Eléments du calcul des probabilités

P. Gribomont

Introduction et motivation

Objectifs d'apprentissage

L'étudiant sera capable de reconnaître des problèmes probabilistes simples et les variables aléatoires correspondantes. Il pourra résoudre ces problèmes et écrire les programmes informatiques appropriés.

Seule la théorie indispensable est introduite, après qu'elle ait été motivée par des exercices.

Références

`http://www.montefiore.ulg.ac.be/~gribomon/cours/cours.html`

Quatre problèmes classiques

Ces problèmes permettent d'appréhender concrètement les notions probabilistes.

Chacun d'eux est résolu de trois manières différentes.

Ils permettent aussi d'introduire les concepts théoriques nécessaires pour la résolution fiable et systématique de problèmes probabilistes plus généraux.

Le problème du casino

On propose à un casino une machine à sous dont le fonctionnement est le suivant. Le joueur introduit sa mise. L'automate jette six pièces. S'il y a trois Pile et trois Face, le joueur récupère trois fois sa mise. Sinon, il perd sa mise. Cette machine est-elle intéressante ?

Nous allons étudier le problème, ainsi que les suivants, par trois méthodes :

- Simulation informatique
- Recherche exhaustive (pour les trois premiers problèmes)
- Analyse combinatoire (pour les trois premiers problèmes)

Problème du casino : simulation informatique

Tous les systèmes de programmation comportent une fonction `random`(n) qui renvoie un nombre choisi au hasard dans l'ensemble des naturels compris entre 0 (inclus) et n (exclu). En codant Pile par 0 et Face par 1, des appels à `random(2)` permettent de simuler des jets de pièces et en particulier de simuler le comportement de l'automate. On dispose d'un programme simulant une série de 300 parties ; il renvoie le nombre de parties gagnées par le joueur. On exécute dix fois le programme et on note les résultats suivants :

$$95, 86, 91, 87, 81, 105, 97, 87, 84, 89\,.$$

Un seul joueur a gagné plus du tiers de parties et a réalisé un bénéfice.

Le jeu est favorable au casino, mais "pas trop" ... ce qui évite le découragement prématuré des joueurs.

Les joueurs ont remporté en tout 902 parties ; la probabilité *approximative* de gain d'une partie par le joueur est de 902/3000, c'est-à-dire 30.07 %.

On *devine* que plus le nombre de parties simulées est grand, meilleure sera l'approximation.

Problème du casino : recherche exhaustive

Le résultat d'une partie se représente par un “mot” de six lettres P (pour Pile) et F (pour Face). On peut, par programme ou même à la main, écrire tous les mots possibles et compter ceux qui comportent trois fois chaque lettre.

Il y a 64 mots possibles, dont 20 comportent trois occurrences de P et trois occurrences de F.

La probabilité *exacte* de gain d'une partie par un joueur est donc de 20/64, soit 31.25 %.

Problème du casino : analyse combinatoire

On sait que le nombre total de mots de n lettres écrits avec un alphabet de p lettres est p^n. Dans notre cas, $n = 6$ et $p = 2$, d'où ce nombre total est $2^6 = 64$.

Le nombre de moyens de choisir trois positions (pour placer des F, par exemple) dans un ensemble de 6 positions est $\mathrm{C}_6^3 = 20$.

On retrouve le résultat précédent $20/64$ ou $31.25\,\%$.

Problème du tricheur

Deux joueurs s'affrontent à pile ou face en utilisant des pièces d'or.

Chaque joueur à son tour fournit la pièce, qui est acquise au gagnant.

A la fin de la partie, durant laquelle 1 000 pièces ont été lancées, le perdant observe qu'il a perdu 643 à 357 (l'écart est donc de 286). Il affirme alors qu'une perte aussi sévère ne peut être due au hasard.

Que penser de cette affirmation ?

Problème du tricheur : simulation informatique

Un programme simule le jet de mille pièces et renvoie l'écart entre les nombres de P et de F. Dix exécutions fournissent les résultats suivants :

$$32, 56, 4, 6, 22, 20, 62, 16, 10, 28\,.$$

Le plus gros écart observé, 62, est loin d'atteindre 286, c'est suspect !

Un second programme appelle m fois le premier et fournit le plus gros écart observé lors des m parties.

Un appel avec $m = 1\,000$ fournit 112 ; un appel avec $m = 10\,000$ fournit 138 ... on est loin du compte ! La probabilité d'obtenir un écart de 286 semble nettement inférieure à 1/10 000 ; la tricherie est donc extrêmement probable !

Problème du tricheur : recherche exhaustive

Même avec un ordinateur puissant, écrire tous les mots de 1 000 lettres sur l'alphabet {F,P} est impossible car $2^{1\,000}$ est un nombre bien plus vaste que le nombre estimé de particules élémentaires dans l'univers observable !

Le nombre $2^{1\,000}$ comporte 302 chiffres ; le nombre de particules élémentaires dans l'univers observable n'est évidemment pas connu mais selon les estimations des astrophysiciens, il comporterait entre 80 et 95 chiffres “seulement”.

Problème du tricheur : analyse combinatoire

Le nombre total de parties distinctes est $2^{1\,000}$

Le nombre de parties comportant exactement i fois Pile (et donc, $1\,000 - i$ fois Face) est $A = \mathrm{C}_{1\,000}^{i}$.

Pour que l'écart soit de 286 ou plus, il faut que $i \leq 357$ ou $i \geq 643$; le nombre de parties "extrêmes" est donc

$$B \;=\; \sum_{i=0}^{357} \mathrm{C}_{1\,000}^{i} + \sum_{i=643}^{1\,000} \mathrm{C}_{1\,000}^{i} \;=\; 2\sum_{i=0}^{357} \mathrm{C}_{1\,000}^{i}\,.$$

La probabilité d'une partie "extrême" est le quotient du second nombre B (qui comporte 283 chiffres) et du premier A (qui en comporte 302). Cette probabilité est donc infime : environ $1.141\,10^{-19}$, soit pas plus que la probabilité de gagner au rang 1 à Euromillions deux fois consécutives !
La tricherie est donc quasi certaine.

Problème de la compétition I

Compétition entre les équipes de M. Lerond (17 membres)
et de Mme Lacroix (13 membres).

Résultat du concours, représenté par un *17-13-mot* :

o o x o x x o x o o o x o o x o o o x o o x o x o x x x x o

Qui a gagné ?

Score de Wilcoxon : des $17 \times 13 = 221$ "duels" simultanés,
les Ronds en remportent 133 et les Croix en remportent 88.

[score des Croix : (+ 1 1 1 1 2 3 5 8 10 13 14 14 15)]

Problème de la compétition II

La victoire des Ronds (133-88 : écart de 45) est-elle significative ?

Un score est temporairement dit *extrême* si l'écart est de 45 au moins ; il est dit *équilibré* si l'écart est de 44 au plus.

Si la proportion de scores extrêmes est faible, on peut penser que la victoire de M. Lerond est significative ; sinon, on est tenté de croire qu'elle est due au hasard.

Compétition : simulation

Un programme construit un 17-13-mot au hasard et calcule son score de Wilcoxon. On constate que, assez fréquemment, ce score est extrême, c'est-à-dire que l'écart entre l'équipe gagnante et l'équipe perdante est de 45 au moins.

Essais : en appelant le programme
- dix fois, on obtient deux scores extrêmes ;
- cent fois, on obtient 43 scores extrêmes ;
- mille fois, on obtient 361 scores extrêmes.

Une probabilité estimée à 36.1 % n'est en rien négligeable.
La victoire des Ronds est “statistiquement non significative”.

Compétition : recherche exhaustive

Un programme génère l'ensemble des 17-13-mots et calcule les scores correspondants ; il y a exactement 119 759 850 mots, parmi lesquels 43 433 256 donnent lieu à des scores “extrêmes”.

Les scores “extrêmes” se produisent fréquemment ; ils ne méritent pas leur nom et ne sont donc pas réellement significatifs. Mme Lacroix peut ainsi prétendre que la victoire de M. Lerond est due au hasard.

Le rapport 43 433 256 / 119 759 850 équivaut à 36.27 %

Compétition : analyse combinatoire I

Méthode exhaustive : impraticable dès que n et p deviennent assez grands : le nombre de n-p-mots devient énorme.
Analyse combinatoire : solution simple et générale.
On définit la fonction numérique w à trois arguments : $w(n, p, i)$ est le nombre de n-p-mots dont le score de Wilcoxon est i pour les Croix (et donc $np - i$ pour les Ronds).

- $w(n, p, i) = 0$ si $i < 0$ ou $i > np$;
- $w(n, 0, 0) = 1 = w(0, p, 0)$;
- $w(n, p, i) = w(n, p, np - i) = w(p, n, i) = w(p, n, np - i)$ (symétries).
- $\sum_{i=0}^{np} w(n, p, i) = \mathrm{C}_{n+p}^{n} = \mathrm{C}_{n+p}^{p}$

Compétition : analyse combinatoire II

Comment calculer les valeurs non triviales $w(n,p,i)$: $n,p>0$, $0<i<np$?

Un n-p-mot de score i (pour les Croix) commence par x ou par o.
Dans le premier cas, le suffixe est un n-$(p-1)$-mot de score $i-n$ et dans le second, un $(n-1)$-p-mot de score i ; on a donc, si $0<i<np$,

$$w(n,p,i) = w(n,p-1,i-n) + w(n-1,p,i)$$

et la fonction w se tabule par "couches" indexées, la couche d'index r contenant les éléments $w(n,p,i)$ tels que $n+p=r$. On observe que les éléments de la couche r s'obtiennent facilement à partir de ceux de la couche $r-1$.

Phénomènes aléatoires continus

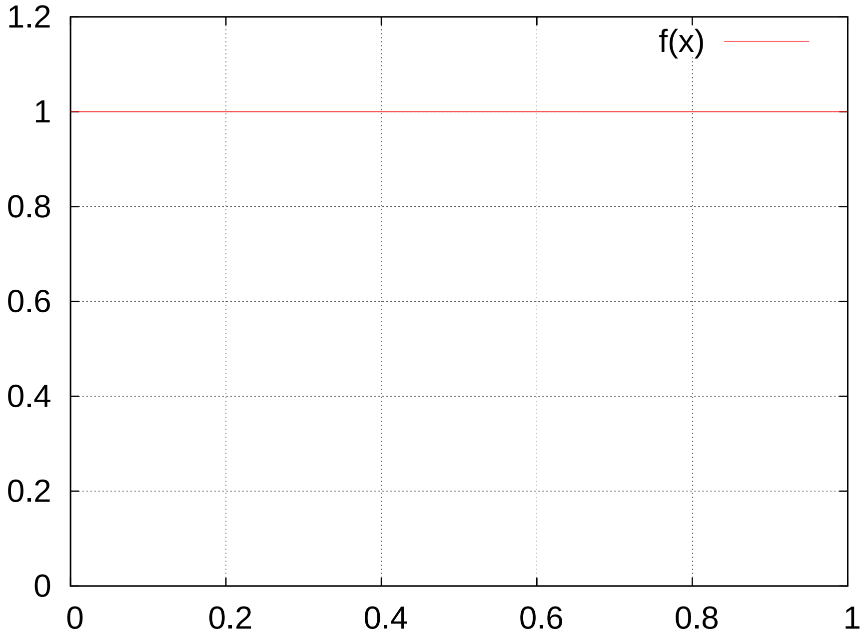
Les trois premiers problèmes concernaient des phénomènes aléatoires discrets, dont chaque occurrence peut être représentée par un nombre entier. Un jet de pièce peut être représenté par exemple par 0 (Pile) ou 1 (Face) ; le jet d'un dé par un nombre de 1 à 6 ; le score de Wilcoxon d'un 17-13-mot par un nombre entre 0 et 221.

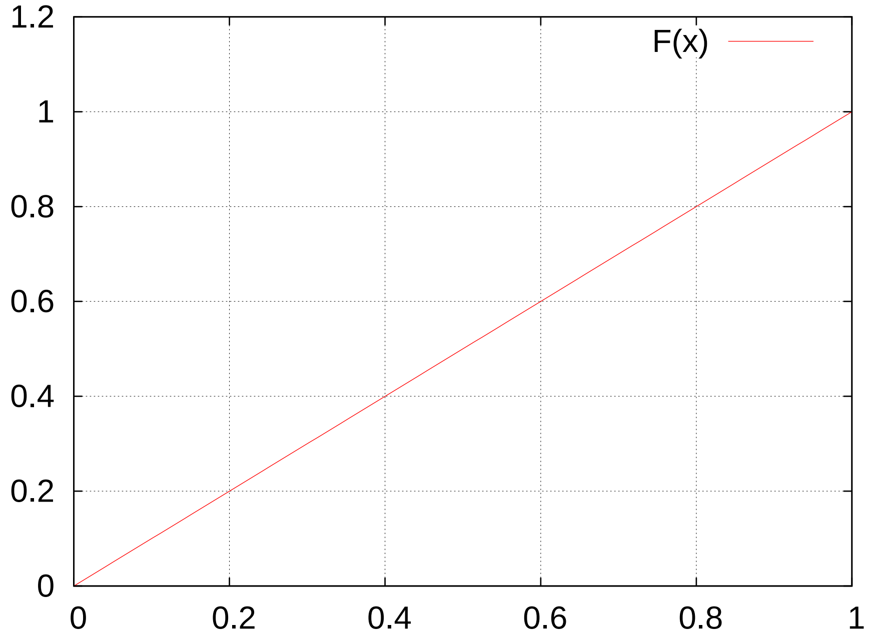
L'issue de certains phénomènes aléatoires est représentée plutôt par un nombre réel, par exemple la taille d'un individu choisi au hasard dans une certaine population, ou la durée de vie d'une lampe d'un certain type, etc.

On peut étudier de tels phénomènes par simulation informatique, en utilisant une fonction `random` sans argument qui renvoie un nombre quelconque r dans l'intervalle semi-ouvert $[0:1[$. On suppose qu'à chaque appel, la probabilité que le nombre renvoyé soit dans le sous-intervalle $[a:b[$ (avec $0 \leq a \leq b < 1$) est exactement $b - a$.

Test du générateur aléatoire I

On se propose de tester le fait que, dans un système donné, `random` se comporte conformément à ses spécifications, résumées en la *distribution de probabilité* et la *fonction de répartition*.





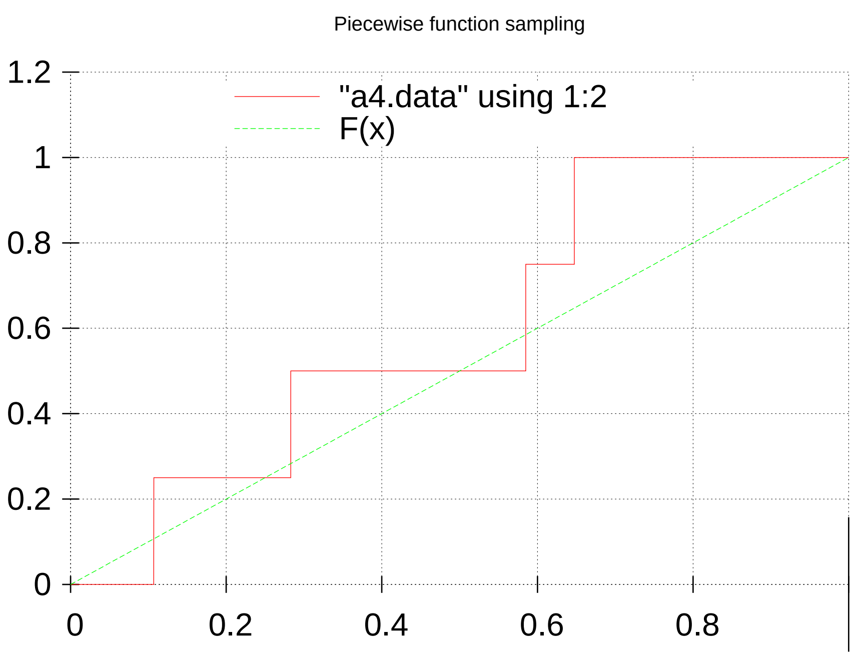
$$\int_a^b f(x)\,\mathrm{d}x \;=\; \Pr(a \leq X < b) \;=\; F(b) - F(a)\,.$$

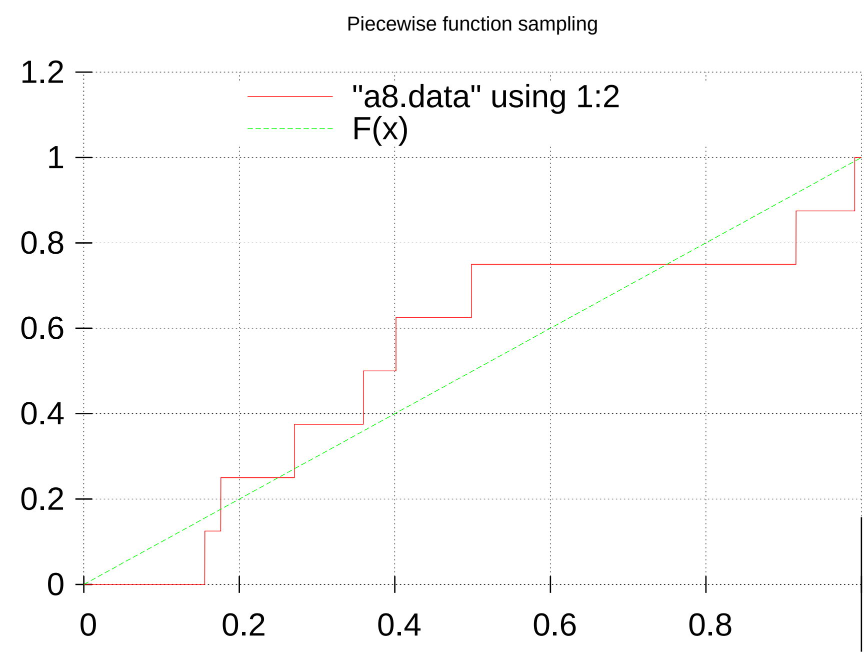
f : distribution de probabilités

F : fonction de répartition

Test du générateur aléatoire II

Fonctions de répartition correpondant à des ensembles de $n = 4$ et $n = 8$ nombres issus du générateur aléatoire. Les abscisses des sauts correspondent aux n nombres ; les ordonnées sont les nombres k/n pour $k = 0, 1, \ldots, n$.

