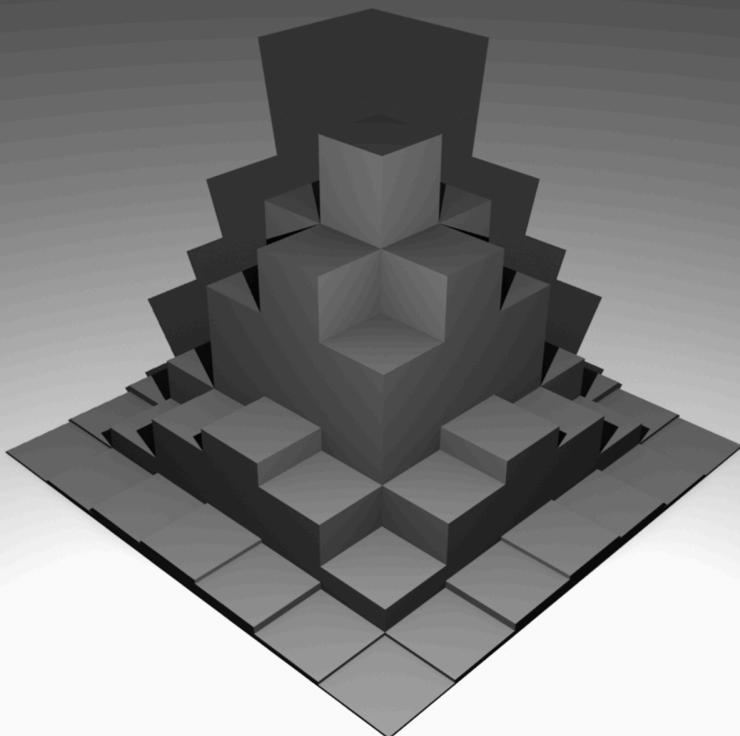


Probabilités, statistiques et analyses multicritères



Mathieu ROUAUD

Probabilités, statistiques et analyses multicritères

*Application
aux sciences expérimentales
(exercices corrigés)*

Mathieu ROUAUD
Professeur Agrégé de Sciences Physiques
en classes préparatoires aux grandes écoles d'ingénieurs
Diplômé en Physique Théorique

Pour un meilleur partage de la connaissance et l'accès au plus grand nombre, le livre est en licence libre, le livre numérique est gratuit et pour minimiser le coût de la version papier, il est imprimé en noir et blanc et sur papier économique.

Livres complets numérique et papier,

avec tous les exercices corrigés :

sur www.lulu.com

Pour contacter l'auteur :

ecrire@incertitudes.fr

Boudiguen 29310 Querrien

Ce livre est sous licence Creative Commons Attribution-Non Commercial 3.0.



Vous êtes libres :



de reproduire, distribuer et communiquer
cette création au public ,



de modifier cette
création .

Selon les conditions suivantes :



Attribution. Vous devez citer le nom de l'auteur original de la manière indiquée par l'auteur de l'œuvre ou le titulaire des droits qui vous confère cette autorisation (mais pas d'une manière qui suggérerait qu'ils vous soutiennent ou approuvent votre utilisation de l'œuvre).



Pas d'Utilisation Commerciale. Vous n'avez pas le droit d'utiliser cette création à des fins commerciales.

- A chaque réutilisation ou distribution de cette création, vous devez faire apparaître clairement au public les conditions contractuelles de sa mise à disposition. La meilleure manière de les indiquer est un lien vers cette page web : <http://creativecommons.org/licenses/by-nc/2.0/deed.fr>
- Chacune de ces conditions peut être levée si vous obtenez l'autorisation du titulaire des droits sur cette œuvre.
- Rien dans ce contrat ne diminue ou ne restreint le droit moral de l'auteur ou des auteurs.

Avant-propos

Cet ouvrage se veut accessible et pédagogique. Il est le fruit d'interrogations personnelles sur la nature probabiliste des mesures en sciences. Dans un cursus classique ces aspects ne sont pas, ou peu, abordés. Il est important que les fondements expérimentaux et pratiques des sciences soient complémentaires d'une science au tableau en cours magistraux. Il existe une beauté scientifique qui naît de l'interaction entre la théorie et l'expérience.

Tout en introduisant les principes fondamentaux de la statistique, cet ouvrage explique comment déterminer les incertitudes dans différentes situations expérimentales. Beaucoup d'exemples sont issus de cours et travaux pratiques réalisés en math sup.

Le dernier chapitre aborde les analyses multicritères telles qu'elles ont été enseignées en deuxième année d'école d'ingénieur (École Hubert Curien / Maîtrise de l'efficacité énergétique).

Bonne lecture !

Remerciements :

Je remercie Éric NOIZET (professeur Agrégé de Chimie en prépa) et Grégoire BONNET (ingénieur charpentier) pour leurs multiples apports à la clarté pédagogique de l'ouvrage. Un grand merci à Reine pour sa relecture précise et consciencieuse. Pleins de mercis, à Aurélien SEMACH (étudiant) et, aux enseignants de sciences-physiques Françoise MARCADET (pour ses contributions en métrologie) et Julien BONVALET.

Merci à la vie et à tous ceux qui m'ont précédé.

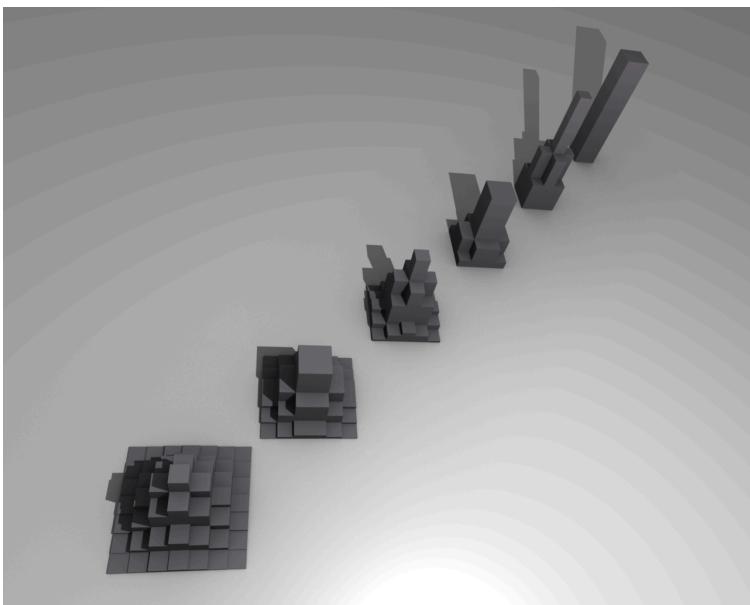


Table des matières

I. VARIABLE ALÉATOIRE.....	1
A. Grandeur et mesures.....	1
B. Centre d'une distribution.....	1
C. Dispersion d'une distribution.....	3
D. Exemples de distributions.....	4
E. Théorème central limite.....	7
1) Population et échantillons.....	7
2) Le théorème central limite.....	10
3) Coefficient de Student et incertitude.....	12
4) Exemples.....	15
F. Distribution de Gauss.....	19
1) Définition d'une distribution continue.....	19
2) Courbe de Gauss.....	20
3) Loi normale standard.....	23
G. Test d'hypothèse.....	24
H. Test du Khi-deux.....	30
I. Sources des incertitudes.....	33
J. Exercices.....	37
II. CORRÉLATIONS ET INDÉPENDANCES.....	47
A. Coefficient de corrélation.....	47
B. Formule de propagation des incertitudes.....	52
1) Formule de propagation des écarts-types.....	52
2) Calcul d'incertitude.....	53
C. Régression linéaire.....	58
1) Principe et formules.....	58
2) Détermination du zéro absolu.....	62
3) Régression avec barres d'erreurs.....	64
4) Linéarisation.....	67
5) Comparaison des méthodes.....	68

D. Régression généralisée.....	75
1) Principe.....	75
2) Régression polynomiale.....	77
3) Régression non linéaire.....	80
E. Exercices.....	86
III. LOIS DE PROBABILITÉS.....	101
A. Lois discrètes.....	102
1) Loi binomiale.....	102
2) Loi géométrique.....	103
3) Loi de Poisson.....	105
B. Lois continues.....	107
1) Loi uniforme.....	107
2) Loi exponentielle.....	109
3) Loi normale.....	110
4) Loi de Student.....	110
5) Loi du Khi-Deux.....	111
C. Fonctions de variables à densité.....	113
D. Simulation numérique.....	116
E. Exercices.....	119
IV. ESTIMATEURS.....	124
A. Qualité d'un estimateur.....	124
1) Biais.....	124
2) Risque.....	125
B. Construction d'estimateurs.....	127
1) Méthode des moments.....	127
2) Méthode du maximum de vraisemblance.....	131
C. Estimation par intervalle.....	135
D. Exercices.....	144
V. APPROCHES MULTICRITÈRES.....	150
A. Méthode par pondération.....	151

B.	Analyse en Composantes Principales.....	154
1)	Principes.....	154
2)	Une deuxième illustration.....	167
3)	Résumé.....	170
C.	Exercices.....	171
VI.	COMPLÉMENTS.....	174
A.	Mesure avec une règle.....	174
B.	Métrologie.....	186
C.	Thermodynamique.....	194
D.	Indépendance des variables.....	200
VII.	DEVOIRS.....	202
A.	Devoir Suricate.....	202
B.	Devoir Narval.....	209
VIII.	TRAVAUX PRATIQUES.....	217
A.	Mesure d'un indice lumineux.....	217
B.	Le miroir sphérique.....	221
C.	Relation de conjugaison d'une lentille.....	224
D.	Dioptrés et lentilles minces sphériques.....	226
IX.	OUTILS MATHÉMATIQUES.....	228
X.	CORRECTIONS.....	233
XI.	Bibliographie / Sources / Logiciels / Illustrations.....	303
XII.	TABLES / Index.....	308
A.	Loi normale centrée réduite.....	308
B.	Coefficients de Student.....	309
C.	Valeurs critiques de Khi-deux.....	310



I. VARIABLE ALÉATOIRE

A. *Grandeurs et mesures*

Soit X une variable aléatoire et n réalisations $\{x_i\}$ de cette variable.

Nous pouvons simplement estimer une grandeur classique : par exemple, combien y-a-t-il de jours dans une semaine ? La réponse est sans ambiguïté. Par contre pour une grandeur statistique l'approche est plus subtile. Imaginons des étudiants qui font des expériences de calorimétrie pour mesurer la capacité thermique de l'eau¹. Les différents groupes mesurent les valeurs suivantes : {5100; 4230; 3750; 4560; 3980} J/K/kg. Que vaut alors la capacité ? Nous donnerons dans ce chapitre une réponse à cette question. Elle sera de nature probabiliste.

B. *Centre d'une distribution*

Nous cherchons une caractéristique du centre de la distribution des observations $\{x_i\}$. Il en existe plusieurs, le *mode*, par exemple, est facile à déterminer, il s'agit de la valeur la plus représentée (illustrations page 4). Nous avons

1 PHYSIQUE : Quantité d'énergie à fournir à un kilogramme d'eau pour que sa température s'élève de 1°C. L'eau emmagasine ainsi de l'énergie et peut la restituer par la suite en diminuant sa température. Tables : $c_{\text{eau}} = 4180$ Joules par degré Celsius et par kilogrammes.

aussi la *médiane* qui correspond à la valeur qui sépare la distribution en deux parties égales. Mais la plus utilisée est la **moyenne** qui représente au mieux le centre d'une distribution :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_i + \dots + x_n}{n} \quad \text{soit} \quad \boxed{\bar{x} = \frac{\sum_{i=1}^n x_i}{n}}^2$$

Pour la capacité thermique de l'eau nous obtenons :

$$\bar{c} = \frac{5100 + 4230 + 3750 + 4560 + 3980}{5} = 4324 \text{ J/K/kg}$$

Nous avons considéré la moyenne arithmétique. Nous aurions pu prendre la moyenne géométrique :

$$\bar{x} = \sqrt[n]{\prod x_i}$$

Par exemple, pour deux températures 20°C et 40°C , la moyenne géométrique est $\sqrt{20^\circ\text{C} \cdot 40^\circ\text{C}} \approx 28,3^\circ\text{C}$ alors que la moyenne arithmétique est 30°C . Dans la pratique on constate que la moyenne arithmétique est mieux adaptée.

² MATH : se dit «la moyenne de x est égale à la somme de 1 à n des x_i , le tout divisé par n ». Pour la moyenne géométrique nous considérons la racine $n^{\text{ième}}$ du produit des x_i . x , "x moyen", se dit aussi "x barre".

C. Dispersion d'une distribution

Il s'agit d'estimer ce que nous pourrions aussi appeler la largeur d'une distribution. La grandeur la plus simple à déterminer est l'*étendue*, différence entre les valeurs maximale et minimale. Mais celle-ci est très sensible aux valeurs extrêmes qui ne sont pas toujours représentatives, et peuvent même parfois être absurdes.

