Consultas - práctica 2

Variables aleatorias (discretas) famosas

Bernoulli: sirve para un experimento con probabilidad de éxito p.

$X \sim Be(p)$ ----> $p_X(X=1) = p$, $p_X(X=0) = 1-p$

$E(X) = p$.

$V(X) = p(p-1)$

Binomial: sirve para n experimentos independientes, cada uno con probabilidad de éxito p.
X = "cuántos de los experimentos son exitosos"

$X \sim B(n,p)$ ----> $p_X(X=k) = (n\ k)\ .\ p^k.\ (1-p)^{n-k}$, $k = 0, \dots, n$.

$E(X) = np$.

$V(X) = np(p-1)$.

Geométrica: Tengo experimentos independientes todos tienen la misma probabilidad p de éxito y cuento el "tiempo" hasta que fui exitoso la primera vez.
X = "el número de experimento donde me salió el primer éxito".
Rango(X) = 1, 2, 3, ...

$X \sim Ge(p)$ ----> $p_X(X=k) = (1-p)^{k-1}p$

$E(X) = 1/p$

$V(X) = (1-p)/p^2$.

Binomial negativa: Número de experimentos independientes, cada uno con proba de éxito *p*, hasta llegar a un número predeterminado *r* de éxitos.
Generaliza la geométrica.
X = "el número de experimento donde tuve éxitos r veces".
Rango(X) = r, r+1, ...

$X \sim BN(r,p)$ ----> $p_X(X=k) = (k-1\ r-1)\ p^r\ (1-p)^{k-r}$

(en otros lugares es $(k+r-1\ k)\ p^r\ (1-p)^k$, hice k+r experimentos, r tuvieron éxito y k fallaron. Ahí el rango es 0, 1, …)

$E(X) = r/p$

$V(X) = r(1-p)/p^2$.

Hipergeométrica:
Tengo una urna con N bolitas en total, K de ellas son de un color y N-K son de otro. Extraigo n bolitas de la urna, sin reposición.
X = "el número de experimentos donde me salió éxitos".
Rango(X) = 0, …, n

$X \sim H(N,K,n)$ N: bolitas en total, K: bolitas "exitosas", n: total de la extracción.

Proba puntual: $p_X(X=k) = (K\ k)\ (N-K\ n-k)\ /\ (N\ n)$ (cada uno de estos es un combinatorio).

$E(X) = n.N/K$

$V(X) = n.N/K.(N-K)/N.(N-n)/(N-1)$

(una hipergeométrica aproxima a una binomial(n,p) con p=N/K, cuando N es grande, o sea "sin reposición" ---> "con reposición").

Poisson: simula la cantidad de llegadas de personas/peticiones aleatorias, asumiendo que estas son independientes una de otra, y llegan con tasa promedio λ.
X = "el número de peticiones que recibo en un intervalo de tiempo".
Rango(X) = 0, 1, 2, 3, …

$X \sim P(\lambda)$

Proba puntual: $p_X(X=k) = e^{-\lambda} . \lambda^k /k!$

$E(X) = \lambda$

$V(X) = \lambda$

Una Poisson de parámetro lambda = n.p aproxima a una binomial de parámetros n, p (cuando tomamos n grande).

Ej. 2.1)

1. De un lote que contiene 15 artículos, de los cuales 4 son defectuosos, se eligen 3 artículos al azar con reposición. Si llamamos X al número de artículos defectuosos entre los seleccionados,

    a) Hallar la función de probabilidad puntual asociada a X y graficarla usando R.

    b) ¿Cuál es la probabilidad de que al menos 2 artículos sean defectuosos?

    c) Hallar la función de distribución acumulada de X y graficarla usando R.

    d) Estimar mediante una simulación las probabilidades calculadas en el item a).

Si tuviera que sacar un artículo defectuoso ("éxito"):
Tengo 15 en total, de los cuales 4 son malos y 11 son buenos.
Probabilidad de un éxito: 4/15

Para la segunda extracción, como es con reposición, la probabilidad de éxito es la misma: 4/15. O sea, la segunda extracción es independiente de la primera.

Para la tercera, lo mismo: es independiente de las demás y proba de éxito: 4/15.

X = número de éxitos que tuve (número de artículos malos que saqué en 3 extracciones).

