Introduction à la statistique

S. LE BORGNE

Table des matières

1 Introduction

Le cours est intitulé Probabilités et Statistique 2 : il portera donc sur ces disciplines et se placera dans le prolongement du cours PS1. PS1 n'aborde quasiment pas la statistique [1] ; PS2 sera essentiellement consacré à la statistique. La première partie du cours ne fera usage d'aucun concept de probabilités. Elle présentera certains outils permettant de décrire un ensemble de données, de tirer une information synthétique d'un grand nombre de chiffres. C'est ce qu'on appelle en général la statistique *descriptive*. La deuxième partie du cours fera appel au calcul des probabilités (en particulier aux théorèmes limite). On postulera qu'un phénomène à étudier suit un modèle probabilistes et on cherchera à déduire certaines caractéristiques du modèle à partir d'une observation du passé (pour faire des prévisions sur l'avenir). C'est ce qu'on appelle en général la statistique *inférentielle*.

1.1 Programmes de l'enseignement secondaire

Les programmes des collèges et lycées font place à la statistique. Le cours reviendra sur des notions vues dans le secondaire.

L'exploitation de séries de mesures, la réflexion sur leur moyenne et leur dispersion, tant dans le domaine des sciences expérimentales que dans celui de la technologie introduisent l'idée de précision de la mesure et conduisent à une première vision statistique du monde. Comprendre les moyens préventifs ou curatifs mis au point par l'homme introduit à la réflexion sur les responsabilités individuelles et collectives dans le domaine de la santé. Une bonne compréhension de la pensée statistique et de son usage conduit à mieux percevoir le lien entre ce qui relève de l'individu et ce qui relève du grand nombre : alimentation, maladies et leurs causes, vaccination. Elle vise aussi, à travers des thèmes tels que la météorologie ou l'énergie mais aussi la pensée statistique, à faire prendre conscience de ce que la science est plus que la simple juxtaposition de ses disciplines constitutives et donne accès à une compréhension globale d'un monde complexe notamment au travers des modes de pensée qu'elle met en œuvre.

Importance du mode de pensée statistique dans le regard scientifique sur le monde. L'aléatoire est présent dans de très nombreux domaines de la vie courante, privée et publique : analyse médicale qui confronte les résultats à des valeurs normales, bulletin météorologique qui mentionne des écarts par rapport aux normales saisonnières et dont les prévisions sont accompagnées d'un indice de confiance, contrôle de qualité d'un objet technique, sondage d'opinion... Or le domaine de l'aléatoire et les démarches d'observations sont intimement liés à la pensée statistique. Il s'avère donc nécessaire, dès le collège, de former les élèves à la pensée statistique dans le regard scientifique qu'ils portent sur le monde, et de doter

1. Le pluriel *les statistiques* ou *des statistiques* désigne des données, le singulier la science qui développe des concepts et des outils permettant l'étude de données.

les élèves d'un langage et de concepts communs pour traiter l'information apportée dans chaque discipline.

Au collège, seule la statistique exploratoire est abordée et l'aspect descriptif constitue l'essentiel de l'apprentissage. Trois types d'outils peuvent être distingués :
- les outils de synthèse des observations : tableaux, effectifs, regroupement en classe, pourcentages, fréquence, effectifs cumulés, fréquences cumulées,
- les outils de représentation : diagrammes à barres, diagrammes circulaires ou semi-circulaires, histogrammes, graphiques divers,
- les outils de caractérisation numériques d'une série statistique : caractéristiques de position (moyenne, médiane), caractéristiques de dispersion (étendue, quartiles).

Nous reviendrons sur ces notions de manière plus formelle qu'au collège et ajouterons quelques nouveaux outils. Une partie du cours sera vraiment élémentaire.

1.2 Problèmes de la statistique

On pourrait définir la statistique comme l'étude de la variabilité. Un ensemble de données est souvent résumé d'une façon ou d'une autre. La façon de résumer l'information peut avoir des conséquences importante. Dans son livre, l'éventail du vivant par exemple Stephen Jay Gould montre sur différents exemples que l'habitude de faire la moyenne arithmétique de données peut cacher des différences essentielles. Il insiste sur l'importance de prendre en compte l'ensemble de la distribution des données. De ce point de vue une représentation graphique simple comme un histogramme peut être très précieuse. Quand on cherche à comparer des données, il arrive que des indicateurs différents donnent des conclusions différentes. Il est important d'être averti de ce type de phénomènes. Cela ne doit pourtant pas mener à des conclusions trop sceptiques du type :« On peut faire dire ce qu'on veut aux chiffres », mais mener à une prudence raisonnée.

1.3 Histoire de la statistique

L'histoire de la statistique est souvent rapprochée de celle des probabilités. Nous pouvons n'en donner ici qu'un très bref aperçu. Nous nous aidons pour cela essentiellement des deux références [1], [4]

Les grands empires de l'Antiquité (Egypte, Mésopotamie, Chine) prenaient appui sur une sorte de statistique et notamment sur des recensements. Les statistiques remontent donc à plusieurs millénaires.

La réflexion sur le hasard et la fortune remonte elle aussi bien loin. La philosophie grecque ancienne en porte témoignage.

Les jeux de hasard sont connus depuis la préhistoire : on a retrouvé des dés préhistoriques

(en os).

Les mots utilisés pour désigner ces notions sont plus récents. Le mot *hasard* vient de Syrie au moyen-âge. L'étymologie n'est pas établie avec certitude ; le mot arabe d'origine pouvait désigner à la fois une fleur et une figure rare au jeu de dé (par exemple un triple six). Le mot statistique vient de l'italien. Il a été introduit en français au 16ème siècle.

Les probabilité modernes sont nées au 16ème siècle et 17ème siècle avec des savants italiens Luca Pacciolo (145-1514), Tartaglia (1499-1557), Cardan (1501-1576), Galilée (1564-1662), puis d'autres européens Fermat(1601-1665), Pascal (1623-1662), Huyghens (1629-1695), Jacques Bernoulli (1654-1705). Il n'est pas exclu que ces auteurs aient été précédés par certains arabes ou persans.

La combinatoire des jeux de hasard est explorée. Le triangle de Pascal est bien connu au 17ème siècle. Mais il l'était déjà mille ans plus tôt en Chine....

J. Bernoulli a établi la loi des grands nombres pour les variables de... Bernoulli ; De Moivre (1667-1754), britannique ancien huguenot français, le théorème limite central (aussi appelé théorème de De Moivre-Laplace) pour ces mêmes variables au début du 18ème siècle.

Bayes (1702-1761)a étudié le problème inverse ou probabilité des causes. Diverses conceptions de la notion de probabilité existent. L'une d'entre elle est dite bayesienne : en l'absence d'information on postule que les probabilités des différentes issues d'une expérience aléatoire sont équiprobables ; on modifie ensuite cette hypothèse au fur et à mesure que l'expérience nous renseigne. Objectivité, subjectivité. Fréquence, information. Symétrie.

À la fin du 19ème siècle les probabilités ont été appliquées (pas toujours avec ...). Quetelet. « Ces deux illustres savants (*Laplace et Poisson*) préconisèrent largement l'application de la théorie des probabilités aux *sciences morales*... L'engouement pour ces soit-disant applications, faites au mépris de la nature des faits, sans égard aux conditions qui les déterminent, pesa lourdement sur le progrès de la théorie des probabilités. Toutes ces applications fondamentalement erronées furent qualifiées dans la suite de *scandale mathématique*. »(Gnedenko)

La statistique moderne s'est nourrie des apports de plusieurs disciplines : la démographie, la théories des erreurs (notamment des observations astronomiques), la théorie cinétique des gaz (loi de Maxwell), la biométrie (mesures des êtres vivants), la psychométrie, l'actuariat (mathématiques de l'assurance).

Kolmogorov.

Benzecri.

1.4 Objectifs du cours

Un étudiant ayant suivi le cours devrait connaître les définitions de statistique descriptive uni- et bidimensionnelle, connaître le principe de la méthode des moindres carrés. Il est ainsi préparé à comprendre les extensions de ce principe à des situations plus complexes. Plusieurs indices sont présentés (espérance de vie, indices de Gini, IDH,...) : il faudra en connaître la définition et la motivation). Les lois classiques de variables aléatoires à valeurs réelles sont aussi à connaître. Le principe du raisonnement statistique inférentiel devra être compris et appliquer dans les cas simples de l'estimation d'une proportion, du test d'hypothèse (toujours pour une proportion). Si aucune démonstration du théorème limite central ne sera présentée, la signification (illustrée grâce à l'outil informatique) devra être comprise et le théorème lui même utilisé.

2 Séries statistiques simples pour caractères quantitatifs

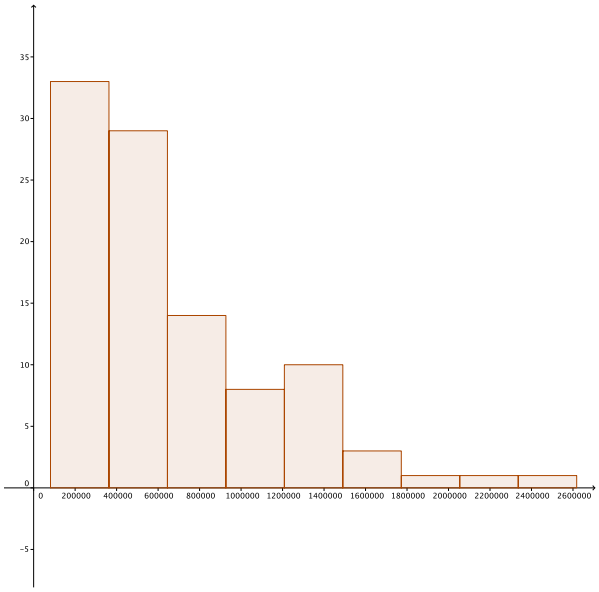
L'économie, la politique, la sociologie utilisent beaucoup de chiffres recueillis de différentes façons : recensement exhaustif, sondage, enquête, archives,... La présentation des données, leur traitement sont essentiels à leur exploitation, leur interprétation. Différents paramètres peuvent être utilisés, calculés pour ce faire.

2.1 Tableaux et représentations graphiques

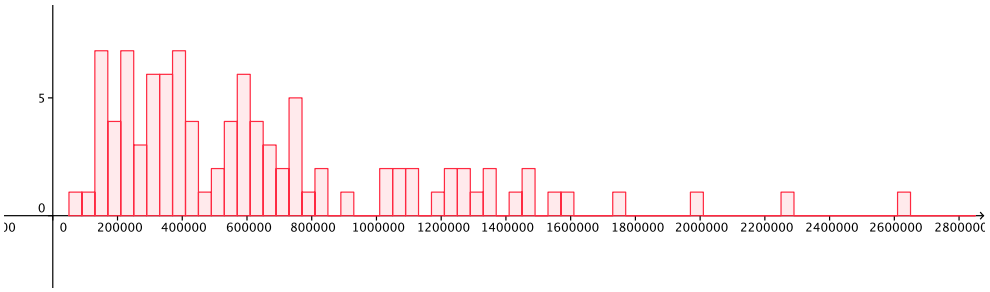
Les données recueillies peuvent être présentées dans des tableaux. Voici par exemple la liste des tailles de populations des départements français :

614331	555094	353124	165155	142312	1094579	324885	291678	157582
311720	365854	288364	2000550	699561	154135	364429	640803	319600
252235	145998	168869	538505	612383	127919	426607	542509	499313
603194	440291	929286	726285	1268370	195489	1479277	1062617	1015470
238261	605819	1233759	271973	397766	340729	766729	231877	1317685
674913	181232	342500	81281	808298	517121	579533	191004	317006
746502	200509	744663	1066667	226997	2617939	823668	301421	1489209
649643	674908	237945	457238	1115226	765634	1756069	247311	574874
579497	428751	760979	2268265	1275952	1347008	1435448	380569	583388
387099	248227	1026222	555240	654096	438566	384781	392846	353366
146475	1233645	1590749	1534895	1340868	1187836	409905	400535	231167
829903								

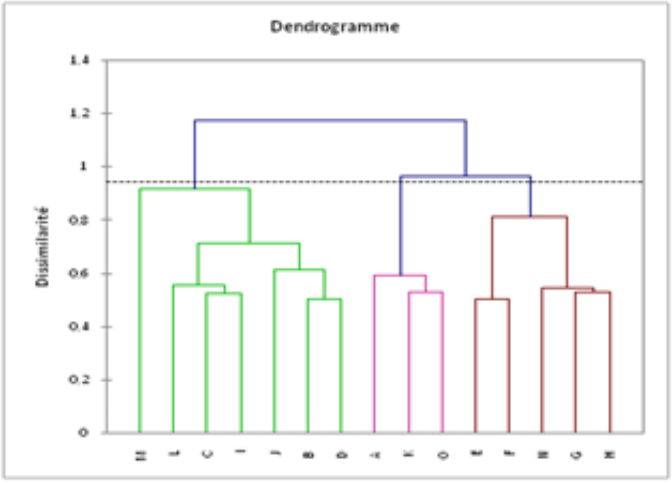
La liste contient toute l'information. Mais un traitement doit être fait pour que l'information soit accessible à qui s'y intéresse. Par exemple la distribution des tailles des populations des départements est assez bien décrite par l'histogramme suivant :



Pour construire un histogramme il faut choisir une taille de classe. De mauvais choix peuvent changer l'aspect obtenu de la distribution. Par exemple :



Les méthodes statistiques sont utilisées pour produire des représentations d'ensemble hiérarchisés. On peut ainsi obtenir des schémas appelés dendrogrammes.



