Notions élémentaires de probabilités utilisées en traitement du signal

Joël Le Roux

13 novembre 2000

Table des matières

Ce document comporte encore pas mal d'erreurs ; merci de me les signaler en me faisant parvenir un message à l'adresse suivante

leroux@essi.fr

D'autres documents de cours sont disponibles sur le site

www.essi.fr/ leroux/

1 Utilité du calcul des probabilités

1.1 Analyse des phénomènes non prédictibles

Lorsqu'un observateur réalise une expérience ou effectue une mesure, il ne peut pas, en général, prédire avec une précision parfaite le résultat de cette mesure. (Si ce résultat était parfaitement prédictible, la mesure ne présenterait pas d'utilité ...) Il y a donc une fluctuation possible des résultats. Parmi les exemples les plus simples d'expériences aux résultats non prévisibles, il y a les jeux de hasard : tirer à pile ou face, lancer un dé, etc ... On est aussi amené à observer des phénomènes imparfaitement prédictibles et présentant des résultats fluctuants lorsqu'on réalise un sondage sur une population, ou bien lorsqu'on effectue une mesure en physique.

Pour caractériser le mieux possible les fluctuations du résultat d'une expérience, il est nécessaire de disposer d'un *formalisme mathématique* permettant de préciser le mieux possible la connaissance qu'on peut déduire du résultat des expériences. Le premier objectif du calcul des probabilités est de donner cet outil mathématique.

Une des utilisation importantes du calcul des probabilités est la *prédiction* aussi précise que possibles de phénomènes malgré les incertitudes qui les caractérisent. C'est fondamental dans différents domaines comme la météorologie, l'analyse des cours de bourses, etc. . . On cherche souvent à établir ou à vérifier une *corrélation* entre différents phénomènes : les fluctuations des uns sont-ils reliés aux variations des autres ?

Dans un grand nombre d'activités, le résultat d'une action ou d'un choix n'est pas parfaitement prévisible. Une mauvaise décision peut avoir des conséquences graves. Il faut donc *évaluer le risque* pris en fonction de ce choix. Par exemple, il est nécessaire d'estimer le pourcentage de pièces défectueuses dans une chaîne de production pour établir le coût d'un produit. Un assureur établira le montant de ses primes en fonction du risque qu'a l'assuré d'avoir un accident. Il est nécessaire de caractériser l'efficacité d'un nouveau médicament même si ses effets sont très variables d'un patient à l'autre. Dans tous ces domaines, le calcul des probabiltés est indispensable.

De plus, le calcul des probabilités est un outil absolument fondamental et incontournable dans l'analyse des phénomènes physiques. D'après les expériences réalisées par les physiciens à partir du début du vingtième siècle, il est impossible de prévoir parfaitement le résultat d'une expérience élémentaire quelle qu'elle soit : si par exemple une particule élémentaire peut parcourir deux chemins différents, on peut donner la probabilité qu'elle a de parcourir l'un ou l'autre chemin mais on ne peut jamais prédire le chemin qu'elle va prendre dans la prochaine expérience. Le calcul des probabilités est ainsi un des outils de base de la *mécanique quantique.*

De même, la théorie darwinienne de l'évolution fait l'hypothèse que les *modifications génétiques* apparaissant lors de la conception d'un nouvel être vivant sont aléatoires. La sélection naturelle ne conservant de ces modifications que celles qui induisent un comportement plus efficace de leur bénéficiaire. Les échanges entre cellules et plus particulièrement entre neurones sont aussi des phénomènes aléatoires. Le calcul des probabilités est ainsi un outil fondamental dans l'analyse des phénomènes biologiques les plus complexes, comme le fonctionnement du système nerveux central.

2 Les axiomes de base importants

On considère un ensemble de N événements possibles, par exemple les six résultats possibles d'un lancer de dé. On assigne un numéro à chacun de ces N événements. On associe à chacun de ces événements élémentaires un nombre positif ou nul $p(i)$ appelé *probabilité* qui est la proportion entre le nombre de fois où ce résultat apparait et le nombre total d'expériences. La somme des $p(i)$

sur l'ensemble des expériences vérifie :

$$\sum_{i=1}^{N} p(i) = 1. \tag{1}$$

Ceci traduit le fait que le résultat de l'expérience fera toujours partie des N événements possibles.

Les N événements sont *mutuellement exclusifs* ou *disjoints* : ils ne peuvent se réaliser simultanément. On ne peut obtenir à la fois un et trois lorqu'on lance un dé. Ou bien, si dans le couple (a, b), a représente le résultat d'un tirage à pile ou face (0 ou 1) et b le résultat d'un lancer de dé, les événements $(a = 0, b = 2)$ et $(a = 1, b = 4)$ sont disjoints. Si deux événements sont disjoints, la probabilité que l'un ou l'autre se réalise est

$$p(a \text{ ou } b) = p(a) + p(b). \tag{2}$$

Les résultats a et b de deux expériences sont *indépendants* si la probabilité de les obtenir simultanément est

$$p(a \text{ et } b) = p(a).p(b). \tag{3}$$

Si par exemple, un lancer de dé et un tirage à pile ou face sont indépendants,

$$p(a = 0 \text{ et } b = 6) = p(a = 0).p(b = 6). \tag{4}$$

La décomposition d'un événement en événements disjoints permet d'écrire le *théorème des probabilités totales*

$$p(a \text{ ou } b) = p(a) + p(b) - p(a \text{ et } b). \tag{5}$$

La mise en évidence de la dépendance au sens des probabilés est fondamentale dans les application : dans de nombreux problèmes, lorsqu'on tente de faire des prédictions, il est essentiel de pouvoir dire si le résutat d'une expérience Y peut dépendre d'un autre phénomène X. Dans l'affirmative, il faut pouvoir préciser l'effet de X sur Y.

Cette notion de dépendance nécessite la définition des *probabilité conditionnelles* : la probabilité que b se réalise sanchant que a s'est réalisé est

$$p(a/b) = \frac{p(a \text{ et } b)}{p(a)}. \tag{6}$$

Cette notion de probabilité conditionnelle est la base de la prédiction au sens statistique du résultat d'une expérience. On peut par exemple donner la probabilité qu'une des "causes" parmi N causes possibles d'un événement b soit celle qui a effectivement produit l'effet mesuré :

On suppose que les N causes possibles $(i = 1 \dots N)$ sont mutuellement exclusives et qu'on les connait toutes (il n'y a pas d'autres causes possibles). On connait la probabilité de chacune d'entre elles, $p(i)$, qui vérifie

$$\sum_{i=1}^{N} p(i) = 1, \tag{7}$$

car on connait toutes les causes. On connait aussi les probabilités conditionnelles $p(b/i)$. La probabilité que b se réalise est

$$p(b) = \sum_{i=1}^{N} p(i \text{ et } b). \tag{8}$$

En utilisant les deux écritures possibles de la formule donnant les probabilités conditionnelles,

$$p(i \text{ et } b) = p(i/b)p(b) = p(b/i)p(i), \tag{9}$$

on obtient d'une part,

$$p(b) = \sum_{i=1}^{N} p(b/i)p(i), \tag{10}$$

et en éliminant $p(b)$ dans l'eq. (9) :

$$p(i/b) = \frac{p(b/i)p(i)}{\sum_{i=1}^{N} p(b/i)p(i)}. \tag{11}$$

Ce résultat est connu sous le nom de formule de Bayes.

