TP 1: Variables discrètes

anna.melnykova@univ-avignon.fr

Le but de cette partie est la prise en main du logiciel R. La commande **runif(n)** simule n variables uniformes de la loi $\mathcal{U}[0,1]$. On va systématiquement utiliser ce générateur de la loi uniforme pour simuler variables aléatoires qui suivent autres lois probabilistes. Pour les variables discrètes, on utilise souvent la transformation des variables uniformes. Pour les variables continues, on utilise plutôt l'inversion de la fonction de répartition ou la méthode du rejet.

Concepts à assimiler:

- Moyenne et variance empiriques
- Lien entre **l'espérance** d'une variable aléatoire X et la **moyenne empirique** d'un échantillon contenant les réalisations de cette variable
- Lien entre la **probabilité d'évènement** E et la **proportion des observations** de cet évènement dans l'échantillon
- Outils graphiques: histogrammes.

En début de chaque TP vous êtes invité(e)s à créer un nouveau script dans le logiciel RStudio. Pour faire ça, vous devez aller dans menu 'File -> New file -> R Script'. Puis, cliquez sur 'Sauvegarder' (ça peut se faire avec Ctrl + S), donnez un nom à votre fichier (par exemple, `Dupont_G1_TP1.R`) et sauvegardez-le dans "Mes documents". Pensez bien à sauvegarder le script régulièrement pour ne pas perdre votre travail en cas de souci informatique!

Exercice 1: Prise en main R

- Exécutez les commandes suivantes et interprétez la sortie.

```
n = 10
?runif
U_n = runif(n)
U_n
?mean
mean(U_n)
?var
var(U_n)
hist(U_n, probability = TRUE)
```

- Prenez $n = 100$ et re-exécutez le code. Qu'est-ce que vous observez? Qu'est-ce que vous observez pour $n = 1000$? Est-ce que l'augmentation de la taille d'échantillon change l'histogramme? La moyenne? La variance?

- Théoriquement, quelle est la probabilité que $U < 0.5$, où $U \sim \mathcal{U}[0,1]$?

- Maintenant, on va créer un nouveau vecteur X_n. Exécutez le code suivant:

```r
X_n = (U_n<0.5)
as.numeric(X_n)
sum(X_n)
```

- X c'est une variable qui est égale à 1, si $U < 0.5$ et 0 sinon. Quelle est la probabilité que $X = 1$? Quelle est la loi de X? Quelle est son espérance? Variance?

- Calculez la moyenne empirique et la variance du vecteur X_n? Est-ce qu'elles correspondent à la moyenne et la variance théoriques?

- Que passe-t-il si on considère une autre variable Y, qui est égale à 1, si $U < p$ et 0 sinon (pour $p \in [0, 1]$)? Quelle est sa loi? Quelle sera sa moyenne? Sa variance?

Exercice 2: Air-France

La liaison Air-France Avignon-Paris est assurée par des Fokker 100 d'une capacité de 104 places. Sachant que la probabilité qu'un passager ayant réservé ne se présente pas à l'embarquement est de 0.07, on cherche à estimer la probabilité qu'AF se trouve en situation embarrassante si elle vend 110 réservation.

- Air-France embauche une ingénieure qui propose de faire une étude numérique. En adaptant le code de la Partie I, elle simule le résultat du vente de 110 tickets (i.e. le nombre de passagers qui se présente à l'embarquement). Répliquez la code de l'ingénieure
- Maintenant, elle décide de répéter l'expérience 1000 fois pour obtenir une estimation plus ou moins fiable de probabilité de se retrouver en situation embarrassante. Complétez la code suivante en utilisant votre algorithme précédent et commentez le résultat. Est-ce qu'on a réussi, à votre avis, à approcher la probabilité théorique?

```r
n = 1000
vol_n = numeric(n) # 0. on déclare un vecteur numérique vide
for (i in 1:n){
  # 1. on simule 110 variables uniformes
  U_n <- runif(110)
  # 2. on crée un vecteur binaire avec 1 et 0 pour les passagers qui sont venus ou pas
  X_n <-
  # 3. on calcule le nombre de passagers venus
  n_pass <-
  # 4. on retourne 1 s'il y a plus de passagers que des places
  vol_n[i] = as.numeric((104-n_pass)<0)
}
# il nous reste de calculer la moyenne
# elle correspond à la probabilité en question pour n suffisament grand!
mean(vol_n)
```