Des tableaux peuvent aussi présenter des données synthétiques. La débauche d'image n'est pas synonyme de présentation claire. L'Insee par exemple propose de nombreux tableaux qui rassemblent certaines paramètres classiques de données. Par exemple :

Distribution du revenu salarial annuel par sexe ou catégorie socioprofessionnelle sur l'ensemble des salariés en 2010

en euros courants

Décile	Ensemble	Femmes	Hommes	Cadres*	Professions intermédiaires	Employés	Ouvriers
1er décile (D1)	2 360	1 970	2 840	10 840	6 310	1 450	1 910
1er quartile (Q1)	9 370	7 930	11 460	24 420	16 130	6 010	7 410
Médiane (D5)	**17 510**	**15 910**	**19 060**	**33 650**	**22 400**	**14 060**	**15 580**
3ème quartile (Q3)	24 590	22 270	26 820	46 350	27 870	18 640	20 000
9ème décile (D9)	34 600	30 070	39 110	66 600	33 660	22 980	24 190
D9/D1	**14,7**	**15,3**	**13,8**	**6,1**	**5,3**	**15,9**	**12,6**
Moyenne	**19 490**	**16 710**	**22 010**	**39 310**	**22 030**	**13 230**	**14 380**

* y compris chefs d'entreprise salariés
Champ : France métropolitaine, ensemble des salariés des secteurs public et privé hors salariés agricoles et apprentis-stagiaires.
Source : Insee, DADS 2010 définitif et SIASP, exploitation au 1/12.

Être capable de lire ces tableaux requiert une connaissance des paramètres qui y sont donnés. Nous allons maintenant en donner quelques uns.

2.2 Paramètres caractéristiques de position

(moyenne ; médiane, quartiles... ; mode)

On se donne $(x_i)_{i=1}^n$ une série statistique à une variable. C'est une suite numérotée de nombres. L'effectif de la série est le nombre de données qu'elle contient ; ici n. La **moyenne arithmétique** de la série, notée $\bar{x}$ est la quantité

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Donnons quelques propriétés de la moyenne.

La moyenne est **linéaire** : soient a et b deux nombres réels, $(x_i)_{i=1}^n$ et $(y_i)_{i=1}^n$ deux séries univariées. On a

$$\overline{ax + by} = a\bar{x} + b\bar{y}.$$

Démontrons cette égalité. Commençons par remarquer que $\overline{ax + by}$ désigne la moyenne de la série définies à partir de x et y par $(ax_i + by_i)_{i=1}^n$. On a, par définition,

$$\overline{ax + by} = \frac{1}{n}\sum_{i=1}^{n}(ax_i + by_i).$$

Dessinons un tableau contenant tous les nombres ax_i et $b_y i$.

i	1	2	$\cdots$	$n-1$	n	**Sommes des lignes**
ax	ax_1	ax_2	$\cdots$	ax_{n-1}	ax_n	$\sum_{i=1}^{n} ax_i$
by	by_1	by_2	$\cdots$	by_{n-1}	by_n	$\sum_{i=1}^{n} by_i$
ax+by	$ax_1 + by_1$	$ax_2 + by_2$	$\cdots$	$ax_{n-1+}by_{n-1}$	$ax_n + by_n$	$\sum_{i=1}^{n}(ax_i + by_i)$

Si on fait la somme des nombres contenus dans les deux premières en faisant d'abord la somme des lignes, on obtient la somme des deux expressions écrites à leur droite, soit

$$\sum_{i=1}^{n} ax_i + \sum_{i=1}^{n} by_i.$$

Si on fait la même somme en sommant d'abord par colonne puis la somme des résultats obtenus (écrits sur la troisième ligne), on obtient

$$\sum_{i=1}^{n}(ax_i + by_i).$$

On a donc

$$\sum_{i=1}^{n}(ax_i + by_i) = \sum_{i=1}^{n} ax_i + \sum_{i=1}^{n} by_i.$$

Maintenant, dans la somme

$$\sum_{i=1}^{n} ax_i = ax_1 + ax_2 + \ldots + ax_{n-1} + ax_n$$

on peut mettre a en facteur. Cela donne

$$\sum_{i=1}^{n} ax_i = a(x_1 + x_2 + \ldots + x_{n-1} + x_n) = a\sum_{i=1}^{n} x_i.$$

On a de la même façon

$$\sum_{i=1}^{n} by_i = b(y_1 + y_2 + \ldots + y_{n-1} + y_n) = b\sum_{i=1}^{n} y_i.$$

On obtient donc finalement

$$\sum_{i=1}^{n}(ax_i + by_i) = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i.$$

Divisons par n cette égalité :

$$\frac{1}{n}\sum_{i=1}^{n}(ax_i + by_i) = a\frac{1}{n}\sum_{i=1}^{n} x_i + b\frac{1}{n}\sum_{i=1}^{n} y_i.$$

Cela s'écrit

$$\overline{ax + by} = a\bar{x} + b\bar{y}.$$

Nous avons donc établi la linearité de la moyenne arithmétique.

Soit $(x_i)_{i=1}^n$ une série statistique et a une constante. On désigne par $x+a$ la série $(x_i+a)_{i=1}^n$. On montre facilement que

$$\overline{x + a} = \bar{x} + a.$$

En particulier, si on prend pour a la constante $\bar{x}$, on a

$$\overline{x - \bar{x}} = \bar{x} - \bar{x} = 0.$$

Je vous conseille de vous entraîner à faire par vous même de différentes façons des calculs comme ceux décrits ci-dessus.

La **médiane** est un nombre m_x qui "sépare en deux la série" : au moins la moitié des données est inférieure à m_x, au moins la moitié supérieure. Vous trouverez différentes définitions de la médiane ; en particulier lorsque les données sont regroupées en classes. Lorsqu'elles ne le sont pas on trouve "la" médiane en rangeant les données par ordre croissant : si n est impair, la médiane est la donnée du milieu ; si n est pair, tout nombre compris entre les deux données centrales convient. On peut alors décider d'un choix. Au collège, vous avez peut-être décidé de choisir la moyenne des deux données centrales. Nous allons faire un autre choix qui sera conforme à celui que nous allons adopter pour définir une notion analogue à la médiane : un quantile.

La courbe des fréquences cumulées définit une fonction appelée fonction de répartition empirique. Désignons par F la fonction ainsi associée à une série. On définit la fonction quantile Q par :

$$\begin{aligned} Q \quad : \quad & [0,1] \to \mathbb{R} \\ & q \mapsto \sup\{x \in \mathbb{R} \ / \ F(x) \leq q\}. \end{aligned}$$

La médiane est ainsi définie par $Q(0,5)$, le premier quartile par $Q(0,25)$, le 43ème centile par $Q(0,43)$, etc...

Plus de détails ont été donnés en cours (avec des exemples et des dessins). En cours j'ai aussi expliqué ce qu'on faisait suivant les différentes présentations des données (en particulier quand elles sont regroupées en classes).

Exemples.
Premier décile des revenus annuels en France (en 2011) : 4700 euros. Neuvième décile : 37180 euros. Médiane : 18280 euros.
Médiane du revenu **mensuel** des notaires : 13000 euros. Moyenne : 17000 euros (toujours pour les notaires, et toujours mensuels).

Il est important de garder en tête que les paramètres dont il a été question ci-dessus ne peuvent prétendre contenir toute l'information contenue dans les données. Seul l'ensemble

des données recèle toute l'information ; les paramètres ne sont que des outils commodes pour décrire certains aspects de l'ensemble des données. Leur utilisation doit s'accompagner d'un examen critique de leur pertinence.

2.3 Paramètres de dispersion

(étendue, variance, écart-type, coefficient de variation ; écart semi-interquartile...) ;

La variance d'une série $(x_i)_{i=1}^n$ est la « moyenne des écarts à la moyenne au carré ». Formellement cela s'écrit :

$$Var(x) = \overline{(x-\bar{x})^2}.$$

L'écart type de la série $(x_i)_{i=1}^n$ est la racine carrée de la variance :

$$\sigma_x = \sigma(x) = \sqrt{Var(x)} = \sqrt{\overline{(x-\bar{x})^2}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}.$$

Définition 2.1. *Soient $(x_i)_{i=1}^n$ et $(y_i)_{i=1}^n$ deux séries statistiques. On appelle covariance de x et y le nombre, noté $Cov(x,y)$, donné par*

$$Cov(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \overline{(x-\bar{x})(y-\bar{y})}.$$

Proposition 2.2. *Soient $(x_i)_{i=1}^n$ et $(y_i)_{i=1}^n$ deux séries statistiques, a et b deux nombres réels. On a :*

$$Var(x) = \overline{x^2} - \overline{x}^2,$$
$$Cov(x,y) = \overline{xy} - \overline{x}\ \overline{y},$$
$$Var(ax+b) = a^2 Var(x),$$
$$Var(x+y) = Var(x) + Var(y) + 2Cov(x,y).$$

Montrons quelques unes de ces égalités. Là encore, je vous encourage à le faire et refaire vous même.

Commençons par faire le calcul en utilisant les propriétés de linéarité de la moyenne. Le calcul est alors une suite de petite transformations formelles. Le passage d'une ligne à l'autre se fait en appliquant la propriété $\overline{ax+by} = a\overline{x} + b\overline{y}$ mais avec une somme de quatre termes plutôt que deux et avec a et b de la forme $-\overline{x}$ ou $-\overline{y}$:

$$\begin{aligned} Cov(x,y) &= \overline{(x-\bar{x})(y-\overline{y})} \\ &= \overline{xy - \bar{x}y - \overline{y}x + \overline{x}\ \overline{y}} \\ &= \overline{xy} + \overline{(-\bar{x}y)} + \overline{(-\overline{y}x)} + \overline{\overline{x}\ \overline{y}} \\ &= \overline{xy} - \overline{x}\ \overline{y} - \overline{y}\ \overline{x} + \overline{x}\ \overline{y} \\ &= \overline{xy} - \overline{x}\ \overline{y} \end{aligned}$$

L'égalité

$$Var(x) = \overline{x^2} - \overline{x}^2,$$

est celle que l'on vient de démontrer dans un cas particulier car $Var(x) = Cov(x, x)$.

Reprenons cette démonstration sous une forme moins compacte.

$$\begin{aligned}
Cov(x,y) &= \overline{(x-\bar{x})(y-\overline{y})} \\
&= \frac{1}{n}\sum_{i=1}^{n}((x_i-\bar{x})(y_i-\bar{y})) \\
&= \frac{1}{n}\sum_{i=1}^{n}(x_iy_i - \bar{x}y_i - \bar{y}x_i + \bar{x}\bar{y}) \\
&= \frac{1}{n}\sum_{i=1}^{n}x_iy_i + \frac{1}{n}\sum_{i=1}^{n}(-\bar{x}y_i) + \frac{1}{n}\sum_{i=1}^{n}(-\bar{y}x_i) + \frac{1}{n}\sum_{i=1}^{n}\bar{x}\bar{y} \\
&= \frac{1}{n}\sum_{i=1}^{n}x_iy_i - \bar{x}\frac{1}{n}\sum_{i=1}^{n}y_i - \bar{y}\frac{1}{n}\sum_{i=1}^{n}x_i + \bar{x}\bar{y}\frac{1}{n}\sum_{i=1}^{n}1 \\
&= \overline{xy} - \overline{x}\ \overline{y} - \overline{y}\ \overline{x} + \overline{x}\ \overline{y} \\
&= \overline{xy} - \overline{x}\ \overline{y}
\end{aligned}$$

Les deux autres égalités ont été démontrées en cours. C'est un bon exercice que de les démontrer seul.

2.4 Courbe de concentration ; médiale, exemples d'indices économiques

2.4.1 Indice de Gini

2.4.2 Indice de développement humain

Principe du calcul. Problème du recueil des données. Choix du critère : est-il facile à améliorer de manière artificielle ou non ? Exemple : taux de mortalité à cinq ans. Amartya Sen. Duflo ?

2.4.3 L'espérance de vie

Termes utilisés pour "espérance" dans différentes langues : italien "valore atteso", allemand "erwartungswert", anglais "expected value", espagnol "esperanza".

L'espérance de vie d'une population n'est pas calculée en faisant la moyenne des âges des personnes mourant dans l'année. On utilise une autre méthode de calcul apparemment

plus compliquée. Commençons par décrire cette méthode. Notons p_i la proportion des personnes ayant l'âge i au premier janvier qui sont encore vivant le 31 décembre. L'espérance de vie d'une population est égale à

$$\sum_{k=0}^{130}\prod_{i=0}^{k} p_i.$$

Expliquons un peu.

