Universidad Marista

ESCUELA DE ACTUARIA

TAREA ESTADISTICA III

José Antonio Sánchez Guevara

Profesor: CLAUDIO CUEVAS PAZOS

30 de abril de 2022

1. **1. El método más utilizado para estimar los parámetros de una serie de tiempo es llamado "Método de Máxima Verosimilitud", el cual consiste en maximizar la función de log-verosimilitud correspondiente a una muestra aleatoria. Como ejemplo, tomemos un proceso AR(1).**

Tenemos la funcion de densidad condicionando las variables $X_t|X_{t-1}$.

$f(X_t|X_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -(\frac{(X_t-\Phi X_{t-1})^2}{2\sigma^2})$
Para la sacar la funcion de verosimilitud usamos la definicion

$L(\Phi) = \prod_{t=1}^{n} f(X_t|X_{t-1})$
Primero calculamos el logaritmo de la funcion $ln(L(\Phi))$

$ln(L(\Phi)) = \sum_{t=1}^{n} ln(f(X_t|X_{t-1}))$

$ln(L(\Phi)) = \sum_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp -(\frac{(X_t-\Phi X_{t-1})^2}{2\sigma^2})$

$\frac{d}{d\sigma} = -\sum_{t=1}^{n} ln(\sqrt{2\pi\sigma^2}) - \sum_{t=1}^{n}(X_t - \Phi X_{t-1})^2(\frac{1}{2\sigma^2})$

$\frac{d}{d\sigma} = -\sum_{t=1}^{n} ln(\sqrt{2\pi\sigma^2}) - \sum_{t=1}^{n}(X_t - \Phi X_{t-1})^2(\frac{4\sigma^2}{\sqrt{2\sigma^2}})$

$\frac{d}{d\sigma} = -\sum_{t=1}^{n} \frac{1}{\sigma} + \sum_{t=1}^{n}(X_t - \Phi X_{t-1})^2(\frac{4\sigma^2}{\sqrt{2\sigma^2}})$

Igualamos a 0 tendriamos

$\sum_{t=1}^{n} \frac{4(X_t-\Phi X_{t-1})^2}{\sqrt{2}} = \frac{1}{\sigma}$

$\therefore \hat{\sigma} = \sum_{t=1}^{n} \frac{\sqrt{2}}{4(X_t-\Phi X_{t-1})^2}$

Ahora para Φ tendriamos

$\frac{d}{d\Phi} = -\sum_{t=1}^{n} ln(\sqrt{2\pi\sigma^2}) - \sum_{t=1}^{n}(X_t - \Phi X_{t-1})^2(\frac{1}{2\sigma^2})$

$\frac{d}{d\Phi} = -\sum_{t=1}^{n} 2(X_t - \Phi X_{t-1})X_{t-1}\frac{1}{2\sigma^2}$

$\frac{d}{d\Phi} = \sum_{t=1}^{n} \frac{(X_t - \Phi X_{t-1})X_{t-1}}{\sigma^2}$

Nuevamente igualamos a cero y despejamos Φ

$\sum_{t=1}^{n}(X_t - \Phi X_{t-1})X_{t-1} = 0$

$\sum_{t=1}^{n} X_t X_{t-1} - \Phi(X_{t-1})^2 = 0$

$\sum_{t=1}^{n} X_t X_{t-1} - \Phi \sum_{t=1}^{n}(X_{t-1})^2 = 0$

$\sum_{t=1}^{n} X_t X_{t-1} = \Phi \sum_{t=1}^{n}(X_{t-1})^2$

$\therefore \hat{\Phi} = \sum_{t=1}^{n} X_t X_{t-1} (\sum_{t=1}^{n}(X_{t-1})^2)^{-1}$

2. Uno de los indicadores más utilizados para determinar si el modelo utilizado en una serie de tiempo es el criterio de información de Akaike (Akaike Information Criterion). Investigue en qué consiste dicho indicador y dé una breve explicación al respecto.

