QÜESTIIÓ, vol. 26, 1-2, p. 123-138, 2002

UN MODELO POISSONIANO PARA PREDECIR LA MATRICULACIÓN DE VEHÍCULOS EN PAÍSES EUROPEOS

M. J. VALDERRAMA
A. M. AGUILERA
P. R. BOUZAS
Universidad de Granada*

En este artículo presentamos una modelización para la matriculación de vehículos en países de Europa mediante un proceso de Poisson Doblemente Estocástico con media aleatoria Normal truncada. Apoyándonos en trabajos previos acerca de este proceso, se amplía el estudio de características de éste. Asímismo, se hace una predicción de este proceso para los años 2000 y 2001.

A Poissonian model to forecast vehicules matriculations in european countries

Palabras clave: Proceso de Poisson Doblemente Estocástico, Distribución Normal truncada, Matriculaciones de Turismos

Clasificación AMS (MSC 2000): 60G55, 60K30

---

* Departamento de Estadística e I.O., Facultad de Farmacia, Campus de Cartuja, Universidad de Granada, 18071 Granada.

– Recibido en julio de 2001.
– Aceptado en noviembre de 2001.

1. INTRODUCCIÓN

El número de vehículos turismo matriculados constituye uno de los indicadores más representativos de la actividad económica de un país. Numerosos trabajos se han centrado en modelizar dicho proceso y tratar de predecir su evolución a lo largo del tiempo. Así, por ejemplo, García-Ferrer y Queralt (1998), desarollan un modelo predictivo de varias series de indicadores económicos españoles, entre ellos el número de matriculaciones, utilizando modelos de componentes no observadas con parámetros variables.

Dado que la matriculación de automóviles constituye un claro ejemplo de proceso de recuento de tipo Poissoniano, en el presente trabajo presentamos la aplicación de metodología original para modelizar dicho fenómeno en los países europeos, en la que se considera el parámetro del proceso como un término aleatorio que debe igualmente modelizarse a partir de las trayectorias muestrales. De hecho, el proceso de Poisson es insufiente en multitud de casos debido al carácter determinístico de su función de intensidad. El proceso de Poisson Doblemente Estocástico (PPDE) es una generalización del proceso de Poisson introducida por Cox (1955).

El trabajo que presentamos trata un caso particular de PPDE, aquél cuya media es una Normal truncada en cada instante de tiempo. Está principalmente enfocado a la ampliación del estudio de características de ese caso particular de PPDE y a su aplicación al caso real de proceso puntual que hemos comentado.

Más concretamente, en la Sección 2 se hace un resumen teórico del PPDE recordando aquellas características que más adelante se necesitarán. Se particulariza el caso en el que la media sea una Normal truncada en cada instante de tiempo, esquematizando los cálculos que conducen a la expresión explícita de la esperanza y varianza de este tipo de PPDE.

En la Sección 3, se muestra un método de predicción del PPDE con Normal truncada en el futuro. El método está basado en la modelización de la función paramétrica del proceso puntual y su estimación para un momento futuro. En la última sección, se aplica lo presentado en las secciones anteriores al caso real del proceso puntual número de automóviles demandados en Europa. Se muestran los resultados de cada uno de los pasos del método de predicción y se finaliza dando una predicción del proceso para el año 2000, que se compara con el dato muestral para ese año, y una predicción para el año 2001.

2. PPDE CON MEDIA ALEATORIA NORMAL

Intuitivamente, un PPDE es un proceso de recuento de Poisson cuya intensidad viene dada en cada instante por un proceso externo al proceso puntual en estudio. Notaremos al PPDE como $\{N(t): t \geq t_0\}$, donde $N(t)$ indica el número de sucesos ocurridos desde t_0 hasta t. Más concretamente, sea $\{x(t): t \geq t_0\}$ un proceso continuo a la izquierda, entonces se dice que el proceso de recuento $\{N(t): t \geq t_0\}$ es un proceso de Poisson doblemente estocástico con proceso de intensidad $\{\lambda(t,x(t)): t \geq t_0\}$, si para cada trayectoria de $\{x(t): t \geq t_0\}$, $N(\cdot)$ es un proceso de Poisson con función de intensidad $\{\lambda(t,x(t)): t \geq t_0\}$. Es decir, si $\{N(t): t \geq t_0\}$ condicionado a $\{x(t): t \geq t_0\}$ es un proceso de Poisson con función de intensidad $\{\lambda(t,x(t)): t \geq t_0\}$. Al proceso $\{x(t): t \geq t_0\}$ se le denomina proceso información. El PPDE ha sido ampliamente estudiado por Daley and Vere-Jones (1988), Snyder and Miller (1991), Valderrama *et al.* (1995), entre otros.

