Chapitre 4 : Probabilités

Julien Reichert

0 Introduction à la combinatoire

Soit un ensemble E à 6 éléments distincts. On souhaite calculer le nombre de façons d'ordonner E. Le nombre de choix possibles pour le premier élément est 6 : tous les éléments de E conviennent. Après cela, le nombre de choix possibles pour le deuxième élément n'est plus que 5, car l'élément retenu pour être le premier n'est plus disponible. Ensuite, le nombre de choix possibles pour le troisième élément est de 4 pour la même raison, et ainsi de suite. Le nombre de façons d'ordonner E, ou nombre de permutations, est alors $6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$.

Dans le cas d'un ensemble à n éléments, ce principe s'étend et le nombre de permutations est de $n \times (n-1) \times \cdots \times 2 \times 1$. On note $n!$ et on appelle « factorielle n » ce nombre.

Reprenons l'ensemble E, on souhaite calculer le nombre de façons de prendre 2 éléments de E en leur donnant un ordre. Cette fois-ci, le calcul s'arrête une fois le deuxième élément sélectionné. Il y a donc $6 \times 5 = 30$ possibilités. En quelque sorte, on a fait $6 \times 5 \times 4 \times 3 \times 2 \times 1$ qu'on a divisé par ce qui n'était pas nécessaire, à savoir $4 \times 3 \times 2 \times 1$.

Dans le cas d'un ensemble à n éléments, le nombre de façons de prendre k éléments de E en leur donnant un ordre, aussi appelé nombre d'arrangements de k parmi n éléments, est $n \times (n-1) \times \cdots \times (n-k+1)$. Il y a k facteurs dans ce produit, qu'on écrit plus simplement $\frac{n!}{(n-k)!}$.

Toujours avec l'ensemble E, on souhaite maintenant calculer le nombre de façons de prendre 2 éléments de E, mais sans ordre cette fois. Les 30 possibilités obtenues avant se rassemblent en 15 paires de façons indistinctibles. En fait, on a divisé 30 par le nombre de façons d'ordonner les 2 éléments retenus afin d'éviter les redondances.

Dans le cas d'un ensemble à n éléments, le nombre de façons de prendre k éléments de E sans ordre, aussi appelé nombre de combinaisons de k parmi n éléments, est $\frac{n \times (n-1) \times \cdots \times (n-k+1)}{k \times (k-1) \times \cdots \times 1}$. Il y a k facteurs au numérateur et k au dénominateur, et on écrit plus simplement ce nombre de combinaisons $\frac{n!}{(n-k)!k!}$. Ce nombre est également appelé coefficient binomial d'indices $(n;k)$ et écrit $\binom{n}{k}$.

1 Probabilités discrètes

1.1 Rappels d'années précédentes (ou définitions de base)

Définition
On considère une expérience aléatoire, qui peut aller du lancer d'une pièce à un mélange de cartes en passant par le tirage du loto. Une **issue** est un résultat possible de l'expérience aléatoire, par exemple « pile », une permutation des cartes ou une combinaison de boules entre 1 et 50 plus un numéro complémentaire. L'ensemble des issues est appelé l'**univers** associé à l'expérience et noté Ω. Il est important de savoir calculer le nombre de ses éléments. Un **événement** A est un sous-ensemble (aussi appelé une partie) de l'univers, donc un ensemble d'issues. On dit qu'une issue **réalise** un événement si elle appartient à l'ensemble correspondant. Des événements correspondant aux expériences sont par exemple $\{\text{face}\}$, l'ensemble des mélanges tels que la troisième carte soit un pique ou l'ensemble des tirages figurant sur une grille de loto qu'on a jouée. L'événement **impossible**, noté $\emptyset$, correspond à l'ensemble vide ; l'événement **certain** est Ω et un événement **élémentaire** est un sous-ensemble de Ω de taille 1.

Définition
L'intersection de deux événements A et B, notée $A \cap B$, est l'intersection au sens ensembliste des ensembles correspondant aux événements. Il s'agit de l'ensemble des issues réalisant à la fois A et B. De même, la réunion de deux événements A et B, notée $A \cup B$, est l'ensemble des issues réalisant A ou B, voire les deux, et le complémentaire d'un événement A, noté $\bar{A}$, est l'ensemble des issues ne réalisant pas l'événément A.

