

Histogrammes

Introduction à la statistique descriptive

Lycée Pierre Corneille – MP

2016-2017

Calculs

But : décrire une série de données réelles $(x_i)_{i \in I}$

Calculs

But : décrire une série de données réelles $(x_i)_{i \in I}$

- Nombre de données : $N = \#(I)$

Calculs

But : décrire une série de données réelles $(x_i)_{i \in I}$

- Nombre de données : $N = \#(I)$
- **Étendue** : $m = \min_{i \in I} x_i$, $M = \max_{i \in I} x_i$

Calculs

But : décrire une série de données réelles $(x_i)_{i \in I}$

- Nombre de données : $N = \#(I)$
- **Étendue** : $m = \min_{i \in I} x_i$, $M = \max_{i \in I} x_i$
- Grouper les données en **classes** :

$$l_0 = [\alpha_0, \alpha_1[, \dots, l_{n-2} = [\alpha_{n-2}, \alpha_{n-1}[, l_{n-1} = [\alpha_{n-1}, \alpha_n]$$

$$\text{où } m = \alpha_0 < \alpha_1 < \dots < \alpha_n = M$$

Calculs

But : décrire une série de données réelles $(x_i)_{i \in I}$

- Nombre de données : $N = \#(I)$
- **Étendue** : $m = \min_{i \in I} x_i$, $M = \max_{i \in I} x_i$
- Grouper les données en **classes** :

$$l_0 = [\alpha_0, \alpha_1[, \dots, l_{n-2} = [\alpha_{n-2}, \alpha_{n-1}[, l_{n-1} = [\alpha_{n-1}, \alpha_n]$$

où $m = \alpha_0 < \alpha_1 < \dots < \alpha_n = M$

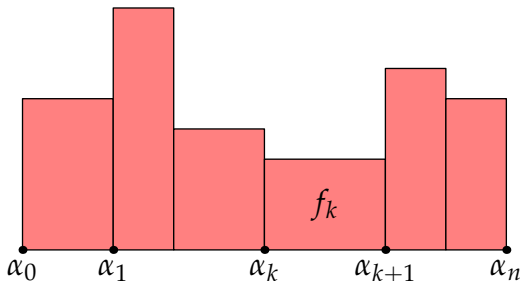
- Calculer la **fréquence** de chaque classe :

$$\forall 0 \leq k < n, \quad f_k = \frac{1}{N} \cdot \#\{i \in I : x_i \in l_k\}$$

Graphique

Chaque classe I_k est représentée par un rectangle de base $(\alpha_{k+1} - \alpha_k)$ et de hauteur h_k telle que les **aires** des rectangles soient proportionnelles aux fréquences des classes.

$$A_k = (\alpha_{k+1} - \alpha_k) \times h_k \propto f_k.$$



Fonction hist du module matplotlib

Arguments nécessaires :

- Série statistique : tableau **unidimensionnel**
- Option **normed=True**

Fonction hist du module matplotlib

Arguments nécessaires :

- Série statistique : tableau **unidimensionnel**
- Option **normed=True**

Argument facultatif : classes

- Par défaut, 10 classes de même taille

Fonction hist du module matplotlib

Arguments nécessaires :

- Série statistique : tableau **unidimensionnel**
- Option **normed=True**

Argument facultatif : classes

- Par défaut, 10 classes de même taille
- Option **bins=n** : n classes de même taille

Fonction hist du module matplotlib

Arguments nécessaires :

- Série statistique : tableau **unidimensionnel**
- Option **normed=True**

Argument facultatif : classes

- Par défaut, 10 classes de même taille
- Option **bins=n** : n classes de même taille
- Option **bins=L** où **L** est une liste strictement croissante de réels $\alpha_0 < \alpha_1 < \dots < \alpha_n$

Fonction hist du module matplotlib

Arguments nécessaires :

