UNIVERSITAT DE BARCELONA

Treball en R. Variables aleatòries i simulació

Noah Márquez
Alejandro Guzman

9 maig 2022

ÍNDEX

1 POBLEMES PRÀCTICA 3

1 Problema 1

1. **La variable aleatòria X té les probabilitats següents:**

Valors x_i	2	3	6	7	8	10
Probabilitats $P[X = x_i]$	0.2	0.1	0.2	0.1	0.2	0.2

   a) **Representeu la funció de massa de probabilitat i la corresponent funció de distribució de probabilitat.**

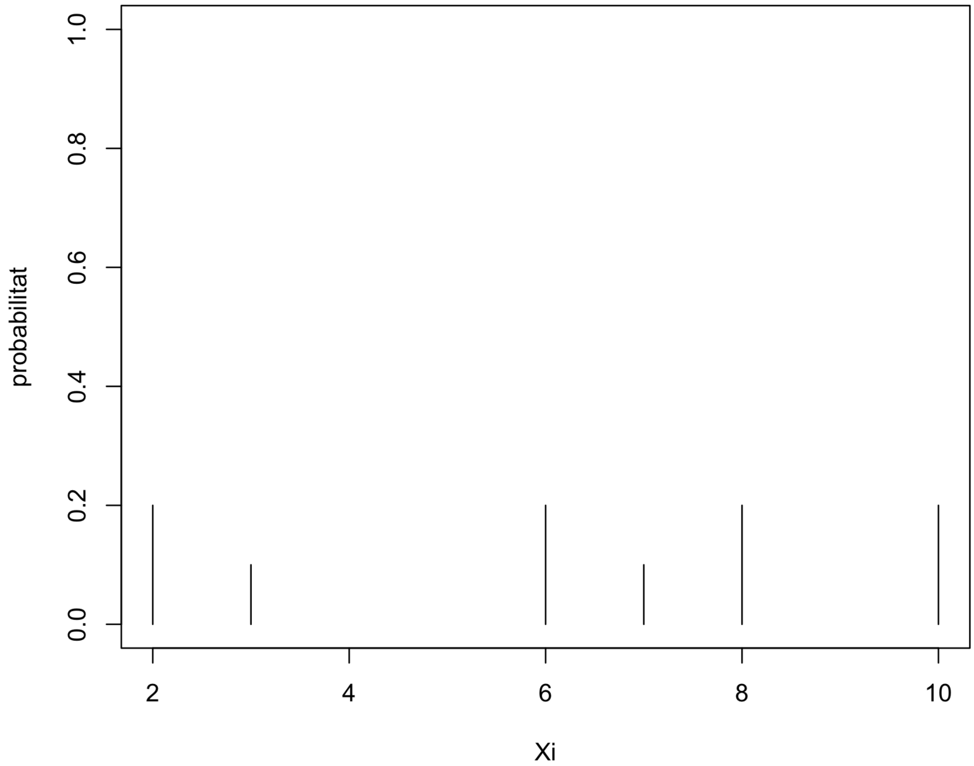

Figura 1.1: Funció de massa de probabilitat

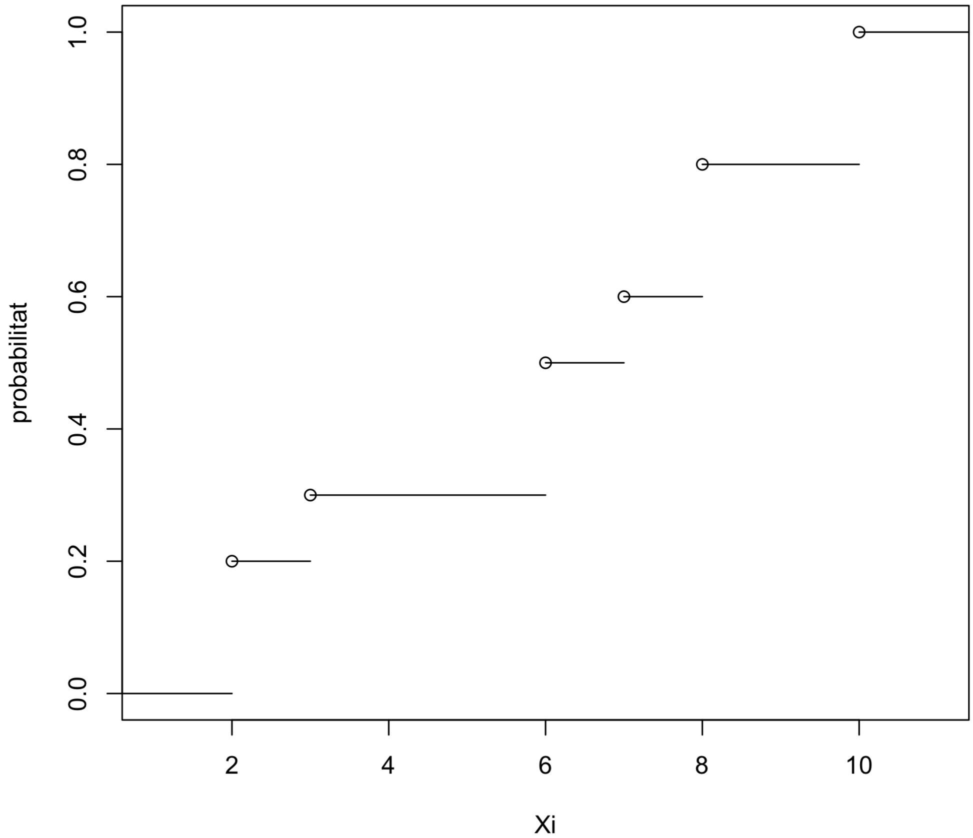

Figura 1.2: Funció de distribució de probabilitat

Aprofitem per comentar que tots els scripts utilitzats al llarg de la pràctica es troben a la secció Scripts en R.

b) **Calculeu l'esperança, la variància i la desviació típica de *X*.**

Recordem que $E[X] = \sum_{i=1}^{n} x_i P(X = x_i)$. Per tant, el su càlcul el podem fer fàcilment mitjançant la comanda: `esp <- sum(x*prob)`.
Llavors tenim:

$$E[X] = \mathbf{6.2}$$

Pel cálcul de la variància seguim la fòrmula: $s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$. Fent el càlcul tenim:

$$s_x^2 = \mathbf{7.\overline{6}}$$

Finalment, pel càlcul de la desviació típica de X recordem que: $s_x = \sqrt{s_x^2}$. Fent el càlcul:

$$s_x = \mathbf{2.768875}$$

2 Problema 8

8. **El nombre mitjà d'automòbils que arriben a una estació de subministrament de gasolina és de 210 per hora. Si aquesta estació pot atendre un màxim de 10 automòbils per minut, determineu la probabilitat que en un minut donat arribin a l'estació més automòbils dels que poden atendre. Suposeu que el nombre d'automòbils que arriben a**

