Teorías demográficas

De acuerdo a @alcalde2010teoria y @mariscal2018tres, una de las teorías demográficas más importantes es la Teoría de la Transición Demográfica (TTD). Esta consiste en una generalización empírica en función de observaciones pasas y establece una conexión entre la evolución demográfica de la población y el crecimiento económico [@mariscal2018tre].

Modelos estadísticos propuestos

Con la finalidad de realizar un pronóstico de la serie de defunciones totales anuales de Costa Rica, se desea implementar el modelo estadístico que mejor se ajuste a los datos.

Para nuestro estudio en cuestión, se ha optado por realizar una implementación de Modelos de Espacio-Estado.Particularmente, Modelos Dinámicos Lineales (DLM).

Tal como lo establece @petrisDLM, estos últimos son una clase de Modelos de Espacio-Estado también llamados Modelos de Espacio-Estado Lineales Gaussianos. Estos modelos son especificados mediante dos ecuaciones, para $t \geq 1$ se tiene:

$$Y_t = F_t\theta_t + v_t,$$
$$\theta_t = G_t\theta_{t-1} + w_t$$

$$(\theta_0|D_0) \sim \mathcal{N}(m_0, C_0)$$

Donde la primer ecuación es llamada ecuación de observación,la segunda ecuación estado o ecuación del sistema y la última información inicial.

Es importante señalar que F_t y G_t son matrices y (v_t) , (w_t) son secuencias de ruidos blancos independientes tales que:

$$v_t \sim \mathcal{N}_m(0, V_t),$$
$$w_t \sim \mathcal{N}_p(0, W_t)$$

Los DLM poseen dos supuestos, la linealidad y el supuesto de distribuciones Gaussianas. @petrisDLM señala que este último supuesto puede ser justificado mediante argumentos del teorema del límite central.

La estimación y pronóstico se pueden resolver calculando las distribuciones condicionales de las cantidades de interés, dada la información disponible. Para estimar el vector de estados es necesario computar la densidad condicional $p(\theta_t|y_1, .., y_t)$. En particular, nos interesa el

problema de filtrado (cuando $s = t$), donde los datos se supone que llegan secuencialmente en el tiempo.

En general, el problema de pronóstico de k-pasos hacia adelante consiste en estimar la evolución del sistema θ_{t+k} para $k \geq 1$ y realizar un pronóstico de k-pasos para Y_{t+k}.

Según @petrisDLM en los DLM, el filtro de Kalman proporciona las fórmulas para actualizar nuestra inferencia actual sobre el vector de estado conforme se disponga de nuevos datos.

Para un DLM, si se cumple que:

$$\theta_t | \mathcal{D}_t \sim \mathcal{N}(m_t, C_t), t \geq 1$$

Se tiene que:

La densidad de predicción de estado de k-pasos con $k \geq 1$ hacia adelante de θ_{t+k} dada la información pasada D_t es Gaussiana con media y varianza condicional dadas respectivamente por:

$$a_t(k) = E[\theta_{t+k}|D_t] = G_{t+k}a_{t,k-1}$$
$$R_t(k) = Var[\theta_{t+k}|D_t] = G_{t+k}R_{t,k-1}G_{t+k}^{'} + W_{t+k}$$

La densidad de predicción de k-pasos con $k \geq 1$ hacia adelante de Y_{t+k} dada la información pasada D_t, es Gaussiana con media y varianza condicional dadas respectivamente por:

$$f_t(k) = E[Y_{t+k}|D_t] = F_{t+k}a_t(k)$$
$$Q_t(k) = Var[Y_{t+k}|D_t] = F_{t+k}R_t(k)F_{t+k}^{'} + V_{t+k}$$

Justificación modelos DLM propuestos

Como se mencionó en **?@fig-defunciones_ano** , la cantidad de defunciones totales siguen una cierta tendencia lineal creciente, en particular para años posteriores a 1980.

Debido a que esta es nuestra variable de interés para realizar un pronóstico, es propicio para nuestro estudio en cuestión la implementación de un modelo con supuesto de linealidad, como se mencionó justamente los DLM siguen este supuesto.

Para llevar a cabo los pronósticos se proponen por tanto tres métodos estadísticos pertenecientes a los DLM, estos son: modelo DLM polinomial de primer orden, modelo DLM polinomial de segundo orden y el modelo ARIMA(p, d, q).

Modelo ARIMA(p, d, q)

Como una primer implementación se utiliza un modelo ARIMA(p, d, q). Tal como lo menciona @petrisDLM un modelo ARIMA(p, d, q) puede ser considerado un DLM, esto ya que es posible representar todo modelo ARIMA(p,d,q) (ya sea univariado o multivariado) como un DLM.

La escogencia de este modelo al ser un DLM, sigue la misma línea de justificación antes mencionado sobre la elección de modelos DLM para nuestro estudio, siendo este un caso particular de estos.

Sin embargo, es importante mencionar que la escogencia de este modelo como primera implementación también se basa en su simplicidad, y en que dada la bibliografía consultada, se observa que en múltiples investigaciones con temáticas relacionadas a nuestro estudio como el de @adekambi y el estudio por @ordorika, se implementa este tipo de modelo.

Tal como lo establece @petrisDLM entre los modelos más utilizados para el análisis de series temporales se encuentra la clase de modelos de media móvil autorregresiva (ARMA). Para enteros, no negativos p y q, un modelo ARMA(p, q) es definido mediante la notación:

$$Y_t = \mu + \sum_{j=1}^{p} \phi_j (Y_{t-j} - \mu) + \sum_{j=1}^{p} \psi_j \epsilon_{t-j} + \epsilon_t$$

Donde (ϵ_t) es una ruido blanco Gaussiano con varianza σ_ϵ^2 y los parámetros $\phi_1, \phi_2, ..., \phi_p$ satisfacen la condición de estacionariedad.

Cuando los datos no presentan estacionariedad, se suele tomar las diferencias hasta que se obtenga esta, una vez obtenida se procede a ajustar el modelo ARMA(p, q) a la data diferenciada.

Un modelo para un proceso cuya d-ésima diferencia sigue un modelo ARMA(p, q) es llamado un ARIMA(p, d, q).

La escogencia de los ordenes p y q pueden ser escogidos de una manera informal, observando la autocorrelación empírica y la autocorrelación parcial, o utilizando un criterio de selección de modelos más formal como lo es el AIC y BIC.

Modelos polinomiales de primer y segundo orden

Se propone un modelo DLM de primer orden ya que como establece @optimalDLM los DLM de primer orden son algoritmos recomendados al lidiar con datos anuales debido a que las series de tiempo es corta y no presentan patrones estacionales.Dado que nuestros datos son anuales, este modelo se presenta como un posible candidato.

Por su parte @optimalDLM, señala que los DLM de segundo orden son útiles para describir tendencias. Dada la tendencia observada de la serie de defunciones totales sugiere por tanto realizar un modelo polinomial de segundo orden.

Es oportuno señalar que el desarrollo teórico de estos modelos se llevará a cabo en bitácoras posteriores, ya que se considera prudente primero realizar la implementación de estos (tal como se realizó el ARIMA(p, d, q)) para ver sus alcances para responder la pregunta de investigación.