TP 3

Statistiques descriptives univariées

Exercice 1 (★★)

La commande `rd.binomial(n,p,r)` renvoie un vecteur contenant r simulations de la loi binomiale $\mathscr{B}(n,p)$.

1. À l'aide de la commande `rd.binomial`, simuler 10000 nombres suivant la loi $\mathscr{B}(10, 0.5)$. On notera `x` le vecteur contenant cette série statistique.
2. Déterminer l'effectif, la fréquence et la fréquence cumulée de la modalité 5.
3. Déterminer la moyenne, la médiane et l'écart-type de `x`. Était-ce prévisible ?
4. Créer un vecteur `m` de taille 11 tel que `m[k]` contient l'effectif de la modalité k. Déterminer le(s) mode(s) de la série `x`.
5. Représenter à l'aide de la commande `plt.bar` les diagrammes en bâtons des effectifs, des fréquences et des fréquences cumulées de la série `x`.
6. Représenter de nouveau le diagramme en bâtons des effectifs et des fréquences de la série `x`, cette fois à l'aide de la commande `plt.hist`.

1. On peut utiliser la commande `x = rd.binomial(10,0.5,10000)`.
2. On peut procéder ainsi :

```
>>> np.sum(x==5)
2413
>>> np.mean(x==5)
0.2413
>>> np.mean(x<=5)
0.6192
```

3. On procède comme suit :

```
>>> np.mean(x)
5.0079
>>> np.median(x)
5.0
>>> np.var(x)
2.51523759
>>> np.std(x)
1.5859500591128335
```

C'était en effet prévisible pour la moyenne et la variance (ou l'écart-type), puisque si $X \hookrightarrow \mathscr{B}(10, 0.5)$ alors $E(X) = 10 \times 0.5 = 5$ et $V(X) = 10 \times 0.5 \times (1 - 0.5) = 2.5$.

4. On peut procéder ainsi :

```python
for k in range (11):
    m[k] = np.sum(x == k)
```

En exécutant cette commande, on obtient en appelant `m` :

```
>>> m
array([  13.,   104.,   434.,  1261.,  2081.,  2407.,  2082.,  1097.,   425.,
          86.,    10.])
```

Le mode (modalité dont l'effectif est le plus important) est donc 5.

5. On peut procéder ainsi :

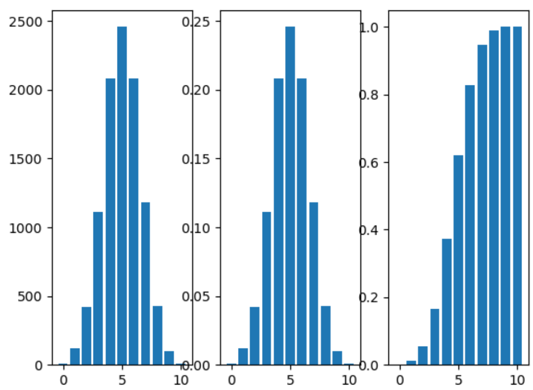
```
u = np.arange(11) # modalités
plt.subplot(1,3,1) #pour placer les hist dans une même fenètre
plt.bar(u,m)

f = m/10000 # fréquences des modalit és
plt.subplot(1,3,2)
plt.bar(u,f)

fcum = np.cumsum(f) # fréquences cumulées des modalit és
plt.subplot(1,3,3)
plt.bar(u,fcum)

plt.show()
```

On obtient les diagrammes en bâtons suivants :



6. On peut procéder ainsi :

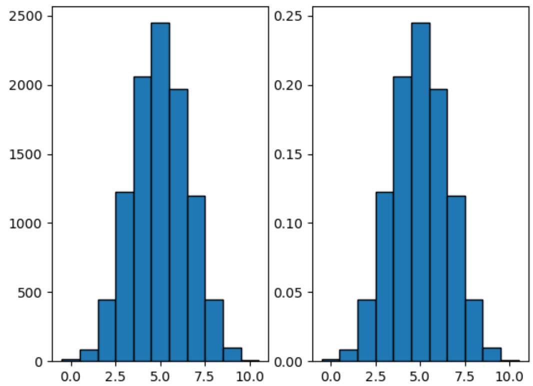
```
u = np.arange(-0.5, 11)

plt.subplot(1,2,1)
plt.hist(x,u,edgecolor = 'k') # diag en batons des effectifs

plt.subplot(1,2,2)
plt.hist(x,u,density = 'True', edgecolor = 'k') # diag en batons des fréquences
```

```
plt.show()
```

On obtient :



Exercice 2 (★★)

La commande `rd.random(n)` permet de simuler un vecteur de taille n dont chaque coefficient est un nombre réel choisi aléatoirement entre 0 et 1.

