INTRODUCTION AUX VARIABLES ALÉATOIRES

L'objectif de cette manipulation de Travaux Pratiques est d'illustrer sous MATLAB les notions théoriques vues en cours sur les variables aléatoires.

I Variables aléatoires monodimensionnelles

On se propose de simuler des variables aléatoires Gaussiennes et de visualiser la répartition de leurs échantillons. Pour cela on utilise la fonction `randn`.

1. Simuler $N = 1000$ réalisations d'une variable aléatoire Gaussienne scalaire X de moyenne nulle et de variance unité, et placer le résultat dans un vecteur `x`. Afficher l'histogramme de ces réalisations en procédant comme suit † :

```
[counts,centers] = hist(x,100);
bar(centers,counts/N/(centers(2)-centers(1)));
```

   Décrire l'histogramme. Augmenter le nombre de réalisations et commenter.

2. Afin de vérifier l'adéquation avec la loi théorique, tracer sur la même figure (en utilisant `hold on`) avec une couleur différente la densité de probabilité théorique, dont on rappelle qu'elle est de la forme

$$p_{\mathsf{X}}(x) = \mathcal{G}(x; m_x, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{(x - m_x)^2}{2\sigma^2}. \tag{1}$$

   Choisir des valeurs de x allant de min(`x`) à max(`x`) avec un pas de 0.1.

Soit la variable aléatoire Gaussienne scalaire Y définie par $\mathsf{Y} = 10 + \sqrt{2}\mathsf{X}$.

3. Établir la moyenne m_y et la variance σ_y^2 de Y. En déduire une manière de générer des réalisations de la loi $\mathcal{G}(y; m_y, \sigma_y^2)$ au moyen de la commande `randn`.
4. Répéter les questions 1 et 2 pour Y. Peut-on retrouver les moments de Y à partir de son histogramme ?

II Variables aléatoires bi-dimensionnelles

On considère des variables aléatoires $\mathsf{X}_i = \binom{\mathsf{x}_{i,1}}{\mathsf{x}_{i,2}}$, $i = 1, 2, \ldots$, à valeurs dans $\mathbb{R}^2$. Leurs réalisations, générées sous MATLAB, sont stockées dans des matrices `X1`,`X2`,... dont chaque colonne correspond à un tir aléatoire.

5. Simuler des réalisations de deux variables aléatoires $\mathsf{x}_{1,1}, \mathsf{x}_{1,2}$ indépendantes, Gaussiennes de moyenne nulle et variance unitaire, et les placer dans une matrice `X1`.

†. La valeur de chaque barre de l'histogramme, une fois normalisée par N, approxime la probabilité $\mathbb{P}(a < \mathsf{x} < b) = \sum_a^b p_{\mathsf{X}}(x)\mathrm{d}x$, avec $p_{\mathsf{x}}(x)$ la densité de probabilité de x. Cette probabilité peut elle-même être approximée par $(b-a)p_{\mathsf{x}}(\frac{a+b}{2})$ si $b - a$ est petit, d'où la division par la largeur d'une barre (approximation de l'intégrale par l'aire d'un rectangle).

6. Afficher le nuage des points 2D obtenus. Commenter.
7. Répéter la même opération pour deux variables aléatoires $\mathsf{x}_{2,1}, \mathsf{x}_{2,2}$ indépendantes, où $\mathsf{x}_{2,1}$ (resp. $\mathsf{x}_{2,2}$) suit une loi Gaussienne de moyenne 10 et de variance 2 (resp. de moyenne 2 et de variance 0.2). Les réalisations seront placées dans une matrice `X2`. Commenter.
8. Former la matrice `X3` définie par les deux variables aléatoires

$$\mathsf{x}_{3,1} = \mathsf{x}_{1,1}; \qquad \mathsf{x}_{3,2} = \mathsf{x}_{1,2} + a\mathsf{x}_{1,1}; \tag{2}$$

   où a admet successivement les valeurs 1 et 5.
9. Afficher le nuage de points correspondant à `X3`. Le comparer à celui correspondant à `X1`. Commenter relativement à l'indépendance des variables aléatoires entrant en jeu.
10. Établir l'expression théorique des moments de $\mathsf{X}_3 = \binom{\mathsf{x}_{3,1}}{\mathsf{x}_{3,2}}$. Les comparer avec leurs estimés empiriques établis par MATLAB au moyen des fonctions `mean`, `cov`, etc.

III Somme de variables aléatoires

On considère K variables aléatoires scalaires $\{\mathsf{x}_k\}_{k=1,\dots,K}$, indépendantes, Gaussiennes centrées de variance unité. K pourra prendre diverses valeurs, e.g., $K = 3$, $K = 6$, etc.

11. Simuler 10 000 réalisations de ces K variables.

Soit la variable aléatoire $\mathsf{y} = \sum_{k=1}^{K} \mathsf{x}_k$.

12. Tracer l'histogramme des réalisations de y.
13. Quelle loi suit la variable y (type, moyenne, variance) ? Superposer la densité de probabilité de y avec son approximation empirique.

Soit la variable aléatoire $\mathsf{z} = \sum_{k=1}^{K} \mathsf{x}_k^2$. La fonction suivante correspond à la densité de probabilité d'une loi de χ^2 à K degrés de liberté :

$$p_{\mathsf{z}}(z) = \chi^2_K(z) = \frac{z^{\frac{K}{2}-1} \exp\left(-\frac{z}{2}\right)}{2^{\frac{K}{2}}\Gamma(\frac{K}{2})}, z \geq 0, \tag{3}$$

où Γ désigne la fonction Gamma.

12. Tracer les histogrammes des réalisations de z pour diverses valeurs de K.
13. Superposer la densité de probabilité de z avec son approximation empirique. Commenter.
14. Établir la moyenne et la variance de z en fonction de K.
15. Renouveler ces calculs pour $K = 100$. Quel type de loi obtient-on ? Quels sont ses moments ? Établir un lien avec le théorème central limite.