Variables aléatoires finies

1 Variables aléatoires

Les variables aléatoires sont aux probabilités ce que les fonctions sont à l'Analyse.

Définition 1 (Variable aléatoire)
Soit Ω un ensemble fini muni d'une loi de probabilité P. X est une variable aléatoire sur Ω si X est une application de Ω dans $\mathbb{R}$. Son image $X(\Omega) = \{x_1, \cdots, x_m\}$ est munie naturellement d'une loi P_X déterminée par la probabilité des événements élémentaires de l'univers image : $P_X(\{x_i\}) := P(X = x_i) = P(X^{-1}(\{x_i\}))$.

On peut considérer aussi que la loi de X est en fait une loi de probabilité sur $\mathbb{R}$. Ce qui est troublant dans cette définition c'est que X n'est pas la variable x de l'analyse mais une application :

$$X : \begin{array}{ccc} \Omega & \to & \mathbb{R} \\ \omega & \mapsto & X(\omega) \end{array}$$

Cependant, en probabilité, on ne s'intéresse guère à connaître précisément le lien qu'il y a entre l'antécédent et l'image, c'est pour cela que l'on voit trés rarement en probabilité l'écriture explicite de l'image de l'antécédent, mais l'on s'intéresse à leur loi, ce qui est beaucoup moins demandé. En effet, pour connaître la loi d'une variable aléatoire il suffit de connatre son image et les "poids" associés aux éléments de son image.

Exercice : Si $\text{card}(\Omega) = N$ combien y-a-t'il de variables aléatoires qui ont la loi uniforme sur $\{1, \cdots, N\}$?

Ainsi, avoir la mme loi consitue une relation d'équivalence sur les applications de Ω dans $\mathbb{R}$.

Un exemple trés important de variable aléatoire est le suivant :

Définition 2 (Loi de Bernoulli)
Si X est une variable aléatoire qui ne prend que deux valeurs 0 ou 1, on dit que X suit une loi de Bernoulli de paramètre $p := P(X = 1)$: $X \sim \mathcal{B}(p)$.

Ainsi une variable de Bernoulli correspond à la notion de fonction indicatrice en Analyse. Elle est trés importante pour compter des évènements. Tout candidat au CAPES se doit de savoir simuler une variable de Bernoulli à l'aide de sa calculatrice.

En géométrie, avec des nombres et des poids on utilise des barycentres, on fait de mme en Mécanique, en Analyse, en Statistiques. En probabilité on calcule des espérances.

2 Moyenne ou Espérance d'une variable aléatoire

Définition 3 (Espérance)
Soit Ω un ensemble fini muni d'une loi de probabilité P.
X est une variable aléatoire sur Ω, $X(\Omega) = \{x_1, \cdots, x_m\}$, on appelle l'espérance de X

$$E(X) := \sum_{i=1}^{m} x_i P(X = x_i).$$

Remarquez que l'espérance ne dépend que de la loi. Par exemple, si $X \sim \mathcal{B}(p)$ on obtient $E(X) = p$. Ce résultat se comprendra mieux à l'aide de la loi des grand nombres.

Le mot espérance nous arrive de l'histoire des Probabilités. Il s'agit de connatre le gain d'un joueur à un jeu de hasard. Comme ce gain ou perte varie en fonction de ses résultats le jeu, on s'intéresse à un nombre qui représentera ce que le joueur peut espérer gagner : son *espérance de gain*. Ce terme "espérance" est resté mais il s'applique désormais à tout les domaines : médecine, espérance de vie, finance, assurances, nombres de reçus au CAPES de Mathématiques, ...

Pour illustrer cette notion, on propose l'exercice suivant :

2.1 Payer ou ne pas payer ?

On s'intéresse à différentes stratégies d'achat de ticket d'un usager des transports en commun. Il faut savoir qu'à Nice :
- un ticket de Bus cote 1 Euro,
- une personne qui n'achète pas son ticket et qui est controlée paye une amende de 10 Euros,
- seulement 10% des usagers sont controlés.

On étudie trois comportements différents d'usagers :

1. la stratégie honnte : j'achète toujours mon ticket.
2. la stratégie voleuse : je n'achète jamais mon ticket.
3. la stratégie aléatoire : j'achète mon ticket au hasard, une fois sur deux en moyenne.

Cherchez à répondre aux questions suivantes :

1. Quelle est la stratégie la plus économique sur l'année de préparation au CAPES ?
2. Que proposez vous à la société de transport en commun pour augmenter son chiffre d'affaire sans léser les honntes gens ?