l'estació durant 1 minut segueix una distribució de Poisson.

Seguint l'enunciat, el nombre mitjà d'automòbils que arriben a una estació de subministrament de gasolina és de 210/hora, d'aquí tenim que $\frac{210}{hora} \cdot \frac{1hora}{60min} = \frac{3{,}5}{min}$.
Per tant, $\lambda = 3,5$.

Volem determinar la probabilitat de que en un minut donat arribin a l'estació més automòbils dels que poden atendre: $P(X \geq 10)$.

En **R** ho podem calcular de vàries maneres, nosaltres ho hem fem de la següent forma:

ppois(10, 3.5, lower.tail=FALSE)
(*lower.tail* ens estalvia haver de fer *1 - ppois(10, 3.5)*)

Per tant, després de fer el càlcul, tenim:

$P(X \geq 10) =$ **0.0010193944...**

O bé

$P(X \geq 10) \simeq$ **0.10193944 %**

2 PROBLEMES PRÀCTICA 4

1 Problema 5

5. **Els coeficients d'intel·ligència d'un grup d'adults d'entre 20 i 34 anys tenen una distribució aproximadament normal de mitjana** $\mu = 110$ **i desviació típica** $\sigma = 25$.

   a) **Quin percentatge de persones entre 20 i 34 anys té coeficient més gran que 100?**

   Volem calcular $P(X \geq 100) = 1 - P(X < 100)$. Amb ajuda de **R** fem el càlcul i obtenim:

   $$P(X \geq 100) = \mathbf{0.6554217}$$

   Com demana el percentatge, multipliquem per 100:

   $$P(X \geq 100) = 0.6554217 \cdot 100 = \mathbf{65.54217\ \%}$$

   b) **Quin percentatge de persones entre 20 i 34 anys té coeficients més petit que 150?**