Rango de X es {0,1,2,3}

$p_X(X=0)$ = casos favorables/casos totales = 11.11.11 / 15.15.15
(caso favorable: cada uno de los tres objetos _ _ _ es bueno, hay 11 formas de elegirlo. El total es 15, hay 15.15.15 ternas de objetos)

$p_X(X=1)$ = casos favorables/casos totales = 3.4.11.11 / 15.15.15

(caso favorable: de los tres objetos _ _ _ uno es malo y dos son buenos. Hay 3 formas de elegir cuál es el malo: el primero, el segundo o el tercero. Asigno uno de los 4 elementos malos a ese lugar, y asigno de entre los 11 buenos a los lugares restantes => 3.4.11.11)

$p_X(X=2)$ = casos favorables/casos totales = 3.4.4.11 / 15.15.15
(caso favorable: de los tres objetos _ _ _ uno es bueno y dos son malos. Hay 3 formas de elegir cuál es el bueno: el primero, el segundo o el tercero. Asigno uno de los 11 elementos buenos a ese lugar, y asigno de entre los 4 malos a los lugares restantes => 3.11.4.4)

$p_X(X=3)$ = casos favorables/casos totales = 4.4.4 / 15.15.15
(caso favorable: de los tres objetos _ _ _ todos son malos. Hay 4.4.4 formas de asignarlos => 4.4.4)

Con eventos: E_i=éxito (“sacar malo”) en el experimento i, i=1,2,3.
X=1 si y sólo si sucede E_1 pero no E_2, E_3, o E_2 pero no E_1, E_3, o E_3 pero no E_1, E_2
Para que X valga tiene que haber habido un éxito en un experimento pero fracaso en los demás.
Elijo uno de los tres experimentos para tener éxito -> combinatorio(3,1) = 3.
$P(E_i)$ = 4/15 = p, y $P(E_i^c)$ = 11/15 = 1-p.
$P(E_1 \cap E_2^c \cap E_3^c) = p\,(1-p)\,(1-p)$ (usando que son independientes estos eventos).
P(X=1) = (3 1)*p*$(1-p)^2$ = 3.4/15.11/15.11/15.
(me da lo mismo por cualquiera de los dos caminos)

Código en R:

```
Lote <- c(rep(1,4), rep(0,11))
#Malos es 1, buenos es 0 (no necesitamos este vector, pero si quisiéramos simular:
Extraccion <- sample(Lote, 3, replace = TRUE)

x <- c(0,1,2,3)
#Este es el rango de X

pX <- rep(0,4)
pX[1] <- 11*11*11/(15*15*15)
pX[2] <- 3*4*11*11/(15*15*15)
pX[3] <- 3*4*4*11/(15*15*15)
pX[4] <- 4*4*4/(15*15*15)
#Definimos proba puntual haciendo la cuenta a mano, para cada valor del rango
#El rango es 0, …, 3, pero 1, …, 4 son los índices del vector.

FX <- c(pX[1], pX[1]+pX[2], pX[1]+pX[2]+pX[3], pX[1]+pX[2]+pX[3]+pX[4])
#Definimos FX a mano para 0,1,2,3, sumando las probas puntuales.

plot(x,pX, ylim=c(0,1))
points(x, FX, ylim=c(0,1), col=2)
#Graficamos con puntos la proba puntual y la acumulada (faltaría extender las líneas
```

```
#a derecha para que sea la función de distribución.

#Para simular este experimento:
vectorX <- rep(0,10000)
pXsim <- rep(0,4)
#Armo vector para los 10000 resultados y las probas puntuales simuladas

for (i in 1:10000){
  vectorX[i] <- sum(sample(Lote, 3, replace = TRUE))
}
for (j in 0:3){
  pXsim[j+1] <- length(vectorX[vectorX==j])/10000
}
pXmat <- matrix(c(pXsim,pX), nrow=2, byrow=TRUE)
pXmat
```

Ej. 2.2)

2. Sea X una variable aleatoria con función de distribución

$$F_X(x) = \begin{cases} 0 & \text{si } x < 1 \\ 0.3 & \text{si } 1 \le x < 3 \\ 0.4 & \text{si } 3 \le x < 6 \\ 0.6 & \text{si } 6 \le x < 12 \\ 1 & \text{si } 12 \le x \end{cases}$$

a) Hallar la función de probabilidad puntual de X.

b) Calcular de dos maneras: utilizando la función de distribución y utilizando la función de probabilidad puntual, las siguientes probabilidades:

$$P(3 < X \le 6) \qquad P(3 \le X \le 6) \qquad P(X \ge 4) \qquad P(X \ge 6)$$

La función de probabilidad puntual es la que me dice cuánto vale la probabilidad de que X sea igual a ciertos puntos. Esos puntos son aquellos donde la función de distribución "pega un salto" (y si no pega un salto y F_X es constante, vale 0):

$$p_X(X = x) = F_X(x) - \lim_{t \to x^-} F_X(t)$$

(o sea, me voy acercando de la izquierda y veo si pega un salto F_X o no. Esta definición es que es la probabilidad de que X sea menor o igual a x fijo, y le voy restando los casos donde sea menor o igual a algo un poquito más chico).

