



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**MODELO DE POISSON NO HOMOGENEO
APLICADO A LA ESTIMACIÓN
DE LA CONTAMINACIÓN
ATMOSFÉRICA**

T E S I N A

**QUE PARA OBTENER EL TÍTULO DE:
ACTUARIO**

**PRESENTA:
RAYMUNDO IBARRA CARRASCO**

**DIRECTOR DE TESINA:
DRA. ELIANE REGINA RODRIGUES**



2011



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

En primer lugar y muy especialmente quiero agradecerle a la Dra. Eliane R. Rodrigues, mi tutora, por toda su dedicación y consejo para conmigo, quien a lo largo de este seminario de titulación me apoyo en todo momento.

A mi familia, Julio, Edgar Lucia, Luis y en especial a mi madre Norma, agradecerle es poco, le debo todo el amor que me dá y el no haberme dejado bajar los brazos nunca y apoyarme siempre.

Índice

Introducción	1
Capítulo 1: Procesos de Poisson	2
1.1 Proceso Estocástico.....	2
1.2 Construcción del Proceso Poisson.....	3
1.3 Proceso Poisson Homogéneo.....	4
1.4 Proceso Poisson no-Homogéneo.....	10
Capítulo 2: Introducción a la Estadística Bayesiana	15
2.1 Función de verosimilitud.....	16
2.2 Distribución a priori.....	16
2.3 Distribución a posteriori.....	18
2.4 Muestreo de Gibbs.....	20
2.5 Construcción de la función de verosimilitud para el modelo de Poisson.....	21
Capítulo 3: Modelo Bayesiano	24
Capítulo 4: Aplicación de las mediciones de Ozono en la ciudad de México	30
Conclusiones	34
Referencias	35

Prefacio

Esta Tesina describe el Modelo de Poisson no Homogéneo aplicado a la Estimación de la Contaminación Atmosférica, tema de gran interés y relación con la salud de los habitantes de la ciudad de México y Área Metropolitana.

Con esta Tesina se cumple el trabajo escrito para el requisito de la opción de seminario de titulación en la licenciatura en Actuaría bajo la dirección de la Doctora Eliane R. Rodrigues.

Sumario

El objetivo de esta tesina es estudiar la aplicación del modelo Poisson no homogéneo a la estimación de la contaminación atmosférica. En el primer capítulo hay definiciones básicas de Procesos Estocásticos, la construcción del Proceso Poisson y el planteamiento del modelo Poisson homogéneo y el no-homogéneo con sus respectivas definiciones así como las demostraciones de las mismas.

En el capítulo 2, se encuentra la introducción a la estadística Bayesiana definiciones de la función de verosimilitud, distribución a priori y posteriori, la demostración de que para el parámetro η (conocido y desconocido) hay una distribución $P_{priori}(\theta|\eta)$ a priori. Se describe el algoritmo de muestreo de Gibbs con la definición y origen de éste además de un ejemplo, se realiza la construcción de la función de verosimilitud del modelo Poisson para el caso homogéneo y no homogéneo.

En el capítulo 3, se presenta el desarrollo del modelo Bayesiano cuya función de verosimilitud sigue el modelo Poisson no homogéneo con la función de intensidad $\lambda(t)$, $t \geq 0$ Weibull-exponenciada que depende de parámetros por calcular. Se plantea la distribución a priori y se hace el planteamiento y obtención de las distribuciones a posteriori marginales condicionales completas que se necesitan para generar los valores α , β y σ en el algoritmo del Muestreo de Gibbs.

En el capítulo 4, se describe la aplicación de las mediciones de ozono en la ciudad de México y se describen los resultados que se obtuvieron en el artículo (Achcar et al. 2008) para los años de 1998 a 2004. Se describe las condiciones (cantidad de datos, valores en las distribuciones, número de pasos, Criterio de información Desviada entre otros) que tomaron para las simulaciones que se realizaron por el programa Winbugs. Se analizan los diferentes valores límite y zonas (incluyendo el caso del AMCM). Por último con los valores obtenidos se plantean de cómo se obtienen las probabilidades y con ello se puede prever el número de días en donde el límite de la NOM es sobrepasado. Finalmente se presentan conclusiones del trabajo.

Introducción

La contaminación atmosférica y los altos niveles de ozono son problemas que afectan a los habitantes de las grandes y medianas ciudades de todo el mundo. Cuando la concentración de ozono está por encima del límite 0.11ppm (ppm=partes por millón), durante un periodo prolongado de tiempo, los individuos expuestos a la contaminación podrían experimentar serios problemas de salud. Por lo tanto es muy importante predecir cuando podrían ocurrir estos episodios. El nivel de concentración o límites permitidos considerados pueden variar de ciudad en ciudad dependiendo las autoridades ambientales. El nivel de concentración de ozono permitido en el aire de México es de 0.11ppm y un individuo no debería estar expuesto a esta concentración por más de 1 hora (Norma Oficial Mexicana NOM, 2002). Sin embargo, en la Ciudad de México el nivel usado para declarar una alerta ambiental es de 0.22ppm.

Existen varios estudios que se enfocan a predecir episodios alarmantes de contaminación ambiental como los de: Roberts (1979a, b), Horowitz (1980) y Smith (1989) que utilizan teoría de valores extremos; Flaum (1996) quien utilizó análisis de series de tiempo; Guardani (1999) redes neuronales y Guardani (2003) a través del análisis multivariado; Huerta y Sansó (2005) desarrollaron un análisis considerando medidas máximas de ozono de la ciudad de México.

En las últimas décadas algunos autores han observado la posibilidad de utilizar cadenas de Markov, para predecir la ocurrencia de niveles de ozono perjudiciales. El supuesto más importante es que los intervalos que contienen sucesivos máximos diarios de ozono siguen una cadena de Markov. Algunos autores que utilizan este supuesto son por ejemplo Austin y Tran (1999) y Larsen (1990). La característica común de su trabajo es que supusieron que el orden de la cadena de Markov es conocido (e igual a 1) y calcularon la probabilidad de transición utilizando el método de la función de máxima verosimilitud. En el trabajo de (Álvarez et al. 2005) también se utilizó el modelo de cadenas de Markov. La diferencia en este es que el orden de la cadena de Markov es considerado como una variable aleatoria y se calcula utilizando el método de máximo a posteriori.

Cuando el objetivo es calcular el número de veces que se rebasa los límites permitidos de calidad atmosférica, Javits (1980) utilizó modelos Bernoulli y Poisson homogéneo y Raftery (1989) utilizó una combinación de modelos Poisson homogéneo. No obstante en todos los casos tenemos modelos homogéneos en el tiempo y la homogeneidad en el tiempo no es una propiedad de una sucesión de mediciones de ozono.

Capítulo 1

Procesos de Poisson

En este capítulo se describen los conceptos básicos del Proceso Poisson. El contenido de lo presentado aquí se tomó de (Ross, 1996 Tudor, 2002).

1.1 Proceso Estocástico

Definición 1: Un **proceso estocástico** en tiempo continuo $N = \{N(t), t \geq 0\}$ es llamado **puntual** o de **conteo** si $N(t)$ representa el número de veces que un evento de interés ocurre hasta el tiempo t . De tal forma, N debe satisfacer.

- i) $N(t) \geq 0$,
- ii) $N(t)$ es entero,
- iii) Si $s < t$, entonces $N(t) - N(s) \geq 0$, es igual al número de eventos que ocurre en el intervalo $(s, t]$.

Definición 2: Un proceso de conteo se dice de **incrementos independientes** si el número de veces que un evento de interés ocurre en intervalos de tiempo disjuntos son independientes, es decir, el número de sucesos $N(t_2) - N(t_1)$ en el intervalo (t_1, t_2) , es independiente del número de sucesos $N(t_4) - N(t_3)$ en (t_3, t_4) , para todo $t_1, t_2, t_3, t_4 \geq 0$ tal que $(t_1, t_2) \cap (t_3, t_4) = \emptyset$.

Definición 3: Un proceso de conteo se dice de **incrementos estacionarios** si la distribución del número de eventos que ocurren en un intervalo de tiempo sólo depende del tamaño del intervalo, es decir, el número de sucesos $N(t_2 + s) - N(t_1 + s)$ que ocurren en el intervalo $(t_1 + s, t_2 + s)$, tiene la misma distribución que el número de sucesos $N(t_2) - N(t_1)$ en (t_1, t_2) , para todo $t_1, t_2 \geq 0$, $t_1 < t_2$ y $s \geq 0$. (Ross, 1996).

1.2 Construcción del Proceso Poisson

Considere una medida de probabilidad \mathbb{P} y una sucesión τ_i de variables aleatorias independientes exponenciales con parámetro λ , $i = 1, 2, \dots$. Tome $\tau_0 = 0$ y suponga que τ_i representa el tiempo entre la realización de $(i - 1)$ -ésimo y el i -ésimo evento $i = 1, 2, \dots$. Sea S_k el tiempo de ocurrencia del k -ésimo evento, $k = 1, 2, \dots$, entonces $S_k = \sum_{i=1}^k \tau_i$.

Note que por hipótesis se tiene que

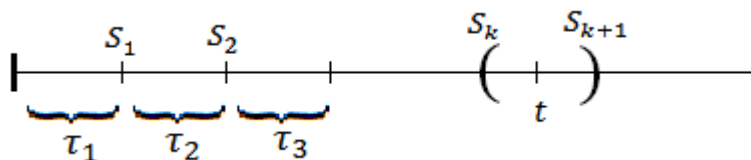
$$\mathbb{P}(\tau_i > s) = e^{-\lambda s}, \quad s > 0,$$

por lo tanto S_k tiene una distribución Gamma (λ, k) , es decir, si $f_{S_k}(s)$ es la función de densidad de S_k , entonces

$$f_{S_k}(s) = \frac{\lambda^k}{(k-1)!} s^{k-1} e^{-\lambda s}, \quad s > 0.$$

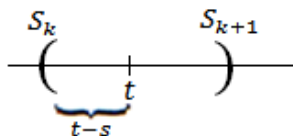
Sea $N(t)$ el número de eventos que ocurren en el intervalo de tiempo $[0, t)$. Entonces, se tiene que

$$\{N(t) = k\} = \{S_k \leq t, S_{k+1} > t - s\} \quad t \geq 0.$$



Por lo consiguiente, para una $k \in \mathbb{N}$ se tiene

$$\mathbb{P}\{N(t) = k\} = \mathbb{P}\{S_k \leq t, S_{k+1} > t - s\} = \mathbb{P}\{S_{k+1} \geq t - s | S_k \leq t\} \mathbb{P}\{S_k \leq t\}.$$



Dado que $\mathbb{P}\{S_k \leq t\} = \int_0^t f_{S_k}(s) ds$ se tiene que

$$\begin{aligned} \mathbb{P}\{N(t) = k\} &= \int_0^t \mathbb{P}\{S_{k+1} > t - s | S_k = s\} f_{S_k}(s) ds \\ &= \int_0^t \frac{\lambda^k}{(k-1)!} s^{k-1} e^{-\lambda s} e^{-\lambda(t-s)} ds \end{aligned}$$

$$\begin{aligned}
&= \frac{\lambda^k}{(k-1)!} \int_0^t s^{k-1} e^{-\lambda s} e^{-\lambda t + \lambda s} ds \\
&= \frac{\lambda^k}{(k-1)!} \int_0^t s^{k-1} e^{-\lambda t} ds = \frac{\lambda^k}{(k-1)!} e^{-\lambda t} \int_0^t s^{k-1} ds \\
&= \frac{\lambda^k}{(k-1)!} e^{-\lambda t} \left. \frac{s^k}{k} \right|_0^t = \frac{\lambda^k}{(k-1)!} e^{-\lambda t} \left[\frac{t^k}{k} - \frac{0^k}{k} \right] \\
&= \frac{\lambda^k}{(k-1)!} e^{-\lambda t} \frac{t^k}{k} = \frac{(\lambda t)^k}{k!} e^{-\lambda t} .
\end{aligned}$$

Por lo tanto se concluye que el número de eventos en un intervalo $[0, t)$, cuyos tiempos entre ocurrencias tienen distribución Poisson (Tudor, 2002).

1.3 Proceso Poisson Homogéneo

Definición 4: Un proceso de conteo $N = \{N(t), t \geq 0\}$ es un **proceso de Poisson** con tasa, $\lambda > 0$ si

- i) $N(0) = 0$.
- ii) El proceso tiene incrementos independientes y estacionarios.
- iii) $\mathbb{P}[N(t+s) - N(s) = n] = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$, $n \in \mathbb{N}$ y $s, t \geq 0$.

Observación: Así, el número medio de sucesos hasta el instante t es:

$$E[N(t)] = E[N(t+0) - N(0)] = \lambda t$$

esto se debe a que $N(t+0) - N(0)$ tiene distribución Poisson con parámetro λt y por lo tanto

$$\begin{aligned}
E[N(t)] &= \sum_{n=0}^{\infty} \frac{n(\lambda t)^n e^{-\lambda t}}{n!} = e^{-\lambda t} (\lambda t) \sum_{n=1}^{\infty} \frac{(\lambda t)^{n-1}}{(n-1)!} \\
&= e^{-\lambda t} (\lambda t) \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} = e^{-\lambda t} (\lambda t) e^{\lambda t} = \lambda t .
\end{aligned}$$

Definición 5: Una función $f(x)$ es $o(h)$ si

$$\lim_{h \rightarrow 0} \frac{f(h)}{h} = 0 .$$

Por ejemplo $f(h) = h^2$ es $o(h)$ dado que $\lim_{h \rightarrow 0} \frac{f(h)}{h} = \lim_{h \rightarrow 0} h = 0$, pero para $f(h) = h$ no lo es ya que $\lim_{h \rightarrow 0} \frac{f(h)}{h} = 1$ (Ross, 1996).