Akaike propone estudiar el problema de la identificación desde la perspectiva de la teoría de la decisión estadística, lo que traslada el problema del ajuste de un modelo a la elección de la función de pérdida más adecuada. La solución dada por Akaike es elegir como función de pérdida el mínimo del criterio de información AIC, que simboliza por MAICE. Este criterio, definido por Akaike como "An Information Criterion", se basa en la medida de información de Kullback-Leibler, la cual permite interpretar la distancia entre dos distribuciones a partir de la log-verosimilitud de un modelo.

AIC = 2 ln(maxima verosimilitud) +2(parametros independientemente ajustados)

3. En clase vimos que un método muy utilizado para determinar el orden de una serie de tiempo (es decir, si el proceso es AR(1), AR(2), etc...), era calculando la ACF y la PACF de la serie de tiempo. No obstante, existe una alternativa que es encontrar el orden lat que el criterio de AIC se maximice. Implemente una función tal que para un orden máximo dado, la función busque sobre todos los órdenes menores a este valor máximo y seleccione el orden con un criterio de AIC máximo.

AIC de un proceso autoregresivo AR(k) es el siguiente:

$AIC(k) = (N - (k+1))ln(\sigma^2) + 2(k+1)$

Suponer que existe una cantidad de candidadps, dados los valores AIC para los modelos: AIC1, AIC2, AIC3, . . . , AICn el minimo de esos valores, entonces la función que maximiza al AIC puede interpretarse como proporcional a la probabilidad de que el k-ésimo modelo minimice la pérdida de la información.

$AIC_c = AIC + \frac{2k^2+2k}{n-k-1}$

4. Implemente una función en Python que calcule la función de autocorrelación parcial para una serie de tiempo dada.

Adujuntado en el script de python

5. Los datos adjuntos en la tarea corresponden al producto interno bruto trimestral de Estados Unidos desde 1947 hasta 1969.

6. Investiguen en qué consiste la gráca de Cuantil-Cuantil ó QQ-plot.

- **Realice un análisis exploratorio de los datos. La serie de tiempo es estacionaria? Qué alternativas propone para "quitarle"la tendencia a la serie de tiempo?**

  Se puede observar que los datos no son estacionarios pues mas bien siguen una tendencia lineal por lo cual el poder .[es]tabilizarla"por asi decirlo funcionaria buscar la función auto arima en python y un forecast calcular las predicciones tambien con el Diff eliminar la dependencia dentro de la serie de tiempo que pueden ser observadas como estacionalidad o como tendencias.

- **Calcule la ACF y la PACF del proceso estacionario utilizando tanto las librerías de Python como sus funciones implementadas. 3/4Qué pueden decir sobre los grácos observados?**

  Se puede observar que hay muchos valores de ACF altos por lo que la la serie es NO estacionaria y habrá que agregar una diferenciación con d pues la ACF si muestra una caida muy lentamente suele ser una serie no estacionaria.

- **Ajuste un modelo, ya sea AR o MA utilizando las librerías de Python. Para determinar el órden de los modelos, utilice tanto el criterio de ACF y PACF como el criterio mencionado en la pregunta 3.**

  Adjuntado en el script en Python

- **Haga un análisis de los resultados del modelo que eligieron. Es buen modelo? Qué pueden decir sobre la signicancia de los parámetros del modelo?**

  Al hacer un ajuste de los datos a un modelo AR podemos observar que en los residuos no se tiene buena acptacion de modelo ya que la grafica de los residuos no vemos una buena adecuacion al modelo ni homocedasticidad

- **Hagan un análisis de los residuales resultantes del modelo? Cómo se distribuyen? Qué dice la QQ-Plot al respecto?**

  En la grafica q-q plot observamos que el comportamiento de los residuos no se apegan con la recta lo que indica que no se ditribuyen normal dos residuos ademas de l pvalue de un 0.52 nos dice que no es un buen ajuste del modelo

- **A partir de su modelo, pueden realizar un pronóstico de la serie de tiempo utilizando la función "forecastçontenida en la paquetería "statsmodels.en Python. Investiguen el funcionamiento de dicha función y hagan un pronóstico de 5**

trimestres adelante. Qué opinan de los resultados?
Como hemos visto la tendencia que sigue esta serie de datos es lineal por lo que los pronosticos dados por el modelo siguen aun asi la tendencia lineal y el mismo comportamiento que esta llevando el modelo