Guia 9 V2

Agustin Muñoz Gonzalez

17/5/2020

Preparamos el entorno.

Limpiamos los registros.

```r
rm(list=ls())
kerrich=read.csv('Kerrich.csv',header = T)
attach(kerrich)
names(kerrich)
```

1. Primera Parte

1. Frecuencia Relativa y Probabilidad. En este ejercicio queremos retomar el estudio de frecuencias relativas de Kerrich, pero dado que no tenemos todas sus secuencias, realizaremos nuestras propias simulaciones. Para ello, consideramos el siguiente escenario:

- Experimento: lanzamos una moneda equilibrada.
- Evento: A = sale cara.

a) Guardar los resultados de simular m = 10 lanzamientos y determinar la frecuencia relativa de A.

Simulamos el lanzamiento de una moneda con el comando sample(c(0,1),1). Vamos a definir una función para cada cosa que pide el ejercicio.

```r
m_lanzamientos=function(m){
  resultados=c()
  for(i in (1:m)){
  resultados=c(resultados,sample(c(0,1),1))
  }
  resultados
}
# frec_rel_m_lanz=function(m){
#   sum(m_lanzamientos(m)==1)/m
# }
frec_rel_experimento=function(resultados){
  sum(resultados==1)/length(resultados)
}
```

b) Repetir utilizando m = 100

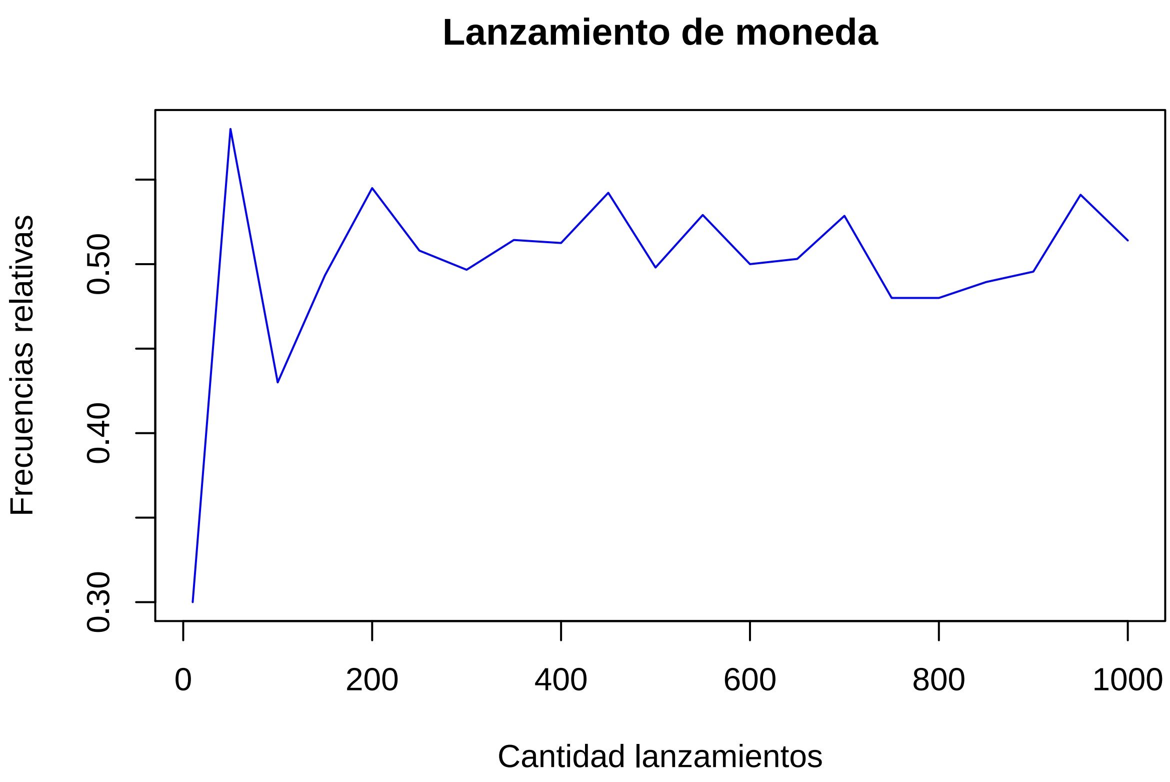
```
lanzamientos=m_lanzamientos(100)
frec_rel_experimento(lanzamientos)
```

```
## [1] 0.51
```

c) Para m en {10, 50, 100, . . . , 950, 1000} registrar la frecuencia relativa de A. Graficar m vs. frecuencia relativa. ¿Qué valor límite espera obtener?

