**M. Sublime**

**M. Sébastien MASCHA**

**ISEP Paris - 17 Septembre 2019**

**Analyse de données – Lab 1**

Sommaire

[I. Séries discrètes 4](#_Toc19832995)

[1. Question 1 - Générez un vecteur contenant 1000 valeurs aléatoires incluses entre 0 et 10 : 4](#_Toc19832996)

[2. Questions 3 à 5 - Déterminez le mode, la moyenne et la médiane de ces données sans utiliser les fonctions Python prédéfinies. 4](#_Toc19832997)

[3. Question 6 - Calculez le domaine de définition, l'écart-type et la variance de ces données 4](#_Toc19832998)

[II. Partie B - Séries discrètes par fréquence 6](#_Toc19832999)

[1. Question 1 - Créez le jeu de données correspondant et représentez le sous forme d'un histogramme. 6](#_Toc19833000)

[III. Partie c - Distribution gaussiennes 8](#_Toc19833001)

[IV. Partie D - Analyse de QI 9](#_Toc19833002)

[1. Générez 3 échantillons contenant 10, 1000 et 100000 observations. 9](#_Toc19833003)

[a. Initialisation 9](#_Toc19833004)

[b. Comparez les valeurs trouvées pour l'écart moyen et l'écart type avec les valeurs théoriques. 9](#_Toc19833005)

[c. Commenter l'écart type et IC95 9](#_Toc19833006)

[2. Utilisez la commande read\_csv(file) de la librairie pandas afin d'ouvrir le fichier malnutrition.csv 10](#_Toc19833007)

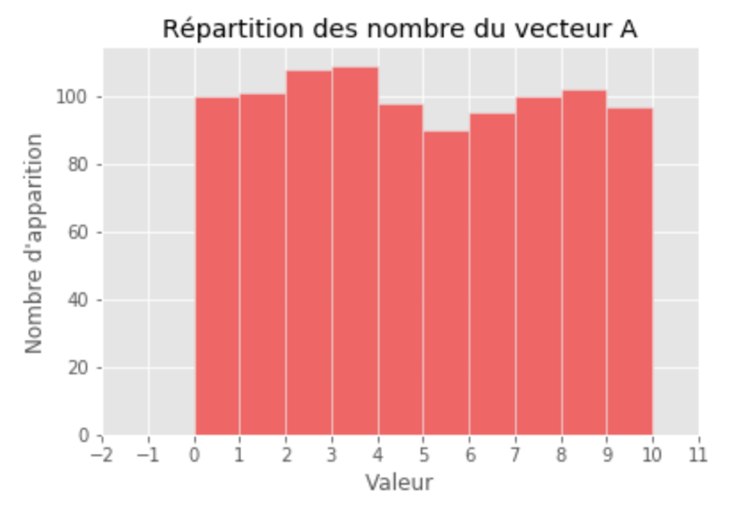
[3. A l'aide des outils statistiques à votre disposition, que pouvez-vous conclure sur l'effet de la malnutrition sur le QI lorsque vous comparez ces résultats avec ceux de votre échantillon de 100000 individus ? 10](#_Toc19833008)

# Séries discrètes

## Question 1 - Générez un vecteur contenant 1000 valeurs aléatoires incluses entre 0 et 10 :

A = np.random.randint(0,10,1000)



Fig. 1 : Histogramme du vecteur A

## Questions 3 à 5 - Déterminez le mode, la moyenne et la médiane de ces données sans utiliser les fonctions Python prédéfinies.

Le mode de A est 1. On restrouve aussi cette valeur avec la fonction prédéfinie stats de la librairie scipy.

La médiane est 4,674.



On remarque que la mediane s'écarte de la moyenne d'autant plus que le jeu de données est grand.

