welling 2017)

GCN mendefinisikan operasi propagasi lapisan ke- $l$  sebagai

$$\mathbf{H}^{(l+1)} = \sigma\left(\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2} \mathbf{H}^{(l)} \mathbf{W}^{(l)}\right), \quad (2.9.5)$$

dengan  $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_n$  matriks ketetanggaan dengan self-loop,  $\tilde{\mathbf{D}} = \text{diag}(\sum_j \tilde{A}_{ij})$  matriks derajat dari  $\tilde{\mathbf{A}}$ ,  $\mathbf{H}^{(l)} \in \mathbb{R}^{n \times d_l}$  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.

Secara eksplisit, propagasi GCN dapat ditulis sebagai

$$\mathbf{h}_u^{(l+1)} = \sigma\left(\sum_{v \in \mathcal{N}_u \cup \{u\}} \frac{1}{\sqrt{d_u d_v}} \mathbf{h}_v^{(l)} \mathbf{W}^{(l)}\right), \quad (2.9.6)$$

dengan  $d_u$  derajat simpul  $u$  dan  $\sigma$  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.

GCN merupakan kasus khusus dari kerangka MPNN dengan fitur sisi diabaikan ( $\mathbf{e}_{uv}$  tidak digunakan), fungsi pesan  $\psi(\mathbf{x}_u, \mathbf{x}_v, \mathbf{e}_{uv}) = \mathbf{x}_v$  (hanya mengambil fitur tetangga, operator agregasi  $\oplus$  berupa penjumlahan atau rata-rata tertimbang (melalui normalisasi Laplacian), dan fungsi update  $\varphi$  adalah komposisi linear dan aktivasi non-linear:  $\varphi(\mathbf{x}_u, \text{AGG}) = \sigma(\text{AGG} \cdot \mathbf{W})$ .

**Definisi 2.9.8** (*Graph attention networks (GAT)*, Veličković dkk. 2018) GAT mengganti normalisasi derajat dengan mekanisme perhatian. Koefisien perhatian untuk sisi  $(i, j)$  adalah

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^\top [\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_j]))}{\sum_{k \in \mathcal{N}(i)} \exp(\text{LeakyReLU}(\mathbf{a}^\top [\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_k]))}, \quad (2.9.7)$$

dengan  $\mathbf{h}_i \in \mathbb{R}^d$  vektor fitur simpul  $i$ ,  $\mathbf{W}$  bobot transformasi linear,  $\mathbf{a} \in \mathbb{R}^{2d}$  vektor bobot perhatian, dan  $\parallel$  operator konkatenasi.

Pembaruan fitur simpul GAT adalah

$$\mathbf{h}'_i = \sigma \left( \sum_{j \in \mathcal{N}(i)} \alpha_{ij} \mathbf{W} \mathbf{h}_j \right). \quad (2.9.8)$$

GAT merupakan kasus khusus dari kerangka MPNN dengan spesifikasi berikut.

- Fitur sisi  $\mathbf{e}_{uv}$  tidak digunakan (atau dapat diabaikan).
- Fungsi pesan  $\psi(\mathbf{x}_u, \mathbf{x}_v, \mathbf{e}_{uv}) = \mathbf{W} \mathbf{x}_v$  (transformasi linear fitur tetangga).
- Operator agregasi  $\oplus$  berupa penjumlahan tertimbang, dengan bobot agregasi  $\alpha_{uv}$  diperoleh dari mekanisme perhatian (*attention*) yang bersifat *permutation invariant*.
- Fungsi update  $\varphi$  adalah komposisi linear dan aktivasi non-linear:  $\varphi(\mathbf{x}_u, \text{AGG}) = \sigma(\text{AGG})$ .

Secara formal, pembaruan fitur simpul  $u$  pada GAT dapat ditulis dalam notasi MPNN sebagai

$$\mathbf{h}'_u = \sigma \left( \sum_{v \in \mathcal{N}_u} \alpha_{uv} \psi(\mathbf{x}_u, \mathbf{x}_v, \mathbf{e}_{uv}) \right) = \sigma \left( \sum_{v \in \mathcal{N}_u} \alpha_{uv} \mathbf{W} \mathbf{x}_v \right), \quad (2.9.9)$$

dengan  $\alpha_{uv}$  adalah skor perhatian yang bergantung pada fitur  $\mathbf{x}_u$  dan  $\mathbf{x}_v$  melalui fungsi  $\text{LeakyReLU}(\mathbf{a}^\top [\mathbf{W} \mathbf{x}_u \parallel \mathbf{W} \mathbf{x}_v])$  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, Hamilton, Ying, dan Leskovec 2017)** *GraphSAGE* atau Graph Sample and Aggregation mendefinisikan pembaruan simpul  $i$  sebagai

$$\mathbf{h}_i^{(l+1)} = \sigma \left( \mathbf{W}^{(l)} \cdot \text{AGGREGATE} \left( \{\mathbf{h}_i^{(l)}\} \cup \{\mathbf{h}_j^{(l)} : j \in \mathcal{N}(i)\} \right) \right), \quad (2.9.10)$$

dengan AGGREGATE fungsi agregasi non-linear yang bisa berupa mean, max-pooling, atau LSTM.



### 2.9.4 Sifat Simetri dan Invariansi Permutasi

Salah satu sifat fundamental yang membedakan Jaringan Saraf Graf dari jaringan saraf konvensional adalah kemampuannya untuk menghormati simetri graf. Secara khusus, representasi graf tidak bergantung pada pelabelan simpul, sehingga model yang sah harus bersifat *permutation equivariant* atau *permutation invariant*.

**Definisi 2.9.10 (*Permutation equivariance*)** Misalkan  $\mathbf{P} \in \mathbb{R}^{n \times n}$  adalah matriks permutasi. Suatu pemetaan  $\mathcal{F} : (\mathbf{A}, \mathbf{X}) \mapsto \mathbf{H}$  disebut *permutation equivariant* jika

$$\mathcal{F}(\mathbf{PAP}^\top, \mathbf{PX}) = \mathbf{P} \mathcal{F}(\mathbf{A}, \mathbf{X}). \quad (2.9.11)$$

Kerangka *message passing neural networks* secara inheren memenuhi sifat ini karena operator agregasi yang digunakan bersifat komutatif dan asosiatif. Oleh karena itu, keluaran GNN tidak bergantung pada urutan simpul, melainkan hanya pada struktur graf itu sendiri, memastikan bahwa representasi yang dipelajari adalah sah dan bermakna dalam konteks graf.

### 2.9.5 Daya Aproksimasi Jaringan Saraf Graf

Sejumlah hasil teoretis dalam literatur menunjukkan bahwa kerangka GNN memiliki sifat aproksimasi universal pada kelas fungsi graf tertentu. Secara khusus, Brüel-Gabrielsson (2020) menunjukkan bahwa GNN yang dibangun dalam kerangka *message passing*, dengan operator agregasi yang bersifat *permutation invariant* dan fungsi pembaruan yang cukup kaya (misalnya *multilayer perceptron*), mampu mengaproksimasi setiap fungsi graf yang kontinu dan invariant terhadap permutasi simpul.

Secara matematis, hasil tersebut dapat dirumuskan sebagai berikut. Misalkan  $\mathcal{G}_n$  adalah himpunan graf beratribut dengan jumlah simpul terbatas  $n$ , dan misalkan  $f : \mathcal{G}_n \rightarrow \mathbb{R}$  adalah fungsi graf yang kontinu dan invariant terhadap permutasi simpul, yaitu

$$f(\mathbf{PAP}^\top, \mathbf{PX}) = f(\mathbf{A}, \mathbf{X}) \quad \text{untuk setiap matriks permutasi } \mathbf{P}.$$

Maka, untuk setiap  $\varepsilon > 0$ , terdapat parameter  $\theta$  dan arsitektur GNN sedemikian sehingga

$$\sup_{(\mathbf{A}, \mathbf{X}) \in \mathcal{G}_n} |f(\mathbf{A}, \mathbf{X}) - \text{GNN}_\theta(\mathbf{A}, \mathbf{X})| < \varepsilon.$$

Lebih lanjut, Xu, Hu, Leskovec, dan Jegelka (2019) menunjukkan bahwa daya aproksimasi GNN berbasis *message passing* dibatasi oleh kemampuan uji isomorfisme graf Weisfeiler–Lehman tingkat pertama (1-WL). Tes Weisfeiler–Lehman diperkenalkan sebagai prosedur efisien untuk membedakan graf berdasarkan struktur lokalnya, sebagai pendekatan praktis terhadap masalah graf isomorfisme yang secara teoretis sulit. Meskipun tidak memberikan solusi lengkap terhadap isomorfisme graf, tes WL cukup kuat untuk membedakan sebagian besar graf dalam praktik. Dalam konteks GNN berbasis *message passing*, daya representasi model secara teoretis dibatasi oleh kemampuan tes WL tingkat pertama. Oleh karena itu, keterbatasan ini bukan merupakan kelemahan implementasi, melainkan konsekuensi struktural dari desain agregasi lokal yang digunakan.

## BAB III

# KERANGKA METODOLOGIS DAN ANALISIS ASIMTOTIK

## *GRAPH ATTENTION-BASED GEOGRAPHICALLY*

## *WEIGHTED REGRESSION*

Pada bab ini diuraikan metode *Graph Attention-based Geographically Weighted Regression* (GA-GWR), sebuah ekstensi dari model *Geographically Weighted Regression* (GWR) yang menggunakan jaringan saraf graf untuk mempelajari fungsi *kernel* spasial secara adaptif dengan struktur pembobotan berbasis konsep perhatian (*attention*). Pembahasan dimulai dengan analisis keterbatasan pembobotan spasial berbasis *kernel* konvensional, dilanjutkan dengan konstruksi *kernel* terpelajari berbasis jaringan saraf graf, integrasi ke dalam kerangka GWR, serta analisis asimtotik dari estimator yang dihasilkan.

### 3.1 Keterbatasan Pembobotan Spasial Berbasis *Kernel* Konvensional

Sebagaimana telah diuraikan pada Bab II, model GWR menggunakan pendekatan pembobotan spasial yang mengandalkan fungsi *kernel* tetap (*fixed kernel*) untuk menangkap struktur dependensi spasial dalam data. Meskipun pendekatan ini telah banyak digunakan dalam berbagai aplikasi, terdapat beberapa keterbatasan fundamental yang berdampak pada akurasi estimasi dan validitas inferensi statistik. Pada bagian ini dibahas prinsip pembobotan spasial dalam GWR klasik, sumber bias induktif yang muncul akibat penggunaan *kernel* tetap, serta implikasinya terhadap inferensi statistik.

#### 3.1.1 Prinsip Pembobotan Spasial dalam GWR

Model GWR menggunakan estimator koefisien lokal yang diperoleh melalui metode *locally weighted least squares* (LWLS) dengan matriks bobot spasial  $W(u_0, v_0)$ . Fungsi bobot pada model GWR ditentukan melalui fungsi *kernel*, di

mana bobot antara observasi  $i$  dan lokasi target  $(u_0, v_0)$  diberikan oleh

$$w_i(u_0, v_0) = K\left(\frac{d_i(u_0, v_0)}{h}\right), \quad (3.1.1)$$

dengan  $d_i(u_0, v_0)$  adalah jarak 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* (Fan dan Gijbels, 1996). Fungsi *kernel* dengan karakteristik demikian selanjutnya disebut sebagai *kernel* spasial konvensional. Fungsi *kernel*  $K : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  dalam GWR harus memenuhi beberapa sifat matematis sesuai dengan Asumsi GWR.8.

Dari definisi di atas, terlihat bahwa *kernel* konvensional memiliki dua karakteristik utama:

- (i) Bobot spasial sepenuhnya ditentukan oleh jarak geografis antara lokasi.
- (ii) Parameter *bandwidth*  $h$  bersifat tetap untuk seluruh wilayah studi.

Kedua karakteristik ini membawa implikasi penting terhadap fleksibilitas representasi struktur dependensi spasial. Kelemahan fundamental dari formulasi di atas adalah bahwa bentuk fungsional  $K(\cdot)$  dan metrik  $d_{ij}$  ditentukan secara *a priori* (sebelum analisis data) dan bersifat kaku. Peneliti memaksakan asumsi bahwa interaksi spasial di dunia nyata mengikuti kurva lonceng Gaussian atau peluruhan polinomial Bi-square yang mulus. Dalam terminologi *machine learning*, hal ini merupakan bentuk bias induktif (*inductive bias*) yang sangat kuat.

### 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. Secara matematis, ini berarti  $K(\|\mathbf{u}_i - \mathbf{u}_j\|)$  hanya bergantung pada magnitudo vektor jarak, bukan arahnya. Geometri dari *kernel* ini adalah lingkaran (dalam 2D) atau bola (dalam 3D). Implikasinya adalah bahwa pengaruh spasial meluruh dengan laju yang sama ke segala arah, yaitu Utara, Selatan, Timur, dan Barat.

**Definisi 3.1.1 (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 (Yan, Wu, dan Duan, 2024).

Secara matematis, misalkan  $\kappa^*(\mathbf{u}_i, \mathbf{u}_0)$  menyatakan *kernel* spasial sejati yang mendasari proses data. Penggunaan *kernel* konvensional menyiratkan aproksimasi

$$\kappa^*(\mathbf{u}_i, \mathbf{u}_0) \approx K\left(\frac{\|\mathbf{u}_i - \mathbf{u}_0\|}{h}\right), \quad (3.1.2)$$

dengan  $K(\cdot)$  dipilih dari kelas fungsi terbatas. Jika  $\kappa^*$  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 (Fotheringham, Yang, dan Kang, 2017).

Parameter  $d_{ij}$  dalam kernel GWR hampir secara eksklusif dihitung sebagai jarak Euclidean garis lurus. Asumsi ini mengimplikasikan ruang yang kontinu, homogen, dan bebas hambatan. Namun, ruang geografis penuh dengan diskontinuitas dan hambatan topologis. Sungai, pegunungan, atau batas administratif dapat memutus interaksi spasial antara dua titik yang secara Euclidean sangat dekat. Dalam GWR konvensional, dua rumah yang berseberangan di tepi sungai tanpa jembatan akan dianggap sebagai tetangga dekat dengan bobot interaksi tinggi ( $w_{ij} \approx 1$ ). Padahal, dalam realitas pasar perumahan atau akses layanan, jarak efektif mereka (jarak jaringan) mungkin sangat jauh. Penggunaan metrik jarak Euclidean mengabaikan faktor-faktor ini, sehingga menghasilkan bobot spasial yang tidak akurat dan bias dalam estimasi koefisien lokal (Lu, Charlton, Harris, dan Fotheringham, 2014). Dalam studi transportasi atau epidemiologi, interaksi terjadi melalui jaringan

jalan atau rute penerbangan. Penelitian menunjukkan bahwa kalibrasi GWR menggunakan jarak jaringan menghasilkan goodness-of-fit yang jauh lebih baik daripada jarak Euclidean. Namun, integrasi jarak non-Euclidean ke dalam GWR konvensional memerlukan perhitungan matriks jarak eksternal yang berat secara komputasi dan statis (tidak dipelajari dari data) (Lu dkk., 2014).

### 3.2 Pembelajaran *Kernel* Spasial dengan Jaringan Saraf Graf

Untuk mengatasi keterbatasan pembobotan spasial berbasis *kernel* konvensional, pendekatan pembelajaran mesin modern menawarkan mekanisme yang lebih fleksibel dalam mempelajari struktur dependensi spasial secara langsung dari data. Salah satu pendekatan yang menonjol adalah penggunaan jaringan saraf graf (*Graph Neural Networks*, GNN), yang dirancang untuk memodelkan relasi antar-unit pengamatan yang direpresentasikan sebagai graf (Kipf dan Welling, 2017; Wu dkk., 2021). Berbeda dengan *kernel* spasial tetap yang ditentukan *a priori*, GNN memungkinkan pembelajaran bobot spasial yang adaptif terhadap struktur graf dan atribut data, sehingga memperluas kelas fungsi *kernel* yang dapat direpresentasikan secara *data-driven*. Model GWR yang menggunakan GNN untuk mempelajari *kernel* spasial selanjutnya disebut sebagai *Graph Attention-based Geographically Weighted Regression* (GA-GWR).

#### 3.2.1 Representasi Graf untuk Data Spasial

Misalkan tersedia sampel acak  $\{(y_i, \mathbf{x}_i, \mathbf{u}_i)\}_{i=1}^n$ , dengan  $y_i \in \mathbb{R}$  adalah variabel respons,  $\mathbf{x}_i \in \mathbb{R}^p$  adalah vektor kovariat, dan  $\mathbf{u}_i = (u_i, v_i)^\top \in \mathbb{R}^2$  menyatakan koordinat lokasi spasial. Data spasial ini dapat direpresentasikan sebagai graf dengan mempertimbangkan struktur ketetanggaan geografis.

**Definisi 3.2.1 (Graf spasial lokal)** Untuk suatu lokasi target  $\mathbf{u}_0 \in \mathbb{R}^2$  dan bandwidth  $h > 0$ , didefinisikan himpunan tetangga lokal

$$\mathcal{N}_h(\mathbf{u}_0) = \{i : d(\mathbf{u}_i, \mathbf{u}_0) \leq h\}. \quad (3.2.1)$$

Graf spasial lokal didefinisikan sebagai  $\mathcal{G}_h(\mathbf{u}_0) = (\mathcal{V}_h, \mathcal{E}_h)$ , dengan  $\mathcal{V}_h = \mathcal{N}_h(\mathbf{u}_0)$

adalah himpunan simpul dan  $\mathcal{E}_h$  adalah himpunan sisi yang dibangun berdasarkan keterhubungan spasial atau kedekatan topologis antarsimpul dalam  $\mathcal{V}_h$ .

Setiap simpul  $i \in \mathcal{V}_h$  diasosiasikan dengan vektor fitur yang merepresentasikan kovariat dan posisi relatif terhadap lokasi target. Salah satu pendekatan yang dapat digunakan adalah representasi kartesian, yaitu

$$\mathbf{z}_i(\mathbf{u}_0) = (\mathbf{x}_i^\top, (\mathbf{u}_i - \mathbf{u}_0)^\top)^\top = (\mathbf{x}_i^\top, \Delta u_i, \Delta v_i)^\top, \quad (3.2.2)$$

dengan  $\Delta u_i = u_i - u_0$  dan  $\Delta v_i = v_i - v_0$ . Representasi ini secara ekuivalen dapat dinyatakan dalam koordinat polar sebagai

$$\mathbf{u}_i - \mathbf{u}_0 = r_i (\cos \phi_i, \sin \phi_i)^\top, \quad (3.2.3)$$

dengan  $r_i = \|\mathbf{u}_i - \mathbf{u}_0\|$  adalah jarak dari lokasi target ke observasi  $i$ , dan  $\phi_i = \arctan(\Delta v_i / \Delta u_i) \in [0, 2\pi)$  adalah sudut arah dari  $\mathbf{u}_0$  menuju  $\mathbf{u}_i$  relatif terhadap sumbu horizontal. Penggunaan posisi relatif  $\mathbf{u}_i - \mathbf{u}_0$  alih-alih koordinat absolut  $\mathbf{u}_i$  memiliki beberapa keunggulan berikut.

**Proposisi 3.2.2** Misalkan  $\mathbf{t} \in \mathbb{R}^2$  adalah vektor translasi sembarang. Jika seluruh lokasi ditranslasikan, yaitu  $\mathbf{u}'_i = \mathbf{u}_i + \mathbf{t}$  dan  $\mathbf{u}'_0 = \mathbf{u}_0 + \mathbf{t}$ , maka vektor fitur spasial tidak berubah:

$$\mathbf{u}'_i - \mathbf{u}'_0 = (\mathbf{u}_i + \mathbf{t}) - (\mathbf{u}_0 + \mathbf{t}) = \mathbf{u}_i - \mathbf{u}_0. \quad (3.2.4)$$

Akibatnya,  $\mathbf{z}_i(\mathbf{u}'_0) = \mathbf{z}_i(\mathbf{u}_0)$ , sehingga bobot GWR yang dihasilkan bersifat invarian terhadap translasi sistem koordinat.

Invariansi translasi ini memastikan bahwa model dapat digeneralisasi ke lokasi target baru tanpa bergantung pada posisi absolut dalam sistem koordinat.

Posisi relatif  $\mathbf{u}_i - \mathbf{u}_0 = (\Delta u_i, \Delta v_i)^\top$  menyimpan informasi arah dan jarak secara terpisah. Misalkan dua observasi  $i$  dan  $j$  memiliki jarak yang sama terhadap

$\mathbf{u}_0$ , yaitu  $\|\mathbf{u}_i - \mathbf{u}_0\| = \|\mathbf{u}_j - \mathbf{u}_0\| = r$ , tetapi berada pada arah yang berbeda:

$$\mathbf{u}_i - \mathbf{u}_0 = r(\cos \phi_i, \sin \phi_i)^\top, \quad \mathbf{u}_j - \mathbf{u}_0 = r(\cos \phi_j, \sin \phi_j)^\top, \quad (3.2.5)$$

dengan  $\phi_i \neq \phi_j$ . Sebab  $\mathbf{z}_i(\mathbf{u}_0) \neq \mathbf{z}_j(\mathbf{u}_0)$ , maka jaringan saraf graf dapat mempelajari pemetaan nilai yang berbeda meskipun jaraknya sama. Hal ini memungkinkan representasi struktur anisotropi, berbeda dengan *kernel* isotropik konvensional yang hanya bergantung pada jarak  $\|\mathbf{u}_i - \mathbf{u}_0\|$ . Selain itu, representasi posisi relatif juga menyediakan informasi jarak serta orientasi yang esensial untuk menentukan bobot spasial adaptif.

### 3.2.2 Fleksibilitas Metrik Jarak dalam Konstruksi Representasi

Pembahasan sebelumnya menggunakan representasi kartesian  $\mathbf{u}_i - \mathbf{u}_0$  yang implisit mengasumsikan jarak Euclidean. Namun, keunggulan pendekatan berbasis GNN adalah fleksibilitas dalam mengakomodasi metrik jarak alternatif yang dapat lebih sesuai dengan struktur data spasial sesungguhnya. Bagian ini membuktikan secara matematis bahwa berbagai metrik jarak dapat diintegrasikan ke dalam kerangka GA-GWR tanpa mengorbankan validitas teoretis.

**Proposisi 3.2.3** Misalkan  $d(\cdot, \cdot) : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^+$  adalah metrik jarak sembarang yang memenuhi:

- (i)  $d(\mathbf{u}, \mathbf{v}) \geq 0$  untuk semua  $\mathbf{u}, \mathbf{v}$ .
- (ii)  $d(\mathbf{u}, \mathbf{v}) = 0 \iff \mathbf{u} = \mathbf{v}$ .
- (iii)  $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$ .
- (iv)  $d(\mathbf{u}, \mathbf{w}) \leq d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w})$ .
- (v)  $d(\mathbf{u} + \mathbf{t}, \mathbf{v} + \mathbf{t}) = d(\mathbf{u}, \mathbf{v})$  untuk semua  $\mathbf{t} \in \mathbb{R}^2$  (translasi-invarian).

Ketetanggaan lokal dapat didefinisikan secara umum sebagai

$$\mathcal{N}_h^d(\mathbf{u}_0) = \{i : d(\mathbf{u}_i, \mathbf{u}_0) \leq h\}. \quad (3.2.6)$$



Beberapa contoh metrik jarak yang memenuhi sifat di atas adalah sebagai berikut.

- (a) Jarak Euclidean, yaitu metrik standar pada ruang  $\mathbb{R}^2$ :

$$d_{\text{Euc}}(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2} = \|\mathbf{u} - \mathbf{v}\|_2. \quad (3.2.7)$$

Metrik ini menghasilkan ketetanggaan berbentuk lingkaran dan bersifat *isotropik* (sama ke segala arah).

- (b) Jarak Manhattan, yaitu metrik norma- $L_1$ :

$$d_{\text{Man}}(\mathbf{u}, \mathbf{v}) = |u_1 - v_1| + |u_2 - v_2| = \|\mathbf{u} - \mathbf{v}\|_1. \quad (3.2.8)$$

Metrik ini menghasilkan ketetanggaan berbentuk belah ketupat dan sesuai untuk lingkungan dengan struktur jalan grid ortogonal.

- (c) Jarak Chebyshev, yaitu metrik norma- $L_\infty$ :

$$d_{\text{Cheb}}(\mathbf{u}, \mathbf{v}) = \max\{|u_1 - v_1|, |u_2 - v_2|\} = \|\mathbf{u} - \mathbf{v}\|_\infty. \quad (3.2.9)$$

Metrik ini menghasilkan ketetanggaan berbentuk persegi dan berguna ketika perpindahan diagonal memiliki biaya sama dengan perpindahan ortogonal.

- (d) Jarak Minkowski, generalisasi norma- $L_p$  untuk  $p \geq 1$ :

$$d_p(\mathbf{u}, \mathbf{v}) = (|u_1 - v_1|^p + |u_2 - v_2|^p)^{1/p}. \quad (3.2.10)$$

Kasus khusus:  $p = 1$  (Manhattan),  $p = 2$  (Euclidean),  $p \rightarrow \infty$  (Chebyshev).

