Pràctica 5

Simulació de variables aleatòries

En les pràctiques anteriors vam utilitzar la funció `rnomvariable` per simular variables aleatòries conegudes. Podeu repassar-ho fent els següents exercicis:

Exercicis:

1. Genereu $B = 20$ nombres aleatoris, seguint la llei $Ber(p)$, on $p = 0.1$, $p = 0.4$, $p = 0.8$.

   Calculeu l'esperança i la variància teòrica d'aquesta llei. Després, calculeu la mitjana i variància empíriques de la llista de nombres aleatoris (observeu que la suma dels valors obtinguts és la freqüència absoluta de 1 i la mitjana empírica és la freqüència relativa).

   Repetiu la simulació amb $B = 1000$.

   Solució: Abans de generar els nombres aleatoris amb la distribució X $\sim$ $Ber(0.1)$, recordem que l'esperança i la variància teòrica en aquest cas han de ser $E(X) = 0.1$ i $Var(X) = 0.1 * (1 - 0.1) = 0.09$.

   Generem amb R una mostra aleàtoria d'una distribució $Ber(0.1)$ i calculem la seva mitjana, variància i freqüències absolutes i relatives.

   ```
   > ber<-rbinom(20,1,0.1)
   > ber
    [1] 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
   > mean(ber)
   [1] 0.05
   > var(ber)
   [1] 0.05
   > table(ber)
   ber
    0  1
   19  1
   > table(ber)/length(ber)
   ber
      0    1
   0.95 0.05
   ```

   Observem que aquests valors s'aproximen als valors que esperàvem però voldríem una aproximació encara millor, per aconseguir-ho només cal repetir les instruccions augmentant la mida de la mostra.

2. Simuleu $B = 15$ valors d'una variable aleatòria X amb llei de Binomial de paràmetres $n = 6$ i $p = 0.3$.

   Calculeu l'esperança i la variància teòriques de X. Després, calculeu la mitjana i variància empíriques de la llista de nombres aleatoris. I calculeu també les freqüències absolutes i relatives, són les que esperàveu?

Repetiu la simulació amb $B = 1000$. Obteniu millors aproximacions?

3. Simuleu $B = 30$ valors d'una variable aleatòria X amb llei de Poisson de paràmetre $\lambda = 3$.

   Calculeu l'esperança i la variància teòriques de X. Després, calculeu la mitjana i variància empíriques de la llista de nombres aleatoris.

   Repetiu la simulació amb $B = 10000$.

4. Simuleu $B = 50$ valors d'una variable aleatòria X amb distribució exponencial de paràmetre $\lambda = 2$. Dibuixeu l'histograma. Superposeu-hi el dibuix de la funció de densitat de probabilitat.

   Calculeu l'esperança i la variància teòriques de X. Després, calculeu la mitjana i variància empíriques de la llista de nombres aleatoris.

   Repetiu la simulació amb $B = 1000$.

Però com ho fem per generar variables aleatòries qualssevol tant si són variables aleàtories discretes com si són absolutament contínues?

5.1 Simulacions de variables discretes

Volem generar ara una mostra de mida n d'una variable aleatòria X discreta que pren valors $a_1, a_2, ..., a_n$ amb probabilitats $p_1, p_2, ..., p_n$ respectivament. És a dir, $P(X = a_i) = p_i, i = 1, ..., n$.

El primer que fem és generar n valors d'una variable uniforme $(0, 1)$: `runif(n)`, aquests valors que acabem de generar seran les probabilitats de manera que el segon pas serà convertir aquests valors en els valors $a_1, \ldots a_n$:

valors	**convertir**
entre 0 i p_1	a_1
entre p_1 i $p_1 + p_2$	a_2
$\ldots$	
entre $p_1 + \ldots + p_{n-1}$ i 1	a_n

El mètode per fer la simulació serà:

- obtenir nombres aleatoris $u_1, u_2, ..., u_n$,

```
>x<-runif(n)
```

- fer la transformació següent: donat u_i, existirà un k tal que $u_i \in (p_1 + ... + p_{k-1}, p_1 + ... + p_k]$. Aleshores caldrà transformar u_i en a_k.

```
>x1<-(x<=p_1)
>x2<-(x>p_1&x<=p_1+p_2)
>x3<-(x>p_1+p_2&x<=p_1+p_2+p_3)
>...
>xn<-(x>p_1+p_2+...+p_{n-1})
>y<-a_1*x1+...+a_n*x_n
```

La variable y és la que contindrà els n valors generats amb la funció de probabilitat donada.

Exemple. Volem generar una mostra de mida 10 d'una variable aleatòria X tal que:

$$\begin{aligned} P(X=5) &= \frac{1}{3}, \\ P(X=7) &= \frac{1}{2}, \\ P(X=9) &= \frac{1}{6}. \end{aligned}$$

Aleshores el mètode serà:

- Obtenir els nombres aleatoris:

```
> u<-runif(10)
> u
 [1] 0.3405060 0.2556905 0.1379111 0.5610426 0.2238937
     0.1841603 0.3919349 0.9773482 0.4329352 0.4750485
```

- fer la transformació:

valors	**convertir a**
entre 0 i $\frac{1}{3}$	5
entre $\frac{1}{3}$ i $\frac{1}{3}+\frac{1}{2}=\frac{5}{6}$	7
entre $\frac{5}{6}$ i 1	9

```
> x1<-(u<=1/3)
> x1
 [1] FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE
> x2<-(u>1/3&u<=1/3+1/2)
> x2
 [1]  TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE
> x3<-(u>1/3+1/2)
> x3
 [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE
> y<-5*x1+7*x2+9*x3
> y
 [1] 7 5 5 7 5 5 7 9 7 7
```

Igualment, el programari **R** porta una funció que permet generar variables aleatòries discretes amb una sola instrucció, per exemple podem repetir l'exemple que acabem de resoldre utilitzant:

```
> sample(c(5,7,9),10,prob=c(1/3,1/2,1/6),replace=TRUE)
 [1] 5 7 5 7 7 9 5 9 7 7
```

5.1.1 Problemes

1. Genereu $B=100$ nombres aleatoris, seguint la llei de probabilitat de la variable aleatòria X que té la funció de massa de probabilitat següent:

Valors x_i de X	2	3	7
Probabilitats $p_i = P[X=x_i]$	0.2	0.5	0.3

Calculeu les freqüències relatives de $[X = 2]$, $[X = 3]$, $[X = 7]$ a la llista de nombres aleatoris, i compareu-les amb les corresponents probabilitats teòriques.

