WIKIPÉDIA

Variable aléatoire réelle

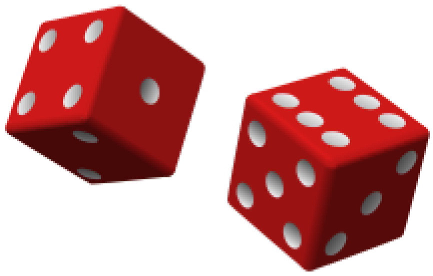
Une **variable aléatoire réelle** est une variable aléatoire à valeurs dans $\mathbb{R}$, ou une partie de $\mathbb{R}$; c'est une fonction définie depuis l'ensemble des résultats possibles d'une expérience aléatoire, dont on doit pouvoir déterminer la probabilité qu'elle prenne une valeur donnée ou un ensemble donné de valeurs. Les variables aléatoires réelles sont les variables aléatoires les plus couramment étudiées, ce qui conduit certains auteurs à omettre l'adjectif *réel*, et à parler de variable aléatoire tout court.

Les variables aléatoires sont très utilisées en théorie des probabilités et en statistiques. Dans les applications, les variables aléatoires sont utilisées pour modéliser le résultat d'un mécanisme non-déterministe ou encore comme le résultat d'une expérience non-déterministe qui génère un résultat aléatoire. En statistique mathématique ou inférentielle, les variables aléatoires servent généralement à modéliser des populations supposées infinies.

Cet article ne traite que les *variables aléatoires réelles* :

- L'article Variable aléatoire généralise cet article au *cas non réel* sous l'angle de la théorie de la mesure ;
- L'article Variables aléatoires élémentaires aborde la *notion de variable aléatoire* d'une manière plus intuitive.

Sommaire



Un exemple de variable aléatoire : la fonction qui associe au résultat du jet de deux dés la somme de leurs valeurs.

Détails

- À l'origine, une variable était une fonction de gain, qui représentait le gain obtenu à l'issue du résultat d'un jeu.[réf. nécessaire] Par exemple, supposons qu'un joueur lance un dé et que celui-ci gagne 1 € s'il amène un six et perde 10 € s'il amène un autre résultat. Alors il est possible de définir la variable aléatoire de gain qui associe 1 au résultat « six » et -10 à un résultat inintéressant. La probabilité pour que la variable aléatoire prenne la valeur 1 correspond exactement à la probabilité pour que le joueur gagne 1 €.

- La variable aléatoire réelle la plus simple est donnée par le résultat d'un lancer au jeu de pile ou face, qui vaut *pile* ou *face*. Un autre exemple simple est donné par le résultat d'un jet de dés, pour lequel les valeurs possibles sont 1, 2, 3, 4, 5 ou 6 (si le dé est classiquement cubique). De telles variables aléatoires réelles sont qualifiées de *discrètes* car elles prennent des valeurs bien séparées. *A contrario*, la mesure de la taille d'un individu pris au hasard dans une population ressemble davantage à un nombre réel positif (cela n'est pas tout à fait vrai non plus, car des questions ergonomiques rendent d'autant plus improbable l'énoncé d'un nombre qu'il comporte de décimales significatives ; l'attracteur est en fait un fractal). Cette variable aléatoire réelle est alors qualifiée par convention de *continue*. L'étude de la répartition des valeurs prises par une variable aléatoire conduit à la notion de loi de probabilité.

- En mathématiques, et plus précisément en théorie des probabilités, une **variable aléatoire** est une fonction mesurable définie sur un espace de probabilités. La mesure image correspondante est appelée loi de la variable aléatoire. Ce type de fonction permet de modéliser un phénomène aléatoire, comme le résultat d'un jet de dés. Une propriété intéressante de l'intégrale de Lebesgue fait qu'un évènement de probabilité strictement nulle n'est pas nécessairement *impossible* au sens strict du terme (ainsi, considérons un réel tiré *au hasard* dans l'intervalle [0, 1] ; la probabilité qu'il soit rationnel est nulle, alors que les rationnels constituent dans cet intervalle un ensemble infini, et même partout dense).

Quelques variables aléatoires réelles

En guise d'introduction aux définitions concernant les variables aléatoires réelles, il semble intéressant de présenter brièvement une famille de variables très utilisées.

