

# Estudio de la frecuencia de ocurrencia de accidentes de tráfico mediante procesos estocásticos de Pascal-Pólya

---

D. Sánchez-Molina<sup>1</sup>, S. García-Vilana<sup>1</sup>, J. Velázquez-Ameijide<sup>1</sup>, C. Arregui-Dalmases<sup>2</sup>

<sup>1</sup>UPC, EEBE-GRABI, Barcelona

<sup>2</sup>Centro Zaragoza, Pedrola

## Resumen

*Objetivos:* (1) Plantear un modelo estocástico adecuado para la frecuencia de accidentes, que permita analizar tanto la peligrosidad, como la exposición al riesgo. (2) Introducir los conceptos de macroaccidentalidad y microaccidentalidad.

*Métodos:* En este artículo se han usado la teoría de procesos estocásticos para estimar el número de accidentes y así como la frecuencia con la que se producen accidentes de tráfico de una red viaria amplia (macroaccidentalidad). Para ello, se han analizado diversas bases de datos oficiales de una región durante un período prolongado de casi una década.

*Resultados:* El principal hallazgo es que un proceso estocástico de Poisson mixto (mezcla probabilista de varios procesos de Poisson) ajusta adecuadamente el número de accidentes, dando lugar a una distribución binomial negativa para el número de accidentes en un período específico. Se propone una interpretación estocástica explicativa de las razones por las cuales la distribución del número de accidentes es de ese tipo.

**Palabras clave:** Distribución binomial negativa, Proceso de Poisson, Accidentes de tráfico

## Abstract

*Objectives:* (1) To propose a stochastic model suitable for the frequency of accidents, which allows to analyze both dangerousness and risk exposure, and (2) Introduce the concepts of aggregate collision analysis (collision macroanalysis) and specific collision rate (collision microanalysis).

*Methods:* In this article, stochastic process theory has been used to estimate the number of traffic collisions and how often traffic collisions occur on a wide road network (collision macroanalysis). To this end, different official databases from the same area have been analyzed over an extended period, covering nearly a decade.

*Results:* The main finding is that a mixed Poisson stochastic process (probabilistic mixture of various Poisson processes) adequately represents the number of traffic collisions, resulting in a Negative Binomial Distribution for the number of collisions in a specific period. An explanatory stochastic interpretation is proposed for explaining why the number of collisions follows such a distribution.

**Keywords:** Negative Binomial Distribution, Poisson Process, Traffic Collision

---

Correspondencia:

David Sánchez.

E-mail: david.sanchez-molina@upc.edu

DOI: 10.5821/sibb.26.1.8765

## Introducción

A nivel mundial, cada año alrededor de 1,25 millones de personas fallecen en accidentes de tráfico y entre 20 y 50 millones más sufren traumatismos de cierta consideración [1]. En muchos casos estas lesiones derivan en una discapacidad adquirida. La misma fuente (OMS) destaca que los accidentes de tráfico son ya la principal causa de muerte en personas de entre 5 y 29 años a nivel mundial. No obstante, debe señalarse que las tasas de mortalidad varían mucho según la región: así, el número de fallecidos en accidente de tráfico por cada 100 mil habitantes es 26,6 en África, mientras que para Europa es sólo de 9,3 por cada 100 mil.

Más específicamente en el contexto europeo, se observa que en España en 2016 se produjeron 1,85 millones de accidentes de tráfico al año, de acuerdo con los datos oficiales registrados [2]. En ese mismo año, la cifra de fallecidos por accidente de tráfico fue 1160 y el número de heridos graves fue de 140 mil.

Las cifras anteriores muestran por sí mismas la importancia epidemiológica, social y económica que los accidentes de tráfico representan y la razón por la cual es importante estudiar las causas de la accidentalidad por regiones, mediante modelos estadísticos fiables que modelicen la frecuencia de ocurrencia de los accidentes de tráfico (macroaccidentalidad) y como se relaciona esta ocurrencia con las características peculiares de conductores, conductas al volante, tecnología de los vehículos, estados de la vía y situaciones particulares de puntos específicos de la red (microaccidentalidad).

En cuanto a la modelización de la frecuencia del número de accidentes se han intentado diversas metodologías. Algunos autores han encontrado el mismo resultado principal que aquí se expone: que el número previsto de accidentes en muchas situaciones puede venir representado por una distribución binomial negativa, la mayor parte de trabajos a partir de mediados de los años 1990 usan ese tipo de distribución para la modelización [3,4]. Otros estudios previos se habían basado en la distribución de Poisson [5,6], aunque esos modelos

parecen haber sido abandonados en gran medida, ya que los estudios recientes con datos mejores encuentran evidencias en favor del uso de un modelo binomial negativo [7,8,9]. En estos últimos trabajos prácticamente la única razón aportada para el modelo binomial negativo es que proporciona un mejor ajuste, pero no se incide en una explicación teórica de por qué cabría esperar ese tipo de distribución, asunto que en este artículo trata de aclararse desde el punto de vista de la teoría de procesos estocásticos.

El artículo se organiza de la siguiente manera: en la sección *Conceptos previos* se aclaran las nociones de peligrosidad, exposición al riesgo, macroaccidentalidad y microaccidentalidad y en las secciones *Procesos estocásticos de Poisson y distribuciones compuestas* y *Proceso de Poisson mixto y distribución binomial negativa* se aporta el núcleo teórico del artículo donde se toma el punto de vista de los procesos estocásticos. La sección *Bases de datos* explica la estructura de la base de datos que se usó para ilustrar el análisis. En la sección *Peligrosidades y Exposiciones al riesgo específicas* se dan las peligrosidades y exposiciones al riesgo bajo diversas circunstancias donde se aprecia qué factores temporales afectan a la frecuencia, y en la sección *Análisis de la evolución de la peligrosidad* se analiza la variación de la peligrosidad según situaciones específicas. En la sección *Discusión* se discute la importancia del resultado y se discuten otros enfoques alternativos que si bien proporcionan buenos ajustes numéricos carecen de profundidad teórica, y se argumenta por qué los procesos estocásticos si tienen un valor explicativo a diferencia de otros enfoques.

