

LAPORAN TUGAS AKHIR I

**PEMBELAJARAN REPRESENTASI SPASIAL-TEMPORAL
BERBASIS JARINGAN SARAF GRAF UNTUK OPTIMASI
PEMBOBOTAN PADA MODEL REGRESI DENGAN
KOEFISIEN BERVARIASI DAN REGRESI LOKAL
TERBOBOTI**



Kurniawan Chandra Wijaya
22/497908/PA/21466

PEMBIMBING : Prof. Dr. Abdurakhman, S.Si., M.Si.

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

2025

DAFTAR ISI

DAFTAR ISI	
BAB I PENDAHULUAN	2
1.1. Latar Belakang.....	2
1.2. Pembatasan Masalah	4
1.3. Tujuan dan Manfaat Penelitian.....	4
1.4. Tinjauan Pustaka	5
1.5. Metodologi Penelitian.....	5
1.6. Sistematika Penulisan	6
BAB II LANDASAN TEORI	7
2.1. Aljabar Matriks	7
2.1.1. Ruang Vektor dan Matriks.....	7
2.1.2. Sifat-Sifat Matriks	13
2.1.3. Invers Matriks.....	23
2.1.4. Determinan dan <i>Rank</i>	31
2.1.5. Nilai Eigen dan Vektor Eigen	38
2.1.6. Perkalian Kronecker.....	41
2.1.7. Diferensial Vektor dan Matriks.....	42
2.2. Teori Ukuran dan Probabilitas	47
2.2.1. Ruang Ukur dan Ukuran Probabilitas.....	48
2.2.2. Variabel Acak sebagai Fungsi Terukur.....	51
2.2.3. Distribusi dan Ekspektasi	53
2.2.4. Momen, Variansi, dan Kovariansi.....	63
2.2.5. Probabilitas Bersyarat dan Independensi	65
2.2.6. Teorema Bayes dan Hukum Probabilitas Total	67
2.2.7. Hukum Ekspektasi Total dan Variansi Total	70
2.3. Konvergensi dan Laju Pertumbuhan	74
2.3.1. Barisan Variabel Acak	75
2.3.2. Konvergensi Barisan Variabel Acak.....	75
2.3.3. Hubungan Antarmode Konvergensi	78
2.4. Teori Probabilitas Asimtotik.....	82
2.4.1. Hukum Bilangan Besar	82
2.4.2. Konsistensi Estimator	83
2.4.3. Fungsi Karakteristik.....	84
2.4.4. Teorema Pemetaan Kontinu dan Teorema Slutsky	86
2.4.5. Notasi Big- \mathcal{O} dan Little- o (Deterministik).....	92
2.4.6. Notasi Big- \mathcal{O} dan Little- o dalam Probabilitas.....	95
2.4.7. Teorema Limit Pusat Klasik	103

2.4.8. Metode Delta	108
2.5. Analisis Regresi Linear.....	110
2.5.1. Regresi Linear Biasa atau <i>Ordinary Least Squares</i> (OLS).....	110
2.5.2. Teori Asimtotik untuk OLS.....	115
2.5.3. Regresi Terboboti	122
2.5.4. Teori Asimtotik untuk GLS/WLS.....	126
2.6. Pemodelan Polinomial Lokal	130
2.6.1. Asumsi Model Polinomial Lokal	131
2.6.2. Pendugaan Model Polinomial Lokal	132
2.6.3. Teori Asimtotik Estimator Polinomial Lokal.....	137
2.7. Analisis Regresi Spasial	139
2.7.1. Regresi dengan Dependensi Spasial.....	140
2.7.2. Regresi dengan Heterogenitas Spasial	142
2.8. Jaringan Saraf Tiruan	150
2.8.1. Model Dasar dan Notasi	150
2.8.2. Fungsi Aktivasi.....	150
2.8.3. Fungsi Kerugian dan Kriteria Pembelajaran	154
2.8.4. Pembelajaran dengan Propagasi Mundur.....	155
2.8.5. Optimisasi Parameter	157
2.8.6. Jaringan Saraf Tiruan sebagai Aproksimasi Universal	160
2.9. Jaringan Saraf Graf	160
2.9.1. Dasar Graf dan Laplacian	160
2.9.2. Kerangka Penyampaian Pesan pada Jaringan Saraf	162
2.9.3. Arsitektur Umum dalam GNN	163
2.9.4. Kekuatan Representasi GNN.....	166
BAB III RENCANA PENELITIAN	168
3.1. Rencana Penelitian	168
3.2. Rencana Pengambilan Data	168
3.3. Rencana Jadwal Penelitian	169
BAB IV PENGEMBANGAN PEMBOBOTAN BERBASIS JARINGAN SARAF GRAF UNTUK REGRESI SPASIAL-TEMPORAL	170
4.1. Kajian Literatur Terkait.....	170
4.1.1. <i>Geographically Weighted Artificial Neural Network</i> (GWANN)...	170
4.1.2. <i>Spatial and Attribute Neural Network Weighted Regression</i> (SAN-NWR).....	171
4.1.3. <i>Spatial Regression Graph Convolutional Neural Networks</i> (SR-GCNN).....	171
4.1.4. <i>Spatio-Temporal Graph Convolutional Networks</i> (ST-GCN)	172
4.1.5. <i>Geographically Neural Network Weighted Regression</i> (GNNWR). .	172

4.2. Kerangka Metodologis	175
4.2.1. Kerangka Umum dan Alur Metodologi	175
4.2.2. Alur Pembelajaran GNN-GTVC dan GNN-GTWR	177
4.2.3. Perbandingan Konseptual Awal	178
4.3. Regresi Spasial-Temporal Berkoeffisien Bervariasi berbasis GNN (GNN-GTVC).....	179
4.3.1. Formulasi Matematis	179
4.3.2. Pembentukan Bobot Koefisien Lokal melalui Representasi Graf ..	180
4.3.3. Estimasi Koefisien Global dan Prediksi	180
4.4. Regresi Spasial-Temporal Terboboti berbasis GNN (GNN-GTWR)	181
4.4.1. Formulasi Estimator WLS pada Konteks Graf	181
4.4.2. Pembobotan Antartetangga dengan Representasi GNN	182
4.5. Arsitektur Jaringan Graf untuk Pembelajaran Bobot	182
4.6. Pilihan Pemetaan Representasi ke Bobot	183
4.6.1. Fungsi Objektif dan Proses Pelatihan	185
4.7. Pembelajaran Semi-Terawasi (<i>Semi-Supervised Learning, SSL</i>)	187
4.7.1. Kerangka Dasar dan Asumsi Fundamental	187
4.7.2. Fungsi Kerugian Semi-Terawasi	187
4.7.3. Integrasi SSL ke dalam Pelatihan GNN-GTVC dan GNN-GTWR.	189
4.8. Hasil Sementara	190
4.8.1. Pengaturan Eksperimen dan Karakteristik Data	190
4.8.2. Analisis Konfigurasi Model GNN-GTVC dan GNN-GTWR	191
4.8.3. Perbandingan Performa dengan Metode Lainnya	192
DAFTAR PUSTAKA	197

BAB I

PENDAHULUAN

1.1. Latar Belakang

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

Dalam literatur ekonometrika spasial, ketidakstasioneran spasial (*spatial heterogeneity*) dan ketergantungan spasial (*spatial dependence*) menjadi dua aspek penting yang perlu diperhatikan (?). Untuk mengakomodasi hal tersebut, Fotheringham, Brunsdon, and Charlton (2002b) mengembangkan model *Geographically Weighted Regression* (GWR) yang memungkinkan koefisien regresi bervariasi antar lokasi. Model ini kemudian diperluas oleh Huang, Wu, and 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 nonlinear yang kompleks antara jarak spasial dan bobot pembobotan.

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

ar secara adaptif. Salah satunya adalah *Geographically Weighted Artificial Neural Network* (GWANN) yang dikembangkan oleh Hagenauer and Helbich (2022) serta *Spatial and Attribute Neural Network Weighted Regression* (SANNWR) oleh Ni, Wu, and 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, D. Zhu, Liu, Yao, and 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 antarunit observasi. Dengan mekanisme *message passing* yang mengagregasi informasi dari tetangga terdekat, SRGCNN tidak hanya mampu menangani struktur data non-Euclidean, tetapi juga mendukung pembelajaran semi-terawasi (*semi-supervised learning*) yang memungkinkan model belajar dari data yang sebagian tidak teramatii. 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., 2002b) 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 & Helbich, 2022), SANNWR (Ni dkk., 2022), dan GNNWR (Yin dkk., 2024). Di sisi lain, pendekatan *graph-based learning* seperti SRGCNN (D. 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 K adalah subhimpunan dari bilangan kompleks \mathbb{C} , K disebut sebagai lapangan apabila memenuhi kondisi berikut.

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

(b) Apabila $x \in K$, maka $-x$ juga merupakan elemen dari K . Lebih lanjut, jika $x \neq 0$, maka x^{-1} merupakan elemen dari K .

(c) 0 dan 1 merupakan elemen dari K .

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

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

$$\mathbb{Q} = \left\{ \frac{m}{n} : m, n \in \mathbb{Z}, n \neq 0 \right\}, \quad (2.1)$$

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

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd} \in \mathbb{Q}, \quad (2.2)$$

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

$$\left(\frac{a}{b} \right)^{-1} = \frac{b}{a} \in \mathbb{Q}. \quad (2.3)$$

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

Dalam ruang vektor dan matriks, dikenal juga istilah sublapangan. Apabila K dan L keduanya merupakan lapangan, serta dimisalkan bahwa $K \subseteq L$, maka K dapat disebut sebagai sublapangan atau *subfield* dari L . Elemen-elemen dari lapangan disebut sebagai bilangan atau skalar (Lang, 1987).

Definisi 2.1.3 (Ruang Vektor, Lang 1987) Suatu ruang vektor V atas suatu lapangan K adalah sebuah himpunan objek yang dapat dijumlahkan dan dikalikan dengan elemen-elemen dari K , sedemikian rupa sehingga hasil penjumlahan dua elemen dari V merupakan elemen V kembali, dan hasil perkalian sebuah elemen V dengan sebuah elemen dari K juga merupakan elemen dari V . Selain itu, sifat-sifat berikut dipenuhi:

- (a) $\forall u, v, w \in V$ berlaku $(u + v) + w = u + (v + w)$.
- (b) $\exists 0 \in V$ sedemikian sehingga $\forall u \in V$ berlaku $u + 0 = u$.
- (c) $\forall u \in V, \exists -u \in V$ sedemikian sehingga $u + (-u) = 0$.
- (d) $\forall u, v \in V$ berlaku $u + v = v + u$.
- (e) $\forall c \in K$ dan $\forall u, v \in V$ berlaku $c(u + v) = cu + cv$.
- (f) $\forall a, b \in K$ dan $\forall v \in V$ berlaku $(a + b)v = av + bv$.
- (g) $\forall a, b \in K$ dan $\forall v \in V$ berlaku $(ab)v = a(bv)$.
- (h) $\forall u \in V$ berlaku $1 \cdot u = u$, dengan 1 adalah elemen identitas pada K .

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

Definisi 2.1.4 (Ruang Matriks, (Lang, 1987)) Misalkan K merupakan lapangan dan $m, n \geq 1 \in \mathbb{Z}$. Matriks adalah suatu larik atau array dari bilangan-bilangan dalam K yang dinotasikan sebagai

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

Notasi matriks pada definisi di atas dapat dipersingkat dengan menulis (a_{ij}) , $i = 1, 2, \dots, m$ 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 seiap i, j .

Definisi 2.1.5 (Kebebasan Linear, (Lang, 1987)) Misalkan V ruang vektor atas lapangan K dan misalkan pula $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ elemen-elemen dari V , maka $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ dikatakan bebas linear jika dan hanya jika kapan pun a_1, a_2, \dots, a_n merupakan bilangan sehingga

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = \mathbf{0}, \quad (2.4)$$

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 K apabila terdapat elemen-elemen a_1, a_2, \dots, a_n di K yang semuanya tidak sama dengan nol, sedemikian sehingga Persamaan (2.4) terpenuhi.

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

berikut.

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

⋮

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

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 (2.4) terpenuhi. Sebab

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n = (a_1, a_2, \dots, a_n), \quad (2.5)$$

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

Definisi 2.1.7 (Span atau Jangkauan Linear, (Axler, 2024)) Misalkan V ruang vektor atas lapangan K dan $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\} \subseteq V$. Span dari S , ditulis $\text{span}(S)$, adalah himpunan semua kombinasi linear dari elemen-elemen S , yaitu

$$\text{span}(S) = \left\{ \sum_{i=1}^m a_i \mathbf{v}_i : a_i \in K \right\}. \quad (2.6)$$

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

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

$$\text{span}(S) = \{a(1, 0) + b(0, 1) : a, b \in \mathbb{R}\} = \mathbb{R}^2. \quad (2.7)$$

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

$$\text{span}(S) = \{a(1, 0, 0) + b(0, 1, 0) : a, b \in \mathbb{R}\} = \{(x, y, 0) : x, y \in \mathbb{R}\}, \quad (2.8)$$

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

Definisi 2.1.9 (Basis, (Axler, 2024)) Suatu basis V adalah himpunan vektor-vektor di V yang bebas linear dan menjangkau (spanning) V .

Contoh 2.1.10 Perhatikan vektor-vektor $\mathbf{v}_1 = (1, 2, -4)$ dan $\mathbf{v}_2 = (7, -5, 6)$ di dalam \mathbb{R}^3 . Jelas bahwa \mathbf{v}_1 dan \mathbf{v}_2 bebas linear, sebab tidak terdapat skalar $c \in \mathbb{R}$ sehingga $\mathbf{v}_2 = c\mathbf{v}_1$. Namun, himpunan $\{\mathbf{v}_1, \mathbf{v}_2\}$ hanya terdiri dari dua vektor, sehingga $\text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ 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 2.1.11 (Karakterisasi Basis, (Axler, 2024)) Sebuah list dari vektor-vektor $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ merupakan basis dari V jika dan hanya jika $\forall \mathbf{v} \in V$ dapat dituliskan secara unik dalam bentuk

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n, \quad (2.9)$$

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

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

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n. \quad (2.10)$$

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

$$0 = (a_1 - c_1)\mathbf{v}_1 + \cdots + (a_n - c_n)\mathbf{v}_n. \quad (2.11)$$

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 (2.9) bersifat unik. Hal ini menyelesaikan pembuktian arah pertama.

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

tuk

$$\mathbf{v} = a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n, \quad (2.12)$$

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

$$0 = a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n. \quad (2.13)$$

Sebab representasi (2.12) bersifat unik, khususnya untuk $\mathbf{v} = 0$, maka harus berlaku $a_1 = \cdots = a_n = 0$. Jadi $\mathbf{v}_1, \dots, \mathbf{v}_n$ bebas linear. Dengan demikian $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ adalah basis V . ■

Definisi 2.1.12 (Ruang Vektor Berdimensi-Hingga (Axler, 2024)) Suatu ruang vektor V atas medan F dikatakan berdimensi-hingga apabila terdapat suatu basis $\{v_1, \dots, v_n\}$ dari V yang terdiri atas sejumlah hingga (n buah) vektor. Dengan kata lain, V berdimensi-hingga jika V dapat dijangkau oleh suatu himpunan vektor berukuran hingga.

Definisi 2.1.13 (Dimensi (Axler, 2024)) Dimensi dari suatu ruang vektor berdimensi-hingga V adalah banyaknya elemen dalam setiap basis V . Sebab semua basis dari V memiliki panjang yang sama, maka bilangan ini terdefinisi dengan baik. Dimensi V dilambangkan dengan $\dim V$.

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

2.1.2. Sifat-Sifat Matriks

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

Definisi 2.1.15 (Kesamaan, Penjumlahan, dan Perkalian dengan Skalar) Untuk

$\mathbf{A} = (a_{ij})$, $\mathbf{B} = (b_{ij}) \in \text{Mat}_{m \times n}(K)$ dan $c \in K$, berlaku beberapa hal berikut.

- (a) $\mathbf{A} = \mathbf{B}$ jika dan hanya jika $a_{ij} = b_{ij}$ untuk semua i, j ;
- (b) $(\mathbf{A} + \mathbf{B})_{ij} = a_{ij} + b_{ij}$ (didefinisikan hanya bila ukuran sama); dan
- (c) $(c\mathbf{A})_{ij} = c a_{ij}$.

Definisi 2.1.16 (Perkalian Matriks, (Axler, 2024)) Jika $\mathbf{A} \in \text{Mat}_{m \times n}(K)$ dan $\mathbf{B} \in \text{Mat}_{n \times p}(K)$, maka hasil kali $\mathbf{AB} \in \text{Mat}_{m \times p}(K)$ didefinisikan oleh

$$(\mathbf{AB})_{ik} = \sum_{j=1}^n a_{ij} b_{jk} \quad (1 \leq i \leq m, 1 \leq k \leq p). \quad (2.14)$$

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

$$\underbrace{\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}}_{3 \times 2} \underbrace{\begin{pmatrix} 6 & 5 & 4 & 3 \\ 2 & 1 & 0 & -1 \end{pmatrix}}_{2 \times 4} = \underbrace{\begin{pmatrix} 10 & 7 & 4 & 1 \\ 26 & 19 & 12 & 5 \\ 42 & 31 & 20 & 9 \end{pmatrix}}_{3 \times 4}. \quad (2.15)$$

Proposisi 2.1.18 (Sifat-Sifat Elementer Perkalian Matriks, (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 $\mathbf{A} \in \text{Mat}_{m \times n}(K)$; dan
- (d) dapat terjadi $\mathbf{AB} \neq \mathbf{BA}$.

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

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

$$\begin{aligned}
 ((\mathbf{AB})\mathbf{C})_{ik} &= \sum_{\ell=1}^p (\mathbf{AB})_{i\ell} c_{\ell k} \\
 &= \sum_{\ell=1}^p \left(\sum_{j=1}^n a_{ij} b_{j\ell} \right) c_{\ell k} \\
 &= \sum_{j=1}^n \sum_{\ell=1}^p a_{ij} b_{j\ell} c_{\ell k}.
 \end{aligned} \tag{2.16}$$

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

$$\begin{aligned}
 (\mathbf{A}(\mathbf{BC}))_{ik} &= \sum_{j=1}^n a_{ij} (\mathbf{BC})_{jk} \\
 &= \sum_{j=1}^n a_{ij} \left(\sum_{\ell=1}^p b_{j\ell} c_{\ell k} \right) \\
 &= \sum_{j=1}^n \sum_{\ell=1}^p a_{ij} b_{j\ell} c_{\ell k}.
 \end{aligned} \tag{2.17}$$

Keduanya merupakan matriks yang sama.

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

$$\begin{aligned}
 (\mathbf{A}(\mathbf{B} + \mathbf{C}))_{ik} &= \sum_{j=1}^n a_{ij} (b_{jk} + c_{jk}) \\
 &= \sum_{j=1}^n a_{ij} b_{jk} + \sum_{j=1}^n a_{ij} c_{jk} \\
 &= (\mathbf{AB})_{ik} + (\mathbf{AC})_{ik}.
 \end{aligned} \tag{2.18}$$

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

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

$$(\mathbf{I}_m \mathbf{A})_{ik} = \sum_{j=1}^m (\mathbf{I}_m)_{ij} a_{jk} = \sum_{j=1}^m \delta_{ij} a_{jk} = a_{ik}, \tag{2.19}$$

dan

$$(\mathbf{A}\mathbf{I}_n)_{ik} = \sum_{j=1}^n a_{ij}(\mathbf{I}_n)_{jk} = \sum_{j=1}^n a_{ij}\delta_{jk} = a_{ik}, \quad (2.20)$$

dengan

$$\delta_{ij} = \begin{cases} 1, & \text{jika } i = j, \\ 0, & \text{jika } i \neq j. \end{cases} \quad (2.21)$$

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

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

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \quad (2.22)$$

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

■

Definisi 2.1.19 (Perkalian Hadamard, (Horn & Johnson, 2012)) Jika diberikan $\mathbf{A} = (a_{ij}) \in \text{Mat}_{m \times n}$ dan $\mathbf{B} = (b_{ij}) \in \text{Mat}_{m \times n}$, maka perkalian Hadamard atau perkalian Schur dari \mathbf{A} dan \mathbf{B} merupakan perkalian elemen per elemen dari matriks tersebut atau dapat dituliskan sebagai berikut.

$$\mathbf{A} \odot \mathbf{B} = (a_{ij}b_{ij}) \in \text{Mat}_{m \times n}. \quad (2.23)$$

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

Contoh 2.1.20 Misalkan

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

Maka hasil perkalian Hadamard adalah

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

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

Contoh 2.1.22 Salah satu contoh *transpose* dari matriks adalah

$$\mathbf{A} = \begin{pmatrix} -5 & 2 & 4 \\ 3 & 6 & -2 \end{pmatrix} \iff \mathbf{A}^\top = \begin{pmatrix} -5 & 3 \\ 2 & 6 \\ 4 & -2 \end{pmatrix}, \quad (2.24)$$

sedangkan untuk vektor adalah

$$\mathbf{a} = \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix} \iff \mathbf{a}^\top = (2, -3, 1). \quad (2.25)$$

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

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

- (a) $(\mathbf{A}^\top)^\top = \mathbf{A}$;
- (b) $(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$;
- (c) $(c \mathbf{A})^\top = c \mathbf{A}^\top$; dan
- (d) $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$.

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

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

$$((\mathbf{A}^\top)^\top)_{ij} = a_{ij} = (\mathbf{A})_{ij}. \quad (2.26)$$

(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

$$\begin{aligned} ((\mathbf{AB})^\top)_{ki} &= c_{ik} \\ &= \sum_j a_{ij}b_{jk} \\ &= \sum_j (\mathbf{B}^\top)_{kj}(\mathbf{A}^\top)_{ji} \\ &= (\mathbf{B}^\top \mathbf{A}^\top)_{ki}. \end{aligned} \quad (2.27)$$

■

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

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

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

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

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

Definisi 2.1.26 (Matriks Idempoten, (Dhrymes, 2013)) Suatu matriks persegi \mathbf{A} dikatakan idempoten jika dan hanya jika

$$\mathbf{AA} = \mathbf{A}^2 = \mathbf{A}. \quad (2.28)$$

Contoh 2.1.27 Matriks proyeksi merupakan contoh klasik dari matriks idempoten.

Misalkan

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

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

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

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

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

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

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

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}. \quad (2.29)$$

Contoh 2.1.29 Misalkan

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

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

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

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 2.1.30 (Sifat-Sifat Trace, (Axler, 2024)) Untuk ukuran yang sesuai, berlaku:

- (a) $\text{tr}(\alpha\mathbf{A} + \beta\mathbf{B}) = \alpha \text{ tr}(\mathbf{A}) + \beta \text{ tr}(\mathbf{B});$
- (b) $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA}),$ lebih umum $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{BCA});$ dan
- (c) jika \mathbf{B} invertibel, maka $\text{tr}(\mathbf{B}^{-1}\mathbf{AB}) = \text{tr}(\mathbf{A}).$

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

(a)

$$\begin{aligned} \text{tr}(\alpha\mathbf{A} + \beta\mathbf{B}) &= \sum_i (\alpha a_{ii} + \beta b_{ii}) \\ &= \alpha \sum_i a_{ii} + \beta \sum_i b_{ii} \\ &= \alpha \text{ tr}(\mathbf{A}) + \beta \text{ tr}(\mathbf{B}). \end{aligned} \tag{2.30}$$

(b) Untuk dua faktor,

$$\begin{aligned} \text{tr}(\mathbf{AB}) &= \sum_i (\mathbf{AB})_{ii} \\ &= \sum_i \sum_j a_{ij}b_{ji} \\ &= \sum_j \sum_i b_{ji}a_{ij} \\ &= \sum_j (\mathbf{BA})_{jj} \\ &= \text{tr}(\mathbf{BA}). \end{aligned} \tag{2.31}$$

Untuk tiga faktor,

$$\begin{aligned}
 \text{tr}(\mathbf{ABC}) &= \sum_i (\mathbf{ABC})_{ii} \\
 &= \sum_i \sum_j \sum_k a_{ij} b_{jk} c_{ki} \\
 &= \sum_j \sum_k \sum_i b_{jk} c_{ki} a_{ij} \\
 &= \text{tr}(\mathbf{BCA}). \tag{2.32}
 \end{aligned}$$

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

$$\begin{aligned}
 \text{tr}(\mathbf{B}^{-1}\mathbf{AB}) &= \text{tr}(\mathbf{ABB}^{-1}) \\
 &= \text{tr}(\mathbf{AI}) \\
 &= \text{tr}(\mathbf{A}). \tag{2.33}
 \end{aligned}$$

■

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

$$\mathbf{x}^\top \mathbf{Ax} \geq 0. \tag{2.34}$$

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

$$\mathbf{x}^\top \mathbf{Ax} > 0. \tag{2.35}$$

Contoh 2.1.32 Pertimbangkan matriks

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

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

$$\begin{aligned}\mathbf{x}^\top \mathbf{A} \mathbf{x} &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= 2x_1^2 + 2x_1x_2 + 2x_2^2 \\ &= x_1^2 + (x_1 + x_2)^2 + x_2^2 \geq 0.\end{aligned}\quad (2.36)$$

Lebih lanjut, jika $\mathbf{x} \neq 0$, 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 $\mathbf{x} \neq 0$.

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

Contoh 2.1.34 Pertimbangkan matriks

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

Matriks $-\mathbf{B}$ adalah

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

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

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

minan dari submatriks pojok kiri atas berukuran $k \times k$, yaitu

$$\Delta_k = \det \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix}, \quad k = 1, 2, \dots, n. \quad (2.37)$$

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

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

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

2.1.3. Invers Matriks

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

Definisi 2.1.35 (Invers Matriks, (Axler, 2024)) Misalkan $\mathbf{A} \in K^{n \times n}$. Matriks \mathbf{A} disebut invertibel (atau non-singular) apabila terdapat matriks $\mathbf{B} \in K^{n \times n}$ sedemikian sehingga

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n, \quad (2.38)$$

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

Contoh 2.1.36 Pertimbangkan matriks

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

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

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

Dapat diverifikasi bahwa

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

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

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

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

$$\mathbf{B} = \mathbf{BI}_n = \mathbf{B}(\mathbf{AC}) = (\mathbf{BA})\mathbf{C} = \mathbf{I}_n\mathbf{C} = \mathbf{C}. \quad (2.39)$$

■

Proposisi 2.1.38 (Sifat-Sifat Invers, (Axler, 2024)) *Misalkan $\mathbf{A}, \mathbf{B} \in K^{n \times n}$ invertibel dan $\alpha \in K$. Berlaku:*

- (a) $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$;
- (b) $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$; pada $K = \mathbb{C}$ juga berlaku $(\mathbf{A}^*)^{-1} = (\mathbf{A}^{-1})^*$;
- (c) $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$;
- (d) *Jika $\mathbf{A} = \text{diag}(\mathbf{B}_1, \dots, \mathbf{B}_k)$ dengan setiap \mathbf{B}_i invertibel (ukuran mungkin*

berbeda), maka

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

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

$$\mathbf{A}^{-1} = \mathbf{E}_k \cdots \mathbf{E}_1. \quad (2.40)$$

Secara praktis, dapat dilakukan lakukan reduksi baris teraugmentasi:

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

Contoh 2.1.39 Ambil

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

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

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

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

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

Maka

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

Definisi 2.1.40 (Kofaktor dan Matriks Adjoin, (Lang, 1987)) Misalkan $\mathbf{A} = (a_{ij}) \in K^{n \times n}$. Minor dari elemen a_{ij} , dinotasikan M_{ij} , adalah determinan submatriks berukuran $(n-1) \times (n-1)$ yang diperoleh dengan menghapus baris ke- i dan kolom ke- j dari \mathbf{A} . Kofaktor dari a_{ij} didefinisikan sebagai

$$C_{ij} = (-1)^{i+j} M_{ij}. \quad (2.41)$$

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

$$\text{adj}(\mathbf{A}) = \mathbf{C}^\top. \quad (2.42)$$

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

0, maka

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \text{adj}(\mathbf{A}). \quad (2.43)$$

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

$$(\mathbf{A} \cdot \text{adj}(\mathbf{A}))_{ij} = \sum_{k=1}^n a_{ik} \cdot (\text{adj}(\mathbf{A}))_{kj} = \sum_{k=1}^n a_{ik} C_{jk}, \quad (2.44)$$

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

Kasus 1: Jika $i = j$, maka

$$\sum_{k=1}^n a_{ik} C_{ik} \quad (2.45)$$

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

$$\det(\mathbf{A}) = \sum_{k=1}^n a_{ik} C_{ik}, \quad (2.46)$$

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

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

$$\sum_{k=1}^n a_{ik} C_{jk} \quad (2.47)$$

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

Dari kedua kasus di atas, diperoleh

$$\mathbf{A} \cdot \text{adj}(\mathbf{A}) = \det(\mathbf{A}) \cdot \mathbf{I}_n. \quad (2.48)$$

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

$$\mathbf{A} \cdot \frac{1}{\det(\mathbf{A})} \text{adj}(\mathbf{A}) = \mathbf{I}_n. \quad (2.49)$$

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

Contoh 2.1.42 Untuk matriks $\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 5 & 6 & 0 \end{pmatrix}$, akan dihitung inversnya menggunakan rumus adjoint. 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:

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

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

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

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

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

Dengan demikian, hasil perhitungan invers terverifikasi benar.

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

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

sehingga

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \quad (2.50)$$

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

2.1.4. Determinan dan *Rank*

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

Definisi 2.1.44 (Pemetaan Linear, (Axler, 2024)) Misalkan V, W ruang vektor atas suatu lapangan K . Suatu pemetaan $T : V \rightarrow W$ disebut pemetaan linear jika untuk semua $\mathbf{u}, \mathbf{v} \in V$ dan $\lambda \in K$ berlaku

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}), \quad T(\lambda\mathbf{v}) = \lambda T(\mathbf{v}). \quad (2.51)$$

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

Contoh 2.1.45 Berikut ini adalah beberapa contoh pemetaan linear.

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

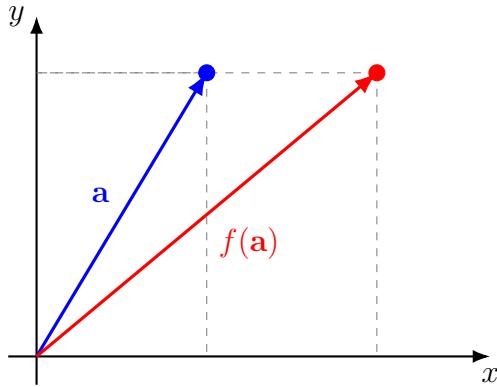
$$T(\mathbf{x}, \mathbf{y}) = (\mathbf{x} + 2\mathbf{y}, 3\mathbf{x} - \mathbf{y}) \quad (2.52)$$

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

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

$$T(x_1, x_2, x_3) = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \quad (2.53)$$

Transformasi linear dapat divisualisasikan sebagai pemetaan yang membawa suatu vektor ke vektor lain dengan cara yang teratur (tanpa membengkokkan ruang). Sebagai contoh, Gambar 2.1 berikut memperlihatkan fungsi linear $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ 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 $f(x, y) = (2x, y)$.
(Sumber: Dokumen penulis)

Definisi 2.1.46 (Null Space, Range, dan Rank, (Axler, 2024)) Untuk $T \in \mathcal{L}(V, W)$, null space dan range didefinisikan sebagai

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

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

Contoh 2.1.47 Misalkan

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

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

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

sehingga

$$\text{null } T = \text{span}\{(5, -4, 1)^\top\} \quad \text{dan} \quad \dim(\text{null } T) = 1. \quad (2.54)$$

(b) Kolom pertama dan kedua bebas linear, maka

$$\text{range } T = \text{span}\{(1, 0)^\top, (2, 1)^\top\} \quad (2.55)$$

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

$$\dim V = \dim(\text{null } T) + \dim(\text{range } T). \quad (2.56)$$

Contoh 2.1.49 Berdasarkan Contoh 2.1.47, 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 2.1.50 (Bentuk Multilinear Anti-Simetri, (Axler, 2024)) *Misalkan V ruang vektor atas lapangan K . Suatu fungsi $\omega : V^n \rightarrow K$ disebut bentuk n -linear jika linear pada setiap argumennya. Bentuk n -linear ω disebut anti-simetri atau alternating jika*

$$\omega(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n) = -\omega(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_i, \dots, \mathbf{v}_n) \quad (2.57)$$

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

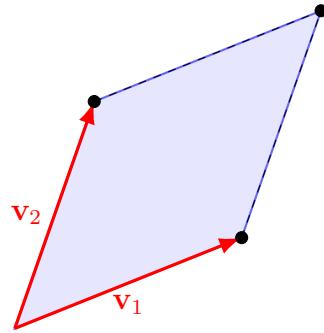
Contoh 2.1.51 Pada $V = \mathbb{R}^2$, fungsi

$$\omega((x_1, y_1), (x_2, y_2)) = x_1y_2 - x_2y_1 \quad (2.58)$$

adalah bentuk bilinear anti-simetri, karena

$$\omega((x_2, y_2), (x_1, y_1)) = x_2y_1 - x_1y_2 = -\omega((x_1, y_1), (x_2, y_2)). \quad (2.59)$$

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



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

Luas dari jajar genjang pada Gambar 2.2 adalah

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

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

$$\begin{aligned} |\mathbf{v}_1||\mathbf{v}_2| \sin(\theta) &= \sqrt{|\mathbf{v}_1|^2|\mathbf{v}_2|^2 - (x_1x_2 + y_1y_2)^2} \\ &= \sqrt{x_1^2x_2^2 + x_1^2y_2^2 + y_1^2x_2^2 + y_1^2y_2^2 - (x_1^2x_2^2 + 2x_1x_2y_1y_2 + y_1^2y_2^2)} \\ &= \sqrt{x_1^2y_2^2 + y_1^2x_2^2 - 2x_1x_2y_1y_2} \\ &= \sqrt{(x_1y_2 - x_2y_1)^2} \\ &= |x_1y_2 - x_2y_1|. \end{aligned} \quad (2.60)$$

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

Contoh 2.1.51 memberikan intuisi penting tentang determinan. Fungsi bilinear anti-simetri $\omega(v_1, v_2) = x_1y_2 - x_2y_1$ tidak lain adalah determinannya pada \mathbb{R}^2 . Secara

geometri, determinan mengukur luas terorientasi dari jajar genjang yang dibentang oleh dua vektor. Nilai absolut $|\det|$ memberikan ukuran luas (atau volume pada dimensi lebih tinggi), sedangkan tanda determinan membedakan orientasi. Jika \mathbf{v}_1 ke \mathbf{v}_2 berputar berlawanan arah jarum jam, maka $\det(\cdot) > 0$ dan sebaliknya jika searah jarum jam, maka $\det(\cdot) < 0$.

Definisi 2.1.52 (Determinan, (Axler, 2024)) Misalkan V berdimensi $n \geq 1$. Ruang semua bentuk n -linear beranti-simetri pada V memiliki dimensi 1. Untuk setiap $T \in \mathcal{L}(V)$, determinan T , ditulis $\det T$, adalah satu-satunya skalar di K dengan sifat

$$\omega(T\mathbf{v}_1, \dots, T\mathbf{v}_n) = (\det T) \omega(\mathbf{v}_1, \dots, \mathbf{v}_n), \quad (2.61)$$

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

Contoh 2.1.53 Ambil $V = \mathbb{R}^2$ dan operator linear $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ 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

$$\omega(\mathbf{v}_1, \mathbf{v}_2) = \det \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = 3. \quad (2.62)$$

Sekarang apabila dilihat bayangan kedua vektor tersebut,

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

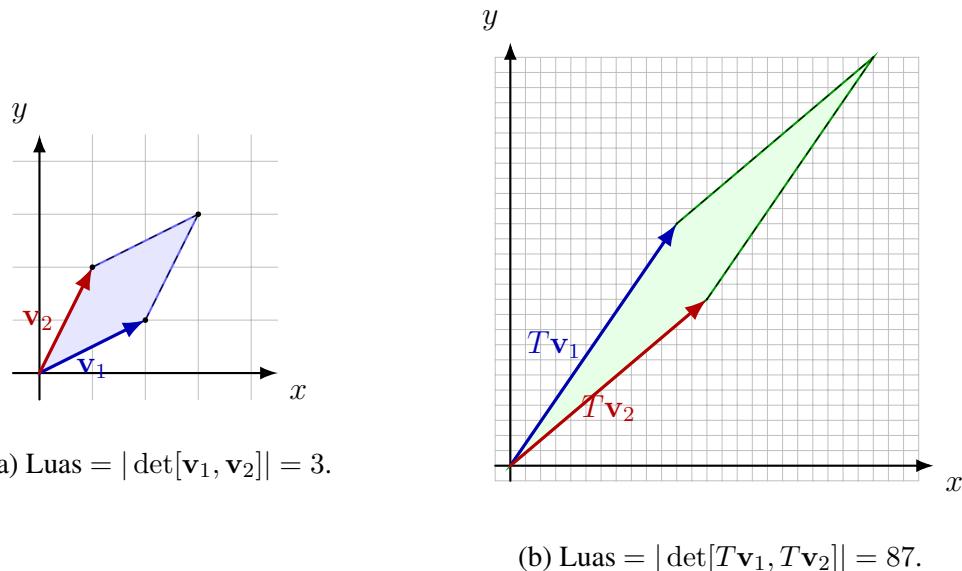
sehingga

$$\omega(T\mathbf{v}_1, T\mathbf{v}_2) = \det \begin{pmatrix} 11 & 13 \\ 16 & 11 \end{pmatrix} = 11 \cdot 11 - 13 \cdot 16 = -87. \quad (2.63)$$

Dengan demikian, sesuai definisi determinan operator,

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

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



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

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

Teorema 2.1.54 (Sifat Dasar Determinan, (Axler, 2024)) Untuk $S, T \in \mathcal{L}(V)$ dan skalar $c \in K$ berlaku:

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

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

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

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

$$\omega(\mathbf{I}_V \mathbf{v}_1, \dots, \mathbf{I}_V \mathbf{v}_n) = \omega(\mathbf{v}_1, \dots, \mathbf{v}_n). \quad (2.64)$$

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

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

$$\begin{aligned} \omega((ST)\mathbf{v}_1, \dots, (ST)\mathbf{v}_n) &= \omega(S(T\mathbf{v}_1), \dots, S(T\mathbf{v}_n)) \\ &= (\det S) \omega(T\mathbf{v}_1, \dots, T\mathbf{v}_n) \\ &= (\det S)(\det T) \omega(\mathbf{v}_1, \dots, \mathbf{v}_n). \end{aligned} \quad (2.65)$$

Di sisi lain, berdasarkan definisi determinan untuk ST :

$$\omega((ST)\mathbf{v}_1, \dots, (ST)\mathbf{v}_n) = (\det(ST)) \omega(\mathbf{v}_1, \dots, \mathbf{v}_n). \quad (2.66)$$

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

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

$$1 = \det \mathbf{I}_V = \det(TT^{-1}) = \det T \cdot \det T^{-1}. \quad (2.67)$$

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

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

maka untuk setiap bentuk n -linear beranti-simetri ω :

$$\omega(T\mathbf{v}_1, \dots, T\mathbf{v}_n) = 0. \quad (2.68)$$

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

$$0 = \omega(T\mathbf{v}_1, \dots, T\mathbf{v}_n) = (\det T) \omega(\mathbf{v}_1, \dots, \mathbf{v}_n), \quad (2.69)$$

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

■

2.1.5. Nilai Eigen dan Vektor Eigen

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

Definisi 2.1.55 (Vektor Eigen dan Nilai Eigen, (Axler, 2024, Lang, 1987)) Misalkan $T : V \rightarrow V$ adalah operator linear pada ruang vektor V atas \mathbb{C} .

