



Puebla, August 31 to September 4, 2020.

Objetivo

Presentar una aplicación del modelo Poisson no homogéneo con puntos de cambio.

Introducción

Se expone un trabajo de investigación donde se utiliza un modelo Poisson no homogéneo considerando puntos de cambio y es aplicado a datos de ozono en la ciudad de Puebla, del periodo 1 de enero del 2001 al 31 de diciembre de 2009.

Algunas definiciones importantes

- 1 Un proceso de Poisson de parámetro $\lambda > 0$, es un proceso a tiempo continuo $\{X_t; t \geq 0\}$ con espacio de estados $\{0, 1, \dots\}$, con trayectorias no decrecientes y que satisfacen:

- 1 $X_0 = 0$.
- 2 Tiene incrementos independientes.
- 3 $X_{t+s} - X_s \sim \text{Poisson}(\lambda t)$, para cualesquiera $s \geq 0, t > 0$.

- 2 Un Proceso de Poisson No Homogéneo es un proceso a tiempo continuo $\{X_t; t \geq 0\}$, con espacio de estados $\{0, 1, \dots\}$, con parámetro la función positiva y localmente integrable $\lambda(t)$, y que cumple las siguientes propiedades [0]:

- 1 $X_0 = 0$.
- 2 Tiene incrementos independientes.
- 3 Para cualquier $t \geq 0$ y cuando $h \searrow 0$,
 - 1 $P(X_{t+h} - X_t \geq 1) = \lambda(t)h + o(h)$.
 - 2 $P(X_{t+h} - X_t \geq 2) = o(h)$.

Planteamiento

Sea N_t , indica el número de veces que un estándar ambiental de cierto contaminante ha sido rebasado en un intervalo de tiempo $[0, t)$, $t \geq 0$. Además asumimos que $\mathcal{N} = \{N_t : t \geq 0\}$ sigue un proceso Poisson no homogéneo con función de intensidad $\lambda(t) > 0$, $t > 0$ y función media $m(t) = \int_0^t \lambda(s) ds$.

Cuando se conoce el comportamiento de $\lambda(t)$, entonces la probabilidad de que el estándar ambiental sea violado en el intervalo de tiempo $[t, t + s)$, podrá ser calculado a través de:

$$P(N_{t+s} - N_t = k) = e^{-[m(t+s)-m(t)]} \frac{[m(t+s) - m(t)]^k}{k!}, \quad t, s \geq 0$$

Cuando se considera la presencia de puntos de cambio sobre el modelo, se tiene los siguientes resultados: Sea $I \geq 0$ y $\tau = \{\tau_1, \tau_2, \dots, \tau_I\}$ el conjunto de puntos de cambio que hay sobre el modelo, de esta manera la función de intensidad para el proceso es dada por:

$$\lambda(t) = \begin{cases} \lambda_1(t), & 0 \leq t < \tau_1 \\ \lambda_i(t), & \tau_{i-1} \leq t < \tau_i, \quad i = 2, 3, \dots, I \\ \lambda_{I+1}(t), & \tau_I \leq t \leq T, \end{cases} \quad (1)$$

mientras que la función media del proceso es:

$$m(t) = \begin{cases} m_1(t), & 0 \leq t < \tau_1 \\ m_1(\tau_1) + m_2(t) - m_2(\tau_1), & \tau_1 \leq t < \tau_2 \\ m_{j+1}(t) - m_{j+1}(\tau_j) \\ + \sum_{i=2}^j [m_i(\tau_i) - m_i(\tau_{i-1})] \\ + m_1(\tau_1), & \tau_j \leq t \leq \tau_{j+1}, \\ & j = 2, 3, \dots, I, \end{cases}$$

donde $\lambda_{I+1} = T$

La formulación del problema de interés se realiza mediante inferencia bayesiana, haciendo uso del método Monte Carlo vía Cadenas de Markov, y considerando la función de intensidad de forma Weibull y Musa-Okumoto