Mediante el método de condicionamiento, se tiene como principal estadístico de recuento que la probabilidad de que el número de puntos o sucesos que ocurran en $[t_0,t)$ sea n es igual a:

$$\begin{aligned} P[N(t)=n] &= E\left\{P[N(t)=n/x(\sigma): t_0 \leqslant \sigma < t]\right\} \\ &= E\left\{\frac{1}{n!}\Lambda(t,x(t))^n \exp[-\Lambda(t,x(t))]\right\} = G_\Lambda^{n)}(-1) \end{aligned} \tag{1}$$

para $n = 0,1,2,\ldots$, donde $\Lambda(t,x(t))$ es la función paramétrica del PPDE, es decir $\Lambda(t,x(t)) = \int_{t_0}^{t} \lambda(\sigma,x(\sigma))d\sigma = E[N(t)/x(\sigma): t_0 \leqslant \sigma < t]$ y $G_\Lambda^{n)}(s)$ es la derivada n-*ésima* de la función generatriz de momentos (que supondremos existe) de $\Lambda(t,x(t))$.

Usando nuevamente el método de condicionamiento, la función característica de un PPDE, usando la notación de Snyder y Miller (1991), resulta ser,

$$\begin{aligned} M_{N(t)}(iu) &= E\left\{\exp\left[\left(e^{iu}-1\right)\int_{t_0}^{t}\lambda(\sigma,x(\sigma))\,d\sigma\right]\right\} \\ &= E\left\{\exp\left[\left(e^{iu}-1\right)\Lambda(t,x(t))\right]\right\} = M_\Lambda(e^{iu}-1) \end{aligned} \tag{2}$$

donde $M_\Lambda(iu)$ es la función característica de $\Lambda(t,x(t))$.

Aunque se tenga esta expresión de la función característica, su evaluación no es siempre posible. Sí que a partir de ella resulta fácil observar que

$$\begin{aligned} E[N(t)/x(\sigma): t_0 \leqslant \sigma < t] &= E[\Lambda(t,x(t))] \\ Var[N(t)/x(\sigma): t_0 \leqslant \sigma < t] &= Var[\Lambda(t,x(t))] + E[\Lambda(t,x(t))] \end{aligned}$$

Después de haber trabajado con distintos procesos de recuento (ej. número de hipotecas, número de efectos bancarios devueltos o impagados, etc.) parece que a menudo su media podría modelizarse mediante un modelo Gaussiano. Puesto que $E[N(t)/x(\sigma):$

$t_0 \leqslant \sigma < t] = \Lambda(t,x(t))$ y la función paramétrica debe ser mayor o igual que cero y creciente, para evitar problemas técnicos sin apartarse demasiado del modelo Gaussiano hemos estudiado el caso de que la media, $\Lambda(t,x(t))$, se distribuya mediante una Normal truncada en cada instante de tiempo. De hecho, ha resultado ser una acertada elección como veremos más adelante.

Si la función paramétrica se distribuye mediante una Normal truncada, que notaremos mediante $\Lambda(t,x(t)) \rightsquigarrow N_T(A(t),B(t),\mu(t),\sigma(t))$, haciendo uso de la expresión de $G_\Lambda^{n)}(-1)$ para esta distribución y la ecuación (1), se puede demostrar que la función masa de probabilidad del PPDE con esa media es (Rodríguez Bouzas *et al.*, 2001)

$$P[N(t)=n] =$$

$$\frac{1}{n!\sigma\sqrt{2\pi}}\left(P\left[\frac{A-\mu}{\sigma} < Z < \frac{B-\mu}{\sigma}\right]\right)^{-1}\int_A^B \Lambda^n e^{-\Lambda} e^{-\frac{1}{2}\left(\frac{\Lambda-\mu}{\sigma}\right)^2} d\Lambda$$

para $n = 0,1,2,\ldots$ donde A, B, μ y σ son los parámetros de $\Lambda(t,x(t))$, todos ellos funciones del tiempo y tal que $0 \leq A < \Lambda < B$, $\mu \in \mathbb{R}$, $\sigma \geq 0$ con A y B los puntos de truncamiento inferior y superior, respectivamente.

Usando la ecuación (11) y la función característica de $\Lambda(t,x(t))$, cuya expresión es

$$M_\Lambda(iu) = \frac{P[\frac{A-\mu-iu\sigma^2}{\sigma} < Z < \frac{B-\mu-iu\sigma^2}{\sigma}]}{P[\frac{A-\mu}{\sigma} < Z < \frac{B-\mu}{\sigma}]} \exp\left[iu\mu + \frac{1}{2}(iu)^2\sigma^2\right]$$

se tiene que la función característica del PPDE que nos ocupa es de la forma

$$M_{N(t)}(iu) = C(iu,t)\cdot\exp\left[\mu(e^{iu}-1) + \frac{1}{2}(e^{iu}-1)^2\sigma^2\right] \tag{3}$$