## Question 6 - Calculez le domaine de définition, l'écart-type et la variance de ces données

Le domaine de définition de A est : [ 0 , 9 ]

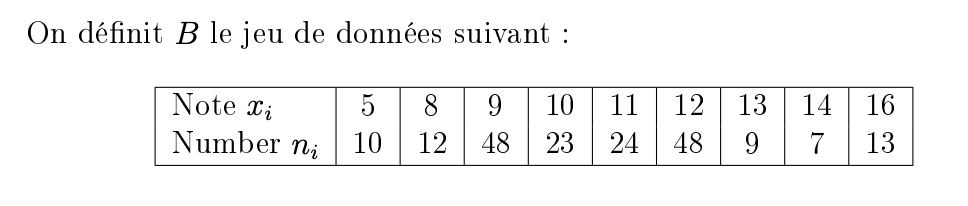
Son écart-type est de : 2.8737130989366317

La variance est de 8.25822697499998

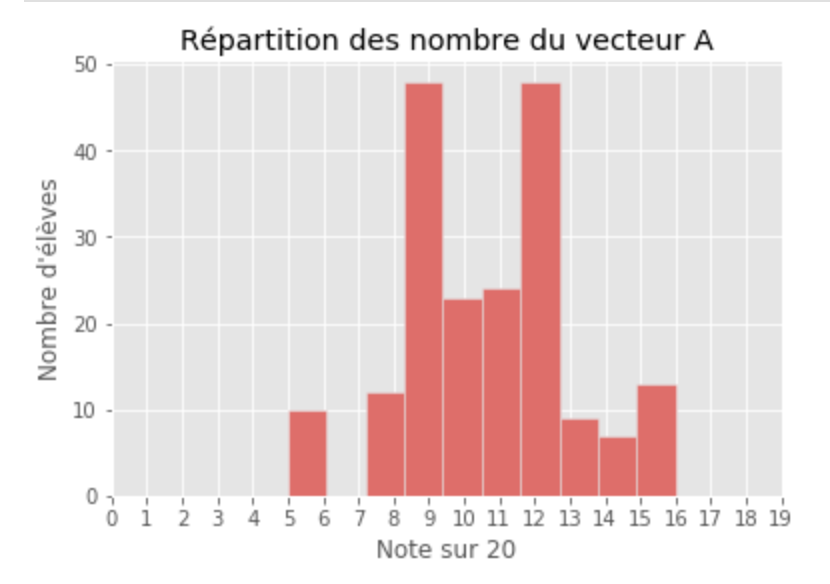
Le domaine de définition sera toujours le même, il est unique.

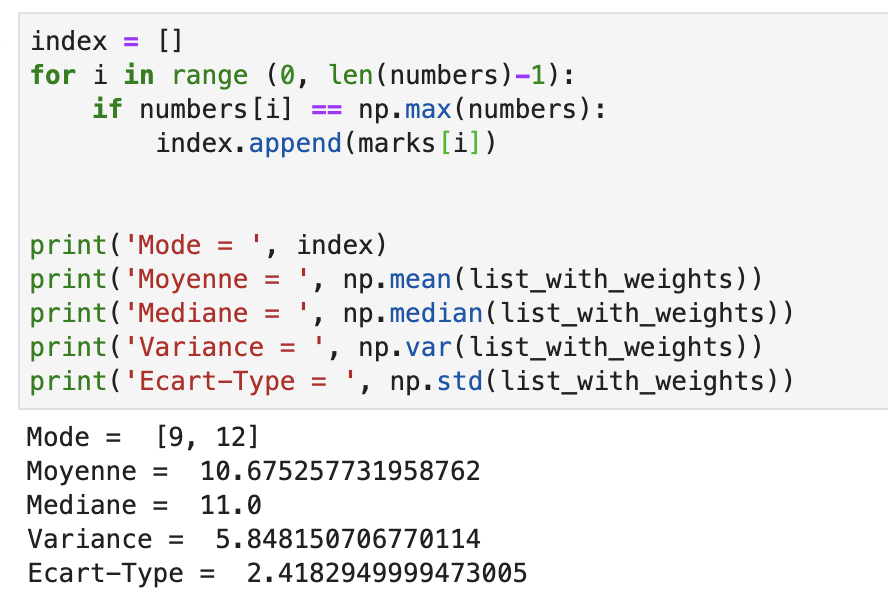
Pour l'ecart type et la variance les valeurs peuvent changer en fonction de la méthode de calcul utilisée par les fonctions pré-définies.