   Volem calcular $P(X \leq 150)$. Amb ajuda de **R** (tot detallat en l'apartat de Scripts en R) fem el càlcul i obtenim:

   $$P(X \leq 150) = \mathbf{0.9452007}$$

   Com demana el percentatge, multipliquem per 100:

   $$P(X \leq 150) = 0.9452007 \cdot 100 = \mathbf{94.52007\ \%}$$

   c) **Quin coeficient mínim tenen els adults entre 20 i 34 anys situats en el 25% que han obtingut millors resultats?**

   Per trobar el valor de la variable (coeficient) en que la funció de distribució assoleix una probabilitat concreta, utilitzem la instrucció de **R**:

   qnorm(0.25,110,25)

   Fent el cálcul obtenim que el coeficient mínim que tenen els adults entre 20 i 34 anys situats en el 25% que han obtingut millors resultats és de **93.13776**.

2 Problema 6

6. **El temps de funcionament d'una bombeta segueix una distribució exponencial de mitjana 2000 hores. Quina probabilitat hi ha que després de 3000 hores segueixi funcionant? Quina és la probabilitat que s'espatlli abans de 2000 hores?**

   De l'enunciat sabem que tenim una distribució exponencial de mitjana 2000 hores, això vol dir que $\lambda = \frac{1}{2000}$.

   La probabilitat de que després de 3000 hores segueixi funcionant és: $P(X \geq 3000) = 1 - P(X < 3000)$.
   Amb la següent instrucció de **R** ho calculem:

   1-pexp(3000, 1/2000)

   I obtenim:

   $$P(X \geq 3000) = \mathbf{0.2231302}$$

La probabilitat de que s'espatlli abans de les 2000 hores és: $P(X < 2000)$.
Amb la següent instrucció de **R** ho calculem:

pexp(2000, 1/2000)

Ja que la distribució exponencial és una distribució continua i llavors $P(X = x) = 0$.

Obtenim:

$$P(X < 2000) = \mathbf{0.6321206}$$

3 PROBLEMES PRÀCTICA 5

1 Problema 2.1

2. **Calculeu $B = 50$ valors d'una variable aleatòria X que té la següent funció de massa de probabilitat:**

$$P[X = 3] = 0.1, P[X = 5] = 0.3, P[X = 7] = 0.2, P[X = 9] = 0.4.$$

Calculeu les freqüències relatives de $[X = 3]$, $[X = 5]$, $[X = 7]$, $[X = 9]$ a la llista de nombres aleatoris, i compareu-les amb les corresponents probabilitats teòriques.

*2.1 **B = 50***

Valors	Freqüències relatives
3	0.08
5	0.34
7	0.20
9	0.38

*2.2 **B = 10000***

Valors	Freqüències relatives
3	0.0976
5	0.2996
7	0.2053
9	0.3975

Podem apreciar com les freqüències relatives s'aproximen als valors de la funció de massa de probabilitat a mesura que augmentem el nombre de valors aleatoris de la mostra.

Calculeu (teòricament) l'esperança i la variància de X.

Per tal de fer el càlcul teòric de l'esperança i la variància recorrem a les seves definicions. En quant a l'esperança tenim:

$$E(X) = \sum_x x \cdot P(X = x)$$

Si substituïm amb els valors de l'enunciat,

$$E(X) = \sum_{3,5,7,9} x \cdot P(X = x) = \frac{34}{8} = \mathbf{6.8}$$

Per la variància, tenim dues formes de calcular-la:

$$Var(X) = E[(X - E(X))^2]$$
$$Var(X) = E(X^2) - (E(X))^2$$

Si substituïm a la primera definició:

$$Var(X) = E[(X - E(X))^2] = \sum_{3,5,7,9} [X - E(X)]^2 \cdot P(X = x) = \frac{109}{25} = \mathbf{4.36}$$

Després, calculeu la mitjana i la variància empíriques de la llista de nombres aleatoris.

2.3 ***B = 50***

Mitjançant la instrucció d'**R**, obtenim que la mitjana és **6.76**.

Fem el càlcul de la variància també en **R** i obtenim: **4.1024**.

2.4 ***B = 10000***

Mitjançant la instrucció d'**R**, obtenim que la mitjana és **6.8054**.

Fem el càlcul de la variància també en **R** i obtenim: **4.312562**.

Podem apreciar com els valors s'apropen als càlculs teòrics a mesura que fem més gran el valor de B.

Repetiu la simulació amb $B = 10000$.
Ho hem fet a mesura que es contestàvem les preguntes.