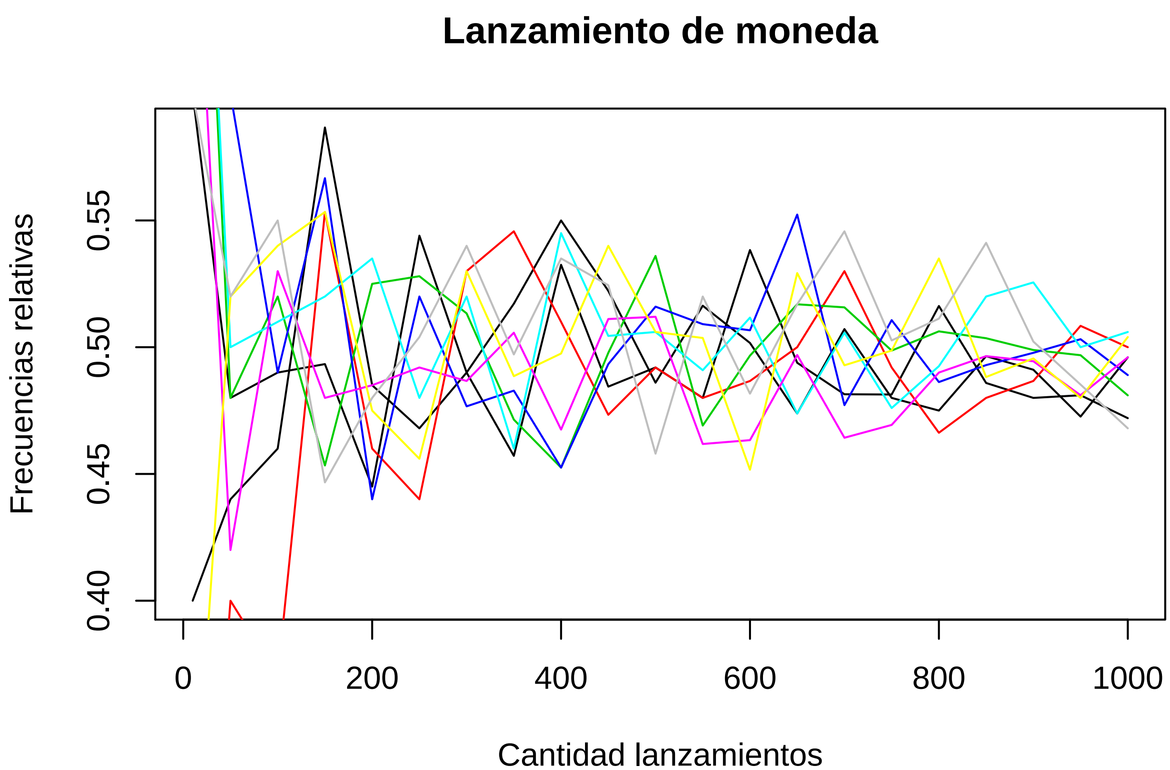
Suponiendo una moneda equilibrada esperaría obtener una frec rel de 0.5.

```
emes=c(10,50,seq(from=100,to=1000,by=50))
varios_experimentos=function(emes){
  experimentos=c()
  for(m in emes){
    experimentos=c(experimentos,list(m_lanzamientos(m)))
  }
  experimentos
}
experimentos=varios_experimentos(emes)
plot(emes,lapply(experimentos,frec_rel_experimento),
     xlab='Cantidad lanzamientos',
     ylab='Frecuencias relativas',
     col='blue', type='l', main='Lanzamiento de moneda')
```



d) Repetir el ítem anterior 10 veces y representar todos los resultados juntos utilizando para cada repetición un color diferente. ¿Qué concluye? ¿Cómo se compara con los reales resultados de Kerrich?

```
experimento_1=varios_experimentos(emes)
plot(emes,lapply(experimento_1,frec_rel_experimento),
     xlab='Cantidad lanzamientos',
     ylab='Frecuencias relativas',
     col='black', type='l', main='Lanzamiento de moneda')
tira_experimentos=c(list(experimento_1))
for(i in (2:10)){
  tira_experimentos=c(tira_experimentos,list(varios_experimentos(emes)))
  points(emes,lapply(tira_experimentos[[i]],frec_rel_experimento),
     col=palette()[i-2], type='l')
}
```



Todos los experimentos tienen una frecuencia relativa que ronda 0.5 y esto coincide con los resultados de Kerrich.