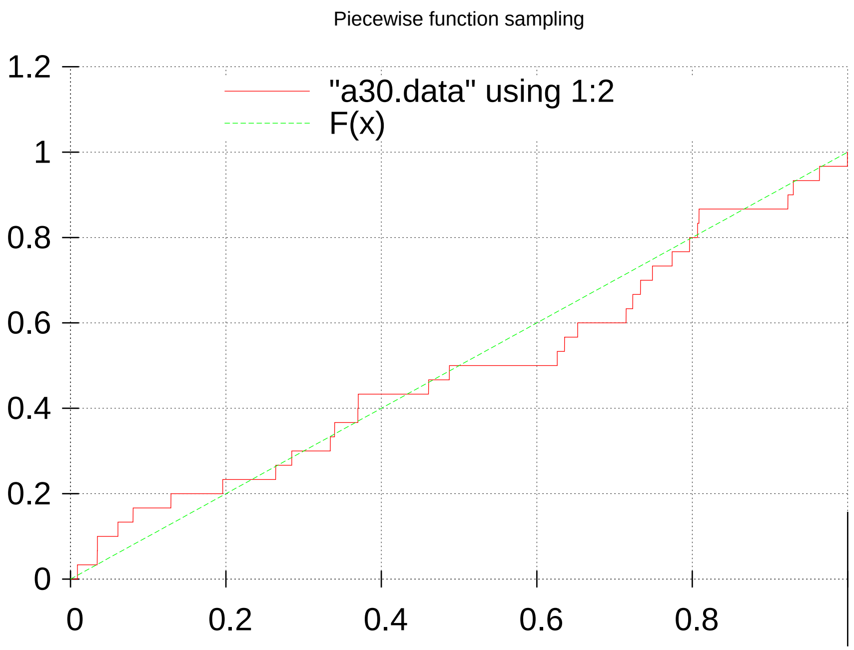


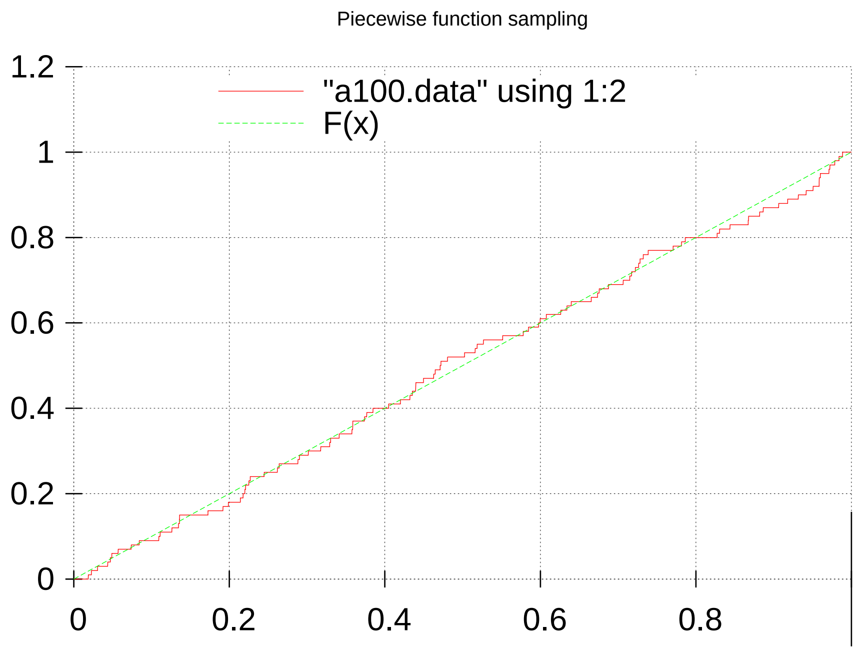


On observe que ces fonctions de répartition approximent grossièrement la fonction de répartition théorique.

Test du générateur aléatoire III

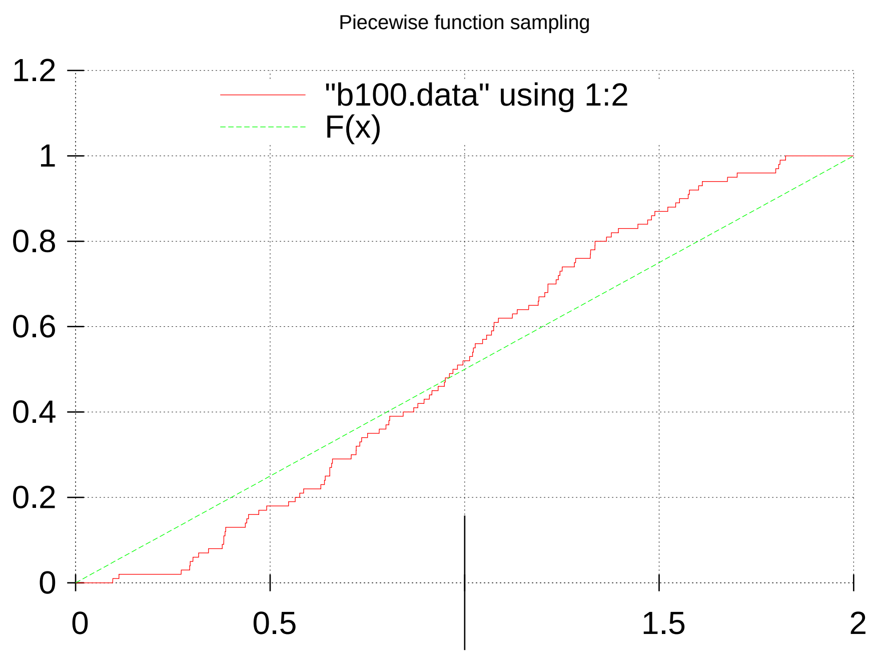
L'approximation devient meilleure quand n augmente (ici, $n = 30$ et $n = 100$).

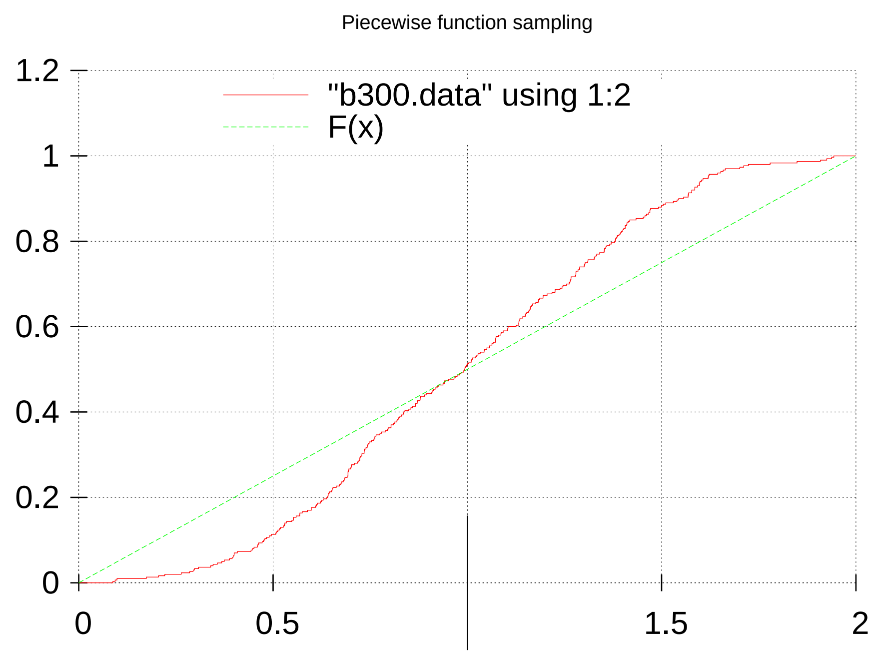




Un problème continu simple I

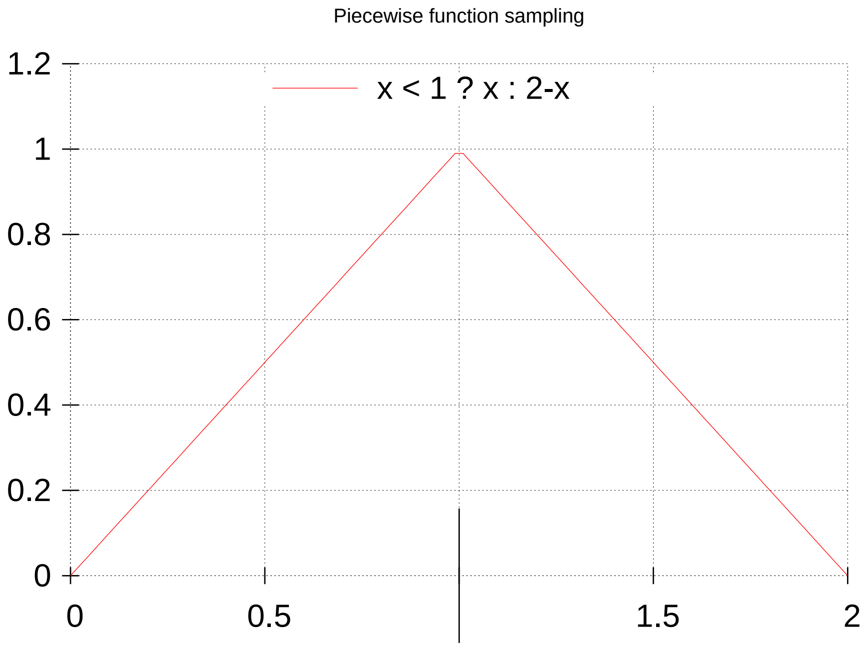
On voudrait obtenir un générateur de nombres aléatoires (toujours uniformément répartis) compris entre 0 et 2. Une solution évidente consiste à doubler les valeurs fournies par `random`. Quelqu'un propose de plutôt additionner deux à deux les valeurs fournies par `random`. Est-ce acceptable ? La simulation avec 100 paires de nombres fournis par `random` semble indiquer que non, ce qu'une nouvelle simulation avec 300 paires confirme :

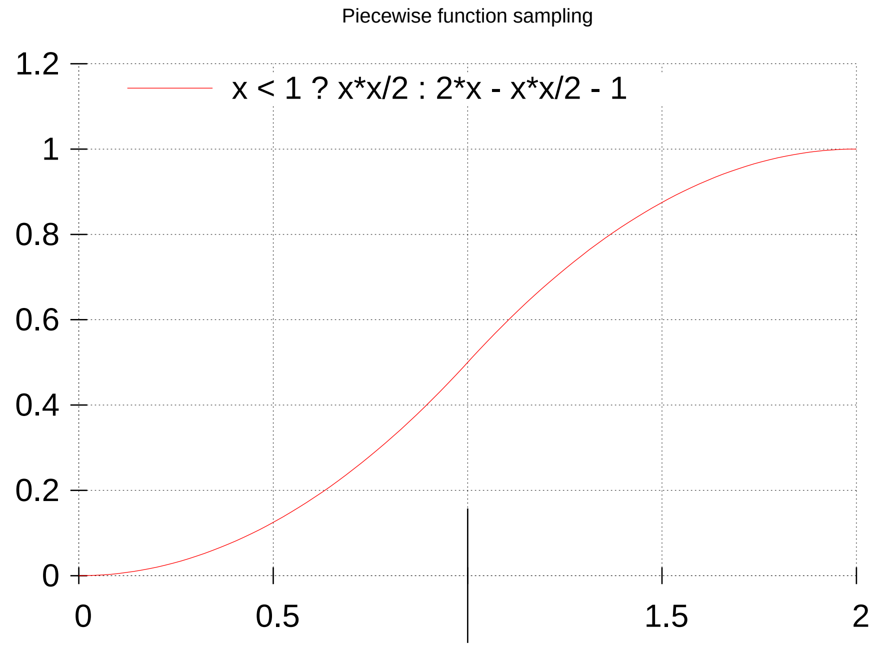




Un problème continu simple II

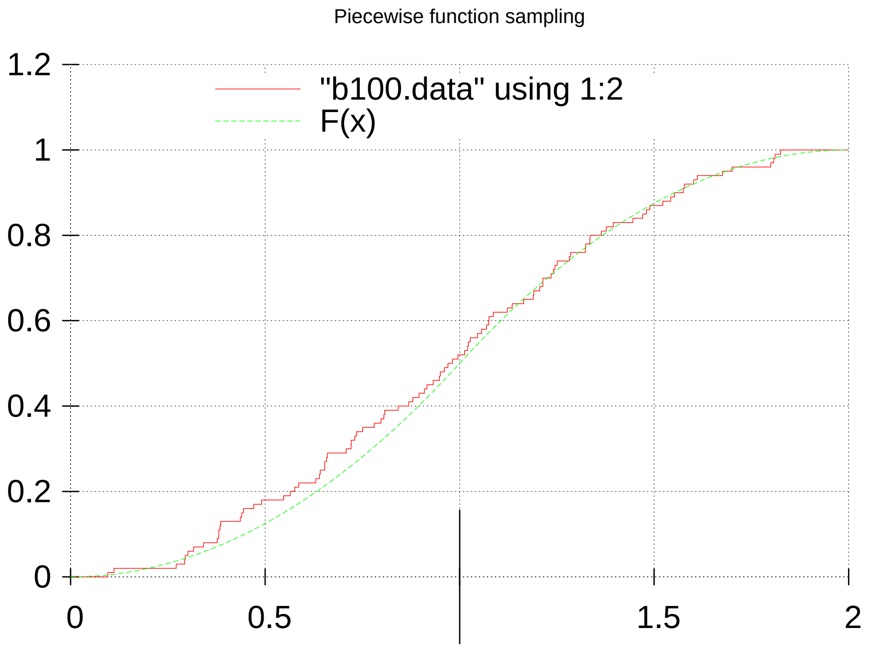
En anticipant sur une étude théorique : la somme de deux variables aléatoires indépendantes uniformément distribuées sur [0 : 1] n'est pas uniformément distribuée mais admet la distribution de probabilité et la fonction de répartition ci-dessous :

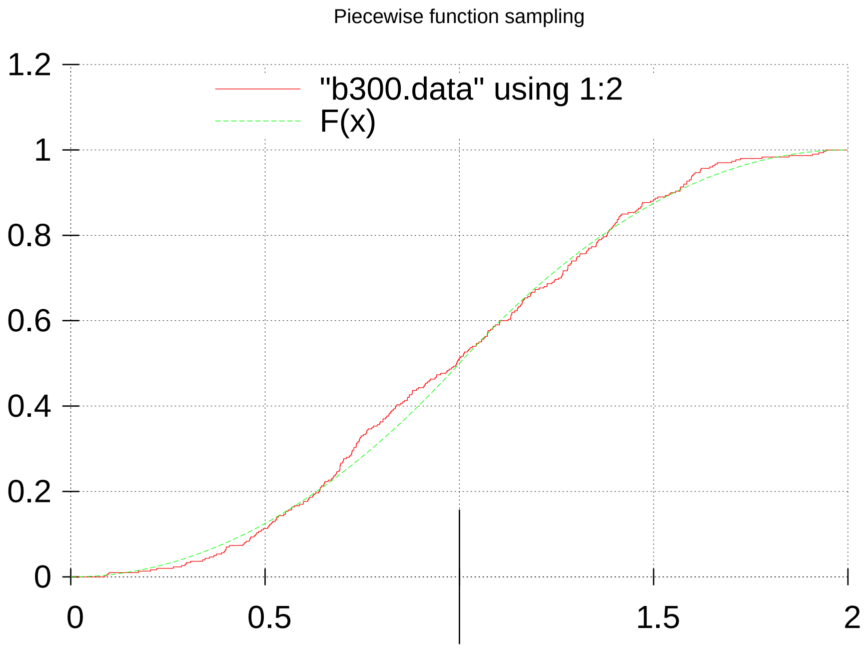




Un problème continu simple III

Ceci est bien en accord avec les simulations :





Bilan : besoins en théorie

— Vocabulaire des probabilités ;
— Analyse combinatoire (problèmes discrets) ;
— Analyse mathématique (problèmes continus) ;
— Distributions de probabilité (théoriques et expérimentales) ;
— Fonctions de répartition (théoriques et expérimentales) ;
— Liens entre théorie et expérimentation/observation.

