# TD 1 : Rappels de probabilités

**Exercice 1**
Soient $\theta > 0$ et $X \sim \mathcal{P}(\theta)$. Calculer l'espérance et la variance de $X$.

**Exercice 2**
On lance $n$ fois un dé équilibré, et on note $X$ la variable aléatoire qui compte le nombre de "1" obtenus.

1. Quelle est la loi de $X$ ?
2. Calculer l'espérance et la variance de $X$.
3. Soit $Y$ une variable aléatoire indépendante et de même loi que $X$. Quelle expérience aléatoire la variable aléatoire $X + Y$ pourrait-elle modéliser ? Quelle est la loi de $X + Y$ ?

**Exercice 3**
Un lac contient une proportion $\theta$ de poissons rouges. L'expérience consiste à pécher des poissons au moyen d'une cane à pêche, jusqu'à ce qu'un poisson rouge ait été capturé. On note $X$ la variable aléatoire qui modélise le nombre total de poissons ainsi péchés.

1. Quelle est la loi de $X$ ?
2. Calculer l'espérance et la variance de $X$.
3. Soit $Y$ une variable aléatoire indépendante et de même loi que $X$.
   (a) Décrire une expérience pour laquelle l'évaluation de $\mathbb{P}(X < Y)$ aurait un intérêt. Que vaut $\mathbb{P}(X < Y)$ ?
   (b) Quelle expérience aléatoire la variable aléatoire $X + Y$ pourrait-elle modéliser ? Quelle est la loi de $X + Y$ ?

**Exercice 4**
1. Soit $\theta > 0$ et pour tout $n \geq 1$, $X_n \sim \mathcal{B}(n, \theta/n)$. Montrer que
$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = e^{-\theta}\frac{\theta^k}{k!},\ \forall k \in \mathbb{N}.$$
2. La probabilité de gagner au tiercé est de $10^{-3}$. Déduire de la question précédente une approximation de la probabilité de gagner 2 fois en jouant 2000 fois.

**Exercice 5**
1. Soient $\theta \in ]0, 1[$ et $X_1, \cdots, X_n$ des variables aléatoires indépendantes et de même loi $\mathcal{B}(\theta)$. On note $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. En utilisant l'inégalité de Bienaymé-Tchebytchev, montrer que pour tout $a > 0$,
$$\mathbb{P}(\theta \in ]\bar{X}_n - a, \bar{X}_n + a[) \geq 1 - \frac{1}{4na^2}.$$
2. On réalise 10000 lancers d'une pièce de monnaie. Parmis ces lancers, 20% sont des piles. Donner un intervalle *raisonnablement fiable* pour la probabilité que la pièce tombe sur pile (un intervalle plus précis que $[0, 1]$ !).

3. Reprendre les questions précédentes, mais en utilisant cette fois l'inégalité de Hoeffding : si $Z_1, \cdots, Z_n$ sont des variables aléatoires indépendantes telles que $\mathbb{P}(\alpha \leq Z_i \leq \beta) = 1$ pour $\alpha, \beta \in \mathbb{R}$, alors pour tout $a > 0$,
$$\mathbb{P}\Big(\Big|\sum_{i=1}^{n} Z_i - \mathbb{E}(\sum_{i=1}^{n} Z_i)\Big| \geq a\Big) \leq 2\exp\Big(-\frac{2a^2}{n(\beta-\alpha)^2}\Big).$$
Comparer l'intervalle ainsi obtenu avec celui de la question précédente.

---

## TD 2 : Modélisation statistique

---

**Exercice 1**
Merlin, qui travaille dans un magasin de bricolage, vient de recevoir une importante commande de vis. Pour estimer la proportion de vis défectueuses, il effectue $n$ expériences indépendantes. Au cours de chaque expérience, il tire des vis avec remise, jusqu'à ce qu'il en tire une défectueuse. Elle est alors remise dans le lot, et Merlin passe à l'expérience suivante.

1. Déterminer le modèle statistique associé à cette expérience.
2. Donner un estimateur de la proportion de vis défectueuses.

**Exercice 2**
Armelle et Bob jouent à pile ou face, chacun avec sa propre pièce. Les pièces ont des probabilités inconnues de donner pile. Simultanément, ils lancent $n$ fois leurs pièces de manière indépendante : à l'issue de chaque lancer, le joueur qui a obtenu pile est déclaré vainqueur si son adversaire a tiré face, les autres cas ne permettant pas de déclarer un vainqueur.