En este caso, queda que $p_X(X=1) = 0,3$; $p_X(X=3) = 0,1$; $p_X(X=6) = 0,2$; $p_X(X=12) = 0,4$.

Fuera de esos puntos, vale 0.

Para calcular las probabilidades indicadas en b), uso la definición de $F_X(x) = P(X \le x)$
=> $P(a < X < b) = P(X \le b) - P(X=b) - P(X \le a) = F_X(b) - F_X(a) - p_X(b)$.

(o sea, la probabilidad de estar entre a y b es la de ser menor a b, pero le resto los casos en los que sea menor a a o igual a a). Si además tenemos menor o igual, etc., tendremos que tener en cuenta $p_X(a)$, etc..

Por otro lado, X sólo tiene como valores con probabilidad positiva a 1, 3, 6 y 12. Calcular con la probabilidad puntual es ver cuáles de estos valores caen en el rango de interés, y sumar esas probabilidades.

Entonces:

$P(3 < X \leq 6) = F_X(6) - F_X(3) = 0,6 - 0,4 = 0,2$

Y si no: $P(3 < X \leq 6) = p_X(6) = 0,2$ porque 6 es el único de los puntos "de interés" que caen en el intervalo (3,6].

$P(3 \leq X \leq 6) = F_X(6) - (F_X(3) - p_X(3)) = 0,6 - 0,4 + 0,1 = 0,3$

Usando que P(X<3) = P(X≤3) - P(X=3) = $F_X(3)$ - $p_X(3)$.

Y si no: $P(3 \leq X \leq 6) = p_X(3) + p_X(6) = 0,2 + 0,1 = 0,3$ porque 3 y 6 pertenecen a [3,6].

$P(X \geq 4) = 1 - P(X < 4) = 1 - F_X(4) = 1 - 0,4 = 0,6$

Usando que P(X<4) = P(X≤4) - P(X=4) = $F_X(4)$ - $p_X(4)$, pero $p_X(4)$=0 (o sea, que P(X<4) = P(X≤4), el 4 "no aporta probabilidad").

Y si no: $P(X \geq 4) = p_X(6) + p_X(12) = 0,2 + 0,4 = 0,6$ porque 6 y 12 pertenecen a $[4, +\infty)$ (o son los únicos valores posibles mayores o iguales a 4).

$P(X \geq 6) = 1 - P(X < 6) = 1 - (F_X(6) - p_X(6)) = 1 - 0,6 + 0,2 = 0,6$

Usando que P(X<6) = P(X≤6) - P(X=6) = $F_X(6)$ - $p_X(6)$.

Y si no: $P(X \geq 4) = p_X(6) + p_X(12) = 0,2 + 0,4 = 0,6$ porque 6 y 12 pertenecen a $[4, +\infty)$ (o son los únicos valores posibles mayores o iguales a 4).

c) Utilizando el comando `sample`, generar 5 realizaciones de esta variable aleatoria en R.

d) Mediante una simulación, estimar las probabilidades del item b).

Para el 2.c) y 2.d)
El comando sample: toma una muestra al azar de un vector (al azar = con proba uniforme).
p_X(X=1) = 0,3; p_X(X=3) = 0,1; p_X(X=6) = 0,2; p_X(X=12) = 0,4.
Estamos diciendo que 1 aparece el triple que 3, el 6 tiene el doble de probabilidad de salir que el 3, …

```
Xprima <- c(rep(1,3), rep(3,1), rep(6,2), rep(12, 4))
sample(Xprima, 5, replace = TRUE)
```

```
muestra <- sample(Xprima, 10000, replace = TRUE)
length(muestra[muestra == 1])/10000
```

Chequear: agregar prob a sample (para ponderar la extracción).
Efectivamente, sample lleva como argumentos sample(rango, número, replace = TRUE/FALSE, vectorproba)
Con vectorproba le puedo asignar probabilidades a los elementos del vector rango.
Por ejemplo, funciona:

```
vectorproba <- c(0.3,0.1,0.2,0.4)
rango <- c(1,3,6,12)
muestra <- sample(rango, 5, replace=TRUE, prob=vectorproba)
```

para hacer una muestra de 5 elementos, como pide el 2.c).
Con esto puedo simular muchas extracciones y ver que las probabilidades de cada valor efectivamente se parecen a las indicadas en vectorproba.

```
muestra10000 <- sample(rango, 10000, replace=TRUE, prob=vectorproba)
k <- 1
proba <- rep(0,4)
for (r in rango){
        proba[k] <- length(muestra10000[muestra10000==r])/10000
         k <- k+1
}
proba
```

Ej. 2.9)

9. El 70% de las consultas de un sistema interactivo de computación requiere de acceso a bases de datos. Un sistema recibe 25 consultas independientes unas de otras,

   a) ¿cuál es la probabilidad de que:

      i. exactamente 20 consultas requieran acceso a una base de datos?

      ii. el número de consultas que requieran acceso a una base de datos esté entre 20 y 24 inclusive?

   b) Calcular el valor esperado y la varianza del número de consultas que requieren acceso a una base de datos.