Dans les faits, la grandeur la plus utilisée est l'**écart-type** :

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Pour l'écart-type de la capacité thermique de l'eau nous obtenons :

$$s_c = \sqrt{\frac{(5100-4324)^2 + (4230-4324)^2 + (3750-4324)^2 + (4560-4324)^2 + (3980-4324)^2}{4}}$$

$$\text{soit } s_c \approx 530 \text{ J/K/kg}$$

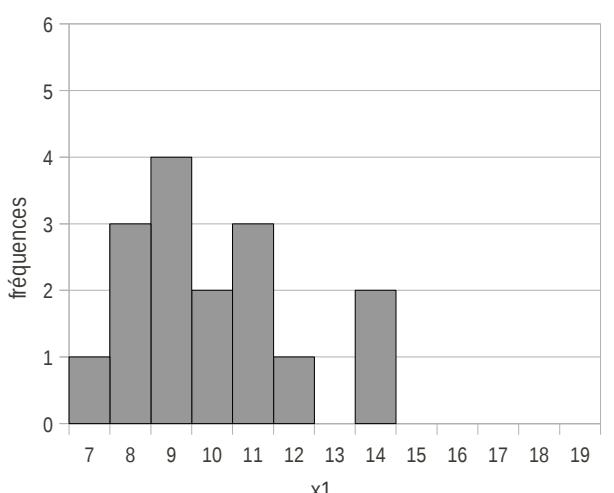
Nous pourrions aussi considérer l'*écart moyen* par rapport à la moyenne (voir l'exercice 1).

Pour l'écart-type si nous divisions par n au lieu de $n-1$, nous obtiendrions l'*écart quadratique moyen*. Le choix de l'écart-type sera justifié par la simplicité des formules qui en découlent (justification complète p126). De plus nous travaillons souvent avec n grand et la différence entre les deux types d'écart quadratiques est alors minime.

D. Exemples de distributions

Cas 1 :

	x_1
x_1^1	11
x_1^2	9
x_1^3	10
x_1^4	14
x_1^5	11
x_1^6	8
x_1^7	9
x_1^8	12
x_1^9	7
x_1^{10}	8
x_1^{11}	8
x_1^{12}	9
x_1^{13}	11
x_1^{14}	14
x_1^{15}	10
x_1^{16}	9



moyenne = 10

mode= 9

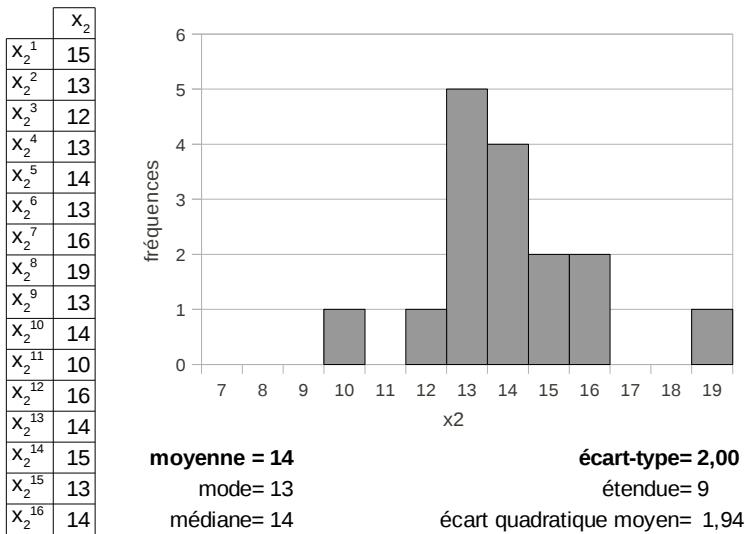
médiane= 9,5

écart-type= 2,07

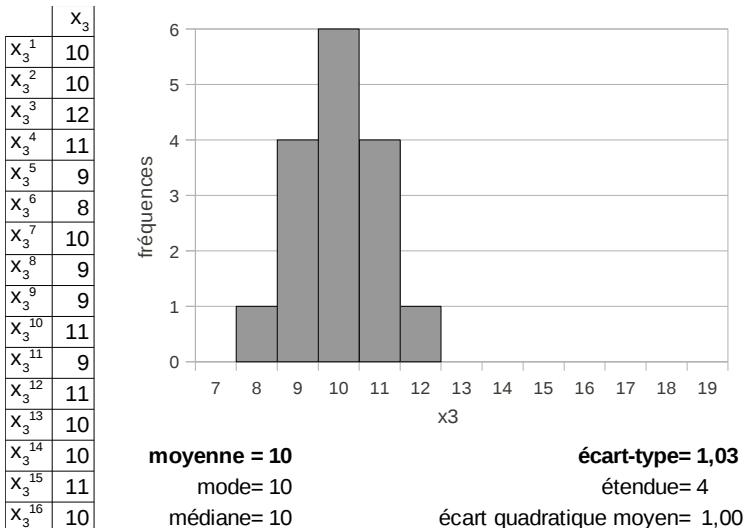
étendue= 7

écart quadratique moyen= 2,00

Cas 2 :



Cas 3 :



La moyenne n'est pas toujours la valeur la plus représentée (*cas 1 et 2*) et elle peut même dans certains cas être absente. Dans le *cas 3* la courbe est symétrique ce qui implique l'égalité de la médiane et de la moyenne.

Sur les trois exemples certaines valeurs sont représentées plusieurs fois, on parle alors de la *fréquence* f_i d'une valeur x_i . Nous avons $n = \sum_{i=1}^c f_i$, où c correspond au nombre de valeurs de x_i différentes auxquelles nous attribuons une fréquence (dans la suite, c sera aussi le nombre de classes). La moyenne et l'écart-type peuvent alors s'exprimer ainsi :

$$\bar{x} = \frac{\sum_{i=1}^c f_i \cdot x_i}{n} = \sum_{i=1}^c \frac{f_i}{n} \cdot x_i \quad s = \sqrt{\frac{\sum_{i=1}^c f_i (x_i - \bar{x})^2}{n-1}}$$

Parfois on rassemble les valeurs par *classe*, par exemple si nous nous intéressons à la taille des habitants d'une ville nous pouvons rassembler tous les habitants qui ont une taille comprise entre 160 cm et 170 cm dans le même ensemble appelé classe. Leur nombre dans cette classe est la fréquence (ou effectif de la classe) et la valeur est prise égale au milieu de la classe, ici 165 cm (démarche illustrée dans l'exercice 5).

Plus la courbe est ramassée sur son centre plus l'écart-type est faible (sur le *cas 3* l'écart est deux fois plus faible que sur les *cas 1 et 2*).

E. Théorème central limite

1) Population et échantillons

Considérons une ville d'un million d'habitants. Pour sonder la population nous pouvons interroger un échantillon de seulement mille personnes tirées au hasard. A partir de cet échantillon de $n=1000$ individus, nous pouvons, grâce à la statistique, avoir des informations sur la population toute entière. Plus la taille de l'échantillon est grand, plus les résultats seront précis. Nous appelons \bar{x} la moyenne de l'échantillon et s son écart-type. Pour la population nous notons μ (lettre grec mu) la moyenne et σ (sigma) l'écart-type. Plus l'échantillon est grand, plus les valeurs de \bar{x} et de s de cet échantillon sont amenées à se rapprocher de celles μ et σ de la population.

Comme dans le cas des sondages d'opinion avec des échantillons de l'ordre de mille personnes, si nous mesurons la taille de mille habitants qui sont choisis au hasard parmi la population d'une ville d'un million d'habitants, la moyenne de la taille sur cet échantillon a de fortes chances d'être proche de celle sur l'ensemble de la population, mais n'a aucune raison de lui être égale.

Illustrons maintenant par le jeu du lancer de pièces. A chaque lancer nous obtenons une réalisation de la variable aléatoire *pile ou face*. Ici la population est infinie, nous pouvons répéter le lancer indéfiniment et avoir une infinité de mesures. De plus, les probabilités étant connues, nous pouvons déterminer à l'avance les caractéristiques de la population.

Tout d'abord, quand la taille de l'échantillon devient très grande et donc infinie, l'échantillon s'identifie à la population : $\mu = \lim_{n \rightarrow \infty} \bar{x}$ ³.

Ensuite nous voyons apparaître la notion de probabilité :
 $p_i = \lim_{n \rightarrow \infty} \frac{f_i}{n}$ où p_i est la probabilité de réalisation de l'événement x_i .

D'après la formule page 6 nous avons ainsi l'expression de la moyenne pour la population : $\mu = \sum p_i \cdot x_i$.

Nous avons $\sum p_i = 1$, car nous considérons tous les événements possibles (1=100%).

Nous associons $x_0=0$ à l'observation *face*, et $x_1=1$ pour *pile*. Comme la pièce est équilibrée $p_0=p_1=1/2=0,5=50\%$ et $\mu=p_0 \cdot x_0 + p_1 \cdot x_1$. Nous avons jugé inexistant l'événement *la pièce reste sur la tranche*.

De même nous avons : $\sigma = \lim_{n \rightarrow \infty} s$ et en prenant la limite de la formule pour s page 6 nous obtenons $\sigma = \sqrt{\sum p_i \cdot (x_i - \mu)^2}$ (avec pour n grand, $n-1$ pris égal à n).

Au final : $\mu = 0,5$ et $\sigma = 0,5$.

³ MATH : se lit « μ est égal à la limite de \bar{x} quand n tend vers l'infini».

Prélevons un échantillon en lançant neuf pièces :

$$\{0; 1; 1; 0; 1; 1; 1; 0; 0\}.$$

Nous avons alors $\bar{x} \simeq 0,56$ et $s \simeq 0,53$.

Supposons que nous tirions de nombreuses fois neuf pièces aléatoirement au sein de la même population. A chaque fois nous aurions un résultat différent pour \bar{x} .

Par exemple pour deux autres échantillons prélevés :

$$\{1; 1; 0; 1; 1; 0; 1; 0; 0\} \text{ avec } \bar{x} \simeq 0,67 \text{ et } s \simeq 0,50$$

$$\text{et } \{0; 1; 0; 0; 0; 0; 1; 0; 0\} \text{ avec } \bar{x} \simeq 0,22 \text{ et } s \simeq 0,44$$

Quelle serait alors la répartition de l'ensemble de ces résultats ? (appelée la distribution d'échantillonnage)

Les valeurs obtenues pour les échantillons sont en général différentes de celles de la population, mais plus l'échantillon est grand plus il est probable que les valeurs soient proches de celle de la population. Cas d'un échantillon de taille $n=50$ où $\bar{x} \simeq 0,520$ et $s \simeq 0,505$:

$$\{001000111111001011110001001101101000110000011101\}$$

2) Le théorème central limite

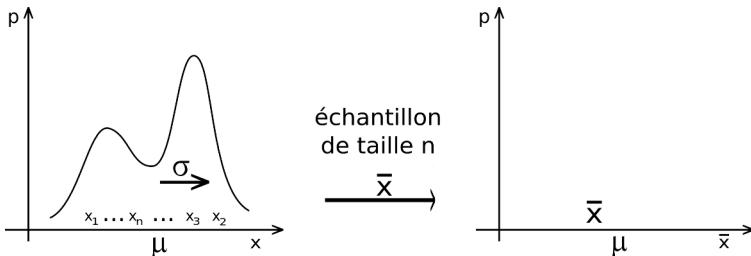
THÉORÈME CENTRAL LIMITÉ :

Nous prélevons au sein d'une population des échantillons aléatoires de taille n , la moyenne de l'échantillon \bar{x} varie autour de la moyenne de la population μ avec un écart-type égal à σ/\sqrt{n} , où σ est l'écart-type de la population.

Quand n croît la distribution d'échantillonnage de \bar{x} est de plus en plus concentrée autour de μ et devient de plus en plus proche d'une distribution de Gauss.

Nous décrirons prochainement ce qu'est une distribution de Gauss, encore appelée loi normale, dans l'immédiat nous considérerons simplement une courbe en cloche. C'est un théorème très important, **quelque soit** la forme de la distribution de la population, la distribution d'échantillonnage tend vers une **gaussienne**, et sa dispersion est donnée par le théorème central limite.

Illustrons sur des schémas :



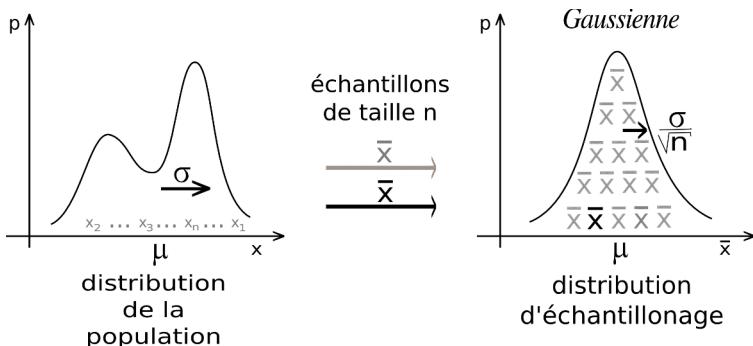
A gauche nous avons la probabilité p d'un événement x (distribution de la population). Par exemple, pour une po-

pulation d'un million d'habitants, pourrait être ici représenté la probabilité p qu'ils aient une taille donnée x . Si nous pouvions mesurer la taille de tous les habitants, nous pourrions déterminer exactement leur taille moyenne μ et son écart-type σ . C'est d'un point de vue pratique difficile, ou même impossible, alors nous décidons de mesurer la taille de mille habitants seulement. Pour que ce soit caractéristique nous tirons ces mille personnes au hasard.

