Bayesian Information Criterion (BIC)

Rudy C Tarumingkeng

Oleh:
Prof Ir Rudy C Tarumingkeng, PhD
Guru Besar Manajemen, NUP: 9903252922
Ketua Senat Akademik IBM-ASMI

rudyct75@gmail.com
Bogor, Indonesia
22 April 2025

Bayesian Information Criterion (BIC) adalah salah satu kriteria yang sering digunakan untuk melakukan pemilihan model (model selection) dalam statistik dan machine learning. Dikembangkan oleh Gideon Schwarz (1978), BIC bertujuan menyeimbangkan dua hal: seberapa baik model menjelaskan data (goodness of fit) dan seberapa sederhana model tersebut (parsimonious). Secara naratif, bayangkan seorang peneliti melakukan regresi linier terhadap data epidemiologi—ia menguji beberapa model dengan jumlah variabel prediktor yang berbeda. Model yang "terlalu kompleks" mungkin cocok pada data sampel, tetapi memiliki risiko overfitting; sebaliknya, model yang "terlalu sederhana" bisa gagal menangkap pola penting. BIC membantu memilih model yang optimal dengan memberikan "hukuman" (penalty) bagi setiap parameter tambahan di dalam model.

Secara matematis, BIC dirumuskan sebagai berikut:

$$\mathrm{BIC} = -2\,\ln\big(\hat{L}\big) \;+\; k\,\ln(n)$$

di mana:

- $\hat{L}$ adalah likelihood maksimum dari model (nilai probabilitas data di bawah model terbaik),
- k adalah jumlah total parameter yang diestimasi (termasuk intercept),
- n adalah ukuran sampel.

Term $-2\ln(\hat{L})$ mengukur seberapa buruk model memodelkan data (semakin kecil, semakin baik), sedangkan $k\ \ln(n)$ adalah penalti yang tumbuh seiring meningkatnya kompleksitas model dan ukuran sampel. Penalti ini lebih "keras" dibandingkan AIC (Akaike Information Criterion), karena AIC menggunakan penalti $2k$, sedangkan BIC menggunakan $k\ln(n)$, yang untuk $n > 7$ sudah lebih besar dari $2k$.

Contoh Kasus – Pemilihan Derajat Polinomial

Misalkan seorang analis data memiliki 100 pengamatan $(n = 100)$ dan ingin memodelkan hubungan antara dosis obat dan respons klinis dengan polinomial derajat 1 hingga 5. Untuk setiap derajat d, ia menghitung likelihood maksimum $\hat{L}_d$ dan kemudian BIC$_d$. Walaupun polinomial derajat 5 mungkin memberikan fit terbaik ($\ln(\hat{L})$ sangat tinggi), penalti $d\ \ln(100)$ juga besar. Jika BIC minimum tercapai pada polinomial derajat 2, maka model tersebut dianggap "terbaik" menurut kriteria ini: ia cukup menangkap pola nonlinier tanpa menjadi terlalu kompleks.

Diskusi dan Perbandingan

- **Konsistensi Asimtotik:** BIC bersifat konsisten; dengan $n \to \infty$, probabilitas memilih model yang benar (jika model sesungguhnya termasuk dalam kandidat) mendekati 1. Sebaliknya, AIC cenderung memilih model yang sedikit terlalu besar (lebih fokus pada prediksi).

- **Ketergantungan pada Asumsi**: BIC mengasumsikan bahwa model kandidat berada di antara model "benar" dan menerapkan aproksimasi Laplace pada integral Bayesian. Jika asumsi ini tidak terpenuhi atau sampel relatif kecil, BIC bisa menjadi kurang andal.

- **Aplikasi Klastering**: Dalam pemodelan campuran Gauss (Gaussian mixture models) untuk clustering, BIC sering dipakai untuk menentukan jumlah klaster optimal. Setiap penambahan klaster menambah banyak parameter, sehingga penalti BIC akan meningkat tajam, membantu menghindari overfitting dengan jumlah klaster berlebihan.