D'abord le 130. Les humains vivent très rarement plus de 120 ans. Imaginons qu'en France le doyen âgé de 118 ans meure dans l'année. Alors $p_{118} = 0$, et tous les produits suivant sont nuls. La somme précédente vaut alors

$$\sum_{k=0}^{117}\prod_{i=0}^{k} p_i.$$

Il suffit qu'il existe un âge tel que toutes les personnes de cet âge meurent dans l'année pour que les produits soient nuls à partir d'un certain rang. Cela ne se produit pas à coup sûr. On peut imaginer par exemple que pour tous les âges représentés une personne au moins survive. Il faudrait alors dire ce que vaut p_i quand aucune personne d'âge i n'existe au premier janvier (p_i est alors une proportion de la quantité nulle). Si on prend 0 (pourquoi pas ?) alors par exemple comme personne en France n'a atteint l'âge de 130 ans $p_{131} = 0$ et on a

$$\sum_{k=0}^{\infty}\prod_{i=0}^{k} p_i = \sum_{k=0}^{130}\prod_{i=0}^{k} p_i.$$

On pourrait trouver que cela pose un problème. Par exemple si le doyen des français a trois ans de moins que celui le plus vieux qui le suit (par exemple le doyen a 121 et le suivant 118) alors comme $p_{119} = 0$ on ne prend pas en compte le doyen. Remarquons qu'on ne le prend pas en compte non plus si tous ceux qui avaient 118 ans meurent mais pas lui. Si on considère que $p_i = 1$ quand l'âge i n'est pas représenté alors on considère que toutes les personnes d'âge i atteignent l'âge $i + 1$; or si une année tous les âges représentés ont des survivants alors on considère que si un certain âge est atteint alors on vit indéfiniment. La somme

$$\sum_{k=0}^{\infty}\prod_{i=0}^{k} p_i$$

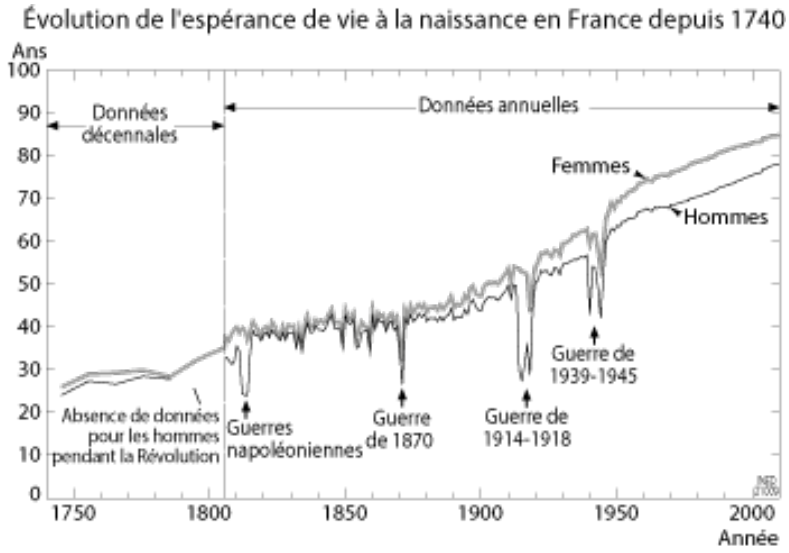
sera alors infini. On considérera donc que $p_i = 0$ si l'âge n'est pas représenté. Cela fait qu'on ne tient pas compte des personnes dont l'âge dépasse 115 ou 120 ans. Elles sont si peu nombreuses que l'influence sur le résultat sera très faible quelque soit la manière acceptable qu'on pourrait trouver de les prendre en compte.

Voici la définition que donne l'INSEE de l'espérance de vie : *L'espérance de vie à la naissance (ou à l'âge 0) représente la durée de vie moyenne - autrement dit l'âge moyen au décès - d'une génération fictive soumise aux conditions de mortalité de l'année. Elle caractérise la mortalité indépendamment de la structure par âge.*

On considère donc une population fictive (disons de 1000 personnes) et on lui applique à chaque âge les taux de mortalité de la population française constatée en 2012 par exemple. Au bout d'un certain temps il ne reste plus personne (les 1000 personnes sont mortes). On fait alors la moyenne des âges de décès constatés sur cette population fictive : on obtient l'espérance de vie à la naissance en France en 2012. Lien sur une animation créée par l'Ined : http://www.ined.fr/fr/tout_savoir_population/animations/esperance_vie/.

Pourquoi fait-on ça ? On obtient ainsi un résultat qui donne une moyenne si les taux de mortalité étaient les mêmes que ceux de l'année 2012. Cela permet de comparer les années : qu'une avancée médicale viennent modifier le taux de mortalité des nourrisson alors l'espérance de vie est modifiée, l'âge moyen de décès des morts de l'année beaucoup moins. D'autre part, comme affirmé dans la définition de l'Insee cela donne un résultat indépendant de la structure par âge. Imaginons par exemple qu'existe une classe démographique creuse dans une population. Alors quand cette classe atteindra un grand âge les décès de personnes très âgées seront sous représentés par rapport à celle de plus jeunes.

Quel est l'effet des guerres ? L'espérance chute brutalement pendant la période de guerre. Elle remonte presque aussi brusquement après la guerre.



Quel rapport entre ce qu'on trouve et la moyenne des âges de décès constatés dans l'année ? L'espérance de vie en 2012 donne la moyenne des âges de décès d'une population si tout se passe comme en 2012. Si tout se passait comme en 2012 pendant disons 250 ans et que chaque année le nombre d'enfants naissant était le même alors au bout d'un certain temps (disons 100 ans) les deux quantités coïncideraient. Remarquez que l'âge moyen des morts d'une année se comporte de manière similaire à l'espérance de vie pour les périodes de guerre. Pour tenir compte des guerres dans l'espérance de vie il faut faire une moyenne sur plusieurs années. On peut aussi calculer l'âge moyen de décès des personnes nées au cours d'une année donnée : par exemple personnes nées en 1895, personnes nées en 1943.

Pourquoi l'espérance de vie est-elle donnée par la formule

$$\sum_{k=0}^{\infty}\prod_{i=0}^{k} p_i ?$$

Notons X la variable aléatoire "durée de vie". On a

$$\mathbb{P}(X \geq 1) = p_0.$$

De même

$$\mathbb{P}(X \geq i+1 | X \geq i) = p_i.$$

Ceci donne

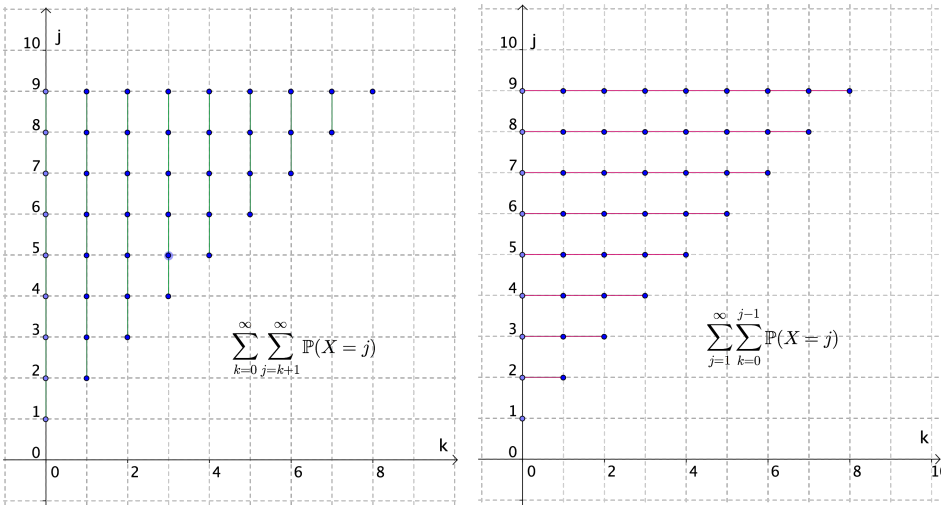
$$\mathbb{P}(X \geq k) = \mathbb{P}(X \geq k | X \geq k-1)\mathbb{P}(X \geq k-1) = p_{k-1}\mathbb{P}(X \geq k-1 | X \geq k-2)\mathbb{P}(X \geq k-2).$$

On obtient ainsi

$$\mathbb{P}(X \geq k) = \prod_{i=0}^{k-1} p_i.$$

Écrivons maintenant la somme

$$\begin{aligned}
\sum_{k=0}^{\infty}\prod_{i=0}^{k} p_i &= \sum_{k=0}^{\infty}\mathbb{P}(X \geq k+1) \\
&= \sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty}\mathbb{P}(X = j) \\
&= \sum_{j=1}^{\infty}\sum_{k=0}^{j-1}\mathbb{P}(X = j) \\
&= \sum_{j=1}^{\infty} j\mathbb{P}(X = j) \\
&= \mathbb{E}(X)
\end{aligned}$$



Prise en compte d'inégalité entre les sexes. Utilisation de l'inégalité isopérimétrique.

3 Séries statistiques doubles pour caractères quantitatifs

: tableaux à double entrée. Distributions marginales et conditionnelles.

3.1 Covariance et coefficient de corrélation linéaire, droites de régression affines

Soient $(x_i, y_i)_{i=1}^n$ une série statistique double. On cherche à approcher le nuage de points défini par cette série par une droite. Comment le faire ? Évidemment il est toujours possible de tracer une droite au jugé. On cherche une méthode systématique. Une méthode qui puisse être programmée. Plusieurs possibilités existent. Une manière raisonnable de procéder : on définit une quantité reflétant la "distance" du nuage de point à une droite quelconque, puis on cherche la meilleure droite au sens de cette distance, celle qui est à "distance" minimale.

Une droite est définie par son équation : $y = ax + b$.

Quelques quantités qui peuvent mesurer l'éloignement du nuage de points à cette droite :

$$F_1(a,b) = \sum_{i=1}^{n} |y_i - ax_i - b|$$

$$F_2(a,b) = \sum_{i=1}^{n} (y_i - ax_i - b)^2$$

$$F_3(a,b) = \sum_{i=1}^{n} (y_i - ax_i - b)^4$$

On pourrait aussi minimiser la somme des distances des points à la droite. Quelle quantité cela donne-t-il ? La distance du point (x_i, y_i) à la droite d'équation $y = ax + b$ est donnée par

$$\frac{|y_i - ax_i - b|}{\sqrt{1+a^2}}.$$

La quantité à minimiser dans ce cas est donc

$$F_4(a,b) = \sum_{i=1}^{n} \frac{|y_i - ax_i - b|}{\sqrt{1+a^2}}.$$

La quantité la plus manipulable est la deuxième. C'est avec celle-ci que nous allons travailler. Retenons que ce n'est pas le seul choix possible. Remarquons aussi la chose suivante : quelle que soit la quantité que vous cherchez à minimiser la droite que vous obtenez est sans intérêt si le nuage de points est très mal approché par une droite, correspond à la droite contenant le nuage de points si les points du nuage sont alignés.

Chercher la droite la plus proche du nuage de points au sens de F_2 c'est chercher a et b tels que F_2 soit minimale. Il s'agit donc de minimiser une fonction de deux variables. On peut attaquer ce type de problème comme pour les fonctions d'une variables en dérivant. Ici on peut dériver par rapport à a ou à b. Comme dans le cas des fonctions d'une variables, trouver les points où la dérivée s'annule n'est pas suffisant pour affirmer que la fonction est minimale en un point donné (par exemple la dérivée est nulle aussi en un point où la fonction est maximale). Nous emploierons donc une autre méthode (plus élémentaire) pour obtenir une réponse complète.

$$\frac{\partial F_2}{\partial a} = \sum_{i=1}^{n} -2x_i(y_i - ax_i - b)$$

$$\frac{\partial F_2}{\partial b} = \sum_{i=1}^{n} -2(y_i - ax_i - b)$$

Pour que ces deux dérivées soient nulles, il faut donc qu'on ait

$$\sum_{i=1}^{n} -2x_i(y_i - ax_i - b) = 0$$

$$\sum_{i=1}^{n} -2(y_i - ax_i - b) = 0$$

soit

$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} y_i = a\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} b$$

En divisant par n on obtient les deux conditions suivantes :

$$\bar{xy} = a\bar{x^2} + b\bar{x} \text{ et } \bar{y} = a\bar{x} + b$$

En multipliant la deuxième par $\bar{x}$ et en la soustrayant à la première on obtient

$$\bar{xy} - \bar{x}\bar{y} = a(\bar{x^2} - \bar{x}^2) \text{ et } \bar{y} = a\bar{x} + b$$

soit

$$Cov(x, y) = aVar(x) = a\sigma_x^2 \text{ et } \bar{y} = a\bar{x} + b$$

Si la série x n'est pas constante alors le système a une solution unique

$$a = \frac{Cov(x, y)}{\sigma_x^2},\ b = \bar{y} - \frac{Cov(x, y)}{\sigma_x^2}\bar{x}.$$

Remarquons que si $\sigma_x^2 = 0$ alors x est constante : le nuage de point est sur une droite verticale ; cette droite a pour équation $x = cste$ (pas une équation de la forme $y = ax+b$).

