6. Variables aleatòries discretes

Universitat de Barcelona

UNIVERSITAT DE BARCELONA

Variables aleatòries

Donat un experiment aleatori, una variable aleatòria és una 'funció' tal que per cada valor de l'espai mostral Ω li fa correspondre un número real.

Una variable aleatòria és:

- discreta si pren una quantitat numerable de valors.
- contínua si pot prendre tots els valors d'un interval.

Exemple

- *Tirem 10 vegades una moneda*
  - $X=$ *'número de creus'*
  - $Y=$ *'Diferència en valor absolut del número de cares i creus'*
- *Seleccionem una empresa de Barcelona a l'atzar*
  - $Z=$ *'número de treballadors'*
  - $T=$ *'guanys mensuals'*
- *Seleccionen un alumne un cop acabats els exàmens del 1r trimestre*
  - $R=$*'número d'assignatures aprovades aquest semestre'*
  - $S=$*'nota mitjana obtinguda'*

Funció de massa de probabilitat (f.m.p.)

X v.a. discreta.

La funció de massa de probabilitat de X, que denotem per $P_X(x)$, representa la probabilitat de que X prengui el valor x, com a funció de x, és a dir

$$P_X(x) = P(X = x)$$

on aquesta funció s'avalua per a tots els possibles valors de x.

Les funcions de massa satisfan,

1. $0 \leq P_X(x) \leq 1$
2. $\sum_x P_X(x) = 1$

Exemple (Rifa)

En una rifa ofereixen un premi de 600 euros, dos de 300 euros i vint de 100 euros.

Sabent que es van vendre 10.000 bitllets al preu de 30 cèntims el bitllet, calcula la funció de massa de la variable aleatòria X = "guany a la rifa"

Solució:

x	$P(X = x)$
$-0,30 + 600 = 599,70$	$\frac{1}{10^4}$
$-0,30 + 300 = 299,70$	$\frac{2}{10^4}$
$-0,30 + 100 = 99,70$	$\frac{20}{10^4}$
$-0,30$	$1 - \frac{1}{10^4} - \frac{2}{10^4} - \frac{20}{10^4} = \frac{9977}{10^4}$

La probabilitat de que no guanyem diners és, $P(X \leq 0) = \frac{9977}{10^4}$

Número de sisos en la tirada de tres daus

Tirem tres daus de sis cares, i observem el número de sisos obtinguts. Definim X com el número de sisos obtinguts en una tirada de tres daus.

1. Valors possibles de X
2. Probabilitat dels valors possibles
3. $P(X > 1)$
4. $P(X \leq 2)$
5. $P(2 < X \leq 3)$
6. $P(X = 0)$

Número de sisos en la tirada de tres daus

Funció de massa de probabilitat:

valors X	$p(x)$
0	$\binom{3}{0}(\frac{5}{6})^3(\frac{1}{6})^0 = 0.5787$
1	$\binom{3}{1}(\frac{5}{6})^2(\frac{1}{6})^1 = 0.3472$
2	$\binom{3}{2}(\frac{5}{6})^1(\frac{1}{6})^2 = 0.0694$
3	$\binom{3}{3}(\frac{5}{6})^0(\frac{1}{6})^3 = 0.0046$

R:

$$\binom{3}{1}\left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right)^1 = \texttt{choose}(3,1) * (5/6)^2 * (1/6)^1 = 0.3472$$

Esperança: definició

Sigui X una v.a. discreta que pren valors $x_1, x_2, \ldots, x_K$ amb probabilitats respectives $p_1, p_2, \ldots, p_K$.