2. En una materia optativa la distribución de alumnos según carrera y género está dada por la tabla que definiremos a continuacion.

```
datos_materia=matrix(c(0.15,0.06,0.12,0.05,0.10,0.10,0.12,0.15,0.10,0.05),ncol=5,byrow = TRUE)
rownames(datos_materia)=c('Femenino','Masculino')
colnames(datos_materia)=c('Biología','Fisica','Computacion','Quimica','Matematica')
materia=data.frame(datos_materia)
attach(materia)
materia
```

```
##           Biología Fisica Computacion Quimica Matematica
## Femenino      0.15   0.06        0.12    0.05       0.10
## Masculino     0.10   0.12        0.15    0.10       0.05
```

Se elige un estudiante al azar. Calcular la probabilidad de que el estudiante elegido sea

(a) de género femenino y de Biología.

(b) de género fenenino.

(c) de Biología..

(d) de Biología o de género femenino.

(e) de Biología sabiendo que es de género femenino.

(f) ¿Es más probable que un estudiante sea Biólogo o Físico?

(g) Sabiendo que un estudiante es de género masculino, ¿es más probable que un estudiante sea Biólogo o Físico?

(h) Comparar los dos últimos items. Observar cómo la cosa cambia cuando tenemos informacion adicional....

En primer lugar si identificamos el genero masculino con el 0 y el femenino con el 1 y a la tira de carreras las asociamos con la tira c(1,2,3,4,5) entonces el espacio muestral podemos pensarlo como el conjunto $\{(g,c)|g \in \{0,1\}, c \in \{1,2,3,4,5\}\}$, con lo cual $\#S = 2^1 * 5^1 = 10$.

Cabe recordar además que la distribución es la probabilidad de que el evento ocurra, con lo cual los valores de la tabla podemos entenderlos efectivamente como probabilidades.

Respondamos cada item.

(a).

Viendo la tabla tenemos que la probabilidad de que el estudiante sea de genero femenino y de biologia es del 15%.

```
p_a=materia['Femenino','Biología']
p_a
```

```
## [1] 0.15
```

(b)

Para este item queremos P(A) con $A = \{(0,c)|c \in \{1,\ldots,5\}\} = \{(0,1),(0,2),\ldots,(0,5)\}$, es decir, debemos sumar todas las probas de la fila Femenino.

```
p_b=sum(materia['Femenino',])
p_b
```

```
## [1] 0.48
```

(c)

Debemos sumar todas las probas de la columna biologia.

```
p_c=sum(materia[,'Biología'])
p_c
```

```
## [1] 0.25
```

(d)

Queremos $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

```
p_d=p_b+p_c-p_a
p_d
```

```
## [1] 0.58
```

(e)

Ahora la cuestion es distinta porque nuestro espacio muestral ya no permite el sexo masculino. Una forma de pensar esto es imaginar que la fila de distribuciones de femenino es el espacio muestral total, con lo cual las distribuciones que alli se muestran deben sumar 1, es decir debemos suponer que la suma de todas esas distribuciones dan el 100% y calcular el porcentaje que representa cada una bajo esa suposición. O sea es la regla de 3 simple sum(materia[1,])—>1, p_e—>x=p_e/sum(materia[1,]).

```
proba_total_cond_fem=p_b
p_e=p_a*1/proba_total_cond_fem
p_e
```

```
## [1] 0.3125
```

(f)

Calculemos la proba de que sea físico y comparemos con p_c.

```
p_fisico=sum(materia[,'Fisica'])
p_fisico
```

```
## [1] 0.18
```

```
if(p_fisico>=p_c){'Es más probable que sea físico'}else{'Es más probable que sea biologo'}
```

```
## [1] "Es más probable que sea biologo"
```

(g)

Sabemos que el estudiante es masculino, con lo cual hay que reescalar las probas suponiendo que el espacio muestral es $\{(0, c)\}$ y luego ver cual es mas grande.

```
proba_total_cond_masc=sum(materia[2,])
p_cond_biol=materia['Masculino','Biología']/proba_total_cond_masc
p_cond_biol
```

```
## [1] 0.1923077
```

```
p_cond_fis=materia['Masculino','Fisica']/proba_total_cond_masc
p_cond_fis
```

```
## [1] 0.2307692
```

```r
if(p_cond_fis>=p_cond_biol){'Es más probable que sea físico'}else{'Es más probable que sea biologo'}
```

```
## [1] "Es más probable que sea físico"
```

(h)

Gracias a los items (f) y (g) observamos que la cosa cambió cuando nos restringimos al genero masculino, con lo cual las probabilidades condicionales no tienen por qué coincidir con la proba general pues la condicion nos cambia el espacio muestral y por tanto la proba final.

