Criterio de información bayesiano (BIC) – Bayesian information criterion (BIC)

Authored by

memjavad

November 5, 2025

RECOMMENDED CITATION

memjavad (2025). *Criterio de información bayesiano (BIC) – Bayesian information criterion (BIC)*. Spanish Psychological Databases. Retrieved from https://spanish.arabpsychology.com/?p=2909

Criterio de Información Bayesiano (BIC)

Primary Disciplinary Field(s): Estadística, **Econometría**, **Aprendizaje Automático** (Machine Learning), **Inferencia Causal**.

1. Definición y Propósito Central

El Criterio de Información Bayesiano (BIC), también conocido como Criterio de Schwarz, es una métrica fundamental utilizada en la estadística para la **selección de modelos**. Su propósito esencial es ofrecer un método riguroso para comparar un conjunto finito de modelos estadísticos, permitiendo identificar cuál de ellos ofrece el mejor equilibrio entre la bondad de ajuste a los datos observados y la complejidad del modelo. En esencia, el BIC aborda el problema central del **sobreajuste** (overfitting), que ocurre cuando un modelo incorpora demasiados parámetros, ajustándose excesivamente al ruido de la muestra sin lograr una verdadera generalización predictiva. Al igual que otros criterios de selección, como el Criterio de Información de Akaike (AIC), el BIC penaliza la inclusión de parámetros adicionales, pero lo hace con una severidad mayor, basando su justificación en la teoría de la probabilidad bayesiana.

La necesidad del BIC surge de la práctica común en la investigación donde múltiples modelos pueden explicar razonablemente un fenómeno. Si un investigador simplemente elige el modelo con la mayor verosimilitud (likelihood), casi siempre seleccionará el modelo más complejo, ya que añadir parámetros nunca empeora la verosimilitud del ajuste. El BIC proporciona un marco formal para cuantificar la evidencia a favor de un modelo sobre otro. Un valor de **BIC más bajo** indica un modelo preferible. Este criterio no solo evalúa qué tan bien se ajusta el modelo a los datos, sino que también estima la probabilidad de que el modelo sea el "verdadero" generador de los datos, bajo ciertas suposiciones asintóticas.

Formalmente, el BIC se deriva como una aproximación asintótica de la integral logarítmica de la evidencia marginal de un modelo. Fue desarrollado por el estadístico **Gideon E. Schwarz en 1978** y publicado en su artículo seminal «Estimating the Dimension of a Model», sentando las bases para su uso generalizado en campos que van desde la biología hasta las finanzas. La robustez del BIC reside en su capacidad para actuar como una herramienta de regularización implícita, promoviendo la **parsimonia**, lo que significa que favorece modelos simples que son suficientes para describir la estructura de los datos sin incorporar complejidad innecesaria.

2. Fundamentos Teóricos: La Perspectiva Bayesiana

Aunque el BIC se utiliza frecuentemente en el contexto de la estadística frecuentista, su nombre y su derivación se arraigan profundamente en la **inferencia bayesiana**. La base teórica del BIC es la aproximación asintótica de la probabilidad posterior del modelo. En el contexto bayesiano, la selección de modelos se realiza comparando las probabilidades posteriores de los modelos candidatos, lo que requiere calcular la evidencia marginal del modelo (o la verosimilitud de los datos dado el modelo, integrando sobre el espacio de parámetros).

La integral de la evidencia marginal suele ser intratable analíticamente y computacionalmente costosa. El BIC ofrece una solución elegante y práctica a este problema. Se demuestra que, bajo la suposición de que la distribución a priori de los parámetros es "razonablemente plana" y que el tamaño de la muestra (n) es grande, el logaritmo de la evidencia marginal puede aproximarse mediante una expansión de Laplace. El BIC es, de hecho, el resultado de esta aproximación de Laplace, ignorando términos de orden inferior. Esta aproximación justifica por qué el BIC es a menudo descrito como un criterio que tiene una base más sólida en la teoría de la probabilidad que el AIC, que se basa en la teoría de la información y la distancia de Kullback-Leibler.

La característica clave de la aproximación bayesiana es que, a medida que el tamaño de la muestra tiende al infinito (asintóticamente), el BIC selecciona consistentemente el modelo verdadero, siempre y cuando el modelo verdadero esté incluido en el conjunto de modelos candidatos. Esta propiedad de **consistencia** es una distinción crucial frente al AIC. La aproximación de Schwarz permite que el BIC penalice los modelos de manera que la penalización crece logarítmicamente con el tamaño de la muestra. Esta dependencia del tamaño de la muestra es el motor de su mayor penalización en comparación con el AIC, asegurando que, con muestras grandes, el BIC tienda a seleccionar modelos más simples y, por lo tanto, más cercanos a la estructura verdadera.

