Taller 1 entrega problema en grupo. MAT3 (estadística) GIN2 2020-2021 - Probabilidad, Variables Aleatorias, Distribuciones Notables 28-03-2020.

nombre1, apellido1_1 apellido1_22; nombre2, apellido2_1 apellido2_2;...

Contenidos

1 Taller1 evaluable. Entrega de problemas

Taller en grupo entregad las soluciones en .Rmd y .html o .pdf. o escribidlas de forma manual y escanear el resultado, en un solo fichero.

1.1 Problema 1

Sean A , B y C tres sucesos tales que $P(A) = 0.4$, $P(B) = 0.4$ y $P(A \cup B) = 0.6$. Calcular $P(A \cap B)$.

1.1.1 Solución

$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.4 + 0.4 - 0.6 = 0.2$.

1.2 Problema 2

Consideremos la v.a. continua X que tiene por función de densidad para a alguna constante $\alpha \in \mathbb{R}$:

$$f_X(t) = \begin{cases} \alpha \cdot t^4, & \text{si } -1 < t < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

1. Calculad α para que f_X sea densidad y especificad su dominio D_X.
2. Calculad la función de distribución de la v.a. X; $F_X(x) = P(X \leq x)$.
3. Calculad $E(X)$ y $Var(X)$.
4. Calcula en cuantil 0.9 de X.

1.2.1 Solución

Apartado 1.

$1 = \int_{-1}^{1} \alpha \cdot t^4 dt = \alpha \frac{t^5}{5}\Big|_{-1}^{1} = \alpha \left(\frac{1^5}{5} - \frac{(-1)^5}{5}\right) = \alpha \cdot \frac{2}{5}.$

luego tenemos que $\alpha = \frac{5}{2}$. y

$$f_X(t) = \begin{cases} \frac{5}{2} \cdot t^4, & \text{si } -1 < t < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Apartado 2.

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{, si } x \leq -1 \\ \displaystyle\int_{-1}^{x} \frac{5}{2} \cdot t^4 dt = \left[\frac{5}{2} \cdot \frac{t^5}{5}\right]_{t=-1}^{t=x} = \frac{5}{2}\left(\frac{x^5}{5} + \frac{1}{5}\right) & \text{, si } -1 < x < 1, \\ 1 & \text{, si } 1 \leq x. \end{cases}$$

$$E(X) = \int_{-1}^{1} x \frac{5}{2} \cdot x^4 dx = \int_{-1}^{1} \frac{5}{2} \cdot x^5 dx = \left[\frac{5}{2} \cdot \frac{x^6}{6}\right]_{-1}^{1} = \frac{5}{2}\left(\frac{1}{5} - \frac{1}{5}\right) = 0.$$

$$E(X^2) = \int_{-1}^{1} x^2 \cdot \frac{5}{2} \cdot x^4 dx = \int_{1}^{1} \frac{5}{2} \cdot x^7 dx = \left[\frac{5}{2} \cdot \frac{x^7}{7}\right]_{-1}^{1} = \frac{5}{2} \cdot \left(\frac{1}{7} + \frac{1}{7}\right) = \frac{5}{7}.$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{5}{7} - 0^2 = \frac{5}{7}.$$

Apartado 4

Nos piden $x_{0.9}$ el cuantil 0.9

$$0.9 = F_X(x_{0.9}) = \frac{5}{2} \cdot \left(\frac{x_{0.9}^5}{5} + \frac{1}{5}\right),$$

despejando de la ecuación

$$x_{0.9}^5 = 5 \cdot (0.9 \cdot \frac{2}{5} - \frac{1}{5}) = 0.8,$$

finalmente

$$x_{0.9} = \sqrt[5]{0.8} = 0.9563525.$$

1.3 Problema 3

Sea Y una variable discreta con función de probabilidad :

$$P_Y(y) = \begin{cases} \alpha \cdot \frac{1}{y^2} & \text{, si } y = -2, -1, 1, 2, \\ 0 & \text{en otro caso.} \end{cases}$$

1. Hallad α para que P_Y sea función de probabilidad.
2. Hallad la función de distribución $F_Y(y) = P(Y \leq Y)$.
3. Calculad $E(Y)$ y $Var(Y)$.
4. Calculad el cuantil 0.5 de de Y.