Vocabulaire des probabilités I

L'étude de tout phénomène aléatoire présuppose un *espace*, c'est-à-dire un ensemble dont les éléments représentent les issues possibles du phénomène. L'espace peut être $\{P, F\}$ (jet d'une pièce), ou l'ensemble des mots de six lettres, ou de mille lettres, sur l'alphabet $\{P, F\}$ (problème du casino, problème du tricheur), ou l'ensemble des 13-17-mots de Wilcoxon (problème de la compétition).

Chaque sous-ensemble de l'espace est un *événement*. Par exemple, dans le problème du casino, l'événement "gain du joueur" est un sous-ensemble de taille 20 d'un espace comportant 64 éléments.

La *probabilité* d'un événement est (souvent) le rapport de la *mesure* de l'événement par rapport à la mesure de l'espace, ou simplement la mesure de l'événement si on suppose que la mesure de l'espace est 1.

Vocabulaire des probabilités II

Un événement est un ensemble : on peut utiliser les opérations ensemblistes (intersection, réunion, différence, complémentation) pour définir des événements compliqués à partir d'événements simples.

Les espaces évoqués jusqu'ici sont finis et les événements *élémentaires* (correspondant à des singletons) ont été supposés *équiprobables* ; ce sera souvent le cas en pratique pour les espaces finis, avec un grand avantage : les problèmes de probabilités se ramènent à des problèmes d'analyse combinatoire.

Si l'espace est infini, dénombrable ou non, cet avantage disparaît.

Vocabulaire des probabilités III

A chaque événement E est associée une *probabilité*, notée $P(E)$. Quelles conditions impose-t-on à la fonction P ? On souhaite que $P(E)$ corresponde à la fréquence de E, censée converger quand le phénomène se répète. On observe une telle convergence en pratique mais il est délicat de l'exiger d'emblée. Cependant, on peut exiger ceci de toute notion de probabilité :

— $0 \leq P(E) \leq 1$;
— si les E_i sont mutuellement exclusifs, deux à deux disjoints, alors $P(\bigcup_i E_i) = \sum_i P(E_i)$;
— si S est l'espace entier, alors $P(S) = 1$.

Conséquences (à démontrer) :

— Si $E_1 \subset E_2$, alors $P(E_1) \leq P(E_2)$;
— $P(\emptyset) = 0$, $P(\overline{E}) = 1 - P(E)$;
— $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$.

Vocabulaire des probabilités IV

L'étude d'un phénomène aléatoire peut donner lieu à l'introduction de diverses *variables aléatoires*, dont la valeur (souvent numérique) dépend fonctionnellement de la réalisation du phénomène (exemples du casino et du tricheur). Plutôt que d'étudier séparément chaque variable aléatoire, on peut développer une théorie qui permettrait, à partir des propriétés (évidentes) du phénomène "jet d'une pièce", d'obtenir les propriétés des v.a. correspondant au jet de six ou mille pièces, ou à la variable "nombre d'essais nécessaires pour obtenir deux fois Face", etc. Cela implique l'étude des *probabilités conditionnelles*.

Analyse combinatoire I

Il s'agit de l'art (science, technique, ...) de dénombrer des ensembles, finis mais souvent très grands, sans devoir énumérer les éléments. Cette branche des mathématiques comporte des techniques difficiles mais nous n'aurons besoin que de techniques faciles dans le cadre de ce cours.

La taille de l'ensemble $A \times B$ est le produit des tailles des ensembles A et B : $|A \times B| = |A|.|B|$.

Un premier résultat de base concerne les permutations : un ensemble de n éléments comporte $n!$ permutations.

Un second résultat de base concerne les combinaisons : un ensemble de n éléments admet C_n^p sous-ensembles de p éléments (on suppose $0 \leq p \leq n$), avec

$$\mathsf{C}_n^p = \binom{n}{p} = \frac{n!}{p!(n-p)!}$$

Analyse combinatoire II

Le second résultat découle du premier : si on coupe une permutation en un préfixe de taille p et un suffixe de taille $n-p$, on détermine un sous-ensemble de taille p, en considérant que les éléments du préfixe sont ceux du sous-ensemble. Cependant, chaque sous-ensemble est obtenu $p!(n-p)!$ fois, puisque le préfixe peut être ordonné de $p!$ manières différentes et le suffixe de $(n-p)!$ manières différentes.

Ces résultats peuvent aussi se démontrer par récurrence. Notamment, à partir d'une permutation d'un ensemble E de n éléments, on obtient $n+1$ permutations de l'ensemble $E \cup \{x\}$ de $n+1$ éléments, puisque x peut s'insérer de $n+1$ manières différentes. Cela correspond à $(n+1)! = n!(n+1)$.

Analyse combinatoire III

Quelques résultats importants sur les combinaisons (ou coefficients binomiaux) :

$$\mathrm{C}_n^0 = \mathrm{C}_n^n = 1\,,\ \mathrm{C}_n^p = \mathrm{C}_n^{n-p}\,;$$

$$\text{Si } 0 < p < n, \text{ alors } \mathrm{C}_n^p = \mathrm{C}_{n-1}^p + \mathrm{C}_{n-1}^{p-1}\,;$$

$$\text{Si } 0 < p \leq n, \text{ alors } \mathrm{C}_n^p = n\mathrm{C}_{n-1}^{p-1}/p\,;$$

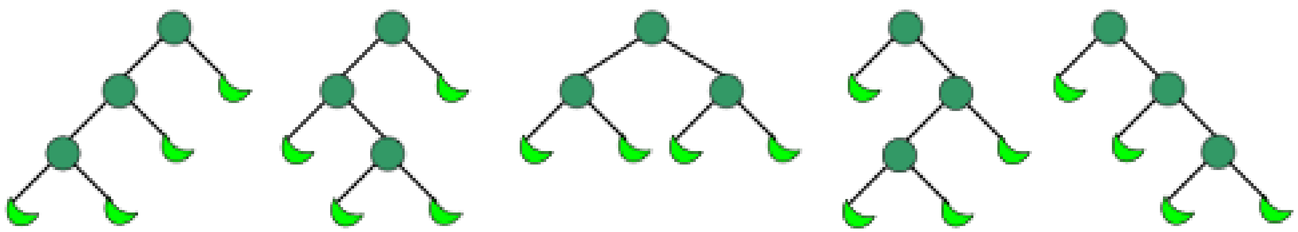
$$(1+x)^n \ = \ \sum_{p=0}^{n} \mathrm{C}_n^p x^p\,.$$

Analyse combinatoire IV

Comptage de parenthèses, de groupements, d'arbres . . .

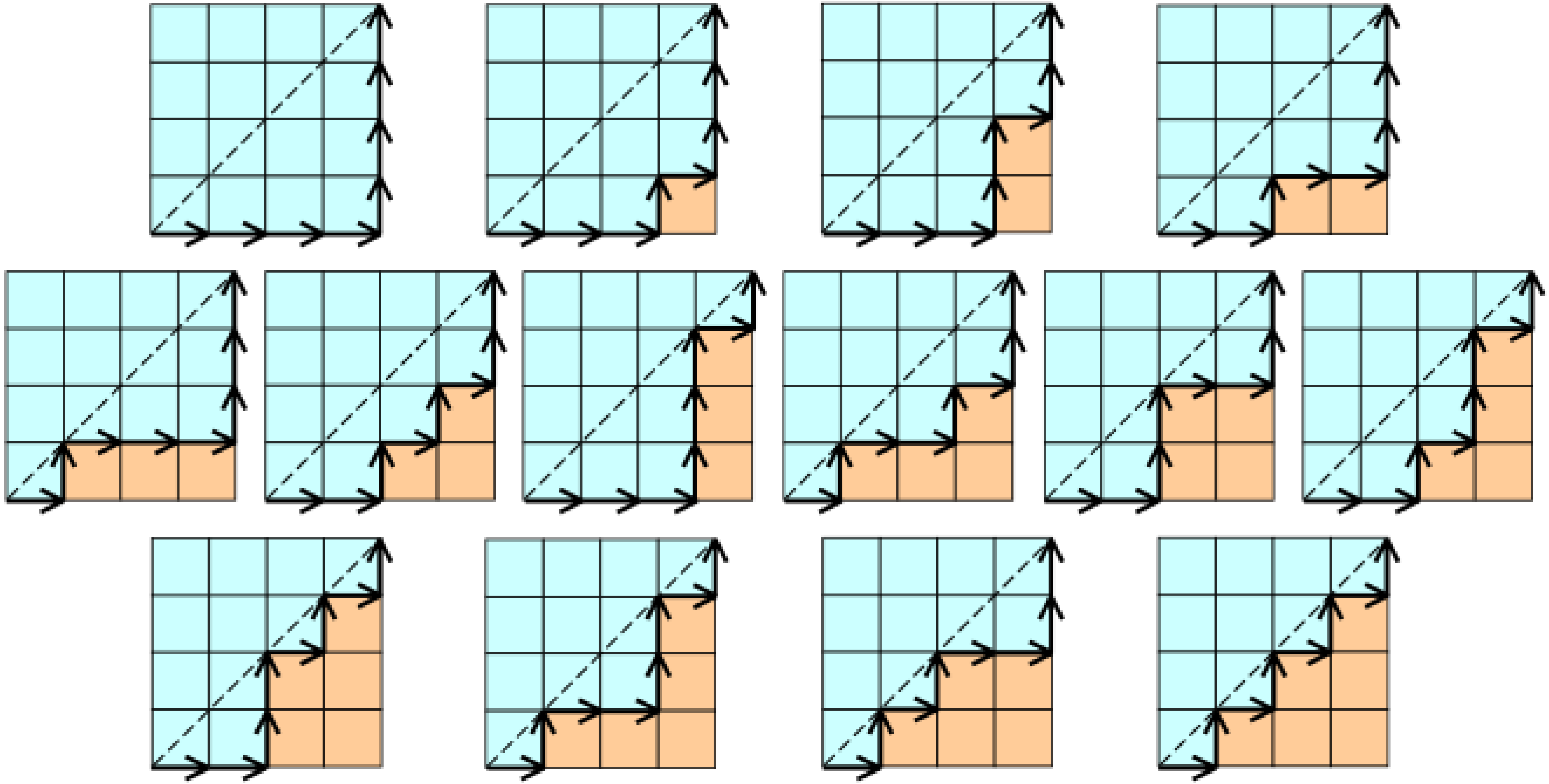
$$((()))\quad ()(())\quad ()()()\quad (())()\quad (()())$$