1. Déterminer le modèle statistique associé à cette expérience aléatoire, les deux paramètres de ce modèle étant les probabilités que les pièces donnent pile.
2. Donner un estimateur de la probabilité qu'un joueur soit déclaré vainqueur à l'issue d'un lancer de pièce d'Armelle et Bob.

**Exercice 3**
Un lac contient une proportion inconnue de poissons rouges. Muni d'une cane à pêche, on réalise $n$ expériences de tirages de poissons dans le lac.

1. Considérons le cas où chacune de ces expériences consiste à tirer avec remise un poisson dans le lac.
   (a) Quel est le modèle statistique associé à cette expérience aléatoire ?
   (b) Construire un estimateur de la proportion de poissons rouges.
2. Considérons maintenant que chacune des $n$ expériences consiste à tirer avec remise des poissons jusqu'à l'obtention d'un poisson rouge.
   (a) Quel est le modèle statistique associé à cette expérience aléatoire ?
   (b) Construire un estimateur de la proportion de poissons rouges.

**Exercice 4**
On dispose d'un sac rempli de boules numérotées de 1 à $K^\star$. Mais $K^\star$ est inconnu, on cherche donc à construire une expérience visant à le déterminer. Pour cela, on répète $n$ fois l'opération suivante : on tire une boule dans le sac, on note son numéro, puis on la remet dans le sac.

1. Quel est le modèle statistique associé à cette expérience aléatoire ?
2. Trouver deux estimateurs pour le nombre de boules, l'un construit avec la moyenne empirique, l'autre avec le maximum des observations.

---

## TD 3 : Erreurs d'estimation

---

**Exercice 1**
Reprenons le cas de Merlin, qui travaille dans un magasin de bricolage. Il vient de recevoir une importante commande de vis, et pour estimer la proportion de vis défectueuses, il effectue $n$ expériences indépendantes. Au cours de chaque expérience, il tire des vis avec remise, jusqu'à ce qu'il en tire une défectueuse. Elle est alors remise dans le lot, et Merlin passe à l'expérience suivante. Il est familier de l'entreprise qui fabrique ces vis, ce qui lui permet d'avoir l'ordre de grandeur $[0.1, 0.5]$ pour la proportion de vis défectueuses. Un $n$-échantillon associé à cette expérience aléatoire est donc $X_1, \cdots, X_n$ i.i.d. de loi $\mathcal{G}(\theta)$, avec $\theta \in [0.1, 0.5]$.

1. Le paramètre d'intérêt est ici l'inverse de la proportion de vis défectueuses.
   (a) Construire un estimateur sans biais du paramètre d'intérêt.
   (b) Calculer le risque quadratique, et le majorer indépendamment de $\theta$.
   (c) Déterminer un intervalle de confiance par excès au niveau de confiance 95% pour le paramètre d'intérêt.
2. Quel intervalle de confiance peut-on en déduire pour $\theta$ ?

**Exercice 2**
Soit $X_1, \cdots, X_n$ un $n$-échantillon i.i.d. de loi $\mathcal{L}_\theta$, $\theta \in \Theta$.

1. On note respectivement $m(\theta)$ et $\sigma(\theta)^2$ la moyenne et la variance de la loi $\mathcal{L}_\theta$.
   (a) Montrer que $\bar{X}_n$ estime le paramètre d'intérêt $m(\theta)$ sans biais.
   (b) Montrer que
   $$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$
   estime le paramètre d'intérêt $\sigma(\theta)^2$ sans biais.
2. On suppose ici que $\mathcal{L}_\theta = \mathcal{B}(\theta)$ avec $\theta \in ]0,1[$, et que le paramètre d'intérêt est $\theta$. On fixe $m < n$, et on considère les estimateurs suivants :
   $$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i,\ \hat{\theta}_2 = \frac{1}{m}\sum_{i=1}^{m} X_i \text{ et } \hat{\theta}_3 = \frac{\hat{\theta}_1 + \hat{\theta}_2}{2}.$$
   (a) Quel est l'estimateur préférable entre $\hat{\theta}_1$, $\hat{\theta}_2$ et $\hat{\theta}_3$ ?
   (b) Quel est l'estimateur sans biais préférable parmi tous les estimateurs de la forme $a\hat{\theta}_1 + b\hat{\theta}_2$, $a, b \in \mathbb{R}$ ?
3. On suppose ici que $\mathcal{L}_\theta = \mathcal{P}(\theta)$, avec $\theta > 0$, et que le paramètre d'intérêt est $\theta$.
   (a) Montrer que les estimateurs $\bar{X}_n$ et $\hat{\sigma}^2$ sont sans biais.