Nous obtenons mille mesures de tailles de x_1 à x_{1000} . De cet échantillon de taille $n=1000$ nous calculons une moyenne \bar{x} et un écart-type s . Nous pensons que \bar{x} est proche de μ , mais à la fois il n'y a aucune raison qu'il soit égal à μ . Nous plaçons cette valeur de \bar{x} sur la partie droite de la figure page 10.

Nous prenons un nouvel échantillon aléatoire de mille personnes et nous plaçons un nouveau \bar{x} .

Nous répétons ensuite cette opération un grand nombre de fois. Nous voyons à droite la distribution des échantillons obtenue :



3) Coefficient de Student et incertitude

Le théorème central limite s'applique dans la limite des grands nombres. Dans le cas particulier où la distribution de la population est normale nous pouvons l'appliquer dès n petit grâce aux coefficients de Student t .

Intervalle de fluctuation :

Si μ et σ sont connus la distribution d'échantillonnage est aussi gaussienne et les fluctuations statistiques attendues se situent avec une probabilité de $p\%$ entre $\mu - t_{\alpha/2} \sigma / \sqrt{n}$ et $\mu + t_{\alpha/2} \sigma / \sqrt{n}$.

Les valeurs de t sont lues dans la table page 309.

Intervalle de confiance :

Dans le cas du calcul d'incertitudes μ et σ ne sont pas connus et nous les estimons à partir de l'échantillon avec \bar{x} et s . Du fait d'une statistique faible il y a alors un élargissement connu donné par la loi de Student :

$$\mu = \bar{x} \pm t \cdot \frac{s}{\sqrt{n}}$$

Le coefficient de Student t dépend de n et inclut la *confiance* que nous voulons donner au résultat. Si la confiance est de 95%, nous avons 95 chances sur 100 que μ soit compris entre $\bar{x} - t \cdot s / \sqrt{n}$ et $\bar{x} + t \cdot s / \sqrt{n}$.

Nous reconnaissons ici la notion d'***incertitude*** Δx ⁴:

$$x = \bar{x} \pm \Delta x \quad \text{avec} \quad \Delta x = t \cdot \frac{s}{\sqrt{n}}$$

Δx est aussi appelée l'*incertitude absolue* et $\Delta x / |\bar{x}|$ l'*incertitude relative*.

Reprendons l'expérience de calorimétrie décrite page 1, supposons que nous voulions maintenant connaître la capacité thermique de l'eau avec une confiance de 95%. Comme souvent en sciences expérimentales nous considérons la distribution des données gaussienne, en effet de par l'influence de nombreux facteurs indépendants sur la valeur des grandeurs mesurées nous nous attendons, toujours en vertu du théorème central limite, à avoir normalité des fluctuations.

Nous trouvons dans la table pour quatre degrés de liberté ($ddl=n-1$) $t=2,78$.

D'où : $c = \bar{c} \pm t \cdot s_c / \sqrt{n} = 4320 \pm 660 \text{ J/K/kg}$ à 95%.

Ici suite à la dispersion des valeurs mesurées par les étudiants $\Delta c / \bar{c} \approx 15\%$. Les mesures en calorimétrie ne sont pas très précises. La valeur attendue, ici connue, est bien dans l'intervalle : $3660 < 4180 < 4980$.

En sciences expérimentales nous nous efforçons de quantifier l'ensemble des phénomènes naturels. Mais, par la nature même de la démarche expérimentale, les différents paramètres qui permettent de décrire une situation ne sont

⁴ MATH : se lit "delta x".

pas parfaitement connus. Nous n'avons pas simplement une valeur numérique associée à chaque caractéristique, mais un intervalle pour une confiance donnée. En toute rigueur, à toute grandeur expérimentale doit être associée son incertitude avec sa confiance.

Cas particuliers :

Cas des grands nombres : la taille de l'échantillon n est suffisamment grande pour pouvoir appliquer directement le théorème de la limite centrée. La distribution d'échantillonnage est normale indépendamment de la distribution de la population. Nous n'avons pas à nous soucier de la loi de Student qui de toute façon s'identifie à une loi de Gauss dans ce cas.

Cas des petits nombres et normalité des données : nous appliquons la loi de Student comme précédemment.

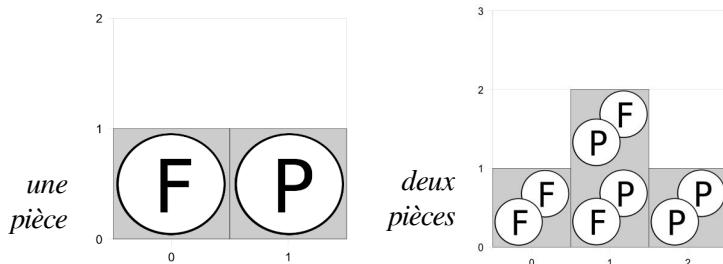
Cas des petits nombres et non normalité des données : par exemple, nous constatons en calculant les coefficients d'asymétrie et aplatissement des données qu'elles ne correspondent pas une distribution normale. Nous sommes bloqué et nous devons faire une étude au cas par cas. Par exemple pour des données qui obéissent à une loi uniforme nous avons l'intervalle de fluctuation donné p12 qui fonctionne dès $n=2$ (se montre en s'appuyant sur les données de l'article *Mesure avec une règle* p174) . Par contre pour un loi de Bernoulli de paramètre $p=0,24$ et un échantillon de taille $n=2$, l'intervalle de fluctuation à 50% comporte 0% des valeurs... Un cas plus complexe p142 montre pour $n=12$, par comparaison avec une simulation numérique, que le théorème central sous-estime l'intervalle de confiance.

4) Exemples

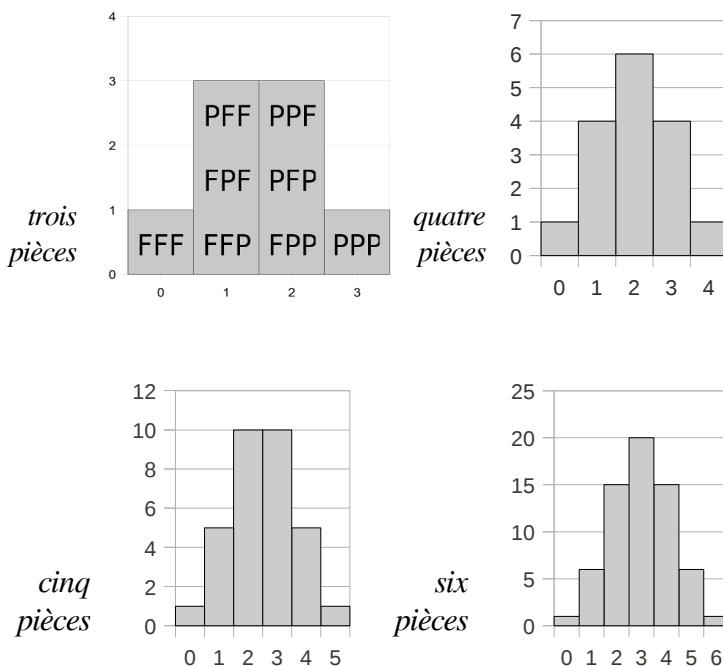
Concrètement un grand nombres de facteurs aléatoires vont influer sur la mesure d'une grandeur, facteurs indépendants, qui, quelque soient leurs natures vont au final générer une distribution gaussienne des valeurs. Prenons deux exemples, le lancer de pièces et celui de dés à six faces.

Pour les pièces nous comptons le nombre de piles à chaque lancer de plusieurs pièces. Nous dénombrons le nombre de possibilités pour un nombre de piles donné. Pour une pièce, une possibilité pour zéro pile (face, F) et une possibilité pour un pile (P).

Pour deux pièces, une possibilité pour zéro pile (F F), deux possibilités pour un pile (F P ou P F) et une possibilité pour deux piles (P P). Plus le nombres de pièces lancées simultanément est grand, plus nous tendons vers une distribution gaussienne.



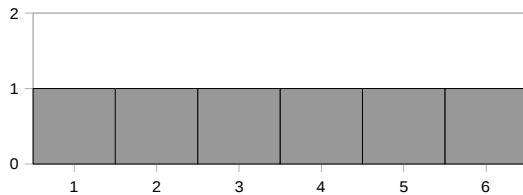
Nous pouvons obtenir en ordonnée la probabilité, en divisant par le nombres de possibilités 2^n , et en abscisse la moyenne pour chaque pièce, en divisant par le nombre de pièces n . Pour $n=1$ nous avons alors la distribution de la population et ensuite les distributions d'échantillonnage pour différentes valeurs de n .



De même pour les dés nous énumérons les possibilités pour leur somme et nous tendons aussi vers une gaussienne.

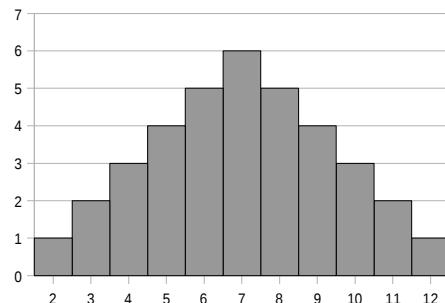
Pour un seul dé, la somme correspond tout simplement à la valeur du dé. Nous avons une possibilité pour chaque valeur :

un dé :

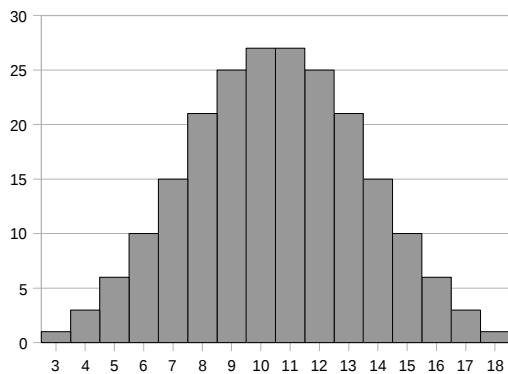


Pour deux dés, il y a une seule possibilité pour que la somme fasse deux : 1 pour le premier dé et 1 pour le deuxième dé. Pour que la somme fasse trois, il y a deux possibilités : 1 puis 2, ou, 2 puis 1. Le plus probable avec deux dés est d'obtenir 7 : (1,6) (6,1) (2,5) (5,2) (3,4) (4,3).

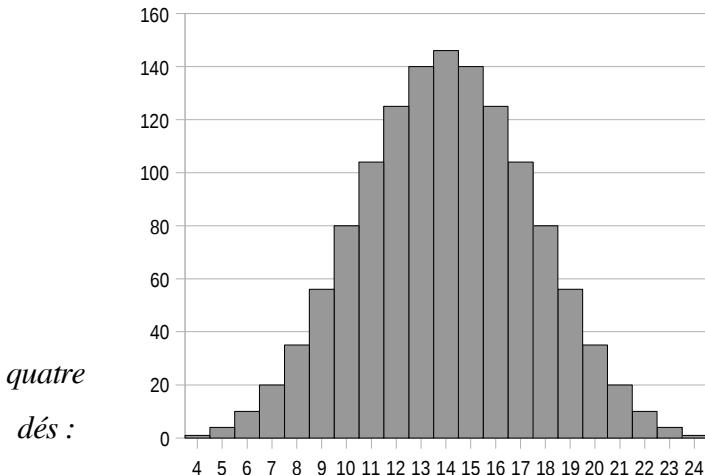
deux dés :



trois dés :



Pour quatre dés nous reconnaissions déjà bien la courbe en cloche et le profil est clairement de type gaussien :



Sur ce dernier exemple vérifions la validité du théorème central limite.

La moyenne pour la population est :

$$\mu_x = (1+2+3+4+5+6)/6 = 3,5$$

La moyenne de la distribution d'échantillonnage est bien la même : $\mu_{\bar{x}} = 14/6 = 3,5$.

L'écart-type de la population :

$$\sigma_x = \sqrt{\sum p_i \cdot (x_i - \mu)^2} \quad \text{d'où}$$

$$\sigma_x = \sqrt{1/6 \cdot [(1-3,5)^2 + (2-3,5)^2 + (3-3,5)^2 + (4-3,5)^2 + (5-3,5)^2 + (6-3,5)^2]}$$

et $\sigma_x \approx 1,71$

d'où pour quatre dés $\sigma_x = \sigma_x / \sqrt{n} = \sigma_x / 2 \approx 0,85$. Or, sur la courbe du dessus, à 40% du maximum (explication page 21), nous avons un écart d'environ 3,5 (entre 3 et 4), soit en moyenne $3,5/4 \approx 0,88$. Il y a bien correspondance.

F. Distribution de Gauss

1) Définition d'une distribution continue

Il existe des grandeurs fondamentalement continues. Le temps en est une. Le temps s'écoule continûment tel un fluide, il ne saute pas d'une valeur à l'autre. S'il semble parfois égrainer les secondes, comme les perles d'un chapelet, cela n'est dû qu'à l'appareil de mesure insuffisamment précis. Par contre quand nous lançons un dé six faces, impossible de dire "j'ai fait 2,35 !", c'est une valeur interdite, seules les valeurs entières de un à six sont permises.