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

$$T(\mathbf{v}) = \lambda\mathbf{v}. \quad (2.70)$$

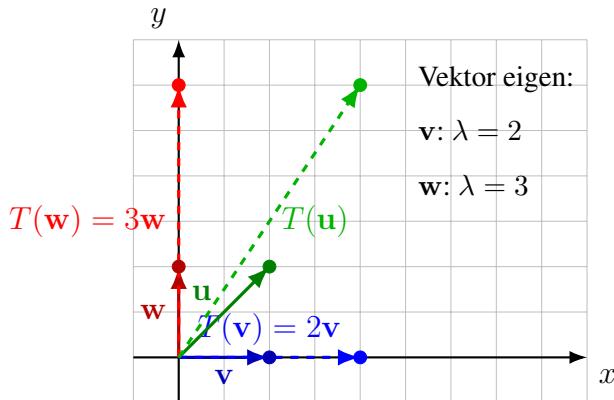
- (b) Skalar λ yang memenuhi persamaan di atas disebut nilai eigen dari T .

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

Contoh 2.1.56 Pertimbangkan $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dengan

$$T(x, y) = (2x, 3y). \quad (2.71)$$

- (a) Untuk vektor $\mathbf{v} = (1, 0)$, berlaku $T(1, 0) = (2, 0) = 2(1, 0)$, sehingga \mathbf{v} adalah vektor eigen dengan nilai $\lambda = 2$.
- (b) Untuk vektor $\mathbf{w} = (0, 1)$, berlaku $T(0, 1) = (0, 3) = 3(0, 1)$, sehingga \mathbf{w} adalah vektor eigen dengan nilai $\lambda = 3$.



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

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

Teorema 2.1.57 (Keberadaan Nilai Eigen pada Ruang Vektor Kompleks, (Axler, 2024)) *Jika V adalah ruang vektor berdimensi-hingga atas \mathbb{C} dan $T: V \rightarrow V$ operator linear, maka T memiliki setidaknya satu nilai eigen.*

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

$$p_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}_n). \quad (2.72)$$

Polinomial ini berderajat n dalam variabel λ , dengan koefisien dalam \mathbb{C} . Teorema dasar aljabar menyatakan bahwa setiap polinomial kompleks nonkonstan memiliki

akar. Sebab $\deg p_{\mathbf{A}}(\lambda) = n \geq 1$, maka terdapat $\lambda_0 \in \mathbb{C}$ sehingga

$$p_{\mathbf{A}}(\lambda_0) = \det(\mathbf{A} - \lambda_0 \mathbf{I}_n) = 0. \quad (2.73)$$

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

$$(\mathbf{A} - \lambda_0 \mathbf{I}_n)\mathbf{v} = \mathbf{0}. \quad (2.74)$$

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

$$T(\mathbf{v}) = \lambda_0\mathbf{v}, \quad \mathbf{v} \neq \mathbf{0}. \quad (2.75)$$

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

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

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

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

$$\begin{aligned} \det \begin{pmatrix} 4 - \lambda & 1 \\ 2 & 3 - \lambda \end{pmatrix} &= 0 \\ (4 - \lambda)(3 - \lambda) - (1)(2) &= 0 \\ 12 - 4\lambda - 3\lambda + \lambda^2 - 2 &= 0 \\ \lambda^2 - 7\lambda + 10 &= 0 \\ (\lambda - 5)(\lambda - 2) &= 0. \end{aligned} \quad (2.76)$$

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

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

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

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

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

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

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

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

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

2.1.6. Perkalian Kronecker

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

Definisi 2.1.59 (Perkalian Kronecker, (Hackbusch, 2012)) Untuk $\mathbf{A} \in \text{Mat}_{m \times n}$ dan $\mathbf{B} \in \text{Mat}_{p \times q}$, perkalian Kronecker dari matriks tersebut dituliskan sebagai $\mathbf{A} \otimes \mathbf{B}$, didefinisikan sebagai

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix} \in \text{Mat}_{mp \times nq}. \quad (2.77)$$

Contoh 2.1.60 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

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} 1 \cdot \mathbf{B} & 2 \cdot \mathbf{B} \\ 3 \cdot \mathbf{B} & 4 \cdot \mathbf{B} \end{pmatrix} = \begin{pmatrix} 0 & 5 & 0 & 10 \\ 6 & 7 & 12 & 14 \\ 0 & 15 & 0 & 20 \\ 18 & 21 & 24 & 28 \end{pmatrix}. \quad (2.78)$$

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

2.1.7. Diferensial Vektor dan Matriks

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

Definisi 2.1.61 (Konvensi Turunan pada Vektor dan Matriks, (Dhrymes, 2013))

Misalkan

$$\mathbf{y} = \psi(\mathbf{x}), \quad (2.79)$$

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

Notasi

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \left[\frac{\partial y_i}{\partial x_j} \right], \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n \quad (2.80)$$

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 x , yaitu

$$\frac{\partial y_i}{\partial x_1}, \frac{\partial y_i}{\partial x_2}, \dots, \frac{\partial y_i}{\partial x_n}. \quad (2.81)$$

Proposisi 2.1.62 *Jika*

$$\mathbf{y} = \mathbf{Ax}, \quad (2.82)$$

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

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{A}. \quad (2.83)$$

Proposisi 2.1.63 *Jika*

$$\mathbf{y} = \mathbf{Ax}, \quad (2.84)$$

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

$$\frac{\partial \mathbf{y}}{\partial \boldsymbol{\alpha}} = (\mathbf{x}^\top \otimes \mathbf{I}_m) \frac{\partial \mathbf{A}}{\partial \boldsymbol{\alpha}} + \mathbf{A} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\alpha}}. \quad (2.85)$$

Proposisi 2.1.64 *Jika*

$$\mathbf{y} = \mathbf{z}^\top \mathbf{Ax}, \quad (2.86)$$

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

$$\frac{\partial \mathbf{y}}{\partial \mathbf{z}} = \mathbf{x}^\top \mathbf{A}^\top, \quad \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{z}^\top \mathbf{A}. \quad (2.87)$$

Proposisi 2.1.65 *Jika*

$$\mathbf{y} = \mathbf{x}^\top \mathbf{Ax}, \quad (2.88)$$

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

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{x}^\top (\mathbf{A} + \mathbf{A}^\top). \quad (2.89)$$

Jika \mathbf{A} adalah matriks simetri, maka

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = 2\mathbf{x}^\top \mathbf{A}. \quad (2.90)$$

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

$$\frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{A}^{-1}. \quad (2.91)$$

Teorema 2.1.67 (Aturan Jejak, (Dhrymes, 2013)) Untuk matriks dengan ukuran yang sesuai berlaku:

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

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

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

$$\text{tr}(\mathbf{AX}) = \sum_{i=1}^m (\mathbf{AX})_{ii} = \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_{ji}. \quad (2.92)$$

Turunan parsial terhadap x_{kl} adalah

$$\frac{\partial}{\partial x_{kl}} \text{tr}(\mathbf{AX}) = \frac{\partial}{\partial x_{kl}} \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_{ji} = a_{lk}, \quad (2.93)$$

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

$$\left(\frac{\partial}{\partial \mathbf{X}} \text{tr}(\mathbf{AX}) \right)_{kl} = a_{lk} = (\mathbf{A}^\top)_{kl}, \quad (2.94)$$

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

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

$$\text{tr}(\mathbf{X}^\top \mathbf{AX}) = \sum_{i,j,k} x_{ji} a_{jk} x_{ki}. \quad (2.95)$$

Turunan parsial terhadap x_{pq} adalah

$$\begin{aligned} \frac{\partial}{\partial x_{pq}} \text{tr}(\mathbf{X}^\top \mathbf{AX}) &= \frac{\partial}{\partial x_{pq}} \sum_{i,j,k} x_{ji} a_{jk} x_{ki} \\ &= \sum_k a_{pk} x_{kq} + \sum_j x_{jq} a_{jp} \\ &= (\mathbf{AX})_{pq} + (\mathbf{X}^\top \mathbf{A}^\top)_{qp}^\top \\ &= (\mathbf{AX})_{pq} + (\mathbf{A}^\top \mathbf{X})_{pq}. \end{aligned} \quad (2.96)$$

Oleh karena itu,

$$\frac{\partial}{\partial \mathbf{X}} \text{tr}(\mathbf{X}^\top \mathbf{AX}) = \mathbf{AX} + \mathbf{A}^\top \mathbf{X} = (\mathbf{A} + \mathbf{A}^\top) \mathbf{X}. \quad (2.97)$$

■

Contoh 2.1.68 Dalam banyak permasalahan estimasi, seringkali ingin dicari vektor parameter β 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 $\mathbf{X}\beta$, yaitu

$$S(\beta) = (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta). \quad (2.98)$$

Fungsi $S(\beta)$ ini mengukur seberapa jauh vektor $\mathbf{X}\beta$ dari vektor \mathbf{y} . Dengan meng-

ekspansi perkalian dalam, diperoleh

$$\begin{aligned}
 S(\boldsymbol{\beta}) &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
 &= \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{X}\boldsymbol{\beta} - (\mathbf{X}\boldsymbol{\beta})^\top \mathbf{y} + (\mathbf{X}\boldsymbol{\beta})^\top \mathbf{X}\boldsymbol{\beta} \\
 &= \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}.
 \end{aligned} \tag{2.99}$$

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

$$S(\boldsymbol{\beta}) = \mathbf{y}^\top \mathbf{y} - 2\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}. \tag{2.100}$$

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

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

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

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{y}^\top \mathbf{X} + 2\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}. \tag{2.101}$$

Untuk menemukan titik stasioner, disamakan gradien dengan nol:

$$\begin{aligned}
 -2\mathbf{y}^\top \mathbf{X} + 2\widehat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{X} &= \mathbf{0}^\top \\
 \widehat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{X} &= \mathbf{y}^\top \mathbf{X}.
 \end{aligned} \tag{2.102}$$

Dengan mengambil *transpose* kedua ruas, diperoleh

$$\mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{y}. \quad (2.103)$$

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

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (2.104)$$

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

$$\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = 2 \mathbf{X}^\top \mathbf{X}. \quad (2.105)$$

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

$$\mathbf{v}^\top \mathbf{X}^\top \mathbf{X} \mathbf{v} = (\mathbf{X}\mathbf{v})^\top (\mathbf{X}\mathbf{v}) = \|\mathbf{X}\mathbf{v}\|^2 \geq 0. \quad (2.106)$$

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

2.2. Teori Ukuran dan Probabilitas

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

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

- (a) $\Omega \in \mathcal{F}$;
- (b) jika $A \in \mathcal{F}$, maka $A^c \in \mathcal{F}$; dan
- (c) jika $A, B \in \mathcal{F}$, maka $A \cup B \in \mathcal{F}$.

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

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

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

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

2.2.1. Ruang Ukur dan Ukuran Probabilitas

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

- (a) $\Omega \in \mathcal{F}$;
- (b) jika $A \in \mathcal{F}$, maka $A^c \in \mathcal{F}$; dan
- (c) jika $\{A_n\}_{n=1}^\infty \subseteq \mathcal{F}$, maka $\bigcup_{n=1}^\infty A_n \in \mathcal{F}$.

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

- (a) σ -aljabar trivial: $\mathcal{F}_1 = \{\emptyset, \Omega\} = \{\emptyset, \{1, 2, 3\}\}$. Koleksi ini merupakan σ -aljabar terkecil pada Ω .
- (b) σ -aljabar power set: $\mathcal{F}_2 = 2^\Omega = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$. Koleksi ini merupakan σ -aljabar terbesar pada Ω yang memuat semua subhimpunan.

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

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

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

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

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

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

Definisi 2.2.7 (Ukuran) Diberikan ruang sampel Ω dan σ -aljabar \mathcal{F} relatif terhadap Ω . Suatu fungsi $\mu : \mathcal{F} \rightarrow \mathbb{R}^+$ 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 2.2.8 (Ukuran Probabilitas, (?)) *Suatu fungsi* $\Pr : \mathcal{F} \rightarrow [0, 1]$ *disebut sebagai ukuran probabilitas apabila:*

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

$$\Pr\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \Pr(A_n). \quad (2.107)$$

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

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

$$\Pr(A) = \frac{|A|}{6} \quad (2.108)$$

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

Akan diperiksa bahwa \Pr memenuhi aksioma ukuran probabilitas.

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

$$\Pr\left(\bigcup_{n=1}^{\infty} A_n\right) = \frac{|\bigcup_{n=1}^{\infty} A_n|}{6} = \frac{\sum_{n=1}^{\infty} |A_n|}{6} = \sum_{n=1}^{\infty} \frac{|A_n|}{6} = \sum_{n=1}^{\infty} \Pr(A_n). \quad (2.109)$$

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

$$\Pr(\{2, 4, 6\}) = \frac{3}{6} = \frac{1}{2}. \quad (2.110)$$

2.2.2. Variabel Acak sebagai Fungsi Terukur

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

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}. \quad (2.111)$$

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

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

$$\Pr_X(B) = \Pr(X^{-1}(B)) = \Pr(\{\omega \in \Omega : X(\omega) \in B\}), \quad \forall B \in \mathcal{B}(\mathbb{R}). \quad (2.112)$$

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

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

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

$$\Pr_X(\mathbb{R}) = \Pr(X^{-1}(\mathbb{R})) = \Pr(\Omega) = 1. \quad (2.113)$$

- (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 (Ω, \mathcal{F}) , berlaku

$$\Pr_X(B) = \Pr(X^{-1}(B)) \geq 0. \quad (2.114)$$

(c) **σ -aditivitas:** Misalkan $\{B_n\}_{n=1}^{\infty}$ adalah barisan himpunan saling lepas dalam $\mathcal{B}(\mathbb{R})$, yaitu $B_i \cap B_j = \emptyset$ 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

$$X^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right) = \bigcup_{n=1}^{\infty} X^{-1}(B_n). \quad (2.115)$$

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

$$\begin{aligned} \Pr_X\left(\bigcup_{n=1}^{\infty} B_n\right) &= \Pr\left(X^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right)\right) \\ &= \Pr\left(\bigcup_{n=1}^{\infty} X^{-1}(B_n)\right) \\ &= \sum_{n=1}^{\infty} \Pr(X^{-1}(B_n)) \\ &= \sum_{n=1}^{\infty} \Pr_X(B_n). \end{aligned} \quad (2.116)$$

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

2.2.3. Distribusi dan Ekspektasi

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

$$\Pr_X(B) = \Pr(X \in B), \quad B \in \mathcal{B}(\mathbb{R}). \quad (2.117)$$

Definisi 2.2.13 (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*

$$\Pr_X(B) = \Pr(X \in B) = \int_B f_X(x) dx, \quad (2.118)$$

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

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

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

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

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

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

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

fungsi kepadatan probabilitas

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}. \quad (2.119)$$

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

$$\Pr(X = k) = p^k(1-p)^{1-k}, \quad k \in \{0, 1\}. \quad (2.120)$$

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

$$\Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \in \{0, 1, 2, \dots\}. \quad (2.121)$$

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

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0. \quad (2.122)$$

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

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

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

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

$$\lambda \left(\bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \lambda(A_n); \quad (2.123)$$

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

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

Contoh 2.2.16 Beberapa contoh perhitungan ukuran Lebesgue adalah sebagai berikut.

- (a) Untuk interval $[2, 5]$, berlaku $\lambda([2, 5]) = 5 - 2 = 3$.
- (b) Untuk himpunan singleton $\{a\}$, berlaku $\lambda(\{a\}) = 0$ karena $\{a\} = \bigcap_{n=1}^{\infty} [a - 1/n, a + 1/n]$ dan $\lambda([a - 1/n, a + 1/n]) = 2/n \rightarrow 0$.
- (c) Untuk himpunan bilangan rasional \mathbb{Q} , berlaku $\lambda(\mathbb{Q}) = 0$ karena \mathbb{Q} terhitung dan merupakan gabungan terhitung dari singleton.

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

$$s(\omega) = \sum_{k=1}^m c_k \mathbf{1}_{A_k}(\omega), \quad (2.124)$$

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

Contoh 2.2.18 Misalkan $\Omega = [0, 3]$ dan didefinisikan fungsi $s : \Omega \rightarrow \mathbb{R}$ 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 2.2.19 (Integral Lebesgue untuk Fungsi Sederhana, (?)) Untuk fungsi sederhana non-negatif $s = \sum_{k=1}^m c_k \mathbf{1}_{A_k}$ pada ruang ukur $(\Omega, \mathcal{F}, \mu)$, integral Lebesgue didefinisikan sebagai

$$\int_{\Omega} s d\mu = \sum_{k=1}^m c_k \mu(A_k). \quad (2.125)$$

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

$$\int_{\Omega} f d\mu = \sup \left\{ \int_{\Omega} s d\mu : s \text{ sederhana}, 0 \leq s \leq f \right\}. \quad (2.126)$$

Integral ini selalu terdefinisi (meskipun mungkin bernilai ∞).

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

$$\int_{\Omega} f d\mu = \int_{\Omega} f^+ d\mu - \int_{\Omega} f^- d\mu, \quad (2.127)$$

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

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

$$\int_{[0,1]} x^2 d\lambda(x) = \int_0^1 x^2 dx = \frac{1}{3}. \quad (2.128)$$

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

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

Definisi 2.2.23 (Ekspektasi Lebesgue, (?)) Misalkan $(\Omega, \mathcal{F}, \Pr)$ adalah ruang probabilitas dan $X : \Omega \rightarrow \mathbb{R}$ 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

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\Pr(\omega). \quad (2.129)$$

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

$$\mathbb{E}[X] = \int_{\mathbb{R}} x d\Pr_X(x). \quad (2.130)$$

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

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx. \quad (2.131)$$

Contoh 2.2.24 Berikut adalah contoh perhitungan ekspektasi untuk beberapa distribusi.

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

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=0}^1 k \cdot P(X = k) \\ &= 0 \cdot (1 - p) + 1 \cdot p = p.\end{aligned}\quad (2.132)$$

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

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) dx = \mu. \quad (2.133)$$

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

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

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!} \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda.\end{aligned}\quad (2.134)$$

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

$$\mathbb{E}[X] = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx = \frac{1}{\lambda}. \quad (2.135)$$

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

Teorema 2.2.25 (Sifat-Sifat Ekspektasi, (?)) Misalkan X dan Y adalah varia-

bel acak pada ruang probabilitas $(\Omega, \mathcal{F}, \Pr)$ dengan ekspektasi terdefinisi. Berlaku sifat-sifat berikut.

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

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]. \quad (2.136)$$

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

$$\mathbb{E}[X] \leq \mathbb{E}[Y]. \quad (2.137)$$

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

$$\mathbb{E}[c] = c. \quad (2.138)$$

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

$$\mathbb{E}[X] \geq 0. \quad (2.139)$$

(e) **Ketaksamaan segitiga:** Berlaku

$$|\mathbb{E}[X]| \leq \mathbb{E}[|X|]. \quad (2.140)$$

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

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]. \quad (2.141)$$

Bukti. Berikut adalah pembuktian untuk setiap sifat ekspektasi.

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

$$\begin{aligned}\mathbb{E}[aX + bY] &= \int_{\Omega} (aX(\omega) + bY(\omega)) d\Pr(\omega) \\ &= a \int_{\Omega} X(\omega) d\Pr(\omega) + b \int_{\Omega} Y(\omega) d\Pr(\omega) \\ &= a\mathbb{E}[X] + b\mathbb{E}[Y].\end{aligned}\tag{2.142}$$

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

$$\mathbb{E}[Y - X] = \mathbb{E}[Y] - \mathbb{E}[X] \geq 0,\tag{2.143}$$

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

$$\mathbb{E}[c] = \int_{\Omega} c d\Pr(\omega) = c \cdot \Pr(\Omega) = c \cdot 1 = c.\tag{2.144}$$

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

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\Pr(\omega) \geq 0.\tag{2.145}$$

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

$$-\mathbb{E}[|X|] \leq \mathbb{E}[X] \leq \mathbb{E}[|X|],\tag{2.146}$$

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 teore-

ma Fubini, berlaku

$$\begin{aligned}
 \mathbb{E}[XY] &= \int_{\mathbb{R}^2} xy d\Pr_{X,Y}(x,y) \\
 &= \int_{\mathbb{R}} x d\Pr_X(x) \cdot \int_{\mathbb{R}} y d\Pr_Y(y) \\
 &= \mathbb{E}[X] \cdot \mathbb{E}[Y].
 \end{aligned} \tag{2.147}$$

■

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

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) dF_X(x). \tag{2.148}$$

Secara khusus:

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

$$\mathbb{E}[g(X)] = \sum_x g(x) p_X(x). \tag{2.149}$$

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

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx. \tag{2.150}$$

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 diin-

duksi oleh X . Secara formal, untuk setiap fungsi terukur g , berlaku

$$\mathbb{E}[g(X)] = \int_{\Omega} g(X(\omega)) d\Pr(\omega) = \int_{\mathbb{R}} g(x) d\Pr_X(x), \quad (2.151)$$

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

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx. \quad (2.152)$$

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

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

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

$$\begin{aligned} \mathbb{E}[X^2] &= \int_0^{\infty} x^2 \cdot e^{-x} dx \\ &= \Gamma(3) = 2! = 2. \end{aligned} \quad (2.153)$$

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:

$$\begin{aligned}
 \mathbb{E}[e^X] &= \int_{-\infty}^{\infty} e^x \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{x-x^2/2} dx \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x^2-2x)/2} dx \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-1)^2/2+1/2} dx \\
 &= e^{1/2} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-1)^2/2} dx \\
 &= e^{1/2} = \sqrt{e}.
 \end{aligned} \tag{2.154}$$

2.2.4. Momen, Variansi, dan Kovariansi

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

$$\mu_k = \mathbb{E}[X^k], \tag{2.155}$$

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

$$\mu'_k = \mathbb{E}[(X - \mathbb{E}[X])^k]. \tag{2.156}$$

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

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]. \tag{2.157}$$

Variansi dapat juga dihitung menggunakan rumus alternatif

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2. \tag{2.158}$$

Contoh 2.2.30 Misalkan X adalah variabel acak diskrit dengan distribusi

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

Momen pertama (rataan) dari X adalah

$$\begin{aligned} \mathbb{E}[X] &= \sum_x x \cdot \Pr(X = x) \\ &= 1 \cdot 0.3 + 2 \cdot 0.5 + 3 \cdot 0.2 = 0.3 + 1.0 + 0.6 = 1.9. \end{aligned} \quad (2.159)$$

Momen kedua dari X adalah

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_x x^2 \cdot \Pr(X = x) \\ &= 1^2 \cdot 0.3 + 2^2 \cdot 0.5 + 3^2 \cdot 0.2 = 0.3 + 2.0 + 1.8 = 4.1. \end{aligned} \quad (2.160)$$

Variansi dari X dapat dihitung dengan rumus alternatif, yaitu

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 4.1 - (1.9)^2 = 4.1 - 3.61 = 0.49. \quad (2.161)$$

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

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]. \quad (2.162)$$

Rumus alternatif untuk kovariansi adalah

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \quad (2.163)$$

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

bersama

$$\begin{aligned}\Pr(X = 0, Y = 0) &= 0.2, & \Pr(X = 0, Y = 1) &= 0.1, \\ \Pr(X = 1, Y = 0) &= 0.3, & \Pr(X = 1, Y = 1) &= 0.4.\end{aligned}\quad (2.164)$$

Rataan dari X dan Y adalah

$$\mathbb{E}[X] = 0 \cdot (0.2 + 0.1) + 1 \cdot (0.3 + 0.4) = 0.7, \quad (2.165)$$

$$\mathbb{E}[Y] = 0 \cdot (0.2 + 0.3) + 1 \cdot (0.1 + 0.4) = 0.5. \quad (2.166)$$

Nilai harapan dari XY adalah

$$\begin{aligned}\mathbb{E}[XY] &= \sum_{x,y} xy \cdot \Pr(X = x, Y = y) \\ &= 0 \cdot 0 \cdot 0.2 + 0 \cdot 1 \cdot 0.1 + 1 \cdot 0 \cdot 0.3 + 1 \cdot 1 \cdot 0.4 = 0.4.\end{aligned}\quad (2.167)$$

Kovariansi dari X dan Y adalah

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0.4 - 0.7 \cdot 0.5 = 0.4 - 0.35 = 0.05. \quad (2.168)$$

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

2.2.5. Probabilitas Bersyarat dan Independensi

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

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

$$\Pr(A | B) = \frac{\Pr(A \cap B)}{\Pr(B)}. \quad (2.169)$$

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

$$\Pr(A) = \frac{3}{6} = \frac{1}{2}, \quad \Pr(B) = \frac{3}{6} = \frac{1}{2}, \quad (2.170)$$

$$\Pr(A \cap B) = \Pr(\{4, 6\}) = \frac{2}{6} = \frac{1}{3}. \quad (2.171)$$

Probabilitas bersyarat dari A diberikan B adalah

$$\Pr(A | B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{1/3}{1/2} = \frac{2}{3}. \quad (2.172)$$

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

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

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B). \quad (2.173)$$

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

Contoh 2.2.36 Pertimbangkan pelemparan dua koin seimbang secara independen.

Definisikan $A = \{\text{koin pertama muncul kepala}\}$ dan $B = \{\text{koin kedua muncul kepala}\}$.

Ruang sampel adalah $\Omega = \{KK, KE, EK, EE\}$ dengan probabilitas seragam.

Maka

$$\Pr(A) = \Pr(\{KK, KE\}) = \frac{2}{4} = \frac{1}{2}, \quad (2.174)$$

$$\Pr(B) = \Pr(\{KK, EK\}) = \frac{2}{4} = \frac{1}{2}, \quad (2.175)$$

$$\Pr(A \cap B) = \Pr(\{KK\}) = \frac{1}{4}. \quad (2.176)$$

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

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

$$\Pr(X \in A, Y \in B) = \Pr(X \in A) \Pr(Y \in B). \quad (2.177)$$

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

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y), \quad \forall x, y \in \mathbb{R}. \quad (2.178)$$

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

$$\Pr(X = 1, Y = 1) = \Pr(X = 1) \cdot \Pr(Y = 1) = 0.5 \cdot 0.5 = 0.25, \quad (2.179)$$

$$\Pr(X = 0, Y = 1) = \Pr(X = 0) \cdot \Pr(Y = 1) = 0.5 \cdot 0.5 = 0.25. \quad (2.180)$$

Distribusi bersama sepenuhnya ditentukan oleh hasil kali distribusi marginal.

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

$$\Pr(A \cap B | C) = \Pr(A | C) \cdot \Pr(B | C). \quad (2.181)$$

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

$$\Pr(X \in A, Y \in B | Z = z) = \Pr(X \in A | Z = z) \cdot \Pr(Y \in B | Z = z). \quad (2.182)$$

2.2.6. Teorema Bayes dan Hukum Probabilitas Total

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

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

$$\Pr(A) = \sum_{i=1}^n \Pr(A | B_i) \cdot \Pr(B_i). \quad (2.183)$$

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

$$\Pr(A) = \int_{-\infty}^{\infty} \Pr(A | Y = y) \cdot f_Y(y) dy. \quad (2.184)$$

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

$$A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^n B_i \right) = \bigcup_{i=1}^n (A \cap B_i). \quad (2.185)$$

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

$$\Pr(A) = \sum_{i=1}^n \Pr(A \cap B_i) = \sum_{i=1}^n \Pr(A | B_i) \cdot \Pr(B_i). \quad (2.186)$$

■

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

$$\begin{aligned} \Pr(\text{Cacat}) &= \Pr(\text{Cacat} | \text{Mesin 1}) \cdot \Pr(\text{Mesin 1}) \\ &\quad + \Pr(\text{Cacat} | \text{Mesin 2}) \cdot \Pr(\text{Mesin 2}) \\ &\quad + \Pr(\text{Cacat} | \text{Mesin 3}) \cdot \Pr(\text{Mesin 3}) \\ &= 0.02 \cdot 0.50 + 0.03 \cdot 0.30 + 0.05 \cdot 0.20 \\ &= 0.01 + 0.009 + 0.01 = 0.029. \end{aligned} \quad (2.187)$$

Jadi, probabilitas suatu komponen adalah cacat adalah 2.9%.

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

$$\Pr(B_j | A) = \frac{\Pr(A | B_j) \cdot \Pr(B_j)}{\sum_{i=1}^n \Pr(A | B_i) \cdot \Pr(B_i)}. \quad (2.188)$$

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

$$\Pr(B | A) = \frac{\Pr(A | B) \cdot \Pr(B)}{\Pr(A)}. \quad (2.189)$$

Bukti. Berdasarkan definisi probabilitas bersyarat:

$$\Pr(B_j | A) = \frac{\Pr(A \cap B_j)}{\Pr(A)} = \frac{\Pr(A | B_j) \cdot \Pr(B_j)}{\Pr(A)}. \quad (2.190)$$

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

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

$$\begin{aligned} \Pr(\text{Mesin 3} | \text{Cacat}) &= \frac{\Pr(\text{Cacat} | \text{Mesin 3}) \cdot \Pr(\text{Mesin 3})}{\Pr(\text{Cacat})} \\ &= \frac{0.05 \cdot 0.20}{0.029} = \frac{0.01}{0.029} \approx 0.345. \end{aligned} \quad (2.191)$$

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

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

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

- (b) $\Pr(A | B)$ 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: Posterior \propto Likelihood \times Prior.

2.2.7. Hukum Ekspektasi Total dan Variansi Total

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

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X | Y]]. \quad (2.192)$$

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

$$\mathbb{E}[X] = \sum_j \mathbb{E}[X | Y = y_j] \cdot \Pr(Y = y_j). \quad (2.193)$$

Untuk kasus kontinu dengan Y memiliki kepadatan f_Y :

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X | Y = y] \cdot f_Y(y) dy. \quad (2.194)$$

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

$$\begin{aligned}
 \mathbb{E}[\mathbb{E}[X | Y]] &= \sum_j \mathbb{E}[X | Y = y_j] \cdot \Pr(Y = y_j) \\
 &= \sum_j \left(\sum_i x_i \cdot \Pr(X = x_i | Y = y_j) \right) \cdot \Pr(Y = y_j) \\
 &= \sum_i \sum_j x_i \cdot \Pr(X = x_i | Y = y_j) \cdot \Pr(Y = y_j) \\
 &= \sum_i \sum_j x_i \cdot \Pr(X = x_i, Y = y_j) \\
 &= \sum_i x_i \cdot \Pr(X = x_i) = \mathbb{E}[X]. \tag{2.195}
 \end{aligned}$$

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:

$$\begin{aligned}
 \mathbb{E}[\mathbb{E}[X | Y]] &= \int_{-\infty}^{\infty} \mathbb{E}[X | Y = y] \cdot f_Y(y) dy \\
 &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x \cdot f_{X|Y}(x | y) dx \right) \cdot f_Y(y) dy \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x | y) \cdot f_Y(y) dx dy. \tag{2.196}
 \end{aligned}$$

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

$$\begin{aligned}
 \mathbb{E}[\mathbb{E}[X | Y]] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f_{X,Y}(x, y) dx dy \\
 &= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \right) dx \\
 &= \int_{-\infty}^{\infty} x \cdot f_X(x) dx = \mathbb{E}[X], \tag{2.197}
 \end{aligned}$$

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

$$\begin{aligned}\mathbb{E}[S] &= \mathbb{E}[\mathbb{E}[S | N]] = \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^N X_i | N\right]\right] \\ &= \mathbb{E}[N \cdot \mu] = \mu \cdot \mathbb{E}[N] = \mu\lambda.\end{aligned}\quad (2.198)$$

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

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X | Y)] + \text{Var}(\mathbb{E}[X | Y]). \quad (2.199)$$

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

$$\text{Var}(X | Y) = \mathbb{E}[X^2 | Y] - (\mathbb{E}[X | Y])^2. \quad (2.200)$$

Dengan mengambil ekspektasi kedua ruas:

$$\mathbb{E}[\text{Var}(X | Y)] = \mathbb{E}[\mathbb{E}[X^2 | Y]] - \mathbb{E}[(\mathbb{E}[X | Y])^2] = \mathbb{E}[X^2] - \mathbb{E}[(\mathbb{E}[X | Y])^2]. \quad (2.201)$$

Di sisi lain:

$$\begin{aligned}\text{Var}(\mathbb{E}[X | Y]) &= \mathbb{E}[(\mathbb{E}[X | Y])^2] - (\mathbb{E}[\mathbb{E}[X | Y]])^2 \\ &= \mathbb{E}[(\mathbb{E}[X | Y])^2] - (\mathbb{E}[X])^2.\end{aligned}\quad (2.202)$$

Dengan menjumlahkan kedua hasil:

$$\begin{aligned}\mathbb{E}[\text{Var}(X | Y)] + \text{Var}(\mathbb{E}[X | Y]) &= \mathbb{E}[X^2] - \mathbb{E}[(\mathbb{E}[X | Y])^2] \\ &\quad + \mathbb{E}[(\mathbb{E}[X | Y])^2] - (\mathbb{E}[X])^2 \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \text{Var}(X).\end{aligned}\quad (2.203)$$

■

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

$$\text{Var}(S | N) = N \cdot \sigma^2, \quad (2.204)$$

$$\mathbb{E}[S | N] = N \cdot \mu. \quad (2.205)$$

Dengan hukum variansi total:

$$\begin{aligned}\text{Var}(S) &= \mathbb{E}[\text{Var}(S | N)] + \text{Var}(\mathbb{E}[S | N]) \\ &= \mathbb{E}[N \cdot \sigma^2] + \text{Var}(N \cdot \mu) \\ &= \sigma^2 \cdot \mathbb{E}[N] + \mu^2 \cdot \text{Var}(N) \\ &= \sigma^2 \lambda + \mu^2 \lambda = \lambda(\sigma^2 + \mu^2).\end{aligned}\quad (2.206)$$

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

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

2.3. Konvergensi dan Laju Pertumbuhan

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

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

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}. \quad (2.207)$$

Bukti. Dengan definisi ekspektasi:

$$\mathbb{E}[X] = \int_0^\infty \Pr(X \geq t) dt. \quad (2.208)$$

Karena X non-negatif, maka

$$\mathbb{E}[X] \geq \int_a^\infty \Pr(X \geq t) dt \geq \int_a^\infty \Pr(X \geq a) dt = \Pr(X \geq a) \cdot (\infty - a). \quad (2.209)$$

Oleh karena itu,

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}. \quad (2.210)$$

■

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

$$\Pr(|X - \mathbb{E}[X]| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}. \quad (2.211)$$

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

$$\Pr(|X - \mathbb{E}[X]| \geq \varepsilon) = \Pr(Y \geq \varepsilon^2) \leq \frac{\mathbb{E}[Y]}{\varepsilon^2} = \frac{\text{Var}(X)}{\varepsilon^2}. \quad (2.212)$$

■

2.3.1. Barisan Variabel Acak

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

2.3.2. Konvergensi Barisan Variabel Acak

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

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

$$\Pr(\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1. \quad (2.213)$$

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

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

sehingga $X_n \rightarrow X$ hampir pasti.

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

$$\lim_{n \rightarrow \infty} \Pr(|X_n - X| > \varepsilon) = 0. \quad (2.214)$$

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

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 ω terdapat tak hingga banyak n sedemikian sehingga $X_n(\omega) = n$ (dengan probabilitas positif).

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

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

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)] \quad (2.215)$$

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

$$\lim_{n \rightarrow \infty} \mathbb{E}[(X_n - X)^2] = 0. \quad (2.216)$$

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, $X_n \xrightarrow{L^2} 0$.

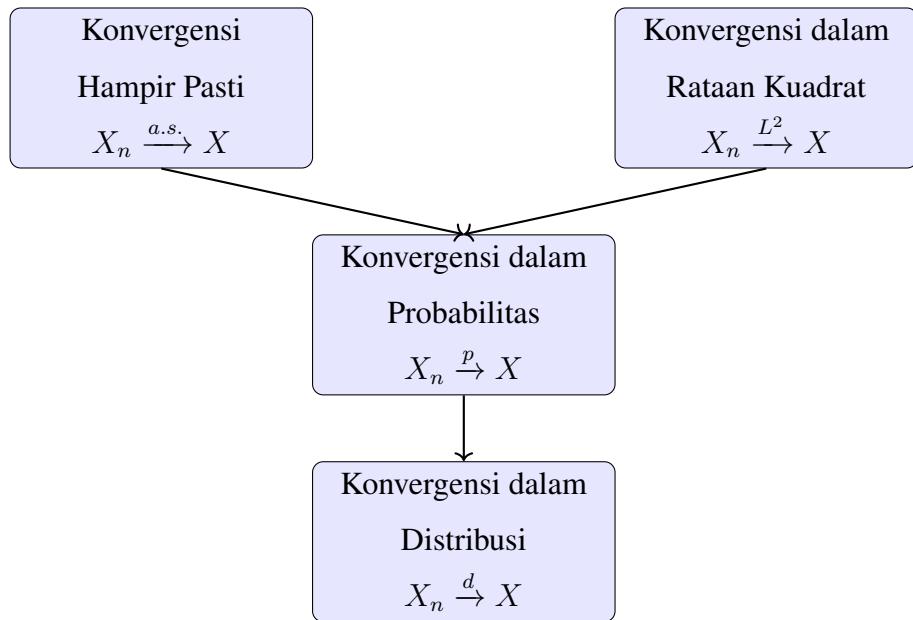
Contoh 2.3.11 Misalkan $X_n \sim \mathcal{N}(0, 1/n)$ untuk $n = 1, 2, \dots$. Akan ditunjukkan bahwa $X_n \xrightarrow{L^2} 0$. Perhatikan bahwa

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

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

$$A = \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\}. \quad (2.217)$$

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,

$$A \subseteq \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} B_n(\varepsilon)^c. \quad (2.218)$$

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

$$\Pr \left(\limsup_{n \rightarrow \infty} B_n(\varepsilon) \right) \leq \Pr(A^c) = 0. \quad (2.219)$$

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

$$\lim_{n \rightarrow \infty} \Pr(|X_n - X| > \varepsilon) = 0, \quad (2.220)$$

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

$$\Pr(|X_n - X| > \varepsilon) = \Pr((X_n - X)^2 > \varepsilon^2) \leq \frac{\mathbb{E}[(X_n - X)^2]}{\varepsilon^2}. \quad (2.221)$$

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

$$\lim_{n \rightarrow \infty} \Pr(|X_n - X| > \varepsilon) \leq \lim_{n \rightarrow \infty} \frac{\mathbb{E}[(X_n - X)^2]}{\varepsilon^2} = 0. \quad (2.222)$$

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

$$\begin{aligned} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| &= |\mathbb{E}[f(X_n) - f(X)]| \\ &\leq \mathbb{E}[|f(X_n) - f(X)|]. \end{aligned} \quad (2.223)$$

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

$$\begin{aligned} \mathbb{E}[|f(X_n) - f(X)|] &= \mathbb{E}[|f(X_n) - f(X)| \cdot \mathbf{1}_{\{|X_n - X| \leq \eta\}}] \\ &\quad + \mathbb{E}[|f(X_n) - f(X)| \cdot \mathbf{1}_{\{|X_n - X| > \eta\}}]. \end{aligned} \quad (2.224)$$

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

$$\mathbb{E}[|f(X_n) - f(X)| \cdot \mathbf{1}_{\{|X_n - X| \leq \eta\}}] < \delta. \quad (2.225)$$

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

$$\mathbb{E}[|f(X_n) - f(X)| \cdot \mathbf{1}_{\{|X_n - X| > \eta\}}] \leq 2M \cdot \Pr(|X_n - X| > \eta). \quad (2.226)$$

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

$$\limsup_{n \rightarrow \infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq \delta. \quad (2.227)$$

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

$$f_\varepsilon(x) = \begin{cases} 1, & \text{jika } |x - c| \geq \varepsilon, \\ \frac{|x - c|}{\varepsilon}, & \text{jika } |x - c| < \varepsilon. \end{cases} \quad (2.228)$$

Fungsi f_ε kontinu dan memenuhi $0 \leq f_\varepsilon(x) \leq 1$ untuk semua x , serta $f_\varepsilon(c) = 0$.