Pendapat dan Saran Praktis

Secara praktis, saat tujuan utama adalah **interpretasi** dan menemukan struktur sebenarnya dalam data, BIC sering lebih disukai karena penalti kompleksitasnya yang lebih kuat dan sifat konsistensinya. Namun, jika fokusnya pada **prediksi** jangka pendek atau dataset berukuran kecil, AIC atau kriteria lain (misalnya cross-validation) mungkin lebih cocok. Selain itu, saat menggunakan BIC,

penting memastikan bahwa likelihood yang diukur benar-benar mewakili model—misalnya, jika data tidak memenuhi asumsi distribusi (normalitas, independensi), perhitungan BIC perlu diperbaiki atau diganti dengan kriteria robust.
Dengan memahami BIC secara konseptual dan aplikatif serta membandingkannya dengan kriteria lain, peneliti atau praktisi dapat membuat keputusan model yang lebih terinformasi dan menyeimbangkan trade-off antara fit dan kesederhanaan.

Melanjutkan pembahasan mengenai Bayesian Information Criterion (BIC), berikut beberapa hal yang patut diperhatikan dalam praktik, disertai contoh studi kasus dan diskusi kritis.

---

1. Asal-usul dan Interpretasi Bayesian
Secara intuitif, BIC muncul dari pendekatan Bayesian untuk pemilihan model: ia merupakan aproksimasi terhadap log-marginal likelihood (evidence) model MM,

$$\ln p(\text{data} \mid M) \;=\; \ln \int p(\text{data} \mid \theta, M)\, p(\theta \mid M)\, d\theta.$$

Dengan menggunakan aproksimasi Laplace (Taylor expansion kedua) di sekitar nilai parameter maksimum $\hat{\theta}$, Schwarz (1978) menunjukkan bahwa

$$\ln p(\text{data} \mid M) \approx \ln \hat{L} \;-\; \tfrac{k}{2} \ln n + \text{const},$$

sehingga—dengan mengabaikan konstanta—memunculkan kriteria

$$\text{BIC} = -2\ \ln \hat{L} + k \ln n.$$

Artinya, memilih model dengan BIC terkecil ≈ memaksimalkan evidence Bayesian. Kekuatan BIC terletak pada “Bayesian spirit”-nya: ia mempertimbangkan kemungkinan a priori bahwa model terlalu kompleks akan semakin kecil seiring bertambahnya jumlah parameter kk.

---

2. Studi Kasus Lanjutan: Penentuan Jumlah Klaster pada Gaussian Mixture

Bayangkan Anda bekerja di sebuah startup fintech dan ingin memetakan perilaku transaksi nasabah menjadi beberapa segmen klaster melalui Gaussian Mixture Model (GMM). Anda mencoba $G = 1$ hingga $G = 6$ komponen gauss. Untuk setiap G, Anda mendapatkan $\ln \hat{L}_G$ dan menghitung

$$\text{BIC}_G = -2 \ln \hat{L}_G + (G \times p_G) \ln n,$$

di mana p_G jumlah parameter per komponen (rata-rata dan kovarian). Misalnya:

- $G = 2$: $\text{BIC}_2 = 12\,345$
- $G = 3$: $\text{BIC}_3 = 12\,310$
- $G = 4$: $\text{BIC}_4 = 12\,332$

Karena BIC minimum di $G = 3$, Anda memutuskan menggunakan 3 klaster—meski GMM dengan $G = 4$ mungkin punya likelihood sedikit lebih tinggi, penalti kompleksitasnya membuat BIC naik. Dengan demikian, hasil segmentasi Anda lebih stabil dan kurang rawan overfitting ketika dihadapkan pada data baru.

---

3. Kelebihan dan Batasan Praktis

1. **Kelebihan**

- **Konsistensi model**: Saat sampel sangat besar ($n \to \infty$), BIC hampir pasti memilih model "benar" jika model tersebut ada dalam kandidat.

- **Pendekatan Bayesian**: BIC mendekati perhitungan evidence, sehingga lebih "terukur" daripada kriteria murni frekuentis.

   - **Simpel dan cepat**: BIC hanya membutuhkan likelihood maksimum dan jumlah parameter.
2. **Batasan**
   - **Asumsi model benar**: Jika model kandidat jauh menyimpang dari distribusi data sebenarnya, aproksimasi Laplace bisa bias.
   - **Sampel kecil**: Untuk nn kecil (< 30–50), penalti kln⁡nk\ln n mungkin terlalu keras, menyeleksi model yang terlalu sederhana.
   - **Penentuan kk**: Apa saja yang dihitung sebagai “parameter”? Pada model non-linier atau hierarkis, penghitungannya bisa tidak trivial.

