Tarea 1 – Estadística para Astrónomos

Alumno: Javier Minniti

Problema 1.

a)

Independencia:

A y B son independientes $\Leftrightarrow P(A \cap B) = P(A)P(B)$

Cuando dos eventos son independientes entre si, esto significa que la probabilidad de que un evento ocurra de ninguna manera afecta la probabilidad de que el otro ocurra.

Para dos eventos A y B cualesquiera, la probabilidad condicional se define como $P(A|B) = P(A \cap B)/P(B)$.

Luego, para dos eventos independientes A y B, se tiene que $P(A|B) = P(A)$.

Esto es consistente con la definición de eventos independientes, la ocurrencia del evento B no influencia de ninguna manera la ocurrencia del evento A, luego la probabilidad de que el evento A ocurra dado que el evento B ha ocurrido o sea verdadero es la misma que la probabilidad del evento A.

Eventos Disjuntos:

Eventos A_1, A_2,… son disjuntos (mutuamente excluyentes) cuando, para $i \neq j$, $A_i \cap A_j = \emptyset$. Luego, si A y B son eventos disjuntos, tenemos que $P(A \cup B) = P(A) + P(B)$.

La propiedad de que dos eventos sean disjuntos se define sobre el espacio muestral, mientras que la independencia se define sobre la medida de probabilidad.

Ejemplo de eventos astronómicos disjuntos: Al observar una galaxia resuelta, los eventos disjuntos podrían ser: A= \`que sea una galaxia espiral´, B=\`que sea una galaxia elíptica´.

Ejemplo de eventos astronómicos independientes: : Al observar una galaxia, los eventos independientes pordrían ser: A=\`que la galaxia tenga un núcleo activo´, B=\`que mientras se la observa se produzca una explosión de supernova Ia´.

b) Probar que si dos eventos A y B son independientes ⇒ A^c y B^c son independientes.

A y B independientes $\Rightarrow P(A \cap B) = P(A)P(B)$

Por otro lado, tenemos que:

$$P(A^c)P(B^c) = \big(1 - P(A)\big)\big(1 - P(B)\big) = 1 + P(A)P(B) - P(A) - P(B)$$

y, utilizando las leyes de Morgan, $A^c \cap B^c = (A \cup B)^c \Longrightarrow P(A^c \cap B^c) = 1 - P(A \cup B)$.

La regla de la suma nos dice que $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, entonces:

$$P(A^c \cap B^c) = 1 - P(A \cup B) = 1 - P(A) - P(B) + P(A \cap B)$$

Usando que A y B son independientes esto resulta en:

$$P(A^c \cap B^c) = 1 - P(A) - P(B) + P(A)P(B)$$

que como vimos antes es igual a $P(A^c)P(B^c)$.

$$\therefore A \text{ } y \text{ } B \text{ } independientes \Longrightarrow A^c \text{ } y \text{ } B^c \text{ } independientes$$

Problema 2.

Intentaremos responder si, efectivamente, podemos concluir probabilísticamente que la mayoría de la gente escoge quedarse con su respuesta original.

Datos: $N_{personas}$=33; x=n° de personas que se quedan con la rta. original=18.

a) Suponemos que la gente escoge quedarse con su respuesta original con probabilidad 'r' y cambiar con probabilidad '1-r'. Nosotros estamos interesados en la distribución de la V.A. X: 'n° de veces que la gente escoge quedarse con su respuesta original'.

En este caso X~Binomial(N_p,r) donde N_p=33 es el único parámetro observado que tenemos de la distribución.

$$\Rightarrow P_X(x) = \binom{N_p}{x} r^x (1-r)^{N_p - x}$$

b) ¿Qué tan raro es observar X=18 si asumimos r=1/2?

La probabilidad de observar X=18 es $P(X = 18|r = 0.5) = \binom{33}{18}(0.5)^{33}$, donde $\binom{33}{18} = 1.037158 \times 10^9$ y $(0.5)^{33} = 1.164153 \times 10^{-10}$.