Perhatikan bahwa $\mathbf{1}_{\{|X_n - c| \geq \varepsilon\}} \leq f_\varepsilon(X_n)$, sehingga

$$\Pr(|X_n - c| \geq \varepsilon) = \mathbb{E}[\mathbf{1}_{\{|X_n - c| \geq \varepsilon\}}] \leq \mathbb{E}[f_\varepsilon(X_n)]. \quad (2.229)$$

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

$$\lim_{n \rightarrow \infty} \Pr(|X_n - c| \geq \varepsilon) = 0, \quad (2.230)$$

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

Contoh 2.3.16 Misalkan $\Omega = [0, 1]$ dengan ukuran probabilitas Lebesgue. Definisikan barisan variabel acak sebagai berikut: untuk setiap $n \geq 1$, tulis $n = 2^k + j$ dengan $0 \leq j < 2^k$, dan definisikan

$$X_n(\omega) = \mathbf{1}_{[j/2^k, (j+1)/2^k]}(\omega).$$

Barisan ini disebut barisan *typewriter*. Untuk setiap $\varepsilon \in (0, 1)$,

$$\Pr(|X_n| > \varepsilon) = \Pr(X_n = 1) = \frac{1}{2^k} \rightarrow 0$$

ketika $n \rightarrow \infty$ (sebab $k \rightarrow \infty$). Jadi, $X_n \xrightarrow{p} 0$.

Namun, untuk setiap $\omega \in [0, 1]$, terdapat tak hingga banyak n sehingga $X_n(\omega) = 1$ (karena setiap titik ω akan tercakup oleh interval $[j/2^k, (j+1)/2^k]$ untuk tak hingga banyak nilai k). Oleh karena itu, $\limsup_{n \rightarrow \infty} X_n(\omega) = 1 \neq 0$ untuk semua ω , sehingga X_n tidak konvergen hampir pasti ke 0.

Contoh 2.3.17 Misalkan $X \sim \mathcal{N}(0, 1)$ dan definisikan $X_n = -X$ untuk semua n . Maka $X_n \sim \mathcal{N}(0, 1) = X$ untuk semua n , sehingga $X_n \xrightarrow{d} X$ secara trivial.

Namun, untuk $\varepsilon = 1$,

$$\Pr(|X_n - X| > 1) = \Pr(|-X - X| > 1) = \Pr(|2X| > 1) = \Pr(|X| > 0.5) \approx 0.617 \neq 0.$$

Oleh karena itu, X_n tidak konvergen dalam probabilitas ke X .

2.4. Teori Probabilitas Asimtotik

Teori probabilitas asimtotik mempelajari perilaku limit dari barisan variabel random dan estimator ketika ukuran sampel menuju tak hingga. Dalam statistika, hasil-hasil asimtotik digunakan sebagai pendekatan terhadap distribusi *sampling* yang umumnya sulit diperoleh secara eksak. Pembahasan dalam bagian ini mengikuti kerangka statistik asimtotik sebagaimana dikembangkan oleh ? dengan fondasi probabilitas dari ?.

2.4.1. Hukum Bilangan Besar

Teorema 2.4.1 (Weak Law of Large Numbers, (?)) Misalkan $\{X_i\}_{i=1}^{\infty}$ adalah barisan variabel acak i.i.d. dengan $\mathbb{E}[X_1] = \mu$ dan $\mathbb{E}[|X_1|] < \infty$. Maka

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} \mu. \quad (2.231)$$

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

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n} \cdot n\mu = \mu. \quad (2.232)$$

Selanjutnya, karena X_i saling independen, berlaku

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}. \quad (2.233)$$

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

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}. \quad (2.234)$$

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

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \lim_{n \rightarrow \infty} \frac{\sigma^2}{n\varepsilon^2} = 0. \quad (2.235)$$

Oleh karena probabilitas selalu non-negatif, berlaku

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0, \quad (2.236)$$

yang ekuivalen dengan $\bar{X}_n \xrightarrow{p} \mu$. Dengan demikian, *Weak Law of Large Numbers* terbukti. ■

Hukum Bilangan Besar memberikan dasar probabilistik bagi konsistensi estimator berbasis rata-rata sampel.

2.4.2. Konsistensi Estimator

Definisi 2.4.2 (Konsistensi, (?)) Suatu estimator $\hat{\theta}_n$ untuk parameter θ dikatakan konsisten apabila

$$\hat{\theta}_n \xrightarrow{p} \theta. \quad (2.237)$$

Konsistensi sering diperoleh sebagai konsekuensi langsung dari Hukum Bilangan

Besar, khususnya untuk estimator yang dapat diekspresikan sebagai fungsi rata-rata sampel.

2.4.3. Fungsi Karakteristik

Fungsi karakteristik merupakan alat fundamental dalam teori probabilitas yang memungkinkan karakterisasi lengkap distribusi suatu variabel acak. Berbeda dengan fungsi pembangkit momen yang mungkin tidak terdefinisi, fungsi karakteristik selalu ada untuk setiap variabel acak.

Definisi 2.4.3 (Fungsi Karakteristik, (?)) Misalkan X adalah variabel acak. Fungsi karakteristik dari X didefinisikan sebagai

$$\phi_X(t) = \mathbb{E}[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} dF_X(x), \quad t \in \mathbb{R}, \quad (2.238)$$

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

$$\phi_X(t) = \exp\left(i\mu t - \frac{\sigma^2 t^2}{2}\right). \quad (2.239)$$

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

$$\phi_X(t) = \exp(\lambda(e^{it} - 1)). \quad (2.240)$$

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

$$\phi_X(t) = \frac{\lambda}{\lambda - it}. \quad (2.241)$$

Proposisi 2.4.5 (Sifat-Sifat Fungsi Karakteristik, (?)) *Fungsi karakteristik $\phi_X(t)$ memenuhi sifat-sifat berikut.*

- (a) $\phi_X(0) = 1$.
- (b) $|\phi_X(t)| \leq 1$ untuk semua $t \in \mathbb{R}$.
- (c) ϕ_X kontinu seragam pada \mathbb{R} .
- (d) $\overline{\phi_X(z)} = \phi_X(\bar{z})$, 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 (Hubungan dengan Momen, (?)) *Jika $\mathbb{E}[|X|^n] < \infty$, maka fungsi karakteristik ϕ_X terdiferensiasi n kali dan*

$$\phi_X^{(k)}(0) = i^k \mathbb{E}[X^k], \quad k = 1, 2, \dots, n. \quad (2.242)$$

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

$$\phi_X(t) = 1 + it\mathbb{E}[X] - \frac{t^2}{2}\mathbb{E}[X^2] + o(t^2). \quad (2.243)$$

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

Teorema 2.4.7 (Teorema Kontinuitas Lévy, (?)) *Misalkan $\{X_n\}_{n=1}^\infty$ adalah barisan variabel acak dengan fungsi karakteristik ϕ_{X_n} , dan misalkan X adalah variabel acak dengan fungsi karakteristik ϕ_X . Maka berlaku dua pernyataan berikut.*

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

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

$$\phi_{X_n}(t) = \exp\left(-\frac{(1+1/n)t^2}{2}\right). \quad (2.244)$$

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

$$\lim_{n \rightarrow \infty} \phi_{X_n}(t) = \exp\left(-\frac{t^2}{2}\right) = \phi_X(t), \quad (2.245)$$

dengan $\phi_X(t)$ adalah fungsi karakteristik dari $X \sim \mathcal{N}(0, 1)$. Sebab limit ini kontinu di $t = 0$, berdasarkan Teorema Lévy, diperoleh $X_n \Rightarrow X \sim \mathcal{N}(0, 1)$.

Teorema Lévy sangat penting dalam pembuktian Teorema Limit Pusat, di mana konvergensi dalam distribusi dibuktikan melalui konvergensi fungsi karakteristik ke fungsi karakteristik distribusi normal.

2.4.4. Teorema Pemetaan Kontinu dan Teorema Slutsky

Teorema pemetaan kontinu (*Continuous Mapping Theorem*) dan teorema Slutsky merupakan alat fundamental dalam statistika asimtotik yang memungkinkan manipulasi limit variabel acak melalui fungsi kontinu dan operasi aritmetika.

Teorema 2.4.9 (Teorema Pemetaan Kontinu (*Continuous Mapping Theorem*), (?) Misalkan $\{X_n\}$ adalah barisan variabel acak dan $g : \mathbb{R}^k \rightarrow \mathbb{R}^m$ 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

$$\mathbb{E}[h(g(X_n))] = \mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)] = \mathbb{E}[h(g(X))]. \quad (2.246)$$

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 ω , 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

$$X_n^2 \xrightarrow{d} X^2 \sim \chi^2(1). \quad (2.247)$$

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

Teorema 2.4.11 (Teorema Slutsky, (?)) Misalkan $\{X_n\}$ dan $\{Y_n\}$ adalah barisan variabel acak. Jika $X_n \xrightarrow{d} X$ dan $Y_n \xrightarrow{p} c$ dengan c adalah konstanta, maka berlaku:

- (a) $X_n + Y_n \xrightarrow{d} X + c$,
- (b) $X_n \cdot Y_n \xrightarrow{d} c \cdot X$, dan
- (c) $X_n/Y_n \xrightarrow{d} X/c$ jika $c \neq 0$.

Bukti. Akan dibuktikan ketiga bagian secara berurutan.

Bagian (a): Akan ditunjukkan bahwa $(X_n, Y_n) \xrightarrow{d} (X, c)$, kemudian menerapkan teorema pemetaan kontinu dengan $g(x, y) = x + y$.

Untuk menunjukkan $(X_n, Y_n) \xrightarrow{d} (X, c)$, digunakan fungsi karakteristik. Fungsi karakteristik bersama dari (X_n, Y_n) adalah

$$\phi_{(X_n, Y_n)}(t_1, t_2) = \mathbb{E}[e^{i(t_1 X_n + t_2 Y_n)}]. \quad (2.248)$$

Perhatikan bahwa

$$\begin{aligned} |\phi_{(X_n, Y_n)}(t_1, t_2) - \phi_X(t_1)e^{it_2 c}| &= |\mathbb{E}[e^{it_1 X_n}(e^{it_2 Y_n} - e^{it_2 c})]| \\ &\leq \mathbb{E}[|e^{it_2 Y_n} - e^{it_2 c}|]. \end{aligned} \quad (2.249)$$

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

$$\phi_{(X_n, Y_n)}(t_1, t_2) \rightarrow \phi_X(t_1)e^{it_2 c} = \phi_{(X, c)}(t_1, t_2). \quad (2.250)$$

Berdasarkan teorema Lévy, $(X_n, Y_n) \xrightarrow{d} (X, c)$.

Dengan teorema pemetaan kontinu dan $g(x, y) = x + y$ yang kontinu, dipe-

oleh $X_n + Y_n \xrightarrow{d} X + c$.

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 rataan μ dan variansi σ^2 , dan S_n^2 adalah variansi sampel. Berdasarkan Teorema Limit Pusat:

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2). \quad (2.251)$$

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

Dengan teorema Slutsky:

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \cdot \frac{\sigma}{S_n} \xrightarrow{d} \mathcal{N}(0, 1) \cdot 1 = \mathcal{N}(0, 1). \quad (2.252)$$

Hasil ini membenarkan penggunaan statistik t dalam inferensi asimtotik.

Teorema 2.4.13 (Teorema Slutsky untuk Vektor dan Matriks, (?) 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:

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

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

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

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

Bukti. Bagian (a): Perkalian matriks $g(\mathbf{A}, \mathbf{x}) = \mathbf{Ax}$ 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

$$\mathbb{E}[h(\mathbf{A}_n, \mathbf{X}_n)] = \mathbb{E}[h(\mathbf{A}_n, \mathbf{X}_n) \mathbf{1}_{\{\|\mathbf{A}_n - \mathbf{A}\| \leq \varepsilon\}}] + \mathbb{E}[h(\mathbf{A}_n, \mathbf{X}_n) \mathbf{1}_{\{\|\mathbf{A}_n - \mathbf{A}\| > \varepsilon\}}]. \quad (2.253)$$

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 ε 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 $g(\mathbf{A}, \mathbf{x}) = \mathbf{Ax}$, diperoleh $\mathbf{A}_n \mathbf{X}_n \xrightarrow{d} \mathbf{AX}$.

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

Bagian (c): Bentuk kuadratik $g(\mathbf{A}, \mathbf{x}) = \mathbf{x}^\top \mathbf{Ax}$ 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 $\mathbf{X}_n^\top \mathbf{A}_n \mathbf{X}_n \xrightarrow{d} \mathbf{X}^\top \mathbf{AX}$.

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 $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma)$ dan $\hat{\Sigma}_n \xrightarrow{p} \Sigma$ dengan Σ definit positif. Maka statistik Wald

$$W_n = n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})^\top \hat{\Sigma}_n^{-1}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \quad (2.254)$$

memenuhi $W_n \xrightarrow{d} \chi^2(p)$, dengan $p = \dim(\boldsymbol{\beta})$.

Hal ini karena $\hat{\Sigma}_n^{-1} \xrightarrow{p} \Sigma^{-1}$ (berdasarkan teorema pemetaan kontinu untuk

fungsi invers matriks), dan

$$\sqrt{n}(\hat{\beta}_n - \boldsymbol{\beta})^\top \hat{\Sigma}_n^{-1} \sqrt{n}(\hat{\beta}_n - \boldsymbol{\beta}) \xrightarrow{d} \mathbf{Z}^\top \Sigma^{-1} \mathbf{Z} \sim \chi^2(p), \quad (2.255)$$

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

Teorema 2.4.15 (Teorema Cramér–Wold, (?)) 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

$$\mathbf{t}^\top \mathbf{X}_n \xrightarrow{d} \mathbf{t}^\top \mathbf{X}. \quad (2.256)$$

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

$$\phi_{\mathbf{X}_n}(\mathbf{t}) = \mathbb{E}[e^{i\mathbf{t}^\top \mathbf{X}_n}]. \quad (2.257)$$

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,

$$\phi_{\mathbf{X}_n}(\mathbf{t}) = \mathbb{E}[e^{i \cdot 1 \cdot (\mathbf{t}^\top \mathbf{X}_n)}] \rightarrow \mathbb{E}[e^{i \cdot 1 \cdot (\mathbf{t}^\top \mathbf{X})}] = \phi_{\mathbf{X}}(\mathbf{t}). \quad (2.258)$$

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 rataan

μ dan matriks kovarians Σ , maka untuk setiap $t \in \mathbb{R}^p$:

$$t^\top (\sqrt{n}(\bar{X}_n - \mu)) = \sqrt{n}(t^\top \bar{X}_n - t^\top \mu) \xrightarrow{d} \mathcal{N}(0, t^\top \Sigma t), \quad (2.259)$$

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

2.4.5. Notasi Big- \mathcal{O} dan Little- o (Deterministik)

Notasi asimtotik Big- \mathcal{O} dan Little- o merupakan alat fundamental dalam analisis matematika untuk menyatakan laju pertumbuhan dan tingkat ketakterabaian suatu besaran deterministik.

Definisi 2.4.17 (Notasi Big- \mathcal{O}) Misalkan $\{a_n\}$ dan $\{b_n\}$ adalah barisan bilangan real dengan $b_n > 0$ untuk semua n cukup besar. Barisan $\{a_n\}$ dikatakan berorde paling banyak b_n , ditulis

$$a_n = O(b_n), \quad (2.260)$$

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

$$|a_n| \leq Cb_n, \quad \forall n \geq n_0. \quad (2.261)$$

Secara ekuivalen, $a_n = O(b_n)$ jika dan hanya jika barisan $\{a_n/b_n\}$ terbatas untuk n cukup besar.

Contoh 2.4.18 Berikut adalah beberapa contoh penggunaan notasi Big- \mathcal{O} .

(a) **Polinomial:** Jika $a_n = 3n^2 + 5n + 7$, maka $a_n = O(n^2)$. Untuk $n \geq 1$, 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 $a_n = O(n^\varepsilon)$ 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 $a_n = O(3^n)$ 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 σ^2 , maka galat standar $\sigma/\sqrt{n} = O(n^{-1/2})$.

Definisi 2.4.19 (Notasi Little-o) Misalkan $\{a_n\}$ dan $\{b_n\}$ adalah barisan bilangan real dengan $b_n > 0$ untuk semua n cukup besar. Barisan $\{a_n\}$ dikatakan berorde lebih kecil dari b_n , ditulis

$$a_n = o(b_n), \quad (2.262)$$

jika

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0. \quad (2.263)$$

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) **Ekspansi 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 (Hubungan antara O dan o) Untuk barisan $\{a_n\}$ dan $\{b_n\}$ dengan $b_n > 0$, berlaku:

- (a) Jika $a_n = o(b_n)$, maka $a_n = O(b_n)$, tetapi tidak sebaliknya.
- (b) $a_n = o(b_n)$ jika dan hanya jika $a_n = O(b_n)$ dan $\limsup_{n \rightarrow \infty} |a_n/b_n| = 0$.

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 $a_n = O(b_n)$ dengan $C = 1$.

Sebaliknya tidak berlaku: $a_n = 1$ dan $b_n = 1$ memberikan $a_n = O(b_n)$ (dengan $C = 1$), tetapi $\lim_{n \rightarrow \infty} a_n/b_n = 1 \neq 0$, sehingga $a_n \neq o(b_n)$.

- (b) Langsung dari definisi: $a_n = o(b_n)$ ekivalen dengan $\lim_{n \rightarrow \infty} |a_n/b_n| = 0$, yang mengimplikasikan barisan $\{|a_n/b_n|\}$ terbatas (sehingga $a_n = O(b_n)$) dan $\limsup_{n \rightarrow \infty} |a_n/b_n| = 0$.

■

Proposisi 2.4.22 (Sifat-Sifat Notasi Asimtotik) Untuk barisan $\{a_n\}$, $\{b_n\}$, $\{c_n\}$ dengan suku positif yang sesuai, berlaku:

- (a) **Transitivitas:** Jika $a_n = O(b_n)$ dan $b_n = O(c_n)$, maka $a_n = O(c_n)$.
- (b) **Penjumlahan:** $O(a_n) + O(b_n) = O(\max\{a_n, b_n\})$.
- (c) **Perkalian:** $O(a_n) \cdot O(b_n) = O(a_n b_n)$.
- (d) **Penyerapan:** $o(a_n) + O(a_n) = O(a_n)$ dan $o(a_n) \cdot O(b_n) = o(a_n b_n)$.

2.4.6. Notasi Big- \mathcal{O} dan Little- o dalam Probabilitas

Notasi \mathcal{O}_p dan o_p merupakan perluasan dari notasi deterministik $O(\cdot)$ dan $o(\cdot)$ ke dalam konteks stokastik. Notasi ini sangat penting dalam perumusan representasi asimtotik estimator dan digunakan secara luas dalam statistika asimtotik (?).

Definisi 2.4.23 (Notasi \mathcal{O}_p (Terbatas dalam Probabilitas), (?)) *Barisan variabel acak $\{X_n\}$ dikatakan terbatas dalam probabilitas dengan orde a_n , ditulis*

$$X_n = \mathcal{O}_p(a_n), \quad (2.264)$$

apabila untuk setiap $\varepsilon > 0$ terdapat $M > 0$ dan $N \in \mathbb{N}$ sehingga

$$\Pr(|X_n| > Ma_n) < \varepsilon, \quad \forall n \geq N. \quad (2.265)$$

Secara ekuivalen, $X_n = \mathcal{O}_p(a_n)$ jika dan hanya jika barisan $\{X_n/a_n\}$ terbatas dalam probabilitas, yaitu

$$\sup_{n \geq N} \Pr \left(\left| \frac{X_n}{a_n} \right| > M \right) < \varepsilon. \quad (2.266)$$

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 (Perhitungan Numerik) 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 (Konvergen ke Nol dalam Probabilitas), (?)) Barisan variabel acak $\{X_n\}$ dikatakan konvergen ke nol dalam probabilitas dengan orde a_n , ditulis

$$X_n = o_p(a_n), \quad (2.267)$$

apabila

$$\frac{X_n}{a_n} \xrightarrow{p} 0, \quad (2.268)$$

yaitu untuk setiap $\varepsilon > 0$ dan $\delta > 0$, terdapat $N \in \mathbb{N}$ sehingga

$$\Pr\left(\left|\frac{X_n}{a_n}\right| > \varepsilon\right) < \delta, \quad \forall n \geq N. \quad (2.269)$$

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 $\bar{X}_n - \mu = o_p(1)$.

Contoh 2.4.28 (Perhitungan Numerik) Misalkan $X_n \sim \mathcal{N}(0, 1/n^2)$ untuk $n = 1, 2, \dots$. Akan ditunjukkan bahwa $X_n = o_p(n^{-1/2})$.

Perhatikan bahwa

$$\frac{X_n}{n^{-1/2}} = n^{1/2} X_n \sim \mathcal{N}\left(0, \frac{n}{n^2}\right) = \mathcal{N}\left(0, \frac{1}{n}\right).$$

Untuk setiap $\varepsilon > 0$, dengan ketaksamaan Chebyshev diperoleh

$$\Pr(|n^{1/2} X_n| > \varepsilon) \leq \frac{\text{Var}(n^{1/2} X_n)}{\varepsilon^2} = \frac{1/n}{\varepsilon^2} = \frac{1}{n\varepsilon^2} \rightarrow 0$$

ketika $n \rightarrow \infty$. Dengan demikian, $n^{1/2} X_n \xrightarrow{p} 0$, sehingga $X_n = o_p(n^{-1/2})$.

Sebagai contoh numerik, untuk $n = 100$ dan $\varepsilon = 0.1$:

$$\Pr(|n^{1/2} X_n| > 0.1) \leq \frac{1}{100 \cdot 0.01} = 1.$$

Batas ini tidak informatif, tetapi dengan menggunakan sifat distribusi normal:

$$\Pr(|n^{1/2} X_n| > 0.1) = 2 \left(1 - \Phi \left(\frac{0.1}{\sqrt{1/100}} \right) \right) = 2(1 - \Phi(1)) \approx 0.317.$$

Untuk $n = 10000$:

$$\Pr(|n^{1/2} X_n| > 0.1) = 2 \left(1 - \Phi \left(\frac{0.1}{\sqrt{1/10000}} \right) \right) = 2(1 - \Phi(10)) \approx 0.$$

Teorema 2.4.29 (Sifat-Sifat Notasi \mathcal{O}_p dan o_p , (?)) Misalkan $\{X_n\}$ dan $\{Y_n\}$ adalah barisan variabel acak, serta $\{a_n\}$ dan $\{b_n\}$ adalah barisan bilangan positif. Berlaku sifat-sifat berikut.

- (a) **Penjumlahan:** $\mathcal{O}_p(a_n) + \mathcal{O}_p(b_n) = \mathcal{O}_p(\max\{a_n, b_n\})$.
- (b) **Perkalian:** $\mathcal{O}_p(a_n) \cdot \mathcal{O}_p(b_n) = \mathcal{O}_p(a_n b_n)$.
- (c) **Relasi o_p dan \mathcal{O}_p :** $o_p(a_n) = \mathcal{O}_p(a_n)$, tetapi tidak sebaliknya.
- (d) **Penyerapan:** $o_p(a_n) + \mathcal{O}_p(a_n) = \mathcal{O}_p(a_n)$.
- (e) **Perkalian dengan o_p :** $o_p(a_n) \cdot \mathcal{O}_p(b_n) = o_p(a_n b_n)$.
- (f) **Konvergensi dalam probabilitas:** $X_n \xrightarrow{p} c$ jika dan hanya jika $X_n = c + o_p(1)$.
- (g) **Konvergensi dalam distribusi:** Jika $X_n \Rightarrow X$ dengan X variabel acak, maka $X_n = \mathcal{O}_p(1)$.

Bukti. Berikut adalah bukti untuk beberapa sifat utama.

- (a) **Penjumlahan:** Misalkan $X_n = \mathcal{O}_p(a_n)$ dan $Y_n = \mathcal{O}_p(b_n)$. Untuk setiap $\varepsilon > 0$, terdapat $M_1, M_2 > 0$ sehingga

$$\Pr(|X_n| > M_1 a_n) < \varepsilon/2, \quad \Pr(|Y_n| > M_2 b_n) < \varepsilon/2.$$

Dengan ketaksamaan segitiga, $|X_n + Y_n| \leq |X_n| + |Y_n|$. Definisikan $c_n = \max\{a_n, b_n\}$ dan $M = M_1 + M_2$. Maka

$$\begin{aligned} \Pr(|X_n + Y_n| > Mc_n) &\leq \Pr(|X_n| + |Y_n| > (M_1 + M_2)c_n) \\ &\leq \Pr(|X_n| > M_1 a_n) + \Pr(|Y_n| > M_2 b_n) \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned} \tag{2.270}$$

- (b) **Perkalian:** Misalkan $X_n = \mathcal{O}_p(a_n)$ dan $Y_n = \mathcal{O}_p(b_n)$. Untuk setiap $\varepsilon > 0$,

terdapat $M_1, M_2 > 0$ sehingga

$$\Pr(|X_n| > M_1 a_n) < \varepsilon/2, \quad \Pr(|Y_n| > M_2 b_n) < \varepsilon/2.$$

Perhatikan bahwa

$$\begin{aligned} \Pr(|X_n Y_n| > M_1 M_2 a_n b_n) &\leq \Pr(|X_n| > M_1 a_n) + \Pr(|Y_n| > M_2 b_n) \\ &< \varepsilon. \end{aligned} \tag{2.271}$$

- (c) **Relasi o_p dan \mathcal{O}_p :** Jika $X_n = o_p(a_n)$, maka $X_n/a_n \xrightarrow{p} 0$. Untuk setiap $\varepsilon > 0$ dan $M > 0$, terdapat N sehingga untuk $n \geq N$:

$$\Pr(|X_n/a_n| > M) < \varepsilon,$$

yang berarti $X_n = \mathcal{O}_p(a_n)$. Sebaliknya tidak berlaku: $X_n \sim \mathcal{N}(0, 1)$ memenuhi $X_n = \mathcal{O}_p(1)$ tetapi bukan $o_p(1)$.

- (f) **Konvergensi dalam probabilitas:** Jika $X_n \xrightarrow{p} c$, maka $X_n - c \xrightarrow{p} 0$, yang ekuivalen dengan $(X_n - c)/1 \xrightarrow{p} 0$, sehingga $X_n - c = o_p(1)$, atau $X_n = c + o_p(1)$. Sebaliknya juga berlaku.
- (g) **Konvergensi dalam distribusi:** Jika $X_n \Rightarrow X$, maka untuk setiap $\varepsilon > 0$, terdapat M sehingga $\Pr(|X| > M) < \varepsilon/2$. Untuk n cukup besar, $\Pr(|X_n| > M) < \varepsilon$ karena konvergensi dalam distribusi. Dengan demikian, $X_n = \mathcal{O}_p(1)$.

■

Teorema 2.4.30 (Dekomposisi \mathcal{O}_p Berdasarkan Momen, (?) Misalkan $\{X_n\}$ adalah barisan variabel acak. Berlaku pernyataan-pernyataan berikut.

- (a) **Dekomposisi ekspektasi:** Jika $\mathbb{E}[|X_n|] = \mathcal{O}(a_n)$ (dalam arti deterministik), maka $X_n = \mathcal{O}_p(a_n)$.
- (b) **Dekomposisi variansi:** Jika $\mathbb{E}[X_n] = \mu_n$ dan $\text{Var}(X_n) = \mathcal{O}(a_n^2)$, maka $X_n - \mu_n = \mathcal{O}_p(a_n)$.

(c) **Dekomposisi momen kedua:** Jika $\mathbb{E}[X_n^2] = \mathcal{O}(a_n^2)$, maka $X_n = \mathcal{O}_p(a_n)$.

Bukti. Berikut adalah bukti untuk ketiga pernyataan.

(a) **Dekomposisi ekspektasi:** Andaikan $\mathbb{E}[|X_n|] \leq Ca_n$ untuk suatu konstanta $C > 0$ dan semua n cukup besar. Dengan ketaksamaan Markov, untuk setiap $M > 0$:

$$\Pr(|X_n| > Ma_n) \leq \frac{\mathbb{E}[|X_n|]}{Ma_n} \leq \frac{Ca_n}{Ma_n} = \frac{C}{M}. \quad (2.272)$$

Untuk setiap $\varepsilon > 0$, pilih $M = C/\varepsilon$, maka $\Pr(|X_n| > Ma_n) < \varepsilon$. Dengan demikian, $X_n = \mathcal{O}_p(a_n)$.

(b) **Dekomposisi variansi:** Andaikan $\text{Var}(X_n) \leq Ca_n^2$ untuk suatu konstanta $C > 0$. Dengan ketaksamaan Chebyshev, untuk setiap $M > 0$:

$$\Pr(|X_n - \mu_n| > Ma_n) \leq \frac{\text{Var}(X_n)}{M^2 a_n^2} \leq \frac{Ca_n^2}{M^2 a_n^2} = \frac{C}{M^2}. \quad (2.273)$$

Untuk setiap $\varepsilon > 0$, pilih $M = \sqrt{C/\varepsilon}$, maka $\Pr(|X_n - \mu_n| > Ma_n) < \varepsilon$. Dengan demikian, $X_n - \mu_n = \mathcal{O}_p(a_n)$.

(c) **Dekomposisi momen kedua:** Andaikan $\mathbb{E}[X_n^2] \leq Ca_n^2$ untuk suatu konstanta $C > 0$. Dengan ketaksamaan Markov pada X_n^2 :

$$\Pr(|X_n| > Ma_n) = \Pr(X_n^2 > M^2 a_n^2) \leq \frac{\mathbb{E}[X_n^2]}{M^2 a_n^2} \leq \frac{Ca_n^2}{M^2 a_n^2} = \frac{C}{M^2}. \quad (2.274)$$

Untuk setiap $\varepsilon > 0$, pilih $M = \sqrt{C/\varepsilon}$, maka $\Pr(|X_n| > Ma_n) < \varepsilon$. Dengan demikian, $X_n = \mathcal{O}_p(a_n)$.

■

Contoh 2.4.31 (Aplikasi Dekomposisi Momen) Misalkan X_1, X_2, \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 $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

Menggunakan dekomposisi variansi:

$$\mathbb{E}[\bar{X}_n] = \mu, \quad (2.275)$$

$$\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n} = \mathcal{O}(n^{-1}) = \mathcal{O}\left((n^{-1/2})^2\right). \quad (2.276)$$

Berdasarkan Teorema Dekomposisi Momen bagian (b), diperoleh

$$\bar{X}_n - \mu = \mathcal{O}_p(n^{-1/2}).$$

Perhitungan numerik: Untuk $n = 100$, $\mu = 0$, dan $\sigma^2 = 1$:

$$\text{Var}(\bar{X}_{100}) = \frac{1}{100} = 0.01, \quad (2.277)$$

$$\Pr(|\bar{X}_{100}| > 0.2) \leq \frac{0.01}{0.04} = 0.25. \quad (2.278)$$

Dengan distribusi eksak $\bar{X}_{100} \sim \mathcal{N}(0, 0.01)$:

$$\Pr(|\bar{X}_{100}| > 0.2) = 2(1 - \Phi(2)) \approx 0.0455.$$

Ini konsisten dengan $\bar{X}_n - \mu = \mathcal{O}_p(n^{-1/2})$ karena $0.2 = 2 \cdot 0.1 = 2 \cdot n^{-1/2}$.

Teorema 2.4.32 (Dekomposisi o_p Berdasarkan Momen, (?)) Misalkan $\{X_n\}$ adalah barisan variabel acak. Berlaku pernyataan-pernyataan berikut.

(a) **Dekomposisi ekspektasi:** Jika $\mathbb{E}[|X_n|] = o(a_n)$ (dalam arti deterministik), maka $X_n = o_p(a_n)$.

(b) **Dekomposisi variansi:** Jika $\mathbb{E}[X_n] = o(a_n)$ dan $\text{Var}(X_n) = o(a_n^2)$, maka $X_n = o_p(a_n)$.

(c) **Dekomposisi momen kedua:** Jika $\mathbb{E}[X_n^2] = o(a_n^2)$, maka $X_n = o_p(a_n)$.

Bukti. Berikut adalah bukti untuk ketiga pernyataan.

(a) **Dekomposisi ekspektasi:** Andaikan $\mathbb{E}[|X_n|]/a_n \rightarrow 0$ ketika $n \rightarrow \infty$. De-

ngan ketaksamaan Markov, untuk setiap $\varepsilon > 0$:

$$\Pr \left(\left| \frac{X_n}{a_n} \right| > \varepsilon \right) \leq \frac{\mathbb{E}[|X_n|/a_n]}{\varepsilon} = \frac{\mathbb{E}[|X_n|]}{\varepsilon a_n} \rightarrow 0. \quad (2.279)$$

Dengan demikian, $X_n/a_n \xrightarrow{p} 0$, sehingga $X_n = o_p(a_n)$.

- (b) **Dekomposisi variansi:** Andaikan $\mathbb{E}[X_n] = o(a_n)$ dan $\text{Var}(X_n) = o(a_n^2)$. Maka $\mathbb{E}[X_n]/a_n \rightarrow 0$ dan $\text{Var}(X_n)/a_n^2 \rightarrow 0$. Dengan ketaksamaan Chebyshev:

$$\Pr \left(\left| \frac{X_n - \mathbb{E}[X_n]}{a_n} \right| > \varepsilon \right) \leq \frac{\text{Var}(X_n)}{\varepsilon^2 a_n^2} \rightarrow 0. \quad (2.280)$$

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) **Dekomposisi momen kedua:** Andaikan $\mathbb{E}[X_n^2]/a_n^2 \rightarrow 0$. Dengan ketaksamaan Markov:

$$\Pr \left(\left| \frac{X_n}{a_n} \right| > \varepsilon \right) = \Pr \left(\frac{X_n^2}{a_n^2} > \varepsilon^2 \right) \leq \frac{\mathbb{E}[X_n^2]}{\varepsilon^2 a_n^2} \rightarrow 0. \quad (2.281)$$

Dengan demikian, $X_n/a_n \xrightarrow{p} 0$, sehingga $X_n = o_p(a_n)$.

■

Contoh 2.4.33 (Aplikasi Dekomposisi o_p) Misalkan $X_n \sim \mathcal{N}(0, 1/n^3)$ untuk $n = 1, 2, \dots$. Akan ditunjukkan bahwa $X_n = o_p(n^{-1})$.

Menggunakan dekomposisi momen kedua:

$$\mathbb{E}[X_n^2] = \text{Var}(X_n) = \frac{1}{n^3}. \quad (2.282)$$

Perhatikan bahwa

$$\frac{\mathbb{E}[X_n^2]}{(n^{-1})^2} = \frac{1/n^3}{1/n^2} = \frac{1}{n} \rightarrow 0. \quad (2.283)$$

Berdasarkan Teorema Dekomposisi o_p bagian (c), diperoleh $X_n = o_p(n^{-1})$.

Verifikasi langsung:

$$\frac{X_n}{n^{-1}} = nX_n \sim \mathcal{N} \left(0, \frac{n^2}{n^3} \right) = \mathcal{N} \left(0, \frac{1}{n} \right). \quad (2.284)$$

Untuk setiap $\varepsilon > 0$:

$$\Pr(|nX_n| > \varepsilon) = 2 \left(1 - \Phi \left(\frac{\varepsilon}{\sqrt{1/n}} \right) \right) = 2 \left(1 - \Phi(\varepsilon\sqrt{n}) \right) \rightarrow 0 \quad (2.285)$$

ketika $n \rightarrow \infty$. Dengan demikian, $nX_n \xrightarrow{p} 0$, sehingga $X_n = o_p(n^{-1})$.

2.4.7. Teorema Limit Pusat Klasik

Teorema 2.4.34 (Teorema Limit Pusat, (?)) Misalkan $\{X_i\}_{i=1}^{\infty}$ adalah barisan variabel acak i.i.d. dengan $\mathbb{E}[X_1] = \mu$ dan $\text{Var}(X_1) = \sigma^2 < \infty$. Maka

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X_i - \mu \right) \Rightarrow \mathcal{N}(0, \sigma^2). \quad (2.286)$$

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

$$S_n = \sum_{i=1}^n Y_i = \sum_{i=1}^n (X_i - \mu). \quad (2.287)$$

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

$$\phi_Y(t) = 1 + it\mathbb{E}[Y_i] + \frac{(it)^2}{2}\mathbb{E}[Y_i^2] + o(t^2) = 1 - \frac{t^2\sigma^2}{2} + o(t^2). \quad (2.288)$$

Sebab Y_1, Y_2, \dots, Y_n saling independen dan berdistribusi identik, fungsi ka-

rakteristik dari $S_n/\sqrt{n\sigma^2}$ adalah

$$\begin{aligned}
 \phi_{S_n/\sqrt{n\sigma^2}}(t) &= \mathbb{E} \left[\exp \left(\frac{itS_n}{\sqrt{n\sigma^2}} \right) \right] \\
 &= \mathbb{E} \left[\exp \left(\sum_{i=1}^n \frac{itY_i}{\sqrt{n\sigma^2}} \right) \right] \\
 &= \prod_{i=1}^n \mathbb{E} \left[\exp \left(\frac{itY_i}{\sqrt{n\sigma^2}} \right) \right] \\
 &= \left[\phi_Y \left(\frac{t}{\sqrt{n\sigma^2}} \right) \right]^n.
 \end{aligned} \tag{2.289}$$

Dengan substitusi ekspansi Taylor, diperoleh

$$\begin{aligned}
 \phi_{S_n/\sqrt{n\sigma^2}}(t) &= \left[1 - \frac{1}{2} \cdot \frac{t^2}{n\sigma^2} \cdot \sigma^2 + o\left(\frac{1}{n}\right) \right]^n \\
 &= \left[1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \right]^n.
 \end{aligned} \tag{2.290}$$

Dengan mengambil limit $n \rightarrow \infty$ dan menggunakan fakta bahwa $\lim_{n \rightarrow \infty} (1 + \frac{a}{n})^n = e^a$, diperoleh

$$\lim_{n \rightarrow \infty} \phi_{S_n/\sqrt{n\sigma^2}}(t) = \exp \left(-\frac{t^2}{2} \right), \tag{2.291}$$

yang merupakan fungsi karakteristik dari distribusi $\mathcal{N}(0, 1)$.