Outre la variable certaine qui prend une valeur donnée avec une probabilité égale à 1, la variable aléatoire réelle la plus simple est appelée variable de Bernoulli. Celle-ci peut prendre deux états, qu'il est toujours possible de coder 1 et 0, avec les probabilités p et $1-p$. Une interprétation simple concerne un jeu de dé dans lequel on gagnerait un euro en tirant le six ($p = 1/6$). Sur une séquence de parties, la moyenne des gains tend vers p lorsque le nombre de parties tend vers l'infini.

Si on considère qu'une partie est constituée par n tirages au lieu d'un seul, le total des gains est une réalisation d'une variable binomiale qui peut prendre toutes les valeurs entières de 0 à n. Cette variable a pour moyenne le produit np. On obtient un exemple moins futile en considérant le score d'un candidat dans un sondage électoral.

Si n est assez grand et p pas trop petit, on peut trouver une approximation convenable en utilisant la variable de Gauss. Dans les sondages cela permet d'associer un intervalle de confiance au résultat brut. Ainsi, il y a 95 chances sur 100 pour qu'une enquête portant sur 1 000 personnes donne un résultat correct à ± 3 % près.

Toujours avec n grand, l'approximation de Poisson est préférable si p est assez petit pour que la moyenne np ne soit pas trop grande, de l'ordre de quelques unités. Dans un sondage ce serait la loi applicable aux « petits » candidats. C'est surtout la loi utilisée dans des problèmes de files d'attente.

La somme des carrés de ν variables de Gauss indépendantes est une variable de χ^2 à ν degrés de liberté (la variable exponentielle en est un cas particulier). Le test du χ^2 est utilisé pour apprécier la valeur de l'adéquation d'une loi de probabilité sur une distribution empirique.

Si on divise une variable de Gauss par une variable de χ (racine carrée de la précédente), on obtient une variable de Student. Le rapport de deux variables de χ^2 indépendantes définit une variable de Snedecor. Ces deux lois sont utilisées dans l'analyse de populations supposées gaussiennes.

Notions de base

Loi de probabilité

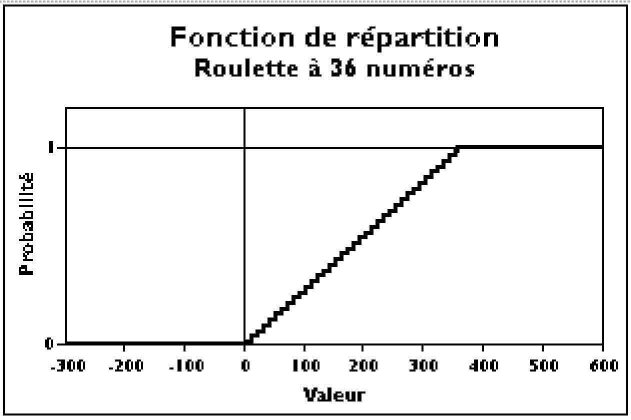
- La loi d'une variable aléatoire réelle décrit en détail la répartition des valeurs de cette variable. La loi de la variable X contient toutes les informations nécessaires pour calculer sa fonction de répartition, son espérance et plus généralement ses moments, sa fonction caractéristique, sa médiane et ses quantiles.

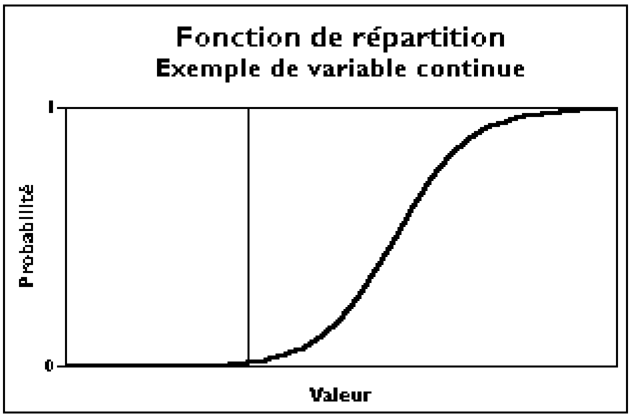
- En d'autres termes, si deux variables aléatoires réelles X et Y ont même loi de probabilité, alors elles ont même fonction de répartition, même espérance et plus généralement mêmes moments, même fonction caractéristique, même médiane et mêmes quantiles.
- La loi de probabilité d'une variable aléatoire réelle X est la donnée des valeurs $\mathbb{P}(X \in A)$ pour une classe très large de parties A de $\mathbb{R}$, la tribu borélienne, incluant en particulier les intervalles. Par exemple, la loi de X donne la valeur de $\mathbb{P}(a \leq X \leq b)$, pour tout choix de nombres réels $a < b$, qui correspond au calcul de $\mathbb{P}(X \in A)$ pour le choix $A = [a,b]$, ou encore la valeur de $\mathbb{P}(X \leq x)$, pour tout choix du nombre réel x, ce qui correspond au calcul de $\mathbb{P}(X \in A)$ pour le choix $A =]-\infty\ , x]$.
- Selon sa définition mathématique (voir l'article variable aléatoire pour plus de détail), la loi de X est la mesure de probabilité $\mathbb{P}_X$ définie, pour tout élément A de la tribu borélienne de $\mathbb{R}$ par la relation