3. Formulación Matemática y Componentes Clave

La fórmula estándar para el Criterio de Información Bayesiano, asumiendo la estimación de máxima verosimilitud (MLE), se expresa mediante dos componentes esenciales: el término de ajuste y el término de penalización. Es vital comprender la función de cada parte para entender cómo el BIC equilibra la bondad de ajuste y la complejidad.

La fórmula general del BIC puede representarse de la siguiente manera:

BIC = + .

Desglosando los componentes:

$ -2 ln(hat{L}) $: Es el logaritmo de la máxima verosimilitud (log-likelihood) del modelo estimado, multiplicado por -2. Este término mide la **bondad de ajuste** del modelo a los datos. Un valor más alto de la máxima verosimilitud (y por lo tanto, un valor menos negativo del término de ajuste) indica un mejor ajuste.
$ k $: Representa el número de **parámetros libres** o grados de libertad que se han estimado en

el modelo. Este es el contador directo de la complejidad del modelo.
$\ln(n)$: Es el logaritmo natural del **tamaño de la muestra** (número de observaciones). Este factor es crucial, ya que introduce la dependencia del BIC en el volumen de datos.

El primer término, $-2 \ln(\hat{L})$, es idéntico al término de ajuste utilizado en el Criterio de Información de Akaike (AIC) y se relaciona directamente con la devianza del modelo. El segundo término, $k \ln(n)$, es el término de penalización por complejidad. La estructura de esta penalización revela por qué el BIC es más estricto que el AIC: mientras que el AIC multiplica la complejidad (k) por una constante fija (2), el BIC la multiplica por una cantidad que crece con el tamaño de la muestra ($\ln(n)$). Para cualquier muestra $n \geq 8$, el factor de penalización del BIC es mayor que el del AIC, lo que garantiza su mayor inclinación hacia la parsimonia.

4. Interpretación y Principios de Penalización

La interpretación del valor del BIC es fundamentalmente relativa; los valores absolutos carecen de significado directo, pero las diferencias entre los valores de BIC de modelos candidatos son altamente informativas. El modelo con el valor de BIC más bajo es el modelo preferido. La diferencia de BIC entre dos modelos, ΔBIC, puede transformarse en una **razón de probabilidad posterior** (posterior odds ratio) o un **factor de Bayes** aproximado, lo que facilita una interpretación directa en términos de la fuerza de la evidencia bayesiana.

Esta transformación a factores de Bayes es una de las grandes fortalezas del BIC. Por ejemplo, si el Modelo 1 tiene un BIC significativamente menor que el Modelo 2, la diferencia puede cuantificarse. Generalmente, los investigadores utilizan las directrices propuestas por Jeffreys o Kass y Raftery para interpretar estas diferencias. Una diferencia de BIC de 2 a 6 se considera evidencia positiva a favor del modelo con el BIC menor; una diferencia de 6 a 10 es evidencia fuerte; y una diferencia mayor a 10 se considera evidencia muy fuerte o decisiva. Esta capacidad de cuantificar la certeza es crucial para la comunicación de resultados estadísticos.

El principio de penalización del BIC está diseñado específicamente para evitar la selección de modelos sobredimensionados, un problema conocido como sobreajuste. En la práctica, esto significa que el BIC tiende a seleccionar modelos que son "demasiado simples" o "subdimensionados" en comparación con los seleccionados por el AIC, especialmente cuando el tamaño de la muestra es grande. Esta tendencia a la parsimonia es una ventaja si el objetivo principal es identificar la estructura subyacente verdadera del proceso generador de datos (propiedad de consistencia en la selección del modelo). Si el objetivo es la explicación, la consistencia del BIC es altamente deseable; si el objetivo fuera únicamente la precisión predictiva fuera de la muestra, un criterio menos penalizador podría ser más adecuado.

5. Ventajas y Desventajas Comparativas (vs. AIC)

La comparación más frecuente en la selección de modelos es entre el BIC y el Criterio de Información de Akaike (AIC). Ambos son criterios basados en la verosimilitud con penalización, pero difieren fundamentalmente en su objetivo filosófico y su comportamiento asintótico. El AIC, formulado como $\text{AIC} = -2 \ln(\hat{L}) + 2k$, tiene una penalización fija, mientras que la penalización del BIC ($k \ln(n)$) aumenta con el tamaño de la muestra.

La **principal ventaja** del BIC es su **consistencia**: si el modelo verdadero existe en el conjunto de candidatos, el BIC lo identificará con probabilidad que tiende a 1 a medida que el tamaño de la muestra tiende al infinito. Esta propiedad lo hace ideal para la inferencia causal, la identificación de la estructura subyacente o la comprobación de hipótesis. Además, su justificación bayesiana proporciona una interpretación teórica sólida, permitiendo la transformación a factores de Bayes, lo que facilita una interpretación probabilística de la evidencia del modelo. Cuando el objetivo es la explicación parsimoniosa del fenómeno, el BIC es generalmente el criterio preferido.