- (e) Jarak jaringan, yaitu metrik jarak pada graf jaringan  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  dengan bobot sisi  $w_{ij}$  berupa

$$d_{\text{net}}(\mathbf{u}, \mathbf{v}) = \min_{\text{path } P: \mathbf{u} \rightarrow \mathbf{v}} \sum_{(i,j) \in P} w_{ij}, \quad (3.2.11)$$

yaitu panjang jalur terdekat (*shortest path*) antara  $\mathbf{u}$  dan  $\mathbf{v}$  dalam jaringan. Metrik ini memenuhi sifat (i)–(iv) secara ketat, dan sifat (v) terpenuhi secara efektif karena jarak hanya bergantung pada struktur topologis jaringan.

- (f) Jarak Haversine, untuk koordinat geografis (*latitude*, *longitude*) pada permukaan bumi:

$$d_{\text{hav}}(\mathbf{u}, \mathbf{v}) = \quad (3.2.12)$$

$$2R \arcsin \left( \sqrt{\sin^2 \left( \frac{\phi_v - \phi_u}{2} \right) + \cos(\phi_u) \cos(\phi_v) \sin^2 \left( \frac{\lambda_v - \lambda_u}{2} \right)} \right), \quad (3.2.13)$$

dengan  $R$  adalah jari-jari bumi,  $\phi$  adalah *latitude*, dan  $\lambda$  adalah *longitude*. Metrik ini sesuai untuk data geografis berskala besar.

- (g) Jarak Mahalanobis, yaitu metrik yang memperhitungkan struktur kovarians data:

$$d_{\text{Mah}}(\mathbf{u}, \mathbf{v}; \Sigma) = \sqrt{(\mathbf{u} - \mathbf{v})^\top \Sigma^{-1} (\mathbf{u} - \mathbf{v})}, \quad (3.2.14)$$

dengan  $\Sigma$  adalah matriks kovarians. Metrik ini menghasilkan ketetanggaan berbentuk elips dan secara eksplisit memodelkan *anisotropi* spasial.

Secara khusus, jarak jaringan  $d_{\text{net}}$  memenuhi sifat-sifat metrik sebagai berikut,

- (i) Panjang *shortest path* selalu  $\geq 0$  karena bobot sisi non-negatif.
- (ii)  $d_{\text{net}}(\mathbf{u}, \mathbf{u}) = 0$  karena tidak ada sisi yang perlu dilalui.
- (iii) Untuk graf tak berarah, *shortest path*  $\mathbf{u} \rightarrow \mathbf{v}$  sama dengan  $\mathbf{v} \rightarrow \mathbf{u}$ .
- (iv) Jika ada jalur  $\mathbf{u} \rightarrow \mathbf{w}$  melalui  $\mathbf{v}$ , maka  $d_{\text{net}}(\mathbf{u}, \mathbf{w}) \leq d_{\text{net}}(\mathbf{u}, \mathbf{v}) + d_{\text{net}}(\mathbf{v}, \mathbf{w})$  karena *shortest path* optimal.

- (v) Jarak jaringan bergantung pada topologi graf, bukan koordinat absolut. Dalam praktik, translasi koordinat tidak mengubah struktur jaringan sehingga jarak tetap sama.

### 3.2.3 Fungsi Skor Spasial dan Jaringan Saraf Graf

Dalam kerangka ini, jaringan saraf graf digunakan untuk mempelajari fungsi skor yang merepresentasikan kekuatan interaksi spasial antarsimpul dalam graf lokal.

**Definisi 3.2.4 (Fungsi skor spasial)** *Fungsi skor spasial adalah pemetaan*

$$s^* : (\mathbf{z}_i(\mathbf{u}_0), \mathcal{G}_h(\mathbf{u}_0)) \mapsto \mathbb{R}, \quad (3.2.15)$$

yang merepresentasikan kontribusi relatif observasi ke- $i$  terhadap estimasi lokal di  $\mathbf{u}_0$ .

Fungsi  $s^*$  bersifat tidak diketahui dan dapat bergantung secara kompleks pada kovariat, posisi relatif, serta struktur graf lokal. Dalam analisis ini, fungsi skor sejati  $s^*$  diasumsikan kontinu pada domain kompak dan invarian terhadap permutasi simpul dalam graf lokal.

**Definisi 3.2.5** *Diberikan graf lokal  $\mathcal{G}_h(\mathbf{u}_0)$ , jaringan saraf graf dengan parameter  $\theta$  mendefinisikan fungsi*

$$s_\theta(i, \mathbf{u}_0) = \text{GNN}_\theta(\mathcal{G}_h(\mathbf{u}_0), \mathbf{z}_i(\mathbf{u}_0)), \quad (3.2.16)$$

yang menghasilkan skor untuk setiap simpul  $i \in \mathcal{N}_h(\mathbf{u}_0)$ .

Berdasarkan teorema 2.8.28 untuk jaringan saraf graf, di bawah asumsi kontinuitas dan invariansi permutasi, untuk setiap  $\varepsilon > 0$  terdapat parameter  $\theta^*$  sedemikian sehingga

$$\sup_{i \in \mathcal{N}_h(\mathbf{u}_0)} |s_{\theta^*}(i, \mathbf{u}_0) - s^*(i, \mathbf{u}_0)| < \varepsilon. \quad (3.2.17)$$

Persamaan di atas menunjukkan bahwa GNN dapat mengaproksimasi fungsi skor

spasial sejati dengan presisi yang diinginkan, sehingga memungkinkan pembelajaran bobot spasial yang adaptif dan kompleks.

### 3.2.4 Konstruksi *Kernel* Spasial Berbasis GNN

Fungsi skor yang dipelajari kemudian dipetakan menjadi bobot *kernel* melalui normalisasi *softmax*. Sebelum mendefinisikan *kernel* terestimasi, perlu ditetapkan terlebih dahulu bahwa setiap *kernel* konvensional dapat direpresentasikan secara ekuivalen dalam bentuk *softmax*. Penggunaan normalisasi *softmax* ini yang memberikan nama “*attention*” pada model GA-GWR karena mekanisme ini mirip dengan mekanisme perhatian dalam jaringan saraf.

Misalkan  $w_i = K(d_i/h) > 0$  adalah bobot *kernel* konvensional untuk observasi  $i \in \mathcal{N}_h(\mathbf{u}_0)$ . Matriks bobot diagonal ditulis sebagai  $\mathbf{W} = \text{diag}(w_1, \dots, w_m)$  dengan  $m = |\mathcal{N}_h(\mathbf{u}_0)|$ . Didefinisikan bobot ternormalisasi sebagai

$$\tilde{w}_i = \frac{w_i}{\sum_{j=1}^m w_j} = \frac{K(d_i/h)}{\sum_{j=1}^m K(d_j/h)}, \quad (3.2.18)$$

sehingga  $\tilde{w}_i > 0$  dan  $\sum_{i=1}^m \tilde{w}_i = 1$ . Dengan demikian,  $\{\tilde{w}_i\}_{i=1}^m$  membentuk distribusi probabilitas diskrit atas himpunan tetangga lokal.

**Proposisi 3.2.6** *Estimator locally weighted least squares bersifat invarian terhadap penskalaan bobot. Secara eksplisit, jika  $\widetilde{\mathbf{W}} = \text{diag}(\tilde{w}_1, \dots, \tilde{w}_m)$  adalah matriks bobot ternormalisasi, maka*

$$(\mathbf{X}^\top \widetilde{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^\top \widetilde{\mathbf{W}} \mathbf{y} = (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{y}. \quad (3.2.19)$$

**Bukti.** Sebab  $\tilde{w}_i = w_i / \sum_j w_j$ , maka  $\widetilde{\mathbf{W}} = c^{-1} \mathbf{W}$  dengan  $c = \sum_{j=1}^m w_j > 0$ . Substitusi memberikan

$$\begin{aligned} (\mathbf{X}^\top \widetilde{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^\top \widetilde{\mathbf{W}} \mathbf{y} &= (c^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} (c^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{y}) \\ &= c \cdot (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \cdot c^{-1} \cdot \mathbf{X}^\top \mathbf{W} \mathbf{y} \\ &= (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{y}. \end{aligned} \quad (3.2.20)$$

Dengan demikian, estimator koefisien lokal identik untuk bobot asli  $\mathbf{W}$  dan bobot ternormalisasi  $\widetilde{\mathbf{W}}$ . ■

Proposisi di atas menunjukkan bahwa normalisasi bobot tidak mengubah estimator koefisien lokal. Oleh karena itu, tanpa mengurangi keumuman, dapat diasumsikan bahwa *kernel* sejati bersifat ternormalisasi, yaitu  $\kappa^*(i \mid \mathbf{u}_0) = \tilde{w}_i$  dengan  $\sum_i \kappa^*(i \mid \mathbf{u}_0) = 1$ . Selanjutnya, dapat ditunjukkan bahwa setiap distribusi probabilitas diskrit dapat dinyatakan sebagai transformasi *softmax*.

**Proposisi 3.2.7** Misalkan  $\{\kappa^*(i \mid \mathbf{u}_0)\}_{i=1}^m$  adalah distribusi probabilitas diskrit positif, yaitu  $\kappa^*(i \mid \mathbf{u}_0) > 0$  dan  $\sum_{i=1}^m \kappa^*(i \mid \mathbf{u}_0) = 1$ . Maka terdapat fungsi skor  $s^* : \{1, \dots, m\} \rightarrow \mathbb{R}$  sedemikian sehingga

$$\kappa^*(i \mid \mathbf{u}_0) = \frac{\exp\{s^*(i, \mathbf{u}_0)\}}{\sum_{j=1}^m \exp\{s^*(j, \mathbf{u}_0)\}} = \text{softmax}(\mathbf{s}^*)_i. \quad (3.2.21)$$

**Bukti.** Didefinisikan  $s^*(i, \mathbf{u}_0) = \log \kappa^*(i \mid \mathbf{u}_0)$  untuk setiap  $i \in \{1, \dots, m\}$ . Sebab  $\kappa^*(i \mid \mathbf{u}_0) > 0$ , logaritma terdefinisi dengan baik dan  $s^*(i, \mathbf{u}_0) \in \mathbb{R}$ . Verifikasi dilakukan dengan substitusi langsung, yaitu

$$\begin{aligned} \frac{\exp\{s^*(i, \mathbf{u}_0)\}}{\sum_{j=1}^m \exp\{s^*(j, \mathbf{u}_0)\}} &= \frac{\exp\{\log \kappa^*(i \mid \mathbf{u}_0)\}}{\sum_{j=1}^m \exp\{\log \kappa^*(j \mid \mathbf{u}_0)\}} \\ &= \frac{\kappa^*(i \mid \mathbf{u}_0)}{\sum_{j=1}^m \kappa^*(j \mid \mathbf{u}_0)} \\ &= \frac{\kappa^*(i \mid \mathbf{u}_0)}{1} \\ &= \kappa^*(i \mid \mathbf{u}_0). \end{aligned} \quad (3.2.22)$$

Perlu dicatat bahwa  $s^*$  bersifat unik hingga konstanta aditif, yaitu jika diketahui  $\tilde{s}(i) = s^*(i) + c$  untuk  $c \in \mathbb{R}$  sembarang, maka

$$\frac{\exp\{\tilde{s}(i)\}}{\sum_j \exp\{\tilde{s}(j)\}} = \frac{e^c \exp\{s^*(i)\}}{e^c \sum_j \exp\{s^*(j)\}} = \frac{\exp\{s^*(i)\}}{\sum_j \exp\{s^*(j)\}}, \quad (3.2.23)$$

sehingga model dapat menggunakan vektor skor yang tidak perlu dinormalisasi secara eksplisit, meningkatkan fleksibilitas dalam desain arsitektur GNN tanpa meng-

orbangkan interpretabilitas bobot kernel. ■

Dengan menggabungkan kedua proposisi di atas, diperoleh rantai kesetaraan untuk *kernel* konvensional, yaitu

$$w_i = K\left(\frac{d_i}{h}\right) \xrightarrow{\text{normalisasi}} \tilde{w}_i = \frac{w_i}{\sum_j w_j} \xrightarrow{\text{log-transform}} s_i^* = \log \tilde{w}_i, \quad (3.2.24)$$

sehingga secara eksplisit, fungsi skor yang merepresentasikan *kernel* konvensional adalah

$$s^*(i, \mathbf{u}_0) = \log K\left(\frac{d_i}{h}\right) - \log \sum_{j=1}^m K\left(\frac{d_j}{h}\right). \quad (3.2.25)$$

Sebab suku kedua konstan untuk semua  $i$  dan *softmax* invarian terhadap translasi konstan, fungsi skor dapat disederhanakan menjadi

$$s^*(i, \mathbf{u}_0) = \log K\left(\frac{\|\mathbf{u}_i - \mathbf{u}_0\|}{h}\right). \quad (3.2.26)$$

**Definisi 3.2.8 (Kernel spasial terestimasi)** Kernel *spasial terestimasi* berbasis GNN didefinisikan sebagai

$$\hat{\kappa}_{\theta}(i \mid \mathbf{u}_0) = \frac{\exp\{s_{\theta}(i, \mathbf{u}_0)\}}{\sum_{j \in \mathcal{N}_h(\mathbf{u}_0)} \exp\{s_{\theta}(j, \mathbf{u}_0)\}}. \quad (3.2.27)$$

**Proposisi 3.2.9** Kernel *terestimasi*  $\hat{\kappa}_{\theta}(\cdot \mid \mathbf{u}_0)$  memenuhi sifat stokastik baris (row stochastic):

(i) *Non-negativitas*:  $\hat{\kappa}_{\theta}(i \mid \mathbf{u}_0) > 0$  untuk semua  $i \in \mathcal{N}_h(\mathbf{u}_0)$ .

(ii) *Normalisasi*:  $\sum_{i \in \mathcal{N}_h(\mathbf{u}_0)} \hat{\kappa}_{\theta}(i \mid \mathbf{u}_0) = 1$ .

Sifat ini merupakan konsekuensi langsung dari transformasi softmax dan memungkinkan interpretasi  $\hat{\kappa}_{\theta}(i \mid \mathbf{u}_0)$  sebagai probabilitas diskrit atas himpunan tetangga lokal.

**Bukti.** Untuk setiap  $i \in \mathcal{N}_h(\mathbf{u}_0)$ , karena fungsi eksponensial bernilai positif untuk semua argumen real, maka  $\exp\{s_{\theta}(i, \mathbf{u}_0)\} > 0$ . Akibatnya,  $\hat{\kappa}_{\theta}(i \mid \mathbf{u}_0) > 0$ ,

yang membuktikan sifat (i). Untuk sifat (ii), dengan menjumlahkan atas seluruh  $i \in \mathcal{N}_h(\mathbf{u}_0)$  diperoleh

$$\begin{aligned} \sum_{i \in \mathcal{N}_h(\mathbf{u}_0)} \hat{\kappa}_{\boldsymbol{\theta}}(i \mid \mathbf{u}_0) &= \sum_{i \in \mathcal{N}_h(\mathbf{u}_0)} \frac{\exp\{s_{\boldsymbol{\theta}}(i, \mathbf{u}_0)\}}{\sum_{j \in \mathcal{N}_h(\mathbf{u}_0)} \exp\{s_{\boldsymbol{\theta}}(j, \mathbf{u}_0)\}} \\ &= \frac{\sum_i \exp\{s_{\boldsymbol{\theta}}(i, \mathbf{u}_0)\}}{\sum_j \exp\{s_{\boldsymbol{\theta}}(j, \mathbf{u}_0)\}} \\ &= 1. \end{aligned} \tag{3.2.28}$$

■

Selanjutnya, ditunjukkan bahwa jika fungsi skor terestimasi mengaproksimasi fungsi skor sejati, maka *kernel* terestimasi juga mengaproksimasi *kernel* sejati.

**Teorema 3.2.10** *Misalkan  $|\mathcal{N}_h(\mathbf{u}_0)| = m$  adalah jumlah tetangga lokal. Jika fungsi skor terestimasi mengaproksimasi fungsi skor sejati secara uniform, yaitu*

$$\sup_{i \in \mathcal{N}_h(\mathbf{u}_0)} |s_{\boldsymbol{\theta}}(i, \mathbf{u}_0) - s^*(i, \mathbf{u}_0)| \leq \delta \tag{3.2.29}$$

*untuk suatu  $\delta > 0$ , maka kernel terestimasi mengaproksimasi kernel sejati dengan batas galat*

$$\sup_{i \in \mathcal{N}_h(\mathbf{u}_0)} |\hat{\kappa}_{\boldsymbol{\theta}}(i \mid \mathbf{u}_0) - \kappa^*(i \mid \mathbf{u}_0)| \leq 2\delta. \tag{3.2.30}$$

*Secara khusus, jika  $\delta \rightarrow 0$ , maka  $\hat{\kappa}_{\boldsymbol{\theta}} \rightarrow \kappa^*$  secara uniform.*

**Bukti.** Didefinisikan  $a_i = s_{\boldsymbol{\theta}}(i, \mathbf{u}_0)$  dan  $b_i = s^*(i, \mathbf{u}_0)$  untuk menyederhanakan notasi. Berdasarkan teorema 2.8.28 pada skor,  $|a_i - b_i| \leq \delta$  untuk semua  $i \in \mathcal{N}_h(\mathbf{u}_0)$ . Untuk setiap  $i$ , selisih *kernel* dapat ditulis sebagai

$$\hat{\kappa}_{\boldsymbol{\theta}}(i \mid \mathbf{u}_0) - \kappa^*(i \mid \mathbf{u}_0) = \frac{e^{a_i}}{\sum_j e^{a_j}} - \frac{e^{b_i}}{\sum_j e^{b_j}}. \tag{3.2.31}$$

Dengan menyamakan penyebut, diperoleh

$$\widehat{\kappa}_{\theta}(i \mid \mathbf{u}_0) - \kappa^*(i \mid \mathbf{u}_0) = \frac{e^{a_i} \sum_j e^{b_j} - e^{b_i} \sum_j e^{a_j}}{\left(\sum_j e^{a_j}\right) \left(\sum_j e^{b_j}\right)}. \quad (3.2.32)$$

Pembilang dapat diuraikan sebagai

$$e^{a_i} \sum_j e^{b_j} - e^{b_i} \sum_j e^{a_j} = \sum_j (e^{a_i+b_j} - e^{b_i+a_j}). \quad (3.2.33)$$

Untuk melanjutkan, diperlukan sifat Lipschitz dari fungsi *softmax* yang dinyatakan dalam lema berikut.

**Lema 3.2.11** *Fungsi softmax :  $\mathbb{R}^m \rightarrow \mathbb{R}^m$  yang didefinisikan oleh*

$$[\text{softmax}(\mathbf{a})]_i = \frac{e^{a_i}}{\sum_{j=1}^m e^{a_j}} \quad (3.2.34)$$

*memenuhi ketaksamaan Lipschitz, yaitu untuk setiap  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$ ,*

$$\|\text{softmax}(\mathbf{a}) - \text{softmax}(\mathbf{b})\|_{\infty} \leq 2\|\mathbf{a} - \mathbf{b}\|_{\infty}. \quad (3.2.35)$$

Kembali ke bukti utama, dengan mengidentifikasi  $\mathbf{a} = (a_1, \dots, a_m)^{\top}$  dengan  $a_i = s_{\theta}(i, \mathbf{u}_0)$  dan  $\mathbf{b} = (b_1, \dots, b_m)^{\top}$  dengan  $b_i = s^*(i, \mathbf{u}_0)$ , serta menggunakan asumsi  $\|\mathbf{a} - \mathbf{b}\|_{\infty} \leq \delta$ , maka berdasarkan lema di atas diperoleh

$$\sup_{i \in \mathcal{N}_h(\mathbf{u}_0)} |\widehat{\kappa}_{\theta}(i \mid \mathbf{u}_0) - \kappa^*(i \mid \mathbf{u}_0)| = \|\text{softmax}(\mathbf{a}) - \text{softmax}(\mathbf{b})\|_{\infty} \leq 2\delta. \quad (3.2.36)$$

■

Dengan demikian, penggunaan GNN memperluas kelas fungsi *kernel* dari keluarga *kernel* tetap menuju kelas *kernel* adaptif yang dipelajari dari data, tanpa mengorbankan struktur lokalitas yang diperlukan untuk analisis asimtotik.



### 3.3 Integrasi *Kernel* Terestimasi ke dalam Kerangka GWR

Pada bagian ini diuraikan integrasi *kernel* spasial terestimasi berbasis GNN ke dalam kerangka inferensi GWR. Pembahasan mencakup definisi formal model, peran *kernel* sebagai *nuisance parameter*, permasalahan endogenitas yang muncul, serta skema *cross-fitting* untuk memulihkan validitas inferensi.

#### 3.3.1 Definisi Model *Graph Attention-based* GWR (GA-GWR)

Model *Graph Attention-based* (GA-GWR) mengintegrasikan *kernel* spasial terestimasi ke dalam kerangka regresi terboboti geografis. Model struktural yang mendasari proses pembangkitan data diasumsikan sebagai berikut.

**Definisi 3.3.1** Misalkan  $\{(y_i, \mathbf{x}_i, \mathbf{u}_i)\}_{i=1}^n$  adalah sampel acak dengan  $y_i \in \mathbb{R}$ ,  $\mathbf{x}_i \in \mathbb{R}^p$ , dan  $\mathbf{u}_i \in \mathbb{R}^2$ . Model struktural diasumsikan sebagai

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta}(\mathbf{u}_i) + \varepsilon_i, \quad (3.3.1)$$

dengan  $\boldsymbol{\beta} : \mathbb{R}^2 \rightarrow \mathbb{R}^p$  adalah fungsi koefisien spasial yang halus dan  $\varepsilon_i$  adalah galat acak dengan mean nol. Target inferensi adalah koefisien lokal  $\boldsymbol{\beta}(\mathbf{u}_0)$  untuk suatu lokasi target tetap  $\mathbf{u}_0 \in \mathbb{R}^2$ .

Berbeda dengan GWR klasik yang menggunakan *kernel* tetap  $K(\cdot)$ , pada GA-GWR bobot spasial diperoleh dari *kernel* terestimasi berbasis GNN. Matriks bobot didefinisikan sebagai berikut.

**Definisi 3.3.2** Matriks bobot untuk lokasi target  $\mathbf{u}_0$  didefinisikan sebagai

$$\mathbf{W}_\theta(\mathbf{u}_0) = \text{diag}(\hat{\kappa}_\theta(1 | \mathbf{u}_0), \dots, \hat{\kappa}_\theta(n | \mathbf{u}_0)), \quad (3.3.2)$$

dengan  $\hat{\kappa}_\theta(i | \mathbf{u}_0) > 0$  jika  $i \in \mathcal{N}_h(\mathbf{u}_0)$  dan  $\hat{\kappa}_\theta(i | \mathbf{u}_0) = 0$  jika  $i \notin \mathcal{N}_h(\mathbf{u}_0)$ .

**Definisi 3.3.3** Estimator koefisien lokal pada lokasi target  $\mathbf{u}_0$  didefinisikan sebagai

$$\hat{\boldsymbol{\beta}}_\theta(\mathbf{u}_0) = (\mathbf{X}^\top \mathbf{W}_\theta(\mathbf{u}_0) \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_\theta(\mathbf{u}_0) \mathbf{y}, \quad (3.3.3)$$

dengan  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$  adalah matriks desain dan  $\mathbf{y} = (y_1, \dots, y_n)^\top$  adalah vektor respons.

Estimator ini memiliki bentuk yang identik dengan estimator GWR klasik, tetapi dengan matriks bobot  $\mathbf{W}_\theta(\mathbf{u}_0)$  yang dihasilkan dari *kernel* terestimasi. Perbedaan fundamental terletak pada sifat stokastik dari matriks bobot, yang membawa implikasi penting terhadap analisis asimtotik.

### 3.3.2 Kernel Terestimasi sebagai Nuisance Parameter

Dalam kerangka GA-GWR, parameter GNN  $\theta$  yang menghasilkan *kernel* terestimasi berperan sebagai *nuisance parameter*, yaitu parameter yang diperlukan untuk estimasi tetapi bukan target inferensi utama. Peran *kernel* sebagai *nuisance parameter* membawa beberapa implikasi penting sebagai berikut.

- (i) Parameter  $\theta$  dapat berdimensi sangat tinggi (ribuan hingga jutaan parameter pada arsitektur GNN modern), jauh melebihi dimensi target  $\beta(\mathbf{u}_0) \in \mathbb{R}^p$ .
- (ii) Estimasi dilakukan dalam dua tahap: pertama mengestimasi  $\theta$  melalui pelatihan GNN, kemudian menggunakan  $\hat{\theta}$  untuk menghitung  $\hat{\beta}_{\hat{\theta}}(\mathbf{u}_0)$ .
- (iii) Estimasi  $\theta$  menambah sumber ketidakpastian yang harus diperhitungkan dalam inferensi.

Kerangka teoretis untuk menangani *nuisance parameter* berdimensi tinggi dalam konteks inferensi semiparametrik telah dikembangkan dalam literatur ekonometrika dan statistik, serta dinamakan *double machine learning* atau *debiased machine learning* (Chernozhukov dkk., 2024). Pendekatan ini akan diadaptasi untuk konteks GA-GWR pada bagian selanjutnya.

### 3.3.3 Permasalahan Endogenitas dan Ketergantungan Data

Ketika *kernel* diestimasi dari data yang sama dengan data yang digunakan untuk mengestimasi koefisien lokal, muncul permasalahan endogenitas yang dapat menginvalidasi inferensi standar. Permasalahan ini bersumber dari ketergantungan

parameter GNN  $\hat{\theta}$  terhadap galat  $\varepsilon$  melalui fungsi kerugian, terlepas dari bentuk spesifik fungsi kerugian yang digunakan.

