Taller 2 entrega problemas - Probabilidad, Variables aleatoria , distribuciones notables y Teorema Central del Límite

SOLUCIONES DEL TALLER 2

Contenidos

1 Taller 2 evaluable. Entrega de problemas

Taller en grupo entregad las soluciones en Rmd y html o pdf, o si lo hacéis mano escaneado ...

1.1 Problema 1

Encuentra un ejemplo de tres sucesos A, B, C tales que A y B sean independientes, pero en cambio no sean condicionalmente independientes dado C.

1.1.1 Solución

Recordemos las definiciones

- Sucesos independientes: los sucesos A y B son independientes si $P(A \cap B) = P(A) \cdot P(B)$.
- Independencia condicional: A y B son sucesos condicionalmente independientes respecto de otro suceso C si $P(A \cap B/C) = P(A/C) \cdot P(B/C)$.

Consideremos el espacio muestral $\Omega = \{1, 2, 3, 4\}$ en el que $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = \frac{1}{4}$.

Sean los sucesos $A = \{1, 2\}$, $B = \{2, 3\}$ y $C = \{1, 3\}$ tenemos que

$P(A) = \frac{2}{4} = P(B) = P(C) = \frac{2}{4} = \frac{1}{2}$. Además tenemos que

$P(A \cap B) = P(\{2\}) = \frac{1}{4}$, $P(A \cap C) = P(\{1\}) = \frac{1}{4}$, $P(B \cap C) = P(\{3\}) = \frac{1}{4}$ y $P(A \cap B \cap C) = P(\emptyset) = 0$.

- Los sucesos A y B son independientes pues $P(A \cap B) = \frac{1}{4}$ y $P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.
- Los sucesos A y B NO son sucesos condicionalmente independientes respecto a C ya que $P(A \cap B/C) \neq P(A/C) \cdot P(B/C)$. Efectivamente $P(A \cap B/C) = \frac{P(A \cap B \cap C)}{P(C)} = \frac{0}{\frac{1}{2}} = 0$ que es distinta de $P(A/C) \cdot P(B/C) = \frac{P(A \cap C)}{P(C)} \cdot \frac{P(B \cap C)}{P(C)} = \frac{\frac{1}{4}}{\frac{1}{2}} \cdot \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

1.2 Problema 2

Verificar que:

$$F_X(t) = \begin{cases} 0, & \text{si } t < -1, \\ \frac{t+1}{2}, & \text{si } -1 \leq t \leq 1, \\ 1, & \text{si } t > 1, \end{cases}$$

es una función de distribución y hallar la función de densidad para X. Calcular también $P\left(-\frac{1}{2} \leq X \leq \frac{1}{2}\right)$.

1.2.1 Solución

Hay que comprobar que:

- $0 \leq F_X(t) \leq 1$: evidente
- $\lim_{t\to -\infty} F_X(t) = 0.$: evidente
- $\lim_{t\to +\infty} F_X(t) = 1.$: evidente
- Y que es continua (por la derecha): es continua así que lo es a derecha e izquierda

La densidad es la derivada de la distribución

$$f_X(t) = (F_X(t))' = \begin{cases} 0, & \text{si } t < -1, \\ \frac{1}{2}, & \text{si } -1 \leq t \leq 1, \\ 0, & \text{si } t > 1, \end{cases} = \begin{cases} \frac{1}{2}, & \text{si } -1 \leq t \leq 1, \\ 0, & \text{en cualquier otro caso.} \end{cases}$$

Nos piden esta probabilidad

$$\begin{aligned} P\left(-\frac{1}{2} \leq X \leq \frac{1}{2}\right) &= P\left(X \leq \frac{1}{2}\right) - P\left(X < -\frac{1}{2}\right) = F_X\left(\frac{1}{2}\right) - F_X\left(-\frac{1}{2}\right) \\ &= \frac{\frac{1}{2}+1}{2} - \frac{-\frac{1}{2}+1}{2} = \frac{\frac{3}{2}}{2} - \frac{\frac{1}{2}}{2} = \frac{1}{2}\cdot\left(\frac{3}{2} - \frac{1}{2}\right) = \frac{1}{2}. \end{aligned}$$