# Partie B - Séries discrètes par fréquence



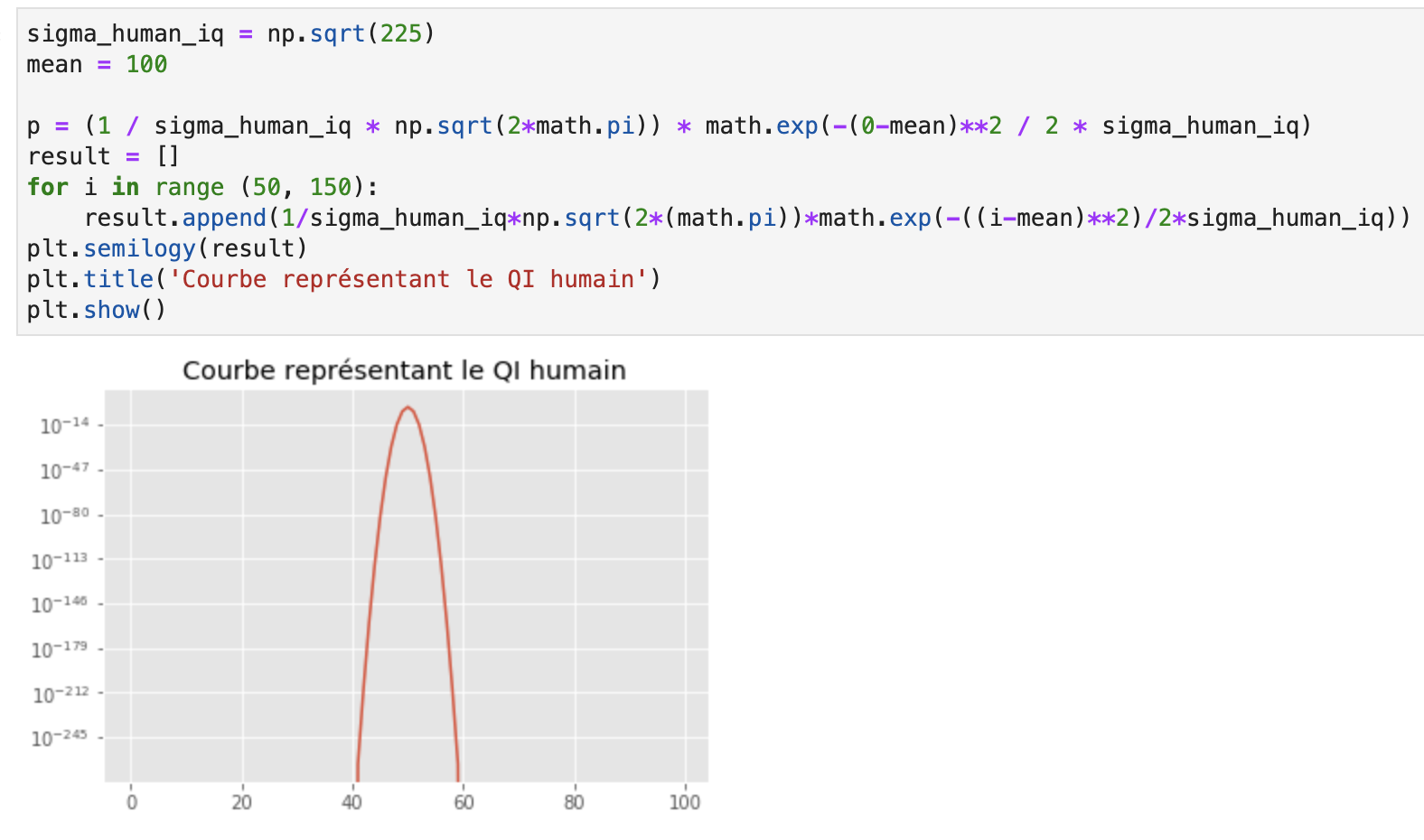
## Question 1 - Créez le jeu de données correspondant et représentez le sous forme d'un histogramme.