Dans le cas où $\sigma_x^2 \neq 0$ les valeurs trouvées de a et b sont effectivement celles qui minimisent F_2. Nous allons maintenant le montrer.

$$\begin{aligned}
F_2(a,b) &= \sum_{i=1}^{n}(y_i - ax_i - b)^2 \\
&= \sum_{i=1}^{n}(y_i - ax_i - b - \bar{y} + a\bar{x} + b + \bar{y} - a\bar{x} - b)^2 \\
&= \sum_{i=1}^{n}((y_i - \bar{y}) - a(x_i - \bar{x}) + \bar{y} - a\bar{x} - b)^2 \\
&= \sum_{i=1}^{n}((y_i - \bar{y}) - a(x_i - \bar{x}))^2 + \sum_{i=1}^{n}(\bar{y} - a\bar{x} - b)^2 \\
&\qquad +2(\bar{y} - a\bar{x} - b)\sum_{i=1}^{n}((y_i - \bar{y}) - a(x_i - \bar{x}))
\end{aligned}$$

La dernière égalité étant obtenue en développant le carré à l'intérieur du signe $\sum$. Comme

$$\sum_{i=1}^{n}(y_i - \bar{y}) = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\bar{y} = n\bar{y} - n\bar{y} = 0$$

(et une égalité analogue pour x) le dernier des trois termes précédents est nul. Transformons maintenant le premier de ces trois termes :

$$\begin{aligned}
\sum_{i=1}^{n}((y_i - \bar{y}) - a(x_i - \bar{x}))^2 &= \sum_{i=1}^{n}(y_i - \bar{y})^2 + a^2\sum_{i=1}^{n}(x_i - \bar{x})^2 - 2a\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) \\
&= n\sigma_y^2 + na^2\sigma_x^2 - 2naCov(x,y) \\
&= n\left(\sigma_y^2 + \left(a\sigma_x - \frac{Cov(x,y)}{\sigma_x}\right)^2 - \frac{Cov(x,y)^2}{\sigma_x^2}\right)
\end{aligned}$$

On a donc obtenu l'égalité

$$F_2(a,b) = n\sigma_y^2 - n\frac{Cov(x,y)^2}{\sigma_x^2} + n\left(a\sigma_x - \frac{Cov(x,y)}{\sigma_x}\right)^2 + n(\bar{y} - a\bar{x} - b)^2.$$

Le deuxième terme contient une partie qui ne dépend ni de a ni de b : $n\sigma_y^2 - n\frac{Cov(x,y)^2}{\sigma_x^2}$ à laquelle on ajoute une somme de carrés. Les carrés étant positifs on peut dire que

$$F_2(a,b) \geq n\sigma_y^2 - n\frac{Cov(x,y)^2}{\sigma_x^2}.$$

De plus il y a un seul choix de a et b qui annule les carrés que l'on ajoute à cette quantité pour obtenir $F_2(a,b)$. Ce sont les nombres a et b vérifiant :

$$a\sigma_x - \frac{Cov(x,y)}{\sigma_x} = 0 \text{ et } \bar{y} - a\bar{x} - b = 0.$$

En conclusion les valeurs de a et b qui rendent $F_2(a,b)$ minimale sont donc données par

$$a = \frac{Cov(x,y)}{\sigma_x^2} = 0 \text{ et } \bar{y} - a\bar{x} - b = 0.$$

La valeur minimale de F_2 est $n\left(\sigma_y^2 - \frac{Cov(x,y)^2}{\sigma_x^2}\right)$. À quelle condition cette valeur est-elle nulle ?

C'est l'occasion de parler de l'inégalité de Cauchy-Schwarz.

Théorème 3.1. *(Inégalité de Cauchy-Schwarz) Soient $(x_1,\dots,x_n)$ et $(y_1,\dots,y_n)$ deux suites de n nombres réels. On a*

$$\left|\sum_{i=1}^{n} x_i y_i\right| \leq \sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2},$$

avec égalité seulement s'il existe un nombre λ tel que, pour tout i on ait, $y_i = \lambda x_i$.

Démonstration Soient x et y deux vecteurs et λ un nombre réel. Le nombre $\langle \lambda x+y, \lambda x+y\rangle$ est positif ou nul. On peut développer ce produit scalaire en appliquant les propriétés de linéarité. On obtient :

$$\begin{aligned}
\langle \lambda x + y, \lambda x + y\rangle &= \langle \lambda x, \lambda x + y\rangle + \langle y, \lambda x + y\rangle \\
&= \lambda\langle x, \lambda x + y\rangle + \langle y, \lambda x + y\rangle \\
&= \lambda\langle x, \lambda x\rangle + \lambda\langle x, y\rangle + \langle y, \lambda x\rangle + \langle y, y\rangle \\
&= \lambda^2\langle x, x\rangle + \lambda\langle x, y\rangle + \lambda\langle y, x\rangle + \langle y, y\rangle \\
&= \lambda^2\langle x, x\rangle + 2\lambda\langle x, y\rangle + \langle y, y\rangle
\end{aligned}$$

On obtient donc un polynôme de degré 2 en λ qui est positif ou nul pour tout λ. Cela signifie que le discriminant de ce polynôme est négatif ou nul. Autrement dit on a

$$\langle x, y\rangle^2 - \langle y, y\rangle\langle x, x\rangle \leq 0.$$

Dire que ce discriminant est nul est équivalent à dire qu'il existe λ pour lequel le polynôme est nul ou encore qu'il existe λ pour lequel $\|\lambda x+y\|^2 = 0$ ce qui est équivalent à $\lambda x+y = 0$. Autrement dit l'égalité ne peut être vérifiée que dans les cas où x et y sont colinéaires. On vérifie facilement que l'égalité est satisfaite lorsque x et y sont colinéaires (on a donc l'équivalence). □

Corollaire 3.2. *Soient $x = (x_1,\dots,x_n)$ et $y = (y_1,\dots,y_n)$ deux suites de n nombres réels. On a*

$$|Cov(x,y)| \leq \sigma_x\sigma_y,$$

avec égalité si et seulement si les points de coordonnées (x_i, y_i) sont alignés.

Démonstration Pour obtenir l'inégalité il suffit d'appliquer l'inégalité de Cauchy-Swcharz aux deux séries $x - \bar{x}$ et $y - \bar{y}$. On obtient

$$\left|\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right| \leq \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2},$$

qui est exactement ce qu'on veut montrer. Le théorème affirme que l'inégalité n'est une égalité que s'il existe λ tel que, pour tout i on ait, $y_i - \bar{y} = \lambda(x_i - \bar{x})$ soit $y_i = \lambda x_i + \bar{y} - \lambda\bar{x}$. Cela signifie que les points (x_i, y_i) sont tous sur la droite d'équation $y = \lambda x + \bar{y} - \lambda\bar{x}$. □

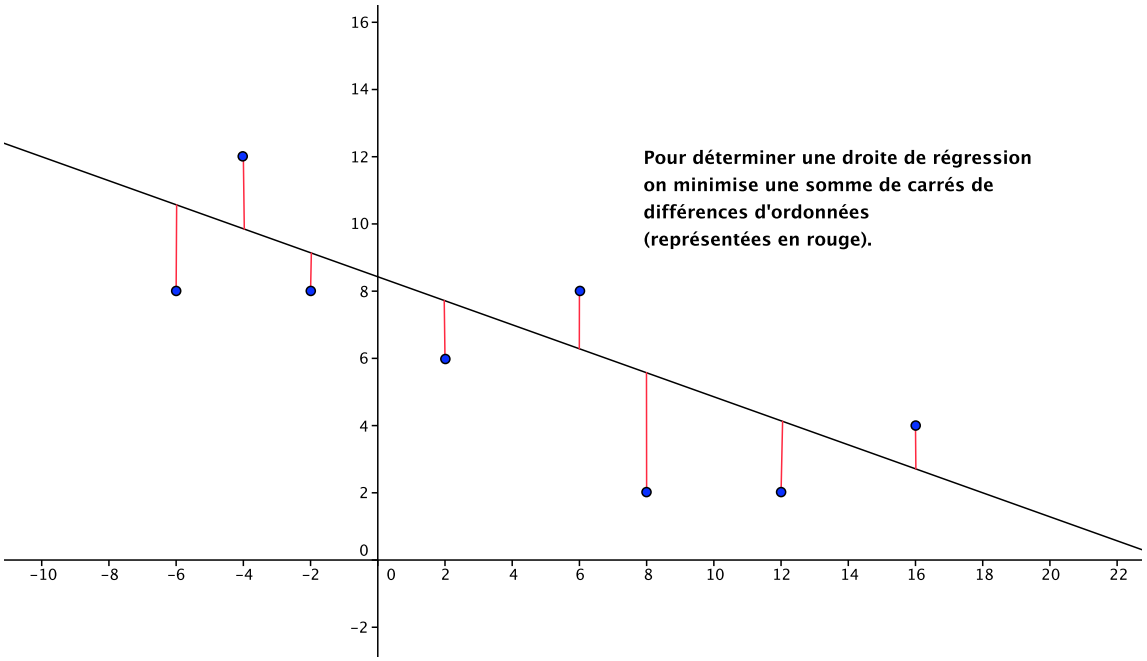
Pour Quelques exemples. Différentes formes de nuages de points.

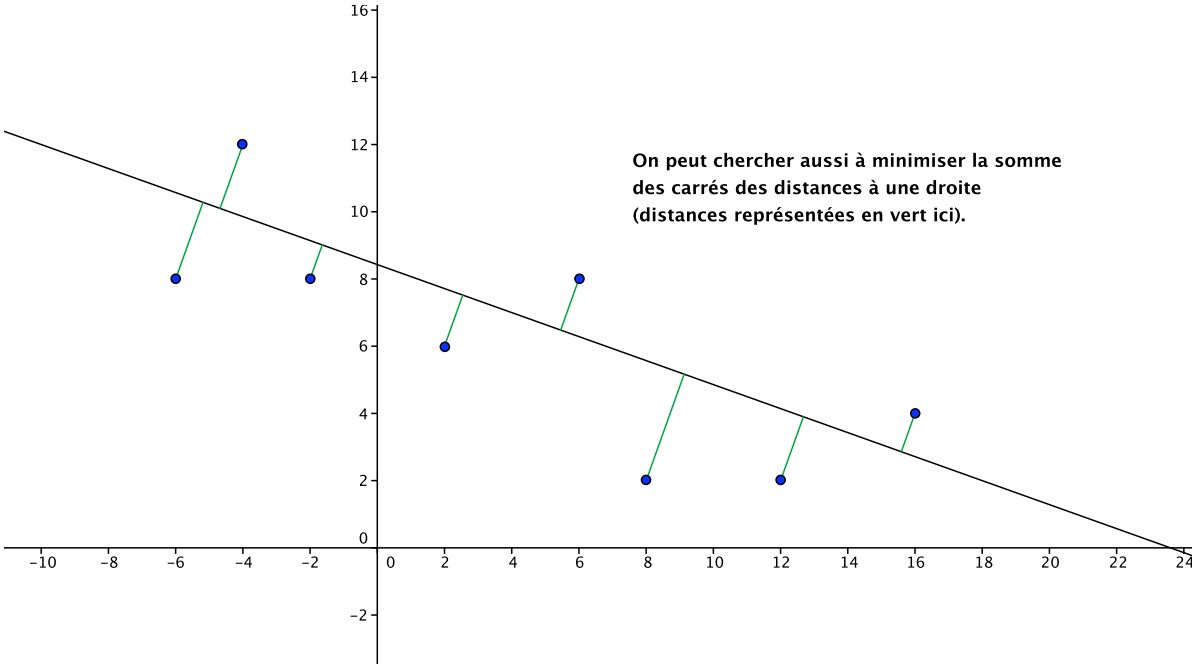
3.2 Un premier pas vers la régression multiple

3.3 Un premier pas vers l'analyse en composantes principales

En grande dimension (quand on associe de nombreux caractères à chaque individu de la population) la représentation du nuage de points est impossible. On cherche alors à en obtenir une image en dimension 2 la moins déformée possible. Cette baisse de dimension est ce que nous avons fait avec la droite de régression : trouver la droite qui représente aussi bien que possible une nuage de points dans le plan (la définition de « aussi bien que possible »peut varier ; nous avons choisi une méthode des moindres carrés). D'autres critères peuvent être utilisés (nous avons par exemple vu qu'existaient deux droites de réfression). Dans tous les cas il convient d'accompagner le calcul de ces approximations d'un indicateur de la qualité de l'approximation obtenue (le coefficient de corrélation dans le cas de la droite de régression). Nous allons présenter une autre façon de procéder qui est très employée en dimension supérieure sous le nom d'analyse en composantes principales. En pratique cette méthode permet d'obtenir la meilleure image plane d'un nuage de points placé dans un espace de grande dimension. Elle n'est donc pas appliquée en dimension 2. Ce qui suit est une présentation de la méthode sur un cas simple.

Nous cherchons à minimiser, non plus des sommes de différences d'ordonnées ou d'abscisse, mais la somme des carrés des distances à une droite (et cherchons une droite qui minimise cette somme).





Nous allons le faire en nous limitant aux droites qui passent par le point moyen du nuage de points. La somme des carrés des distances au point point moyen est

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 + (y_i - \bar{y})^2.$$

Grâce au théorème de Pythagore on peut écrire cette somme comme la somme des carrés des distances des points à leurs projetés orthogonaux sur la droite plus la somme des

carrés des distances de ces projetés au point moyen. Minimiser la somme des carrés des distances des points à la droite revient donc à maximiser la somme des carrés des distances des projetés au point moyen. Soit $(\cos(t), \sin(t))$ un vecteur directeur unitaire de la droite. La somme à maximiser (on cherche t pour que cette somme soit maximale) est :

$$\sum_{i=1}^{n} (\cos(t)(x_i - \bar{x}) + \sin(t)(y_i - \bar{y}))^2.$$

Réécrivons cette somme :

$$\begin{aligned}
&\sum_{i=1}^{n} (\cos(t)(x_i - \bar{x}) + \sin(t)(y_i - \bar{y}))^2 \\
&= \sum_{i=1}^{n} \cos^2(t)(x_i - \bar{x})^2 + \sum_{i=1}^{n} \sin^2(t)(y_i - \bar{y})^2 + 2\sum_{i=1}^{n} \sin(t)\cos(t)(x_i - \bar{x})(y_i - \bar{y}) \\
&= n\left(\cos^2(t)\sigma_x^2 + \sin^2(t)\sigma_y^2 + 2\sin(t)\cos(t)cov(x,y)\right).
\end{aligned}$$

Il suffit maintenant d'étudier les variations de cette fonction de t pour déterminer ses extrema (oublions le facteur n qui n'a aucune importance pour l'étude des variations). Sa dérivée vaut

$$-2\sin(t)\cos(t)\sigma_x^2 + 2\sin(t)\cos(t)\sigma_y^2 + 2\cos(2t)cov(x,y) = \sin(2t)(\sigma_y^2 - \sigma_x^2) + 2\cos(2t)cov(x,y).$$

Cette dérivées est nulle si

$$\tan(2t) = -\frac{2cov(x,y)}{\sigma_y^2 - \sigma_x^2},$$

à condition que $\sigma_y^2 - \sigma_x^2$ ne soit pas nul. Cette égalité définit deux nombres t entre $-\pi/2$ et $\pi/2$ où la fonction qui nous intéresse est minimale et maximale (différentes possibilités suivant les signes). Si la covariance $cov(x,y)$ et la différence $\sigma_y^2 - \sigma_x^2$ sont nulles alors la fonction est constante. Si la différence $\sigma_y^2 - \sigma_x^2$ est nulle mais la covariance ne l'est pas alors la valeur maximale est atteinte pour t égal à $\pi/4$ si la covariance est positive, $-\pi/4$ si la covariance est négative.

