Rappels de probabilités et statistique + quelques applications

Évènements, probabilités : définitions

- **Épreuve ou expérience aléatoire** : processus dont le résultat est incertain (tirage au hasard , jets de pièces, de dès,…)
- **Évènement** : ensemble de résultats d'une épreuve aléatoire
    - **Évènement élémentaire ou simple** : événement constitué d'un seul élément (*e.g.* avoir un 6 pour le lancer d'un seul dé)
    - **Évènement composé** : union de plusieurs évènements élémentaires (*e.g.* avoir un multiple de 3 lors du lancer d'un seul dé ou de plusieurs dés)
    - **Évènement certain** : union de tous les évènements élémentaires. Sa probabilité vaudra 1
    - **Évènement impossible** : ensemble vide d'évènement {} dont la probabilité sera nulle
    - **Événements incompatibles** : 2 événements sont incompatibles si et seulement si leur réalisation simultanée est impossible
    - **Événements compatibles** : événements dont la réalisation simultanée est possible
    - **Événements indépendants** : évènements tels que la réalisation de l'un n'influence pas la probabilité de réalisation de l'autre
        - Deux événements incompatibles, de probabilité non nulle ne sont jamais indépendants
- **Espace des observables** Ω : ensemble de tous les évènements élémentaires d'une expérience aléatoire (*e.g.* {1,2,3,4,5,6} pour le lancer d'un seul dé)
- **Probabilité** : quantification du caractère probable d'un évènement, nombre réel compris entre 0 et 1. Plus ce nombre est grand, plus le risque, ou la chance, que l'événement se produise est grand.
    - À tout événement élémentaire, E_i correspond une probabilité d'obtenir cet événement, $p(E_i)$
        - $0<p(E_i) < 1$
        - $p(E_i \ ou \ E_j) = p(E_i) + p(E_j) $
        - $\sum_{E_i} p(E_i) =p(\Omega) = 1$
- **Évènements équiprobables** : évènements élémentaires ou composés, dont la probabilité est strictement équivalente
- **Théorie des ensembles et logique booléenne** (dans le cadre des probabilités élémentaires). Soient A et B deux évènements *a priori* composés d'un ensemble d'évènements élémentaires. On aura :
    - si $A=\Omega$ alors $p(A)=1$: évènement certain
    - si $A=\left\{\right\}$ alors $p(A)=0$: évènement impossible (ex. faire 0 au dé)
    - si $A \subset B$ ou en écriture logique $A \Rightarrow B$, alors $p(A) \le p(B)$ *e.g.*faire 2 implique un nombre pair
    - **Loi de multiplication** $A \cap B$ (ET) :
        - A et B sont incompatibles alors $A \cap B = 0$ et $p(A \cap B)=0$: faire un 2 ET un nombre impair
        - A et B sont indépendants alors $p(A \cap B)=p(A) p(B)$: tirer une dame de coeur dans un jeu de 52 cartes = obtenir une dame **ET** avoir la couleur coeur, donc $p=1/13 \ 1/4= 1/52$

    - **Loi d'addition** $A \cup B$ (OU) : $p(A \cup B) = p(A) + p(B) - p(A \cap B)$
        - *e.g.* avoir un nombre pair ou un multiple de 3, réponse: 2, 4, 6 et 3 ($p = 4/6$). En terme de probabilités : $p = 1/2 + 1/3 - 1/6 = 2/3$)
        - **Attention!** si A et B sont incompatibles $A \cap B =0$ et $p(A \cup B) = p(A) + p(B)$}
- **Définition expérimentale de la probabilité d'un événement** :
    - Si on réalise N évènements **indépendants** lors d'une expérience (*e.g.* pile ou face), et si on observe n_i résultats du type E_i, alors $p(E_i) = \lim_{N \rightarrow \infty} {n_i \over N}$ (*N.B.* : cela peut-être testé facilement par des expériences “pile ou face')

Analyse combinatoire (éléments)

L'analyse combinatoire étudie les configurations de collections finies d'objets, et les dénombrements.