La distribution des notes indique deux maxima atteints pour les notes 9 et 12. Il existe donc par définition deux modes pour cette série. C'est alors bien une distribution bimodale.

# Partie c - Distribution gaussiennes





# Partie D - Analyse de QI

## Générez 3 échantillons contenant 10, 1000 et 100000 observations.

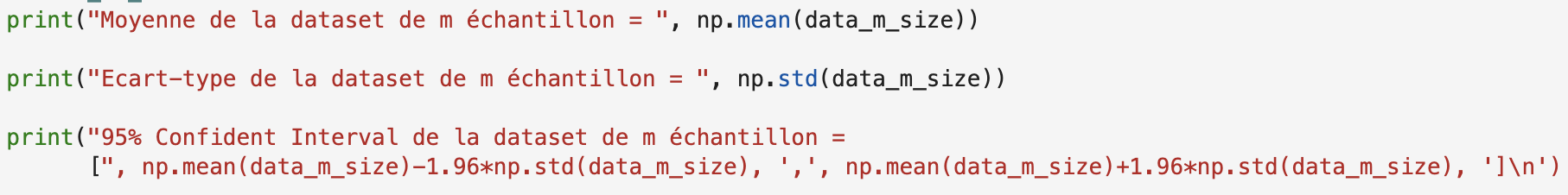
Vos 3 échantillons devront suivre une loi normale de moyenne 100 et d'écart-type 15 ( fonction np.random.normal(·)).

Nous avons généré trois échantillons de tailles différentes suivant une distribution normale avec une valeur moyenne de 100 et un écart type de 15.

### Initialisation

En nommant m la taille de l’échantillon :



Nous avons alors :

### Comparez les valeurs trouvées pour l'écart moyen et l'écart type avec les valeurs théoriques.

Les valeurs que nous obtenons sont proches des valeurs théoriques mais elles ne sont pas identiques car nous générons notre échantillon de manière aléatoire suivant la distribution normale. Nous observons qu'une taille d'échantillon plus élevée conduit à un écart moyen et standard plus proche de la valeur théorique.

### Commenter l'écart type et IC95

Nous observons que lorsque la taille des échantillons augmente, les valeurs de la moyenne et de l'écart type tend à se rapprocher de la valeur théorique attendue. Nous constatons une diminution nette de l’erreur type à mesure que la taille des échantillons augmente car elle est inversement proportionnelle à la racine carrée de la taille de l’échantillon. La conséquence est que l'intervalle de confiance devient plus mince à mesure que la taille de l'échantillon augmente.

## Utilisez la commande read\_csv(file) de la librairie pandas afin d'ouvrir le fichier malnutrition.csv



La moyenne est de 87.969 points de QI.

L’écart type est de 9.6.

## A l'aide des outils statistiques à votre disposition, que pouvez-vous conclure sur l'effet de la malnutrition sur le QI lorsque vous comparez ces résultats avec ceux de votre échantillon de 100000 individus ?

La moyenne des mesures de QI sur les personnes souffrant de malnutrition est inférieure de 12,03 points à la moyenne de l'échantillon que nous avons généré.

L'écart-type des mesures de QI chez les personnes souffrant de malnutrition est inférieur de 5,32 points à l'écart-type de l'échantillon généré.

Lorsque nous considérons l'intervalle de confiance à 95%, cette différence varie entre 10,13 et 13,93 points de QI en faveur de l'échantillon sain.