Problema 3

Sea Y una variable continua con función de densidad:

$$f_Y(y) = \begin{cases} 2(1-y), & \text{si } 0 < y < 1, \\ 0, & \text{en los otros casos.} \end{cases}$$

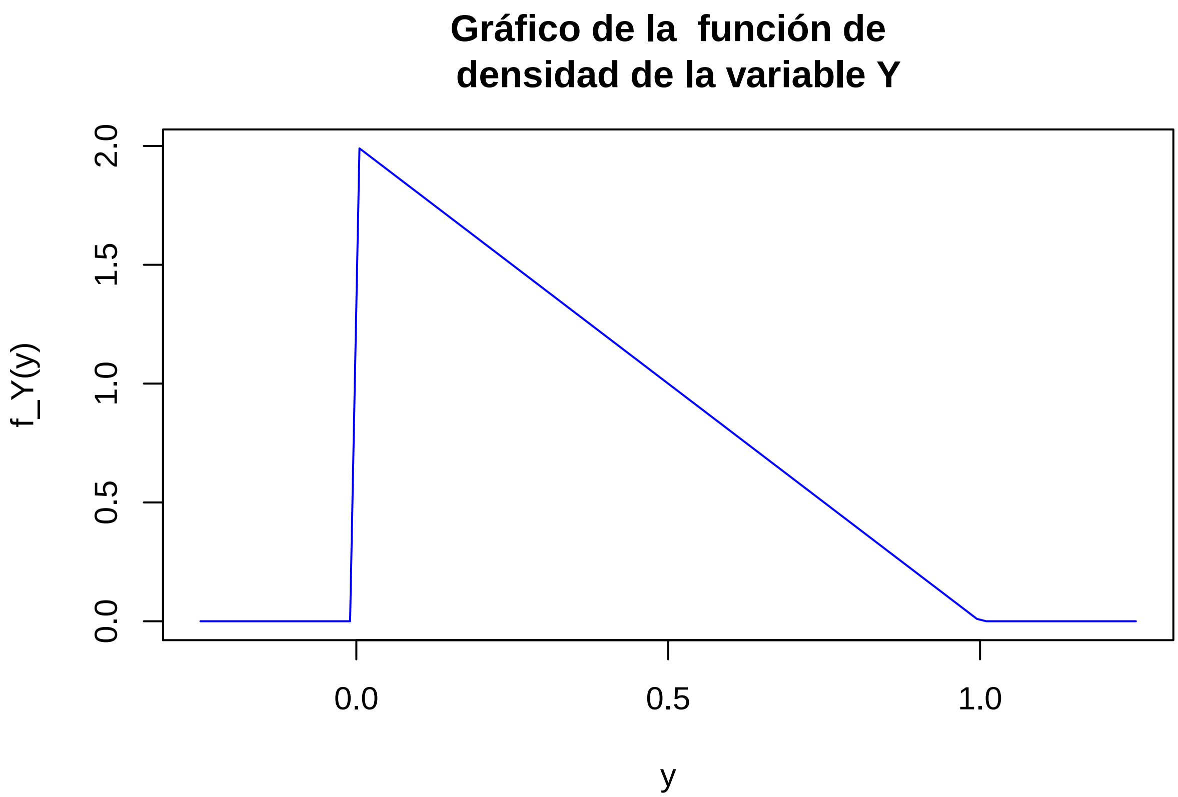
Hallar la función de distribución $F_Y(t)$.

1.2.2 Solución

Hagamos el dibujo con R

```
fY=function(y) { sapply(y, FUN=
                          function(x) {
                            if(x>0 & x<1){2*(1-x)} else {0}
                            }
                        )}

### cuidado curve pide que la función tenga la variable de nombre x
curve(fY(x),xlim=c(-0.25,1.25),xlab="y",
      ylab=expression("f_Y(y)"),
      main="Gráfico de la  función de\n  densidad de la variable Y",
      col="blue")
```



La punción de distribución es $F_Y(t) = P(Y \leq t) = \int_{-\infty}^{t} f_Y(y)dy$.