3 Problema 2.2

2. **Simuleu B = 50 valors d'una variable aleatòria X, absolutament contínua, amb funció de densitat de probabilitat:**

$$f(x) = \frac{5}{32}x^4,\ 0 < x < 2.$$

Dibuixeu l'histograma. Superposeu-hi el dibuix de la funció de densitat de probabilitat.

Primer de tot hem de calcular la funció de distribució:

$$F(x) = \begin{cases} 0, & x < 0 \\ \int_0^x \frac{5}{32}y^4\,dy = \frac{x^5}{32}, & 0 \leq x < 2 \\ 1, & 2 \leq x \end{cases}$$

de manera que la inversa és $F^{-1}(u) = 2\sqrt[5]{u}$.

Ara amb l'**R** ja podem obtenir l'histograma amb la funció de densitat de probabilitat superposada:

2.1 B = 50

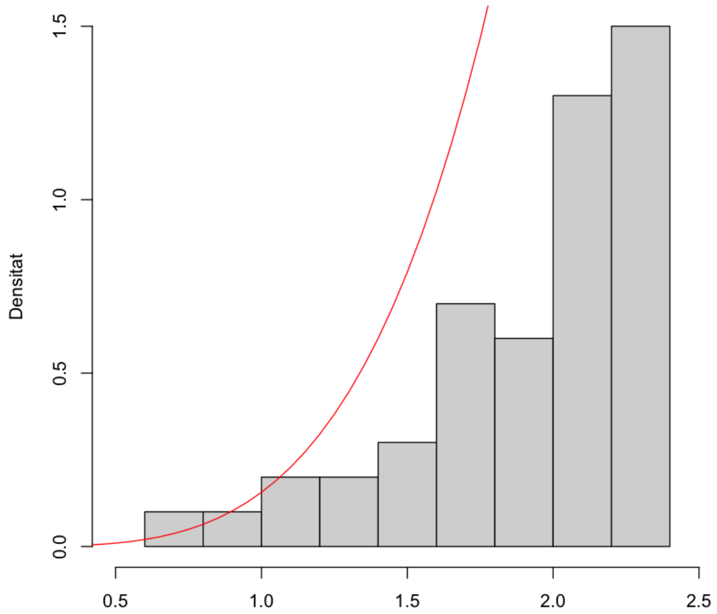

Figura 3.1: Histograma amb funció de densitat de probabilitat superposada

2.2 B = 1000

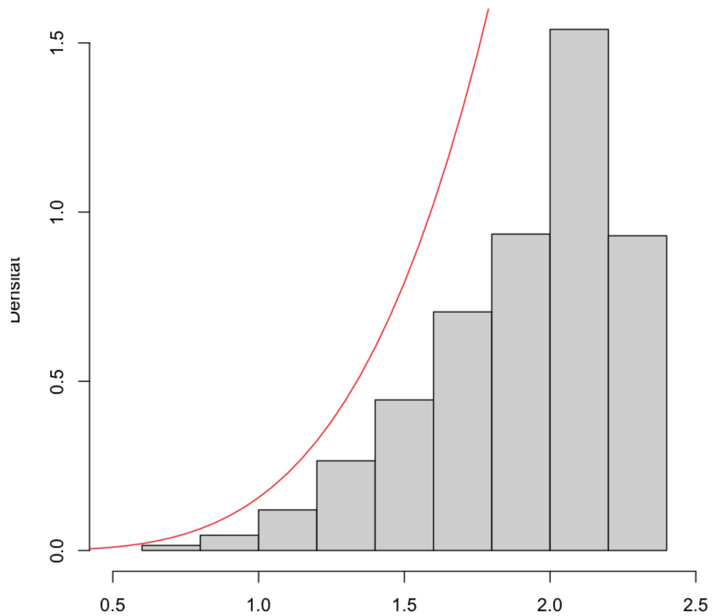

Figura 3.2: Histograma amb funció de densitat de probabilitat superposada

Calculeu (teòricament) l'esperança i la variància de X.

Per tal de fer el càlcul teòric de l'esperança i la variància recorrem a les seves definicions per a les distribucions uniformes contínues. En quant a l'esperança tenim:

$$E(X) = \frac{a+b}{2}$$

Si substituïm amb els valors de l'enunciat,

$$E(X) = \frac{0+2}{2} = \mathbf{1}$$

Per la variància, tenim la següent fórmula:

$$Var(X) = \frac{(b-a)^2}{12}$$

Si substituïm:

$$Var(X) = \frac{(2-0)^2}{12} = \frac{1}{3} = \mathbf{0.\overline{3}}$$

Després, calculeu la mitjana i la variància empíriques de la llista de nombres aleatoris.