Diríamos que para una consulta dada la probabilidad de tener que acceder a la base de datos es 0,7 = p.
Experimento: hago 25 consultas independientes.

X = “número de consultas de estas 25 que requirieron acceso a una base de datos”.
X tiene distribución binomial con parámetros n=25, p=0,7.
(25 k) es el combinatorio 25, k. k=0,...,25.

$P_X(X=k) = (25\ k).0,7^k.0,3^{25-k}$.

Puedo pensar que hacer **una consulta** es como tirar una moneda con probabilidad(cara)=0,7, y probabilidad(ceca)=0,3 ----> Bernoulli(0,7) (una binomial es como sumar los resultados de n variables aleatorias Bernoulli).

a) $P_X(X=20)=(25\ 20).0,7^{20}.0,3^5$ ----> probabilidad de obtener exactamente 20 éxitos.
$P_X(X=20)+P_X(X=21)+P_X(X=22)+P_X(X=23)+P_X(X=24)$

R ----> comando pbinom

b) X es binomial ----> fórmulas para la esperanza y la varianza.
E(X) = n.p = 25.0,7 = 17,5.
V(X) = n.p.(p-1)

Ej. 2.10)

10. Se tienen dos dados, uno equilibrado y el otro cargado en el cual los números 1 y 2 tienen probabilidad 1/3 y el resto 1/12. Se elige un dado al azar y se lo arroja tres veces (independientemente). Sea X el número de veces que sale 1 ó 2.

   a) ¿Cuál es la distribución de X condicional a que se eligió el dado cargado?
   b) Hallar una expresión general para $p_X(k)$.

Dado 1 "normal" : 1,2, …, 6 con proba ⅙.
Dado 2 "cargado": 1, 2 con proba ⅓, 3, …, 6 con proba 1/12.

Elegimos un dado al azar, y lo arrojamos 3 veces. X = "número de veces que salió 1 o 2".

Rango de X = {0, 1, 2, 3}

a) La distribución de X condicional a que se eligió el dado cargado P(X menor o igual que t | dado cargado). Definimos D = "elegimos dado cargado", P(D) = 0,5.

Ahora paso a vivir en el universo donde pasó D.
Éxito = "sacar 1 o 2", E_i = éxito en la tirada i-ésima, i = 1, 2, 3.
Sale un 1 con proba ⅓, sale un 2 con proba ⅓ (estamos en el caso donde pasó D), entonces la probabilidad de 1 o 2 es ⅓ + ⅓ = ⅔.

$P(E_1|D)$ = ⅔, $P(E_1^c|D)$ = ⅓. Los eventos de éxito o de fracaso son independientes.

Vamos a calcular la proba puntual $p_X(X=x \mid D)$ ((3 1) es el combinatorio 3,1)
$p_X(X=0 \mid D)$ = P(tres fracasos) = ⅓ * ⅓ * ⅓ = 1/27.
$p_X(X=1 \mid D)$ = P(elijo un éxito de los 3 experimentos, los otros dos son fracasos)
= (3 1) * ⅔ * ⅓ * ⅓ = 3*2 /27 = 6/27.

$p_X(X=1 \mid D)$ = P(elijo dos éxito de los 3 experimentos, el otro es fracasos)
= (3 2) * ⅔ * ⅔ * ⅓ = 3*2*2 /27 = 12/27.
$p_X(X=3 \mid D)$ = P(tres éxitos) = ⅔ * ⅔ * ⅔ = 8/27.

(más adelante diremos que X|D (X condicionada a que salió D) tiene una distribución binomial B(3, ⅔)).

La función de distribución acumulada $F_{X|D}(t) = P(X<= t |D)$ (= sumar $p_X(X=x|D)$ hasta llegar a t) es

$F_{X\|D}(t) = 0$	$t < 0$
1/27	$0 \leq t < 1$
7/27	$1 \leq t < 2$
19/27	$2 \leq t < 3$
1	$3 \leq t$

b) $p_x(X=k)$ pero ahora no es la condicional.

Pensamos en términos de probabilidad total. D y D^c son una partición. Sumo las probabilidades de X=1 intersecado con D y D^c.

