Probabilidad y análisis de datos

Daniel Fraiman

Probabilidad Condicional e Independencia

1. Se tiene la siguiente distribución de 300 estudiantes, según sexo y área de estudio:

	Biología	Exactas	Humanas
Masculino	52	40	58
Femenino	38	32	80

   Un estudiante es elegido al azar.

   (a) ¿Cuál es la probabilidad de que sea de sexo femenino y del área de humanas?

   (b) ¿Cuál es la probabilidad de que sea de sexo masculino y no sea del área de biología?

   (c) Dado que fue elegido un estudiante del área de humanas, ¿cuál es la probabilidad de que sea de sexo femenino?

   (d) Dado que fue elegido un estudiante de sexo masculino, ¿cuál es la probabilidad de que no sea del área de biología?

2. Supongamos que la probabilidad de vivir 70 o más años es 0.6 y que la probabilidad de vivir 80 o más años es 0.2. Si una persona cumple 70 años, ¿cuál es la probabilidad de que festeje su cumpleaños número 80?

3. Supongamos que se testean los chips para un circuito integrado y que la probabilidad de que sean declarados con fallas cuando efectivamente las tienen de 0.95 y que la probabilidad de que sean declarados en buen estado si efectivamente lo están es de 0.97. Si el 0.5 % de los chips tienen fallas, ¿cuál es la probabilidad de que un chip que ha sido declarado con fallas sea bueno?

4. La urna A contiene cuatro bolas rojas, tres azules y dos bolas verdes. La urna B tiene dos bolas rojas, tres azules y cuatro bolas verdes. Se elige una bolilla de la urna A y se la coloca en la urna B y luego se extrae una bolilla de la urna B.

   (a) ¿Cuál es la probabilidad de que la bolilla extraída de la urna B resulte roja?

   (b) Si una bolilla roja es extraída de la urna B, ¿cuál es la probabilidad de que una bolilla roja haya sido extraída de la urna A?

5. Una compañía de seguros contra incendios tiene clientes de alto, mediano y bajo riesgo, que tienen probabilidades de presentar reclamos de 0.02, 0.01 y 0.0025 respectivamente, en un año dado. Las proporciones de los números de clientes en las tres categorías son 0.10, 0.20 y 0.70 respectivamente. ¿Qué proporción de los reclamos recibidos por año por la compañía proviene de los clientes de alto riesgo?

6. El mejor algoritmo actual de clasificación supervisada (supervised machine learning) para etiquetar en forma automática perros de gatos (a partir de imágenes) clasifica en forma perfecta a los perros y gatos de casi todas las razas salvo Pomerania y Caracal.

El 30 % de las veces que se presenta una foto de un perro Pomerania el algoritmo lo clasifica como un gato, y el 20 % de las fotos de gatos Caracal son mal clasificadas. Sabiendo que la base de imágenes a analizar contiene un 10 % de perros Pomerania y un 25 % de gatos Caracal, ¿qué porcentaje de las imágenes se espera que estén bien clasificadas? Dicho de otra manera, si elegimos una imagen al azar, ¿cuál es la probabilidad que el algoritmo la clasifique bien?

7. Probabilidades condicionales en el número π. Utilizando calcule la probabilidad condicional empírica de que inmediatamente despúes del dígito i aparezca el dígito j ($i, j = \{0, 1, \ldots, 9\}$) en la secuencia de números correspondiente a π. Por ejemplo, ¿cuál es la probabilidad de que aparezca el 2 inmediatamente luego del 1? Estudie esta matriz de "transiciones" (matriz de 10×10). Curioso, ¿no?, ¿valdrá lo mismo al condicionar a secuencias dobles?

```

url=readLines("https://www.piday.org/million/")
pii=url[21]
sal=strsplit(pii,"</div></div></div></article>")
sal2= strsplit(sal[[1]][1],"data-page=\"1\">")
pii=sal2[[1]][2]
num_pii=strsplit(pii,split="")
class(num_pi)
num_pi=num_pi[[1]]
num_pi=num_pi[-2]  # sacamos el punto  3.149
num_pi=as.numeric(num_pi)
```

8. Sean A y B eventos independientes tales que $P(A) = \frac{1}{4}$ y $P(B) = \frac{1}{3}$.

   (a) Hallar la probabilidad de que ninguno de estos dos eventos ocurra.

   (b) Hallar la probabilidad de que exactamente uno de los dos eventos ocurra.

9. Un sistema tiene n unidades independientes, cada una de ellas con probabilidad p de fallar. El sistema completo falla si 1 o más unidades fallan. ¿Cuál es la probabilidad de que el sistema falle?

10. Probar que si $A \subset B$ entonces $P(B|A) = 1$ y $P(A^c|B^c) = 1$. Interpretar el resultado.

11. Si A y B son disjuntos, ¿pueden ser independientes?

12. Probar que si A y B son eventos independientes entonces A y B^c también.