

TP n° 1 : Représentation graphique et analyse de lois de variables aléatoires et d'échantillons

Exercice 1. Le but de cet exercice est de visualiser quelques lois continues importantes : les lois gaussiennes, Cauchy, et Gamma.

1. Nous souhaitons tracer les fonctions de densité et les fonctions de répartition des lois gaussiennes, Cauchy et Gamma. Pour cela, diviser la fenêtre en 6 cases (3 lignes et 2 colonnes) avec `plt.subplot`. Puis tracer (à l'aide de `plt.plot`) :
 - Dans la première ligne, la fonction de densité et la fonction de répartition de la loi gaussienne de moyenne nulle et écart-type $\sigma = 0.5, 1$ et 2 (superposer les trois tracés), sur l'intervalle $[-5, 5]$.
 - Dans la seconde ligne, la fonction de densité et la fonction de répartition de la loi de Cauchy, sur l'intervalle $[-5, 5]$.
 - Dans la troisième ligne, la fonction de densité et la fonction de répartition de la loi Gamma de premier paramètre $k = 0.5, 1$ et 2 et de deuxième paramètre 1 (superposer les trois tracés), sur l'intervalle $[0, 6]$.
2. Superposer sur l'intervalle $[-10, 10]$ les densités des lois gaussiennes standard et Cauchy. Comparer les deux : quels sont les points communs, quelles sont les différences ?
3. Superposer sur l'intervalle $[-5, 5]$ les densités des lois gaussiennes de moyenne nulle et d'écart-type $1, 0.5, 0.25, 0.1$ et 0.05 . Faire de même pour la fonction de répartition. Qu'observe-t-on quand l'écart-type tend vers zéro ?
4. *Bonus* : Superposer sur l'intervalle $[0, 6]$ les densités des lois Gamma de paramètres $\alpha = \beta = k$ pour $k \in \{1, 2, 4, 8, 16, 32, 64\}$. Qu'observe-t-on quand k tend vers l'infini ? Pour $k = 100$, superposer la densité avec la densité d'une autre loi que l'on connaît et qui y est proche.

Exercice 2. Le but de cet exercice est de visualiser quelques lois discrètes importantes : les lois de Poisson, binomiale et géométrique.

1. Comme dans l'Exercice 1.1, diviser la fenêtre en 6 cases (3 lignes et 2 colonnes) et tracer :
 - Dans la première ligne, la fonction de masse (avec `plt.stem`) et la fonction de répartition (avec `plt.step`) de la loi de Poisson de paramètre $\lambda = 0.5, 2$ et 5 (superposer les trois tracés), sur l'intervalle $[0, 10]$.
 - Dans la deuxième ligne, la fonction de masse et la fonction de répartition de la loi binomiale de paramètres $n = 20$ et $p = 0.2, 0.5, 0.8$ (superposer les trois tracés), sur l'intervalle $[0, 20]$. Utiliser `scs.binom`.
 - Dans la troisième ligne, la fonction de masse et la fonction de répartition de la loi géométrique de paramètre $p = 0.7, 0.5, 0.2$ (superposer les trois tracés), sur l'intervalle $[1, 10]$. Utiliser `scs.geom`.
2. Superposer sur l'intervalle $[0, 10]$ les fonctions de masse des lois binomiales de paramètres n et $p = 1/n$ pour $n \in \{10, 20, 40\}$. Faire de même pour la fonction de répartition. Qu'observe-t-on quand n tend vers l'infini ? Peut-on approcher la loi par une autre loi discrète que l'on connaît ?
3. Répéter la partie précédente avec $n = 10, 30, 100$ et $p = 0.5$, sur l'intervalle $[0, 100]$. Peut-on cette fois-ci approcher la loi par une loi continue ?

Exercice 3. (Analyse d'échantillon) Marie rapporte d'une randonnée en montagne un bloc de pierre contenant des substances radioactives qui émettent des rayons α . Sa mère, chimiste à l'Université Paul Sabatier, lui prête un compteur Geiger pour étudier la radioactivité de cette pierre. Marie branche alors le compteur à son ordinateur et enregistre les espaces de temps entre deux rayons α . Elle remarque que ces espaces de temps ont l'air aléatoires. Aidez-la à découvrir la loi sous-jacente.

1. Charger le contenu du fichier `TP1_Fission.txt` dans un vecteur `Fission`. Il contient les espaces de temps (en ms) entre deux rayons α . Afficher la taille de l'échantillon, moyenne, quartiles, écart interquartile, écart-type, maximum et minimum.
2. Diviser la fenêtre en deux parties et afficher dans la première un histogramme de l'échantillon avec 30 classes (avec `plt.hist`), et dans la deuxième la fonction de répartition empirique de l'échantillon (avec `plt.step`).
3. Emettre une conjecture sur la loi de l'espace de temps entre deux rayons α . Superposer la densité de cette loi à l'histogramme normalisé. Superposer la fonction de répartition de cette loi à la fonction de répartition empirique.

Exercice 4. (Bonus)

En 1968, à la station ornithologique du col de Golèze situé dans les Alpes françaises, 48 merles à plastron ont été capturés au filet pour être marqués, durant les 89 jours d'ouverture de la station. Les merles sont réputées pour vivre en solitude. On s'intéresse au nombre de merles capturés par jour. Les données sont reproduites dans la table suivante :

Nombre de merles	0	1	2	3	4	5	6
Fréquences	56	22	9	1	0	1	0

1. Représenter graphiquement ces données par un diagramme en bâtons en annotant les axes.
2. Argumenter s'il fait sens de modéliser ces données par un échantillon de loi Poisson d'un certain paramètre λ .
3. Calculer (en Python, avec les fonctions `np.dot` ou `np.average`), le nombre moyen de merles capturés par jour. Appelons ce nombre $\hat{\lambda}$.
4. Dans une nouvelle fenêtre, superposer au diagramme de la première partie la fonction de masse de la loi de Poisson de paramètre $\hat{\lambda}$, convenablement mise à échelle.