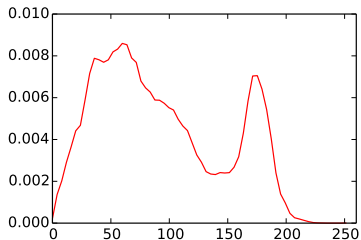
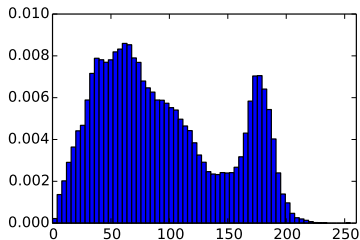
- Série statistique : tableau **unidimensionnel**
- Option **normed=True**

Argument facultatif : classes

- Par défaut, 10 classes de même taille
- Option **bins=n** : n classes de même taille
- Option **bins=L** où **L** est une liste strictement croissante de réels $\alpha_0 < \alpha_1 < \dots < \alpha_n$: classes de tailles variables.

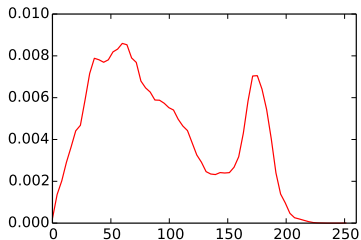
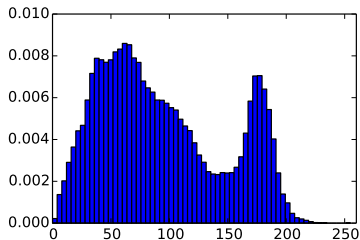
Graphique

Avec un grand nombre de classes :



Graphique

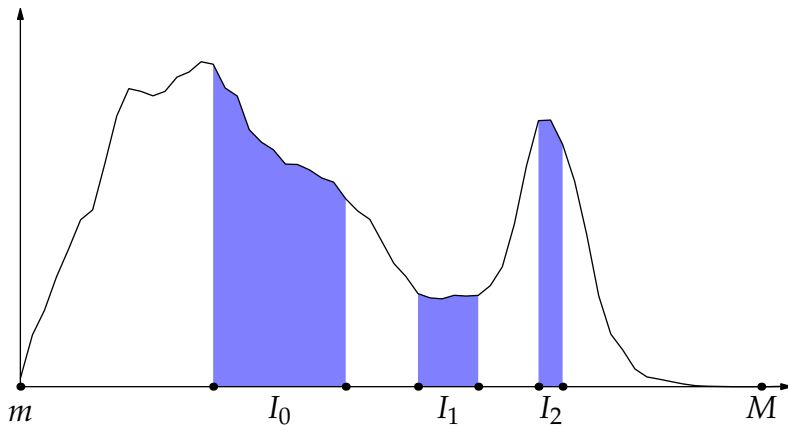
Avec un grand nombre de classes :



Notion de **densité de probabilité**

Interprétation

La fréquence d'un intervalle est estimée par un calcul d'aire.