**Proposisi 3.3.4** Misalkan  $\mathcal{L} : \Theta \times \mathbb{R}^n \rightarrow \mathbb{R}$  adalah fungsi kerugian sembarang yang memenuhi:

- (a)  $\mathcal{L}(\theta; \mathbf{y})$  bergantung pada vektor respons  $\mathbf{y}$ ; dan
- (b)  $\hat{\theta} = \arg \min_{\theta \in \Theta} \mathcal{L}(\theta; \mathbf{y})$  ada dan terukur.

Jika parameter kernel  $\theta$  diestimasi menggunakan seluruh data  $\{(y_i, \mathbf{x}_i, \mathbf{u}_i)\}_{i=1}^n$ , maka matriks bobot  $\mathbf{W}_{\hat{\theta}}(\mathbf{u}_0)$  bergantung pada vektor galat  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^\top$ . Akibatnya, kondisi

$$\mathbb{E}[\mathbf{X}^\top \mathbf{W}_{\hat{\theta}}(\mathbf{u}_0) \varepsilon] \neq \mathbf{0} \quad (3.3.4)$$

dapat terjadi, yang melanggar Asumsi GWR.2 yang diperlukan untuk konsistensi estimator. Dengan kata lain, eksogenitas galat tidak terjamin jika kernel diestimasi dari data yang sama.

**Bukti.** Dari model struktural  $y_i = \mathbf{x}_i^\top \beta(\mathbf{u}_i) + \varepsilon_i$ , vektor respons dapat ditulis sebagai

$$\mathbf{y} = \mathbf{f}(\mathbf{X}, \mathbf{U}) + \varepsilon, \quad (3.3.5)$$

dengan  $\mathbf{f}(\mathbf{X}, \mathbf{U}) = (\mathbf{x}_1^\top \beta(\mathbf{u}_1), \dots, \mathbf{x}_n^\top \beta(\mathbf{u}_n))^\top$  adalah komponen deterministik. Sebab  $\mathcal{L}(\theta; \mathbf{y})$  bergantung pada  $\mathbf{y}$  (kondisi (a)), maka

$$\mathcal{L}(\theta; \mathbf{y}) = \mathcal{L}(\theta; \mathbf{f}(\mathbf{X}, \mathbf{U}) + \varepsilon) =: \tilde{\mathcal{L}}(\theta; \mathbf{X}, \mathbf{U}, \varepsilon) \quad (3.3.6)$$

adalah fungsi eksplisit dari  $\varepsilon$ .

Selanjutnya, sebab  $\hat{\theta} = \arg \min_{\theta} \tilde{\mathcal{L}}(\theta; \mathbf{X}, \mathbf{U}, \varepsilon)$  (kondisi (b)), maka  $\hat{\theta}$  adalah fungsi terukur dari  $(\mathbf{X}, \mathbf{U}, \varepsilon)$ :

$$\hat{\theta} = g(\mathbf{X}, \mathbf{U}, \varepsilon) \quad (3.3.7)$$

untuk suatu fungsi  $g$  yang bergantung pada bentuk  $\mathcal{L}$  dan prosedur optimisasi. Kon-

sekuensinya, bobot  $w_i = \hat{\kappa}_{\hat{\theta}}(i \mid \mathbf{u}_0)$  juga bergantung pada  $\varepsilon$ :

$$w_i = h_i(\mathbf{X}, \mathbf{U}, \varepsilon). \quad (3.3.8)$$

Ekspektasi suku galat tertimbang dapat didekomposisi menggunakan identitas kovarians  $\mathbb{E}[AB] = \mathbb{E}[A]\mathbb{E}[B] + \text{Cov}(A, B)$ :

$$\mathbb{E}[w_i \varepsilon_i] = \mathbb{E}[w_i] \cdot \underbrace{\mathbb{E}[\varepsilon_i]}_{=0} + \text{Cov}(w_i, \varepsilon_i) = \text{Cov}(w_i, \varepsilon_i). \quad (3.3.9)$$

Sebab  $w_i = h_i(\mathbf{X}, \mathbf{U}, \varepsilon)$  bergantung pada  $\varepsilon_i$  melalui rantai  $\varepsilon \rightarrow \hat{\theta} \rightarrow w_i$ , maka dapat terjadi  $\text{Cov}(w_i, \varepsilon_i) \neq 0$ . Dengan demikian, mungkin untuk:

$$\mathbb{E}[\mathbf{X}^\top \mathbf{W}_{\hat{\theta}} \varepsilon] = \sum_{i=1}^n \mathbf{x}_i \cdot \text{Cov}(w_i, \varepsilon_i) \neq \mathbf{0}, \quad (3.3.10)$$

sehingga eksogenitas tidak terjamin. ■

Untuk memperjelas mekanisme kovarians secara kuantitatif, berikut adalah karakterisasi menggunakan ekspansi Taylor.

**Lema 3.3.5** Misalkan  $\hat{\theta}$  memiliki ekspansi Taylor di sekitar  $\varepsilon = \mathbf{0}$ , yaitu

$$\hat{\theta}(\varepsilon) = \hat{\theta}(\mathbf{0}) + \mathbf{J}_{\theta} \varepsilon + \mathcal{O}(\|\varepsilon\|^2), \quad (3.3.11)$$

dengan  $\mathbf{J}_{\theta} = \frac{\partial \hat{\theta}}{\partial \varepsilon^\top} \Big|_{\varepsilon=\mathbf{0}}$  adalah matriks Jacobian, maka kovarians antara bobot dan galat adalah

$$\text{Cov}(w_i, \varepsilon_j) = \sigma^2 \nabla_{\theta} w_i^\top [\mathbf{J}_{\theta}]_{:,j} + \mathcal{O}(\sigma^4), \quad (3.3.12)$$

dengan  $[\mathbf{J}_{\theta}]_{:,j}$  adalah kolom ke- $j$  dari  $\mathbf{J}_{\theta}$ .

**Bukti.** Ekspansi bobot di sekitar  $\varepsilon = \mathbf{0}$  menghasilkan

$$w_i(\varepsilon) = w_i(\mathbf{0}) + \nabla_{\theta} w_i^\top \cdot \mathbf{J}_{\theta} \varepsilon + \mathcal{O}(\|\varepsilon\|^2). \quad (3.3.13)$$

Sebab  $\mathbb{E}[\varepsilon_j] = 0$  dan  $\mathbb{E}[\varepsilon\varepsilon_j] = \sigma^2\mathbf{e}_j$  (dengan  $\mathbf{e}_j$  adalah vektor unit ke- $j$ ), maka

$$\begin{aligned}
 \text{Cov}(w_i, \varepsilon_j) &= \mathbb{E}[(w_i - \mathbb{E}[w_i])\varepsilon_j] \\
 &= \mathbb{E}[(\nabla_{\boldsymbol{\theta}} w_i^\top \mathbf{J}_{\boldsymbol{\theta}} \boldsymbol{\varepsilon})\varepsilon_j] + \mathcal{O}(\sigma^4) \\
 &= \nabla_{\boldsymbol{\theta}} w_i^\top \mathbf{J}_{\boldsymbol{\theta}} \mathbb{E}[\varepsilon\varepsilon_j] \\
 &= \sigma^2 \nabla_{\boldsymbol{\theta}} w_i^\top [\mathbf{J}_{\boldsymbol{\theta}}]_{:,j}.
 \end{aligned} \tag{3.3.14}$$

■

Lema di atas menunjukkan bahwa  $\text{Cov}(w_i, \varepsilon_j) \neq 0$  selama  $\mathbf{J}_{\boldsymbol{\theta}} \neq \mathbf{0}$  (estimator  $\hat{\boldsymbol{\theta}}$  sensitif terhadap *noise*) dan  $\nabla_{\boldsymbol{\theta}} w_i \neq \mathbf{0}$  (bobot responsif terhadap parameter). Kedua kondisi ini hampir selalu terpenuhi dalam praktik karena:

- (i) pelatihan GNN menggunakan respons  $\mathbf{y}$  yang mengandung galat  $\boldsymbol{\varepsilon}$ ;
- (ii) optimisasi menyesuaikan  $\boldsymbol{\theta}$  untuk meminimalkan kerugian, sehingga  $\hat{\boldsymbol{\theta}}$  sensitif terhadap perubahan dalam  $\mathbf{y}$  (dengan demikian terhadap  $\boldsymbol{\varepsilon}$ );
- (iii) bobot yang dihasilkan berkorelasi dengan galat, yaitu observasi dengan  $\varepsilon_i$  besar cenderung diberi bobot yang disesuaikan untuk mengakomodasi nilai  $y_i$  yang ekstrem; dan
- (iv) korelasi antara bobot dan galat merusak sifat *mean-zero* dari suku galat terimbang.

Permasalahan ini analog dengan *overfitting* dalam pembelajaran mesin, tetapi dengan konsekuensi yang lebih serius, bukan hanya prediksi yang buruk, tetapi *inferensi yang tidak valid*. Uji statistik dan interval konfidensi yang dihasilkan tidak dapat dipercaya.

### 3.3.4 Skema *Cross-Fitting* untuk Inferensi Valid

Untuk mengatasi permasalahan endogenitas, digunakan skema *cross-fitting* yang memisahkan data untuk estimasi *kernel* dan estimasi koefisien lokal. Pende-

katan ini mengadaptasi teknik *double/debiased machine learning* (Chernozhukov dkk., 2024) ke konteks GWR.

**Definisi 3.3.6 (Skema *cross-fitting*)** Partisi himpunan indeks  $\{1, \dots, n\}$  menjadi  $K$  fold yang saling lepas, yaitu

$$\{1, \dots, n\} = \mathcal{I}_1 \cup \mathcal{I}_2 \cup \dots \cup \mathcal{I}_K, \quad \mathcal{I}_k \cap \mathcal{I}_\ell = \emptyset \text{ untuk } k \neq \ell. \quad (3.3.15)$$

Untuk setiap fold  $k$ , didefinisikan:

(i) fold pelatihan, yaitu  $\mathcal{I}_{-k} = \{1, \dots, n\} \setminus \mathcal{I}_k$ ; dan

(ii) fold estimasi, yaitu  $\mathcal{I}_k$ .

Parameter GNN  $\hat{\boldsymbol{\theta}}^{(-k)}$  diestimasi hanya menggunakan data pada  $\mathcal{I}_{-k}$ , kemudian digunakan untuk menghitung bobot pada observasi di  $\mathcal{I}_k$ .

Untuk setiap fold  $k$ , dapat didefinisikan estimator parsial yang hanya menggunakan observasi pada fold tersebut.

**Definisi 3.3.7** Untuk setiap fold  $k \in \{1, \dots, K\}$ , estimator parsial didefinisikan sebagai

$$\hat{\boldsymbol{\beta}}^{(k)}(\mathbf{u}_0) = (\mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \mathbf{X}_k)^{-1} \mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \mathbf{y}_k, \quad (3.3.16)$$

dengan  $\mathbf{X}_k$  dan  $\mathbf{y}_k$  adalah matriks desain dan vektor respons untuk observasi di  $\mathcal{I}_k$ , serta  $\mathbf{W}_k^{(-k)}(\mathbf{u}_0)$  adalah matriks bobot diagonal yang dihitung menggunakan parameter  $\hat{\boldsymbol{\theta}}^{(-k)}$ .

Estimator parsial  $\hat{\boldsymbol{\beta}}^{(k)}(\mathbf{u}_0)$  bersifat valid dalam arti eksogenitas terpenuhi, namun hanya menggunakan sebagian kecil data (sekitar  $n/K$  observasi). Untuk memanfaatkan seluruh data sambil mempertahankan validitas inferensi, estimator-estimator parsial digabungkan melalui skema rata-rata tertimbang.

**Teorema 3.3.8** Misalkan  $\mathbf{H}_k(\mathbf{u}_0) = \mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \mathbf{X}_k$  adalah matriks informasi

lokal untuk fold  $k$ . Estimator cross-fitted yang didefinisikan sebagai rata-rata terimbang dari estimator parsial

$$\hat{\beta}_{\text{CF}}(\mathbf{u}_0) = \left( \sum_{k=1}^K \mathbf{H}_k(\mathbf{u}_0) \right)^{-1} \left( \sum_{k=1}^K \mathbf{H}_k(\mathbf{u}_0) \hat{\beta}^{(k)}(\mathbf{u}_0) \right) \quad (3.3.17)$$

ekuivalen dengan solusi masalah weighted least squares gabungan

$$\hat{\beta}_{\text{CF}}(\mathbf{u}_0) = \arg \min_{\beta \in \mathbb{R}^p} \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) (y_i - \mathbf{x}_i^\top \beta)^2, \quad (3.3.18)$$

yang memiliki bentuk tertutup

$$\hat{\beta}_{\text{CF}}(\mathbf{u}_0) = \left( \sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \mathbf{X}_k \right)^{-1} \left( \sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \mathbf{y}_k \right). \quad (3.3.19)$$

**Bukti.** Didefinisikan fungsi objektif

$$L(\beta) = \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) (y_i - \mathbf{x}_i^\top \beta)^2. \quad (3.3.20)$$

Sebab partisi  $\{\mathcal{I}_1, \dots, \mathcal{I}_K\}$  bersifat saling lepas dan lengkap, maka penjumlahan ganda di atas melingkupi seluruh observasi  $i \in \{1, \dots, n\}$ , dengan setiap observasi  $i$  diberi bobot  $w_i^{(-k(i))}(\mathbf{u}_0)$  dengan  $k(i)$  adalah *fold* yang memuat  $i$ .

Untuk memperoleh kondisi orde pertama, dihitung turunan parsial terhadap  $\beta$  sebagai berikut.

$$\frac{\partial L}{\partial \beta} = -2 \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) \mathbf{x}_i (y_i - \mathbf{x}_i^\top \beta). \quad (3.3.21)$$

Dengan menyamakan turunan ke nol dan menyusun ulang, diperoleh

$$\sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) \mathbf{x}_i y_i = \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) \mathbf{x}_i \mathbf{x}_i^\top \hat{\beta}. \quad (3.3.22)$$

Suku di ruas kiri dapat ditulis dalam notasi matriks sebagai

$$\sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) \mathbf{x}_i y_i = \sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \mathbf{y}_k, \quad (3.3.23)$$

dan suku di ruas kanan sebagai

$$\sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) \mathbf{x}_i \mathbf{x}_i^\top = \sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \mathbf{X}_k. \quad (3.3.24)$$

Dengan demikian, persamaan normal menjadi

$$\left( \sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \mathbf{X}_k \right) \hat{\boldsymbol{\beta}} = \sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \mathbf{y}_k. \quad (3.3.25)$$

Dengan asumsi bahwa matriks  $\sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \mathbf{X}_k$  invertibel, solusinya adalah

$$\hat{\boldsymbol{\beta}}_{\text{CF}}(\mathbf{u}_0) = \left( \sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \mathbf{X}_k \right)^{-1} \left( \sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \mathbf{y}_k \right). \quad (3.3.26)$$

Dari definisi estimator parsial, berlaku

$$\hat{\boldsymbol{\beta}}^{(k)}(\mathbf{u}_0) = \mathbf{H}_k(\mathbf{u}_0)^{-1} \mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \mathbf{y}_k, \quad (3.3.27)$$

sehingga

$$\mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \mathbf{y}_k = \mathbf{H}_k(\mathbf{u}_0) \hat{\boldsymbol{\beta}}^{(k)}(\mathbf{u}_0). \quad (3.3.28)$$

Substitusi ke Persamaan (3.3.26) menghasilkan

$$\hat{\boldsymbol{\beta}}_{\text{CF}}(\mathbf{u}_0) = \left( \sum_{k=1}^K \mathbf{H}_k(\mathbf{u}_0) \right)^{-1} \left( \sum_{k=1}^K \mathbf{H}_k(\mathbf{u}_0) \hat{\boldsymbol{\beta}}^{(k)}(\mathbf{u}_0) \right). \quad (3.3.29)$$

Bentuk ini merupakan *rata-rata tertimbang matriks (matrix-weighted average)* dari estimator parsial, dengan bobot berupa matriks informasi lokal  $\mathbf{H}_k(\mathbf{u}_0)$ .

■



Perlu dicatat bahwa jika  $K = 1$  (tanpa *cross-fitting*), maka  $\hat{\beta}_{\text{CF}}(\mathbf{u}_0) = \hat{\beta}^{(1)}(\mathbf{u}_0)$ , yang merupakan estimator GA-GWR standar tanpa pemisahan data.

Skema *cross-fitting* memulihkan eksogenitas melalui mekanisme berikut.

**Proposisi 3.3.9** Untuk setiap fold  $k$ , karena  $\hat{\theta}^{(-k)}$  diestimasi tanpa menggunakan data pada  $\mathcal{I}_k$ , maka kondisional pada  $\hat{\theta}^{(-k)}$  berlaku

$$\mathbb{E}[\mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \varepsilon_k \mid \hat{\theta}^{(-k)}] = \mathbf{0}, \quad (3.3.30)$$

dengan  $\varepsilon_k$  adalah vektor galat untuk observasi di  $\mathcal{I}_k$ . Hal ini karena  $\mathbf{W}_k^{(-k)}(\mathbf{u}_0)$  independen dari  $\varepsilon_k$ .

**Bukti.** Langkah pertama adalah menunjukkan independensi antara bobot dan galat. Berdasarkan konstruksi *cross-fitting*, parameter  $\hat{\theta}^{(-k)}$  diestimasi menggunakan data pada *fold* pelatihan  $\mathcal{I}_{-k}$ , sehingga  $\hat{\theta}^{(-k)}$  merupakan fungsi terukur dari  $\{(y_i, \mathbf{x}_i, \mathbf{u}_i) : i \in \mathcal{I}_{-k}\}$ . Secara eksplisit, dapat ditulis

$$\hat{\theta}^{(-k)} = g(\{(y_i, \mathbf{x}_i, \mathbf{u}_i)\}_{i \in \mathcal{I}_{-k}}) \quad (3.3.31)$$

untuk suatu fungsi terukur  $g$ . Sebab galat  $\{\varepsilon_i\}_{i=1}^n$  diasumsikan independen antarobservasi, maka himpunan galat pada *fold* pelatihan  $\{\varepsilon_i : i \in \mathcal{I}_{-k}\}$  independen dari himpunan galat pada *fold* estimasi  $\{\varepsilon_i : i \in \mathcal{I}_k\}$ . Sebab  $\hat{\theta}^{(-k)}$  hanya bergantung pada data di  $\mathcal{I}_{-k}$  dan matriks bobot  $\mathbf{W}_k^{(-k)}(\mathbf{u}_0)$  merupakan fungsi deterministik dari  $\hat{\theta}^{(-k)}$  dan lokasi  $\{\mathbf{u}_i : i \in \mathcal{I}_k\}$ , maka

$$\mathbf{W}_k^{(-k)}(\mathbf{u}_0) \perp\!\!\!\perp \varepsilon_k \mid \{\mathbf{x}_i, \mathbf{u}_i\}_{i \in \mathcal{I}_k}. \quad (3.3.32)$$

Matriks bobot  $\mathbf{W}_k^{(-k)}(\mathbf{u}_0)$  berbentuk diagonal dengan elemen  $w_i^{(-k)}(\mathbf{u}_0) = \hat{\kappa}_{\hat{\theta}^{(-k)}}(i \mid \mathbf{u}_0)$  untuk  $i \in \mathcal{I}_k$ . Dengan demikian, produk  $\mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \varepsilon_k$  dapat

diuraikan sebagai

$$\begin{aligned} \mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \boldsymbol{\varepsilon}_k &= \begin{pmatrix} \mathbf{x}_{i_1} & \mathbf{x}_{i_2} & \cdots & \mathbf{x}_{i_{|k|}} \end{pmatrix} \begin{pmatrix} w_{i_1}^{(-k)} & & & \\ & \ddots & & \\ & & w_{i_{|k|}}^{(-k)} & \\ & & & \end{pmatrix} \begin{pmatrix} \varepsilon_{i_1} \\ \vdots \\ \varepsilon_{i_{|k|}} \end{pmatrix} \\ &= \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) \mathbf{x}_i \varepsilon_i, \end{aligned} \quad (3.3.33)$$

dengan  $\{i_1, i_2, \dots, i_{|k|}\}$  adalah enumerasi elemen-elemen  $\mathcal{I}_k$ . Dengan menggunakan linearitas ekspektasi kondisional dan independensi yang telah ditunjukkan, diperoleh

$$\begin{aligned} \mathbb{E}[\mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \boldsymbol{\varepsilon}_k \mid \hat{\boldsymbol{\theta}}^{(-k)}] &= \mathbb{E}\left[\sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) \mathbf{x}_i \varepsilon_i \mid \hat{\boldsymbol{\theta}}^{(-k)}\right] \\ &= \sum_{i \in \mathcal{I}_k} \mathbb{E}[w_i^{(-k)}(\mathbf{u}_0) \mathbf{x}_i \varepsilon_i \mid \hat{\boldsymbol{\theta}}^{(-k)}]. \end{aligned} \quad (3.3.34)$$

Sebab  $w_i^{(-k)}(\mathbf{u}_0)$  merupakan fungsi deterministik dari  $\hat{\boldsymbol{\theta}}^{(-k)}$  dan  $\mathbf{u}_i$ , maka kondisional pada  $\hat{\boldsymbol{\theta}}^{(-k)}$ , bobot  $w_i^{(-k)}(\mathbf{u}_0)$  bersifat konstan. Dengan demikian, Persamaan (3.3.34) dapat ditulis ulang sebagai

$$\sum_{i \in \mathcal{I}_k} \mathbb{E}[w_i^{(-k)}(\mathbf{u}_0) \mathbf{x}_i \varepsilon_i \mid \hat{\boldsymbol{\theta}}^{(-k)}] = \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) \mathbb{E}[\mathbf{x}_i \varepsilon_i \mid \hat{\boldsymbol{\theta}}^{(-k)}]. \quad (3.3.35)$$

Selanjutnya, karena  $\hat{\boldsymbol{\theta}}^{(-k)} \perp\!\!\!\perp (\mathbf{x}_i, \varepsilon_i)$  untuk setiap  $i \in \mathcal{I}_k$ , maka berlaku

$$\mathbb{E}[\mathbf{x}_i \varepsilon_i \mid \hat{\boldsymbol{\theta}}^{(-k)}] = \mathbb{E}[\mathbf{x}_i \varepsilon_i]. \quad (3.3.36)$$

Dengan menggunakan hukum ekspektasi bersyarat (*tower property*) dan

asumsi eksogenitas  $\mathbb{E}[\varepsilon_i \mid \mathbf{x}_i, \mathbf{u}_i] = 0$ , diperoleh

$$\begin{aligned}\mathbb{E}[\mathbf{x}_i \varepsilon_i] &= \mathbb{E}[\mathbb{E}[\mathbf{x}_i \varepsilon_i \mid \mathbf{x}_i, \mathbf{u}_i]] \\ &= \mathbb{E}[\mathbf{x}_i \mathbb{E}[\varepsilon_i \mid \mathbf{x}_i, \mathbf{u}_i]] \\ &= \mathbb{E}[\mathbf{x}_i \cdot 0] = \mathbf{0}.\end{aligned}\tag{3.3.37}$$

Dengan menggabungkan hasil-hasil di atas, diperoleh

$$\mathbb{E}[\mathbf{X}_k^\top \mathbf{W}_k^{(-k)}(\mathbf{u}_0) \varepsilon_k \mid \hat{\boldsymbol{\theta}}^{(-k)}] = \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) \cdot \mathbf{0} = \mathbf{0}.\tag{3.3.38}$$

■

Dengan pemulihan eksogenitas ini, hukum bilangan besar dapat diterapkan kembali dan konsistensi estimator dapat dijamin. Analisis asimtotik lengkap akan diuraikan pada bagian selanjutnya.

### 3.4 Analisis Asimtotik Koefisien Lokal pada GA-GWR

Bagian ini mengembangkan teori asimtotik untuk estimator *cross-fitted* GA-GWR. Analisis mencakup dekomposisi bias-variansi, konsistensi, dan distribusi asimtotik koefisien lokal. Untuk menyederhanakan notasi, dependensi terhadap lokasi target  $\mathbf{u}_0$  akan dihilangkan bila tidak menimbulkan ambiguitas.

#### 3.4.1 Asumsi-Asumsi Regularitas

Analisis asimtotik memerlukan sejumlah asumsi regularitas yang mencakup struktur data, properti fungsi koefisien, serta karakteristik *kernel* terestimasi. Asumsi-asumsi ini dikelompokkan menjadi tiga kategori, yaitu asumsi data dan model, asumsi kehalusan fungsi koefisien, serta asumsi lokalitas dan *kernel*.

Asumsi data dan model menetapkan kondisi dasar pada observasi, galat, dan kovariat. Di bawah ini adalah keempat asumsi utama dalam kategori ini.

**Asumsi GA-GWR.1 (Independensi)** *Observasi*  $\{(y_i, \mathbf{x}_i, \mathbf{u}_i)\}_{i=1}^n$  bersifat indepen-

den dan berdistribusi identik (i.i.d.).

Asumsi independensi diperlukan untuk penerapan hukum bilangan besar dan teorema limit pusat. Dalam konteks spasial, asumsi ini dapat direlaksasi menjadi dependensi spasial yang melemah (*mixing conditions*), namun pembahasan tersebut berada di luar cakupan analisis ini.