Pou neutraliser les effets dûs au choix des unités il est fréquent de réduire et normaliser les données x et y. On se ramène alors au nuage de points défini par

$$\left(\frac{x_i - \bar{x}}{\sigma_x}, \frac{y_i - \bar{y}}{\sigma_y}\right)$$

pour lequel les variances apparaissant dans le calcul précédent sont toutes deux égales à 1 et la covariance vaut $\rho(x,y)$ le coefficient de corrélation de x et y. Ce que nous avons vu est que (si $\rho(x,y) < 0$ par exemple) la droite minimisant la somme des carrés distances à la droite est la deuxième bissectrice. Autrement dit on approche le nuage de points par la droite d'équation

$$\frac{y - \bar{y}}{\sigma_y} = -\frac{x - \bar{x}}{\sigma_x}.$$

Revenant au nuage de points de départ cela donne la droite

$$y = -\frac{\sigma_y}{\sigma_x}x + \frac{\sigma_y}{\sigma_x}\bar{x} - \bar{y}.$$

Nous obtenons ainsi une quatrième droite approchant le nuage de point (après la droite de régression de y sur x, la droite de régression de x sur y, la droite minimisant la somme des carrés des distances). L'inégalité de Cauchy-Schwarz donne ici ($\rho(x, y) < 0$)

$$0 > \rho(x, y) > -\sigma_x\sigma_y.$$

On en déduit l'encadrement

$$\frac{\sigma_y^2}{\rho(x, y)} < -\frac{\sigma_y}{\sigma_x} < \frac{\rho(x, y)}{\sigma_x^2}.$$

Autrement dit la quatrième droite dont nous avons parlé est entre les deux droites de régression.

4 Théorèmes limite en probabilités

4.1 Variables aléatoires (loi, indépendance)

TABLE 1 – **Lois discrètes classiques**

Dénomination	**Loi**	**Espérance**	**Variance**
Loi Uniforme $X \sim \mathcal{U}(\{1,2,\ldots,n\})$	$X(\Omega) = \{1,2,\ldots,n\}$ $\mathbb{P}(X=k) = \frac{1}{n}$	$\mathbb{E}(X) = \frac{n+1}{2}$	$\mathbb{V}(X) = \frac{n^2-1}{12}$
Loi de Bernoulli $X \sim \mathcal{B}(1,p)$	$X(\Omega) = \{0,1\}$ $\mathbb{P}(X=0) = 1-p$ $\mathbb{P}(X=1) = p$	$\mathbb{E}(X) = p$	$\mathbb{V}(X) = p(1-p)$
Loi Binomiale $X \sim \mathcal{B}(n,p)$	$X(\Omega) = \{0,1,2,\ldots,n\}$ $\mathbb{P}(X=k) = C_n^k p^k (1-p)^{n-k}$	$\mathbb{E}(X) = np$	$\mathbb{V}(X) = np(1-p)$
Loi Géométrique $X \sim \mathcal{G}(p)$	$X(\Omega) = \mathbb{N}^*$ $\mathbb{P}(X=k) = p(1-p)^{k-1}$	$\mathbb{E}(X) = \frac{1}{p}$	$\mathbb{V}(X) = \frac{1-p}{p^2}$
Loi de Poisson $X \sim \mathcal{P}(\lambda)$	$X(\Omega) = \mathbb{N}$ $\mathbb{P}(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$	$\mathbb{E}(X) = \lambda$	$\mathbb{V}(X) = \lambda$

TABLE 2 – **Lois continues classiques**

Dénomination	**Densité**	**Espérance**	**Variance**
Loi Uniforme $X \sim \mathcal{U}([a,b])$	$f(x) = \frac{1}{b-a}, \quad x \in [a,b]$ $f(x) = 0, \quad x \notin [a,b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Loi Exponentielle $X \sim \mathcal{E}(\lambda)$	$f(x) = \lambda e^{-\lambda x}, \quad x \geqslant 0$ $f(x) = 0, \quad x < 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Loi Normale $X \sim \mathcal{N}(\mu, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ \forall x \in \mathbb{R}$	μ	σ^2

4.2 La loi faible des grands nombres

La loi faible des grands nombres est une conséquence de l'inégalité de Bienaymé-Tchébychev.

Théorème (inégalité de Bienaymé-Tchébychev) : Soit X une variable aléatoire d'espérance $\mathbb{E}(X)$ et d'écart type $\sigma(X)$. Pour tout $\epsilon > 0$,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon) \leq \frac{\sigma^2(X)}{\epsilon^2}.$$

Démonstration :

$$Var(X) = \sum_{i=1}^{n} p_i (x_i - \mathbb{E}(X))^2.$$

On note I l'ensemble des indices i tels que $|x_i - \mathbb{E}(X)| \geq \epsilon$ et J l'ensemble des indices i tels que $|x_i - \mathbb{E}(X)| < \epsilon$. On peut séparer la somme en deux en utilisant les deux ensembles I et J :

$$Var(X) = \sum_{i \in I} p_i (x_i - \mathbb{E}(X))^2 + \sum_{i \in J} p_i (x_i - \mathbb{E}(X))^2.$$

On en déduit les minorations suivantes

$$Var(X) \geq \sum_{i \in I} p_i (x_i - \mathbb{E}(X))^2 \geq \sum_{i \in I} p_i \epsilon^2 = \epsilon^2 \sum_{i \in I} p_i$$

Or $\sum_{i \in I} p_i = \mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon)$; on a ainsi

$$Var(X) \geq \epsilon^2 \mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon).$$

d'où le résultat. □

Définition : Soit (X_n) une suite de variables aléatoires réelles et soit X une variable aléatoire réelle. On dit que (X_n) converge en probabilité vers X si et seulement si

$$\lim_n \mathbb{P}(|\overline{X_n} - X| \geq \epsilon) = 0.$$

Théorème (loi faible des grands nombres) Soit (X_n) une suite de variables aléatoires réelles de même espérance m et de même écart type σ. On suppose de plus que les sont deux à deux indépendantes. On pose

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

La suite $(\overline{X_n})$ converge en probabilité vers la constante m et on a pour tout, $\epsilon > 0$:

$$\mathbb{P}(|\overline{X_n} - m| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}.$$

Ce qui a pour conséquenceă : pour tout $\epsilon > 0$,

$$\mathbb{P}(|\overline{X_n} - m| \geq \epsilon) \to 0$$

ou encore

$$\mathbb{P}(|\overline{X_n} - m| < \epsilon) \to 1.$$

Démonstration : L'espérance de $\overline{X_n}$ est

$$\mathbb{E}(\overline{X_n}) = \mathbb{E}(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(X_i) = \frac{1}{n} nm = m.$$

D'après l'inégalité de Bienaymé-Tchébychev, on a

$$\mathbb{P}(|\overline{X_n} - m| \geq \epsilon) = \mathbb{P}(|\overline{X_n} - \mathbb{E}(\overline{X_n})| \geq \epsilon) \leq \frac{\sigma^2(\overline{X_n})}{\epsilon^2}.$$

Calculons $\sigma^2(\overline{X_n})$: en utilisant les propriétés de la variance et l'indépendance des X_i, il vient

$$Var(\overline{X_n}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i),$$

et comme $\sum_{i=1}^{n} Var(X_i) = nVar(X_1) = n\sigma^2$, on a

$$Var(\overline{X_n}) = \frac{\sigma^2}{n},$$

d'où le résultat.

Exemple : On lance n fois un dé équilibré. Pour le i-ème lancer on considère la variable aléatoire X_i qui vaut 1 si le résultat est un multiple de 3 et 0 sinon. On pose $X =$

$\frac{1}{n}\sum_{i=1}^{n} X_i$. Déterminer n pour que la probabilité de l'événement « X s'éloigne de son espérance de plus de 0,02 »soit inférieure à 0,1.

Il faut rechercher n tel que $\mathbb{P}(|X - \frac{1}{3}| > 0,02) < 0,1$. Comme $\mathbb{P}(|X - \frac{1}{3}| \leq \frac{\sigma^2(X)}{n0,02^2} \leq 0,1$. Comme $Var(X_i) = \frac{2}{9}$, la majoration voulue sera satisfaite lorsque $\frac{2}{n.9.0,02^2}$ sera inférieur à 0,1, soit $n > \frac{1}{0,0018.0,1}$ ou $n \geq 5556$. Il est bon de remarquer que cette majoration est de très mauvaise qualité car le calcul donne pour $n = 5556$, $\mathbb{P}(|X - \frac{1}{3}| > 0,02) \simeq 0,0015.???$ Il faut en fait beaucoup moins de lancers pour obtenir la majoration souhaitée.

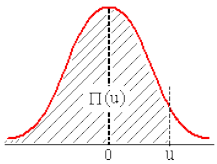
4.3 La loi normale

Table de Loi Normale

Fonction de répartition Π de la loi normale centrée réduite.

Probabilité de trouver une valeur inférieure à u .