$$((ab)c)d\quad (a(bc))d\quad (ab)(cd)\quad a((bc)d)\quad a(b(cd))$$

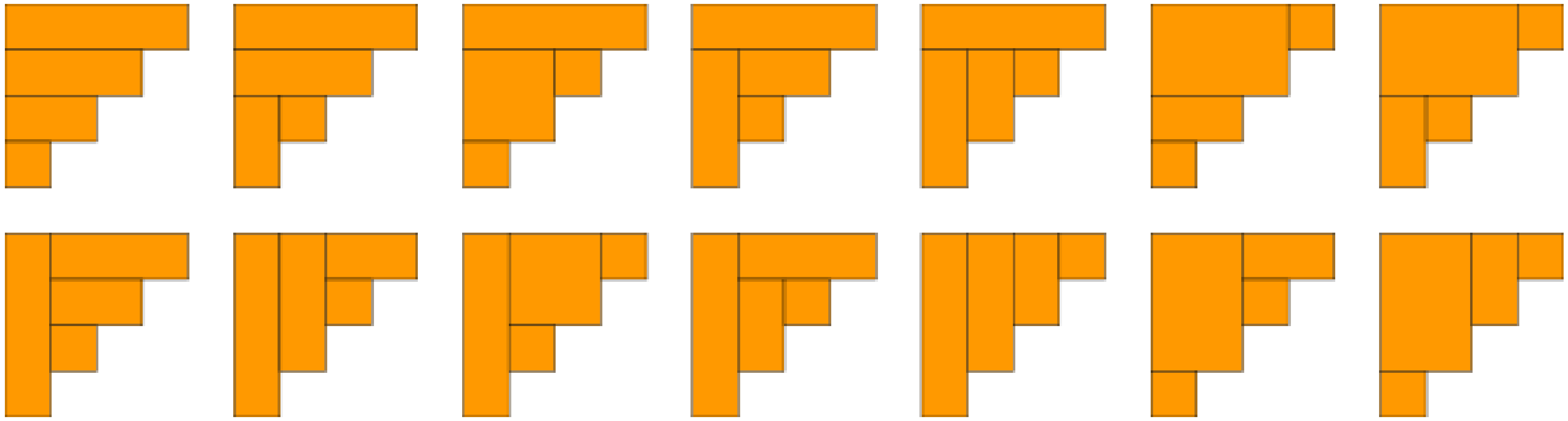


`http://en.wikipedia.org/wiki/Catalan_number`

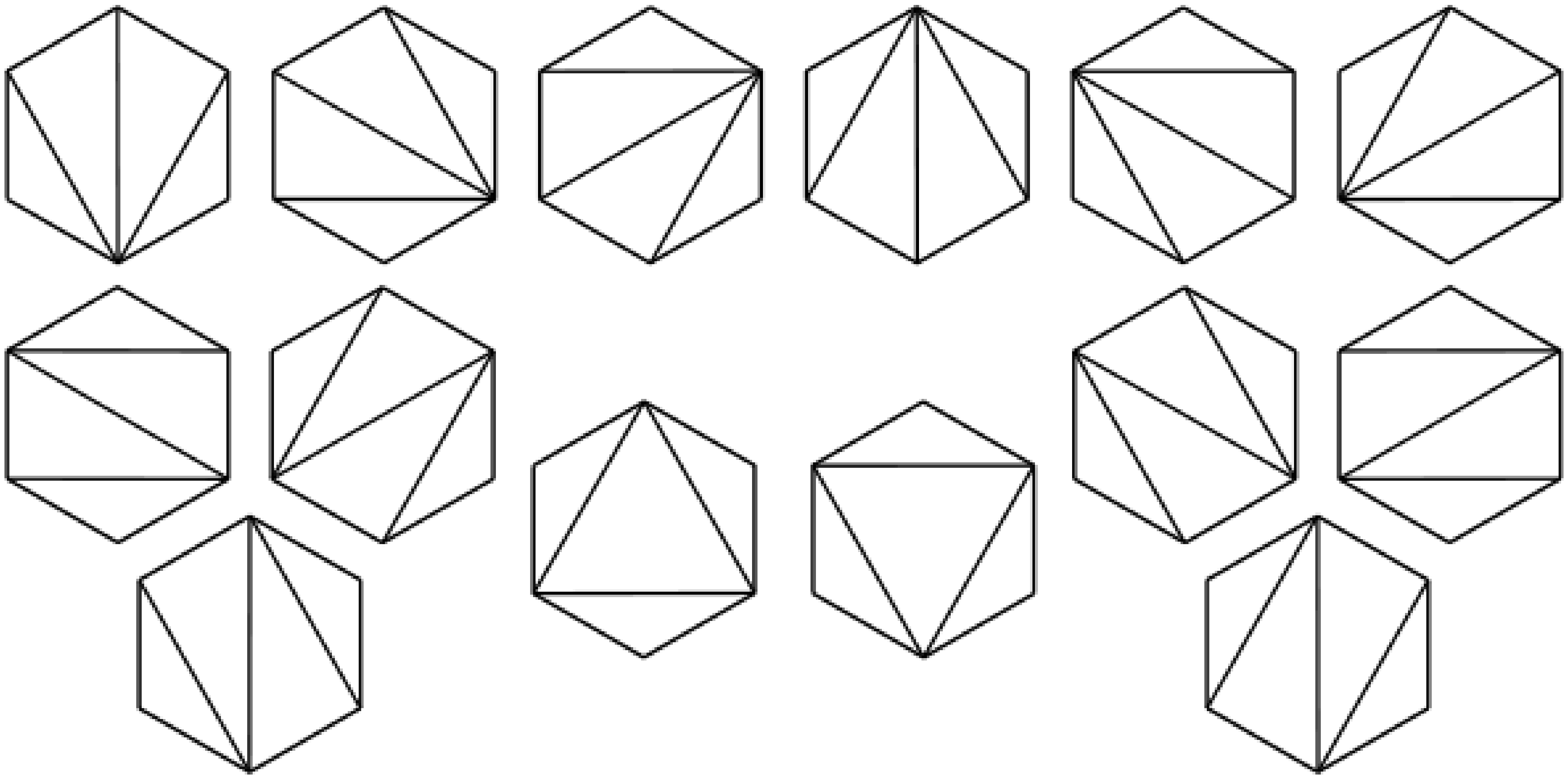
Analyse combinatoire V



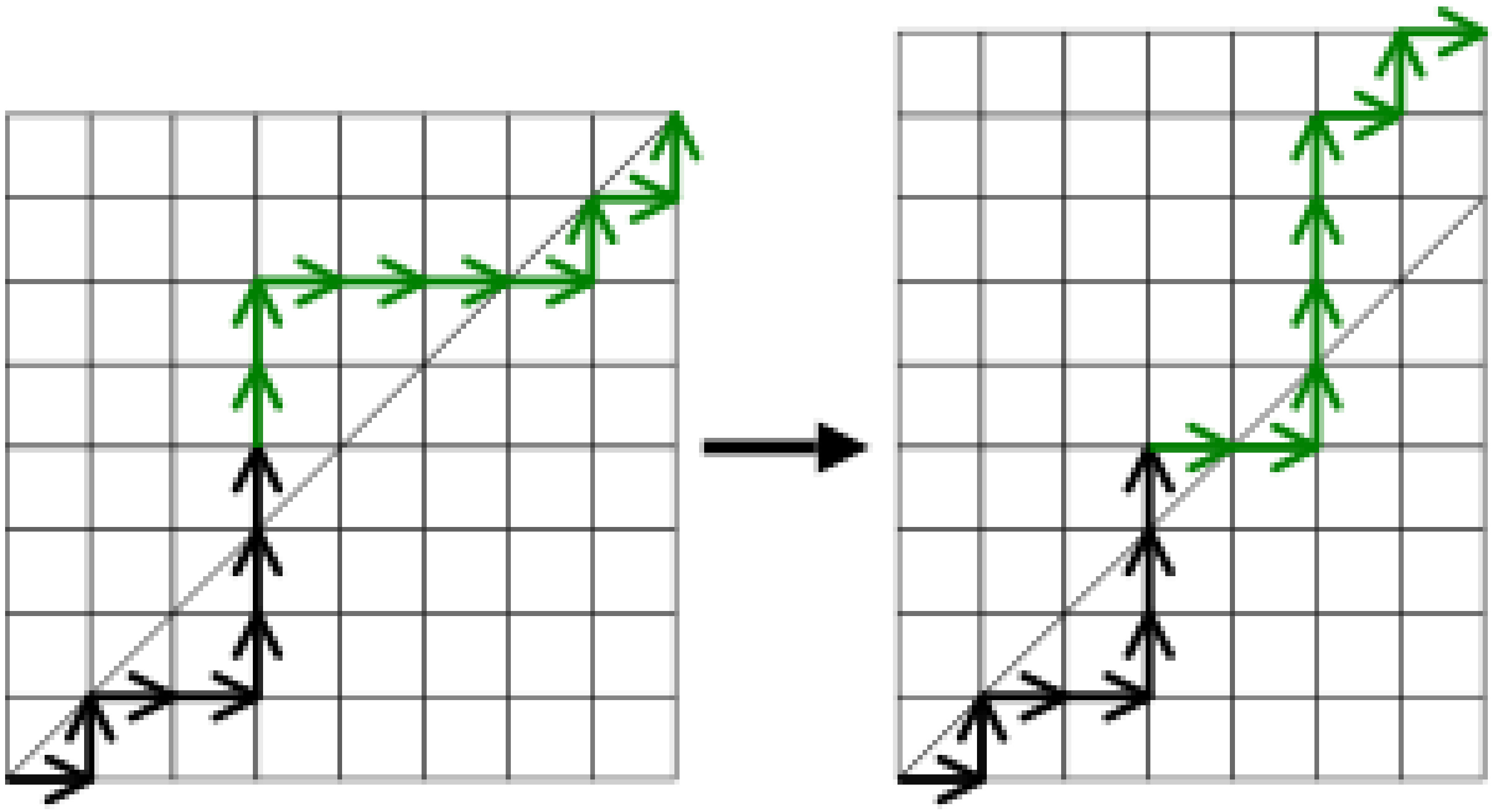
Analyse combinatoire VI



Analyse combinatoire VII



Analyse combinatoire VIII



$$C_n = C_{2n}^{n} - C_{2n}^{n-1}.$$

Analyse combinatoire IX

Exercices :

$$\mathrm{C}_n \;=\; \mathrm{C}_{2n}^{n}/(n+1)\,;$$

$$\mathrm{C}_0 \;=\; 1\,,\; \mathrm{C}_{n+1} \;=\; (4n+2)\mathrm{C}_n/(n+2)\,;$$

$$\mathrm{C}_{n+1} \;=\; \sum_{i=0}^{n} \mathrm{C}_i \mathrm{C}_{n-i}\,.$$

Analyse combinatoire X

S'il existe des billes de n couleurs, C_n^p est le nombre de possibilités de composer un sac de p billes de couleurs deux à deux distinctes.
Si on autorise les répétitions de couleurs, le nombre de possibilités, noté K_n^p, devient

$$\mathsf{C}_{n+p-1}^p = \mathsf{C}_{n+p-1}^{n-1}.$$

En effet, un sac de billes peut être représenté par une chaîne de bits comportant p '0' (les billes) et $n-1$ '1' (les séparateurs), le nombre de '0' précédant immédiatement le $i$ème '1' représentant le nombre de billes de la $i$ème couleur. (NB. On n'impose plus $p \leq n$.)

Le nombre K_n^p est aussi le nombre de listes de naturels de longueur n dont la somme vaut p, ou encore le nombre de solutions dans N de l'équation $x_1 + \cdots + x_n = p$. Si on impose que chaque x_i soit non nul (ce qui implique $p \geq n$), le nombre de solutions est $\mathsf{K}_n^{p-n} = \mathsf{C}_{p-1}^{p-n} = \mathsf{C}_{p-1}^{n-1}$.

Analyse combinatoire XI

Le nombre de manières de répartir un ensemble de n objets (distincts) en p groupes de tailles respectives $n_1, \ldots, n_p$ (avec $n = n_1 + \cdots + n_p$) est donné par les coefficients multinomiaux :

$$\mathsf{C}_n^{n_1,\ldots,n_p} = \frac{n!}{n_1! \ldots n_p!}.$$

La démonstration directe est évidente ; on peut aussi développer la relation

$$\mathsf{C}_n^{n_1,\ldots,n_p} = \mathsf{C}_n^{n_1} \mathsf{C}_{n-n_1}^{n_2,\ldots,n_p},$$

elle aussi évidente.

Le cas $p = 2$ correspond aux coefficients binomiaux ; $\mathsf{C}_n^{k,n-k}$ est simplement abrégé en C_n^k.

Analyse combinatoire XII

De combien de manières peut-on répartir un ensemble de n objets (distincts) en p groupes (non vides) ? Les nombres $D(n,p)$ vérifient clairement ceci :

$$D(n,0) = 0 \ [n > 0] \ , \ D(n,n) = 1 \ ;$$

$$D(n,k) \ = \ D(n-1,k-1) + kD(n-1,k) \ [n \geq k > 0] \, .$$

ce qui permet un calcul facile.

Quel est le nombre B_n de partitions d'un ensemble de n objets ? On a :

$$B_0 = 1 \ , \ B_n \ = \ \sum_{k=1}^{n} D(n,k) \quad [n > 0] \, .$$

$$B_0 = 1 \ , \ B_{n+1} \ = \ \sum_{k=0}^{n} \mathrm{C}_n^k B_{n-k} \ = \ \sum_{k=0}^{n} \mathrm{C}_n^k B_k \, .$$

Principe d'inclusion-exclusion I

La formule $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$ se généralise à un nombre quelconque d'événements. On a par exemple

$$\begin{aligned} P(E_1 \cup E_2 \cup E_3) = & \quad [P(E_1) + P(E_2) + P(E_3)] \\ & - \; [P(E_1 \cap E_2) + P(E_1 \cap E_3) + P(E_2 \cap E_3)] \\ & + \; P(E_1 \cap E_2 \cap E_3)\,. \end{aligned}$$

Les diagrammes de Venn rendent ces formules "graphiquement évidentes".

La forme générale peut s'écrire :

$$\begin{aligned} P(E_1 \cup E_2 \cup \cdots \cup E_n) = & \quad [P(E_1) + \cdots + P(E_n)] \\ & - \; [P(E_1 \cap E_2) + P(E_1 \cap E_3) + \cdots + P(E_{n-1} \cap E_n)] \\ & + \; [P(E_1 \cap E_2 \cap E_3) + \cdots + P(E_{n-2} \cap E_{n-1} \cap E_n)] \\ & \;\vdots \quad \vdots \\ & - \; (-1)^n P(E_1 \cap E_2 \cap \cdots \cap E_n)\,. \end{aligned}$$

On observe que, dans le membre de droite, la $i$ème ligne comporte C_n^i termes ; elle est positive si i est impair et négative si i est pair.

Principe d'inclusion-exclusion II

La preuve par récurrence sur n est facile mais laborieuse à cause des notations, aussi, plutôt que de traiter le cas inductif général (passage de n à $n+1$) on montre seulement le passage de 2 à 3 :

$$P(E_1 \cup E_2 \cup E_3) = P(E_1 \cup E_2) + P(E_3) - P((E_1 \cup E_2) \cap E_3) =$$
$$P(E_1) + P(E_2) - P(E_1 \cap E_2) + P(E_3) - P((E_1 \cap E_3) \cup (E_2 \cap E_3)) =$$
$$P(E_1)+P(E_2)-P(E_1 \cap E_2)+P(E_3)-P(E_1 \cap E_3)-P(E_2 \cap E_3)+P(E_1 \cap E_2 \cap E_3)\,.$$

Probabilités conditionnelles

$$P(A|B) \;=\; P(A\cap B)/P(B)\,;\; P(B)P(A|B) = P(A\cap B)\,.$$

$$P(E_1\cap E_2\cap\cdots\cap E_n) \;=\; P(E_1)P(E_2|E_1)P(E_3|E_1\cap E_2)\cdots P(E_n|E_1\cap\cdots\cap E_{n-1})\,.$$

$$P(E) = P(E|F)P(F) + P(E|\overline{F})P(\overline{F})\,.$$

Si les F_i partitionnent l'espace,

$$P(E) = \sum_i P(E|F_i)P(F_i)\,.$$

Formule de Bayes

Si les F_i partitionnent S, alors

$$P(E) \;=\; \sum_i P(E|F_i)P(F_i)\,.$$

On en déduit la formule

$$P(F_j|E) \;=\; \frac{P(F_j \cap E)}{P(E)} \;=\; \frac{P(E|F_j)P(F_j)}{\sum_i P(E|F_i)P(F_i)}\,.$$

Evénements indépendants

E et F *indépendants* si $P(E \cap F) = P(E)P(F)$.

Dans ce cas, $P(E|F) = P(E)$, $P(F|E) = P(F)$.

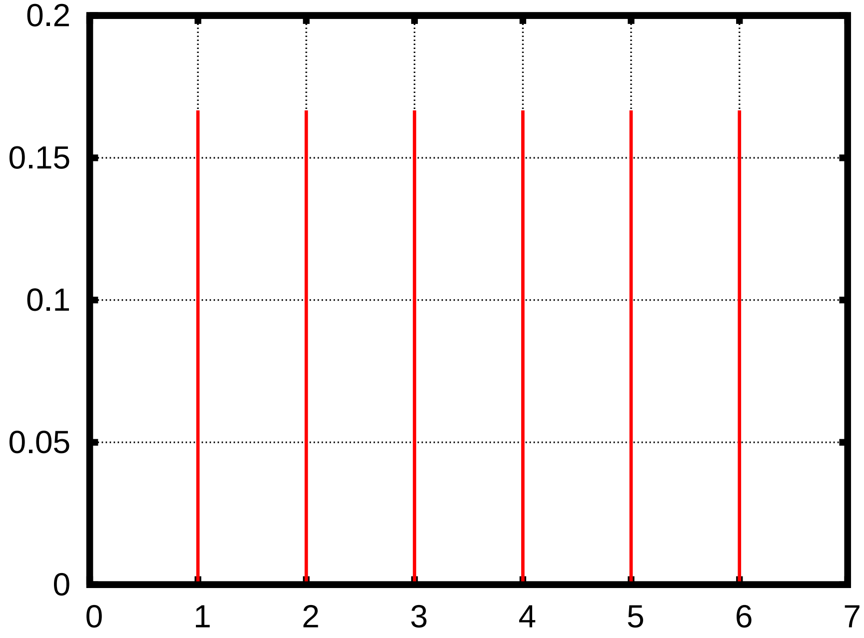
Les trois cartes

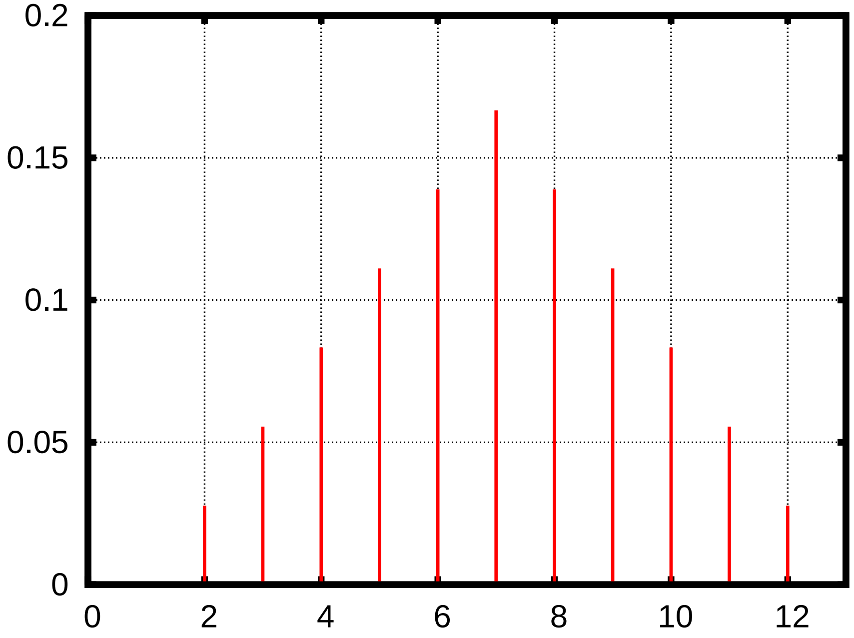
Trois cartes : une à deux faces rouges, une à deux faces noires et une à une face de chaque couleur. Une carte est choisie au hasard et posée. On note une face rouge. Quelle est la probabilité que l'autre face soit noire ?

$$
\begin{aligned}
P(RB|R) &= \frac{P(RB \cap R)}{P(R)} \\
&= \frac{P(R|RB)P(RB)}{P(R|RR)P(RR)+P(R|RB)P(RB)+P(R|BB)P(BB)} \\
&= \frac{\frac{1}{2} \cdot \frac{1}{3}}{1\frac{1}{3}+\frac{1}{2} \cdot \frac{1}{3}+0\frac{1}{3}} \\
&= \frac{1}{3}
\end{aligned}
$$

Distributions de probabilité, cas discret I

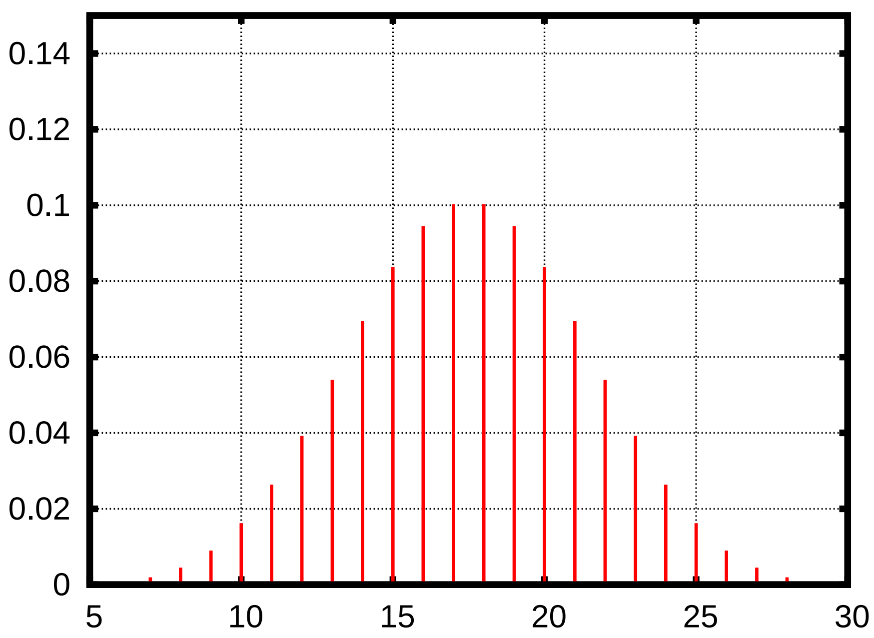
Variables aléatoires, points relatifs au jet d'un et de deux dés.