*2.1 **B = 50***

Mitjançant la instrucció d'**R**, obtenim que la mitjana és **1.10187**.

Fem el càlcul de la variància també en **R** i obtenim: **0.3167086**.

*2.2 **B = 1000***

Mitjançant la instrucció d'**R**, obtenim que la mitjana és **1.037323**.

Fem el càlcul de la variància també en **R** i obtenim: **0.3330103**.

Podem apreciar com els valors s'apropen als càlculs teòrics a mesura que fem més gran el valor de B.

Repetiu la simulació amb $B = 1000$**.**
Ho hem fet a mesura que es contestàvem les preguntes.

4 Estudi i simulació d'una variable discreta

a) **Donar la funció de massa de probabilitat i la funció de distribució de la variable Y corresponent a la resta dels resultats del llançament de dos daus (estudi teòric). Fer els gràfics de les dues funcions. Calcular l'esperança i la variància.**

Prenem la variable Y com el resultat de la resta del llançament de dos daus. Per tal de donar les funcions que es demanen, calcularem la probabilitat de cada possible resultat d'aquesta resta.

Prenem Y = *dau1* - *dau2*. La taula següent mostra el càlcul dels possibles resultats:

		DAU 1					
		1	**2**	**3**	**4**	**5**	**6**
DAU 2	**1**	0	1	2	3	4	5
	2	-1	0	1	2	3	4
	3	-2	-1	0	1	2	3
	4	-3	-2	-1	0	1	2
	5	-4	-3	-2	-1	0	1
	6	-5	-4	-3	-2	-1	0

Sabent que el número de resultats possibles és 36 (*6 · 6 cares de cada dau*), tenim la funció de massa de probabilitat següent:

$$P(Y=x)=\begin{cases} -5, & \frac{1}{36} \\ -4, & \frac{2}{36} \\ -3, & \frac{3}{36} \\ -2, & \frac{4}{36} \\ -1, & \frac{5}{36} \\ 0, & \frac{6}{36} \\ 1, & \frac{5}{36} \\ 2, & \frac{4}{36} \\ 3, & \frac{3}{36} \\ 4, & \frac{2}{36} \\ 5, & \frac{1}{36} \end{cases}$$

Un cop indicades les probabilitats de cada cas, sabem que la funció de distribució associa a cada valor de la variable aleatoria la probabilitat acumulada fins aquest valor.

Per tant, la funció de distribució quedaria de la següent manera:

$$F(Y) = \begin{cases} 0, & Y < -5 \\ \frac{1}{36}, & -5 \leq Y < -4 \\ \frac{3}{36}, & -5 \leq Y < -3 \\ \frac{6}{36}, & -5 \leq Y < -2 \\ \frac{10}{36}, & -5 \leq Y < -1 \\ \frac{15}{36}, & -5 \leq Y < -0 \\ \frac{21}{36}, & -5 \leq Y < 1 \\ \frac{26}{36}, & -5 \leq Y < 2 \\ \frac{30}{36}, & -5 \leq Y < 3 \\ \frac{33}{36}, & -5 \leq Y < 4 \\ \frac{35}{36}, & -5 \leq Y < 5 \\ 1, & Y \leq 5 \end{cases}$$

A continuació es mostren els gràfics de les dues funcions:

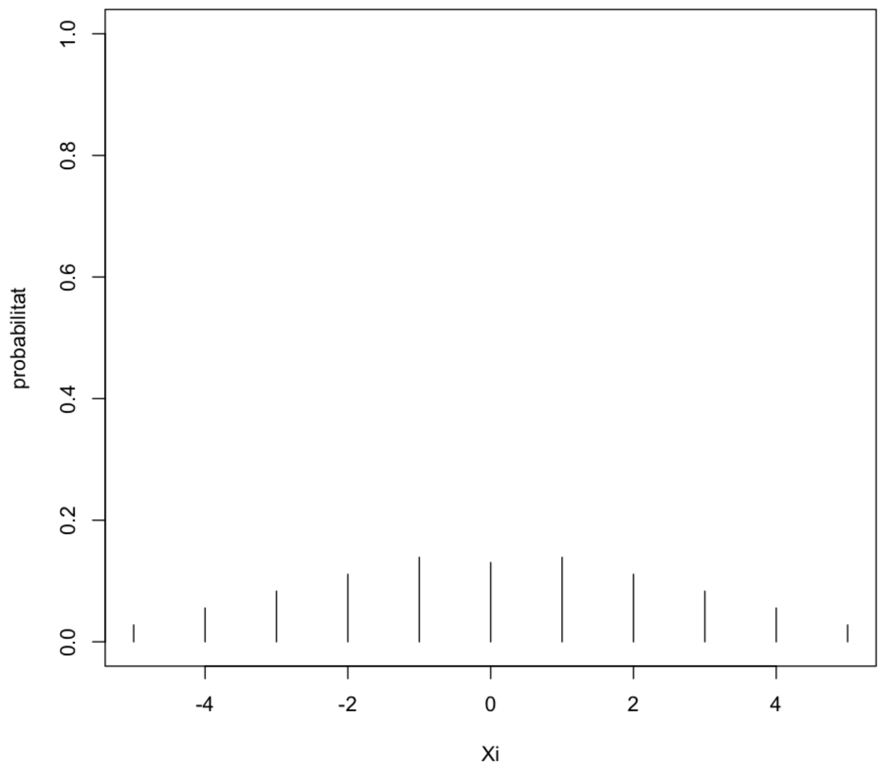

Figura 4.1: Funció de massa de probabilitat

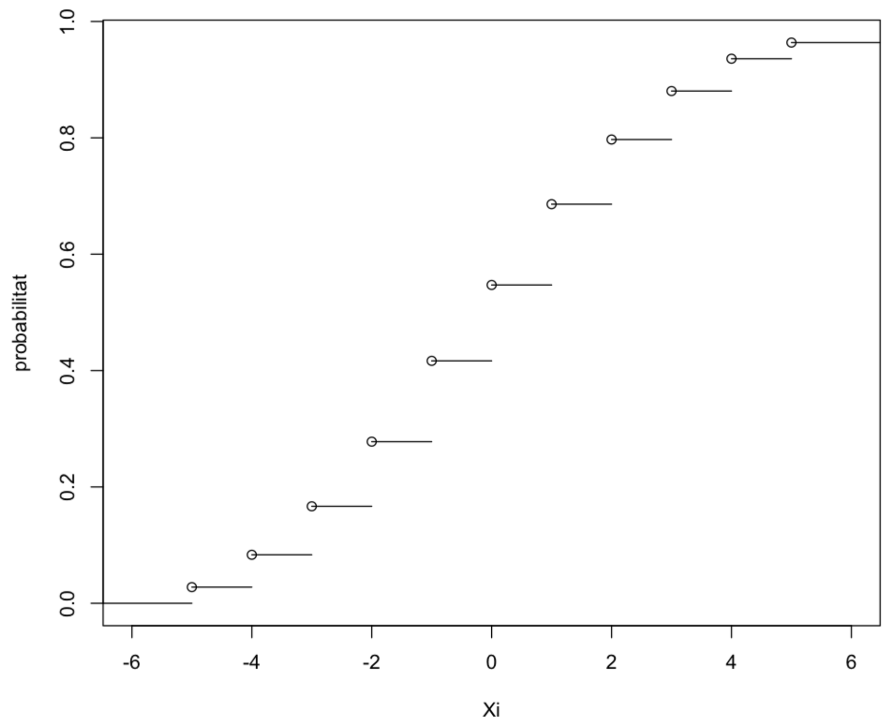

Figura 4.2: Funció de distribució de probabilitat

Per tal de fer el càlcul teòric de l'esperança i la variància recorrem a les seves definicions. En quant a l'esperança tenim:

$$E(X) = \sum_x x \cdot P(X = x)$$

Si substituïm amb els valors de l'enunciat,

$$E(X) = \sum_{-5:5} x \cdot P(X = x) = \mathbf{0}$$

Per la variància, tenim dues formes de calcular-la:

$$Var(X) = E[(X - E(X))^2]$$
$$Var(X) = E(X^2) - (E(X))^2$$

Si substituïm a la segona definició:

$$Var(X) = E(X^2) - (E(X))^2 = \sum_{-5:5} x^2 \cdot P(X = x) - (E(X))^2 = 5.83 - 0 = \mathbf{5.83}$$

b) **Simular una mostra de la variable discreta Y i calcular la mitjana i la variància i comparar-les amb les teòriques.**
Simuleu les mostres per $n = 50, 100$ i 1000.

