Probabilités et simulations.
Feuille de TD 4.

Exercice 1

(1) Ecrire un programme qui, en fonction de la matrice M de taille $m \times n$, et des paramètres $a_{\min} < a_{\max}, j \in \mathbb{N}^*$ renvoie un histogramme des m valeurs
$$\{\frac{1}{n}\sum_{j=1}^{n} M_{1j}, \dots \frac{1}{n}\sum_{j=1}^{n} M_{mj}\}$$
sur l'intervalle $[a_{\min}, a_{\max}]$ découpé en j intervalles de classe de même taille.

(2) Utiliser ce programme avec les matrices M suivantes, pour $n = 1, 10, 100, 1000, 10000$, et un intervalle $[a_{\min}, a_{\max}]$ ainsi qu'un nombre de classes j appropriés.
- `M=grand(1000,n,"exp",1/2)`
- `M=grand(1000,n,"gam",3,1/2)`
- `M=grand(1000,n,"nor",1,1)`
- `M=grand(1000,n,"bin",4,1/2)`
- `M=grand(1000,n,"poi",3)`
- `M=grand(1000,n,"unf",0,%pi)`

Exercice 2 On considère une variable binomiale $X \sim \mathrm{Bin}(n,p)$. Quelle(s) valeur(s) de j maximise(nt) l'expression $\mathbb{P}(X = j)$? Vérifier numériquement.
Indication : On pourra commencer par calculer $\frac{\mathbb{P}(X=k)}{\mathbb{P}(X=k-1)}$.

Exercice 3 Dans tout l'exercice, X suit une loi de Poisson(λ), et Y_n est indépendante de X et suit une loi $\mathrm{Bin}(n, p_n)$. On suppose de plus que $np_n \to \lambda > 0$ lorsque $n \to \infty$.

(1) Montrer que pour chaque k fixé on a :
$$\lim_{n\to\infty} \mathbb{P}(Y_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$
Dans la suite de l'exercice on considère le cas particulier $np_n = \lambda$ et dans les simulations on s'intéresse plus particulièrement aux valeurs particulières $\lambda = 1/4, \lambda = 1, \lambda = 3$.

(2) Sur un même graphe (on veillera à bien choisir la taille de la fenêtre), représenter les diagrammes en bâtons des distributions de X, Y_{1000}.

(3) Comment simuler le couple de variables (X, Y_n) ?

(4) Observer la valeur $|X - Y_{1000}|$ pour 1000 réalisations du couple de variables. Tracer le graphe de la fréquence de l'événement $\{|X - Y_n| \le \varepsilon\}$ en fonction de ε.
Observe-t-on que
$$Y_n \xrightarrow[n\to\infty]{(\mathbb{P})} X?$$

(5) Sur un même graphe, tracer les histogrammes de 10000 réalisations indépendantes de chacune de ces variables (toujours pour $n = 1000$, et les diffrentes valeurs de λ). Qu'observe-t-on?

Exercice 4 Comme dans l'exercice précédent on considère $X_n \sim \mathrm{Bin}(n, p_n)$, et $\lambda = np_n$.

(1) Soit $A_n = \{X_n \geq 1\}$, et soit Y une v.a. de Poisson de paramètre λ. Montrer que
$$\lim_{n\to\infty} \mathbb{P}(X_n = j|A_n) = \mathbb{P}(Y = j|Y \geq 1).$$
(2) Simuler, en fonction de n, et pour plusieurs valeurs de λ, 1000 réalisations de X_n sachant A_n, et de Y sachant $\{Y \geq 1\}$. Pour $n = 1000$, tracer les histogrammes correspondants. Qu'observe-t-on?

Exercice 5 Soit X une variable de Poisson de paramètre λ.

(1) Quelle(s) valeur(s) de j maximise(nt) $\mathbb{P}(X = j)$? Vérifier votre réponse sur le diagramme en bâtons de la loi considérée.
(2) Soit $j \in \mathbb{N}$ fixé. Quelle(s) valeur(s) de λ maximise(nt) $\mathbb{P}(X = j)$? Vérifier votre réponse sur le graphe de $\lambda \to \mathbb{P}(X = j)$.

Exercice 6 Si X prend ses valeurs dans $\mathbb{N}$, montrer que
$$\mathbb{E}(X) = \sum_{n=0}^{\infty} \mathbb{P}(X > n)$$

Exercice 7 Soit X une variable de Poisson de paramètre $\lambda > 0$.