Berdasarkan teorema kontinuitas Lévy, konvergensi fungsi karakteristik mengimplikasikan konvergensi dalam distribusi. Oleh karena itu,

$$\frac{S_n}{\sqrt{n\sigma^2}} = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \Rightarrow \mathcal{N}(0, 1). \tag{2.292}$$

Dengan mengalikan kedua ruas dengan σ , diperoleh

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X_i - \mu \right) \Rightarrow \mathcal{N}(0, \sigma^2). \tag{2.293}$$

■

Teorema Limit Pusat merupakan hasil fundamental yang memungkinkan pendekatan distribusi normal terhadap rata-rata sampel, dan menjadi dasar utama inferensi statistik asimtotik.

Dalam banyak aplikasi statistik, khususnya estimator berbobot dan estimator lokal, observasi tidak lagi identik terdistribusi. Hal ini memotivasi penggunaan CLT untuk *triangular arrays*. *Triangular arrays* adalah kumpulan variabel acak yang disusun dalam bentuk segitiga, dengan setiap baris dapat memiliki distribusi yang berbeda.

Teorema 2.4.35 (CLT untuk *Triangular Arrays*, (?)) Misalkan $\{X_{n,i} : i = 1, \dots, n\}$ adalah *triangular array* variabel acak dengan $\mathbb{E}[X_{n,i}] = 0$ dan varians total

$$\sum_{i=1}^n \text{Var}(X_{n,i}) \rightarrow \sigma^2. \quad (2.294)$$

Jika kondisi Lindeberg terpenuhi, maka

$$\sum_{i=1}^n X_{n,i} \Rightarrow \mathcal{N}(0, \sigma^2). \quad (2.295)$$

Hasil ini sangat relevan untuk analisis asimtotik estimator lokal, di mana kontribusi masing-masing observasi bergantung pada n .

Teorema 2.4.36 (Teorema Lindeberg–Lévy, (?)) Misalkan $\{X_i\}_{i=1}^\infty$ adalah barisan variabel acak i.i.d. dengan $\mathbb{E}[X_i] = \mu$ dan $\text{Var}(X_i) = \sigma^2 < \infty$. Maka

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (2.296)$$

Bukti. Teorema Lindeberg–Lévy merupakan kasus khusus dari Teorema Limit Pusat klasik yang telah dibuktikan sebelumnya. Definisikan $Y_i = X_i - \mu$ sehingga $\mathbb{E}[Y_i] = 0$ dan $\text{Var}(Y_i) = \sigma^2$. Dengan menggunakan fungsi karakteristik, ekspansi Taylor dari $\phi_Y(t)$ di sekitar $t = 0$ memberikan

$$\phi_Y(t) = 1 - \frac{t^2\sigma^2}{2} + o(t^2). \quad (2.297)$$

Fungsi karakteristik dari $\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n Y_i$ adalah

$$\left[\phi_Y \left(\frac{t}{\sigma\sqrt{n}} \right) \right]^n = \left[1 - \frac{t^2}{2n} + o \left(\frac{1}{n} \right) \right]^n \rightarrow e^{-t^2/2} \quad (2.298)$$

ketika $n \rightarrow \infty$. Berdasarkan teorema kontinuitas Lévy, konvergensi ini mengimplikasikan konvergensi dalam distribusi ke $\mathcal{N}(0, 1)$. ■

Teorema 2.4.37 (Teorema Lindeberg–Feller, (?)) Misalkan $\{X_{n,i} : i = 1, \dots, n\}$ adalah triangular array variabel acak independen (tidak harus identik) dengan $\mathbb{E}[X_{n,i}] = 0$ untuk semua n dan i . Definisikan

$$s_n^2 = \sum_{i=1}^n \text{Var}(X_{n,i}) = \sum_{i=1}^n \sigma_{n,i}^2. \quad (2.299)$$

Jika kondisi Lindeberg terpenuhi, yaitu untuk setiap $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^n \mathbb{E} [X_{n,i}^2 \cdot \mathbf{1}_{\{|X_{n,i}| > \varepsilon s_n\}}] = 0, \quad (2.300)$$

maka

$$\frac{1}{s_n} \sum_{i=1}^n X_{n,i} \xrightarrow{d} \mathcal{N}(0, 1). \quad (2.301)$$

Bukti. Pembuktian dilakukan dengan menggunakan fungsi karakteristik. Definisikan $S_n = \sum_{i=1}^n X_{n,i}$ dan $Z_n = S_n / s_n$. Fungsi karakteristik dari Z_n adalah

$$\phi_{Z_n}(t) = \prod_{i=1}^n \phi_{X_{n,i}} \left(\frac{t}{s_n} \right). \quad (2.302)$$

Untuk setiap i , dengan ekspansi Taylor fungsi karakteristik diperoleh

$$\phi_{X_{n,i}} \left(\frac{t}{s_n} \right) = 1 + i \frac{t}{s_n} \mathbb{E}[X_{n,i}] - \frac{t^2}{2s_n^2} \mathbb{E}[X_{n,i}^2] + R_{n,i}(t), \quad (2.303)$$

dengan $\mathbb{E}[X_{n,i}] = 0$ dan sisa $R_{n,i}(t)$ memenuhi $|R_{n,i}(t)| \leq \frac{|t|^3}{6s_n^3} \mathbb{E}[|X_{n,i}|^3]$.

Sehingga,

$$\phi_{X_{n,i}}\left(\frac{t}{s_n}\right) = 1 - \frac{t^2 \sigma_{n,i}^2}{2s_n^2} + R_{n,i}(t). \quad (2.304)$$

Kondisi Lindeberg menjamin bahwa kontribusi suku-suku dengan $|X_{n,i}|$ besar dapat diabaikan. Secara lebih rinci, kondisi ini memastikan bahwa

$$\max_{1 \leq i \leq n} \frac{\sigma_{n,i}^2}{s_n^2} \rightarrow 0 \quad (2.305)$$

ketika $n \rightarrow \infty$, yang berarti tidak ada variabel tunggal yang mendominasi variansi total.

Dengan menggunakan fakta bahwa $\log(1 + x) \approx x$ untuk $|x|$ kecil dan

$$\log \phi_{Z_n}(t) = \sum_{i=1}^n \log \phi_{X_{n,i}}\left(\frac{t}{s_n}\right), \quad (2.306)$$

diperoleh

$$\begin{aligned} \log \phi_{Z_n}(t) &= \sum_{i=1}^n \log \left(1 - \frac{t^2 \sigma_{n,i}^2}{2s_n^2} + R_{n,i}(t)\right) \\ &\approx \sum_{i=1}^n \left(-\frac{t^2 \sigma_{n,i}^2}{2s_n^2}\right) + \sum_{i=1}^n R_{n,i}(t) \\ &= -\frac{t^2}{2s_n^2} \sum_{i=1}^n \sigma_{n,i}^2 + \sum_{i=1}^n R_{n,i}(t) \\ &= -\frac{t^2}{2} + \sum_{i=1}^n R_{n,i}(t). \end{aligned} \quad (2.307)$$

Kondisi Lindeberg menjamin bahwa $\sum_{i=1}^n R_{n,i}(t) \rightarrow 0$ ketika $n \rightarrow \infty$.

Oleh karena itu,

$$\lim_{n \rightarrow \infty} \phi_{Z_n}(t) = e^{-t^2/2}, \quad (2.308)$$

yang merupakan fungsi karakteristik dari $\mathcal{N}(0, 1)$. Berdasarkan teorema kontinuitas Lévy, $Z_n \xrightarrow{d} \mathcal{N}(0, 1)$. ■

Contoh 2.4.38 Misalkan $X_{n,i} = Y_i/\sqrt{n}$ untuk $i = 1, \dots, n$, 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 $\varepsilon > 0$,

$$\begin{aligned} \frac{1}{s_n^2} \sum_{i=1}^n \mathbb{E} [X_{n,i}^2 \cdot \mathbf{1}_{\{|X_{n,i}| > \varepsilon s_n\}}] &= \frac{1}{\sigma^2} \sum_{i=1}^n \mathbb{E} \left[\frac{Y_i^2}{n} \cdot \mathbf{1}_{\{|Y_i| > \varepsilon \sigma \sqrt{n}\}} \right] \\ &= \frac{1}{\sigma^2} \mathbb{E} [Y_1^2 \cdot \mathbf{1}_{\{|Y_1| > \varepsilon \sigma \sqrt{n}\}}]. \end{aligned} \quad (2.309)$$

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.

2.4.8. Metode Delta

Dalam banyak aplikasi statistik, sering kali diperlukan distribusi asimtotik dari suatu fungsi transformasi estimator, bukan hanya distribusi estimator itu sendiri. Sebagai contoh, jika $\hat{\theta}_n$ adalah estimator untuk parameter θ dan diperlukan inferensi terhadap $g(\theta)$ untuk suatu fungsi g , maka perlu diketahui distribusi asimtotik dari $g(\hat{\theta}_n)$. Metode Delta menyediakan alat yang elegan untuk menurunkan distribusi limit dari transformasi halus suatu estimator yang asimtotik normal. Metode ini memanfaatkan ekspansi Taylor orde pertama untuk mengaproksimasi fungsi nonlinear secara lokal sebagai fungsi linear, sehingga sifat normalitas asimtotik dapat dipertahankan melalui transformasi.

Teorema 2.4.39 (Metode Δ , (?) Misalkan

$$\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow \mathcal{N}(0, \Sigma), \quad (2.310)$$

dan g adalah fungsi terdiferensialkan di θ . Maka

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \Rightarrow \mathcal{N}(0, g'(\theta)\Sigma g'(\theta)^\top). \quad (2.311)$$

Bukti. Dengan menggunakan ekspansi Taylor orde pertama dari g di sekitar θ , di-

peroleh

$$g(\hat{\theta}_n) = g(\theta) + g'(\theta)(\hat{\theta}_n - \theta) + o(|\hat{\theta}_n - \theta|). \quad (2.312)$$

Sebab $\hat{\theta}_n \xrightarrow{p} \theta$, suku sisa $o(|\hat{\theta}_n - \theta|)$ memenuhi

$$\frac{o(|\hat{\theta}_n - \theta|)}{|\hat{\theta}_n - \theta|} \xrightarrow{p} 0. \quad (2.313)$$

Dengan mengalikan kedua ruas ekspansi Taylor dengan \sqrt{n} , diperoleh

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) = g'(\theta) \cdot \sqrt{n}(\hat{\theta}_n - \theta) + \sqrt{n} \cdot o(|\hat{\theta}_n - \theta|). \quad (2.314)$$

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,

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) = g'(\theta) \cdot \sqrt{n}(\hat{\theta}_n - \theta) + o_p(1). \quad (2.315)$$

Berdasarkan teorema Slutsky, sebab $\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow \mathcal{N}(0, \Sigma)$ dan $o_p(1) \xrightarrow{p} 0$, maka

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \Rightarrow g'(\theta) \cdot \mathcal{N}(0, \Sigma) = \mathcal{N}\left(0, g'(\theta)\Sigma g'(\theta)^\top\right). \quad (2.316)$$

Kesamaan terakhir mengikuti dari sifat transformasi linear variabel acak normal: jika $Z \sim \mathcal{N}(0, \Sigma)$, maka $AZ \sim \mathcal{N}(0, A\Sigma A^\top)$ untuk matriks $A = g'(\theta)$. ■

Contoh 2.4.40 Misalkan X_1, X_2, \dots, X_n adalah sampel acak i.i.d. dari distribusi dengan rataan $\mu > 0$ dan variansi σ^2 . Akan dicari distribusi asimtotik dari $\log(\bar{X}_n)$.

Berdasarkan Teorema Limit Pusat, $\sqrt{n}(\bar{X}_n - \mu) \Rightarrow \mathcal{N}(0, \sigma^2)$. Dengan $g(x) = \log(x)$ dan $g'(x) = 1/x$, metode Delta memberikan

$$\sqrt{n}(\log(\bar{X}_n) - \log(\mu)) \Rightarrow \mathcal{N}\left(0, \frac{\sigma^2}{\mu^2}\right). \quad (2.317)$$

Dengan demikian, $\log(\bar{X}_n)$ adalah asimtotik normal dengan rataan $\log(\mu)$ dan variansi asimtotik $\sigma^2/(n\mu^2)$.

2.5. Analisis Regresi Linear

Analisis regresi merupakan suatu teknik statistik untuk menginvestigasi dan memodelkan hubungan antarvariabel. Aplikasi dari analisis regresi sangat luas dan digunakan dalam hampir semua bidang, seperti teknik atau *engineering*, ilmu fisika dan kimia, ekonomi, manajemen, ilmu biologi, dan sosiologi. Dalam analisis regresi, variabel respons atau variabel dependen (dinotasikan dengan variabel acak Y) dimodelkan dengan fungsi dalam variabel penjelas atau variabel independen (dinotasikan dengan variabel acak X). (Montgomery, Peck, & 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*

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x} + \boldsymbol{\varepsilon}, \quad (2.318)$$

dengan intersep β_0 dan koefisien β_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 β_0 dan β_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*

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \cdots + \beta_p \mathbf{x}_p + \boldsymbol{\varepsilon}. \quad (2.319)$$

Model di atas dapat dituliskan dalam notasi matriks dan vektor, yaitu

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \quad (2.320)$$

atau

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (2.321)$$

Model OLS mengasumsikan bahwa galat memiliki rataan nol dan variansi konstan σ^2 yang tidak diketahui nilainya. Lebih lanjut, model ini juga mengasumsikan bahwa komponen galat tidak memiliki autokorelasi yang berarti bahwa nilai dari suatu galat tidak bergantung pada galat lainnya. Dengan kata lain $\boldsymbol{\varepsilon} \stackrel{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma^2)$, untuk suatu σ^2 yang bersifat konstan.

Model regresi linear memandang vektor acak \mathbf{y} sebagai fungsi dari matriks acak \mathbf{X} . Oleh karena itu, rataan dari distribusi \mathbf{y} dapat dituliskan sebagai

$$\mathbb{E}[\mathbf{y} | \mathbf{X}] = \mathbb{E}[\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}] = \mathbf{X}\boldsymbol{\beta} \quad (2.322)$$

dan variansinya

$$\text{Var}(\mathbf{y} | \mathbf{X}) = \text{Var}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n, \quad (2.323)$$

karena $\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$ dan $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n$. Oleh karena itu, rataan dari \mathbf{y} adalah fungsi linear dari \mathbf{X} meskipun variansinya tidak bergantung pada nilai \mathbf{X} .

Definisi 2.5.3 (Asumsi Regresi Linear Klasik, (Greene, 2002; Gujarati & Porter, 2009; Montgomery dkk., 2012)) *Model regresi linear klasik didasarkan pada sejumlah asumsi yang dikenal sebagai asumsi Gauss–Markov. Asumsi-asumsi ini diperlukan agar penduga kuadrat terkecil memiliki sifat optimal sebagai best linear unbiased estimator (BLUE). Asumsi-asumsi tersebut adalah sebagai berikut.*

(a) *Model bersifat linear dalam parameter, yaitu $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.*

(b) *Nilai tengah galat adalah nol bersyarat pada \mathbf{X} , yaitu $\mathbb{E}[\boldsymbol{\varepsilon} | \mathbf{X}] = \mathbf{0}$.*

(c) Matriks desain \mathbf{X} memiliki rank penuh sehingga $\mathbf{X}^\top \mathbf{X}$ nonsingular.

(d) Galat memiliki ragam konstan, $\text{Var}(\varepsilon_i | \mathbf{X}) = \sigma^2$ untuk semua i .

(e) Komponen galat tidak berkorelasi, $\text{Cov}(\varepsilon_i, \varepsilon_j | \mathbf{X}) = 0$ untuk $i \neq j$.

(f) Untuk keperluan inferensi, diasumsikan $\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$.

Estimasi parameter dalam regresi linear dilakukan dengan metode kuadrat terkecil atau *ordinary least squares* (OLS). Prinsip dasarnya adalah mencari nilai $\boldsymbol{\beta}$ yang meminimumkan jumlah kuadrat galat

$$S(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (2.324)$$

Perhatikan bahwa fungsi kuadrat terkecil dari persamaan di atas dapat dituliskan sebagai

$$S(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}^\top \mathbf{y} - 2\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}. \quad (2.325)$$

Dengan diferensial matriks, gradien terhadap $\boldsymbol{\beta}$ adalah

$$\nabla_{\boldsymbol{\beta}} S(\boldsymbol{\beta}) = -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}. \quad (2.326)$$

Syarat orde pertama meminimumkan S adalah $\nabla_{\boldsymbol{\beta}} S(\boldsymbol{\beta}) = \mathbf{0}$, sehingga diperoleh

$$\mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{y}, \quad (2.327)$$

yang disebut persamaan normal. Selanjutnya, matriks Hessian atau orde kedua adalah

$$\nabla_{\boldsymbol{\beta}}^2 S(\boldsymbol{\beta}) = 2\mathbf{X}^\top \mathbf{X} \succeq \mathbf{0}, \quad (2.328)$$

sehingga S konveks dalam $\boldsymbol{\beta}$. 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 peminimum global.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (2.329)$$

Jika $\mathbf{X}^\top \mathbf{X}$ singular, himpunan pemimum tidak tunggal. Salah satu solusi adalah solusi *norma-minimum* $\hat{\boldsymbol{\beta}} = \mathbf{X}^+ \mathbf{y}$ dengan \mathbf{X}^+ merupakan *Moore–Penrose pseudoinverse*.

Teorema 2.5.4 (Teorema Gauss–Markov, (Greene, 2002; Montgomery dkk., 2012)) *Di bawah asumsi regresi linear klasik, penduga OLS $\hat{\boldsymbol{\beta}}$ bersifat tak bias, yaitu $\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$, dan memiliki kovariansi*

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}. \quad (2.330)$$

Selain itu, $\hat{\boldsymbol{\beta}}$ merupakan best linear unbiased estimator (BLUE), artinya $\hat{\boldsymbol{\beta}}$ memiliki variansi terkecil di antara semua penduga linear tak bias.

Bukti. Penduga OLS dapat dituliskan sebagai

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (2.331)$$

Substitusi $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ menghasilkan

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\varepsilon}, \quad (2.332)$$

sehingga rataan dari $\hat{\boldsymbol{\beta}}$ adalah

$$\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta} + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbb{E}[\boldsymbol{\varepsilon}] = \boldsymbol{\beta}, \quad (2.333)$$

yang menunjukkan bahwa $\hat{\boldsymbol{\beta}}$ tak bias. Kovariansnya diperoleh dari

$$\text{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \text{Var}(\boldsymbol{\varepsilon}) \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}. \quad (2.334)$$

Untuk sifat BLUE, dipertimbangkan penduga linear tak bias umum $\tilde{\boldsymbol{\beta}} = \mathbf{C}\mathbf{y}$ dengan $\mathbf{C}\mathbf{X} = \mathbf{I}$. Variansi dari $\tilde{\boldsymbol{\beta}}$ adalah

$$\text{Var}(\tilde{\boldsymbol{\beta}}) = \mathbf{C}\sigma^2 \mathbf{I}\mathbf{C}^\top = \sigma^2 \mathbf{C}\mathbf{C}^\top. \quad (2.335)$$

Melalui dekomposisi $\mathbf{C} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top + \mathbf{D}$ dengan $\mathbf{D}\mathbf{X} = \mathbf{0}$, diperoleh

$$\text{Var}(\tilde{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} + \sigma^2 \mathbf{D}\mathbf{D}^\top \succeq \text{Var}(\hat{\boldsymbol{\beta}}), \quad (2.336)$$

yang membuktikan bahwa $\hat{\boldsymbol{\beta}}$ memiliki variansi minimum di antara semua penduga linear tak bias. ■

Teorema 2.5.5 (Kepadan OLS dan MLE, (Montgomery dkk., 2012)) *Jika asumsi normalitas galat terpenuhi, maka penduga kemungkinan maksimum (PKM) atau maximum likelihood estimation (MLE) untuk $\boldsymbol{\beta}$ identik dengan penduga kuadrat terkecil (OLS), yaitu*

$$\hat{\boldsymbol{\beta}}_{MLE} = \hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \quad (2.337)$$

Penduga MLE untuk varians galat adalah

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \quad (2.338)$$

Bukti. Dengan asumsi $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, fungsi *likelihood* untuk $\boldsymbol{\beta}$ dan σ^2 adalah

$$L(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right), \quad (2.339)$$

yang memiliki log-*likelihood*

$$\ell(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (2.340)$$

Maksimisasi terhadap β dengan turunan pertama didapatkan

$$\begin{aligned}
 \frac{\partial \ell(\beta, \sigma^2)}{\partial \beta} &= \frac{2\mathbf{X}^\top}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta) \\
 &= \frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{y} - \frac{1}{\sigma^2} \mathbf{X}^\top \mathbf{X} \beta \\
 0 &= \mathbf{X}^\top \mathbf{y} - \mathbf{X}^\top \mathbf{X} \hat{\beta} \\
 \mathbf{X}^\top \mathbf{X} \hat{\beta} &= \mathbf{X}^\top \mathbf{y} \\
 \hat{\beta} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}. \tag{2.341}
 \end{aligned}$$

Perhatikan bahwa turunan kedua terhadap β 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

$$\hat{\beta}_{\text{MLE}} = \hat{\beta}_{\text{OLS}}. \tag{2.342}$$

Selanjutnya, turunan pertama ℓ terhadap σ^2 akan menghasilkan

$$\begin{aligned}
 \frac{\partial \ell(\beta, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta) \\
 0 &= -n\hat{\sigma}^2 + (\mathbf{y} - \mathbf{X}\hat{\beta})^\top (\mathbf{y} - \mathbf{X}\hat{\beta}) \\
 \hat{\sigma}^2 &= \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\beta})^\top (\mathbf{y} - \mathbf{X}\hat{\beta}). \tag{2.343}
 \end{aligned}$$

■

2.5.2. Teori Asimtotik untuk OLS

Teori asimtotik mempelajari perilaku penduga OLS ketika ukuran sampel n menuju tak hingga. Hasil-hasil asimtotik sangat penting karena memberikan justifikasi teoritis untuk inferensi statistik pada sampel besar, bahkan ketika asumsi normalitas galat tidak terpenuhi.

Teorema 2.5.6 (Konsistensi Penduga OLS, (Greene, 2002)) Misalkan model regresi linear $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ memenuhi kondisi-kondisi berikut.

- (i) $\mathbb{E}[\varepsilon | \mathbf{X}] = \mathbf{0}$ (eksogenitas ketat).

(ii) $\frac{1}{n}\mathbf{X}^\top \mathbf{X} \xrightarrow{p} \mathbf{Q}$ dengan \mathbf{Q} matriks definit positif.

Maka penduga OLS $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ bersifat konsisten, yaitu

$$\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}. \quad (2.344)$$

Bukti. Penduga OLS dapat dituliskan sebagai

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \\ &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{\beta} + (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\varepsilon} \\ &= \boldsymbol{\beta} + \left(\frac{1}{n} \mathbf{X}^\top \mathbf{X} \right)^{-1} \left(\frac{1}{n} \mathbf{X}^\top \boldsymbol{\varepsilon} \right). \end{aligned} \quad (2.345)$$

Berdasarkan kondisi (ii), $\frac{1}{n}\mathbf{X}^\top \mathbf{X} \xrightarrow{p} \mathbf{Q}$ dengan \mathbf{Q} definit positif. Dengan teorema pemetaan kontinu, $(\frac{1}{n}\mathbf{X}^\top \mathbf{X})^{-1} \xrightarrow{p} \mathbf{Q}^{-1}$.

Selanjutnya, perlu ditunjukkan bahwa $\frac{1}{n}\mathbf{X}^\top \boldsymbol{\varepsilon} \xrightarrow{p} \mathbf{0}$. Perhatikan bahwa

$$\frac{1}{n}\mathbf{X}^\top \boldsymbol{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \varepsilon_i. \quad (2.346)$$

Sebab $\mathbb{E}[\mathbf{x}_i \varepsilon_i] = \mathbb{E}[\mathbb{E}[\mathbf{x}_i \varepsilon_i \mid \mathbf{X}]] = \mathbb{E}[\mathbf{x}_i \mathbb{E}[\varepsilon_i \mid \mathbf{X}]] = \mathbf{0}$ (dari kondisi (i)) dan dengan asumsi kondisi regularitas tambahan (seperti $\mathbb{E}[\|\mathbf{x}_i\|^2 \varepsilon_i^2] < \infty$), Hukum Bilangan Besar menjamin bahwa $\frac{1}{n}\mathbf{X}^\top \boldsymbol{\varepsilon} \xrightarrow{p} \mathbf{0}$.

Dengan teorema Slutsky, diperoleh

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\frac{1}{n} \mathbf{X}^\top \mathbf{X} \right)^{-1} \left(\frac{1}{n} \mathbf{X}^\top \boldsymbol{\varepsilon} \right) \xrightarrow{p} \mathbf{Q}^{-1} \cdot \mathbf{0} = \mathbf{0}, \quad (2.347)$$

sehingga $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$. ■

Teorema 2.5.7 (Konsistensi Penduga Variansi, (Greene, 2002)) Di bawah kondisi-kondisi Teorema konsistensi OLS dan dengan tambahan asumsi $\mathbb{E}[\varepsilon_i^2 \mid \mathbf{X}] = \sigma^2$

(homoskedastisitas), penduga variansi

$$s^2 = \frac{1}{n-p-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \frac{1}{n-p-1} \sum_{i=1}^n \hat{\varepsilon}_i^2 \quad (2.348)$$

bersifat konsisten, yaitu $s^2 \xrightarrow{p} \sigma^2$.

Bukti. Definisikan residual $\hat{\varepsilon} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{M}\mathbf{y}$ dengan $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top$ adalah matriks proyeksi idempoten. Maka

$$\hat{\varepsilon} = \mathbf{M}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{M}\boldsymbol{\varepsilon}, \quad (2.349)$$

sebab $\mathbf{M}\mathbf{X} = \mathbf{0}$. Oleh karena itu,

$$\hat{\varepsilon}^\top \hat{\varepsilon} = \boldsymbol{\varepsilon}^\top \mathbf{M}^\top \mathbf{M}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^\top \mathbf{M}\boldsymbol{\varepsilon}. \quad (2.350)$$

Dengan menggunakan sifat jejak dan ekspektasi, dapat ditunjukkan bahwa

$$\mathbb{E}[\hat{\varepsilon}^\top \hat{\varepsilon}] = \mathbb{E}[\text{tr}(\boldsymbol{\varepsilon}^\top \mathbf{M}\boldsymbol{\varepsilon})] = \mathbb{E}[\text{tr}(\mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top)] = \text{tr}(\mathbf{M}\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top]) = \sigma^2 \text{tr}(\mathbf{M}) = \sigma^2(n-p-1). \quad (2.351)$$

Dengan Hukum Bilangan Besar, $\frac{1}{n}\hat{\varepsilon}^\top \hat{\varepsilon} \xrightarrow{p} \sigma^2$. Sebab $\frac{n}{n-p-1} \rightarrow 1$, maka $s^2 \xrightarrow{p} \sigma^2$. ■

Teorema 2.5.8 (Normalitas Asimtotik Penduga OLS, (Greene, 2002)) Misalkan kondisi-kondisi konsistensi terpenuhi dan tambahan kondisi berikut dipenuhi.

- (i) $\mathbb{E}[\varepsilon_i^2 | \mathbf{X}] = \sigma^2$ (homoskedastisitas).
- (ii) $\frac{1}{\sqrt{n}}\mathbf{X}^\top \boldsymbol{\varepsilon} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q})$ (kondisi Lindeberg–Feller terpenuhi).

Maka penduga OLS memiliki distribusi asimtotik

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}^{-1}). \quad (2.352)$$

Bukti. Dari pembuktian konsistensi, dapat dituliskan

$$\begin{aligned}\sqrt{n}(\hat{\beta} - \beta) &= \sqrt{n} \left(\frac{1}{n} \mathbf{X}^\top \mathbf{X} \right)^{-1} \left(\frac{1}{n} \mathbf{X}^\top \boldsymbol{\varepsilon} \right) \\ &= \left(\frac{1}{n} \mathbf{X}^\top \mathbf{X} \right)^{-1} \left(\frac{1}{\sqrt{n}} \mathbf{X}^\top \boldsymbol{\varepsilon} \right).\end{aligned}\quad (2.353)$$

Berdasarkan kondisi (ii), $\frac{1}{\sqrt{n}} \mathbf{X}^\top \boldsymbol{\varepsilon} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q})$. Dengan teorema Slutsky, sebab $\left(\frac{1}{n} \mathbf{X}^\top \mathbf{X} \right)^{-1} \xrightarrow{p} \mathbf{Q}^{-1}$, maka

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathbf{Q}^{-1} \cdot \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}) = \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}^{-1}). \quad (2.354)$$

■

Akibat 2.5.9 (Distribusi Asimtotik dengan Variansi Diestimasi) *Dengan penduga variansi konsisten $s^2 \xrightarrow{p} \sigma^2$, berlaku*

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N} \left(\mathbf{0}, \sigma^2 \left(\frac{1}{n} \mathbf{X}^\top \mathbf{X} \right)^{-1} \right), \quad (2.355)$$

dan penduga matriks kovarians asimtotik yang konsisten adalah

$$\widehat{\text{Var}}(\hat{\beta}) = s^2 (\mathbf{X}^\top \mathbf{X})^{-1}. \quad (2.356)$$

Teorema 2.5.10 (Distribusi Eksak dan Asimtotik Uji-t, (Greene, 2002; Montgomery dkk., 2012)) *Untuk menguji hipotesis $H_0 : \beta_k = \beta_k^0$ terhadap $H_1 : \beta_k \neq \beta_k^0$, statistik uji-t didefinisikan sebagai*

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{\text{se}(\hat{\beta}_k)}, \quad (2.357)$$

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, $t_k \sim t_{n-p-1}$.*

- (b) **Distribusi asimtotik:** Di bawah kondisi regularitas asimtotik dan H_0 benar,
 $t_k \xrightarrow{d} \mathcal{N}(0, 1)$.

Bukti.

- (a) **Distribusi eksak:** Di bawah normalitas, $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1})$, sehingga
 $\hat{\beta}_k \sim \mathcal{N}(\beta_k, \sigma^2[(\mathbf{X}^\top \mathbf{X})^{-1}]_{kk})$. Oleh karena itu,

$$Z_k = \frac{\hat{\beta}_k - \beta_k}{\sigma \sqrt{[(\mathbf{X}^\top \mathbf{X})^{-1}]_{kk}}} \sim \mathcal{N}(0, 1). \quad (2.358)$$

Selain itu, $(n - p - 1)s^2/\sigma^2 \sim \chi^2_{n-p-1}$ dan independen dari $\hat{\beta}$. Dengan demikian,

$$t_k = \frac{Z_k}{\sqrt{(n - p - 1)s^2/\sigma^2/(n - p - 1)}} = \frac{Z_k}{\sqrt{s^2/\sigma^2}} \sim t_{n-p-1}. \quad (2.359)$$

- (b) **Distribusi asimtotik:** 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,

$$t_k = \frac{\hat{\beta}_k - \beta_k^0}{s \sqrt{[(\mathbf{X}^\top \mathbf{X})^{-1}]_{kk}}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (2.360)$$

■

Teorema 2.5.11 (Distribusi Eksak dan Asimtotik Uji-F, (Greene, 2002; Montgomery dkk., 2012)) 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

$$F = \frac{(\mathbf{R}\hat{\beta} - \mathbf{r})^\top [\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r})}{q \cdot s^2}. \quad (2.361)$$

- (a) **Distribusi eksak:** Di bawah asumsi normalitas galat dan H_0 benar, $F \sim F_{q, n-p-1}$.

- (b) **Distribusi asimtotik:** Di bawah kondisi regularitas asimtotik dan H_0 benar,
 $qF \xrightarrow{d} \chi_q^2$.

Bukti.

(a) **Distribusi eksak:** Di bawah H_0 dan normalitas, $\mathbf{R}\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\mathbf{r}, \sigma^2 \mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top)$.

Oleh karena itu,

$$\frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^\top [\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})}{\sigma^2} \sim \chi_q^2. \quad (2.362)$$

Sebab $(n - p - 1)s^2/\sigma^2 \sim \chi_{n-p-1}^2$ dan independen dari pembilang, maka

$$F = \frac{\chi_q^2/q}{\chi_{n-p-1}^2/(n - p - 1)} \sim F_{q, n-p-1}. \quad (2.363)$$

(b) **Distribusi asimtotik:** Dengan normalitas asimtotik dan $s^2 \xrightarrow{p} \sigma^2$,

$$qF = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^\top [\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})}{s^2} \xrightarrow{d} \chi_q^2. \quad (2.364)$$

■

Teorema 2.5.12 (Uji Wald, (Greene, 2002; ?)) Untuk menguji hipotesis H_0 : $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ dengan restriksi linear, statistik Wald didefinisikan sebagai

$$W = (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^\top \left[\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) \mathbf{R}^\top \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}), \quad (2.365)$$

dengan $\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = s^2(\mathbf{X}^\top \mathbf{X})^{-1}$. Di bawah H_0 dan kondisi regularitas asimtotik,

$$W \xrightarrow{d} \chi_q^2. \quad (2.366)$$

Bukti. Dari normalitas asimtotik, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}^{-1})$. Dengan transformasi linear,

$$\sqrt{n}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{R}\mathbf{Q}^{-1} \mathbf{R}^\top). \quad (2.367)$$

Di bawah H_0 , $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, sehingga

$$\sqrt{n}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{R}\mathbf{Q}^{-1} \mathbf{R}^\top). \quad (2.368)$$

Statistik Wald dapat dituliskan sebagai

$$W = n(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^\top \left[\mathbf{R}(n\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}))\mathbf{R}^\top \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}). \quad (2.369)$$

Sebab $n\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = s^2 n(\mathbf{X}^\top \mathbf{X})^{-1} \xrightarrow{p} \sigma^2 \mathbf{Q}^{-1}$, maka dengan teorema Slutsky dan sifat bentuk kuadratik normal, $W \xrightarrow{d} \chi_q^2$. ■

Teorema 2.5.13 (Uji Rasio Kemungkinan (*Likelihood Ratio Test*), (Greene, 2002))

Untuk menguji $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, statistik rasio kemungkinan didefinisikan sebagai

$$\text{LR} = n \left(\log \hat{\sigma}_R^2 - \log \hat{\sigma}_U^2 \right), \quad (2.370)$$

dengan $\hat{\sigma}_R^2$ dan $\hat{\sigma}_U^2$ adalah penduga MLE variansi di bawah model terestriksi dan tidak terestriksi. Di bawah H_0 dan kondisi regularitas,

$$\text{LR} \xrightarrow{d} \chi_q^2. \quad (2.371)$$

Bukti. Fungsi log-likelihood untuk model regresi normal adalah

$$\ell(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (2.372)$$

Penduga MLE untuk σ^2 pada model tidak terestriksi adalah $\hat{\sigma}_U^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_U)^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_U)$, sedangkan pada model terestriksi adalah $\hat{\sigma}_R^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_R)^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_R)$.

Statistik rasio kemungkinan didefinisikan sebagai $\text{LR} = 2(\ell_U - \ell_R)$, dengan ℓ_U dan ℓ_R adalah nilai log-likelihood maksimum pada model tidak terestriksi dan terestriksi. Substitusi memberikan

$$\begin{aligned} \text{LR} &= 2 \left[-\frac{n}{2} \log(\hat{\sigma}_U^2) - \frac{n}{2} + \frac{n}{2} \log(\hat{\sigma}_R^2) + \frac{n}{2} \right] \\ &= n \left(\log \hat{\sigma}_R^2 - \log \hat{\sigma}_U^2 \right). \end{aligned} \quad (2.373)$$

Di bawah H_0 , dengan ekspansi Taylor dari $\log(\hat{\sigma}_R^2/\hat{\sigma}_U^2)$ di sekitar 1 dan menggunakan fakta bahwa $\hat{\sigma}_R^2 - \hat{\sigma}_U^2 = \frac{1}{n}(\mathbf{R}\hat{\boldsymbol{\beta}}_U - \mathbf{r})^\top [\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{R}^\top]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}_U - \mathbf{r})$,

$\mathbf{r}) + o_p(n^{-1})$, dapat ditunjukkan bahwa

$$\text{LR} = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}}_U - \mathbf{r})^\top [\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}}_U - \mathbf{r})}{\hat{\sigma}_U^2} + o_p(1). \quad (2.374)$$

Sebab $\sqrt{n}(\mathbf{R}\hat{\boldsymbol{\beta}}_U - \mathbf{r}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{RQ}^{-1} \mathbf{R}^\top)$ di bawah H_0 dan $\hat{\sigma}_U^2 \xrightarrow{p} \sigma^2$, maka berdasarkan sifat bentuk kuadratik normal, $\text{LR} \xrightarrow{d} \chi_q^2$. ■

Teorema 2.5.14 (Uji Pengali Lagrange (Lagrange Multiplier Test), (Greene, 2002))

Untuk menguji $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, statistik pengali Lagrange didefinisikan sebagai

$$\text{LM} = \frac{1}{\hat{\sigma}_R^2} \hat{\boldsymbol{\varepsilon}}_R^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \hat{\boldsymbol{\varepsilon}}_R, \quad (2.375)$$

dengan $\hat{\boldsymbol{\varepsilon}}_R$ adalah residual dari model terestriksi. Di bawah H_0 dan kondisi regularitas,

$$\text{LM} \xrightarrow{d} \chi_q^2. \quad (2.376)$$

Proposisi 2.5.15 (Kesetaraan Asimtotik Uji-Uji Klasik, (Greene, 2002)) Di bawah hipotesis nol $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ dan kondisi regularitas, ketiga statistik uji—Wald (W), rasio kemungkinan (LR), dan pengali Lagrange (LM)—bersifat setara secara asimtotik dalam arti

$$W - \text{LR} \xrightarrow{p} 0, \quad \text{LR} - \text{LM} \xrightarrow{p} 0, \quad W - \text{LM} \xrightarrow{p} 0. \quad (2.377)$$

Ketiga statistik ini memiliki distribusi limit yang sama, yaitu χ_q^2 , sehingga memberikan kesimpulan inferensi yang identik secara asimtotik.

2.5.3. Regresi Terboboti

Regresi terboboti dibentuk apabila model regresi linear klasik melanggar asumsi homoskedastisitas galat. Heteroskedastisitas terjadi apabila komponen galat regresi tidak memiliki variansi yang konstan. Apabila terjadi heteroskedastisitas, maka Teorema Gauss–Markov tidak berlaku sehingga penduga dari model regresi tidak lagi bersifat BLUE.

Definisi 2.5.16 (Model Regresi Linear dengan Heteroskedastisitas, (Greene, 2002; Gujarati & Porter, 2009; Montgomery dkk., 2012)) Pertimbangkan model regresi linear

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathbb{E}[\boldsymbol{\varepsilon} | \mathbf{X}] = \mathbf{0}, \quad (2.378)$$

dengan struktur ragam–kovarians galat

$$\text{Var}(\boldsymbol{\varepsilon} | \mathbf{X}) = \sigma^2 \boldsymbol{\Omega}, \quad (2.379)$$

dengan $\boldsymbol{\Omega}$ adalah matriks simetri definit-positif berukuran $n \times n$. Model yang memperkenalkan struktur variansi-kovarians galat disebut sebagai model kuadrat terkecil tergeneralisasi atau generalized least squares (GLS).