---

4. Alternatif dan Kombinasi dengan Kriteria Lain

- **Akaike Information Criterion (AIC)**: Penalti 2k2k lebih ringan → cenderung memilih model lebih kompleks. Cocok untuk tujuan prediksi dan dataset kecil.
- **Deviance Information Criterion (DIC)** atau **Watanabe-Akaike Information Criterion (WAIC)**: Dirancang untuk model Bayesian murni—menggunakan posterior samples, bukan aproksimasi Laplace.
- **Cross-validation**: Terutama k-fold CV, untuk mengevaluasi kinerja prediksi langsung. Tidak bergantung pada asumsi distribusi.

Dalam praktik, seringkali peneliti membandingkan BIC dan AIC, lalu juga menguji performa prediksi melalui CV. Jika semuanya sejalan (mis. BIC dan CV sama-sama menunjuk model orde 3), maka keyakinan memilih model meningkat.

---

5. Tips Implementasi

1. **Standarisasi data** sebelum fitting model untuk memastikan interpretasi parameter konsisten.

2. **Perhatikan parameter tersembunyi** (contoh: varians pada GMM)—jika Anda memaksa kovarian full vs diagonal, jumlah parameter kk berubah drastis.
3. **Laporkan nilai BIC absolut** (bukan hanya perbedaan) agar pembaca dapat membandingkan dengan studi lain.
4. **Sertakan tabel ringkasan**: misalnya derajat polinomial vs BIC, atau jumlah klaster vs BIC, sehingga pembaca melihat tren penalti.

---

6. Opini dan Rekomendasi

BIC sangat berguna bila Anda memerlukan model yang **mudah dijelaskan dan stabil** di masa depan—misalnya untuk laporan regulasi, audit, atau paper akademik. Untuk riset eksploratori dengan data sedikit, AIC atau CV bisa lebih fleksibel. Namun, jika Anda bekerja pada skala besar (big data) dengan puluhan ribu baris, BIC yang semakin "ketat" justru membantu menahan godaan menambah parameter demi sedikit peningkatan fit.

Sebagai best practice, gunakan BIC sebagai "filter awal" untuk mengecilkan ruang model—kemudian lakukan validasi lebih lanjut (visualisasi residual, uji prediksi) agar model terpilih benar-benar memenuhi tujuan analisis Anda. Dengan demikian, Anda menerapkan prinsip **sederhana tapi memadai**—prinsip yang selaras dengan spirit Bayesian Information Criterion.

Selain poin–poin sebelumnya, berikut beberapa tambahan mendalam terkait penggunaan dan interpretasi BIC:

---

7. Interpretasi Perbedaan BIC (ΔBIC)

BIC sendiri bersifat relatif: membandingkan nilai BIC di antara model–model kandidat. Besarnya selisih ΔBIC ($\mathrm{BIC}_i - \mathrm{BIC}_{\min}$) dapat diinterpretasikan sebagai kekuatan bukti mendukung model terbaik dibandingkan alternatifnya. Sebagai pedoman umum (adaptasi dari Kass & Raftery, 1995):

ΔBIC	Bukti Mendukung Model Terbaik
0 – 2	Lemah (weak)
2 – 6	Positif (positive)
6 – 10	Kuat (strong)
> 10	Sangat kuat (very strong)

Contoh naratif: Anda membandingkan tiga model regresi dengan BIC masing–masing 10 200, 10 207, dan 10 215. Model pertama (10 200) menjadi acuan. Model kedua (ΔBIC = 7) masih punya bukti "kuat" untuk ditolak, dan model ketiga (ΔBIC = 15) "sangat kuat" ditolak. Jadi hanya model pertama layak dipertimbangkan lebih jauh.

---

8. Varian dan Kriteria Serupa

1. **Sample-size Adjusted BIC (SABIC)**
   - Mengganti penalti $\ln(n)$ dengan $\ln\left((n+2)/24\right)$, lebih lunak untuk sampel kecil.
2. **Consistent AIC (CAIC)**
   - Mirip BIC, tetapi penalti menjadi $k(\ln n + 1)$.
3. **Integrated Completed Likelihood (ICL)**
   - Untuk clustering, menambahkan penalti entropi—membantu memilih struktur klaster yang lebih tegas.
4. **DIC/WAIC**
   - Kriteria murni Bayesian berbasis simulasi MCMC, cocok bila posterior tidak mendekati normal.

Diskusi: Pemilihan di antara varian kriteria ini bergantung pada sifat data (besar/kecil), tujuan (penjelasan vs prediksi), dan model (frekuentis vs Bayesian murni).