Desarrollo

Se consideran la función de riesgo Weibull y Musa Okumoto cuyas funciones media y de intensidad son:

Modelo	Función de intensidad $\lambda(t \theta)$	Función media $m(t \theta)$
Weibull(W)	$(\alpha/\sigma)(t/\sigma)^{\alpha-1}$	$(t/\sigma)^\alpha$
Musa-Okumoto(MO)	$\frac{\beta}{t+\sigma}$	$\beta \log(t/\sigma + 1)$

Cuadro: Función media y de intensidad

θ hace referencia al vector de parámetros y al punto de cambio a estimar en cada una de las funciones anteriores, por tanto, el problema se reduce a estimar el vector de parámetros.

De acuerdo al desarrollo Bayesiano, para determinar la distribución a posteriori del vector de parámetros dados los datos, se tiene la siguiente ecuación,

$$p(\theta|D) \propto p(\theta)l(\theta|D) \quad (2)$$

donde, $l(\theta|D)$ es La función de verosimilitud, D el conjunto de observaciones, $p(\theta)$ es la función a priori del vector de parámetros que se consideran a priori independientes.

Nota: En la aplicación del modelo, se realiza programas en Openbugs considerando un, dos, tres cuatro y cinco puntos de cambio. Estos puntos de cambio son observados mediante el gráfico de media observado, en este trabajo solo presentamos los resultados con un punto de cambio.

La función de verosimilitud cuando se consideran puntos de cambio es dada por:

$$l(\theta|D) = \prod_{i=1}^{N_{\tau_1}} \lambda(d_i|\theta_1) e^{-m(\tau_1|\theta_1)} \left[\prod_{j=2}^J \left(\prod_{i=N_{\tau_{j-1}}}^{N_{\tau_j}} \lambda(d_i|\theta_j) e^{-[m(\tau_j|\theta_j) - m(\tau_{j-1}|\theta_{j-1})]} \right) \right] \\ \times \prod_{i=N_{\tau_J}+1}^{N_T} \lambda(d_i|\theta_{J+1}) e^{-[m(T|\theta_{J+1}) - m(\tau_J|\theta_{J+1})]},$$

donde N_{τ_j} es el número de excedencias después del punto de cambio τ_j , $j = 1, 2, \dots, I$. y el conjunto $D = \{n : d_1, \dots, d_{N_{\tau_1}}, d_{N_{\tau_1}+1}, \dots, d_n; T\}$, donde d_k , $k = 1, 2$, es el tiempo de ocurrencia del k -ésimo evento ocurrido.

Observemos la distribución a posteriori del proceso considerando la función de intensidad de Weibull y un punto de cambio, en este caso considere a $\alpha = (\alpha_1, \alpha_2)$ y $\sigma = (\sigma_1, \sigma_2)$.

$$P(\alpha, \sigma, \tau|D) \propto \left(\prod_{i=1}^{N_{\tau_1}} \left(\frac{\alpha_1}{\sigma_1} \right) \left(\frac{d_i}{\sigma_1} \right)^{\alpha_1-1} \right) e^{-(\tau)/\sigma_1} \left(\prod_{i=N_{\tau_1}+1}^{N_T} \left(\frac{\alpha_2}{\sigma_2} \right) \left(\frac{d_i}{\sigma_2} \right)^{\alpha_2-1} \right) \\ \times (e^{-(T/\sigma_2) + (\tau/\sigma_2)} P(\alpha) P(\sigma) P(\tau)).$$

Aplicación y resultados.

Tomando en consideración la base de datos respecto a una de las estaciones que conforman la Red de Monitoreo Atmosférico del Estado de Puebla, se seleccionan los máximos diarios de 01 de enero de 2001 a 31 de diciembre de 2009 y se completa los datos faltantes. De está forma se obtiene una base de datos para el municipio de Puebla.