**Asumsi GA-GWR.2 (Eksogenitas Lokal)** Galat memenuhi kondisi ekspektasi kondisional nol, yaitu

$$\mathbb{E}[\varepsilon_i \mid \mathbf{x}_i, \mathbf{u}_i] = 0. \quad (3.4.1)$$

Asumsi eksogenitas menjamin bahwa galat tidak berkorelasi dengan kovariat dan lokasi. Bersama dengan skema *cross-fitting*, asumsi ini memastikan bahwa suku galat tertimbang memiliki ekspektasi nol.

**Asumsi GA-GWR.3 (Homoskedastisitas Kondisional)** Variansi galat bersyarat bersifat konstan, yaitu

$$\text{Var}(\varepsilon_i \mid \mathbf{x}_i, \mathbf{u}_i) = \sigma^2 < \infty. \quad (3.4.2)$$

Asumsi homoskedastisitas menyederhanakan analisis variansi. Kasus heteroskedastisitas dapat ditangani dengan modifikasi pada estimator variansi, namun tidak mengubah hasil konsistensi.

**Asumsi GA-GWR.4 (Momen Kovariat Terbatas)** Kovariat memiliki momen kedua yang terbatas:

$$\mathbb{E}[\|\mathbf{x}_i\|^2] < \infty. \quad (3.4.3)$$

Asumsi ini menjamin eksistensi dan keterbatasan matriks  $\mathbf{X}^\top \mathbf{W} \mathbf{X}$  serta inversnya.

Asumsi selanjutnya berkaitan dengan kehalusan fungsi koefisien spasial. Asumsi ini penting untuk mengendalikan bias yang muncul akibat aproksimasi lokal. Berikut adalah asumsi kehalusan yang digunakan.

**Asumsi GA-GWR.5 (Diferensiabilitas Fungsi Koefisien)** Fungsi koefisien  $\beta : \mathbb{R}^2 \rightarrow \mathbb{R}^p$  dua kali terdiferensiasikan kontinu di sekitar lokasi target  $\mathbf{u}_0$ . Secara eks-

plisit, terdapat lingkungan  $\mathcal{B}_\delta(\mathbf{u}_0)$  sedemikian sehingga untuk setiap  $\mathbf{u} \in \mathcal{B}_\delta(\mathbf{u}_0)$ :

$$\beta(\mathbf{u}) = \beta(\mathbf{u}_0) + \nabla\beta(\mathbf{u}_0)^\top(\mathbf{u} - \mathbf{u}_0) + \frac{1}{2}(\mathbf{u} - \mathbf{u}_0)^\top \mathbf{H}_\beta(\mathbf{u}_0)(\mathbf{u} - \mathbf{u}_0) + \quad (3.4.4)$$

$$o(\|\mathbf{u} - \mathbf{u}_0\|^2), \quad (3.4.5)$$

dengan  $\nabla\beta(\mathbf{u}_0) \in \mathbb{R}^{2 \times p}$  adalah matriks gradien dan  $\mathbf{H}_\beta(\mathbf{u}_0)$  adalah tensor Hessian.

Kehalusan fungsi koefisien memungkinkan ekspansi Taylor yang menjadi dasar analisis bias. Orde kehalusan menentukan laju konvergensi bias terhadap nol.

Asumsi terakhir berkaitan dengan karakteristik *kernel* terestimasi dan lokalitas estimasi. Asumsi-asumsi ini penting untuk mengendalikan bias dan variansi estimator. Berikut adalah asumsi-asumsi tersebut.

**Asumsi GA-GWR.6 (Shrinking Neighborhood)** Parameter bandwidth  $h = h_n$  memenuhi  $h_n \rightarrow 0$  ketika  $n \rightarrow \infty$ .

Asumsi ini menjamin bahwa observasi yang digunakan dalam estimasi lokal semakin terkonsentrasi di sekitar lokasi target, sehingga aproksimasi  $\beta(\mathbf{u}_i) \approx \beta(\mathbf{u}_0)$  semakin akurat.

**Asumsi GA-GWR.7 (Ukuran Sampel Lokal Divergen)** Jumlah observasi efektif di dalam neighborhood divergen:

$$nh_n^{d_{\text{eff}}} \rightarrow \infty \quad \text{ketika } n \rightarrow \infty, \quad (3.4.6)$$

dengan  $d_{\text{eff}} \in (0, 2]$  adalah dimensi fraktal efektif dari metrik jarak yang digunakan.

Untuk metrik Euclidean pada ruang homogen,  $d_{\text{eff}} = d = 2$ . Untuk metrik non-Euclidean (misalnya jarak jaringan),  $d_{\text{eff}}$  dapat berbeda dan perlu diestimasi dari struktur data lokal.

**Proposisi 3.4.1** Dimensi fraktal efektif dapat diestimasi dari distribusi jumlah observasi dalam ball lokal. Untuk suatu lokasi target  $\mathbf{u}_0$  dan rangkaian radius  $h_1 <$

$h_2 < \dots < h_K$ , didefinisikan

$$m_k = |\mathcal{N}_{h_k}^d(\mathbf{u}_0)| = \#\{i : d(\mathbf{u}_i, \mathbf{u}_0) \leq h_k\}. \quad (3.4.7)$$

*Dimensi fraktal diestimasi melalui regresi log-log:*

$$\hat{d}_{\text{eff}} = \frac{\sum_k (\log h_k - \log \bar{h})(\log m_k - \log \bar{m})}{\sum_k (\log h_k - \log \bar{h})^2}, \quad (3.4.8)$$

dengan bar menunjukkan rata-rata. Nilai  $\hat{d}_{\text{eff}}$  umumnya stabil untuk lokasi dalam wilayah yang relatif homogen.

**Asumsi GA-GWR.8 (Keterbatasan Bobot Kernel)** Bobot kernel terestimasi memenuhi kondisi keterbatasan:

$$\sup_{i \in \mathcal{N}_h(\mathbf{u}_0)} w_i(\mathbf{u}_0) \leq \frac{C}{nh^d} \quad (3.4.9)$$

untuk suatu konstanta  $C > 0$  yang tidak bergantung pada  $n$ .

Asumsi ini mencegah satu observasi mendominasi estimator dan diperlukan untuk penerapan teorema limit pusat Lindeberg–Feller.

**Asumsi GA-GWR.9 (Simetri Lokal Asimtotik)** Bobot kernel memenuhi kondisi simetri lokal, yaitu

$$\sum_{i=1}^n w_i(\mathbf{u}_0)(\mathbf{u}_i - \mathbf{u}_0) \xrightarrow{p} \mathbf{0}. \quad (3.4.10)$$

Asumsi simetri lokal menjamin bahwa bobot tidak bias secara sistematis ke satu arah. Hal ini dicapai melalui desain arsitektur GNN yang menggunakan koordinat relatif  $\mathbf{u}_i - \mathbf{u}_0$  sebagai input.

**Asumsi GA-GWR.10 (Kontinuitas Bobot)** Fungsi bobot bersifat Lipschitz kontinu terhadap input, yaitu

$$|w_i(\mathbf{u}_0; \mathbf{z}) - w_i(\mathbf{u}_0; \mathbf{z}')| \leq L \|\mathbf{z} - \mathbf{z}'\| \quad (3.4.11)$$

untuk suatu konstanta Lipschitz  $L > 0$ .

Kontinuitas diperlukan untuk penerapan hukum bilangan besar *uniform* dan kontrol suku sisa dalam ekspansi Taylor.

### 3.4.2 Analisis Bias Asimtotik Koefisien Lokal

Analisis bias dimulai dengan dekomposisi estimator *cross-fitted* ke dalam suku target, suku bias, dan suku galat.

**Teorema 3.4.2** *Estimator cross-fitted GA-GWR dapat didekomposisikan sebagai*

$$\hat{\beta}_{CF}(\mathbf{u}_0) = \beta(\mathbf{u}_0) + \mathbf{B}_n(\mathbf{u}_0) + \mathbf{V}_n(\mathbf{u}_0), \quad (3.4.12)$$

dengan suku bias

$$\mathbf{B}_n(\mathbf{u}_0) = \left( \sum_{k=1}^K \mathbf{H}_k \right)^{-1} \left( \sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)} \mathbf{X}_k (\beta(\mathbf{U}_k) - \beta(\mathbf{u}_0)) \right) \quad (3.4.13)$$

dan suku galat (noise term)

$$\mathbf{V}_n(\mathbf{u}_0) = \left( \sum_{k=1}^K \mathbf{H}_k \right)^{-1} \left( \sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)} \boldsymbol{\varepsilon}_k \right), \quad (3.4.14)$$

dengan  $\beta(\mathbf{U}_k) - \beta(\mathbf{u}_0)$  adalah vektor yang elemen ke- $i$ -nya (untuk  $i \in \mathcal{I}_k$ ) adalah  $\beta(\mathbf{u}_i) - \beta(\mathbf{u}_0)$ .

**Bukti.** Dari model struktural, untuk setiap observasi  $i \in \mathcal{I}_k$  berlaku

$$y_i = \mathbf{x}_i^\top \beta(\mathbf{u}_i) + \varepsilon_i. \quad (3.4.15)$$

Dengan menambah dan mengurangi  $\mathbf{x}_i^\top \beta(\mathbf{u}_0)$ , diperoleh

$$y_i = \mathbf{x}_i^\top \beta(\mathbf{u}_0) + \mathbf{x}_i^\top (\beta(\mathbf{u}_i) - \beta(\mathbf{u}_0)) + \varepsilon_i. \quad (3.4.16)$$

Dalam notasi matriks untuk *fold k*:

$$\mathbf{y}_k = \mathbf{X}_k \boldsymbol{\beta}(\mathbf{u}_0) + \mathbf{X}_k (\boldsymbol{\beta}(\mathbf{U}_k) - \boldsymbol{\beta}(\mathbf{u}_0)) + \boldsymbol{\varepsilon}_k. \quad (3.4.17)$$

Substitusi ke estimator *cross-fitted*:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\text{CF}} &= \left( \sum_{k=1}^K \mathbf{H}_k \right)^{-1} \left( \sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)} \mathbf{y}_k \right) \\ &= \left( \sum_{k=1}^K \mathbf{H}_k \right)^{-1} \left( \sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)} \left[ \mathbf{X}_k \boldsymbol{\beta}(\mathbf{u}_0) + \mathbf{X}_k (\boldsymbol{\beta}(\mathbf{U}_k) - \boldsymbol{\beta}(\mathbf{u}_0)) + \boldsymbol{\varepsilon}_k \right] \right). \end{aligned} \quad (3.4.18)$$

Dengan linearitas penjumlahan dan fakta bahwa  $\mathbf{X}_k^\top \mathbf{W}_k^{(-k)} \mathbf{X}_k = \mathbf{H}_k$ , diperoleh

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\text{CF}} &= \left( \sum_{k=1}^K \mathbf{H}_k \right)^{-1} \left( \sum_{k=1}^K \mathbf{H}_k \right) \boldsymbol{\beta}(\mathbf{u}_0) \\ &\quad + \left( \sum_{k=1}^K \mathbf{H}_k \right)^{-1} \left( \sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)} \mathbf{X}_k (\boldsymbol{\beta}(\mathbf{U}_k) - \boldsymbol{\beta}(\mathbf{u}_0)) \right) \\ &\quad + \left( \sum_{k=1}^K \mathbf{H}_k \right)^{-1} \left( \sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)} \boldsymbol{\varepsilon}_k \right) \\ &= \boldsymbol{\beta}(\mathbf{u}_0) + \mathbf{B}_n(\mathbf{u}_0) + \mathbf{V}_n(\mathbf{u}_0). \end{aligned} \quad (3.4.19)$$

■

Untuk menganalisis perilaku asimtotik suku bias, digunakan ekspansi Taylor dari fungsi koefisien.

**Proposisi 3.4.3** *Di bawah Asumsi GA-GWR.5, untuk setiap  $\mathbf{u}_i \in \mathcal{N}_h(\mathbf{u}_0)$ :*

$$\boldsymbol{\beta}(\mathbf{u}_i) - \boldsymbol{\beta}(\mathbf{u}_0) = \nabla \boldsymbol{\beta}(\mathbf{u}_0)^\top (\mathbf{u}_i - \mathbf{u}_0) + \mathbf{R}_i, \quad (3.4.20)$$

dengan suku sisa  $\mathbf{R}_i$  memenuhi  $\|\mathbf{R}_i\| \leq C_\beta \|\mathbf{u}_i - \mathbf{u}_0\|^2$  untuk suatu konstanta  $C_\beta > 0$ .



Dengan ekspansi ini, suku bias dapat diuraikan menjadi dua komponen.

**Teorema 3.4.4** *Suku bias dapat ditulis sebagai*

$$\mathbf{B}_n(\mathbf{u}_0) = \mathbf{B}_n^{(1)}(\mathbf{u}_0) + \mathbf{B}_n^{(2)}(\mathbf{u}_0), \quad (3.4.21)$$

*dengan suku bias orde pertama*

$$\mathbf{B}_n^{(1)}(\mathbf{u}_0) = \left( \sum_{k=1}^K \mathbf{H}_k \right)^{-1} \left( \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top \nabla \beta(\mathbf{u}_0)^\top (\mathbf{u}_i - \mathbf{u}_0) \right) \quad (3.4.22)$$

*dan suku bias orde kedua*

$$\mathbf{B}_n^{(2)}(\mathbf{u}_0) = \left( \sum_{k=1}^K \mathbf{H}_k \right)^{-1} \left( \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top \mathbf{R}_i \right). \quad (3.4.23)$$

**Bukti.** Dengan mensubstitusikan ekspansi Taylor ke dalam ekspresi bias:

$$\begin{aligned} & \sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)} \mathbf{X}_k (\beta(\mathbf{U}_k) - \beta(\mathbf{u}_0)) \\ = & \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top (\beta(\mathbf{u}_i) - \beta(\mathbf{u}_0)) \\ = & \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top \left[ \nabla \beta(\mathbf{u}_0)^\top (\mathbf{u}_i - \mathbf{u}_0) + \mathbf{R}_i \right]. \end{aligned} \quad (3.4.24)$$

Pemisahan menjadi dua suku memberikan hasil yang diklaim. ■

**Proposisi 3.4.5** *Di bawah Asumsi GA-GWR.9, suku bias orde pertama memenuhi*

$$\mathbf{B}_n^{(1)}(\mathbf{u}_0) \xrightarrow{p} \mathbf{0}. \quad (3.4.25)$$

**Bukti.** Perhatikan bahwa

$$\sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top \nabla \beta(\mathbf{u}_0)^\top (\mathbf{u}_i - \mathbf{u}_0) \quad (3.4.26)$$

$$= \left( \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top \right) \nabla \beta(\mathbf{u}_0)^\top \left( \sum_{j=1}^n \tilde{w}_j (\mathbf{u}_j - \mathbf{u}_0) \right), \quad (3.4.27)$$

dengan  $\tilde{w}_j$  adalah bobot yang sesuai untuk observasi  $j$ . Berdasarkan Asumsi GA-GWR.9, suku  $\sum_j \tilde{w}_j (\mathbf{u}_j - \mathbf{u}_0) \xrightarrow{p} \mathbf{0}$ , sehingga  $\mathbf{B}_n^{(1)}(\mathbf{u}_0) \xrightarrow{p} \mathbf{0}$ . ■

**Teorema 3.4.6** *Di bawah Asumsi GA-GWR.5–GA-GWR.9, suku bias memenuhi*

$$\mathbf{B}_n(\mathbf{u}_0) = \mathcal{O}_p(h^2). \quad (3.4.28)$$

Secara khusus, karena  $h \rightarrow 0$ , maka  $\mathbf{B}_n(\mathbf{u}_0) \xrightarrow{p} \mathbf{0}$ .

**Bukti.** Dari Proposisi 3.4.5,  $\mathbf{B}_n^{(1)}(\mathbf{u}_0) \xrightarrow{p} \mathbf{0}$ . Untuk suku bias orde kedua, diperhatikan bahwa

$$\begin{aligned} \left\| \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top \mathbf{R}_i \right\| &\leq \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \|\mathbf{x}_i\|^2 \|\mathbf{R}_i\| \\ &\leq C_\beta \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \|\mathbf{x}_i\|^2 \|\mathbf{u}_i - \mathbf{u}_0\|^2. \end{aligned} \quad (3.4.29)$$

Sebab  $\mathbf{u}_i \in \mathcal{N}_h(\mathbf{u}_0)$  menyiratkan  $\|\mathbf{u}_i - \mathbf{u}_0\| \leq h$ , maka

$$\|\mathbf{u}_i - \mathbf{u}_0\|^2 \leq h^2. \quad (3.4.30)$$

Dengan demikian,

$$\left\| \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top \mathbf{R}_i \right\| \leq C_\beta h^2 \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \|\mathbf{x}_i\|^2. \quad (3.4.31)$$

Sebab bobot ternormalisasi ( $\sum_i w_i = 1$ ) dan  $\mathbb{E}[\|\mathbf{x}_i\|^2] < \infty$ , maka suku terakhir terbatas dalam probabilitas. Dengan mengalikan dengan  $(\sum_k \mathbf{H}_k)^{-1}$  yang juga ter-

batas, diperoleh

$$\|\mathbf{B}_n^{(2)}(\mathbf{u}_0)\| = \mathcal{O}_p(h^2). \quad (3.4.32)$$

Dengan demikian,  $\mathbf{B}_n(\mathbf{u}_0) = \mathbf{B}_n^{(1)}(\mathbf{u}_0) + \mathbf{B}_n^{(2)}(\mathbf{u}_0) = o_p(1) + \mathcal{O}_p(h^2) = \mathcal{O}_p(h^2)$ . ■

Hasil  $\mathbf{B}_n(\mathbf{u}_0) = \mathcal{O}_p(h^2)$  berlaku secara umum untuk setiap metrik jarak yang memenuhi Asumsi GA-GWR.1–GA-GWR.10. Secara khusus:

- (i) Jika fungsi koefisien  $\beta(\cdot)$  bersifat dua kali terdiferensiasikan kontinu (Asumsi GA-GWR.5), maka ekspansi Taylor orde dua

$$\beta(\mathbf{u}_i) - \beta(\mathbf{u}_0) = \mathcal{O}(\|\mathbf{u}_i - \mathbf{u}_0\|^2) \quad (3.4.33)$$

tetap berlaku tanpa memperhatikan metrik yang digunakan, asalkan observasi dalam  $\mathcal{N}_h^d(\mathbf{u}_0)$  memenuhi  $\|\mathbf{u}_i - \mathbf{u}_0\| = \mathcal{O}(h)$  dalam norma Euclidean lokal.

- (ii) Untuk metrik yang sangat berbeda dari Euclidean (misalnya jarak jaringan dengan  $d_{\text{eff}} \ll 2$ ), meskipun laju bias tetap  $\mathcal{O}_p(h^2)$  dalam parameter  $h$ , interpretasi geometris berubah.  $h$  sekarang menyatakan radius dalam metrik non-Euclidean, bukan Euclidean.
- (iii) Orde bias tidak bergantung pada metrik, melainkan pada kehalusan fungsi  $\beta$  berdasarkan Asumsi GA-GWR.5 dan Asumsi GA-GWR.9.

**Akibat 3.4.7** *Estimator cross-fitted GA-GWR bersifat asymptotically unbiased, yaitu*

$$\mathbb{E}[\hat{\beta}_{\text{CF}}(\mathbf{u}_0)] - \beta(\mathbf{u}_0) = \mathcal{O}(h^2) \rightarrow 0 \quad \text{ketika } n \rightarrow \infty \text{ dan } h \rightarrow 0. \quad (3.4.34)$$

### 3.4.3 Analisis Bias Asimtotik Variansi Galat

Selain estimasi koefisien lokal, diperlukan pula estimasi variansi galat  $\sigma^2$  untuk konstruksi interval konfidensi dan uji statistik.

**Teorema 3.4.8** *Estimator variansi galat cross-fitted yang diperoleh dari minimisasi*

fungsi log-likelihood tertimbang Gaussian diberikan oleh

$$\hat{\sigma}_{\text{CF}}^2(\mathbf{u}_0) = \frac{\sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) (y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\text{CF}}(\mathbf{u}_0))^2}{\sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) - p}, \quad (3.4.35)$$

dengan  $p$  adalah jumlah parameter (dimensi  $\boldsymbol{\beta}$ ).

**Bukti.** Diasumsikan galat berdistribusi normal  $\varepsilon_i \mid \mathbf{x}_i, \mathbf{u}_i \sim \mathcal{N}(0, \sigma^2)$ . Fungsi log-likelihood tertimbang lokal untuk lokasi  $\mathbf{u}_0$  diberikan oleh

$$\ell(\boldsymbol{\beta}, \sigma^2; \mathbf{u}_0) = -\frac{1}{2} \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) \left[ \log(2\pi\sigma^2) + \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\sigma^2} \right]. \quad (3.4.36)$$

Dengan mendefinisikan  $W_{\text{tot}} = \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0)$  dan

$$\text{RSS}_w = \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) (y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2,$$

fungsi log-likelihood dapat ditulis sebagai

$$\ell(\boldsymbol{\beta}, \sigma^2; \mathbf{u}_0) = -\frac{W_{\text{tot}}}{2} \log(2\pi) - \frac{W_{\text{tot}}}{2} \log(\sigma^2) - \frac{\text{RSS}_w}{2\sigma^2}. \quad (3.4.37)$$

Untuk memperoleh estimator  $\sigma^2$ , dihitung turunan parsial terhadap  $\sigma^2$  dan disamakan dengan nol:

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{W_{\text{tot}}}{2\sigma^2} + \frac{\text{RSS}_w}{2(\sigma^2)^2} = 0. \quad (3.4.38)$$

Dengan mengalikan kedua ruas dengan  $2(\sigma^2)^2$  dan menyusun ulang, diperoleh

$$\text{RSS}_w = W_{\text{tot}} \cdot \sigma^2, \quad (3.4.39)$$

sehingga estimator *maximum likelihood* tertimbang adalah

$$\hat{\sigma}_{\text{ML}}^2 = \frac{\text{RSS}_w}{W_{\text{tot}}} = \frac{\sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) (y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\text{CF}}(\mathbf{u}_0))^2}{\sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0)}. \quad (3.4.40)$$

Namun, estimator MLE ini bersifat bias ke bawah karena tidak memperhitungkan derajat kebebasan yang hilang akibat estimasi  $p$  parameter dalam  $\boldsymbol{\beta}$ . Koreksi bias dilakukan dengan mengganti penyebut  $W_{\text{tot}}$  dengan  $W_{\text{tot}} - p$ , yang menghasilkan estimator tak-bias (atau mendekati tak-bias)

$$\hat{\sigma}_{\text{CF}}^2(\mathbf{u}_0) = \frac{\sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) (y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\text{CF}}(\mathbf{u}_0))^2}{\sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)}(\mathbf{u}_0) - p}. \quad (3.4.41)$$

Koreksi ini analog dengan pembagi  $(n - p)$  pada estimator variansi OLS klasik  $\hat{\sigma}^2 = \text{RSS}/(n - p)$ . ■

Estimator ini merupakan generalisasi dari estimator variansi GWR klasik dengan koreksi derajat kebebasan. Sebab  $\sum_i w_i = 1$ , penyebut menjadi  $1 - p \cdot \bar{w}$  dengan  $\bar{w}$  adalah rata-rata bobot, yang untuk sampel besar mendekati 1.

Untuk menganalisis sifat asimtotik  $\hat{\sigma}_{\text{CF}}^2$ , residual didekomposisikan sebagai berikut.