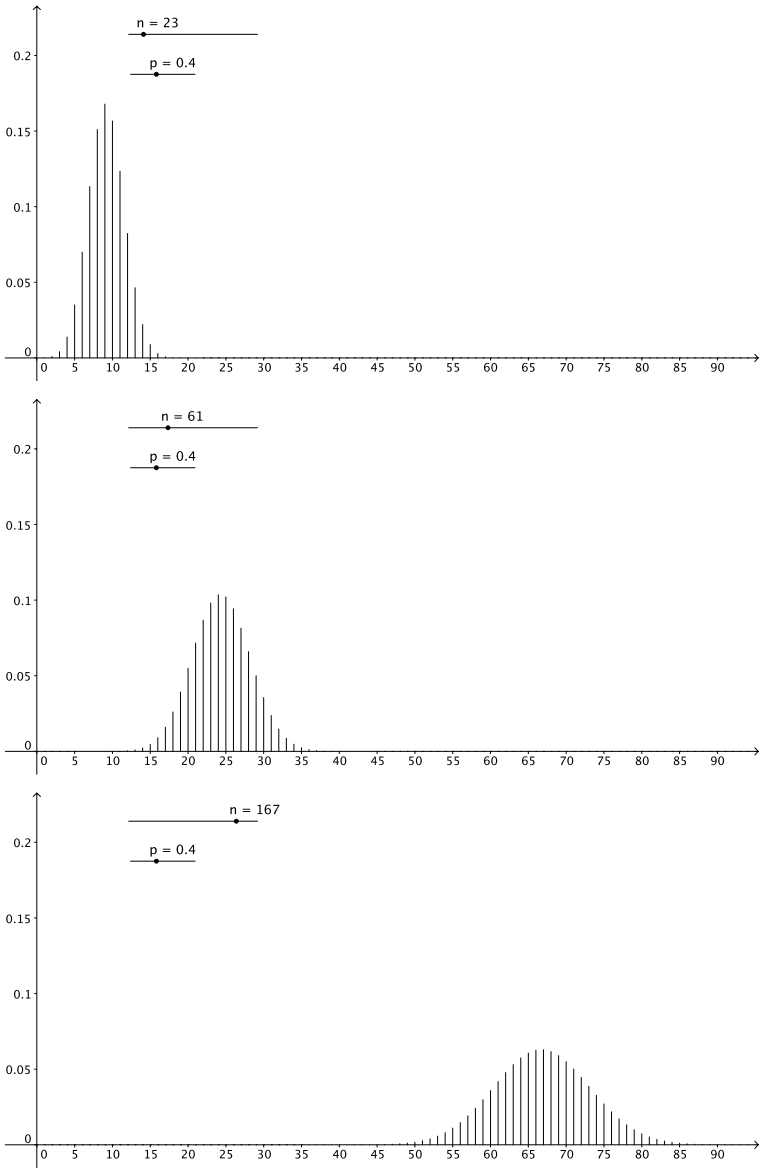
$\Pi(-u) = 1 - \Pi(u)$



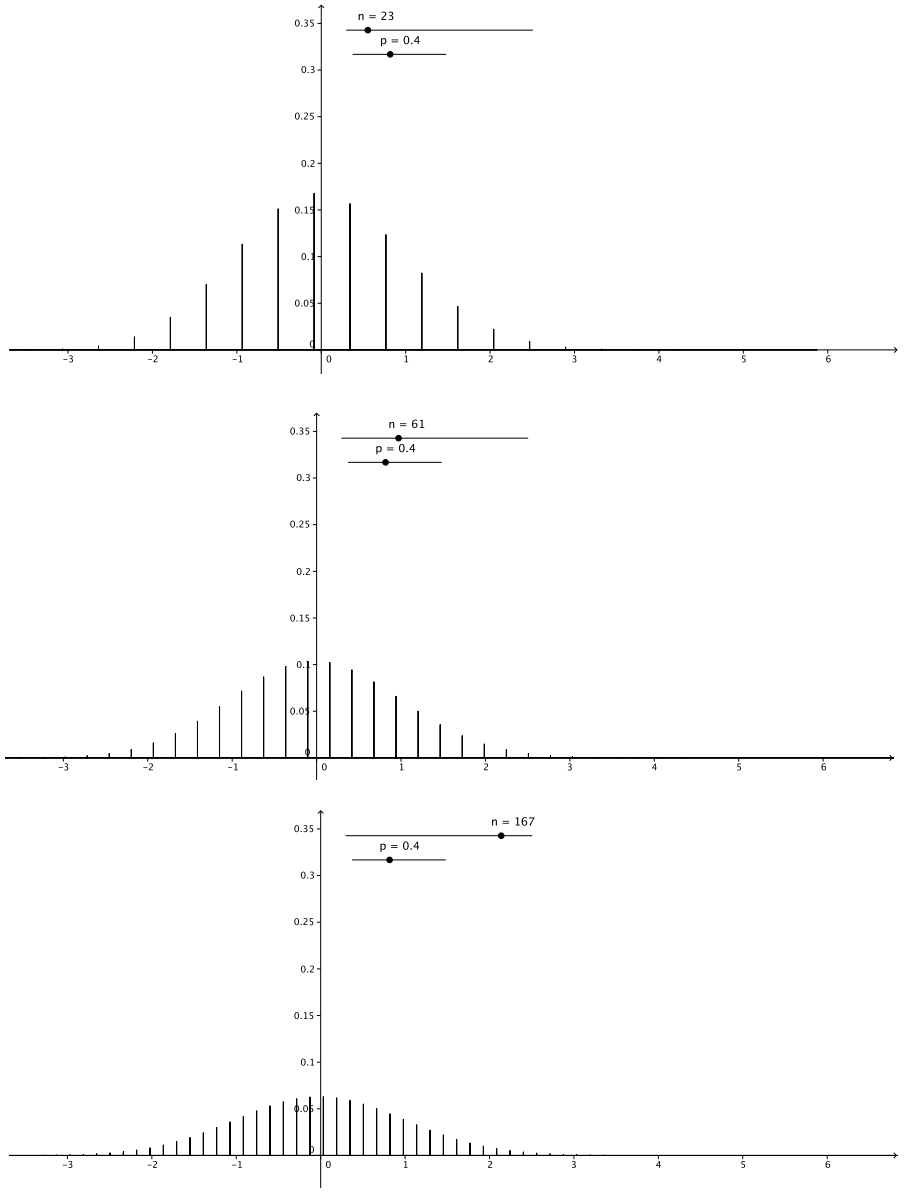
u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

4.4 Le théorème limite central

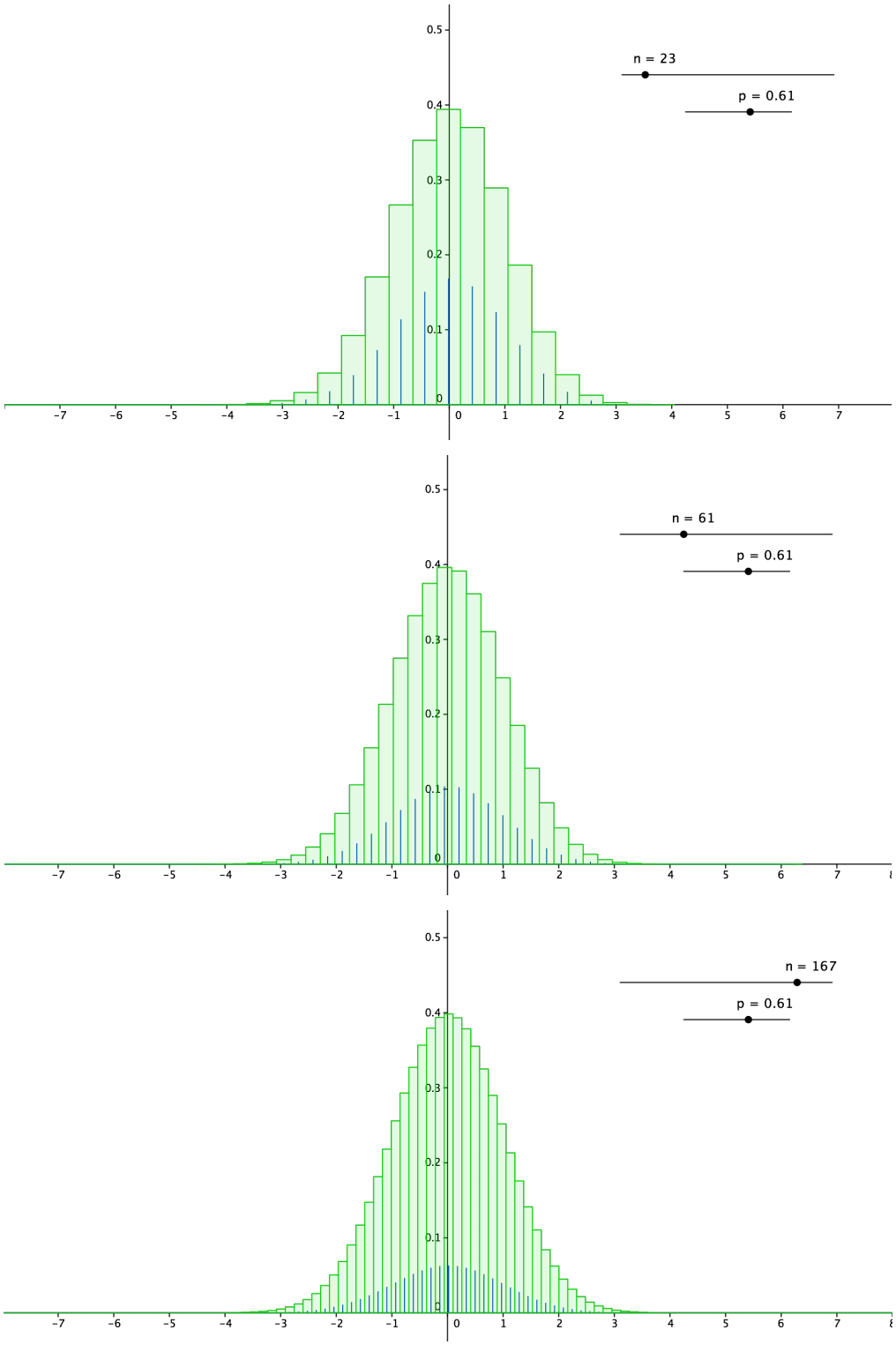
Observons le comportement de la loi binomiale $\mathcal{B}(n,p)$ lorsque n grandit. Sur les figures suivantes nous avons représenté les diagrammes en bâtons décrivant pour $p = 0,4$ et n successivement égal à 23, 61 et 167 les probabilités de prendre les valeurs entières d'une variable de loi $\mathcal{B}(n,p)$.



On observe que la figure obtenue se déplace vers la droite, a la forme d'une cloche qui s'écrase quand n grandit. Si on décale vers la gauche le graphique (en enlevant np) on obtient des dessins centrés. De plus on resserre les bâtons de sorte que la distance entre deux bâtons soit $\frac{1}{\sqrt{np(1-p)}}$. Autrement dit on dessine le diagramme en bâtons de la loi de $\frac{X-np}{\sqrt{np(1-p)}}$.



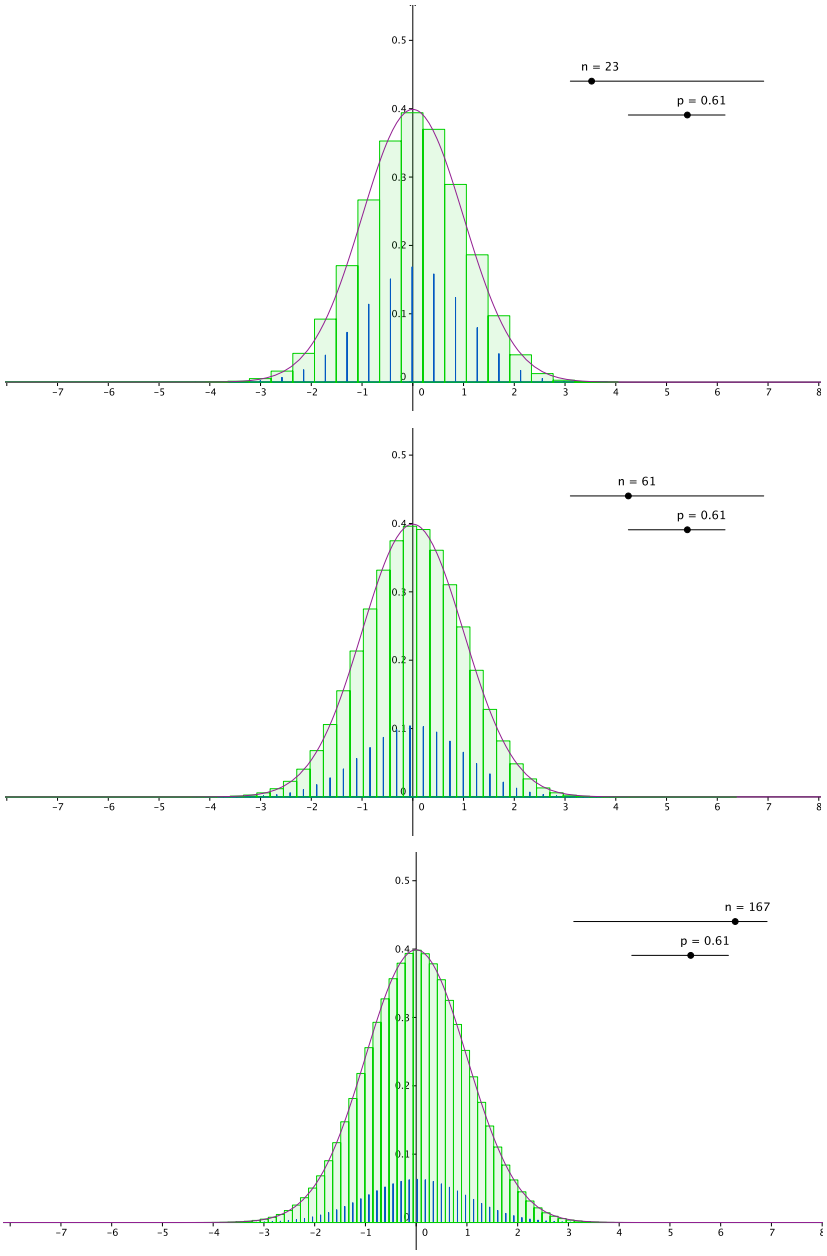
À ces diagrammes en bâtons associons des histogrammes dont les rectangles ont des bases de longueur $\frac{1}{\sqrt{np(1-p)}}$ centrées sur les bâtons et dont l'aire vaut la hauteur des bâtons correspondants (autrement dit les hauteurs des rectangles sont celles des bâtons multipliées par $\sqrt{np(1-p)}$). On obtient les dessins suivants.



Traçons maintenant la courbe représentative de la fonction

$$x \mapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

sur chacun de ces graphiques.



Les dessins que nous venons de tracer suggèrent que les histogrammes construits de cette façons s'approchent de la courbe de la fonction de densité de la loi normale centrée réduite, lorsque n grandit. C'est ce qu'affirme effectivement le théorème de De Moivre Laplace (démontré par De Moivre pour les variables binomiales, par Laplace dans un cas plus général).

Théorème 4.1. *Soit $(X_j)_{j\in\mathbb{N}^*}$ une suite de variables de Bernoulli de paramètre p indépendantes. Pour tous nombres réels a et b (avec $a < b$), on a*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_1 + X_2 + \ldots + X_n - np}{\sqrt{np(1-p)}} \in [a,b]\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Prenez un peu de temps pour bien comprendre que la procédure décrite pour obtenir les dessins, et les dessins eux mêmes, incitent bien à penser que cet énoncé est correct. De Moivre a obtenu ce résultat en utilisant ce qui est aujourd'hui appelé la formule de Stirling :

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

Pour les courageux nous donnons une démonstration plus moderne et générale en annexe.

5 Tests et estimation

5.1 Fluctuations

5.2 Raisonnement statistique

Dans les cours de statistique sur l'intervalle de confiance ou les tests on trouve des phrases comme les suivantes :

Un intervalle de confiance pour une proportion p à un niveau de confiance $1-\alpha$ est la réalisation, à partir d ?un échantillon, d'un intervalle aléatoire contenant la proportion p avec une probabilité supérieure ou égale à $1-\alpha$. Cet intervalle aléatoire est déterminé à partir de la variable aléatoire $F_n = \frac{X_n}{n}$ qui, à tout échantillon de taille n, associe la fréquence.

Il est intéressant de démontrer que, pour une valeur de p fixée, l'intervalle $[F_n - \frac{1}{\sqrt{n}}, F_n + \frac{1}{\sqrt{n}}]$ contient, pour n assez grand, la proportion p avec une probabilité au moins égale à 0,95.

