Tarea # 1

Jonathan Quirola Vásquez
Estadística para Astrónomos

Preguntas cortas

1. ¿Cuál es la diferencia entre eventos disjuntos y eventos independientes? De un ejemplo astronómico de un evento disjunto y un evento independiente.

   Sea A_i y A_j dos eventos cualquiera, talque $\mathbb{P}(A_i) \geq 0$ y $\mathbb{P}(A_j) \geq 0$. Se definen como eventos disjuntos si la probabilidad de que suceda A_i y A_j es cero, es decir,

   $$\mathbb{P}(A_i \cap A_j) = 0.$$

   En otras palabras, los conjuntos A_i y A_j no comparten elementos en común, $A_i \cap A_j = \emptyset$.

   • Si explota una supernova, la probabilidad que sea una supernova Ia (SNIa) y que tenga lineas de Hidrógeno (H) es cero. Los eventos son disjuntos ya que no se ha observado la emisión de Hidrógeno en este tipo de supernovas, por lo tanto es imposible:

   $$\mathbb{P}(\text{SN Ia} \cap \text{H}) = 0.$$

   Se definen como eventos independientes si la probabilidad de que suceda A_i y A_j sea igual a la probabilidad de A_i y la probabilidad de A_j, expresado matemáticamente como

   $$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j).$$

   • Si explota una supernova, la probabilidad que sea tipo Ia (SNIa) y que se encuentre en una galaxia espiral (espiral) es diferente de cero. Los eventos son independientes ya que las supernovas Ia han sido observadas en todo tipo de galaxias, es decir:

   $$\mathbb{P}(\text{SN Ia} \cap \text{espiral}) = \mathbb{P}(\text{SN Ia})\mathbb{P}(\text{espiral}).\blacksquare$$

2. Suponga que dos eventos, A y B son independientes. Pruebe que A^c y B^c son independientes.

   Supongamos que Ω es todo el espacio muestral, entonces podemos decir que

   $$A^c \cap B^c = \Omega - A \cup B.$$

   Si obtenemos la probabilidad de la relación anterior vamos a tener

   $$\mathbb{P}(A^c \cap B^c) = \mathbb{P}(\Omega) - \mathbb{P}(A \cup B).$$

   donde $\mathbb{P}(\Omega) = 1$. Podemos expresar el segundo término de la derecha como $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$, entonces la expresión anterior será

   $$\mathbb{P}(A^c \cap B^c) = 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A \cap B),$$

   como los eventons A y B son independientes, podemos expresar $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, entonces

   $$\mathbb{P}(A^c \cap B^c) = 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(B),$$

   tomando como factor común a $\mathbb{P}(B)$ en los últimos dos elementos de la expresión anterior, tenemos

   $$\mathbb{P}(A^c \cap B^c) = 1 - \mathbb{P}(A) - \mathbb{P}(B)(1 - \mathbb{P}(A)),$$

y tomando como factor común a $1 - \mathbb{P}(A)$, vamos a obtener

$$\mathbb{P}(A^c \cap B^c) = (1 - \mathbb{P}(A))(1 - \mathbb{P}(B)),$$

y como $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$, demostramos que

$$\mathbb{P}(A^c \cap B^c) = \mathbb{P}(A^c)\mathbb{P}(B^c).\blacksquare$$

Estudiando encuestas

1. En el juego de Monty Hall, una persona tiene dos probabilidades: mantener su respuesta inicial (probabilidad r) o cambiarla (probabilidad $1 - r$). Si X es el número de veces que la gente escoge mantener su respuesta inicial en una encuesta, entonces X podría describirse como una distribución binomial:

$$X \sim \text{binomial}(n, r).$$

Se escogió esta distribución ya que solamente existen dos respuestas posibles, con probabilidad r y $1 - r$, además el evento se repite n número de veces. La distribución tendrá la forma:

$$f(X = x) = \begin{cases} \begin{pmatrix} n \\ x \end{pmatrix} r^x (1-r)^{n-x} & \text{con } x = 0, 1, \cdots, n \\ \\ 0 & \text{otros valores} \end{cases}$$

donde x es cualquier valor posible que puede tomar la variable aleatoria X, es decir, el número de personas que mantienen su respuesta inicial, n es el número total de eventos (o personas encuestadas) y r la probabilidad que X personas mantengan su respuesta inicial.$\blacksquare$