Definición 6: Un proceso de conteo $N = \{N(t), t \geq 0\}$ se dice que es un **proceso de Poisson homogéneo** con tasa $\lambda > 0$ si:

- i) $N(0) = 0$.
- ii) N tiene incrementos independientes y estacionarios.
- iii) $\mathbb{P}[N(h) = 1] = \lambda h + o(h)$.
- iv) $\mathbb{P}[N(h) \geq 2] = o(h)$.

Teorema 1. Las definiciones 4 y 6 son equivalentes.

Demostración: Primero se demostrará que la Definición 6 implica en la Definición 4. Defina $\mathbb{P}_n(t) = \mathbb{P}[N(t) = n]$.

Se derivará una ecuación diferencial para $\mathbb{P}_0(t)$ de la siguiente manera:

$$\begin{aligned} \mathbb{P}_0(t+h) &= \mathbb{P}[N(t+h) = 0] = {}^1\mathbb{P}[N(t) = 0, N(t+h) - N(t) = 0] \\ &= {}^2\mathbb{P}[N(t) = 0]\mathbb{P}[N(t+h) - N(t) = 0] = {}^3\mathbb{P}[N(t) = 0]\mathbb{P}[N(h) = 0] \\ &= {}^4\mathbb{P}_0(t)[1 - \lambda h + o(h)]. \end{aligned}$$

Por lo tanto, se tiene que

$$\mathbb{P}_0(t+h) = \mathbb{P}_0(t) - \lambda h \mathbb{P}_0(t) + o(h) \mathbb{P}_0(t).$$

Note que $o(h) \mathbb{P}_0(t) = o(h)$, de esta forma se puede escribir

$$\mathbb{P}_0(t+h) - \mathbb{P}_0(t) = -\lambda h \mathbb{P}_0(t) + o(h)$$

y por lo tanto

$$\frac{\mathbb{P}_0(t+h) - \mathbb{P}_0(t)}{h} = -\lambda \mathbb{P}_0(t) + o(h).$$

Haciendo $h \rightarrow 0$, se obtiene que

$$\mathbb{P}_0'(t) = -\lambda \mathbb{P}_0(t),$$

y por lo tanto

$$\frac{\mathbb{P}_0'(t)}{\mathbb{P}_0(t)} = -\lambda.$$

¹No hay ocurrencia de eventos de 0 a t ni de t a $t+h$.

²Por incrementos independientes.

³Por incrementos estacionarios $\mathbb{P}[N(h+t) - N(h) = 0] = \mathbb{P}[N(h) = 0]$.

⁴Por definición, es decir. $\mathbb{P}[N(h) = 0] = 1 - \mathbb{P}[N(h) = 1] - \mathbb{P}[N(h) \geq 2]$.

Integrando en ambos lados de la igualdad, se tienen que

$$\int_0^t \frac{\mathbb{P}'_0(s)}{\mathbb{P}_0(s)} ds = - \int_0^t \lambda ds .$$

De esta forma,

$$\log \mathbb{P}_0(t) = -\lambda t + c \quad \text{o} \quad \mathbb{P}_0(t) = Ke^{-\lambda t} .$$

Note que $\mathbb{P}_0(t) = \mathbb{P}_0[N(0) = 0] = 1$ así se tiene que

$$1 = \mathbb{P}_0(0) = Ke^{-\lambda \cdot 0}$$

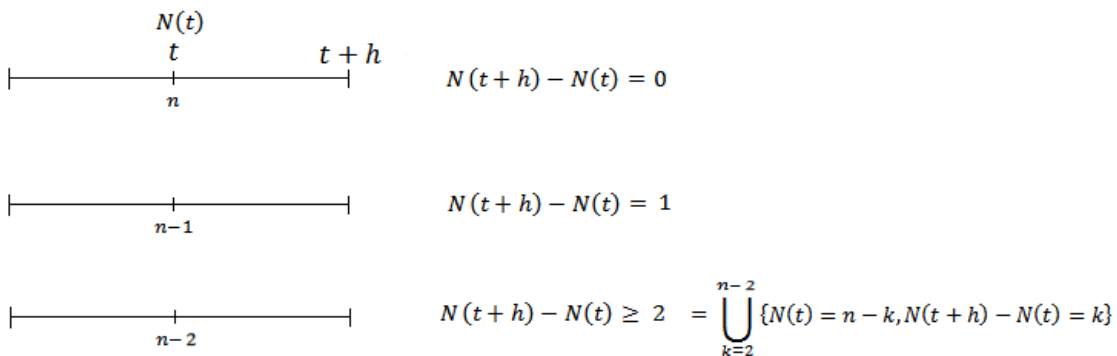
y por tanto $K = 1$. De esta forma, se obtiene que

$$\mathbb{P}_0(t) = e^{-\lambda t}. \tag{1.3.1}$$

Similarmente para $n \geq 1$, se tiene que

$$\begin{aligned} \mathbb{P}_n(t+h) &= \mathbb{P}[N(t+h) = n] \\ &= \mathbb{P}[N(t) = n, N(t+h) - N(t) = 0] \\ &\quad + \mathbb{P}[N(t) = n-1, N(t+h) - N(t) = 1] \\ &\quad + \mathbb{P}[N(t+h) = n, N(t+h) - N(t) \geq 2]. \end{aligned}$$

Ejemplificando gráficamente



De esta forma, se puede escribir

$$\mathbb{P}_n(t+h) = \mathbb{P}_n(t)\mathbb{P}_0(h) + \mathbb{P}_{n-1}(t)\mathbb{P}_1(h) + o(h),$$

y por los incisos *ii*), *iii*) y *iv*) de la Definición 6, se tiene

$$\begin{aligned} \mathbb{P}_n(t+h) &= \mathbb{P}_n(t)\mathbb{P}_0(h) + \mathbb{P}_{n-1}(t)\mathbb{P}_1(h) + o(h) \\ &= \mathbb{P}_n(t)(1 - \lambda h) + \mathbb{P}_{n-1}(t)\lambda h + o(h), \end{aligned}$$

y por lo tanto

$$\frac{\mathbb{P}_n(t+h) - \mathbb{P}_n(t)}{h} = -\lambda \mathbb{P}_n(t) + \lambda \mathbb{P}_{n-1}(t) + o(h).$$

Haciendo $h \rightarrow 0$ se obtiene que

$$\mathbb{P}_n'(t) = -\lambda \mathbb{P}_n(t) + \lambda \mathbb{P}_{n-1}(t).$$

Reacomodando los términos de la ecuación y multiplicando por el factor de integración $e^{\lambda t}$ se obtiene que

$$e^{\lambda t} [\mathbb{P}_n'(t) + \lambda \mathbb{P}_n(t)] = \lambda e^{\lambda t} \mathbb{P}_{n-1}(t).$$

Por lo tanto

$$\frac{d}{dt} [e^{\lambda t} \mathbb{P}_n(t)] = \lambda e^{\lambda t} \mathbb{P}_{n-1}(t). \quad (1.3.2)$$

Se usará inducción matemática en n para demostrar que *iii*) de la Definición 4 es verdadera.

Primero se verificará que vale para $n = 1$. Integrando (1.3.2) con respecto a t se tiene

$$\int \frac{d}{dt} [e^{\lambda t} \mathbb{P}_1(t)] dt = \int \lambda e^{\lambda t} \mathbb{P}_0(t) dt.$$

Sustituimos el valor de $\mathbb{P}_0(t)$, dado por (1.3.1) se tiene que

$$e^{\lambda t} \mathbb{P}_1(t) + c_1 = \int \lambda e^{\lambda t} e^{-\lambda t} dt = \int \lambda dt = \lambda t + c_2.$$

Por lo tanto,

$$\mathbb{P}_1(t) = e^{-\lambda t} (\lambda t + c).$$

Note que vale $\mathbb{P}_1(0) = 0$ (por el inciso *i*) de la Definición 6).

De esta forma, se tiene que

$$\mathbb{P}_1(0) = e^{-\lambda * 0} (\lambda * 0 + c) = 0$$

de donde se obtiene que $c = 0$. Entonces, $\mathbb{P}_1(t) = e^{-\lambda t} \lambda t$.

Demostraremos que se cumple para $\mathbb{P}_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$. Por inducción matemática suponga que vale para $n - 1$.

Sustituimos $\mathbb{P}_{n-1}(t) = e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}$ en (1.3.2) se obtiene que

$$\frac{d}{dt} [e^{\lambda t} \mathbb{P}_n(t)] = \lambda e^{\lambda t} \mathbb{P}_{n-1}(t) = \lambda e^{\lambda t} e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} = \frac{(\lambda t)^{n-1}}{(n-1)!}. \quad (1.3.3)$$

Integrando (1.3.3) con respecto a t se tiene que

$$\int \frac{d}{dt} [e^{\lambda t} \mathbb{P}_n(t)] dt = \int \lambda \frac{t^{n-1}}{(n-1)!} dt.$$

Por consiguiente

$$e^{\lambda t} \mathbb{P}_n(t) + c_1 = \lambda^n \int \frac{t^{n-1}}{(n-1)!} dt + c = \lambda^n \frac{t^n}{n!} + c = \frac{(\lambda t)^n}{n!} + c.$$

Dado que $\mathbb{P}_n(0) = 0$, para $n \neq 0$ se tiene que

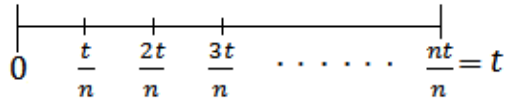
$$0 = e^{\lambda t} \mathbb{P}_n(0) = \frac{(\lambda * 0)^n}{n!} + c,$$

De esta manera $c = 0$ y el resultado sigue (Ross, 1996).

Se demostrará ahora que la Definición 4 implica la Definición 6.

Por la Definición 4 se tiene que el número de eventos que ocurren en un intervalo de longitud t es una variable aleatoria con distribución Poisson (λ).

Sea $[0, t]$ el intervalo considerado y $N(t)$ el número de eventos que ocurren en ese intervalo. Se divide el intervalo $[0, t]$ en n subintervalos disjuntos, cada uno de longitud $\frac{t}{n}$.



Tome una $k \leq n$, y considere la siguiente partición:

A: k subintervalos que contienen exactamente 1 evento y los otros $n - k$ no contienen ninguno.

B: Al menos un subintervalo contiene 2 o más eventos.

Entonces

$$\mathbb{P}[N(t) = k] = \mathbb{P}[(N(t) = k) \cap A] + \mathbb{P}[(N(t) = k) \cap B]. \quad (1.3.4)$$

Primero se obtendrá una expresión para $\mathbb{P}[(N(t) = k) \cap B]$. Sea C_i el suceso “el i –ésimo subintervalo contiene 2 o más eventos”, entonces.

$$\mathbb{P}[(N(t) = k) \cap B] \leq \mathbb{P}(B) \leq \mathbb{P}\left(\bigcup_{i=1}^n C_i\right). \quad (1.3.5)$$

Sin embargo, note que

$$\mathbb{P}\left(\bigcup_{i=1}^n C_i\right) \leq \sum_{i=1}^n \mathbb{P}(C_i) = \sum_{i=0}^n o\left(\frac{t}{n}\right) = n o\left(\frac{t}{n}\right) = \frac{t o\left(\frac{t}{n}\right)}{\frac{t}{n}}.$$

Cuando $n \rightarrow \infty$ se tiene que $\frac{t}{n} \rightarrow 0$ entonces $\frac{o\left(\frac{t}{n}\right)}{\frac{t}{n}} \rightarrow 0$. Por lo tanto

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{i=1}^n C_i\right) = 0,$$

y por (1.3.5) se tiene que $\mathbb{P}[(N(t) = k) \cap B] = 0$. Por otro lado $\mathbb{P}[(N(t) = k) \cap A]$ tiene una distribución Binomial con n subintervalos, t éxitos y con una probabilidad de éxito de $p = \left[\frac{\lambda t}{n} + o\left(\frac{t}{n}\right)\right]$. De esta forma,

$$\begin{aligned} \mathbb{P}[(N(t) = k) \cap A] &= \mathbb{P}(A) = \binom{n}{k} \left[\frac{\lambda t}{n} + o\left(\frac{t}{n}\right)\right]^k \left[1 - \frac{\lambda t}{n} - o\left(\frac{t}{n}\right)\right]^{n-k} \\ &= \underbrace{\frac{n(n-1) \dots (n-k+1)}{n^k k!}}_I \left[\underbrace{\lambda t + t \frac{o\left(\frac{t}{n}\right)}{\frac{t}{n}}}_II \right]^k \\ &\quad \underbrace{\left[1 - \frac{\lambda t}{n} - o\left(\frac{t}{n}\right)\right]^n}_{III} \underbrace{\left[1 - \frac{\lambda t}{n} - o\left(\frac{t}{n}\right)\right]^{-k}}_{IV}. \end{aligned}$$