donde

$$C(iu,t) = \frac{P\left[A-(e^{iu}-1)\sigma^2 < \xi < B-(e^{iu}-1)\sigma^2\right]}{P[A<\xi<B]}$$

y ξ es una variable con distribución $N(\mu,\sigma)$, recordando que μ y σ son funciones del tiempo. La función $C(iu,t)$ que aparece en la ecuación (12) es sólo una función del tiempo dados unos μ y σ particulares. Observando además la exponencial que aparece en la misma expresión, se puede escribir abreviadamente

$$M_{N(t)}(iu) = C(iu,t)\,M_\xi\left(e^{iu}-1\right) \tag{4}$$

También se puede interpretar la función característica del proceso como el producto de $C(iu,t)$ por la función característica de un proceso de Poisson $\{N^*(t); t \geq t_0\}$ con intensidad gaussiana $\{\lambda^*(t,x(t));t\geq t_0\}$ tal que

$$E\left[\lambda^*(t,x(t))\right]=\mu'(t) \text{ y } \int_{t_0}^{t}\int_{t_0}^{t}R(u,v)\,du\,dv=\frac{1}{2}\sigma^2(t)$$

donde $R(u,v)$ es la función de covarianza de $\lambda^*(t,x(t))$, ver el artículo de Valderrama *et al.* (1995), por lo que tendríamos

$$M_{N(t)}(iu)=C(iu,t)\,M_{N^*(t)}(iu) \tag{5}$$

Ambas formas de escribir la función característica del PPDE con media Normal truncada, nos ayudan notablemente a la hora de hacer los cálculos para hallar la media y varianza del proceso, que serían imposibles de realizar usando sus correspondientes definiciones, puesto que los momentos de orden 1 y 2 de la variable ξ son ya por todos conocidos, así como los del proceso $N^*(t)$ calculados en el citado artículo.

Usando (4), tenemos que la esperanza se calcularía

$$\begin{aligned}E\left[N(t)\right]=E\left[\Lambda(t)\right]&=\left.\frac{\partial M_{N(t)}(iu)}{\partial iu}\right|_{i0}=\left.\frac{\partial M_{\Lambda(t)}(e^{iu}-1)}{\partial iu}\right|_{i0}\\&=\left.\frac{\partial C(iu,t)}{\partial iu}\right|_{i0}M_\xi\left(e^{i0}-1\right)+C(i0,t)\left.\frac{\partial M_\xi(e^{iu}-1)}{\partial iu}\right|_{i0}\end{aligned}$$

y para la varianza del proceso necesitamos el momento de segundo orden

$$\begin{aligned}E\left[N^2(t)\right]&=\left.\frac{\partial^2 M_{N(t)}(iu)}{\partial iu^2}\right|_{i0}=\\&\left.\frac{\partial^2 C(iu,t)}{\partial iu^2}\right|_{i0}M_\xi\left(e^{i0}-1\right)+2\left.\frac{\partial C(iu,t)}{\partial iu}\right|_{i0}\left.\frac{\partial M_\xi(e^{iu}-1)}{\partial iu}\right|_{i0}+\\&\qquad+C(i0,t)\left.\frac{\partial^2 M_\xi(e^{iu}-1)}{\partial iu^2}\right|_{i0}\end{aligned}$$

Puesto que $\xi \rightsquigarrow N(\mu,\sigma)$, es fácil observar que

$$M_\xi\left(e^{i0}-1\right)=1;\quad \left.\frac{\partial M_\xi(e^{iu}-1)}{\partial iu}\right|_{i0}=\mu \quad \text{y} \quad \left.\frac{\partial^2 M_\xi(e^{iu}-1)}{\partial iu^2}\right|_{i0}=\mu^2+\mu+\sigma^2$$

se comprueba también que

$$C(i0,t) = 1; \qquad \left.\frac{\partial C(iu,t)}{\partial iu}\right|_{i0} = \frac{\sigma}{\sqrt{2\pi}} \frac{e^{-\frac{1}{2}\left(\frac{B-\mu}{\sigma}\right)^2} - e^{-\frac{1}{2}\left(\frac{A-\mu}{\sigma}\right)^2}}{P[A < \xi < B]} \quad \text{y}$$

$$\left.\frac{\partial^2 C(iu,t)}{\partial iu^2}\right|_{i0} = \frac{\sigma}{\sqrt{2\pi}} \frac{(A-\mu+1)\,e^{-\frac{1}{2}\left(\frac{A-\mu}{\sigma}\right)^2} - (B-\mu+1)\,e^{-\frac{1}{2}\left(\frac{B-\mu}{\sigma}\right)^2}}{P[A < \xi < B]}$$