Ainsi, certaines distributions de probabilité sont discrètes et d'autres continues. Pour le lancer de dés la proba-

bilité est discrète : $p_1 = \dots = p_6 = 1/6$ et $\sum_{i=1}^{n=6} p_i = 1$. Le résultat

du lancer n'a que six valeurs possibles.

Par contre si nous nous intéressons à la taille des habitants d'une ville, celles-ci sont toutes possibles, la distribution est continue. Nous parlons alors de densité de probabilité $p(x)$ avec $p(x)dx$ la probabilité d'un événement d'être entre x et $x+dx$. dx est une petite variation, et $x+dx$ est infiniment voisin de x .

La probabilité que l'évènement se réalise sur l'ensemble des valeurs possibles est de 100% :

$$\int_{x=-\infty}^{+\infty} p(x) dx = 1 \quad ^5$$

Calcul de la moyenne et de l'écart-type d'une distribution continue :

$$\mu = \int_{-\infty}^{+\infty} x \cdot p(x) dx \quad \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 p(x) dx \quad ^6$$

2) Courbe de Gauss

Le théorème central limite s'intéresse au cas où la taille de l'échantillon n est grande, et dans le cas limite où n tend vers l'infini nous considérons une distribution continue. Celle-ci est une gaussienne, l'expression mathématique est connue :

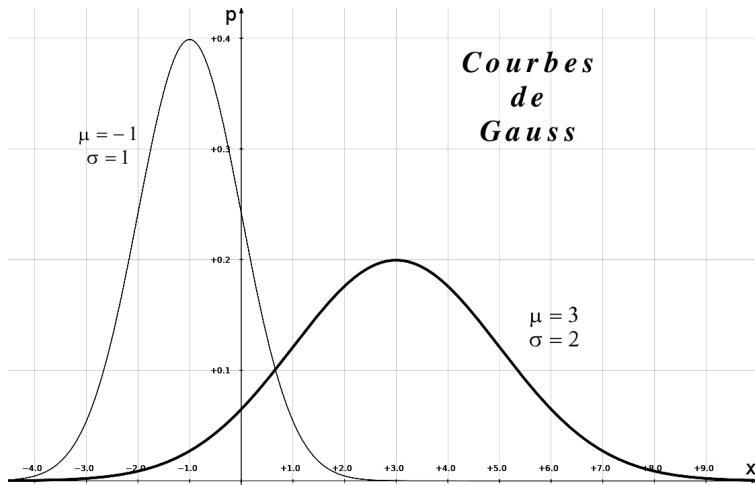
$$p(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2}$$

Dans le complément mathématique page 230 différents termes sont justifiés.

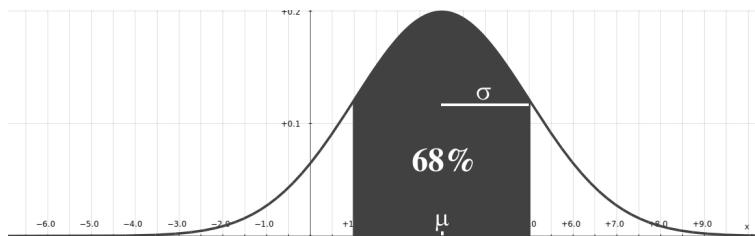
5 MATH : se lit «l'intégrale de p de x , x allant de moins l'infini à $+\infty$, est égale à 1 ».

6 MATH : La moyenne est aussi appelée $E(X)$, espérance de X . $\sigma^2 = V(X)$ est appelée variance de la variable aléatoire X . Propriétés : $E(aX+b) = aE(X)+b$, $V(aX) = a^2V(X)$ et $V(X) = E(X^2) - E(X)^2$.

Nous avons représenté, sur le graphe suivant, deux cas :



L'aire totale sous la courbe vaut toujours 1. Mais, si maintenant nous ne nous écartons que d'un écart-type par rapport à la moyenne, nous n'avons que 68% des possibilités :

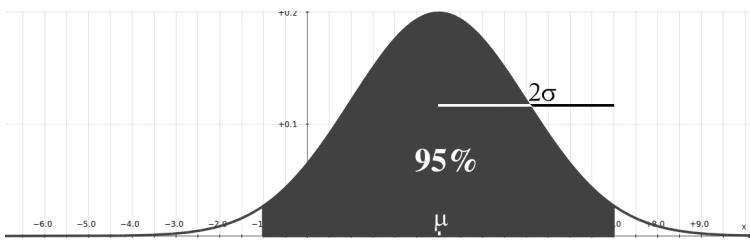


$$\int_{\mu - \sigma}^{\mu + \sigma} p(x) dx = 0,683\dots \simeq 68\%$$

L'écart-type peut s'évaluer ici, en effet à $60\%.p_{max}$:

$$p(\mu \pm \sigma) / p_{max} = 1/\sqrt{e} \simeq 0,607$$

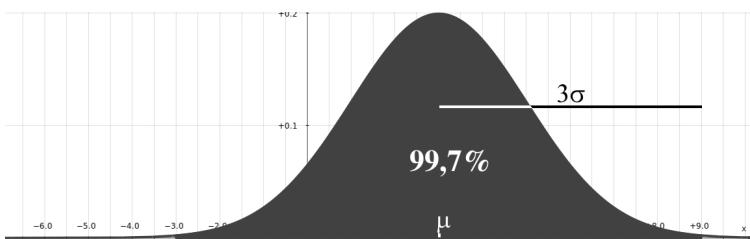
Élargissons à deux écart-types :



$$\int_{\mu - 2\sigma}^{\mu + 2\sigma} p(x) dx = 0,954 \dots \simeq 95\%$$

On dit 95% de confiance, ou, une confiance à deux sigmas.
On travaille souvent avec cette confiance.

Puis pour trois sigmas :



$$\int_{\mu - 3\sigma}^{\mu + 3\sigma} p(x) dx = 0,997 \dots > 99\%$$

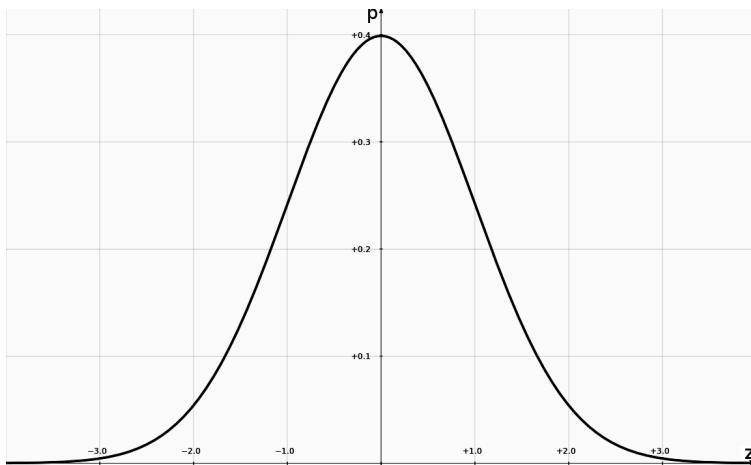
3) Loi normale standard

C'est la distribution normale de Gauss centrée et réduite. Pour le recentrage nous soustrayons la moyenne : $x' = x - \mu$. Pour la réduction nous divisons par l'écart-type :

$$z = \frac{x - \mu}{\sigma}$$

d'où : $p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

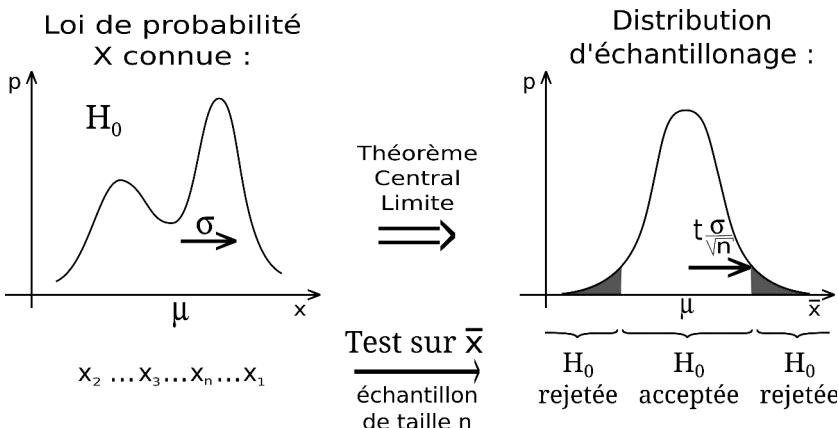
Nous avons alors une distribution normale de moyenne nulle et d'écart-type égale à un :



G. Test d'hypothèse

Nous allons à nouveau utiliser le théorème central limite. Nous l'avions précédemment utilisé pour estimer la moyenne d'une loi de probabilité inconnue. A l'aide d'un échantillon prélevé nous utilisons les propriétés de la distribution d'échantillonnage pour déterminer une grandeur et son incertitude. Ici, pour les tests d'hypothèses nous procérons dans l'autre sens, la loi de probabilité est supposée connue, ainsi la distribution d'échantillonnage est parfaitement déterminée et en prélevant un échantillon nous l'utilisons pour définir un critère de décision permettant d'accepter ou de rejeter l'hypothèse.

En utilisant la moyenne nous testons une hypothèse H_0 faite sur une loi de probabilité X . Soient μ et σ sa moyenne et son écart-type. Nous prélevons au sein de la population un échantillon de taille n suffisamment grand. La valeur moyenne sur l'échantillon est égale à \bar{x} . Si \bar{x} est comprise entre $\mu - t_\alpha \cdot \sigma / \sqrt{n}$ et $\mu + t_\alpha \cdot \sigma / \sqrt{n}$ alors l'hypothèse H_0 est acceptée, sinon l'hypothèse est rejetée (test bilatéral).



Nous considérons la valeur t_∞ de la loi normale adaptée à la confiance p désirée.

Nous pouvons utiliser aussi d'autres intervalles caractéristiques de la distribution d'échantillonnage. De manière générale, l'hypothèse H_0 implique une propriété A de la distribution d'échantillonnage. Ici, l'implication n'est pas déterministe mais statistique et la prise de décision procède différemment.

Cas d'un test déterministe : $H_0 \Rightarrow A$

- Si nous observons A alors H_0 ne peut être rejetée.
- Si nous n'observons pas A alors H_0 est rejetée⁷.

Cas d'un test statistique :

Dans $p\%$ des cas : $H_0 \Rightarrow A$

Dans $(1-p)\%$ des cas : $H_0 \not\Rightarrow A$

- Si nous observons A alors H_0 ne peut être rejetée avec $p\%$ de confiance.
- Si nous n'observons pas A alors H_0 est rejetée avec $p\%$ de confiance.

Ce protocole permet de prendre une décision, mais il comporte en même temps des risques de se tromper. Nous pouvons rejeter l'hypothèse alors quelle est vraie ou nous pouvons accepter l'hypothèse alors qu'elle est fausse.

⁷ Proposition contraposée : si $P \Rightarrow Q$ alors $\bar{Q} \Rightarrow \bar{P}$.

	H_0 est acceptée	H_0 est rejetée
H_0 est vraie	Bonne décision	Mauvaise décision (risque de première espèce α)
H_0 est fausse	Mauvaise décision (risque de deuxième espèce β)	Bonne décision

Nous devons chercher à minimiser les risques liés aux probabilités α et β . Si nous acceptons H_0 , alors α (la probabilité de rejeter l'hypothèse alors quelle est vraie) doit être importante, et β (la probabilité de d'accepter l'hypothèse alors quelle est fausse) doit être faible. Si au contraire nous rejetons H_0 c'est que α est faible et β importante.

Dans le cas où H_0 est vraie nous pouvons faire le calcul du risque, et α vaut tout simplement $1-p$. Par contre si H_0 est fausse, nous ne pouvons faire le calcul de β , à moins de disposer d'une hypothèse alternative H_1 connue.

Pour un test classique on fixe $1-p$ à l'avance, par exemple nous pourrons considérer le test statistiquement discernable au seuil de 5%, et suivant le résultat trouvé nous rejetons ou acceptons l'hypothèse. Nous pouvons aussi calculer les probabilités critiques α et β qui correspondent à la valeur \bar{x} trouvée avec notre échantillon. Nous mesurerons ainsi la crédibilité de H_0 et nous ferons ensuite le choix d'accepter ou pas notre hypothèse.

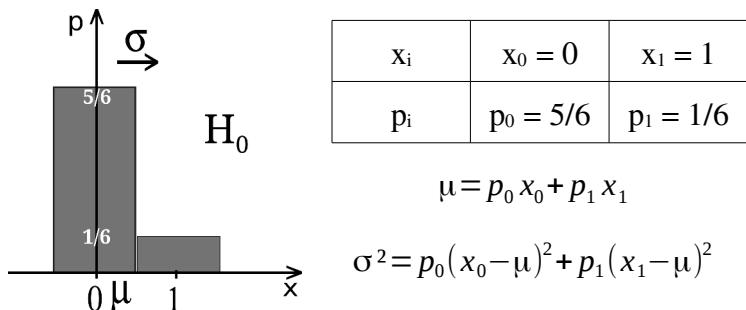
Par exemple imaginons que nous disposions de plusieurs dés. Tous les dés sont équilibrés sauf un, conçu pour

tomber sur le six avec une probabilité double. Malheureusement le dé pipé s'est mélangé aux autres et il n'a pas de signe distinctif. Nous en choisissons un à emmener pour une soirée jeux et nous voulons tout d'abord nous assurer qu'il est bien équilibré.