1.3.1 Solución

Apartado 1.

$$
\begin{aligned}
1=\sum_{y=-2,-1,1,2} P(Y=y) &= \sum_{y=-2,-1,1,2} \alpha \frac{1}{y^2} = \alpha \sum_{y=-2,-1,1,2} \frac{1}{y^2}\\
&= \alpha\left(\frac{1}{(-2)^2}+\frac{1}{(-1)^2}+\frac{1}{1^2}+\frac{1}{2^2}\right)\\
&= \alpha\left(\frac{1}{2}+2\right)=\alpha\cdot\frac{5}{2}.
\end{aligned}
$$

Luego $\alpha=\frac{2}{5}=0.4$ y la función de probabilidad es

$$
P_Y(y)=P(Y=y)=\begin{cases}
0.4\cdot\frac{1}{(-2)^2}=0.1 & \text{, si } y=-2,\\
0.4\cdot\frac{1}{(-1)^2}=0.4 & \text{, si } y=1,\\
0.4\cdot\frac{1}{1^2}=0.4 & \text{, si } y=-1,\\
0.4\cdot\frac{1}{2^2}=0.1 & \text{, si } y=2,\\
0 & \text{ en otro caso.}
\end{cases}
$$

Apartado 2.

$$
F_Y(y)=P(Y\leq y)=\begin{cases}
0 & \text{, si } y<-2,\\
0.1 & \text{, si } -2\leq y<-1,\\
0.1+0.4=0.5 & \text{, si } -1\leq y<1,\\
0.5+0.4=0.9 & \text{, si } 1\leq y<2,\\
0.9+0.1=1 & \text{, si } 2\leq y.
\end{cases}
$$

Apartado 3.

$$
E(Y)=\sum_{y=-2,-1,1,2} y\cdot P(Y=y)=-2\cdot 0.1+-1\cdot 0.4+1\cdot 0.4+2\cdot 0.1=0.
$$

$$
E(Y^2)=\sum_{y=-2,-1,1,2} y^2\cdot P(Y=y)=(-2)^2\cdot 0.1+(-1)^2\cdot 0.4+1^2\cdot 0.4+2^2\cdot 0.1=1.6.
$$

$$
Var(Y)=E(Y^2)-E(Y)^2=1.6-0^2=1.6.
$$

1.4 Problema 4

Tenemos un dado, bien equilibrado, de doce caras numeradas del 1 al 12 (dodecaedro dados de rol).

1. Calcular la función de probabilidad de la variables $X=$ número de la cara superior del dado en un lanzamiento, calcular E(X) y $Var(X)$.
2. Calcular la función de distribución de X y el cuantil 0.4.
3. Si Y es al v.a. que cuenta el número de veces que tiramos el dado hasta obtener el primer 5 calcular la función de distribución de Y.

1.4.1 Solución

Apartado 1.

$$P_X(x) = P(X = x) = \left\{ \begin{array}{ll} \frac{1}{12} & \text{, si } x = 1, 2, 3, \ldots, 11, 12, \\ 0 & \text{en otro caso.} \end{array} \right.$$

$$E(X) = \sum_{x=1}^{12} x \cdot P_X(x) = \sum_{x=1}^{12} x \cdot \frac{1}{12} = \frac{1}{12} \sum_{x=1}^{12} x = \frac{1}{12} \cdot \frac{12 \cdot (12+1)}{2} = 6.5.$$

$$E(X^2) = \sum_{x=1}^{12} x^2 \cdot P_X(x) = \sum_{x=1}^{12} x^2 \cdot \frac{1}{12} = \frac{1}{12} \sum_{x=1}^{12} x^2 = \frac{1}{12} \cdot \frac{12 \cdot (12+1) \cdot (2 \cdot 12 + 1)}{6} = \frac{13 \cdot 25}{6} = 54.16667.$$

$$Var(X) = E(X) - E(X)^2 = 54.16667 - 6.5^2 = 11.91667.$$

1.5 Problema 5

La proporción de niños pelirrojos es 1 cada 100. En una ciudad se produjeron 500 nacimientos (independientes) nacimientos en 2020, modelad mediante una distribución binomial la variable $X =$ número de niños pelirrojos nacidos entre los 500 niños. Utilizad R para calcular de forma exacta

1. La probabilidad de que ninguno de los nacidos ese año sea pelirrojo.
2. La probabilidad de que nazcan más de 2 niños pelirrojos
3. Repetir los cálculos con R utilizando una aproximación Poisson

1.5.1 Solución

La propbabilidad de pelirrojo es $p = \frac{1}{100}$, tener un hijo con el pelo pelirrojo se puede interperetar como un experimento Bernouilli $Ber(p = 0.01)$, en el que el éxito es que el niño sea pelirrojo. Si consideramos los nacimienro sucesos inpedendientes del mimo experimento $Ber(p = 0.01)$ tenemos que $X =$ número de niños pelirrojos entre $n = 500$ nacimientos seguirá una ley de distribución $B(n = 500, p = 0.01)$.