Une notion couramment utilisée quand on étudie les probabilités conditionnelles est la maximisation de *fonction de vraisemblance* : on recherche le maximum sur tous les i de $p(b/i)$.

3 Les variables aléatoires réelles

Dans la grande majorité des cas, on peut associerà un événement une mesure qui est un nombre réel, par exemple une mesure de longueur ou de température. Pour caractériser la variable, on commence par décourper l'intervalle des réels en segments $[x_k, x_{k+1}[$. Ensuite on effectue N mesures et on construit un *histogramme* : on compte le nombre de fois n_k où le résultat de la mesure, v appartient à chacun de ces segments (fig. 1). On tient compte du fait que les segments couvrent l'axe des réels et sont disjoints et on définit la probabilité

$$proba(v \in [x_k, x_{k+1}[= lim_{N\to\infty} \frac{n_k}{N}. \tag{12}$$

On réduit alors la taille des intervalles en faisant tendre leur nombre vers l'infini.

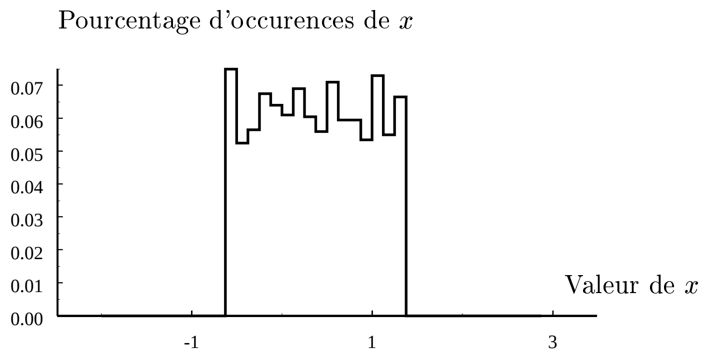

FIG. 1: Exemple d'histogramme. Le nombre de valeurs quantifiées de la variable x est 16

3.1 Fonction de répartition

On définit la *fonction de répartition* qui est la fonction $F(x)$ donnant la probabilité pour que la valeur mesurée x' soit strictement inférieure à x.

$$F(x) = proba(x' < x). \tag{13}$$

La figure 2 donne un exemple de fonction de répartition pour la loi de Cauchy où $F(x) = \frac{1}{2} + \frac{arctg(x)}{\pi}$.

C'est une fonction non décroissante qui peut présenter des paliers et des discontinuités (à droite). Elle est continue à gauche. Elle tend vers 0 lorque $x \to -\infty$ et vers 1 lorsque $x \to +\infty$.

Il est parfois intéressant de penser à étudier la fonction de répartition pour trouver la solution de certains problèmes élémentaires. Nous en verrons un exemple pour traiter le cas des changements de variables.

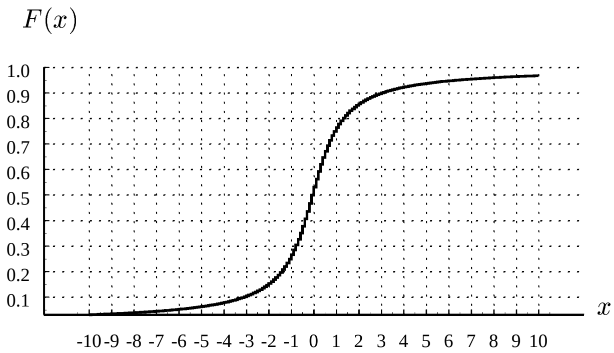

FIG. 2: Exemple de fonction de répartition, la loi de Cauchy

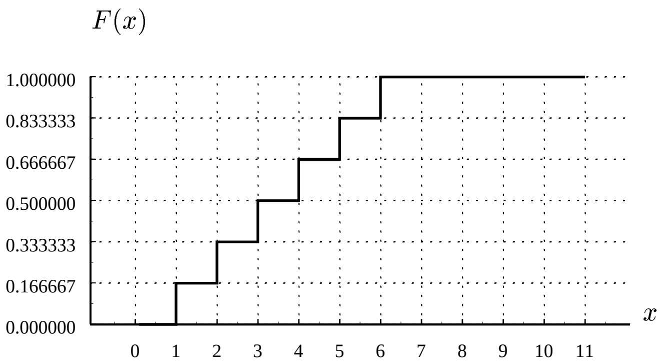

FIG. 3: Fonction de répartition pour le lancer de dé

3.2 Densité de probabilité

La probabilité que x appartienne à l'intervalle $[x, x+dx[$ est

$$proba(v \in [x, x+dx[) = dF(x), \tag{14}$$

et si $F(x)$ est différentiable,

$$proba(v \in [x, x+dx[) = p(x)dx. \tag{15}$$

$p(x)$, ***densité de probabilité*** de la variable aléatoire x est ainsi la dérivée de la fonction de répartition $F(x)$. C'est une fonction non négative mais ce n'est pas une probabilité, elle n'est pas nécessairement

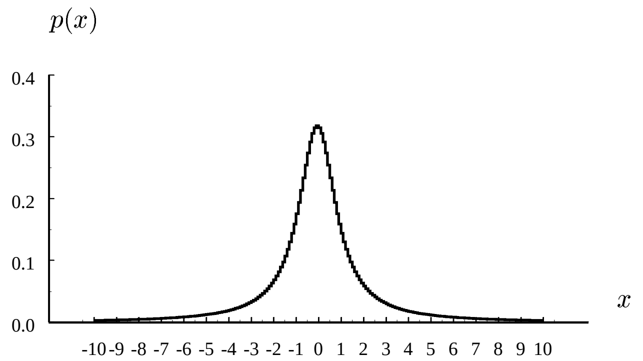

FIG. 4: Exemple de densité de probabilité, la loi de Cauchy : $p(x) = \frac{1}{\pi}\frac{1}{1+x^2}$

inférieure à 1. (Dans de nombreux ouvrages, on conserve la notation $dF(x)$, plu scohérente, qui tient compte des éventuelles discontinuités de la fonction de répartition. On peut estimer la densité de probabilité en construisant un histogramme pour lequel le pas d'échantillonnage de l'axe des réels est infiniment petit et le nombre d'expériences infiniment grand. Dans l'exemple de la figure 5, le nombre d'échantillons est de 2000, ensuite 20 000 et enfin 200 000.