Nous effectuons 92 lancers :

3151353256243365313441354244652632465436616546
2241154636433165514444241456414316555146362534

Pour l'hypothèse H_0 nous définissons la variable aléatoire discrète X qui vaut zéro lorsqu'on ne fait pas un six et qui vaut un sinon avec $P(X=0)=5/6$ et $P(X=1)=1/6$.

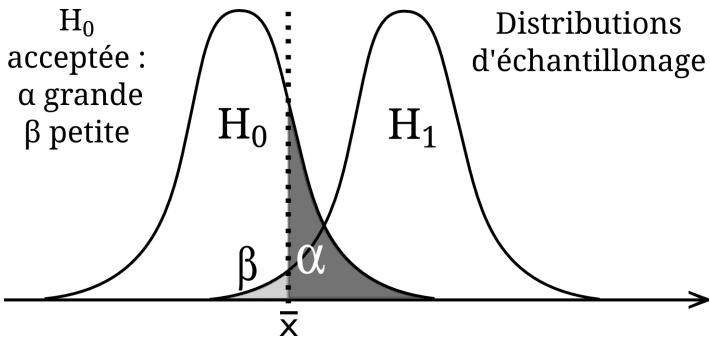


Cette loi a pour moyenne $\mu=1/6 \approx 0,167$ et pour écart-type $\sigma=\sqrt{5}/6 \approx 0,373$.

Dans notre échantillon il y a seize 6 d'où la moyenne $\bar{x}=16/92 \approx 0,174$ ainsi $\bar{x}-\mu=t_\infty \cdot \sigma/\sqrt{n}$ et nous avons un élargissement : $t_\infty=(\bar{x}-\mu)/\sigma \approx 0,187$.

L'aire sous la courbe de Gauss pour des valeurs supérieures à 0,187 vaut alors 0,43 et $\alpha \approx 43\%$ ⁸.

8 Ici σ est connu et $n=92$ est suffisamment grand utiliser le théorème central limite dans la limite des grands nombres. Pour le calcul sur LibreOffice : 1- LOI.NORMALE.STANDARD(0,187).



Si H_0 est fausse alors H_1 est vraie. Si l'hypothèse alternative H_1 était que le dé n'était pas équilibré, sans spécifier dans quel sens, le test aurait été bilatéral et nous aurions pris pour l'aire α celle des deux queues de distribution à gauche et à droite. Ici le test est unilatéral, si H_0 était fausse la probabilité d'observer six serait doublée et nous observerions des valeurs supérieures.

Pour H_1 nous définissons la variable aléatoire Y avec $P(Y=0)=2/3$ et $P(Y=1)=1/3$. Cette loi a pour moyenne $\mu'=1/3 \approx 0,333$ et pour écart-type $\sigma'=\sqrt{2}/3 \approx 0,471$. D'où $\bar{x}-\mu'=t_\infty \cdot \sigma'/\sqrt{n}$ et $t_\infty'=(\mu'-\bar{x})\sqrt{n}/\sigma' \approx 3,24$.

L'aire de la queue inférieure vaut alors $\beta \approx 0,06\%$.

Nous pouvons donc très confortablement accepter l'hypothèse que le dé choisi est équilibré, dans le cas de rejet on aurait 43% de chance de se tromper (on essaye en priorité de minimiser cette erreur, classiquement ce n'est qu'en dessous du seuil de 5% que l'on commence à remettre en cause l'hypothèse de départ). Au regard de l'hypothèse alternative il y a moins d'une chance sur 1000 qu'on l'on ai considéré le dé équilibré alors qu'il est pipé (on parle aussi

de puissance du test : $\eta=1-\beta$).

Il est à noter que nous ne calculons jamais la probabilité qu'une hypothèse soit vraie mais la probabilité de la rejeter alors quelle est vraie (risque de première espèce).

Dans le cadre juridique on fait une erreur de première espèce si une personne innocente est condamnée et de deuxième espèce si une personne coupable est relaxée. On demande au jury que la culpabilité soit prouvée au-delà d'un doute raisonnable, si la personne est condamnée α doit être suffisamment petit ^[vi]. Nous cherchons à minimiser la probabilité de le condamner alors qu'il est innocent. Nous ne considérons pas directement la probabilité qu'il soit coupable, la personne jugée est présumée innocente (H_0).

Les exercices p39 traitent différentes configurations possibles pour ce test.

H. Test du Khi-deux

Il s'agit d'un test d'hypothèse simple basé sur les différences entre les effectifs observés et ceux espérés :

Effectifs observés :

O_1	O_2	...	O_j	...	O_c
-------	-------	-----	-------	-----	-------

Effectifs espérés :

E_1	E_2	...	E_j	...	E_c
-------	-------	-----	-------	-----	-------

Nous calculons la sommes des écarts au carré par rapport à la valeur théorique attendue :

$$\chi^2 = \sum_{j=1}^c \frac{(O_j - E_j)^2}{E_j} \quad (\text{nommé "khi-deux"})$$

Nous avons ensuite une table (page 310) qui permet d'estimer la probabilité que l'hypothèse H_0 soit rejetée alors quelle est vraie. Suivant la valeur de χ^2 et le nombre de degrés de liberté nous déterminons si l'hypothèse est acceptable. Les degrés de liberté se calculent ainsi :

$$ddl = c - 1 \quad (\text{nombre de valeurs moins une unité})$$

Illustrons avec les expériences réalisées par le botaniste Mendel. Il effectue des croisements entre des plants, ici des hybrides. Il croise entre eux des pois à fleurs roses. Sa théorie lui indique qu'il doit obtenir 25% de pois à fleurs rouges, 25% de pois à fleurs blanches et 50% de pois à fleurs roses. Ceci résulte de la rencontre aléatoire des gamètes. Imaginons qu'il observe sur mille fleurs les valeurs suivantes : 27% de blanches, 24% de rouges et 49% de roses. Doit-il continuer à croire en son hypothèse ?

Effectifs observés :

270	240	490
-----	-----	-----

Effectifs espérés :

250	250	500
-----	-----	-----

d'où

$$\chi^2 = \frac{(270-250)^2}{250} + \frac{(240-250)^2}{250} + \frac{(490-500)^2}{500} \approx 2,2 \text{ et}$$

$$ddl = 3 - 1 = 2 .$$

Donc, d'après la table, il y a plus de 30% de chance que l'hypothèse soit rejetée alors qu'elle est juste.

Nous décidons alors d'accepter l'hypothèse. En général on prend une probabilité critique α de 5%, en dessous de laquelle on envisage de rejeter l'hypothèse.

Le test se généralise facilement pour un tableau :

Effectifs observés :

$$\begin{pmatrix} O_{11} & O_{12} & \dots & O_{1j} & \dots & O_{1c} \\ O_{21} & O_{22} & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ O_{i1} & \dots & \dots & O_{ij} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ O_{l1} & \dots & \dots & \dots & \dots & O_{lc} \end{pmatrix}$$

Effectifs espérés :

$$\begin{pmatrix} E_{11} & E_{12} & \dots & E_{1j} & \dots & E_{1c} \\ E_{21} & E_{22} & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ E_{i1} & \dots & \dots & E_{ij} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ E_{l1} & \dots & \dots & \dots & \dots & E_{lc} \end{pmatrix}$$

Les ddl tiennent compte du nombre de colonnes c et de lignes l : $ddl = (c-1)(l-1)$

Le χ^2 se calcul avec le même type de formule :

$$\chi^2 = \sum_{(i,j)} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

De plus nous utilisons la même table pour déterminer la validité de l'hypothèse.

I. Sources des incertitudes

Une variable aléatoire a une incertitude d'autant plus faible que la mesure est fidèle, juste et que le système d'acquisition a une bonne résolution.

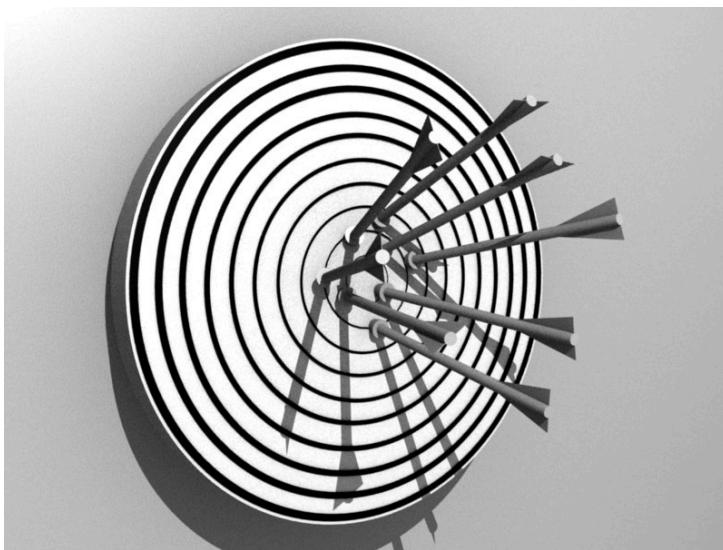
La justesse est assurée par l'absence d'erreurs systématiques. Il peut exister un biais qui rend la mesure inexacte (même si la dispersion est faible). Erreurs de lecture, absence de contrôle et de corrections de facteurs influents, incertitude due à la modélisation, etc. Tous les biais doivent être identifiés et estimés afin d'être ajoutés à la dispersion, le système devient alors juste.

La fidélité provient de la répétabilité et de la reproductibilité des mesures. Les valeurs d'un système fidèle sont peu dispersées. La dispersion peut provenir d'erreurs accidentelles ou d'un phénomène physique par essence aléatoire (comme par exemple la radioactivité). Les expérimentateurs par un travail propre, conscientieux et selon un protocole bien défini et réfléchi, pourront minimiser la dispersion. Les sources peuvent être innombrables, nous essaierons d'en identifier un maximum afin de les évaluer.

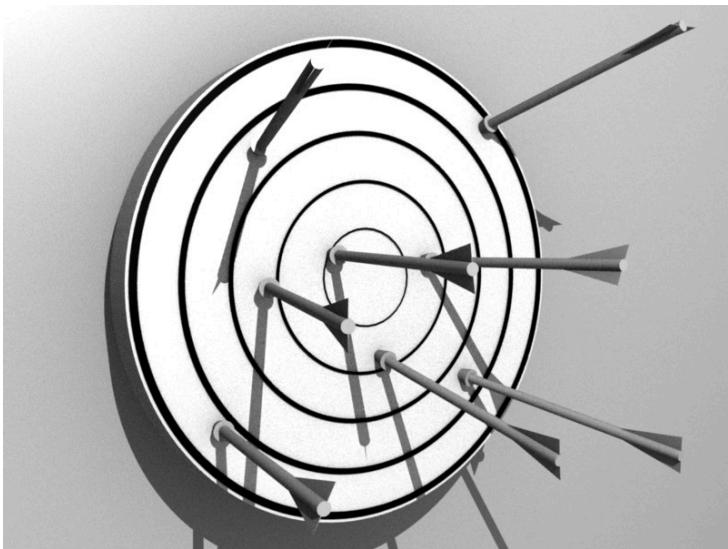
La résolution de l'instrument de mesure dépend de la taille des graduations, du type du vernier ou du nombre de digits de l'affichage. A l'incertitude due à la discrétisation des mesures peuvent s'ajouter d'autres facteurs. Il faudra se référer à la notice ou contacter le fabricant pour connaître au mieux la précision de votre appareil. On peut aussi effectuer un étalonnage avec un instrument de haute précision qui sert de référence.

L'influence de ces différentes sources d'incertitude peut être illustrée par une cible et des flèches. Le centre de la cible correspond à la grandeur à mesurer et les flèches représentent les différentes mesures. Si les flèches, dans leur ensemble, ne sont pas correctement centrées, la justesse n'est pas assurée. Le resserrement des flèches représente la fidélité. La distance entre les cercles sur la cible indique la résolution. La valeur notée est celle du cercle dont la flèche est le plus proche. L'expérimentateur voit les flèches et les cercles, par contre il ne sait pas où est la cible et son centre. Il tient l'arc et son désir d'être au plus proche du centre de la cible montre la qualité et la rigueur de son travail.