1. Créer un vecteur `x` contenant 10000 nombres réels choisis aléatoirement entre 1 et 5.
2. Calculer la moyenne, la médiane, l'écart-type et l'étendue de la série statistique `x`.
3. Faut-il mieux regrouper cette série statistique par modalités ou par classes ? Pourquoi ?
4. Tracer l'histogramme associé à cette série statistique en la regroupant par classes (choisir 100 classes de même amplitude). Que remarque-t-on ?

1. On peut procéder ainsi (en se souvenant que si $U \hookrightarrow \mathscr{U}([0,1])$, alors $(b-a)U + a \hookrightarrow \mathscr{U}([a,b])$) :

```
x = 4*rd.random(10000)+1
```

2. On peut procéder ainsi :

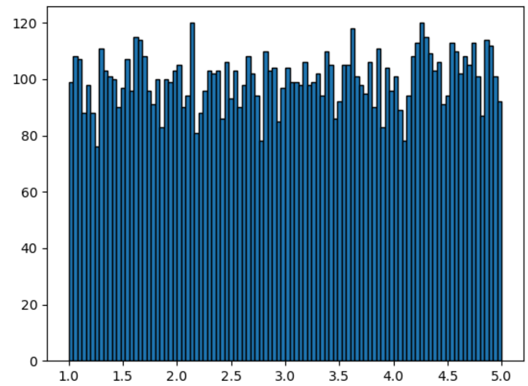
```
>>> np.mean(x)
2.9905154186575342
>>> np.median(x)
2.9768217953857503
>>> np.std(x)
1.150978263442839
>>> np.max(x)-np.min(x)
3.9993452821628632
```

3. Il vaut mieux regrouper cette série statistique par classes. En effet, on peut s'attendre à avoir un très grand nombre de modalités, 10000, étant donné qu'on génère notre échantillon à l'aide d'une loi à densité. Un tri par modalités n'aurait aucun intérêt : l'effectif de chaque modalité sera de 1.

4. On procède ainsi :

```
plt.hist(x,100, edgecolor = 'k')
plt.show()
```

On obtient :



Exercice 3 (Interrogation de la base de données)

Écrire une fonction `donnees(n)` qui affiche le nom, la superficie, le nombre d'habitants et la densité de population du pays d'index n.

Exercice 4

1. Calculer la surface terrestre mondiale, le nombre d'habitants mondial et la densité moyenne d'habitants au km^2.
2. Calculer la surface terrestre, le nombre d'habitants et la densité moyenne d'habitants au km^2 pour chaque continent.
3. Représenter la densité moyenne d'habitants au km^2 pour chaque continent en utilisant un diagramme en bâtons (on mettra en abscisse des entiers de 1 à 5).
4. Faire de même pour la répartition de la surface terrestre par continent, puis du nombre d'habitants par continent.

Exercice 5

On considère l'espérance de vie des hommes (ou des femmes) par pays.

1. Calculer la moyenne sur l'ensemble des pays. Ce résultat correspond-il à l'espérance de vie mondiale des hommes (ou des femmes) ?
2. Calculer l'écart-type et la médiane.
3. Calculer les espérances de vie minimale et maximale en précisant les pays correspondant à ces valeurs extrémales.
4. Représenter l'histogramme de l'espérance de vie des hommes sur l'intervalle $[0, 100]$ avec 20 classes. Quelle est la classe modale de l'espérance de vie des hommes ?
5. On s'intéresse au tableau `homme`.
   (a) Vérifier que ses premier et troisième quartiles valent respectivement 64 et 76.

(b) Vérifier que ses premier et neuvième déciles valent respectivement 59 et 80. Donner la liste des pays dont l'espérance de vie des hommes est inférieure au premier décile ou supérieure au neuvième décile.

Exercice 6

On rappelle que le taux d'accroissement naturel est la différence entre la natalité et la mortalité.

1. Quels sont les accroissements minimaux et maximaux ? Préciser les pays.
2. Faire afficher la liste des pays pour lesquels l'accroissement est négatif.
3. Déterminer l'accroissement mondial moyen.
4. Dans ses projections, l'INED prévoit une population mondiale de 9731 millions d'habitants en 2050. Cela est-il conforme à l'hypothèse d'un taux d'accroissement constant ?
