

Introduccio a la Probabilitat

GEA

Pràctica 1: Simulant probabilitats d'esdeveniments amb R

Generació de nombres aleatoris

Com a software estadístic, R disposa d'una llibreria molt completa de funcions que permeten generar nombres aleatoris segons diferents distribucions de probabilitat (que anirem veient al llarg de l'assignatura). De moment, veurem la més simple de totes, que ens permet generar un nombre aleatori (en realitat és pseudo-aleatori) entre 0 i 1 o, més en general, entre dos nombres qualssevol.

La funció **runif**

Si escrivim:

```
x ← runif(1)
x
```

veurem que a x s'ha guardat un nombre entre 0 i 1. Si volem 10 nombres entre 0 i 1 fem

```
runif(10)
```

i si volem 3 nombres aleatoris (reals) entre 5 i 10, per exemple, fem

```
runif(3,5,10).
```

IMPORTANT!

Cada vegada que executeu alguna d'aquestes instruccions, obteniu valors diferents. Si voleu obtenir els mateixos valors (per a poder repetir exactament la mateixa simulació), heu de fixar la **llavor aleatòria** que es fa servir per a generar els nombres (pseudo-)aleatoris. Si no es fixa, aquesta llavor és triada per R (no entrarem en com ho fa) de tal manera que cada vegada que s'executa una instrucció, obtenim un resultat diferent. La instrucció per a fixar la llavor és:

```
set.seed(99)
```

(hem posat 99, però qualsevol nombre **ENTER** és vàlid).

La funció **sample**

Aquesta funció ens permet generar un o més nombres aleatoris **ENTERS** entre dos **ENTERS** donats.

Per exemple,

```
sample(1:10,2)
```

genera 2 nombres enters entre 1 i 10 **diferents** (és a dir, fent **mostreig sense reemplaçament**). Si feu

```
sample(1:3,5)
```

veureu que apareix un missatge d'error perquè no pot triar 5 nombres enters entre 1 i 3 sense repetir. Si voleu fer això, s'ha d'afegir l'opció per a fer-ho:

```
sample(1:3,5,replace=TRUE).
```

Observeu que ara es fa **mostreig amb reemplaçament**.

L'opció per defecte, si no s'indica el contrari, és **sense** reemplaçament, és a dir, per defecte, a la funció `sample` es té l'opció `replace=FALSE`. Per tant, si per exemple volem generar 5 nombres a l'atzar entre 1 i 10 permetent que hi hagi repeticions, haurem de fer servir

```
sample(1:10,5,replace=TRUE).
```

Però com que 5 és més petit que el nombre de possibilitats, que són 10, podem obtenir repeticions... o no!

Problema 1

Quina és la probabilitat que en executar la comanda

```
sample(1:10,5,replace=TRUE)
```

no obtinguem cap repetició en els 5 nombres que generem?

Heu de resoldre el problema de dues maneres:

- Analíticament (resultat **exacte**).
- Simulant amb R l'execució de la comanda 100 000 vegades (resultat **aproximat**).

Indicació: per a fer la simulació, la idea és crear un comptador que comenci en 0 i vagi augmentant cada vegada que en l'execució de la comanda, en els 5 nombres generats **no** hi hagi cap de repetit. La funció **for()** de R ens permet fer un bucle per a les 100 000 repeticions, i la funció **if()** ens permet afegir un 1 al comptador si i només si en els 5 nombres generats en cada execució **no** hi ha cap de repetit. Aquestes dues funcions d'R són molt útils.

Problema 2

Volem calcular la probabilitat de guanyar jugant a una Lotto, que per a poder fer les simulacions amb R més ràpid, per comptes de ser la Lotto 6/49 serà la **Lotto 4/10**.

Consisteix a triar 4 nombres (NATURALS) diferents, de l'1 al 10.

Es guanya el primer premi si s'encerta els 4 (l'ordre no importa). Es guanya el segon premi si s'encerta **exactament** 3 dels 4.

Heu de calcular la probabilitat de guanyar el primer premi, i també la de guanyar el segon premi, si només es fa una aposta (es trien els 4 nombres amb els que es juga).

- Analíticament (resultats **exactes**).
- Simulant amb R el sorteig 100 000 vegades (resultats **aproximat**s).