autorregresión – autoregression

Authored by
memjavad

November 3, 2025

Autoregresión

Primary Disciplinary Field(s): Estadística, Econometría, Análisis de Series Temporales, Aprendizaje Automático

1. Definición Central y Fundamentos Matemáticos

La autoregresión, o **AR** (por sus siglas en inglés, Autoregressive), es un concepto fundamental y un tipo de modelo estadístico utilizado para analizar y predecir series temporales. Un modelo autoregresivo asume que el valor de una variable en un momento dado, Y_t, depende linealmente de sus propios valores anteriores (o rezagos). Esta dependencia interna distingue a los modelos AR de la regresión lineal tradicional, donde la variable dependiente se explica por variables exógenas e independientes.

El modelo más básico y comúnmente estudiado es el modelo autoregresivo de orden p, denotado como AR(p). Matemáticamente, el modelo AR(p) se define mediante la ecuación: $Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t$, donde c es una constante, ϕ_i son los coeficientes autoregresivos que cuantifican la magnitud de la influencia de cada rezago, y ϵ_t es el término de error o ruido blanco, que representa las perturbaciones aleatorias no explicadas por el modelo. Es crucial que el término de error ϵ_t cumpla con las propiedades de ser independiente, idénticamente distribuido (i.i.d.), y que posea media cero y varianza constante (homoscedasticidad). La naturaleza lineal de esta relación permite una interpretación relativamente sencilla de los coeficientes: ϕ_1, por ejemplo, indica cómo una unidad de cambio en Y en el período anterior ($t-1$) afecta el valor actual de Y_t, manteniendo constantes los demás factores.

La importancia central de la autoregresión radica en su capacidad para capturar la **inercia** o persistencia temporal inherente a muchas series de datos económicos, financieros y naturales. Si una serie temporal muestra una correlación significativa entre sus valores actuales y pasados (autocorrelación), un modelo AR(p) puede proporcionar una representación parsimoniosa de la estructura dinámica subyacente. La selección del orden p es una tarea crítica, generalmente guiada por el análisis de la Función de Autocorrelación Parcial (FACP), que identifica el número de rezagos necesarios para explicar la variación actual, después de considerar la influencia de los rezagos intermedios.

2. Etimología y Desarrollo Histórico

El concepto de autoregresión surgió como una extensión natural de la teoría de la regresión lineal, adaptada al dominio de las series temporales. El desarrollo histórico se remonta a principios del siglo XX, impulsado por la necesidad de modelar fenómenos cíclicos y de persistencia en la naturaleza y la economía. El término y la formalización inicial de los modelos autoregresivos se atribuyen en gran medida al estadístico británico George Udny Yule.

En la década de 1920, Yule introdujo formalmente el concepto de autoregresión en su trabajo sobre la periodicidad de las manchas solares y las fluctuaciones económicas. Su artículo seminal de 1927, "On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers," sentó las bases al proponer que una variable podía ser predicha por sus propios valores pasados más un término de error aleatorio. Esto marcó un alejamiento de los modelos puramente deterministas hacia modelos estocásticos, que reconocen la influencia del azar y las perturbaciones no modeladas.

El concepto de AR se integró y floreció dentro del marco más amplio de los modelos de series temporales conocidos como ARMA (Autoregressive Moving Average) y ARIMA (Autoregressive Integrated Moving Average). Este desarrollo fue consolidado y popularizado por los estadísticos George Box y Gwilym Jenkins en su influyente libro de 1970, *Time Series Analysis: Forecasting and Control*. La metodología Box-Jenkins proporcionó un proceso sistemático y riguroso (identificación, estimación, diagnóstico) para seleccionar y aplicar modelos AR, MA o combinados, transformando la modelización de series temporales en una disciplina científica estandarizada. Este marco garantizó que la autoregresión se estableciera firmemente como una herramienta indispensable en econometría y estadística aplicada.

3. Tipos y Clasificación del Orden AR(p)

La clasificación de los modelos autoregresivos se basa principalmente en el orden p, que indica el número de rezagos que se incluyen como predictores en el modelo. La elección correcta de p es vital, ya que un modelo con p demasiado bajo puede subestimar la dependencia temporal (sesgo), mientras que un p demasiado alto puede capturar ruido y reducir la precisión de la predicción (sobreajuste).