El valor esperat de X (esperança de X, $E(X)$, μ_X) és la suma ponderada dels valors de la variable per les seves probabilitats

$$E(X) = x_1 \cdot p_1 + x_2 \cdot p_2 + \ldots x_K \cdot p_K$$

Escriurem,

$$\boxed{E(X) = \sum_x x \cdot P(X = x)}$$

Exemples

Exemple (Rifa)

X='guany a la rifa'.

$$E(X) = 599.7 \cdot \frac{1}{10^4} + 299.7 \cdot \frac{2}{10^4} + 99.7 \cdot \frac{20}{10^4} + (-0.3) \cdot \frac{9977}{10^4} = 0,02$$

Exemple (Joc)

En una capsa tenim 4 boles etiquetades amb $1, 1, 2, 4$. Traiem una bola a l'atzar i guanyem en euros el valor X de la bola extreta. Aleshores,

$$E(X) = 1 \cdot \frac{2}{4} + 2 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} = 2$$

Simulacions

```
# Exemple 1
> guanys<- sample(c(1,2,4),1000,replace = T, prob=c(0.5, 0.25, 0.25))
> guanys
   [1] 1 2 1 2 1 2 4 4 1 1 1 4 2 4 2 2 1 1 2 2 1 4 1 1 1 4 1
       4 2 4 1 1 1 4 1 1 2 4 1 1 2 1 1 1 1 1 4
....
 [988] 4 4 1 2 2 1 1 4 4 1 1 2 1
> mean(guanys)
 1.947
# Exemple 2
> capsa = c(1,1,2,4)
> simul=sample(capsa,1000,replace=T)
> simul
  [1]  1 1 1 4 1 1 2 4 2 1 4 1 1 1 1 1 4 4 4 1 1 4 4 1 1 1
       1 1 4 1 4 2 4 4 1 1 4 1 1 1 2 1 2 2 4 4 1
     .....
  [988] 1 4 1 1 1 1 4 2 1 1 4 4 2
> mean(simul)
[1] 1.949
```

Joc just

Vols que el joc sigui "just" és a dir que un jugador no guanyi ni perdi diners encara que jugui molt de temps.

Quin és el preu just per participar en el joc?

És $E(X)$!!!

Exemple

Tenim un valor en borsa que pot cotitzar a 100 amb probabilitat 0.3 i a 400 amb probabilitat 0.7. Quant pagaries per aquest valor? No més del seu preu just, és a dir

$$E(X) = 100 \cdot 0.3 + 400 \cdot 0.7 = 310$$

Propietats de l'esperança

Propietats

1. Si $X = a$, aleshores $E(X) = a$
2. $E(a \cdot X + b \cdot Y + c) = a \cdot E(X) + b \cdot E(Y) + c$

L'esperança de $g(X)$ es defineix com

$$\boxed{E[g(X)] = \sum_x g(x) \cdot P_X(x)}$$

Exemple

Tirem un dau de sis cares X, i guanyem $Y = 3X^2 - 46$. Quina és $E(Y)$?

$$E(Y) = (3 \cdot 1^2 - 46) \cdot \frac{1}{6} + (3 \cdot 2^2 - 46) \cdot \frac{1}{6} + \ldots (3 \cdot 6^2 - 46) \cdot \frac{1}{6} = -0.5$$

Mesura de Risc. La variància.

La variància de X [σ_X^2, $Var(X)$] es defineix (si $\exists$) com,

$$\boxed{\begin{aligned} Var(X) &= E\left[(X - E(X))^2\right] \\ &= E(X^2) - (E(X))^2 \end{aligned}}$$

Propietats

1. $Var(X) \geq 0$
2. Si $X = a$, aleshores $Var(X) = 0$
3. $Var(aX + b) = a^2\, Var(X)$
4. $\sigma_{aX} = \sqrt{Var(aX)} = |a|\, \sigma_X$

Exemple

1. Triem a l'atzar de la caixa $[2,2,2,2] \to X$. Tenim que $E(X) = 2$
2. Triem a l'atzar de la caixa $[0,0,4,4] \to Y$. Tenim que $E(Y) = 2$
3. Triem a l'atzar de la caixa $[-100,-100,104,104] \to Z$. Aquí també $E(Z) = 2$

Quina diferència hi ha entre aquestes variables?