Kasus umum dari GLS yang sering dipakai adalah kuadrat terkecil terboboti atau *weighted least squares* (WLS), yaitu $\boldsymbol{\Omega} = \text{diag}(\omega_1, \dots, \omega_n)$ yang merupakan heteroskedastisitas diagonal dengan bobot $\mathbf{W} = \boldsymbol{\Omega}^{-1} = \text{diag}(w_1, \dots, w_n)$, $w_i = 1/\omega_i$.

Penduga GLS diperoleh dengan meminimumkan jumlah kuadrat terboboti

$$S_{\mathbf{W}}(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad (2.380)$$

dengan $\mathbf{W} = \boldsymbol{\Omega}^{-1}$. Diferensiasi terhadap $\boldsymbol{\beta}$ dan disamakan dengan nol memberikan

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} S_{\mathbf{W}}(\boldsymbol{\beta}) &= -2\mathbf{W}\mathbf{X}^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (\text{karena } \mathbf{W} \text{ simetris}) \\ \mathbf{0} &= -2\mathbf{X}^\top \mathbf{W} \mathbf{y} - 2\mathbf{X}^\top \mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}} \\ \mathbf{X}^\top \mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}} &= \mathbf{X}^\top \mathbf{W} \mathbf{y}, \end{aligned} \quad (2.381)$$

sehingga pada asumsi $\mathbf{X}^\top \mathbf{W} \mathbf{X}$ nonsingular,

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{y}. \quad (2.382)$$

Penduga ini tak bias jika \mathbf{W} tidak acak atau independen dari $\boldsymbol{\varepsilon}$ dan $\mathbb{E}[\boldsymbol{\varepsilon} | \mathbf{X}] = \mathbf{0}$,

sebab $\mathbb{E}[\hat{\beta}_{\text{GLS}} | \mathbf{X}] = \boldsymbol{\beta}$.

Teorema 2.5.17 (Teorema Aitken atau Generalisasi Gauss–Markov, (Greene, 2002; Montgomery dkk., 2012)) Misalkan model regresi linear

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathbb{E}[\boldsymbol{\varepsilon} | \mathbf{X}] = 0, \quad \text{Var}(\boldsymbol{\varepsilon} | \mathbf{X}) = \sigma^2 \boldsymbol{\Omega},$$

dengan $\boldsymbol{\Omega}$ simetris definit-positif, maka penduga

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{y} \quad (2.383)$$

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

$$\mathbb{E}[\tilde{\boldsymbol{\beta}} | \mathbf{X}] = \mathbf{A}\mathbb{E}[\mathbf{y} | \mathbf{X}] = \mathbf{A}\mathbf{X}\boldsymbol{\beta}. \quad (2.384)$$

Supaya sama dengan $\boldsymbol{\beta}$ untuk semua $\boldsymbol{\beta}$, syarat yang harus dipenuhi adalah $\mathbf{A}\mathbf{X} = \mathbf{I}_p$. Selanjutnya, kovarians dari $\tilde{\boldsymbol{\beta}}$ diberikan oleh

$$\text{Var}(\tilde{\boldsymbol{\beta}} | \mathbf{X}) = \mathbf{A} \sigma^2 \boldsymbol{\Omega} \mathbf{A}^\top. \quad (2.385)$$

Di antara semua matriks \mathbf{A} yang memenuhi $\mathbf{A}\mathbf{X} = \mathbf{I}_p$, pilih

$$\mathbf{A}_0 = (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1}. \quad (2.386)$$

Jelas bahwa $\mathbf{A}_0 \mathbf{X} = \mathbf{I}_p$, sehingga \mathbf{A}_0 menghasilkan penduga tak bias. Penduga inilah yang dikenal sebagai *generalized least squares*.

Sekarang, ambil sembarang penduga tak bias lain \mathbf{A} . Matriks \mathbf{A} tersebut dapat ditulis sebagai $\mathbf{A} = \mathbf{A}_0 + \mathbf{D}$ dengan $\mathbf{D}\mathbf{X} = \mathbf{0}$, sebab $\mathbf{A}_0 \mathbf{X} = \mathbf{I}_p$ dan syarat $\mathbf{A}\mathbf{X} = \mathbf{I}_p$ harus tetap dipenuhi. Dengan demikian,

$$\tilde{\boldsymbol{\beta}} = \mathbf{A}\mathbf{y} = \mathbf{A}_0\mathbf{y} + \mathbf{D}\mathbf{y}. \quad (2.387)$$

Variansi dari penduga ini adalah

$$\text{Var}(\tilde{\boldsymbol{\beta}} \mid \mathbf{X}) = \text{Var}(\mathbf{A}_0 \mathbf{y} \mid \mathbf{X}) + \sigma^2 \mathbf{D} \boldsymbol{\Omega} \mathbf{D}^\top. \quad (2.388)$$

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 2.5.18 (Kepadanan GLS dan MLE di Bawah Normalitas, (Greene, 2002; Montgomery dkk., 2012)) Jika $\varepsilon \sim \mathcal{N}(0, \sigma^2 \boldsymbol{\Omega})$, maka penduga maksimum likelihood (MLE) untuk $\boldsymbol{\beta}$ adalah

$$\hat{\boldsymbol{\beta}}_{\text{MLE}} = \hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{y}, \quad (2.389)$$

dan MLE untuk σ^2 (dengan $\boldsymbol{\Omega}$ diketahui) adalah

$$\hat{\sigma}_{\text{MLE}}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top \boldsymbol{\Omega}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \quad (2.390)$$

Bukti. Di bawah normalitas galat, fungsi *likelihood* untuk $\boldsymbol{\beta}$ dan σ^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 σ^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 ℓ untuk β dengan turunan pertama menghasilkan

$$\begin{aligned} \frac{\partial \ell(\beta, \sigma^2 | \mathbf{y})}{\partial \beta} &= \frac{2}{2\sigma^2} \mathbf{X}^\top \Omega^{-1} (\mathbf{y} - \mathbf{X}\beta) \\ 0 &= \frac{1}{\sigma^2} \mathbf{X}^\top \Omega^{-1} \mathbf{y} - \frac{1}{\sigma^2} \mathbf{X}^\top \Omega^{-1} \mathbf{X} \hat{\beta} \\ \frac{1}{\sigma^2} \mathbf{X}^\top \Omega^{-1} \mathbf{X} \hat{\beta} &= \frac{1}{\sigma^2} \mathbf{X}^\top \Omega^{-1} \mathbf{y} \\ \hat{\beta} &= (\mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Omega^{-1} \mathbf{y}, \end{aligned} \quad (2.391)$$

sehingga $\hat{\beta}_{MLE} = \hat{\beta}_{GLS}$. Selanjutnya, turunan terhadap σ^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 $\hat{\sigma}_{MLE}^2$ di atas. ■

Dalam analisis nyata, Ω biasanya jarang diketahui. *Feasible GLS* (FGLS) memperkirakan Ω , misalnya dari residu OLS atau model varian parametrik, lalu menyusun $\widehat{\mathbf{W}} = \widehat{\Omega}^{-1}$ dan menghitung $\hat{\beta}_{FGLS} = (\mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^\top \widehat{\mathbf{W}} \mathbf{y}$. Di bawah kondisi reguler, $\hat{\beta}_{FGLS}$ 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 2.5.19 (Konsistensi Penduga GLS, (Greene, 2002)) 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 Ω 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}_Ω matriks definit positif,

(ii) $\frac{1}{n} \mathbf{X}^\top \Omega^{-1} \varepsilon \xrightarrow{p} \mathbf{0}$,

maka penduga GLS $\hat{\beta}_{GLS} = (\mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Omega^{-1} \mathbf{y}$ bersifat konsisten, yaitu

$$\hat{\beta}_{GLS} \xrightarrow{p} \boldsymbol{\beta}. \quad (2.392)$$

Bukti. Penduga GLS dapat dituliskan sebagai

$$\begin{aligned}\hat{\beta}_{GLS} &= (\mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Omega^{-1} \mathbf{y} \\ &= (\mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Omega^{-1} (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{\beta} + (\mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Omega^{-1} \boldsymbol{\varepsilon} \\ &= \boldsymbol{\beta} + \left(\frac{1}{n} \mathbf{X}^\top \Omega^{-1} \mathbf{X} \right)^{-1} \left(\frac{1}{n} \mathbf{X}^\top \Omega^{-1} \boldsymbol{\varepsilon} \right).\end{aligned} \quad (2.393)$$

Berdasarkan kondisi (i), $\frac{1}{n} \mathbf{X}^\top \Omega^{-1} \mathbf{X} \xrightarrow{p} \mathbf{Q}_\Omega$ dengan \mathbf{Q}_Ω definit positif. Dengan teorema pemetaan kontinu, $(\frac{1}{n} \mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1} \xrightarrow{p} \mathbf{Q}_\Omega^{-1}$. Berdasarkan kondisi (ii), $\frac{1}{n} \mathbf{X}^\top \Omega^{-1} \boldsymbol{\varepsilon} \xrightarrow{p} \mathbf{0}$. Dengan teorema Slutsky, diperoleh

$$\hat{\beta}_{GLS} - \boldsymbol{\beta} = \left(\frac{1}{n} \mathbf{X}^\top \Omega^{-1} \mathbf{X} \right)^{-1} \left(\frac{1}{n} \mathbf{X}^\top \Omega^{-1} \boldsymbol{\varepsilon} \right) \xrightarrow{p} \mathbf{Q}_\Omega^{-1} \cdot \mathbf{0} = \mathbf{0}, \quad (2.394)$$

sehingga $\hat{\beta}_{GLS} \xrightarrow{p} \boldsymbol{\beta}$. ■

Teorema 2.5.20 (Konsistensi Penduga Variansi GLS, (Greene, 2002)) Di bawah kondisi-kondisi konsistensi GLS, penduga variansi

$$s_{GLS}^2 = \frac{1}{n-p-1} (\mathbf{y} - \mathbf{X}\hat{\beta}_{GLS})^\top \Omega^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta}_{GLS}) \quad (2.395)$$

bersifat konsisten, yaitu $s_{GLS}^2 \xrightarrow{p} \sigma^2$.

Bukti. Definisikan residual GLS $\hat{\boldsymbol{\varepsilon}}_{GLS} = \mathbf{y} - \mathbf{X}\hat{\beta}_{GLS} = \mathbf{M}_\Omega \mathbf{y}$ dengan $\mathbf{M}_\Omega = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Omega^{-1}$. Maka

$$\hat{\boldsymbol{\varepsilon}}_{GLS} = \mathbf{M}_\Omega (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{M}_\Omega \boldsymbol{\varepsilon}, \quad (2.396)$$

sebab $M_\Omega \mathbf{X} = \mathbf{0}$. Oleh karena itu,

$$\hat{\boldsymbol{\varepsilon}}_{\text{GLS}}^\top \boldsymbol{\Omega}^{-1} \hat{\boldsymbol{\varepsilon}}_{\text{GLS}} = \boldsymbol{\varepsilon}^\top M_\Omega^\top \boldsymbol{\Omega}^{-1} M_\Omega \boldsymbol{\varepsilon}. \quad (2.397)$$

Dengan menggunakan sifat ekspektasi dan fakta bahwa $\text{tr}(M_\Omega^\top \boldsymbol{\Omega}^{-1} M_\Omega \cdot \sigma^2 \boldsymbol{\Omega}) = \sigma^2(n - p - 1)$, serta Hukum Bilangan Besar, diperoleh $s_{\text{GLS}}^2 \xrightarrow{p} \sigma^2$. ■

Teorema 2.5.21 (Normalitas Asimtotik Penduga GLS, (Greene, 2002)) Misalkan kondisi-kondisi konsistensi terpenuhi dan tambahan kondisi berikut dipenuhi:

- (i) $\frac{1}{\sqrt{n}} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}_\Omega)$ (kondisi Lindeberg–Feller terpenuhi).

Maka penduga GLS memiliki distribusi asimtotik

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{\text{GLS}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}_\Omega^{-1}). \quad (2.398)$$

Bukti. Dari pembuktian konsistensi, dapat dituliskan

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\beta}}_{\text{GLS}} - \boldsymbol{\beta}) &= \sqrt{n} \left(\frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X} \right)^{-1} \left(\frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} \right) \\ &= \left(\frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X} \right)^{-1} \left(\frac{1}{\sqrt{n}} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} \right). \end{aligned} \quad (2.399)$$

Berdasarkan kondisi (i), $\frac{1}{\sqrt{n}} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}_\Omega)$. Dengan teorema Slutsky, sebab $\left(\frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X} \right)^{-1} \xrightarrow{p} \mathbf{Q}_\Omega^{-1}$, maka

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{\text{GLS}} - \boldsymbol{\beta}) \xrightarrow{d} \mathbf{Q}_\Omega^{-1} \cdot \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}_\Omega) = \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Q}_\Omega^{-1}). \quad (2.400)$$

■

Akibat 2.5.22 (Distribusi Asimtotik dengan Variansi Diestimasi) Dengan penduga variansi konsisten $s_{\text{GLS}}^2 \xrightarrow{p} \sigma^2$, berlaku

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{\text{GLS}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \sigma^2 \left(\frac{1}{n} \mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X} \right)^{-1}\right), \quad (2.401)$$

dan penduga matriks kovarians asimtotik yang konsisten adalah

$$\widehat{\text{Var}}(\hat{\beta}_{GLS}) = s_{GLS}^2 (\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1}. \quad (2.402)$$

Teorema 2.5.23 (Distribusi Uji Statistik untuk GLS/WLS, (Greene, 2002)) *Distribusi asimtotik dari uji-uji statistik untuk model GLS/WLS analog dengan kasus OLS. Secara spesifik:*

(a) **Uji-t:** Statistik uji

$$t_k = \frac{\hat{\beta}_{k,GLS} - \beta_k^0}{s_{GLS} \sqrt{[(\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1}]_{kk}}} \quad (2.403)$$

memiliki distribusi t_{n-p-1} secara eksak di bawah normalitas galat, dan $\mathcal{N}(0, 1)$ secara asimtotik.

(b) **Uji-F:** Statistik uji untuk hipotesis linear $H_0 : \mathbf{R}\beta = \mathbf{r}$

$$F = \frac{(\mathbf{R}\hat{\beta}_{GLS} - \mathbf{r})^\top [\mathbf{R}(\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{R}^\top]^{-1} (\mathbf{R}\hat{\beta}_{GLS} - \mathbf{r})}{q \cdot s_{GLS}^2} \quad (2.404)$$

memiliki distribusi $F_{q,n-p-1}$ secara eksak dan $qF \xrightarrow{d} \chi_q^2$ secara asimtotik.

(c) **Uji Wald:** Statistik Wald

$$W = (\mathbf{R}\hat{\beta}_{GLS} - \mathbf{r})^\top \left[\mathbf{R} \widehat{\text{Var}}(\hat{\beta}_{GLS}) \mathbf{R}^\top \right]^{-1} (\mathbf{R}\hat{\beta}_{GLS} - \mathbf{r}) \quad (2.405)$$

memiliki distribusi $W \xrightarrow{d} \chi_q^2$ di bawah H_0 .

Pembuktian distribusi-distribusi ini mengikuti pola yang serupa dengan kasus OLS (lihat Teorema distribusi uji-t, uji-F, dan uji Wald pada Subbab Teori Asimtotik untuk OLS), dengan mengganti $(\mathbf{X}^\top \mathbf{X})^{-1}$ menjadi $(\mathbf{X}^\top \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1}$ dan s^2 menjadi s_{GLS}^2 .

Proposisi 2.5.24 (Kesetaraan Asimtotik Uji-Uji Klasik untuk GLS) *Di bawah hipotesis nol $H_0 : \mathbf{R}\beta = \mathbf{r}$ dan kondisi regularitas, ketiga statistik uji—Wald (W), rasio kemungkinan (LR), dan pengali Lagrange (LM)—untuk model GLS bersifat*

setara secara asimtotik. Ketiganya memiliki distribusi limit χ_q^2 dan memberikan kesimpulan inferensi yang identik secara asimtotik.

2.6. Pemodelan Polinomial Lokal

Pemodelan polinomial lokal (*local polynomial modeling*) merupakan teknik nonparametrik yang fleksibel untuk mengestimasi fungsi regresi tanpa mengasumsikan bentuk fungsional tertentu secara global. Pendekatan ini menjadi fondasi penting bagi metode regresi terboboti geografis (GWR) yang akan dibahas pada bagian selanjutnya. Pembahasan dalam bagian ini mengikuti kerangka yang dikembangkan oleh ? dan ?.

Definisi 2.6.1 (Model Regresi Nonparametrik) Pertimbangkan model regresi nonparametrik

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (2.406)$$

dengan $m(\cdot)$ adalah fungsi regresi yang tidak diketahui, X_i adalah variabel prediktor, Y_i adalah variabel respons, dan ε_i adalah galat acak dengan $\mathbb{E}[\varepsilon_i | X_i] = 0$ dan $\text{Var}(\varepsilon_i | X_i) = \sigma^2(X_i)$.

Definisi 2.6.2 (Aproksimasi Polinomial Lokal) Untuk mengestimasi $m(x_0)$ pada titik target x_0 , fungsi regresi $m(x)$ diaproksimasi secara lokal menggunakan eksponsi Taylor orde- p di sekitar x_0 :

$$m(x) \approx m(x_0) + m'(x_0)(x - x_0) + \frac{m''(x_0)}{2!}(x - x_0)^2 + \dots + \frac{m^{(p)}(x_0)}{p!}(x - x_0)^p. \quad (2.407)$$

Dengan mendefinisikan $\beta_j = \frac{m^{(j)}(x_0)}{j!}$ untuk $j = 0, 1, \dots, p$, aproksimasi lokal dapat dituliskan sebagai

$$m(x) \approx \sum_{j=0}^p \beta_j (x - x_0)^j. \quad (2.408)$$

Definisi 2.6.3 (Estimator Polinomial Lokal) Estimator polinomial lokal $\hat{m}(x_0)$

diperoleh dengan menyelesaikan masalah optimisasi berikut:

$$\min_{\beta_0, \beta_1, \dots, \beta_p} \sum_{i=1}^n K_h(X_i - x_0) \left[Y_i - \sum_{j=0}^p \beta_j (X_i - x_0)^j \right]^2, \quad (2.409)$$

dengan $K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$ adalah fungsi kernel dengan parameter lebar pita (bandwidth) $h > 0$. Estimator polinomial lokal pada titik x_0 didefinisikan sebagai $\hat{m}(x_0) = \hat{\beta}_0(x_0)$, di mana $\hat{\beta}_0(x_0)$ adalah solusi optimal dari masalah di atas.

Contoh 2.6.4 Dua kasus khusus yang paling umum digunakan adalah sebagai berikut.

- (a) **Estimator Nadaraya–Watson** ($p = 0$): Dengan $p = 0$, diperoleh estimator rata-rata lokal

$$\hat{m}(x_0) = \frac{\sum_{i=1}^n K_h(X_i - x_0) Y_i}{\sum_{i=1}^n K_h(X_i - x_0)}. \quad (2.410)$$

- (b) **Estimator Linear Lokal** ($p = 1$): Dengan $p = 1$, model lokal adalah $m(x) \approx \beta_0 + \beta_1(x - x_0)$, dan estimator $\hat{m}(x_0) = \hat{\beta}_0(x_0)$ diperoleh dari regresi linear tertimbang lokal.

2.6.1. Asumsi Model Polinomial Lokal

Untuk memperoleh sifat-sifat asimtotik estimator polinomial lokal, diperlukan asumsi-asumsi regularitas berikut.

- (A1) **Kehalusan fungsi regresi:** Fungsi regresi $m(x)$ memiliki turunan kontinu hingga orde $(p + 1)$ di sekitar x_0 .
- (A2) **Kepadatan prediktor:** Variabel prediktor X memiliki fungsi kepadatan $f(x)$ yang kontinu dan positif di x_0 , yaitu $f(x_0) > 0$.
- (A3) **Struktur galat:** Galat ε_i memenuhi $\mathbb{E}[\varepsilon_i | X_i] = 0$ dan $\text{Var}(\varepsilon_i | X_i) = \sigma^2(X_i)$ dengan $\sigma^2(x)$ kontinu di x_0 .
- (A4) **Sifat kernel:** Fungsi kernel $K(\cdot)$ adalah fungsi kepadatan simetris dengan momen hingga, yaitu $K(u) \geq 0$, $\int K(u) du = 1$, $\int uK(u) du = 0$, dan

$$\mu_2(K) = \int u^2 K(u) du < \infty.$$

(A5) **Kondisi bandwidth:** 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. Pendugaan Model Polinomial Lokal

Definisi 2.6.5 (Regresi Tertimbang Lokal) *Regresi tertimbang lokal* (locally weighted least squares, LWLS) pada titik x_0 menyelesaikan masalah optimisasi berikut:

$$\min_{\beta} \sum_{i=1}^n K_h(X_i - x_0) [Y_i - z_i^\top \beta]^2, \quad (2.411)$$

dengan $\beta = (\beta_0, \beta_1, \dots, \beta_p)^\top$ dan $z_i = (1, X_i - x_0, (X_i - x_0)^2, \dots, (X_i - x_0)^p)^\top$. Solusi optimal $\hat{\beta}(x_0)$ memberikan estimator polinomial lokal $\hat{m}(x_0) = \hat{\beta}_0(x_0)$.

Teorema 2.6.6 Definisikan matriks desain lokal \mathbf{X}_{x_0} dan matriks bobot \mathbf{W}_{x_0} sebagaimana

$$\mathbf{X}_{x_0} = \begin{pmatrix} 1 & (X_1 - x_0) & (X_1 - x_0)^2 & \cdots & (X_1 - x_0)^p \\ 1 & (X_2 - x_0) & (X_2 - x_0)^2 & \cdots & (X_2 - x_0)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (X_n - x_0) & (X_n - x_0)^2 & \cdots & (X_n - x_0)^p \end{pmatrix} \in \mathbb{R}^{n \times (p+1)}, \quad (2.412)$$

dan

$$\mathbf{W}_{x_0} = \text{diag}(K_h(X_1 - x_0), K_h(X_2 - x_0), \dots, K_h(X_n - x_0)) \in \mathbb{R}^{n \times n}. \quad (2.413)$$

Maka estimator polinomial lokal diberikan oleh

$$\hat{\beta}(x_0) = (\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{y}, \quad (2.414)$$

dengan $\mathbf{y} = (Y_1, Y_2, \dots, Y_n)^\top$. Khususnya, estimator untuk $m(x_0)$ adalah

$$\hat{m}(x_0) = \mathbf{e}_1^\top \hat{\beta}(x_0) = \mathbf{e}_1^\top (\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{y}, \quad (2.415)$$

dengan $\mathbf{e}_1 = (1, 0, \dots, 0)^\top \in \mathbb{R}^{p+1}$.

Bukti. Masalah minimisasi LWLS dapat dituliskan dalam bentuk matriks sebagai

$$\min_{\beta} (\mathbf{y} - \mathbf{X}_{x_0} \beta)^\top \mathbf{W}_{x_0} (\mathbf{y} - \mathbf{X}_{x_0} \beta). \quad (2.416)$$

Turunan pertama terhadap β adalah

$$\frac{\partial}{\partial \beta} (\mathbf{y} - \mathbf{X}_{x_0} \beta)^\top \mathbf{W}_{x_0} (\mathbf{y} - \mathbf{X}_{x_0} \beta) = -2 \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} (\mathbf{y} - \mathbf{X}_{x_0} \beta). \quad (2.417)$$

Menyamakan dengan nol memberikan persamaan normal tertimbang

$$\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0} \hat{\beta}(x_0) = \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{y}. \quad (2.418)$$

Jika $\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0}$ nonsingular, solusinya adalah

$$\hat{\beta}(x_0) = (\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{y}. \quad (2.419)$$

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 2.6.7 (Representasi Bobot Ekuivalen) Estimator $\hat{m}(x_0)$ dapat dituliskan sebagai kombinasi linear tertimbang dari respons:

$$\hat{m}(x_0) = \sum_{i=1}^n w_i(x_0) Y_i, \quad (2.420)$$

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

$$\hat{m}(x_0) = \mathbf{e}_1^\top (\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{y}. \quad (2.421)$$

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

$$\mathbf{y} = \sum_{i=1}^n Y_i \mathbf{e}_i, \quad (2.422)$$

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

$$\begin{aligned} \hat{m}(x_0) &= \mathbf{e}_1^\top (\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \left(\sum_{i=1}^n Y_i \mathbf{e}_i \right) \\ &= \sum_{i=1}^n Y_i \cdot \mathbf{e}_1^\top (\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{e}_i, \end{aligned} \quad (2.423)$$

dengan pertukaran urutan penjumlahan dan perkalian matriks dijamin oleh linearitas operasi matriks.

Dengan mendefinisikan bobot efektif sebagai

$$w_i(x_0) = \mathbf{e}_1^\top (\mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{X}_{x_0})^{-1} \mathbf{X}_{x_0}^\top \mathbf{W}_{x_0} \mathbf{e}_i, \quad (2.424)$$

diperoleh representasi yang diinginkan

$$\hat{m}(x_0) = \sum_{i=1}^n w_i(x_0) Y_i. \quad (2.425)$$

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. ■

Teorema 2.6.8 (Bias Estimator Polinomial Lokal) *Di bawah asumsi regularitas (A1)–(A5), bias bersyarat dari estimator polinomial lokal orde- p adalah*

$$\text{Bias}(\hat{m}(x_0) \mid X_1, \dots, X_n) = \frac{m^{(p+1)}(x_0)}{(p+1)!} h^{p+1} \mu_{p+1}(K) + o(h^{p+1}), \quad (2.426)$$

dengan $\mu_j(K) = \int u^j K(u) du$ adalah momen ke- j dari kernel K .

Khususnya, untuk estimator linear lokal ($p = 1$) dengan kernel simetris, berlaku

$$\text{Bias}(\hat{m}(x_0)) = \frac{1}{2} m''(x_0) h^2 \mu_2(K) + o(h^2). \quad (2.427)$$

Bukti. Untuk estimator linear lokal ($p = 1$), ekspansi Taylor dari $m(X_i)$ di sekitar x_0 adalah

$$m(X_i) = m(x_0) + m'(x_0)(X_i - x_0) + \frac{1}{2} m''(x_0)(X_i - x_0)^2 + O((X_i - x_0)^3). \quad (2.428)$$

Substitusi ke dalam ekspektasi estimator dan menggunakan sifat *kernel* simetris memberikan

$$\mathbb{E}[\hat{m}(x_0)] = m(x_0) + \frac{1}{2} m''(x_0) \mathbb{E} \left[\sum_{i=1}^n w_i(x_0)(X_i - x_0)^2 \right] + O(h^3). \quad (2.429)$$

Dengan analisis asimtotik dan fakta bahwa $\sum_{i=1}^n w_i(x_0)(X_i - x_0)^2 \approx h^2 \mu_2(K)$ untuk n besar dan h kecil, diperoleh hasil yang diinginkan. ■

Teorema 2.6.9 (Variansi Estimator Polinomial Lokal) *Di bawah asumsi regularitas (A1)–(A5), variansi bersyarat dari estimator polinomial lokal orde- p adalah*

$$\text{Var}(\hat{m}(x_0) \mid X_1, \dots, X_n) = \frac{\sigma^2(x_0) R(K)}{nh f(x_0)} + o\left(\frac{1}{nh}\right), \quad (2.430)$$

dengan $R(K) = \int K^2(u) du$ adalah kekasaran (roughness) dari kernel K .

Bukti. Variansi estimator diberikan oleh

$$\text{Var}(\hat{m}(x_0) \mid \mathbf{X}) = \sum_{i=1}^n w_i^2(x_0) \text{Var}(Y_i \mid X_i) = \sum_{i=1}^n w_i^2(x_0) \sigma^2(X_i). \quad (2.431)$$

Dengan analisis asimtotik dan menggunakan aproksimasi integral Riemann, dapat ditunjukkan bahwa

$$\begin{aligned} \sum_{i=1}^n w_i^2(x_0) \sigma^2(X_i) &\approx \frac{\sigma^2(x_0)}{n^2 h^2 f^2(x_0)} \sum_{i=1}^n K^2 \left(\frac{X_i - x_0}{h} \right) \\ &\approx \frac{\sigma^2(x_0)}{nhf(x_0)} \int K^2(u) du = \frac{\sigma^2(x_0) R(K)}{nhf(x_0)}. \end{aligned} \quad (2.432)$$

■

Teorema 2.6.10 (Rataan Kuadrat Galat (Mean Squared Error, MSE)) *MSE estimator polinomial lokal orde-p di titik x_0 adalah*

$$\text{MSE}(\hat{m}(x_0)) = \text{Bias}^2(\hat{m}(x_0)) + \text{Var}(\hat{m}(x_0)). \quad (2.433)$$

Untuk estimator linear lokal ($p = 1$), MSE asimtotik diberikan oleh

$$\text{MSE}(\hat{m}(x_0)) = \frac{1}{4} (m''(x_0))^2 h^4 \mu_2^2(K) + \frac{\sigma^2(x_0) R(K)}{nhf(x_0)} + o\left(h^4 + \frac{1}{nh}\right). \quad (2.434)$$

Akibat 2.6.11 (Bandwidth Optimal) Bandwidth optimal yang meminimumkan MSE asimtotik untuk estimator linear lokal diperoleh dengan menyeimbangkan bias dan variansi. Dengan mengambil turunan MSE terhadap h dan menyamakan dengan nol, diperoleh

$$h_{\text{opt}} = \left(\frac{\sigma^2(x_0) R(K)}{nf(x_0) (m''(x_0))^2 \mu_2^2(K)} \right)^{1/5} = O(n^{-1/5}). \quad (2.435)$$

Dengan bandwidth optimal ini, MSE optimal adalah

$$\text{MSE}_{\text{opt}}(\hat{m}(x_0)) = O(n^{-4/5}). \quad (2.436)$$

2.6.3. Teori Asimtotik Estimator Polinomial Lokal

Teorema 2.6.12 (Konsistensi Estimator Polinomial Lokal) *Di bawah asumsi regularitas (A1)–(A5), estimator polinomial lokal $\hat{m}(x_0)$ bersifat konsisten, yaitu*

$$\hat{m}(x_0) \xrightarrow{P} m(x_0) \quad \text{ketika } n \rightarrow \infty. \quad (2.437)$$

Bukti. Dari hasil bias dan variansi, berlaku

$$\text{Bias}(\hat{m}(x_0)) = O(h^{p+1}) \rightarrow 0 \quad \text{sebab } h \rightarrow 0, \quad (2.438)$$

$$\text{Var}(\hat{m}(x_0)) = O\left(\frac{1}{nh}\right) \rightarrow 0 \quad \text{sebab } nh \rightarrow \infty. \quad (2.439)$$

Dengan demikian, $\text{MSE}(\hat{m}(x_0)) \rightarrow 0$, yang mengimplikasikan konvergensi dalam probabilitas. ■

Teorema 2.6.13 (Normalitas Asimtotik Estimator Polinomial Lokal) *Di bawah asumsi regularitas (A1)–(A5) dan dengan asumsi tambahan bahwa galat memiliki momen keempat terbatas, estimator polinomial lokal memenuhi*

$$\sqrt{nh}(\hat{m}(x_0) - m(x_0) - \text{Bias}(\hat{m}(x_0))) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2(x_0)R(K)}{f(x_0)}\right). \quad (2.440)$$

Secara ekuivalen, dengan mendefinisikan bias asimtotik $b(x_0) = \frac{m^{(p+1)}(x_0)}{(p+1)!}h^{p+1}\mu_{p+1}(K)$, berlaku

$$\sqrt{nh}(\hat{m}(x_0) - m(x_0) - b(x_0)) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2(x_0)R(K)}{f(x_0)}\right). \quad (2.441)$$

Bukti. Estimator $\hat{m}(x_0)$ dapat dituliskan sebagai

$$\hat{m}(x_0) = \sum_{i=1}^n w_i(x_0)Y_i = \sum_{i=1}^n w_i(x_0)(m(X_i) + \varepsilon_i). \quad (2.442)$$

Dengan memisahkan komponen deterministik dan stokastik, diperoleh

$$\hat{m}(x_0) - m(x_0) = \underbrace{\sum_{i=1}^n w_i(x_0) (m(X_i) - m(x_0))}_{\text{bias}} + \underbrace{\sum_{i=1}^n w_i(x_0) \varepsilon_i}_{\text{komponen stokastik}} . \quad (2.443)$$

Komponen stokastik $\sum_{i=1}^n w_i(x_0) \varepsilon_i$ adalah kombinasi linear dari variabel acak independen dengan rataan nol. Dengan Teorema Lindeberg–Feller, dapat ditunjukkan bahwa setelah normalisasi dengan \sqrt{nh} , komponen ini konvergen dalam distribusi ke normal.

Secara spesifik, definisikan $Z_n = \sqrt{nh} \sum_{i=1}^n w_i(x_0) \varepsilon_i$. Variansi dari Z_n adalah

$$\text{Var}(Z_n) = nh \sum_{i=1}^n w_i^2(x_0) \sigma^2(X_i) \rightarrow \frac{\sigma^2(x_0) R(K)}{f(x_0)} . \quad (2.444)$$

Kondisi Lindeberg dapat diverifikasi dengan menggunakan fakta bahwa bobot $w_i(x_0)$ mengecil secara seragam dan momen keempat galat terbatas. Dengan demikian, Teorema Lindeberg–Feller menjamin

$$Z_n \xrightarrow{d} \mathcal{N} \left(0, \frac{\sigma^2(x_0) R(K)}{f(x_0)} \right) . \quad (2.445)$$

■

Teorema 2.6.14 (Interval Konfidensi Asimtotik) *Interval konfidensi asimtotik tingkat $(1 - \alpha) \times 100\%$ untuk $m(x_0)$ diberikan oleh*

$$\hat{m}(x_0) \pm z_{\alpha/2} \sqrt{\frac{\hat{\sigma}^2(x_0) R(K)}{nh \hat{f}(x_0)}}, \quad (2.446)$$

dengan $z_{\alpha/2}$ adalah kuantil $(1 - \alpha/2)$ dari distribusi normal standar, $\hat{\sigma}^2(x_0)$ adalah estimator konsisten untuk $\sigma^2(x_0)$, dan $\hat{f}(x_0)$ adalah estimator kepadatan kernel untuk $f(x_0)$.

Interval ini valid secara asimtotik jika bias diabaikan, yang terjadi ketika $nh^{2(p+1)+1} \rightarrow 0$ (undersmoothing) atau ketika koreksi bias eksplisit diterapkan.

Proposisi 2.6.15 (Pemilihan Bandwidth dengan Cross-Validation) Bandwidth optimal dapat dipilih secara adaptif menggunakan leave-one-out cross-validation dengan meminimumkan

$$\text{CV}(h) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{m}_{-i}(X_i; h))^2, \quad (2.447)$$

dengan $\hat{m}_{-i}(X_i; h)$ adalah estimator leave-one-out yang dihitung tanpa observasi ke- i . Secara asimtotik, bandwidth yang dipilih dengan cross-validation mendekati bandwidth optimal yang meminimumkan integrated mean squared error (IMSE).

2.7. Analisis Regresi Spasial

Analisis regresi spasial merupakan perluasan dari analisis regresi klasik yang secara eksplisit memperhitungkan aspek geografis atau lokasi dalam pemodelan. Terdapat dua fenomena utama dalam data spasial yang perlu diperhatikan, yaitu dependensi spasial (*spatial dependence*) dan keragaman spasial (*spatial heterogeneity*). Kedua konsep ini memiliki karakteristik yang berbeda dan memerlukan pendekatan pemodelan yang berbeda pula (Fotheringham dkk., 2002b; ?; ?).

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 β tidak konstan di seluruh wilayah pengamatan, melainkan merupakan fungsi dari lokasi geografis, yaitu $\beta = \beta(s)$ dengan s menyatakan lokasi. Fenomena ini mengin-

dikasikan bahwa proses yang mendasari hubungan antarvariabel berbeda di setiap wilayah.

Tabel 2.1 Perbandingan Dependensi Spasial dan Keragaman Spasial

IYY
Aspek Dependensi Spasial Keragaman Spasial
Definisi Korelasi antarlokasi pada variabel yang sama Variasi hubungan struktural antarlokasi
Fokus Hubungan antar-observasi Hubungan antar-variabel
Manifestasi $\text{Cov}(Y_i, Y_j) \neq 0$ untuk lokasi berdekatan $\beta_i \neq \beta_j$ untuk lokasi berbeda
Asumsi yang dilanggar Independensi galat Homogenitas parameter Model penanganan SAR, SEM, SDM GWR, GTWR
Contoh Harga rumah di lokasi berdekatan saling memengaruhi Pengaruh pendidikan terhadap pendapatan berbeda di kota dan desa

Kedua fenomena ini dapat terjadi secara bersamaan dalam data spasial. Sebagai ilustrasi, pada analisis harga properti, dependensi spasial terjadi ketika harga rumah di suatu lokasi dipengaruhi oleh harga rumah di sekitarnya (efek *spillover*), sedangkan keragaman spasial terjadi ketika faktor-faktor yang memengaruhi harga rumah (seperti luas tanah atau jarak ke pusat kota) memiliki pengaruh yang berbeda di wilayah perkotaan dibandingkan dengan wilayah pinggiran.

2.7.1. Regresi dengan Dependensi Spasial

Dependensi spasial (*spatial dependence*) atau autokorelasi spasial terjadi ketika nilai suatu variabel di suatu lokasi berkorelasi dengan nilai variabel yang sama di lokasi lain yang berdekatan secara geografis. Konsep ini melanggar asumsi independensi dalam regresi linear klasik dan memerlukan perlakuan khusus dalam pemodelan ekonometrika spasial (??).

Definisi 2.7.1 (Dependensi Spasial) Misalkan Y_i merupakan nilai variabel di lokasi i dan Y_j merupakan nilai variabel di lokasi tetangga j . Dependensi spasial

terjadi jika

$$\text{Cov}(Y_i, Y_j) \neq 0, \quad \text{untuk lokasi } i \neq j \text{ yang berdekatan}, \quad (2.448)$$

yang dapat diukur menggunakan statistik Moran's I , yaitu

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_{i=1}^n (Y_i - \bar{Y})^2}, \quad (2.449)$$

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

$$y_i = \rho \sum_{j=1}^n w_{ij} y_j + \sum_{k=1}^p \beta_k x_{ik} + \varepsilon_i, \quad (2.450)$$

atau dalam bentuk matriksnya adalah

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2.451)$$

dengan ρ adalah parameter autokorelasi spasial dan \mathbf{W} adalah matriks bobot spasial.

Definisi 2.7.3 (Spatial Error Model (SEM)) Model SEM mengasumsikan bahwa dependensi spasial terjadi pada komponen galat, yaitu

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{u}, \quad \mathbf{u} = \lambda \mathbf{W} \mathbf{u} + \boldsymbol{\varepsilon}, \quad (2.452)$$

dengan λ adalah parameter autokorelasi spasial pada galat dan $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$.