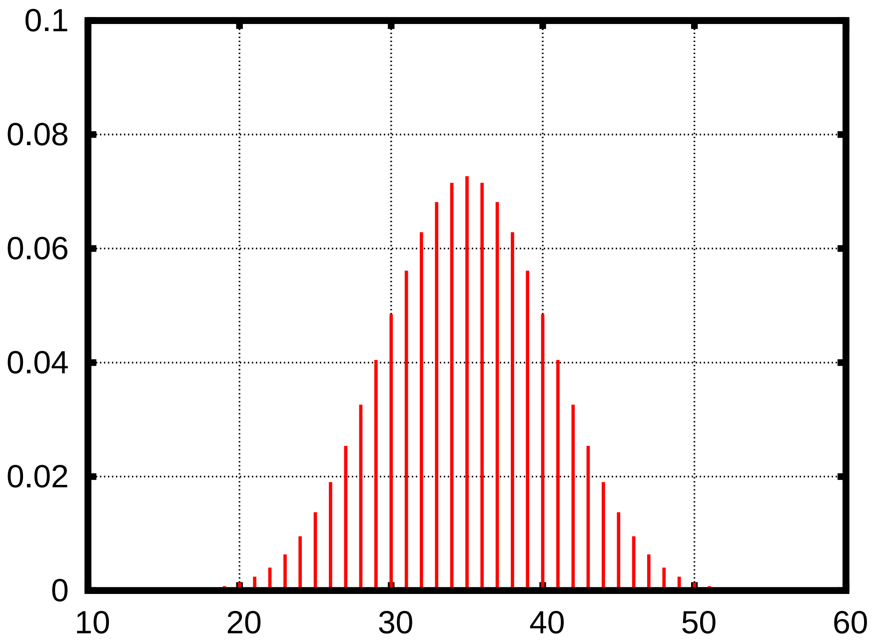




Distributions de probabilité, cas discret II

Variables aléatoires, points relatifs au jet de cinq ou de dix dés.





Distribution binomiale (Bernoulli)

Elle dépend de deux paramètres, n et p.

Probabilité de succès d'un essai : p.

Nombre de succès en n essais indépendants : X

$$P(X = i) \;=\; \mathrm{C}_n^i \, p^i (1-p)^{n-i} \, .$$

L'expression "distribution de Bernouilli" est employée pour le cas $n = 1$;
On a alors $P(X = 0) = 1 - p$ et $P(X = 1) = p$.

Distribution de Poisson I

"Evénements rares"

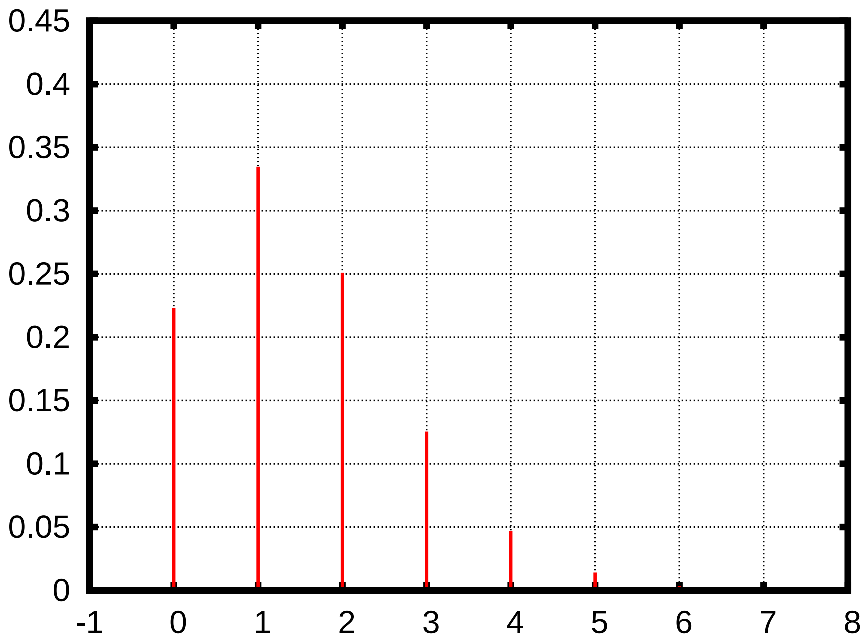
$$P(X = i) \;=\; e^{-\lambda}\,\frac{\lambda^i}{i!}\,.$$

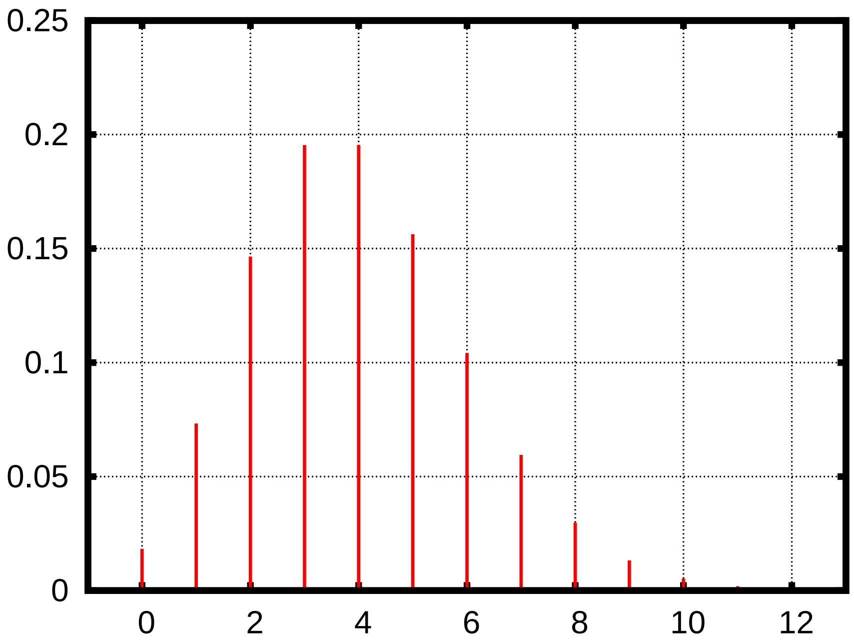
X désigne par exemple le nombre d'accidents par an sur telle route, si en moyenne ce nombre est λ, le paramètre (réel positif) de la distribution.

Additivité. Si X et Y sont indépendantes et admettent des distributions de Poisson de paramètres respectifs λ et μ, alors $X + Y$ admet une distribution de Poisson de paramètre $\lambda + \mu$.

Distribution de Poisson II

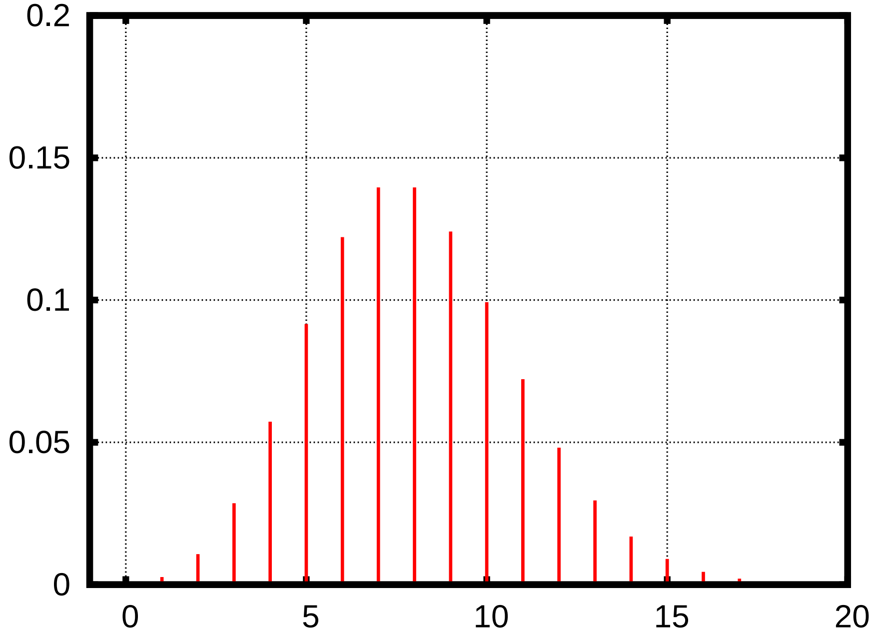
Paramètre 1.5 et paramètre 4

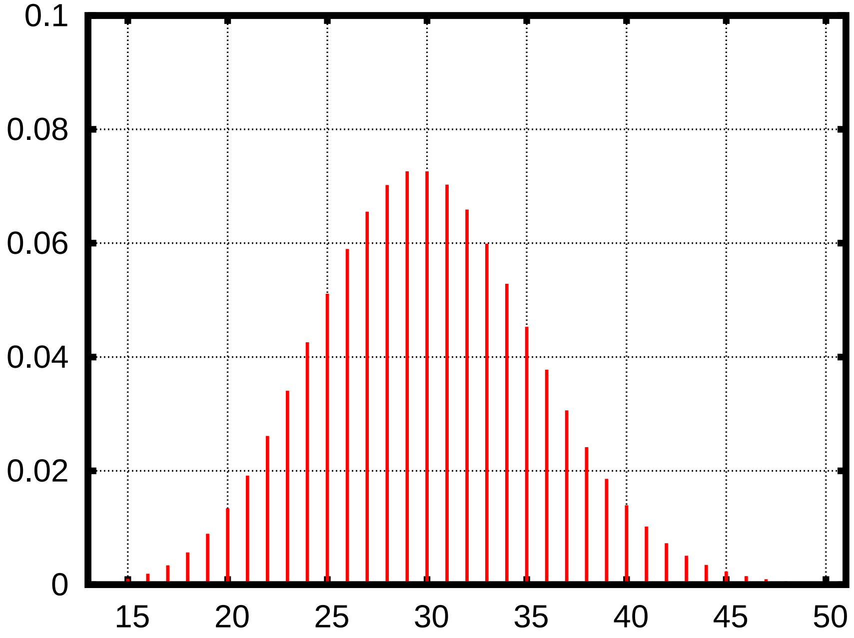




Distribution de Poisson III

Paramètre 8 et paramètre 30





Produit de convolution de distributions

Si deux variables aléatoires indépendantes X et Y admettent les distributions de probabilité f et g, alors la variable aléatoire $X + Y$ admet la distribution de probabilité $h = f * g$ (produit de convolution). Dans le cas particulier où les distributions f_X et g_Y admettent pour support l'ensemble des naturels, avec $P(X = n) = f_n$ et $P(Y = n) = g_n$, on a $P(X + Y = n) = h_n$ avec

$$h_n = \sum_{i=0}^{n} f_i g_{n-i} \,.$$

La preuve est immédiate. Ce résultat permet de prouver facilement l'additivité de la distribution de Poisson.

Espérance mathématique et moyenne I

Définition.

$$E[X] \;=\; \sum_i x_i P(X = x_i)\,.$$

Binomiale, paramètres n et p. L'ensemble des valeurs possibles de X est $\{0, \ldots, n\}$; $P(X = i)$ est $\mathsf{C}_n^i\, p^i(1-p)^{n-i}$.

$$E[X] \;=\; \sum_{i=0}^{n} i\, \mathsf{C}_n^i\, p^i(1-p)^{n-i} \;=\; np\,.$$

Espérance mathématique et moyenne II

$$
\begin{aligned}
E[f(X)] &= \sum_j y_j P(f(X) = y_j) \\
&= \sum_j y_j \sum_{i: f(x_i) = y_j} P(X = x_i) \\
&= \sum_j \sum_{i: f(x_i) = y_j} y_j P(X = x_i) \\
&= \sum_j \sum_{i: f(x_i) = y_j} f(x_i) P(X = x_i) \\
&= \sum_i f(x_i) P(X = x_i)\,.
\end{aligned}
$$

En général, $E[f(X)] \neq f(E[X])$ mais $E[aX + b] = aE[X] + b$.

Espérance mathématique et moyenne III

Binomiale, paramètres n et p.

$$
\begin{aligned}
E[X^k] &= \sum_{i=0}^{n} i^k \, \mathsf{C}_n^i \, p^i (1-p)^{n-i} \\
&= \sum_{i=1}^{n} i^{k-1} n \, \mathsf{C}_{n-1}^{i-1} \, p^i (1-p)^{n-i} \\
&= np \sum_{i=1}^{n} i^{k-1} \, \mathsf{C}_{n-1}^{i-1} \, p^{i-1} (1-p)^{n-i} \\
&= np \sum_{j=0}^{n-1} (j+1)^{k-1} \, \mathsf{C}_{n-1}^{j} \, p^j (1-p)^{n-1-j} \\
&= npE((Y+1)^{k-1}) \quad \text{où } Y \text{ est binomiale de paramètres } n-1 \text{ et } p.
\end{aligned}
$$

En particulier,

$$
E[X^2] = np[(n-1)p+1] = (np)^2 + np(1-p)\,.
$$

Espérance mathématique et moyenne IV

Poisson, paramètre λ

$$
\begin{aligned}
E[X] &= \sum_{i=0}^{\infty} i e^{-\lambda} \frac{\lambda^i}{i!} \\
&= \lambda \sum_{i=1}^{\infty} i e^{-\lambda} \frac{\lambda^{i-1}}{i!} \\
&= \lambda \sum_{i=1}^{\infty} e^{-\lambda} \frac{\lambda^{i-1}}{(i-1)!} \\
&= \lambda \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^j}{j!} \\
&= \lambda\,.
\end{aligned}
$$

Espérance mathématique et moyenne V

Poisson, paramètre λ

$$\begin{aligned}
E[X^2] &= \sum_{i=0}^{\infty} i^2 e^{-\lambda}\frac{\lambda^i}{i!} \\
&= \lambda \sum_{i=1}^{\infty} i e^{-\lambda}\frac{\lambda^{i-1}}{(i-1)!} \\
&= \lambda \sum_{j=0}^{\infty} (j+1) e^{-\lambda}\frac{\lambda^{j}}{j!} \\
&= \lambda \left[\sum_{j=0}^{\infty} j e^{-\lambda}\frac{\lambda^{j}}{j!} \; + \; \sum_{j=0}^{\infty} e^{-\lambda}\frac{\lambda^{j}}{j!}\right] \\
&= \lambda(\lambda+1)\,.
\end{aligned}$$

Variance I

$$\text{var}(X) \;=\; E[(X - E[X])^2]\,.$$

On a aussi, en posant $E[X] = \mu$,

$$\begin{aligned}
\text{var}(X) &= E[(X-\mu)^2] \\
&= \sum_i (x_i - \mu)^2 P(X = x_i) \\
&= \sum_i (x_i^2 - 2\mu x_i + \mu^2) P(X = x_i) \\
&= \sum_i x_i^2 P(X = x_i) - 2\mu \sum_i x_i P(X = x_i) + \mu^2 \sum_i P(X = x_i) \\
&= E[X^2] - 2\mu^2 + \mu^2 \\
&= E[X^2] - (E[X])^2\,.
\end{aligned}$$

L'*écart-type* est la racine carrée de la variance.