1. $Var(X) = E[(X-2)^2] = 0^2 \cdot 1 = 0$
2. $Var(Y) = E[(Y-2)^2] = (0-2)^2 \cdot 0.5 + (4-2)^2 \cdot 0.5 = 4 \cdot 1 = 4$
3. $Var(Z) = E[(Z-2)^2] = (-100-2)^2 \cdot 0.5 + (104-2)^2 \cdot 0.5 = 10816$

X, Y i Z tenen la mateixa mitjana però difereixen molt en variàncies.

Exemple

Exemple

Sabem que els ingressos anuals esperats de les famílies de la ciutat C són de 42000$ amb una desviació estàndard de 8000$.

Volem expressar la informació en euros enlloc de dòlars: el ràtio de canvi és de 1.24$ per euro. Quina serà la nova mitjana i la nova variància?

Solució: *Realitzem una transformació multiplicativa que consisteix en multiplicar les observacions en $ (X) per* $1/1.24$ *per a obtenir el resultat en €(Y).*

$$
\begin{aligned}
E(Y) &= \frac{42000}{1.24} = 33870.97€ \\
Var(Y) &= \left(\frac{8000}{1.24}\right)^2 = 41623309.05 \ €^2
\end{aligned}
$$

Centrar i estandarditzar

Centrar: Variable X centrada és $\boxed{Y = X - \mu_X}$.
Tenim que $m_Y = 0$

Estandarditzar: La variable X estandarditzada és $\boxed{Y = \dfrac{X - \mu_X}{\sigma_X}}$

Tenim que $E(Y) = 0$ i $Var(Y) = 1$.

Funció R, `scale` : Si tenim una llista de números, observacions de X,

```
X= c(1,6,2,1,6,9)
scale(X) és la llista de números estandarditzada
```

Exemple

Si $E(X) = 3$ i $Var(X) = 4$, aleshores la variable X estandarditzada és $Y = \dfrac{X-3}{2}$. Demostra que el valor esperat de Y és 0 i la variància de Y és 1.

Exemple

En un examen tipus test tenim amb N preguntes i k possibles respostes per pregunta. Per cada pregunta es marca una resposta o es deixa en blanc. Quant hem de restar a les respostes fallades per tal que si es contesten les preguntes a l'atzar es tingui una puntuació esperada de 0.

Sigui X_i, $i = 1, \dots N$, la v.a. que val 1 si la resposta a la pregunta i és correcta i $-a$ si és falsa.

x_i	1	$-a$
$P(X = x_i)$	$\frac{1}{k}$	$\frac{k-1}{k}$

Aleshores, $E(X_1 + \cdots + X_N) = E(X_1) + \cdots + E(X_N) = N \, \dfrac{1 - a(k-1)}{k}$.
Volem,

$$N \, \frac{1 - a(k-1)}{k} = 0 \Leftrightarrow \boxed{a = \frac{1}{k-1}}$$

Distribució de Bernoulli

Context: Experiment aleatori amb **dos resultats possibles**: *èxit* i *fracàs*.

Valors: Sigui X la v.a.,

$$X = \begin{cases} 1 & \text{si } \textit{èxit} \\ 0 & \text{si } \textit{fracàs} \end{cases}$$

Definició: Si $P(\textit{èxit}) = p$, diem que X té una distribució. Bernoulli de paràmetre p, $X \sim Bern(p)$.

Fun. Massa:

$$P(X = x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \end{cases}$$

o equivalentment $P(X = x) = p^x (1-p)^{1-x}, \; x = 0, 1$

E(X) i V(X):

$$E(X) = p, \qquad Var(X) = p(1-p)$$

Exemple

Tirem un dau no trucat i observem si surt o no 6. Definim

$$X = \begin{cases} 1 & si\ s'observa\ 6 \\ 0 & altrament \end{cases}$$

Aleshores,

- $X \sim Bern\left(\frac{1}{6}\right)$.
- $E(X) = \frac{1}{6} = 0.1667$ *i* $Var(X) = \frac{5}{36} = 0.1389$.