Tenemos ya los cálculos necesarios para obtener la expresión de la esperanza y la varianza. Aunque hemos hecho los cálculos para la descomposición de la función característica del proceso como la función $C(iu,t)$ por la función característica de ξ en $e^{iu} - 1$, resultaría igual haber tomado la segunda interpretación (ecuación 5) que hacíamos anteriormente. Queda, por tanto, que la esperanza y la varianza de un PPDE con media aleatoria Normal truncada son los siguientes:

$$E\,[N(t)] = \frac{\sigma}{\sqrt{2\pi}} \frac{f(B) - f(A)}{P[A < \xi < B]} + \mu \tag{6}$$

y

$$Var\,[N(t)] = \frac{\sigma\,[(A-\mu+1)\ f(A) - (B-\mu+1)\ f(B)]}{\sqrt{2\pi}P[A < \xi < B]} - \left[\frac{\sigma\,[f(B) - f(A)]}{\sqrt{2\pi}P[A < \xi < B]}\right]^2 + \sigma^2 + \mu \tag{7}$$

donde ξ se distribuye como una $N(\mu,\sigma)$ y $f(x) = \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$. Recordamos de nuevo que A, B, μ y σ son todas ellas funciones del tiempo.

3. PREDICCIÓN DEL PPDE CON MEDIA NORMAL TRUNCADA

Una vez conocidos de forma teórica los estadísticos del PPDE, esta sección trata el problema de modelizar y predecir este tipo de procesos puntuales teniendo conocimiento de r trayectorias muestrales. Se propone un método de estimación de la distribución de $\Lambda(t, x(t))$ en un instante t_1, conocido el proceso en un intervalo de tiempo pasado $[t_0, T]$, $T < t_1$, a partir de lo cual se puede predecir la distribución del proceso puntual en ese instante futuro. Esto nos permitirá estudiar las características del proceso en ese mismo instante de tiempo.

El método de predicción de la distribución del PPDE en t_1 se basa en la idea de modelizar la evolución de $\Lambda(t, x(t))$ en el tiempo. En el primer paso del método, se necesitará conocer la estimación de la distribución Normal truncada a partir de una muestra. Por tanto, estudiaremos la estimación de sus parámetros antes de adentrarnos en el método de predicción del PPDE.

El artículo de Mittal y Dahiya (1987) explica que los estimadores máximo verosímiles (EMV) para los parámetros de la distribución Normal truncada, $N_T(A,B,\mu,\sigma)$, pueden ser infinito con probabilidad positiva. Por esta razón, proponen otra forma de estimar esos parámetros, la estimación modal bayesiana, usando una densidad a priori para el parámetro $\frac{1}{\sigma^2}$. En la literatura bayesiana, es común usar una densidad chi-cuadrado cuando no hay información acerca del parámetro, así la densidad sería

$$f(\sigma^2) = f(\theta) = c(\nu)\,\theta^{-\frac{\nu-2}{2}} \exp\left(\frac{-1}{2\theta}\right), \theta > 0$$

donde $c(\nu) = 2^{-\frac{\nu}{2}}\left[\Gamma(\frac{\nu}{2})\right]^{-1}$ es la densidad a priori para μ. La función de verosimilitud modificada es ahora

$$L_m(\mu,\sigma/\Lambda_{1,\dots,}\Lambda_r) = \frac{c_2\,\sigma^{2-\nu} \exp\left[-\left(\sum_{i=1}^{r}(\Lambda_i-\mu)+1\right)/2\sigma^2\right]}{\left[\int_A^B \exp\left(-(y-\mu)^2/2\sigma^2\right)dy\right]^r}$$

donde c_2 es una función de ν. En Mittal and Dahiya (1987) se propone $\nu = 4$ como valor óptimo para realizar la estimación.

La estimación modal bayesiana da como resultado los estimadores máximo verosímiles modificados (EMVM). En el artículo citado, se prueba que los EMVM existen con probabilidad uno y notablemente mejores que los EMV. Estos estimadores modificados se desarrollan en la siguiente proposición.

Proposición 1. *Sea Λ una variable aleatoria distribuida según una Normal truncada, $N_T(A,B,\mu,\sigma)$, con μ y σ desconocidos.*

Entonces, los EMVM de la distribución Normal truncada verifican el siguiente sistema de ecuaciones no lineales:
(8)

$$\begin{aligned}
\overline{\Lambda} &= \frac{\int_A^B y \exp[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2]\,dy}{\int_A^B \exp[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2]\,dy}\\
S^2+\frac{1}{r} &= \frac{2\sigma^2}{r} + \frac{\int_A^B y^2 \exp[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2]\,dy}{\int_A^B \exp[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2]\,dy} - \left(\frac{\int_A^B y \exp[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2]\,dy}{\int_A^B \exp[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2]\,dy}\right)^2
\end{aligned}$$

donde $\overline{\Lambda}$ y S son la media y la desviación típica muestrales, respectivamente, y r el tamaño muestral.