(b) Lequel de ces deux estimateurs est préférable ?

**Exercice 3**
Reprenons le jeu d'Armelle et Bob, qui jouent à pile ou face, chacun avec sa propre pièce. Les pièces ont des probabilités inconnues de donner pile. Simultanément, ils lancent $n$ fois leurs pièces de manière indépendante : à l'issue de chaque lancer, le joueur qui a obtenu pile est déclaré vainqueur si son adversaire a tiré face, les autres cas ne permettant pas de déclarer un vainqueur.

1. Construire un intervalle de confiance par excès au niveau de confiance 95% pour la probabilité qu'une seule des deux pièces donne pile.

2. Sur les $n = 500$ parties effectuées par Alex et Bob, 350 d'entre elles ont vu un vainqueur.

   (a) Soit $f$ la fonction sur $[0,1]^2$ telle que $f(x,y) = x(1-y) + y(1-x)$. Montrer que $f(x,y) > \frac{1}{2}$ implique $x \neq y$ (penser à utiliser la contraposée).

   (b) Déduire des questions précédentes que, avec un niveau de confiance supérieur à 95%, les pièces d'Armelle et Bob n'ont pas la même probabilité de donner pile.

**Exercice 4**
Selon la sécurité routière, le nombre d'accidents survenus chaque jour sur une portion d'autoroute suit une loi de Poisson, dont le paramètre est inconnu. On observe chaque semaine, pendant $n$ semaines, le nombre d'accidents de la portion. Dans la suite, le paramètre d'intérêt est la probabilité qu'aucun accident n'ait eu lieu pendant une semaine.

1. Construire le modèle statistique associé à cette expérience aléatoire.

2. Quel est l'estimateur naturel pour le paramètre d'intérêt ? Est-il sans biais ? Calculer son risque quadratique.

3. Construire un intervalle de confiance par excès de niveau au moins $1-\alpha$, $\alpha \in ]0,1[$.

4. L'objet de cette question est de trouver un autre estimateur pour le paramètre d'intérêt.

   (a) Montrer par récurrence que la somme de $n$ variables aléatoires indépendantes de même loi de Poisson de paramètre $\lambda$ suit une loi de Poisson de paramètre $n\lambda$.

   (b) Si $(X_1, \cdots, X_n)$ est un échantillon du modèle statistique, calculer $\mathbb{E}(e^{-\bar{X}_n})$.

   (c) Quel autre estimateur peut-on proposer ? Pourquoi ?

**Exercice 5**
Soit $X_1, \cdots, X_n$ un $n$-échantillon i.i.d. de loi $\mathcal{U}(\{1, \cdots, k\})$, $k \in \mathbb{N}^\star$, avec le paramètre d'intérêt $k \in \mathbb{N}^\star$. On introduit les estimateurs :

$$\hat{k}_1 = 2\bar{X}_n - 1 \text{ et } \hat{k}_2 = \max_{i=1,\cdots,n} X_i.$$

1. Montrer que $\hat{k}_1$ est sans biais, et calculer son risque quadratique.

2. Calculer avec l'estimateur $\hat{k}_1$ un intervalle de confiance par excès pour $k$ au niveau de confiance 95%.

3. Soit $t \in \mathbb{R}$. Calculer $\mathbb{P}(X_1 \leq t)$. En déduire $\mathbb{P}(\hat{k}_2 \leq t)$.

4. On rappelle que pour toute variable aléatoire positive $Z$, on a $\mathbb{E}(Z) = \int_0^{+\infty} \mathbb{P}(Z > t)\mathrm{d}t$. Utiliser cette formule pour calculer le biais de $\hat{k}_2$. Commenter le résultat.

5. Construire avec l'estimateur $\hat{k}_2$ un intervalle de confiance au niveau de confiance 95%.

6. Comparer les deux intervalles de confiance ainsi obtenus.