Simulem 10000 vegades l'experiment amb `R`:

```
simBern=rbinom(10000, 1, 1/6)
table(simBern)/10000
dbinom(0:1, 1, 1/6)
mean(simBern)
var(simBern)
```

Distribució Uniforme discreta

Context: Experiment aleatori a on **tots els possibles valors tenen la mateixa probabilitat** de ser observats.

Valors: Sigui X la v.a. que pren els valors $x_1, x_2, \ldots, x_k$.

Definició: Escriurem $X \sim Unif_d(\{x_1, x_2, \ldots, x_k\})$.

Fun. Massa:

$$P(X = x) = \begin{cases} \frac{1}{k} & \text{si } x \in \{x_1, x_2, \ldots, x_k\} \\ 0 & \text{altrament} \end{cases}$$

E(X) i V(X): Si $X \sim Unif_d(\{1, 2, \ldots, n\})$, aleshores [1]

$$E(X) = \frac{n+1}{2}, \qquad Var(X) = \frac{(n+1)\cdot(n-1)}{12}.$$

[1] fem servir que: $1 + 2 + \ldots n = \frac{n(n+1)}{2}$ i $1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$

Exemple

Un jugador té 2 monedes. Cada vegada pot mostrar 0, 1 o 2 monedes amb igual probabilitat. Calculeu l'esperança i la variància.

X='*nombre de monedes a la mà*' $\sim Unif_d(\{0,1,2\})$

$E(X) = P(X=0)\cdot 0 + P(X=1)\cdot 1 + P(X=2)\cdot 2 = 1.$

$E(X^2) = P(X=0)\cdot 0^2 + P(X=1)\cdot 1^2 + P(X=2)\cdot 2^2 = \frac{5}{3}.$

Per tant,

$$Var(X) = E(X^2) - (E(X))^2 = \frac{2}{3}.$$

Exemple

X='*resultat de la tirada d'un dau de 10 cares*', $X \sim Unif_d(\{1,\dots,10\})$. *Aleshores, podem calcular aplicant les fórmules*

$$E(X) = \frac{11}{2} = 5.5 \qquad Var(X) = \frac{11\cdot 9}{12} = 8.25$$

Distribució Binomial

Context: Realitzem n d'experiments de Bernoulli independents amb la **mateixa probabilitat d'èxit** p. Comptem el **nombre d'èxits observats** en n experiments aleatoris.

Valors: Sigui X la v.a. que pren els valors $0, 1, 2, \ldots, n$.

Definició: Diem que X segueix una distribució binomial de paràmetres n i p, es denota per $X \sim B(n, p)$.

Fun. Massa:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n$$

E(X) i V(X):

$$E(X) = np, \qquad Var(X) = np(1-p)$$

Funcions de massa i distribució

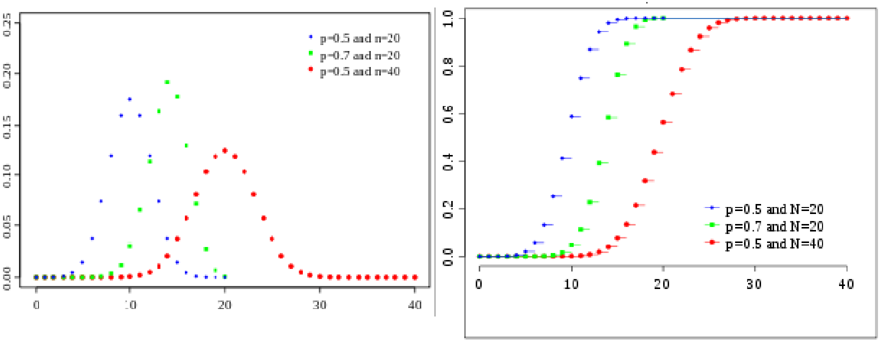


Figura: Funcions de massa de probabilitat i de distribució de binomials

Exemple

El 5% dels individus de la població pesen més de 90kg.

Quina és la probabilitat exacta de trobar com a mínim 2 persones per sobre de 90kg en un grup de 10 persones?