Cuando $n \rightarrow \infty$ en la ecuación (I) se tiene que

$$\begin{aligned} &\lim_{n \rightarrow \infty} \frac{n(n-1) \dots (n-k+1)}{n^k k!} \\ &= \lim_{n \rightarrow \infty} \frac{n(n-1) \dots (n-(k-1))}{n n n \dots n k!} \\ &= \lim_{n \rightarrow \infty} \frac{1}{k!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) = \frac{1}{k!}. \end{aligned}$$

Para la ecuación (II) se obtiene

$$\lim_{n \rightarrow \infty} \left[\lambda t + t \frac{o\left(\frac{t}{n}\right)}{\frac{t}{n}} \right]^k = (\lambda t)^k .$$

Para la ecuación (III) se tiene que

$$\begin{aligned} \lim_{n \rightarrow \infty} \left\{ \left[1 - \frac{\lambda t}{n} - o\left(\frac{t}{n}\right) \right]^{\frac{n}{t}} \right\}^t &= \lim_{n \rightarrow \infty} \left\{ \left[1 - \frac{\lambda}{\frac{t}{n}} - \frac{t}{n} \frac{o\left(\frac{t}{n}\right)}{\frac{t}{n}} \right]^{\frac{n}{t}} \right\}^t = \lim_{n \rightarrow \infty} \left\{ \left[1 - \frac{\lambda}{\frac{t}{n}} - \frac{t}{n} \frac{o\left(\frac{t}{n}\right)}{\frac{t}{n}} \right]^{\frac{n}{t}} \right\}^t \\ &= e^{-\lambda t} . \end{aligned}$$

En la ecuación (IV) se tiene que

$$\lim_{n \rightarrow \infty} \left[1 - \frac{\lambda t}{n} - o\left(\frac{t}{n}\right) \right]^{-k} = 1 .$$

De esta manera, sustituyendo los valores obtenidos anteriormente en (I), (II), (III) y (IV) en la ecuación (1.3.4) se tiene que

$$\mathbb{P}[N(t) = k] = \frac{1}{k!} (\lambda t)^k e^{-\lambda t} 1 + 0 = \frac{(\lambda t)^k}{k!} e^{-\lambda t} .$$

Por consiguiente se concluye que las Definiciones 4 y 6 son equivalentes.

1.4 Proceso Poisson No-Homogéneo

Definición 7: Un proceso de conteo $N = \{N(t), t \geq 0\}$ es un **proceso de Poisson no-homogéneo** con tasa o función de intensidad $\lambda(t) \geq 0$ donde $t \geq 0$ si

- i) $N(0) = 0$.
 - ii) $\{N(t), t \geq 0\}$ tiene incrementos independientes.
 - iii) $\mathbb{P}[N(t+h) - N(t) \geq 2] = o(h)$.
 - iv) $\mathbb{P}[N(t+h) - N(t) = 1] = \lambda(t)h + o(h)$.
- (Ross, 1996).

Teorema 2: Si denotamos $m(t) = \int_0^t \lambda(s) ds$ se tiene que

$$\mathbb{P}[N(t+s) - N(t) = n] = e^{-[m(t+s)-m(t)]} \frac{[m(t+s)-m(t)]^n}{n!}, \quad n \geq 0,$$

es decir, $N(t+s) - N(t)$, tiene una distribución de Poisson con parámetro $m(t+s) - m(t)$. A $m(t)$ se le llama función de valor medio de N .

Demostración: Se utilizará inducción matemática. Defina $\mathbb{P}_n(s) = \mathbb{P}[N(t+s) - N(t) = n]$.

Tome $n = 0$, entonces,

$$\begin{aligned}\mathbb{P}_0(s+h) &= \mathbb{P}[N(t+s+h) - N(t) = 0] \\ &= \mathbb{P}[0 \text{ eventos en } (t, t+s), 0 \text{ eventos en } (t+s, t+s+h)] \\ &= \mathbb{P}[0 \text{ eventos en } (t, t+s)]\mathbb{P}[0 \text{ eventos en } (t+s, t+s+h)] \\ &= \mathbb{P}_0(s)[1 - \lambda(t+s)h + o(h)].\end{aligned}$$

De esta forma

$$\mathbb{P}_0(s+h) = \mathbb{P}_0(s) - \lambda(t+s)h\mathbb{P}_0(s) + o(h),$$

y por lo tanto

$$\mathbb{P}_0(s+h) - \mathbb{P}_0(s) = -\lambda(t+s)h\mathbb{P}_0(s) + o(h),$$

o de igual forma

$$\frac{\mathbb{P}_0(s+h) - \mathbb{P}_0(s)}{h} = -\lambda(t+s)\mathbb{P}_0(s) + o(h). \quad (1.4.1)$$

Tomando el límite en (1.4.1) cuando $h \rightarrow 0$ se tiene que

$$\mathbb{P}_0'(s) = -\lambda(t+s)\mathbb{P}_0(s),$$

o de forma equivalente

$$\frac{\mathbb{P}_0'(s)}{\mathbb{P}_0(s)} = -\lambda(t+s). \quad (1.4.2)$$

Integrando los dos lados de la igualdad (1.4.2) se tiene que

$$\int_0^s \frac{\mathbb{P}_0'(u)}{\mathbb{P}_0(u)} du = -\int_0^s \lambda(t+u) du,$$

o igualmente, que

$$\log \mathbb{P}_0(s) = -\int_0^s \lambda(t+u) du.$$

Por lo tanto,

$$\mathbb{P}_0(s) = e^{-\int_0^s \lambda(t+u) du} = e^{-\int_t^{t+s} \lambda(u) du} = e^{-[\int_0^{t+s} \lambda(u) du - \int_0^t \lambda(u) du]}.$$

De esta forma, se tiene que

$$\mathbb{P}_0(s) = e^{-[m(t+s) - m(t)]}.$$

Por lo tanto se cumple que,

$$\begin{aligned}\mathbb{P}_0(s) &= \mathbb{P}[N(t+h) - N(t) = 0] = e^{-[m(t+s)-m(t)]} \frac{[m(t+s) - m(t)]^0}{0!} \\ &= e^{-[m(t+s)-m(t)]}.\end{aligned}$$

Ahora se demostrará para $n \geq 1$, es decir, se demostrará que

$$\mathbb{P}_n(s) = \mathbb{P}[N(t+s) - N(t) = n] = e^{-[m(t+s)-m(t)]} \frac{[m(t+s) - m(t)]^n}{n!}.$$

De esta forma

$$\begin{aligned}\mathbb{P}_n(s+h) &= \mathbb{P}[N(t+s+h) - N(t) = n] \\ &= \mathbb{P}[n \text{ eventos en } (t, t+s), 0 \text{ eventos en } (t+s, t+s+h)] \\ &\quad + \mathbb{P}[1 \text{ evento en } (t, t+s), n-1 \text{ eventos en } (t+s, t+s+h)] \\ &\quad + \sum_{k=2}^{n-2} \mathbb{P}[n-k \text{ eventos en } (t, t+s), k \text{ eventos en } (t+s, t+s+h)].\end{aligned}$$

Usando la propiedad de incrementos independientes se tiene que

$$\begin{aligned}\mathbb{P}_n(s+h) &= \mathbb{P}[N(t+s) - N(t) = n] \mathbb{P}[N(t+s+h) - N(t+s) = 0] \\ &\quad + \mathbb{P}[N(t+s) - N(t) = n-1] \mathbb{P}[N(t+s+h) - N(t+s) = 1] \\ &\quad + \sum_{k=2}^{n-2} \mathbb{P}[N(t+s) - N(t) = n-k] \mathbb{P}[N(t+s+h) - N(t+s) = k].\end{aligned}$$

Para $k \geq 2$, $\mathbb{P}[N(t+s) - N(t) = n-k] \mathbb{P}[N(t+s+h) - N(t+s) = k] = o(h)$ entonces

$$\begin{aligned}\mathbb{P}_n(s+h) &= \mathbb{P}[N(t+s) - N(t) = n] \mathbb{P}[N(t+s+h) - N(t+s) = 0] \\ &\quad + \mathbb{P}[N(t+s) - N(t) = n-1] \mathbb{P}[N(t+s+h) - N(t+s) = 1] + o(h).\end{aligned}$$

Por la Definición 7 se tiene que

$$\begin{aligned}\mathbb{P}_n(s+h) &= \mathbb{P}_n(s)[1 - \lambda(t+s)h] + \mathbb{P}_{n-1}(s)[\lambda(t+s)h] + o(h) \\ &= \mathbb{P}_n(s) - \mathbb{P}_n(s)\lambda(t+s)h + \mathbb{P}_{n-1}(s)\lambda(t+s)h + o(h)\end{aligned}$$

y por lo tanto,

$$\frac{\mathbb{P}_n(s+h) - \mathbb{P}_n(s)}{h} = -\lambda(t+s)\mathbb{P}_n(s) + \lambda(t+s)\mathbb{P}_{n-1}(s) + o(h). \quad (1.4.3)$$

Tomando el límite en (1.4.3) cuando $h \rightarrow 0$ se tiene que

$$\mathbb{P}_n'(s) = -\lambda(t+s)\mathbb{P}_n(s) + \lambda(t+s)\mathbb{P}_{n-1}(s)$$

o de igual forma, que

$$\mathbb{P}_n'(s) + \lambda(t+s)\mathbb{P}_n(s) = \mathbb{P}_{n-1}(s)\lambda(t+s).$$

Multiplicamos ambos lados de la igualdad por el factor de integración $e^{-\int_0^s \lambda(t+w)dw}$ se obtiene que

$$e^{\int_0^s \lambda(t+w)dw} [\mathbb{P}_n'(s) + \lambda(t+s)\mathbb{P}_n(s)] = e^{\int_0^s \lambda(t+w)dw} [\mathbb{P}_{n-1}(s)\lambda(t+s)]. \quad (1.4.4)$$

Note que

$$\frac{d}{ds} \left[e^{\int_0^s \lambda(t+w)dw} \mathbb{P}_n(s) \right] = e^{\int_0^s \lambda(t+w)dw} [\mathbb{P}_n'(s) + \lambda(t+s)\mathbb{P}_n(s)].$$

Esto se debe a lo siguiente:

$$\begin{aligned} \frac{d}{ds} \left[e^{\int_0^s \lambda(t+w)dw} \mathbb{P}_n(s) \right] &= e^{\int_0^s \lambda(t+w)dw} \mathbb{P}_n'(s) + \mathbb{P}_n(s) \frac{d}{ds} \left[e^{\int_0^s \lambda(t+w)dw} \right] \\ &= e^{\int_0^s \lambda(t+w)dw} \mathbb{P}_n'(s) + \mathbb{P}_n(s) e^{\int_0^s \lambda(t+w)dw} \frac{d}{ds} \left[\int_0^s \lambda(t+w)dw \right] \\ &= e^{\int_0^s \lambda(t+w)dw} \mathbb{P}_n'(s) + \mathbb{P}_n(s) e^{\int_0^s \lambda(t+w)dw} \lambda(t+s) \\ &= e^{\int_0^s \lambda(t+w)dw} [\mathbb{P}_n'(s) + \lambda(t+s)\mathbb{P}_n(s)]. \end{aligned}$$

Por lo tanto (1.4.4) equivale a

$$\frac{d}{ds} \left[e^{\int_0^s \lambda(t+w)dw} \mathbb{P}_n(s) \right] = e^{\int_0^s \lambda(t+w)dw} [\mathbb{P}_{n-1}(s)\lambda(t+s)],$$

Integrando en ambos lados de la desigualdad se tiene que

$$\int_0^s \frac{d}{dv} \left[e^{\int_0^v \lambda(t+w)dw} \mathbb{P}_n(v) \right] dv = \int_0^s e^{\int_0^v \lambda(t+w)dw} [\mathbb{P}_{n-1}(v)\lambda(t+v)] dv$$

y por lo tanto

$$e^{\int_0^s \lambda(t+w)dw} \mathbb{P}_n(s) = \int_0^s e^{\int_0^v \lambda(t+w)dw} [\mathbb{P}_{n-1}(v)\lambda(t+v)] dv. \quad (1.4.5)$$

Suponga que el teorema vale para $n-1$, es decir,

$$\begin{aligned} \mathbb{P}_{n-1}(s) &= \mathbb{P}[N(t+h) - N(t) = n-1] \\ &= e^{-[m(t+s)-m(t)]} \frac{[m(t+s)-m(t)]^{n-1}}{(n-1)!}. \end{aligned} \quad (1.4.6)$$

Sustituyendo (1.4.6) en (1.4.5) y se tiene que

$$e^{\int_0^s \lambda(t+w)dw} \mathbb{P}_n(s) = \int_0^s e^{\int_0^v \lambda(t+w)dw} \lambda(t+v) e^{-[m(t+s)-m(t)]} \frac{[m(t+s)-m(t)]^{n-1}}{(n-1)!} dv. \quad (1.4.7)$$

Sabemos que $e^{\int_0^s \lambda(t+w)dw} = e^{[m(t+s)-m(t)]}$. De esta forma (1.4.7) equivalente a

$$\begin{aligned} e^{[m(t+s)-m(t)]} \mathbb{P}_n(s) &= \int_0^s e^{[m(t+s)-m(t)]} \lambda(t+v) e^{-[m(t+s)-m(t)]} \frac{[m(t+s)-m(t)]^{n-1}}{(n-1)!} dv \\ &= \int_0^s \lambda(t+v) \frac{[m(t+s)-m(t)]^{n-1}}{(n-1)!} dv \\ &= \frac{1}{(n-1)!} \int_0^s \lambda(t+v) [m(t+s)-m(t)]^{n-1} dv. \end{aligned}$$

Sean $u = m(t+s) - m(t)$ y $dv = \lambda(t+v)dv$. Note que

- i) $\frac{d}{dv} [m(t+v) - m(t)] = \lambda(t+v)$
- ii) Cuando $v = 0$ se tiene que $u = m(t+v) - m(t) = 0$ y cuando $v = s$, entonces $u = m(t+s) - m(t)$.