(1) Calculer $\mathbb{E}(X)$, $\mathrm{var}(X)$.
(2) Lorsque $\lambda = 1$, calculer $\mathbb{E}[|X - 1|]$.
(3) Montrer que pour $r = 2, 3, 4, \ldots$ on a
$$\mathbb{E}(X(X-1)\ldots(X-r+1)) = \lambda^r.$$

Exercice 8 Soit X une variable aléatoire géométrique de paramètre p, $\mathbb{P}(X = k) = p(1-p)^k$, $k = 0, 1, 2, \ldots$.

(1) Représenter les diagrammes en bâtons de cette ditribution pour $p = 1/10, 1/3, 1/2, 5/6$.
(2) Ecrire un programme qui simule une variable géométrique de paramètre p.
    Note : par la suite on pourra aussi utiliser la fonction `grand((m,n,"geom", p))`.
(3) Calculer $\mathbb{E}[X]$, puis $\mathbb{E}(\frac{1}{1+X})$.

Exercice 9 Soit X une variable aléatoire géométrique de paramètre p, $\mathbb{P}(X = k) = p(1-p)^k$,$k = 0, 1, 2, \ldots$. Montrer que si $i, j > 0$, on a $\mathbb{P}(X > i + j|X > i) = \mathbb{P}(X > j)$.

Exercice 10 Soit X une variable aléatoire géométrique de paramètre p, $\mathbb{P}(X = k) = p(1-p)^k$,$k = 0, 1, 2, \ldots$. Montrer que pour $r = 2, 3, 4, \ldots$ on a
$$\mathbb{E}(X(X-1)\ldots(X-r+1)) = \frac{r!p^r}{(1-p)^r}$$

Exercice 11 Si $g : [0, \infty[\to [0, \infty[$ est strictement croissante montrer que
$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(g(|X|))}{g(a)}$$
pour $a > 0$.

Exercice 12 En tout point $t \in \mathbb{R}_+$ où $\mathbb{E}[t^X] < \infty$ on définit

$$G_X(t) = \mathbb{E}[t^X].$$

Déterminer G_X (et son intervalle de définition) lorsque

(1) $X \sim \text{Ber}(p)$.
(2) $\mathbb{P}(X = 0) = \mathbb{P}(X = 2) = 1/2$
(3) $\mathbb{P}(X = 0) = \mathbb{P}(X = 3) = 1/4, \mathbb{P}(X = 2) = 1/2$.
(4) $X \sim \text{Unif}\{1, ..., n\}$, où $n \in \mathbb{N}^*$.
(5) $X \sim \text{Bin}(n, p)$, où $n \in \mathbb{N}^*, p \in [0, 1]$.
(6) $X \sim \text{Geom}(p)$, où $p \in (0, 1)$.
(7) $X \sim \text{Poisson}(\lambda)$, où $\lambda > 0$.

Pour les variables des questions 1 à 7, que dire de la fonction ϕ_X définie par $\phi_X(u) := \mathbb{E}[\exp(uX)]$, en tout point $u \in \mathbb{R}$ tel que $\mathbb{E}[\exp(uX)] < \infty$?

Exercice 13 Soit $(X_1, ..., X_k) \sim \text{Multi}(n, p_1, ..., p_k)$, avec k et n des entiers strictement positifs et $p_i, 1 \leq i \leq k$ des réels de $[0, 1]$ tels que $\sum_{i=1}^k p_i = 1$. On définit, pour des réels $t_1, ..., t_k$ tous positifs,

$$G(t_1, ..., t_k) = \mathbb{E}\left[t_1^{X_1} ... t_k^{X_k}\right].$$

Montrer que G est bien définie, puis la calculer.

Exercice 14 Soient $\lambda_1, ..., \lambda_n$ des réels strictement positifs et $X_1, ..., X_n$ des variables indépendantes, avec $X_i \sim \text{Poisson}(\lambda_i)$.

On note $S = \sum_{i=1}^n X_i$. Calculer $G_S(t) = \mathbb{E}[t^S]$ pour $t \geq 0$ et en déduire la loi de S.

Exercice 15 Soit une variable $X \in \mathbb{N}$ d'espérance m et de variance σ^2.

(1) On suppose que la fonction génératrice G_X de la variable X est définie dans un voisinage du point 1.
Exprimer le DL de G_X à l'ordre 2 au point 1.
(2) On suppose que la fontion $\phi_X : u \to \mathbb{E}[\exp(uX)]$ est définie dans un voisinage de 0.
Exprimer le DL de ϕ_X à l'ordre 2, en l'origine. Que dire du DL de ϕ_X à l'ordre $n \geq 1$, toujours en 0 ?