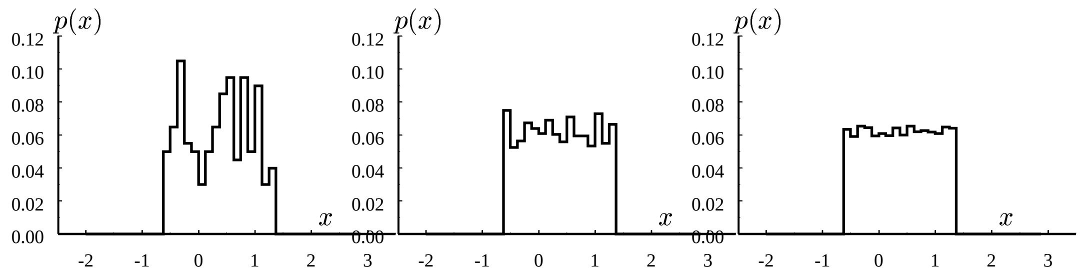

FIG. 5: Convergence de l'histogramme vers la densité de probabilité

3.3 Changement de variables

Soit une variable aléatoire x à partir de laquelle on construit une seconde variable $y = g(x)$. $g(x)$ est une fonction strictement monotone et admet une fonction inverse $h(y)$. Si la fonction de répartition de x est $F_x(x)$ et celle de y est $F_y(y)$, on a

$$F_y[g(x)] = F_x(x) \tag{16}$$

En dérivant cette relation, en admettant que $F(x)$ le permet,

$$p_y[g(x)]\frac{dg}{dx} = p_x(x), \tag{17}$$

ou encore, en exprimant ce résultat en fonction de la nouvelle variable

$$p_y[y] = \frac{dh}{dy} p_x(h(y)). \tag{18}$$

En particulier dans une transformation de la forme $y = ax + b$

$$p_y[y] = \frac{1}{a} p_x(\frac{y-b}{a}). \tag{19}$$

3.4 Loi de probabilités couramment utilisées

La loi de Bernouilli x prend la valeur 0 avec la probabilité r et la valeur 1 avec la probabilité $(1-r)$. Comme toute variable aléatoire discrète, elle ne possède pas de densité de probabilité, il faut avoir recours à la distribution de Dirac

$$p(x) = r\delta(x) + (1-r)\delta(1-x). \tag{20}$$

Loi uniforme

x prend des valeurs uniquement dans l'intervalle $[a, b[$. Sa densité de probabilité est nulle en dehors de cet intervalle et y vaut

$$p(x) = \frac{1}{b-a}. \tag{21}$$

Sa fonction de répartition est composée de trois segments de droites

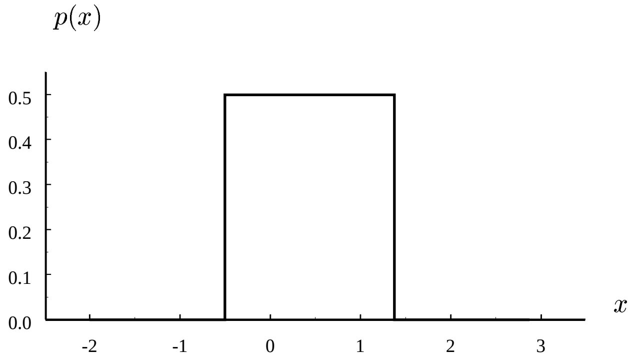

FIG. 6: Densité de probabilité de la loi uniforme :$a = -0.5$,$b = 1.5$

$$\begin{aligned} F(x) &= 0 \quad \text{pour} \quad x < a, \\ F(x) &= \frac{x-a}{b-a} \quad \text{pour} \quad a \leq x < b, \\ F(x) &= 1 \quad \text{pour} \quad x \leq b. \end{aligned} \tag{22}$$

Loi normale ou gaussienne

Sa densité de probabilité est

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]. \tag{23}$$

La densité de probabilité passe par son maximum pour $x = m$. Elle a ses points d'inflexion en $m + \sigma$ et $m - \sigma$. Cette loi présente trois caractéristiques importantes :

• Elle permet des développements mathématiques efficaces, en particulier en ce qui concerne les variables aléatoires indépendantes parce qu'elle s'exprime comme l'exponentielle d'une forme quadratique (ceci sera précisé lorsque nous présenterons l'analyse des couples de variables aléatoires.

• Toute l'information est donnée directement par les paramètres m et σ qui caractérisent respectivement la valeur moyenne et la dispersion autour de cette valeur moyenne.

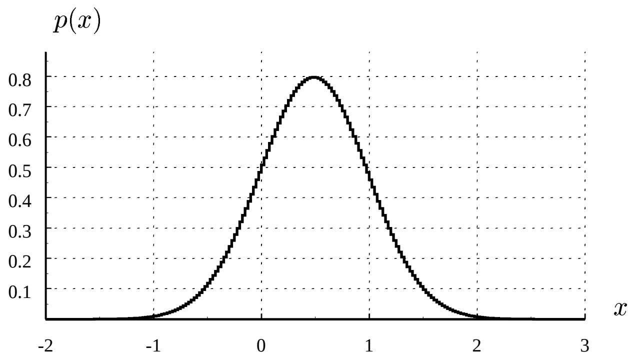

FIG. 7: Densité de probabilité de la loi gaussienne

• Enfin, c'est la loi qu'on obtient naturellement en additionnant un grand nombre de variables aléatoires indépendantes ; elle représente bien un grand nombre de phénomènes physiques aléatoires comme le bruit de fond, les erreurs entâchant les mesures, etc...

Loi de Poisson

C'est la loi qu'on obtient en comptant des évènements aléatoires indépendants : arrivée d'une particule sur un capteur, décompte des appels sur un central téléphonique, comptage de voitures sur une route, etc.. C'est la loi centrale des études sur les files d'attente dans les réseaux de communication. Elle dépend d'un paramètre λ qui caractérise le débit moyen du flux. La probabilité d'observer k événements pendant une durée t est donnée par

$$p(k,t) = \frac{(\lambda t)^k}{k!}\exp(-\lambda t), \tag{24}$$

pour $k \geq 0$ et $t \geq 0$, où on prend $0! = 1$.