Por lo tanto, se puede escribir

$$\begin{aligned} e^{[m(t+s)-m(t)]} \mathbb{P}_n(s) &= \frac{1}{(n-1)!} \int_0^{m(t+s)-m(t)} u^{n-1} du \\ e^{[m(t+s)-m(t)]} \mathbb{P}_n(s) &= \frac{1}{(n-1)!} \frac{u^n}{n} \Big|_0^{m(t+s)-m(t)} \\ &= \frac{u^n}{n!} \Big|_0^{m(t+s)-m(t)} \\ &= \frac{[m(t+s)-m(t)]^n}{n!}, \end{aligned}$$

y por lo tanto

$$\mathbb{P}_n(s) = e^{-[m(t+s)-m(t)]} \frac{[m(t+s)-m(t)]^n}{n!},$$

y el resultado se sigue.

Capítulo 2

Introducción a la Estadística Bayesiana

En este capítulo se describirán algunos conceptos básicos de la estadística Bayesiana, como los son: la función de verosimilitud y las distribuciones a priori y a posteriori. También se presentará la forma de la función de verosimilitud en el caso de modelos de Poisson tanto en el caso homogéneo como en el no-homogéneo. Lo presentado aquí se tomó de Reyes (2008), López (2005) y varios otros trabajos que se mencionan más adelante.

Antes de presentar las definiciones de interés, damos algunos conceptos preliminares (Reyes, 2008).

Definición 8: Considere un experimento que puede ser descrito mediante un modelo matemático (estocástico o no). La variable que determina dicho modelo se llama **parámetro** y se denotará por θ . Un experimento es llamado aleatorio, si se puede especificar los posibles resultados del mismo pero no se puede decir de antemano cual será el resultado en cada realización del experimento.

Definición 9: El conjunto de posibles valores que el parámetro de un modelo puede tomar se llama **espacio parametral** y se denotará por Θ . En el espacio parametral se especifica una familia de funciones de densidad, es decir, se especifica

$$F = \{f(x, \theta); \theta \in \Theta\},$$

donde $f(x, \theta)$ es una función de densidad con parámetro $\theta \in \Theta$ y x es un elemento del espacio donde f toma valores.

En este trabajo se considerara que θ es una variable aleatoria a la que se le puede asignar una función de probabilidad.

Definición 10: Suponga que se efectúan N repeticiones de un experimento aleatorio. Al conjunto de valores $x = \{x_1, x_2, \dots, x_N\}$ obtenidos a partir de las variables aleatorias idénticamente distribuidas X_1, X_2, \dots, X_N , donde X_i , $i = 1, 2, \dots, N$ registra el resultado de la i – ésima repetición del experimento aleatorio, se le llama **muestra aleatoria**.

Se supondrá que los experimentos son realizados independientemente, es decir, la muestra aleatoria mencionada estará compuesta de valores obtenidos independientemente.

Ahora podemos enunciar las definiciones básicas de la estadística Bayesiana.

2.1 Función de verosimilitud

Definición 11: Sea $x = (x_1, x_2, \dots, x_N)$ una muestra aleatoria y $\theta \in \Theta$ el parámetro desconocido del modelo que produjo x . La función que describe el modelo que produjo x vista como función de θ es llamada **función de verosimilitud** que se denotará por $\mathcal{L}(x|\theta)$. En el caso de que n es tomada independientemente e idénticamente distribuidas $f(\cdot|\theta)$ se tiene que la función de verosimilitud es dada por

$$\mathcal{L}(x|\theta) \propto \prod_{i=1}^N f(x_i|\theta).$$

Observaciones:

- a) La función de verosimilitud resume la información del modelo contenida en los datos y toda inferencia acerca del parámetro debe hacerse tomando en cuenta dicha función.
- b) Lo primordial de la función de verosimilitud establece que si para dos experimentos distintos se obtienen funciones de verosimilitud proporcionales, entonces las inferencias obtenidas de ella son las mismas.

2.2 Distribución a priori

En este apartado se presentará la definición de la distribución a priori y se planteará la necesidad de considerar estas distribuciones mencionadas al realizar inferencias.

Definición 12: Sea $x = (x_1, x_2, \dots, x_N)$ una muestra aleatoria obtenida de la realización de un experimento aleatorio, $\theta \in \Theta$ el parámetro desconocido del modelo que describe el experimento que produjo la muestra x . La **distribución a priori** de θ es la distribución del parámetro θ que se le asigna sin tener en cuenta la muestra obtenida del experimento y se denotará por $P_{priori}(\theta)$.

Nota: Las distribuciones a priori son especificadas por intuición, por medio de estudios previos o de opiniones de expertos en el área, pero si no se tiene intuición o información alguna generalmente se usa una distribución aleatoria uniforme sobre un espacio muestral suficientemente grande.

Se ejemplificará la necesidad de tomar en cuenta la información previa en el momento de hacer inferencias, considerar lo siguiente (Reyes, 2008).

- 1) Un músico afirma que puede saber si una página de una obra musical pertenece a una obra de Hayden o una de Mozart. Imagine que el músico clasificó correctamente diez páginas.
- 2) Una mujer ebria afirma que es capaz de saber, con un solo trago, la manera en que un café con leche fue preparado: si la leche fue agregada antes o después del café. Imagine que la mujer acertó en diez tazas de dicha bebida.

Tomar en cuenta las siguientes variables aleatorias

X = Número de páginas seleccionadas correctamente por el músico.

Y = Número de aciertos de la mujer.

Por lo que se tiene $\mathbb{P}(X|10, \theta)$ y $\mathbb{P}(Y|10, \theta)$ son Binomial $(10, \theta)$ donde θ es la probabilidad de éxito en cada experimento. Como la mujer y el músico acertaron las diez veces que intentaron sus experimentos se infiere que, $\theta > \frac{1}{2}$, para los dos casos, aunque el grado de creencia para la mujer ebria es muy bajo y para el músico es medianamente alto. Así, tomando en cuenta la información personal y los datos observados a la hora de inferir se evitarán tales subjetividades.

En algunas ocasiones será difícil encontrar una distribución a priori para el parámetro desconocido basándose solamente en la muestra, ya sea porque la información sobre él no es muy confiable o porque no es suficiente. En otras ocasiones se deberá hacer inferencia sólo tomando en cuenta la muestra, para no interferir en la información proporcionada por los datos. En estos casos es necesario encontrar una distribución a priori que aporte poca información o no favorezca a ningún valor.

Definición 13: Sea $x = (x_1, x_2, \dots, x_N)$ una muestra aleatoria y $\theta \in \Theta$ el parámetro del modelo que produjo la muestra x . Se dice que una distribución a priori es **no informativa** para θ si no favorece a algún valor de θ sobre otros.

Observación: En general, las distribuciones no informativas son impropias, es decir,

$$\sum_{\theta \in \Theta} P_{\text{priori}}(\theta) = \infty,$$

cuando Θ es discreto, o

$$\int_{\Theta} P_{\text{priori}}(\theta) = \infty,$$

cuando Θ es continuo. Sin embargo, muchas veces son tomadas distribuciones uniformes definidas en un conjunto suficientemente grande.

2.3 Distribución a posteriori

En esta sección se definirán, entre otros conceptos, la distribución a posteriori.

Definición 14: Sea $x = (x_1, x_2, \dots, x_N)$ una muestra aleatoria y $\theta \in \Theta$ el parámetro del modelo que produjo la muestra X , $P_{priori}(\theta)$ la función a priori de θ y $\mathcal{L}(x|\theta)$ la función de verosimilitud. La **distribución a posteriori** es la distribución del parámetro θ cuando se toma en cuenta la información producida por la muestra obtenida, es decir, es $\mathbb{P}(\theta|x)$.

Observación: A veces las distribuciones a priori dependen de otros parámetros que pueden ser conocidos o desconocidos y que son llamados hiperparámetros. Cuando son desconocidos pueden tratarse de dos formas:

- 1) El método propiamente Bayesiano consiste en asignarle al hiperparámetro una distribución a priori (también llamado hiperpriori) y a partir de ella encontrar la distribución a posteriori. Por otro lado, la distribución hiperpriori podría depender de una colección de parámetros desconocidos, a la especificación de un modelo sobre varios niveles se le llama modelo jerárquico, en donde cada nueva distribución forma un nivel de la jerarquía.
- 2) El segundo método llamado análisis empírico de Bayes consiste en encontrar un valor η que maximice la distribución de $\mathbb{P}(\eta|x)$ pero visto como función de η . Si $\eta = \hat{\eta}$ fuera dicho valor entonces la distribución a posteriori que se trabajará será $\mathbb{P}(\theta|x, \hat{\eta})$. La expresión para $\mathbb{P}(\theta|x, \eta)$ puede obtenerse sustituyendo $\hat{\eta}$ en la expresión (1) dada a continuación.

Teorema 3: Sean $x = (x_1, x_2, \dots, x_N)$ una muestra de un experimento aleatorio, $\theta \in \Theta$ el parámetro del modelo que produjo la muestra x y $P_{priori}(\theta|\eta)$ la distribución a priori que depende de un hiperparámetro η .

- a) Si η es conocido la distribución a posteriori es

$$\mathbb{P}(\theta|x, \eta) = \frac{\mathbb{P}(x|\theta)P_{priori}(\theta|\eta)}{\sum_{\gamma \in \Theta} \mathbb{P}(x|\gamma)P_{priori}(\gamma|\eta)}. \quad (1)$$

- b) Si η es desconocido la distribución a posteriori es

$$\mathbb{P}(\theta|x) = \frac{\sum_{\gamma \in \Theta} \mathbb{P}(x|\theta)P_{priori}(\theta|\gamma)P_{priori}(\gamma)}{\sum_{\eta} \sum_{\gamma \in \Theta} \mathbb{P}(x|\gamma)P_{priori}(\gamma|\eta)P_{priori}(\eta)} \quad (2)$$

Demostración

a) Note que

$$\begin{aligned}
 \mathbb{P}(\theta|x, \eta) &\stackrel{1}{=} \frac{\mathbb{P}(\theta, x, \eta)}{\mathbb{P}(x, \eta)} \\
 &\stackrel{2}{=} \frac{\mathbb{P}(x|\theta, \eta)\mathbb{P}(\theta, \eta)}{\mathbb{P}(x, \eta)} \frac{\mathbb{P}(\eta)}{\mathbb{P}(\eta)} \\
 &\stackrel{3}{=} \frac{\mathbb{P}(x|\theta, \eta)\mathbb{P}(\theta|\eta)}{\mathbb{P}(x|\eta)} \\
 &\stackrel{4}{=} \frac{\mathbb{P}(x|\theta, \eta)\mathbb{P}(\theta|\eta)}{\sum_{\gamma \in \Theta} \mathbb{P}(x, \gamma|\eta)} \\
 &= \frac{\mathbb{P}(x|\theta)\mathbb{P}(\theta|\eta)}{\sum_{\gamma \in \Theta} \mathbb{P}(x|\gamma, \eta)\mathbb{P}(\gamma|\eta)} \\
 &\stackrel{5}{=} \frac{\mathbb{P}(x|\theta)P_{\text{priori}}(\theta|\eta)}{\sum_{\gamma \in \Theta} \mathbb{P}(x|\gamma)P_{\text{priori}}(\gamma|\eta)}
 \end{aligned}$$

b) Si η es desconocido, también se le asignará una distribución a priori $P_{\text{priori}}(\eta)$.

$$\begin{aligned}
 \mathbb{P}(\theta|x) &\stackrel{6}{=} \frac{\mathbb{P}(\theta, x)}{\mathbb{P}(x)} \\
 &\stackrel{7}{=} \frac{\sum_{\eta} \mathbb{P}(\theta, x, \eta)}{\sum_{\gamma \in \Theta} \mathbb{P}(x, \gamma)} \\
 &\stackrel{8}{=} \frac{\sum_{\eta} \mathbb{P}(x|\theta, \eta)\mathbb{P}(\theta, \eta)}{\sum_{\eta} \sum_{\gamma \in \Theta} \mathbb{P}(x, \gamma, \eta)} \\
 &\stackrel{9}{=} \frac{\sum_{\eta} \mathbb{P}(x|\theta, \eta)\mathbb{P}(\theta|\eta)\mathbb{P}(\eta)}{\sum_{\eta} \sum_{\gamma \in \Theta} \mathbb{P}(x|\gamma, \eta)\mathbb{P}(\gamma, \eta)} \\
 &\stackrel{10}{=} \frac{\sum_{\eta} \mathbb{P}(x|\theta)P_{\text{priori}}(\theta|\eta)P_{\text{priori}}(\eta)}{\sum_{\eta} \sum_{\gamma \in \Theta} \mathbb{P}(x|\gamma)P_{\text{priori}}(\gamma|\eta)P_{\text{priori}}(\eta)} .
 \end{aligned}$$