Haremos tres casos

- Si $t \leq 0$ entonces $F_Y(t) = P(Y \leq t) = \int_{-\infty}^{t} f_Y(y)dy = \int_{-\infty}^{t} 0dy = 0$.
- Si $0 < t < 1$ entonces $F_Y(t) = P(Y \leq t) = \int_{-\infty}^{t} f_Y(y)dy = \int_{-\infty}^{0} 0dy + \int_{0}^{t} 2\cdot(1-y)dy = 0 + \left[2\cdot y - y^2\right]_0^t = 2\cdot t - t^2 - (2\cdot 0 + 0^2) = 2\cdot t - t^2$.
- Si $1 \leq t$ entonces

$$\begin{aligned} F_Y(t) &= P(Y \leq t) = \int_{-\infty}^{t} f_Y(y)dy \\ &= \int_{-\infty}^{0} 0dy + \int_{0}^{1} 2\cdot(1-y)dy + \int_{1}^{+\infty} 0dy \\ &= 0 + \left[2\cdot y - y^2\right]_0^1 + 0 = 2\cdot 1 - 1^2 - (2\cdot 0 + 0^2) \\ &= 2 - 1 = 1. \end{aligned}$$

En resumen la función de distribución de Y es

$$F_Y(t) = \begin{cases} 0 & \text{si } t \leq 0, \\ 2\cdot t - t^2 & \text{si } 0 < t < 1, \\ 1 & \text{si } 1 \leq t. \end{cases}$$

1.3 Problema 4

Se lanza una moneda al aire hasta que sale cara. Supongamos que cada tirada es independiente de las otras y que la probabilidad de que salga cara cada vez es p.

- a) Demostrar que la probabilidad de que hagan falta un número impar de lanzamientos es $\frac{p}{1-q^2}$ donde $q = 1 - p$.

- b) Encontrar el valor de p tal que la probabilidad de que necesitemos un número impar de intentos sea 0.6.
- c) ¿Existe un valor de p tal que la probabilidad de que haga falta un nombre impar de intentos sea 0.5?

1.3.1 Solución

La variable X = numero de intentos PARA conseguir la primera cara sigue una ley $Ge(p)$ pero empezando en 1 pues contamos el intento en el que se obtiene cara. Entonces sabemos que su función de probabilidad es $P_X(x) = P(X = x) = (1-p)^{x-1} \cdot p$ para $x = 1, 2, 3, \ldots$ y cero en el resto de casos.

En el apartado a) nos piden la probabilidad de lanzamientos impares que es

$$\begin{aligned} P(X \text{ impar}) &= \sum_{k=0}^{\infty} P_X(X = 2\cdot k+1) = \sum_{k=0}^{\infty} (1-p)^{(2\cdot k+1)-1} \cdot p \\ &= p \cdot \sum_{k=0}^{\infty} \left((1-p)^2\right)^k = p \cdot \frac{1}{1-(1-p)^2} = \frac{p}{1-q^2} \\ &= \tfrac{p}{1-1+2p-p^2} = \tfrac{p}{2p-p^2} = \tfrac{1}{2-p}. \end{aligned}$$

En el apartado b) nos piden el valor de p tal que $P(X \text{ impar}) = \frac{1}{2-p} = 0.6$; resolviendo la ecuación obtenemos que $p = 2 - \frac{1}{0.6} = 2 - \frac{10}{6} = \frac{12-10}{6} = \frac{1}{3}$.

En el apartado c) nos piden p para que la probabilidad de impar y par sea la misma 0.5 despejando $\frac{1}{2-p} = 0.5$ obtenemos que $p = 0$ que NO es un valor posible para la variable; dejaría de ser variable aleatoria. Así que si contamos el intento del éxito/cara no es posible que la probabilidad de par e impar sean iguales.

1.4 Problema 5

La proporción de niños pelirrojos es 1 cada 10.000. En una gran ciudad se produjeron 5.000 nacimientos en 2020, aproximar por la distribución de Poisson la probabilidad que ninguno de los nacidos ese año sea pelirrojo. Aproximar la probabilidad de que nazca exactamente 1 niño pelirrojo y la de que hayan nacido al menos 2 pelirrojos.