3. Teorema de la Probabilidad Total y de Bayes. En una población el 20 % de los adultos mayores practican actividad física con baja intensidad, el 55 % con intensidad media y el resto con intensidad alta.

- La probabilidad de que un adulto mayor que practica actividad física con intensidad baja sea hipertenso es 0.70
- La probabilidad de que un adulto mayor que practica actividad física con intensidad media sea hipertenso es 0.50
- La probabilidad de que un adulto mayor que practica actividad física con intensidad media sea hipertenso es 0.20

Si se elige al azar un adulto mayor de dicha población:

(a) calcular la probabilidad de que sea hipertenso.

(b) calcular la probabilidad de que practique actividad física con baja intensidad sabiendo que es hipertenso.

En primer lugar guardemos cada dato en distintas variables.

```
p_baja=0.2
p_med=0.55
p_alta=1-0.2-0.55
p_hiper_dado_baja=0.7
p_hiper_dado_med=0.5
p_hiper_dado_alta=0.2
```

(a)

Debemos calcular la proba de que sea hipertenso. Por el teo de la proba total podemos descomponer

$$P(hiper) = P(hiper \cap baja) + P(hiper \cap med) + P(hiper \cap alta).$$

A su vez usando la proba condicional sabemos que

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

y entonces nos queda

$$P(hiper) = P(hiper|baja)P(baja) + P(hiper|med)P(med) + P(hiper|alta)P(alta)$$

.

```
p_hiper=p_hiper_dado_baja*p_baja+p_hiper_dado_med*p_med+p_hiper_dado_alta*p_alta
p_hiper
```

```
## [1] 0.465
```

(b)

Quiero $P(baja|hiper)$ que por el teo de Bayes nos queda

$$P(baja|hiper) = \frac{P(hiper|baja) * P(baja)}{P(hiper)}.$$

```
p_baja_dado_hiper=p_hiper_dado_baja*p_baja/p_hiper
p_baja_dado_hiper
```

```
## [1] 0.3010753
```

4. Sensibilidad y Especificidad. Se realiza un análisis de laboratorio para diagnosticar cierta enfermedad. El test utilizado puede dar dos resultados: positivo, indicando que la persona está enferma; o negativo, sugiriendo que no hay enfermedad. En tal caso, se está frente a la posibilidad de cometer dos tipos de error en el diagnóstico:

(i) Falso positivo: diagnosticar como enferma a una persona sana.

(ii) Falso negativo: diagnosticar como sana a una persona enferma.

En este contexto, se definen los siguientes conceptos:

- Especificidad: es la probabilidad de que el análisis resulte negativo en un paciente sano.
- Sensibilidad: es la probabilidad de que el análisis resulte positivo en un paciente enfermo.
- Prevalencia: es la proporción de la población que padece la enfermedad.

En el siguiente enlace se encuentra información hipotética sobre una enfermedad y sobre la prueba utilizada para su detección. La información brindada queda determinada por su número de libreta o DNI. Utilizando tud datos, calcular la probabilidad de que un paciente esté enfermo sabiendo que el resultado de la prueba es positivo, para luego completar el formulario.

Guardo los datos asociados a mi numero de libreta.

```
Prevalencia =  0.02
Sensibilidad =  0.95
Especificidad =  0.95
```

Queremos entonces calcular $P(enfermo|positivo)$. Sabemos que

$$\begin{aligned} Prevalencia &= \frac{Enfermos}{Poblacion\_Total} = P(enfermo); \\ Sensibilidad &= P(positivo|enfermo); \\ Especificidad &= P(negativo|sano). \end{aligned}$$

Tenemos que

$$P(enfermo|positivo) = \frac{P(positivo|enfermo) * P(enfermo)}{P(positivo)},$$

donde

$$P(positivo) = 1 - P(negativo).$$

A su vez

$$P(negativo) = P(negativo|sano)P(sano) + P(negativo|enfermo)P(enfermo),$$

con

$$\begin{aligned} P(sano) &= 1 - P(enfermo); \\ P(negativo|enfermo) &= 1 - P(positivo|enfermo). \end{aligned}$$

Tenemos entonces

```r
p_positivo_dado_enfermo=Sensibilidad
p_negativo_dado_sano=Especificidad
p_enfermo=Prevalencia
p_negativo=p_negativo_dado_sano*(1-p_enfermo)+(1-p_positivo_dado_enfermo)*p_enfermo
p_enfermo_dado_positivo=p_positivo_dado_enfermo*p_enfermo/(1-p_negativo)
p_enfermo_dado_positivo
```

```
## [1] 0.2794118
```

¿No me dio muy bajo?¿Esta bien lo que hice?