No obstante, el BIC presenta ciertas **desventajas**. Primero, asume que el modelo verdadero existe y está incluido en el conjunto de modelos candidatos, una suposición que a menudo no se cumple en la práctica. Segundo, en muestras pequeñas o moderadas, la aproximación asintótica que fundamenta el BIC puede ser inexacta, llevando a resultados menos fiables. En estas situaciones, el BIC puede penalizar en exceso la complejidad, favoreciendo modelos demasiado simples. Tercero, el BIC no está orientado a la predicción. Debido a su fuerte penalización, el BIC tiende a seleccionar modelos que son subóptimos para la precisión predictiva fuera de la muestra, ya que rechaza modelos ligeramente más complejos que podrían capturar mejor las variaciones futuras.

En contraste, el AIC es asintóticamente eficiente en la predicción, buscando minimizar el error de predicción fuera de la muestra (relacionado con la distancia de Kullback-Leibler). El AIC tiende a seleccionar modelos más complejos (sobredimensionados) que el BIC. La elección entre BIC y AIC se reduce, por lo tanto, a una elección filosófica: si se prioriza la identificación de la estructura verdadera y la parsimonia (BIC), o si se busca la mejor capacidad predictiva (AIC), reconociendo que el modelo predictivo óptimo puede no ser el modelo causalmente "verdadero".

6. Aplicaciones Prácticas en la Modelización

El BIC se emplea extensamente en cualquier campo que requiera la comparación de modelos estadísticos jerárquicos o anidados. Es una herramienta estándar en la **econometría** para la selección de variables en modelos de regresión lineal y no lineal, donde la parsimonia es esencial para la interpretabilidad económica y la estabilidad de los coeficientes. Por ejemplo, al construir un modelo de series de tiempo, como un modelo de media móvil integrada autorregresiva (ARIMA), el BIC ayuda a determinar el orden óptimo (p, d, q) al penalizar los órdenes altos que solo ofrecen ganancias marginales en el ajuste.

En el ámbito del **Aprendizaje Automático** y la **Minería de Datos**, el BIC se utiliza para la selección de características y la determinación del número óptimo de grupos en algoritmos de clustering, como la mezcla gaussiana (Gaussian Mixture Models - GMM). Al aplicar el BIC a los GMM, se busca el número de componentes que mejor equilibre la verosimilitud de los datos con el número de parámetros necesarios para definir esos componentes. Esto es crucial para prevenir la sobresegmentación de los datos y asegurar que los grupos identificados sean robustos y significativos.

Además, el BIC es crucial en la **biología y la genética**, particularmente en la filogenética y la modelización de la evolución molecular. Los investigadores utilizan el BIC para seleccionar entre diferentes modelos de sustitución de nucleótidos o aminoácidos, decidiendo cuál de ellos explica mejor los patrones de variación observados en las secuencias de ADN o proteínas. Su uso garantiza que las conclusiones sobre la estructura o la causalidad no se basen en modelos innecesariamente inflados, promoviendo la máxima interpretabilidad biológica.

7. Críticas y Limitaciones Metodológicas

A pesar de su amplia aceptación y su sólida base teórica, el BIC enfrenta varias críticas y limitaciones que deben considerarse al aplicarlo. Una limitación fundamental, compartida con el AIC, es que el BIC solo es válido para la comparación de modelos que han sido estimados utilizando la **misma muestra de datos** y la misma variable dependiente. No puede utilizarse para comparar modelos no anidados o modelos que abordan preguntas de investigación fundamentalmente diferentes.

Una crítica metodológica importante se centra en la dependencia de la **aproximación asintótica**. El BIC se deriva bajo la estricta suposición de que el tamaño de la muestra (n) es grande. En situaciones donde n es pequeño o moderado, la aproximación de Laplace puede ser pobre, y el BIC puede seleccionar modelos con menos precisión que otros criterios. Para muestras pequeñas, a menudo se prefiere el AIC corregido (AIC_c), que introduce una penalización adicional para compensar el sesgo en muestras finitas.

Finalmente, existe un debate sobre la idoneidad del BIC cuando el objetivo no es la identificación del "verdadero" modelo, sino la **predicción**. Críticos señalan que la consistencia del BIC, si bien es teóricamente deseable, a menudo resulta en la selección de modelos que son subóptimos para la predicción en el mundo real, especialmente cuando se trabaja con modelos no paramétricos o semi-paramétricos donde el concepto de un "modelo verdadero" simple es ambiguo o inexistente. En estos contextos, el fuerte sesgo del BIC hacia la simplicidad puede limitar su utilidad práctica en comparación con criterios orientados a la predicción como la validación cruzada o el AIC.

Lecturas Adicionales

Criterio de información bayesiano (Wikipedia)
Schwarz, Gideon E. "Estimating the dimension of a model." The Annals of Statistics 6.2 (1978): 461-464.
Selección de modelos (Wikipedia)