2. Si asumimos que $r = 0.50$ estamos suponiendo que es igualmente probable que elijan quedarse con su respuesta inicial o cambiarla. Si analizamos los resultados de la encuesta, vemos que la probabilidad que no cambien de respuesta es ~ 0.54 $(18/33 = 0.54)$, por lo tanto no es incorrecto suponer que $r = 0.50$.
$X = 18$ quiere decir que 18 personas de la encuesta decidieron mantener su decisión inicial. Si definimos a $r = 0.50$ estamos suponiendo una probabilidad fija, por lo tanto si utilizamos la distribución binomial definida anteriormente podemos determinar la probabilidad de $X = 18$ dado que $r = 0.50$, es decir, $\mathbb{P}(X = 18|r = 0.50)$.
Si tomamos que $n = 33$, tal como la encuesta realizada en Twitter, vamos a tener que la probabilidad que 18 personas decidan mantener su opción inicial dado que $r = 0.50$ es

$$\mathbb{P}(X = 18|r = 0.50) = \begin{pmatrix} 33 \\ 18 \end{pmatrix} 0.50^{18} 0.50^{15} = 0.1207.$$

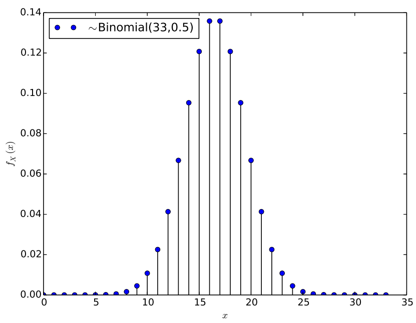

Figure 1: Distribución binomial para el caso de $n = 33$, $X = 18$ y $r = 0.50$.

En la Figura 1 podemos ver la distribución binomial ($\sim$ Binomial$(33, 0.5)$) considerada para el cálculo de $\mathbb{P}(X = 18|r = 0.5)$.
Adicionalmente, se escribió un código en `Python` para simular la encuesta 1000 veces y calcule

cuantas veces se obtiene $X = 18$. Los resultados que se obtuvieron de las simulaciones muestran que $X = 18$ en 102, 137, 117, ... de las 1000 encuestas. Este resultado es consistente con $\mathbb{P}(X = 18|r = 0.5)$ de la distribución binomial. Si multiplicamos la $\mathbb{P}(X = 18|r = 0.5)$ por 1000 encuestas, esperamos que en ~ 120 encuestas vamos a tener $X = 18$ dado que $r = 0.5$. Por lo tanto, los resultados obtenidos de las simulaciones son consistentes con la probabilidad de la distribución binomial. El código de la simulación y del histograma puede ser encontrado en el siguiente link: Github.

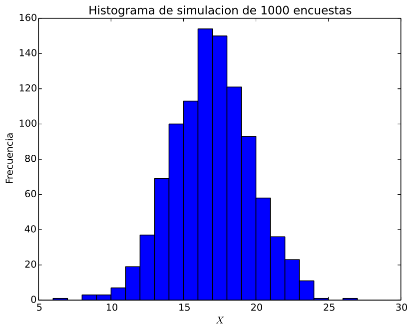

Figure 2: Histograma del X (número de personas que mantienen su respuesta) en 1000 encuestas simuladas.

En la Figura 2 se aprecia un histograma del número de personas que mantienen su primera opción (X) en las 1000 encuestas simuladas. Se observa que cuando $X = 18$ la frecuencia de suceso es de ~ 120. El código para obtener la gráfica de la distribución binomial se encuentra en el siguiente link: Github.$\blacksquare$

3. Partiendo de la definición de probabilidad condicional, tenemos que la probabilidad de r dado un $X = x$ será

$$\mathbb{P}(r|X = x) = \frac{\mathbb{P}(r \cap X)}{\mathbb{P}(X)}.$$

Utilizando el teorema de Bayes sobre todos los valores que puede tomar r, tenemos

$$\mathbb{P}(r|X = x) = \frac{\mathbb{P}(X|r)\mathbb{P}(r)}{\sum_r \mathbb{P}(X|r)\mathbb{P}(r)}.$$