**Proposisi 3.4.9** *Residual tertimbang dapat ditulis sebagai*

$$y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\text{CF}}(\mathbf{u}_0) = \varepsilon_i - \mathbf{x}_i^\top (\hat{\boldsymbol{\beta}}_{\text{CF}}(\mathbf{u}_0) - \boldsymbol{\beta}(\mathbf{u}_i)). \quad (3.4.42)$$

*Dengan menambah dan mengurangi  $\boldsymbol{\beta}(\mathbf{u}_0)$ :*

$$y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\text{CF}}(\mathbf{u}_0) = \varepsilon_i + \mathbf{x}_i^\top (\boldsymbol{\beta}(\mathbf{u}_i) - \boldsymbol{\beta}(\mathbf{u}_0)) - \mathbf{x}_i^\top (\hat{\boldsymbol{\beta}}_{\text{CF}}(\mathbf{u}_0) - \boldsymbol{\beta}(\mathbf{u}_0)). \quad (3.4.43)$$

**Bukti.** Dari model struktural  $y_i = \mathbf{x}_i^\top \boldsymbol{\beta}(\mathbf{u}_i) + \varepsilon_i$ , maka

$$\begin{aligned} y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\text{CF}} &= \mathbf{x}_i^\top \boldsymbol{\beta}(\mathbf{u}_i) + \varepsilon_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\text{CF}} \\ &= \varepsilon_i + \mathbf{x}_i^\top (\boldsymbol{\beta}(\mathbf{u}_i) - \hat{\boldsymbol{\beta}}_{\text{CF}}) \\ &= \varepsilon_i + \mathbf{x}_i^\top (\boldsymbol{\beta}(\mathbf{u}_i) - \boldsymbol{\beta}(\mathbf{u}_0)) - \mathbf{x}_i^\top (\hat{\boldsymbol{\beta}}_{\text{CF}} - \boldsymbol{\beta}(\mathbf{u}_0)). \end{aligned} \quad (3.4.44)$$

■

Apabila didefinisikan notasi ringkas, yaitu  $\delta_i := \boldsymbol{\beta}(\mathbf{u}_i) - \boldsymbol{\beta}(\mathbf{u}_0)$  adalah deviasi koefisien lokal dari target dan  $\hat{\boldsymbol{\Delta}} := \hat{\boldsymbol{\beta}}_{\text{CF}}(\mathbf{u}_0) - \boldsymbol{\beta}(\mathbf{u}_0) = \mathbf{B}_n + \mathbf{V}_n$  adalah galat estimasi, residual menjadi

$$\hat{e}_i = y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\text{CF}} = \varepsilon_i + \mathbf{x}_i^\top \delta_i - \mathbf{x}_i^\top \hat{\boldsymbol{\Delta}}. \quad (3.4.45)$$

**Teorema 3.4.10** *Jumlah kuadrat residual tertimbang dapat diekspansi sebagai*

$$\begin{aligned} \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \hat{e}_i^2 &= \underbrace{\sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \varepsilon_i^2}_{(I): \text{ suku utama}} + \underbrace{\sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} (\mathbf{x}_i^\top \delta_i)^2}_{(II): \text{ bias lokal}} \\ &\quad + \underbrace{\hat{\boldsymbol{\Delta}}^\top \left( \sum_{k=1}^K \mathbf{H}_k \right) \hat{\boldsymbol{\Delta}}}_{(III): \text{ galat estimasi}} + (\text{suku silang}). \end{aligned} \quad (3.4.46)$$

**Bukti.** Dengan mengkuadratkan ekspresi residual akan didapatkan

$$\begin{aligned} \hat{e}_i^2 &= (\varepsilon_i + \mathbf{x}_i^\top \delta_i - \mathbf{x}_i^\top \hat{\boldsymbol{\Delta}})^2 \\ &= \varepsilon_i^2 + (\mathbf{x}_i^\top \delta_i)^2 + (\mathbf{x}_i^\top \hat{\boldsymbol{\Delta}})^2 \\ &\quad + 2\varepsilon_i \mathbf{x}_i^\top \delta_i - 2\varepsilon_i \mathbf{x}_i^\top \hat{\boldsymbol{\Delta}} - 2(\mathbf{x}_i^\top \delta_i)(\mathbf{x}_i^\top \hat{\boldsymbol{\Delta}}). \end{aligned} \quad (3.4.47)$$

Selanjutnya, dengan menjumlahkan bobot

$$\begin{aligned}
\sum_{k,i} w_i^{(-k)} \tilde{e}_i^2 &= \sum_{k,i} w_i^{(-k)} \varepsilon_i^2 + \sum_{k,i} w_i^{(-k)} (\mathbf{x}_i^\top \delta_i)^2 + \sum_{k,i} w_i^{(-k)} (\mathbf{x}_i^\top \hat{\Delta})^2 \\
&\quad + 2 \sum_{k,i} w_i^{(-k)} \varepsilon_i \mathbf{x}_i^\top \delta_i - 2 \sum_{k,i} w_i^{(-k)} \varepsilon_i \mathbf{x}_i^\top \hat{\Delta} \\
&\quad - 2 \sum_{k,i} w_i^{(-k)} (\mathbf{x}_i^\top \delta_i) (\mathbf{x}_i^\top \hat{\Delta}).
\end{aligned} \tag{3.4.48}$$

Untuk suku (III), perhatikan bahwa

$$\begin{aligned}
\sum_{k,i} w_i^{(-k)} (\mathbf{x}_i^\top \hat{\Delta})^2 &= \sum_{k,i} w_i^{(-k)} \hat{\Delta}^\top \mathbf{x}_i \mathbf{x}_i^\top \hat{\Delta} \\
&= \hat{\Delta}^\top \left( \sum_{k,i} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top \right) \hat{\Delta} \\
&= \hat{\Delta}^\top \left( \sum_{k=1}^K \mathbf{H}_k \right) \hat{\Delta}.
\end{aligned} \tag{3.4.49}$$

■

**Teorema 3.4.11** *Di bawah Asumsi GA-GWR.1–GA-GWR.10, estimator variansi galat cross-fitted bersifat konsisten:*

$$\hat{\sigma}_{\text{CF}}^2(\mathbf{u}_0) \xrightarrow{p} \sigma^2. \tag{3.4.50}$$

**Bukti.** Pada suku utama, berdasarkan hukum bilangan besar tertimbang dan Asumsi GA-GWR.3:

$$\sum_{k,i} w_i^{(-k)} \varepsilon_i^2 \xrightarrow{p} \mathbb{E}[\varepsilon_i^2] = \sigma^2. \tag{3.4.51}$$

Selanjutnya, pada bias lokal, karena  $\|\delta_i\| = \|\beta(\mathbf{u}_i) - \beta(\mathbf{u}_0)\| = \mathcal{O}(h)$  untuk  $\mathbf{u}_i \in \mathcal{N}_h(\mathbf{u}_0)$ , maka

$$\sum_{k,i} w_i^{(-k)} (\mathbf{x}_i^\top \delta_i)^2 \leq \sum_{k,i} w_i^{(-k)} \|\mathbf{x}_i\|^2 \|\delta_i\|^2 = \mathcal{O}_p(h^2) \rightarrow 0. \tag{3.4.52}$$

Dari hasil sebelumnya,  $\hat{\Delta} = \mathcal{O}_p(h^2) + \mathcal{O}_p((nh^d)^{-1/2})$ . Maka

$$\hat{\Delta}^\top \left( \sum_k \mathbf{H}_k \right) \hat{\Delta} = \mathcal{O}_p(h^4 + (nh^d)^{-1}) \rightarrow 0. \quad (3.4.53)$$

Dengan argumen serupa, semua suku silang konvergen ke nol dalam probabilitas karena melibatkan produk dari suku yang konvergen ke nol. Dengan demikian,

$$\sum_{k,i} w_i^{(-k)} \hat{e}_i^2 \xrightarrow{p} \sigma^2, \quad (3.4.54)$$

dan karena penyebut  $\sum_{k,i} w_i^{(-k)} - p \rightarrow 1 - 0 = 1$  (dengan koreksi derajat bebas yang dapat dihilangkan secara asimtotik), maka  $\hat{\sigma}_{\text{CF}}^2 \xrightarrow{p} \sigma^2$ . ■

**Akibat 3.4.12** *Bias estimator variansi galat memenuhi*

$$\mathbb{E}[\hat{\sigma}_{\text{CF}}^2(\mathbf{u}_0)] - \sigma^2 = \mathcal{O}(h^2) + \mathcal{O}((nh^d)^{-1}), \quad (3.4.55)$$

yang konvergen ke nol di bawah kondisi asimtotik GA-GWR.6 dan GA-GWR.7.

### 3.4.4 Konsistensi Koefisien Lokal

Bagian ini membuktikan konsistensi estimator *cross-fitted* GA-GWR secara formal. Konsistensi merupakan syarat minimal untuk validitas estimator, yaitu estimator harus konvergen ke nilai parameter sejati ketika ukuran sampel menuju tak hingga.

**Teorema 3.4.13** *Di bawah Asumsi GA-GWR.1–GA-GWR.10, estimator cross-fitted GA-GWR bersifat konsisten, yaitu*

$$\hat{\beta}_{\text{CF}}(\mathbf{u}_0) \xrightarrow{p} \beta(\mathbf{u}_0) \quad \text{ketika } n \rightarrow \infty. \quad (3.4.56)$$

**Bukti.** Dari Teorema 3.4.2, berlaku

$$\hat{\beta}_{\text{CF}}(\mathbf{u}_0) - \beta(\mathbf{u}_0) = \mathbf{B}_n(\mathbf{u}_0) + \mathbf{V}_n(\mathbf{u}_0). \quad (3.4.57)$$



Untuk membuktikan konsistensi, perlu ditunjukkan bahwa  $\mathbf{B}_n(\mathbf{u}_0) \xrightarrow{p} \mathbf{0}$  dan komponen galat  $\mathbf{V}_n(\mathbf{u}_0) \xrightarrow{p} \mathbf{0}$ . Dari Teorema 3.4.6, telah ditunjukkan bahwa

$$\mathbf{B}_n(\mathbf{u}_0) = \mathcal{O}_p(h^2). \quad (3.4.58)$$

Sebab  $h \rightarrow 0$ , berdasarkan Asumsi GA-GWR.6, maka  $\mathbf{B}_n(\mathbf{u}_0) \xrightarrow{p} \mathbf{0}$ .

Suku galat diberikan oleh

$$\mathbf{V}_n(\mathbf{u}_0) = \left( \sum_{k=1}^K \mathbf{H}_k \right)^{-1} \left( \sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)} \boldsymbol{\varepsilon}_k \right). \quad (3.4.59)$$

Untuk menganalisis konvergensi, ditulis dalam bentuk yang lebih eksplisit:

$$\mathbf{V}_n(\mathbf{u}_0) = \left( \frac{1}{nh^d} \sum_{k=1}^K \mathbf{H}_k \right)^{-1} \left( \frac{1}{nh^d} \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \varepsilon_i \right). \quad (3.4.60)$$

Didefinisikan matriks

$$\mathbf{Q}_n(\mathbf{u}_0) := \frac{1}{nh^d} \sum_{k=1}^K \mathbf{H}_k = \frac{1}{nh^d} \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top. \quad (3.4.61)$$

Berdasarkan hukum bilangan besar tertimbang lokal, dengan menggunakan Asumsi GA-GWR.1, GA-GWR.4, GA-GWR.6, dan GA-GWR.10:

$$\mathbf{Q}_n(\mathbf{u}_0) \xrightarrow{p} \mathbf{Q}(\mathbf{u}_0), \quad (3.4.62)$$

dengan  $\mathbf{Q}(\mathbf{u}_0) = \mathbb{E}[\mathbf{x}_i \mathbf{x}_i^\top \mid \mathbf{u}_i \approx \mathbf{u}_0]$  adalah matriks momen lokal yang positif definit. Akibatnya,

$$\mathbf{Q}_n(\mathbf{u}_0)^{-1} \xrightarrow{p} \mathbf{Q}(\mathbf{u}_0)^{-1}. \quad (3.4.63)$$

Selanjutnya, didefinisikan vektor

$$\mathbf{S}_n(\mathbf{u}_0) := \frac{1}{nh^d} \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \varepsilon_i. \quad (3.4.64)$$

Berdasarkan Proposisi 3.3.9, untuk setiap *fold*  $k$ :

$$\mathbb{E}\left[w_i^{(-k)} \mathbf{x}_i \varepsilon_i \mid \widehat{\boldsymbol{\theta}}^{(-k)}\right] = \mathbf{0} \quad \text{untuk } i \in \mathcal{I}_k. \quad (3.4.65)$$

Dengan menggunakan hukum ekspektasi iterasi:

$$\mathbb{E}[\mathbf{S}_n(\mathbf{u}_0)] = \frac{1}{nh^d} \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} \mathbb{E}\left[\mathbb{E}\left[w_i^{(-k)} \mathbf{x}_i \varepsilon_i \mid \widehat{\boldsymbol{\theta}}^{(-k)}\right]\right] = \mathbf{0}. \quad (3.4.66)$$

Untuk variansi, dengan menggunakan independensi antarobservasi GA-GWR.1 dan Asumsi GA-GWR.3:

$$\begin{aligned} \text{Var}(\mathbf{S}_n(\mathbf{u}_0)) &= \frac{1}{(nh^d)^2} \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} \mathbb{E}\left[(w_i^{(-k)})^2 \mathbf{x}_i \mathbf{x}_i^\top \varepsilon_i^2\right] \\ &= \frac{\sigma^2}{(nh^d)^2} \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} \mathbb{E}\left[(w_i^{(-k)})^2 \mathbf{x}_i \mathbf{x}_i^\top\right]. \end{aligned} \quad (3.4.67)$$

Berdasarkan Asumsi GA-GWR.7,  $w_i^{(-k)} \leq C/(nh^d)$ , sehingga  $(w_i^{(-k)})^2 \leq C^2/(nh^d)^2$ . Dengan demikian,

$$\begin{aligned} \text{Var}(\mathbf{S}_n(\mathbf{u}_0)) &\leq \frac{\sigma^2 C^2}{(nh^d)^4} \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} \mathbb{E}[\mathbf{x}_i \mathbf{x}_i^\top] \\ &= \frac{\sigma^2 C^2}{(nh^d)^4} \cdot n \cdot \mathbb{E}[\mathbf{x}_i \mathbf{x}_i^\top] \\ &= \mathcal{O}((nh^d)^{-3}) \rightarrow 0. \end{aligned} \quad (3.4.68)$$

Sebab  $\mathbb{E}[\mathbf{S}_n] = \mathbf{0}$  dan  $\text{Var}(\mathbf{S}_n) \rightarrow 0$ , maka berdasarkan ketaksamaan Chebyshev:

$$\mathbf{S}_n(\mathbf{u}_0) \xrightarrow{p} \mathbf{0}. \quad (3.4.69)$$

Sebab  $\mathbf{Q}_n^{-1} \xrightarrow{p} \mathbf{Q}^{-1}$  dan  $\mathbf{S}_n \xrightarrow{p} \mathbf{0}$ , maka berdasarkan lemma Slutsky:

$$\mathbf{V}_n(\mathbf{u}_0) = \mathbf{Q}_n^{-1} \cdot (nh^d) \cdot \mathbf{S}_n \xrightarrow{p} \mathbf{Q}^{-1} \cdot \mathbf{0} = \mathbf{0}. \quad (3.4.70)$$

Dengan menggabungkan hasil untuk suku bias dan suku galat:

$$\widehat{\beta}_{\text{CF}}(\mathbf{u}_0) - \beta(\mathbf{u}_0) = \mathbf{B}_n + \mathbf{V}_n \xrightarrow{p} \mathbf{0} + \mathbf{0} = \mathbf{0}. \quad (3.4.71)$$

■

**Akibat 3.4.14** *Laju konvergensi estimator cross-fitted ditentukan oleh*

$$\widehat{\beta}_{\text{CF}}(\mathbf{u}_0) - \beta(\mathbf{u}_0) = \mathcal{O}_p(h^2) + \mathcal{O}_p((nh^d)^{-1/2}), \quad (3.4.72)$$

dengan suku pertama adalah kontribusi bias dan suku kedua adalah kontribusi variansi.

### 3.4.5 Konsistensi Variansi Koefisien Lokal

Untuk melakukan inferensi statistik, diperlukan estimasi variansi dari estimator koefisien lokal. Bagian ini mendefinisikan variansi teoretis, menurunkan estimatornya, dan membuktikan konsistensinya.

**Teorema 3.4.15** *Di bawah Asumsi GA-GWR.1–GA-GWR.10, variansi kondisional dari estimator cross-fitted memiliki bentuk sandwich:*

$$\text{Var}(\widehat{\beta}_{\text{CF}}(\mathbf{u}_0) \mid \{\widehat{\theta}^{(-k)}\}_{k=1}^K) = \mathbf{Q}_n(\mathbf{u}_0)^{-1} \mathbf{\Omega}_n(\mathbf{u}_0) \mathbf{Q}_n(\mathbf{u}_0)^{-1}, \quad (3.4.73)$$

dengan

$$\mathbf{Q}_n(\mathbf{u}_0) = \sum_{k=1}^K \mathbf{H}_k = \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top, \quad (3.4.74)$$

$$\mathbf{\Omega}_n(\mathbf{u}_0) = \sigma^2 \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} (w_i^{(-k)})^2 \mathbf{x}_i \mathbf{x}_i^\top. \quad (3.4.75)$$

**Bukti.** Dari dekomposisi estimator, suku galat diberikan oleh

$$\mathbf{V}_n(\mathbf{u}_0) = \left( \sum_{k=1}^K \mathbf{H}_k \right)^{-1} \left( \sum_{k=1}^K \mathbf{X}_k^\top \mathbf{W}_k^{(-k)} \boldsymbol{\varepsilon}_k \right) = \mathbf{Q}_n^{-1} \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \varepsilon_i. \quad (3.4.76)$$

Kondisional pada  $\{\widehat{\boldsymbol{\theta}}^{(-k)}\}$ , bobot  $w_i^{(-k)}$  bersifat deterministik. Dengan menggunakan independensi antarobservasi dan Asumsi GA-GWR.3:

$$\begin{aligned} \text{Var}\left(\sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \varepsilon_i \mid \{\widehat{\boldsymbol{\theta}}^{(-k)}\}\right) &= \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} (w_i^{(-k)})^2 \mathbf{x}_i \mathbf{x}_i^\top \text{Var}(\varepsilon_i) \\ &= \sigma^2 \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} (w_i^{(-k)})^2 \mathbf{x}_i \mathbf{x}_i^\top \\ &= \boldsymbol{\Omega}_n(\mathbf{u}_0). \end{aligned} \quad (3.4.77)$$

Dengan menggunakan sifat variansi transformasi linear:

$$\begin{aligned} \text{Var}(\mathbf{V}_n \mid \{\widehat{\boldsymbol{\theta}}^{(-k)}\}) &= \mathbf{Q}_n^{-1} \text{Var}\left(\sum_{k,i} w_i^{(-k)} \mathbf{x}_i \varepsilon_i\right) \mathbf{Q}_n^{-1} \\ &= \mathbf{Q}_n^{-1} \boldsymbol{\Omega}_n \mathbf{Q}_n^{-1}. \end{aligned} \quad (3.4.78)$$

Sebab suku bias  $\mathbf{B}_n$  bersifat deterministik kondisional pada  $\{\widehat{\boldsymbol{\theta}}^{(-k)}\}$ , maka

$$\text{Var}(\widehat{\boldsymbol{\beta}}_{\text{CF}} \mid \{\widehat{\boldsymbol{\theta}}^{(-k)}\}) = \text{Var}(\mathbf{V}_n \mid \{\widehat{\boldsymbol{\theta}}^{(-k)}\}) = \mathbf{Q}_n^{-1} \boldsymbol{\Omega}_n \mathbf{Q}_n^{-1}. \quad (3.4.79)$$

■

**Akibat 3.4.16** *Variansi koefisien lokal memiliki orde*

$$\text{Var}(\widehat{\boldsymbol{\beta}}_{\text{CF}}(\mathbf{u}_0)) = \mathcal{O}((nh^d)^{-1}). \quad (3.4.80)$$

*Secara intuitif, effective local sample size adalah  $nh^d$ , sehingga variansi berbanding terbalik dengannya.*

**Bukti.** Dari Asumsi GA-GWR.7,  $w_i^{(-k)} \leq C/(nh^d)$ . Dengan demikian,

$$\boldsymbol{\Omega}_n = \sigma^2 \sum_{k,i} (w_i^{(-k)})^2 \mathbf{x}_i \mathbf{x}_i^\top \leq \frac{\sigma^2 C}{nh^d} \sum_{k,i} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top = \frac{\sigma^2 C}{nh^d} \mathbf{Q}_n. \quad (3.4.81)$$

Sebab  $\mathbf{Q}_n = \mathcal{O}(1)$  dan  $\mathbf{\Omega}_n = \mathcal{O}((nh^d)^{-1})$ , maka

$$\mathbf{Q}_n^{-1} \mathbf{\Omega}_n \mathbf{Q}_n^{-1} = \mathcal{O}(1) \cdot \mathcal{O}((nh^d)^{-1}) \cdot \mathcal{O}(1) = \mathcal{O}((nh^d)^{-1}). \quad (3.4.82)$$

■

**Teorema 3.4.17** *Estimator variansi koefisien lokal cross-fitted yang diperoleh dengan mengganti parameter populasi dengan estimator konsistennya diberikan oleh*

$$\widehat{\text{Var}}(\widehat{\beta}_{\text{CF}}(\mathbf{u}_0)) = \widehat{\mathbf{Q}}_n(\mathbf{u}_0)^{-1} \widehat{\mathbf{\Omega}}_n(\mathbf{u}_0) \widehat{\mathbf{Q}}_n(\mathbf{u}_0)^{-1}, \quad (3.4.83)$$

dengan

$$\widehat{\mathbf{Q}}_n(\mathbf{u}_0) = \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top = \sum_{k=1}^K \mathbf{H}_k, \quad (3.4.84)$$

$$\widehat{\mathbf{\Omega}}_n(\mathbf{u}_0) = \widehat{\sigma}_{\text{CF}}^2(\mathbf{u}_0) \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} (w_i^{(-k)})^2 \mathbf{x}_i \mathbf{x}_i^\top. \quad (3.4.85)$$

**Bukti.** Dari Teorema 3.4.15, variansi kondisional sejati diberikan oleh

$$\text{Var}(\widehat{\beta}_{\text{CF}}(\mathbf{u}_0) \mid \{\widehat{\theta}^{(-k)}\}_{k=1}^K) = \mathbf{Q}_n(\mathbf{u}_0)^{-1} \mathbf{\Omega}_n(\mathbf{u}_0) \mathbf{Q}_n(\mathbf{u}_0)^{-1}, \quad (3.4.86)$$

dengan

$$\mathbf{Q}_n(\mathbf{u}_0) = \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top, \quad (3.4.87)$$

$$\mathbf{\Omega}_n(\mathbf{u}_0) = \sigma^2 \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} (w_i^{(-k)})^2 \mathbf{x}_i \mathbf{x}_i^\top. \quad (3.4.88)$$

Perhatikan bahwa  $\mathbf{Q}_n(\mathbf{u}_0)$  dapat dihitung langsung dari data karena hanya melibatkan bobot  $w_i^{(-k)}$  dan kovariat  $\mathbf{x}_i$  yang keduanya teramati. Dengan demikian,

estimator untuk  $\mathbf{Q}_n(\mathbf{u}_0)$  adalah dirinya sendiri:

$$\hat{\mathbf{Q}}_n(\mathbf{u}_0) = \mathbf{Q}_n(\mathbf{u}_0) = \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top = \sum_{k=1}^K \mathbf{H}_k. \quad (3.4.89)$$

Untuk  $\Omega_n(\mathbf{u}_0)$ , satu-satunya kuantitas yang tidak teramati adalah variansi galat  $\sigma^2$ . Berdasarkan Teorema 3.4.11, estimator konsisten untuk  $\sigma^2$  adalah  $\hat{\sigma}_{\text{CF}}^2(\mathbf{u}_0)$ . Dengan mensubstitusikan  $\sigma^2$  dengan estimatornya, diperoleh

$$\hat{\Omega}_n(\mathbf{u}_0) = \hat{\sigma}_{\text{CF}}^2(\mathbf{u}_0) \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} (w_i^{(-k)})^2 \mathbf{x}_i \mathbf{x}_i^\top. \quad (3.4.90)$$

Dengan menggabungkan kedua estimator ke dalam bentuk *sandwich*, diperoleh estimator variansi koefisien lokal:

$$\widehat{\text{Var}}(\hat{\beta}_{\text{CF}}(\mathbf{u}_0)) = \hat{\mathbf{Q}}_n(\mathbf{u}_0)^{-1} \hat{\Omega}_n(\mathbf{u}_0) \hat{\mathbf{Q}}_n(\mathbf{u}_0)^{-1}. \quad (3.4.91)$$

Bentuk *sandwich* ini memiliki interpretasi geometris bahwa matriks  $\hat{\mathbf{Q}}_n^{-1}$  mentransformasi variabilitas dalam ruang kovariat tertimbang  $\hat{\Omega}_n$  menjadi variabilitas dalam ruang parameter  $\beta$ . ■

Estimator ini menggantikan variansi galat sejati  $\sigma^2$  dengan estimator konsistennya  $\hat{\sigma}_{\text{CF}}^2$ . Bentuk *sandwich* menjamin *robustness* terhadap spesifikasi yang tidak tepat pada struktur variansi.

**Teorema 3.4.18** *Di bawah Asumsi GA-GWR.1–GA-GWR.10, estimator variansi bersifat konsisten dalam arti*

$$(nh^d) \cdot \widehat{\text{Var}}(\hat{\beta}_{\text{CF}}(\mathbf{u}_0)) \xrightarrow{p} \mathbf{Q}(\mathbf{u}_0)^{-1} \Omega(\mathbf{u}_0) \mathbf{Q}(\mathbf{u}_0)^{-1}, \quad (3.4.92)$$

dengan  $\mathbf{Q}(\mathbf{u}_0)$  dan  $\Omega(\mathbf{u}_0)$  adalah limit probabilitas dari versi ternormalisasi  $\mathbf{Q}_n$  dan  $\Omega_n$ .