Definisi 2.7.4 (Spatial Durbin Model (SDM)) Model SDM menggabungkan lag spasial dari variabel dependen dan independen, yaitu

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \mathbf{W} \mathbf{X} \boldsymbol{\delta} + \boldsymbol{\varepsilon}, \quad (2.453)$$

dengan $\boldsymbol{\delta}$ adalah vektor parameter untuk lag spasial variabel independen.

2.7.2. Regresi dengan Heterogenitas 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., 2002b).

Definisi 2.7.5 (Heterogenitas Spasial) Heterogenitas spasial terjadi ketika parameter regresi $\boldsymbol{\beta}$ bervariasi menurut lokasi geografis, yaitu

$$\boldsymbol{\beta} = \boldsymbol{\beta}(s_i) = \boldsymbol{\beta}(u_i, v_i), \quad (2.454)$$

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.

Model utama untuk menangkap heterogenitas spasial adalah *Geographically Weighted Regression* (GWR) yang memungkinkan koefisien regresi bervariasi secara kontinu di ruang geografis.

Definisi 2.7.6 (Geographically Weighted Regression (GWR)) Model GWR memungkinkan koefisien regresi bervariasi menurut lokasi:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i)x_{ik} + \varepsilon_i, \quad (2.455)$$

dengan $\beta_k(u_i, v_i)$ adalah koefisien regresi untuk variabel k di lokasi (u_i, v_i) .

Estimasi parameter GWR dilakukan menggunakan *locally weighted least squares* (LWLS), yang memberikan bobot lebih besar pada observasi yang berdekatan secara geografis.

Teorema 2.7.7 (Estimator LWLS untuk GWR) Untuk setiap lokasi i , koefisien lokal diestimasi dengan:

$$\hat{\beta}(u_i, v_i) = (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i \mathbf{y}, \quad (2.456)$$

dengan \mathbf{W}_i adalah matriks diagonal bobot dengan elemen

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{h^2}\right), \quad (2.457)$$

dengan d_{ij} adalah jarak antara lokasi i dan j , dan h adalah bandwidth.

Bukti. Pertama-tama didefinisikan masalah optimisasi lokal. Untuk setiap lokasi (u_i, v_i) , akan diminimumkan jumlah kuadrat residual tertimbang lokal. Misalkan model GWR diberikan oleh $y_j = \mathbf{x}_j^\top \beta(u_i, v_i) + \varepsilon_j$ untuk observasi $j = 1, 2, \dots, n$.

Fungsi objektif yang akan diminimumkan adalah

$$S_i(\beta) = \sum_{j=1}^n w_{ij} (y_j - \mathbf{x}_j^\top \beta)^2, \quad (2.458)$$

dengan w_{ij} adalah bobot yang mencerminkan kedekatan geografis antara lokasi i dan j .

Untuk mencari nilai β yang meminimumkan $S_i(\beta)$, diambil turunan pertama terhadap β dan disamakan dengan nol. Dapat menulis ulang $S_i(\beta)$ dalam bentuk matriks sebagai

$$S_i(\beta) = (\mathbf{y} - \mathbf{X}\beta)^\top \mathbf{W}_i (\mathbf{y} - \mathbf{X}\beta), \quad (2.459)$$

dengan $\mathbf{W}_i = \text{diag}(w_{i1}, w_{i2}, \dots, w_{in})$ adalah matriks diagonal berukuran $n \times n$.

Turunan pertama dari $S_i(\boldsymbol{\beta})$ terhadap $\boldsymbol{\beta}$ adalah

$$\frac{\partial S_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}} [(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{W}_i (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})] \quad (2.460)$$

$$= -2\mathbf{X}^\top \mathbf{W}_i (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (2.461)$$

Selanjutnya, dengan menyamakan turunan pertama dengan nol, diperoleh

$$-2\mathbf{X}^\top \mathbf{W}_i (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0}, \quad (2.462)$$

sehingga

$$\mathbf{X}^\top \mathbf{W}_i \mathbf{y} - \mathbf{X}^\top \mathbf{W}_i \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{0} \quad (2.463)$$

$$\mathbf{X}^\top \mathbf{W}_i \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^\top \mathbf{W}_i \mathbf{y} \quad (2.464)$$

Dengan asumsi bahwa matriks $\mathbf{X}^\top \mathbf{W}_i \mathbf{X}$ adalah non-singular (invertibel), kita dapat menyelesaikan untuk $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_i \mathbf{y} \quad (2.465)$$

■

Perlu diperhatikan bahwa bobot dalam GWR merupakan bobot diagonal yang berarti bahwa model ini merupakan perluasan dari model regresi linear ter-boboti (WLS) yang telah dibahas sebelumnya.

Definisi 2.7.8 (Fungsi *kernel* untuk Pembobotan Spasial) Beberapa fungsi kernel yang umum digunakan dalam GWR adalah sebagai berikut.

(a) Kernel Gaussian, yaitu $w_{ij} = \exp\left(-\frac{1}{2} \left(\frac{d_{ij}}{h}\right)^2\right)$.

(b) Kernel Bi-square, yaitu $w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h}\right)^2\right)^2 & \text{jika } d_{ij} < h \\ 0 & \text{jika } d_{ij} \geq h. \end{cases}$

$$(c) \text{ Kernel } Tri\text{-}cube, \text{ yaitu } w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{h}\right)^3\right)^3 & \text{jika } d_{ij} < h \\ 0 & \text{jika } d_{ij} \geq h. \end{cases}$$

Pemilihan *bandwidth* h sangat krusial dalam GWR dan dapat dilakukan melalui *cross-validation* atau validasi silang.

Definisi 2.7.9 (Cross-Validation untuk Bandwidth) Bandwidth optimal h^* dipilih dengan meminimalkan fungsi cross-validation:

$$\text{CV}(h) = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(h))^2, \quad (2.466)$$

dengan $\hat{y}_{\neq i}(h)$ adalah prediksi untuk observasi i menggunakan model yang di-fit tanpa observasi i . Hal ini berarti bahwa

$$h^* = \arg \min_h \text{CV}(h). \quad (2.467)$$

Teorema 2.7.10 (Uji Heterogenitas Spasial) Untuk menguji keberadaan heterogenitas spasial, dapat digunakan uji Breusch-Pagan lokal:

$$\text{BP} = \frac{1}{2} \mathbf{f}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{f}, \quad (2.468)$$

dengan $\mathbf{f} = \mathbf{X}^\top \mathbf{A} \mathbf{e}^2$ dengan \mathbf{A} adalah matriks transformasi spasial dan \mathbf{e}^2 adalah vektor residual kuadrat. Di bawah hipotesis nol bahwa terdapat homoskedastisitas spasial, statistik BP mengikuti distribusi χ^2 dengan derajat bebas p .

Bukti. Dimulai dengan model regresi linear $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ untuk menguji apakah terdapat heterogenitas spasial. Hipotesis yang akan diuji adalah

$$H_0 : \text{Var}(\varepsilon_i) = \sigma^2 \text{ untuk semua } i \text{ (homoskedastisitas)}, \quad (2.469)$$

$$H_1 : \text{Var}(\varepsilon_i) = \sigma^2 h(\mathbf{z}_i^\top \boldsymbol{\alpha}) \text{ (heteroskedastisitas spasial)}, \quad (2.470)$$

dengan \mathbf{z}_i adalah vektor yang berisi karakteristik spasial lokasi i , dan $h(\cdot)$ adalah fungsi yang menghubungkan karakteristik spasial dengan variansi galat. Langkah

pertama adalah menentukan bentuk fungsi $h(\cdot)$. Dipilih $h(t) = \exp(t)$ sehingga variansi galat dapat dituliskan sebagai $\text{Var}(\varepsilon_i) = \sigma^2 \exp(\mathbf{z}_i^\top \boldsymbol{\alpha})$. Dengan mengambil logaritma, didapatkan

$$\log(\text{Var}(\varepsilon_i)) = \log(\sigma^2) + \mathbf{z}_i^\top \boldsymbol{\alpha}.$$

Di bawah hipotesis nol H_0 , parameter $\boldsymbol{\alpha} = \mathbf{0}$, yang berarti tidak ada heterogenitas spasial. Dalam kasus ini, nilai harapan dari kuadrat galat adalah

$$\mathbb{E}[\varepsilon_i^2] = \sigma^2 + \mathbf{z}_i^\top \boldsymbol{\alpha}. \quad (2.471)$$

Sebab ε_i tidak dapat diamati secara langsung, akan digunakan residual OLS e_i sebagai *proxy*. Langkah selanjutnya adalah melakukan regresi *auxiliary* dengan menggunakan kuadrat residual sebagai variabel dependen, yaitu

$$e_i^2 = \gamma_0 + \mathbf{z}_i^\top \boldsymbol{\gamma} + v_i, \quad (2.472)$$

dengan \mathbf{z}_i dapat berupa koordinat spasial, *lag* spasial dari variabel, atau fungsi jarak antar lokasi. Selanjutnya dikonstruksikan statistik uji. Didefinisikan vektor $\mathbf{f} = \mathbf{X}^\top \mathbf{A} \mathbf{e}^2$ dengan \mathbf{A} adalah matriks transformasi yang mengkodekan struktur spasial dan \mathbf{e}^2 adalah vektor yang berisi kuadrat residual. Statistik Breusch-Pagan lokal kemudian didefinisikan sebagai

$$\text{BP}_{\text{Lokal}} = \frac{1}{2} \mathbf{f}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{f}. \quad (2.473)$$

Untuk menunjukkan bahwa statistik ini mengikuti distribusi χ^2 , perlu dibuktikan sifat-sifat distribusinya di bawah H_0 . Dengan asumsi regularitas dan normalitas galat, diperoleh tiga sifat penting, yaitu

- (i) nilai harapan dari \mathbf{f} adalah nol, yaitu $\mathbb{E}[\mathbf{f}] = \mathbf{0}$, karena di bawah H_0 terdapat $\mathbb{E}[e_i^2] = \sigma^2$ yang konstan untuk semua i ;
- (ii) matriks kovarians dari \mathbf{f} adalah $\text{Var}(\mathbf{f}) = 2\sigma^4(\mathbf{X}^\top \mathbf{X})$ ketika galat mengikuti

distribusi normal; dan

- (iii) secara asimtotik, vektor \mathbf{f} konvergen dalam distribusi ke $\mathcal{N}(\mathbf{0}, 2\sigma^4(\mathbf{X}^\top \mathbf{X}))$.

Dengan menggunakan sifat-sifat di atas, dapat ditunjukkan konvergensi distribusi statistik BP_{Lokal} , yaitu

$$\text{BP}_{\text{Lokal}} = \frac{1}{2} \mathbf{f}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{f}, \quad (2.474)$$

$$= \frac{1}{4\sigma^4} \mathbf{f}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{f} \cdot 2\sigma^4, \quad (2.475)$$

$$\xrightarrow{d} \frac{1}{2} \mathbf{Z}^\top \mathbf{Z}, \quad (2.476)$$

dengan $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ adalah vektor acak normal standar berdimensi p .

Langkah terakhir adalah menggunakan sifat distribusi normal standar. Karena $\mathbf{Z}^\top \mathbf{Z} = \sum_{j=1}^p Z_j^2$ dan setiap Z_j^2 mengikuti distribusi χ^2 dengan satu derajat bebas, maka jumlahnya mengikuti $\chi^2(p)$. Oleh karena itu,

$$\text{BP}_{\text{Lokal}} \xrightarrow{d} \frac{1}{2} \chi^2(p). \quad (2.477)$$

Dengan normalisasi yang tepat, diperoleh hasil akhir bahwa $\text{BP}_{\text{Lokal}} \sim \chi^2(p)$ di bawah hipotesis nol H_0 . Dengan demikian, uji heterogenitas spasial dapat dilakukan dengan membandingkan nilai statistik BP_{Lokal} dengan nilai kritis dari distribusi χ^2 dengan derajat bebas p . ■

Selain uji heterogenitas spasial yang telah dibahas sebelumnya, terdapat beberapa uji hipotesis penting lainnya dalam konteks model GWR dan GTWR yang perlu dipertimbangkan untuk validasi model (Fotheringham dkk., 2002b; ?).

Teorema 2.7.11 (Uji Signifikansi Koefisien Lokal GWR) *Untuk menguji signifikansi koefisien lokal $\beta_k(s_i)$ pada lokasi tertentu, dapat digunakan statistik uji*

$$t_k(s_i) = \frac{\hat{\beta}_k(s_i)}{\sqrt{\hat{\sigma}^2 e_k^\top (\mathbf{X}^\top \mathbf{W}_i \mathbf{X})^{-1} e_k}}, \quad (2.478)$$

dengan e_k adalah vektor unit ke- k . Di bawah $H_0 : \beta_k(s_i) = 0$, statistik ini meng-

ikuti distribusi t dengan derajat bebas $n_{\text{eff}} - p$, dengan n_{eff} adalah ukuran sampel efektif lokal.

Bukti. Mengikuti dari teorema interval konfidensi yang telah dibuktikan sebelumnya, estimator $\hat{\beta}_k(s_i)$ mengikuti distribusi normal asimtotik dengan mean $\beta_k(s_i)$ dan varians $\sigma^2 \mathbf{e}_k^T (\mathbf{X}^T \mathbf{W}_i \mathbf{X})^{-1} \mathbf{e}_k$. Dengan menstandarisasi dan menggunakan estimator konsisten $\hat{\sigma}^2$ untuk σ^2 , diperoleh statistik t yang dapat digunakan untuk pengujian hipotesis. Derajat bebas efektif $n_{\text{eff}} - p$ merefleksikan pengurangan informasi akibat pembobotan lokal (Fotheringham dkk., 2002b). ■

Teorema 2.7.12 (Uji Stasioneritas Spasial) Untuk menguji apakah koefisien regresi bersifat stasioner secara spasial (tidak bervariasi antar lokasi), dapat digunakan uji F dengan statistik

$$F = \frac{(RSS_{OLS} - RSS_{GWR})/df_1}{RSS_{GWR}/df_2}, \quad (2.479)$$

dengan RSS_{OLS} dan RSS_{GWR} adalah jumlah kuadrat residual dari model OLS global dan GWR, sedangkan df_1 dan df_2 adalah derajat bebas yang sesuai. Di bawah H_0 bahwa koefisien stasioner, $F \sim F(df_1, df_2)$.

Uji ini didasarkan pada perbandingan kemampuan prediksi model global (OLS) dibandingkan model lokal (GWR). Jika tidak ada heterogenitas spasial yang signifikan, maka peningkatan kemampuan prediksi dari GWR tidak akan substansial. Statistik F mengukur apakah pengurangan RSS dari GWR secara statistik signifikan dibandingkan dengan kompleksitas tambahan yang diintroduksi oleh parameter lokal (Fotheringham dkk., 2002b).

Teorema 2.7.13 (Diagnostik Model GWR) Beberapa diagnostik penting untuk evaluasi model GWR meliputi:

(a) **Akaike Information Criterion (AIC) terkoreksi:**

$$AIC_c = 2n \log(\hat{\sigma}) + n \log(2\pi) + n \left\{ \frac{n + \text{tr}(\mathbf{S})}{n - 2 - \text{tr}(\mathbf{S})} \right\}, \quad (2.480)$$

dengan \mathbf{S} adalah matriks hat dan $tr(\mathbf{S})$ adalah jejak matriks yang mengukur derajat bebas efektif.

(b) **Effective degrees of freedom:**

$$df = tr(\mathbf{S}), \quad (2.481)$$

yang mengukur kompleksitas model efektif.

(c) **Indeks determinasi lokal:**

$$R^2_{lokal}(s_i) = 1 - \frac{RSS_{lokal}(s_i)}{TSS_{lokal}(s_i)}, \quad (2.482)$$

yang mengukur kemampuan prediksi model di setiap lokasi.

Teorema 2.7.14 (Uji Multikolinearitas Lokal) Untuk mendeteksi multikolinearitas dalam konteks lokal, dapat digunakan indeks kondisi lokal

$$\kappa_i = \sqrt{\frac{\lambda_{\max}(\mathbf{X}^\top \mathbf{W}_i \mathbf{X})}{\lambda_{\min}(\mathbf{X}^\top \mathbf{W}_i \mathbf{X})}}, \quad (2.483)$$

dengan λ_{\max} dan λ_{\min} adalah nilai eigen terbesar dan terkecil dari matriks $\mathbf{X}^\top \mathbf{W}_i \mathbf{X}$. Nilai $\kappa_i > 30$ mengindikasikan masalah multikolinearitas yang serius di lokasi i .

Teorema 2.7.15 (Uji Autokorelasi Residual Spasial untuk GWR) Untuk menguji keberadaan autokorelasi spasial pada residual model GWR, dapat digunakan statistik Moran's I yang dimodifikasi:

$$I_{GWR} = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} e_i e_j}{\sum_{i=1}^n e_i^2}, \quad (2.484)$$

dengan e_i adalah residual GWR di lokasi i . Di bawah H_0 bahwa tidak ada autokorelasi spasial, I_{GWR} mengikuti distribusi normal asimtotik.

Pengujian-pengujian ini memberikan kerangka diagnostik yang komprehensif untuk memvalidasi model heterogenitas spasial dan memastikan bahwa asumsi-

asumsi yang mendasari analisis GWR dan GTWR terpenuhi dengan baik.

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, & 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

$$y = \phi(\mathbf{w}^\top \mathbf{x} + b), \quad (2.485)$$

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

$$\mathbf{z}^{(l)} = \mathbf{W}^{(l)} \mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}, \quad \mathbf{a}^{(l)} = \phi^{(l)}(\mathbf{z}^{(l)}), \quad l = 1, \dots, L, \quad (2.486)$$

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 ber-

lapis.

Definisi 2.8.4 (Fungsi Sigmoid) *Fungsi sigmoid didefinisikan sebagai*

$$\varphi(z) = \frac{1}{1 + e^{-z}}. \quad (2.487)$$

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*

$$\varphi(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}. \quad (2.488)$$

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 & Hinton, 2010)) *Fungsi ReLU (Rectified Linear Unit) didefinisikan sebagai*

$$\varphi(z) = \max(0, z). \quad (2.489)$$

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*

$$\varphi(z) = \begin{cases} z, & z \geq 0, \\ \alpha z, & z < 0, \end{cases} \quad (2.490)$$

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*

$$\varphi(z) = z \cdot \Phi(z), \quad (2.491)$$

dengan $\Phi(z)$ fungsi distribusi normal standar. Aproksimasi praktisnya adalah sebagai berikut.

$$\varphi(z) \approx 0.5z \left(1 + \tanh \left[\sqrt{\frac{2}{\pi}}(z + 0.0447z^3) \right] \right). \quad (2.492)$$

Contoh 2.8.13 Untuk $z = 1$, $\varphi(1) \approx 0.84$. Untuk $z = -1$, $\varphi(-1) \approx -0.16$.

Lema 2.8.14 (Sifat Fungsi Aktivasi) *Fungsi aktivasi $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ yang digunakan pada jaringan saraf tiruan sebaiknya memenuhi:*

(i) φ non-linear,

(ii) φ terdiferensiasi hampir di semua titik,

(iii) φ tidak konstan, dan

(iv) φ memiliki range memadai, tidak menyebabkan saturasi total.

Bukti.

(i) *Nonlinearitas.* Jika φ linear, misalnya $\varphi(z) = az + b$, maka suatu multilayer

perceptron dengan L lapisan menghasilkan

$$f(\mathbf{x}) = \mathbf{W}^{(L)} \varphi(\mathbf{W}^{(L-1)} \varphi(\dots \varphi(\mathbf{W}^{(1)} \mathbf{x}))) = \mathbf{W}' \mathbf{x} + \mathbf{b}',$$

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) *Kediferensialan.* 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, φ setidaknya harus terdiferensiasi hampir di semua titik. Fungsi seperti ReLU tetap sah karena ketakdiferensialan hanya terjadi pada $z = 0$ (himpunan titik ukuran nol).

- (iii) *Nonkonstansi.* Jika $\varphi(z) = c$ konstan, maka $\varphi'(z) = 0$ untuk semua z . Subsitusi ke rumus backpropagation menghasilkan $\delta^{(l)} = \mathbf{0}$ untuk semua l , sehingga

$$\frac{\partial L}{\partial \mathbf{W}^{(l)}} = \mathbf{0}.$$

Bobot tidak akan pernah diperbarui, sehingga proses pelatihan tidak berjalan.

- (iv) *Rentang keluaran yang memadai.* Jika φ cepat jenuh, misalnya sigmoid dengan limit $\varphi(z) \rightarrow 0$ 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 menggagalk-

an 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

$$L(\Theta) = \frac{1}{n} \sum_{i=1}^n \ell(\hat{y}_i, y_i), \quad (2.493)$$

dengan fungsi kerugian ℓ dipilih sesuai dengan tugas yang dihadapi.

Contoh 2.8.16 Untuk regresi skalar, sebagai contoh digunakan $\ell(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$. Hal ini berarti jika $\hat{y} = 2.3$ dan $y = 1.5$, maka $\ell = \frac{1}{2}(0.8)^2 = 0.32$.

Contoh 2.8.17 Untuk klasifikasi biner, digunakan probabilitas prediksi $\hat{p} = \sigma(z)$ dan fungsi kerugian

$$\ell(\hat{p}, y) = -[y \log \hat{p} + (1 - y) \log(1 - \hat{p})], \quad y \in \{0, 1\}. \quad (2.494)$$

Fungsi ini merupakan entropi silang antara distribusi target $p(y)$ dan prediksi \hat{p} , yaitu

$$\begin{aligned} H(p, \hat{p}) &= - \sum_{y \in \{0,1\}} p(y) \log \hat{p}(y) \\ &= -[p(1) \log \hat{p} + p(0) \log(1 - \hat{p})]. \end{aligned} \quad (2.495)$$

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 $\hat{\mathbf{p}}$:

$$\begin{aligned} H(p, \hat{\mathbf{p}}) &= -\sum_{k=1}^K p(k) \log \hat{p}_k \\ &= -\log \hat{p}_y, \end{aligned} \tag{2.496}$$

dengan $p(k) = 1$ jika $k = y$ dan 0 selainnya.

Sebagai contoh, untuk $K = 3$ kelas, jika $\mathbf{z} = (2.0, 1.0, 0.1)$, maka softmax $\hat{\mathbf{p}} \approx (0.659, 0.242, 0.099)$. Jika label benar $y = 1$, maka $\ell = -\log 0.659 \approx 0.417$. Jika $y = 3$, maka $\ell = -\log 0.099 \approx 2.313$.

2.8.4. Pembelajaran dengan Propagasi Mundur

Teorema 2.8.19 (Aturan Rantai pada Jaringan) Misalkan keluaran akhir didefinisikan sebagai komposisi $y = f^{(L)} \circ f^{(L-1)} \circ \dots \circ f^{(1)}(\mathbf{x})$. Maka turunan terhadap parameter pada lapisan ke- l memenuhi

$$\frac{\partial y}{\partial \mathbf{W}^{(l)}} = \frac{\partial f^{(L)}}{\partial f^{(L-1)}} \cdots \frac{\partial f^{(l+1)}}{\partial f^{(l)}} \frac{\partial f^{(l)}}{\partial \mathbf{W}^{(l)}}. \tag{2.497}$$

Interpretasi teorema ini adalah bahwa gradien dihitung dengan mengalikan turunan setiap lapisan dari keluaran ke lapisan yang diinginkan. Hal ini berarti akan dihitung efek perubahan bobot pada lapisan tersebut terhadap keluaran akhir melalui semua lapisan berikutnya. Apabila efek ini kecil, maka gradien akan mengecil secara eksponensial (fenomena *vanishing gradient*). Apabila efek ini besar, maka gradien akan membesar secara eksponensial (fenomena *exploding gradient*). Secara umum, pembelajaran menginginkan untuk memperbaiki bobot yang memiliki

efek signifikan terhadap keluaran, tetapi tidak terlalu besar sehingga menyebabkan ketidakstabilan.

Teorema 2.8.20 (Propagasi Mundur atau *Backpropagation*, (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

$$\frac{\partial \ell}{\partial \mathbf{W}^{(l)}} = \delta^{(l)} (\mathbf{a}^{(l-1)})^\top, \quad (2.498)$$

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 Θ dioptimalkan untuk meminimalkan fungsi risiko $L(\Theta)$. Beberapa algoritma optimisasi yang akan dibahas adalah *Gradient Descent* (GD), *Stochastic Gradient Descent* (SGD), dan *Adaptive Moment Estimation* (Adam).

Definisi 2.8.22 (Gradient Descent) Diberikan fungsi risiko empiris

$$L(\Theta) = \frac{1}{n} \sum_{i=1}^n \ell(f_\Theta(\mathbf{x}_i), y_i),$$

pembaruan parameter dengan laju belajar $\eta > 0$ adalah

$$\Theta_{t+1} = \Theta_t - \eta \nabla_\Theta L(\Theta_t). \quad (2.499)$$

Metode ini konvergen ke minimum lokal untuk fungsi non-konveks seperti jaringan saraf.

Intuisi dari metode berbasis gradien ini adalah memperbarui parameter model ke arah yang menurunkan fungsi kerugian paling cepat. Pada setiap iterasi, gradien $\nabla_\Theta L(\Theta)$ menunjukkan arah perubahan parameter yang paling efektif untuk mengurangi kerugian. Dengan memilih laju pembelajaran η yang sesuai, proses optimisasi bergerak menuju minimum lokal dari fungsi kerugian. Jika $L(\Theta)$ konveks, metode ini menjamin konvergensi ke minimum global. Namun, pada kasus 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.

$$\Theta_{t+1} = \Theta_t - \eta \frac{1}{|\mathcal{B}_t|} \sum_{(\mathbf{x}_i, y_i) \in \mathcal{B}_t} \nabla_{\Theta} \ell(f_{\Theta}(\mathbf{x}_i), y_i), \quad (2.500)$$

dengan \mathcal{B}_t adalah batch kecil pada iterasi ke- t , η adalah laju pembelajaran, dan ℓ adalah fungsi kerugian. Metode ini sangat efisien untuk data berukuran besar dan memberikan regularisasi implisit melalui pengacakan batch.

Contoh 2.8.25 Misalkan fungsi kerugian $L(w) = \frac{1}{2}(w - 3)^2$ dan data terdiri dari $n = 1000$ sampel, masing-masing x_i dengan target $y_i = 3$ untuk semua i . Dengan batch kecil berukuran $m = 10$, pada setiap iterasi SGD dilakukan langkah berikut.

- (a) Pilih acak $m = 10$ sampel dari data, misal indeks $\mathcal{B}_t = \{i_1, \dots, i_{10}\}$.
- (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 η :

$$w_{t+1} = w_t - \eta(w_t - 3)$$

(d) Ulangi langkah (a)–(c) hingga w_t konvergen ke 3.

Dengan batch acak, setiap iterasi hanya menggunakan sebagian kecil data, sehingga proses lebih efisien dan tetap menuju minimum global $w = 3$.

Definisi 2.8.26 (Adaptive Moment Estimation (Adam)) Adam adalah algoritma optimisasi berbasis gradien yang menggabungkan momentum (rata-rata bergerak gradien) dan skala adaptif (rata-rata kuadrat gradien). Adam memperbarui parameter dengan estimasi momen pertama dan kedua yang dikoreksi bias. Dengan gradien $\mathbf{g}_t = \nabla_{\Theta} L_t(\Theta_t)$, inisialisasi $\mathbf{m}_0 = \mathbf{0}$, $\mathbf{v}_0 = \mathbf{0}$, lalu untuk setiap iterasi $t \geq 1$:

$$\mathbf{m}_t = \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1) \mathbf{g}_t, \quad (2.501)$$

$$\mathbf{v}_t = \beta_2 \mathbf{v}_{t-1} + (1 - \beta_2) (\mathbf{g}_t \odot \mathbf{g}_t), \quad (2.502)$$

$$\hat{\mathbf{m}}_t = \frac{\mathbf{m}_t}{1 - \beta_1^t}, \quad \hat{\mathbf{v}}_t = \frac{\mathbf{v}_t}{1 - \beta_2^t}, \quad (2.503)$$

$$\Theta_{t+1} = \Theta_t - \eta \frac{\hat{\mathbf{m}}_t}{\sqrt{\hat{\mathbf{v}}_t} + \varepsilon}. \quad (2.504)$$

Parameter umum: $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$.

Contoh 2.8.27 Misalkan gradien pada iterasi pertama $\mathbf{g}_1 = (0.1, -0.2)$, inisialisasi $\mathbf{m}_0 = (0, 0)$, $\mathbf{v}_0 = (0, 0)$, $\eta = 0.01$. Maka:

$$\mathbf{m}_1 = 0.9 \cdot (0, 0) + 0.1 \cdot (0.1, -0.2) = (0.01, -0.02),$$

$$\mathbf{v}_1 = 0.999 \cdot (0, 0) + 0.001 \cdot (0.01, 0.04) = (0.00001, 0.00004),$$

$$\hat{\mathbf{m}}_1 = (0.01, -0.02)/(1 - 0.9^1) = (0.1, -0.2),$$

$$\hat{\mathbf{v}}_1 = (0.00001, 0.00004)/(1 - 0.999^1) = (0.01, 0.04),$$

$$\Theta_2 = \Theta_1 - 0.01 \cdot \frac{(0.1, -0.2)}{(\sqrt{0.01}, \sqrt{0.04}) + 10^{-8}} = \Theta_1 - 0.01 \cdot (1, -1).$$

Jadi, parameter diperbarui dengan langkah adaptif pada setiap komponen.

2.8.6. Jaringan Saraf Tiruan sebagai Aproksimasi Universal

Salah satu hasil teoretis paling penting tentang jaringan saraf tiruan adalah Teorema Aproksimasi Universal (UAT). Teorema ini menyatakan bahwa jaringan saraf dengan satu lapisan tersembunyi yang cukup besar dapat mengaproksimasi fungsi kontinu pada himpunan kompak dengan presisi setinggi apapun, asalkan fungsi aktivasi memenuhi syarat tertentu.

Teorema 2.8.28 (Teorema Aproksimasi Universal (UAT), (Cybenko, 1989)) *Jika $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ adalah fungsi sigmoid kontinu tak konstan (misal, $\varphi(z) = 1/(1 + e^{-z})$ atau $\varphi(z) = \tanh(z)$), maka untuk setiap fungsi kontinu f pada $[0, 1]^p$ dan setiap $\varepsilon > 0$, ada jaringan satu lapis dengan jumlah neuron cukup banyak sehingga*

$$\sup_{\mathbf{x} \in [0,1]^p} |f(\mathbf{x}) - g(\mathbf{x})| < \varepsilon,$$

dengan $g(\mathbf{x}) = \sum_{j=1}^m c_j \varphi(\mathbf{w}_j^\top \mathbf{x} + b_j)$.

2.9. Jaringan Saraf Graf

Jaringan Saraf Graf (*Graph Neural Networks*, GNN) merupakan kelas arsitektur pembelajaran representasi yang dirancang khusus untuk data terstruktur dalam bentuk graf. Berbeda dengan jaringan saraf tiruan (JST) yang bekerja pada vektor atau matriks, GNN memanfaatkan informasi topologi melalui graf $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ dengan simpul \mathcal{V} dan sisi \mathcal{E} .

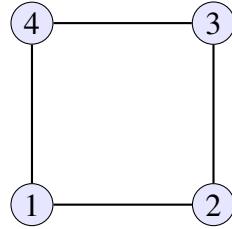
2.9.1. Dasar Graf dan Laplacian

Definisi 2.9.1 (Graf dan Matriks Ketetanggaan) *Graf $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ dengan $|\mathcal{V}| = n$ direpresentasikan oleh:*

$$\mathbf{A}_{ij} = \begin{cases} 1 & \text{jika } (i, j) \in \mathcal{E}, \\ 0 & \text{lainnya,} \end{cases} \quad (2.505)$$

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 $\mathbf{L} = \mathbf{D} - \mathbf{A}$.

A. Normalisasi simetris diberikan oleh

$$\mathbf{L}_{\text{sym}} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}. \quad (2.506)$$

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

$$\mathbf{L}_{\text{sym}} = \mathbf{I} - \frac{1}{2}\mathbf{A},$$

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.

$$\mathbf{L} = \mathbf{U}\Lambda\mathbf{U}^{\top}, \quad (2.507)$$

dengan \mathbf{U} ortogonal dan Λ diagonal berisi nilai eigen $\lambda_1, \dots, \lambda_n \geq 0$. Selalu berlaku $\lambda_1 = 0$ dengan vektor eigen $\mathbf{1}$.

2.9.2. Kerangka Penyampaian Pesan pada Jaringan Saraf

Definisi 2.9.6 (Message Passing Neural Networks (MPNN), (?)) Misalkan graf $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ dengan \mathcal{V} adalah himpunan simpul dan \mathcal{E} adalah himpunan sisi. Misalkan pula \mathcal{N}_u adalah ketetanggaan simpul $u \in \mathcal{V}$. Lebih lanjut, misalkan \mathbf{x}_u adalah fitur simpul u dan e_{uv} adalah fitur sisi $(u, v) \in \mathcal{E}$. Suatu lapisan MPNN

didefinisikan sebagai berikut.

$$\mathbf{h}_u = \varphi \left(\mathbf{x}_u, \bigoplus_{v \in \mathcal{N}_u} \psi(\mathbf{x}_u, \mathbf{x}_v, e_{uv}), \right) \quad (2.508)$$

dengan φ dan ψ adalah fungsi yang terdiferensiasi (seperti jaringan saraf) dan \bigoplus adalah operator agregasi yang bersifat komutatif dan asosiatif (seperti penjumlahan atau rata-rata) atau permutation invariant.

Kerangka MPNN ini sangat fleksibel dan dapat disesuaikan dengan berbagai jenis data graf. Fungsi ψ dapat dirancang untuk menangkap interaksi spesifik antara simpul dan sisi, sedangkan fungsi φ dapat berupa jaringan saraf yang kompleks untuk pembaruan fitur. Operator agregasi \bigoplus memastikan bahwa model tetap invariant terhadap permutasi tetangga, yang penting dalam konteks graf. Beberapa arsitektur umum GNN, seperti *Graph Convolutional Networks* (GCN) dan *Graph Attention Networks* (GAT), dapat dianggap sebagai kasus khusus dari kerangka MPNN ini.

2.9.3. Arsitektur Umum dalam GNN

Definisi 2.9.7 (Graph Convolutional Networks (GCN), (Kipf & Welling, 2017))
GCN mendefinisikan operasi propagasi lapisan ke- l sebagai

$$\mathbf{H}^{(l+1)} = \sigma \left(\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2} \mathbf{H}^{(l)} \mathbf{W}^{(l)} \right), \quad (2.509)$$

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.

GCN merupakan kasus khusus dari kerangka MPNN dengan:

- Fitur sisi diabaikan (e_{uv} tidak digunakan).
- Fungsi pesan $\psi(\mathbf{x}_u, \mathbf{x}_v, e_{uv}) = \mathbf{x}_v$ (hanya mengambil fitur tetangga).
- Operator agregasi \oplus berupa penjumlahan atau rata-rata tertimbang (melalui normalisasi Laplacian).
- Fungsi update φ adalah komposisi linear dan aktivasi non-linear: $\varphi(\mathbf{x}_u, \text{AGG}) = \sigma(\text{AGG} \cdot \mathbf{W})$.

Secara eksplisit, propagasi GCN dapat ditulis sebagai

$$\mathbf{h}_u^{(l+1)} = \sigma \left(\sum_{v \in \mathcal{N}_u \cup \{u\}} \frac{1}{\sqrt{d_u d_v}} \mathbf{h}_v^{(l)} \mathbf{W}^{(l)} \right),$$

dengan d_u derajat simpul u dan σ fungsi aktivasi. Matriks normalisasi simetris $\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2}$ merepresentasikan agregasi pesan dari tetangga dengan bobot yang sesuai, sehingga GCN adalah MPNN dengan agregasi rata-rata dan update linear.

Definisi 2.9.8 (Graph Attention Networks (GAT), (Veličković dkk., 2018)) GAT mengganti normalisasi derajat dengan mekanisme perhatian. Koefisien perhatian untuk sisi (i, j) adalah

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^\top [\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_j]))}{\sum_{k \in \mathcal{N}(i)} \exp(\text{LeakyReLU}(\mathbf{a}^\top [\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_k]))}, \quad (2.510)$$

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
- $\|$ operator konkatenasi.

Pembaruan fitur simpul adalah

$$\mathbf{h}'_i = \sigma \left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} \mathbf{W} \mathbf{h}_j \right). \quad (2.511)$$

GAT merupakan kasus khusus dari kerangka MPNN dengan spesifikasi berikut.

- Fitur sisi e_{uv} tidak digunakan (atau dapat diabaikan).
- Fungsi pesan $\psi(\mathbf{x}_u, \mathbf{x}_v, e_{uv}) = \mathbf{W} \mathbf{x}_v$ (transformasi linear fitur tetangga).
- Operator agregasi \oplus berupa penjumlahan tertimbang, dengan bobot agregasi α_{uv} diperoleh dari mekanisme perhatian (attention) yang bersifat permutation invariant.
- Fungsi update φ adalah komposisi linear dan aktivasi non-linear: $\varphi(\mathbf{x}_u, \text{AGG}) = \sigma(\text{AGG})$.

Secara formal, pembaruan fitur simpul u pada GAT dapat ditulis dalam notasi MPNN sebagai

$$\mathbf{h}'_u = \sigma \left(\sum_{v \in \mathcal{N}_u} \alpha_{uv} \psi(\mathbf{x}_u, \mathbf{x}_v, e_{uv}) \right) = \sigma \left(\sum_{v \in \mathcal{N}_u} \alpha_{uv} \mathbf{W} \mathbf{x}_v \right), \quad (2.512)$$

dengan α_{uv} adalah skor perhatian yang bergantung pada fitur \mathbf{x}_u dan \mathbf{x}_v melalui fungsi LeakyReLU($\mathbf{a}^\top [\mathbf{W} \mathbf{x}_u \| \mathbf{W} \mathbf{x}_v]$) 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, & Leskovec, 2017)) GraphSAGE atau Graph Sample and Aggregation mendefinisikan pembaruan simpul i sebagai

$$\mathbf{h}_i^{(l+1)} = \sigma \left(\mathbf{W}^{(l)} \cdot \text{AGGREGATE} \left(\{\mathbf{h}_i^{(l)}\} \cup \{\mathbf{h}_j^{(l)} : j \in \mathcal{N}(i)\} \right) \right), \quad (2.513)$$

dengan AGGREGATE fungsi agregasi non-linear yang bisa berupa mean, max-pooling, atau LSTM.

2.9.4. Kekuatan Representasi GNN

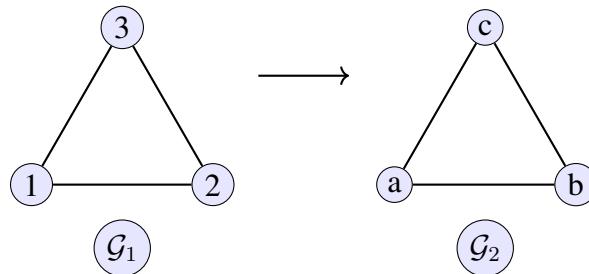
Definisi 2.9.10 (Isomorfisme Graf) Dua graf $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ dan $\mathcal{G}_2 = (\mathcal{V}_2, \mathcal{E}_2)$ disebut isomorfik jika terdapat bijeksi $\pi : \mathcal{V}_1 \rightarrow \mathcal{V}_2$ sehingga

$$(i, j) \in \mathcal{E}_1 \iff (\pi(i), \pi(j)) \in \mathcal{E}_2. \quad (2.514)$$

Sebuah fungsi f dikatakan invarian terhadap isomorfisme graf jika $f(\mathcal{G}_1) = f(\mathcal{G}_2)$ untuk setiap graf isomorfik $\mathcal{G}_1 \cong \mathcal{G}_2$.