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}\left(X^{-1}(A)\right).$$

- La loi d'une variable aléatoire réelle peut être discrète ou posséder une densité (par rapport à la mesure de Lebesgue sur $\mathbb{R}$). Ce sont les deux cas les plus fréquents dans les applications.

Fonction de répartition





Il serait possible d'introduire cette notion à partir de l'une quelconque des variables précédemment considérées mais il paraît plus clair d'étudier le cas du dé sous un angle différent. En effet, il définit une variable aléatoire X qui prend avec la même probabilité d'apparition (1/6) des valeurs dans l'ensemble {1,2,3,4,5,6}. On peut alors associer à toute valeur réelle x la probabilité d'obtenir un tirage inférieur ou égal à x, ce qui définit une courbe en escalier dont les marches ont une hauteur égale à 1/6.

Formellement, cela conduit à une fonction de répartition

$$F_X(x) = \mathbb{P}(X \leq x)$$

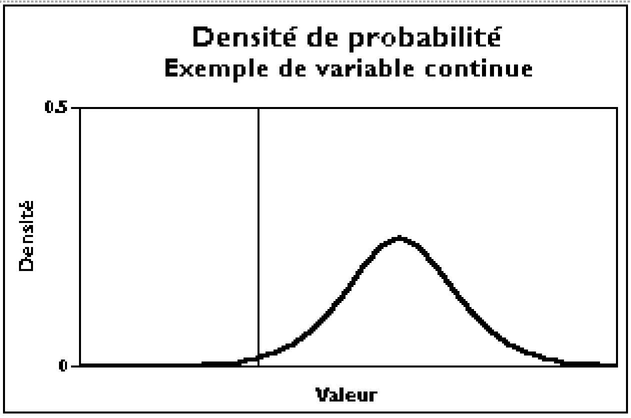
Dans celle-ci, la majuscule X représente la variable aléatoire réelle, ensemble de valeurs numériques, et la minuscule x représente la variable d'état, variable au sens usuel du terme.

Si les événements ne sont plus équiprobables, cela ne fait que déformer la courbe. Pour introduire une notion nouvelle, on peut commencer par remplacer le dé par une roulette à six numéros (ce qui conduit à un problème rigoureusement identique). Ensuite, on ne change rien de fondamental si on remplace les six nombres entiers par les repères des centres d'arcs de 60 degrés. À partir de là il est possible d'augmenter le nombre de secteurs en réduisant leur taille : les échelons deviendront de plus en plus petits jusqu'à être indiscernables sur un dessin. Le passage à la limite remplace la variable discrète par une variable continue qui prend toutes les valeurs réelles dans l'intervalle]0,360] : c'est une variable uniforme.

Une fonction de répartition est croissante (au sens large) sur l'intervalle $]-\infty\ , +\infty[$, et continue à droite en tout point ; elle tend vers 0 en $-\infty$ et vers 1 en $+\infty$. Réciproquement, toute fonction vérifiant les propriétés (caractéristiques) précédentes peut être considérée comme la fonction de répartition d'une variable aléatoire.

L'intérêt de la fonction de répartition réside dans le fait qu'elle est bien définie aussi bien pour les variables continues définies sur un ensemble continu que pour les variables discrètes définies sur un ensemble dénombrable (dans la plupart des cas pratiques il se réduit à un ensemble de valeurs équidistantes que l'on peut ramener à un ensemble d'entiers). Le remplacement progressif (l'approximation) d'une fonction de répartition dont la courbe est en escalier par une fonction de répartition dont la courbe est continue permet de voir intuitivement comment une variable continue peut fournir une approximation souvent plus facile à manipuler que la variable discrète originale. Voir l'article Convergence en loi pour une formulation plus mathématique de ce type d'approximation de variables discrètes par des variables continues.