En esta proposición, se han dado estimadores para μ y σ pero no para A ni B. No existe en la literatura solución para la estimación de todos los parámetros de una distribución Normal truncada, siempre se suponen conocidos los límites de truncamiento. Nosotros proponemos escoger A y B como el mínimo y máximo de los valores de la muestra, de esta forma se hará en la aplicación. Se ha implementado la resolución del sistema de ecuaciones no lineales para la estimación de μ y σ (ecuación 8) usando el método de Newton-Raphson para su utilización en distintas aplicaciones como la presentada en este trabajo.

El método de predicción del PPDE en un instante de tiempo futuro que en este trabajo vamos a utilizar para la modelización y predicción del número de vehículos matriculados en países europeos puede ya exponerse de forma abreviada en los siguientes tres pasos.

1. Estimar los parámetros de la distribución Normal truncada perteneciente a la media del PPDE para cada $t \in [t_0, T]$, aplicando la implementación de la estimación de parámetros de la Normal truncada mencionada anteriormente.
2. Comprobar, para cada $t \in [t_0, T]$, que la media del proceso puntual se modeliza adecuadamente por la Normal truncada encontrada en el primer paso. Para ello ha sido también implementado el test de Kolmogorov-Smirnov para la bondad de ajuste de la distribución Normal truncada.
3. Una vez aceptado, que $\Lambda(t, x(t))$ es una particular Normal truncada para cada $t \in [t_0, T]$, tenemos entonces unos valores diferentes para A, B, μ and σ para cada t. Se estudia ahora cómo esos parámetros evolucionan en el tiempo estableciendo las funciones $A(t), B(t), \mu(t)$ y $\sigma(t)$. Con el conocimiento de estas funciones se puede extrapolar los parámetros de $N_T(A(t_1), B(t_1), \mu(t_1), \sigma(t_1))$, que es la distribución de $\Lambda(t_1, x(t_1))$. Esta predicción de $\Lambda(t_1, x(t_1))$ nos permite calcular la distribución del propio PPDE, así como sus características en ese instante.

Los tres pasos del método de predicción del PPDE han sido implementados en notebooks de *Mathematica 3.0*, el cual usa el método gaussiano adaptativo para calcular numéricamente las integrales que aparecen en la función de densidad de $\Lambda(t, x(t))$, el estadístico de recuento del PPDE, el sistema de ecuaciones no lineales, etc.

4. MODELIZACIÓN DEL NÚMERO DE MATRICULACIONES EN PAÍSES EUROPEOS

Como ya adelantamos anteriormente, vamos a modelizar el proceso puntual real número de matriculaciones en países europeos usando como modelo el PPDE con media aleatoria un Normal truncada. Poseemos los datos del número de matriculaciones en dieciséis países europeos desde 1979 a 1999, pero para que éstos sean comparables trabajaremos con ese número de matriculaciones por cada 10000 habitantes.

Tabla 1. N.º medio de matriculaciones acumuladas de 1979 a 1999 en España.

1979 13.9	1980 26.8	1981 37.9	1982 49.7	1983 61.8	1984 73.1	1985 85.5
1986 100.4	1987 120.2	1988 143.6	1989 168.3	1990 189.8	1991 209.4	1992 230.8
1993 247.2	1994 267.3	1995 285.7	1996 306.2	1997 329.3	1998 356.4	1999 388.1

Tabla 2. Normales truncadas estimadas para los 21 años y sus estadísticos experimentales de K-S.

Λ en el año	Distribución	D_{15}
1979	$N_T(8.3, 35.9, 22.7, 7.5)$	0.095
1980	$N_T(11.9, 70.1, 43.9, 15.3)$	0.081
1981	$N_T(16.6, 100.7, 65.3, 22.7)$	0.102
1982	$N_T(24.4, 131.3, 87.4, 29.6)$	0.079
1983	$N_T(30.7, 161.6, 111.0, 37.5)$	0.078
1984	$N_T(37.9, 192.7, 133.2, 44.3)$	0.073
1985	$N_T(47.0, 224.6, 158.6, 51.8)$	0.088
1986	$N_T(55.1, 258.8, 186.8, 61.2)$	0.082
1987	$N_T(61.2, 294.3, 214.3, 70.6)$	0.094
1988	$N_T(67.6, 514.8, 235.2, 93.7)$	0.135
1989	$N_T(76.2, 544.4, 260.9, 100.6)$	0.132
1990	$N_T(86.9, 578.4, 285.5, 107.9)$	0.131
1991	$N_T(101.3, 611.7, 307.8, 114.3)$	0.122
1992	$N_T(117.9, 645.1, 330.1, 120.7)$	0.130
1993	$N_T(130.4, 669.6, 349.3, 126.2)$	0.135
1994	$N_T(139.9, 693.7, 371.5, 132.3)$	0.137
1995	$N_T(150.7, 717.5, 393.9, 138.1)$	0.156
1996	$N_T(162.4, 742.8, 419.4, 144.3)$	0.154
1997	$N_T(175.6, 778.2, 446.0, 150.2)$	0.138
1998	$N_T(190.1, 812.5, 474.9, 156.8)$	0.137
1999	$N_T(211.1, 846.8, 504.2, 161.8)$	0.147