Demostración inciso a)

¹ Por definición de probabilidad condicional

² Propiedad $\mathbb{P}(\theta, x, \eta) = \mathbb{P}(x|\theta, \eta)\mathbb{P}(\theta, \eta)$

³ Por $\mathbb{P}(x|\eta) = \frac{\mathbb{P}(x, \eta)}{\mathbb{P}(\eta)}$, para $\mathbb{P}(\eta) > 0$

⁴ Por probabilidad total en divisor $\mathbb{P}(x|\eta) = \sum_{\gamma \in \Theta} \mathbb{P}(x, \gamma|\eta)$

⁵ Por definición $\mathbb{P}(\gamma|\eta) = P_{\text{priori}}(\gamma|\eta)$ en dividendo y divisor

inciso b)

⁶ Por definición de probabilidad condicional

⁷ Por probabilidad total $\mathbb{P}(\theta, x) = \sum_{\eta} \mathbb{P}(\theta, x, \eta)$ y en el divisor

$\mathbb{P}(x) = \sum_{\gamma \in \Theta} \mathbb{P}(x, \gamma)$

⁸ Por probabilidad condicional en el dividendo $\sum_{\eta} \mathbb{P}(\theta, x, \eta) = \sum_{\eta} \mathbb{P}(x|\theta, \eta)\mathbb{P}(\theta, \eta)$ y

por probabilidad total en el divisor $\sum_{\gamma \in \Theta} \mathbb{P}(x, \gamma) = \sum_{\eta} \sum_{\gamma \in \Theta} \mathbb{P}(x, \gamma, \eta)$

⁹ Probabilidad condicional en divisor y divisor

¹⁰ Por función a priori en el dividendo $\sum_{\eta} \mathbb{P}(x|\theta, \eta)\mathbb{P}(\theta|\eta) = \sum_{\eta} \mathbb{P}(x|\theta)P_{\text{priori}}(\theta|\eta)$ y

$\mathbb{P}(\eta) = P_{\text{priori}}(\eta)$ y en el divisor $\sum_{\gamma \in \Theta} \mathbb{P}(x|\gamma, \eta) = \sum_{\gamma \in \Theta} \mathbb{P}(x|\gamma)P_{\text{priori}}(\gamma|\eta)$ y

$\mathbb{P}(\gamma, \eta) = P_{\text{priori}}(\eta)$

Observación: Tomando a η como conocido no es necesario escribirla en la distribución a posteriori, y como el denominador en la expresión (1) ya no depende de θ entonces se puede simplificar y se obtiene

$$\mathbb{P}(\theta|x) \propto \mathcal{L}(x|\theta)P_{\text{priori}}(\theta), \quad (2.3.1)$$

donde $\mathcal{L}(x|\theta)$ es la función de verosimilitud del modelo y la constante de proporcionalidad, que sólo depende de x , hará que $\mathbb{P}(\theta|x)$ integre o sume uno.

2.4 Muestreo de Gibbs

Dada la función de densidad $f(x_1, x_2, \dots, x_K)$ una forma de obtener información al respecto, por ejemplo, de la media o de la varianza es realizar los cálculos correspondientes, ya sea analíticamente o numéricamente. Sin embargo eso puede ser complicado. Una forma de obtener una estimación sobre la esperanza, es obtener una muestra de tamaño N a partir de la distribución cuya densidad es f y utilizar la Ley de los Grandes Números. Otras cantidades de interés como varianza, moda, percentiles pueden ser estimadas, a través de una muestra obtenida a partir de la distribución con densidad f (López, 2005).

El muestreo de Gibbs o muestreador de Gibbs es un algoritmo iterativo el cual tiene una funcionalidad y estructura muy simple ya que simplifica dificultades técnicas con los cálculos de distribuciones multivariadas complejas. La idea principal consiste en utilizar distribuciones condicionales univariadas (más fáciles de simular que la distribución conjunta). Por lo tanto, el método transforma un problema de simulación de vectores aleatorios en un problema de simulación unidimensional.

Este método recibe su nombre del físico Willard Gibbs en referencia a sus trabajos en física estadística, aunque fue introducido formalmente por los hermanos Stuart y Donald Geman en 1984, en un artículo sobre el estudio de restauración de imágenes, ochenta años después de la muerte de Gibbs.

El muestreo de Gibbs (Geman y Geman, 1984; Gelfand y Smith, 1990; Casella y Edward, 1992; Stephens y Dellaportas, 1992), nos permite simular una cadena de Markov a partir de cualquier valor inicial $\theta^{(0)}$. Se genera valores sucesivos de esa cadena de Markov hasta que alcanza la estacionaridad, digamos en el paso m . Los siguientes valores generados $\{\theta^{(i)}; i > m\}$ pueden constituir una muestra.

Sea $\theta = (\theta_1, \theta_2, \dots, \theta_K)$ el vector paramétrico y sean $\pi(\theta_1|\theta_2, \dots, \theta_K)$, $\pi(\theta_2|\theta_1, \theta_3, \dots, \theta_K)$, ..., $\pi(\theta_K|\theta_1, \dots, \theta_{K-1})$, las distribuciones condicionales completas obtenidas a partir de la distribución $\pi(\theta)$. Partiendo de un punto inicial $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_K^{(0)})$ se genera $\theta_1^{(1)}$ a partir de $\pi(\theta_1|\theta_2^{(0)}, \dots, \theta_K^{(0)})$, $\theta_2^{(1)}$ a partir de $\pi(\theta_2|\theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_K^{(0)})$, y así hasta $\theta_K^{(1)}$, generado a partir de $\pi(\theta_K|\theta_1^{(1)}, \dots, \theta_{K-1}^{(1)})$, con lo que se obtiene $\theta^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_K^{(1)})$ y así sucesivamente. Bajo condiciones generales, la sucesión $\{\theta^{(i)}\}$ es una realización de una cadena de Markov con la distribución final $\pi(\theta)$ como distribución estacionaria (Gelfand y Smith, 1990).

A continuación se ejemplifica el muestreo de Gibbs. Sea $\theta^{(n)} = (\theta_1^{(n)}, \theta_2^{(n)}, \theta_3^{(n)}, \theta_4^{(n)}) = X_n$ y sea la distribución final $\pi(\theta|\text{Datos})$, queremos obtener el vector $\theta^{(n+1)} = (\theta_1^{(n+1)}, \theta_2^{(n+1)}, \dots, \theta_K^{(n+1)}) = X_{n+1}$ a través del muestreo de Gibbs. Se considerará el caso donde $K = 4$.

Se utiliza $\pi(\theta|\text{Datos})$ para muestrear el primer valor $\theta_1^{(n+1)}$ a partir de $\pi(\theta_1|\theta_2^{(n)}, \theta_3^{(n)}, \theta_4^{(n)})$. Por lo que se obtiene $(\theta_1^{(n+1)}, \theta_2^{(n)}, \theta_3^{(n)}, \theta_4^{(n)})$ se simula $\theta_2^{(n+1)}$ a partir de $\pi(\theta_2|\theta_1^{(n+1)}, \theta_3^{(n)}, \theta_4^{(n)})$. Se obtiene $(\theta_1^{(n+1)}, \theta_2^{(n+1)}, \theta_3^{(n)}, \theta_4^{(n)})$ se simula $\theta_3^{(n+1)}$ a partir de $\pi(\theta_3|\theta_1^{(n+1)}, \theta_2^{(n+1)}, \theta_4^{(n)})$. Se obtiene $(\theta_1^{(n+1)}, \theta_2^{(n+1)}, \theta_3^{(n+1)}, \theta_4^{(n)})$ se simula $\theta_4^{(n+1)}$ a partir $\pi(\theta_4|\theta_1^{(n+1)}, \theta_2^{(n+1)}, \theta_3^{(n+1)})$. Por lo tanto, $\theta^{(n+1)} = (\theta_1^{(n+1)}, \theta_2^{(n+1)}, \dots, \theta_K^{(n+1)}) = X_{n+1}$.

2.5 Construcción de la función de verosimilitud para un modelo de Poisson.

Sean d_1, d_2, \dots, d_K el tiempo de ocurrencia de un determinado evento. Suponga que X_1, X_2, \dots , los tiempos entre las ocurrencias, sean independientes y con distribución exponencial, entonces podemos describir el modelo que cuenta el número de eventos en un determinado intervalo tiempo por un modelo de Poisson.

Para la construcción de la función de verosimilitud nos basamos en la siguiente ilustración donde se observa la no ocurrencia de un evento al término del intervalo y de interés, es decir,

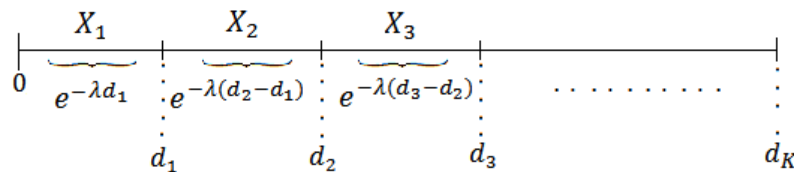


Figura 1

Note que por definición de proceso de Poisson se tiene

$$\mathbb{P}(X_i > t_i) = e^{-\lambda t_i}, \quad y \quad f(t_i) = -\frac{d}{dt_i} \mathbb{P}(X_i > t_i) = \lambda e^{-\lambda t_i}.$$

Caso Homogéneo

Sea $X = (X_1, X_2, \dots, X_K)$ una muestra aleatoria producida por variables aleatorias exponenciales X_i , $i = 1, 2, \dots, K$, con $0 < X_i < \infty$ y $0 < \lambda < \infty$, es decir,

$$\mathbb{P}(X_i > t_i | \lambda) = e^{-\lambda t_i}.$$

Tomando la función de densidad con la partición presentada en la *Figura 1*, se tiene que

$$-\frac{d}{dd_i} \mathbb{P}(X_i > d_i - d_{i-1} | X_{i-1} = d_{i-1} - d_{i-2}; \lambda) = \lambda e^{-\lambda(d_i - d_{i-1})}.$$

Para obtenerse la función de verosimilitud note que se tiene

$$\mathcal{L}(d_1, d_2, \dots, d_K | \lambda) \propto f(d_1) \prod_{i=2}^K f(d_i - d_{i-1} | d_{i-1} - d_{i-2}; \lambda). \quad (2.5.1)$$

Sustituyendo la función de densidad para los X_i $i = 1, 2, \dots, K$ y simplificando la expresión (2.5.1) se obtiene

$$\begin{aligned} \mathcal{L}(d_1, d_2, \dots, d_K | \lambda) &\propto \lambda e^{-\lambda d_1} \prod_{i=2}^K f(d_i - d_{i-1} | d_{i-1}; \lambda) \\ &= (\lambda e^{-\lambda d_1}) (\lambda e^{-\lambda d_2 + \lambda d_1}) \dots (\lambda e^{-\lambda d_K + \lambda d_{K-1}}) = \lambda^K e^{-\lambda d_K}. \end{aligned} \quad (2.5.2)$$

Caso No-Homogéneo

Se desarrolla de manera análoga que para el caso homogéneo pero tomaremos la definición del proceso Poisson no homogéneo. De esta forma se tiene que por definición vale

$$\mathbb{P}_n(s) = \mathbb{P}[N(t+h) - N(t) = n] = e^{-[m(t+s) - m(t)]} \frac{[m(t+s) - m(t)]^n}{n!}, \quad n = 0, 1, 2, \dots$$

Note que por definición se tiene que

$$\mathbb{P}(X_1 > d_1) = \mathbb{P}(N(d_1) = 0) = e^{-[m(d_1)]} \frac{[m(d_1)]^0}{0!} = e^{-[m(d_1)]}.$$

Por lo tanto la función de densidad es

$$f(d_1) = -\frac{d}{dd_1} \mathbb{P}(X_1 > d_1) = \left(\frac{d}{dd_1} m(d_1) \right) e^{-[m(d_1)]}.$$

Entonces,

$$f(d_1) = \lambda(d_1) e^{-[m(d_1)]}.$$

Para el caso de dos o más observaciones se tiene que

$$\begin{aligned} \mathbb{P}(X_2 > d_2 - d_1 | X_1 = d_1; \lambda) &= \mathbb{P}[N(d_2) - N(d_1) = 0] \\ &= e^{-[m(d_2) - m(d_1)]} \frac{[m(d_2) - m(d_1)]^0}{0!} \\ &= e^{-[m(d_2) - m(d_1)]}. \end{aligned}$$