Las observaciones que se consideran, son los rebases del contaminante ozono que sobrepasan el umbral de las 0.11ppm, a partir de ella se obtiene un resumen de estadísticas descriptivas y se usa el Open Bus para estimar los parámetros y se observan los siguientes resultados.

Función de riesgo	α_1	α_2	σ_1	σ_2	τ
Weibull	U(0,3)	U(0,2)	U(10,70)	Gamma(6.56,0.18)	U(150,300)
Musa-Okumoto	U(0,50000)	U(0,2000)	U(10,10000)	U(50,200)	U(50,150)

Cuadro: A priori usadas para las estimaciones de los parámetros.

Modelo	Parámetro	media	sd	intervalo de credibilidad 95 %
Weibull	α_1	2.889	1.12	(0.932, 4.859)
	σ_1	43.86	12.84	(19.5, 73.24)
	α_2	1.956	0.324	(1.29, 2.458)
	σ_2	40.16	12.43	(14.38, 58.72)
	τ	65	11.34	(50.44, 88.5)
M-O	β_1	39280	13.930	(9571, 59100)
	σ_1	2412	1151	(466, 7, 4689)
	β_2	1410	639.7	(206, 3, 2690)
	σ_2	304.2	103.2	(85, 51, 444.9)
	τ	104	14.47	(77.03, 126.1)

Cuadro: Parámetros estimados.

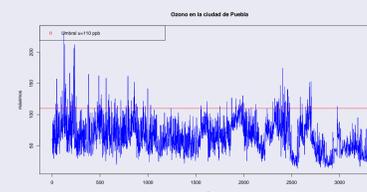


Figura: Serie de tiempo

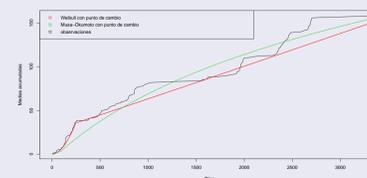


Figura: Medias acumuladas

Conclusiones.

En el gráfico de las medias acumuladas con respecto a la media de los datos, se puede verificar que el modelo que mejor se ajusta a nuestras observaciones considerando un punto de cambio es el proceso con función de intensidad de Weibull, sin embargo cabe resaltar que para la estimación del punto de cambio nos basamos de acuerdo al gráfico de medias acumuladas sin puntos de cambio.

Bibliografía

- Achcar A. Jorge, Fernández-Bremauntz Adrián A., Rodrigues Eliane, Tzintzun Guadalupe, *Estimating the number of ozone peaks in Mexico City using a non-homogeneous Poisson model*, Environmetrics, 19: 469-485, 2008.
- Achcar A. Jorge, Barrios M. Juan, Rodrigues Eliane R., *Comparing the Adequacy of Some Non-Homogeneous Poisson Models to Estimate Ozone Exceedances in Mexico City*, Journal of Environmental Protection, 3, 1213-1227, 2012.
- Achcar A. Jorge, Rodrigues Eliane, Tzintzun Guadalupe, *Using non-homogeneous Poisson models with multiple change-points to estimate the number of ozone exceedances in Mexico City*, Environmetrics; 22: 1-12, 2011.
- Rincón L.; *Introducción a los procesos estocásticos*, Departamento de Matemáticas, Facultad de ciencias UNAM, 2012.
- Rodrigues Eliane R., Achcar Jorge A., *Applications of Discrete-time Markov Chain and Poisson Processes to Air Pollution Modeling and studies*, Editorial Springer, 2013

Agradecimientos.

- Secretaría de Desarrollo Rural, Sustentabilidad Ambiental y Ordenamiento Territorial, (SDRSOT Puebla), por proporcionar las observaciones requeridas para el estudio.
- Act. Guadalupe Tzintzun, mediante el apoyo referente a información de los datos.
- Este trabajo recibió apoyo financiero de los proyectos PAPIIT-IN102713 y PAPIIT-IN102416 de la Dirección General de Apoyo al Personal Académico de la Universidad Nacional Autónoma de México, UNAM.