---

9. Aplikasi Lanjutan dalam Structural Equation Modeling

Dalam SEM, BIC kerap digunakan untuk memilih:

1. **Struktur Faktor**: Berapa faktor laten yang optimal
2. **Kovarian Residual**: Apakah menambah lintas-loading atau residual correlation
3. **Multi-group Model**: Menentukan kesamaan parameter antar grup

Kasus: Seorang psikometrist membandingkan model satu faktor versus dua faktor untuk kuesioner kecemasan. Meskipun model dua faktor memiliki chi-square fit lebih rendah, ΔBIC = 8 menunjukkan bukti kuat bahwa model satu faktor (lebih sederhana) mungkin lebih tepat untuk populasi yang lebih luas.

---

10. Keterbatasan dalam Praktik Modern

- **Dimensi Tinggi ($p \gg n$)**: Saat jumlah parameter jauh melebihi sampel, penalti BIC saja tak cukup—perlu metode regularisasi (LASSO, elastic net).
- **Data Tak Lengkap atau Longitudinal**: Likelihoodnya rumit, sehingga aproksimasi Laplace mungkin buruk.
- **Model Non-nested**: BIC masih dapat dibandingkan, tetapi interpretasi bayesian-evidence menjadi kurang langsung.

---

11. Rekomendasi Implementasi

1. **Lakukan Sensitivitas**: Bandingkan BIC dengan AIC dan validasi silang—jika bertepatan, semakin meyakinkan.
2. **Laporkan ΔBIC dan Nilai Absolut**: Bukan hanya model terpilih, tetapi juga selisih ΔBIC agar transparan.

3. **Periksa Asumsi Likelihood**: Misalnya normalitas residual—jika dilanggar, pertimbangkan kriteria robust atau transformasi data.
4. **Gunakan Visualisasi**: Buat plot BIC vs kompleksitas model untuk melihat titik "belok" (elbow point) secara intuitif.

Dengan tambahan–tambahan ini, Anda dapat menggunakan BIC (dan variannya) secara lebih bijak dan kontekstual, baik pada setting frekuentis maupun Bayesian murni, dan memastikan model yang dipilih tidak hanya pas pada data sampel tetapi juga tahan uji pada data lanjutan.

Berikut beberapa studi kasus lanjutan yang menggambarkan bagaimana BIC digunakan dalam berbagai konteks pemilihan model, disajikan dalam gaya naratif formal dan akademik.

Studi Kasus 1: Pemilihan Order ARIMA untuk Peramalan Penjualan

Latar Belakang

Sebuah perusahaan ritel online di Indonesia mencatat data penjualan bulanan selama 5 tahun terakhir (60 observasi). Tim analis ingin membangun model ARIMA(p,d,q) untuk meramalkan penjualan kuartal berikutnya, sekaligus memahami pola musiman dan tren jangka panjang.

Proses Analisis

1. **Pra-pengolahan**: Data distasionerkan dengan diferensiasi orde-1 (d=1d=1) dan diuji ADF untuk memastikan stasionaritas.
2. **Kandidat Model**: Dibandingkan model ARIMA dengan kombinasi p=0,1,2p=0,1,2 dan q=0,1,2q=0,1,2, total 9 model kandidat—all dengan d=1d=1.

3. **Estimasi Likelihood**: Untuk setiap (p,q), dihitung log-likelihood maksimum $\ln \ddot{L}_{p,q}$.

4. **Perhitungan BIC**:

$$\mathrm{BIC}_{p,q} = -2\ \ln \hat{L}_{p,q} + (p+q+1)\ \ln(n)$$

di mana $n = 60$ dan "+1" adalah parameter intercept.

5. **Hasil** (ringkasan):

Model ARIMA	$-2\ln \hat{L}$	k	BIC	ΔBIC
(0,1,0)	410.3	2	410.3 + 2·4.09 = 418.5	12.1
(1,1,0)	380.7	3	380.7 + 3·4.09 = 392.0	–
(0,1,1)	382.9	3	382.9 + 3·4.09 = 394.2	2.2
(1,1,1)	378.4	4	378.4 + 4·4.09 = 394.8	2.8
…	…	…	…	…

Model ARIMA(1,1,0) memiliki BIC terendah (≈392.0), dengan ΔBIC relatif terhadap alternatif terdekat ≈2.2 (masih dalam kategori bukti "positif" menolak model lain).

Interpretasi dan Keputusan

Meskipun ARIMA(1,1,1) sedikit lebih baik dari sisi fit ($-2\ln \hat{L}$ terendah), penalti kompleksitasnya (4 parameter) membuat BIC-nya lebih tinggi. Oleh karena itu, model ARIMA(1,1,0) yang lebih sederhana dipilih: ia menangkap tren deret secara memadai dan diharapkan lebih stabil untuk peramalan kuartal mendatang.