Calculeu (teòricament) l'esperança i la variància de X. Després, calculeu la mitjana i variància empíriques de la llista de nombres aleatoris.

Repetiu la simulació amb $B = 1000$.

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

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

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.

Calculeu (teòricament) l'esperança i la variància de X. Després, calculeu la mitjana i variància empiriques de la llista de nombres aleatoris.

Repetiu la simulació amb $B = 10000$.

3. Simuleu el llançament d'un dau 1000 vegades. Quines freqüències absolutes i relatives obteniu? Són les que us esperàveu?

Ara, el que volem simular és la variable aleatòria que s'obté quan sumem el resultat del llançament de dos daus. Quins valors pren aquesta variable? I amb quines probabilitats? De quina distribució es tracta?

Responeu les mateixes preguntes si en lloc de sumar els resultats del llançament de dos daus els restem.

5.2 Simulacions de variables absolutament contínues

En el cas que vulguem simular una variable aleatòria absolutament contínua que no sigui cap de les conegudes, utilitzarem, si podem el **mètode de la funció inversa**. La idea és que volem generar una mostra de mida n d'una variable aleatòria X que té funció de distribució **contínua** i **invertible** F_X. Hi ha un resultat que ens diu que si considerem U una variable aleatòria $U(0, 1)$ aleshores la variable aleatòria

$$V = F_X^{-1}(U) \text{ té la mateixa distribució que la variable } X.$$

En efecte, mirem la seva funció de distribució:

$$F_V(x) = P(V \leq x) = P\left(F_X^{-1}(U) \leq x\right) = P\left(U \leq F_X(x)\right) = F_X(x).$$

Aleshores, el mètode per fer la simulació serà:

- obtenir nombres aleatoris $u_1, u_2, ..., u_n$,

```
>u<-runif(n)
```

- calcular F_X^{-1},
- aplicar-la als nombres aleatoris i obtenir la mostra $x_1 = F_X^{-1}(u_1), ..., x_n = F_X^{-1}(u_n)$.

$\boxed{\text{Exemple.}}$ Volem generar una mostra de mida 10 d'una variable aleatòria amb densitat

$$f(x) = 3x^2 \mathbb{1}_{(0,1)}(x).$$

Aleshores:

- Obtenim els nombres aleatoris:

```
> u<-runif(10)
> u
 [1] 0.3513617 0.4313612 0.7389351 0.6115086 0.8457035
     0.5707821 0.4215648 0.7860317 0.4976643 0.6077936
```

- Calculem la funció de distribució

$$F(x) = \begin{cases} 0, & x < 0, \\ \int_0^x 3y^2 dy = x^3, & 0 \leq x < 1, \\ 1, & 1 \leq x, \end{cases}$$

de manera que la inversa és $F^{-1}(u) = \sqrt[3]{u}$.

- Apliquem F^{-1} :

```
> y<-u^{1/3}
> y
 [1] 0.7056426 0.7555798 0.9040701 0.8487912 0.9456695
     0.8295135 0.7498161 0.9228831 0.7924627 0.8470689
```

- Dibuixem l'histograma amb el dibuix de la funció de densitat.

```
> hist(y,freq=FALSE)
> z<-seq(0,1,by=0.05)
> t<-3*z^2
> lines(z,t)
```

5.2.1 Problemes

1. Simuleu $B = 50$ valors d'una variable aleatòria X, absolutament contínua, amb funció de densitat de probabilitat:
$$f(x) = 2x, \quad 0 < x < 1.$$
Dibuixeu l'histograma. Superposeu-hi el dibuix de la funció de densitat de probabilitat.

   Calculeu (teòricament) l'esperança i la variància de X. Després, calculeu la mitjana i variància empíriques de la llista de nombres aleatoris.

   Repetiu la simulació amb $B = 1000$.

2. Simuleu $B = 50$ valors d'una variable aleatòria X, absolutament contínua, amb funció de densitat de probabilitat:
$$f(x) = \frac{5}{32}\,x^4, \quad 0 < x < 2.$$
Dibuixeu l'histograma. Superposeu-hi el dibuix de la funció de densitat de probabilitat.

   Calculeu (teòricament) l'esperança i la variància de X. Després, calculeu la mitjana i variància empíriques de la llista de nombres aleatoris.

   Repetiu la simulació amb $B = 1000$.

3. Simuleu $B = 50$ valors d'una variable aleatòria X, absolutament contínua, amb funció de distribució
$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{7}(x+1)^3 - \frac{1}{7} & \text{si } 0 < x < 1 \\ 1 & \text{si } x > 1 \end{cases}$$
Dibuixeu l'histograma. Superposeu-hi el dibuix de la funció de densitat de probabilitat.

Calculeu (teòricament) l'esperança i la variància de X. Després, calculeu la mitjana i variància empíriques de la llista de nombres aleatoris.

Repetiu la simulació amb $B = 1000$.