

Cine y estadística (y 2)

JOSÉ MARÍA SORANDO MUZÁS

La Estadística en *Numb3rs*

81

Continuamos¹ con la presencia estadística en el cine y en las series de televisión, donde *Numb3rs* ocupa un lugar preferente. Diría que en los 119 episodios que componen sus seis temporadas hay casi tantas matemáticas como en todo el resto de la ficción cinematográfica y televisiva; y, desde luego, más solventes. Algunas escenas son utilizables en el aula y en la mayoría de los casos nos ofrecen ejemplos insospechados e interesantes de aplicación matemática, ejemplos de los que al profesorado le conviene estar provisto para poder utilizarlos en el momento oportuno.

Charlie, el protagonista matemático, recibe datos de cada caso policial, a veces en gran cantidad. Para su interpretación a menudo recurre al estudio estadístico y suele concluir haciendo predicciones en términos de probabilidad. Terminaba el anterior artículo con un inquietante caso orientado al control social. Hay otros más aceptables.

En «Punto de vista» (episodio 9 de la primera temporada), usa la regresión a la media para justificar qué disparos corresponden a un mismo francotirador.



NOVIEMBRE
2012

En «Convergencia» (episodio 7 de la 2.^a temporada), recurre a la minería de datos, técnica de uso creciente que consiste en la extracción no trivial de información que reside de manera implícita en los datos. La explica así:

—Yo diseño un algoritmo y, a partir de los datos de esos crímenes, el algoritmo busca correlaciones. Necesito las estadísticas de delitos del último año en el condado.

—Pero ¿tantos datos no añadirán complejidad a la búsqueda?

—Cuanto más datos, más probabilidad de encontrar algo. Es como cuando quieres armar un puzzle.... Pero resolver un problema de la vida real es como armar un rompecabezas cuando todas las piezas que necesitas están mezcladas con piezas de otros rompecabezas... Tienes que revisar la caja entera y separar las piezas que son de tu puzzle. El algoritmo revisa todas las piezas y entresaca aquellas que encajan.

En «El corredor» (episodio 15 de la 2.^a temporada), considera que unos datos son muy improbables en base a la curiosa Ley de Benford:

—Si observas cualquier tabla de valores de amplio espectro: datos del censo, tamaños de terrenos... el 1 aparece como primer dígito en más del 30% de los valores, el 2 aparece un 17,6%, el 3 un 12,5% y así sucesivamente (en orden decreciente).

—Parece ser que Simon Newcomb lo descubrió ojeando libros de tablas de logaritmos. Observó que las primeras

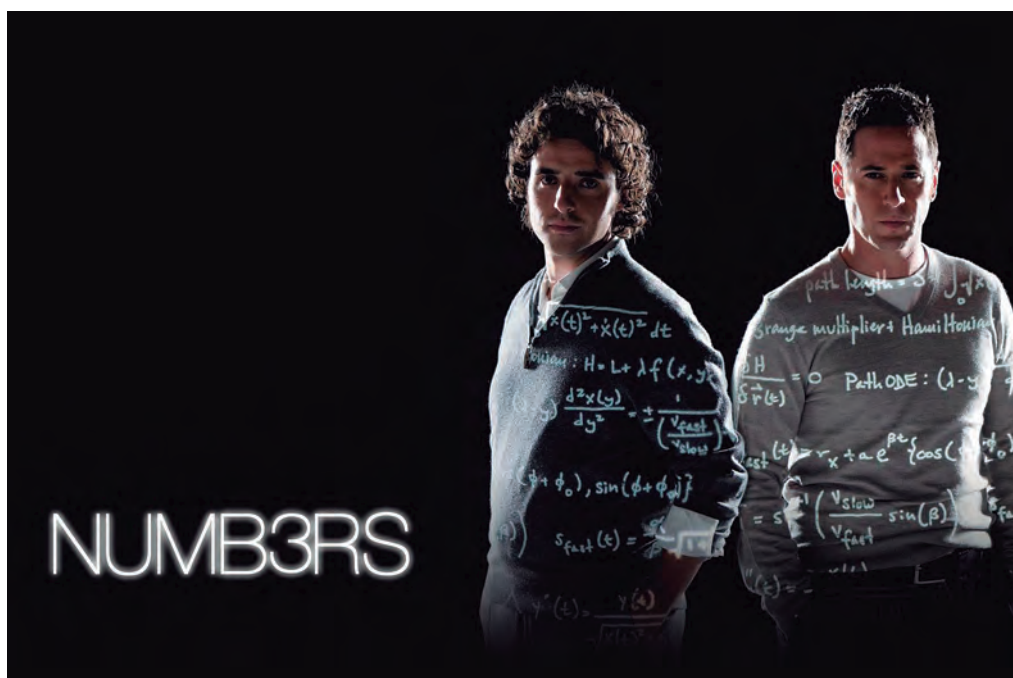
páginas están más desgastadas que las demás. La Ley de Benford no trata de toma consciente de decisiones, es un fenómeno estadístico.