**Bukti.** Perhatikan bahwa

$$\frac{1}{nh^d} \widehat{\mathbf{Q}}_n = \frac{1}{nh^d} \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top. \quad (3.4.93)$$

Berdasarkan hukum bilangan besar tertimbang lokal, didapatkan bahwa

$$\frac{1}{nh^d} \widehat{\mathbf{Q}}_n \xrightarrow{p} \mathbf{Q}(\mathbf{u}_0) = \mathbb{E}[\mathbf{x}_i \mathbf{x}_i^\top \mid \mathbf{u}_i \approx \mathbf{u}_0]. \quad (3.4.94)$$

Didefinisikan

$$\widetilde{\mathbf{\Omega}}_n = (nh^d) \cdot \widehat{\mathbf{\Omega}}_n = (nh^d) \cdot \widehat{\sigma}_{\text{CF}}^2 \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} (w_i^{(-k)})^2 \mathbf{x}_i \mathbf{x}_i^\top. \quad (3.4.95)$$

Dari Teorema 3.4.11,  $\widehat{\sigma}_{\text{CF}}^2 \xrightarrow{p} \sigma^2$ . Untuk suku penjumlahan, dengan menggunakan Asumsi GA-GWR.7:

$$(nh^d) \sum_{k,i} (w_i^{(-k)})^2 \mathbf{x}_i \mathbf{x}_i^\top \xrightarrow{p} \mathbf{\Omega}'(\mathbf{u}_0), \quad (3.4.96)$$

dengan  $\mathbf{\Omega}'(\mathbf{u}_0)$  adalah limit yang terkait dengan momen bobot kuadrat lokal. Didefinisikan  $\mathbf{\Omega}(\mathbf{u}_0) = \sigma^2 \mathbf{\Omega}'(\mathbf{u}_0)$ . Oleh karena itu,

$$\widetilde{\mathbf{\Omega}}_n \xrightarrow{p} \sigma^2 \mathbf{\Omega}'(\mathbf{u}_0) = \mathbf{\Omega}(\mathbf{u}_0). \quad (3.4.97)$$

Estimator variansi ternormalisasi dapat ditulis sebagai

$$(nh^d) \cdot \widehat{\text{Var}}(\widehat{\beta}_{\text{CF}}) = \left( \frac{1}{nh^d} \widehat{\mathbf{Q}}_n \right)^{-1} \widetilde{\mathbf{\Omega}}_n \left( \frac{1}{nh^d} \widehat{\mathbf{Q}}_n \right)^{-1}. \quad (3.4.98)$$

Selanjutnya, sebab  $\frac{1}{nh^d} \widehat{\mathbf{Q}}_n \xrightarrow{p} \mathbf{Q}(\mathbf{u}_0)$  yang invertibel,  $\widetilde{\mathbf{\Omega}}_n \xrightarrow{p} \mathbf{\Omega}(\mathbf{u}_0)$ , dan fungsi inversi matriks kontinu pada himpunan matriks invertibel, maka berdasarkan *continuous mapping theorem* didapatkan

$$(nh^d) \cdot \widehat{\text{Var}}(\widehat{\beta}_{\text{CF}}) \xrightarrow{p} \mathbf{Q}(\mathbf{u}_0)^{-1} \mathbf{\Omega}(\mathbf{u}_0) \mathbf{Q}(\mathbf{u}_0)^{-1}. \quad (3.4.99)$$



**Akibat 3.4.19** Galat standar untuk komponen ke- $j$  dari koefisien lokal diestimasi sebagai

$$\widehat{\text{se}}(\widehat{\beta}_{\text{CF},j}(\mathbf{u}_0)) = \sqrt{[\widehat{\text{Var}}(\widehat{\boldsymbol{\beta}}_{\text{CF}}(\mathbf{u}_0))]_{jj}}, \quad (3.4.100)$$

dengan  $[\cdot]_{jj}$  menandakan elemen diagonal ke- $j$  dari matriks variansi-kovariansi.

### 3.4.6 Distribusi Asimtotik Koefisien Lokal

Bagian ini menurunkan distribusi asimtotik dari estimator *cross-fitted* GA-GWR. Hasil utama menunjukkan bahwa setelah dinormalisasi dengan  $\sqrt{nh^d}$ , estimator berdistribusi normal asimtotik di bawah kondisi *undersmoothing*.

Berdasarkan Akibat 3.4.16, variansi estimator memiliki orde  $\mathcal{O}((nh^d)^{-1})$ . Hal ini menunjukkan bahwa *effective local sample size* adalah  $nh^d$ , sehingga normalisasi yang tepat adalah  $\sqrt{nh^d}$ .

**Proposisi 3.4.20 (Dekomposisi Terstandarisasi)** Estimator *cross-fitted* yang dinormalisasi dapat didekomposisikan sebagai

$$\sqrt{nh^d}(\widehat{\boldsymbol{\beta}}_{\text{CF}}(\mathbf{u}_0) - \boldsymbol{\beta}(\mathbf{u}_0)) = \underbrace{\sqrt{nh^d}\mathbf{B}_n(\mathbf{u}_0)}_{\text{bias terstandarisasi}} + \underbrace{\sqrt{nh^d}\mathbf{V}_n(\mathbf{u}_0)}_{\text{galat terstandarisasi}}. \quad (3.4.101)$$

Distribusi limit ditentukan oleh perilaku kedua suku ini.

Suku galat terstandarisasi dapat ditulis dalam bentuk yang memfasilitasi penerapan Teorema Limit Pusat.

**Proposisi 3.4.21** Suku galat terstandarisasi dapat ditulis sebagai

$$\sqrt{nh^d}\mathbf{V}_n(\mathbf{u}_0) = \left( \frac{1}{nh^d} \sum_{k=1}^K \mathbf{H}_k \right)^{-1} \left( \frac{1}{\sqrt{nh^d}} \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \varepsilon_i \right), \quad (3.4.102)$$

yang merupakan produk dari dua faktor, yaitu satu yang konvergen dalam probabilitas dan satu yang konvergen dalam distribusi.



**Teorema 3.4.22** Di bawah Asumsi GA-GWR.1–GA-GWR.10, didefinisikan

$$\mathbf{S}_n(\mathbf{u}_0) = \frac{1}{\sqrt{nh^d}} \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \varepsilon_i. \quad (3.4.103)$$

Maka

$$\mathbf{S}_n(\mathbf{u}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{\Omega}(\mathbf{u}_0)), \quad (3.4.104)$$

dengan  $\mathbf{\Omega}(\mathbf{u}_0)$  adalah limit dari  $\mathbf{\Omega}_n(\mathbf{u}_0)$  ternormalisasi.

**Bukti.** Bukti diuraikan menggunakan Teorema Limit Pusat Lindeberg-Feller untuk *triangular array* dari variabel acak. Berdasarkan Proposisi 3.3.9, untuk setiap *fold*  $k$  dan  $i \in \mathcal{I}_k$  berlaku

$$\mathbb{E}[w_i^{(-k)} \mathbf{x}_i \varepsilon_i \mid \hat{\boldsymbol{\theta}}^{(-k)}] = w_i^{(-k)} \mathbf{x}_i \mathbb{E}[\varepsilon_i \mid \mathbf{x}_i, \mathbf{u}_i] = \mathbf{0}, \quad (3.4.105)$$

karena bobot  $w_i^{(-k)}$  dilatih pada data yang tidak termasuk observasi  $i$ , sehingga bersifat deterministik kondisional pada  $\hat{\boldsymbol{\theta}}^{(-k)}$ . Dengan demikian,  $\mathbb{E}[\mathbf{S}_n] = \mathbf{0}$ .

Variansi kondisional diberikan oleh

$$\begin{aligned} \text{Var}(\mathbf{S}_n \mid \{\hat{\boldsymbol{\theta}}^{(-k)}\}) &= \frac{1}{nh^d} \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} (w_i^{(-k)})^2 \mathbb{E}[\mathbf{x}_i \mathbf{x}_i^\top \varepsilon_i^2 \mid \mathbf{u}_i] \\ &= \frac{\sigma^2}{nh^d} \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} (w_i^{(-k)})^2 \mathbf{x}_i \mathbf{x}_i^\top. \end{aligned} \quad (3.4.106)$$

Selanjutnya, didefinisikan

$$\mathbf{\Omega}'_n := \frac{nh^d}{1} \cdot \frac{\sigma^2}{nh^d} \sum_{k,i} (w_i^{(-k)})^2 \mathbf{x}_i \mathbf{x}_i^\top = \sigma^2 \sum_{k,i} (w_i^{(-k)})^2 \mathbf{x}_i \mathbf{x}_i^\top. \quad (3.4.107)$$

Berdasarkan hukum bilangan besar,  $\mathbf{\Omega}'_n \xrightarrow{p} \mathbf{\Omega}(\mathbf{u}_0)$  yang positif definit dan terbatas.

Untuk CLT Lindeberg-Feller, perlu diverifikasi bahwa untuk setiap  $\delta > 0$

harus terpenuhi kondisi Lindeberg, yaitu

$$\lim_{n \rightarrow \infty} \frac{1}{\text{Var}(\mathbf{S}_n)} \sum_{k,i} \mathbb{E}[\|w_i^{(-k)} \mathbf{x}_i \varepsilon_i\|^2 \mathbf{1}_{\{\|w_i^{(-k)} \mathbf{x}_i \varepsilon_i\| > \delta \sqrt{nh^d}\}}] = 0. \quad (3.4.108)$$

Berdasarkan Asumsi GA-GWR.7,  $w_i^{(-k)} \leq C/(nh^d)$ . Dengan demikian,

$$\|w_i^{(-k)} \mathbf{x}_i \varepsilon_i\| \leq \frac{C}{nh^d} \|\mathbf{x}_i\| |\varepsilon_i| \rightarrow 0 \quad (3.4.109)$$

untuk  $n$  besar, karena  $nh^d \rightarrow \infty$  (Asumsi GA-GWR.7). Ini berarti untuk  $n$  cukup besar,  $\|w_i^{(-k)} \mathbf{x}_i \varepsilon_i\| < \delta \sqrt{nh^d}$  hampir pasti, sehingga kondisi Lindeberg terpenuhi.

Dengan demikian, karena suku-suku  $w_i^{(-k)} \mathbf{x}_i \varepsilon_i$  independen antarobservasi, rata-rata nol, variansi total terbatas dan definit positif, dan kondisi Lindeberg terpenuhi, maka berdasarkan Teorema Limit Pusat Lindeberg-Feller akan dihasilkan

$$\mathbf{S}_n(\mathbf{u}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{\Omega}(\mathbf{u}_0)). \quad (3.4.110)$$

■

**Teorema 3.4.23** *Di bawah Asumsi GA-GWR.1–GA-GWR.10:*

$$\sqrt{nh^d} \mathbf{V}_n(\mathbf{u}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{Q}(\mathbf{u}_0)^{-1} \mathbf{\Omega}(\mathbf{u}_0) \mathbf{Q}(\mathbf{u}_0)^{-1}). \quad (3.4.111)$$

**Bukti.** Berdasarkan Teorema 3.4.22, didapatkan  $\frac{1}{nh^d} \sum_k \mathbf{H}_k \xrightarrow{p} \mathbf{Q}(\mathbf{u}_0)$  yang invertibel (definit positif) dan  $\mathbf{S}_n \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{\Omega}(\mathbf{u}_0))$ . Sebab satu faktor konvergen dalam probabilitas ke matriks invertibel dan faktor lainnya konvergen dalam distribusi, maka berdasarkan Lemma Slutsky:

$$\begin{aligned} \sqrt{nh^d} \mathbf{V}_n &= \left( \frac{1}{nh^d} \sum_k \mathbf{H}_k \right)^{-1} \mathbf{S}_n \\ &\xrightarrow{d} \mathbf{Q}(\mathbf{u}_0)^{-1} \cdot \mathcal{N}(\mathbf{0}, \mathbf{\Omega}(\mathbf{u}_0)) \\ &= \mathcal{N}(\mathbf{0}, \mathbf{Q}(\mathbf{u}_0)^{-1} \mathbf{\Omega}(\mathbf{u}_0) \mathbf{Q}(\mathbf{u}_0)^{-1}). \end{aligned} \quad (3.4.112)$$

■

Perilaku suku bias terstandarisasi  $\sqrt{nh^d}\mathbf{B}_n(\mathbf{u}_0)$  menentukan apakah distribusi asimtotik berpusat di nol atau tidak.

**Proposisi 3.4.24** *Dari Teorema 3.4.6,  $\mathbf{B}_n(\mathbf{u}_0) = \mathcal{O}_p(h^2)$ . Dengan demikian,*

$$\sqrt{nh^d}\mathbf{B}_n(\mathbf{u}_0) = \mathcal{O}_p(\sqrt{nh^d} \cdot h^2) = \mathcal{O}_p(\sqrt{nh^{d+4}}). \quad (3.4.113)$$

Terdapat dua skenario fundamental bergantung pada laju relatif antara  $nh^d$  dan  $h^2$ .

**Teorema 3.4.25** *Jika kondisi undersmoothing*

$$\sqrt{nh^d} \cdot h^2 \rightarrow 0 \quad \text{ketika } n \rightarrow \infty \quad (3.4.114)$$

*dipenuhi, maka suku bias terstandarisasi lenyap secara asimtotik:*

$$\sqrt{nh^d}\mathbf{B}_n(\mathbf{u}_0) \xrightarrow{p} \mathbf{0}. \quad (3.4.115)$$

**Bukti.** Sebab  $\mathbf{B}_n(\mathbf{u}_0) = \mathcal{O}_p(h^2)$ , maka  $\sqrt{nh^d}\mathbf{B}_n(\mathbf{u}_0) = \mathcal{O}_p(\sqrt{nh^d}h^2)$ . Di bawah kondisi  $\sqrt{nh^d}h^2 \rightarrow 0$ , berlaku  $\sqrt{nh^d}\mathbf{B}_n \xrightarrow{p} \mathbf{0}$ . ■

**Akibat 3.4.26** *Kondisi undersmoothing yang dapat dinyatakan dengan  $\sqrt{nh^d}h^2 \rightarrow 0$  ekuivalen dengan*

$$h = o(n^{-1/(d+4)}). \quad (3.4.116)$$

*Hal ini berarti bandwidth harus menyusut lebih cepat dari laju optimal untuk estimasi titik (yang biasanya  $h \asymp n^{-1/(d+4)}$ ).*

**Bukti.** Kondisi  $\sqrt{nh^d}h^2 \rightarrow 0$  ekuivalen dengan  $nh^{d+4} \rightarrow 0$ , yang ekuivalen dengan  $h^{d+4} = o(n^{-1})$ , atau  $h = o(n^{-1/(d+4)})$ . ■

**Proposisi 3.4.27** *Ketika menggunakan metrik alternatif dengan dimensi fraktal efektif  $d_{\text{eff}} \neq 2$ , kondisi undersmoothing menjadi*

$$\sqrt{nh^{d_{\text{eff}}}} \cdot h^2 \rightarrow 0 \quad \Longleftrightarrow \quad h = o(n^{-1/(d_{\text{eff}}+4)}). \quad (3.4.117)$$

Implikasi praktis untuk pemilihan *bandwidth*:

- (i) Metrik Euclidean dengan  $d_{\text{eff}} = 2$  dapat dipilih  $h = o(n^{-1/6}) \approx n^{-0.167}$
- (ii) Jarak jaringan urban dengan  $d_{\text{eff}} \approx 1$  dapat dipilih  $h = o(n^{-1/5}) \approx n^{-0.2}$
- (iii) Metrik Intermediat dengan  $d_{\text{eff}} = 1.5$  dapat dipilih  $h = o(n^{-1/5.5}) \approx n^{-0.182}$

**Teorema 3.4.28** *Di bawah Asumsi GA-GWR.1–GA-GWR.10 dan kondisi undersmoothing  $\sqrt{nh^d}h^2 \rightarrow 0$ :*

$$\sqrt{nh^d}(\hat{\beta}_{\text{CF}}(\mathbf{u}_0) - \beta(\mathbf{u}_0)) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{Q}(\mathbf{u}_0)^{-1}\Omega(\mathbf{u}_0)\mathbf{Q}(\mathbf{u}_0)^{-1}), \quad (3.4.118)$$

dengan

$$\mathbf{Q}(\mathbf{u}_0) = \lim_{n \rightarrow \infty} \frac{1}{nh^d} \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top, \quad (3.4.119)$$

$$\Omega(\mathbf{u}_0) = \lim_{n \rightarrow \infty} \sigma^2(nh^d) \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} (w_i^{(-k)})^2 \mathbf{x}_i \mathbf{x}_i^\top. \quad (3.4.120)$$

**Bukti.** Dari dekomposisi terstandarisasi:

$$\sqrt{nh^d}(\hat{\beta}_{\text{CF}} - \beta) = \sqrt{nh^d}\mathbf{B}_n + \sqrt{nh^d}\mathbf{V}_n. \quad (3.4.121)$$

Berdasarkan Teorema 3.4.25,  $\sqrt{nh^d}\mathbf{B}_n \xrightarrow{p} \mathbf{0}$ . Berdasarkan Teorema 3.4.23 didapatkan

$$\sqrt{nh^d}\mathbf{V}_n \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1}\Omega\mathbf{Q}^{-1}). \quad (3.4.122)$$

Dengan menggunakan Lemma Slutsky untuk penjumlahan suku yang konvergen dalam probabilitas ke nol dengan suku yang konvergen dalam distribusi, maka diperoleh

$$\sqrt{nh^d}(\hat{\beta}_{\text{CF}} - \beta) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1}\Omega\mathbf{Q}^{-1}). \quad (3.4.123)$$

■

Jika kondisi *undersmoothing* tidak dipenuhi, yaitu  $\sqrt{nh^d}h^2 \rightarrow c \neq 0$  untuk suatu konstanta  $c$ , maka distribusi asimtotik menjadi

$$\sqrt{nh^d}(\hat{\beta}_{\text{CF}}(\mathbf{u}_0) - \beta(\mathbf{u}_0)) \xrightarrow{d} \mathcal{N}(\mu(\mathbf{u}_0), \mathbf{Q}^{-1}\Omega\mathbf{Q}^{-1}), \quad (3.4.124)$$

dengan  $\mu(\mathbf{u}_0) \neq \mathbf{0}$  adalah vektor bias asimtotik. Dalam kasus ini, interval konfideni standar dan uji Wald tidak valid karena bias tidak diabaikan.

**Akibat 3.4.29** Untuk setiap komponen ke- $j$  dari vektor koefisien:

$$\sqrt{nh^d}(\hat{\beta}_{\text{CF},j}(\mathbf{u}_0) - \beta_j(\mathbf{u}_0)) \xrightarrow{d} \mathcal{N}(0, [\mathbf{Q}^{-1}\Omega\mathbf{Q}^{-1}]_{jj}). \quad (3.4.125)$$

### 3.4.7 Distribusi Asimtotik Uji Statistik

Bagian ini menurunkan distribusi asimtotik dari statistik uji yang digunakan untuk inferensi pada koefisien lokal, termasuk statistik- $t$  untuk pengujian individual dan statistik Wald untuk pengujian gabungan.

Bagian pertama menurunkan distribusi asimtotik statistik- $t$  untuk pengujian hipotesis nol individual pada koefisien lokal dengan hiptotesis umum berupa

$$H_0 : \beta_j(\mathbf{u}_0) = \beta_{j,0}, \text{ lawan}$$

$$H_1 : \beta_j(\mathbf{u}_0) \neq \beta_{j,0},$$

dengan  $\beta_{j,0}$  adalah nilai tertentu, yang biasanya adalah nol untuk menguji signifikansi.

**Teorema 3.4.30** Di bawah Asumsi GA-GWR.1–GA-GWR.10, kondisi *undersmoothing*, dan hipotesis nol  $H_0 : \beta_j(\mathbf{u}_0) = \beta_{j,0}$ , statistik- $t$

$$t_j(\mathbf{u}_0) = \frac{\hat{\beta}_{\text{CF},j}(\mathbf{u}_0) - \beta_{j,0}}{\widehat{\text{se}}(\hat{\beta}_{\text{CF},j}(\mathbf{u}_0))} \quad (3.4.126)$$

berdistribusi normal standar asimtotik:

$$t_j(\mathbf{u}_0) \xrightarrow{d} \mathcal{N}(0, 1) \quad \text{di bawah } H_0. \quad (3.4.127)$$

**Bukti.** Di bawah  $H_0 : \beta_j(\mathbf{u}_0) = \beta_{j,0}$ , statistik- $t$  dapat ditulis sebagai

$$t_j = \frac{\widehat{\beta}_{\text{CF},j} - \beta_j(\mathbf{u}_0)}{\widehat{\text{se}}(\widehat{\beta}_{\text{CF},j})}. \quad (3.4.128)$$

Berdasarkan Teorema 3.4.23 didapatkan

$$\sqrt{nh^d}(\widehat{\beta}_{\text{CF},j} - \beta_j) \xrightarrow{d} \mathcal{N}(0, \sigma_j^2), \quad (3.4.129)$$

dengan  $\sigma_j^2 = [\mathbf{Q}^{-1}\mathbf{\Omega}\mathbf{Q}^{-1}]_{jj}$ . Selanjutnya, dari Teorema ?? juga didapatkan

$$(nh^d) \cdot \widehat{\text{Var}}(\widehat{\beta}_{\text{CF},j}) \xrightarrow{p} \sigma_j^2. \quad (3.4.130)$$

Oleh karena itu,

$$\widehat{\text{se}}(\widehat{\beta}_{\text{CF},j}) = \sqrt{\widehat{\text{Var}}(\widehat{\beta}_{\text{CF},j})} = \frac{\sigma_j + o_p(1)}{\sqrt{nh^d}}, \quad (3.4.131)$$

sehingga statistik- $t$  dapat ditulis sebagai

$$\begin{aligned} t_j &= \frac{\widehat{\beta}_{\text{CF},j} - \beta_j}{\widehat{\text{se}}(\widehat{\beta}_{\text{CF},j})} \\ &= \frac{\sqrt{nh^d}(\widehat{\beta}_{\text{CF},j} - \beta_j)}{\sqrt{nh^d} \cdot \widehat{\text{se}}(\widehat{\beta}_{\text{CF},j})} \\ &= \frac{\sqrt{nh^d}(\widehat{\beta}_{\text{CF},j} - \beta_j)}{\sigma_j + o_p(1)}. \end{aligned} \quad (3.4.132)$$

Berdasarkan Lemma Slutsky, karena pembilang konvergen dalam distribusi ke  $\mathcal{N}(0, \sigma_j^2)$  dan penyebut konvergen dalam probabilitas ke  $\sigma_j$ :

$$t_j \xrightarrow{d} \frac{\mathcal{N}(0, \sigma_j^2)}{\sigma_j} = \mathcal{N}(0, 1). \quad (3.4.133)$$

■

**Akibat 3.4.31** *Interval konfidensi  $(1 - \alpha) \times 100\%$  untuk koefisien lokal  $\beta_j(\mathbf{u}_0)$*

diberikan oleh

$$CI_{1-\alpha}(\beta_j(\mathbf{u}_0)) = \hat{\beta}_{CF,j}(\mathbf{u}_0) \pm z_{1-\alpha/2} \cdot \widehat{se}(\hat{\beta}_{CF,j}(\mathbf{u}_0)), \quad (3.4.134)$$

dengan  $z_{1-\alpha/2}$  adalah kuantil  $(1 - \alpha/2)$  dari distribusi normal standar.

**Bukti.** Dari distribusi asimtotik  $t_j \xrightarrow{d} \mathcal{N}(0, 1)$ , didapatkan

$$\Pr(-z_{1-\alpha/2} \leq t_j \leq z_{1-\alpha/2}) \rightarrow 1 - \alpha. \quad (3.4.135)$$

Oleh karena itu, dengan mensubstitusikan definisi  $t_j$  dan menyusun ulang:

$$\Pr(\hat{\beta}_{CF,j} - z_{1-\alpha/2} \cdot \widehat{se} \leq \beta_j \leq \hat{\beta}_{CF,j} + z_{1-\alpha/2} \cdot \widehat{se}) \rightarrow 1 - \alpha. \quad (3.4.136)$$

■

Untuk pengujian hipotesis gabungan yang melibatkan beberapa komponen koefisien secara simultan, digunakan statistik Wald. Uji ini dibangun atas hipotesis berupa

$$H_0 : \mathbf{R}\beta(\mathbf{u}_0) = \mathbf{r}, \quad \text{lawan}$$

$$H_1 : \mathbf{R}\beta(\mathbf{u}_0) \neq \mathbf{r},$$

dengan  $\mathbf{R}$  adalah matriks  $q \times p$  yang menentukan kombinasi linier dari koefisien yang diuji, dan  $\mathbf{r}$  adalah vektor  $q \times 1$  dari nilai tertentu.

**Teorema 3.4.32** *Di bawah Asumsi GA-GWR.1–GA-GWR.10, kondisi undersmoothing, dan hipotesis nol linear  $H_0 : \mathbf{R}\beta(\mathbf{u}_0) = \mathbf{r}$  dengan  $\mathbf{R}$  matriks  $q \times p$  berpangkat penuh baris, statistik Wald*

$$W(\mathbf{u}_0) = (\mathbf{R}\hat{\beta}_{CF}(\mathbf{u}_0) - \mathbf{r})^\top \left[ \mathbf{R}\widehat{\text{Var}}(\hat{\beta}_{CF}(\mathbf{u}_0))\mathbf{R}^\top \right]^{-1} (\mathbf{R}\hat{\beta}_{CF}(\mathbf{u}_0) - \mathbf{r}) \quad (3.4.137)$$

berdistribusi  $\chi^2$  asimtotik dengan  $q$  derajat kebebasan, yaitu

$$W(\mathbf{u}_0) \xrightarrow{d} \chi_q^2 \quad \text{di bawah } H_0, \quad (3.4.138)$$

dengan  $q = \text{rank}(\mathbf{R})$  adalah jumlah restriksi.