Densité de probabilité d'une variable continue



Une variable continue possède souvent une fonction de répartition continue en tout point et dérivable par morceaux. Il est alors commode de la dériver pour obtenir la densité de probabilité, vérifiant :

$$p_X(x) = \frac{\mathrm{d}F_X}{\mathrm{d}x}(x)$$

qui est définie et à valeurs positives (ou nulles) sur]–∞ , +∞[, telle que $\int_{-\infty}^{+\infty} p_X(x)\,\mathrm{d}x = 1$.

On reconstruit la fonction de répartition par la relation :

$$F_X(x) = \int_{-\infty}^{x} p_X(\xi)\,\mathrm{d}\xi$$

Dans les raisonnements généraux, il est souvent commode d'écrire ces formules sous forme différentielle :

$$\mathbb{P}(x < X \leq x + \mathrm{d}x) = p_X(x)\,\mathrm{d}x$$

Si l'on effectue un changement de variable selon la formule $Y = f(X)$, la nouvelle densité de probabilité se calcule par :

$$p_Y(y)\mathrm{d}y = p_X(x)\mathrm{d}x$$

Fonction de probabilité et densité de probabilité d'une variable discrète

La loi d'une variable aléatoire discrète X est déterminée par l'ensemble des probabilités de ses valeurs nommé **fonction de probabilité** (*mass function* en anglais). Si l'on suppose qu'elle prend des valeurs entières (de signe quelconque), cela s'écrit :

$$\forall k \in \mathbb{Z}, p_X(k) = \mathbb{P}(X = k)$$

En implicitant la variable aléatoire discrète X, on peut alléger la notation comme suit :

$$\forall k \in \mathbb{Z}, p_k = \mathbb{P}(X = k)$$

On reconstruit la fonction de répartition (dont les valeurs sont alors appelées *probabilités cumulées*) par la relation :

$$\forall n \in \mathbb{Z}, \forall x \in [n; n+1[, F_X(x) = \sum_{k=-\infty}^{n} p_k$$

En considérant la fonction de répartition comme une somme d'échelons ou fonctions de Heaviside, sa dérivée peut s'interpréter comme une somme d'impulsions ou fonctions de Dirac et s'écrit donc :

$$\forall x \in \mathbb{R}, F'_X(x) = \sum_{k\in\mathbb{Z}} p_k\, \delta_k(x)$$

On vérifie alors bien que :

$$\forall x \in \mathbb{R}, p_X(x) = F'_X(x)$$

Cette « densité de probabilité » présente un intérêt dans un problème particulier : lorsqu'une intégrale porte sur une densité de probabilité, la propriété fondamentale de la fonction de Dirac permet de transformer l'intégrale en une simple somme impliquant la fonction de probabilité.

Espérance mathématique

Définitions

L'espérance mathématique d'une variable aléatoire réelle se définit comme la "moyenne" des valeurs de cette variable, pondérées par leurs probabilités de réalisation. Pour une variable continue, la formule différentielle donnée précédemment s'intègre, sous réserve d'intégrabilité, en

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x\, p_X(x)\; \mathrm{d}x.$$

Cette quantité est plus connue sous le nom de moyenne.

X étant une variable aléatoire réelle, une fonction f supposée régulière définit une nouvelle variable aléatoire $f \circ X$ notée $f(X)$ dont l'espérance, lorsqu'elle existe, s'écrit en remplaçant x par $f(x)$ dans la formule précédente (théorème de transfert).

$$\mathbb{E}[f(X)] = \int_{-\infty}^{+\infty} f(x)\, p_X(x)\; \mathrm{d}x.$$

Pour une variable discrète, la « densité de probabilité » conduit, sous réserve de sommabilité, à

$$\mathbb{E}[f(X)] = \sum_{k=-\infty}^{+\infty} f(k)\; P_X(k).$$

Fonction caractéristique

Si la densité de probabilité d'une variable aléatoire réelle X possède une transformée de Fourier, celle-ci (ou, plus précisément, la transformée inverse), fonction à valeurs complexes définie sur $\mathbb{R}$

$$\phi_X(t) = \mathbb{E}[\mathrm{e}^{itX}]$$

s'appelle *fonction caractéristique* de la variable.

Fonction génératrice des moments

La **fonction génératrice des moments** d'une variable aléatoire X est définie par

$$M_X(t) = \mathbb{E}\left(\mathrm{e}^{tX}\right), \quad t \in \mathbb{R},$$

lorsque son espérance existe. Cette fonction, comme son nom l'indique, est utilisée afin de générer les moments associés à la distribution de probabilités de la variable aléatoire X. Elle permet en outre de déterminer l'additivité d'une loi.