Nos piden la probabilidad de ningún pelirrojo entre 500

$$P(X = 0) = \binom{500}{0} 0.01^0 \cdot (1 - 0.01)^{500} \approx 0.00657.$$

Con R hacemos

```
n=500
p=0.01
choose(5,0)*0.01^0*(1-0.01)^500
```

```
## [1] 0.006570483
```

```
dbinom(0,size=n,prob=p)
```

```
## [1] 0.006570483
```

se observa que, de cualquier forma, obtenemos el mismo resultado.

La probabilidad de que nazcan más de 2 niños pelirrojos es

$$P(X > 2) = 1 - P(X \leq 2) = 1 - (P(X = 0) + P(X = 1) + P(X = 2))$$

```r
n=500
p=0.01
1-pbinom(2,size=n,p=p,lower.tail = TRUE)# por complementario
```

```
## [1] 0.8766142
```

```r
pbinom(2,size=n,p=p,lower.tail = FALSE)# agrupando a derecha
```

```
## [1] 0.8766142
```

```r
1-sum(dbinom(c(0,1,2),size=n,prob=p))
```

```
## [1] 0.8766142
```

La aproximación por Poisson consiste en suponer que X sigue, aproximadamente, una ley $Po(\lambda = n \cdot p = 500 \cdot 0.01 = 5)$

```r
dpois(0,lambda=5)
```

```
## [1] 0.006737947
```

```r
1-ppois(2,lambda=n*p,lower.tail = TRUE)# por complementario
```

```
## [1] 0.875348
```

```r
ppois(2,lambda=n*p,lower.tail = FALSE)# agrupando a derecha
```

```
## [1] 0.875348
```

```r
1-sum(dpois(c(0,1,2),lambda=n*p))
```

```
## [1] 0.875348
```

Como se ve los resultados son similares a los exactos con la distribución binomial.

1.6 Problema 6

Las consultas a una base datos llegan a un ritmo de medio $\lambda = 5$ peticiones por segundo. Sabemos que el nombre de peticiones que llegan en un segundo es una variable aleatoria que aproximadamente tienen una distribución de Poisson.

1. Calcular la probabilidad que lleguen más de 10 peticiones en 3 segundos utilizad R.
2. Calcular que entre una consulta y la siguiente pasen más de 0.5 segundos.
3. Calcular el cuantil 0.5 de $X_{t=10}$ numero de peticiones en 10 segundos utilizad R.

1.6.1 Solución

Tenemos que $\lambda = 5$ es el rimo de peticiones **por segundo**. Así en $t > 0$ segundos el número de peticiones en el intervalo $(0, t]$ tendrá un ritmo $\lambda \cdot t = 5 \cdot t$ peticiones.

En general X_t = número de peticiones en el intervalo $(0, t]$ seguirá una ley $Po(\lambda \cdot t = 5 \cdot t)$.

Apartado 1. Aquí nos piden $P(X_3 > 10)$ donde X_3 sigue una ley $Po(5 \cdot 3 = 15)$, $P(X_3 > 10) = 1 - P(X_3 \leq 10) = 1 - 0.1184644 = 0.8815356$.

Para el cálculo final hemo utilizado R

```r
ppois(10,lambda=15)# P(X<=10)
```

```
## [1] 0.1184644
```

```r
1-ppois(10,lambda=15) # 1-P(X<=10) por complementario
```

```
## [1] 0.8815356
```

```r
ppois(10,lambda=15,lower.tail=FALSE)# sumando la "cola" superior
```

```
## [1] 0.8815356
```

Apartado 2.

Sabemos que el tiempo T en segundos entre dos eventos consecutivos de un proceso Poisson de parámetro $\lambda = 5$ eventos por segundo sigue una ley $Exp(\lambda = 5)$ con proporción el ritmo o la ratio (*rate*) por segundo $\lambda = 5$ (para la exponencial en R el parámetro es `rate=5`)

Nos piden

$$P(T > 0.5) = 1 - P(T \leq 0.5) = 1 - \left(1 - ^{-5 \cdot 0.5}\right) `1 - (1 - exp(-5 * 0.5))` = `r1 - (1 - exp(-5 * 0.5)).$$

```r
1-(1-exp(-5*0.5))
```

```
## [1] 0.082085
```

```r
pexp(0.5,rate = 5)
```

```
## [1] 0.917915
```

```r
1-pexp(0.5,rate=5)# por complementario, rate=lambda=5
```

```
## [1] 0.082085
```

```r
pexp(0.5,rate=5,lower.tail=FALSE)# acumulado a derecha
```

```
## [1] 0.082085
```

Apartado 3.