Luego $P(X = 18|r = 0.5) = 0.120741$

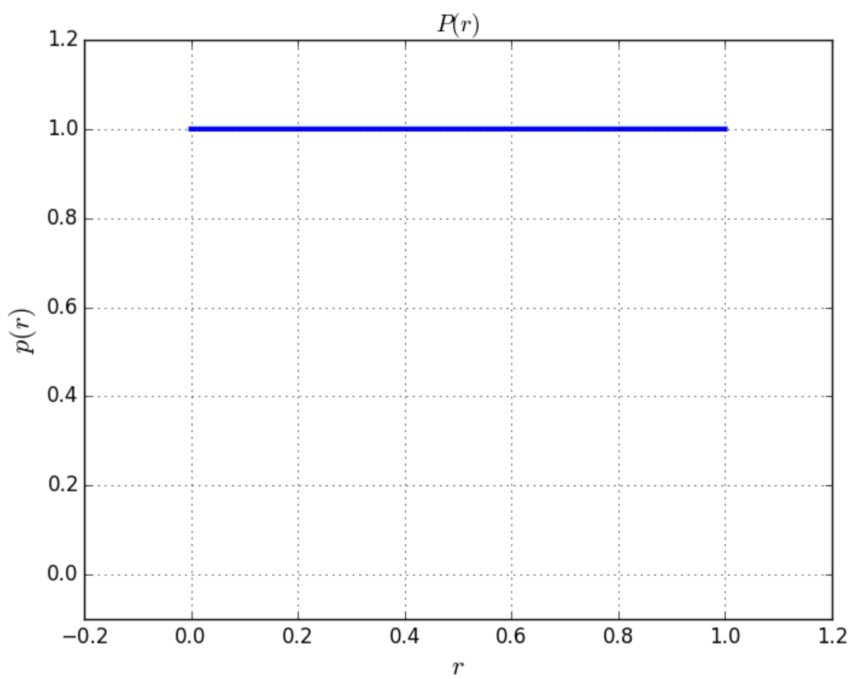
Del código escrito que simula 1000 votaciones (ver https://github.com/J-Minniti/Estadistica/blob/master/Problema_2_Tarea1.ipynb) obtenemos que el número de veces que observamos X=18 es 118. Esto nos da una probabilidad de ocurrencia dado r=0.5 de 118/1000=0.118. Para corroborar si esto es consistente con el resultado obtenido para $P(X = 18|r = 0.5) = 0.120741$, repetimos el cálculo del código anterior 300 veces para tener la distribución de la cantidad de veces que se observa X=18. Al hacer el cómputo, obtenemos que la media de esta distribución es 120.86, i.e., $P(X = 18|r = 0.5) \approx 0.12086$ con una desviación estándar $\sigma = 10.36$. Entonces el resultado obtenido en el código que simula 1000 votaciones ($P(X = 18|r = 0.5) = 0.118$) está a menos de 1σ de dicho valor. Los resultados son consistentes.

c) Buscamos comparar dos hipótesis: r>0.5 vs. r<0.5.

Queremos la distribución $P(r|X)$, es decir, la distribución de 'r' dado nuestros datos observados. Para obtenerla, utilizamos el Teorema de Bayes:

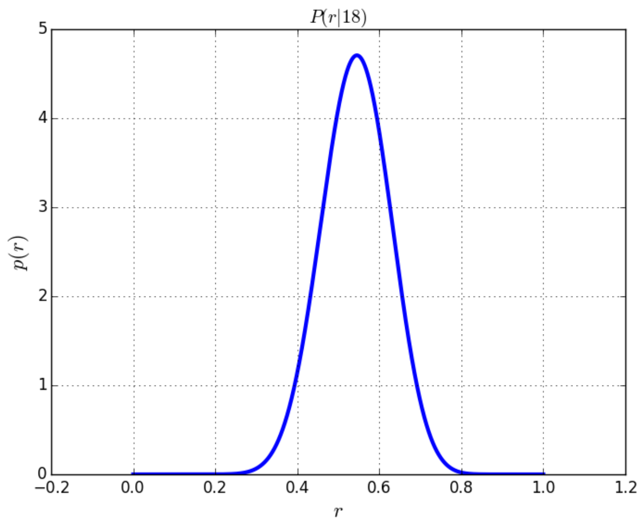
$$P(r|X) = \frac{P(X|r)P(r)}{\int_0^1 P(X|r')P(r')dr'} = \frac{P(X|r)}{\int_0^1 P(X|r')dr'}$$

Donde en la última igualdad usamos que $P(r) \sim Uniforme(0,1)$ (máxima incertidumbre) y $\int_0^1 P(X|r')P(r')dr' = \int_0^1 P(X|r')dr'$.



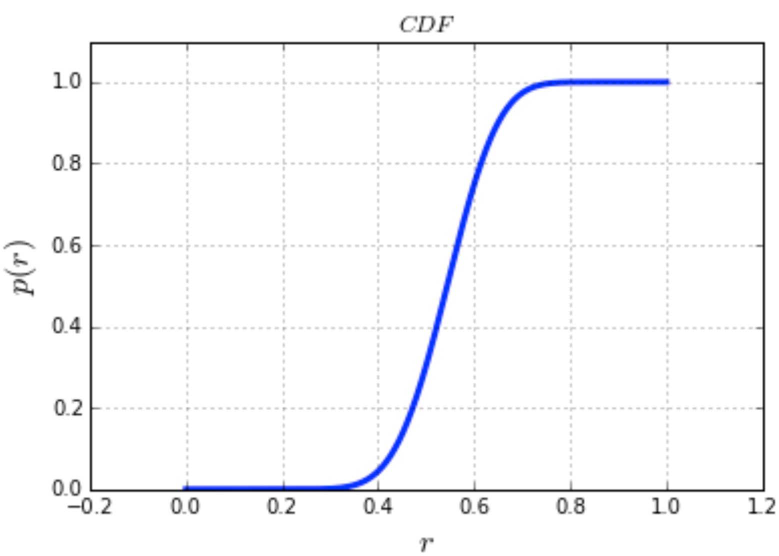
$$P(r|X = 18) = \frac{\binom{33}{18} r^{18}(1-r)^{15}}{\binom{33}{18}\int_0^1 r'^{18}(1-r')^{15}dr'}$$

i) El denominador fue calculado utilizando un integrador numérico (ver https://github.com/J-Minniti/Estadistica/blob/master/Problema_2_Tarea1.ipynb). El valor obtenido para el mismo es 0.0294117647059. La distribución obtenida para $P(r|X = 18)$ puede verse en la siguiente figura:



ii) El máximo de esta distribución se encuentra en r=0.54545

iii) Gráfico de la CDF:



iv) Hemos verificado (ver GitHub) que $\int_0^1 P(r|X = 18)dr = 1$.

d) Utilizando un integrador numérico, hemos calculado $P(r > 0.5|X = 18) \approx 0.70$ y $P(r < 0.5|X = 18) \approx 0.30$

El resultado obtenido en b), sumado a estos valores sugiere con cierto grado de confianza que, dados los datos obtenidos de la encuesta en Twitter, la mayoría de la gente escoge la peor estrategia, esto es, quedarse con la puerta original una vez que Monty Hall revela una puerta sin premio. Sin embargo, hay una probabilidad, no despreciable, de que en realidad la mayoría de la gente elija cambiarse a pesar de que la encuesta realizada en una muestra de 33 personas sugiera lo contrario. Para mejorar la confiabilidad de nuestro resultado deberíamos intentar aumentar la cantidad de gente que responda la encuesta.