Cas élémentaire

- Donnée

Événements $(A_n)_{n \in \mathbb{N}}$ indépendants et de même probabilité p

Cas élémentaire

- Donnée

Événements $(A_n)_{n \in \mathbb{N}}$ indépendants et de même probabilité p

- Étude statistique

Proportion d'événements réalisés

$$f_n = \frac{1}{n} \sum_{0 \leq k < n} \mathbb{1}_{A_k}$$

Cas élémentaire

- Donnée

Événements $(A_n)_{n \in \mathbb{N}}$ indépendants et de même probabilité p

- Étude statistique

Proportion d'événements réalisés

$$f_n = \frac{1}{n} \sum_{0 \leq k < n} \mathbb{1}_{A_k}$$

- Prédiction

La fréquence empirique f_n tend vers la probabilité p

Cas élémentaire

- Donnée

Événements $(A_n)_{n \in \mathbb{N}}$ indépendants et de même probabilité p

- Étude statistique

Proportion d'événements réalisés

$$f_n = \frac{1}{n} \sum_{0 \leq k < n} \mathbb{1}_{A_k}$$

- Prédiction

La **fréquence empirique** f_n tend vers la probabilité p

→ *interprétation fréquentiste*

Probabilité = **fréquence théorique**

Généralisation

- Donnée

Variables aléatoires $(X_n)_{n \in \mathbb{N}}$ indépendantes et de même loi

Généralisation

- Donnée

Variables aléatoires $(X_n)_{n \in \mathbb{N}}$ indépendantes et de même loi

- Étude statistique

Moyenne des valeurs observées

$$M_n = \frac{1}{n} \sum_{0 \leq k < n} X_k$$

Généralisation

- Donnée

Variables aléatoires $(X_n)_{n \in \mathbb{N}}$ indépendantes et de même loi

- Étude statistique

Moyenne des valeurs observées

$$M_n = \frac{1}{n} \sum_{0 \leq k < n} X_k$$

- Prédiction

La **moyenne empirique** M_n tend vers $\mathbf{E}(X_0)$

Généralisation

- Donnée

Variables aléatoires $(X_n)_{n \in \mathbb{N}}$ indépendantes et de même loi

- Étude statistique

Moyenne des valeurs observées

$$M_n = \frac{1}{n} \sum_{0 \leq k < n} X_k$$

- Prédiction

La **moyenne empirique** M_n tend vers $\mathbf{E}(X_0)$

→ Espérance = **moyenne théorique**

Exemple

Un dé à 6 faces lancé indéfiniment

Exemple

Un dé à 6 faces lancé indéfiniment

Modélisation du résultat :

- Variables indépendantes
- Loi uniforme sur $\{1, 2, 3, 4, 5, 6\}$

Exemple

Un dé à 6 faces lancé indéfiniment

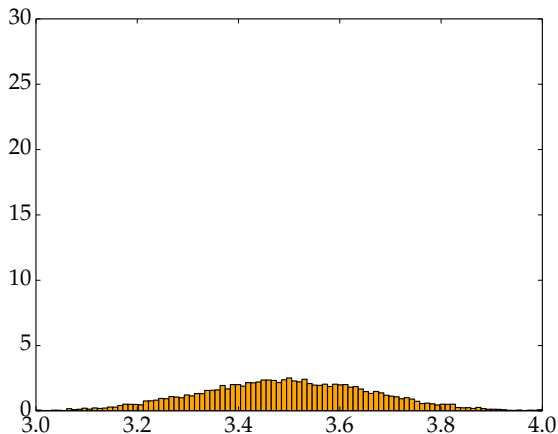
Modélisation du résultat :