Fácilmente se calcula a partir de los datos las medias acumuladas por año para cada país, es decir $\Lambda(1979), \ldots, \Lambda(1999)$, con lo que preparamos los datos para realizar el primer paso del método de predicción. Los países estudiados han sido Alemania, Austria, Bélgica, Checoslovaquia, Dinamarca, España, Finlandia, Francia, Grecia, Holanda, Irlanda, Italia, Luxemburgo, Noruega, Gran Bretaña y Suecia. Mostramos como ejemplo en la Tabla 1 las medias acumuladas para España desde 1979 a 1999.

Tomando los valores de las medias acumuladas de todos los países en un cierto año, t, tenemos una muestra de 16 datos de $\Lambda(t,x(t))$, que intentaremos modelizar mediante una Normal truncada que estimamos mediante la implementación de la resolución de las ecuaciones de los EMVM (ecuación 8). La Tabla 2 muestra las Normales truncadas estimadas, $N_T(A,B,\mu,\sigma)$, para cada uno de los veintiún años.

Una vez concluido el primer paso del método propuesto en la sección anterior, cada una de estas distribuciones ha sido contrastada mediante el test de bondad de ajuste de Kolmogorov-Smirnov según el segundo de los pasos. Todas ellas daban un buen ajuste a los datos muestrales por lo que se han aceptado como estimadoras de las distribuciones de cada $\Lambda(t,x(t))$, $t \in [t_0,T]$. El estadístico experimental del test de Kolmogorov-Smirnov (D_r) para cada Normal truncada puede también observarse en la Tabla 2. Puede comprobarse que todos los estadísticos experimentales son menores que el valor crítico, 0.338, para $\alpha = 0.05$.

Una vez estimadas todas las Normales truncadas, tenemos veintiún valores «observados» de las cuatro funciones del tiempo, $A(t), B(t), \mu(t)$ y $\sigma(t)$, parámetros de la distribución de la media del PPDE. Para completar el tercer paso del método se hace una regresión para cada una de esas funciones sobre los datos que nos han proporcionado las Normales truncadas encontraremos la expresión general de la media, $\Lambda(t,x(t)) \rightsquigarrow N_T(A(t),B(t),\mu(t),\sigma(t))$. Las regresiones más adecuadas para los cuatro parámetros han resultado ser:

$$\begin{aligned} A(t) &= (2.57+0.57z)^2 \\ B(t) &= z/\left(-z\,15\cdot 10^{-5}+0.02\right) \\ \mu(t) &= z/\left(-z\,65\cdot 10^{-6}+0.04\right) \\ \sigma(t) &= z/\left(-z\,69\cdot 10^{-5}+0.13\right) \end{aligned} \tag{9}$$

donde $z = t - 1978$. Las gráficas de los ajustes pueden verse en las Figuras 1, 2, 3 y 4.

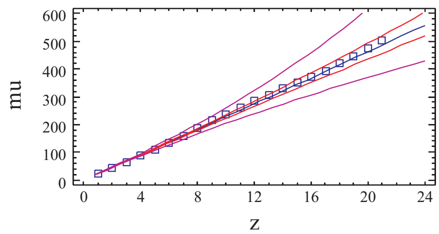

Figura 1. Regresión del parámetro A ($R^2 = 0.9972$).

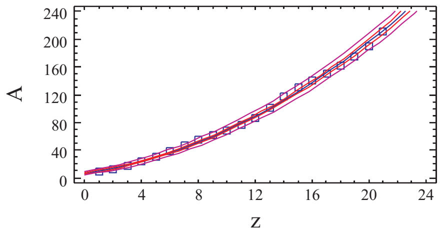

Figura 2. Regresión del parámetro B ($R^2 = 0.9948$).

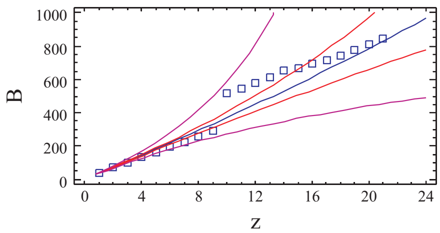

Figura 3. Regresión del parámetro μ ($R^2 = 0.9993$).

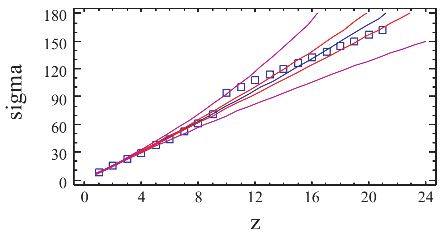

Figura 4. Regresión del parámetro σ ($R^2 = 0.9992$).