On passe ensuite à des énoncés du type :

On fait un sondage auprès de 21521 personnes au sujet de leur préférences OUI ou NON. Sur ces 21521 personnes, 14268 disent préférer OUI. Donner un intervalle de confiance de niveau 0,95 pour la proportion de OUI dans la population totale.

La réponse attendue est : $[\frac{14268}{21521} - \frac{1}{\sqrt{21521}}, \frac{14268}{21521} + \frac{1}{\sqrt{21521}}]$. Mais alors que dans la phrase reprise plus haut $[F_n - \frac{1}{\sqrt{n}}, F_n + \frac{1}{\sqrt{n}}]$ désigne un intervalle aléatoire, ici $[\frac{14268}{21521} - \frac{1}{\sqrt{21521}}, \frac{14268}{21521} + \frac{1}{\sqrt{21521}}]$ n'est pas aléatoire. C'est une réalisation de l'intervalle précédent. De deux choses l'une, soit p appartient à l'intervalle, soit il n'y appartient pas.

Quel sens a le niveau 0,95 ? La probabilité que p appartienne à l'intervalle aléatoire $[F_n - \frac{1}{\sqrt{n}}, F_n + \frac{1}{\sqrt{n}}]$ est 0,95. Donc quand on calcule cet intervalle à chaque fois qu'on fait le sondage, si on répète le sondage, dans 95% des cas on obtiendra un intervalle contenant p. Quand on fait le sondage une fois, on se trompe ou non.

Quand on décide de considérer que p appartient à l'intervalle trouvé pour éventuellement prendre certaines mesures on fait donc un pari. De quel type est ce pari ? Dans 5% des

cas les intervalles qu'on me donne ne contienne pas p. Dans 5% des cas en considérant que p est dedans je me tromperai.

Imaginons qu'on ait une urne contenant cent boules : des boules rouges et des boules blanches. On ignore combien. On sait seulement que deux compositions sont possibles : une boule blanche, quatre-vingt-dix-neuf boules rouges ou bien une boule rouge, quatre-vingt-dix-neuf boules blanches. Comme d'habitude il est impossible d'ouvrir l'urne etc... et on ne peut faire qu'un chose : tirer une boule dans cette urne une seule fois. Question posée : quelle est la composition de l'urne ? Vous pouvez faire plusieurs choses sans doute (tirer à pile ou face votre réponse par exemple). Y-a-t-il une façon de procéder plus maligne ? Tirer une boule dans l'urne et regarder sa couleur. Que faire de cette information ? Elle est insuffisante pour répondre à la question à coup sûr. Vous avez très bien pu tirer la boule rouge unique parmi quatre-vingt-dix-neuf boules blanches. Avez-vous intérêt à parier sur la composition 99 rouges, 1 blanche si vous tirez une boule rouge ? En le faisant vous pariez que vous n'avez pas assisté à l'événement peu fréquent tirer une boule rouge parmi 99 blanches. Évidemment si c'est ce qui s'est produit vous vous trompez. C'est le sens du pari fait avec l'intervalle de confiance.

Il faut ajouter des remarques.

Si vous répétez l'expérience en adoptant cette stratégie (pour différentes urnes rencontrées) vous vous tromperez dans environ 95% des cas (loi des grands nombres). Si vous adoptez le pile ou face, dans environ la moitié des cas. La stratégie "statistique" semble meilleure de ce point de vue. Cela ne signifie pas qu'il n'existe pas de meilleure stratégie : par exemple ouvrir l'urne dans le cas qui nous occupe ou bien moins exigent mais aussi efficace ici obtenir le droit de tirer trois boules sans remise.

Vous pouvez aussi décider de ne rien décider dans de telles situations. Il est très probable alors que ne vous preniez que très rarement une décision quelconque ; je ne connais personne qui n'agisse qu'à coup sûr. Il est facile d'imaginer en revanche des situations où la décision est inévitable alors que les informations disponibles sont partielles. C'est même la règle. Et l'abstention est une décision, une option qui a elle aussi ses conséquences.

Imaginez que l'on vous donne le droit de répéter l'expérience de tirer une boule avec remise pour décider de sa composition. Disons deux fois. Si vous tirer deux fois une boule rouge. Là encore il se peut que vous ayez tiré deux fois de suite une boule rouge parmi 99 blanches. Cela se produit une fois sur 10000. Vous pouvez refuser de parier [2]. Trois fois une boule rouge. Il est possible d'avoir tiré trois fois de suite une boule rouge parmi 99 blanches. Cela se produit une fois sur 1000000. Vous pouvez refuser de parier. Ne seriez-vous pas de plus en plus tenté de parier ?

Il faut maintenant insister sur une différence entre l'exemple de l'urne et l'intervalle de

2. Si on tire une boule rouge et une blanche, on se retrouve bien embêté. Amusant : on accepte facilement de parier après un tirage, le tirage suivant peut nous faire hésiter.

confiance. Si vous prenez une décision à partir du pari que p appartient à $[0,25; 0,29]$, vous pouvez vous tromper. Mais ce n'est sans doute pas la même erreur si p vaut 0,2901 ou bien s'il vaut 0,56. Dans l'exemple de l'urne il y a une seule façon de se tromper. Pas pour l'intervalle de confiance. La longueur de l'intervalle de confiance dépend du niveau de confiance.

Reprenons l'exemple du sondage auprès de 21521 personnes au sujet de leur préférences OUI ou NON. Un intervalle asymptotique au niveau $1-\alpha$ pour la proportion de OUI dans la population totale est [3]

$$\left[\hat{p} - u_\alpha \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{21521}}, \hat{p} + u_\alpha \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{21521}}\right]$$

où $\hat{p} = \frac{14268}{21521}$ et u_α tel que $\Phi(u_\alpha) = 1 - \alpha/2$ (Φ fonction de répartition de la loi normale centrée réduite). On obtient différents intervalles pour différents niveaux :

$1-\alpha$	u_α	Intervalle de niveau $1-\alpha$
0,9	1,64	[0,658 ;0,668]
0,95	1,96	[0,657 ;0,669]
0,99	2,58	[0,655 ;0,671]
0,9999	3,89	[0,650 ;0,676]
0,99999	4,42	[0,649 ;0,677]

Parier que p se trouve dans l'intervalle $[0,658; 0,668]$ c'est parier à 1 contre 10. Parier qu'il se trouve dans $[0,649; 0,677]$ c'est parier à 1 contre 100000. On peut très bien avoir tort pour le premier intervalle mais raison pour le deuxième. Imaginons que vous ne vous intéressiez qu'à la réponse à la question : "p est-t-il supérieur à 0,5 ?". Alors si vous êtes dans l'intervalle $[0,649; 0,677]$ la réponse est oui. Vous pouvez parier à 1 contre 100000 que p est supérieur à 0,5. C'est le cas aussi bien sûr si p appartient à $[0,658; 0,668]$. La situation est donc plus compliquée que dans le cas de l'urne. Ce n'est pas parce que vous ne disposez que de l'intervalle à 90% que le sondage ne donne pas des informations à des niveaux de confiance supérieurs.

Imaginons qu'un intervalle de confiance à 95% soit $[0,49; 0,61]$. Le nombre 0,5 est dans cet intervalle. Il ne faut pas en conclure qu'on ne peut rien dire sur la probabilité que p soit supérieur à 0,5.

5.3 Intervalle de confiance ; sondages

3. Nous ne discutons du problème de l'approximation normale faite ici qui n'est qu'asymtotiquement vérifiée.

6 Quelques problèmes faisant intervenir les notions du cours

6.1 Probabilités et décision : l'exemple du pari de Pascal

Les raisonnements que Pascal a tenus pour répondre à des problèmes au sujet de jeux de hasard que lui a posés le chevalier de Méré sont considérés comme les premiers exemples de ce qu'on appelle aujourd'hui le calcul mathématique des probabilités et l'espérance mathématique (voir l'article de Derriennic [4]). Le fameux pari de Pascal est une application (discutable ?) de ces outils au domaine religieux. L'histoire littéraire a accordé une très grande place au fragment 397 des pensées où l'argument du pari est exposé.

Un extrait des pensées

Extrait du fragment 397 des pensées. Suivez l'argumentation de Pascal en gardant en tête la notion d'espérance mathématique associée à un jeu de pile ou face ou de dé.

Examinons donc ce point, et disons : "Dieu est, ou il n'est pas." Mais de quel côté pencherons-nous ? La raison n'y peut rien déterminer : il y a un chaos infini qui nous sépare. Il se joue un jeu, à l'extrémité de cette distance infinie, où il arrivera croix ou pile. Que gagerez-vous ? Par raison, vous ne pouvez faire ni l'un ni l'autre ; par raison, vous ne pouvez défendre nul des deux. Ne blâmez donc pas de fausseté ceux qui ont pris un choix ; car vous n'en savez rien.

- "Non ; mais je les blâmerai d'avoir fait, non ce choix, mais un choix ; car, encore que celui qui prend croix et l'autre soient en pareille faute, ils sont tous deux en faute : le juste est de ne point parier."

- Oui ; mais il faut parier. Cela n'est pas volontaire, vous êtes embarqué. Lequel prendrez-vous donc ? Voyons. Puisqu'il faut choisir, voyons ce qui vous intéresse le moins. Vous avez deux choses à perdre : le vrai et le bien, et deux choses à engager : votre raison et votre volonté, votre connaissance et votre béatitude ; et votre nature a deux choses à fuir : l'erreur et la misère. Votre raison n'est pas plus blessée, en choisissant l'un que l'autre, puisqu'il faut nécessairement choisir. Voilà un point vidé. Mais votre béatitude ? Pesons le gain et la perte, en prenant croix que Dieu est. Estimons ces deux cas : si vous gagnez, vous gagnez tout ; si vous perdez, vous ne perdez rien. Gagez donc qu'il est, sans hésiter.

- "Cela est admirable. Oui, il faut gager ; mais je gage peut-être trop."

- Voyons. puisqu'il y a pareil hasard de gain et de perte, si vous n'aviez qu'à gagner deux vies pour une, vous pourriez encore gagner ; mais s'il y en avait trois à gagner, il faudrait encore jouer (puisque vous êtes dans la nécessité de jouer), et vous seriez imprudent,

4. http://smf4.emath.fr/Publications/Gazette/2003/97/

lorsque vous êtes forcé de jouer, de ne pas hasarder votre vie pour en gagner trois, à un jeu où il y a pareil hasard de perte et de gain. Mais il y a une éternité de vie et de bonheur. Et cela étant, quand il aurait une infinité de hasards, dont un seul serait pour vous, vous auriez encore raison de gager un pour avoir deux ; et vous agiriez de mauvais sens, en étant obligé à jouer, de refuser de jouer une vie contre trois à un jeu où d'une infinité de hasards il y en a un pour vous, s'il y avait une infinité de vie infiniment heureuse à gagner Mais il y a ici une infinité de vie infiniment heureuse à gagner, un hasard de gain contre un nombre fini de hasards de perte, et ce que vous jouez est fini. Cela ôte tout parti : partout où est l'infini, et où il n'y a pas infinité de hasards de perte contre celui du gain, il n'y a point à balancer, il faut tout donner. Et ainsi, quand on est forcé à jouer, il faut renoncer à la raison pour garder la vie, plutôt que de la hasarder pour le gain infini aussi prêt à arriver que la perte du néant. Car il ne sert de rien de dire qu'il est incertain si on gagnera, et qu'il est certain qu'on hasarde, et que l'infinie distance qui est entre la certitude de ce qu'on s'expose, et l'incertitude de ce qu'on gagnera, égale le bien fini, qu'on expose certainement, à l'infini, qui est incertain. Cela n'est pas ; aussi tout joueur hasarde avec certitude pour gagner avec incertitude ; et néanmoins il hasarde certainement le fini pour gagner incertainement le fini, sans pécher contre la raison. Il n'y a pas infinité de distance entre cette certitude de ce qu'on s'expose et l'incertitude du gain ; cela est faux. Il y a, à la vérité, infinité entre la certitude de gagner et la certitude de perdre. Mais l'incertitude de gagner est proportionnée à la certitude de ce qu'on hasarde, selon la proportion des hasards de gain et de perte. Et de là vient que, s'il y a autant de hasards d'un côté que de l'autre, le parti est à jouer égal contre égal ; et alors la certitude de ce qu'on s'expose est égale à l'incertitude du gain : tant s'en faut qu'elle en soit infiniment distante. Et ainsi, notre proposition est dans un force infinie, quand il y a le fini à hasarder à un jeu où il y a pareils hasards de gain que de perte, et l'infini à gagner. Cela est démonstratif ; et si les hommes sont capables de quelque vérité, celle-là l'est.

Ma nuit chez Maud

Version 68+1 du Pari de Pascal, inventée par Eric Rohmer pour « Ma nuit chez Maud ».

L'histoire se passe à Clermont-Ferrand. Celui qui dit « je »le premier est ingénieur. L'autre (Vidal) est professeur de philosophie à la fac. Ils ont été amis il y a une douzaine d'années lorsqu'ils étaient étudiants. Ils se rencontrent par hasard dans un café.
– Ah, tiens ! dis-je, Pascal !
– Ca t'étonne ?
– C'est curieux. Je suis justement en train de le relire, en ce moment.
– Et alors ?
– Je suis très déçu.
– Dis, continue, ça m'intéresse.
– Ben, je ne sais pas. D'abord, j'ai l'impression de le connaître presque par c ?ur. Et puis ça ne m'apporte rien : je trouve ça assez vide. Dans la mesure où je suis catholique, ou tout au moins j'essaie de l'être, ça ne va pas du tout dans le sens de mon catholicisme actuel. C'est justement parce que je suis chrétien que je m'insurge contre ce rigorisme. Ou alors, si le christianisme c'est ça, moi je suis athée !... Tu es toujours marxiste ?