- Air France a embauché en plus une probabiliste qui décide de vérifier les résultats empiriques fournis par l'ingénieure. Elle décide maintenant de calculer la probabilité que tous les passagers viennent à l'embarquement. Fournissez une formule mathématique que vous avez vu en cours. Calculez dans R (ça fonctionne comme une calculatrice aussi).

- Quelle est la loi de probabilité d'un évènement 'k passagers parmi 110 viennent à l'embarquement'?

- Calculez la probabilité que le vol est surchargé (au moins 105 passagers viennent). Comparez aux résultats empiriques.

- On peut vérifier notre calcul par une des commandes R suivantes:

```
k = 105:110
sum(choose(110,k)*(0.93^k)*(0.07^(110-k)))
sum(dbinom(k,110,0.93))
pbinom(104,110,0.93, lower.tail = FALSE)
```

- **Question bonus:** estimez empiriquement le nombre des billets que la compagnie aérienne puisse vendre en limitant le risque de la surréservation du vol à 5%. Pour ça, vous pouvez varier le nombre des billets vendus jusqu'à que le pourcentage des vols surréservés reste inférieur à 5%. Quel est ce nombre?

Exercice 3: jeux de dés à 6 faces

Proposez un algorithme qui permets de simuler un dé équilibré à 6 faces à partir d'un générateur d'une variable uniforme. Vous pouvez utiliser, par exemple, le double conditionnement en développant votre code: pour vérifier que 2 conditions sont remplies en même temps, on utilise le syntax $A\&B$ (par exemple, $(X > 10)\&(X <= 12)$ va retourner `TRUE` si et seulement si la variable X est comprise entre 10 et 12 inclus).

Comparez le résultat de votre algorithme avec le résultat de la commande suivante: `floor(6*runif(1)+1)`. Quel algorithme est plus facile à généraliser?

Exercice 4: La décomposition radioactive

On considéra la variable X qui compte le nombre d'émissions des α-particules de Carbon-14 qui sont comptés par le compteur de Geiger chaque seconde. On supposera que X suit une distribution de Poisson avec le paramètre $\lambda = 16$. Rappelez-vous la fonction de masse pour une variable de Poisson:

Pour simuler la variable de Poisson, on utilisera l'algorithme proposé en 1968 par Donald Knuth:

```
lmd = 16        # Paramètre lambda
n = 100         # Nombre des réalisations
X = numeric(n)  # Vecteur pour stocker les observations
for (i in 1:100){
  p = 1           # Probabilité
  k = 0           # Compteur
  while (p>exp(-lmd)){
    U <- runif(1)
    p = p*U
    k = k + 1
  }
 X[i] = k-1
}
```

- Construisez un histogramme *en probabilités* de vecteur X. Notez la commande utilisée.

- Calculez la moyenne et le mode du vecteur. Est-ce qu'ils s'approchent aux valeurs théoriques? C'est quoi les valeurs théoriques, d'ailleurs (regardez dans le cours)?

- Augmentez la taille d'échantillon (en mettant, par exemple, $n = 1000$) et répétez. Est-ce que la moyenne s'approche à la moyenne théorique?

- Estimez la probabilité (empiriquement) que dans une seconde on observe plus de 20 émissions de Carbon-14. Moins de 10?

- Pour ne pas écrire l'algorithme, on peut simuler un échantillon des variables de Poisson avec la commande `rpois`. Exécutez la ligne du code suivante et comparez les vecteurs X et Y (i.e., leur moyenne, variance, les histogrammes):

```
Y = rpois(n, lmd)
```