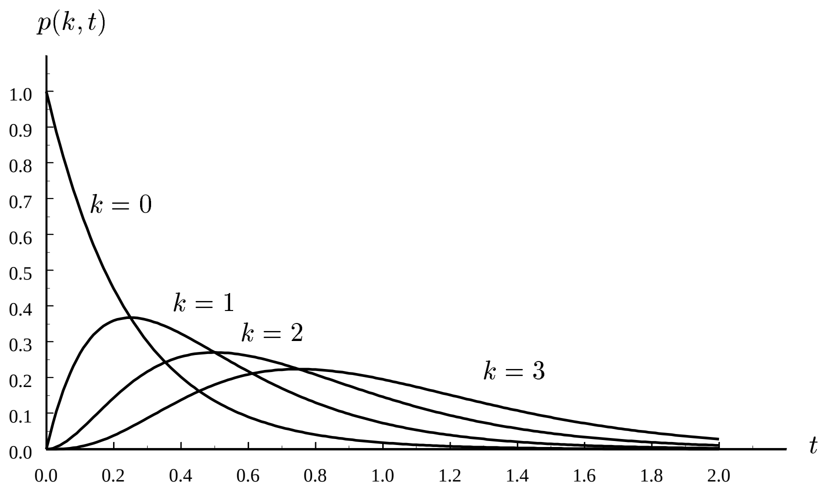

FIG. 8: Densités de probabilités $p(k,t)$ de la loi de Poisson

3.5 exercices

1. Tracer la fonction de répartition associée au jeu de pile ou face (0 avec la probabilité p et 1 avec la probabilité $(1-p)$.

2. L'aiguille de Buffon : on trace sur le sol des lignes parallèles équidistantes séparées d'une distance ℓ. On laisse tomber au hasard une aiguille de même longueur ℓ sur le sol. Enoncer les hypothèses permettant de bien formuler le problème et trouver la probabilité pour que l'aiguille touche ou ne touche pas les droites.

3. On peut engendrer grâce à un générateur de nombres pseudo-aléatoires une variable aléatoire y équirépartie entre 0 et 1. Montrer qu'il est possible, en appliquant à y la transformation

$$x = F^{-1}(y), \tag{25}$$

d'obtenir une variable aléatoire dont la fonction de répartition est $F(x)$.

4 Caractérisation de variables aléatoires par les moments

Il est souvent trop coûteux et guère utile de caractériser une variable aléatoire ou un ensemble de variables aléatoires par leur fonction de répartition ou leur densité de probabilité. On peut se contenter d'informations partielles comme la valeur moyenne de la variable aléatoire et sa dispersion autour de cette valeur moyenne. Ces notions sont quantifiées par la définition des moments.

4.1 Moment du premier ordre, Moyenne

Lorsqu'on dispose d'un certain nombre N de résultats x_n pour $n = 1, \ldots, N$, obtenus en répétant plusieurs fois une mesure, il paraît naturel d'estimer la valeur moyenne du résultat en calculant

$$m_N = \frac{1}{N} \sum_{n=1}^{N} x_n. \tag{26}$$

Si on construit un histogramme de la variable x, le nombre de fois où le résultat de la mesure appartient à l'intervalle $[x, x + dx[$ sera donné par $Np(x)dx$. En supposant que dx est petit, on peut dans l'expression précédente de la moyenne remplacer x_n par x lorsque x_n est dans l'intervalle $[x, x + dx[$. Alors, quand $N \to \infty$ m_N tend vers

$$E(x) = m = \int_{-\infty}^{\infty} xp(x)dx. \tag{27}$$

Lorsque la variable aléatoire est discrète, et prend les valeurs x_k avec les probabilités p_k, la moyenne est

$$E(x) = \sum_{k=0}^{\infty} x_k p_k. \tag{28}$$

Si la fonction de répartition de x présente des discontinuités, on peut réécrire la formule donnant la moyenne à partir de l'intégrale (dite de Stiljès) faisant apparaître cette fonction de répartition et non la densité de probabilité :

$$E(x) = \int_{-\infty}^{\infty} xdF(x). \tag{29}$$

La première des formules donne la méthode pratique pour calculer une estimation de la valeur moyenne à partir d'un nombre fini de mesures tandis que les trois autres donnent des expressions en fonction des densités de probabilité, expressions nécessaires pour les développements théoriques.

Lorsqu'on peut calculer ou estimer la valeur moyenne d'une variable aléatoire, il est souvent judicieux d'en déduire une nouvelle variable en lui soustrayant cette valeur moyenne de manière à ne pas alourdir les calculs. Cfette seconde variable est appelée variable *centrée*. La figure 9 montre deux exemples de variables aléatoires gaussiennes dont les moyennes m sont respectivement 2 et 1.5 et les écarts-type σ 1 et 0.5. Les marques indiquées sont aux ordonnées $m - \sigma$, m et $m + \sigma$.

4.2 Moment du deuxième ordre, variance, écart-type

La deuxième caractérisation importante d'une variable aléatoire est sa dispersion autour de la valeur moyenne. On peut l'estimer en calculant un "moment d'inertie" de la variable par rapport à cette valeur moyenne. Par exemple à partir de N mesures, on calculera

$$\sigma_N^2 = \frac{1}{N} \sum_{n=1}^{N} (x_n - m)^2. \tag{30}$$

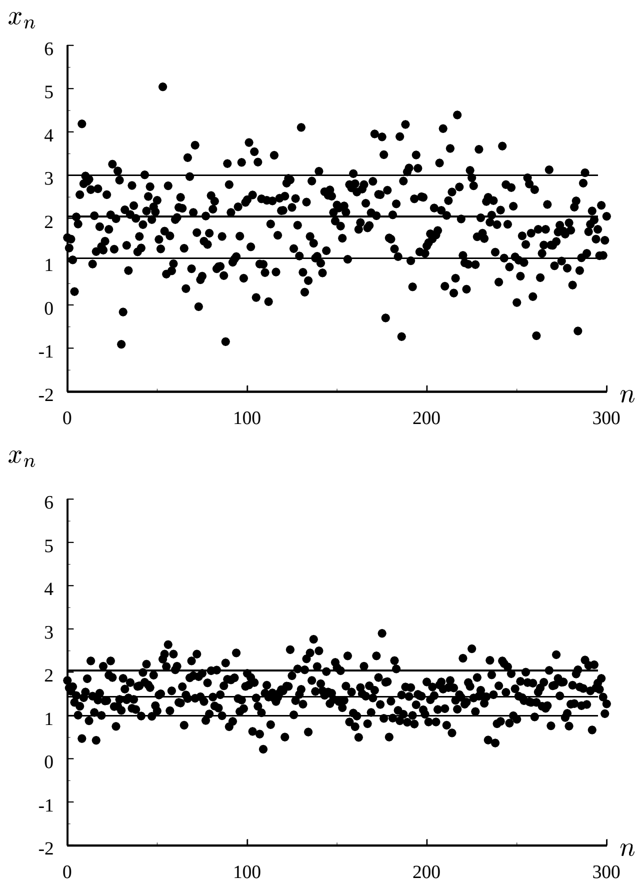

FIG. 9: Illustration de la moyenne et de la dispersion d'une variable aléatoire gaussienne. Sur chacune des figures, les droites horizontales indiquent la valeur moyenne μ, la valeur moyenne après soustraction et addition de l'écart-type, soit $\mu - \sigma$ et $\mu + \sigma$.