5. Un poquito de independencia. . . .

a) Considerar n = 5 repeticiones independientes de un ensayo con dos posibles resultados: éxito (1), con probabilidad p y fracaso (0) con probabilidad 1 - p. En este archivo cada uno va a encontrar una secuencia de 1's y 0's resultantes de la realización de los 5 ensayos y un valor de p. Completar en la columna correspondiente con el cálculo de la probabilidad de observar la secuencia dada.

La secuencia que me tocó fue rep(0,5) con p=0.18. Debemos ver la probabilidad de que vuelva a tocar la misma secuencia con esa proba de exito. Es decir

$$P(x_1 = 0, x_2 = 0, \dots, x_5 = 0).$$

Como son experimentos independientes sabemos que $P(A \cap B) = P(A|B)P(B) = P(A)P(B)$. Con lo cual nuestra proba es

```r
p=0.18
proba_secuencia=(1-p)^5
proba_secuencia
```

```
## [1] 0.3707398
```

b) Caso Particular. Considerar n = 5 repeticiones independientes de un ensayo con dos posibles resultados: éxito (1), con probabilidad p = 0,8 y fracaso (0)con probabilidad 1 - p = 0,2. Se registra el valor del resultado obtenido en cada repetición; por ejemplo, el elemento (0, 1, 1, 0, 0) indica que en el segundo y tercer ensayo se obtuvo un éxito y en los demás un fracaso.

1) Describir el espacio muestral asociado a este experimento. Calcular el cardinal del espacio muestral.

El espacio muestral es $S = \{(x_1, x_2, \dots, x_5) | x_i \in \{0, 1\}\}$ y donde $\#S = 2^5$.

2) Calcular la probabilidad de cada elemento del espacio muestral y determine si es equiprobable.

Por el ítem (a) tenemos que la probabilidad de un elemento cualquiera es

$$P(x_i = 1, x_j = 0, i \in \{i_1, \dots, i_r\}, j \neq i) = 0.8^r 0.2^{5-r}.$$

El espacio no es equiprobable pues depende de las potencias r y $5 - r$, por ejemplo $0.8^5 \neq 0.2^5$.

2. Segunda Parte

Discretas y famosas...

6. La Binomial

a) Caso Particular. Considerar n = 5 repeticiones independientes de un ensayo con dos posibles resultados: éxito (1), con probabilidad p = 0,8 y fracaso (0)con probabilidad 1 - p = 0,2.

(i) Calcular la probabilidad de cada elemento del espacio muestral y determine si es equiprobable.

Tenemos 5 experimentos independientes de exito/fracaso con lo cual el espacio muestral es $S = \{(x_1, x_2, \ldots, x_5) | x_i \in \{0, 1\}\}$ y donde $\#S = 2^5$. Como los experimentos son independientes la proba de un elemento cualquiera de S es

$$P(x_i = 1, x_j = 0, 1 \leq i \leq r, r + 1 \leq j \leq 5) = 0.8^r 0.2^{5-r}.$$

No es un espacio equiprobable.

Sea X=cantidad de éxitos en las n = 5 repeticiones.

(ii) ¿Cuál es la probabilidad de que ocurra al menos un éxito? P (X > 0)

El suceso de que ocurra al menos 1 exito puede pensarse como el evento $A = \{(x_1, \ldots, x_5) | x_i = 1 \text{ para algun i}\}$. Notemos que $\#A = \sum_{r \geq 1} \binom{5}{r}$Entonces la proba es

$$P(A) = \sum_{r \geq 1} \binom{5}{r} 0.8^r 0.2^{5-r}.$$

Otra forma de calcular $P(A)$ es

$$P(A) = 1 - P(\text{ningun exito}) = 1 - (0.2)^5$$

(iii) ¿Cuál es la probabilidad de ocurra exactamente un éxito? P (X = 1)