Exercice 16 Soit X une v.a.r. et $\phi : \mathbb{R} \to \mathbb{R}_+$ telles que $\mathbb{E}[\phi(X)] < \infty$. Pour $A \in \mathcal{B}(\mathbb{R})$, on note $i_A := \inf\{\phi(x), x \in A\}$. Montrer que

$$i_A \mathbb{P}(X \in A) \leq \mathbb{E}[\phi(X)\mathbf{1}_A] \leq \mathbb{E}[\phi(X)].$$

En déduire que lorsque $i_A > 0$ on trouve finalement que

$$\mathbb{P}(X \in A) \leq i_A^{-1}\mathbb{E}[\phi(X)].$$

Dans la suite on fixe deux réels a, m et on considère $A = \{x : |x - m| \geq a\}$. Qu'obtient-on dans les cas particuliers qui suivent? Que dire de la valeur particulière $m = \mathbb{E}[X]$?

(1) $\phi(x) = |x - m|$.
(2) $\phi(x) = (x - m)^2$.
(3) $\phi(x) = |x - m|^p$ où $p \geq 1$.
(4) $\phi(x) = \exp(\lambda|x - m|)$, où $\lambda \in \mathbb{R}_+^*$.

Exercice 17 Dans les simulations des feuilles précédentes on a souvent utilisé une fréquence empirique pour estimer la probabilité d'un événement. On note $F(n)$ la fréquence empirique d'un événement A pour n réalisations indépendantes d'une certaine expérience.

(1) A l'aide de l'inégalité obtenue dans la question 2 de l'exercice précédent, majorer $\mathbb{P}(|F(n) - \mathbb{P}(A)| > \varepsilon)$ en fonction de n, ε
(2) On fixe $\varepsilon = 0.01$. Comment faut-il choisir n_ε pour que $\mathbb{P}(|F(n_\varepsilon) - \mathbb{P}(A)| > \varepsilon) \leq 0.05$?
(3) On conduit une simulation de n_ε réalisations indépendantes de notre expérience qui renvoie la valeur de la fréquence empirique. Au regard de la question précédente, que peut-on dire de la valeur obtenue lors de cette simulation ?

Exercice 18

Expliquer tout d'abord comment estimer l'espérance d'une variable aléatoire X dont on sait simuler n réalisations indépendantes Dans la suite on notera E_n l'estimation de l'espérance obtenue pour une simulation faisant usage de n réalisations indépendantes, et on fera usage de l'exercice 16.

(1) On suppose dans cette question que $\mathrm{var}[X] = A$. Comment peut-on choisir n_ε pour que $\mathbb{P}(|E_{n_\varepsilon} - \mathbb{E}(X)| > \varepsilon) \leq 0.05$?
(2) (*) On suppose dans cette question que $\mathbb{E}[|X - \mathbb{E}[X]|^4] = A$. Montrer que pour une constante $C > 0$ que l'on déterminera
$$\mathbb{E}[(E_n - \mathbb{E}[X])^4] \leq CAn^{-2}.$$
Comment peut-on alors choisir n_ε pour que $\mathbb{P}(|E_{n_\varepsilon} - \mathbb{E}(X)| > \varepsilon) \leq 0.05$?
(3) (**) On suppose dans cette question que $\mathbb{E}[X] = m$ avec $m \in \mathbb{R}$, et que
$$\Phi(\lambda) := \mathbb{E}[\exp(\lambda X)]$$
est bien définie sur un intervalle $[-a, a]$, avec $a > 0$. Rappeler pourquoi on a
$$\ln(\Phi(\lambda)) - \lambda m = \frac{\lambda^2}{2}\mathrm{Var}(X) + o(\lambda^2).$$
Etablir, en utilisant l'exercice 16, que pour tout $\lambda \in [-a, a]$,
$$\mathbb{P}(|E_n - m| > \varepsilon) \leq \exp(-n\lambda\varepsilon)\exp(n(\ln(\Phi(\lambda)) - \lambda m)).$$
Sans chercher le calculer explicitement, expliquer pourquoi, lorsque $\epsilon << 1$, ceci permet d'amliorer les n_ε trouvés aux deux premières questions.
Discuter plus précisément la valeur de n_ε dans le cas où $X \sim \mathrm{Poisson}(\mu)$, avec $\mu > 0$.