On étend la notion d'intersection ou de réunion à plus de deux événements.

Proposition
Pour deux événéments A et B, on a :
- $\overline{A \cup B} = \bar{A} \cap \bar{B}$;
- $\overline{A \cap B} = \bar{A} \cup \bar{B}$;
- $\bar{\bar{A}} = A$.

Définition
Deux événéments sont dits **disjoints** ou **incompatibles** lorsque leur intersection est l'événement impossible. Les événements A et $\bar{A}$ sont toujours disjoints, par définition.

Définition
On considère une expérience aléatoire, avec le vocabulaire défini précédemment. Une **loi de probabilité** sur l'univers (fini) Ω associe à chaque issue une probabilité, qui est un réel positif (potentiellement nul, forcément inférieur ou égal à 1 d'après ce qui suit), de sorte que la somme des probabilités de toutes les issues soit exactement 1.

Dans le cas où les issues sont notées $e_1, e_2, \ldots, e_n$, on peut noter p_i la probabilité de chaque issue e_i, appelée probabilité de l'événement élémentaire $\{e_i\}$. **La probabilité d'un événement est la somme des probabilités des issues qui le composent.** On note $P(A)$ la probabilité de l'événément A.

Proposition
Pour deux événéments A et B, on a :
- $P(\bar{A}) + P(A) = 1$. En effet, toute issue réalise exactement un des événements A ou $\bar{A}$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. En particulier, si A et B sont incompatibles, la formule devient $P(A \cup B) = P(A) + P(B)$ car on retire la probabilité de l'événément impossible. En fait, pour trouver la probabilité de $A \cup B$, on additionne les probabilités de toutes les issues de A et de celles de B, mais celles de $A \cap B$ sont alors comptées une fois de trop et il faut les retirer.

Définition
Soit X une variable aléatoire réelle. L'**espérance** de X est la moyenne des valeurs de X pondérées par leur probabilité. Elle s'obtient par la formule $E(X) = \sum_{k \in \Omega} k \times P(X = k)$.

Définition
Soit X une variable aléatoire réelle. La **variance** de X est la moyenne des carrés des écarts entre X et $E(X)$, pondérés par leur probabilité. Elle s'obtient par la formule $V(X) = E((X - E(X))^2)$, et on a aussi $V(X) = E(X^2) - E(X)^2$.

L'écart-type, comme en statistiques, est la racine carrée de la variance.

Définition
Une **partition** de l'univers Ω est une liste d'événéments $(A_1, A_2, \ldots, A_n)$ disjoints deux à deux et de réunion totale Ω. Par conséquent, on a $P(A_1) + P(A_2) + \cdots + P(A_n) = 1$.

La notation pour une partition est relativement libre, avec ou sans parenthèses, éventuellement avec des accolades à la place.

Soit A un événément. Les événements A et $\bar{A}$ forment une partition de l'univers.

Une autre partition de base est l'ensemble des événements élémentaires.

1.2 Lois de probabilité usuelles

Il y en a trois au programme de la section ES : la loi uniforme, la loi de Bernoulli et la loi binomiale.

La loi uniforme (dite « équirépartie ») associe à tout élément de l'univers la même probabilité. Dans le cas d'un univers fini de taille n, la probabilité de chaque issue sera donc $\frac{1}{n}$. L'espérance d'une variable aléatoire suivant la loi uniforme est la moyenne simple de toutes les issues.

La loi de Bernoulli est la loi de toute variable aléatoire représentant une épreuve de Bernoulli, donc quand l'univers est limité à deux issues.

La loi binomiale de paramètres n et p associe à tout entier k entre 0 et n la probabilité d'obtenir k fois l'issue favorable en répétant n fois une épreuve de Bernoulli pour laquelle la probabilité de l'issue favorable est toujours p. En notant X la variable aléatoire associée, les formules à retenir sont : $P(X = k) = \binom{n}{k}p^k(1-p)^{n-k}$, $E(X) = np$ et $\sigma(X) = \sqrt{np(1-p)}$ (cette dernière valeur étant l'écart-type).