Un cop simulades les mostres de la variable discreta Y amb **R** (recordem que està tot detallat a l'apartat Scripts en R), calculem la mitjana i la variància d'aquestes per cada n que especifica l'enunciat.

1.1 ***n = 50***

Mitjançant la instrucció d'**R**, obtenim que la mitjana és **0.36**.

Fem el càlcul de la variància també en **R** i obtenim: **4.969796**.

1.2 ***n = 100***

Mitjançant la instrucció d'**R**, obtenim que la mitjana és **0.12**.

Fem el càlcul de la variància també en **R** i obtenim: **4.914747**.

1.3 ***n = 1000***

Mitjançant la instrucció d'**R**, obtenim que la mitjana és **0.01**.

Fem el càlcul de la variància també en **R** i obtenim: **5.75966**.

Podem apreciar com els valors s'apropen als càlculs teòrics a mesura que fem més gran el valor de B.

c) **Simular dues variables aleatòries *T* i *V* corresponents al llançament de dos daus i definim $Z = T - V$. Fer la taula de freqüències relatives de *Z*. Comparar amb els resultats teòrics de (a) (Funció de massa de probabilitat).**
Simuleu les mostres per $n = 50, 100$ **i** 1000.

Un cop simulades les dues variables aleatòries T i V corresponents al llançament de dos daus, hem definit $Z = T - V$. Mitjançant R hem fet la taula de freqüències relatives de Z. A continuació les mostrem per cada valor de n que indica l'enunciat:

1.4 ***n = 50***

Valors	Freqüències relatives
-5	0.04
-4	0.02
-3	0.14
-2	0.04
-1	0.08
0	0.14
1	0.14
2	0.16
3	0.10
4	0.10
5	0.04

1.5 ***n = 100***

Valors	Freqüències relatives
-5	0.03
-4	0.05
-3	0.04
-2	0.10
-1	0.21
0	0.12
1	0.15
2	0.10
3	0.06
4	0.09
5	0.05

1.6 ***n = 1000***

Valors	Freqüències relatives
-5	0.037
-4	0.051
-3	0.083
-2	0.107
-1	0.153
0	0.157
1	0.155
2	0.118
3	0.085
4	0.038
5	0.016

Podem apreciar com els valors s'apropen als resultats teòrics de l'apartat **(a)** a mesura que fem més gran el valor de n.

5 Conclusions

En aquest treball hem pogut treballar en l'àmbit de les variable aleatòries i la simulació, fent tant exercicis de pràctiques realitzades a classe com un estudi d'una variable discreta. Ens ha ajudat a treballar d'una forma molt més visual i descriptiva i alhora tenir més en contacte amb el funcionament del llenguatge **R**.

Ens va ser bastant fàcil començar la pràctica degut a que els primers dos apartats ja els teníem fets de les hores de laboratori. És per això que hem pogut dedicar-hi més temps a repassar els conceptes i alhora fer els dos apartats restants en profunditat.

En general la realització d'aquest treball ens ha ajudat a tenir un bon contacte i entendre molt millor els conceptes de teoria i el potencial que té el llenguatge **R** en l'àmbit de l'estadística.

6 Scripts en R

1 Pràctica 3

1.1 Problema 1

```
# Creació dels vectors de valors xi
x <- c(2,3,6,7,8,10)
prob <- c(0.2,0.1,0.2,0.1,0.2,0.2)

# Representació de la funció de massa de probabilitat
plot(x,prob,type="h", xlim=c(2,10), ylim=c(0,1), xlab="Xi",
   ylab="probabilitat")

# Representació de la funció de distribució de probabilitat
acum <- cumsum(prob)
s <- stepfun(x, c(0,acum))
plot(s,verticals=FALSE)

# Esperanca
esp <- sum(x*prob)
esp

# Variancia
var <- 1/6 * sum((x-mean(x))**2)
var

# Desviació típica de X
sd <- sqrt(var)
sd
```

1.2 Problema 8

```
x <- ppois(10, 3.5, lower=FALSE)
x <- x*100

print(paste0("Probabilitat de que arribin mes cotxes dels que
   es poden atendre: ", x))
```