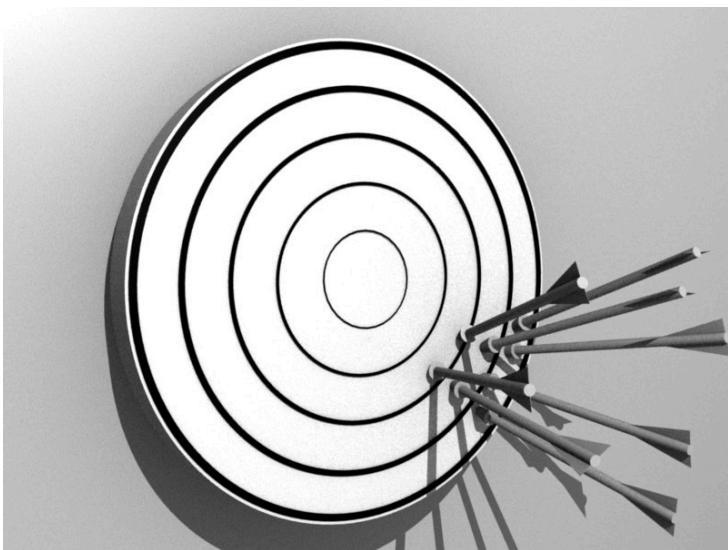
Mesure juste, fidèle et avec une bonne résolution :



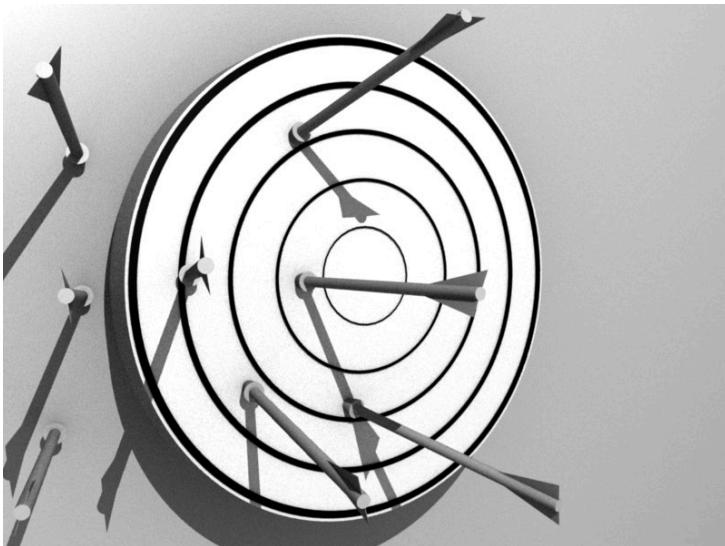
Mesure juste, mais peu fidèle et avec une faible résolution :



Mesure peu dispersée mais avec un biais, et mal résolue :



*Mesure avec un biais, fortement dispersée
et une faible résolution :*



L'écart-type complet sera déterminé à partir des écarts de chaque source en ajoutant les carrés (dus aux compensations explicitées au chapitre 2) :

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \dots}$$

J. Exercices

Exercice 1 : Âges corrigé p233

Soient les âges des étudiants d'une classe {18; 20; 18; 19; 18; 18; 18; 17; 18; 19; 17; 19; 17; 21; 18}. Déterminez le mode, la médiane, la moyenne arithmétique, la moyenne géométrique, l'étendue, l'écart-type, l'écart quadratique moyen et l'écart moyen⁹.

Exercice 2 : Cartes corrigé p233

Soit un jeu de 32 cartes. Nous tirons au hasard cinq cartes. Déterminez la probabilité d'avoir un carré d'as, puis celle d'avoir une couleur.

Exercice 3 : Champ de pesanteur corrigé p233

Des étudiants mesurent l'intensité g du champ de pesanteur terrestre au laboratoire. Les huit binômes mesurent les valeurs suivantes en m/s² : 6,20 ; 8,35 ; 13,00 ; 8,37 ; 8,54 ; 9,67 ; 9,75 ; 10,66.

- Quel commentaire général feriez-vous en voyant ces résultats.
- Calculez la moyenne et l'écart-type.
- La dispersion des valeurs donne quelle incertitude sur la moyenne (confiance de 95%) ? Le résultat est-il cohérent avec la valeur attendue ?
- Un neuvième binôme fait une nouvelle mesure dans les mêmes conditions expérimentales. Évaluez la probabilité qu'ils obtiennent un résultat entre 8 et 12 m/s².

9 $\text{Écart moyen} = \left(\sum |x_i - \bar{x}| \right) / n = \left(\sum \sqrt{|x_i - \bar{x}|^2} \right) / n$

Exercice 4 : Ascenseur corrigé p234

La charge maximale d'un ascenseur est de 300 kg, la masse totale des occupants est de 280 ± 10 kg à σ . Quelle est la probabilité d'être en surcharge ?

Exercice 5 : Devoir corrigé p234

Pour un devoir les notes obtenues par des étudiants sont les suivantes (notation sur 20) :

10	5	13	7	6	9	5	5	10	15	5	3
15	12	11	1	3	13	11	10	2	7	2	8
2	15	4	11	11	5	8	12	10	18	6	

- Déterminez la moyenne et l'écart-type.
- Réalisez un diagramme avec les notes en abscisse et les fréquences en ordonnée.
- Faire un autre diagramme des fréquences en regroupant les données en classes : [0, 1, 2], [3, 4, 5] ... , [18, 19, 20]. On indiquera aussi les fréquences relatives. Comparez la lisibilité avec le premier diagramme.

Exercice 6 : Yams corrigé p235

Le Yams se joue avec cinq dés équilibrés à six faces.

- On lance les cinq dés, quelle est la probabilité de faire un Yams directement (mêmes valeurs sur les cinq dés).
- Pour un lancer quelle est la probabilité d'avoir un total inférieur à dix ?

3) Nous effectuons une série de lancers et nous obtenons les totaux suivants : 18, 15, 17, 22, 16, 12, 14, 22, 23, 14, 23, 14, 18, 21, 12, 15, 18, 13, 15, 18, 17, 15, 17, 21, 25, 16, 8, 15, 15, 13.

- a) Calculez la moyenne et l'écart-type.
- b) Quelle est l'estimation de la moyenne avec un intervalle de confiance de 95% ? Est-ce que cela correspond à la valeur théorique ?
- c) Réalisez un graphe des valeurs avec leurs fréquences correspondantes.
- d) Si nous effectuons un nouveau lancer, quelle est, selon vous, la probabilité d'avoir un résultat supérieur à 24 ?

Exercice 7 : Élastiques corrigé p236

Un fabricant d'élastiques indique, que sur mille élastiques vendus, dix en moyenne ne sont pas fonctionnels. Un acheteur veut tester les lots livrés avant d'accepter la livraison. Il décide de refuser la livraison si le nombre d'élastiques endommagés est trop élevé et il veut avoir moins de 1% de chance de se tromper en réfutant l'indication du fabricant. L'acheteur prélève n élastiques au hasard, à partir de quel nombre n_o d'élastiques endommagés doit-il refuser la livraison ? Indiquer ce nombre pour trois cas différents : un échantillon de 1000, 200 puis 50 élastiques.

Exercice 8 : Test d'un isolant corrigé p237

Un fabricant indique pour un panneau isolant en cellulose une conductivité thermique de $0,039 \text{ W/m.K}$. La valeur est certifiée à 5% près. Vous voulez vérifier si c'est vrai, pour cela vous prenez dix panneaux au hasard et vous mesurez leurs conductivités respectives ($\text{mW.m}^{-1}\text{.K}^{-1}$) :

39,1	38,8	39,5	39,2	38,9	39,1	39,2	41,1	38,6	39,3
------	------	------	------	------	------	------	------	------	------

Les valeurs sont-elles conformes à celles annoncées par le fabricant (il considère une confiance de 95% sur la marge donnée) ? Pourrait-il, selon vos résultats, annoncer une autre marge ?

Exercice 9 : Pièces corrigé p237

Nous effectuons un grand nombre de lancers d'une pièce, nous cherchons à déterminer si les probabilités d'obtenir un pile ou un face sont égales. Nous effectuons l'expérience avec trois pièces différentes, sont-elles équilibrées ? (réponses avec une confiance de 95%)

- 1) 42 piles et 58 faces.
- 2) 510 piles et 490 faces.
- 3) 420 piles et 580 faces.

Exercice 10 : Parité corrigé p238

La parité hommes-femmes est-elle respectée dans les deux chambres et le conseil constitutionnel ?

	Hommes	Femmes
Assemblée nationale	470	107
Sénat	272	76
Conseil constitutionnel	10	2

Exercice 11 : Naissances corrigé p238

Testons l'hypothèse suivante : les naissances en Suède se répartissent uniformément tout au long de l'année. Supposons que nous avons un échantillon aléatoire de 88 naissances. Les résultats sont regroupés selon des saisons de longueur variable : 27 naissances au printemps (avril-juin), 20 en été (juillet/août), 8 en automne (septembre/octobre) et 33 en hiver (novembre-mars).

Au seuil de 5% l'hypothèse peut-elle être rejetée ?

Sur le même exemple on collecte maintenant un très grand échantillon : 23385, 14978, 14106 et 35804.

Quelle est alors la conclusion ?

Théorie

Exercice 12 : Gaussiennes dans le plan et l'espace corrigé p238

Gaussienne à une dimension :

Soit la loi de probabilité de Gauss centrée et réduite $p(x)$.

1- Calculez et comparez la moyenne de x et de $|x|$. Laquelle correspond à la distance moyenne au point d'origine ? Que dire de σ_x et de $\sigma_{|x|}$?

2- Calculez numériquement $P(|x| \leq 1)$, $P(|x| \leq 2)$ et $P(|x| \leq 3)$.

Gaussienne bidimensionnelle :

Soit la loi de probabilité qui généralise une gaussienne à deux dimensions suivante : $p(x, y) = p(x)p(y)$ avec $p(x)$ et $p(y)$ des lois normales centrées et réduites.

Aides :

Soit des intégrales multiples avec des bornes indépendantes des variables et des fonctions continues, les variables sont alors séparables :

$$\iint f(x, y) dx dy = \int f(x) dx \int f(y) dy$$

Passage en coordonnées polaires :

$$\rho^2 = x^2 + y^2 \text{ et } dx dy = 2\pi \rho \, d\rho \text{ (symétrie par rotation)}$$

1- Quelle est l'expression de $p(x, y)$? Montrez que

$p(x, y)$ vérifie les deux conditions nécessaires pour une loi de probabilité.

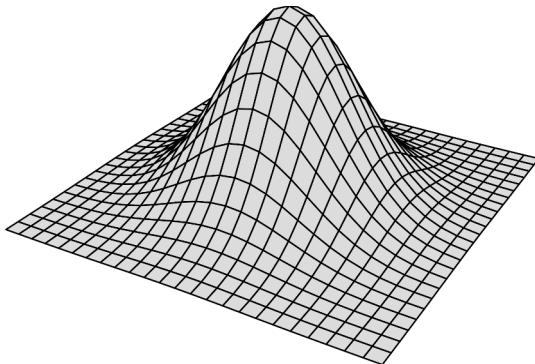
2- En introduisant les coordonnées polaires vérifiez que la probabilité sur tout le plan vaut bien l'unité. Vous exprimerez $p(\rho)$ définie tel que :

$$\iint p(x, y) dx dy = \int p(\rho) d\rho .$$

$p(\rho)$ est la densité de probabilité par rapport à ρ .
 $p(\rho) d\rho$ correspond à la probabilité d'un événement d'être entre ρ et $\rho + d\rho$.

Que valent la moyenne de la distance au point d'origine $\bar{\rho}$ et l'écart-type σ_ρ pour cette loi de probabilité ?

3- Calculez $P(\rho \leq \sigma_\rho)$, $P(\rho \leq 2\sigma_\rho)$ et $P(\rho \leq 3\sigma_\rho)$.



Gaussienne tridimensionnelle :

Soit la loi de probabilité qui généralise une gaussienne à trois dimensions suivante :

$p(x, y, z) = p(x)p(y)p(z)$ avec $p(x)$, $p(y)$ et $p(z)$ des lois normales centrées et réduites.

Aides :

Nous avons avec les intégrales triples les mêmes conditions de séparation des variables qu'avec les intégrales doubles.

Passage en coordonnées sphériques :

$$r^2 = x^2 + y^2 + z^2 \quad \text{et} \quad dx dy dz = 4\pi r^2 dr$$

(symétrie sphérique)

1- Quelle est l'expression de $p(x, y, z)$? Exprimez $p(r)$ définie tel que : $\iiint p(x, y, z) dx dy dz = \int p(r) dr$.

$p(r)$ est la densité de probabilité par rapport à r .

$p(r) dr$ correspond à la probabilité d'un événement d'être entre r et $r+dr$.

2- Avec cette nouvelle expression vérifiez que la probabilité sur tout l'espace vaut bien l'unité.

3- Que valent la moyenne \bar{r} et l'écart-type σ_r pour cette loi de probabilité ?

4- Calculez $P(R \leq \sigma_r)$, $P(R \leq 2\sigma_r)$ et $P(R \leq 3\sigma_r)$.

5- Comparez les trois cas d'une gaussienne 1D, 2D ou 3D. Vous devez retrouver les résultats du tableau qui suit.