- Variables indépendantes
- Loi uniforme sur $\{1, 2, 3, 4, 5, 6\}$

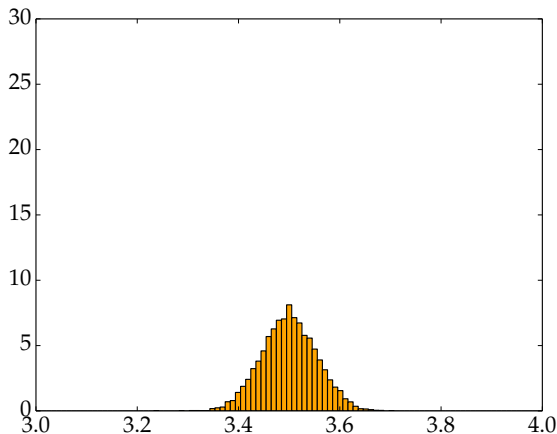
Espérance

$$\frac{1}{6} \cdot (1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2}$$

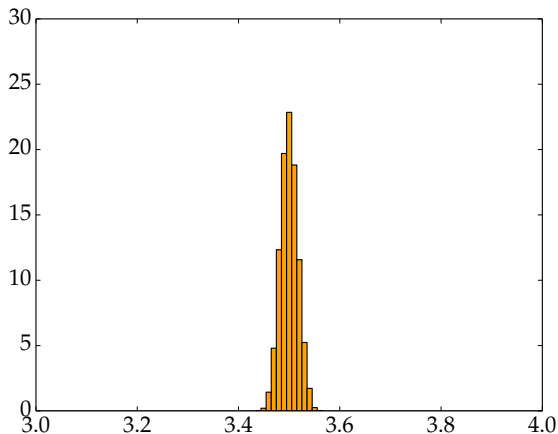
Exemple : histogramme de M_{100}



Exemple : histogramme de M_{1000}



Exemple : histogramme de M_{10000}



Hypothèses

- Données

Échantillon $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires indépendantes et de même loi

Hypothèses

- Données

Échantillon $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires indépendantes et de même loi

- Calculs

Hypothèses

- Données

Échantillon $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires indépendantes et de même loi

- Calculs

- Moyennes empiriques

$$M_n = \frac{1}{n} \sum_{0 \leq k < n} X_k$$

Hypothèses

- **Données**

Échantillon $(X_n)_{n \in \mathbb{N}}$ de variables aléatoires indépendantes et de même loi

- **Calculs**

- Moyennes empiriques

$$M_n = \frac{1}{n} \sum_{0 \leq k < n} X_k$$

- Centrées et réduites

$$M_n^* = \frac{M_n - \mathbf{E}(M_n)}{\sqrt{\mathbf{V}(M_n)}}$$

Répartition asymptotique

Pour $n \rightarrow +\infty$,

$$\forall a < b, \quad \mathbf{P}(a \leq M_n^* \leq b) \approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

Répartition asymptotique

Pour $n \rightarrow +\infty$,

$$\forall a < b, \quad \mathbf{P}(a \leq M_n^* \leq b) \approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

Indépendamment de la loi de X_0

Répartition asymptotique

Pour $n \rightarrow +\infty$,

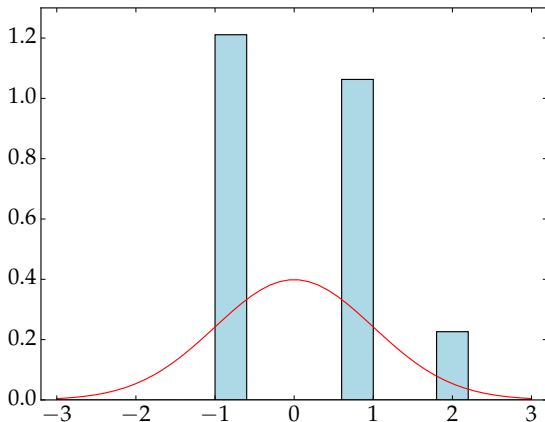
$$\forall a < b, \quad \mathbf{P}(a \leq M_n^* \leq b) \approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

Indépendamment de la loi de X_0

→ *Central limit theorem*

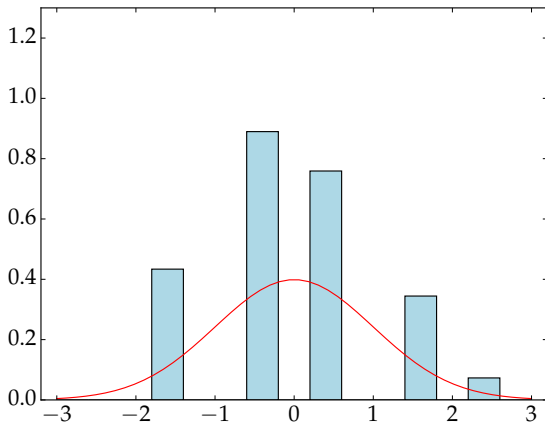
Exemple : loi de Bernoulli $\mathcal{B}(0, 3)$