Los modelos más comunes son:

AR(1): Es el modelo más simple, donde el valor actual de la serie depende únicamente del valor inmediatamente anterior: $Y_t = c + phi_1 Y_{t-1} + epsilon_t$. El AR(1) es fundamental para entender conceptos como la persistencia de choques. Si phi_1 es cercano a 1, la serie muestra una alta persistencia; si es cercano a 0, la serie es casi ruido blanco.

AR(2): Este modelo incluye los dos rezagos anteriores: $Y_t = c + phi_1 Y_{t-1} + phi_2 Y_{t-2} + epsilon_t$. Los modelos AR(2) son capaces de capturar dinámicas más complejas, incluyendo patrones cuasi-cíclicos u oscilatorios que no pueden ser explicados por un AR(1).

AR(p): La generalización para cualquier número p de rezagos. La determinación del valor óptimo de p se realiza típicamente utilizando criterios de información, como el Criterio de

Información de Akaike (AIC) o el Criterio de Información Bayesiano (BIC). Estos criterios penalizan la complejidad del modelo (aumentando p) mientras recompensan el buen ajuste a los datos.

Además de los modelos AR puros, la autoregresión se utiliza como componente esencial en modelos híbridos más avanzados. Por ejemplo, el modelo ARMA(p, q) combina la parte autoregresiva (AR(p)) con la parte de media móvil (MA(q)), donde q representa el número de rezagos del término de error. Si la serie temporal no es estacionaria, se requiere una diferenciación previa, llevando al modelo ARIMA(p, d, q), donde d es el orden de diferenciación necesario para inducir la estacionariedad. Estos modelos compuestos permiten abordar una gama mucho más amplia de estructuras de datos.

4. Propiedades Clave: Estacionariedad y Causalidad

Para que un modelo autoregresivo sea interpretable y útil para la predicción a largo plazo, debe cumplir con la propiedad fundamental de la **estacionariedad**. Una serie temporal es estacionaria en sentido débil si su media, varianza y estructura de autocorrelación no cambian con el tiempo. La estacionariedad asegura que la estructura probabilística del proceso que genera los datos es constante, permitiendo que los parámetros estimados a partir de datos históricos sean válidos para la predicción futura.

En el contexto del modelo AR(p), la estacionariedad está determinada por los valores de los coeficientes ϕ_i. Específicamente, las raíces de la ecuación característica (polinomio autoregresivo) deben caer fuera del círculo unitario en el plano complejo. Si una o más raíces caen sobre o dentro del círculo unitario, el proceso es no estacionario, resultando en una serie con varianza creciente o una deriva constante. El caso más crítico de no estacionariedad es la presencia de una **raíz unitaria** (una raíz igual a uno), lo que implica que los choques son permanentes. La detección de raíces unitarias requiere pruebas estadísticas específicas, siendo la Prueba de Dickey-Fuller Aumentada (ADF) la más utilizada.

Otra propiedad crucial es la **causalidad**. Un modelo AR(p) es causal si la serie actual (Y_t) puede expresarse como una función de los valores pasados del término de error ($\epsilon_{t}, \epsilon_{t-1}, \dots$). Para los modelos AR($p$) estacionarios, la condición de estacionariedad generalmente implica la causalidad. La causalidad garantiza que la predicción de Y_t solo dependa de la información disponible hasta el momento $t-1$, lo cual es fundamental para la modelización y la previsión.

5. Estimación de Parámetros y Diagnóstico

La estimación de los coeficientes ϕ_i en un modelo AR(p) se realiza típicamente utilizando el método de Mínimos Cuadrados Ordinarios (MCO) o el método de Máxima Verosimilitud (MV). Si el proceso es estacionario, la estimación por MCO es consistente y asintóticamente eficiente. Sin embargo, en muestras pequeñas, especialmente en modelos con alto grado de persistencia (coeficientes cercanos a la no estacionariedad), pueden surgir sesgos, lo que a veces favorece el uso de la Máxima Verosimilitud.

Una vez que los parámetros han sido estimados, es obligatorio realizar un diagnóstico riguroso del modelo. El diagnóstico se centra en el análisis del término de error, $hat{epsilon}_t$ (los residuos). Los requisitos clave para un modelo AR bien especificado son:

Ruido Blanco: Los residuos deben ser ruido blanco, es decir, no deben presentar autocorrelación significativa a ningún rezago. Esto se verifica mediante la inspección del correlograma de los residuos y el uso de pruebas formales, como la Prueba de Ljung-Box.