A partir del conocimiento de las funciones parámetro (9) de la Normal truncada general, se puede extrapolar a un instante futuro. Por ejemplo, se tiene que el PPDE en el año 2000 tiene una media que se distribuye mediante la Normal truncada, $N_T(229.7, 875.4, 510.4, 187.3)$. Teniendo en cuenta las ecuaciones (13) y (14) se obtiene también que la esperanza del PPDE en el año 2000 y su varianza son

$$E[N(2000)] = 524.89 \text{ y } Var[N(2000)] = 23376.7.$$

Mediante la desigualdad de Chebyschev podemos decir que la probabilidad de que el número de matriculaciones en un país Europeo esté en el intervalo $[219.102, 830.678]$ es menor o igual que 0.75. Puesto que el número medio acumulado de matriculaciones hasta 1999 fue de 481.36 y la $E[N(2000)] = 524.89$, el número de matriculaciones en el año 2000 puede estimarse en aproximadamente 43.53.

Utilizando los datos disponibles para el año 2000, hemos calculado el valor muestral del número medio de matriculaciones hasta el 2000, resultando ser 540.92. Dado que el valor de la desviación típica que obteníamos mediante el método de predicción para el PPDE en el 2000 era 152.89, consideramos que la estimación obtenida de la media (524.89) es muy precisa. Debido a esto y al hecho de que las regresiones (ecuación 9) son tan buenas y suaves podemos extender la predicción al año 2001. La distribución de $\Lambda(2001)$ es la Normal truncada $N_T(247.4, 920.7, 534.4, 197.04)$, y a partir de ella y la ecuaciones (13) y (14) se obtiene que

$$E[N(2001)] = 551.84 \text{ y } Var[N(2001)] = 25495.1.$$

AGRADECIMIENTOS

Los autores queremos expresar nuestra gratitud a los dos referees que han revisado nuestro trabajo por sus interesantes sugerencias.

Este trabajo ha sido realizado dentro del Proyecto BFM2000-1466, financiado por la Dirección General de Investigación del Ministerio de Ciencia y Tecnología.

REFERENCIAS

Cox, D. R. (1955). «Some Statistical Methods Connected with Series of Events». *J. Royal Statistical Society B.* 17, 129-164.

Daley, D. J. and Vre-Jones, D. (1988). *An Introduction to Point Processes*, Springer-Verlag, New York.

García Ferrer, A. and Queralt, R. A. (1998). «Can univariate models forecast turning points in seasonal economic time series?». *International Journal of Forecasting*, 14, 433-446.

Mittal, M. M. y Dahiya, R. C. (1987). «Estimating the parameters of a Doubly Truncated Normal Distribution ». *Communications in Statistics. Simula*, 16, 141-159.

Rodríguez Bouzas, P., Aguilera, A. M. y Valderrama, M. J. (2001). «Forecasting a Class of Doubly Stochastic Poisson Processes». *Statistical Papers. En prensa.*

Snyder, D. L. and Miller, M. I. (1991). *Random Point Processes in Time and Space*, 2nd edition. Springer Verlag, New York.

Valderrama, M. J., Jiménez, F., Gutiérrez, R. y Martínez-Almécija, A. (1995). «Estimation and Filtering on a Doubly Stochastic Poisson Process». *Aplied Stochastic Model and Data Analysis*, 11, 13-24.

ENGLISH SUMMARY

A POISSONIAN MODEL TO FORECAST VEHICULES MATRICULATIONS IN EUROPEAN COUNTRIES

M. J. VALDERRAMA
A. M. AGUILERA
P. R. BOUZAS
Universidad de Granada*

In this paper, we present a methodology to model the number of domestic car registrations in European countries by a doubly stochastic Poisson process which mean is a truncated Normal variable. On the basis of previous work about this statistical model, we extend the study of its moments. Also, we forecast the mentioned real process for 2000 and 2001.

Keywords: Doubly stochastic Poisson process, truncated Gaussian distribution, car registrations

AMS Classification (MSC 2000): 60G55, 60K30

---

* Departamento de Estadística e I.O., Facultad de Farmacia, Campus de Cartuja, Universidad de Granada, 18071 Granada.

– Received July 2001.
– Accepted November 2001.

The number of domestic passenger car registrations is one of the most representative indicators of a country economic activity. This paper presents an original methodology to model this indicator in european countries. The statistical model is a doubly stochastic Poisson process (DSPP) that is a Poisson process in which the parameter is considered to be random. We study the particular case where its mean is a truncated Normal distribution in each instant of time. We also show a forecasting method for this process.

We will note the DSPP as $\{N(t); t \geq t_0\}$, where $N(t)$ is the number of ocurrences from t_0 to t. $\{N(t); t \geq t_0\}$ is called a DSPP with intensity $\{\lambda(t, x(t)); t \geq t_0\}$ where $\{x(t); t \geq t_0\}$ is an external process iff for any sample path of $x(\cdot)$, $N(\cdot)$ is a Poisson process with intensity $\lambda(\cdot)$.