- **Permutations sans répétition d'objets discernables** : si vous avez *e.g.* 5 objets différents à placer à 5 emplacements, vous avez 5 possibilités pour placer un des objets en premier lieu, puis, indépendamment 4 autres pour le second objet, 3 pour le troisième, et 2 pour le quatrième. Le cinquième n'a plus qu'une seule possibilité. L'indépendance des placements induit une multiplication des différentes possibilités, soit ici $5\times4\times3\times2\times1$ possibilités. Par généralisation, le nombre de permutations d'un nombre n d'objets discernables vaut factorielle de n et est noté $n!$, avec $n! = \prod_{1\le i\le n} i = 1\times 2\times 3\times \cdots \times (n-1) \times n$
- **Permutations avec répétition d'objets discernables** : si on veut déterminer le nombre total de dispositions de 9 lettres dont précisément 4 A, 3 B et 2 C, il faut réduire les $9!$ permutations en tenant compte qu'il s'agit de 3 lettres discernables, mais qu'entre elles, les lettres A, B et C sont absolument indiscernables. Dans l'exemple, il faut donc diviser par $4!$, $3!$ et $2!$. En généralisant, le nombre de permutations de *n* éléments, répartis dans *k* classes dont n_1 sont de classe 1, n_2 sont de classe 2, …, n_k sont de classe *k*, indiscernables dans chaque classe, ou le nombre de permutations de *n* éléments avec n_1, n_2, …, n_k répétitions, avec $\left(\sum^k_{i=1} n_i = n \right)$, est égal à : $\frac{n!}{n_1! n_2! \ldots n_k!}$
- **Arrangements** de p éléments parmi n : si on effectue p tirages successifs parmi un ensemble de p objets différents, le nombre d'arrangements vaut $n (n-1) (n-2) ... (n-p+1)$ et est noté A_n^p. En utilisant les notations factorielles, on a $A_n^p= {n! \over (n-p)}\quad\mbox{avec }p\leq n!$
    - Par exemple, au Lotto, sept boules sont prélevées une par une parmi 45 boules numérotées (de 1 à 45). Cela donne donc $45\times 44\times 43\times 42\times 41\times 40\times 39 = {45! \over 38!}=228713284800$ ou deux cent vingt-huit milliards sept cent treize millions deux cent quatre-vingt-quatre mille huit cents arrangements ou tirages possibles !
- **Combinaisons sans répétition**
    - En poursuivant l'exemple du lotto, ce jeu attribue des gains [1] en fonction des numéros des 6 premières boules tirées quelque soit l'ordre de tirage de ces boules, et du numéro de la septième (numéro bonus). Les joueurs remplissent des grilles dont l'unité de base (combinaison de jeu) consiste à indiquer 6 numéros dans une série de 45 avec une mise de 1€. Si on s'intéresse uniquement au “gros lot”, le “rang 1” pour lequel les 6 numéros de la grille correspond exactement au 6 boules tirées, le nombre de combinaisons possibles est obtenu en divisant le nombre d'arrangements de 6 boules tirées parmi 45, divisé par le nombre de permutations de ces 6 boules qui donnent toutes la même

    combinaison. On a donc ${45! \over {39! \times 6!}}=(45\times 44\times 43\times 42\times 41\times 40)/(6\times 5\times 4\times 3\times 2) = 8145060$ possibilités !
  - En généralisant, le nombre de combinaison est donc obtenu en divisant le nombre d'arrangement par la factorielle du nombre d'objets tirés. On utilise les relations et notations suivantes : $C_n^p = {A_n^p \over p!} = {n! \over (n-p)! p!} = \binom{n}{p}$
  - L'exemple du Lotto permet d'introduire une **interprétation des combinaisons sans répétition comme des permutations avec répétitions de 2 objets discernables**. C'est très clair si on considère qu'une grille reprenant les 45 possibilités indique une combinaison de jeu par 6 “croix”, et 39 cases vierges. Le nombre de permutation de 6 croix et 39 cases vides vaut bien ${45! \over {39! \times 6!}}$.
  - **Propriétés de** C_n^p :
    - Symétrie : $C_n^p = C_n^{n-p}$
    - Triangle de Pascal : $C_n^p = C_{n-1}^{p} + C_{n-1}^{p-1}$
    - Binôme de Newton : $(x + y)^n = \sum_{p=0}^n C_n^p \ x^p \ y^{n-p}$
      - Démonstration par récurrence, considérant que la relation est vraie pour $n=0, 1$ et en utilisant la formule du triangle de Pascal, \begin{eqnarray*} (x + y)^{n+1} &=& (x+y) \ \sum_{p=0}^n C_n^p \ x^p \ y^{n-p}\\ &=& x^{n+1} + x \sum_{p=0}^{n-1} C_n^p \ x^p \ y^{n-p} + y^{n+1} + y \sum_{p=1}^n C_n^p \ x^p \ y^{n-p}\\ &=& x^{n+1} + y^{n+1} + \sum_{p=1}^n \left[C_n^p + C_n^{p-1}\right] \ x^p \ y^{n-p+1} \\ &=& x^{n+1} + y^{n+1} + \sum_{p=1}^n C_{n+1}^p \ x^p \ y^{n+1-p} \\ &=& \sum_{p=0}^{n+1} C_{n+1}^p \ x^p \ y^{n+1-p} \end{eqnarray*}