1.3 Probabilités conditionnelles

Soient A et B deux événements. On considère que A est réalisé (donc réalisable !). Il est alors possible que la probabilité de B en soit affectée. On appelle cela la probabilité conditionnelle : la notation $P_A(B)$ se lit « probabilité de B sachant A », sa signification étant intuitive.

Par exemple, en prenant un mois de l'année au hasard (suivant la loi uniforme), la probabilité que ce soit mars sachant qu'il a 31 jours n'est plus de $\frac{1}{12}$, ce qui était la probabilité sans information supplémentaire, mais $\frac{1}{7}$, car il y a sept mois de 31 jours. Au passage, la probabilité que ce soit par exemple avril sachant qu'il y a 31 jours tombe alors à 0.

La nouvelle notation donne lieu à plusieurs formules :

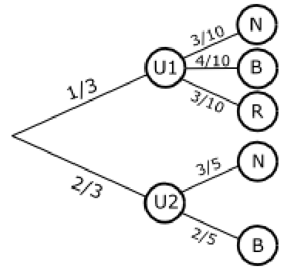
$$P_A(B) = \frac{P(A \cap B)}{P(A)} \ ;$$

$$\text{donc } P(A \cap B) = P_A(B) \times P(A) = P_B(A) \times P(B).$$

Quand une probabilité est inconnue, elle peut se déduire de deux probabilités connues à partir de ces formules.

Grâce aux probabilités conditionnelles, il est possible de construire un arbre de probabilités pour plusieurs partitions de l'univers d'une ou plusieurs variables aléatoires. À chaque nœud de l'arbre, des branches partent le long d'une partition, et sur chaque branche, la probabilité conditionnelle dans le contexte de la branche est notée. Dans l'exemple ci-après, on dispose de trois urnes indiscernables, dont une de type 1 (U_1) et deux de type 2 (U_2). Les urnes de type 1 contiennent trois boules noires (N), quatre boules blanches (B) et trois boules rouges (R) ; les urnes de type 2 contiennent trois boules noires et deux boules blanches.

Il y a donc trois partitions de trois univers différents, et la boule tirée se lit sur la partie droite de l'arbre en suivant un chemin. On retrouve la formule de la probabilité de l'intersection à partir d'une probabilité conditionnelle.



En gardant le même exemple, on peut se demander la probabilité de tirer une boule noire. Il y a deux cas possibles et incompatibles : c'est une boule noire d'urne de type 1 ou une boule noire de type 2. Le premier cas a pour probabilité $P(U_1 \cap N) = P(U_1) \times P_{U_1}(N) = \frac{1}{3} \times \frac{3}{10} = 0,1$; le second cas a pour probabilité $P(U_2 \cap N) = P(U_2) \times P_{U_2}(N) = \frac{2}{3} \times \frac{3}{5} = 0,4$. La probabilité totale est donc la somme, puisque les deux cas sont incompatibles, soit $0,5$.

On en déduit la formule dans le cas général d'une partition $(A_1, \ldots, A_n)$:

$$P(B) = \sum_{i=1}^{n} P(A_i \cap B) = \sum_{i=1}^{n} P(A_i) \times P_{A_i}(B).$$

Elle s'exprime plus simplement quand la partition se limite à un événement et son contraire :

$$P(B) = P(A \cap B) + P(\bar{A} \cap B) = P(A) \times P_A(B) + P(\bar{A}) \times P_{\bar{A}}(B).$$

En outre, à chaque nœud de l'arbre, on voit apparaître la formule suivante :

$$\sum_{i=1}^{n} P_B(A_i) = 1.$$

Attention, dans ce dernier cas, l'ordre d'appartition des A_i et de B dans l'arbre est inversé.

Définition
Deux événements A et B sont dits **indépendants** si la réalisation ou non de l'un ne change pas la probabilité de l'autre. En d'autres termes, si $P_A(B) = P_{\bar{A}}(B) = P(B)$ et $P_B(A) = P_{\bar{B}}(A) = P(A)$.