1.4.1 Solución

Con los datos del problema sabemos que $P(\text{Pelirrojo}) = \frac{1}{1000}$. Suponiendo independencia entre el color del pelo de los 5000 nacimientos la variable X = número de pelirrojos entre 5000 nacimientos sigue una distribución binomial $B(n = 5000, p = \frac{1}{10000})$. Esta distribución se puede aproximar por una $Po\left(\lambda = n \cdot p = \frac{1}{2} = 0.5\right)$.

Aproximando por una Poisson tenemos que

$P(X = 1) = \binom{5000}{1}(1 - \frac{1}{10000})^{4999} \cdot \frac{1}{10000} \approx \frac{0.5^1}{1!} \cdot e^{-1} = 0.5 \cdot e^{-1} \approx 0.1839.$

Con R podemos calcular el valor exacto y el aproximando

```
dbinom(1,size = 5000,prob =1/10000)
```

```
## [1] 0.3032881
```

```
dpois(1,5000/10000)
```

```
## [1] 0.3032653
```

Ahora probabilidad de al menos dos Pelirrojos es $P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1) = 1 - (P(X = 0) + P(X = 1)) \approx 1 - \left(\frac{0.5}{0!} \cdot e^{-0.5} + \frac{0.5^1}{1!} \cdot e^{-0.5} \approx\right) \approx 0.9098.$

Con R podemos calcular el valor exacto y el aproximando

```
1-pbinom(1,size = 5000,prob =1/10000)
```

```
## [1] 0.09019643
```

```
1-ppois(1,5000/10000)
```

```
## [1] 0.09020401
```

1.5 Problema 6

Las peticiones a un servidor informático llegan a un ritmo de medio de 15 peticiones por segundo. Sabemos que el nombre de peticiones que llegan en un segundo es una variable aleatoria de Poisson.

- a) Calcular la probabilidad que no lleguen peticiones en un segundo.
- b) Calcular la probabilidad que lleguen más de 10 peticiones en un segundo.

1.5.1 Solución

La variables $X=$ número de peticiones por segundo y nos dicen sigue una ley $P0(\lambda = 15)$ luego la función de probabilidad es $P(X=x)=\frac{15^x}{x!}\cdot e^{-15}$ para $x=0,1,2,\ldots$.

En el apartado a) nos piden $P(X=0)=\frac{15^0}{0!}\cdot e^{-15}=e^{-15}\approx 3\times 10^{-7}$.

Con R

```
dpois(0,15)
```

```
## [1] 3.059023e-07
```

```
round(dpois(0,15),7)
```

```
## [1] 3e-07
```

En el apartado a) nos piden $P(X>10)=1-P(X\leq 10)=\sum_{x=0}^{1} 0\frac{15^x}{x!}\cdot e^{-15}$.

Es un cálculo farragoso, lo podemos hacer con las tablas de la Poisson o con R

```
1-ppois(10,15)
```

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

El cálculo manual

```
x=0:10
x
```

```
##  [1]  0  1  2  3  4  5  6  7  8  9 10
```

```
probabilidades=(15^x)/factorial(x)*exp(-15)
probabilidades
```

```
##  [1] 3.059023e-07 4.588535e-06 3.441401e-05 1.720701e-04 6.452627e-04
##  [6] 1.935788e-03 4.839470e-03 1.037029e-02 1.944430e-02 3.240717e-02
## [11] 4.861075e-02
```

```
1-sum(probabilidades)
```

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

Como se observa se obtiene el mismo resultado.

1.6 Problema 7

Tenemos que elegir entre dos tarjetas gráficas (TG1 y TG2) para entrenar su red neuronal. El tiempo de vida del la TG1 se ha modelado según una $N(\mu_1 = 120000, \sigma_1 = 140000)$ (la probabilidad de un tiempo de vida negativo es despreciable) y en TG2 según una $N(\mu_2 = 22000, \sigma_2 = 1000)$.