Exercice : Démontrez le lemme suivant

Lemme 1 (Espérance, intégrale de Lebesgue et intégrale de Riemann)
Soit Ω un ensemble fini muni d'une loi de probabilité P.
X est une variable aléatoire sur Ω, $X(\Omega) = \{x_1, \cdots, x_m\}$, on appelle l'espérance de X

$$E(X) := \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}).$$

Ce lemme élémentaire mais profond permet de démontrer simplement les deux résultats suivants :

– **linéarité de l'espérance :** pour tout $\lambda \in \mathbb{R}$, et pour tout X, Y, deux variables aléatoires sur (Ω, P),

$$E(\lambda X) = \lambda E(X), \qquad E(X+Y) = E(X) + E(Y)$$

– **formule de tranfert :** pour tout $\varphi : \mathbb{R} \to \mathbb{R}$ et pour toute variable aléatoire X sur (Ω, P),

$$E(\varphi(X)) = \sum_{i=1}^{m} \varphi(x_i) P(X = x_i).$$

3 Variance, écart type

Comme en statistique on définit la variance comme étant l'écart quadratique moyen à la moyenne :

Définition 4 (Variance, écart type)
Soit Ω un ensemble fini muni d'une loi de probabilité P.
X est une variable aléatoire sur Ω, $X(\Omega) = \{x_1, \cdots, x_m\}$.
On appelle variance de X notée $Var(X)$ et écart-type de X noté $\sigma(X)$:

$$Var(X) := E\left((X - E(X))^2\right), \qquad \sigma(X) := \sqrt{Var(X)}.$$

Donnons tout de suite quelques propriétés dont les démonstrations sont laissées en exercices.

Propriétés 1 (Variance, écart type) *Soit X une variable aléatoire alors on a :*

1. $Var(X) \geq 0$,
2. $Var(X) = 0$ *si et seulement si* $P(X = E(X)) = 1$,
3. $Var(X) = E(X^2) - E(X)^2$,
4. $Var(aX + b) = a^2 Var(X)$, $\sigma(aX + b) = |a|\sigma(X)$.

Exemples :
Si $X \sim \mathcal{B}(p)$ alors $Var(X) = p(1-p)$ et $\sigma(X) = \sqrt{p(1-p)}$.
Si $X \sim \mathcal{B}(1/2)$ alors $\sigma(X) = 1/2$. Dans ce cas $\sigma(X)$ représente exactement l'écart de X à $E(X)$.
De mme, si X ne prend que deux valeurs $\pm a$ avec la mme probabilité alors $E(X) = 0$ et $\sigma(X) = |a|$.
Ainsi, on voit déjà que l'écart type mesure la dispersion autour de la moyenne.

Cependant vérifier que l'on a toujours : $\max_{\Omega} |X - E(X)| \geq \sigma(X)$.
On a en fait le résultat général suivant :

Proposition 1 (Inégalité de Bienaymé-Tchebicheff) *Soit X une variable aléatoire d'espérance μ et d'écart type $\sigma > 0$. Alors, pour tout $\alpha > 0$ on a :*

$$P(|X - \mu| \geq \alpha\sigma) \leq \frac{1}{\alpha^2}.$$

La démonstration est élémentaire, il suffit de minorer la variance. On note $p_i := P(X = x_i)$.

$$\sigma^2 = \sum_{i=1}^{m}(x_i - \mu)^2 p_i \geq \sum_{|x_i-\mu|\geq\alpha\sigma}(x_i - \mu)^2 p_i = \alpha^2\sigma^2 \sum_{|x_i-\mu|\geq\alpha\sigma} p_i = \alpha^2\sigma^2 P(|X - \mu| \geq \alpha\sigma)$$

En simplifiant par la variance, on obtient le résultat demandé.

La première conséquence de cette inégalité et qu'elle précise mieux le fait que l'écart type est une mesure de dispersion autour de la moyenne. Notons les deux exemples suivant à méditer :

$$\begin{aligned} P(X \in]\mu - 2\sigma, \mu + 2\sigma[) &\geq 75\% \\ P(X \in]\mu - 10\sigma, \mu + 10\sigma[) &\geq 99\%. \end{aligned}$$

Ainsi, plus l'écart-type est petit, plus la variable aléatoire est centré autour de sa moyenne. Mutatis mutandis, on a le mme résultat pour une série statistique.

4 Variance d'une somme

Cette partie abstraite va nous permettre d'arriver rapidement à avoir les outils nécessaire pour démontrer la loi des grands nombres. Les démonstrations des énoncés suivants constituent de bons exercices calculatoires de manipulation de la variance.

On note, comme en statistique, $cov(X, Y) := E((X-E(X))\times(Y-E(Y))) = E(XY)-E(X)E(Y)$. Ainsi, on remarque que $Var(X) = cov(X, X)$ et surtout que :

$$Var(X + Y) = Var(X) + Var(Y) + 2cov(X, Y).$$

On reconnat une formule bien connu en géométrie sur la norme euclidienne au carré de la somme de deux vecteurs. On obtient l'analogue de la formule de Pythagore si $cov(X, Y) = 0$. En probabilité, au lieu de parler d'orthogonalité, on dit plutt :

Définition 5 (Variables décorrélées)
Deux variables aléatoires sont dites decorrélées si leur covariance est nulle.