– Oui, précisément : pour un communiste, ce texte du pari est extrêmement actuel. Au fond, moi, je doute profondément que l'histoire ait un sens. Pourtant, je parie pour le sens de l'histoire, et je me trouve dans la situation pascalienne. Hypothèse A : la vie sociale et toute action politique sont totalement dépourvues de sens. Hypothèse B : l'histoire a un sens. Je ne suis absolument pas sûr que l'hypothèse B ait plus de chances d'être vraie que l'hypothèse A. Je vais même dire qu'elle en a moins. Admettons que l'hypothèse B n'a que dix pour cent de chances et l'hypothèse A quatre-vingt-dix pour cent. Néanmoins, je ne peux pas ne pas parier pour l'hypothèse B, parce qu'elle est la seule qui me permette de vivre. Admettons que j'aie parié pour l'hypothèse A et que l'hypothèse B se vérifie, malgré ses dix pour cent de chances, seulement : alors j'ai absolument perdu ma vie... Donc je

dois choisir l'hypothèse B, parce qu'elle est la seule qui justifie ma vie et mon action. Naturellement, il y a quatre-vingt-dix chances pour cent que je me trompe, mais ça n'a aucune importance.

– C'est ce qu'on appelle l'espérance mathématique, c'est-à-dire le produit du gain par la probabilité. Dans le cas de ton hypothèse B, la probabilité est faible, mais le gain est infini, puisque c'est pour toi le sens de ta vie, et pour Pascal le salut éternel.

Deux jours plus tard Dès qu'elles sont sorties, Vidal se lève et va vers la bibliothèque.

– Il doit bien y avoir un Pascal, ici. On a beau être franc-maçon... Il s'accroupit et découvre sur le rayon inférieur, une édition scolaire des Pensées. Il la feuillette. Je me suis levé et m'approche de lui.

– Pourrais-tu me dire, me demande-t-il, s'il y a une référence précise aux mathématiques dans le texte sur le pari. (Il lit) : « Partout où est l'infini et où il n'y a pas infinité de hasard de perte contre celui de gain, il n'y a point à balancer : il faut tout donner... et ainsi, quand on est forcé à jouer, il faut renoncer à la raison pour garder la vie », etc.

Il me tend le livre. J'y jette un coup d'oeil.

– C'est exactement ça, « l'espérance mathématique », dis-je. Dans le cas de Pascal, elle est toujours infinie... à moins que la probabilité de salut ne soit nulle. Puisque l'infini multiplié par zéro égale zéro. Donc l'argument ne vaut rien pour quelqu'un qui est absolument incroyant.

Critique

L'intérêt de jouer quand l'espérance est positive est lié à la possibilité de répéter le jeu. Cet intérêt est-il le même lorsque la répétition est impossible ?

Prenons un exemple.

Supposez que vous et votre famille disposiez d'un revenu annuel de 50 000 euros après impôts (selon l'INSEE le revenu moyen en France en 2007 avant impôts par ménage est de 39 000 euros environ, 54 000 pour les familles ayant deux enfants) et qu'aucune autre source de revenu ne soit imaginable. Et que je vous propose le jeu suivant. Vous misez 500 000 euros. Avec probabilité 1/100 je vous donne 101 fois votre mise, avec probabilité 99/100 vous perdez 500 000 euros.

L'espérance du gain est-elle positive ? Jouez-vous ?

Si je vous faisais crédit pendant autant de parties que vous voulez vous auriez peut-être intérêt à jouer. Mais je ne fais pas crédit. Vous devez déposez l'argent sur mon compte et je n'accepte de jouer qu'une fois. Il vous faut emprunter l'argent. Disons que vous l'empruntez à 3%. Faites le calcul (utilisant les suites géométriques) des sommes que vous devrez rembourser par mois pendant 40 ans si vous perdez (et pendant 20 ans).

Jouez-vous ?

Mais il est aussi possible de parler autrement du pari de Pascal, comme Jacques Prévert, par exemple (dans *Paroles*) :

LES PARIS STUPIDES

Un certain Blaise Pascal
etc... etc...

6.2 Métrologie, incertitude

6.3 Mathématiques de l'assurance (actuariat)

7 Annexe : démonstrations du théorème limite central

Plusieurs démonstrations existent. Pour les variables de lois binomiales on peut utiliser la formule de Stirling dont la démonstration fait appel à la notion de développement limité. Mais l'un des aspects remarquables du théorème limite central est qu'il est valable pour toute suite de variables indépendantes de même loi de carré intégrable. Une façon élégante de le voir est de considérer la fonction caractéristique et d'appliquer le théorème de Lévy. Nous proposons ici une preuve (ou plutôt une esquisse de preuve) n'utilisant que les développements limités et quelques manipulations élémentaires. La preuve complète serait assez longue : nous n'en donnerons que les idées principales.

Soit donc une suite X_i de variables aléatoires indépendantes de même loi de carré intégrable que nous considérerons centrées pour simplifier. Notons $\sigma^2 = \mathbb{E}(X_1^2)$ la variance de chacune des variables X_i. Notons S_n la somme $X_1 + \ldots + X_n$. Nous voulons montrer que la probabilité $\mathbb{P}(S_n/\sqrt{n} \in [a,b])$ converge vers

$$\int_a^b \exp(-\frac{x^2}{2\sigma^2})\frac{dx}{\sqrt{2\pi}\sigma}.$$

Première idée : montrer la convergence précédente revient à montrer la convergence de l'espérance de $f(S_n/\sqrt{n})$ vers l'espérance de $f(Y)$ où Y est une variable gaussienne centrée de variance σ^2 et f est l'application qui vaut 1 sur $[a,b]$, 0 sur son complémentaire.

Deuxième idée : si on arrive à montrer la même chose pour les fonctions f C^∞ à supports dans des intervalles finis, alors le théorème est vrai par approximation de l'indicatrice de $[a,b]$ par des fonctions régulières.

Troisième idée : si les variables Y_i sont indépendantes de loi gaussienne centrée de variance σ^2 alors $(Y_1 + \ldots + Y_n)/\sqrt{n}$ est gaussienne centrée de variance σ^2.

Il nous suffit donc de montrer que

$$\mathbb{E}\left(f\left(\sum_{i=1}^{n}\frac{X_i}{\sqrt{n}}\right)\right)-\mathbb{E}\left(f\left(\sum_{i=1}^{n}\frac{Y_i}{\sqrt{n}}\right)\right)$$

tend vers 0 quand n tend vers l'infini.

Quatrième idée : on écrit cette différence comme une somme de différences :

$$\begin{aligned}
&\mathbb{E}\left(f\left(\sum_{i=1}^{n}\frac{X_i}{\sqrt{n}}\right)\right)-\mathbb{E}\left(f\left(\sum_{i=1}^{n}\frac{Y_i}{\sqrt{n}}\right)\right)\\
&\quad=\mathbb{E}\left(f\left(\sum_{i=1}^{n}\frac{X_i}{\sqrt{n}}\right)-f\left(\sum_{i=1}^{n}\frac{Y_i}{\sqrt{n}}\right)\right)\\
&\quad=\sum_{k=1}^{n}\mathbb{E}\left(f\left(\sum_{i=1}^{k-1}\frac{X_i}{\sqrt{n}}+\sum_{i=k}^{n}\frac{Y_i}{\sqrt{n}}\right)-f\left(\sum_{i=1}^{k}\frac{X_i}{\sqrt{n}}+\sum_{i=k+1}^{n}\frac{Y_i}{\sqrt{n}}\right)\right)
\end{aligned}$$

Les sommes apparaissant comme variables de f à k donné diffèrent de $\frac{X_k}{\sqrt{n}}-\frac{Y_k}{\sqrt{n}}$.

Cinquième idée : on fait des développements limités de f à k fixé au point $\sum_{i=1}^{k-1}\frac{X_i}{\sqrt{n}}+\sum_{i=k+1}^{n}\frac{Y_i}{\sqrt{n}}$. On obtient :

$$\begin{aligned}
f\left(\sum_{i=1}^{k-1}\frac{X_i}{\sqrt{n}}+\sum_{i=k}^{n}\frac{Y_i}{\sqrt{n}}\right) &= f\left(\sum_{i=1}^{k-1}\frac{X_i}{\sqrt{n}}+\sum_{i=k+1}^{n}\frac{Y_i}{\sqrt{n}}\right)+f'\left(\sum_{i=1}^{k-1}\frac{X_i}{\sqrt{n}}+\sum_{i=k+1}^{n}\frac{Y_i}{\sqrt{n}}\right)\frac{Y_k}{\sqrt{n}}\\
&\quad+f''\left(\sum_{i=1}^{k-1}\frac{X_i}{\sqrt{n}}+\sum_{i=k+1}^{n}\frac{Y_i}{\sqrt{n}}\right)\frac{Y_k^2}{2n}+O(n^{-3/2})
\end{aligned}$$

et

$$\begin{aligned}
f\left(\sum_{i=1}^{k}\frac{X_i}{\sqrt{n}}+\sum_{i=k+1}^{n}\frac{Y_i}{\sqrt{n}}\right) &= f\left(\sum_{i=1}^{k-1}\frac{X_i}{\sqrt{n}}+\sum_{i=k+1}^{n}\frac{Y_i}{\sqrt{n}}\right)+f'\left(\sum_{i=1}^{k-1}\frac{X_i}{\sqrt{n}}+\sum_{i=k+1}^{n}\frac{Y_i}{\sqrt{n}}\right)\frac{X_k}{\sqrt{n}}\\
&\quad+f''\left(\sum_{i=1}^{k-1}\frac{X_i}{\sqrt{n}}+\sum_{i=k+1}^{n}\frac{Y_i}{\sqrt{n}}\right)\frac{X_k^2}{2n}+O(n^{-3/2})
\end{aligned}$$

En faisant la différence on obtient :

$$\begin{aligned}
&f\left(\sum_{i=1}^{k-1}\frac{X_i}{\sqrt{n}}+\sum_{i=k}^{n}\frac{Y_i}{\sqrt{n}}\right)-f\left(\sum_{i=1}^{k}\frac{X_i}{\sqrt{n}}+\sum_{i=k+1}^{n}\frac{Y_i}{\sqrt{n}}\right)\\
&\quad=f'\left(\sum_{i=1}^{k-1}\frac{X_i}{\sqrt{n}}+\sum_{i=k+1}^{n}\frac{Y_i}{\sqrt{n}}\right)\left(\frac{Y_k}{\sqrt{n}}-\frac{X_k}{\sqrt{n}}\right)\\
&\quad+f''\left(\sum_{i=1}^{k-1}\frac{X_i}{\sqrt{n}}+\sum_{i=k+1}^{n}\frac{Y_i}{\sqrt{n}}\right)\left(\frac{Y_k^2}{2n}-\frac{X_k^2}{2n}\right)+O(n^{-3/2})
\end{aligned}$$

Comme les variables sont indépendantes, centrées et de mêmes variances, quand on prend l'espérance, il ne reste plus que le terme $O(n^{-3/2})$:

$$\begin{aligned}
&\mathbb{E}\left(f\left(\sum_{i=1}^{k-1}\frac{X_i}{\sqrt{n}}+\sum_{i=k}^{n}\frac{Y_i}{\sqrt{n}}\right)-f\left(\sum_{i=1}^{k}\frac{X_i}{\sqrt{n}}+\sum_{i=k+1}^{n}\frac{Y_i}{\sqrt{n}}\right)\right)\\
&=\mathbb{E}\left(f'\left(\sum_{i=1}^{k-1}\frac{X_i}{\sqrt{n}}+\sum_{i=k+1}^{n}\frac{Y_i}{\sqrt{n}}\right)\right)\mathbb{E}\left(\frac{Y_k}{\sqrt{n}}-\frac{X_k}{\sqrt{n}}\right)\\
&\quad+\mathbb{E}\left(f''\left(\sum_{i=1}^{k-1}\frac{X_i}{\sqrt{n}}+\sum_{i=k+1}^{n}\frac{Y_i}{\sqrt{n}}\right)\right)\mathbb{E}\left(\frac{Y_k^2}{2n}-\frac{X_k^2}{2n}\right)+O(n^{-3/2})
\end{aligned}$$

On a un tel terme pour chaque k donc n tels termes, dont on fait la somme. Reste une quantité de l'ordre de $n^{-1/2}$ qui tend vers 0. Le théorème est démontré.

Références

[1] J.-P. Benzecri, *Histoire et préhistoire de l'analyse des données*

[2] D. Perrin

[3] M. Gromov, *Introduction aux mystères*, .

[4] I. Hacking, L'émergence de la probabilité

[5] B. Pascal, *Les pensées*, .

[6] *Images des mathématiques.*

[7] Y. Chevallard

[8] C. Schwartz, *La preuve par les chiffres : de quoi s'agit-il ?*, http ://publications-sfds.math.cnrs.fr/index.php/StatEns/article/view/125/115.

[9] A. Vessereau, *La statistique*, Presses universitaires de France, Collection Que sais-je ?