Remarque : Si A a une probabilité strictement comprise entre 0 et 1, le fait que $P_A(B) = P(B)$ entraîne forcément que $P_{\bar{A}}(B) = P(B)$. Cela se voit à partir des formules précédentes. En outre, si $P_A(B) = P(B)$, alors on a $P(A \cap B) = P_A(B) \times P(A) = P(B) \times P(A)$ mais aussi $P(A \cap B) = P_B(A) \times P(B)$, d'où $P_B(A) = P(A)$. Finalement, une seule condition est nécessaire et suffisante.

Proposition
La probabilité de l'intersection de deux événements indépendants est le produit de leur probabilité respective.

Par convention, un événement impossible est indépendant de tout autre événement, un événement certain aussi. En-dehors de cela, aucun événement n'est indépendant de lui-même ni, clairement, de son complémentaire.

Définition
Deux variables aléatoires sont indépendantes si tous les événements associés à l'une sont indépendants des événements associés à l'autre.

Proposition
Si deux variables aléatoires sont indépendantes, l'espérance de leur somme est la somme de leurs espérances respectives et la variance de leur somme est la somme de leurs variances respectives.

Ceci permet de retrouver rapidement l'espérance et la variance d'une variable aléatoire suivant une loi binomiale.

2 Probabilités continues

Dans cette section, les valeurs possibles pour les variables aléatoires ne formeront plus un ensemble fini, ni même dénombrable, il s'agira plutôt d'intervalles de $\mathbb{R}$. Dans ce cas, ce qui est valable pour les probabilités discrètes doit être adapté.

Définition
Une fonction f définie sur un intervalle $[a; b]$ est une fonction **densité de probabilité** à condition que : f soit continue et positive sur $[a; b]$ et que $\int_a^b f(x)\,\mathrm{d}x = 1$.

Remarque : Il n'est pas du tout nécessaire que f soit partout inférieure à 1, elle peut même ne l'être nulle part.

Définition
Soit X une variable aléatoire. On dit que X suit la loi de probabilité continue à fonction de densité f si les valeurs possibles de X forment l'ensemble de définition de f. Dans ce cas, la probabilité que X soit dans un sous-intervalle $[c;d]$ de $[a;b]$ est $P(c \leq X \leq d) = \int_c^d f(x)\,\mathrm{d}x$.

Remarque : Par définition, la probabilité que X vaille précisément un réel est forcément 0.

Définition
Une variable aléatoire X suit la loi uniforme continue si la fonction de densité associée est une fonction constante. La valeur de la constante est nécessairement $\frac{1}{b-a}$. De plus, on a alors $P(c \leq X \leq d) = \frac{d-c}{b-a}$.

Définition
Une variable aléatoire X suit la loi normale $\mathcal{N}(0;1)$ si la fonction de densité associée est la fonction dite de Gauß $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, définie sur tout $\mathbb{R}$. La courbe de la fonction φ est en forme de cloche. L'espérance de X est alors 0 et son écart-type est 1.

Remarque : Il est impensable de chercher une primitive de la fonction de Gauß, mais une calculatrice scientifique permet cependant de faire des calculs de probabilité avec la loi normale, les fonctions figurant même dans le catalogue.

Proposition
Soit X une variable aléatoire qui suit la loi normale $\mathcal{N}(0;1)$. D'après la courbe de la fonction de Gauß, on sait que $P(X \leq 0) = P(X \geq 0) = \frac{1}{2}$. On a de plus les valeurs à retenir :
- $P(-1 \leq X \leq 1) \simeq 0,68$;
- $P(-2 \leq X \leq 2) \simeq 0,954$;
- $P(-3 \leq X \leq 3) \simeq 0,997$;
- Le réel k tel que $P(-k \leq X \leq k) = 0,95$ vaut environ $1,96$;

Remarque : Beaucoup d'autres valeurs de ce type se déduisent à l'aide de la relation de Chasles pour les intégrales.

Définition
Une variable aléatoire X suit la loi normale $\mathcal{N}(\mu;\sigma^2)$ si la variable aléatoire $\frac{X-\mu}{\sigma}$ suit la loi normale $\mathcal{N}(0;1)$. La fonction de densité associée est alors $\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. L'espérance de X est alors μ et son écart-type est σ.