Newcomb fue un astrónomo del siglo XIX y Frank Benford un físico del siglo XX que observaron el mismo fenómeno de manera independiente. Benford lo comprobó empíricamente sobre un total de 20 229 datos de procedencias diversas (sociales, naturales, de laboratorio, etc). Su formulación final fué que «la probabilidad de que n ($n = 1, 2, \dots, 9$) sea la primera cifra no nula de un número tomado al azar, es $\log(n + 1) - \log n$ ».

Dicha ley concluye en la citada escala decreciente de probabilidades para los dígitos de 1 a 9 y de ella se deriva también la mayor probabilidad de que la primera cifra sea impar.

La extrañeza que provoca de entrada esta ley es menor cuando empezamos a revisar cuántas medidas, nuestra propia edad por ejemplo, o no alcanzan un dígito inicial alto o, si lo alcanzan, para

82
SUMO 71



llegar a él debieron pasar antes por los anteriores, que eventualmente ostentaron la primacía.

En «El chat de la muerte» (episodio 11 de la 3.^a temporada), se aplica el análisis estadístico lingüístico para establecer el perfil de las intervenciones del sospechoso.

En «Democracia» (episodio 18 de la 3.^a temporada), se detecta una falsificación de datos en base a un sorprendente criterio tras el análisis de frecuencias:

—Hay demasiados 7 y 3 para que esta lista sea aleatoria. Por alguna razón, cuando la gente inventa informes que tienen números, siempre ponen demasiados 3 y 7 y muy pocos 1 y 2. Así, no puedes saber qué describen los números, pero puedes saber si alguien miente.

No he encontrado información que lo corrobore, pero dada la potente asesoría matemática con que ha contado *Numb3rs*, les concedo el beneficio de mi ignorancia².

La muestra perfecta

En cualquier investigación, policial o no, hay que saber escoger las muestras de datos. *Ciudad Mágica* (*Magic Town*, William A. Wellman, 1947) es una película cuya trama gira en torno a la representatividad de dichas muestras.

Rip Smith (interpretado por James Stewart) dirige una empresa que se dedica a los sondeos de opinión. Al borde de la bancarrota, casualmente descubre que los resultados obtenidos en la pequeña ciudad de Grandview son una réplica exacta de los obtenidos para todos los Estados Unidos. Es lo que repetidamente llama el «milagro matemático», la salvación para su empresa, que le permitirá ofrecer resultados válidos con un coste y rapidez muy ventajosos frente a la competencia. Rip



NOVIEMBRE
2012

83
sumat
1

planea trasladarse con sus colaboradores a Grandview y establecerse bajo una falsa identidad de vendedores de seguros, para así conseguir información de forma permanente sin que los vecinos sospechen cuál es su finalidad.

Rip se integra rápidamente en la vida de la ciudad, llegando a ser muy apreciado y se enamora de una periodista (interpretada por Jane Wyman). Ésta descubre el propósito de Rip y lo publica en el diario local. La noticia corre por todo el país y Grandview pasa a ser conocida como la ciudad modelo de la nación. Nada volverá a ser lo mismo. La codicia de unos y otros traerán el rápido ascenso y la estrepitosa caída de la ciudad. Como no podía faltar en una película de Hollywood de posguerra, hay un final esperanzador.

Es ésta una película de las de antes: en blanco y negro, donde los personajes son elegantes y corteses, la honradez triunfa, el amor redime... En el aspecto social, llama la atención la democracia asamblearia



NOVIEMBRE
2012

que rige el destino de la comunidad. Pasemos a sus aspectos matemáticos.

