



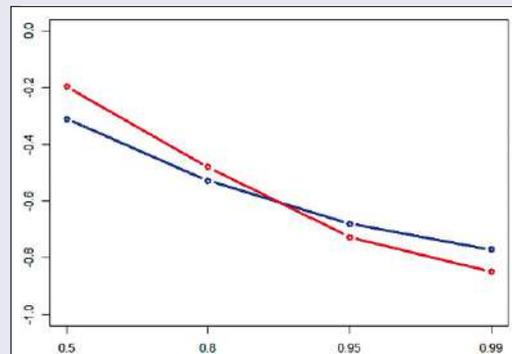
Azares del Clima

POR JOSÉ ANTONIO LÓPEZ DÍAZ

LA PARADOJA DE BERKSON Y LA PERCEPCIÓN DE EXTREMOS

Cuando dos variables aleatorias son probabilísticamente independientes su correlación vale cero. Pero resulta que partiendo de dos variables aleatorias independientes aparece dependencia entre ellas con tal de que seleccionemos aquellos casos en que ambas satisfagan una condición, incluso aunque esa selección se haga de forma que ambas sean tratadas de la misma forma. Para explicarlo de una forma sencilla, imaginemos un número grande N de objetos que pueden ser marcados como pertenecientes a una clase A, o a otra clase B, o a ambas o a ninguna de ellas. Marquemos ahora como de la clase A un subconjunto de N_A objetos escogidos al azar, a continuación marquemos como de la clase B otro subconjunto de N_B objetos escogido también al azar sin tener en cuenta si es de tipo A o no, resultando un número N_{AB} de objetos con las dos marcas, y N_0 objetos sin ninguna marca. En estas condiciones es fácil calcular el número N_{AB} aproximadamente: los N_B objetos tipo B se habrán repartido entre los marcados con A y los sin marcar de forma simplemente proporcional a sus números respectivos. Como la proporción total de B es N_B/N , los A que también se marquen como B serán aproximadamente $N_{AB} = N_A N_B/N$ (el lector con conocimientos de probabilidad reconocerá, dividiendo ambos miembros por N , que esta igualdad es la expresión probabilística de la independencia entre los dos sucesos marca A y marca B). Pues bien, la paradoja de Berkson surge cuando nos planteamos qué ocurre si nos limitamos a los objetos con alguna marca A o B, de los cuales habrá $N' = N_A + N_B - N_{AB}$ (hay que sumar a los tipo A aquellos $N_B - N_{AB}$ sin marca A que son marcados como B). Limitándonos a estos N' objetos resulta que ahora las marcas A y B ya no están repartidas aleatoriamente como al principio, lo cual a primera vista puede parecer sorprendente. Para convencerse de ello, calculemos cuántos objetos con ambas marcas esperaríamos en estos N' objetos si ambas marcas se hubiesen repartido aleatoriamente. Repitiendo el razonamiento anterior esperaríamos $N'_{AB} = N_A N_B/N'$ objetos con las dos marcas, pero los que tenemos realmente son, por supuesto, N_{AB} . Como $N_{AB} = N_A N_B/N$ y $N' < N$ tenemos que $N_{AB} < N'_{AB}$ y por tanto hay menos objetos con dos marcas que lo que esperaríamos si hubiera aleatoriedad. Dicho de otro modo, restringiéndonos al subconjunto de los objetos con alguna marca, se aprecia por así decirlo una tenden-

cia a la oposición entre las marcas A y B, lo que estadísticamente se expresa en forma de una correlación negativa entre ambas. Es posible hacer una traducción climatológica de estas conclusiones si tomamos como marca A y B el hecho de ser extrema una variable climática en dos sentidos posibles, y la paradoja de Berkson sirve para caer en la cuenta de posibles distorsiones en la percepción de la relación entre variables extremas. Por ejemplo, supongamos que una variable térmica en un año toma un máximo y un mínimo que se puedan considerar independientes uno de otro, por ejemplo un índice de extremosidad de la ola de calor y frío más intensas. Si estudiamos la correlación entre ambos índices a lo largo de varios años, de la independencia se deduce que deben estar incorrelacionados. Pero si seleccionamos aquellos años en que al menos uno de los índices de extremosidad exceda un umbral determinado, la correlación en esos años entre los índices será negativa en virtud de lo expuesto.



Correlación entre parejas de variables aleatorias independientes tras condicionar a que una, al menos, supere la probabilidad acumulada indicada en abscisas; en azul variables normales, en rojo exponenciales.

Una ilustración numérica aparece en la figura, para la que se ha supuesto que ambos índices extremos son variables aleatorias con distribución normal en un caso y exponencial en otro. En abscisas se representan valores de umbrales extremos crecientes expresados en términos de su probabilidad de excedencia. Por ejemplo, los valores de la correlación que corresponden a $x = 0.8$ (para las dos distribuciones próximos a -0.45) se obtienen tras condicionar a que los dos índices superen el percentil 80% (o sea, estén entre el 20% más extremo). Por tanto si solo centramos nuestra atención en aquellos años extremos al menos por calor o frío, percibiremos una tendencia a la oposición entre ambos extremos, de forma que concurren ambos simultáneamente significativamente menos de lo que indicaría una independencia entre ambos.

La percepción será de una tendencia clara al zigzaguo, como si el clima mostrara una tendencia a oscilaciones marcadas. Creo que este fenómeno puede darse cuando uno confía en su memoria, dado que los años que quedarán más marcados en la memoria son precisamente aquellos que han sido extremos en algún sentido, cuando en realidad puede que no haya realmente esa tendencia al zigzaguo si incluimos los años más anodinos también. Análogamente, dado que la atención mediática es selectivamente atraída por lo extremo, puede difundirse una percepción poco rigurosa de un clima desquiciado en exceso.