Variance II

$$\begin{aligned}\text{var}(aX+b) &= E[(aX+b-a\mu-b)^2]\\ &= E[a^2(X-\mu)^2]\\ &= a^2E[(X-\mu)^2]\\ &= a^2\text{var}(X)\,.\end{aligned}$$

Pour la distribution binomiale (paramètres n et p) :

$$\text{var}(X) \;=\; E[X^2]-(E[X])^2 \;=\; np(1-p)\,.$$

Pour la distribution de Poisson (paramètre λ) :

$$\text{var}(X) \;=\; E[X^2]-(E[X])^2 \;=\; \lambda\,.$$

Distribution géométrique I

Séquence d'essais indépendants, probabilité de succès de chaque essai : p.
Soit X, la v.a. correspondant au rang du premier succès. On a
$P(X=1)=p$, $P(X=2)=(1-p)p$, $P(X=3)=(1-p)^2p$, etc.

$$P(X=n) \;=\; (1-p)^{n-1}p\,; \quad P(X\geq n) \;=\; (1-p)^{n-1}\,.$$

$$\begin{aligned}
E[X] &= \sum_{n=1}^{\infty} n(1-p)^{n-1}p \\
&= \sum_{n=1}^{\infty}(n-1)(1-p)^{n-1}p + \sum_{n=1}^{\infty}(1-p)^{n-1}p \\
&= \left(\sum_{k=0}^{\infty} k(1-p)^k p\right) + 1 \\
&= (1-p)\left(\sum_{k=1}^{\infty} k(1-p)^{k-1}p\right) + 1 \;=\; (1-p)E[X]+1
\end{aligned}$$

d'où on tire

$$E[X] \;=\; \frac{1}{p}\,.$$

Distribution géométrique II

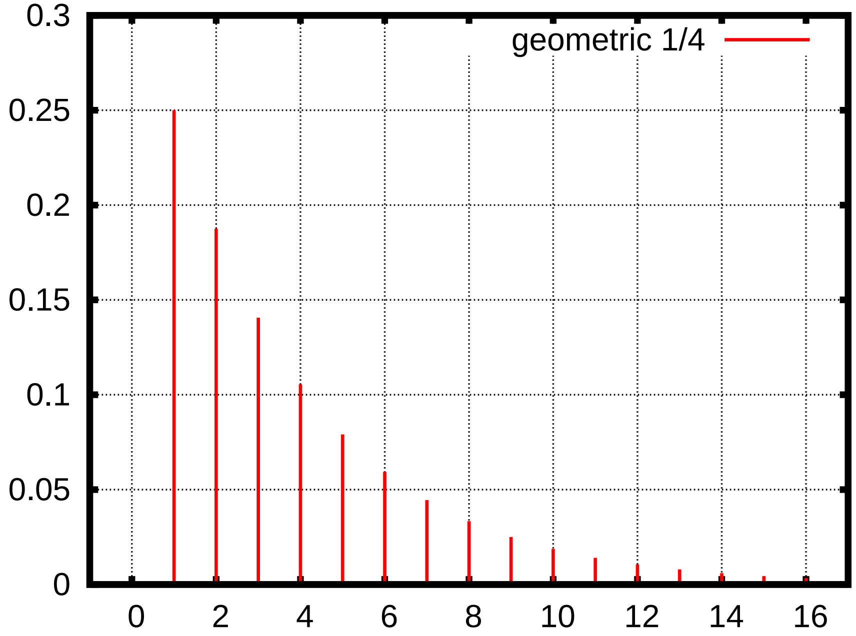
$$\begin{aligned}
E[X^2] &= \sum_{n=1}^{\infty} n^2(1-p)^{n-1}p \\
&= \sum_{n=1}^{\infty} (n-1)^2(1-p)^{n-1}p + \sum_{n=1}^{\infty} 2(n-1)(1-p)^{n-1}p + \sum_{n=1}^{\infty} (1-p)^{n-1}p \\
&= \left(\sum_{k=0}^{\infty} k^2(1-p)^k p\right) + \left(\sum_{k=0}^{\infty} 2k(1-p)^k p\right) + 1 \\
&= (1-p)E[X^2] + 2(1-p)E[X] + 1
\end{aligned}$$

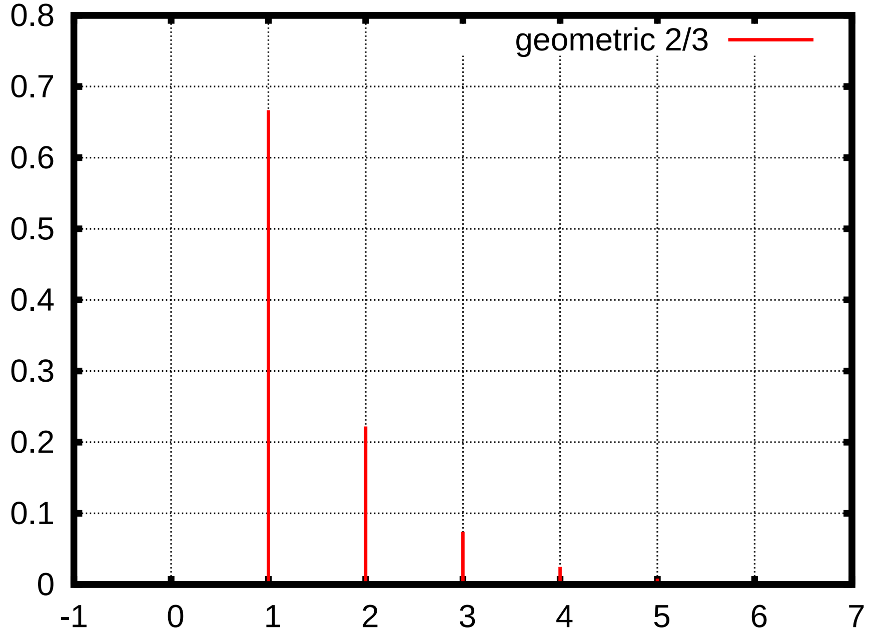
d'où on tire

$$E[X^2] \;=\; \frac{2-p}{p^2} \quad \text{et} \quad \mathrm{var}(X) \;=\; \frac{1-p}{p^2}.$$

Distribution géométrique III

Paramètre 1/4 et paramètre 2/3





Distribution binomiale négative I

On généralise : X est la v.a. correspondant au rang du $r$ième succès.

$$P(X = n) \;=\; \mathsf{C}_{n-1}^{r-1} p^r (1-p)^{n-r}\,.$$

Une v.a. binomiale négative de paramètres r et p est la somme de r v.a. binomiales négatives indépendantes de paramètres 1 et p, c'est-à-dire de r v.a. indépendantes géométriques (pourquoi ?).

Attention, plusieurs définitions alternatives se trouvent dans la littérature.

Distribution binomiale négative II

$$
\begin{aligned}
E[X^k] &= \sum_{n=r}^{\infty} n^k \mathsf{C}_{n-1}^{r-1} p^r (1-p)^{n-r} \\
&= \frac{r}{p} \sum_{n=r}^{\infty} n^{k-1} \mathsf{C}_{n}^{r} p^{r+1} (1-p)^{n-r} \\
&= \frac{r}{p} \sum_{m=r+1}^{\infty} (m-1)^{k-1} \mathsf{C}_{m-1}^{r} p^{r+1} (1-p)^{m-(r+1)} \\
&= \frac{r}{p} E[(Y-1)^{k-1}]
\end{aligned}
$$

où Y est une v.a. binomiale négative de paramètres $r+1$ et p.

Distribution binomiale négative III

Pour $k = 1$ on obtient

$$E[X] \;=\; \frac{r}{p}\,.$$

Pour $k = 2$ on obtient

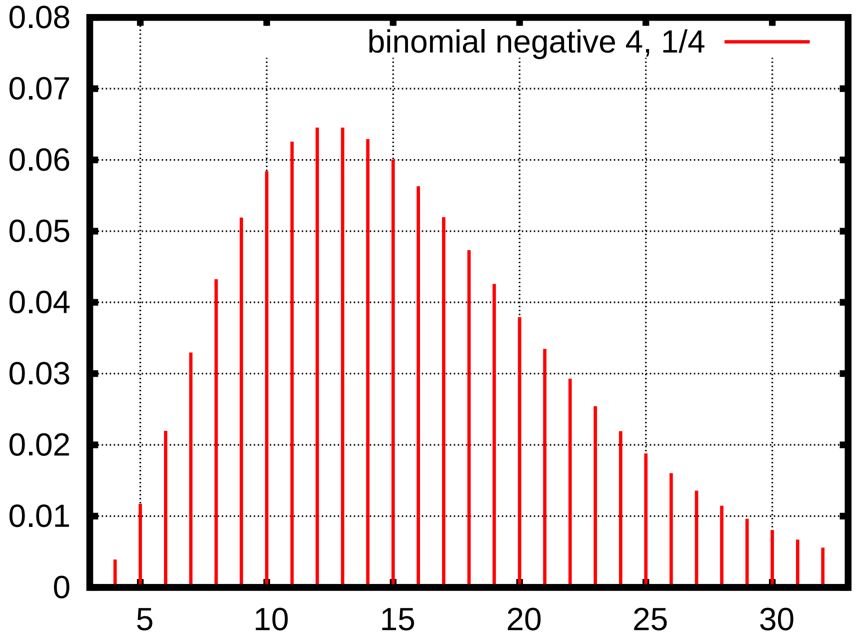
$$E[X^2] \;=\; \frac{r}{p}E[Y-1] \;=\; \frac{r}{p}\left(\frac{r+1}{p} - 1\right)\,.$$

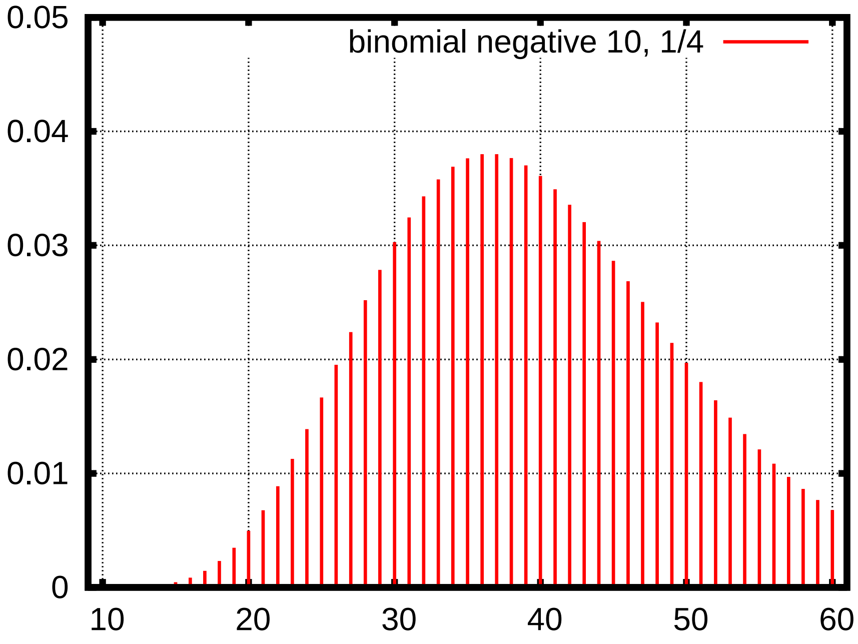
On en déduit

$$\mathrm{var}(X) \;=\; \frac{r}{p}\left(\frac{r+1}{p} - 1\right) - \left(\frac{r}{p}\right)^2 \;=\; \frac{r(1-p)}{p^2}\,.$$

Distribution binomiale négative IV

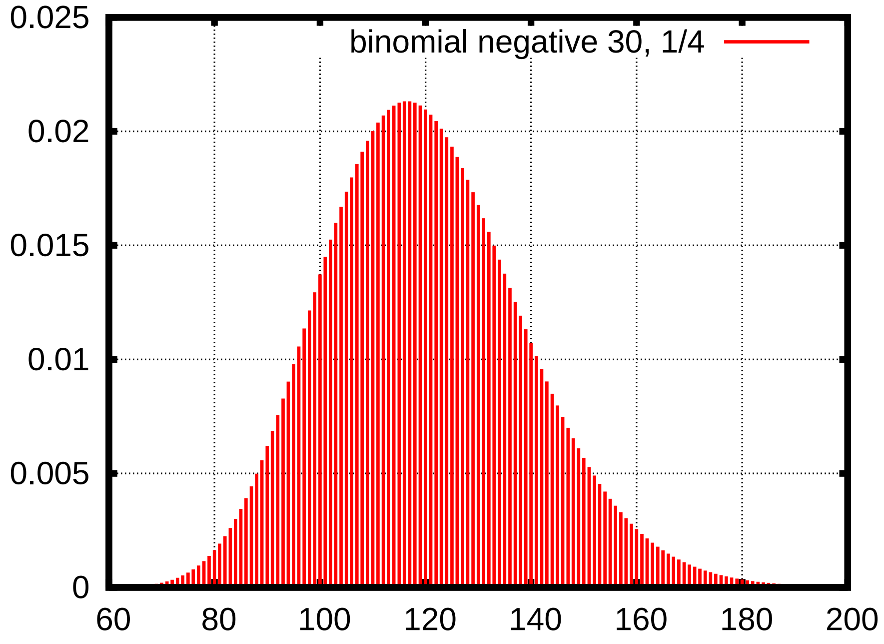
Paramètre r : 4 et 10 ; paramètre p : 1/4

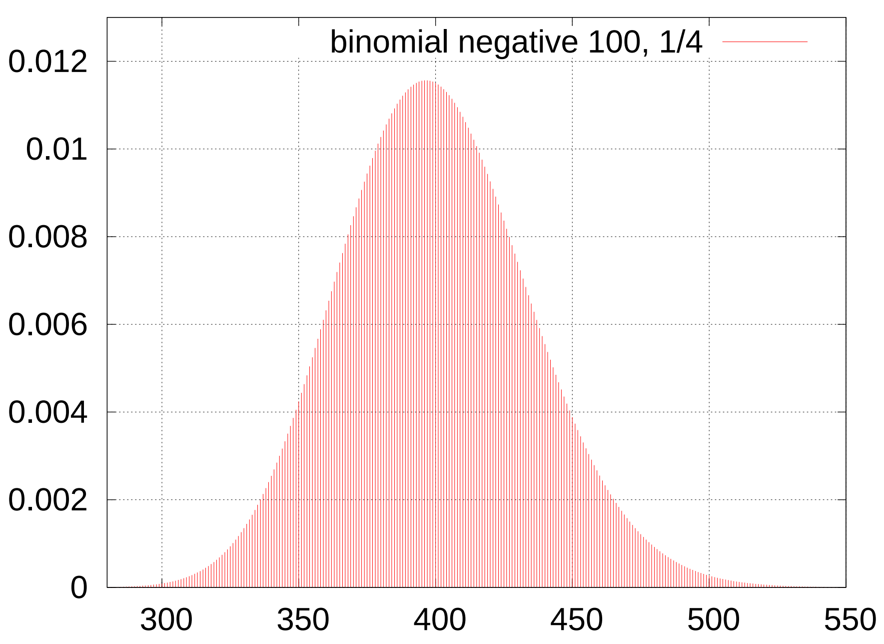




Distribution binomiale négative V

Paramètre r : 30 et 100 ; paramètre p : 1/4





Distribution hypergéométrique I

Une urne comporte N billes dont m blanches et $N - m$ noires. On tire n billes ; soit X le nombre de billes blanches. On a, pour tout i compris entre 0 et n,

$$P(X = i) \;=\; \frac{\mathsf{C}_m^i \mathsf{C}_{N-m}^{n-i}}{\mathsf{C}_N^n}\,.$$

Convention : C_u^k est nul si k n'est pas compris entre 0 et u.

Les paramètres sont n, N et m.

$P(X = i)$ est nul si la condition $n - (N - m) \leq i \leq \min(n, m)$ n'est pas respectée.

Distribution hypergéométrique II

$$
\begin{aligned}
E[X^k] &= \sum_{i=0(1)}^{n} i^k \frac{C_m^i C_{N-m}^{n-i}}{C_N^n} \\
&= \frac{nm}{N} \sum_{i=1}^{n} i^{k-1} \frac{C_{m-1}^{i-1} C_{N-m}^{n-i}}{C_{N-1}^{n-1}} \\
&= \frac{nm}{N} \sum_{j=0}^{n-1} (j+1)^{k-1} \frac{C_{m-1}^{j} C_{N-m}^{n-1-j}}{C_{N-1}^{n-1}} \\
&= \frac{nm}{N} E[(Y+1)^{k-1}]
\end{aligned}
$$

où Y est une v.a. hypergéométrique de paramètres $n-1$, $N-1$ et $m-1$.

Distribution hypergéométrique III

Pour $k = 1$ on obtient

$$E[X] \;=\; \frac{nm}{N}\,.$$

Pour $k = 2$ on obtient

$$E[X^2] \;=\; \frac{nm}{N}E[Y+1] \;=\; \frac{nm}{N}\left(\frac{(n-1)(m-1)}{N-1}+1\right)\,.$$

On en déduit

$$\mathrm{var}(X) \;=\; \frac{nm}{N}\left(\frac{(n-1)(m-1)}{N-1}+1-\frac{nm}{N}\right)\,.$$

En posant $p = m/N$, cela se récrit en

$$\mathrm{var}(X) \;=\; np(1-p)\left(1-\frac{n-1}{N-1}\right)\,.$$

Test exact de Fisher I

Chez X patients ayant subi le traitement I*,*
on a observé A succès et B échecs, et
chez Y patients ayant subi le traitement II*,*
on a observé C succès et D échecs.
(On a naturellement $A + B = X$ et $C + D = Y$;
on pose $N = X + Y$.)

Peut-on, en fonction de A, B, C et D, mettre en évidence une différence statistiquement significative entre les traitements I *et* II *?*

Test exact de Fisher II

Pour répondre à cette question, on va, selon l'expression consacrée, "tester l'hypothèse nulle (H_0)", et déterminer dans quelle mesure la situation observée, représentée par la matrice $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$, suggère statistiquement le rejet de H_0. On note S le nombre de succès et E le nombre d'échecs ; on a donc $S+E = X+Y = N$, $S = A+C$ et $E = B+D$.

Si réellement le choix du traitement n'a pas d'influence sur la proportion de guérisons, la matrice $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ est une variable aléatoire correspondant à l'expérience du tirage exhaustif de $X = A + B$ boules hors d'une urne contenant $S = A + C$ boules blanches et $E = B + D$ boules rouges ; les malades ayant subi le traitement I correspondent aux boules sorties de l'urne, et la guérison correspond à la couleur blanche.