2 Pràctica 4

2.1 Problema 5

```
#a) Percentatge de persones entre 20 i 34 anys té coeficient m
   és gran que 100?
# P(X>=100) = 1 - P(X<100)
100*(1-pnorm(100,110,25))

#b) Percentatge de persones entre 20 i 34 anys té coeficient m
   és petit que 150?
# P(X<150)
```

```
100*(pnorm(150,110,25))

#c) Coeficient mínim del adults entre 20 i 34 anys situats en
   el 25% que han obtingut millors resultats?
qnorm(0.25,110,25)
```

2.2 Problema 6

```
# Probabilitat de que després de 3000 hores segueixi
   funcionant?
# P(X >= 3000) = 1 - P(X<3000)
1-pexp(3000,1/2000)

# Probabilitat de que s'espatlli abans de 2000 hores?
# P (X<2000)
pexp(2000,1/2000)
```

3 Pràctica 5

3.1 Problema 2.1

```
# Obtenim els nombres aleatoris:
u <- runif(50)

# Fer la transformació per a les probabilitats
y <- sample(c(3,5,7,9), 50, prob=c(0.1, 0.3, 0.2, 0.4),
   replace=TRUE)

# Calcul frequencies relatives
table(y)/length(y)

# Calcul esperanca
valors <- c(3,5,7,9)
probs <- c(0.1, 0.3, 0.2, 0.4)
esp_teorica <- sum(valors * probs)

# Calcul variancia (teorica)
var_teorica <- sum((valors - esp_teorica)^2 * probs)

# Mitjana i variancia (empíriques)
mean(y)

var(y)

# El mateix per B = 10000

# Obtenim els nombres aleatoris:
t <- runif(10000)

# Fer la transformació per a les probabilitats
```

```
x <- sample(c(3,5,7,9), 10000, prob=c(0.1, 0.3, 0.2, 0.4),
   replace=TRUE)

# Calcul frequencies relatives
table(x)/length(x)

# Mitjana i variancia (empíriques)
mean(x)

var(x)
```

3.2 Problema 2.2

```
# Obtenim els nombres
u <- runif(50, min=0, max=2)

# Apliquem la inversa
y <- 2 * (u^{1/5})

# Dibuixem l'histograma amb el dibuix de la funció de densitat
hist(y, freq=FALSE, main="", ylab="Densitat", xlim=range(c
   (0.5:3)))
z <- seq(0,2,by=0.05)
t <- 5/32 * z^4
lines(z,t, col = "red")

# Calcul mitjana
mean(u)

# Calcul variancia
var(u)

# El mateix per B = 1000
u <- runif(1000, min=0, max=2)

# Apliquem la inversa
y <- 2 * (u^{1/5})

# Dibuixem l'histograma amb el dibuix de la funció de densitat
hist(y, freq=FALSE, main="", ylab="Densitat",xlim=range(c
   (0.5:2.5)))
z <- seq(0,2,by=0.05)
t <- 5/32 * z^4
lines(z,t, col = "red")

# Calcul mitjana
mean(u)

# Calcul variancia
var(u)
```

4 Estudi i simulació d'una variable discreta

```r
# Apartat a)
# Creació dels vectors de valors xi
x <- c(-5:5)
prob <- c(1/36, 2/36, 3/36, 4/36, 5/36, 6/46, 5/36, 4/36, 3/
   36, 2/36, 1/36)

# Representació de la funció de massa de probabilitat
plot(x,prob,type="h", xlim=c(-5,5), ylim=c(0,1), xlab="Xi",
   ylab="probabilitat")

# Representació de la funció de distribució de probabilitat
acum <- cumsum(prob)
s <- stepfun(x, c(0,acum))
plot(s,verticals=FALSE, xlab="Xi", ylab="probabilitat")

# Calcul esperanca
esp <- sum(x*prob)

# Apartat b)
# Simulem les mostres per n = 50, n = 100 i n = 1000
n = 50
n = 100
n = 1000

x <- c(-5:5)
probs <- c(1/36, 2/36, 3/36, 4/36, 5/36, 6/46, 5/36, 4/36, 3/
   36, 2/36, 1/36)

Y <- sample(x,n,prob = probs, replace=TRUE)
mean(Y)
var(Y)

# Apartat c)
# Taula de frequencies relatives (32 casos possibles)
x <- c(1:6)
probs = c(1/6, 1/6, 1/6, 1/6, 1/6, 1/6)

t <- sample(x, n, prob=probs, replace=TRUE)
v <- sample(x, n, prob=probs, replace=TRUE)

z <- t - v
table(z)/n
```