$p_X(X = 1) = P(X = 1 \cap D) + P(X = 1 \cap D^c)$

Cada una de estas intersecciones las escribo en función de una proba condicional:

$P(X = 1 \mid D) = P(X = 1 \cap D) / P(D) \Rightarrow P(X = 1 \cap D) = P(X = 1 |D) P(D)$

$P(X = 1 \mid D^c) = P(X = 1 \cap D^c) / P(D^c) \Rightarrow P(X = 1 \cap D^c) = P(X = 1 |D^c) P(D^c)$

$p_X(X = 1) = P(X = 1 |D) P(D) + P(X = 1 |D^c) P(D^c)$

$= (3\ 1). 2/3. (1/3)^2 . 1/2 + (3\ 1).\ 1/3 . (2/3)^2 . 1/2 = \ldots$

…

Si en vez de 1 lo hacemos para k genérico, k=0,1,2,3:

$$p_X(X = k) =$$
$$(3\ k). (2/3)^k. (1/3)^{3-k}. 1/2 + (3\ k).\ (1/3)^k . (2/3)^{3-k} . 1/2$$

k es el número de éxitos, la probabilidad de éxito si salía D era ⅔, y si salía D^c es ⅓.

Como antes, diríamos que X|D tiene una distribución binomial B(3, ⅔), pero que $X|D^c$ (X condicionada a que salió D^c) tiene una distribución binomial B(3, ⅓).

Ej. 2.11)

11. Para verificar si se cumplen las normas establecidas para arrojar residuos al río Reconquista, un inspector visita al azar 10 de las 50 industrias establecidas a orillas de dicho río.

    a) Si en realidad 35 industrias no cumplen con alguna de las normas, ¿cuál es la distribución del número de industrias visitadas que están en infracción? Calcular la probabilidad de que 6 de las industrias visitadas estén en infracción.

    b) Si hay 500 industrias de las cuales 350 están en infracción, aproximar la distribución de (a) por una más simple. Calcular nuevamente la probabilidad de que 6 de las industrias visitadas estén en infracción.

    c) Sea X el número de fábricas que están en infracción entre las 10 visitadas. Calcular $E(X)$ y $V(X)$ para las distribuciones exacta (a) y aproximada (b).

a) Tenemos 50 industrias, 35 que no cumplen, elegimos sin reposición 10 al azar para visitar. Planteamos X = “número de industrias que no cumplen alguna regulación”. X tiene distribución hipergeométrica de parámetros 50 (el total), 35 (“exitosas”), 10 (tamaño de la extracción sin rep.).

   P(X=6) = (35 6) (50-35 4) / (50 10) = …

b) Es el mismo planteo, pero ahora X_b tiene distribución H(500, 350, 10).
   P(X_b=6) = (350 6) (500-350 4) / (500 10) = …

   Yo espero que la probabilidad de visitar una industria y que no cumpla sea la proporción 35/50 o 350/500 = 0,7.
   Podemos plantear una binomial Y con distribución B(10, p=0,7). Y = “el número de industrias que no cumplen, haciendo 10 visitas con reposición”.
   Y aproxima a X_b, porque en 500 industrias la chance de volver a repetir una que ya visité haciendo 10 visitas es baja.
   P(Y=6) = (10 6) $0,7^6$ $0,3^4$ = …

c) Aplico las fórmulas de E(X) y V(X).

Ej. 2.12)

12. Una rueda de ruleta está dividida en 38 secciones, de las cuales 18 son rojas, 18 son negras y las 2 restantes son verdes. Sea X el número necesario de juegos hasta obtener una sección verde en jugadas independientes.

    a) ¿Cuál es la probabilidad de que sean necesarias al menos 4 jugadas?
    b) Hallar la función de distribución acumulada de la v.a. X.
    c) Si fueron necesarias 7 o más jugadas, ¿cuál es la probabilidad de que se necesiten al menos 10 jugadas? Comparar con (a).
    d) ¿Cuál es la probabilidad de que sea necesario un número impar de jugadas?
    e) Hallar $E(X)$ y $V(X)$.

X = "número de juegos hasta obtener un verde". Juegos sucesivos son independientes. Éxito: "sale verde". Fracaso: "rojo o negro". Diremos que la probabilidad de éxito en una tirada es 2/38 = 1/19.
$P(X = k) = (1-p)^{k-1}p$, k= 1, 2, ...

a) P(X >= 4) = 1 - P(X < 4) = 1 - P(X=1) - P(X=2) - P(X=3) = 0,850.
   = 1 - P(X <= 3) = 1 - $F_X(3)$.
b) $F_X(k)$ = P(X<=k) = P(X=1) + P(X=2) + … + P(X=k)
   Lo más fácil es calcular P(X>k) y usar que es el complemento de P(X<=k).

   $$P(X > k) = \sum_{i=k+1}^{\infty} (1-p)^{i-1} p = p \sum_{i=k+1}^{\infty} (1-p)^{i-1}$$

   Cambio de variables para empezar en 1 y no en k+1. Uso una variable j que empieza en 1. j = 1 pero empezamos en i=k+1 => i=j+k.