Por lo tanto, la función de densidad es

$$f(d_2 - d_1 | d_1) = -\frac{d}{dd_2} \mathbb{P}(X_2 > d_2 - d_1 | X_1 = d_1)$$

$$\begin{aligned}
&= \left(\frac{d}{dd_2} [m(d_2) - m(d_1)] \right) e^{-[m(d_2) - m(d_1)]} \\
&= \lambda(d_2) e^{-[m(d_2) - m(d_1)]}.
\end{aligned}$$

Para el caso general se tiene

$$\begin{aligned}
f(d_i - d_{i-1} | d_{i-1} - d_{i-2}) &= -\frac{d}{dd_i} \mathbb{P}(X_i > d_i - d_{i-1} | X_{i-1} = d_{i-1}) \\
&= \lambda(d_i) e^{-[m(d_i) - m(d_{i-1})]}.
\end{aligned}$$

De esta forma la función de verosimilitud esta dada por

$$\begin{aligned}
\mathcal{L}(d_1, d_2, \dots, d_k | \lambda) &\propto f(d_1) \prod_{i=2}^K f(d_i - d_{i-1} | d_{i-1} - d_{i-2}; \lambda) \\
&= f(d_1) \prod_{i=2}^K (\lambda(d_i) e^{-[m(d_i) - m(d_{i-1})]}) \\
&= (\lambda(d_1) e^{-m(d_1)}) \left(\prod_{i=2}^K \lambda(d_i) \right) \left(e^{-\sum_{i=2}^K m(d_i) + \sum_{i=2}^K m(d_{i-1})} \right).
\end{aligned}$$

Note que vale lo siguiente,

$$e^{-m(d_1)} e^{-\sum_{i=2}^K m(d_i) + \sum_{i=2}^K m(d_{i-1})} = e^{-m(d_K)}.$$

Esto se debe a que podemos expresar $-\sum_{i=2}^K m(d_i)$ como $-\sum_{i=2}^{K-1} m(d_i) - m(d_K)$. Además de que podemos expresar $\sum_{i=2}^K m(d_{i-1})$ como $m(d_1) + \sum_{i=2}^{K-1} m(d_i)$. De esta forma,

$$\begin{aligned}
e^{-m(d_1)} e^{-\sum_{i=2}^K m(d_i) + \sum_{i=2}^K m(d_{i-1})} &= e^{-m(d_1)} e^{-\sum_{i=2}^{K-1} m(d_i) - m(d_K) + m(d_1) + \sum_{i=2}^{K-1} m(d_i)} \\
&= e^{-m(d_K)}.
\end{aligned}$$

Por lo tanto,

$$\mathcal{L}(d_1, d_2, \dots, d_K | \lambda) \propto \prod_{i=1}^K \lambda(d_i) e^{-m(d_K)}$$

Capítulo 3

Modelo Bayesiano

En este capítulo se presenta la formulación del modelo Bayesiano cuya función de verosimilitud sigue el modelo Poisson no homogéneo con función de intensidad Weibull exponenciada $\lambda(t)$, dada por:

$$\lambda(t) = \frac{\alpha\beta \left[1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}\right]^{\beta-1} \left(\frac{t}{\sigma}\right)^{\alpha-1} e^{-\left(\frac{t}{\sigma}\right)^\alpha}}{\sigma \left[1 - \left(1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}\right)^\beta\right]}. \quad 0 < t < \infty \quad (3.1)$$

donde $\alpha > 0$ y $\beta > 0$ parámetros de forma y $\sigma > 0$ parámetro de dispersión.

La descripción del modelo Bayesiano es la siguiente: Sea $K > 0$ un número natural. Suponga que existen K días en donde la NOM ambiental de ozono ha sido sobrepasado. Sean d_1, d_2, \dots, d_K los días que este suceso ocurrió. Indique por $\mathbf{D} = \{d_1, d_2, \dots, d_K\}$ el conjunto de datos observados. Dada la relación entre la distribución a posteriori y a priori y la función de verosimilitud del modelo (ver ecuación 2.3.1) se tiene que,

$$P(\theta|\mathbf{D}) = P(\alpha, \beta, \sigma|\mathbf{D}) \propto \mathcal{L}(\mathbf{D}|\alpha, \beta, \sigma)P(\alpha)P(\beta)P(\sigma), \quad (3.2)$$

donde $P(\theta|\mathbf{D})$ es la distribución a posteriori de $\theta = (\alpha, \beta, \sigma)$ dado el conjunto de datos \mathbf{D} ; $P(\alpha)$, $P(\beta)$ y $P(\sigma)$ son las distribuciones a priori de α , β y σ respectivamente; y $\mathcal{L}(\mathbf{D}|\alpha, \beta, \sigma)$ es la función de verosimilitud. Los componentes de la ecuación (3.2) se dan a continuación.

Observación: Note que se supone que los parámetros son independientes a priori.

1.- *La función de verosimilitud.* Dado que por hipótesis tenemos un modelo Poisson no-homogéneo se tiene que la función de verosimilitud tomará la forma (ver Cox y Lewis, 1966).

$$\mathcal{L}(\mathbf{D}|\alpha, \beta, \sigma) = \mathcal{L}(\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_K|\alpha, \beta, \sigma) = \left[\prod_{i=1}^K \lambda(d_i) \right] \exp(-m(d_K)). \quad (3.3)$$

Dado que $\lambda(t)$ (3.1) y la media $m(t) = -\log[1 - F(t)]$, cuya función de distribución acumulativa Weibull exponenciada $F(t) = \left(1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}\right)^\beta$ donde $0 < t < \infty$ y $\alpha > 0$, $\beta > 0$ y $\sigma > 0$, se tiene lo siguiente.

Teorema 3: La función de verosimilitud con $\lambda(t)$ dada por (3.1) toma la siguiente forma,

$$\mathcal{L}(\mathbf{D}|\alpha, \beta, \sigma) = \frac{(\alpha\beta)^K \left[\prod_{i=1}^K d_i^{\alpha-1} e^{-\left(\frac{d_i}{\sigma}\right)^\alpha} \left(1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right)^{\beta-1} \right]}{\sigma^{\alpha K} \left[\prod_{i=1}^{K-1} \left(1 - \left[1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right]^\beta\right) \right]}. \quad (3.4)$$

Demostración: Primero se obtiene la expresión para $\prod_{i=1}^K \lambda(d_i)$ tomando la ecuación (3.1). De esta forma,

$$\prod_{i=1}^K \lambda(d_i) = \frac{(\alpha\beta)^K \left[\prod_{i=1}^K \left(1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right)^{\beta-1} \right] \left[\prod_{i=1}^K \left(\frac{d_i}{\sigma}\right)^{\alpha-1} \right] \left[\prod_{i=1}^K e^{-\left(\frac{d_i}{\sigma}\right)^\alpha} \right]}{\sigma^K \left[\prod_{i=1}^K \left(1 - \left[1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right]^\beta\right) \right]}.$$

Para $\exp\{-m(d_K)\}$ se tiene

$$\exp\{-m(d_K)\} = \exp\left\{-\left[-\log\left[1 - \left(1 - e^{-\left(\frac{d_K}{\sigma}\right)^\alpha}\right)^\beta\right]\right]\right\} = \left[1 - \left(1 - e^{-\left(\frac{d_K}{\sigma}\right)^\alpha}\right)^\beta\right].$$

Sustituyendo en $\mathcal{L}(\mathbf{d}_1 \mathbf{d}_2, \dots, \mathbf{d}_K|\alpha, \beta, \sigma) = \left[\prod_{i=1}^K \lambda(d_i)\right] \exp(-m(d_K))$ se tiene

$$\mathcal{L}(\mathbf{d}_1 \mathbf{d}_2, \dots, \mathbf{d}_K|\alpha, \beta, \sigma) = \left(\frac{(\alpha\beta)^K \left[\prod_{i=1}^K \left(1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right)^{\beta-1} \right] \left[\prod_{i=1}^K \left(\frac{d_i}{\sigma}\right)^{\alpha-1} \right] \left[\prod_{i=1}^K e^{-\left(\frac{d_i}{\sigma}\right)^\alpha} \right]}{\sigma^K \left[\prod_{i=1}^K \left(1 - \left[1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right]^\beta\right) \right]} \right) \left[1 - \left(1 - e^{-\left(\frac{d_K}{\sigma}\right)^\alpha}\right)^\beta\right],$$

y por lo tanto,

$$\mathcal{L}(\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_K | \alpha, \beta, \sigma) = \left(\frac{(\alpha\beta)^K \left[\prod_{i=1}^K \left(1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha} \right)^{\beta-1} \right] \left[\prod_{i=1}^K \left(\frac{d_i}{\sigma}\right)^{\alpha-1} \right] \left[\prod_{i=1}^K e^{-\left(\frac{d_i}{\sigma}\right)^\alpha} \right]}{\sigma^K \left[\prod_{i=1}^{K-1} \left(1 - \left[1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha} \right]^\beta \right) \right]} \right). \quad (I)$$

Note que

$$\prod_{i=1}^K \left(\frac{d_i}{\sigma}\right)^{\alpha-1} = \frac{\prod_{i=1}^K d_i^{\alpha-1}}{\prod_{i=1}^K \sigma^{\alpha-1}} = \frac{\prod_{i=1}^K d_i^{\alpha-1}}{\sigma^{K(\alpha-1)}}.$$

Sustituyendo en (I) se tiene que

$$\begin{aligned} \mathcal{L}(\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_K | \alpha, \beta, \sigma) &= \frac{(\alpha\beta)^K \left[\prod_{i=1}^K \left(1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha} \right)^{\beta-1} \right] \left[\prod_{i=1}^K d_i^{\alpha-1} \right] \left[e^{-\sum_{i=1}^K \left(\frac{d_i}{\sigma}\right)^\alpha} \right]}{\sigma^K \left[\prod_{i=1}^{K-1} \left(1 - \left[1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha} \right]^\beta \right) \right]} \sigma^{\alpha K - K} \\ &= \frac{(\alpha\beta)^K \left[\prod_{i=1}^K \left(1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha} \right)^{\beta-1} \right] \left[\prod_{i=1}^K d_i^{\alpha-1} e^{-\left(\frac{d_i}{\sigma}\right)^\alpha} \right]}{\sigma^{\alpha K} \left[\prod_{i=1}^{K-1} \left(1 - \left[1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha} \right]^\beta \right) \right]}. \end{aligned}$$

Por lo tanto llegamos a la ecuación (3.4). ■

2.- *La distribución a priori:* Debido a la falta de información más precisa sobre el comportamiento de los parámetros α , β y σ , se considera lo siguiente:

- a) En un primer caso se considera a $\beta = 1$ y también que α y σ son uniformemente distribuidos.
- b) En el segundo caso, se hace uso de la información proporcionada por el primer caso (cuando $\beta = 1$) para obtener distribuciones a priori con más información para los parámetros α , β y σ . Por lo que se tiene la siguiente selección.
 - i) $\alpha \sim \text{Gamma}(a_1, b_1)$
 - ii) $\sigma \sim \text{Gamma}(a_2, b_2)$
 - iii) $\beta \sim \text{Uniforme}(a_3, b_3)$ $a_3 < b_3$, $a_3, b_3 \geq 0$,

donde los hiperparámetros a_1, a_2, a_3, b_1, b_2 y b_3 son cantidades conocidas y serán especificadas posteriormente.

La selección de valores de los hiperparámetros de la distribución a priori para α y σ en el segundo caso se basan en los resultados obtenidos del primer análisis (cuando se tiene $\beta = 1$).

A continuación se realiza el planteamiento para la obtención de las distribuciones a posteriori marginales condicionales completas.

Note que la distribución Gamma (α, λ) de una variable aleatoria X , tiene la siguiente expresión.

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad \text{para } x > 0$$

$$\text{donde } \Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1} dy \quad \text{y } \alpha, \lambda > 0$$

$$\text{y es tal que } E(X) = \frac{\alpha}{\lambda} \quad \text{y} \quad \text{Var}(X) = \frac{\alpha}{\lambda^2}.$$

Así i) $\alpha \sim \text{Gamma}(a_1, b_1)$

$$f(\alpha) = \frac{b_1^{a_1}}{\Gamma(a_1)} \alpha^{a_1-1} e^{-b_1 \alpha}, \quad \text{para } a_1, b_1 > 0$$

ii) $\sigma \sim \text{Gamma}(a_2, b_2)$

$$f(\sigma) = \frac{b_2^{a_2}}{\Gamma(a_2)} \sigma^{a_2-1} e^{-b_2 \sigma}, \quad \text{para } a_2, b_2 > 0$$

iii) $\beta \sim \text{Uniforme}(a_3, b_3)$.