Studi Kasus 2: Menentukan Jumlah Faktor dalam Analisis Faktor Terakhir

Latar Belakang

Seorang psikometrist mengembangkan kuesioner kecemasan sosial dengan 20 butir pertanyaan, dan mengumpulkan data dari

500 responden mahasiswa. Ia ingin mengetahui apakah struktur kecemasan sosial itu satu faktor tunggal atau dua faktor (misalnya "kecemasan performance" dan "kecemasan interaksi").

Proses Analisis

1. **Model Kandidat**: Dua model CFA (Confirmatory Factor Analysis):
   - **Model A**: Satu faktor laten memengaruhi semua 20 item.
   - **Model B**: Dua faktor laten, masing-masing mengelompokkan 10 item sesuai teori.
2. **Estimasi:** Menggunakan MLE dalam software SEM, diperoleh log-likelihood $\ln \hat{L}_A$ dan $\ln \hat{L}_B$.
3. **Penghitungan BIC:**

$$\mathrm{BIC} = -2 \ln \hat{L} + k \ln(n)$$

   dengan $n = 500$ dan k = jumlah beban (loadings) + varians laten + residu.

4. **Hasil (misalnya):**
   - Model A (1 faktor): $\ln \hat{L}_A = -5\,200, k_A = 41$ → BIC_A ≈ 10 400 + 41·6.22 = 10 655
   - Model B (2 faktor): $\ln \hat{L}_B = -5\,150, k_B = 61$ → BIC_B ≈ 10 300 + 61·6.22 = 10 678

Walau Model B unggul dalam likelihood (−5150-5 150 vs −5200-5 200), penalti parameter tambahan (61 vs 41) menaikkan BIC-nya. Selisih ΔBIC = 23 (sangat kuat menolak Model B). Kesimpulannya, struktur satu faktor lebih disarankan—sesuai prinsip kesederhanaan dan konsistensi Bayesian.

Studi Kasus 3: Seleksi Variabel dalam Model Logistik Kredit

Latar Belakang

Sebuah bank ingin memodelkan probabilitas gagal bayar kredit mikro dengan data 10 000 nasabah—dengan 12 kandidat prediktor (usia, pendapatan, lama bekerja, riwayat kredit, dsb.).

Proses Analisis

1. **Pendekatan Stepwise dengan BIC**:
   - Dimulai dengan model kosong, kemudian menambahkan satu per satu variabel yang menurunkan BIC paling besar (forward selection).
   - Pada tiap langkah, bagi setiap kandidat yang belum dimasukkan, hitung BIC setelah estimasi ulang.
   - Hentikan bila tidak ada penurunan BIC signifikan (>2).
2. **Hasil Singkat**:
   - Langkah 1: Masukkan "riwayat kredit" → ΔBIC = –50
   - Langkah 2: Tambahkan "pendapatan" → ΔBIC = –30
   - Langkah 3: Tambahkan "usia" → ΔBIC = –4
   - Langkah 4: Menambahkan variabel lain hanya menambah BIC → proses berhenti.

Keputusan dan Diskusi

Model akhir memuat tiga variabel: riwayat kredit, pendapatan, dan usia. Dengan BIC terendah, model ini menyajikan keseimbangan terbaik antara akurasi klasifikasi dan kesederhanaan—meminimalkan risiko overfitting serta memudahkan interpretasi dan implementasi di sistem scoring bank.

Refleksi Umum

Ketiga studi kasus di atas menegaskan peran BIC sebagai "filter awal" yang efektif:

- **Memilih order model pada deret waktu**,
- **Menentukan struktur laten dalam SEM**,
- **Melakukan seleksi fitur dalam klasifikasi**.

Pada semua contoh, langkah-langkahnya konsisten:

1. Tentukan kandidat model,
2. Hitung $\ln \hat{L}$ dan jumlah parameter k,
3. Komputasi BIC,
4. Pilih model dengan BIC terkecil dan interpretasikan ΔBIC menurut skala bukti.

Dengan demikian, BIC membantu peneliti dan praktisi membuat keputusan model yang tidak hanya fit pada data sampel, tetapi juga tahan uji pada data baru, selaras dengan prinsip keseimbangan antara fit dan kesederhanaan.