Variable aléatoire et distributions

- **Variable aléatoire** : une variable aléatoire X est définie sur l'espace des observables (espace des événements possibles). À chaque valeur possible x correspond une probabilité $P(x)$ que X soit égale à x
  - Variable aléatoire discrète : si $x_1, x_2, x_3, ...$ constitue l'ensemble discret des valeurs possibles de X, les $P(x_i)$ forment la **distribution de probabilité** de la variable aléatoire X
  - Variable aléatoire continue : si x peut varier continûment, $P(x)$ est la densité de probabilité que la variable prenne une valeur comprise entre x et $x+dx$. L'unité de $P(x)$ est donc en inverse de celle de l'espace des x et seul $P(x) dx$ a la dimension d'une probabilité (nombre) : $P(x) dx = P(x \le X < x+dx)$
  - Positivité :
    - $P(x_i) \ge 0$ pour tout x_i (variable aléatoire discrète)
    - $P(x) \ge 0$ pour tout x (variable aléatoire continue)
  - Normalisation :
    - $\sum_{x_i} P(x_i) =1$ (variable aléatoire discrète)
    - $\int_{\Omega} P(x) dx = 1$ (variable aléatoire continue)
- Toute l'information sur une expérience est contenue dans la distribution $P(x)$}
- Une description **équivalente** est donnée par l'ensemble de toutes les grandeurs caractéristiques appelées {\bf moments de la distribution} :
  - $<X^n> = \sum_i x_i^n P(x_i)$ (variable aléatoire discrète, avec n fini)
  - $<X^n> = \int_{\Omega} x^n P(x) dx$ (variable aléatoire continue, avec n infini)
- Une description **simplifiée** est obtenue en ne tenant compte que de quelques plus petites valeurs de n :
  - Premier moment: moyenne $<X>$ (ou espérance mathématique)

    - Second moment: largeur de la distribution (variance σ^2)
    - Troisième moment : asymétrie (skewness)
    - Quatrième moment : aplatissement (kurtosis)
    - …
- Les deux premiers moments
    - **Valeur moyenne ou espérance**
        - $<X> = \sum_i x_i \ P(x_i)$ ou $<X> = \int_{{\Omega}} x \ P(x) dx$ avec ${\Omega}$ le volume de l'espace des phases/observables
    - **Variance**
        - La variance $Var(X)$ ou σ^2 caractérise la largeur de la distribution (ou l'écart à la moyenne) : $\sigma^2 = <(X - <X>)^2> = <X^2> - <X>^2$. La racine carrée est l'écart type, σ.

Références

- A Concrete Introduction to Probability (using Python)
- 6 ways to test for a Normal Distribution — which one to use? 13/12/2019
- Le paradoxe de Simpson — Science étonnante #7
- Fix Me! loi de Poisson (ajouter)
    - Twitter https://twitter.com/ChatSceptique/status/1301540074178465793 → “Il y a 60% de chance d'observer au moins une étoile filante sur une heure. Quelle probabilité d'en observer au moins une sur 15 minutes ?”

[1]
en fait statistiquement des “pertes” puisque le taux de redistribution est d'environ 50%