Comme dans le cas de la moyenne, l'affinement de l'histogramme lorsque N tend vers l'infini conduit à une formule faisant apparaître la densité de probabilité

$$\sigma^2 = \int_{-\infty}^{\infty} (x-m)^2 p(x) dx. \tag{31}$$

On l'écrit en fonction du *moment du deuxième ordre*

$$E(x^2) = \int_{-\infty}^{\infty} x^2 p(x) dx, \tag{32}$$

soit

$$\sigma^2 = E(x^2) - [E(x)]^2, \tag{33}$$

résultat obtenu par un développement de (31) où on fait apparaitre l'expression (27) de la moyenne $E(x)$. σ^2 est la “variance” de la variable aléatoire. Sa racine carrée, σ est son écart-type. Cette quantité caractérise la dispersion de la variable autour de la moyenne. Lorsque la variance est nulle, la densité de probabilité est concentrée en un point et la variable aléatoire est connue avec certitude.

La notion de variance est fondamentale dans un bon nombre d'applications en particulier en traitement du signal, et aussi pour l'analyse des variables gaussiennes.

Un résultat utile concernant la dispersion d'une variable aléatoire est l'inégalité de Bienaymé-Tchébichev : la probabilité pour que l'écart à la moyenne soit inférieure à une borne $|a|\sigma$ est minorée par

$$p(|x-m| < |a|\sigma) \geq 1 - \frac{1}{a^2}. \tag{34}$$

4.3 Moments d'ordre supérieur à deux

De manière générale, on peut être amené à calculer des quantités de la forme

$$E(x^k) = \int_{-\infty}^{\infty} x^k p(x) dx. \tag{35}$$

Ce sont les moments de la variable aléatoire. Dans le cas des variables aléatoires gaussiennes ils se déduisent des moments du premier et du deuxième ordre. Il présentent un grand intérêt pour la séparation de variables aléatoires indépendantes non gaussiennes. Une autre caractérisation importante d'une densité de probabilité est son *entropie*

$$\int_{-\infty}^{\infty} p(x) \log[p(x)] dx, \tag{36}$$

dont le rôle est fondamental en théorie de l'information.

4.4 Fonction caractéristique

Certains développements sont considérablement facilités par le calcul de la transformée de Fourier de la densité de probabilité. C'est en particulier le cas lorsqu'on cherche à mettre en évidence l'indépendance de variables aléatoires. On appelle cette transformée de Fourier la “fonction caractéristique” et on la note ($j^2 + 1 = 0$)

$$C(u) = E[\exp(jux)] = \int_{-\infty}^{\infty} exp(jux) p(x) dx = \int_{-\infty}^{\infty} exp(jux) dF(x). \tag{37}$$

On notera que cette fonction caractéristique permet d'étudier aussi bien les variables aléatoires à valeurs discrètes que les variable aléatoires réelles.

Si la densité de probabilité est une fonction paire (et donc de moyenne nulle) sa fonction caractéristique est réelle.

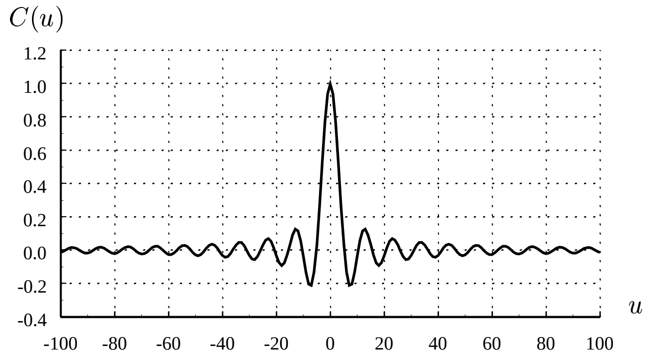

FIG. 10: Fonction caractéristique de la variable équirépartie

Si on effectue un développement en série de Taylor ou de MacLaurin au voisinage de l'origine de la fonction $\exp(jux)$, on voit apparaitre dans le développement de la fonction caractéristique que le facteur de $\frac{(ju)^k}{k!}$ est le moment d'ordre k de la variable aléatoire

$$E[\exp(jux)] = \int_{-\infty}^{\infty} \sum_{k=0}^{\infty} \frac{(ju)^k}{k!} x^k p(x) dx, \tag{38}$$

$$E[\exp(jux)] = \sum_{k=0}^{\infty} \frac{(ju)^k}{k!} E(x^k). \tag{39}$$

La fonction caractéristique d'une variable aléatoire gaussienne s'écrit

$$C(u) = \exp(jum - \frac{\sigma^2 u^2}{2}). \tag{40}$$

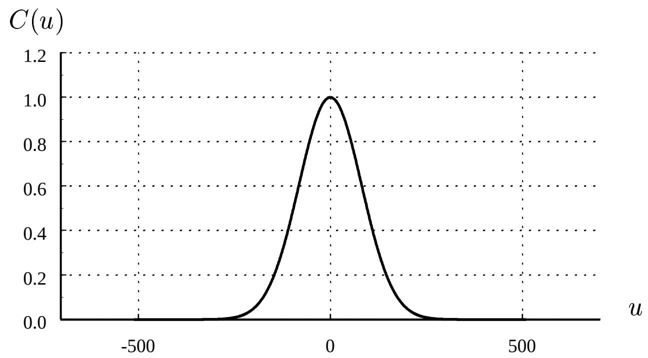

FIG. 11: Fonction caractéristique de la variable gaussienne centrée

Si on multiplie une variable aléatoire par une constante

$$y = ax, \tag{41}$$

alors

$$p_y(y) = \frac{1}{a} p_x(\frac{y}{a}). \tag{42}$$

En en prenant la fonction caractéristique

$$C_y(u) = \int_{-\infty}^{\infty} e^{juy} \frac{1}{a} p_x(\frac{y}{a}) dy, \tag{43}$$

$$C_y(u) = \int_{-\infty}^{\infty} e^{juax} p_x(x) dx, \tag{44}$$

$$C_y(u) = C_x(au). \tag{45}$$

Ce résultat sera utilisé lors de l'étude de la tendance vers la loi gaussienne de la somme d'un grand nombre de variables aléatoires.

En réécrivant cette intégrale en fonction de la variable x On peut aussi effectuer le développement du *logarithme* $\log[C(u)]$de la fonction caractéristique. On met alors en évidence les *cumulants* de la variable aléatoire. Les cumulants d'ordre supérieur à deux sont nuls lorsque la variable aléatoire est gaussienne, ils jouent un rôle important dans l'étude des variables aléatoires non gaussiennes, en particulier pour l'analyse de leur indépendance.

4.5 Exercices

1. Montrer que si une variable aléatoire a pour moyenne m, la variable aléatoire $y = x - m$ a une moyenne nulle.

2. Montrer que si une densité de probabilité $p(x)$ vérifie

$$p(m + x) = p(m - x), \tag{46}$$

sa moyenne vaut m. En déduire la moyenne d'une variable aléatoire gaussienne et d'une variable aléatoire uniforme.