Teorema 4: La función a posteriori del modelo esta dada por:

$$P(\theta|D) \propto \alpha^{a_1-1} \alpha^K \sigma^{a_2-1} \sigma^{-\alpha K} e^{-(\alpha b_1 + \sigma b_2)} \beta^K \frac{\prod_{i=1}^K \left[d_i^{\alpha-1} e^{-\left(\frac{d_i}{\sigma}\right)^\alpha} \left(1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right)^{\beta-1} \right]}{\prod_{i=1}^{K-1} \left(1 - \left[1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right]^\beta\right)}.$$

Demostración: Utilizando la fórmula (3.2) se tiene que

$$\begin{aligned}
P(\theta|D) &\propto P(\alpha)P(\beta)P(\sigma)\mathcal{L}(D|\alpha, \beta, \sigma) \\
&\propto \alpha^{a_1-1}\alpha^K\sigma^{a_2-1}\sigma^{-\alpha K}e^{-(ab_1+\sigma b_2)}\beta^K \frac{\prod_{i=1}^K \left[d_i^{\alpha-1} e^{-\left(\frac{d_i}{\sigma}\right)^\alpha} \left(1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right)^{\beta-1} \right]}{\prod_{i=1}^{K-1} \left(1 - \left[1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right]^\beta\right)} \\
&= \alpha^{a_1+K-1}\sigma^{a_2-\alpha K-1}\beta^K e^{-(ab_1+\sigma b_2)} \frac{\prod_{i=1}^K \left[d_i^{\alpha-1} e^{-\left(\frac{d_i}{\sigma}\right)^\alpha} \left(1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right)^{\beta-1} \right]}{\prod_{i=1}^{K-1} \left(1 - \left[1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right]^\beta\right)}. \quad (3.5)
\end{aligned}$$

Observación: Cuando se supone β conocida e igual a 1, con α y σ uniformemente distribuidas se tiene que

$$\begin{aligned}
P(\theta|D) \propto \mathcal{L}(D|\alpha, \beta, \sigma) &= \frac{\prod_{i=1}^K \left[d_i^{\alpha-1} e^{-\sum_{i=1}^K \left(\frac{d_i}{\sigma}\right)^\alpha} \right]}{\prod_{i=1}^{K-1} \left(e^{-\left(\frac{d_i}{\sigma}\right)^\alpha} \right)} = \frac{\prod_{i=1}^K d_i^{\alpha-1} e^{-\sum_{i=1}^K \left(\frac{d_i}{\sigma}\right)^\alpha}}{e^{-\sum_{i=1}^{K-1} \left(\frac{d_i}{\sigma}\right)^\alpha}} \\
&= \left(\prod_{i=1}^K d_i^{\alpha-1} \right) e^{-\left(\frac{d_K}{\sigma}\right)^\alpha} \quad (3.6)
\end{aligned}$$

Teorema 5: Las distribuciones a posteriori marginales condicionales completas están dadas por:

$$\begin{aligned}
P(\alpha|\beta, \sigma, D) &\propto \alpha^{a_1-1} e^{-ab_1} f_2(\alpha, \beta, \sigma), \\
P(\sigma|\alpha, \beta, D) &\propto \sigma^{a_2-1} e^{-\sigma b_2} f_3(\alpha, \beta, \sigma) \\
\text{y } P(\beta|\alpha, \sigma, D) &\propto e^{K \log \beta} f_1(\alpha, \beta, \sigma).
\end{aligned}$$

Las funciones $f_1(\alpha, \beta, \sigma)$, $f_2(\alpha, \beta, \sigma)$ y $f_3(\alpha, \beta, \sigma)$, se explican en la demostración.

Demostración: Para obtener las distribuciones a posteriori marginales condicionales completas se usará una forma alternativa de (3.5). De esta forma,

$$\log P(\theta|D) = \log \left[\alpha^{a_1+K-1} \sigma^{a_2-\alpha K-1} \beta^K e^{-(ab_1+\sigma b_2)} \frac{\prod_{i=1}^K \left[d_i^{\alpha-1} e^{-\left(\frac{d_i}{\sigma}\right)^\alpha} \left(1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right)^{\beta-1} \right]}{\prod_{i=1}^{K-1} \left(1 - \left[1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right]^\beta\right)} \right].$$

Por lo tanto,

$$\begin{aligned} \log[P(\alpha|\beta, \sigma, D)] &\propto \log \alpha^{a_1-1} - \alpha b_1 + K \log(\alpha) - K \alpha \log(\sigma) + \sum_{i=1}^K (\alpha - 1) \log(d_i) \\ &\quad - \sum_{i=1}^K \left(\frac{d_i}{\sigma}\right)^\alpha + (\beta - 1) \log\left(1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right) - \sum_{i=1}^{K-1} \log\left(1 - \left[1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right]^\beta\right) \end{aligned}$$

Aplicando la función exponencial

$$\begin{aligned} P(\alpha|\beta, \sigma, D) &\propto \alpha^{a_1-1} e^{-\alpha b_1} e^{K \log(\alpha) - K \alpha \log(\sigma)} \exp\left(\sum_{i=1}^K (\alpha - 1) \log(d_i) - \sum_{i=1}^K \left(\frac{d_i}{\sigma}\right)^\alpha\right. \\ &\quad \left. + (\beta - 1) \log\left(1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right) - \sum_{i=1}^{K-1} \log\left(1 - \left[1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right]^\beta\right)\right). \end{aligned}$$

Esta expresión se simplifica quedando de la siguiente manera:

$$P(\alpha|\beta, \sigma, D) \propto \alpha^{a_1-1} e^{-\alpha b_1} f_2(\alpha, \beta, \sigma)$$

donde

$$f_2(\alpha, \beta, \sigma) = \exp[K \log(\alpha) - K \log(\sigma)] f_1(\alpha, \beta, \sigma)$$

y

$$\begin{aligned} f_1(\alpha, \beta, \sigma) &= \exp\left(\sum_{i=1}^K (\alpha - 1) \log(d_i) - \sum_{i=1}^K \left(\frac{d_i}{\sigma}\right)^\alpha + (\beta - 1) \log\left(1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right)\right. \\ &\quad \left. - \sum_{i=1}^{K-1} \log\left(1 - \left[1 - e^{-\left(\frac{d_i}{\sigma}\right)^\alpha}\right]^\beta\right)\right). \end{aligned}$$

Para obtener $P(\sigma|\alpha, \beta, D)$ se toma en cuenta que $\sigma \sim \text{Gamma}(a_2, b_2)$.

$$P(\sigma|\alpha, \beta, D) \propto \sigma^{a_2-1} e^{-\sigma b_2} f_3(\alpha, \beta, \sigma)$$

donde

$$f_3(\alpha, \beta, \sigma) = e^{-K \log(\sigma)} f_1(\alpha, \beta, \sigma).$$

Finalmente para obtener $P(\beta|\alpha, \sigma, D)$ se toma en cuenta solamente que $\beta \sim \text{Uniforme}(a_3, b_3)$, $a_3, b_3 \geq 0$, $a_3 < b_3$ y los parámetros a_3, b_3 , por lo tanto,

$$P(\beta|\alpha, \sigma, D) \propto \log \beta^K \log f_1(\alpha, \beta, \sigma) = K \log \beta \log f_1(\alpha, \beta, \sigma).$$

Aplicando la función exponencial

$$P(\beta|\alpha, \sigma, D) \propto e^{K \log \beta} f_1(\alpha, \beta, \sigma).$$

Estas son las distribuciones a posteriori marginales condicionales completas, que se utilizan para generar los valores α , β y σ en el algoritmo del Muestreo de Gibbs que esta implementado en el software WinBugs.

Capítulo 4

Aplicación de las mediciones de Ozono en la Ciudad de México

En este capítulo se describirán los resultados obtenidos en el artículo (Achcar et al. 2008) para las mediciones de los niveles de ozono en el Área Metropolitana de la Ciudad de México (AMCM). Las autoridades ambientales de la ciudad de México declaran situaciones de emergencia a nivel local debido a que el comportamiento de los contaminantes varían de una zona a otra de la ciudad, de tal manera que el Área Metropolitana la dividen en cinco zonas: Noreste (NE), Noroeste (NO), Centro (CE), Sureste (SE) y Suroeste (SO) y las estaciones de monitoreo de ozono están ubicadas por toda la ciudad. Cuando el límite ambiental establecido es sobrepasado en una ó más zonas, la emergencia se declara de manera local. Las medidas preventivas son tomadas solo en estas zonas en vez de hacerse en toda la ciudad de México.

Los datos que se utilizaron en el análisis se pueden descargar en la pagina de la Secretaria del Medio Ambiente* que corresponden a los años de 1998 al 2004, de las mediciones máximas diarias de ozono en cada zona y de toda la ciudad. Las mediciones se obtienen minuto a minuto y se promedian cada hora y estos resultados se reportan a cada estación de monitoreo de ozono. La medición máxima diaria para una zona es el valor máximo obtenido de todos los valores reportados, hora por hora durante un periodo de 24 horas por cada estación en cada zona. En el caso del AMCM, las mediciones máximas diarias se obtienen tomando el valor máximo diario proporcionado por las zonas NE, NO, CE, SE y SO. El promedio de las mediciones de los siete años, en las zonas de NE, NO, CE, SE y SO son 0.097, 0.121, 0.126, 0.143 y 0.12, respectivamente con las desviaciones estándar 0.035, 0.049, 0.046, 0.052 y 0.042. Para el AMCM se tiene una medición promedio de 0.154 con una desviación estándar de 0.05. Si se toma en cuenta las mediciones dadas por la AMCM durante el periodo 1998-2004, se observa que el valor de la NOM de ozono en México, que es de 0.11ppm (0.11 partes por millón), fue sobrepasado en 2063 días y los picos diarios sobrepasaron el valor 0.22ppm durante 237 días.

Los autores consideran que los límites son $L = 0.11\text{ppm}$, 0.17ppm y 0.22ppm (donde ppm = partes por millón). El primer valor es el que representa la NOM de ozono. El último se utiliza porque es el doble del valor de la NOM además de que $L = 0.22\text{ppm}$ es para cuando se declara emergencia en la ciudad de México y el valor $L = 0.17\text{ppm}$ es un valor intermedio entre los otros dos valores. El análisis se realizó para cada valor límite y zona (incluyendo el caso del AMCM) de forma separada.

* Sistema de Monitoreo Atmosférico de la ciudad de México www.sma.df.gob.mx/simat2

Las simulaciones se realizaron utilizando el software WingBugs. Cuando se considera a cualquiera de los límites $L = 0.11\text{ppm}, 0.17\text{ppm}, 0.22\text{ppm}$ se toma en cuenta lo siguiente: para el primer caso de simulación, se toma a $\beta = 1$ y la distribución a priori de α y σ se toma como uniforme en los intervalos $(0,2)$ y $(0,100)$, respectivamente.

En el segundo caso β tiene una distribución Uniforme definida en el intervalo $(0,100)$. Los parámetros de las distribuciones a priori Gamma de α y σ varían de acuerdo a la zona y al límite considerado. En el caso de la distribución a priori de α se considera lo siguiente: cuando el límite $L = 0.11\text{ppm}$, se tiene $a_1 = 0.09$ para todas las zonas excepto la zona Noreste (NE) donde $a_1 = 0.08$. Si el límite es $L = 0.17\text{ppm}$, se tiene $a_1 = 0.06$ para el Noroeste (NO), Sureste (SE) y Suroeste (SO), y $a_1 = 0.07$ para las zonas Centro (CE) y el AMCM, para la zona del Noreste se obtiene $a_1 = 0.04$. Para todos los casos se tiene $b_1 = 0.1$. Cuando se considera el caso de la distribución a priori de σ se tiene $b_2 = 0.1$ en todos los casos excepto cuando $L = 0.22\text{ppm}$ y cuando se considera la zona Noreste (NE), los valores para el parámetro a_2 se describen a continuación.

Tabla 1. Valores de a_2 para las distribución a priori de las cantidades variables de σ utilizados en el segundo caso de la simulación para cada región y límite.

	0.11ppm	0.17ppm	0.22ppm
NO	0.06	0.02	0.25
NE	0.05	0.12	0.57
CE	0.08	0.05	0.7
SO	0.08	0.04	0.09
SE	0.07	0.04	0.6
AMCM	0.08	0.03	0.027

La muestra utilizada para estimar los parámetros fue tomada cada 30 valores generados. Cuando se toma el límite $L = 0.11\text{ppm}$, se tiene que el periodo de calentamiento del algoritmo del muestreo de Gibbs, es de 1000 pasos para la zona del Suroeste (SO) y 3000 pasos para cualquier otra zona. En el segundo caso, el periodo de calentamiento fue de 1000 pasos en el AMCM y 2000 pasos en cualquier otra zona. La estimación de la distribución a posteriori de α , σ y β (siempre que el último dato sea una cantidad aleatoria) se realiza utilizando una muestra de 2000 datos, en el primer caso para las regiones Suroeste (SO), Noroeste (NO) y el AMCM y en el segundo caso para las regiones: Sureste (SE), Noreste (NO) y Centro (CE), el tamaño de la muestra fue de 3000 datos. Finalmente, en el segundo caso la muestra para la zona Noroeste (NO) fue de 4000 datos y para el caso del AMCM fue de 2308 datos. Con el límite $L = 0.17\text{ppm}$ se tiene que en el periodo de calentamiento para el primer caso de la simulación fueron 3000 pasos y en el segundo caso de la simulación fueron 2000 pasos.

