

Ji- cuadrado

Supongamos que tenemos una serie de cincuenta números, cada uno de ellos entre cero y nueve. ¿Está uniformemente distribuida? Es decir, todos los números del 0 al 9 ¿tienen la misma probabilidad -0,10- de aparecer en cada posición de la serie? Estas son las preguntas que corresponde hacer para determinar el valor de los números generados pseudo-aleatoriamente. Si puede pasar el test estadístico que responda a esas preguntas, cualquiera que sea, entonces, a todo efecto práctico, la serie se comporta como si fuera aleatoria.

Para entender los tests estadísticos vamos a examinar uno de los más usados: el χ^2 (usualmente llamado chi -o ji- cuadrado). Si cada elemento de la serie es un número entre 0 y 9, y la serie está uniformemente distribuida, entonces, en 50 posiciones (la longitud de la serie) cada número del 0 al 9 debería aparecer más o menos cinco veces. Es decir, el valor esperado de repetición de cada número es 5. Naturalmente, nadie pensaría que cada número saldrá con seguridad exactamente cinco veces, pero uno esperaría que la cantidad de veces que sale cada número se acerque a 5.

Si en lugar de tener una serie de longitud 50 tuviéramos una de longitud 5000, esperaríamos que cada número saliera aproximadamente 500 veces. Además, nos llamaría aún más la atención que cada número saliera exactamente 500 veces en una serie de longitud 5000 que 5 veces en una serie de longitud 50. Por otro lado, también nos llamaría más la atención que un número saliera 200 veces en una serie de 5000 que 2 veces en una serie de 50. Estos razonamientos intuitivos quedan reflejados en un número: el valor χ^2 , que es igual a la suma de los cuadrados de las diferencias entre las veces que efectivamente salió cada número y los respectivos valores esperados, divididos por dichos valores esperados; para una serie de 50 números, si cada uno de ellos puede tomar valores del 0 al 9 con la misma probabilidad, tenemos

$$\chi^2 = \frac{(n_0 - 5)^2}{5} + \frac{(n_1 - 5)^2}{5} + \dots + \frac{(n_9 - 5)^2}{5}$$

donde n_0, n_1, \dots, n_9 son las veces que efectivamente salieron el 0, el 1, ..., el 9, respectivamente.

Los estadísticos han calculado, para ciertas probabilidades, los valores de χ^2 que tienen esas probabilidades de ser sobrepasados, o de no ser alcanzados, según la cantidad de elementos distintos (10 en nuestro ejemplo) que pueden componer cada miembro de la serie. Por ejemplo, el valor 21,67 tiene una probabilidad del 1% de ser superado. Si el valor obtenido de χ^2 para nuestros 50 valores de la serie es mayor que 21,67, entonces la serie es muy sospechosa: casi con seguridad no estará uniformemente distribuida, pues sólo una serie de cada 100 uniformemente distribuidas, en promedio, podría dar valores tan altos: convendría generar más números y probar de nuevo para poder, si sigue dando valores tan altos, rechazar la hipótesis de que está uniformemente distribuida.