   $$P(X > k) = p \sum_{j=1}^{\infty} (1-p)^{j+k-1} = p\,(1-p)^k \sum_{j=1}^{\infty} (1-p)^{j-1} = p(1-p)^k \,.\, 1/(1-(1-p)) = (1 -$$

   Uso la fórmula de la suma geométrica:

   $$\sum_{i=0}^{\infty} r^i = 1/(1-r)$$

   Aplicando complemento,

   $$P(X \le k) = 1 - P(X < k) = 1 - (1-p)^k.$$

   Para armar la f de distribución yo tengo que dar su valor para todo t real.
   $F_X(t)$ = 0 si t<1
   $1-(1-p)^k$ si t está en el intervalo [k, k+1), k natural.

c) Fueron necesarias 7 o más tiradas => X>=7.
   P(X>=10 | X>=7) = P({X>=10}) / P(X>=7) = P(X>=4).
   Propiedad de la falta de memoria: P(X>9| X>6) = P(X>9)/P(X>6) = P(X>3).
   Chequear si vale la falta de memoria.

d) Probabilidad de que X sea impar? Repetir el argumento de la suma geométrica con un exponente 2k+1
P(X sea impar) = P(X=1) + P(X=3) + …
= suma de las probabilidades P(X=2k-1), k natural
$= \sum_{k=1}^{\infty} (1-p)^{(2k-1)-1} p = ...$

e) Reemplazar en la fórmula.

Ej. 2.14)

14. Con el fin de encontrar una palabra clave, un motor de búsqueda de internet explora una secuencia de sitios de la WEB en orden aleatorio. Al iniciar la búsqueda, el motor elige, al azar y con igual probabilidad, una entre dos secuencias posibles de sitios. Se sabe que el 10% de los sitios de la primera secuencia contienen esta palabra clave, mientras que sólo el 5% de los sitios de la segunda contienen dicha palabra.

    a) Si la búsqueda termina ni bien se encuentra un sitio que contenga la palabra clave, ¿cuál es la probabilidad de que más de 5 sitios deban ser explorados?
    b) Si se sabe que el motor de búsqueda encontró la palabra clave en la sexta visita ¿cuál es la probabilidad de que la haya encontrado en la segunda secuencia?
    c) Si la búsqueda termina cuando se encuentran 2 sitios que contenga la palabra clave ¿cuál es la probabilidad de que deban explorarse exactamente 10 sitios?

Parado sobre una secuencia dada, el número de sitios a explorar hasta encontrar la palabra clave se distribuye como una variable aleatoria geométrica de parámetro 0,1 (si es la primera secuencia) y 0,05 (si es la segunda).
X = “número de sitios a explorar hasta encontrar la palabra clave”
E = “elijo la primera secuencia de sitios”. P(E) = ½.
X|E tiene distribución Ge(0,1) y X|E^c tiene distribución Ge(0,05).

a) P(X>5) = P(X>5|E) P(E) + P(X>5| E^c) P(E^c) = $(1-0,1)^5$.½ + $(1-0,05)^5$. ½.

P(Y > 5) = 1 - P(Y <= 5) = 1 - $F_Y(5)$ = 1 - $(1-(1-p)^5)$ = $(1-p)^5$ si Y tiene dist. Ge(p) (Y es genérica).

b) P(E^c|X=6)
c) Z = “número de sitios a explorar hasta encontrar la palabra clave dos veces”.
Z|E tiene distribución BN(2, 0,1), y Z|E^c tiene distribución BN(2, 0,05).

P(Z=10) = P(Z=10|E) P(E) + P(Z=10| E^c) P(E^c)

P(Z=10|E) = (10-1 2-1) p^2 $(1-p)^{10-2}$ = (9 1) $0,1^2$ $0,9^8$ = $9.0,1^2.0,9^8$
P(Z=10|E^c) = (10-1 2-1) p^2 $(1-p)^{10-2}$ = (9 1) $0,05^2$ $0,95^8$ = $9.0,05^2.0,95^8$

Ej. 2.15)

15. Un minorista ha verificado que la demanda semanal de cajones de cierto producto es una v.a. con distribución de Poisson de parámetro $\lambda = 2$. Completa su existencia los lunes por la mañana de manera de tener 4 cajones al principio de la semana. Al efectuar un análisis de la actividad de su negocio, se le plantean las siguientes preguntas:

    a) ¿Cuál es la probabilidad de vender todo su stock durante la semana?

    b) ¿Cuál es la probabilidad de que sea incapaz de cumplir con un pedido por lo menos?

    c) ¿Cuál es la distribución del número de cajones vendidos en una semana?

    d) ¿Con cuántos cajones debería iniciar la semana a fin de que la probabilidad de cumplir con todos sus pedidos fuese mayor o igual que 0.99?