Test exact de Fisher III

La loi de distribution hypergéométrique s'applique donc ici et on a

$$p\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) = \frac{\mathsf{C}_X^A \, \mathsf{C}_Y^C}{\mathsf{C}_N^S} = \frac{\mathsf{C}_{A+B}^{\;A} \, \mathsf{C}_{C+D}^{\;C}}{\mathsf{C}_{A+B+C+D}^{\;A+C}} \,.$$

On pourrait tout aussi bien poser que les malades ayant subi le traitement I correspondent aux boules blanches, et la guérison au fait d'être sorti de l'urne. On aurait alors

$$p\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) = \frac{\mathsf{C}_S^A \, \mathsf{C}_E^B}{\mathsf{C}_N^X} = \frac{\mathsf{C}_{A+C}^{\;A} \, \mathsf{C}_{B+D}^{\;B}}{\mathsf{C}_{A+B+C+D}^{\;A+B}} \,.$$

On note que, nécessairement

$$max(0, S - Y) \leq A \leq min(S, X) \,.$$

Test exact de Fisher IV

Ces deux modèles sont naturellement équivalents. En choisissant le premier, on reconnaît (pour la variable A) la distribution hypergéométrique de paramètres n, N et m :

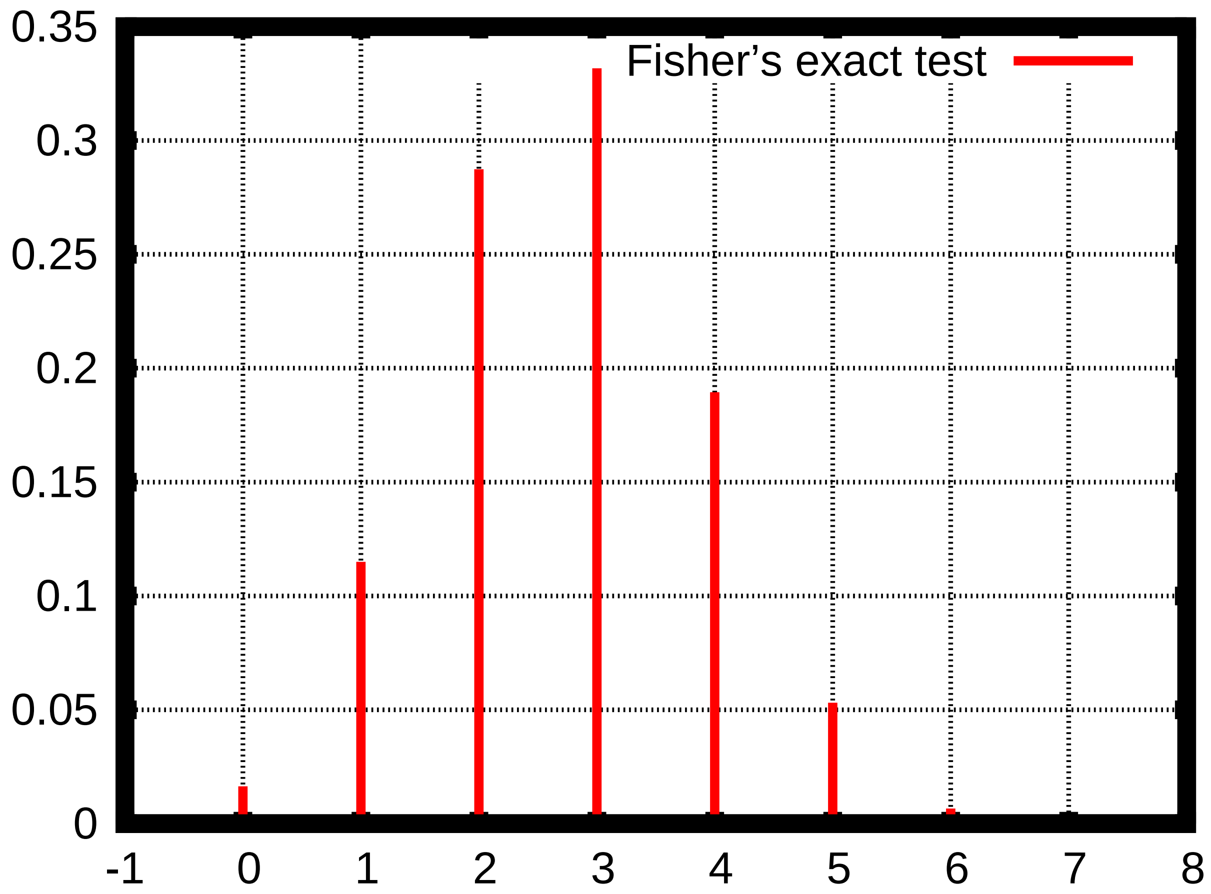
$$P(A = i) \; = \; \frac{\mathrm{C}_m^i \mathrm{C}_{N-m}^{n-i}}{\mathrm{C}_N^n},$$

à condition de choisir $n = S = A + C$ et $m = A + B$.

Exemple. On a $N = 28$ patients, dont $X = 11$ subissent le traitement I et $Y = 17$ subissent le traitement II ; on obtient $S = 7$ réponses favorables et 21 échecs. Il y a huit répartitions possibles des succès ; la répartition observée est $A = 4$, c'est-à-dire $\left(\begin{smallmatrix} 4 & 7 \\ 3 & 14 \end{smallmatrix}\right)$. Que peut-on en penser ?

Clairement, le traitement I a été plus favorable (taux de succès 4/11) que le traitement II (taux de succès 3/17).

Test exact de Fisher V



Matrice	0 11 7 10	1 10 6 11	2 9 5 12	3 8 4 13	4 7 3 14	5 6 2 15	6 5 1 16	7 4 0 17
Probabilité (%)	1.64	11.50	28.74	33.17	18.95	5.31	0.66	0.03
Cumul (%)	1.64	13.14	41.88	75.05	94.00	99.31	99.97	100.
Cumul inverse (%)	100.	98.36	86.86	58.12	24.95	6.00	0.69	0.03

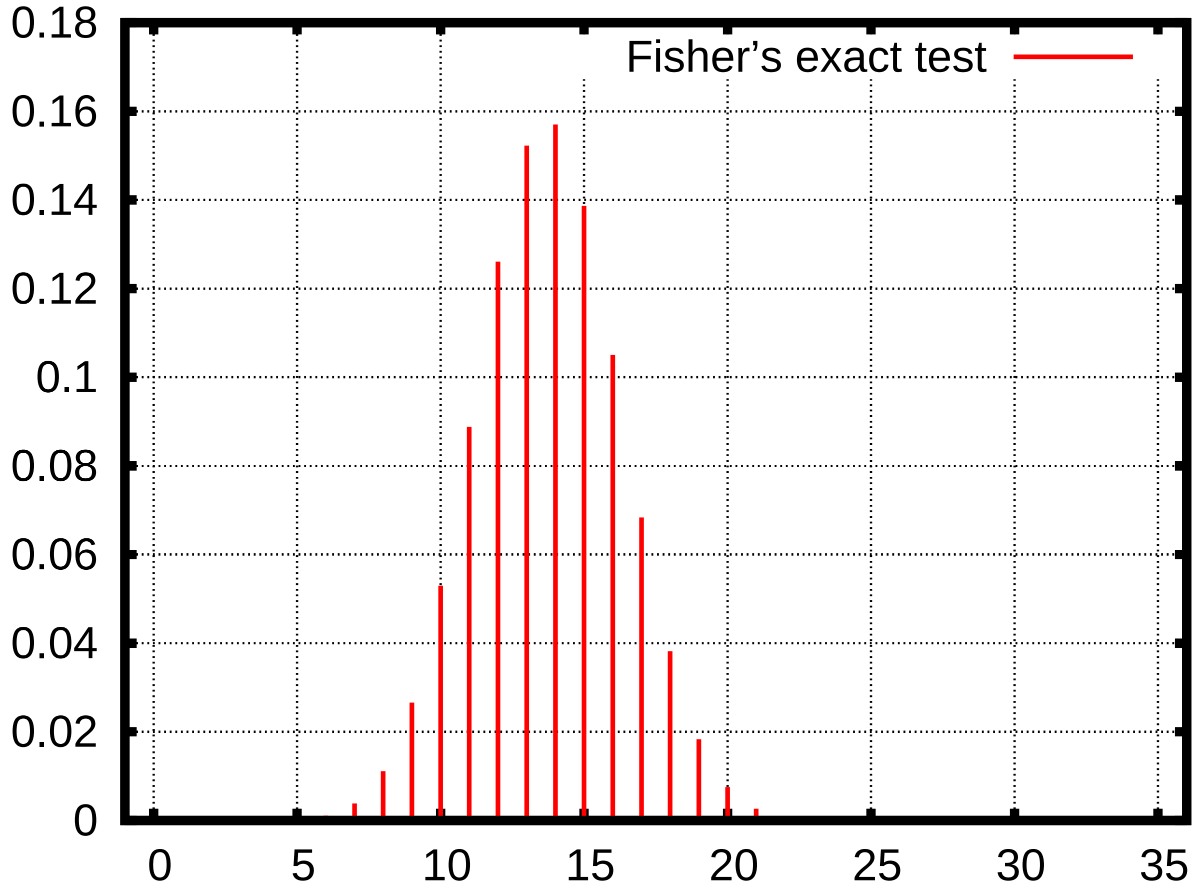
Test exact de Fisher VI

On constate que la situation $\left(\begin{smallmatrix} 4 & 7 \\ 3 & 14 \end{smallmatrix}\right)$ est loin d'indiquer un avantage significatif pour le traitement I puisque, à supposer que les deux traitements soient identiques, un tel "avantage" apparaîtrait dans près de 25 % des cas !

En fait, seuls les résultats extrêmes $\left(\begin{smallmatrix} 6 & 5 \\ 1 & 16 \end{smallmatrix}\right)$ et $\left(\begin{smallmatrix} 7 & 4 \\ 0 & 17 \end{smallmatrix}\right)$ auraient permis de conclure à un avantage statistiquement significatif du traitement I ; à l'inverse, seul le résultat $\left(\begin{smallmatrix} 0 & 11 \\ 7 & 10 \end{smallmatrix}\right)$ aurait indiqué un avantage significatif du traitement II.

Cependant, des données plus abondantes (par exemple cinq fois plus nombreuses) permettent de tirer plus facilement une conclusion.

Test exact de Fisher VII



La matrice $\begin{pmatrix} 4 & 7 \\ 3 & 14 \end{pmatrix}$ n'était pas significative ($\alpha = 24.95\,\%$) mais la matrice $\begin{pmatrix} 20 & 35 \\ 15 & 70 \end{pmatrix}$ est significative ($\alpha = 1.12\,\%$).

Espérance et variance d'une somme de v.a.

L'espérance d'une somme de variables aléatoires vaut la somme des espérances des variables aléatoires (même si elles ne sont pas indépendantes).

$$E[\sum_i X_i] \;=\; \sum_i E[X_i]\,.$$

Ce résultat, admis sans démonstration, est déjà connu dans divers cas particuliers. On admet aussi que *si les variables aléatoires sont indépendantes*,

$$\mathrm{var}[\sum_i X_i] \;=\; \sum_i \mathrm{var}[X_i]\,.$$

Variables aléatoires réelles

Les variables aléatoires *discrètes* prennent leurs valeurs dans un ensemble (espace) fini ou dénombrable, souvent (assimilable à) un ensemble de nombres entiers. Les variables aléatoires *continues* prennent leurs valeurs dans un ensemble non dénombrable, en général un intervalle (fini ou non) de l'ensemble des nombres réels. Une v.a. continue X admet la fonction f comme distribution de probabilité si, pour tout sous-intervalle $[a : b]$ de l'espace,

$$P(a \leq X \leq b) \;=\; \int_a^b f(x)\mathrm{d}x\,.$$

Nous nous limitons aux v.a. continues pour lesquelles une telle fonction f, nécessairement positive, existe. La primitive de f, notée F, est la fonction de répartition ; plus précisément :

$$F(x) = P(X < x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\mathrm{d}u\,.$$

Espérance, moyenne, variance

Définition.

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

Théorème (sera démontré dans le cas particulier $g(x) \geq 0$).

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

Définition.

$$\mathrm{var}(X) \;=\; E[(X - E[X])^2] \;=\; E[X^2] - (E[X])^2 \,.$$

Un lemme utile

Lemme. Si Y est une v.a. réelle positive,

$$E[Y] \;=\; \int_0^{+\infty} P(Y > y)\mathrm{d}y\,.$$

Démonstration. Si f est la d.p. de Y,

$$\int_0^{+\infty} P(Y > y)\mathrm{d}y = \int_0^{+\infty} \left(\int_y^{+\infty} f(x)\mathrm{d}x \right) \mathrm{d}y$$

$$= \int_0^{+\infty} \left(\int_0^{x} \mathrm{d}y \right) f(x)\mathrm{d}x = \int_0^{+\infty} x f(x)\mathrm{d}x = E[Y]\,.$$

Utile pour démontrer le (cas particulier du) théorème précédent.

Démonstration du (cas particulier du) théorème

La fonction g est supposée positive.

$$\begin{aligned}
E[g(X)] &= \int_0^{+\infty} P(g(X) > y)\mathrm{d}y \\
&= \int_0^{+\infty} \left(\int_{x: g(x) > y} f(x)\mathrm{d}x \right) \mathrm{d}y \\
&= \int_{x: g(x) > 0} \int_0^{g(x)} \mathrm{d}y f(x)\mathrm{d}x \\
&= \int_{x: g(x) > 0} g(x) f(x)\mathrm{d}x
\end{aligned}$$

Variable aléatoire uniforme sur un intervalle

Sa d.p. f vérifie $f(x) = 1$ si $0 < x < 1$ et $f(x) = 0$ sinon.

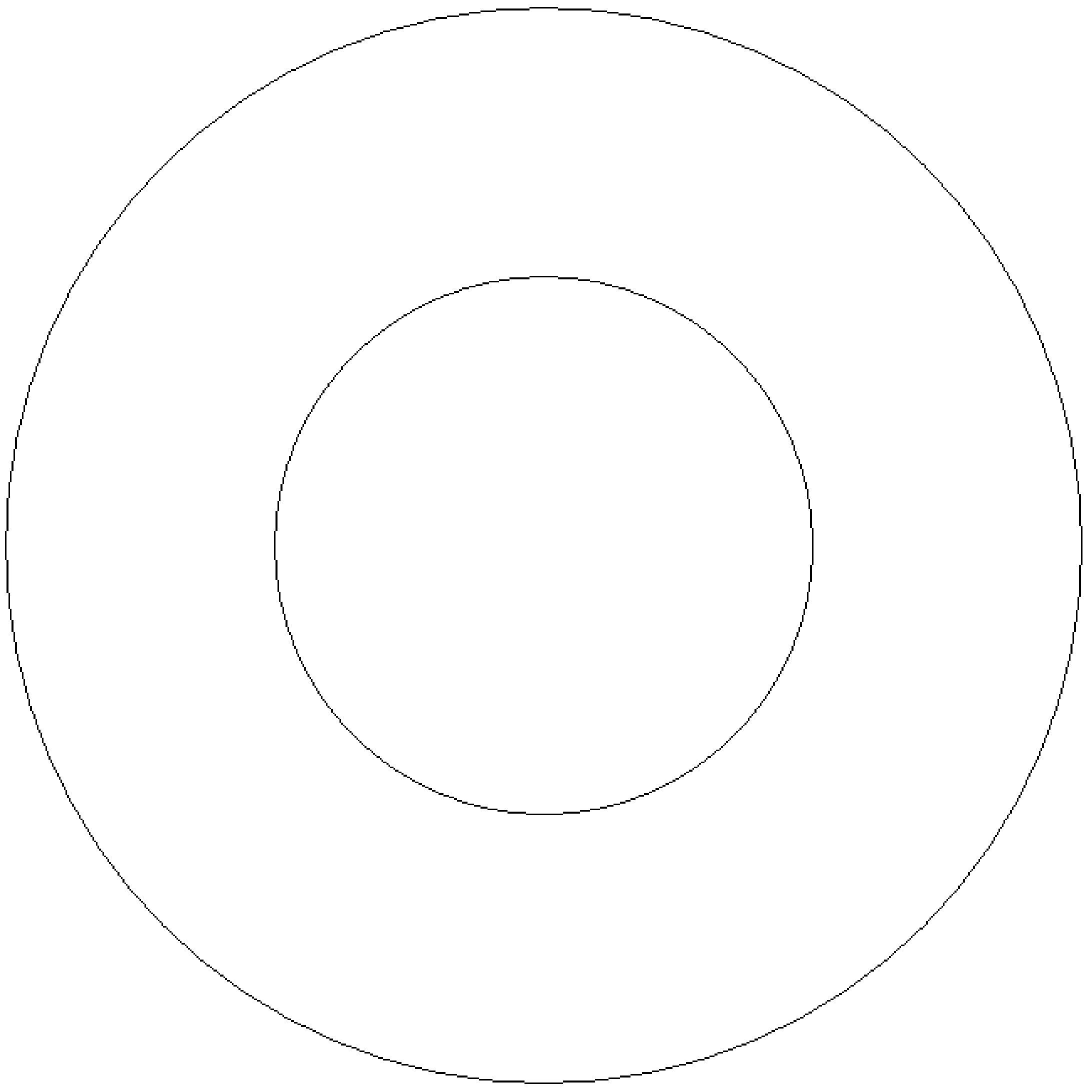
Généralisation : v.a. uniforme sur $[a : b]$ $(a < b)$; sa d.p. f vérifie $f(x) = \frac{1}{b-a}$ si $a < x < b$ et $f(x) = 0$ sinon.

$$E[X] = \int_{-\infty}^{+\infty} x f(x) \mathrm{d}x = \int_a^b \frac{x}{b-a} \mathrm{d}x = \frac{b+a}{2}.$$

$$E[X^2] = \int_a^b \frac{x^2}{b-a} \mathrm{d}x = \frac{b^2 + ab + a^2}{3}.$$

$$\mathrm{var}(X) \;=\; E[(X - E[X])^2] \;=\; \frac{(b-a)^2}{12}.$$

Paradoxe de Bertrand



Deux cercles concentriques, de rayons 1 et 2. Quelle est la probabilité qu'une droite coupant le grand cercle coupe aussi le petit ? Les droites sont uniformément réparties.