Frente a la estadística descriptiva, que analiza los datos obtenidos de todos los elementos de una población, la estadística inferencial extrae conclusiones a partir de muestras de esa población, con un grado de confianza conocido. Se hace así cuando la observación destruye el objeto observado (por ejemplo, algunos controles industriales de calidad) o por economía de medios, que es el caso que corresponde a la película y también el de

los habituales sondeos preelectorales. En la selección de las muestras es fundamental lograr que sean representativas de la población.

Es célebre el caso de las elecciones presidenciales de EE.UU. de 1936, en las que ganó Franklin D. Roosevelt. La revista *The Literary Digest* hizo una encuesta de intención de voto en la que participaron más de cuatro millones de sus lectores y se equivocó en el pronóstico. Otra encuesta realizada por George Gallup sólo a 5.000 personas anunció el éxito de Roosevelt con mucha exactitud. La razón era que en el primer caso la muestra no era representativa de la sociedad norteamericana, pues entre los lectores de esa revista había mayoría republicana, siendo además el método de encuesta el envío por correo de unos cupones, lo cual suponía un alto grado de implicación en la campaña. Sin embargo, en las 5 000 personas seleccionadas por Gallup estaban bien representados todos los sectores e ideologías de dicha sociedad.

Lograr esa representatividad exige conseguir una distribución geográfica y social de las encuestas muy cuidadosa y extensa, es decir, costosa. De ahí que localizar una muestra reunida de forma natural como Grandview fuera un «chollo» para la empresa de sondeos. En la película se dice que en dicha ciudad los porcentajes de hombres y mujeres, de blancos y negros, de republicanos y demócratas, etc. son réplicas exactas de los que se dan para el conjunto de la nación. De esta forma se justifica aceptablemente la coincidencia de opiniones.

Cuando la cámara se detiene en el estadijo de datos podemos comprobar que los porcentajes que atienden a cada criterio suman 100 (¡qué menos!... aunque, como ya hemos visto en otros casos, no siempre se cuidan esos detalles numéricos).

84
SUMO
71



THE MIRACLE OF GRANDVIEW									
Men	Women	White	Colored	Democrat	Republican	Other	Protestant	Catholic	Jewish
60	40	56	44	52	48	49	11	53	45
45	55	46	54	62	38	9	91	79	79
9	11	55	45	46	54	62	38	9	91
70	30	53	47	54	46	34	25	70	10
49	51	60	40	35	65	77	74	87	11
46	54	52	48	53	47	55	45	11	89
48	52	64	36	40	60	40	36	64	36
52	48	48	52	11	55	45	46	39	61
62	38	76	24	74	26	77	61	39	22
78	22	78	22	61	39	29	71	60	40
64	36	46	54	76	24	78	22	61	39
67	33	77	23	68	32	70	30	70	30



Sorprende en esta era digital ver que los cálculos se realizan sobre un gran cuadrante por un contable a la vieja usanza: a lápiz, con manguitos y visera. Pensemos que en 1947 todavía faltaban algunos años para la expansión del uso de las calculadoras mecánicas de oficina (con rollo de papel y manivela) y 25 años para las calculadoras electrónicas de bolsillo.

Las predicciones de la película se obtienen como porcentajes fijos (estimadores puntuales), cuando lo habitual en los estudios demoscópicos es darlas en forma de intervalos de confianza, acompañados de su fiabilidad en términos de probabilidad (nivel de confianza, habitualmente del 95% o el 99%).

Por último, el guión es acertado cuando plantea cómo la consciencia de estar siendo observados, en ese caso con una presión mediática excesiva, altera el comportamiento de los ciudadanos de Grandview. Es una traslación a las Ciencias Sociales del Principio de Incertidumbre de Heisemberg de la Física Cuántica.

La neutralidad de los cuestionarios

Claro que, además de una muestra representativa, es necesario que el cuestionario no sea inductor de las respuestas. En la serie de la BBC *Yes, Prime Minister*, traducida al catalán y emitida por TV3 bajo el título *Sí, Primer Ministre*, lo ejemplifica el siguiente diálogo entre asesores del Primer Ministro:

- El Partido ha hecho una encuesta y se ve que los votantes están a favor de restablecer el Servicio Militar.
- Tenemos otra encuesta que demuestra que los votantes están en contra del Servicio.
- No pueden estar a favor y en contra.
- Claro que sí, ¿es que no te han hecho nunca encuestas?

—Sí, pero no una encuesta política. No sabría qué contestar.