- a) ¿Qué tarjeta elegimos si el tiempo de duración objetivo del sistemas es de 20000 horas?
- b) ¿Y si es de 24000 horas?

1.6.1 Solución

Tenemos dos variables X_1 que sigue una ley $N(\mu_1 = 120000, \sigma_1 = 140000)$ y otra X_2 que tiene una distribución $N(\mu_2 = 120000, \sigma_2 = 1000)$

En el apartado a) nos piden que comparemos $P(X_1 > 20000)$ contra que $P(X_2 > 20000)$

En ambos casos $P(X_i < 20000) = 1 - P(X_i \leq 20000)$ para $i = 1, 2$

Con R la probabilidad de que $P(X_1 < 20000) = 1 - P(X_1 \leq 20000)$ es

```
1-pnorm(20000,mean=120000, sd=140000)
```

```
## [1] 0.7624747
```

y en el segundo caso $P(X_2 < 20000) = 1 - P(X_2 \leq 20000)$

```
1-pnorm(20000,mean=22000, sd=1000)
```

```
## [1] 0.9772499
```

Así que la opción 2 es mejor.

Para el apartado b). procediendo de manera similar, la opción 1 es mejor

la probabilidad de que $P(X_1 < 24000) = 1 - P(X_1 \leq 24000)$ es

```
1-pnorm(24000,mean=120000, sd=140000)
```

```
## [1] 0.7535534
```

y en el segundo caso $P(X_2 < 20000) = 1 - P(X_2 \leq 20000)$

```
1-pnorm(24000,mean=22000, sd=1000)
```

```
## [1] 0.02275013
```

1.7 Problema 8

La probabilidad de que un jugador de básquet enceste es p. ¿Cuántos lanzamientos tiene que hacer como mínimo (aproximadamente) para que la probabilidad de que la media de aciertos esté a distancia 0.01 de p sea de 0.99?

1.7.1 Solución

Sea

$$X_i = \begin{cases} 1 & \text{si acierta el i.-ésimo lanzamiento} \\ 0 & \text{si falla el iésimo lanzamiento} \end{cases}$$

Entonces $X_n = \sum_{i=1}^{n} X_i$ sigue una ley Binomial $B(n, p)$ la proporción de éxitos para X_n es $\hat{p_n} = \frac{X_n}{n}$. Nos piden encontrar n tal que

$$P(|\hat{p_n} - p| < 0.01) \geq 0.99$$

Ahora utilizaremos las propiedades básicas de la probabilidad y algunas manipulaciones algebraicas

$$\begin{aligned}
P(|\hat{p_n} - p| < 0.01) = & \quad P(-0.001 < \hat{p_n} - p < 0.01) = P(p - 0.01 < \hat{p_n} < p + 0.01) \\
= & \quad P\left(p - 0.01 < \tfrac{X_n}{n} < p + 0.01\right) = P\left(n \cdot (p - 0.01) < X_n < n \cdot (p + 0.01)\right) \\
= & \quad P\left(\tfrac{n \cdot (p-0.01) - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} < \tfrac{X_n - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} < \tfrac{n \cdot (p+0.01) + n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}}\right) \\
= & \quad P\left(\tfrac{-n \cdot 0.01}{\sqrt{n \cdot p \cdot (1-p)}} < Z < \tfrac{n \cdot 0.01}{\sqrt{n \cdot p \cdot (1-p)}}\right).
\end{aligned}$$

Por el Teorema central del límite sabemos que $Z = \frac{X_n - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}}$ se aproxima a una distribución $N(0,1)$ cuando $n \to +\infty$. Utilizando esta aproximación y las propiedades una Z normal estándar tenemos que

$$\begin{aligned}
P(|\hat{p_n} - p| < 0.01) \quad & = P\left(\tfrac{-n \cdot 0.01}{\sqrt{n \cdot p \cdot (1-p)}} < Z < \tfrac{n \cdot 0.01}{\sqrt{n \cdot p \cdot (1-p)}}\right) \\
& = P\left(Z \leq \tfrac{n \cdot 0.01}{\sqrt{n \cdot p \cdot (1-p)}}\right) - P\left(Z \leq \tfrac{-n \cdot 0.01}{\sqrt{n \cdot p \cdot (1-p)}}\right) \\
& = 2 \cdot P\left(Z \leq \tfrac{n \cdot 0.01}{\sqrt{n \cdot p \cdot (1-p)}}\right) - 1.
\end{aligned}$$