3. Vérifier que la valeur moyenne de la variable de Bernouilli est $(1 - p)$.

4. Calculer la moyenne d'une variable aléatoire x positive dont la densité de probabilité vaut $\exp -\lambda x$.

5. Vérifier que la valeur moyenne pour t donné de la loi de Poisson $(\sum_{k=0}^{\infty} kp(k,t))$ vaut λt.

6. Quelle est la moyenne de la variable aléatoire de densité de probabilité

$$p(x) = \frac{2}{\pi(1+x)\sqrt{x}} \text{ pour } x > 0 \tag{47}$$

Montrer que la fonction de répartition de x est $\frac{2}{\pi} \arctan(\sqrt{x})$. Où se trouve concentrée la densité de probabilité de cette variable.

7. Quel est le moment du deuxième ordre d'une variable de loi uniforme entre 0 et a ? Quelle est sa variance ?

8. Quelle est la fonction caractéristique de la variable aléatoire équirépartie entre 0 et 1 ? En déduire la moyenne et la variance de cette loi en effectuant son développement en série de Taylor.

9. Quelle est la fonction caractéristique de la variable aléatoire prenant les deux valeur -1 et $+1$ avec la probabilité $1/2$?

5 Analyse de couples de variables aléatoires centrées

Un des objectifs fondamentaux du calcul des probabilités est d'établir un lien éventuel entre plusieurs résultats d'expériences. On est ainsi amené à étudier le comportement conjoint de plusieurs variables aléatoires comme le poids et la taille des individus d'un groupe, ou bien l'age des conjoints d'un couple, etc...

Nous étudierons le cas des couples de variables aléatoires x et y. Tout comme dans le cas des variables aléatoires réelles, on peut construire un histogramme en comptant le nombre de fois où l'une d'entr'elles appartient à l'intervalle $[x_0, x_0 + dx[$ et l'autre à l'intervalle $[y_0, y_0 + dy[$. En affinant cet histogramme lorsque le nombre de mesures tend vers l'infini et lorsque la taille des intervalles tend vers 0, on obtient la densité de probabilité du couple de variables aléatoires $p(x, y)$. Cette densité est positive ou nulle et vérifie

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) dx dy = 1. \tag{48}$$

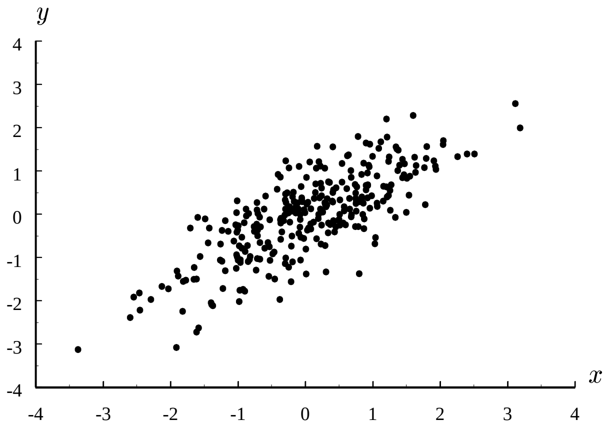

FIG. 12: Visualisation d'un couple de variables aléatoires sous la forme d'un nuage de points

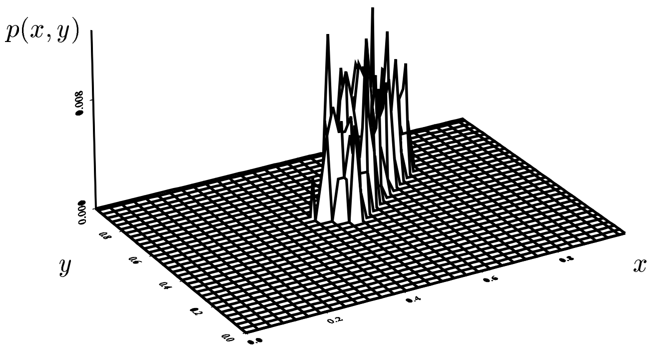

FIG. 13: Histogramme d'un couple de variables aléatoires

Chacune des deux variables aléatoire a une densité de probabilité, soit $p_x(x)$ et $p_y(y)$:

$$p_x(x) = \int_{-\infty}^{\infty} p(x,y)dy, \tag{49}$$

$$p_y(y) = \int_{-\infty}^{\infty} p(x,y)dx, \tag{50}$$

et sa propre variance soit σ_x et σ_y.

Si les deux variables aléatoires sont indépendantes :

$$p(x,y) = p_x(x)p_y(y). \tag{51}$$

5.1 Moment du deuxième ordre

Afin de simplifier la présentation, nous supposerons que les variables aléatoires sont centrées ce qui peut pratiquement toujours s'obtenir en effectuant une translation. La manière la plus élémentaire de caractériser une éventuelle relation entre x et y est de calculer les trois moments du deuxième ordre des deux variables aléatoires soit σ_x, σ_y et

$$E(xy) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyp(x,y)dxdy. \tag{52}$$

Si $E(xy)$ n'est pas nul, alors que x et y ont une moyenne nulle, il y a une corrélation entre les deux variables. Cependant le fait que $E(xy)$ soit nul ne veut pas dire qu'il y a indépendance entre les deux variables (on parle d'orthogonalité) : l'indépendance suppose la factorisation de la densité de probabilité du couple sous la forme (51). En pratique, si on dispose de N observations x_n et y_n, on calcule

$$E(xy) = \frac{1}{N}\sum_{n=1}^{N} x_n y_n. \tag{53}$$

On définit ainsi le *coefficient de corrélation* entre x et y :

$$r = \frac{E(xy)}{\sigma_x \sigma_y}. \tag{54}$$

D'après l'inégalité de Schwarz, le coefficient r est compris entre -1 et 1. Si $|r| = 1$, on peut affirmer qu'il y a une dépendance linéaire entre x et y :

$$y = ax. \tag{55}$$

Plus $|r|$ est proche de 1, plus la corrélation est forte. Dans ce cas les courbes de niveau de la densité de probabilité du couple de variables aléatoires ont une forme d'ellipse très allongées. Au contraire lorsqu'on a normalisé les variables de sorte que $\sigma_x = \sigma_y$ lorsque $r = 0$, ces courbes de niveau ont en général (mais pas toujours) une forme circulaire.

Remarque : affirmer qu'une variable aléatoire est corrélée à une autre ne permet pas de dire que l'une est la cause de l'autre. On peut raisonnablement dire que parmi les nombreuses causes qui influent sur leur comportement il y en a de communes.

On peut définir de la même manière les différents moments du couple de variables aléatoires ainsi que la fonction caractéristique bidimensionnelle, qui a un rôle utile dans l'étude des couples de variables aléatoires.