$n = 2$



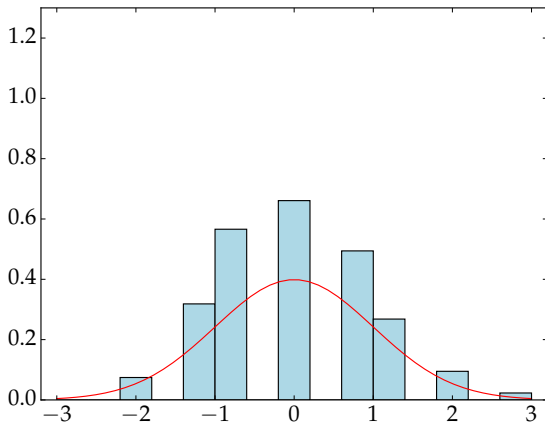
Exemple : loi de Bernoulli $\mathcal{B}(0, 3)$

$n = 5$



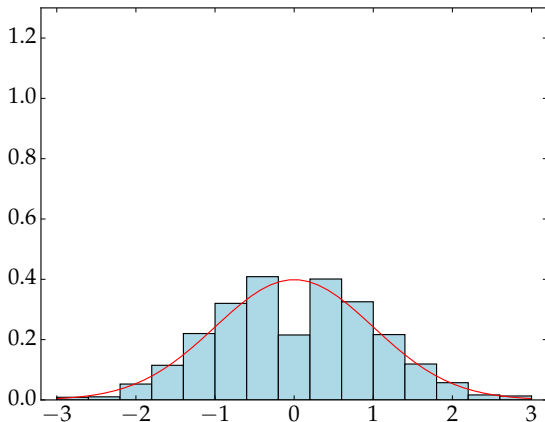
Exemple : loi de Bernoulli $\mathcal{B}(0, 3)$

$n = 10$



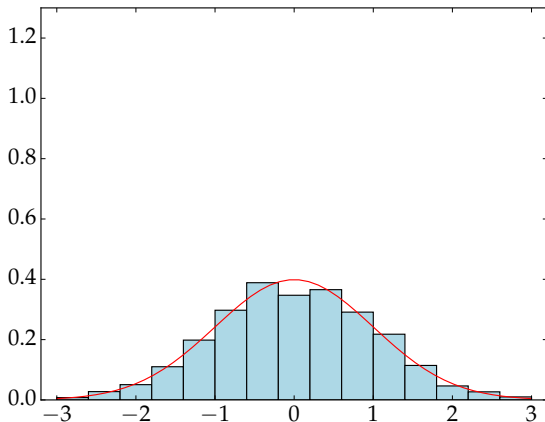
Exemple : loi de Bernoulli $\mathcal{B}(0, 3)$

$n = 100$



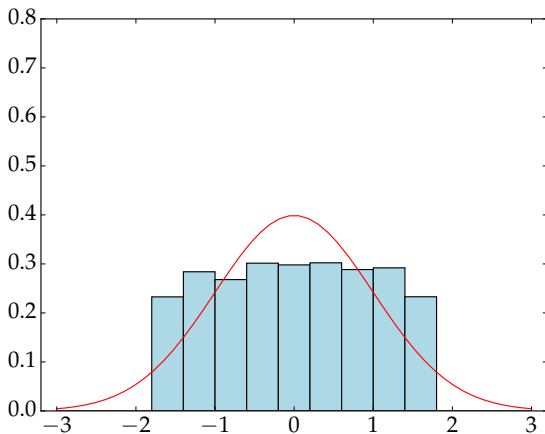
Exemple : loi de Bernoulli $\mathcal{B}(0, 3)$