Nos piden el cuantil 0.5 de $X_{t=10}$ numero de peticiones en 10 segundos. Al ser la v.a. discreta el cuantil $x_{0.5}$ es el primer valor entero tal que $P(X_{10} \leq x_{0.5}) \geq 0.5$, donde X_{10} sigue una ley $Po(5 \cdot 10 = 50)$.

Si lo calculamos con R $x_{0.5} = 50$, el código es el siguiente:

```r
qpois(p = 0.5,lambda=50)
```

```
## [1] 50
```

EXTRA

Podemos también calcular una lista grande de valores de $P(X_{10} \leq x)$ por ejemplo de $x = 30$ a $x = 60$

```r
# El primer valor que supere 0.5  es  el cuantil 0.5
round(ppois(0:60,lambda=50),4)
```

```
##  [1] 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [11] 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [21] 0.0000 0.0000 0.0000 0.0000 0.0000 0.0001 0.0001 0.0003 0.0005 0.0009
## [31] 0.0016 0.0027 0.0044 0.0070 0.0108 0.0162 0.0238 0.0340 0.0474 0.0646
## [41] 0.0861 0.1123 0.1435 0.1798 0.2210 0.2669 0.3167 0.3697 0.4249 0.4812
## [51] 0.5375 0.5927 0.6458 0.6959 0.7423 0.7845 0.8221 0.8551 0.8836 0.9077
## [61] 0.9278
```

```r
# Ejercicio interpretar el siguiente código
c(0:60)[which(round(ppois(0:60,lambda=50),4)>=0.5)][1]
```

```
## [1] 50
```

1.7 Problema 7

Tenemos que elegir entre dos programas (Prog1 y Prog2), el objetivo es elegir el programa más rápido en tiempo de respuesta en nuestro cluster de ordenadores. El tiempo de ejecución del Prog1 se ha modelado según una $N(\mu_1 = 100, \sigma_1 = 300)$ (la probabilidad de un tiempo de ejecución negativo es despreciable) y en Prog2 según una $N(\mu_2 = 90, \sigma_2 = 300)$. Utilizad R para el cálculo final de las probabilidades de la normal. (Utilizad R para el cálculo final)

1. ¿Qué Programa elegimos si queremos que el el 90% de los casos el tiempo de respuesta sea menor ?
2. Calcular la probabilidad de que el tiempo de ejecución sea mayor que 130 para cada algoritmo.

1.7.1 Solución

Apartado 1.

Para responder tenemos que calcular el cuantil $p = 0.9$ para X_1 y X_2 y que escoger el algoritmo con el menor cuantil 0.9

El cuantil para X_1 es

```
qnorm(p=0.9,mean=100,sd=300)
```

```
## [1] 484.4655
```

en el 90% de los casos el tiempo e ejecución del algoritmo 1 es menor que 484.4655.

El cuantil para X_2 es

```
qnorm(p=0.9,mean=90,sd=300)
```

```
## [1] 474.4655
```

en el 90% de los casos el tiempo e ejecución del algoritmo 1 es menor que 474.4655.

Así que respecto al cuantil 90% el algoritmo con menor cuantil 0.9 y por lo tanto mejor será el algoritmo 2.

Apartado 2.

Nos piden $P(X_1 > 130) = 1 - P(X_1 \leq 130) = 0.4601722$.

Lo hemos calculado con R de diversas formas:

```
pnorm(130,100,300)# P(X_1 <= 130), función de distribución
```

```
## [1] 0.5398278
```

```
1-pnorm(130,100,300)# 1_P(X_1 <= 130), cálculo por complementario
```

```
## [1] 0.4601722
```

```
pnorm(130,100,300,lower.tail=FALSE)# P(X_1 >130), cálculo directo por cola derecha
```

```
## [1] 0.4601722
```

Nos piden $P(X_2 > 130) = 1 - P(X_2 \leq 130) = 0.4469649$.

```
pnorm(130,90,300)# P(X_2 <= 130), función de distribución
```

```
## [1] 0.5530351
```

```
1-pnorm(130,90,300)# 1_P(X_2 <= 130), cálculo por complementario
```

```
## [1] 0.4469649
```

```
pnorm(130,90,300,lower.tail=FALSE)# P(X_2 >130), cálculo directo por cola derecha
```

```
## [1] 0.4469649
```

1.8 Problema 8

En la NBA el José Calderón fue en la temporada 2008-09 el jugador de baloncesto con mejor porcentaje tiros libres anotados un 98.05%.