**Bukti.** Di bawah  $H_0$ , berlaku  $\mathbf{R}\beta(\mathbf{u}_0) = \mathbf{r}$ . Didefinisikan

$$\delta_n := \mathbf{R}\hat{\beta}_{\text{CF}} - \mathbf{r} = \mathbf{R}(\hat{\beta}_{\text{CF}} - \beta). \quad (3.4.139)$$

Dari Teorema 3.4.23 didapatkan

$$\sqrt{nh^d}(\hat{\beta}_{\text{CF}} - \beta) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1}\Omega\mathbf{Q}^{-1}). \quad (3.4.140)$$

Sebab transformasi linear mempertahankan normalitas, maka

$$\sqrt{nh^d}\delta_n = \mathbf{R}\sqrt{nh^d}(\hat{\beta}_{\text{CF}} - \beta) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{R}\mathbf{Q}^{-1}\Omega\mathbf{Q}^{-1}\mathbf{R}^\top). \quad (3.4.141)$$

Didefinisikan  $\Sigma_R := \mathbf{R}\mathbf{Q}^{-1}\Omega\mathbf{Q}^{-1}\mathbf{R}^\top$ . Dari Teorema ?? didapatkan

$$(nh^d)\mathbf{R}\widehat{\text{Var}}(\hat{\beta}_{\text{CF}})\mathbf{R}^\top \xrightarrow{p} \Sigma_R. \quad (3.4.142)$$

Statistik Wald dapat ditulis sebagai

$$\begin{aligned} W &= \delta_n^\top [\mathbf{R}\widehat{\text{Var}}(\hat{\beta}_{\text{CF}})\mathbf{R}^\top]^{-1} \delta_n \\ &= (nh^d)\delta_n^\top [(nh^d)\mathbf{R}\widehat{\text{Var}}(\hat{\beta}_{\text{CF}})\mathbf{R}^\top]^{-1} (nh^d)\delta_n \cdot (nh^d)^{-1} \\ &= (\sqrt{nh^d}\delta_n)^\top [(nh^d)\mathbf{R}\widehat{\text{Var}}\mathbf{R}^\top]^{-1} (\sqrt{nh^d}\delta_n). \end{aligned} \quad (3.4.143)$$

Selanjutnya, berdasarkan *continuous mapping theorem* dan Lemma Slutsky:

$$W \xrightarrow{d} \mathbf{Z}^\top \Sigma_R^{-1} \mathbf{Z}, \quad (3.4.144)$$

dengan  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \Sigma_R)$ . Bentuk kuadratik ini berdistribusi  $\chi_q^2$  karena  $\Sigma_R^{-1/2}\mathbf{Z} \sim$



$\mathcal{N}(\mathbf{0}, \mathbf{I}_q)$  dan

$$\mathbf{Z}^\top \Sigma_R^{-1} \mathbf{Z} = \|\Sigma_R^{-1/2} \mathbf{Z}\|^2 \sim \chi_q^2. \quad (3.4.145)$$

■

**Akibat 3.4.33** Untuk menguji  $H_0 : \beta_j(\mathbf{u}_0) = 0$ , pilih  $\mathbf{R} = \mathbf{e}_j^\top$  (vektor satuan ke- $j$ ) dan  $\mathbf{r} = 0$ . Maka statistik Wald menjadi

$$W_j = \frac{\widehat{\beta}_{\text{CF},j}^2}{\widehat{\text{Var}}(\widehat{\beta}_{\text{CF},j})} = t_j^2 \xrightarrow{d} \chi_1^2. \quad (3.4.146)$$

Hal ini konsisten dengan hubungan  $\chi_1^2 = (\mathcal{N}(0, 1))^2$ .

**Akibat 3.4.34** Untuk menguji hipotesis bahwa semua koefisien kecuali konstanta bernilai nol,  $H_0 : \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$  dengan  $\beta_0$  adalah konstanta, pilih

$$\mathbf{R} = \begin{pmatrix} \mathbf{0}_{(p-1) \times 1} & \mathbf{I}_{p-1} \end{pmatrix}, \quad \mathbf{r} = \mathbf{0}_{p-1}. \quad (3.4.147)$$

Statistik Wald yang dihasilkan berdistribusi  $\chi_{p-1}^2$  di bawah  $H_0$ .

Prosedur inferensi-inferensi di atas valid secara asimtotik jika dan hanya jika terpenuhi kondisi:

1. Asumsi GA-GWR.1–GA-GWR.10 terpenuhi,
2. kondisi *undersmoothing*  $\sqrt{nh^d}h^2 \rightarrow 0$  dipenuhi, dan
3. *cross-fitting* diimplementasikan dengan benar sehingga bobot kernel independen dari galat pada *fold* estimasi.

Tanpa *undersmoothing*, interval konfidensi akan memiliki *coverage probability* yang lebih rendah dari tingkat nominal karena bias yang tidak terabaikan.

### 3.5 Analisis Komputasional GA-GWR

Bagian ini menganalisis aspek komputasional dari GA-GWR yang diturunkan secara langsung dari formulasi teoretis pada bagian sebelumnya. Pembahasan

meliputi struktur graf komputasional terdiferensiasi, analisis propagasi gradien (termasuk kondisi *vanishing* dan *exploding gradient*), kompleksitas algoritmik, serta kelayakan komputasional model secara keseluruhan.

### 3.5.1 Graf Komputasional dan Diferensiabilitas

Berdasarkan formulasi pada bagian sebelumnya, *pipeline* komputasional GA-GWR dapat direpresentasikan sebagai graf arah asiklik (*directed acyclic graph*, DAG) yang sepenuhnya terdiferensiasi. Subbagian ini menguraikan setiap komponen graf komputasional dan menganalisis propagasi gradien melalui setiap tahapan.

**Definisi 3.5.1 (Graf komputasional GA-GWR)** Pipeline komputasional GA-GWR adalah komposisi pemetaan

$$\boldsymbol{\theta} \xrightarrow{\phi_1} \mathbf{s} \xrightarrow{\phi_2} \mathbf{w} \xrightarrow{\phi_3} \hat{\boldsymbol{\beta}} \xrightarrow{\phi_4} \mathcal{L}, \quad (3.5.1)$$

dengan komponen-komponen sebagai berikut.

- (i) Pemetaan  $\phi_1 : \boldsymbol{\theta} \mapsto \mathbf{s}$  adalah fungsi GNN yang menghasilkan vektor skor  $\mathbf{s} = (s_1, \dots, s_m)^\top$  dengan  $s_i = s_{\boldsymbol{\theta}}(i, \mathbf{u}_0)$  untuk setiap  $i \in \mathcal{N}_h(\mathbf{u}_0)$ .
- (ii) Pemetaan  $\phi_2 : \mathbf{s} \mapsto \mathbf{w}$  adalah transformasi softmax yang memetakan vektor skor ke bobot kernel.
- (iii) Pemetaan  $\phi_3 : \mathbf{w} \mapsto \hat{\boldsymbol{\beta}}$  adalah solusi LWLS dengan  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{y}$ .
- (iv) Pemetaan  $\phi_4 : \hat{\boldsymbol{\beta}} \mapsto \mathcal{L}$  adalah fungsi kerugian kuadrat  $\mathcal{L} = \sum_i (y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}(\mathbf{u}_i))^2$ .

Setiap pemetaan dalam graf komputasional bersifat terdiferensiasi, sehingga gradien dapat dipropagasikan dari fungsi kerugian ke parameter GNN melalui aturan rantai (*chain rule*).

**Proposisi 3.5.2** Komposisi  $\mathcal{L} \circ \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1$  terdiferensiasi hampir di mana-mana

terhadap  $\theta$ , dengan gradien

$$\frac{\partial \mathcal{L}}{\partial \theta} = \frac{\partial \mathcal{L}}{\partial \hat{\beta}} \cdot \frac{\partial \hat{\beta}}{\partial \mathbf{w}} \cdot \frac{\partial \mathbf{w}}{\partial \mathbf{s}} \cdot \frac{\partial \mathbf{s}}{\partial \theta}. \quad (3.5.2)$$

**Bukti.** Diferensiabilitas setiap komponen ditunjukkan sebagai berikut.

- (i) Pemetaan GNN  $\phi_1$  merupakan komposisi transformasi linear dan fungsi aktivasi yang terdiferensiasi (misalnya ReLU, yang terdiferensiasi kecuali di titik nol dengan ukuran Lebesgue nol).
- (ii) Transformasi *softmax*  $\phi_2$  terdiferensiasi di  $\mathbb{R}^m$  karena fungsi eksponensial dan penjumlahan terdiferensiasi di mana-mana.
- (iii) Solusi LWLS  $\phi_3$  terdiferensiasi ketika matriks  $\mathbf{X}^\top \mathbf{W} \mathbf{X}$  invertibel.
- (iv) Fungsi kerugian kuadrat  $\phi_4$  terdiferensiasi di mana-mana sebagai fungsi polinomial.

Dengan demikian, komposisi keempat pemetaan terdiferensiasi hampir di mana-mana, dan gradien diperoleh melalui aturan rantai. ■

Analisis propagasi gradien memerlukan karakterisasi eksplisit matriks Jacobian dari setiap komponen. Transformasi *softmax* memiliki struktur Jacobian khusus yang penting untuk memahami dinamika gradien.

**Proposisi 3.5.3** *Matriks Jacobian dari transformasi softmax  $\mathbf{w} = \text{softmax}(\mathbf{s})$  adalah*

$$\frac{\partial \mathbf{w}}{\partial \mathbf{s}} = \text{diag}(\mathbf{w}) - \mathbf{w} \mathbf{w}^\top, \quad (3.5.3)$$

*dengan elemen-elemen*

$$\frac{\partial w_i}{\partial s_j} = \begin{cases} w_i(1 - w_i) & \text{jika } i = j, \\ -w_i w_j & \text{jika } i \neq j. \end{cases} \quad (3.5.4)$$

**Bukti.** Untuk  $i = j$ , dengan menggunakan aturan hasil bagi diperoleh

$$\begin{aligned}\frac{\partial w_i}{\partial s_i} &= \frac{\partial}{\partial s_i} \left( \frac{e^{s_i}}{\sum_k e^{s_k}} \right) = \frac{e^{s_i} \sum_k e^{s_k} - e^{s_i} \cdot e^{s_i}}{(\sum_k e^{s_k})^2} \\ &= \frac{e^{s_i}}{\sum_k e^{s_k}} - \frac{e^{2s_i}}{(\sum_k e^{s_k})^2} = w_i - w_i^2 = w_i(1 - w_i).\end{aligned}\quad (3.5.5)$$

Untuk  $i \neq j$ , diperoleh

$$\frac{\partial w_i}{\partial s_j} = \frac{\partial}{\partial s_j} \left( \frac{e^{s_i}}{\sum_k e^{s_k}} \right) = \frac{0 - e^{s_i} \cdot e^{s_j}}{(\sum_k e^{s_k})^2} = -w_i w_j. \quad (3.5.6)$$

Dalam bentuk matriks,  $\frac{\partial \mathbf{w}}{\partial \mathbf{s}} = \text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}^\top$ . ■

Struktur Jacobian *softmax* memiliki implikasi penting terhadap efek gradien.

Komponen kritis dalam propagasi gradien adalah diferensiasi melalui solusi LWLS. Solusi ini merupakan fungsi implisit dari bobot, sehingga memerlukan penanganan khusus.

**Proposisi 3.5.4** Untuk solusi LWLS  $\hat{\beta} = (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{y}$  dengan  $\mathbf{W} = \text{diag}(w_1, \dots, w_m)$ , gradien terhadap bobot tunggal  $w_i$  adalah

$$\frac{\partial \hat{\beta}}{\partial w_i} = \mathbf{M}^{-1} \mathbf{x}_i (y_i - \mathbf{x}_i^\top \hat{\beta}), \quad (3.5.7)$$

dengan  $\mathbf{M} = \mathbf{X}^\top \mathbf{W} \mathbf{X}$  adalah matriks informasi lokal tertimbang.

**Bukti.** Identitas diferensiasi matriks invers menyatakan  $\frac{\partial \mathbf{M}^{-1}}{\partial w_i} = -\mathbf{M}^{-1} \frac{\partial \mathbf{M}}{\partial w_i} \mathbf{M}^{-1}$ . Dengan menggunakan aturan perkalian, diperoleh

$$\frac{\partial \hat{\beta}}{\partial w_i} = \frac{\partial \mathbf{M}^{-1}}{\partial w_i} \mathbf{X}^\top \mathbf{W} \mathbf{y} + \mathbf{M}^{-1} \frac{\partial (\mathbf{X}^\top \mathbf{W} \mathbf{y})}{\partial w_i}. \quad (3.5.8)$$

Sebab  $\mathbf{M} = \sum_j w_j \mathbf{x}_j \mathbf{x}_j^\top$ , maka  $\frac{\partial \mathbf{M}}{\partial w_i} = \mathbf{x}_i \mathbf{x}_i^\top$ . Demikian pula,  $\frac{\partial (\mathbf{X}^\top \mathbf{W} \mathbf{y})}{\partial w_i} = \mathbf{x}_i y_i$ .

Substitusi memberikan

$$\begin{aligned}
 \frac{\partial \hat{\beta}}{\partial w_i} &= -\mathbf{M}^{-1} \mathbf{x}_i \mathbf{x}_i^\top \mathbf{M}^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{y} + \mathbf{M}^{-1} \mathbf{x}_i y_i \\
 &= -\mathbf{M}^{-1} \mathbf{x}_i \mathbf{x}_i^\top \hat{\beta} + \mathbf{M}^{-1} \mathbf{x}_i y_i \\
 &= \mathbf{M}^{-1} \mathbf{x}_i (y_i - \mathbf{x}_i^\top \hat{\beta}).
 \end{aligned} \tag{3.5.9}$$

■

Bentuk gradien di atas menunjukkan bahwa kontribusi gradien dari observasi ke- $i$  proporsional terhadap residualnya  $r_i = y_i - \mathbf{x}_i^\top \hat{\beta}$ . Hal ini memberikan interpretasi bahwa optimasi akan menyesuaikan bobot berdasarkan seberapa baik model menyesuaikan setiap observasi.

### 3.5.2 Analisis *Vanishing* dan *Exploding Gradient*

Analisis stabilitas numerik memerlukan karakterisasi kondisi di mana gradien dapat mengalami *vanishing* (menghilang) atau *exploding* (meledak). Kedua fenomena ini merupakan masalah fundamental dalam pelatihan jaringan saraf dalam yang dapat mengganggu konvergensi optimasi.

**Definisi 3.5.5 (*Vanishing gradient*)** Fenomena *vanishing gradient* terjadi ketika magnitudo gradien menjadi sangat kecil selama propagasi balik (backpropagation), sehingga pembaruan parameter menjadi tidak signifikan. Secara formal, jika  $\mathbf{g}_t = \frac{\partial \mathcal{L}}{\partial \theta}$  adalah gradien pada iterasi  $t$ , maka *vanishing gradient* terjadi ketika

$$\|\mathbf{g}_t\| \rightarrow 0 \quad \text{meskipun } \mathcal{L}(\theta_t) \gg \mathcal{L}^*, \tag{3.5.10}$$

dengan  $\mathcal{L}^*$  adalah nilai optimal. Fenomena ini menyebabkan pembelajaran terhenti sebelum mencapai solusi optimal.

**Definisi 3.5.6 (*Exploding gradient*)** Fenomena *exploding gradient* terjadi ketika magnitudo gradien tumbuh secara eksponensial selama propagasi balik, menyebabkan pembaruan parameter yang tidak stabil. Secara formal, *exploding gradient*

terjadi ketika

$$\|\mathbf{g}_t\| \rightarrow \infty, \quad (3.5.11)$$

yang mengakibatkan divergensi parameter dan ketidakstabilan numerik.

Untuk mendeteksi kedua fenomena ini dalam praktik, dapat digunakan metrik berikut.

**Definisi 3.5.7 (Indikator stabilitas gradien)** Stabilitas gradien dapat dideteksi melalui rasio norma gradien terhadap norma parameter:

$$\rho_t = \frac{\|\mathbf{g}_t\|}{\|\boldsymbol{\theta}_t\| + \epsilon}, \quad (3.5.12)$$

dengan  $\epsilon > 0$  adalah konstanta kecil untuk stabilitas numerik. Kondisi  $\rho_t \ll 1$  mengindikasikan potensi vanishing gradient, sedangkan  $\rho_t \gg 1$  mengindikasikan potensi exploding gradient. Dalam praktik, pelatihan stabil dicapai ketika  $\rho_t \in [10^{-7}, 10^3]$ .

Dalam konteks GA-GWR, kedua fenomena ini dapat terjadi pada komponen softmax dan solusi LWLS.

**Teorema 3.5.8** Gradien melalui transformasi softmax mengalami vanishing ketika distribusi bobot mendekati distribusi degenerasi. Secara kuantitatif, jika  $w_{i^*} \rightarrow 1$  untuk suatu indeks  $i^*$  (sehingga  $w_j \rightarrow 0$  untuk semua  $j \neq i^*$ ), maka

$$\left\| \frac{\partial \mathbf{w}}{\partial \mathbf{s}} \right\|_F \rightarrow 0. \quad (3.5.13)$$

**Bukti.** Norma Frobenius dari Jacobian softmax dapat dihitung sebagai

$$\begin{aligned} \left\| \frac{\partial \mathbf{w}}{\partial \mathbf{s}} \right\|_F^2 &= \sum_i \left( \frac{\partial w_i}{\partial s_i} \right)^2 + \sum_{i \neq j} \left( \frac{\partial w_i}{\partial s_j} \right)^2 \\ &= \sum_i w_i^2 (1 - w_i)^2 + \sum_{i \neq j} w_i^2 w_j^2. \end{aligned} \quad (3.5.14)$$

Suku kedua dapat disederhanakan sebagai

$$\sum_{i \neq j} w_i^2 w_j^2 = \left( \sum_i w_i^2 \right)^2 - \sum_i w_i^4. \quad (3.5.15)$$

Ketika  $w_{i^*} \rightarrow 1$  dan  $w_j \rightarrow 0$  untuk  $j \neq i^*$ , diperoleh  $\sum_i w_i^2 \rightarrow 1$  dan setiap suku  $w_i(1 - w_i) \rightarrow 0$  karena:

- Untuk  $i = i^*$ :  $w_{i^*}(1 - w_{i^*}) \rightarrow 1 \cdot 0 = 0$ .
- Untuk  $i \neq i^*$ :  $w_i(1 - w_i) \rightarrow 0 \cdot 1 = 0$ .

Dengan demikian,  $\left\| \frac{\partial \mathbf{w}}{\partial \mathbf{s}} \right\|_F \rightarrow 0$ . ■

Fenomena *vanishing gradient* pada *softmax* terjadi ketika fungsi skor GNN menghasilkan nilai yang sangat berbeda antarobservasi, menyebabkan satu observasi mendominasi bobot. Dalam konteks spasial, ini terjadi jika GNN terlalu yakin bahwa satu tetangga jauh lebih relevan dibandingkan yang lain.

**Teorema 3.5.9** *Gradien melalui solusi LWLS mengalami exploding ketika matriks informasi lokal  $\mathbf{M} = \mathbf{X}^\top \mathbf{W} \mathbf{X}$  mendekati singularitas. Secara kuantitatif, jika  $\lambda_{\min}(\mathbf{M}) \rightarrow 0$  (nilai eigen terkecil mendekati nol), maka*

$$\left\| \frac{\partial \hat{\boldsymbol{\beta}}}{\partial w_i} \right\| = \mathcal{O} \left( \frac{1}{\lambda_{\min}(\mathbf{M})} \right) \rightarrow \infty. \quad (3.5.16)$$

**Bukti.** Berdasarkan Proposisi sebelumnya, gradien solusi LWLS terhadap bobot adalah  $\frac{\partial \hat{\boldsymbol{\beta}}}{\partial w_i} = \mathbf{M}^{-1} \mathbf{x}_i (y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}})$ . Norma gradien dapat dibatasi oleh

$$\left\| \frac{\partial \hat{\boldsymbol{\beta}}}{\partial w_i} \right\| \leq \|\mathbf{M}^{-1}\|_2 \cdot \|\mathbf{x}_i\| \cdot |y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}|. \quad (3.5.17)$$

Karena norma spektral matriks invers adalah  $\|\mathbf{M}^{-1}\|_2 = 1/\lambda_{\min}(\mathbf{M})$ , maka

$$\left\| \frac{\partial \hat{\boldsymbol{\beta}}}{\partial w_i} \right\| \leq \frac{\|\mathbf{x}_i\| \cdot |r_i|}{\lambda_{\min}(\mathbf{M})}, \quad (3.5.18)$$

dengan  $r_i = y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}$  adalah residual. Ketika  $\lambda_{\min}(\mathbf{M}) \rightarrow 0$ , norma gradien divergen menuju tak hingga. ■

Matriks  $\mathbf{M}$  akan mendekati singularitas dalam situasi-situasi berikut.

- (i) Jika bobot terkonsentrasi pada kurang dari  $p$  observasi, maka  $\text{rank}(\mathbf{M}) < p$ .
- (ii) Jika kovariat pada observasi berbobot tinggi bersifat kolinear, maka  $\mathbf{M}$  mendekati singular.
- (iii) Jika lokasi target memiliki sedikit tetangga dalam  $\mathcal{N}_h(\mathbf{u}_0)$ , ukuran sampel efektif menjadi kecil.

### 3.5.3 Konstanta Lipschitz dan Stabilitas *Pipeline*

Untuk mengkarakterisasi propagasi perturbasi melalui *pipeline*, diperlukan analisis konstanta Lipschitz dari setiap komponen. Konstanta Lipschitz mengukur sensitivitas *output* terhadap perubahan *input* dan merupakan kunci untuk memahami stabilitas numerik.

**Definisi 3.5.10 (Konstanta Lipschitz)** Suatu fungsi  $f : \mathcal{X} \rightarrow \mathcal{Y}$  dikatakan Lipschitz kontinu dengan konstanta  $L \geq 0$  jika untuk semua  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$  berlaku

$$\|f(\mathbf{x}_1) - f(\mathbf{x}_2)\| \leq L \|\mathbf{x}_1 - \mathbf{x}_2\|. \quad (3.5.19)$$

Konstanta Lipschitz terkecil yang memenuhi pertidaksamaan di atas disebut konstanta Lipschitz dari  $f$  dan dinotasikan  $\text{Lip}(f)$ .

Konstanta Lipschitz penting dalam analisis GA-GWR karena alasan berikut.

- (i) Konstanta Lipschitz kecil menjamin bahwa perturbasi kecil pada *input* tidak menyebabkan perubahan besar pada *output*.
- (ii) Untuk fungsi dengan konstanta Lipschitz terbatas, algoritma *gradient descent* dijamin konvergen dengan laju tertentu.



- (iii) Batas Lipschitz berkontribusi pada batas generalisasi model melalui teori kompleksitas Rademacher.

**Teorema 3.5.11** Misalkan  $L_{\text{GNN}}$ ,  $L_{\text{softmax}}$ ,  $L_{\text{LWLS}}$ , dan  $L_{\mathcal{L}}$  adalah konstanta Lipschitz dari masing-masing komponen pipeline. Konstanta Lipschitz keseluruhan memenuhi

$$L_{\text{total}} \leq L_{\mathcal{L}} \cdot L_{\text{LWLS}} \cdot L_{\text{softmax}} \cdot L_{\text{GNN}}. \quad (3.5.20)$$

Berdasarkan hasil sebelumnya,  $L_{\text{softmax}} \leq 2$  dari sifat Lipschitz transformasi softmax. Nilai  $L_{\text{LWLS}}$  bergantung pada kondisioning matriks  $\mathbf{M}$ :

$$L_{\text{LWLS}} = \mathcal{O} \left( \frac{\kappa(\mathbf{M})}{\lambda_{\min}(\mathbf{M})} \right), \quad (3.5.21)$$

dengan  $\kappa(\mathbf{M}) = \lambda_{\max}(\mathbf{M})/\lambda_{\min}(\mathbf{M})$  adalah condition number matriks  $\mathbf{M}$ .

**Bukti.** Untuk komposisi fungsi  $f = f_L \circ \dots \circ f_1$ , konstanta Lipschitz memenuhi  $\text{Lip}(f) \leq \prod_{i=1}^L \text{Lip}(f_i)$ . Hal ini dapat dibuktikan secara induktif: untuk  $\mathbf{x}, \mathbf{y}$  sembarang,

$$\begin{aligned} \|f(\mathbf{x}) - f(\mathbf{y})\| &= \|f_L(f_{L-1}(\dots f_1(\mathbf{x}))) - f_L(f_{L-1}(\dots f_1(\mathbf{y})))\| \\ &\leq L_L \|f_{L-1}(\dots f_1(\mathbf{x})) - f_{L-1}(\dots f_1(\mathbf{y}))\| \\ &\leq \dots \leq \prod_{i=1}^L L_i \|\mathbf{x} - \mathbf{y}\|. \end{aligned} \quad (3.5.22)$$

Untuk komponen LWLS, perhatikan bahwa  $\hat{\beta} = \mathbf{M}^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{y}$ . Sensitivitas terhadap perubahan bobot  $\mathbf{W}$  melibatkan invers matriks  $\mathbf{M}$ , yang norma spektralnya adalah  $1/\lambda_{\min}(\mathbf{M})$ . Dengan analisis perturbasi matriks, diperoleh batas  $L_{\text{LWLS}} = \mathcal{O}(\kappa(\mathbf{M})/\lambda_{\min}(\mathbf{M}))$ . ■

Teorema di atas menunjukkan bahwa stabilitas *pipeline* secara keseluruhan bergantung kritis pada matriks informasi lokal  $\mathbf{M}$ . Matriks dengan *condition number* tinggi (misalnya  $\kappa(\mathbf{M}) > 10^6$ ) menyebabkan amplifikasi perturbasi dan ketidakstabilan numerik, yang mengindikasikan perlunya regularisasi atau peningkatan ukuran sampel lokal.

### 3.5.4 Algoritma Pelatihan dan Kompleksitas Komputasional

Bagian ini menyajikan algoritma komputasional untuk pelatihan GNN dan estimasi koefisien lokal dengan skema *cross-fitting*, serta analisis kompleksitas algoritmiknya.