$n = 1000$



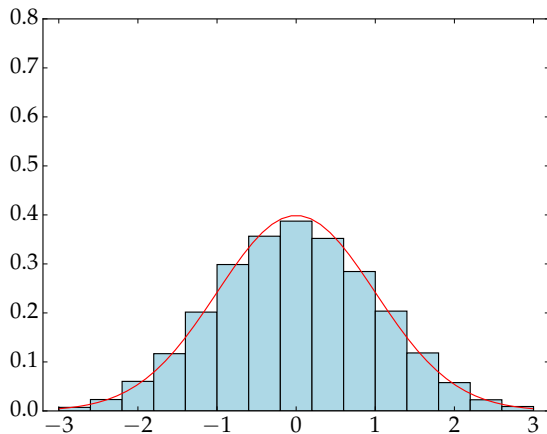
Exemple : loi uniforme sur $[0, 1]$

$n = 1$



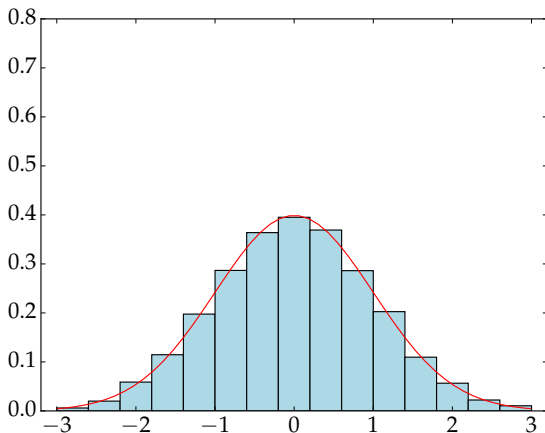
Exemple : loi uniforme sur $[0, 1]$

$n = 5$



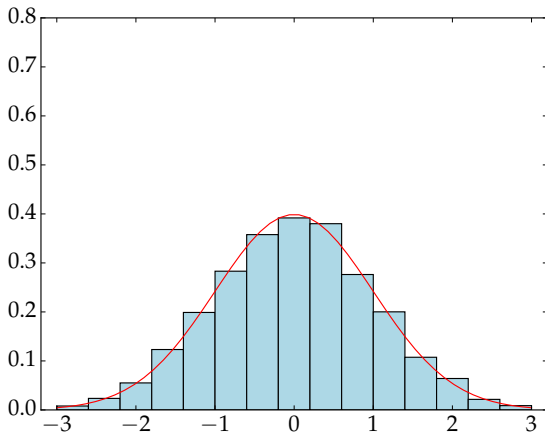
Exemple : loi uniforme sur $[0, 1]$

$n = 10$



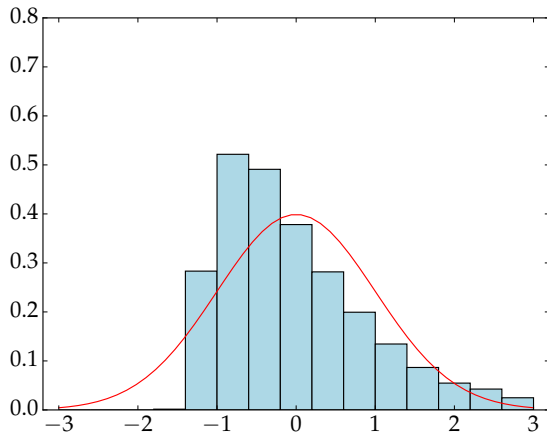
Exemple : loi uniforme sur $[0, 1]$

$n = 50$



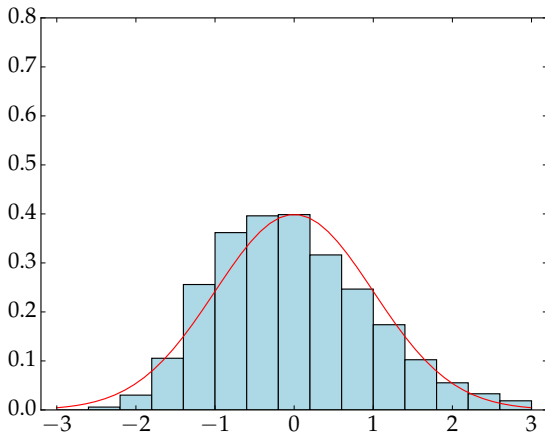
Exemple : loi exponentielle $\mathcal{E}(0,5)$