Como la distribución de r es $\sim u(0, 1)$ (uniforme de 0 a 1), entonces $\mathbb{P}(r)$ es 1 para todo el rango de valores de r (esto porque en una distribución uniforme la probabilidad es $1/(b - a)$ con $b > a$, en este caso $a = 0$ y $b = 1$). Por lo tanto, la probabilidad de r dado un valor de $X = x$ será

$$\mathbb{P}(r|X = x) = \frac{\mathbb{P}(X|r)}{\sum_r \mathbb{P}(X|r)}.$$

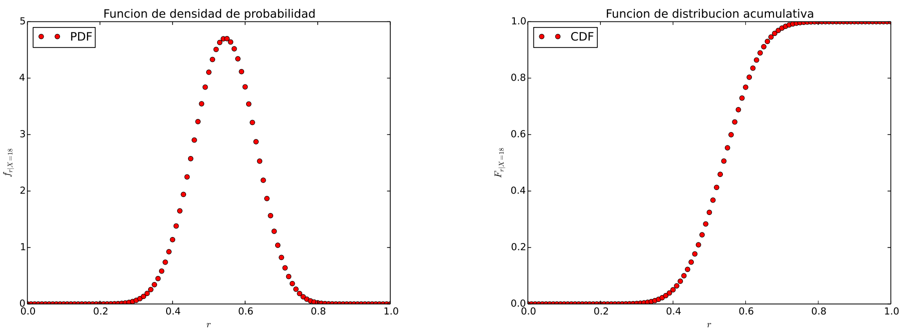

Figure 3: Izquierda: PDF de la $\mathbb{P}(r|X = x)$ con $x = 18$. Derecha: CDF de la $\mathbb{P}(r|X = x)$ con $x = 18$.

Hay que recordar que $\mathbb{P}(X|r)$ representa una distribución binomial. r es un parámetro que va entre 0 y 1, por lo tanto la sumatoria puede ser aproximada como una "integral" (de hecho esto se podría argumentar como una integral de Riemann), entonces con un código de integración numérica se puede obtener el denominador de la distribución, el cual es

$$\sum_r \mathbb{P}(X|r) = 0.029.$$

Para graficar la distribución (PDF) de $\mathbb{P}(r|X = x)$ se fijó el valor de $X = 18$, tal como la encuesta en Twitter, y se integró en el rango de $0 \leq r \leq 1$. Adicionalmente, para obtener los valor de la CDF se integró de forma numérica de $0 \leq r \leq 1$ donde la CDF converge a 1, por lo tanto se verifica que la PDF y la CDF fueron definidas correctamente. En la Figura 3 se observa los resultados obtenidos de la PDF (izquierda) y CDF (derecha). El máximo de la distribución definida se encuentra aproximadamente en $r = 0.5456$. El código en `Python` para obtener el denominador de la PDF y las gráficas de la CDF y PDF se lo puede encontrar en el siguiente link: Github. ■

4. Lo que necesitamos compara es la $\mathbb{P}(r > 0.50|X)$ y $\mathbb{P}(r < 0.50|X)$. Utilizando un integrador numérico en `python` se obtuvo que:

$$\mathbb{P}(r > 0.50|X) = 0.696 \text{ y } \mathbb{P}(r < 0.50|X) = 0.304,$$

por lo tanto, la relación entre estas dos probabilidades es

$$\frac{\mathbb{P}(r > 0.50|X)}{\mathbb{P}(r < 0.50|X)} = 2.292.$$

Este resultado nos dice que la probabilidad de que $r > 0.5$ dado que $X = 18$ es el doble que la probabilidad de que $r < 0.5$ dado $X = 18$. Por lo tanto, es más probable que r sea mayor a cero dado que $X = 18$. Esta conclusión concuerda con lo obtenido en la encuesta de Twitter ya que se encontró que $r = 0.545$.
Este resultado es lógico, ya que si $X < 18$ esperaríamos que la probabilidad de $r > 0.50$ dado X sea pequeña, mientras que para el caso contrario ($X > 18$) la probabilidad de $r > 0.50$ dado X sería grande. Esto también se puede apreciar en el resultado de la encuesta en Twitter.
El código en `Python` para obtener las probabilidades se lo puede encontrar en el siguiente link: Github. ■

Las gráficas presentadas en este documentas se las puede encontrar en Github.