En el caso del límite $L = 0.22\text{ppm}$ una muestra de tamaño 2000 datos se tomo para todas las zonas y casos, exceptuando para el segundo caso la zona del Noroeste (NO), donde se tomo una muestra de 1756 datos. Cuando el límite es $L = 0.22\text{ppm}$, el primer periodo de calentamiento para el primer caso fue de 5000 pasos para las zonas Suroeste (SO), Sureste (SE) y AMCM y 3000 pasos para las zonas del Noroeste (NO), Noreste (NE) y Centro (CE). En el segundo caso de periodo de calentamiento fue de 5000 pasos para las zonas del Suroeste (SO) y Sureste (SE) y para las otras zonas fueron 3000 pasos. En el segundo caso para la zona Suroeste (SO) se tomo una muestra de 2000 datos y de 3000 datos para cualquier otra zona.

Para seleccionar el mejor modelo de cada caso (límites y zonas) se utilizo el “Criterio de Información Desviada (CID)” (Spiegelhalter *et al.*, 2002) donde se ocupan subrutinas internas del programa WinBugs y una modificación en la fórmula del “Criterio de Información de Bayes” (CIB) dado por

$$\text{CIB} = 2E(\log[\mathcal{L}(D|\theta)]) - d\log(n), \quad (4.1)$$

donde d es el número de parámetros en el modelo y n es el tamaño de la muestra (Carlin y Louis, 2000). Las mejores estimaciones de valores en la simulación del algoritmo de muestreo de Gibbs en el CID y en el CIB fueron: 3434.5 (NE), 4464.5 (NO), 4717.1 (CE), 4998.6 (SO), 4661.8 (SE) y 5000.2 (AMCM) del límite $L = 0.11\text{ppm}$ y para el modelo $\beta = 1$.

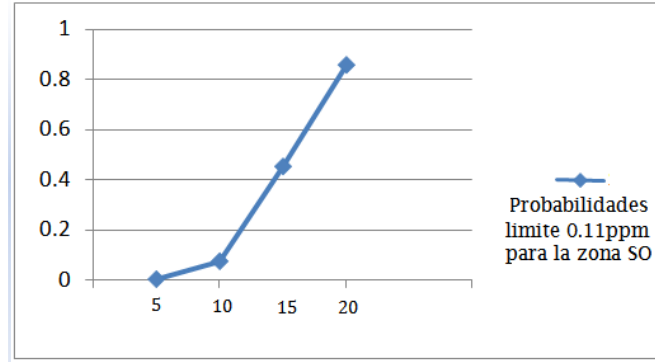
Como ejemplo de los resultados obtenidos, consideramos las zonas del centro (CE) y Suroeste (SO) de la ciudad de México, donde la importancia radica en la dirección que toma el viento en la ciudad (de Noreste (NE) al Suroeste (SO)) y en que algunos precursores de ozono se ubican en el Noroeste (NE) que son llevados por el viento de las zonas centro (CE) y Suroeste (SO).

Tabla 2. Media a posteriori para los parámetros α , β y σ .

		Media		
		0.11 ppm	0.17 ppm	0.22 ppm
CE	α	0.91	0.7	0.65
	σ	0.79	0.47	6.72
	β	1	1	1
SO	α	0.93	0.76	0.64
	σ	0.82	0.42	0.93
	β	1	1	1

Tomando en cuenta los primeros 25 días de Enero de 2005 si se quiere obtener las probabilidades de tener 5, 10, 15 y 20 días o menos, donde se sobrepasa el valor de la NOM de ozono, se utiliza la distribución Poisson con función de intensidad $\lambda(t)$, $t \geq 0$ Weibull exponenciada donde los parámetros α y σ en $\lambda(t)$ son las medias de la Tabla 2 de las zonas Centro (CE) y Suroeste (SO) tomados.

Las probabilidades para los límites $L = 0.11\text{ppm}$, 0.17ppm y 0.22ppm , en la zona Suroeste (SO) cuando $K \leq 5$ son: 0.001, 0.52 y 1, cuando $K \leq 10$ son: 0.073, 0.97 y 1, cuando $K \leq 15$ son: 0.453, 1 y 1 y para el valor $K \leq 20$ son: 0.86, 1 y 1. Para la zona del centro (CE) cuando $K \leq 5$ son: 0.006, 0.93 y 1, cuando $K \leq 10$ son: 0.182, 1 y 1, cuando $K \leq 15$ son: 0.678, 1 y 1 y para el valor $K \leq 20$ son: 0.65, 1 y 1.



El procedimiento para obtener estas probabilidades es el siguiente:

$$\mathbb{P}[N(t+s) - N(t) \leq n] = e^{-[m(t+s)-m(t)]} \frac{[m(t+s) - m(t)]^n}{n!}$$

con

$$m(t) = -\log \left[1 - \left(1 - \exp \left(- \left(\frac{t}{\sigma} \right)^\alpha \right) \right)^\beta \right]$$

Sustituyendo los valores para la zona Suroeste (SO) en el caso de $K \leq 5$, para el límite de 17ppm se tiene,

$$\mathbb{P}[N(2557 + 25) - N(25) \leq 5] = \sum_{n=0}^5 \exp \left[\left(\frac{2582}{0.42} \right)^{0.76} - \left(\frac{2557}{0.42} \right)^{0.76} \right] \frac{\left[\left(\frac{2582}{0.42} \right)^{0.76} - \left(\frac{2557}{0.42} \right)^{0.76} \right]^n}{n!}$$

$$\mathbb{P}[N(2582) - N(25) \leq 5] = 0.52,$$

donde $t = 2557$ y $n = 25$.

Cuando se toma la zona Centro (CE) en el caso de $K \leq 5$, para el límite de 17ppm.

$$\mathbb{P}[N(2557 + 25) - N(25) \leq 5] = \sum_{n=0}^5 \exp \left[\left(\frac{2582}{0.47} \right)^{0.7} - \left(\frac{2557}{0.47} \right)^{0.7} \right] \frac{\left[\left(\frac{2582}{0.47} \right)^{0.7} - \left(\frac{2557}{0.47} \right)^{0.7} \right]^n}{n!}$$

$$\mathbb{P}[N(2582) - N(25) \leq 5] = 0.93.$$

Para calcular las otras probabilidades de las zonas suroeste (SO) y centro (CE) con los diferentes límites se realiza el mismo procedimiento.

Nota: Para calcular estas probabilidades los autores utilizaron el programa Mathematica.

Conclusiones

Los autores utilizan métodos Bayesianos empíricos, así que se toma el caso mas sencillo, primero ($\beta = 1$) y con la información obtenida de este paso se utilizan simulaciones con el modelo mas general, sin embargo el modelo seleccionado es el mas sencillo.

Al considerar los primero 25 días del año 2005, el límite de $L = 0.11\text{ppm}$ se rebasa en 14 días para la zona Centro (CE) y 20 días para la zona Suroeste (SO). La probabilidad de que en un intervalo de 10 a 20 días se sobrepase el límite $L = 0.11$, es de 0.8 en cambio la probabilidad de que en cinco o menos días se rebase el límite es de 0.001. De tal manera que hay mayor probabilidad de que existan más número de días en donde se supere el límite $L = 0.11\text{ppm}$.

Un factor importante a considerar es la dirección del viento en la ciudad de México que va de Noreste (NE) a Suroeste (SO) y la producción de ozono que se origina en las zonas Noroeste (NO) y Centro (CE). Para la zona Centro (CE) en ningún día se supero el límite $L = 0.22\text{ppm}$ y en la zona Suroeste (SO) solamente en un día. La probabilidad de que se sobrepase el límite $L = 0.22\text{ppm}$ en un intervalo de 5 a 10 días para los primeros 25 días del año 2005 es de cero, sin embargo hay una probabilidad de uno, de que haya menos de 5 días en donde se sobrepase el límite $L = 0.22\text{ppm}$, si la cantidad de datos aumenta, los resultados obtenidos para el análisis de la estadística Bayesiana mejorarían. Es importante puntualizar que el ajuste del modelo Poisson no homogéneo con la función de intensidad Weibull-Exponenciada aplicada al incumplimiento de los niveles de ozono en la ciudad de México fue muy satisfactorio, especialmente en las muestras grandes.

Los valores acumulados y esperados de incumplimiento contra el tiempo para el caso del límite $L = 0.11\text{ppm}$ para las zonas Noroeste (NO), Suroeste (SO), Centro (CE), y AMCM resultaron con similitudes muy aceptables y para las zonas Noreste (NE) y Sureste (SE) hay pequeñas desviaciones pero aún así el ajuste es muy satisfactorio. En el caso del límite $L = 0.17\text{ppm}$ el ajuste es razonable pero cabe la posibilidad de mejorar si se considera la distribución a priori informativa para los parámetros así como también tomar en cuenta otra fórmula paramétrica para la función de intensidad o tasa del modelo Poisson no homogéneo. Para el límite $L = 0.22\text{ppm}$ los valores esperados aparecen por muy debajo de los valores acumulados por lo que en este caso no es aceptable el ajuste. Es importante señalar que la cantidad de datos para este caso fue limitado en especial para las zonas: Centro (CE) y Suroeste (SE) donde hubo 38 y 57 días donde se rebaso el límite.

Referencias

- Achcar JA, Fernández-Bremauntz AA, Rodrigues ER, y Tzintzun G. (2008) "Estimating the number of ozone peaks in México City using a non-homogeneous Poisson model" *Environmetrics* 19: 469-485.
- Álvarez LJ, Fernández-Bremauntz AA, Rodrigues ER, Tzintzun G. 2005. "Maximum a posteriori estimation of the daily ozone peaks in Mexico City". *Journal of Agricultural, Biological, and Environmental Statistics* 10: 276-290.
- Austin J, Tran H. 1999. "A characterization of the weekday-weekend behavior of ambient ozone concentration in California". *Air Pollution VII*. WIT Press: Ashurst Lodge, Ashurst, Southampton, UK; 645-661.
- Carlin BP, Louis TA. 2000. "Bayes and Empirical Bayes Methods for Data Analysis" 2nd edn, Chapman and Hall: USA.
- Casella G, Edward IG. 1992. "Explaining the Gibbs Sampler" *The American Statistician*, Vol. 46, No. 3. pp. 167-174.
- Cox DR, Lewis PA. 1996 "Statistical Analysis of Series of Events". Methuen: UK.
- Flaum JB, Rao ST, Zurbenko IG. 1996. "Moderating influence of meteorological conditions on ambient ozone concentrations". *Journal of the Air and Waste Management Association* 46: 33-46.
- Gelfand AE, Smith AFM. 1990. "Sampling-Based Approaches to Calculating Marginal Densities" *Journal of the American Statistical Association*, Vol. 85, No. 410., pp. 398-409.
- Geman S, Geman D. (1984). "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6: 721-741.
- Guardani R, Aguilar JL, Nascimento CAO, Lacava CIV, Yanagi Y. 2003. "Ground-level ozone mapping in large urban areas using multivariate analysis": application to the São Paulo Metropolitan Area. *Journal of the Air and Waste Management Association* 53: 553-559.
- Guardani R, Nascimento CAO, Guardani MLG, Martins MHRB, Romano J. 1999. "Study of atmospheric ozone formation by means of a neural network based model". *Journal of the Air and Waste Management Association* 49: 316-323.
- Horowitz J. 1980. "Extreme values from a nonstationary stochastic process: an application to air quality analysis". *Technometrics* 22: 469-482.

- Huerta G, Sansó B. 2005. "Time-varying models for extreme values". Technical Report 2005-4. Department of Applied Mathematics and Statistics. University of California: Santa Cruz. USA.
- Javits JS, 1980. "Statistical interdependencies en the ozone national ambient air quality standard". Journal of air Pollution Control Association 30:58-59.
- Larsen LC, Bradley RA, Honcoop GL. 1990 "A new method of characterizing the variability of air quality-related indicators". Air and Waste Management Association's International Specialty Conference of Tropospheric Ozone and the Environment. Los Angeles EUA.
- López J. 2005. "El uso del muestreo de Gibbs para identificar segmentos homogéneos en una secuencia de ADN heterogénea". Tesis de Licenciatura Facultad de Ciencias, UNAM México.
- Raftery AE. 1989. "Are ozone exceedance rate decreasing?" Comment of paper "Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone" by R. L. Smith. Statistical Sciences 4: 378-381.
- Reyes M. 2008. "El método de Montecarlo vía cadena de Markov y su uso en problemas de Epidemiología". Tesis de Licenciatura Facultad de Ciencias, UNAM México.
- Roberts EM. 1979a. "Review of statistics of extreme values with applications to air quality data". Part I. Review. Journal of the Air Pollution Control Association 29: 623-637.
- Roberts EM. 1979b. "Review of statistics of extreme values with applications to air quality data". Part II. Applications. Journal of the Air Pollution Control Association 29: 733-740.
- Ross SM. 1996. "Stochastic Processes" 2nd edn, Capítulo 2, 59-63, 78-79. ed. John Wiley & Sons Inc: USA.
- Smith RL. 1989. "Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone". Statistical Sciences 4: 367-393.
- Spiegelhalter DJ, Carlin BP, van der Linde A. 2002. "Bayesian measures of model complexity and fit". Journal of the Royal Statistical Society Series B 64: 583-639.
- Stephens DA, Dellaportas P. 1992. "Bayesian analysis of generalized linear models with covariate measurement error". Clarendon Press, Oxford, UK.
- Tudor C. 2002. "Procesos Estocásticos" 3ra ed. Sociedad Matemática Mexicana. Aportaciones Matemáticas. Textos Nivel Avanzado.