Ainsi la variance de la somme de deux variables aléatoires décorrelées est égale à la somme des variances.

Pour reconnatre la décorrélation de deux variables on utilisera souvent la notion d'indépendance.

Définition 6 (Variables indépendantes)
Deux variables aléatoires X et Y, définies sur le mme univers, sont dites indépendantes si leurs événements élémentaires le sont :
$\forall a, b \in \mathbb{R}, \qquad P((X = a) \text{ et } (Y = b)) = P(X = a) \times P(Y = b).$

Proposition 2 (Variables indépendantes) *Soient deux variables aléatoires X et Y indépendantes, φ et ψ deux fonctions de $\mathbb{R}$ dans $\mathbb{R}$. Alors on a :*
- $E(XY) = E(X)E(Y)$,
- $cov(XY) = 0$,
- $Var(X+Y) = Var(X) + Var(Y)$, $\sigma(X+Y) = \sqrt{\sigma(X)^2 + \sigma(Y)^2}$,
- $\varphi(X)$ *et* $\psi(Y)$ *sont indépendantes.*

Ainsi, deux variables indépendantes sont toujours décorrélées. Mais la réciproque est fausse. il suffit de prendre $b > a > 0$, X qui suit la loi uniforme sur $\{-b, -a, a, b\}$ et $Y := X^2$.

On généralise ces résultats à n variables aléatoires $X = (X_1, X_2, \cdots, X_n)$. On appelle $Cov(X)$ la matrice $n \times n$ définie par : $Cov(X)_{i,j} := E((X_i - E(X_i)) \times (X_j - E(X_j)))$, $1 \leq i, j \leq n$. On a toujours :

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2 \sum_{1 \leq i < j \leq n} cov(X_i, X_j).$$

Les variables aléatoires $X_1, \cdots, X_n$ sont dites indépendantes si pout tout $a = (a_1, \cdots, a_n) \in \mathbb{R}^n$, les évènements $(X_1 = a_1), \cdots, (X_n = a_n)$ sont indépendants.

Les variables $X_1, \cdots, X_n$ sont dites décorélées si la matrice de covariance associée est une matrice diagonale. Cela est le cas si les $X_1, \cdots, X_n$ sont indépendantes, (ou seulement deux à deux indépendantes).

On retiendra le résultat suivant :

Proposition 3 (Variance de la somme de variables indépendantes) *Si $X_1, \cdots, X_n$, sont n variables aléatoires indépendantes alors $Var(X_1 + \cdots + X_n) = Var(X_1) + \cdots + Var(X_n)$.*

5 Loi des grands nombres

Nous donnons une première version abstraite et simplifiée de ce théorème de Mathématiques. Historiquement, il a été démontré pour la première fois par Jean Bernoulli, (1654-1705), dans une oeuvre posthume (1713). La démonstration la plus rapide d'aujourdhui repose sur l'inégalité de Bienaymé-Tchebicheff (et donc l'existence d'un moment d'ordre 2). Enonçons et démontrons ce théorème.

Rappelons qu'une suite de variables aléatoires indépendantes est une suite de variables aléatoires telles que n'importe quel choix d'un nombre fini d'entre elles forment une famille finie de variables aléatoires indépendantes.

Théorème 1 (Loi faible des grands nombres)
Soit $(X_n)_{n\geq 1}$, une suite de variables aléatoires indépendantes de mme moyenne μ et de mme écart type σ. Soit $M_n := \frac{1}{n}\sum_{i=1}^{n} X_i$, alors, pour tout $\varepsilon > 0$ on a :

$$\lim_{n \to +\infty} P(|M_n - \mu| > \varepsilon) = 0.$$

<u>démonstration</u> : Comme $E(M_n) = \mu$ et $Var(M_n) = \sigma^2/n \to 0$ il suffit d'appliquer l'inégalité de Bienaymé-Tchebicheff pour conclure.

<u>Exercice</u> : Rédiger les étapes intermédiares de la démonstration. Où intervient l'indépendance ?

Ainsi, se théorème dit d'une certaine manière que la moyenne arithmétique des variables X_i "converge" vers leur espérance commune.

Le lien avec les statistiques sera fait dans la partie **??**.

6 Schéma de Bernouilli et loi binomiale

Un schéma de Bernouilli consiste en l'épreuve répétée de manière indépendante d'une expérience alétoire ne comportant que deux issues. Par exemple, le jeu de pile ou face, ou le tirage répété avec remise dans une urne à deux catégories. C'est ce dernier exemple que Jean Bernoulli a présenté pour la première fois.

