Azares del Clima

POR JOSÉ ANTONIO LÓPEZ DÍAZ

Una distribución de probabilidad inmune al promediado

Se ha dicho que la utilidad de la estadística se basa en que la desviación típica de la media de N datos es igual a la desviación típica de un dato cualquiera dividido por la raíz de N (técnicamente los datos tienen que tener todos la misma desviación estándar y ser independientes). Esta reducción de la incertidumbre de las medias valorada por su desviación estándar explica probabilísticamente el beneficio derivado de recopilar datos. A un nivel más básico la ley de los grandes números, demostrada por Jacobo Bernoulli a comienzos del siglo XVIII, sentó las bases de la interpretación frecuentista de las probabilidades, que es la más generalizada en la ciencia. Esta ley viene a decir que el resultado de promediar realizaciones independientes de una variable aleatoria converge (en un sentido preciso) hacia la media de la distribución. Este resultado permite interpretar con precisión la idea de que a base de lanzar un dado muchas veces, el número medio de veces que se obtiene una cara determinada se aproxima cada vez más a 1/6 (interpretación de von Mises de la probabilidad). Pues bien, la distribución de Cauchy se rebela contra estas leyes fundamentales. Cuando en la primera mitad del siglo XIX Poisson se propuso estudiar las propiedades de una función de densidad de probabilidad (pdf) de forma aparentemente inocua, $f(x) = C/(1+x^2)$, siendo C la constante de normalización para que su integral valga la unidad, en este caso $1/\pi$, se encontró con que la media de variables de Cauchy no tiene error finito. En lenguaje moderno la varianza de esta pdf es infinita, e incluso su esperanza matemática da una integral impropia (que no converge en valor absoluto). Es más, se puede demostrar que la media aritmética de variables de Cauchy tiene exactamente la misma pdf que cada una de las variables. Podemos decir que una variable aleatoria con pdf de Cauchy es inmune al proceso de extracción de la media de alguna manera, no se reduce en absoluto su incertidumbre, ni siquiera cambia ninguna de sus propiedades estadísticas. La razón de este comportamiento “patológico” estriba en la ya mencionada falta de los momentos de la distribución, empezando por su esperanza matemática.

Por otra parte, esto ilustra muy bien que tanto la media como la desviación estándar, de presencia ubicua en las aplicaciones, no dejan de ser magnitudes estadísticas con aspectos delicados y a veces incluso paradójicos. En último término tiene esto que ver con que se obtienen por integración de la pdf, y las integrales no son objetos matemáticos tan inmediatos y simples como, digamos, una probabilidad. En este sentido, la mediana de una distribución, es decir, aquel valor que tiene un 50 % de probabilidad de ser superado, la misma que de no ser alcanzado (en el caso de variables continuas), es más simple y está libre de estas sutilidades matemáticas. En el caso de la variable de Cauchy, si en lugar de hallar el promedio de N realizaciones hallamos su mediana, entonces sí que esta mediana tiene un error (en términos de rango intercuartílico, por ejemplo) que sigue la ley de reducción inversamente proporcional a la raíz de N. Lo mismo sucede con cualquier percentil, cuya incertidumbre también decrece según la misma ley. Así que, por suerte para la ciencia estadística, hay siempre formas de extraer más precisión de N datos que de uno solo, solo hay que saber que hay determinadas magnitudes como la popular media que pueden no ser la mejor opción.

Por otra parte la problemática de la media y desviación estándar como sumarios de información aparece de forma más inmediata al tratar con pdf asimétricas, circunstancia esta bien habitual en climatología cuando se estudia por ejemplo la precipitación en climas secos. En estos casos la media puede ser engañosa, pues puede ser significativamente superior a la mediana por el sesgo positivo de la distribución, pero en general la media se tiende a interpretar como dividiendo aproximadamente en dos la probabilidad de excedencia, lo cual solo es cierto en general en distribuciones simétricas. Como hemos visto, si la precipitación se distribuyera como el valor absoluto de una variable de Cauchy tendría media…. ¡infinita!. Es importante darse cuenta de que la pdf de Cauchy no tiene en apariencia nada sospechoso, no es físicamente imposible que por ejemplo la precipitación se ajustara a una pdf de Cauchy (en valor absoluto), lo único que pasa es que la cola de la derecha de esta distribución decae según una ley potencial $1/x^2$ en lugar de una exponencial como la gamma. Podríamos pensar que es físicamente imposible un valor medio infinito de la precipitación, pero también lo es un dato de precipitación total anual de, digamos, 100 m de altura de columna, y según cualquier distribución de las usadas habitualmente, como la gamma o la lognormal, este valor tiene una probabilidad pequeñísima pero no nula. Si truncamos la pdf para soslayar esta última imposibilidad física, entonces la pdf de Cauchy truncada se comporta tan bien como cualquier otra pdf.

Otras curiosidades de la pdf de Cauchy es que está emparentada con la distribución normal ya que se puede obtener a partir del cociente de dos variables gaussianas independientes, con lo cual no es tan artificial como parece. Además es muy estable en sentido probabilístico, en el sentido de que tanto la suma de dos variables con pdf de Cauchy como el inverso de una variable de Cauchy vuelven a tener una pdf de la familia de Cauchy, aunque con distintos parámetros de escala y localización (la normal también es estable para la suma, pero no para el inverso).

Por último hay que mencionar que la ley matemática que define la distribución de Cauchy es idéntica al perfil de Lorentz de la física, que aparece al estudiar el ancho de banda de los espectros de absorción, por ejemplo de los gases de efecto invernadero. Resulta que de los efectos que ensanchan las líneas de absorción, dos tienen ese perfil: el debido a la incertidumbre cuántica en la energía/tiempo de absorción del fotón, y el debido a la presión del gas. El hecho de que la distribución de Cauchy sea estable respecto a la suma hace que estos dos efectos se combinen produciendo un nuevo perfil de Lorentz de mayor anchura. Curiosamente, el otro efecto importante es el ensanchamiento Doppler por la velocidad de las moléculas gaseosas, que sigue una ley que es otra pdf, en este caso la más conocida, la normal gaussiana.