Hemos reducido el problema a encontrar n tal que

$$2 \cdot P\left(Z \leq \frac{n \cdot 0.01}{\sqrt{n \cdot p \cdot (1-p)}}\right) - 1 \geq 0.99.$$

operando

$$P\left(Z \leq \frac{n \cdot 0.01}{\sqrt{n \cdot p \cdot (1-p)}}\right) = \frac{0.99 + 1}{2} \geq 0.995.$$

Entonces lo que buscamos es un cuantil de la normal que es igual al menos a 0.995.

```r
qnorm(0.995)
```

```
## [1] 2.575829
```

Luego tengo la ecuación $\frac{n \cdot 0.01}{\sqrt{n \cdot p \cdot (1-p)}} = 2.575829$ elevando al cuadrado toda la ecuación

$$\frac{n^2 \cdot 0.01^2}{n \cdot p \cdot (1-p)} = 2.575829^2$$

operando obtenemos que

$$n = \frac{2.575829^2 \cdot p \cdot (1-p)}{0.01^2}$$

Ahora resulta que p es desconocida pero como ya vimos en las presentaciones esta función $p \cdot (1-p)$ alcanza su máximo en $p = 0.5$.

Así que en el peor de los casos

$$n = \frac{2.575829^2 \cdot 0.5 \cdot (1-0.5)}{0.01^2} \approx 16587.24.$$

Así que tomando el entero superior $n = 16588$ es el número de canastas que el jugador debe lanzar para, en el peor de los casos, estimar su proporción de tiros acertados con un error inferior a 0.001 y con una probabilidad superior al 0.99. Son MUCHOS TIROS, realmente el TCL hace una aproximación bastante grosera, y en el peor de los casos, pero es un sencillo ejemplo para comprender la utilizad de esta aproximación.

1.8 Problema 9

Sea $X_1, \ldots, X_n$ con $n = 48$, una muestra aleatoria simple de una variable aleatoria uniforme en el intervalo $(0, a)$. Aplicando el Teorema Central del Límite, hallar la probabilidad aproximada de que $\sum\limits_{i=1}^{n} X_i > a$.

1.8.1 Solución

Tenemos una muestra aleatoria simple $X_1, \ldots X_{n=48}$ de una v.a. X con distribución uniforme $U(0, a)$ entonces $\mu_X = E(X) = \frac{a}{2}$ y $\sigma_X^2 = \frac{(0-a)^2}{12} = \frac{a^2}{12}$ por lo tanto $\sigma_x^2 = \frac{a^2}{12} = \frac{a}{\sqrt{12}}$.

Entonces $E\left(\sum_{i=1}^{48} X_i\right) = 48 \cdot \frac{a}{2} = 24 \cdot a$, $\sigma^2_{\sum_{i=1}^{48} X_i} = 48 \cdot \frac{a^2}{12} = 4 \cdot a^2$ y por lo tanto la desviación típica de la suma es $\sigma_{\sum_{i=1}^{48} X_i} = +\sqrt{\sigma^2_{\sum_{i=1}^{48} X_i}} = \sqrt{4 \cdot a^2} = 2 \cdot a$.

Límite sabemos que Bajo estas condiciones y por el Teorema Central del Límite sabemos que $\sum_{i=1}^{48} X_i$ sigue aproximadamente una distribución $N(24 \cdot a, 2 \cdot a)$.

Luego $P\left(\sum_{i=1}^{48} X_i > a\right) = 1 - P\left(\sum_{i=1}^{48} \leq a\right) \approx 1 - P\left(Z \leq \frac{a - 24 \cdot a}{2 \cdot a}\right) = 1 - P\left(Z \leq \frac{-23}{2}\right) \approx 1 - 0 = 1.$