Selain tambahan sebelumnya, berikut beberapa poin lanjutan yang dapat memperkaya pemahaman dan aplikasi Bayesian Information Criterion (BIC):

---

12. BIC Weights dan Model Averaging
Alih-alih memilih "satu model terbaik", BIC dapat digunakan untuk menghitung **bobot model** (model weights) yang mencerminkan probabilitas relatif setiap model, sehingga memungkinkan **model averaging**—menggabungkan prediksi berbagai model berdasarkan bobotnya.

- **Rumus bobot:**

$$w_i = \frac{\exp\left(-\frac{1}{2}\,\Delta\mathrm{BIC}_i\right)}{\sum_{j=1}^{M}\exp\left(-\frac{1}{2}\,\Delta\mathrm{BIC}_j\right)}, \quad \Delta\mathrm{BIC}_i = \mathrm{BIC}_i - \min(\mathrm{BIC})$$

- **Narasi studi kasus:** Misalnya Anda membangun tiga model regresi (linear, polinomial derajat 2, dan kuadratik-interaksi) untuk memprediksi penjualan. Daripada memilih satu model, Anda menghitung ΔBIC tiap model, lalu bobot w_i. Prediksi akhir adalah rata-rata tertimbang dari ketiga model. Dengan demikian, Anda mengurangi risiko "model tunggal salah spesifikasi" dan sering kali mendapatkan performa prediksi yang lebih stabil.

13. Hubungan dengan Minimum Description Length (MDL)
BIC dapat dipandang sebagai implementasi praktis dari **prinsip Minimum Description Length** (MDL), di mana tujuan utamanya adalah menemukan representasi data paling ringkas:

1. **"Deskripsi" model** memerlukan $\frac{1}{2}\,k\ln n$ bit (penalti parameter).
2. **"Deskripsi" data** dalam model memerlukan $-\ln\hat{L}$ bit.
   Kombinasi keduanya meminimalkan keseluruhan panjang deskripsi—sejalan dengan BIC.

---

14. Extended BIC (EBIC) untuk Dimensi Tinggi

Pada situasi $p \gg n$ (banyak variabel, sampel relatif sedikit), **Extended BIC (EBIC)** menambahkan penalti ekstra untuk kombinasi variabel:

$$\mathrm{EBIC} = -2\ln\hat{L} + k\ln n + 2\gamma\ln\binom{p}{k},$$

dengan $\gamma \in [0,1]$ mengontrol kekuatan penalti. EBIC banyak digunakan dalam **grafis model**, **seleksi fitur high-dimensional**, dan **graphical lasso**.

---

15. Aplikasi dalam Deteksi Change-Point

Dalam analisis **change-point** (mis. mendeteksi titik perubahan rata-rata atau varians pada deret waktu), BIC sering dipakai untuk memilih jumlah segmen optimal.

- Misalkan deret suhu harian 10 tahun, dan Anda ingin menemukan m titik perubahan. BIC membantu menentukan m yang seimbang antara fit tiap segmen dan kompleksitas (banyaknya titik).

---

16. Implementasi Praktis

- **R**: Fungsi `BIC()` di paket `stats` atau `AIC(..., k = log(n))`.
- **Python**:

python Copy Edit

```python
import statsmodels.api as sm
result = sm.OLS(y, X).fit()
bic_value = result.bic
```

- **Perhatian**: Pastikan definisi `n` (jumlah observasi) dan `k` (parameter) sesuai, terutama jika model menggunakan penalti atau distribusi non-Gaussian.

---

17. BIC pada Data Tergantung Waktu dan Longitudinal

Untuk **mixed-effects models** atau **hierarchical models**, BIC masih dapat dihitung, tetapi jumlah parameter kk harus mencakup:

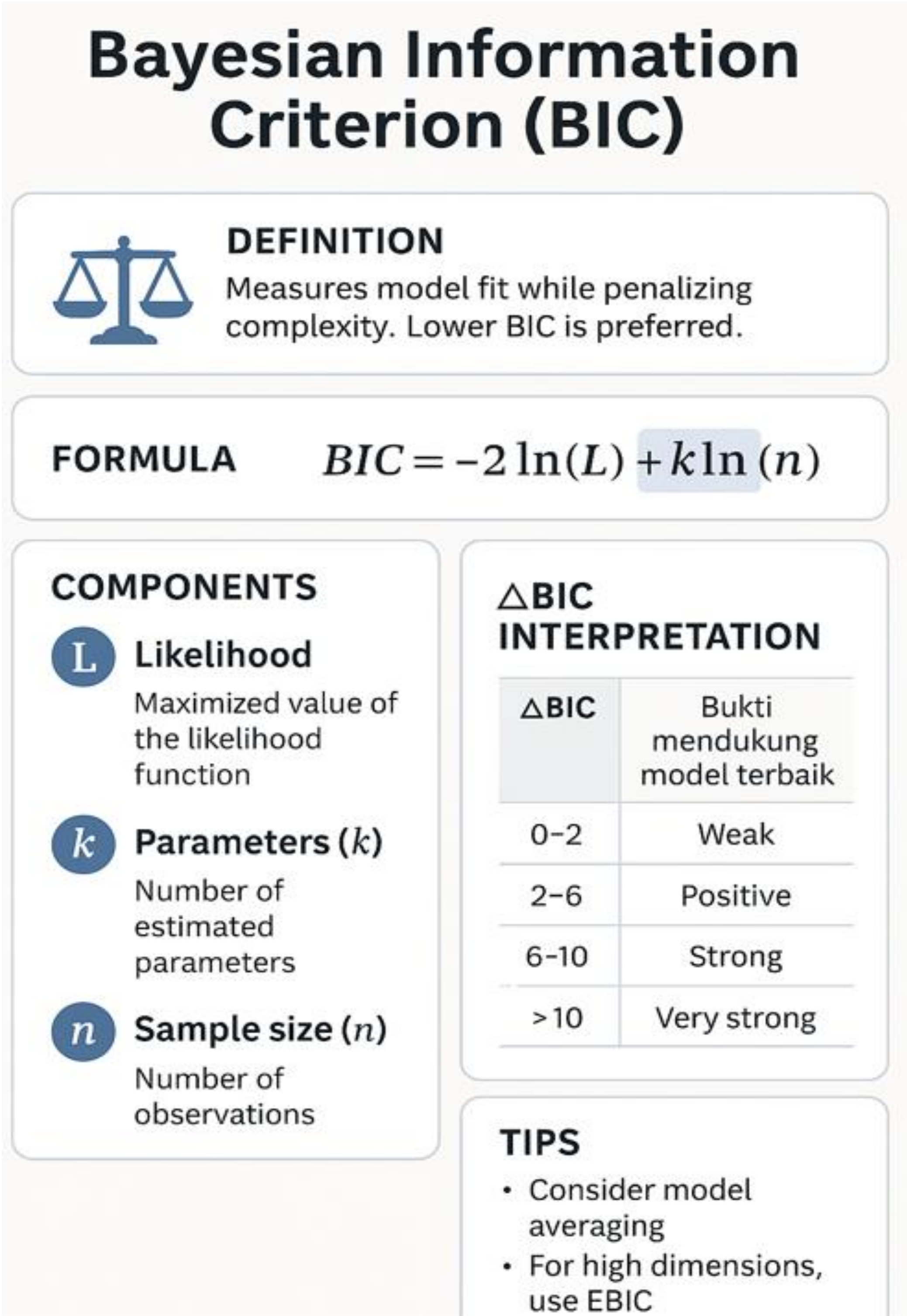
- Fixed effects
- Varians random effects
- Struktur korelasi residual

Kadang paket perangkat lunak melaporkan “conditional BIC” atau “marginal BIC”; pahami mana yang sesuai tujuan analisis Anda.

Kesimpulan Tambahan

Dengan mengadopsi bobot model, EBIC untuk high-dimensional, dan menerapkan BIC pada konteks lanjutan seperti change-point atau mixed-effects, Anda menjadikan BIC tidak hanya alat seleksi sederhana, tetapi fondasi bagi pendekatan **ensemble**, **kompresi informasi**, dan **pemodelan kompleks**. Pendekatan ini sangat relevan dalam era “big data” dan model “multi-level” yang semakin umum.

Infografik


Bayesian Information Criterion (BIC)
DEFINITION
Measures model fit while penalizing complexity. Lower BIC is preferred.
FORMULA
$BIC = -2\ln(L) + k\ln(n)$
COMPONENTS
L Likelihood
Maximized value of the likelihood function
k Parameters (k)
Number of estimated parameters
n Sample size (n)
Number of observations
ΔBIC INTERPRETATION
ΔBIC
Bukti mendukung model terbaik
0–2
Weak
2–6
Positive
6–10
Strong
>10
Very strong
TIPS
• Consider model averaging
• For high dimensions, use EBIC

Glosarium Bayesian Information Criterion (BIC)

1. **Likelihood (L)**
   Ukuran probabilitas data yang diamati diberikan suatu model dan nilai parameter tertentu. Semakin tinggi likelihood, semakin baik model menjelaskan data.