## Datos y Métodos

### *Conceptos previos*

El número de accidentes varía en función de la época del año y los días de la semana, siendo los períodos vacacionales los de menor siniestralidad y los viernes los días de mayor accidentalidad. Estas variaciones particulares tienen que ver con la intensidad del tráfico que resulta ser menor en períodos vacacionales y

fin de semana, por lo que cualquier análisis debe considerar esa variación estacional dentro del año y de la semana. En esta sección seguiremos la terminología establecida en [10].

Para representar la siniestralidad agregada, número de accidentes de cada tipo, se recurre a definir la **peligrosidad**. Esta peligrosidad es una medida numérica directamente relacionada con la probabilidad de ocurrencia de accidentes, por tanto, este parámetro tendrá un valor instantáneo más alto cuando más accidentes se producen; de hecho, una medida adecuada de la peligrosidad debe satisfacer dos condiciones:

1. Debe ser calculable a partir de la distribución del número de accidentes producidos bajo una situación concreta (año, mes, día de la semana, tipo de vehículo, etc.)

2. Debe dar una idea de la distribución de probabilidad de los accidentes cuando la propia peligrosidad aumenta o disminuye.

Por tanto, para construir una media de peligrosidad se necesita estudiar las distribuciones de probabilidad del número de accidentes totales, lo cual da una *peligrosidad general*  $\pi_0$ . La *peligrosidad específica*  $\pi_e$  es el resultado de modificar la peligrosidad general mediante modificadores de macroaccidentalidad  $\varphi_i$ , usando una relación del tipo:

$$\pi_e = \pi_0 (1 - \varphi_a) (1 - \varphi_m) (1 - \varphi_d) (1 - \varphi_h) \quad (1)$$

donde  $\varphi_a$  es un factor que modifica la peligrosidad general que modifica la peligrosidad según el año,  $\varphi_m$  según el mes del año,  $\varphi_d$  es la peligrosidad debida al día de la semana y  $\varphi_h$  debida a la franja horaria. Calculando las peligrosidades específicas por meses, días y franjas horarias es posible determinar sin dificultad las  $\varphi_i$ . En este estudio usamos como medida de la peligrosidad el valor esperado condicionado a ciertas circunstancias del conjunto  $\Sigma$  (mes, día, hora), es decir  $\pi_e = \mathbb{E}(N_T) < \text{var}(N_t)$

Las medidas de peligrosidad son interesantes, pero ocultan un factor importante que es la varianza del número de accidentes, por usar

únicamente la peligrosidad de una situación concreta puede ser engañoso si no se considera que situaciones de igual o muy parecida peligrosidad pueden presentar varianzas muy diferentes y, por tanto, existe la posibilidad de que en un período concreto el número de accidentes observados se aleje mucho del número estimado sólo a partir de la peligrosidad. Por esa razón se introduce la noción de **exposición al riesgo** que aquí definiremos como una combinación lineal de peligrosidad y desviación estándar:

$$ER_T = \mathbb{E}(N_t) + \alpha \sigma(N_t) \quad (2)$$

donde  $N_t$  es la variable aleatoria “número de accidentes”,  $\mathbb{E}(N_t)$  es el valor esperado del número de accidentes (que es función de la peligrosidad),  $\sigma(N_t)$  es la desviación tipo del número de accidentes y  $\alpha$  es un parámetro real arbitrario, aquí hemos considerado  $\alpha=1,65$ .

Este enfoque concentrado en las cifras agregadas de accidentes de todo tipo en una red viaria amplia, se conoce como *macroaccidentalidad* donde se hace abstracción de circunstancias específicas de los accidentes en particular de las características de los conductores y los puntos de la vía, mientras que el ajuste de la peligrosidad específica ajustada según el punto de la vía, poniendo atención al tipo de maniobras, perfiles de conductores y situaciones idiosincráticas del accidente se conoce como *microaccidentalidad* (la terminología comparte algunas de las características que diferencian el estudio agregado de cuestiones económicas o *macroeconomía* y el estudio pormenorizado de las decisiones de los agentes económicos en situaciones específicas o *microeconomía*).

#### *Procesos estocásticos de Poisson y distribuciones compuestas*

La teoría de *procesos estocásticos* se ocupa de *variables aleatorias que “evolucionan” en el tiempo*. Dada una magnitud aleatoria dependiente del tiempo es conveniente considerar una sucesión continua de variables aleatorias sobre el mismo espacio de probabilidad

$(\Omega, \mathcal{F}, \mathbb{P})$ , es decir, consideramos un conjunto indexado por el tiempo  $\mathcal{X} = \{X_t \mid t \geq 0, X_t: \Omega \rightarrow \mathbb{R}\}$ . El número de accidentes  $N_t$  producidos en el intervalo  $(0, t]$  es un ejemplo de proceso estocástico, ya que para un tiempo  $t$  dicho número es una variable aleatoria y además se cumple que con probabilidad uno,  $N_t$  es una magnitud no-decreciente.