Moments

Si la fonction caractéristique (ou la fonction génératrice) d'une variable aléatoire est développable en série, celle-ci fait apparaître les *moments* de celle-ci, le moment d'ordre k étant défini comme

$$m_k \equiv \mathbb{E}[X^k]\,.$$

Dans le cas, important pratiquement, d'une variable assez régulière, celle-ci peut donc être caractérisée par la suite de ses moments, sa fonction caractéristique ou sa fonction génératrice, sa densité de probabilité ou, éventuellement, sa fonction de probabilité ou par sa fonction de répartition.

Dans le cas général, seuls les premiers moments peuvent exister.

Outils pratiques

Moments et moments centrés

Le moment d'ordre un, espérance ou *moyenne* de la variable,

$$\mu \equiv m_1 = \mathbb{E}[X],$$

est un *indicateur de tendance centrale*.

Les moments d'ordre supérieur éliminent ce paramètre de position en considérant la variable centrée par soustraction de sa moyenne.

Le moment centré d'ordre deux,

$$\sigma^2 \equiv m_2' = \mathbb{E}\left[(X-\mu)^2\right],$$

est un *indicateur de dispersion* appelé variance. Sa racine carrée σ, grandeur homogène à la grandeur de base, s'appelle écart type. Lorsque la variable aléatoire est une valeur à un instant donné d'un processus aléatoire, l'expression *moyenne quadratique* est généralement préférée.

Ces deux moments fournissent une partie importante de l'information sur la variable, la totalité si celle-ci peut être considérée comme normale.

Les moments d'ordre supérieur, qui apportent pour les autres variables des précisions supplémentaires sur la forme de la distribution, portent sur la variable centrée réduite, rendue adimensionnelle par division par son écart type.

Le moment d'ordre trois de la variable centrée réduite,

$$m_3' = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right],$$

est un *indicateur d'asymétrie*.

Le moment d'ordre quatre de la variable centrée réduite,

$$m_4' = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right],$$

est un *indicateur d'aplatissement des extrêmes des distributions* appelé kurtosis.

Médiane et quantiles

On appelle médiane d'une variable aléatoire X, un réel m tel que

$$\mathbb{P}(X \leq m) \geq 1/2 \leq \mathbb{P}(X \geq m)$$

Dans le cas d'une variable aléatoire discrète, cette définition est peu intéressante car elle permet l'existence de plusieurs médianes

> si X est le numéro apparaissant sur la face supérieure d'un dé à 6 faces parfaitement équilibré, pour tout réel m strictement compris entre 3 et 4, on a :
>
> $$\mathbb{P}(X \leq m) = \mathbb{P}(X \geq m) = 1/2$$

ou bien l'existence d'une médiane qui ne donne pas une probabilité de 0,5.

> Si X est la somme obtenue en lançant deux dés à 6 faces parfaitement équilibrés. X ne possède qu'une seule médiane 7 mais $\mathbb{P}(X \leq 7) = 21/36$

Dans le cas d'une variable continue, si la fonction de répartition est strictement croissante, la définition est équivalente à la suivante :

> **la** médiane de X est le réel unique m tel que $F_X(m) = 0{,}5$.

Le fait que la fonction de répartition soit continue, et supposée strictement croissante, à valeurs dans $]0\ ;\ 1[$, assure l'existence et l'unicité de la médiane.

Si la médiane a comme valeur $m = 0{,}5$, il est possible cependant de s'intéresser à d'autres valeurs de m (que l'on nomme les quantiles) :

- Quartile : m = 0,25, 0,75
- Décile : m = 0,1, 0,2, 0,3...
- Centile : m = 0,01, 0,02...

Simulation d'une variable aléatoire

On utilise souvent des générateurs pseudo aléatoires pour simuler le hasard. Il existe également des moyens d'exploiter l'indétermination de phénomènes physiques, par exemple en analysant les variations d'un film de lampe à lave, en analysant le bruit thermique, ou mieux encore, en demandant à la nature quantique de jeter des dés pour nous.

Voir aussi

- Convergence de variables aléatoires
- Loi de probabilité à plusieurs variables
- Générateur de nombres pseudo-aléatoires
- Générateur de nombres aléatoires
- Loi de probabilité
- Variables aléatoires élémentaires

Sur les autres projets Wikimedia :

Variables aléatoires continues, sur Wikiversity