Variable aléatoire normale (ou gaussienne)

La distribution de probabilité normale, notée $N(\mu, \sigma^2)$ est

$$f(x) \; = \; \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \, .$$

où μ et σ^2 sont des paramètres réels ($\sigma > 0$).

Cette définition est acceptable car

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-(x-\mu)^2/2\sigma^2} \mathrm{d}x \; = \; 1 \, .$$

Calcul de l'intégrale

Substitution $y = (x - \mu)/\sigma$:

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} e^{-(x-\mu)^2/2\sigma^2}\mathrm{d}x \;=\; \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-y^2/2}\mathrm{d}y\,.$$

et

$$\left(\int_{-\infty}^{+\infty} e^{-y^2/2}\mathrm{d}y\right)^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-(x^2+y^2)/2}\mathrm{d}x\mathrm{d}y$$

$$= \int_0^{+\infty}\int_0^{2\pi} e^{-r^2/2} r\mathrm{d}r\mathrm{d}\theta = 2\pi\int_0^{+\infty} re^{-r^2/2}\mathrm{d}r = 2\pi\,.$$

Transformation linéaire

Si X est $N(\mu, \sigma^2)$, alors $Y = aX + b$ est $N(a\mu + b, a^2\sigma^2)$.
($a > 0$; $a < 0$ analogue.)

$$F_Y(x) \;=\; P(Y \leq x) \;=\; P(X \leq \frac{x-b}{a}) \;=\; F_X(\frac{x-b}{a})\,.$$

$$\begin{aligned} f_Y(x) \;&=\; \frac{1}{a} f_X(\frac{x-b}{a}) \;=\; \frac{1}{\sqrt{2\pi}a\sigma} \exp\{-(\frac{x-b}{a} - \mu)^2/2\sigma^2\} \\ &=\; \frac{1}{\sqrt{2\pi}a\sigma} \exp\{-(x-b-a\mu)^2/2(a\sigma)^2\}\,. \end{aligned}$$

Conséquence : il suffit d'étudier $N(0,1)$.

Calcul de la moyenne

$$
\begin{aligned}
E[Z] &= \int_{-\infty}^{+\infty} x f_Z(x)\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-x^2/2}\mathrm{d}x \\
&= 0
\end{aligned}
$$

La moyenne de $N(0,1)$ est 0 ; la moyenne de $N(\mu,\sigma^2)$ est μ.

Calcul de la variance

$$
\begin{aligned}
\mathrm{var}(Z) &= E[Z^2] \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-x^2/2} \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \left(-xe^{-x^2/2} \Big]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-x^2/2} \mathrm{d}x \right) \\
&= \frac{1}{\sqrt{2\pi}} \left(0 + \int_{-\infty}^{+\infty} e^{-x^2/2} \mathrm{d}x \right) \\
&= 1
\end{aligned}
$$

$$\mathrm{var}(X) \;=\; \sigma^2 \mathrm{var}(Z) \;=\; \sigma^2$$

La variance de $N(0,1)$ est 1 ; la variance de $N(\mu, \sigma^2)$ est σ^2.

Distribution exponentielle (v.a. réelle positive)

Paramètre : réel positif λ.

Distribution de probabilité : $f(x) \;=\; \lambda e^{-\lambda x}$.

Fonction de répartition : $F(x) \;=\; P(X \leq x) \;\; (x \geq 0)$.

$$\begin{aligned} F(x) &= \int_0^x \lambda e^{-\lambda u} \mathrm{d}u \\ &= -e^{-\lambda u}\Big]_0^x \\ &= 1 - e^{-\lambda x} \end{aligned}$$

Calcul de la moyenne et de la variance

$$\begin{aligned}
E[X^n] &= \int_0^{+\infty} x^n \lambda e^{-\lambda x} \mathrm{d}x \\
&= -x^n e^{-\lambda x}\Big]_0^{+\infty} + \int_0^{+\infty} n x^{n-1} e^{-\lambda x} \mathrm{d}x \\
&= 0 + \frac{n}{\lambda} \int_0^{+\infty} \lambda x^{n-1} e^{-\lambda x} \mathrm{d}x \\
&= \frac{n}{\lambda} E[X^{n-1}]
\end{aligned}$$

Pour $n = 1$ et $n = 2$ on a :

$$E[X] = \frac{1}{\lambda}; \qquad E[X^2] = \frac{2}{\lambda} E[X] = \frac{2}{\lambda^2}$$

et donc

$$\mathrm{var}(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

Loi jointe pour deux v.a. discrètes

Fonction de répartition jointe, distribution de probabilité jointe :

$$F(x,y) = P(X \leq x, Y \leq y)\,, \qquad\qquad p(x,y) = P(X = x, Y = y)\,.$$

Distributions de probabilité marginales :

$$p_X(x) = P(X = x) = \sum_{y:p(x,y)>0} p(x,y)\,,$$

$$p_Y(y) = P(Y = y) = \sum_{x:p(x,y)>0} p(x,y)\,.$$

Loi jointe pour deux v.a. discrètes, exemple

(Sheldon Ross, *A First Course in Probability*.)
Une urne contient 3 boules rouges, 4 blanches, 5 noires ;
trois boules sont tirées au hasard ; X, Y : nombres de rouges, de blanches.

TABLE 6.1: $P\{X = i, Y = j\}$

$i \backslash j$	0	1	2	3	Row sum $= P\{X = i\}$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
Column sum $= P\{Y = j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

Variables continues, distribution conjointe

On suppose qu'il existe f telle que

$$P((X,Y) \in C) = \iint_C f(x,y)\mathrm{d}x\mathrm{d}y\,,$$

quel que soit le sous-ensemble mesurable C de $\mathbb{R}^2$. Si $C = A \times B$, on a

$$P(X \in A, Y \in B) = \int_A \int_B f(x,y)\mathrm{d}x\mathrm{d}y$$

et en particulier

$$F(a,b) = P(X < a, Y < b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x,y)\mathrm{d}x\mathrm{d}y\,;$$

on a donc

$$f(a,b) = \frac{\partial^2}{\partial a \partial b} F(a,b)\,.$$

$$P(X \in A) = \int_A \int_{-\infty}^{\infty} f(x,y)\mathrm{d}x\mathrm{d}y = \int_A f_X(x)\mathrm{d}x\,.$$

Variables aléatoires indépendantes

Définition : X et Y respectent la condition

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)\,,$$

pour tous sous-ensembles $A, B \subset \mathbb{R}$.

Il suffit (cela se démontre...) d'avoir

$$P(X < a, Y < b) = P(X < a)P(Y < b)\,,$$

pour tous réels a, b, ou encore

$$F(a, b) = F_X(a)F_Y(b)\,.$$

On a aussi $f(a, b) = f_X(a)f_Y(b)$.

Somme de deux v.a. indépendantes I

$$
\begin{aligned}
F_{X+Y}(a) &= P(X+Y<a) \\
&= \iint_{x+y<a} f_X(x) f_Y(y) \mathsf{d}x \mathsf{d}y \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x) f_Y(y) \mathsf{d}x \mathsf{d}y \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x) \mathsf{d}x f_Y(y) \mathsf{d}y \\
&= \int_{-\infty}^{\infty} F_X(a-y) f_Y(y) \mathsf{d}y
\end{aligned}
$$

Somme de deux v.a. indépendantes II

$$\begin{aligned} f_{X+Y}(a) &= \frac{\mathsf{d}}{\mathsf{d}a}\int_{-\infty}^{\infty} F_X(a-y) f_Y(y)\mathsf{d}y \\ &= \int_{-\infty}^{\infty} \frac{\mathsf{d}}{\mathsf{d}a} F_X(a-y) f_Y(y)\mathsf{d}y \\ &= \int_{-\infty}^{\infty} f_X(a-y) f_Y(y)\mathsf{d}y \\ &= (f_X * f_Y)(a) \end{aligned}$$

Produit de convolution.

Somme de deux v.a. indépendantes, uniformément distribuées sur $]0:1[$

$f_X(a) = f_Y(a) = 1$ si $0 < a < 1$, 0 sinon.

$$(f_X * f_Y)(a) = \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) \mathrm{d}y = \int_0^1 f_X(a-y) \mathrm{d}y\,.$$

Pour $0 \leq a \leq 1$, on a

$$(f_X * f_Y)(a) = \int_0^a \mathrm{d}y = a\,;$$

pour $1 \leq a \leq 2$, on a

$$(f_X * f_Y)(a) = \int_{a-1}^1 \mathrm{d}y = 2 - a\,.$$

On retrouve le résultat obtenu précédemment par discrétisation.

Inégalité de Markov

Si X est une v.a. positive on a, pour tout $a > 0$,

$$P(X \geq a) \leq \frac{E[X]}{a}\,.$$

Démonstration. Soit $I = 1$ si $X \geq a$ et $I = 0$ sinon.
On a $E[I] = P(X \geq a)$ et aussi

$$I \leq \frac{X}{a} \text{ et donc } E[I] \leq \frac{E[X]}{a}\,.$$

Inégalité de Tchebychev

Soit X une v.a. de moyenne μ et de variance σ^2. Pour tout $k > 0$ on a

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

Démonstration. On applique l'inégalité de Markov à la v.a. $(X - \mu)^2$, avec $a = k^2$; cela donne

$$P((X - \mu)^2 \geq k^2) \leq \frac{E[(X - \mu)^2]}{k^2}\,,$$

c'est-à-dire la thèse.

Loi faible des grands nombres

Si $X_1, X_2, \ldots$ est une suite de v.a. indépendantes et identiquement distribuées, de moyenne μ et de variance σ^2, alors, pour tout $\varepsilon > 0$, on a

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right) = 0\,.$$

Démonstration. On applique l'inégalité de Tchebychev à la v.a.

$$Y_n = \frac{X_1 + \cdots + X_n}{n}\,,$$

dont la moyenne est $\frac{n\mu}{n} = \mu$ et la variance $\frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$, ce qui donne

$$P(|Y_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}\,.$$

Théorème central limite

Dans les mêmes conditions, la distribution de la variable aléatoire

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tend vers la loi de Gauss $N(0,1)$.

(On l'admet sans démonstration.)

Test de Wilcoxon, approximation normale I

Les résultats précédents suggèrent – et on pourrait démontrer – que pour des grandes valeurs de n et p, la distribution d'une variable aléatoire de Wilcoxon W est proche de la distribution normale.

Ce résultat est important en pratique. La taille de la table de Wilcoxon est de l'ordre de $(n+p)^4$, ce qui rend la construction de cette table lourde pour $n+p > 200$ et complètement impraticable pour $n+p > 1\,000$.

Test de Wilcoxon, approximation normale II

Variable W : distribuée symétriquement entre 0 et np. On a donc

$$E[X] = \mu = \frac{np}{2}\,.$$

Le calcul de $E[X^2]$ et donc de la variance n'est pas évident mais on peut supposer – sous réserve de vérification – que le résultat sera un polynôme symétrique en n et p, de degré n'excédant pas 4, soit

$$a(n^4+p^4)+b(n^3p+np^3)+cn^2p^2+d(n^3+p^3)+e(n^2p+np^2)$$

$$+f(n^2+p^2)+gnp+h(n+p)+i\,;$$

Le calcul direct du polynôme pour dix couples de petites valeurs de n et p permet de trouver les dix coefficients inconnus, au moyen d'un programme de résolution d'équations linéaires.

Test de Wilcoxon, approximation normale III

On trouve

$$E[W^2] \;=\; \mu_2 \;=\; \frac{np(3np+n+p+1)}{12};$$

$$\mathrm{var}(W) \;=\; \mu_2 - \mu^2 \;=\; \frac{np(n+p+1)}{12}.$$

On peut démontrer par récurrence la validité de la formule donnant μ_2. La récurrence porte sur $k = n + p$. Le calcul direct est facile pour les petites valeurs de k.

Test de Wilcoxon, approximation normale IV

Hypothèse de récurrence : la formule est vraie pour $k-1$ et donc, en particulier, pour les couples $(n-1,p)$ et $(n,p-1)$.

(Si $n=0$ ou $p=0$, la formule à démontrer est évidente, pourquoi ?)

Les C_k^n mots de Wilcoxon se répartissent en deux classes, selon que leur dernière "lettre" est "O" ou "X". Les effectifs des deux classes sont en proportion n et p. En faisant abstraction de la dernière lettre, les mots de la première classe sont de type $(n-1,p)$ et ceux de la seconde de type $(n,p-1)$. Dans le premier cas, la dernière lettre ne change pas W, dans le deuxième cas, W devient $W+n$ et donc W^2 devient $W^2+2nW+n^2$.

On en déduit que la valeur de $\mu_2(n,p)$ est

$$\frac{n}{n+p}\,\mu_2(n-1,p) \;+\; \frac{p}{n+p}\left(\mu_2(n,p-1)+2n\mu(n,p-1)+n^2\right).$$

Test de Wilcoxon, approximation normale V

En appliquant l'hypothèse de récurrence, le premier terme se récrit en

$$\frac{n}{n+p}\frac{(n-1)p(3(n-1)p+n+p)}{12}$$

et le second en

$$\frac{p}{n+p}\left(\frac{n(p-1)(3n(p-1)+n+p)}{12} + 2n\frac{n(p-1)}{2} + n^2\right)$$

et leur somme donne bien le résultat attendu, c'est-à-dire

$$\mu_2(n,p) \;=\; \frac{np(3np+n+p+1)}{12}\,.$$

Test de Wilcoxon, approximation normale VI

Données de test.

Les valeurs (communes) de n et p sont 5, 10, 15, 20, 25, 30 et 35.

```
(1 1 3 3 3)
(2 2 4 4 8)

(1 1 3 3 3 5 5 7 11 15)
(2 4 4 4 8 8 8 8 16 18)

(1 1 3 3 3 5 5 7 11 15 19 23 23 23 23)
(2 4 4 4 8 8 8 8 16 18 26 26 28 30 30)

(1 1 3 3 3 5 5 7 11 15 19 23 23 23 23 29 33 35 37 37)
(4 6 8 8 8 8 8 8 10 10 10 30 30 30 32 34 36 36 38 38)

(1 1 3 3 3 5 5 5 7 7 9 15 19 23 23 23 23 23 23 23 23 29 33 35 37)
(2 4 4 6 6 8 8 8 8 8 8 12 14 26 26 28 30 30 30 32 34 36 36 38 38)

(1 1 1 1 3 3 3 3 3 5 5 5 7 7 9 9 11 15 19 23 23 23 23 23 23 23 29 33 35 37)
(2 2 2 4 4 4 6 6 6 8 8 8 8 8 8 8 10 12 20 26 26 28 30 30 32 34 36 36 38 38)

(1 1 1 1 3 3 3 3 3 5 5 5 7 7 9 9 11 11 13 13 15 17 19 23 23 23 23 23 23 23 29 33 35 37 37)
(2 2 2 4 4 4 6 6 6 8 8 8 8 8 8 8 10 10 12 14 16 18 20 26 26 28 30 30 32 34 36 36 38 38 40)
```

Test de Wilcoxon, approximation normale VII

Résultats exprimés en pourcentage :

size	wil-	nwil-	diff
5	11.11111	10.50375	5.782306
10	8.27470	8.09862	2.174159
15	10.08397	9.92529	1.598735
20	13.83135	13.66426	1.222877
25	9.11344	9.03184	0.903381
30	12.08989	11.99250	0.812077
35	16.88197	16.77287	0.650445

Ces résultats tendent à montrer que l'approximation est envisageable si $n, p \geq 10$, et largement acceptable si $n, p \geq 25$.