Contoh 2.9.11 Misalkan \mathcal{G}_1 adalah graf dengan simpul $\{1, 2, 3\}$ dan sisi $\{(1, 2), (2, 3), (3, 1)\}$ (graf segitiga). Ambil \mathcal{G}_2 dengan simpul $\{a, b, c\}$ dan sisi $\{(a, b), (b, c), (c, a)\}$. Bijeksi π dapat dipilih misal $\pi(1) = a$, $\pi(2) = b$, $\pi(3) = c$. Maka \mathcal{G}_1 dan \mathcal{G}_2 isomorfik.

Sebagai contoh fungsi invarian, jumlah simpul $|\mathcal{V}|$ dan jumlah sisi $|\mathcal{E}|$ adalah invarian terhadap isomorfisme graf. Demikian juga, spektrum Laplacian graf \mathbf{L} adalah invarian terhadap isomorfisme.



Gambar 2.6 Ilustrasi dua graf isomorfik, yaitu \mathcal{G}_1 dan \mathcal{G}_2 memiliki struktur yang sama walaupun label simpul berbeda. (Sumber: Dokumen penulis)

Definisi 2.9.12 (Tes Weisfeiler–Lehman (WL) Satu Dimensi) Algoritma WL satu dimensi, juga dikenal sebagai color refinement, merupakan prosedur iteratif untuk membedakan graf berdasarkan struktur lokal simpul. Pada setiap iterasi, label simpul $h_i^{(t)}$ diperbarui dengan cara:

$$h_i^{(t+1)} = \text{HASH}\left(h_i^{(t)}, \{h_j^{(t)} : j \in \mathcal{N}(i)\}\right), \quad (2.515)$$

dimulai dari label awal $h_i^{(0)}$. Dua graf dikatakan dapat dibedakan oleh tes WL

jika multiset label akhirnya berbeda. Dalam konteks pembelajaran mesin, tes WL memberikan kerangka formal untuk mengukur kemampuan model, seperti GNN, dalam membedakan struktur graf yang berbeda melalui proses agregasi informasi dari tetangga.

Fungsi HASH pada algoritma WL berperan sebagai pemetaan yang menggabungkan label simpul saat ini dengan label-label tetangganya menjadi label baru yang unik. Intuisi utamanya adalah memastikan bahwa jika dua simpul memiliki lingkungan yang berbeda, hasil HASH juga berbeda. Dalam implementasi praktis, HASH dapat berupa fungsi injektif seperti pengkodean *string*, *concatenation*, atau fungsi neural yang mampu membedakan multiset masukan secara efektif.

Teorema 2.9.13 (Keterbatasan GNN Berbasis Agregasi, (Xu, Hu, Leskovec, & Jegelka, 2019)) *GNN berbasis agregasi dengan pembaruan umum*

$$h_i^{(l+1)} = \varphi \left(W^{(l)} \cdot \text{AGGREGATE} \left(\{h_i^{(l)}\} \cup \{h_j^{(l)} : j \in \mathcal{N}(i)\} \right) \right) \quad (2.516)$$

tidak lebih kuat daripada tes WL satu dimensi dalam membedakan graf yang tidak isomorfik. Artinya, kemampuan GNN untuk membedakan struktur graf dibatasi oleh kekuatan agregasi lokal yang serupa dengan tes WL.

Teorema 2.9.14 (Hubungan dengan Tes WL, (Xu dkk., 2019)) *Jika fungsi agregasi AGGREGATE bersifat injektif terhadap multiset, maka GNN memiliki kekuatan representasi setara dengan tes WL satu dimensi. Dengan kata lain, GNN dapat membedakan graf sejauh tes WL mampu membedakannya, asalkan agregasi dan pembaruan fitur dilakukan secara unik untuk setiap lingkungan simpul.*

BAB III

RENCANA PENELITIAN

3.1. Rencana Penelitian

Penelitian ini berfokus pada analisis spasial-temporal inflasi di Indonesia dengan unit observasi sebanyak 38 provinsi. Pendekatan yang digunakan adalah regresi spasial-temporal dengan koefisien bervariasi dan regresi terboboti yang dikembangkan dengan pembobotan berbasis jaringan saraf graf. Kedua model ini diberi nama *Graph Neural Network-Geographically and Temporally Varying Coefficient* (GNN-GTVC) dan *Graph Neural Network-Geographically and Temporally Weighted Regression* (GNN-GTWR). Tujuannya untuk memahami disparitas inflasi antarprovinsi, faktor-faktor global maupun lokal yang memengaruhinya, serta potensi *spillover* antarwilayah.

3.2. Rencana Pengambilan Data

Data diambil pada periode Januari 2024 hingga Agustus 2025. Data yang digunakan dalam penelitian ini adalah sebagai berikut.

- (a) **Inflasi Provinsi:** Indeks Harga Konsumen (IHK) per provinsi dengan tahun dasar 2022. Satuan: persen (bulanan). Sumber: BPS RI.
- (b) **Suku Bunga Kebijakan:** BI 7-Day Reverse Repo Rate. Satuan: persen (bulanan). Sumber: Bank Indonesia.
- (c) **Nilai Tukar Rupiah/USD:** Kurs JISDOR, diagregasi bulanan. Satuan: Rupiah/USD. Sumber: Bank Indonesia.
- (d) **Uang Beredar (M2):** Jumlah uang beredar luas. Satuan: triliun rupiah. Sumber: Bank Indonesia.

- (e) **Harga Pangan Strategis:** Harga beras, cabai, bawang, minyak goreng, dan daging ayam (rata-rata bulanan). Satuan: rupiah/kg atau liter. Sumber: Pusat Informasi Harga Pangan Nasional.
- (f) **Variabel Struktural Provinsi:** PDRB, tingkat kemiskinan, UMP. Satuan: beragam (juta rupiah, persen, rupiah). Sumber: BPS RI.
- (g) **Statistik Ekonomi dan Keuangan Daerah (SEKDA):** Kredit dan simpanan per provinsi. Satuan: miliar rupiah. Sumber: Bank Indonesia.

3.3. Rencana Jadwal Penelitian

Jadwal penelitian dirangkum pada Tabel 3.1 berikut.

Tabel 3.1 Jadwal Penelitian Agustus 2025 – Januari 2026

Bulan	Kegiatan 1	Kegiatan 2	Kegiatan 3
Ags 2025	✓		
Sep 2025	✓	✓	
Okt 2025	✓	✓	
Nov 2025		✓	✓
Des 2025			✓
Jan 2026			✓

Dengan keterangan kegiatan sebagai berikut.

- (a) Kegiatan 1 adalah studi literatur, penyusunan proposal, pengumpulan dan pembersihan data, serta analisis deskriptif awal.
- (b) Kegiatan 2 adalah pengembangan model GNN-GTVC dan GNN-GTWR, serta estimasi model.
- (c) Kegiatan 3 adalah analisis menyeluruh dan penulisan laporan akhir penelitian.

BAB IV

PENGEMBANGAN PEMBOBOTAN BERBASIS JARINGAN SARAF GRAF UNTUK REGRESI SPASIAL-TEMPORAL

4.1. Kajian Literatur Terkait

Kajian mengenai regresi berbobot spasial dan spasial-temporal telah berkembang pesat dalam beberapa dekade terakhir. Evolusi model dapat ditelusuri mulai dari regresi klasik berbasis kernel, regresi berbobot geografis, hingga integrasi dengan pembelajaran mesin dan *neural network* serta *graph neural network*. Bagian ini memaparkan secara kronologis penelitian-penelitian terdahulu yang relevan dengan metodologi yang diusulkan.

4.1.1. *Geographically Weighted Artificial Neural Network (GWANN)*

Sebagai upaya mengintegrasikan *deep learning* ke dalam kerangka geografis, Hagenauer and Helbich (2022) mengusulkan *Geographically Weighted Artificial Neural Network* (GWANN). Pada GWANN, setiap lokasi memiliki model *artificial neural network* (ANN) lokal:

$$y_i = f(\mathbf{x}_i; \boldsymbol{\theta}(\mathbf{s}_i)) + \varepsilon_i, \quad (4.1)$$

dengan parameter $\boldsymbol{\theta}(\mathbf{s}_i)$ dipengaruhi oleh bobot kernel spasial. ANN lokal ini mengantikan model linear pada GWR, sehingga mampu menangkap hubungan non-linear antara kovariat dan respon. Kekurangannya adalah biaya komputasi yang tinggi karena ANN dilatih berulang pada tiap lokasi, serta ketidakstabilan parameter jika data lokal terbatas.

4.1.2. Spatial and Attribute Neural Network Weighted Regression (SANNWR)

Ni dkk. (2022) memperkenalkan *Spatial and Attribute Neural Network Weighted Regression* (SANNWR), yang memperluas GWR dengan bobot berbasis gabungan jarak spasial dan atribut. Estimasi regresi tetap menggunakan formulasi WLS, tetapi bobot w_{ij} dihitung melalui *neural network* yang mempelajari fungsi non-linear dari $(d_{ij}^{\text{spatial}}, d_{ij}^{\text{attribute}})$:

$$w_{ij} = \frac{\exp(\text{NN}(d_{ij}^{\text{spatial}}, d_{ij}^{\text{attribute}}))}{\sum_k \exp(\text{NN}(d_{ik}^{\text{spatial}}, d_{ik}^{\text{attribute}}))}. \quad (4.2)$$

Dengan demikian, bobot tidak lagi berupa fungsi kernel yang kaku, melainkan dipelajari secara adaptif. Inovasi ini menegaskan peran *deep learning* dalam merepresentasikan kedekatan spasial dan atribut secara simultan.

4.1.3. Spatial Regression Graph Convolutional Neural Networks (SRGCNN)

D. Zhu dkk. (2022) mengusulkan *Spatial Regression Graph Convolutional Neural Network* (SRGCNN), yang menggabungkan regresi spasial dengan konvolusi graf. Model ini merumuskan dependensi spasial dengan operator *graph convolutional network* (GCN), yaitu

$$\mathbf{H}^{(\ell+1)} = \sigma\left(\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-1/2}\mathbf{H}^{(\ell)}\mathbf{W}^{(\ell)}\right), \quad (4.3)$$

dan kemudian menghubungkannya dengan regresi linear autoregresif spasial (SAR), yaitu

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon}, \quad (4.4)$$

sehingga efek lag spasial dan heterogenitas dapat ditangkap sekaligus. SRGCNN menandai transisi eksplisit penggunaan *graph neural network* dalam konteks regresi spasial. Sebagai catatan, SRGCNN tidak mengadopsi bobot yang dipelajari melalui GNN, melainkan menggunakan GNN untuk mengekstrak fitur spasial yang memiliki kemiripan dengan efek lag spasial pada model SAR.

4.1.4. Spatio-Temporal Graph Convolutional Networks (ST-GCN)

Dalam domain peramalan deret waktu spasiotemporal, Yu, Yin, and Zhu (2018) mengusulkan *Spatio-Temporal GCN* (ST-GCN). Model ini mengintegrasikan konvolusi graf spasial dengan konvolusi temporal berbasis deret:

$$\mathbf{H}^{(\ell+1)} = \sigma(\text{ConvTemporal}(\tilde{\mathbf{A}}\mathbf{H}^{(\ell)}\mathbf{W}^{(\ell)})). \quad (4.5)$$

ST-GCN dan variannya (ASTGCN, T-GCN) banyak digunakan untuk memprediksi lalu lintas, konsumsi energi, dan dinamika jaringan kompleks. Kontribusinya adalah menunjukkan bahwa *graph neural network* dapat diperluas secara alami ke dimensi waktu.

4.1.5. Geographically Neural Network Weighted Regression (GNNWR)

Salah satu perkembangan mutakhir dalam literatur regresi spasial adalah pengenalan *Geographically Neural Network Weighted Regression* (GNNWR) oleh Yin dkk. (2024). Model ini berangkat dari keterbatasan *Geographically Weighted Regression* (GWR) yang menggunakan fungsi *kernel* parametrik untuk menghitung bobot spasial. Pada GWR, estimator koefisien lokal diberikan oleh

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = (\mathbf{X}^\top \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}(u_i, v_i) \mathbf{y}, \quad (4.6)$$

dengan $\mathbf{W}(u_i, v_i)$ matriks bobot *kernel* spasial berbasis jarak Euclidean antara titik observasi i dan titik lainnya. Namun, pemilihan *kernel* yang kaku sering gagal menangkap heterogenitas non-linear. Untuk itu, GNNWR mengganti $\mathbf{W}(u_i, v_i)$ dengan bobot yang dipelajari oleh *spatial weighted neural network* (SWNN). Secara umum, model GNNWR memandang koefisien lokal sebagai skala dari koefisien global,

$$\beta_k^{\text{Lokal}}(u_i, v_i) = w_k(u_i, v_i) \beta_k^{\text{Global}}, \quad \sum_{k=1}^p w_k(u_i, v_i) = 1, \quad w_k \geq 0, \quad (4.7)$$

dengan β_k^{Global} diperoleh dari model OLS *baseline*. Hubungan ini memastikan bahwa koefisien global dapat diinterpretasikan sebagai rata-rata tertimbang dari koefisien lokal.

Bobot $w_k(u_i, v_i)$ dipelajari menggunakan *neural network* yang menerima input berupa jarak spasial antar unit:

$$\mathbf{W}(u_i, v_i) = \text{SWNN}(d_{i1}^S, d_{i2}^S, \dots, d_{in}^S; \boldsymbol{\theta}), \quad (4.8)$$

dengan d_{ij}^S jarak spasial antara unit i dan j , serta $\boldsymbol{\theta}$ parameter jaringan saraf. Oleh karena itu, prediksi \hat{y}_i dapat dituliskan sebagai

$$\hat{y}_i = \mathbf{x}_i^\top \mathbf{W}(u_i, v_i) (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}, \quad i = 1, \dots, n. \quad (4.9)$$

Persamaan (4.9) dapat dipadatkan menjadi bentuk matriks

$$\hat{\mathbf{y}} = \mathbf{S} \mathbf{y}, \quad (4.10)$$

dengan \mathbf{S} bergantung pada bobot $\mathbf{W}(u_i, v_i)$ yang dilatih melalui *neural network*.

Yin dkk. (2024) kemudian memperluas kerangka ini menjadi *Geographically and Temporally Neural Network Weighted Regression* (GTNNWR), yang memasukkan dimensi waktu ke dalam pembobotan. Estimator GTNNWR ditulis sebagai

$$\hat{y}_i = \sum_{k=1}^p w_k(u_i, v_i, t) \beta_k^{\text{Global}} x_{ik} + \varepsilon_i, \quad (4.11)$$

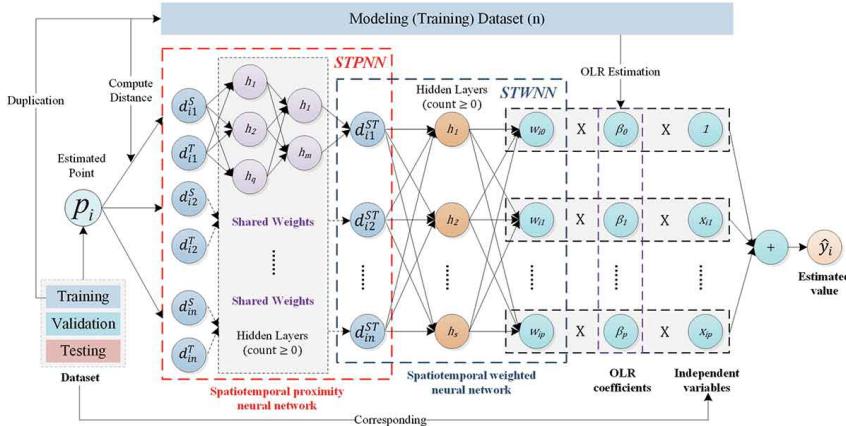
dengan bobot spatiotemporal

$$\mathbf{W}(u_i, v_i, t) = \text{STWNN}(d_{i1}^{ST}, d_{i2}^{ST}, \dots, d_{in}^{ST}; \boldsymbol{\theta}_1), \quad (4.12)$$

di mana jarak spatiotemporal d_{ij}^{ST} diperoleh melalui *spatiotemporal proximity neural network* (STPNN), yaitu

$$d_{ij}^{ST} = \text{STPNN}(d_{ij}^S, d_{ij}^T; \boldsymbol{\theta}_2). \quad (4.13)$$

Dengan demikian, GTNNWR melibatkan dua tahap pembelajaran, yaitu STPNN untuk menyintesis jarak spatiotemporal dari komponen spasial dan temporal dan STWNN untuk mengubah jarak tersebut menjadi bobot regresi. Diagram kerangka kerja GTNNWR ditunjukkan pada Gambar 4.1.



Gambar 4.1 Diagram Kerangka Kerja GTNNWR (Yin dkk., 2024)

Sebagaimana pada GWR, prediksi \hat{y} dalam GNNWR dan GTNNWR tetap dapat dipandang melalui matriks Sy . Akan tetapi, W kini dihasilkan secara endogen melalui *neural network*. Fungsi kerugian yang umum digunakan adalah *mean squared error* (MSE), yaitu

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (4.14)$$

serta kriteria informasi seperti *Akaike Information Criterion* (AIC) dan AICc untuk menghindari *overfitting*, yaitu

$$\text{AIC}_c = 2k - 2 \ln(L) + \frac{2k(k+1)}{n-k-1}, \quad (4.15)$$

dengan k jumlah parameter efektif dan L adalah *likelihood* model.

Metode GNNWR dan GTNNWR telah diaplikasikan pada berbagai domain, antara lain pemodelan kualitas udara ($PM_{2.5}$ dan AOD), pemetaan nutrien laut (*dissolved silicate*), serta estimasi emisi karbon perkotaan. Studi-studi tersebut menunjukkan bahwa dengan menggantikan kernel spasial konvensional dengan *ne-*

ural network, model mampu menangkap heterogenitas spasiotemporal yang lebih kompleks, sekaligus mempertahankan interpretabilitas koefisien lokal sebagaimana pada GWR.

4.2. Kerangka Metodologis

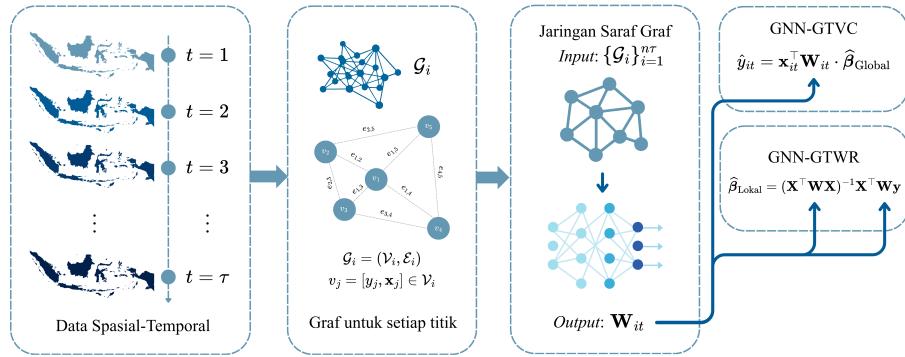
Bagian ini menjelaskan kerangka metodologis yang digunakan dalam pengembangan dua model utama, yaitu *Graph Neural Network-Geographically and Temporally Varying Coefficient* (GNN-GTVC) dan *Graph Neural Network-Geographically and Temporally Weighted Regression* (GNN-GTWR). Kedua model dikembangkan untuk menangkap heterogenitas spasial dan temporal yang kompleks dengan mengintegrasikan kemampuan representasi *Graph Neural Network* (GNN) ke dalam kerangka regresi spasiotemporal berbobot.

Keterbatasan utama pendekatan kernel pada GWR/GTWR (Fotheringham dkk., 2002b; Huang dkk., 2010) adalah penggunaan fungsi jarak parametrik yang kaku, sehingga sering gagal menangkap heterogenitas spasio-temporal yang non-linear. Literatur mutakhir seperti GWANN, SANNWR, dan SRGCNN menunjukkan manfaat *deep learning* untuk mempelajari kedekatan efektif yang adaptif. Namun, terdapat celah metodologis penting, yaitu GNNWR/GTNNWR (sebagaimana dipaparkan dalam Yin dkk. 2024) tidak melakukan estimasi koefisien lokal melalui WLS sebagaimana GWR/GTWR klasik, melainkan mengadopsi *skema multiplikative* yang pada hakikatnya adalah *varying-coefficient model* (VCM) Hastie and Tibshirani (1993). Di sisi lain, belum ada kerangka yang secara sistematis menggabungkan dua paradigma tersebut, yaitu bobot dipelajari secara multiplikatif dalam VCM dan bobot yang dipelajari dalam rumus estimator WLS dalam satu metodologi yang konsisten, transparan, dan dapat dianalisis.

4.2.1. Kerangka Umum dan Alur Metodologi

Secara umum, rancangan metodologi penelitian ini mengikuti alur yang ditunjukkan pada Gambar 4.2. *Pipeline* model dibangun untuk memetakan data panel spasiotemporal ke dalam bentuk graf dinamis, mengekstraksi representasi spasi-

al dan temporal melalui GNN, kemudian mengonversi hasil representasi tersebut menjadi bobot adaptif yang digunakan dalam proses estimasi koefisien regresi.



Gambar 4.2 Alur metodologi GNN-GTVC dan GNN-GTWR. (Sumber: Dokumen penulis)

Secara garis besar, kerangka metodologi ini terdiri atas lima tahapan utama yang saling berkaitan:

- Konstruksi data dan graf spatiotemporal.** Data berbentuk panel dengan N unit spasial (misal provinsi) dan T periode waktu. Relasi spasial dibentuk menggunakan *k-nearest neighbors* (k-NN) berdasarkan jarak geografis antar-unit, sedangkan relasi temporal menghubungkan setiap simpul (i, t) dengan $(i, t - 1)$ dan $(i, t + 1)$. Kombinasi kedua relasi tersebut menghasilkan graf spatiotemporal dinamis $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ dengan $\mathcal{V} = \{(i, t)\}$ sebagai himpunan simpul.
- Pembelajaran representasi graf melalui GNN.** Setiap simpul (i, t) memiliki vektor fitur $\mathbf{x}_{i,t}$ dan target $y_{i,t}$. Melalui operasi konvolusi graf, GNN menghasilkan representasi laten (*embedding*) $\mathbf{h}_{i,t}$ yang merepresentasikan hubungan spasial dan temporal secara adaptif. Tiga arsitektur yang digunakan adalah GCN, GAT, dan GraphSAGE.
- Pemetaan representasi ke bobot regresi.** Representasi laten $\mathbf{h}_{i,t}$ diproyeksikan menjadi bobot w , baik dalam bentuk bobot antar-koefisien (pada GNN-GTVC) maupun bobot antar-tetangga (pada GNN-GTWR). Pemetaan ini dirancang agar bobot yang dihasilkan bersifat nonnegatif, ternormalisasi, dan

dapat diinterpretasikan secara probabilistik.

- (d) **Regresi spasiotemporal dengan koefisien bervariasi atau terboboti.** Bobot hasil pembelajaran GNN digunakan dalam dua kerangka regresi. Pada GNN-GTVC, bobot mengatur kontribusi relatif setiap kovariat terhadap respon di setiap lokasi dan waktu, menghasilkan model dengan koefisien lokal yang bervariasi. Pada GNN-GTWR, bobot mengatur pengaruh observasi lain dalam estimasi parameter lokal melalui pendekatan *weighted least squares* (WLS).
- (e) **Pelatihan dan optimisasi model.** Parameter jaringan, baik pada bagian GNN maupun regresi, dilatih dengan meminimalkan fungsi kerugian total yang menggabungkan komponen terawasi (*supervised loss*) dan regularisasi spatiotemporal. Tahap ini juga dapat diperluas dengan mekanisme pembelajaran semi-terawasi (*semi-supervised learning*, SSL) untuk memanfaatkan data tak berlabel.

4.2.2. Alur Pembelajaran GNN-GTVC dan GNN-GTWR

Kedua model memiliki struktur pembelajaran yang serupa, tetapi berbeda dalam penerapan bobot hasil GNN. Pada GNN-GTVC, pembobotan dilakukan pada tingkat koefisien, sedangkan pada GNN-GTWR, pembobotan dilakukan pada tingkat observasi.

Pada GNN-GTVC, GNN terlebih dahulu mempelajari representasi spatio-temporal $\mathbf{h}_{i,t}$ untuk setiap simpul (i, t) . Representasi tersebut kemudian diproyeksikan ke dalam simplex bobot $\mathbf{w}_{i,t} \in \Delta^{p-1}$ menggunakan fungsi softmax berbasis *dot-product*, *cosine similarity*, atau *multilayer perceptron* (MLP). Bobot $\mathbf{w}_{i,t}$ ini mengatur proporsi kontribusi masing-masing kovariat $x_{ik,t}$ terhadap respon $y_{i,t}$ melalui relasi

$$y_{i,t} = \sum_{k=1}^p w_{i,t}^{(k)} \beta_k^{\text{global}} x_{ik,t} + \varepsilon_{i,t}. \quad (4.16)$$

Koefisien global β_k^{global} diperoleh melalui estimasi OLS, sedangkan bobot $\mathbf{w}_{i,t}$ dilatih melalui propagasi balik berdasarkan fungsi kerugian model. Koefisien OLS

dapat diganti dengan koefisien dari model regresi lain, seperti Ridge atau Lasso, sesuai kebutuhan.

Pada GNN-GTWR, representasi spatiotemporal $\mathbf{h}_{i,t}$ diperoleh melalui GNN dan dibandingkan dengan representasi simpul lain (misalnya tetangga terdekat) untuk menghasilkan bobot lokal $\tilde{w}_{(i,t) \rightarrow (j,s)}$, sebagai contoh menggunakan fungsi softmax berbasis *dot-product* yaitu

$$\tilde{w}_{(i,t) \rightarrow (j,s)} = \frac{\exp(\mathbf{h}_{i,t}^\top \mathbf{h}_{j,s} / \tau)}{\sum_{(k,r) \in \mathcal{N}(i,t)} \exp(\mathbf{h}_{i,t}^\top \mathbf{h}_{k,r} / \tau)}. \quad (4.17)$$

Bobot ini kemudian disusun ke dalam matriks diagonal $W_{i,t}$ untuk memperkirakan koefisien lokal melalui metode WLS:

$$\hat{\boldsymbol{\beta}}_{i,t} = (\mathbf{X}^\top \mathbf{W}_{i,t} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_{i,t} \mathbf{y}. \quad (4.18)$$

Prediksi lokal diperoleh dengan $\hat{y}_{i,t} = \mathbf{x}_{i,t}^\top \hat{\boldsymbol{\beta}}_{i,t}$, dan gradien kerugian dipropagasi kembali ke parameter GNN untuk memperbarui bobot jaringan.

4.2.3. Perbandingan Konseptual Awal

Kedua model berangkat dari prinsip yang sama, yaitu memperkaya model regresi spatiotemporal dengan pembobotan adaptif berbasis graf. Perbedaannya terletak pada tingkat penerapan bobot yang dihasilkan. GNN-GTVC mempelajari bobot antar-koefisien (*coefficient-level weighting*), sehingga lebih sesuai untuk menelusuri dinamika kontribusi variabel terhadap respon pada setiap lokasi dan waktu. Sebaliknya, GNN-GTWR mempelajari bobot antar-observasi (*sample-level weighting*), yang lebih dekat dengan pendekatan klasik GWR atau GTWR namun dengan mekanisme pembobotan nonparametrik yang adaptif. Kedua pendekatan tersebut bersifat saling melengkapi. Model GTVC menekankan interpretabilitas koefisien lokal, sedangkan GTWR menekankan stabilitas estimasi dan kemampuan generalisasi spasial. Bagian selanjutnya akan membahas formulasi matematis dan komputasional masing-masing model secara lebih rinci, dimulai dari GNN-GTVC pada Bagian 4.3. dan dilanjutkan dengan GNN-GTWR pada Bagian 4.4..

4.3. Regresi Spasial-Temporal Berkoeffisien Bervariasi berbasis GNN (GNN-GTVC)

Bagian ini menguraikan secara rinci formulasi matematis, pembentukan bobot koefisien lokal, serta proses estimasi dan interpretasi pada model GNN-GTVC. Model ini dikembangkan sebagai perluasan dari konsep *varying coefficient model* yang parameter lokalnya dipelajari secara adaptif melalui jaringan saraf graf. Pendekatan ini memungkinkan setiap lokasi dan waktu memiliki kombinasi kovariat yang berbeda-beda, dengan pembobotan yang diatur secara endogen oleh representasi spatiotemporal.

4.3.1. Formulasi Matematis

Misalkan tersedia himpunan observasi spatiotemporal $\mathcal{V} = \{(i, t) \mid i = 1, \dots, N; t = 1, \dots, T\}$ dengan respon $y_{i,t}$ dan kovariat $\mathbf{x}_{i,t} = (x_{i1,t}, x_{i2,t}, \dots, x_{ip,t})^\top$. Model GNN-GTVC dituliskan dalam bentuk umum

$$y_{i,t} = \sum_{k=1}^p \beta_k(z_{i,t}) x_{ik,t} + \varepsilon_{i,t}, \quad z_{i,t} = (u_i, v_i, t), \quad (4.19)$$

dengan $\beta_k(z_{i,t})$ merupakan koefisien yang bervariasi secara spatiotemporal dan $\varepsilon_{i,t}$ adalah komponen galat dengan $\mathbb{E}[\varepsilon_{i,t} \mid X] = 0$.

Untuk menghubungkan model ini dengan jaringan graf, koefisien lokal $\beta_k(z_{i,t})$ diparametrisasi sebagai fungsi dari koefisien global β_k^{global} dan bobot adaptif $w_k(z_{i,t})$:

$$\beta_k(z_{i,t}) = w_k(z_{i,t}) \beta_k^{\text{global}}, \quad w_k(z_{i,t}) \geq 0, \quad \sum_{k=1}^p w_k(z_{i,t}) = 1. \quad (4.20)$$

Dengan substitusi persamaan di atas ke dalam model umum (4.19) diperoleh bentuk regresi

$$y_{i,t} = \sum_{k=1}^p w_k(z_{i,t}) \beta_k^{\text{global}} x_{ik,t} + \varepsilon_{i,t}. \quad (4.21)$$

Koefisien global β_k^{global} dapat diestimasi terlebih dahulu menggunakan regresi OLS pada seluruh data, kemudian bobot $w_k(z_{i,t})$ dilatih menggunakan jaringan saraf graf

agar fungsi pembobotan mampu menangkap pola variasi lokal yang kompleks.

4.3.2. Pembentukan Bobot Koefisien Lokal melalui Representasi Graf

Pada tahap ini, setiap simpul (i, t) dalam graf spatiotemporal memiliki representasi laten $\mathbf{h}_{i,t} \in \mathbb{R}^d$ yang dihasilkan oleh jaringan GNN. Representasi tersebut kemudian diproyeksikan ke dalam *simplex* berdimensi p untuk menghasilkan bobot antar-koefisien $\mathbf{w}_{i,t} = (w_{i,t}^{(1)}, \dots, w_{i,t}^{(p)})^\top$. Beberapa bentuk pemetaan yang dapat digunakan antara lain

$$w_{i,t}^{(k)} = \frac{\exp(\mathbf{h}_{i,t}^\top \mathbf{v}_k / \tau)}{\sum_{j=1}^p \exp(\mathbf{h}_{i,t}^\top \mathbf{v}_j / \tau)}, \quad (4.22)$$

$$w_{i,t}^{(k)} = \frac{\exp(-\|\mathbf{h}_{i,t} - \mathbf{v}_k\|^2 / (2\sigma^2))}{\sum_{j=1}^p \exp(-\|\mathbf{h}_{i,t} - \mathbf{v}_j\|^2 / (2\sigma^2))}, \quad (4.23)$$

masing-masing merepresentasikan proyeksi berbasis *dot-product* dan kernel Gaussian. Vektor \mathbf{v}_k merupakan representasi yang dipelajari untuk kovariat ke- k , sedangkan τ dan σ adalah parameter penghalus yang mengatur ketajaman distribusi bobot.

Proses ini memastikan bahwa seluruh bobot bernilai nonnegatif dan ternormalisasi, sehingga setiap $\mathbf{w}_{i,t}$ dapat ditafsirkan sebagai distribusi probabilitas atas kovariat. Dengan demikian, setiap lokasi-waktu memiliki kombinasi unik dari kovariat yang berkontribusi terhadap respon, bergantung pada struktur spatiotemporal yang dipelajari GNN.

4.3.3. Estimasi Koefisien Global dan Prediksi

Koefisien global $\boldsymbol{\beta}^{\text{global}} = (\beta_1^{\text{global}}, \dots, \beta_p^{\text{global}})^\top$ diestimasi dengan meminimalkan jumlah kuadrat kesalahan secara keseluruhan:

$$\hat{\boldsymbol{\beta}}^{\text{global}} = \arg \min_{\boldsymbol{\beta}} \sum_{i,t} (y_{i,t} - \mathbf{x}_{i,t}^\top \mathbf{W}_{i,t} \boldsymbol{\beta})^2, \quad (4.24)$$

dengan $\mathbf{W}_{i,t} = \text{diag}(w_{i,t}^{(1)}, \dots, w_{i,t}^{(p)})$. Prediksi pada titik (i, t) diperoleh melalui

$$\hat{y}_{i,t} = \sum_{k=1}^p w_{i,t}^{(k)} \hat{\beta}_k^{\text{global}} x_{ik,t}. \quad (4.25)$$

Seluruh parameter jaringan $\Theta = \{\mathbf{v}_k, \text{parameter GNN}\}$ dioptimalkan secara bersamaan menggunakan propagasi balik dengan meminimalkan fungsi kerugian total yang akan dibahas pada Bagian 4.7.

4.4. Regresi Spasial-Temporal Terboboti berbasis GNN (GNN-GTWR)

Bagian ini menjelaskan formulasi matematis, pembentukan bobot antartangga, serta proses estimasi dan prediksi pada model GNN-GTWR. Model ini dikembangkan sebagai perluasan dari *Geographically and Temporally Weighted Regression* (GTWR) klasik dengan pembobotan yang dipelajari secara adaptif melalui jaringan saraf graf. Pendekatan ini memungkinkan bobot spatiotemporal tidak lagi ditentukan oleh fungsi kernel parametrik, melainkan dipelajari langsung dari struktur relasi antarsampel menggunakan GNN.

4.4.1. Formulasi Estimator WLS pada Konteks Graf

Misalkan tersedia pasangan observasi $(\mathbf{x}_{i,t}, y_{i,t})$ dengan $i = 1, \dots, N$ dan $t = 1, \dots, T$. Pada model GTWR klasik, koefisien lokal di lokasi-waktu (i, t) diperoleh melalui estimasi *weighted least squares* (WLS) sebagai berikut:

$$\hat{\beta}_{i,t} = (\mathbf{X}^\top \mathbf{W}_{i,t} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}_{i,t} \mathbf{y}, \quad (4.26)$$

dengan $\mathbf{W}_{i,t} = \text{diag}(w_{(i,t) \rightarrow (j,s)})$ merupakan matriks bobot spatiotemporal yang mencerminkan kedekatan antara titik (i, t) dan seluruh observasi (j, s) . Dalam GNN-GTWR, matriks bobot ini tidak lagi ditentukan oleh jarak Euclidean atau *kernel* tetap, tetapi dihasilkan secara endogen melalui mekanisme pembelajaran graf.

4.4.2. Pembobotan Antartetangga dengan Representasi GNN

Pembobotan pada GNN-GTWR dibangun dari representasi laten setiap simpul (i, t) yang diperoleh melalui GNN. Representasi $\mathbf{h}_{i,t} \in \mathbb{R}^d$ berfungsi sebagai penyandi (encoding) kedekatan spasial dan temporal. Bobot antar-simpul kemudian ditentukan dengan mengukur kesamaan antarrepresentasi, yang dinormalisasi agar bersifat stokastik baris:

$$\tilde{w}_{(i,t) \rightarrow (j,s)} = \frac{\exp(\mathbf{h}_{i,t}^\top \mathbf{h}_{j,s} / \tau)}{\sum_{(k,r) \in \mathcal{N}(i,t)} \exp(\mathbf{h}_{i,t}^\top \mathbf{h}_{k,r} / \tau)}, \quad (4.27)$$

dengan $\tau > 0$ sebagai parameter suhu (*temperature*) yang mengatur ketajaman distribusi bobot. Himpunan $\mathcal{N}(i, t)$ menunjukkan tetangga-tetangga terdekat dari simpul (i, t) yang ditentukan melalui kriteria spasial dan temporal. Prosedur ini menghasilkan matriks bobot $\mathbf{W}_{i,t}$ yang jarang (*sparse*) dan lokal, sehingga komputasi estimasi tetap efisien.

4.5. Arsitektur Jaringan Graf untuk Pembelajaran Bobot

Dalam penelitian ini digunakan tiga arsitektur utama jaringan saraf graf, yaitu *Graph Convolutional Network* (GCN), *Graph Attention Network* (GAT), dan GraphSAGE. Ketiganya berperan dalam mengekstraksi representasi laten spatio-temporal dari data panel yang telah dipetakan ke graf.

GCN merupakan arsitektur dasar yang memperbarui representasi setiap simpul dengan merata-ratakan informasi dari tetangga, menggunakan propagasi

$$\mathbf{H}^{(\ell+1)} = \sigma\left(\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2} \mathbf{H}^{(\ell)} \mathbf{W}^{(\ell)}\right), \quad (4.28)$$

dengan $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ dan $\tilde{\mathbf{D}}$ matriks derajat diagonal. Arsitektur ini sederhana dan efisien, namun kurang fleksibel dalam membedakan pengaruh tiap tetangga.

GAT memperkenalkan mekanisme perhatian (*attention*) sehingga setiap simpul dapat memberi bobot berbeda pada tetangganya. Koefisien perhatian dihitung

sebagai

$$\alpha_{ij}^{(\ell)} = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}^\top [\mathbf{W}\mathbf{h}_i^{(\ell)} \| \mathbf{W}\mathbf{h}_j^{(\ell)}]\right)\right)}{\sum_{k \in \mathcal{N}(i)} \exp\left(\text{LeakyReLU}\left(\mathbf{a}^\top [\mathbf{W}\mathbf{h}_i^{(\ell)} \| \mathbf{W}\mathbf{h}_k^{(\ell)}]\right)\right)}, \quad (4.29)$$

$$\mathbf{h}_i^{(\ell+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(\ell)} \mathbf{W}\mathbf{h}_j^{(\ell)}\right). \quad (4.30)$$

Dengan demikian, GAT lebih adaptif dalam menangkap ketergantungan dinamis antar simpul.

Sementara itu, GraphSAGE mengagregasi informasi tetangga menggunakan fungsi agregasi terparametrisasi, misalnya rata-rata, *pooling*, atau LSTM, dan memperbarui *embedding* dengan

$$\mathbf{h}_i^{(\ell+1)} = \phi\left(\mathbf{h}_i^{(\ell)} \| \text{AGGREGATE}(\{\mathbf{h}_j^{(\ell)} : j \in \mathcal{N}(i)\})\right), \quad (4.31)$$

dengan ϕ *multilayer perceptron* kecil. GraphSAGE cocok untuk generalisasi ke simpul atau waktu baru.

