BE Probabilités avec MATLAB - Compléments

1 Génération d'une variable discrète

On considère une variable aléatoire discrète Y qui prend les valeurs $\{0, \pi/2, \pi, 3\pi/2\}$ avec les probabilités suivantes :

$$\begin{aligned} P[Y=0] &= 1/8 \\ P[Y=\pi/2] &= 1/3 \\ P[Y=\pi] &= 1/4 \\ P[Y=3\pi/2] &= 7/24 \end{aligned}$$

On a alors $E[Y] = \frac{41}{48}\pi$.

1. Générer N réalisations d'une variable X uniforme (continue) sur $[0, 1]$.
2. On peut alors poser :
$$\begin{aligned} &Y = 0 \text{ si } X \in [0; a] \\ &Y = \pi/2 \text{ si } X \in [a; b] \\ &Y = \pi \text{ si } X \in [b; c] \\ &Y = 3\pi/2 \text{ si } X \in [c; 1] \end{aligned}$$
   En choisissant correctement les bornes a, b, et c, générer alors N réalisations de la variable Y à partir des réalisations de X (on utilisera de préférence les opérateurs booléens appliqués vectoriellement, plutôt que des boucles).
   Rappel : sous Matlab, l'opérateur "et" s'obtient avec le caractère &.
3. Calculer à partir de ces réalisations les estimations des probabilités de Y, ainsi qu'une estimation de sa moyenne. Tester ce programme pour $N = 100, 200, ...10000$. Commenter.

2 Génération de variables discrètes

Dans un grand nombre d'applications, on peut être amené à générer des variables discrètes qui ne suivent pas une loi standard, comme la loi uniforme ou la loi de Poisson.

On cherche dans cet exercice à créer une fonction `variables_discretes` permettant de renvoyer une matrice $N \times M$ d'échantillons générés indépendamment suivant une loi définie par

$$P[X = x_k] = p_k \ , \quad k = 1, \ldots, K$$

où les valeurs $(x_k)_{k=1,\ldots,K}$ et les probabilités $(p_k)_{k=1,\ldots,K}$ sont spécifiées par l'utilisateur. Les variables d'entrée de la fonction `variables_discretes` sont donc :
— le vecteur $x = [x_1, \ldots, x_K]^T$ des valeurs prises par la variable aléatoire X ;
— le vecteur $p = [p_1, \ldots, p_K]^T$ des probabilités associées aux valeurs $(x_k)_{k=1,\ldots,K}$;
— les dimensions N et M de la matrice à générer.

Une méthode consiste à créer un vecteur $y = (y_1, y_2, \ldots, y_{NM})$ de longueur NM, où les y_i sont tirés suivant une loi uniforme sur $[0; 1]$ (variables continues). A partir du vecteur y, on génère le vecteur des réalisations $X = (X_1, X_2, \ldots, X_{NM})$ en posant

$$X_i = x_k \text{ si } y_i \in [P_{k-1}; P_k[,$$

où $P_k = \sum_{l=1}^{k} p_l$ pour $k = 1, \ldots, K$ et $P_0 = 0$.

1. Vérifier que pour tout i, on a bien $P[X_i = x_k] = p_k$ (dans le cas où on a bien $\sum_{l=1}^{K} p_l = 1$).
2. Dans le cas où $\sum_{l=1}^{K} p_l \neq 1$, c'est-à-dire que l'utilisateur n'indique pas les probabilités absolues, mais des "poids" p_k (supposés toujours positifs) aux différentes valeurs x_k, comment peut-on procéder ?
3. Créer la fonction `variables_discretes` de variables d'entrée x, p, N et M de la façon suivante :
   — Construire le vecteur $P = (P_0, P_1, \ldots, P_K)$ défini précédemment (on pourra utiliser la fonction `cumsum`).
   — Générer le vecteur y, et construire à partir de y et de P le vecteur X. Construire alors la matrice souhaitée.
   — Retourner alors cette matrice, ainsi que la moyenne et la variance de la loi discrète utilisée.
4. Créer un fichier `test.m` permettant de tester la fonction ainsi créée sur des lois discrètes standard (comme la loi uniforme, la loi binômiale, ou la loi de Poisson) en précisant la méthode ainsi que les paramètres utilisés pour valider la fonction.

3 Générateur de triplets de nombres entiers aléatoires

Un générateur de nombres aléatoires produit des séquences de triplets (indépendants) d'entiers équiprobables (i, j, k) avec $i, j, k \in \{1, \ldots, 10\}$. On veut estimer par simulations :
— la probabilité que ce générateur produise un triplet de la forme (i, i, i) ;
— la probabilité qu'au moins un triplet de la forme (i, i, i) apparaisse dans une séquence de 10 triplets ;
— la probabilité d'avoir exactement 4 triplets de la forme (i, i, i) dans une séquence de 20 triplets.

1. Générer sous Matlab de telles séquences de triplets, et estimer les probabilités ci-dessus en calculant la fréquence relative (on pourra utiliser les fonctions `unidrnd`, `find`, `mean`, `sum`). Comparer avec les probabilités théoriques (respectivement : 0.01, 0.0956, $4.125e^{-5}$). On pourra utiliser la fonction `hist` pour compter les triplets de la forme (i, i, i) dans des séries de 10 ou 20 triplets.
2. Observer la précision des probabilités estimées en fonction du nombre de tirages.

4 Loi du maximum pour une loi uniforme $\mathcal{U}([0, \theta])$

On dispose d'un échantillon $x_1, ..., x_N$ où les x_i sont des réalisations de variables aléatoires réelles $X_1, ..., X_N$ indépendantes et de même loi uniforme sur $[0, \theta]$. On s'intéresse à la loi asymptotique du maximum des X_i. Lors du premier TD de statistiques, on calculera la densité de $Z_N = \max_{1 \leqslant i \leqslant N} X_i$, son espérance et sa variance :

$$
\begin{aligned}
f(z) &= \begin{cases} N\frac{z^{N-1}}{\theta^N} \text{ pour } z \in [0,\theta] \\ 0 \text{ sinon} \end{cases}, \\
E[Z_N] &= \frac{N}{N+1}\theta, \\
var[Z_N] &= \left(\frac{N}{N+2} - \left(\frac{N}{N+1}\right)^2\right)\theta^2.
\end{aligned}
$$

1. Pour $N = 100, \theta = 10$, générer $K = 500$ réalisations du vecteur $[X_1 X_2 ... X_N]$. Ces K réalisations seront stockées dans la matrice X de taille $N \times K$. En utilisant la fonction `max` de MATLAB, générer K réalisations de la variable aléatoire Z_N.
2. Pour K variant de 500 à 1000 avec un pas de 100, estimer la moyenne et la variance de Z_N en utilisant les fonctions MATLAB `mean` et `var`. Tracer la moyenne et la variance estimées en fonction de K. Comparer graphiquement aux valeurs théoriques.