Existen muchos tipos de procesos estocásticos, en función de que los eventos representados sean independientes o no, se distribuyan uniformemente en el tiempo o no, etc. Tal vez el tipo más simple de proceso estocástico sea un proceso de conteo (por tanto, el resultado será siempre un número natural) y donde la aparición de un nuevo evento (o accidente) es independiente de la ocurrencia de los anteriores. Los accidentes de tráfico parecen ser de ese tipo, dado lo esporádico de los mismos la ocurrencia de un accidente en una vía un martes, no afecta significativamente a la probabilidad de que haya otro accidente en el mismo punto o uno cercano una semana después (es decir, podemos pensar que los accidentes no están causalmente relacionados entre sí, lo cual es un buen punto de partida). Si las condiciones fueran siempre homogéneas (clima, intensidad de tráfico, tipo de conductor, mismo día de la semana, etc.), sería razonable pensar que el número de accidentes viene dado por un proceso de Poisson, cuyo nombre se debe a que al estudiar un período amplio  $T$  el número de accidentes se ajusta a una distribución de Poisson del tipo:

(3)

$$p_n(T) = \mathbb{P}(N_T = n) = \dots$$

$$\dots = \mathbb{P}(\{\omega \in \Omega \mid N(T, \omega) = n\}) = \frac{(\lambda T)^n}{n!} e^{-\lambda T}$$

donde  $\lambda$  resulta ser la probabilidad por unidad de tiempo de ocurrencia de un accidente, que puede identificarse con la peligrosidad general. Puede demostrarse que esta distribución surge de suponer que el número de accidentes  $N_t$  viene definido por un *proceso estocástico de Poisson*  $N_t = N(t, \omega)$  que se basa en las asunciones siguientes [11]:

1. El proceso empieza en cero  $N_0 = N(0, \omega) = 0$ , para todo  $\omega$ .

2. El proceso tiene incrementos independientes, es decir, para una sucesión cualquiera de instantes  $0 < t_0 < t_1 < \dots < t_n$ , el número [aleatorio] de eventos en cada intervalo  $(t_{i-1}, t_i]$  es una variable aleatoria denotada como  $V_i = N(t_{i-1}, t_i]$ , resultando que para todo  $1 \leq i \leq n$  estas variables son mutuamente independientes (desde un punto de vista probabilista).

3. Si se define  $N_t := N(0, t)$ , entonces el límite  $\lim_{\Delta t \rightarrow 0} \mathbb{P}[N_{t+\Delta t} - N_t \mid N_t, \tau < t] / \Delta t$  está bien

definido y describe la probabilidad por unidad de tiempo de que se produzca un nuevo evento  $d\mathbb{P}_t / dt = \lambda_t$

4. Con probabilidad uno, los caminos  $(N(t, \omega))_{t \geq 0}$  son continuos por la derecha y admiten límites por la izquierda para  $t > 0$ , esta condición define un proceso càdlàg (continue à droite, limites à gauche).

El proceso se llama homogéneo en el tiempo si  $\lambda_t = \lambda_0$ , es decir, la tasa de aparición de nuevos eventos es constante. La figura 1 presenta un ejemplo de proceso de Poisson y su probabilidad asociada (Figura 1).

En los primeros tiempos de la investigación de la frecuencia se empleó este tipo de proceso ya que el correlato físico de las condiciones anteriores es bastante razonable para la producción de accidentes de tráfico (si suponemos una población homogénea de conductores y vehículos, circulando por una red sin variaciones notorias de las condiciones de tráfico). Sin embargo, muy rápidamente la evidencia empírica mostró que la distribución de Poisson dada por la ecuación (3) no representaba adecuadamente los datos, ya que según esa distribución el valor esperado  $\mathbb{E}(N_t)$  y la varianza  $\text{var}(N_t) = \mathbb{E}(N_t^2) - \mathbb{E}(N_t)^2$  para una distribución así coinciden  $\mathbb{E}(N_t) = \text{var}(N_t) = \lambda$  mientras que los datos experimentales mostraban una sobre-dispersión respecto a lo esperado, es decir,  $\mathbb{E}(N_t) < \text{var}(N_t)$ . Eso es lo que llevó a buscar otro tipo de distribuciones que ajustaran mejor los datos. Por ejemplo, es conocido que

una distribución de probabilidad compuesta donde se usa una distribución de cierto tipo, pero donde los parámetros que la definen se consideran en lugar de fijos distribuidos aleatoriamente puede resolver el problema de ajustar mejor unos datos que presenten sobredispersión respecto a la distribución en la que se basa el modelo. Es decir, si se tiene una variable aleatoria  $X$  distribuida de acuerdo a cierta distribución  $X \sim F_\theta$  donde los parámetros  $\theta$  que caracterizan dicha distribución se asume que están distribuidos aleatoriamente  $\theta \sim G$  entonces la distribución compuesta por  $F$  y  $G$  dará lugar a una distribución dada por:

$$p_{F*G}(x) = \int p_F(x|\theta)p_G(\theta) d\theta \quad (4)$$

La distribución resultante, para distribuciones  $G$  suficientemente concentradas, será similar a la distribución  $F$  pero tendrá *sobredispersión* ya que la varianza de la nueva distribución será:

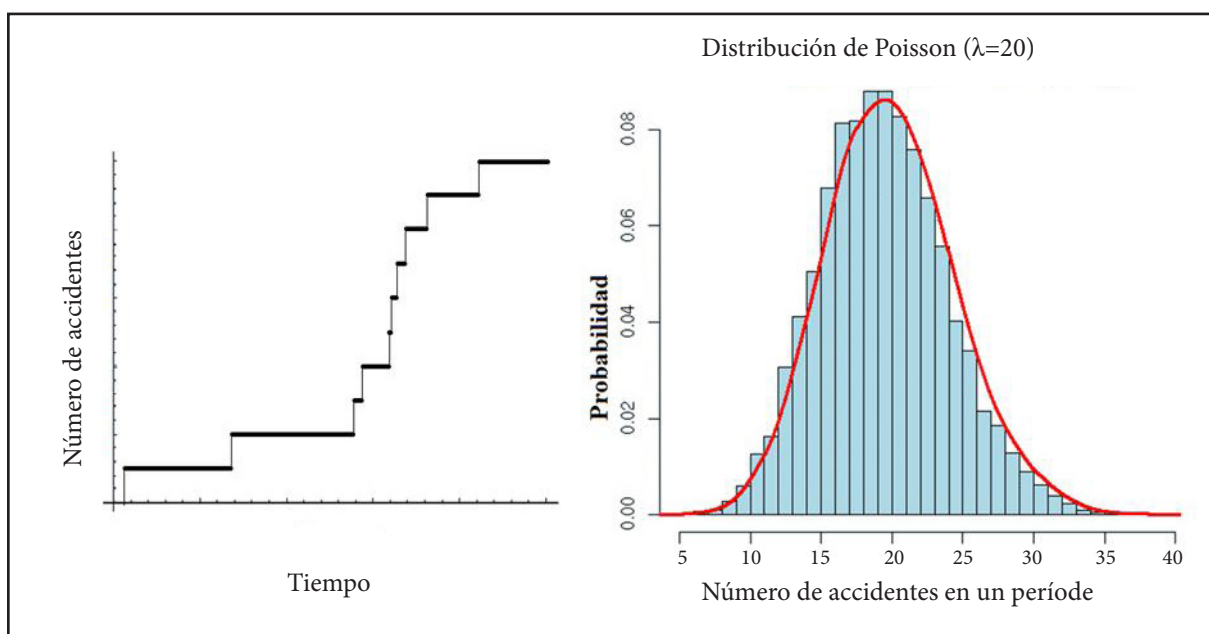
$$\begin{aligned} \text{var}_{F*G}(X) &= \mathbb{E}_G\left(\text{var}_F(X|\theta)\right) + \dots \\ \dots + \text{var}_G\left(\mathbb{E}_F(X|\theta)\right) &> \text{var}_F(X|\theta) \end{aligned} \quad (5)$$

La mayor parte de estudios recientes sobre frecuencias de accidentes de tráfico han usado una *distribución binomial negativa* [7,8,9], que corresponde a considerar una distribución auxiliar  $G$  que sea una *distribución gamma* cuya densidad de probabilidad viene dada por:

$$p_G(\theta) = \frac{\beta^r}{\Gamma(r)}\theta^{r-1}e^{-\beta\theta} \quad (6)$$

Y algunos autores han llegado a usar una modificación conocida como *distribución binomial negativa-Lindley* [12,13] que es una generalización de la binomial negativa que pertenece a la *familia Lindley de distribuciones* descrita en [14]. Sin embargo, la mayor parte de estos trabajos simplemente pretenden salvar el problema de sobredispersión dar otra justificación de la distribución para la frecuencia de accidentes que el buen ajuste empírico a una distribución. En la siguiente sección se da una justificación teórica por la cual la distribución binomial negativa da un buen ajuste, a partir de la noción la noción de *proceso de Poisson mixto*.

A modo de resumen de esta sección se tiene que, aunque los procesos estocásticos de Poisson cualitativamente si dan una idea del



**Figura 1.** Izquierda: número de sucesos (accidentes) generados aleatoriamente por un proceso de Poisson homogéneo. Derecha: Distribución de probabilidad para un periodo fijo dada por un proceso de Poisson homogéneo.

ritmo al que se producen accidentes, cuantitativamente, como se ha discutido anteriormente no reproducen bien la sobredispersión de los datos observados. Por otra parte, aunque las distribuciones compuestas como la binomial negativa reproducen bien la sobredispersión, la justificación teórica de la misma en la literatura ha sido insuficientemente tratada. En la siguiente sección se pretender resolver ambas deficiencias con una justificación teórica más sólida.

#### Proceso de Poisson mixto y distribución binomial negativa

Un proceso de Poisson mixto  $\{\tilde{N}(t) \mid t > 0, \tilde{N}(t) = N(t, \omega) \in \mathbb{N}\}$  es un tipo de proceso estocástico más general donde la distribución de conteo viene dada por [15]:

$$(7) \quad p_n(t) = \mathbb{P}(\{N_T = n\}) = \int_0^\infty \frac{(\lambda t)^n}{n!} e^{-\lambda t} d\mu(\lambda)$$

donde  $\mu(\lambda) := \mathbb{P}[0, \lambda]$  es la llamada *distribución de estructura*. Claramente, un proceso de Poisson homogéneo, es un caso particular de (7), para verlo basta considerar una medida  $\mu(\lambda)$  expresable en términos de la *delta de Dirac*  $d\mu(\lambda) = \delta(\lambda - \lambda_0) d\lambda$ . Varios procesos estocásticos interesantes pueden ser representados de esta manera, para diferentes elecciones de la medida  $d\mu(\lambda)$ , en particular escogiendo una medida asociada a la *distribución gamma*:

$$(8) \quad d\mu(\lambda) = \frac{\beta^r}{\Gamma(r)} \lambda^{r-1} e^{-\beta\lambda} d\lambda$$

que depende de dos parámetros  $r \in \mathbb{N}$  y  $\beta \in \mathbb{R}^+$ . En este caso el proceso de Poisson mixto se puede entender como una mezcla de procesos de Poisson en los que cada uno tiene un peso diferente dado por  $d\mu(\lambda)$ . Así, la accidentalidad en condiciones de lluvia es diferente, así como la accidentalidad en condiciones de poca visibilidad o la accidentabilidad de los conductores noveles respecto a los experimentados, el peso de cada tipo de condición es lo que la medida  $d\mu(\lambda)$  tiene en cuenta. Para la medida específica (8) la integral de la ecuación (7) puede llevarse a cabo de manera explícita resultando la forma cerrada de una binomial negativa:

$$(9) \quad \mathbb{P}(\{N_T = n\}) = \frac{\Gamma(r+n)}{\Gamma(r)n!} \left(\frac{t}{t+\beta}\right)^n \dots$$

$$\dots \left(\frac{\beta}{t+\beta}\right)^r = \binom{n+r-1}{n} p^n (1-p)^r$$

donde el parámetro  $p = t / (t + \beta)$ . El proceso estocástico mixto dado por (7) y (8) se denomina *proceso de Pascal* [15] o *proceso de Pólya* [16,17].