**Definisi 3.5.12 (Fungsi kerugian pelatihan)** Untuk himpunan pelatihan  $\mathcal{D}_{\text{latih}}$ , fungsi kerugian didefinisikan sebagai

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{|\mathcal{D}_{\text{latih}}|} \sum_{i \in \mathcal{D}_{\text{latih}}} (y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}(\mathbf{u}_i))^2, \quad (3.5.23)$$

dengan  $\hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}(\mathbf{u}_i)$  adalah estimator koefisien lokal yang bergantung pada parameter GNN  $\boldsymbol{\theta}$  melalui bobot kernel  $\hat{\kappa}_{\boldsymbol{\theta}}(\cdot \mid \mathbf{u}_i)$ .

Untuk menjamin stabilitas pelatihan berdasarkan analisis gradien pada subbagian sebelumnya, ditambahkan regularisasi entropi yang mencegah *vanishing gradient*.

**Definisi 3.5.13 (Fungsi kerugian teregularisasi)** Fungsi kerugian teregularisasi didefinisikan sebagai

$$\mathcal{L}_{\text{reg}}(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}) - \lambda_{\text{ent}} \cdot \frac{1}{n} \sum_{i=1}^n H(\mathbf{w}_i), \quad (3.5.24)$$

dengan  $H(\mathbf{w}_i) = -\sum_j w_{ij} \log w_{ij}$  adalah entropi distribusi bobot untuk lokasi  $i$ , dan  $\lambda_{\text{ent}} > 0$  adalah parameter regularisasi. Suku entropi mendorong distribusi bobot yang lebih merata, mencegah konsentrasi bobot yang menyebabkan *vanishing gradient*.

Algoritma berikut mengimplementasikan skema *cross-fitting* yang telah diuraikan sebelumnya.

---

**Algoritma 3.4** Estimasi Koefisien Lokal GA-GWR dengan *Cross-Fitting*


---

**Require:** Data  $\{(\mathbf{x}_i, y_i, \mathbf{u}_i)\}_{i=1}^n$ , jumlah *fold*  $K$ , lokasi target  $\mathbf{u}_0$

**Ensure:** Estimator koefisien lokal  $\hat{\beta}_{\text{CF}}(\mathbf{u}_0)$

- 1: **Tahap 1: Partisi Data**
  - 2: Partisi  $\{1, \dots, n\}$  menjadi  $K$  *fold* saling lepas:  $\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_K$
  - 3:
  - 4: **Tahap 2: Pelatihan GNN per Fold**
  - 5: **for**  $k = 1$  **to**  $K$  **do**
  - 6:     Definisikan data pelatihan:  $\mathcal{D}^{(-k)} = \{i : i \notin \mathcal{I}_k\}$
  - 7:     Latih GNN:  $\hat{\theta}^{(-k)} = \arg \min_{\theta} \mathcal{L}_{\text{reg}}(\theta; \mathcal{D}^{(-k)})$
  - 8: **end for**
  - 9:
  - 10: **Tahap 3: Konstruksi Graf Lokal**
  - 11: Bangun  $\mathcal{N}_h(\mathbf{u}_0) = \{i : d(\mathbf{u}_i, \mathbf{u}_0) \leq h\}$
  - 12: Untuk setiap  $i \in \mathcal{N}_h(\mathbf{u}_0)$ : hitung  $\mathbf{z}_i(\mathbf{u}_0) = (\mathbf{x}_i^\top, (\mathbf{u}_i - \mathbf{u}_0)^\top)^\top$
  - 13:
  - 14: **Tahap 4: Komputasi Bobot Cross-Fitted**
  - 15: **for**  $k = 1$  **to**  $K$  **do**
  - 16:     **for**  $i \in \mathcal{I}_k \cap \mathcal{N}_h(\mathbf{u}_0)$  **do**
  - 17:         Hitung skor:  $s_i = \text{GNN}_{\hat{\theta}^{(-k)}}(\mathcal{G}_h(\mathbf{u}_0), \mathbf{z}_i(\mathbf{u}_0))$
  - 18:     **end for**
  - 19:     Normalisasi *softmax*:  $w_i^{(-k)} = \exp(s_i) / \sum_{j \in \mathcal{I}_k \cap \mathcal{N}_h} \exp(s_j)$
  - 20: **end for**
  - 21:
  - 22: **Tahap 5: Agregasi Estimator Cross-Fitted**
  - 23: **for**  $k = 1$  **to**  $K$  **do**
  - 24:      $\mathbf{H}_k \leftarrow \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i \mathbf{x}_i^\top$
  - 25:      $\mathbf{g}_k \leftarrow \sum_{i \in \mathcal{I}_k} w_i^{(-k)} \mathbf{x}_i y_i$
  - 26: **end for**
  - 27:  $\hat{\beta}_{\text{CF}}(\mathbf{u}_0) \leftarrow (\sum_{k=1}^K \mathbf{H}_k)^{-1} (\sum_{k=1}^K \mathbf{g}_k)$
  - 28: **return**  $\hat{\beta}_{\text{CF}}(\mathbf{u}_0)$
-

Analisis kompleksitas komputasional penting untuk memahami skalabilitas algoritma GA-GWR. Notasi  $\mathcal{O}(\cdot)$  menyatakan batas atas asimtotik, dengan kompleksitas yang lebih rendah mengindikasikan algoritma yang lebih efisien.

**Teorema 3.5.14 (Kompleksitas GA-GWR)** Misalkan  $n$  adalah jumlah observasi,  $p$  adalah dimensi kovariat,  $m = |\mathcal{N}_h(\mathbf{u}_0)|$  adalah ukuran rata-rata tetangga lokal,  $K$  adalah jumlah fold,  $T$  adalah jumlah iterasi pelatihan, dan  $|\boldsymbol{\theta}|$  adalah jumlah parameter GNN. Kompleksitas waktu untuk setiap tahap adalah sebagai berikut.

- (i) **Pelatihan GNN** (Tahap 2):  $\mathcal{O}(K \cdot T \cdot n \cdot m \cdot |\boldsymbol{\theta}|)$  — kompleksitas linear terhadap ukuran data dan parameter (baik).
- (ii) **Konstruksi graf lokal** (Tahap 3):  $\mathcal{O}(n \log n)$  dengan struktur data spasial seperti k-d tree (baik), atau  $\mathcal{O}(n^2)$  dengan pencarian naif (moderat).
- (iii) **Komputasi bobot** (Tahap 4):  $\mathcal{O}(K \cdot m \cdot |\boldsymbol{\theta}|)$  — linear terhadap jumlah tetangga (baik).
- (iv) **Solusi LWLS** (Tahap 5):  $\mathcal{O}(m \cdot p^2 + p^3)$  untuk setiap lokasi target — kubik terhadap  $p$  namun linear terhadap  $m$  (baik untuk  $p$  kecil).

Kompleksitas total untuk estimasi di  $n_0$  lokasi target adalah

$$\mathcal{O}(K \cdot T \cdot n \cdot m \cdot |\boldsymbol{\theta}| + n_0 \cdot (m \cdot p^2 + p^3)). \quad (3.5.25)$$

**Bukti.** Untuk Tahap 2, setiap iterasi pelatihan memproses  $n(1 - 1/K)$  observasi dengan  $K$  model. Setiap observasi memerlukan *forward pass* melalui GNN dengan kompleksitas  $\mathcal{O}(m \cdot |\boldsymbol{\theta}|)$  untuk agregasi tetangga dan transformasi parameter.

Untuk Tahap 3, pencarian tetangga dalam radius  $h$  dapat dilakukan dalam  $\mathcal{O}(\log n + m)$  per query menggunakan *k-d tree*, atau  $\mathcal{O}(n)$  dengan pencarian linear.

Untuk Tahap 4, komputasi skor untuk setiap observasi memerlukan satu *forward pass* GNN.

Untuk Tahap 5, pembentukan matriks  $\mathbf{H}_k$  memerlukan  $\mathcal{O}(m \cdot p^2)$  operasi (produk luar  $\mathbf{x}_i \mathbf{x}_i^\top$  untuk  $m$  observasi), dan inversi matriks  $p \times p$  memerlukan  $\mathcal{O}(p^3)$

operasi. ■

Kompleksitas GA-GWR bersifat baik untuk aplikasi praktis karena:

- (i) Linear terhadap ukuran data  $n$  (skalabel untuk dataset besar).
- (ii) Kubik terhadap dimensi  $p$ , namun  $p$  biasanya kecil dalam aplikasi GWR ( $p < 20$ ).
- (iii) Pelatihan GNN hanya dilakukan sekali, sedangkan estimasi di lokasi baru hanya memerlukan  $\mathcal{O}(m \cdot p^2 + p^3)$ .

Sebagai perbandingan, GWR klasik memerlukan  $\mathcal{O}(n \cdot (n \cdot p^2 + p^3))$  untuk seluruh estimasi (kuadrat terhadap  $n$ ), sehingga GA-GWR lebih efisien untuk dataset besar.

### 3.5.5 Diagnostik Model dan Validasi Inferensi

Diagnostik model diperlukan untuk mengevaluasi kualitas estimasi dan memvalidasi prosedur inferensi yang telah diuraikan pada bagian sebelumnya. Subbagian ini menyajikan metrik diagnostik dalam kerangka definisi-teorema yang sesuai dengan struktur teoretis GA-GWR.

Metrik evaluasi prediksi digunakan untuk mengukur kemampuan model dalam memprediksi nilai respons.

**Definisi 3.5.15 (Root Mean Squared Error (RMSE))** *RMSE adalah akar kuadrat dari rata-rata kuadrat galat prediksi, didefinisikan sebagai*

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\text{CF}}(\mathbf{u}_i))^2}. \quad (3.5.26)$$

*RMSE memiliki satuan yang sama dengan variabel respons  $y$ . Nilai RMSE yang lebih rendah mengindikasikan prediksi yang lebih akurat.*

**Definisi 3.5.16 (Mean Absolute Error (MAE))** *MAE adalah rata-rata nilai absolut*

galat prediksi, didefinisikan sebagai

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{\varepsilon}_i| = \frac{1}{n} \sum_{i=1}^n |y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\text{CF}}(\mathbf{u}_i)|. \quad (3.5.27)$$

MAE lebih robust terhadap outlier dibandingkan RMSE karena tidak mengkuadratkan galat. Nilai MAE yang lebih rendah mengindikasikan prediksi yang lebih akurat.

**Definisi 3.5.17 (Koefisien determinasi ( $R^2$ ))** Koefisien determinasi mengukur proporsi variansi dalam variabel respons yang dapat dijelaskan oleh model, didefinisikan sebagai

$$R^2 = 1 - \frac{\text{SS}_{\text{res}}}{\text{SS}_{\text{tot}}} = 1 - \frac{\sum_{i=1}^n (y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\text{CF}}(\mathbf{u}_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (3.5.28)$$

dengan  $\bar{y} = n^{-1} \sum_{i=1}^n y_i$  adalah rata-rata sampel,  $\text{SS}_{\text{res}}$  adalah jumlah kuadrat residual, dan  $\text{SS}_{\text{tot}}$  adalah jumlah kuadrat total. Nilai  $R^2 \in [0, 1]$ , di mana  $R^2 = 1$  menunjukkan prediksi sempurna dan  $R^2 = 0$  menunjukkan model tidak lebih baik dari prediksi menggunakan rata-rata.

Diagnostik residual diperlukan untuk memvalidasi asumsi model, khususnya Asumsi GA-GWR.1 dan normalitas galat.

**Definisi 3.5.18 (Uji Shapiro-Wilk)** Uji Shapiro-Wilk digunakan untuk menguji normalitas distribusi residual dengan hipotesis sebagai berikut.

$H_0$  : Residual berdistribusi normal, lawan

$H_1$  : residual tidak berdistribusi normal.

Statistik uji Shapiro-Wilk didefinisikan sebagai

$$W = \frac{\left( \sum_{i=1}^n a_i \hat{\varepsilon}_{(i)} \right)^2}{\sum_{i=1}^n (\hat{\varepsilon}_i - \bar{\varepsilon})^2}, \quad (3.5.29)$$

dengan  $\hat{\varepsilon}_{(i)}$  adalah statistik urutan ke- $i$  dari residual (residual yang diurutkan dari

terkecil ke terbesar),  $\bar{\varepsilon} = n^{-1} \sum_{i=1}^n \hat{\varepsilon}_i$  adalah rata-rata residual, dan  $a_i$  adalah koefisien yang diturunkan dari momen statistik urutan distribusi normal standar.

Untuk tingkat signifikansi  $\alpha$ , hipotesis nol  $H_0$  ditolak jika  $W < W_{\alpha,n}$ , dengan  $W_{\alpha,n}$  adalah nilai kritis dari tabel distribusi Shapiro-Wilk untuk ukuran sampel  $n$ . Secara ekuivalen, jika  $p\text{-value} < \alpha$ , maka terdapat cukup bukti untuk menolak  $H_0$ .

**Definisi 3.5.19 (Indeks Moran)** Indeks Moran digunakan untuk mendeteksi autokorelasi spasial pada residual dengan hipotesis sebagai berikut.

$H_0$  : Tidak terdapat autokorelasi spasial pada residual ( $I = \mathbb{E}[I]$ ), lawan

$H_1$  : terdapat autokorelasi spasial pada residual ( $I \neq \mathbb{E}[I]$ ).

Statistik Indeks Moran didefinisikan sebagai

$$I = \frac{n}{S_0} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}^* (\hat{\varepsilon}_i - \bar{\varepsilon})(\hat{\varepsilon}_j - \bar{\varepsilon})}{\sum_{i=1}^n (\hat{\varepsilon}_i - \bar{\varepsilon})^2}, \quad (3.5.30)$$

dengan  $w_{ij}^*$  adalah elemen matriks bobot spasial (biasanya biner:  $w_{ij}^* = 1$  jika lokasi  $i$  dan  $j$  bertetangga, 0 sebaliknya), dan  $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}^*$  adalah total bobot.

Di bawah  $H_0$ , nilai harapan dan variansi Indeks Moran adalah

$$\mathbb{E}[I] = -\frac{1}{n-1}, \quad (3.5.31)$$

$$\text{Var}(I) = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{S_0^2 (n^2 - 1)} - \mathbb{E}[I]^2, \quad (3.5.32)$$

dengan  $S_1 = \frac{1}{2} \sum_{i \neq j} (w_{ij}^* + w_{ji}^*)^2$  dan  $S_2 = \sum_i (\sum_j w_{ij}^* + \sum_j w_{ji}^*)^2$ .

Statistik- $z$  untuk pengujian adalah

$$z_I = \frac{I - \mathbb{E}[I]}{\sqrt{\text{Var}(I)}}, \quad (3.5.33)$$

yang berdistribusi asimtotik  $\mathcal{N}(0, 1)$  di bawah  $H_0$ . Untuk uji dua sisi dengan tingkat signifikansi  $\alpha$ ,  $H_0$  ditolak jika  $|z_I| > z_{\alpha/2}$ . Secara ekuivalen, jika  $p\text{-value} < \alpha$ , maka terdapat cukup bukti untuk menolak  $H_0$ .

Untuk memvalidasi bahwa bobot *kernel* yang dipelajari GNN memenuhi kondisi regularitas pada Asumsi GA-GWR.7 dan GA-GWR.8, dilakukan evaluasi berikut.

**Definisi 3.5.20 (Entropi bobot *kernel*)** *Entropi distribusi bobot untuk lokasi target  $\mathbf{u}_0$  didefinisikan sebagai*

$$H(\mathbf{u}_0) = - \sum_{i \in \mathcal{N}_h(\mathbf{u}_0)} \hat{\kappa}_{\theta}(i \mid \mathbf{u}_0) \log \hat{\kappa}_{\theta}(i \mid \mathbf{u}_0). \quad (3.5.34)$$

*Entropi mengukur ketidakpastian atau keragaman distribusi bobot.*

Entropi tinggi berarti bobot terdistribusi merata di antara banyak observasi sehingga baik untuk stabilitas, tetapi dapat menyebabkan *oversmoothing*. Entropi rendah berarti bobot terkonsentrasi pada sedikit observasi sehingga dapat menyebabkan variansi tinggi dan fenomena *vanishing gradient*.

**Definisi 3.5.21 (Effective number of neighbors)** *Jumlah tetangga efektif untuk lokasi target  $\mathbf{u}_0$  didefinisikan sebagai*

$$n_{\text{eff}}(\mathbf{u}_0) = \exp(H(\mathbf{u}_0)). \quad (3.5.35)$$

*Metrik ini mengukur berapa banyak observasi yang secara efektif berkontribusi pada estimasi lokal, dengan interpretasi sebagai berikut.*

- (i) *Jika bobot seragam, maka  $n_{\text{eff}} = |\mathcal{N}_h(\mathbf{u}_0)|$  (maksimum).*
- (ii) *Jika bobot terkonsentrasi pada satu observasi, maka  $n_{\text{eff}} \rightarrow 1$ .*

**Teorema 3.5.22 (Kondisi kecukupan sampel lokal)** *Untuk menjamin identifikasi parameter lokal (Asumsi GA-GWR.7), jumlah tetangga efektif harus memenuhi*

$$n_{\text{eff}}(\mathbf{u}_0) \geq p + 1, \quad (3.5.36)$$



dengan  $p$  adalah dimensi kovariat. Jika kondisi ini tidak terpenuhi, matriks  $\mathbf{X}^\top \mathbf{W} \mathbf{X}$  dapat menjadi singular atau ill-conditioned.

**Bukti.** Untuk solusi LWLS, matriks Hessian  $\mathbf{X}^\top \mathbf{W} \mathbf{X}$  berukuran  $p \times p$ . Agar matriks ini memiliki *full rank*, diperlukan minimal  $p$  observasi dengan bobot tak nol yang kovariat-nya *linearly independent*. ■

Untuk memastikan validitas inferensi statistik, dilakukan validasi berikut.

**Definisi 3.5.23 (Kondisi *undersmoothing*)** Kondisi undersmoothing terpenuhi jika bandwidth  $h$  memenuhi

$$\sqrt{nh^d} \cdot h^2 \rightarrow 0 \quad \text{saat } n \rightarrow \infty, \quad (3.5.37)$$

dengan  $d$  adalah dimensi ruang koordinat spasial. Kondisi ini menjamin bahwa bias asimtotik dapat diabaikan relatif terhadap galat baku, sehingga interval kepercayaan memiliki coverage yang valid.

Dalam praktik, validitas inferensi dapat diverifikasi dengan cara sebagai berikut.

- (i) Membandingkan distribusi empiris statistik- $t$  dengan distribusi  $\mathcal{N}(0, 1)$  menggunakan Q-Q plot.
- (ii) Mengevaluasi sensitivitas hasil terhadap variasi jumlah *fold*  $K$  dalam skema *cross-fitting*.
- (iii) Memastikan *coverage* empiris interval kepercayaan mendekati level nominal melalui simulasi atau *bootstrap*.

## **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](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

$$\begin{aligned}
 (\mathbf{A})_{22} &= (\mathbf{C})_{22} + \min\{(\mathbf{A})_{2(2-1)}, (\mathbf{A})_{(2-1)2}, (\mathbf{A})_{(2-1)(2-1)}\} \\
 &= (\mathbf{C})_{22} + \min\{(\mathbf{A})_{21}, (\mathbf{A})_{12}, (\mathbf{A})_{11}\} \\
 &= 2 + \min\{3, 5, 2\} \\
 &= 2 + 2 = 4.
 \end{aligned}$$

$$(\mathbf{A})_{11} = (\mathbf{C})_{11}. \quad (4.2.1)$$

$$A = (5, 6, 5, 7, 6, 6, 6, 6) \quad B = (7, 8, 6, 10, 10, 10, 8, 8)$$

- 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

## DAFTAR PUSTAKA

- Amri, M. Z. A. (2020). Perbandingan model star dan gstar untuk peramalan indeks harga konsumen di kota padang, pekanbaru, jambi, dan Palembang. *MEDI-AN*, 4(1), 30–37.
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers. doi: 10.1007/978-94-015-7799-1
- Ash, R. B. (1972). *Real analysis and probability*. New York: Academic Press.
- Axler, S. (2024). *Linear Algebra Done Right* (4th ed.). Cham: Springer. doi: 10.1007/978-3-031-41026-0
- Billingsley, P. (1995). *Probability and measure* (3rd ed.). New York: Wiley.
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford: Oxford University Press.
- Bronstein, M. M., Bruna, J., Cohen, T., dan Veličković, P. (2021). Geometric deep learning: Grids, groups, graphs, geodesics, and gauges.. Diambil dari <https://arxiv.org/abs/2104.13478>
- Brüel-Gabrielsson, R. (2020). *Universal function approximation on graphs*. Diambil dari <https://arxiv.org/abs/2003.06706>
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., dan Robins, J. (2024). *Double/debiased machine learning for treatment and causal parameters*. Diambil dari <https://arxiv.org/abs/1608.00060>
- Chung, F. R. K. (1997). *Spectral graph theory*. Providence, RI: American Mathematical Society.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4), 303–314. doi: 10.1007/BF02551274
- Dhrymes, P. J. (2013). *Mathematics for Econometrics* (4th ed.). New York: Spri-

- nger. doi: 10.1007/978-1-4614-8145-4
- Dudley, R. M. (2002). *Real analysis and probability*. Cambridge: Cambridge University Press.
- Fan, J., dan Gijbels, I. (1996). *Local polynomial modelling and its applications*. London: Chapman and Hall/CRC.
- Fotheringham, A. S., Brunsdon, C., dan Charlton, M. (2002). . Chichester: Wiley.
- Fotheringham, A. S., Yang, W., dan Kang, W. (2017). Multiscale geographically weighted regression (mgwr). *Annals of the American Association of Geographers*, 107(6), 1247–1265.
- Goodfellow, I., Bengio, Y., dan Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press. Diambil dari <http://www.deeplearningbook.org>
- Greene, W. H. (2002). *Econometric Analysis* (5th ed.). Upper Saddle River, New Jersey: Pearson Education.
- Hagenauer, J., dan Helbich, M. (2022). Geographically weighted artificial neural networks: A framework for non-linear local modelling. *International Journal of Geographical Information Science*, 36(2), 231–252.
- Hamilton, W., Ying, R., dan Leskovec, J. (2017). Inductive representation learning on large graphs. In *31st conference on neural information processing systems (neurips)*. Diambil dari <https://arxiv.org/abs/1706.02216>
- Haykin, S. (2009). *Neural networks and learning machines* (3rd ed.). New York: Prentice Hall.
- Horn, R. A., dan Johnson, C. R. (2012). *Matrix Analysis* (2nd ed.). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139020411
- Huang, B., Wu, B., dan Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24(3), 383–401.
- Kipf, T. N., dan Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *5th international conference on learning representations (iclr)*. Diambil dari <https://arxiv.org/abs/1609.02907>

- Lang, S. (1987). *Linear Algebra* (3rd ed.). New York: Springer. doi: 10.1007/978-1-4757-1949-9
- LeSage, J. P., dan Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Boca Raton, FL: CRC Press.
- Leung, Y., Mei, C.-L., dan Zhang, W.-X. (2000). Statistical tests for spatial non-stationarity based on the geographically weighted regression model. *Environment and Planning A*, 32, 9–32.
- Lu, B., Charlton, M., Harris, P., dan Fotheringham, A. S. (2014). Geographically weighted regression with a non-euclidean distance metric: a case study using hedonic house price data. *International Journal of Geographical Information Science*, 28(4), 660–681.
- Montgomery, D. C., Peck, E. A., dan Vining, G. G. (2012). (5th ed.). Hoboken, NJ: Wiley.
- Nair, V., dan Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (icml)* (pp. 807–814).
- Ni, X., Wu, H., dan Zhang, X. (2022). Sannwr: Spatial and attribute neural network weighted regression for modeling complex spatial heterogeneity. *International Journal of Geographical Information Science*, 36(7), 1401–1426.
- Pishro-Nik, H. (2014). *Introduction to probability, statistics, and random processes*. Kappa Research, LLC.
- Ross, S. M. (1972). *Introduction to probability models* (6th ed.). USA: Academic Press.
- van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge: Cambridge University Press.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., dan Bengio, Y. (2018). Graph attention networks. In *6th international conference on learning representations (iclr)*. Diambil dari <https://arxiv.org/abs/1710.10903>
- Wooldridge, J. M. (2020). *Introductory Econometrics: A Modern Approach* (7th

- ed.). Boston, MA: Cengage.
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., dan Yu, P. S. (2021, January). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4–24. Diambil dari <http://dx.doi.org/10.1109/TNNLS.2020.2978386> doi: 10.1109/tnnls.2020.2978386
- Xu, K., Hu, W., Leskovec, J., dan Jegelka, S. (2019). How powerful are graph neural networks? In *7th international conference on learning representations (iclr)*. Diambil dari <https://arxiv.org/abs/1810.00826>
- Yan, J., Wu, B., dan Duan, X. (2024). Modeling spatial anisotropic relationships using gradient-based geographically weighted regression. *Annals of the American Association of Geographers*, 114, 697–718.
- Yin, P., Huang, B., Du, Z., Wu, H., et al. (2024). Gnnwr: An r package for geographically neural network weighted regression and its spatiotemporal extension. *Journal of Geographical Systems*.
- Zhu, D., Liu, Y., Yao, X., dan Fischer, M. M. (2022). Spatial regression graph convolutional neural networks (srcnns). *GeoInformatica*, 26(3), 557–582.



## **LAMPIRAN A**

### **Data**

## LAMPIRAN B

### Syntax R

1 SYNTAX R ALALALALALLALA