Solució: X="*nombre persones >90 kg en el grup de 10*"$\sim B(10, 0.05)$

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - (0.5987 + 0.3151) = 1 - 0.9139 = 0.0861$$

Amb R,

$$P(X \geq 2) = 1 - P(X \leq 1) = \texttt{1-pbinom(1, 10, 0.05)}$$
$$= \texttt{1-dbinom(0:1,10,0.05)}$$

Exemple

Quan una xinxeta cau a terra té probabilitat 0.7 de caure amb la punta enlaire. Cauen a terra 50 xinxetes d'una capsa ben escampades i anomenem X al nombre de xinxetes que han quedat amb la punta enlaire.

1. *Quantes xinxetes s'espera que quedin amb la punta enlaire?*

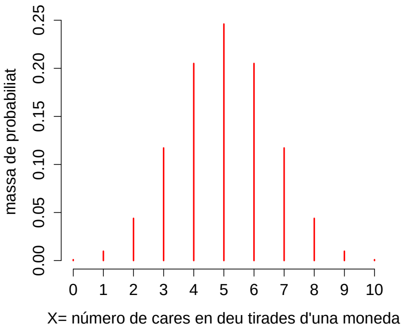
   $$X \sim B(50, 0.7). \quad E(X) = 50 \cdot 0.7 = 35.$$

   S'esperen sobre 35 xinxetes amb la punta enlaire.

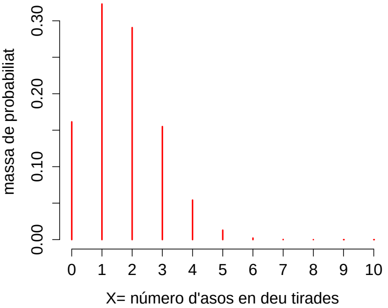
2. *Com calcularies la probabilitat exacta que quedin entre 30 i 40 xinxetes (tots dos inclosos) amb la punta enlaire?*

   $$\begin{aligned} P(30 \leq X \leq 40) &= \sum_{i=30}^{40} P(X = i) = \sum_{i=30}^{40} \binom{50}{i} (0.7)^i (0.3)^{50-i} \\ &= \texttt{pbinom(40, 50, 0.7)-pbinom(29, 50, 0.7)} \end{aligned}$$

Funcions de massa: $B(10,\frac{1}{2})$ vs $B(10,\ \frac{1}{6})$



Tirada 10 monedes



Tirada 10 daus

Distribució Geomètrica

Context: Realitzem una sèrie d'experiments de Bernoulli independents amb la **mateixa probabilitat d'èxit** p.
Comptem el **nombre d'experiments necessaris per tal d'observar el primer èxit**.

Valors: Sigui X la v.a. que pren els valors $1, 2, \ldots,$.

Definició: X segueix una distribució geomètrica de paràmetre p, es denota per $X \sim Geom(p)$.

Fun. Massa:

$$P(X = x) = (1-p)^{x-1} p, \quad x = 1, 2, \ldots$$

E(X) i V(X):

$$E(X) = \frac{1}{p}, \qquad Var(X) = \frac{1-p}{p^2} = \frac{q}{p^2}$$

Compte!!!

Hi ha dues versions de la distribució $Geom(p)$,

1. X="Nombre d'intents necessaris per tal d'observar el primer èxit".
   L'èxit es compta!!!

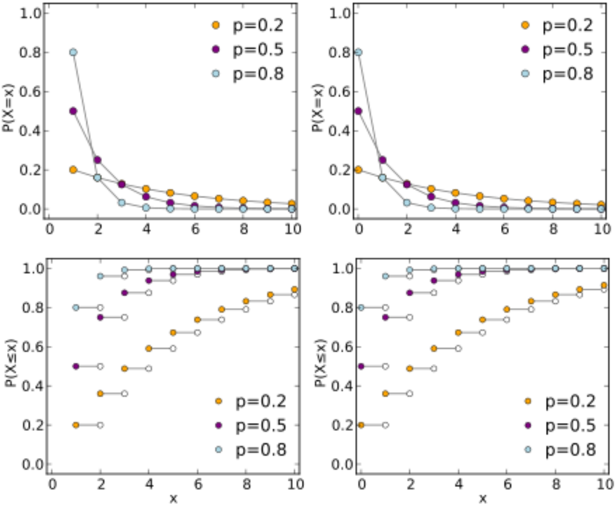
$$E(X) = \frac{1}{p} \quad Var(X) = \frac{q}{p^2}$$