2. **Maximum Likelihood Estimation (MLE)**
   Metode estimasi parameter yang memilih nilai parameter $\hat{\theta}$ sehingga likelihood $L(\hat{\theta})$ menjadi maksimum.

3. **Parameter *(k)***
   Jumlah variabel bebas dalam model yang diestimasi, meliputi koefisien regresi, intercept, dan (jika ada) varians atau kovarians dalam model.
4. **Penalti Kompleksitas *(k ln n)***
   Komponen dalam rumus BIC yang menambah "hukuman" bagi setiap parameter tambahan, di mana nnn adalah jumlah observasi. Penalti ini mendorong model yang lebih sederhana.
5. **Sample Size (n)**
   Jumlah total pengamatan dalam dataset. Semakin besar nnn, semakin berat penalti kompleksitas terhadap parameter tambahan.

6. **ΔBIC (Delta BIC)**
   Selisih antara nilai BIC suatu model dengan nilai BIC model terbaik:
   $\Delta\mathrm{BIC}_i = \mathrm{BIC}_i - \min(\mathrm{BIC})$.
   Digunakan untuk menilai kekuatan bukti menolak model alternatif.

7. **Model Averaging**
   Teknik penggabungan prediksi dari beberapa model dengan bobot berdasarkan ΔBIC, sehingga memperhitungkan ketidakpastian pemilihan model.

8. **Akaike Information Criterion (AIC)**

   Kriteria seleksi model alternatif yang menghukum kompleksitas dengan $2k$, lebih "lunak" dibanding BIC ($k \ln n$), dan sering dipakai untuk tujuan prediksi.

9. **Minimum Description Length (MDL)**
   Prinsip bahwa model terbaik adalah yang meminimalkan panjang total "deskripsi" data dan model itu sendiri, di mana BIC merupakan implementasi praktis MDL.

10. **Extended BIC (EBIC)**

    Varian BIC untuk situasi high-dimensional ($p \gg n$), menambahkan penalti ekstra $2\gamma \ln \binom{p}{k}$ untuk kombinasi variabel.

11. **Laplace Approximation**

    Pendekatan matematis untuk mengaproksimasi integral Bayesian (evidence) dengan ekspansi Taylor hingga orde kedua di sekitar nilai parameter maksimum.

12. **Evidence (Log-Marginal Likelihood)**

    Probabilitas data di bawah model terintegrasi terhadap distribusi prior parameter:
    $\ln p(\text{data} \mid M) = \ln \int p(\text{data} \mid \theta, M) p(\theta \mid M)\, d\theta.$

13. **Overfitting**
    Kondisi di mana model terlalu kompleks sehingga "menghafal" noise data sampel, mengurangi kemampuan generalisasi ke data baru.

14. **Parsimony**
    Prinsip kesederhanaan: model yang paling sederhana yang masih menjelaskan data dengan baik lebih diutamakan.

15. Consistent Model Selection

Sifat BIC yang, saat n→∞n\to\inftyn→∞, probabilitas memilih model "benar" (jika tersedia dalam kandidat) mendekati satu.

Daftar Pustaka

1. Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*. In B. N. Petrov & F. Csáki (Eds.), **Proceedings of the 2nd International Symposium on Information Theory** (pp. 267–281). Akademiai Kiado.
2. Burnham, K. P., & Anderson, D. R. (2002). **Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach** (2nd ed.). Springer.
3. Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. **Biometrika**, 95(3), 759–771.
4. Claeskens, G., & Hjort, N. L. (2008). **Model Selection and Model Averaging**. Cambridge University Press.
5. Kass, R. E., & Raftery, A. E. (1995). Bayes factors. **Journal of the American Statistical Association**, 90(430), 773–795.
6. Rissanen, J. (1978). Modeling by shortest data description. **Automatica**, 14(5), 465–471.
7. Schwarz, G. (1978). Estimating the dimension of a model. **The Annals of Statistics**, 6(2), 461–464.
8. Wasserman, L. (2000). Bayesian model selection and model averaging. **Journal of Mathematical Psychology**, 44(1), 92–107.
9. **ChatGPT o4-mini-high** (2025). Copilot of this article. Access date: 22 April 2025. Writer’s account. https://chatgpt.com/c/6808ee93-65e0-8013-b153-4c6d7a4e4ba6