Con el fin de hacer comparaciones con la distribución de Poisson se define el parámetro real  $\rho = r/T$ , donde  $r$  es el parámetro entero que aparece en la distribución (8). En términos de este parámetro el número de accidentes durante un período  $T$  viene dado por:

$$(10) \quad p_n(T) = \frac{\Gamma(\rho T + n)}{\Gamma(r)n!} p^n (1-p)^{\rho T}$$

Esta distribución vendrá dada por tanto por dos parámetros reales  $0 < p < 1$  y  $\rho$  que se pueden obtener mediante estimación de máxima verosimilitud a partir de los datos empíricos de la distribución de accidentes. Además, para la binomial negativa, el valor esperado y la varianza son:

$$(11) \quad \mathbb{E}(N_T) = \frac{p}{1-p} \rho T$$

$$\text{var}(N_T) = \frac{p}{(1-p)^2} \rho T$$

Obsérvese que ahora para el proceso de Pascal-Pólya se tiene  $\text{var}(N_t)/\mathbb{E}(N_t) > 1$ , mientras que para un proceso de Poisson se tenía  $\text{var}(N_t)/\mathbb{E}(N_t) = 1$ , lo cual permite dar cuenta de la sobredispersión observada en los datos.

Además, la peligrosidad y la exposición al riesgo estimadas por un proceso de Poisson, y un proceso de Pascal-Pólya pueden ser comparadas con facilidad. Si se compara la distribución binomial  $BN(p, \rho)$ , dada por (10), con una distribución de Poisson  $(\lambda)$ , dada por (3), del mismo valor esperado se tiene la relación  $p = \lambda / (\rho + \lambda)$ , por lo que la tasa de accidentes por unidad de tiempo o peligrosidad de ambos procesos resultan ser iguales:

$$(12) \quad \pi_{\text{Poi}} = \lambda = \frac{p}{1-p} \rho = \pi_{\text{BN}}$$

Para las exposiciones al riesgo, la sobredispersión del proceso de Pascal-Pólya da estimaciones algo mayores ( $ER_{\text{BN}} > ER_{\text{Poi}}$ ):

$$ER_{\text{Poi}} = \lambda \left( 1 + \frac{a}{\lambda} \right) = \frac{p\rho}{1-p} \left( 1 + \frac{a(1-p)}{\sqrt{p\rho}} \right) \dots$$

$$(13) \quad \dots < \frac{p\rho}{1-p} \left( 1 + \frac{a}{\sqrt{p\rho}} \right) = ER_{\text{BN}}$$

#### Bases de datos

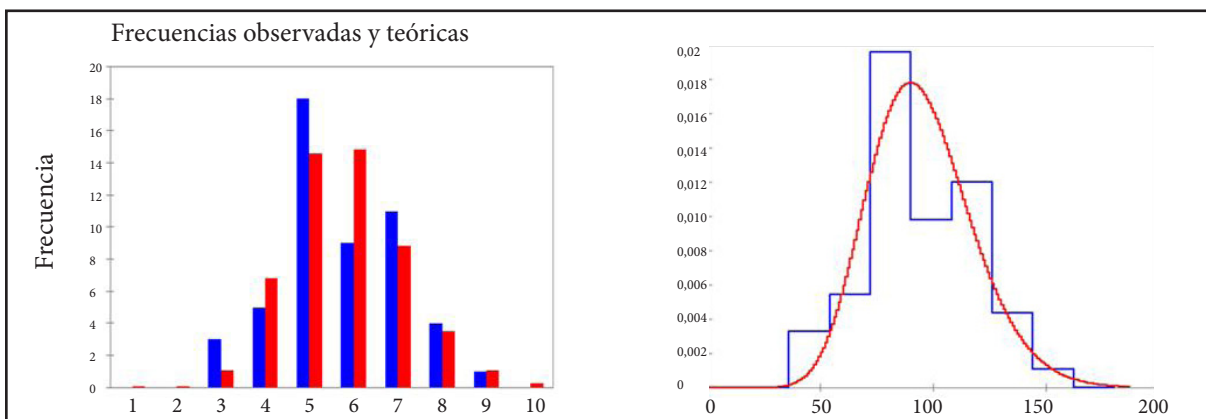
Para el estudio empírico se utilizó la base de datos SIDAT, mantenida por el Servei Català del Transit (SCT), organismo dependiente del gobierno autónomo de Cataluña. Esta es una base extensa que contiene 132 variables y está enfocada a la descripción exhaustiva de cada accidente registrado. Entre estas variables encontramos la localización del accidente, las características de la vía, la descripción de las unidades, además de la fecha, día y hora del accidente. También dependiendo de la gravedad del accidente se recogen detalles complementarios. En este estudio se consideraron todos los accidentes registrados entre 2007 y 2015 sobre la parte de la red viaria que es competencia del SCT. El número de registros encontrados en ese período es de 267 mil accidentes registrados, lo cual da más de 33 mil accidentes por año, que constituye una buena base para cualquier estudio estadístico.

## Resultados

De los datos contenidos en la base de datos SIDAT, para el área de Cataluña se concluye que los meses más peligrosos son el período de mayo a julio y el período de noviembre a diciembre. El día más peligroso es el viernes y el de menos peligrosidad el domingo (esto claramente parece estar relacionado con el número de desplazamientos). En la muestra analizada, debido a la crisis económica española (2008-2014) se aprecia en la primera parte de ese período de 2008, 2009 y 2010 un número de accidente menor que la media del período 2007-2015. La franja horaria de entre las 11:00 y las 15:00 es la que concentra mayor número de accidentes.