X = "demanda de cajones" tiene dist. Poisson(2).

a) P(X>=4)
b) P(X>4) (demanda mayor que stock)
c) Tengo que dar la proba puntual para caracterizar la v.a. Y "número de cajones vendidos con stock de 4 cajones". Rango de Y es {0,1,2,3,4}.
   P(Y=0) = P(X=0),
   P(Y=1) = P(X=1),
   P(Y=2) = P(X=2),
   P(Y=3) = P(X=3),
   P(Y=4) = P(X>=4) = p
   La distribución de la v.a. Y está dada por esta función de probabilidad puntual.
   $P(Y = 4) = \sum_{k=4}^{\infty} P(X = k) = \sum_{k=4}^{\infty} \lambda^k e^{-\lambda}/k!$ = 1 - P(X=0) - P(X=1) - P(X=2) - P(X=3) = p

d) La misma que con Y: Llamo Y_n "número de cajones vendidos si empiezo la semana con n cajones".
   $P(Y_n=0) = P(X=0)$, $P(Y_n=1) = P(X=1)$, …, $P(Y_n=n) = P(X>=n)$
   $P(Y_n=n) = P(X>=n) = 1 - P(X=0) - P(X=1) - … - P(X=n-1)$

   Cumplo con todos los pedidos si la demanda es menor o igual al stock presente. Quiero que la probabilidad del evento "cumplo con todos los pedidos", o sea, "demanda menor o igual al stock" = {X <= n} si tengo n en el stock. n = parámetro: con cuántos cajones arranco la semana (en el caso anterior era 4).

   P(X <= n) = P(X=0) + P(X=1) + … + P(X=n-1) + P(X=n).
   A mano: Ya conozco por a) P(X <= 4). Le sumo P(X=5). Le sumo P(X=6), y así.
   Queremos que P(X=0) + P(X=1) + … + P(X=n-1) + P(X=n) > 0,99

   Respuesta: 6.

Ej. 2.16)

16. Un bibliotecario ubica 1000 libros en un cierto día. Si la probabilidad de que un libro cualquiera sea mal ubicado es 0.001 y los libros se ubican en forma independiente, ¿cuál es la distribución aproximada del número de libros mal ubicados en ese día?

Utilizando esta distribución, calcular la probabilidad de que

a) por lo menos un libro sea mal ubicado ese día.

b) exactamente 3 libros sean mal ubicados ese día. Comparar con el valor exacto.

X = "número de libros mal ubicados en un día"
La tasa promedio de libros mal puestos, 0,001 es la probabilidad de poner mal un libro dado. Espero que la probabilidad de poner mal un libro de los mil sea 0,001*1000 = 1.
Si n = número de libros total, p = 0,001 probabilidad de poner mal un libro => X tiene distribución B(1000, 0,001).

Y = "el número de libros mal puestos en un día" y Y tiene dist. Poisson con parámetro 1, Y debería aproximar a X.

a)
P(X>=1) = 1 - P(X=0) = 1 - (1000 0) $0{,}001^{0}$ $0{,}999^{1000}$ = …
Aproximo:
P(Y>=1) = 1- P(Y=0) = 1 - 1^{0}. e^{-1} / 0! = …

b) P(X=3) = (1000 3) $0{,}001^{3}$ $0{,}999^{997}$ = 0,06130
P(Y=3) = 1^{3}. e^{-3} / 3! = 0,06139

Ej. 2.19)

19. Las tareas llegan a una cola de un sistema de computación con un solo servidor de acuerdo con un proceso de Poisson de parámetro $\lambda = 5$ tareas por minuto. Llamemos a dicho proceso X_t

  a) ¿Cuál es la probabilidad de que en 30 segundos lleguen menos de 5 tareas?

  b) Sea X_1 la cantidad de tareas recibidas en un minuto. Calcular:

$$\begin{array}{ll} P(X_1 \leq 6) & P(3 \leq X_1 \leq 6) \\ P(X_1 = 6) & P(3 < X_1 < 6) \\ P(X_1 \geq 5) & P(3 \leq X_1 \leq 6 | X_1 \geq 4) \end{array}$$

  Calcularlas exactamente utilizando los comandos `dpois` o `ppois` y luego estimarlas mediante una simulación (utilizar el comando `rpois`).

  c) ¿Cuál es el número esperado de tareas que se reciben en media hora?

  d) Estimar la esperanza calculada en el item anterior usando R.