—Tú ya sabes cómo funciona eso. Te visita una señorita muy guapa y quieres quedar bien, no quieres hacer el ridículo.

—No.

—Y te comienza a hacer preguntas: ¿Le preocupa el gran número de jóvenes que están sin trabajo?

—Sí

—Y ¿le preocupa también el aumento de la delincuencia juvenil?

—Sí.

—¿Le parece que falta disciplina en las escuelas públicas?

—Sí

—¿Piensa que los jóvenes aceptarían un mayor nivel de liderazgo y de autoridad? ¿Qué son capaces de aceptar un reto?

—Sí

—¿Estaría a favor de restablecer el servicio Militar?

—Sí

—Está claro que sí, después de lo que me has dicho no puedes contestar que no. Ellos no publican las primeras preguntas y sólo mencionan la última.

—¿Es eso lo que hacen realmente?

—Hombre, los que son honrados tal vez no, pero no hay muchos de esos. Y la señorita habría conseguido el resultado opuesto si hubiera querido.

—¿Cómo?

—¿Le preocupa el peligro de una guerra?

—Sí.

—¿Le preocupa la carrera de armamentos?

—Sí

—¿Encuentra peligroso dar armas a jóvenes y enseñarlos a matar?



- Sí.
- ¿Le parece mal obligar a la gente a usar armas si ellos no quieren?
- Sí.
- ¿Usted se opondría al restablecimiento del Servicio Militar?
- Sí.
- Ya ve, es así como se encargan las encuestas para el Ministerio de Defensa.

El anterior diálogo sería jocoso sin más si no fuera porque se parece demasiado al tipo de encuestas que realizan algunos medios de comunicación políticamente beligerantes. Un caso notorio es el de un diario de tirada nacional cuyos sondeos de opinión son dudosos para el propio sector demoscópico. La razón es que han sido realizados, de forma reincidente, por empresas desconocidas.

Y otro sesgo no despreciable es el introducido por la presión de las empresas sobre los encuestadores, que conduce en ocasiones a muestras repletas de familiares y amigos. También lleva a veces a resultados que agraden a un cliente que no busca conocer la realidad para decidir, sino informes con números que justifiquen decisiones ya tomadas. En todos esos casos, no se culpe a la Estadística sino a quienes la usan de forma perversa.

Videos estadísticos para el aula

Con ocasión del Día Mundial de la Estadística (2010), el Instituto Nacional de Estadística (INE) publicó en la red un video divulgativo titulado «Un día en cifras» en el que se muestran las grandes di-



ferencias existentes según países entre los indicadores de diversas variables sociales. Por ejemplo: «Partos atendidos por personal cualificado por cada 100 personas: Cuba 99,9; Egipto 75; Pakistán 31». Es de mayor aplicación en la clase de Ciencias Sociales que en la de Matemáticas, pero queda muy bien plasmada la conexión entre ambas áreas.

En la misma línea está otro recomendable video del INE: «Si España fuese un pueblo de 100 habitantes» (febrero 2012). Concluye con este rótulo: «El objetivo de este video es transmitir de forma sencilla la utilidad de las estadísticas oficiales para reflejar la sociedad en la que vivimos». Loable propósito, cuyo logro precisa, como tantos otros esenciales, mantener un servicio público profesional y libre de presiones.

Los enlaces para ver en Internet las escenas de éste y anteriores artículos, se encuentran en:

http://catedu.es/matematicas_mundo/Cinemateca.htm

JOSÉ MARÍA SORANDO MUZÁS
 IES Elaios, Zaragoza
 <decine@revistasuma.es>

1 En *Suma 70* nos extrañábamos de que no se estén aplicando las técnicas saberométricas al fútbol. Hemos conocido un interesante blog que trabaja en esa dirección. Es obra del profesor Francisco J. García Cubero y se titula *Aspectos matemáticos del juego del football*.

URL: <<http://www.futmath.blogspot.com>>

2 En una reciente película de acción, *Safe* (Boaz Yakin, 2012), también se sospecha de la supuesta aleatoriedad de una lista de números «porque

hay demasiados 7 y 3». Pero en ese caso no se alude al criterio de tipo psicológico citado en *Numbers*, sino que se supone que se trata de un mensaje en clave, donde la mayor frecuencia de esas cifras se correspondería con las frecuencias relativas de ciertas letras en el idioma, al estilo del célebre descifrado que narra Edgar Allan Poe en el relato *El escarabajo de oro*.