2. Y="Nombre de fallides necessàries per tal d'observar el primer èxit".
   L'èxit no es compta!!!

$$E(Y) = \frac{1}{p} - 1 \quad var(Y) = var(X) = \frac{q}{p^2}$$

doncs $Y = X - 1.$, $Y = 0, 1, 2 \dots$

Funcions de massa i de distribució d'una v.a geomètrica



Exemple

La probabilitat que una màquina produeixi un article defectuós és de 0.03.

Calculeu la probabilitat que en un control de qualitat, s'hagin d'inspeccionar 20 articles fins a trobar el primer defectuós.

Solució: *L'experiment de Bernoulli és el fet de provar la qualitat d'un article on èxit representa el fet de ser defectuós.*

Sigui $X =$ *" v.a. que compta el nombre d'articles inspeccionats fins a trobar el primer defectuós"* $\sim Geom(0.03)$.

Aleshores,

$$P(X = 20) = (1 - 0.03)^{19} \cdot 0.03 = \texttt{dgeom(19,0.03)} \approx 0.0168$$

Exemple

El 40% dels individus d'una població són del grup sanguini A, el 45% són del grup O, el 5% són AB i el 10% són del grup B.

Arriben persones al banc de sang de manera independent i a l'atzar.

Aleshores, quina és la probabilitat que siguin necessàries com a mínim 4 donacions per a obtenir la primera del grup sanguini **AB***?*

Solució: $X =$ *"v.a. que mesura el nombre d'extraccions necessàries per observar la primera del grup AB".* $X \sim Geom(0.05)$

Aleshores,

$$\begin{aligned} P(X \geq 4) &= 1 - P(X < 4)? = 1 - P(X \leq 3) = \texttt{pgeom(2,0.05)} \\ &= 1 - \left((0.95)^0 \cdot 0.05 + (0.95)^1 \cdot 0.05 + (0.95)^2 \cdot 0.05\right) \\ &= 1 - 0.1426 = 0.8574 \end{aligned}$$

Distribució de Poisson

Context: Aquesta distribució sorgeix en el recompte d'esdeveniments rars, poc freqüents, sota certes condicions...

Valors: Sigui X la v.a. que pren els valors $0, 1, 2, \ldots,$

Definició: X segueix la distribució Poisson de paràmetre λ, que denotarem $X \sim Poiss(\lambda)$

Fun. Massa:

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

E(X) i V(X):

$$E(X) = \lambda, \qquad Var(X) = \lambda$$

Propietat. Si $X_1 \sim Poiss(\lambda_1), \ldots, X_k \sim Poiss(\lambda_k)$ i són independents, aleshores

$$\boxed{X_1 + X_2 + \cdots + X_k \sim Poiss(\lambda_1 + \lambda_2 + \cdots + \lambda_k)}$$

Funcions de massa de probabilitat i de distribució d'una v.a Poisson

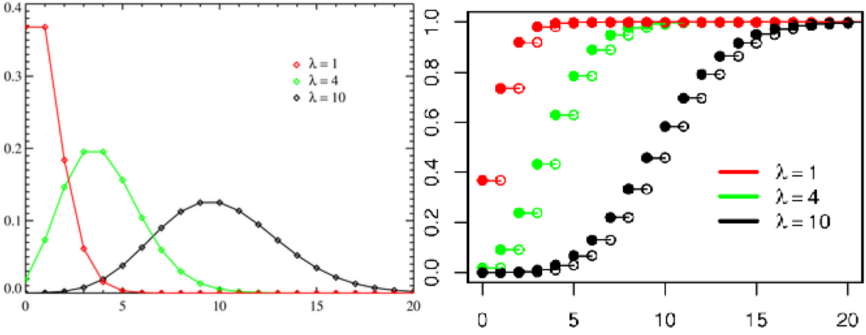


Figura: Funcions de massa de probabilitat i de distribució de Poisson

Exemple

El nombre d'errors X per cada hora de funcionament d'una màquina és una v.a amb funció de massa:

$$P(X = x) = e^{-0.1} \frac{0.1^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

1. *Trobeu la probabilitat que es produeixi algun error.*

$$P(X > 0) = 1 - P(X = 0) = 1 - e^{-0.1} = \texttt{1-dpois(0,0.1)} = 1 - 0.0952$$

2. *Trobeu la probabilitat que es produeixin 5 errors en 5 hores. Definim $Y = X_1 + \cdots + X_5$, per tant volem calcular*

$$P(Y = 5) = e^{-0.5} \frac{0.5^5}{5!} = \texttt{dpois(5,0.5)}$$

Exemple

El promig de trucades de mòbil en una classe és 0.1.

1. *Calculeu la probabilitat que avui tinguem una trucada de mòbil*

$$P(X > 0) = 1 - P(X = 0) = 1 - exp(-0.1) = 0.0952$$

2. *Quina és la probabilitat que un mòbil truqui en una setmana (dues classes)? Definim* $Y = X1 + X2$ *per tant,* $Y \sim Pois(0.1 + 0.1)$

$$P(Y > 0) = 1 - P(Y = 0) = 1 - exp(-0.2) = 1 - exp(-0.2) = 0.1813.$$

Exemple

El nombre d'accidents per setmana en un punt segueix una distribució Poisson amb $\lambda = 2$. *Simula la distribució d'accidents durant 100 setmanes.*

Solució:

```
simPois<-rpois(100, 2)
table(simPois)/100
```

Problema

Exemple

En un país hi ha dues regions A i B on es produeix raïm. Se sap que per què el raïm sigui de bona qualitat el nombre de dies de pluja durant la temporada de camp ha de ser de 4 a 6.

Segons les observacions meteorològiques, el nombre de dies de pluja segueix una distribució de Poisson de mitjana 4 dies a la zona A i 5 dies a la zona B.

1. *Quina és la probabilitat que en un any determinat el raïm produït a cadascuna de les dues zones sigui de bona qualitat?*
2. *A una fruiteria arriben 500 caixes, 200 de la zona A i 300 de la zona B. Si agafem una caixa a l'atzar, quina és la probabilitat que el raïm d'aquesta caixa sigui de bona qualitat?*

Solució

1. Definim les variables
X_A= " nombre de dies de pluja a la zona A" $\sim Poiss(4)$.
X_B= " nombre de dies de pluja a la zona B" $\sim Poiss(5)$.
Aleshores,

$$\begin{aligned} P(4 \leq X_A \leq 6) &= e^{-4}\frac{4^4}{4!} + e^{-4}\frac{4^5}{5!} + e^{-4}\frac{4^6}{6!} = 0.4558 \\ P(4 \leq X_B \leq 6) &= e^{-5}\frac{5^4}{4!} + e^{-5}\frac{5^5}{5!} + e^{-5}\frac{5^6}{6!} = 0.4971 \end{aligned}$$

2. Definim les variables Q="El raïm és de bona qualitat", A="El raïm és de la zona A", B="El raïm és de la zona B".
Sabem, de l'apartat anterior que $P(Q|A) = 0.4558$ i $P(Q|B) = 0.4971$
Segons la informació de la fruiteria $P(A) = 2/5$, $P(B) = 3/5$
Aleshores, pel teorema de les probabilitats totals,

$$P(Q) = P(Q|A) \cdot P(A) + P(Q|B) \cdot P(B) = 0.4806$$