Using the conditioning method, it is obtained the probability mass function:

$$P[N(t) = n] = E\left\{P[N(t) = n/x(\sigma) : t_0 \leqslant \sigma < t]\right\} = G_{\Lambda}^{n)}(-1) \tag{10}$$

for $n = 0, 1, 2, \ldots$ where $\Lambda(t, x(t))$ is the parametric function of the DSPP, it is $\Lambda(t, x(t))$ $= \int_{t_0}^{t} \lambda(\sigma, x(\sigma))\, d\sigma$ and $G_{\Lambda}^{n)}(s)$ is the n$-th$ derivative of the moment generating function of $\Lambda(\cdot)$. Using again the conditioning method, the characteristic function of $N(\cdot)$ results

$$M_{N(t)}(iu) = E\left\{\exp\left[\left(e^{iu} - 1\right) \Lambda(t, x(t))\right]\right\} = M_{\Lambda}(e^{iu} - 1) \tag{11}$$

where $M_{\Lambda}(iu)$ is the characteristic function of $\Lambda(\cdot)$. The evaluation of $M_{N(t)}(iu)$ is habitually difficult if not imposible.

Studying different examples of counting processes, it usually seems that their means could be modeled by a Gaussian model but as $\Lambda(\cdot)$ has to be a non-decreasing function equal or greater than zero, we have considered a truncated Normal variable for each instant of time. We will note that $\Lambda(\cdot)$ is distributed as a truncated Normal like $\Lambda(t, x(t)) \rightsquigarrow N_T(A(t), B(t), \mu(t), \sigma(t))$. In this case, equation (10) becomes

$$P[N(t) = n] = \frac{1}{n!\sigma\sqrt{2\pi}} \left(P\left[\frac{A-\mu}{\sigma} < Z < \frac{B-\mu}{\sigma}\right]\right)^{-1} \int_{A}^{B} \Lambda^n e^{-\Lambda} e^{-\frac{1}{2}\left(\frac{\Lambda-\mu}{\sigma}\right)^2} d\Lambda$$

for $n = 0, 1, 2, \ldots$ and A, B, μ and σ are the parameters of $\Lambda(\cdot)$, all of them functions of time with $0 \leq A < \Lambda < B$, $\mu \in \mathbb{R}$ and $\sigma \geq 0$.

It can also be shown that equation (11) has the form

$$M_{N(t)}(iu) = C(iu, t)\, M_{\xi}\left(e^{iu} - 1\right) \text{ with} \tag{12}$$

$$C(iu, t) = \frac{P\left[A - (e^{iu} - 1)\sigma^2 < \xi < B - (e^{iu} - 1)\sigma^2\right]}{P[A < \xi < B]}$$

where $\xi \rightsquigarrow N(\mu,\sigma)$, μ and σ functions of time. This interpretation of $M_{N(t)}(iu)$ allows us to calculate the mean and variance of the process. Due to (12) and after some manipulations, we have that

$$E[N(t)] = E[\Lambda(t)] = \frac{\sigma}{\sqrt{2\pi}} \frac{f(B)-f(A)}{P[A<\xi<B]} + \mu \tag{13}$$

$$Var[N(t)] = \frac{\sigma[(A-\mu+1)\ f(A)-(B-\mu+1)\ f(B)]}{\sqrt{2\pi}P[A<\xi<B]} -$$

$$\left[\frac{\sigma[f(B)-f(A)]}{\sqrt{2\pi}P[A<\xi<B]}\right]^2 + \sigma^2 + \mu \tag{14}$$

where again $\xi \rightsquigarrow N(\mu,\sigma)$ and $f(x) = \exp[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2]$.

In this paper, we propose a method for estimating $N(\cdot)$ in t_1, known r sample paths of the counting process in $[t_0,T]$, $T<t_1$.

The forecasting method can be exposed now in these three steps:

1. Estimate the truncated Normal distributions of the means of the DSPP for each $t \in [t_0,T]$ using the modified maximum likelihood estimators.
2. Prove, for each $t \in [t_0,T]$, the truncated Normal estimated in Step 1 fits the data by a Kolmogorov-Smirnov test.
3. Once accepted all the truncated Normals in Step 2, it should be studied the evolution in time of A, B,μ and σ, stablishing the functions $A(t)$, $B(t)$, $\mu(t)$ and $\sigma(t)$. These functions are used to forecast $\Lambda(t_1,x(t_1))$ and so, $N(t_1)$.

The model studied has been applied to the number of domestic passenger car registrations per 10000 inhabitants in sixteen european countries from 1979 to 1999. The countries are Germany, Austria, Belgium, Checoslovaquia, Denmark, Spain, Finland, France, Greece, Holland, Ireland, Italy, Luxemburg, France, Great Britain and Sweede.