Proposition
Les mêmes valeurs se retrouvent en s'adaptant aux nouveaux paramètres de la loi normale :
- $P(X \leq \mu) = P(X \geq \mu) = \frac{1}{2}$.
- $P(\mu - \sigma \leq X \leq \mu + \sigma) \simeq 0,68$;
- $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \simeq 0,954$;
- $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \simeq 0,997$;
- Le réel k tel que $P(\mu - k\sigma \leq X \leq \mu + k\sigma) = 0,95$ vaut environ $1,96$;

3 Échantillonnage

La notion d'échantillonnage est à la jonction entre probabilités et statistiques. Il s'agit en quelque sorte de considérer que les fréquences globales et locales sont relativement proches lorsqu'on étudie un caractère particulier.

On admet qu'on est en situation d'utilisation de la loi binomiale. Dans toute cette section, le nombre n représente la taille d'un échantillon, p est la proportion de personnes présentant un caractère parmi toute la population (de taille quelconque), f est la fréquence d'apparition du caractère dans l'échantillon, c'est donc le nombre de personnes présentant le caractère dans l'échantillon, divisé par n. On appelle par ailleurs f une estimation ponctuelle de p.

Définition
Un intervalle de fluctuation au seuil de 95 % est un intervalle dans lequel la fréquence f a une probabilité d'au moins 95 % de se trouver.

Par exemple, si on considère que la moitié de la population humaine est de sexe masculin, en prenant un échantillon représentatif de taille suffisante (disons 500 personnes), on peut supposer qu'il y a une probabilité supérieure à 95 % que la fréquence du sexe masculin soit entre $0,3$ et $0,7$.

Remarque : L'intervalle de fluctuation dépend de n. Plus n est grand, plus on peut trouver des intervalles resserrés.

Définition
L'intervalle de fluctuation asymptotique au seuil de 95 % est l'intervalle $[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}]$.

Les conditions auxquelles cet intervalle est pertinent sont les suivantes : $n > 30$, $np > 5$ et $n(1-p) > 5$. Il faut donc que l'échantillon soit assez grand pour que la présence et l'absence d'un caractère potentiellement trop rare ou trop fréquent soient observées.

Remarque : On retrouve le nombre $1,96$ qu'on avait repéré dans le cadre de la loi normale. Le nombre $\sqrt{p(1-p)}$ est l'écart-type de toute variable aléatoire suivant la loi de Bernoulli de paramètre p.

On utilise l'intervalle de fluctuation asymptotique ainsi : on émet une hypothèse sur la proportion p, donc on lui attribue une valeur sans en être sûr, et on prend un échantillon de taille n. On détermine l'intervalle de fluctuation asymptotique associé et on calcule la fréquence sur l'échantillon (il s'agit de compter). Si la fréquence est dans l'intervalle, alors l'hypothèse n'est pas remise en cause (ce qui ne prouve rien sur la vraie valeur de p, mais rassure quant à la précision). Si la fréquence n'est pas dans l'intervalle, alors on rejette l'hypothèse, mais avec un risque d'erreur de 5 %.

Remarque : Lorsque p est proche de $0,5$, le produit $1,96\sqrt{p(1-p)}$ est proche de 1 et on se permet d'utiliser une autre version de l'intervalle de fluctuation : $[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}]$.

Définition
L'intervalle de confiance de p au niveau de confiance 95 % est $[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}}]$.

On utilise l'intervalle de confiance lorsqu'on cherche à encadrer (avec probabilité 95 %) la valeur de p sans poser d'hypothèse. Les conditions sont cette fois $n > 30$, $nf > 5$ et $n(1-f) > 5$.

Remarque : Dire que p se situe dans l'intervalle de confiance avec une probabilité d'au moins 95 % est un abus, car l'échantillon est fixé, de même que la probabilité, il n'y a alors plus rien d'aléatoire. Rigoureusement, on pourrait se dire qu'on a une probabilité d'au moins 95 % de tomber sur un échantillon tel que p soit dans l'intervalle de confiance.