Stat Inf 2

Liste d'exercices Nro 2. – Simulation

L3 IDS – Univ Lumière Lyon 2 – Hiver 2022

1 Représentation binaire

1. Écrire 45 en base 2, ainsi que $45 \pmod{16}$ et $45 \pmod{17}$.
2. Que remarquez-vous lorsque vous regardez ces trois représentations ?
3. Écrire un code permettant de générer la représentation binaire d'un nombre.

2 Simulation d'un échantillon

Considérons la variable aléatoire discrète avec la fonction de masse suivante :

$$\mathbb{P}[X=1]=0.1, \quad \mathbb{P}[X=2]=0.3, \quad \mathbb{P}[X=5]=0.6.$$

1. Représenter la fonction de répartition de cette variable aléatoire X.
2. Écrire un programme qui permet de générer un échantillon suivant cette distribution à l'aide de la la loi uniforme.

3 Simulation d'une loi de Poisson

Soit X décrivant une loi de Poisson de paramètre λ, $\mathcal{P}(\lambda)$. On note $F(x)=\mathbb{P}[X\leq x]$ et $p(x)=\mathbb{P}[X=x]$.

1. Montrer que la fonction de masse vérifie $p(x+1)=\dfrac{\lambda}{x+1}p(x)$.
2. A partir de cette relation, écrire une fonction permettant de calculer les valeurs des fonctions de masse et de répartition.
3. Si $X\in\mathbb{N}$ est une variable aléatoire et $F.x$ est une fonction qui retourne la valeur de F pour la variable aléatoire X, alors on peut simuler X avec le programme suivant

```
R.rand <- function(){
  u = runif(1)
  x = 0
  while(F(x) < u){
    x = x+1
  }
  return(x)
}
```

   Dans le cas d'une distribution de Poisson, ce programme peut largement être améliorer en utilisant deux nouvelles variables $p.x$ et $F.x$. Améliorer la fonction précédente en tenant compte de ces deux informations.

4 Méthode de l'inverse I

On considère la variable aléatoire continue dont la densité est donnée par

$$f_X(x)=\begin{cases}3(x-1)^2 & \text{si } 1<x\leq 2,0 \\ \text{sinon.} & \end{cases}$$

1. Tracer le graphe de la fonction de répartition associée.
2. Montrer comment simuler une variable aléatoire ayant cette loi avec la méthode de l'inverse et écrire le code `R` correspondant.

5 Méthode de l'inverse II

On considère la variable aléatoire continue X ayant la densité suivante pour tout réel x :

$$f_X(x) = \frac{\exp(-x)}{(1+\exp(-x))^2}.$$

On dit que la variable aléatoire X a une distribution logistique (souvenez de la régression logistique qui a été présentée au semestre précédent).

1. Déterminer la fonction de réparation de cette variable aléatoire.
2. Déterminer l'inverse de la fonction de répartition et écrire un code permettant de simuler un échantillon ayant cette loi à l'aide de la méthode de l'inverse.

6 Un cas concret

Des personnes arrivent dans un magasin de chaussures. Chaque personne regarde un certain nombre de chaussures avant de décider quelle paire acheter.

1. Soit N le nombre de personnes qui entre dans le magasin en une heure. Sachant que $\mathbb{E}[N] = 10$, quelle serait une bonne distribution pour la variable aléatoire N ?
2. Le client i essaie X_i paires de chaussures qu'il n'aime pas avant de trouver une paire qui lui convienne, *i.e.* $X_i \in \{0,1,2,\cdots\}$. Supposons que la probabilité p qu'ils aiment une paire de chaussures est de 0.8, indépendamment des autres paires regardées. Quelle serait une bonne distribution de X_i.
3. Soit Y le nombre total de paires de chaussures qui ont été essayées, sans tenir compte de celles achetées. Supposons que chaque client agisse indépendamment des autres, donner une expression de Y en fonction de N et de X_i et écrire une fonction permettant de simuler N,X_i et Y.
4. Quelle est la valeur de $\mathbb{P}[Y = 0]$? Essayer de retrouver cette valeur à l'aide du code précédent.

7 Teorème Central Limite

Dans cet exercice nous utiliserons la simulation numérique pour illustrer les phénomènes de convergence déjà étudiés dans le cours de Stat Inf 1. Nous nous intéresserons en particulier aux situations pathologiques, c'est-à-dire quand la convergence n'est pas observée lors de l'expérience numérique.

Pour rappel, la loi de grands nombres prédit une convergence de la moyennes empirique $\hat{X}_n$suite de variables aléatoires iid $X_1,X_2,...$ vers leur espérance $\mu = E[X]$. Le TCL prédit la loi limite.

1. Obtenir n observations par tirage d'une va normale de moyenne $\mu = 3$ et variance égal à 9. Représenter les réalisations dans l'ordre du tirage, puis de la plus petite à la plus grande.
2. Soit $p_i = \frac{i-1/2}{n}$ les fréquences empiriques (ou "points de probabilités, c.f. `ppoints`) liées aux données simulées. Obtenir les représentations `qqplot` et `qqnorm` de données simulées.
3. Proposer une expérience par simulation capable de montrer la convergence en loi de la moyenne empirique. Utiliser d'abord une loi normale pour les tirages. Puis, modifiez la loi par une où la convergence n'est pas garantie.