Es la proba del evento $A = \{(x_1, \ldots, x_5) | x_i = 1 \text{ para un único i}\}$ con $\#A = \binom{5}{1}$. La proba entonces es

$$P(A) = \binom{5}{1} 0.8 0.2^4.$$

(iv) ¿Cuál es la probabilidad de que ocurran exactamente dos éxitos? P (X = 2)

Es la proba del evento $A = \{(x_1, \ldots, x_5) | x_i = 1 \text{ para exactamente dos i}\}$ con $\#A = \binom{5}{2}$. La proba entonces es

$$P(A) = \binom{5}{2} 0.8^2 0.2^3.$$

(v) ¿Cuál es la probabilidad de que ocurran a lo sumo dos éxitos? P (X <= 2)

Es la proba del evento $A = \{(x_1, \ldots, x_5) | x_i = 1 \text{ para a lo sumo dos i}\}$ con $\#A = \binom{5}{1} + \binom{5}{2}$. La proba entonces es

$$P(A) = \binom{5}{1} 0.8 0.2^4 + \binom{5}{2} 0.8^2 0.2^3.$$

b) Caso general. Considerar n repeticiones independientes de un ensayo con dos posibles resultados: éxito (1), con probabilidad p y fracaso 0) con probabilidad 1 - p.

Ahora tenemos n ensayos independientes, con lo cual en los eventos anteriores lo que cambia es el 5 por el n, i.e. $\binom{n}{r}$.

Sea X=cantidad de éxitos en las n repeticiones.

(i) ¿Cuál es la probabilidad de que no ocurra ningún exito? P (X = 0)

Es la proba del evento $A = \{(0, \dots, 0)\}$. La proba es

$$P(A) = (1-p)^n.$$

(ii) ¿Cuál es la probabilidad de que ocurra exactamente un éxito? P (X = 1)

Es la proba del evento $A = \{(x_1, \dots, x_n) | x_i = 1 \text{ para un único i}\}$ con $\#A = \binom{n}{1}$. La proba es

$$P(A) = \binom{n}{1} p(1-p)^{n-1}.$$

(iii) ¿Cuál es la probabilidad de que ocurran todos éxitos? P (X = n).

Es la proba del evento $A = \{(1, \dots, 1)\}$. La proba es

$$P(A) = p^n.$$

(iv) Indicar la cantidad mínima y máxima de éxitos que pueden ocurrir.

Pueden ocurrir al menos 1 exito y a lo sumo n.

(v) Considerar $k \in \{0, 1, \dots, n\}$, ¿cuál es la probabilidad de que ocurran exactamente k éxitos? P (X = k)

Es la proba del evento $A = \{(x_1, \dots, x_n) | x_i = 1 \text{ para exactamente k ies}\}$ con $\#A = \binom{n}{k}$. La proba es

$$P(A) = \binom{n}{k} p^k (1-p)^{n-k}.$$

(vi) ¿Cómo se llama X?

7. La Geométrica. Considerar el experimento que consiste en repeticiones independientes de un ensayo con dos posibles resultados, éxito y fracaso, hasta obtener el primer éxito.

c) Segundo Movimiento. Considerar el experimento del ítem anterior y definir la variable Y =cantidad de repeticiones del ensayo hasta obtener el primer éxito.

(i) ¿Qué valores puede tomar la variable Y ? Llamemos R_Y a este conjunto de valores posibles.

Puede tomar numeros naturales mayores a 0.

(ii) ¿Cuál es la probabilidad del resultado salió éxito en la k-ésima repetición? (P (Y = k), k en R_Y)

$P(Y = k) = (1 - p)^{k-1}p$

(iii) ¿Cómo se llama Y ?

d) Finale. Considerar el experimento del item anterior y definir la variable Z=cantidad de fracasos hasta obtener el primer éxito.

(i) ¿Qué valores puede tomar la variable Z? Llamemos R_Z a este conjunto de valores posibles.

Naturales y el 0.

(ii) ¿Qué relación hay entre Z e Y ? ¿Cuál es la probabilidad del resultado salieron k fracasos? (P (Z = k), k en R_Z)

Z=Y-1 porque hice Y repeticiones hasta que encontre el primer exito y el exito estuvo en el paso Y-esimo, con lo cual obtuve antes (Y-1) fracasos.

$P(Z = a) = P(Y = a + 1)$

e)

N=cantidad total de ensayos hasta k exitos=sum Y_k

T=total de fracasos hasta k exitos = N-k

T se llama binomial negativa. R_T=son los naturales y el cero.