5.2 Somme de variables aléatoires indépendantes

Soient deux variables aléatoires indépendantes x et y prenant des valeurs entières. La porbabilité pour que $x = m$ et $y = n$ simultanément est $p_x(m)p_y(n)$. Soit $z = x + y$. Sa densité de probabilité est

$$p_z(k) = \sum_{m=-\infty}^{\infty} p_x(m)p_y(k-m). \tag{56}$$

De manière générale la densité de probabilité de la somme s'écrit sous la forme d'une convolution.

$$p_z(z) = \int_{-\infty}^{\infty} p_x(x)p_y(z-x)dx. \tag{57}$$

Par exemple si sont des variables aléatoires équiréparties entre 0 et 1 La fonction caractéristique

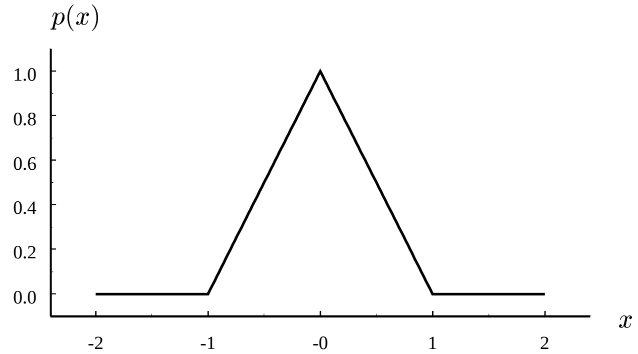

FIG. 14: Densité de probabilité de la somme de deux variables aléatoires indépendantes équiréparties entre -1/2 et 1/2

de z est la transformée de Fourier d'un produit de convolution. C'est donc le produit des fonctions caractéristiques de x et z :

$$E[exp(juz)] = E[exp(jux)]E[exp(juy)]. \tag{58}$$

On en déduit que la moyenne des sommes la somme des moyennes est égale à et que la variance de la somme est égale à la somme des variances

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2. \tag{59}$$

Cependant si les deux variables ne sont pas indépendantes, on peut seulement écrire que

$$p_z(k) = \int_{-\infty}^{\infty} p(x, z-x)dx. \tag{60}$$

5.3 Cas des variables aléatoires gaussiennes

La densité de probabilté d'un couple de variables aléatoires centrées gaussiennes dans leur ensemble est donnée par

$$p(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \exp -\frac{1}{2(1-r^2)}(\frac{x^2}{\sigma_x^2} - \frac{2rxy}{\sigma_x\sigma_x} + \frac{y^2}{\sigma_y^2}) \tag{61}$$

formule qui fait apparaître les variances de chacune des deux variables ainsi que leur coefficient de corrélation ; on notera que l'exposant de e est une forme quadratique définie positive lorsque $r \neq 1$.

Si le coefficient r est nul, $p(x,y)$ s'écrit comme un produit

$$p(x,y) = \left(\frac{1}{\sigma_x\sqrt{2\pi}} \exp -\frac{1}{2}\frac{x^2}{\sigma_x^2}\right) \times \left(\frac{1}{\sigma_y\sqrt{2\pi}} \exp -\frac{1}{2}\frac{y^2}{\sigma_y^2}\right). \tag{62}$$

Les variables sont donc indépendantes.

Deux variables aléatoires gaussiennes orthogonales sont nécessairement indépendantes. On peut toujours représenter un couple de variables aléatoires gaussiennes comme une combinaison linéaire de variables aléatoires gaussiennes indépendantes. Un changement de variables par rotation

$$\begin{aligned} X &= x\cos\theta + y\sin\theta, \\ Y &= -x\sin\theta + y\cos\theta, \end{aligned} \tag{63}$$

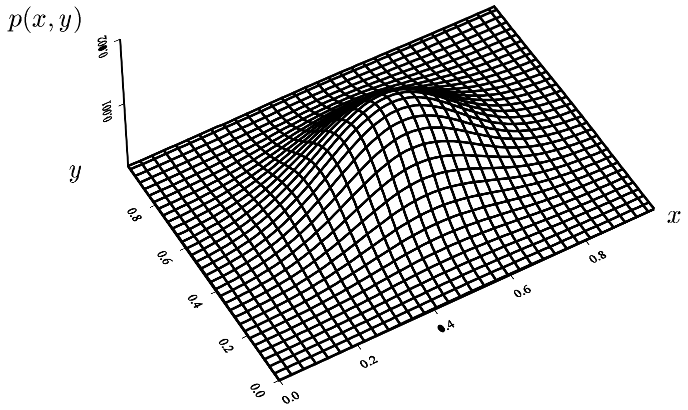

FIG. 15: Densité de probabilité d'un couple de variables aléatoires gaussiennes

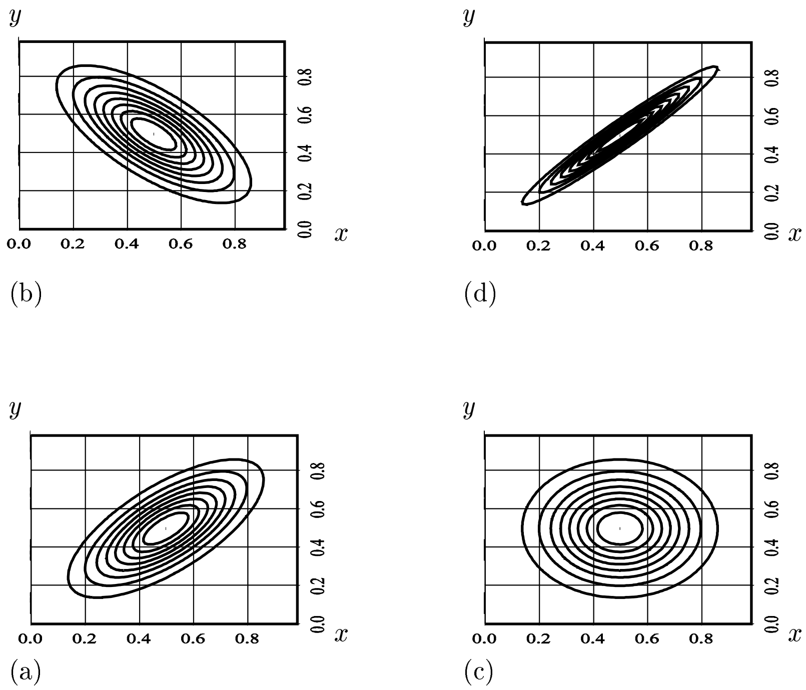

FIG. 16: Courbes de niveau de la densité de probabilité d'un couple de variables aléatoires gaussiennes : le coefficient de corrélation r prend pour valeur (a) 0.7 , (b) -0.7 (corrélation négative) , (c) 0 (indépendance), (d) 0.99 (corrélation forte)

permet d'obtenir deux variables orthogonales ($E(XY)=0$). Par conséquent la densité du couple de variables aléatoires (X,Y) p_{XY} se factorise sous la forme

$$p_{XY}(X,Y)=p_X(X)p_Y(Y). \tag{64}$$

La densité de probabilité conditionnelle de y connaissant x s'écrit

$$p(y/x)=\frac{1}{\sqrt{2\pi(1-r^2)}\sigma_y}\exp-\frac{1}{2(1-r^2)}(\frac{y}{\sigma_y}-r\frac{x}{\sigma_x})^2. \tag{65}$$