Decir que tengo un proceso de Poisson X_t (t es real) es decir que a cada momento lo que estamos considerando sigue una distribución Poisson de parámetro λ*t.

a) $X_{30\ segundos}=X_{0,5}$ tiene distribución Poisson(5*0,5) o sea, Poisson(2,5).
   $P(X_{0,5} < 5)$ y uso la dist. conocida de $X_{0,5}$.

b) X_1 tiene distribución Poisson(5).
   $P(3 <= X_1 <=6) = P(X_1=3) + P(X_1=4) + P(X_1=5) + P(X_1=6)$
   $= 5^3.e^{-5}/3! + 5^4.e^{-5}/4! + 5^5.e^{-5}/5! + 5^6.e^{-5}/6! = \ldots$
   $P(3 < X_1 <6) = P(X_1=4) + P(X_1=5) = 5^4.e^{-5}/4! + 5^5.e^{-5}/5! = \ldots$
   $P(3 <= X_1 <=6 \mid X>=4) = P(4 <= X_1 <=6) / P(X>=4)$
   $= P(4 <= X_1 <=6) / (1- P(X=0) - P(X=1) - P(X=2) - P(X=3))$.

c) $X_{30\ minutos}=X_{30}$ tiene distribución Poisson(5*30).

rpois tiene argumentos rpois(n,lambda)

Ej. de clase práctica 4.

Urna con 3 bolitas rojas, 3 blancas, 5 negras: R -1, B 0, N +1.
Y = ganancia:

P(Y=-2) = P(RR) = (3 2) / (11 2) = 3 / 55
P(Y=-1) = P(RB) = (3 1) (3 1) / (11 2) = 9 /55
P(Y=-1) = P(BB) + P(RN) = (3 2) / (11 2) + (3 1) (5 1) / (11 2) = 18 /55
P(Y=1) = P(NB) = (5 1) (3 1) / (11 2) = 15 / 55
P(Y=2) = P(NN) = (5 2) / (11 2) = 10 /55.

Resolución - primera entrega de ejercicios.

Primer ejercicio:

Una determinada gripe está presente en un 6 % de la población. Existe un test para detectar la presencia de dicha gripe y se sabe que la probabilidad de que el test de positivo cuando se aplica a una persona sana es 0,12 y que la probabilidad de que el test de positivo cuando se aplica a un persona enferma es 0,82. Se elige una persona al azar de la población y se le realiza el test.
¿Cuál es la probabilidad de que el test dé positivo?
Sabiendo que el test dio positivo, ¿cuál es la probabilidad de que el individuo esté sano?

Definimos los eventos G “la persona está enferma de gripe” y T “el test dio positivo”. G^c quiere decir que la persona está sana, y T^c, que el test dio positivo. Que el 6 % de la población esté enferma de gripe me indica que P(G) = 0,06. Tenemos las siguientes cuatro situaciones (anotamos las probabilidades condicionales):

	G = persona enferma	G^c = persona sana
T = test positivo	0,82	0,12
N = test negativo	?	?

Para calcular P(T), consideramos que G y G^c constituyen una partición de las personas que hay. Entonces, podemos aplicar probabilidad total:

$$P(T) = P(T \cap G) + P(T \cap G^c) = P(T|G)\,P(G) + P(T|G^c)P(G^c) = 0,82.\,0,06 + 0,12.\,0,94 = 0,162$$

En la segunda pregunta nos piden que averigüemos $P(G^c|T)$, la probabilidad de que la persona esté sana dado que dio positivo el test. Por definición de probabilidad condicional:

$$P(G^c|T) = P(T \cap G^c)/P(T) = P(T|G^c)\,P(G^c)/P(T) = 0,12.\,0,94/0,162 = 0,6963$$

Segundo ejercicio:

Laura está cursando Probabilidad y Estadística (C) y Algoritmos I este segundo cuatrimestre de 2020. Ella considera que la probabilidad de desaprobar Probabilidad y Estadística (C) es 0,31 y la probabilidad de aprobar Algoritmos I es 0,53. Al mismo tiempo estima que la probabilidad de aprobar ambas materias es 0,24.
¿Cuál es la probabilidad de aprobar Probabilidad y Estadística (C) pero no Algoritmos I?
¿Cuál es la probabilidad de desaprobar ambas materias?

Consideremos los eventos E = “aprobar Probabilidad y Estadística (C)" y A = “aprobar Algoritmos I“. Sabemos que P(E) = 1 - 0,31 = 0,69, P(A) = 0,53 y $P(E \cap A) = 0,24$.
Aprobar Proba pero no Algo es E-A, y desaprobar ambas es la intersección de E^c y A^c (o el complemento de la unión de E y A).

$$P(E - A) = P(E) - P(E \cap A) = 0,69 - 0,24 = 0,45$$

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

Y $P(E \cup A) = P(E) + P(A) - P(E \cap A) = 0,69 + 0,53 - 0,24 = 0,98$, por lo que $P(E^c \cap A^c) = 0,02$.