Normalidad: Los residuos deben seguir una distribución normal, lo cual es un supuesto necesario para la validez de las inferencias y los intervalos de confianza, aunque no siempre es estrictamente necesario para la consistencia de los estimadores MCO.

Homoscedasticidad: La varianza de los residuos debe ser constante a lo largo del tiempo. La heteroscedasticidad (varianza no constante) puede invalidar la precisión de los errores estándar estimados, requiriendo modelos más complejos como los de la familia ARCH/GARCH.

Si el diagnóstico revela problemas (por ejemplo, autocorrelación residual significativa), el modelo AR(p) original es inadecuado y debe ser reespecificado, ya sea aumentando el orden p o incorporando una parte de media móvil (MA) para formar un modelo ARMA.

6. Aplicaciones Prácticas y Extensiones

La autoregresión es una herramienta versátil con amplias aplicaciones en diversas disciplinas que manejan datos secuenciales o temporales. Su simplicidad y robustez la hacen ideal como punto de partida para el análisis de previsión.

En **Econometría y Finanzas**, los modelos AR se utilizan para predecir variables macroeconómicas como el Producto Interno Bruto (PIB), la inflación, y las tasas de interés. Aunque los precios de los activos financieros (como las acciones) a menudo se modelan mejor con modelos más complejos (como los GARCH para la volatilidad), la autoregresión sigue siendo fundamental en el análisis de retornos de activos ajustados o en la modelización de la persistencia de los choques en los mercados. La extensión multivariante de la autoregresión, conocida como Vector Autoregression (VAR), es indispensable para analizar las interrelaciones dinámicas entre múltiples series económicas, como la relación entre la política monetaria y el desempleo.

En **Ingeniería y Procesamiento de Señales**, los modelos AR se utilizan extensamente para la compresión de datos, la cancelación de ecos y la predicción lineal. Un proceso de autoregresión puede utilizarse para modelar la forma en que una señal de audio o una señal de radar se genera a partir de un proceso aleatorio subyacente, facilitando la identificación de patrones y la eliminación de ruido.

En **Ciencias Naturales**, la autoregresión ayuda a modelar fenómenos climáticos, como la temperatura o la precipitación, donde el valor actual está fuertemente influenciado por las condiciones de los períodos anteriores. También se utiliza en epidemiología para modelar la propagación de enfermedades, considerando cómo el número de casos actuales depende de los casos observados en semanas previas.

7. Limitaciones y Críticas

A pesar de su utilidad, el modelo autoregresivo puro (AR) presenta varias limitaciones importantes que deben ser consideradas por el analista:

Supuesto de Linealidad: Los modelos AR asumen una relación lineal entre el valor actual y sus rezagos. Muchas series temporales reales, especialmente en finanzas, exhiben dinámicas no lineales (umbrales, asimetrías, cambios de régimen) que un modelo AR lineal no puede capturar adecuadamente.

Sensibilidad a la No Estacionariedad: El modelo AR es extremadamente sensible a la no estacionariedad. Si la serie no es estacionaria y el analista no aplica la diferenciación (pasando a un modelo ARIMA), los estimadores de MCO serán inconsistentes, y las pruebas de hipótesis serán inválidas. El riesgo de regresión espuria es alto si se modelan series de tendencia sin diferenciación.

Dependencia Exógena Limitada: El modelo AR puro solo utiliza información interna (rezagos de la propia serie). Si el comportamiento de la serie está influenciado significativamente por variables externas (shocks de política, precios de materias primas, etc.), un modelo AR puro no será el más adecuado. En tales casos, se debe recurrir a extensiones como el modelo ARX (Autoregresión con variables Exógenas) o modelos VAR.

Pronóstico a Largo Plazo: Los modelos AR son generalmente más efectivos para el pronóstico a corto plazo. A medida que el horizonte de predicción se extiende, la dependencia de los valores pasados disminuye y los intervalos de confianza del pronóstico se amplían rápidamente, convergiendo la predicción hacia la media incondicional del proceso estacionario.

Lecturas Adicionales

Autoregressive model - Wikipedia

Yule, G. U. (1927). On a method of investigating periodicities in disturbed series. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or

Physical Character, 226, 267-298.

Box-Jenkins model - Wikipedia