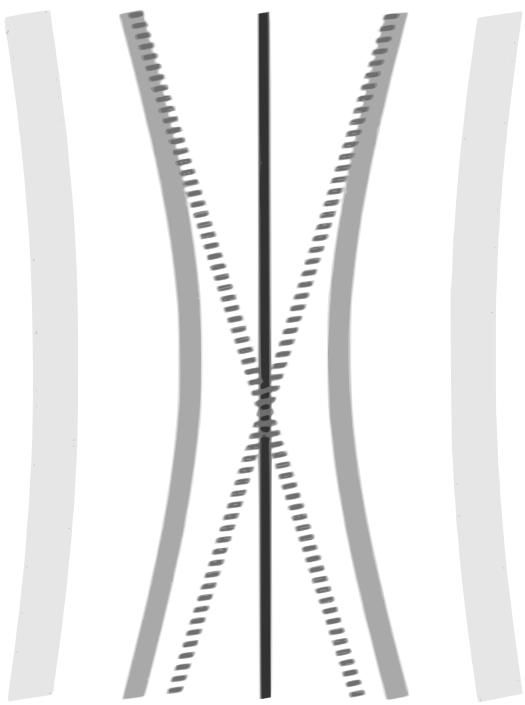
Gaussienne	1D	2D	3D
Distance à l'origine	$ x ^{10}$	ρ	r
Moyenne	$\sqrt{\frac{2}{\pi}}$	$\sqrt{\frac{\pi}{2}}$	$2\sqrt{\frac{2}{\pi}}$
Écart-types	1	$\sqrt{2}$	$\sqrt{3}$
P à σ	68,3%	1-1/e=63,2%	60,8%
P à 2σ	95,4%	98,2%	99,3%
P à 3σ	99,7%	99,988%	99,9994%

Pour vérifier vos calculs sur un tableur vous pouvez utiliser les fonctions suivantes :

sur OpenOffice :

- somme d'une zone sélectionnée * (ex.: B43:B53) :
=SOMME(*)
- valeur fixée à la cellule B3 : \$B\$3
- =MOYENNE(*)
- carré : *^2
- racine carrée : *^(1/2)
- =ECARTYPE(*)
- coefficient de Student pour une confiance de 95% et n=20 : =LOI.STUDENT.INVERSE(0,05;19)
- probabilité que l'hypothèse soit vraie, tableau des valeurs observées * (ex.: B71:E72), celui des valeurs estimées ** : =TEST.KHIDEUX(*,**).

10 MATH : Nous pourrions nous étonner de la différence de l'expression en 1D avec la valeur absolue, ce n'est qu'une question de définitions en coordonnées cylindriques et sphériques. Par exemple, $\rho \in [0;+\infty]$ et $\theta \in [0;2\pi]$, mais nous pourrions aussi prendre $\rho \in [-\infty;+\infty]$ et $\theta \in [0;\pi]$, la moyenne de ρ serait alors nulle et nous aurions considéré sa valeur absolue.



II. CORRÉLATIONS ET INDÉPENDANCES

Au chapitre précédent nous n'avions qu'une grandeur aléatoire X avec ses n réalisations $\{x_i\}$. Maintenant nous avons plusieurs grandeurs et un nouvel indice permet de les distinguer : X_j et ses mesures $\{x_{jk}\}$. X_j est la $j^{\text{ème}}$ grandeur et x_{jk} est la $k^{\text{ème}}$ observation de cette grandeur. Nous allons nous intéresser aux interactions entre différentes grandeurs.

Pour illustrer, considérons un échantillon de quatre individus qui possèdent trois caractéristiques, la taille X_1 , le poids X_2 et le mois de naissance X_3 . A priori, nous nous attendons à une corrélation entre la taille et le poids : plus on est grand plus on a, en général, une masse importante (corrélation positive). Par contre, nous pouvons penser que le mois de naissance n'a aucune incidence sur le poids et la taille (X_3 non corrélée avec X_1 et X_2).

A. Coefficient de corrélation

Le coefficient de corrélation r permet d'identifier s'il y a une relation linéaire entre deux variables X_i et X_j :

$$r_{ij} = \frac{\sum_k [(x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)]}{\sqrt{\sum_k [(x_{ik} - \bar{x}_i)^2]} \cdot \sqrt{\sum_k [(x_{jk} - \bar{x}_j)^2]}}$$

r varie entre -1 et +1. Si $|r|=1$ les variables sont parfaitement corrélées : $r=1$ même sens de variation, $r=-1$ sens opposé. Si $r=0$ il n'y a pas la moindre corrélation, les variables sont parfaitement indépendantes.

Calcul de r_{12} , r_{13} et r_{23} :

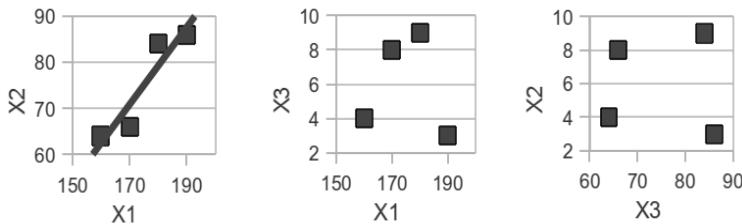
	X_1 (cm)	X_2 (kg)	X_3	$x_1 - \bar{x}_1$	$x_2 - \bar{x}_2$	$x_3 - \bar{x}_3$	$(x_1 - \bar{x}_1)^2$
1	160	64	4	-15	-11	-2	225
2	170	66	8	-5	-9	2	25
3	180	84	9	5	9	3	25
4	190	86	3	15	11	-3	225
\bar{x}	175	75	6			$\Sigma =$	500

et la suite des calculs :

$(x_2 - \bar{x}_2)^2$	$(x_3 - \bar{x}_3)^2$	$(x_1 - \bar{x}_1) \cdot (x_2 - \bar{x}_2)$	$(x_1 - \bar{x}_1) \cdot (x_3 - \bar{x}_3)$	$(x_2 - \bar{x}_2) \cdot (x_3 - \bar{x}_3)$
121	4	165	30	22
81	4	45	-10	-18
81	9	45	15	27
121	9	165	-45	-33
404	26	420	-10	-2

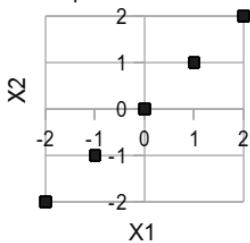
d'où : $r_{12} = \frac{420}{\sqrt{500}\sqrt{404}} \approx 0,93$, $r_{13} \approx -0,09$ et
 $r_{23} \approx -0,02$.

r_{12} est proche de +1, nous avons donc une corrélation positive importante. r_{13} et r_{23} sont proches de zéro : X_3 est indépendante de X_1 et X_2 :

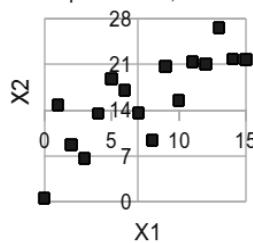


Exemples de nuages de points entre deux variables :

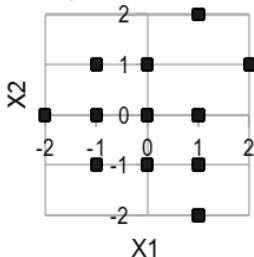
Exemple 1 : $r=1$



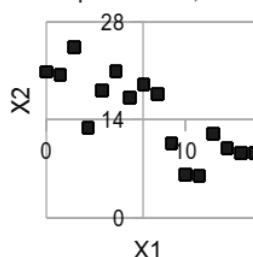
Exemple 2 : $r=0,8$



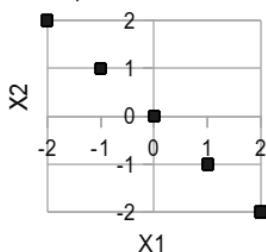
Exemple 3 : $r=0$



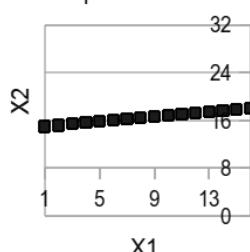
Exemple 4 : $r=-0,8$



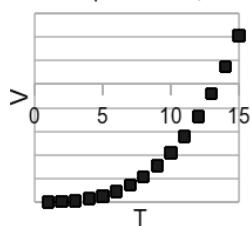
Exemple 5 : $r=-1$



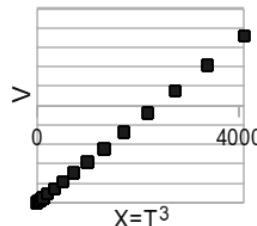
Exemple 6 : $r=1$



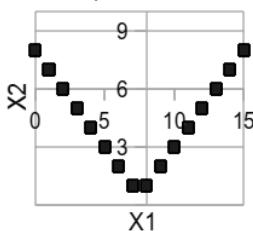
Exemple 7 : $r=0,92$



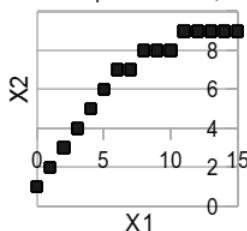
Exemple 8 : $r=1$



Exemple 9 : $r=0$



Exemple 10 : $r=0,94$



Nous voyons sur les exemples 7, 9 et 10 une forte corrélation entre les deux variables. Pourtant le coefficient de corrélation n'est pas aussi proche de -1 ou +1 que nous pourrions l'imaginer, il est même nul sur l'exemple 9. Ceci provient du fait que les corrélations ne sont pas linéaires.

Il peut y avoir des phénomènes de saturation (exemple 10) ou de seuil (un produit peut être bénéfique à faible dose et nocif pour des doses plus importantes - exemple 9). Il est préférable d'avoir identifié ces phénomènes en amont. Pour cela il est important de bien réfléchir à la pertinence des variables que l'on choisit avant de commencer une étude statistique.

Autre exemple : si nous étudions le volume V de différents objets en fonction de leur taille T , nous trouvons une corrélation positive. Mais celle-ci sera bien plus forte entre V et $X=T^3$ (graphes 7 et 8).

Le fait que deux variables soient corrélées n'implique pas nécessairement une relation de cause à effet entre les deux variables (la variation d'une variable entraîne la variation de l'autre). Les variations des deux variables peuvent être attribuables à une cause commune extérieure.

Par exemple, on peut supposer qu'il y a corrélation entre la consommation d'huile à bronzer et celle de crèmes glacées. Il n'y a évidemment aucun liens de cause à effet entre les deux, mais une cause commune d'ordre météorologique. Des raisons physiques permettent de montrer une causalité, pas une étude statistique.

B. Formule de propagation des incertitudes

Soit un seau rempli d'un million de grains de sable. La masse d'un grain est de 10 mg à 1mg près. Qu'elle est la masse de sable contenue dans le seau ?

1) Formule de propagation des écart-types

Pour une approche générale du problème, soit une fonction f qui dépend de p variables indépendantes :

$$f(x_1, x_2, \dots, x_j, \dots, x_p)$$

A chacune de ces variables aléatoires est associée une valeur moyenne \bar{x}_j et un écart-type σ_j .

Que valent f et σ_f ?

La statistique donne la réponse et démontre la formule de propagation des écart-types :

$$\sigma_f^2 = \sum_{j=1}^p \left[\left(\frac{\partial f}{\partial x_j} \right)^2 \sigma_j^2 \right]^{11}$$

11 Nous obtenons la formule de la variance en remplaçant σ^2 par V.

2) Calcul d'incertitude

Pour les incertitudes (définies page 13) nous avons aussi une formule de propagation :

$$\boxed{\Delta f^2 = \sum_{j=1}^p \left[\left(\frac{\partial f}{\partial x_j} \right)^2 \Delta x_j^2 \right]}$$

La formule de propagation des incertitudes n'est pas exacte comme pour les écart-types, mais elle est très pratique et le plus souvent très proche du résultat exact.

Pour notre seau : $M(m_1, m_2, \dots, m_j, \dots, m_p)$

avec $M = \sum_{j=1}^p m_j$

où nous appelons M la masse totale de sable dans le seau, m_j la masse de chaque grain et p le nombre de grains.

$$\Delta M^2 = \sum_{j=1}^p \left(\frac{\partial M}{\partial m_j} \right)^2 \Delta m_j^2$$

$$\begin{aligned}\frac{\partial M}{\partial m_j} &= \frac{\partial m_1}{\partial m_j} + \dots + \frac{\partial m_j}{\partial m_j} + \dots + \frac{\partial m_p}{\partial m_j} \\ \frac{\partial M}{\partial m_j} &= 0 + \dots + 1 + \dots + 0 = 1\end{aligned}$$

(les calculs de dérivées partielles sont expliqués page 229)

alors $\Delta M^2 = \left(\sum_{j=1}^p 1^2 \right) \Delta m^2$

d'où $\Delta M^2 = p \cdot \Delta m^2$ avec $\Delta m = \Delta m_j$ quelque soit j .

Finalement : $\Delta M = \sqrt{p \cdot \Delta m} = \sqrt{1000000} \cdot 0,001 g$.

Le seau pèse donc dix kilos à un gramme près. La précision sur la masse du seau est donc de 0,01%. Naïvement, nous aurions pu penser que l'incertitude globale sur la masse du seau était la somme des incertitudes de chaque grain, nous aurions alors une incertitude absolue d'un kilo et une relative de 10%, ce qui est très différent de la réalité et ignorerait les compensations.

Ici la formule de propagation est très précise car nous avons un très grand nombre de grains. Elle est même exacte, dès les petits nombres, si la distribution de la masse des grains est gaussienne¹².