#### Peligrosidades y Exposiciones al riesgo específicas

En esta sección se calcula la peligrosidad general por año y las peligrosidades específicas por mes y día de la semana, usando un modelo basado en procesos estocásticos. Como se hizo notar en la ecuación (12), las peligrosidades estimadas según un proceso de Poisson o un proceso de Pascal-Pólya resultan ser idénticas, no así la exposición al riesgo que se muestra en las tablas. Esto se hizo comparando las frecuencias observadas con las distribuciones de Poisson y binomial negativa, en todos los casos se realizó un test  $\chi^2$  para la bondad del ajuste. En general dicho test no fue superado por el proceso de Poisson, pero en todos los casos el test  $\chi^2$  fue superado por el proceso de Pascal-Pólya que lleva a la binomial negativa. La figura 2 muestra un mejor ajuste estimado mediante el paquete estadístico XI-Stat®.



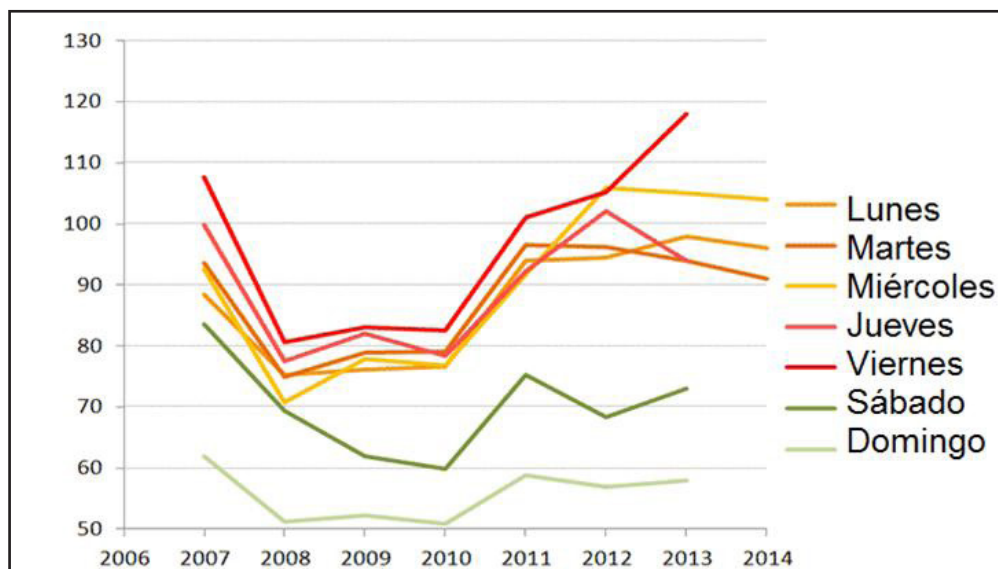
**Figura 2.** Izquierda: Número de accidentes observados y predichos por una binomial negativa. Derecha: Comparación entre observaciones y la envolvente de una binomial negativa.

Las peligrosidades generales por año y por mes se dan en la tabla 1, mientras que las peligrosidades por día de la semana en la figura 3.

#### Análisis de la evolución de la peligrosidad

En esta sección se usan las peligrosidades

generales y específicas de la sección anterior para entender cómo ha ido evolucionando la macroaccidentalidad general en Cataluña en el período (2007-2015). Esto se hace introduciendo los *modificadores de macroaccidentalidad específicos*  $\varphi_i$  que se definen para cada



**Figura 3.** Peligrosidades calculadas para cada día de la semana, la figura muestra tan la mayor peligrosidad de los viernes, como la variación de la peligrosidad en años de mayor desempleo.

		Peligrosidad específica		Exposición al riesgo específica	
		Poisson y Binomial negativa	Poisson	Poisson	Binomial negativa
Año	Peligrosidad general				
	Exposición al riesgo				
		Poisson y Binomial negativa	Poisson	Poisson	Binomial negativa
2007		87,1	93,0	100,3	104,8
2008		71,7	76,9	103,4	83,8
2009		72	77,3	99,2	83,1
2010		72,4	77,7	102,5	88,0
2011		86,2	92,0	102,7	101,4
2012		89,8	95,7	103,6	104,4
2013		90,5	96,5	82,8	108,9
2014		85,3	91,1	95,3	107,3
2015		88,0	93,8	101,8	110,4
Todos		81,2	96,1	97,3	128,2
		Peligrosidad específica		Exposición al riesgo específica	
Mes		Poisson y Binomial negativa	Poisson	Poisson	Binomial negativa
Enero		85,1	100,3	100,3	115,7
Febrero		87,9	103,4	103,4	110,9
Marzo		84,1	99,2	99,2	103,9
Abril		81,7	96,6	96,6	117,8
Mayo		87,1	102,5	102,5	118,6
Junio		87,3	102,7	102,7	114,4
Julio		88,1	103,6	103,6	110,9
Agosto		69,1	82,8	82,8	107,4
Sept.		80,5	95,3	95,3	110,0
Octubre		85,5	100,8	100,8	111,0
Noviem.		86,5	101,8	101,8	118,1
Diciem.		82,3	97,3	97,3	153,2

**Tabla 1.** Peligrosidades calculadas para cada año y exposición al riesgo. Los años de mayor desempleo se registró un descenso del número de vehículos circulantes y del número de accidentes.



año, cada mes y cada día en la forma sugerida por la ecuación (1). La ventaja de estos factores es que cuando son positivos indican que esa situación presenta una peligrosidad específica mayor que la general y cuando son negativos indican una disminución del riesgo. Más específicamente usaremos los modificadores de macroaccidentalidad adoptados a la estructura de los datos, más específicamente cumplirán las siguientes condiciones:

$$(14) \begin{cases} \pi_a = \pi_0(1 - \varphi_a) \\ \pi_m = \pi_0(1 - \varphi_a)(1 - \varphi_{am}) \\ \pi_d = \pi_0(1 - \varphi_a)(1 - \varphi_{ad}) \end{cases}$$

Donde  $\pi_a$ ,  $\pi_m$  y  $\pi_d$  son las peligrosidades específicas consideradas en la sección anterior. En cuanto a los modificadores se tiene que  $\varphi_a$  es el factor de modificación para el año,  $\varphi_{am}$  para el mes y  $\varphi_{ad}$  para el día, más específicamente si se usa la teoría de procesos estocásticos para calcular las peligrosidades específicas estos factores se pueden calcular simplemente como:

$$(15) \quad \varphi_a = \frac{\pi_a}{\pi_0} - 1, \quad \varphi_{am} = \frac{\pi_{am}}{\pi_a} - 1, \quad \varphi_{ad} = \frac{\pi_{ad}}{\pi_a} - 1$$

Estos modificadores de accidentalidad pueden ser calculados para cada año y cada situación específica y se presentan en la tabla 2.