Lorsque x est finé, y est une variable aléatoire gaussienne non centrée de moyenne $r\frac{\sigma_y}{\sigma_x}$ et de variance $\sigma_y^2(1-r^2)$. La somme de deux variables aléatoires gaussiennes dont les variances sont respectivement σ_x^2 et σ_y^2 et dont le coefficient de corrélation est r est une variable aléatoire gaussienne de variance $(\sigma_x^2+2r\sigma_x\sigma_y+\sigma_y^2)$.

6 Somme d'un grand nombre de variables aléatoires indépendantes centrées

Lorsqu'on somme deux variables aléatoires indépendante la densité de probabilité du résultat est la convolution des densités de probabilité des deux termes de la somme. La fonction caractéristique de la somme est le produit des deux fonctions caractéristiques. Ce résultat est fondamental pour l'étude de somme d'un grand nombre de variables que nous supposerons indépendantes.

6.1 Décroissance de la dispersion de l'estimation d'une variable aléatoire par calcul d'une moyenne

Quand on estime une valeur, par exemple la moyenne d'une variable aléatoire à partir du résultat de N expériences différentes, on calcule une nouvelle variable aléatoire sous la forme

$$m_N=\frac{1}{N}\sum_{n=1}^{N}x_n. \tag{66}$$

Nous supposerons que x est une variable aléatoire centrée et que tous les x_n suivent la même loi de probabilité. Pour étudier la variance de m_N, nous utiliserons la fonction caractéristique : la variable

$$y_N=\sum_{n=1}^{N}x_n, \tag{67}$$

a pour fonction caractéristique

$$C_y(u)=E(e^{juy_N})=\left[E(e^{jux})\right]^N=[C_x(u)]^N \tag{68}$$

qu'on peut écrire en utilisant les logarithmes

$$\log C_y(u)=N\log C_x(u). \tag{69}$$

Nous supposons que x est une variable aléatoire centrée. Le premier terme du développement en série de Taylor donne

$$-\sigma_y^2\frac{u^2}{2}=-N\sigma_x\frac{u^2}{2}, \tag{70}$$

et la variance de y_N est $N\sigma_x^2$. Comme m_N est obtenue en divisant y_n par N la variance de m_N vaut

$$\sigma_{m_N}^2=\frac{1}{N^2}\sigma_{y_N}^2=\frac{1}{N}\sigma_x^2. \tag{71}$$

La relation entre les écarts-type est

$$\sigma_{m_N}=\frac{1}{\sqrt{N}}\sigma_x. \tag{72}$$

C'est une quantité qui tend vers 0 lorsque $N\to\infty$: l'incertitude sur le résultat décroit avec le nombre d'expériences. Il faut cependant remarquer que cette décroissance est très lente : pour améliorer la précision d'un facteur dix, il faut multiplier le nombre d'expériences par cent.

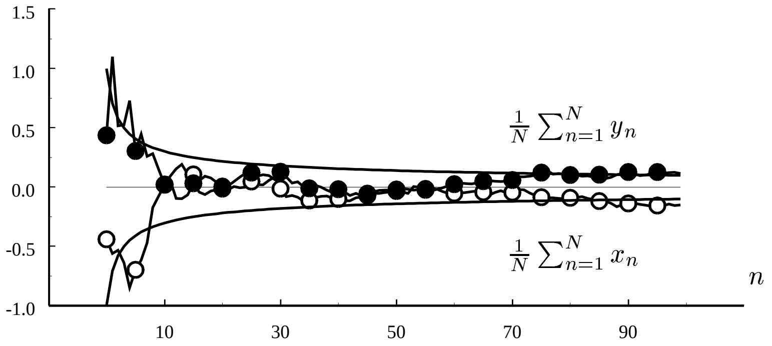

FIG. 17: Evolution de la moyenne de N réalisations de deux variables aléatoires, x et y, en fonction de N

6.2 Tendance vers la loi gaussienne

Si dans l'exemple précédent, au lieu de calculer m_N, on étudie

$$z = \frac{1}{\sqrt{N}} y_N, \tag{73}$$

la relation entre les fonctions caractéristiques devient

$$C_z(u) = C_y(\frac{u}{\sqrt{N}}) = \left[C_x(\frac{u}{\sqrt{N}})\right]^N. \tag{74}$$

Le développement du logarithme de $C_y(u)$ donne

$$\log C_y(u) = -N\sigma_x^2 \frac{u^2}{2} + O(u^2), \tag{75}$$

et donc

$$\log C_z(u) = -\sigma_x^2 \frac{u^2}{2} + \frac{1}{N} O(u^2). \tag{76}$$

Lorsque $N \to \infty$

$$\log C_z(u) \to -\sigma_x^2 \frac{u^2}{2}, \tag{77}$$

$$C_z(u) \to \exp -\sigma_x^2 \frac{u^2}{2}, \tag{78}$$

et en termes de densité de probabilité,

$$p_z(z) \to \frac{1}{\sigma_x \sqrt{2\pi}} \exp -\frac{z^2}{2\sigma_x^2}. \tag{79}$$

La somme d'un grand nombre de variables aléatoires indépendantes de même loi tend vers une variable gaussienne .

On peut généraliser ce résultat au cas où les variables ne sont pas de même loi et partiellement indépendantes. Ce qui justifie l'utilisation des variables aléatoires gaussiennes pour représenter les phénomènes aléatoires relativement complexes.

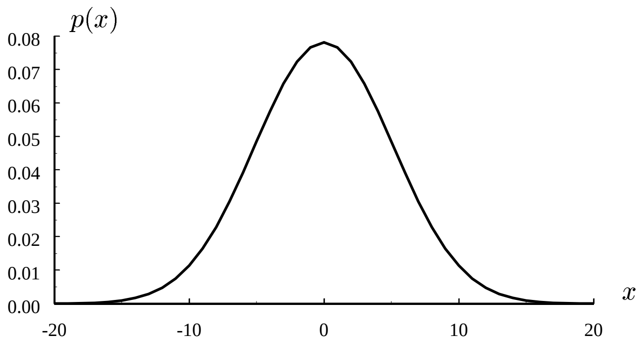

FIG. 18: La densité de probabilité de la somme de 32 variables aléatoires indépendantes équiréparties est très similaire à celle d'une gaussienne