Pour les applications, nous nous intéresserons particulièrement au jeu de pile ou face en le simulant à l'aide d'une calculatrice.

Si l'on fait N lancers. On note 0 si l'on obtient face et 1 si l'on obtient pile. L'univers Ω_N qui correspond à toute les parties possibles de N lancers consécutifs est $\{0,1\}^N$. Ainsi, il y a 2^N parties possibles. La description d'une partie est entièrement codée par $\omega = (\omega_1, \cdots, \omega_N) \in \Omega$. On note X_i le résultat du lancer numéro i, ainsi X_i n'est rien d'autre qu'une projection : $X_i(\omega) = \omega_i$. Si l'on ne s'intéresse qu'au nombre de succés, il suffit de considérer $S_N := X_1 + \cdots + X_N$. Notez que S_N est application parfaitement déterminée de Ω_N dans $\{0, 1, \cdots, N\}$. Où est passé le hasard ? Il est caché dans ω, la partie que l'on va effectivement jouer. Avant de commencer la partie, on ne sait pas laquelle des 2^N parties on va réllement jouer. Si la pièce est parfaitement symétrique on s'attend à ce que toute les parties ont la mme chance de survenir. Ainsi, on se trouve dans le cas équiprobale. Par les calculs usuels de dénombrements on obtient donc aisément la loi de S_N :

$$P(S_N) = \frac{C_N^k}{2^N}, \qquad , k \in \{0, 1, \cdots, N\}.$$

On dit alors que S_N suit la loi binomiale $\mathcal{B}(N, 1/2)$.

Ainsi, on peut faire le lien avec les statistiques. Lorsque l'on va faire, pour de vrai, nos lancers on va noter le résultat de chaque lancer dans une série statistique $(x_i)_{1 \le i \le N}$. On modélise cette expérience par les variables aléatoires $(X_i)_{1 \le i \le N}$. Ainsi, le lien entre la série statistique et la suite de variables aléatoires correspond à la partie réllement jouée codée par un seul élément ω de Ω. Ou encore, c'est à dire, qu'il n'existe qu'un unique $\omega \in \Omega$ tel que

$$x_i = X_i(\omega).$$

Ainsi, si on fait 100 lancers on a environ 8% de chances d'obtenir exactement autant de fois pile que face. C'est à dire, par définition des probabilités, que 8% de toute les parties possibles de pile ou face conduisent à ce résultat. Il y a quand mme

$$2^{100} = 1267650600228229401496703205376 \simeq 1.3 \times 10^{30}$$

parties possibles. Le nombre de parties conduisant à l'égalité parfaite $S_{100} = 50$ est

$$C_{100}^{50} = 100891344545564193334812497256 \simeq 10^{29}.$$

D'autre part, il y a un autre point de vue tout aussi important, celui des Statistiques. En effet, la loi (forte) des grands nombres nous assurent que sur toutes les parties jouées (et à jouer) de 100 lancers d'une pièce parfaite 8% d'entre elles ont donné autant de fois le cté pile que le cté face.

Plus généralement, si la pièce n'est pas parfaite,

$$S_N = X_1 + \cdots + X_N,$$

où $X_k \sim \mathcal{B}(p)$, et les $(X_k)_{1\leq k\leq N}$ sont des variables aléatoires indépendantes. On dit que S_N suit la loi binomiale de paramètre N et p :

$$S_N \sim \mathcal{B}(N,p) \text{ si et seulement si } P(S_N = k) = C_N^k p^k (1-p)^{N-k}.$$

Cette loi fondamentale compte le nombre de succés sur N épreuves répétées, indépendantes ne comportant que deux issues : succés ou échec. Il est vivement conseillé au candidat de s'approprier la signification de cette loi à l'aide d'exercices du secondaire. En particulier, vous montrerez que :

$$E(S_N) = Np, \qquad Var(S_N) = Np(1-p), \qquad \sigma(S_N) = \sqrt{Np(1-p)}.$$

6.1 Simulation pour le jeu de pile ou face

A vos calculatrices : random, suite, somme, ...

1. Simulez 100 lancers d'une pièce non trucquée à l'aide de votre calculatrice. Comptez le nombre de fois que vous avez obtenu le côté pile. Modélisez cette expérience à l'aide de la variable aléatoire X.
2. Tout le monde obtient entre 40 et 60 piles ? Calculez la probabilité de cet évenement et commentez les résultats obtenus avec votre calculatrice.
3. Tout le monde obtient entre 37 et 63 piles ? Calculez la probabilité de cet évenement et commentez les résultats obtenus avec votre calculatrice.
4. Quelqu'un vous dit qu'il a obtenu 64 piles, qu'en pensez vous ?