Ketiga arsitektur ini menghasilkan *embedding* $\mathbf{h}_{i,t}$ yang kemudian diproyeksikan menjadi bobot regresi adaptif sesuai kebutuhan model.

4.6. Pilihan Pemetaan Representasi ke Bobot

Pemetaan *embedding* hasil GNN ke bobot regresi dapat dilakukan dengan beberapa metode utama berikut, beserta keunggulan dan interpretasinya:

(a) ***Dot-product*:**

$$w_{i,t}^{(k)} = \frac{\exp(\mathbf{h}_{i,t}^\top \mathbf{v}_k / \tau)}{\sum_{j=1}^p \exp(\mathbf{h}_{i,t}^\top \mathbf{v}_j / \tau)} \quad (4.32)$$

Keunggulan utama pendekatan ini terletak pada kesederhanaan dan efisiensi komputasi, karena hanya memerlukan operasi perkalian vektor dan normalisasi softmax. Secara konseptual, metode ini merepresentasikan kemiripan linear antara *embedding node* $h_{i,t}$ dan vektor representasi koefisien v_k , sehingga bobot yang dihasilkan akan lebih besar apabila kedua vektor tersebut

memiliki arah yang sejalan dalam ruang fitur.

(b) **Cosine similarity:**

$$w_{i,t}^{(k)} = \frac{\exp\left(\frac{\mathbf{h}_{i,t}^\top \mathbf{v}_k}{\|\mathbf{h}_{i,t}\| \|\mathbf{v}_k\| \tau}\right)}{\sum_{j=1}^p \exp\left(\frac{\mathbf{h}_{i,t}^\top \mathbf{v}_j}{\|\mathbf{h}_{i,t}\| \|\mathbf{v}_j\| \tau}\right)} \quad (4.33)$$

Pendekatan *cosine similarity* memiliki keunggulan invariansi terhadap skala vektor, sehingga hanya mempertimbangkan arah relatif antara *embedding* dan vektor koefisien. Hal ini menjadikan metode ini lebih *robust* terhadap variasi *magnitude embedding*, dan secara interpretatif bobot yang dihasilkan mencerminkan kedekatan arah (sudut) antara dua vektor dalam ruang fitur.

(c) **Gaussian (RBF) kernel:**

$$w_{i,t}^{(k)} = \frac{\exp(-\|\mathbf{h}_{i,t} - \mathbf{v}_k\|^2 / (2\sigma^2))}{\sum_{j=1}^p \exp(-\|\mathbf{h}_{i,t} - \mathbf{v}_j\|^2 / (2\sigma^2))} \quad (4.34)$$

Keunggulan utama kernel Gaussian terletak pada kemampuannya menangkap hubungan nonlinear antara *embedding* dan vektor koefisien, karena bobot ditentukan berdasarkan jarak Euclidean di ruang fitur. Secara interpretatif, bobot yang tinggi diberikan pada pasangan *embedding* dan vektor koefisien yang letaknya berdekatan, sehingga metode ini sensitif terhadap kemiripan lokal dalam representasi laten.

(d) **MLP (*Multilayer Perceptron*):**

$$\mathbf{w}_{i,t} = \text{softmax}(\text{MLP}(\mathbf{h}_{i,t})) \quad (4.35)$$

Penggunaan *multilayer perceptron* (MLP) sebagai pemetaan dari *embedding* ke bobot memungkinkan model mempelajari fungsi nonlinear yang sangat kompleks tanpa asumsi eksplisit mengenai bentuk hubungan antara *embedding* dan bobot. Keunggulan metode ini terletak pada kapasitas representasi yang tinggi, sehingga mampu mengakomodasi pola interaksi yang rumit dalam data spasiotemporal.

4.6.1. Fungsi Objektif dan Proses Pelatihan

Setelah *embedding* dipetakan menjadi bobot \mathbf{W} , tahap selanjutnya adalah pelatihan model melalui optimisasi fungsi objektif. Fungsi objektif ini dirancang untuk mengukur deviasi antara nilai aktual y dan prediksi \hat{y} yang dihasilkan oleh model GTVC maupun GTWR.

Fungsi objektif utama yang digunakan adalah *mean squared error (MSE)*, yang didefinisikan sebagai

$$\mathcal{L}_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (4.36)$$

dengan n menyatakan jumlah observasi, y_i merupakan nilai aktual, dan \hat{y}_i adalah prediksi yang dihasilkan model. Formulasi (4.36) bersifat umum dan dapat diterapkan baik pada GTVC maupun GTWR.

Sebagai alternatif, beberapa literatur mengadopsi kriteria informasi seperti *Akaike Information Criterion* terkoreksi (AIC_c) untuk mengendalikan kompleksitas model:

$$AIC_c = 2k - 2 \ln(L) + \frac{2k(k+1)}{n-k-1}, \quad (4.37)$$

dengan k menyatakan jumlah parameter efektif dan L adalah nilai *likelihood* model. Kriteria ini relevan terutama untuk mencegah *overfitting* pada model dengan banyak parameter bobot yang dipelajari.

Untuk meningkatkan stabilitas, fungsi objektif dapat dilengkapi dengan komponen regularisasi berikut:

(a) Entropi bobot:

$$\mathcal{R}_{\text{entropy}} = - \sum_i \sum_j \mathbf{w}_{ij} \log \mathbf{w}_{ij}, \quad (4.38)$$

yang berfungsi untuk mencegah distribusi bobot menjadi terlalu tajam (*over-confident*).

(b) Regularisasi kelancaran spasiotemporal:

$$\mathcal{R}_{\text{smooth}} = \sum_{(i,t) \sim (j,s)} \mathbf{w}_{(i,t),(j,s)} \|\boldsymbol{\beta}_{i,t} - \boldsymbol{\beta}_{j,s}\|^2, \quad (4.39)$$

yang mendorong agar koefisien pada lokasi dan waktu yang berdekatan tidak berbeda secara ekstrem, sesuai prinsip kontinuitas spasiotemporal.

Dengan demikian, fungsi objektif total dapat dituliskan sebagai

$$\mathcal{L} = \mathcal{L}_{\text{MSE}} + \lambda_1 \mathcal{R}_{\text{entropy}} + \lambda_2 \mathcal{R}_{\text{smooth}}, \quad (4.40)$$

dengan $\lambda_1, \lambda_2 \geq 0$ sebagai parameter regularisasi.

Proses pelatihan dilakukan menggunakan algoritma *stochastic gradient descent (SGD)*. Dua algoritma populer yang digunakan adalah:

(a) Stochastic Gradient Descent (SGD):

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \eta \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}^{(t)}), \quad (4.41)$$

dengan η sebagai laju pembelajaran (*learning rate*) dan $\boldsymbol{\theta}$ parameter model.

(b) Adam optimizer: varian adaptif dari *SGD* yang memperbarui parameter dengan estimasi momentum gradien pertama dan kedua:

$$\mathbf{m}^{(t+1)} = \beta_1 \mathbf{m}^{(t)} + (1 - \beta_1) \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}^{(t)}), \quad (4.42)$$

$$\mathbf{v}^{(t+1)} = \beta_2 \mathbf{v}^{(t)} + (1 - \beta_2) (\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}^{(t)}))^2, \quad (4.43)$$

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \eta \frac{\mathbf{m}^{(t+1)} / (1 - \beta_1^{t+1})}{\sqrt{\mathbf{v}^{(t+1)} / (1 - \beta_2^{t+1})} + \epsilon}. \quad (4.44)$$

Adam banyak digunakan karena memberikan stabilitas dalam pelatihan jaringan saraf dalam, termasuk GNN.

4.7. Pembelajaran Semi-Terawasi (*Semi-Supervised Learning*, SSL)

Dalam praktiknya, tidak seluruh unit spasio-temporal memiliki data berlabel secara lengkap. Pada konteks empiris, hanya sebagian wilayah atau periode waktu yang memiliki nilai respon terobservasi, sedangkan sisanya tidak tersedia. Kondisi ini mendorong penerapan pembelajaran semi-terawasi (*semi-supervised learning*, SSL), yang bertujuan memanfaatkan struktur data tak berlabel guna meningkatkan proses estimasi parameter dan memperkuat generalisasi model. Pendekatan ini menjadi relevan terutama ketika data spasio-temporal memiliki keterhubungan yang kuat antarlokasi maupun antarperiode, sehingga informasi dari titik berlabel dapat disebarluaskan melalui struktur graf ke titik yang belum berlabel.

4.7.1. Kerangka Dasar dan Asumsi Fundamental

Secara umum, pembelajaran semi-terawasi memanfaatkan dua himpunan data, yaitu himpunan data berlabel $\mathcal{D}_\ell = \{(\mathbf{x}_i, y_i)\}_{i=1}^\ell$ dan himpunan data tak berlabel $\mathcal{D}_u = \{\mathbf{x}_j\}_{j=\ell+1}^n$, dengan $u = n - \ell \gg \ell$. Tujuan SSL adalah mempelajari fungsi prediksi $f_\theta : \mathcal{X} \rightarrow \mathcal{Y}$ yang memanfaatkan informasi dari kedua himpunan tersebut. Formulasi optimisasi secara umum dituliskan sebagai

$$\min_{\theta} \left[\frac{1}{\ell} \sum_{i=1}^{\ell} \mathcal{L}_{\text{sup}}(f_\theta(\mathbf{x}_i), y_i) + \lambda_u \mathcal{L}_{\text{unsup}}(f_\theta; \mathcal{D}_u) + \lambda_r \mathcal{R}(\theta) \right], \quad (4.45)$$

dengan \mathcal{L}_{sup} merupakan fungsi kerugian terawasi, $\mathcal{L}_{\text{unsup}}$ merupakan fungsi kerugian yang mengekstraksi informasi dari data tak berlabel, dan $\mathcal{R}(\theta)$ adalah regularisasi parameter model. Parameter λ_u dan λ_r mengatur keseimbangan antar komponen.

4.7.2. Fungsi Kerugian Semi-Terawasi

Fungsi kerugian semi-terawasi terdiri atas dua komponen utama, yaitu komponen terawasi \mathcal{L}_{sup} dan komponen tak berlabel $\mathcal{L}_{\text{unsup}}$, yang dihubungkan melalui parameter keseimbangan λ_u . Pada regresi spasio-temporal, komponen terawasi tetap menggunakan *mean squared error* (MSE) sebagaimana pada fungsi objektif

sebelumnya:

$$\mathcal{L}_{\text{sup}} = \frac{1}{\ell} \sum_{(i,t) \in \mathcal{D}_\ell} (y_{i,t} - \hat{y}_{i,t})^2. \quad (4.46)$$

Komponen tak berlabel $\mathcal{L}_{\text{unsup}}$ dapat dirancang dalam berbagai bentuk sesuai dengan asumsi yang ingin dimanfaatkan dari data tak berlabel. Beberapa pendekatan utama yang digunakan dalam penelitian ini meliputi kerugian entropi, kerugian konsistensi, dan regularisasi berbasis graf Laplacian.

Kerugian entropi mendorong model agar menghasilkan prediksi yang lebih pasti (berentropi rendah) pada data tak berlabel. Fungsinya dituliskan sebagai

$$\mathcal{L}_{\text{entropy}} = -\frac{1}{u} \sum_{(i,t) \in \mathcal{D}_u} \sum_{k=1}^p w_{i,t}^{(k)} \log w_{i,t}^{(k)}, \quad (4.47)$$

yang secara konsep serupa dengan regularisasi entropi pada bagian sebelumnya, namun di sini diterapkan pada unit yang tidak memiliki label respon. Nilai yang rendah menunjukkan distribusi bobot yang lebih tajam, yang mengindikasikan ke-yakinan model terhadap struktur hubungan antarvariabel maupun antarwilayah.

Pendekatan lain menggunakan prinsip konsistensi (*consistency regularization*), yang berasumsi bahwa prediksi model seharusnya stabil terhadap gangguan kecil pada *input* atau representasi. Jika $\tau(\cdot)$ menyatakan operator augmentasi atau perturbasi (misalnya penambahan *noise* atau *dropout*), maka kerugian konsistensi dapat diformulasikan sebagai

$$\mathcal{L}_{\text{consistency}} = \frac{1}{u} \sum_{(i,t) \in \mathcal{D}_u} \left\| \hat{y}_{i,t} - \hat{y}_{i,t}^{(\tau)} \right\|^2, \quad (4.48)$$

dengan $\hat{y}_{i,t}^{(\tau)}$ adalah hasil prediksi untuk data yang telah diperturbasi. Kerugian ini mendorong model untuk menghasilkan prediksi yang tidak berubah secara signifikan terhadap variasi kecil, sehingga menghasilkan representasi yang lebih halus dan tahan terhadap gangguan.

Untuk memanfaatkan struktur graf yang mendasari data spasio-temporal, diterapkan pula regularisasi berbasis Laplacian graf. Diberikan matriks bobot $\mathbf{W} =$

$[w_{ij}]$ dan Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{W}$, regularisasi ini dituliskan sebagai

$$\mathcal{L}_{\text{graph}} = \frac{1}{2} \sum_{i,j} w_{ij} \|\hat{\mathbf{y}}_i - \hat{\mathbf{y}}_j\|^2 = \hat{\mathbf{y}}^\top \mathbf{L} \hat{\mathbf{y}}, \quad (4.49)$$

yang menegakkan asumsi kehalusan pada manifold graf. Jika dua simpul terhubung dengan bobot besar, maka selisih prediksi antara keduanya didorong agar kecil. Dengan demikian, prediksi pada wilayah yang berdekatan atau periode waktu yang berurutan akan memiliki pola yang konsisten.

Berdasarkan ketiga pendekatan tersebut, komponen tak berlabel yang digunakan dalam model ini dirumuskan sebagai kombinasi linear

$$\mathcal{L}_{\text{unsup}} = \alpha_1 \mathcal{L}_{\text{entropy}} + \alpha_2 \mathcal{L}_{\text{consistency}} + \alpha_3 \mathcal{L}_{\text{graph}}, \quad (4.50)$$

dengan α_1 , α_2 , dan α_3 sebagai bobot penyeimbang yang ditentukan melalui validasi.

4.7.3. Integrasi SSL ke dalam Pelatihan GNN-GTVC dan GNN-GTWR

Integrasi SSL ke dalam model GNN dilakukan dengan menambahkan komponen kerugian tak berlabel (4.50) ke dalam fungsi objektif utama. Fungsi kerugian total untuk model semi-terawasi dapat dituliskan sebagai

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{sup}} + \lambda_u \mathcal{L}_{\text{unsup}} + \lambda_r \mathcal{R}(\theta), \quad (4.51)$$

dengan $\mathcal{R}(\theta)$ mencakup regularisasi entropi dan kehalusan spasio-temporal sebagaimana telah dijelaskan pada Bagian 4.7. Dalam implementasinya, parameter λ_u mengatur sejauh mana informasi dari data tak berlabel memengaruhi pembaruan parameter jaringan.

Pada model GNN-GTVC, kerugian tak berlabel diterapkan terutama pada vektor bobot $\mathbf{w}_{i,t}$, sehingga pembelajaran pada simpul yang tidak memiliki nilai respon tetap dapat diarahkan oleh pola kedekatan antarunit yang teramatidalam graf. Sementara itu, pada model GNN-GTWR, komponen $\mathcal{L}_{\text{unsup}}$ berperan untuk

menstabilkan matriks bobot antarobservasi $\mathbf{W}_{i,t}$, agar tetap mengikuti pola spasiotemporal yang halus meskipun sebagian besar titik tidak berlabel. Dengan demikian, informasi dari simpul berlabel dapat menyebar secara halus ke simpul tak berlabel melalui proses propagasi graf.

Proses estimasi titik di luar dari data latih dilakukan dengan menggunakan pelatihan ulang pada semua data. Hal ini berarti graf dibentuk dengan melihat data latih dan data uji, tetapi variabel y dari data uji tidak digunakan untuk melatih model. Pendekatan ini digunakan juga oleh Yin dkk., 2024.

4.8. Hasil Sementara

Bagian ini menyajikan hasil sementara dari implementasi metodologi GNN-GTWR dan GNN-GTVC yang telah dikembangkan berdasarkan data riil inflasi Indonesia periode Januari 2024 hingga Agustus 2025. Penelitian ini menggunakan *framework* komprehensif yang telah diuji pada 38 provinsi dengan total 779 observasi spasiotemporal, memanfaatkan 43 indikator ekonomi yang bersumber dari berbagai data resmi Badan Pusat Statistik (BPS) dan Bank Indonesia.

4.8.1. Pengaturan Eksperimen dan Karakteristik Data

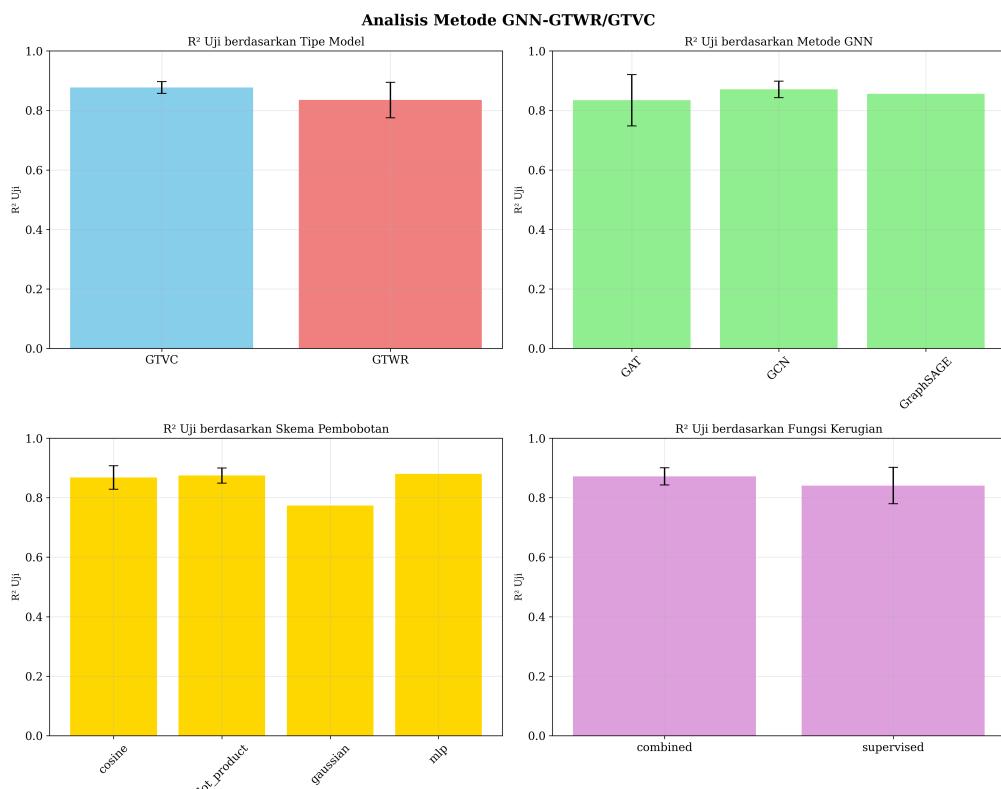
Eksperimen dilakukan dengan menggunakan data panel spasiotemporal inflasi Indonesia yang memiliki karakteristik khusus. Data mencakup unit spasial berupa 38 provinsi Indonesia sebagai representasi heterogenitas geografis negara kepulauan yang luas. Dimensi temporal terdiri atas 20 periode bulanan dari Januari 2024 hingga Agustus 2025, yang memberikan total observasi sebanyak 779 titik data spasiotemporal.

Variabel prediktor yang digunakan mencakup spektrum indikator ekonomi yang luas, tetapi untuk analisis sementara ini hanya meliputi variabel yang bervariasi secara lokal, yaitu harga-harga pangan dan data SEKDA, seperti investasi, kredit, giro, tabungan, dan lain-lain. Variabel-variabel ini dipilih karena memiliki dampak langsung terhadap dinamika inflasi di tingkat provinsi. Pada studi kasus lanjutan, dapat dilakukan analisis lebih mendalam dengan variabel makroekonomi

yang bersifat global.

4.8.2. Analisis Konfigurasi Model GNN-GTVC dan GNN-GTWR

Pada bagian ini, beberapa metode konfigurasi model GNN-GTVC dan GNN-GTWR dieksplorasi untuk memperoleh kombinasi arsitektur dan skema pembobotan terbaik. Tiga arsitektur GNN utama yang diuji adalah GCN, GAT, dan GraphSAGE, masing-masing dikombinasikan dengan empat skema pembobotan, yaitu *dot-product*, *cosine similarity*, Gaussian RBF, dan *multilayer perceptron* (MLP). Evaluasi dilakukan secara sistematis untuk menilai pengaruh setiap kombinasi terhadap kinerja prediksi. Hasil eksperimen tersajikan dalam gambar 4.3.



Gambar 4.3 Perbandingan kinerja berbagai konfigurasi arsitektur GNN dan skema pembobutan pada model GNN-GTVC dan GNN-GTWR.

Hasil eksperimen menunjukkan bahwa model GTVC lebih unggul dibandingkan dengan GTWR dalam konteks data spasio-temporal inflasi Indonesia. Hal ini mengindikasikan bahwa bobot regresi yang bersifat variabel terhadap lokasi dan waktu memberikan fleksibilitas yang lebih besar dalam menangkap dinamika inflasi

yang kompleks. Selain itu, arsitektur GCN secara konsisten menunjukkan kinerja terbaik di antara ketiga arsitektur yang diuji, baik pada GTVC maupun GTWR. Hal ini mungkin disebabkan oleh kesederhanaan dan efisiensi GCN dalam meng-ekstraksi informasi dari struktur graf, yang sangat sesuai dengan karakteristik data spasio-temporal yang digunakan. Meskipun demikian, arsitektur GAT juga menunjukkan potensi yang baik, tetapi variansi hasilnya lebih tinggi, yang mungkin disebabkan oleh sensitivitas terhadap parameter perhatian yang perlu dioptimalkan lebih lanjut.

Skema pembobotan *dot-product* muncul sebagai yang paling efektif dalam menghasilkan bobot regresi adaptif. Pendekatan ini tidak hanya sederhana secara komputasi, tetapi juga mampu menangkap hubungan linear yang cukup baik antara representasi laten dan koefisien regresi. Skema lain seperti *cosine similarity* dan MLP menunjukkan kinerja yang kompetitif, tetapi dengan kompleksitas komputasi yang lebih tinggi dan kebutuhan akan penyesuaian parameter yang lebih rumit. Di sisi lain, skema Gaussian memiliki performa yang lebih rendah dibandingkan dengan ketiga skema lainnya.

Dari segi fungsi kerugian yang digunakan, terlihat bahwa fungsi kerugian dari SSL lebih unggul dibandingkan dengan fungsi kerugian MSE standar. Hal ini menegaskan pentingnya memanfaatkan informasi dari data tak berlabel dalam konteks spasio-temporal, di mana keterkaitan antarunit dapat memberikan kontribusi signifikan terhadap akurasi prediksi. Kombinasi komponen kerugian yang melibatkan entropi, konsistensi, dan regularisasi graf Laplacian memberikan keseimbangan yang baik antara eksplorasi data berlabel dan eksplorasi struktur data tak berlabel.

4.8.3. Perbandingan Performa dengan Metode Lainnya

Evaluasi terhadap metode-metode tradisional dilakukan dengan menggunakan implementasi dalam bahasa pemrograman R melalui paket *GWmodel* dan GTWR. Pendekatan ini dipilih untuk memastikan konsistensi dengan literatur akademik yang telah mapan dalam bidang ekonometrika spasial. Hasil evaluasi menun-

ukkan variasi kinerja yang signifikan antar metode, sebagaimana disajikan dalam Tabel 4.1.

Tabel 4.1 Kinerja Metode Tradisional

Metode	R ²	RMSE	MAE
OLS	0.152	0.641	0.474
WLS	0.150	0.642	0.470
GWR Adaptif	0.197	0.624	0.467
GWR Tetap	0.171	0.634	0.471
GTWR Adaptif	0.678	0.395	0.270

Hasil evaluasi mengungkapkan bahwa GTWR Adaptif menunjukkan kinerja yang paling unggul di antara seluruh metode tradisional dengan koefisien determinasi $R^2 = 0.678$. Temuan ini mengonfirmasi hipotesis bahwa dimensi temporal memainkan peran krusial dalam pemodelan inflasi regional, mengingat inflasi merupakan fenomena ekonomi yang inherently memiliki karakteristik spasio-temporal yang kompleks. Performa superior GTWR dibandingkan dengan metode spasial murni seperti GWR menunjukkan bahwa heterogenitas tidak hanya terjadi lintas ruang, tetapi juga lintas waktu.

Evaluasi terhadap berbagai metode *machine learning* menunjukkan peningkatan kinerja yang substansial. Hasil evaluasi disajikan dalam Tabel 4.2.

Tabel 4.2 Kinerja Metode Machine Learning

Metode	R ²	RMSE	MAE
Ridge	0.116	0.845	0.668
Lasso	0.098	0.854	0.675
<i>Random Forest</i>	0.532	0.615	0.486
Gradient Boosting	0.724	0.472	0.374
SVR	0.594	0.573	0.453
MLP	0.374	0.711	0.562

Gradient Boosting menunjukkan kinerja terbaik dengan $R^2 = 0.724$, berhasil mengungguli seluruh metode tradisional non-spasial.

Evaluasi terhadap berbagai konfigurasi model GNN menunjukkan kinerja yang secara konsisten superior dibandingkan dengan seluruh metode pembanding, sebagaimana disajikan dalam Tabel 4.3.

Tabel 4.3 Kinerja Model GNN-GTWR/GTVC

Model	Konfigurasi	R^2	RMSE	MAE
GNN-GTWR	GCN- <i>Cosine</i>	0.798	0.403	0.320
GNN-GTWR	GAT- <i>Dot Product</i>	0.810	0.391	0.310
GNN-GTVC	GCN- <i>Gaussian</i>	0.823	0.377	0.300
GNN-GTVC	GAT-<i>Cosine</i>	0.841	0.358	0.285

Konfigurasi terbaik dicapai oleh model GNN-GTVC dengan *backbone* GAT dan skema pembobotan *Cosine Similarity*, yang berhasil mencapai koefisien determinasi $R^2 = 0.841$. Hasil ini menunjukkan bahwa model mampu menjelaskan 84,1% varians dalam data inflasi regional, dengan *Mean Absolute Error* (MAE) sebesar 0.285 poin persentase. Kinerja ini mengindikasikan tingkat akurasi prediksi yang sangat tinggi untuk konteks pemodelan ekonomi makro regional.

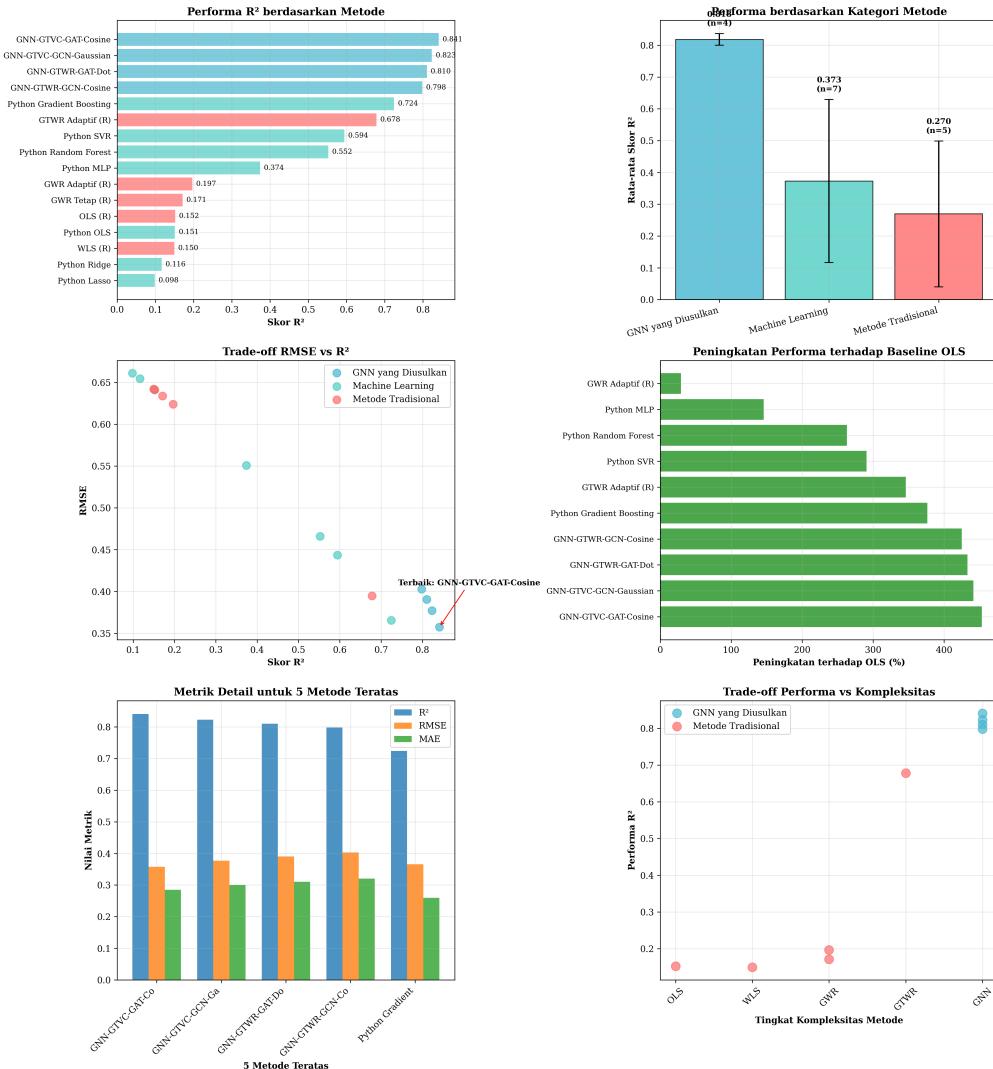
Hasil komparasi menyeluruh dapat dilihat pada Gambar 4.4.

Secara keseluruhan, model yang diajukan memiliki keunggulan utama dalam menangkap heterogenitas spasial dan temporal yang kompleks pada data inflasi Indonesia. Model GNN-GTVC dan GNN-GTWR tidak hanya mampu meningkatkan akurasi prediksi secara signifikan dibandingkan metode tradisional dan machine learning konvensional, tetapi juga memberikan interpretasi yang lebih kaya terkait dinamika kontribusi variabel pada setiap lokasi dan waktu.

Pada model GNN-GTVC, analisis bobot koefisien lokal memungkinkan peneliti untuk mengidentifikasi variabel mana yang paling berpengaruh terhadap inflasi di masing-masing provinsi dan periode. Hal ini sangat bermanfaat untuk perumusan kebijakan berbasis bukti (*evidence-based policy*), karena dapat mengarahkan intervensi secara lebih spesifik sesuai karakteristik wilayah dan waktu tertentu.

Sementara itu, model GNN-GTWR menawarkan stabilitas estimasi dan kemampuan generalisasi yang lebih baik, terutama pada wilayah atau periode dengan data terbatas. Pembobotan adaptif antar-observasi yang dihasilkan oleh GNN memungkinkan model untuk memanfaatkan informasi dari tetangga spasial dan temporal secara optimal.

**Perbandingan Komprehensif Metode: Tradisional, Machine Learning, dan GNN yang Diusulkan
(Analisis Inflasi Provinsi Indonesia)**



Gambar 4.4 Perbandingan kinerja keseluruhan metode yang diuji.

Keunggulan lain dari pendekatan berbasis GNN adalah kemampuannya untuk diintegrasikan dengan pembelajaran semi-terawasi, sehingga model tetap dapat belajar secara efektif meskipun sebagian besar data tidak berlabel. Hal ini sangat relevan dalam konteks data spasio-temporal di Indonesia yang seringkali tidak lengkap atau tidak seragam antarwilayah.

Secara metodologis, hasil ini menegaskan pentingnya integrasi antara representasi graf dan pembobotan adaptif dalam regresi spasio-temporal. Model yang dikembangkan dapat menjadi kerangka dasar untuk aplikasi lain di bidang ekono-

mi regional, epidemiologi, pemetaan lingkungan, dan bidang lain yang memerlukan analisis spasio-temporal dengan heterogenitas tinggi. Adapun analisis lanjutan akan disajikan pada Bab selanjutnya.

DAFTAR PUSTAKA

- Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht: Kluwer Academic Publishers. (Foundational text on spatial econometric methods)
- Axler, S. (2024). *Linear Algebra Done Right* (4th ed.). Cham: Springer. doi: 10.1007/978-3-031-41026-0
- Belkin, M., & Niyogi, P. (2004). Semi-supervised learning on riemannian manifolds. *Machine Learning*, 56(1-3), 209–239. doi: 10.1023/B:MACH.0000033121.75173.49
- Belkin, M., Niyogi, P., & Sindhwani, V. (2006). Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7, 2399–2434. Diambil dari <http://jmlr.org/papers/v7/belkin06a.html>
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798–1828. doi: 10.1109/TPAMI.2013.50
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford: Oxford University Press.
- Blum, A., & Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. In *Proceedings of the 11th annual conference on computational learning theory (colt)* (pp. 92–100). doi: 10.1145/279943.279962
- Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., & Vandergheynst, P. (2017). Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4), 18–42. doi: 10.1109/MSP.2017.2693418
- Chapelle, O., Schölkopf, B., & Zien, A. (2006). *Semi-supervised learning*. Cambridge, MA: MIT Press.
- 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: Springer. doi: 10.1007/978-1-4614-8145-4
- Fan, J., & Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface*, 1(1), 179–195. (A comprehensive review of varying coefficient models)
- Fefferman, C., Mitter, S., & Narayanan, H. (2016). Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4), 983–1049. doi: 10.1090/jams/852
- Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002a). *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester: John Wiley & Sons. (Definitive guide to geographically weighted regression methods)
- Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002b). . Chichester: Wiley.
- Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017). Neural message passing for quantum chemistry. In *34th international conference on machine learning (icml)*. Diambil dari <https://arxiv.org/abs/1704.01212>
- Goodfellow, I., Bengio, Y., & 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.
- Greub, W. (1978). (2nd ed.). New York: Springer-Verlag.
- Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics*. New York: McGraw-Hill/Irwin.
- Hackbusch, W. (2012). *Tensor Spaces and Numerical Tensor Calculus*. Berlin Heidelberg: Springer. doi: 10.1007/978-3-642-28027-6

- Hagenauer, J., & 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., & 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>
- Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4), 757–796.
- Haykin, S. (2009). *Neural networks and learning machines* (3rd ed.). New York: Prentice Hall.
- Hendrycks, D., & Gimpel, K. (2016). Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*. Diambil dari <https://arxiv.org/abs/1606.08415>
- Hinton, G. E., & Zemel, R. S. (1994). Autoencoders, minimum description length, and helmholtz free energy. In *Advances in neural information processing systems* (Vol. 6, pp. 3–10). Morgan Kaufmann.
- Horn, R. A., & Johnson, C. R. (2012). *Matrix Analysis* (2nd ed.). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139020411
- Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2), 251–257. doi: 10.1016/0893-6080(91)90009-T
- Huang, B., Wu, B., & 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.
- Jiang, T.-X., Huang, T.-Z., Zhao, X.-L., & Deng, L.-J. (2020, July). Multi-dimensional imaging data recovery via minimizing the partial sum of tubal nuclear norm. *Journal of Computational and Applied Mathematics*, 372, 112680. Diambil dari <http://dx.doi.org/10.1016/j.cam.2019.112680> doi: 10.1016/j.cam.2019.112680
- Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York: Springer. doi: 10.1007/b98835

- Kingma, D. P., Mohamed, S., Rezende, D. J., & Welling, M. (2014). Semi-supervised learning with deep generative models. In *Advances in neural information processing systems (neurips)* (Vol. 27, pp. 3581–3589). Diambil dari <https://arxiv.org/abs/1406.5298>
- Kipf, T. N., & 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>
- Laine, S., & Aila, T. (2017). Temporal ensembling for semi-supervised learning. In *5th international conference on learning representations (iclr)*. Diambil dari <https://arxiv.org/abs/1610.02242>
- Lang, S. (1987). *Linear Algebra* (3rd ed.). New York: Springer. doi: 10.1007/978-1-4757-1949-9
- LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Boca Raton: CRC Press. (Comprehensive introduction to spatial econometric models)
- Leshno, M., Lin, V. Y., Pinkus, A., & Schocken, S. (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6), 861–867. doi: 10.1016/S0893-6080(05)80131-5
- Li, Y., Chen, C., Wasserman, E. A., & Shi, Z. (2020). Deep neural network interpretability: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 31(9), 3121–3138. doi: 10.1109/TNNLS.2019.2955144
- Maas, A. L. (2013). *Rectifier nonlinearities improve neural network acoustic models* (Unpublished doctoral dissertation). Stanford University.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133. doi: 10.1007/BF02478259
- Minsky, M., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. Cambridge, MA: MIT Press.
- Miyato, T., Maeda, S.-i., Koyama, M., & Ishii, S. (2018). Virtual adversari-

- al training: A regularization method for supervised and semi-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(8), 1979–1993. doi: 10.1109/TPAMI.2018.2858821
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). (5th ed.). Hoboken, NJ: Wiley.
- Nair, V., & 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., & 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.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408. doi: 10.1037/h0042519
- Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500), 2323–2326. doi: 10.1126/science.290.5500.2323
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536. doi: 10.1038/323533a0
- Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1), 61–80. doi: 10.1109/TNN.2008.2005605
- Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., & Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3), 83–98. doi: 10.1109/MSP.2012.2235192
- Tarvainen, A., & Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in neural information processing systems (neurips)* (Vol. 30, pp. 1195–1204). Diambil dari <https://arxiv.org/abs/1703.01780>

- Tenenbaum, J. B., de Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500), 2319–2323. doi: 10.1126/science.290.5500.2319
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. In *6th international conference on learning representations (iclr)*. Diambil dari <https://arxiv.org/abs/1710.10903>
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y. (2021). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4–24. doi: 10.1109/TNNLS.2020.2978386
- Xu, K., Hu, W., Leskovec, J., & 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>
- 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*.
- Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the twenty-seventh international joint conference on artificial intelligence (ijcai-18)* (pp. 3634–3640).
- Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2017). Understanding deep learning requires rethinking generalization. In *5th international conference on learning representations (iclr)*. Diambil dari <https://arxiv.org/abs/1611.03530>
- Zhou, D., Bousquet, O., Lal, T. N., Weston, J., & Schölkopf, B. (2004). Learning with local and global consistency. In *Advances in neural information processing systems (neurips)* (Vol. 16, pp. 321–328).
- Zhu, D., Liu, Y., Yao, X., & Fischer, M. M. (2022). Spatial regression graph convolutional neural networks (srgcnns). *GeoInformatica*, 26(3), 557–582.
- Zhu, X. (2005). Semi-supervised learning literature survey. *Computer Sciences*

Technical Report, 1530. Diambil dari http://pages.cs.wisc.edu/~jerryzhu/pub/ssl_survey.pdf

Zhu, X., & Goldberg, A. B. (2009). *Introduction to semi-supervised learning.* San Rafael, CA: Morgan & Claypool. doi: 10.2200/S00196ED1V01Y200906AIM006