$n = 2$



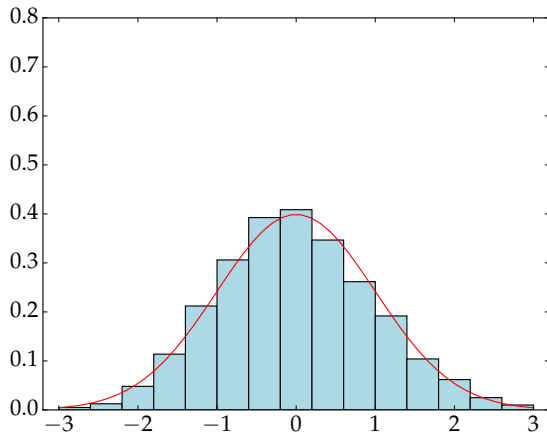
Exemple : loi exponentielle $\mathcal{E}(0,5)$

$n = 10$



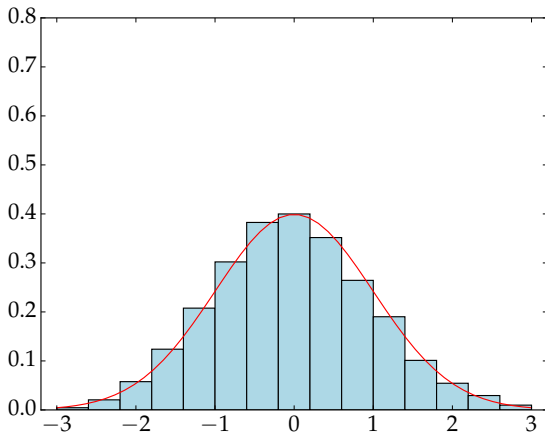
Exemple : loi exponentielle $\mathcal{E}(0,5)$

$n = 50$



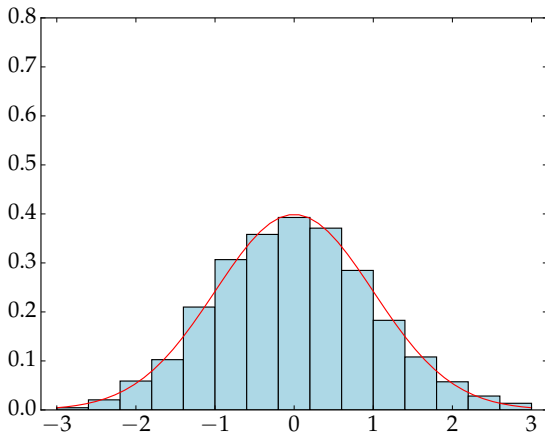
Exemple : loi exponentielle $\mathcal{E}(0,5)$

$n = 100$



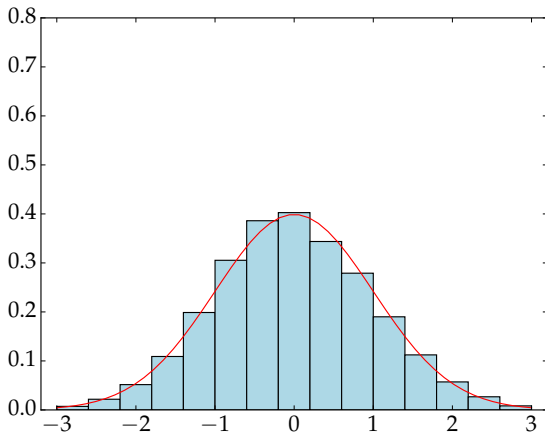
Exemple : loi exponentielle $\mathcal{E}(0,5)$

$n = 250$



Exemple : loi exponentielle $\mathcal{E}(0,5)$

$n = 500$



Exemple : loi exponentielle $\mathcal{E}(0,5)$

$n = 1000$