<i>Factor de modificación específica de la peligrosidad específica según el año</i>									
<b>Año</b>	<b>2007</b>	<b>2008</b>	<b>2009</b>	<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>	<b>2015</b>
$\varphi_a$	0,073	-0,117	-0,113	-0,108	0,062	0,106	0,115	0,050	0,084
<i>Factor de modificación específica de la peligrosidad específica según el mes</i>									
$\varphi_{am}$	<b>2007</b>	<b>2008</b>	<b>2009</b>	<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>	<b>2015</b>
<b>Enero</b>	0,024	-0,157	-0,154	-0,149	0,013	0,055	0,063	0,002	0,034
<b>Febrero</b>	-0,009	-0,184	-0,181	-0,176	-0,019	0,022	0,030	-0,030	0,001
<b>Marzo</b>	0,036	-0,147	-0,144	-0,139	0,025	0,068	0,076	0,014	0,046
<b>Abril</b>	0,066	-0,122	-0,119	-0,114	0,055	0,099	0,108	0,044	0,077
<b>Mayo</b>	0,000	-0,177	-0,173	-0,169	-0,010	0,031	0,039	-0,021	0,010
<b>Junio</b>	-0,002	-0,179	-0,175	-0,171	-0,013	0,029	0,037	-0,023	-0,008
<b>Julio</b>	-0,011	-0,186	-0,183	-0,178	-0,022	0,019	0,027	-0,032	-0,001
<b>Agosto</b>	0,260	0,038	0,042	0,048	0,247	0,300	0,310	0,234	0,274
<b>Septiembre</b>	0,082	-0,109	-0,106	-0,101	0,071	0,116	0,124	0,060	0,093
<b>Octubre</b>	0,019	-0,161	-0,158	-0,153	0,008	0,050	0,058	-0,002	0,029
<b>Noviembre</b>	0,007	-0,171	-0,168	-0,163	-0,003	0,038	0,046	0,014	0,017
<b>Diciembre</b>	0,058	-0,129	-0,125	-0,120	0,047	0,091	0,100	0,036	0,069
<i>Factor de modificación específica de la peligrosidad específica según el día</i>									
$\varphi_{ad}$	<b>2007</b>	<b>2008</b>	<b>2009</b>	<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>	<b>2015</b>
<b>Lunes</b>	0,0161	0,0502	0,0583	0,0594	0,0905	0,0523	0,0829	0,0161	0,0502
<b>Martes</b>	0,0758	0,046	0,0958	0,0912	0,1195	0,0713	0,0387	0,0758	0,046
<b>Jueves</b>	0,0631	-0,0139	0,0819	0,0622	0,065	0,1793	0,1602	0,0631	-0,0139
<b>Viernes</b>	0,2354	0,0809	0,1403	0,0829	0,0708	0,173	0,0387	0,147	0,0809
<b>Sábado</b>	-0,0402	-0,0335	-0,1403	-0,174	-0,1276	-0,2383	-0,1934	-0,0402	-0,0335
<b>Domingo</b>	-0,2893	-0,2859	-0,2736	-0,2983	-0,3167	-0,3653	-0,3591	-0,2893	-0,2859

**Tabla 2.** Modificadores de macroaccidentalidad específicos, para el día según el año.

## Discusión

En este artículo se han analizado datos de una base amplia de más de 250 mil accidentes registrados, mediante la teoría de procesos estocásticos. Se ha encontrado que dicha teoría permite entender la distribución de accidentes y explicar a su vez la sobredispersión observada respecto a un modelo de Poisson. Resulta que la distribución binomial aparece porque la población de conductores puede entenderse como una mezcla no homogénea de conductores con diferentes capacidades, otros estudios que se limitaron a buscar un ajuste y encontraron evidencia de la que binomial negativa ajustaban bien los datos no aportan explicación de cuál es la causa subyacente de que dicha distribución parezca funcionar. Además el análisis propuesto permite determinar cómo es la distribución de destrezas efectivas de los conductores (aunque esta puede cambiar según el grado de cansancio, las condiciones atmosféricas y el estado del vehículo y la vía). Resulta que estas destrezas medidas por la peligrosidad intrínseca de cada conductor pueden representarse razonablemente por una distribución gamma de dos parámetros, ver ecuación (8). Esos dos parámetros dan directamente los dos parámetros que dan lugar a la distribución binomial, por lo cual los parámetros de la distribución de accidentes quedan explicados en términos de una distribución más fundamental (la de la distribución de destrezas/peligrosidades intrínsecas de los conductores). Esto constituye una profundización en la comprensión de qué parámetros estadísticos regulan la producción de accidentes.