Justificar los cálculos con notación matemática y haced el cálculo final con R.

1. ¿Cual es el valor esperando y la varianza del número tiros hasta acertar 10 tiros libres?
2. ¿Cuál es la probabilidad de que acierte al menos 40 tiros libres de forma consecutiva?
3. ¿Cuál es la probabilidad de que haga una serie de 100 tiros hasta obtener el tercer fallo?

1.8.1 Solución

Apartado 1 La probabilidad de que Calderón acierte canasta es $p = 0.9805$.

Sea X el número de tiros fallados (así que $X + 10$ es el número de tiros) hasta acertar 10 tiros libres, esta variable tendrá una distribución binomial negativa $BN(n = 10, p = 0.9805)$ pues cuanta el número de intentos para obtener 10 "éxitos" y se para. Si la definimos así sabemos que

$$E(X) = n \cdot \frac{1-p}{p} = 10 \cdot \frac{1-0.9805}{0.9805} = 0.199898.$$

Luego de media, incluidos los 10 aciertos, necesitará $E(10 + X) = 10 + 0.199898 = 10.199898$ no llega a 0.2 tiros antes de acertar los 10. Dicho de otra manera la media de tiros que necesita no llega 10.2 tiros libres para acertar 10.

$$Var(X) = n \cdot \frac{1-p}{p^2} = 10 \cdot \frac{1-0.9805}{0.9805^2} = 0.2038735.$$

Si hacemos la varianza del número de tiros es

$$Var(10 + X) = Var(X) = 0.2038735.$$

Apartado 2

En este caso la v.a. X = número de tiros acertados hasta el primer fallo sigue una distribución geimétrica $Ge(1 - 0.9805)$ que cuenta el número de aciertos ("fracasos") antes del primer fallo ("éxito" con probabilidad $1 - 0.9805 = 0.0195$).

Bajo estas condiciones sabemos que

$$P(X \leq x) = 1 - (1 - 0.195)^{x+1} = 1 - (0.9805)^{x+1}$$

si $x = 0, 1, 2, 3, \ldots$ (no contamos el fallo).

Nos piden

$$P(X > 40) = 1 - P(X \leq 40) = 1 - (1 - (1 - 0.0195)^{40+1}) = (0.9805)^{40+1} = 0.4460171.$$

Con R

```
0.9805^(40+1) #P(X>40)P(X>=41)  acertar más de 40 es acertar 41 o más X>=41
```

```
## [1] 0.4460171
```

```r
1-(1-(1-0.0195)^(40+1)) #P(X>40)=1-P(X<= 40)=1-(1-(1-0.0195)^(40+1))
```

```
## [1] 0.4460171
```

```r
1-pgeom(40,prob=1-0.9805) #1-P(X<=40)
```

```
## [1] 0.4460171
```

```r
pgeom(40,prob=1-0.9805,lower.tail=FALSE) # P(X>40)=P(X=41)+P(X=42)+.....
```

```
## [1] 0.4460171
```

```r
# .... acumulado la cola superior
```

Apartado 3

Ahora nos piden la probabilidad de que haga una serie de 100 tiros antes de tener tres fallos. Es decir que tire 100 veces acabando en fallo en el 100 y que el resto de de los 99 primero tiros tengan 2 fallos y 97 aciertos.

La variable que puede responder a esta pregunta es $X=$ número de tiros hasta obtener el tercer fallo. Esta variable sigue una ley $BN(n=3, p=1-0.9805)$.

Bajo estas condiciones sabemos que

$$P(X=k) = \binom{k-3+1}{k-1} \cdot 0.9805^k \cdot (1-0.9805)^3$$

para $k=0,1,2,3,\ldots$.

La probabilidad pedida es

$$P(X=97) = \binom{97-3+1}{3-1} \cdot 0.9805^{97} \cdot (1-0.9805)^3 = 0.0046205.$$

Con R, se puede hacer de varias maneras:

```r
choose(97+3-1,3-1)*0.9805^97*(1-0.9805)^3
```

```
## [1] 0.005325399
```

```r
dnbinom(97,size=3,prob=1-0.9805)
```

```
## [1] 0.005325399
```