Otro punto importante que es interesante discutir son las mejoras en los ajustes si en lugar de una distribución binomial negativa pura se usa una distribución binomial con modificación de Lindley (Binomial Negativa-Lindley). Si bien el ajuste del número de accidentes es mucho mejor, debe notarse que es están usando más parámetros para ajustar los mismos datos, con lo cual es esperable que el ajuste mejore como muchos autores detectan. Nótese que una distribución binomial negativa está caracterizada por dos parámetros  $(r, \beta)$

donde  $r \in \mathbb{N}$  y  $\beta \in \mathbb{R}^+$ , mientras que una binomial negativa-Lindley viene dada por tres parámetros, dos como los anteriores y uno adicional que da una mezcla de tipo Lindley de binomiales (ver [14]). Sin embargo, ese procedimiento no tiene un fundamento teórico sólido y previsiblemente otras formas de ajustar los datos a una binomial negativa modificada, que contenga más de los dos parámetros asociados a una binomial negativa simple. Además, en el caso de usar una binomial negativa-Lindley no es fácil ver cuál es la distribución que da la distribución de peligrosidades de los conductores, por lo que los tres parámetros ya no son interpretables en términos de algo más simple, además la mezcla para producir sobredispersión se aplica sobre los datos de frecuencias de accidentes y no sobre algo que explique cómo se producen esos accidentes. Por todas estas razones aunque una Binomial Negativa-Lindley puede obtener mejores ajustes en realidad, hace uso de un procedimiento de fuerza bruta aumentando el número de parámetros ajustables, cosa que se podría hacer de muchas maneras diferentes con resultados similares, y teóricamente no mejora la comprensión del proceso estocástico subyacente a la producción de los accidentes. Por esa razón, consideramos que esa no es la vía en la que debería procederse para entender la manera estocástica en que se producen los accidentes.

## Conclusiones

Se ha analizado la macroaccidentalidad de un área red viaria extensa durante casi una década, lo cual incluye más de 250 mil accidentes de diferentes tipos. Se ha encontrado evidencia empírica que bajo circunstancias específicas el número de accidentes en un período de tiempo sigue una distribución binomial negativa y en todos los casos analizados se ve que test  $\chi^2$  es superado, por lo que el ajuste puede considerarse adecuado. Además se ha propuesto una explicación estocástica, en la que se considera que en todo momento la población de conductores es una mezcla heterogénea (la peligrosidad intrínseca de cada conductor es diferente y la población puede

representarse mediante una distribución gamma). Esto lleva a que el ritmo de accidentes se producirá según un proceso estocástico de Poisson mixto, en particular un proceso de Pascal-Pólya que da lugar a la mencionada *distribución binomial negativa*. El artículo presenta la originalidad de explicar la razón de la aparición de este tipo de distribución, a diferencia de otros trabajos similares que se limitan a señalar que ese tipo de distribuciones parecen modelizar adecuadamente los datos, sin más explicación.

## Referencias

- 1 **Organización Mundial de la Salud** (2015) Global Status Report On Road Safety, p. 2.
- 2 **Dirección General de Tráfico** (2017) Anuario estadístico general de 2016, Archivo pdf.
- 3 **Shankar, V., Mannering, F., & Barfield, W.** (1995) "Effect of roadway geometrics and environmental factors on rural accident frequencies." *Accident Analysis & Prevention*, 27(3), pp. 371-389.
- 4 **Poch, M., & Mannering, F.** (1996) "Negative binomial analysis of intersection-accident frequencies", *Journal of transportation engineering*, 122(2), pp. 105-113.
- 5 **Hammerslag, R., Roos, J. P., & Kwakernaak, M.** (1982) "Analysis of accidents in traffic situations by means of multiproportional weighted Poisson model." *Transp. Res. Rec.* 847, *Transp. Res. Board*, Washington, D.C., pp. 29-36.
- 6 **Joshua, S. C., & Garber, N. J.** (1990) "Estimating truck accident rate and involvements using linear and Poisson regression models." *Transp. Ping. and Technol.*, 15(1), pp. 41-58.
- 7 **Chin, H. C., & Quddus, M. A.** (2003) "Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections." *Accident Analysis & Prevention*, 35(2), pp. 253-259.
- 8 **Ladron de Guevara, F., Washington, S. P., & Oh, J.** (2004) "Forecasting crashes at the planning level: simultaneous negative binomial crash model applied in Tucson, Arizona." *Transportation Research Record*, 1897(1), pp. 191-199.
- 9 **Wei, F. & Lovegrove, G.** (2013) "An empirical tool to evaluate the safety of cyclists: Community based, macro-level collision prediction models using negative binomial regression." *Accident Analysis & Prevention*, 61, pp. 129-137.
- 10 **Arregui-Dalmases, C., García-Vilana, S., Sánchez-Molina, D., & Velázquez-Ameijide, J.** "Àmbit de la Mobilitat i el trànsit." Informe final de la Càtedra de Seguretat Viària de Motocicletes del Servei Català del Trànsit, 2017.
- 11 **Mikosch, Thomas** (2013) *Non-life insurance mathematics: an introduction with the Poisson process*, Springer Science & Business Media. ISBN 978-3-540-88233-6.
- 12 **Lord, D. & Geedipally, S. R.** (2011) "The negative binomial-Lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros." *Accident Analysis & Prevention*, 43(5), pp. 1738-1742.
- 13 **Geedipally, S. R., Lord, D. & Dhavala, S. S.** (2012) "The negative binomial-Lindley generalized linear model: Characteristics and application using crash data." *Accident Analysis & Prevention*, 45, pp. 258-265.
- 14 **Cakmakyapan, S. & Ozel, G.** (2016) "The Lindley family of distributions: properties and applications." *Hacettepe Journal of Mathematics and Statistics*, 46(6), pp. 1113-1137.
- 15 **Teugels, J. L & Vynckier, P.** (1996) "The structure distribution in a mixed Poisson process." *International Journal of Stochastic Analysis*, 9(4), pp. 489-496.
- 16 **Konno, H.** (2010): "On the exact solution of a generalized Pólya process." *Advances in Mathematical Physics*. Article ID 504267 doi:10.1155/2010/504267.
- 17 **Badía, F.G., Sangüesa C., & Cha, J. H.** (2018) "Univariate and multivariate stochastic comparisons and ageing properties of the generalized Pólya process." *Journal of Applied Probability*, 55(1), pp. 233-253.