¹² MATH : une combinaison linéaire de grandeurs gaussiennes est elle-même gaussienne (s'applique ici à une somme). Et dans la formule de propagation des incertitudes, si f et les x_i ont des lois de probabilités de même nature, la formule est tout aussi exacte que la formule de propagation des écart-types.

En pratique, certains cas se rencontrent souvent :

- Pour des sommes ou des différences les incertitudes absolues au carré s'ajoutent :

$$\Delta f^2 = \sum_{j=1}^p \Delta x_j^2$$

Par exemple si $d=x_2-x_1$ avec $\Delta x_2=\Delta x_1=1\text{cm}$ alors $\Delta d \simeq 1,4\text{ cm}$.

- pour des produits ou des quotients les incertitudes relatives au carré s'ajoutent :

$$\left(\frac{\Delta f}{f}\right)^2 = \sum_{j=1}^p \left(\frac{\Delta x_j}{x_j}\right)^2$$

Par exemple si $R=U/I$ avec U et I à 1% alors R est connu à 1,4%.

Dans les cas plus complexes, il faut réaliser explicitement le calcul aux dérivées partielles.

A l'aide d'un ordinateur il est aussi possible de faire un calcul numérique. On peut générer des nombres aléatoires ou utiliser des paquets décorrélés globalement gaussiens. Cette dernière méthode est illustrée dans l'exercice 2 de ce

chapitre. Un tableur peut permettre de faire automatiquement les calculs (par exemple feuille 4 du tableur IncertitudesLibres sur www.incertitudes.fr).

On peut aussi donner des méthodes qui donnent des idées générales sur l'incertitude. Elles développent une certaine intuition mais il y a un risque de se tromper.

Par exemple comme les incertitudes s'ajoutent au carré, on peut considérer que l'incertitude la plus grande va l'emporter rapidement sur les autres. Dans l'exemple où $R=U/I$ si U est connu 1% et I à 0,1% alors R est connu à $1,005\% \approx 1\%$, on peut négliger l'incertitude sur I .

Pour des sommes ou des différences, on considère parfois que le terme qui a le dernier chiffre significatif le moins précis indique la précision du dernier chiffre significatif du résultat. Mais sur notre exemple de la masse du seau rempli de grains de sable, ça ne marche pas : la masse d'un grain est $m = 10$ mg or la masse du seau M est connue au gramme et non au milligramme !

Pour les produits ou des quotients, on considère parfois que le facteur qui a le nombre de chiffres significatifs le plus faible indique le nombre de chiffres significatifs du résultat,

mais là aussi il faut être très attentif.

Une illustration très simple, si $H=2h$ avec $h=5,00m$ (h connu au *cm*), que vaut H ? D'après la règle du dessus H serait connu avec trois chiffres significatifs, donc $H=10,0m$. H ne serait connu qu'à *10cm* près, il va de soi que l'on est plus vers le *cm*...

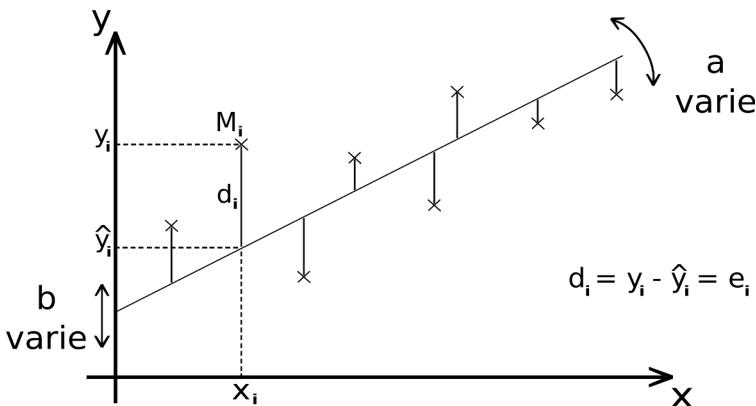
N'utilisez pas ces recettes si vous n'avez pas conscience des pièges possibles.

C. Régression linéaire

Si maintenant nous avons deux variables corrélées nous pourrions vouloir déterminer la relation la plus adaptée entre elles. Les variables aléatoires seront nommées X et Y et nous chercherons la droite qui passe au mieux par le nuage de points $y(x)$. Par exemple, quelle est la relation la mieux adaptée entre la taille X et le poids Y dans notre exemple initial : $y = ax + b$?

Quelles sont les incertitudes Δa et Δb ?

1) Principe et formules



La méthode choisie est celle des moindres carrés : la droite considérée la meilleure est celle qui minimise la somme des carrés des distances à la droite, distances prises selon y (écart).

L'ensemble des points se note $M_i(x_i, y_i)$. Pour x_i donné,

l'ordonnée estimée sur la droite s'écrit $\hat{y}_i = a x_i + b$.

D'où la somme des distances au carré :

$$\sum d^2 = \sum_i (y_i - \hat{y}_i)^2$$

Les dérivées partielles de cette quantité selon a et b s'annulent pour la meilleure droite et nous obtenons les équations suivantes : $\sum_i (y_i - a x_i - b) x_i = 0$ et

$$\sum_i (y_i - a x_i - b) = 0 .$$

Nous obtenons ainsi les résultats désirés :

$$a = \frac{\bar{x} \bar{y} - \bar{x} \bar{y}}{\bar{x}^2 - \bar{x}^2} \quad \text{et} \quad b = \bar{y} - a \bar{x} .$$

On nomme e_i le résidu tel que $y_i = \hat{y}_i + e_i$.

On trouve les différents écart-types suivants¹³:

- pour les résidus $s_r = \sqrt{\frac{\sum_i e_i^2}{n-2}}$
- pour la pente $s_a = \frac{s_r}{\sqrt{\sum_i (x_i - \bar{x})^2}}$
- pour l'ordonnée à l'origine

$$s_b = s_r \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}}$$

13 Démonstrations p97.

Puis $\Delta a = t_{n-2} s_a$ et $\Delta b = t_{n-2} s_b$.

t_{n-2} : coefficients de Student pour $n-2$ degrés de liberté.

Vous êtes ainsi à même de mener tous les calculs.

Effectuons les calculs pour le poids en fonction de la taille:

$$\bar{x}y = (160x64 + 170x66 + 180x84 + 190x86)/4$$

$$\bar{x^2} = (160^2 + 170^2 + 180^2 + 190^2)/4$$

$$a = (13230 - 175x75)/(30750 - 175^2) = 0,84 \text{ et}$$

$$b = 75 - 0,84x175 = -72$$

$$s_r = \sqrt{[(64 - (0,84x160 - 72))^2 + (-4,8)^2 + 4,8^2 + (-1,6)^2]/2} \approx 5,06$$

$$s_a \approx 5,06 / \sqrt{(160 - 175)^2 + (-5)^2 + 5^2 + 15^2} \approx 0,226$$

$$\Delta a \approx 2,92x0,226 \approx 0,66 \text{ avec une confiance de } 90\%$$

$$s_b \approx 5,06 \sqrt{(160^2 + 170^2 + 180^2 + 190^2) / [4(15^2 + 5^2 + 25 + 225)]} \approx 39,7$$

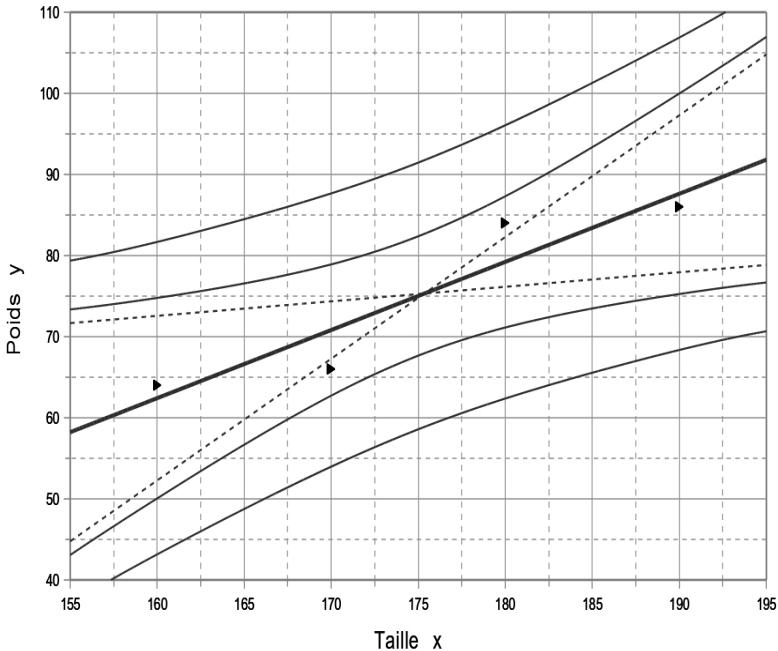
$$\Delta b \approx 2,92x39,7 \approx 116 \text{ à } 90\%$$

d'où : $Poids = (0,84 \pm 0,66) Taille - (72 \pm 116)$ à 90%.

Formule ici très imprécise, ce qui n'est pas étonnant vu le peu de points et la dispersion des données. Mais la méthode de calcul est maintenant explicite et compréhensible.

Sur le graphique qui suit nous avons :

- au milieu la droite interpolée (le meilleur équilibre entre les points du dessus et ceux du dessous de cette droite).



- En pointillés sont représentés les deux droites extrêmes ($y = a_{\min}x + b_{\max}$ et $y = a_{\max}x + b_{\min}$).
- La première enveloppe correspond aux valeurs estimées de y . Intervalle de confiance de la moyenne de y_o pour une valeur x_o :

$$\Delta y_o = t_{n-2} s_r \sqrt{\frac{1}{n} + \sum \frac{(x_i - \bar{x})^2}{(x_i - \bar{x})^2}}$$

Par exemple pour $x_o=175 \text{ cm}$ nous avons $y_o=75,0 \pm 7,4 \text{ kg}$. On peut même avoir une estimation en dehors de l'intervalle par exemple pour $x_o=195 \text{ cm}$ nous avons $y_o=92 \pm 15 \text{ kg}$.

- La deuxième enveloppe correspond à une prédition si nous effectuons une nouvelle mesure. Intervalle de prédition pour une observation y_o :

$$\Delta y_o = t_{n-2} s_r \sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{\sum (x_i - \bar{x})^2} + 1}$$

Par exemple, il y a 90% de chance pour une personne de 175 cm que sa masse soit entre 58 et 92 kg (en moyenne 90% des points sont dans cette seconde enveloppe et 10% en dehors).

2) Détermination du zéro absolu

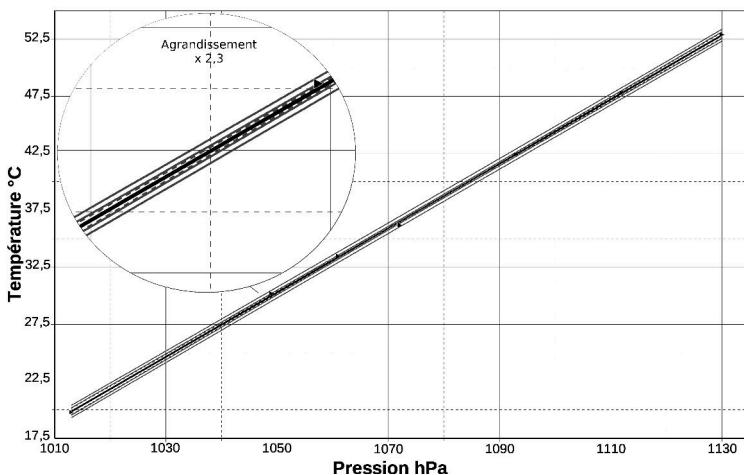
Nous étudions un gaz enfermé dans un récipient rigide de volume constant. Nous avons des sondes pour mesurer sa température et sa pression. Initialement le gaz est à température et pression ambiante, ensuite nous immergeons l'ensemble dans de l'eau chaude et nous laissons évoluer en prenant des mesures à la volée¹ :



¹ Cette expérience a réellement été réalisée le mardi 17 octobre 2006 à Bourges par M. ROUAUD et O. LEROY au Lycée Alain-Fournier.

heure	10h 15	10h 30	10h 40	10h 55	non notée	non notée	12h
température Θ ($^{\circ}$ C)	19,8	52,9	47,8	42,4	36,2	33,5	30,2
pression P (hPa)	1013	1130	1112	1093	1072	1061	1049

Nous supposons que le gaz obéit à l'équation d'état du gaz parfait $PV = nRT = nR(\Theta - \Theta_{0K})$. En traçant $\Theta(P)$ nous pouvons obtenir la température du zéro absolu : l'ordonnée à l'origine donne Θ_{0K} .



Nous voyons que la régression est bonne ($r=0,99991$) mais les points mesurés sont éloignés du zéro absolu. Par prolongement nous obtenons avec une confiance de 95% :

$$\Theta_{0K} = -266,0 \pm 4,8 \text{ } ^{\circ}\text{C}$$

Or, nous savons que le zéro absolu est à -273,15 °C, ce qui n'est pas très cohérent. Nous pouvons donc supposer qu'il y a un biais et que nous n'avons pas considéré toutes les sources d'incertitudes.

3) Régression avec barres d'erreurs

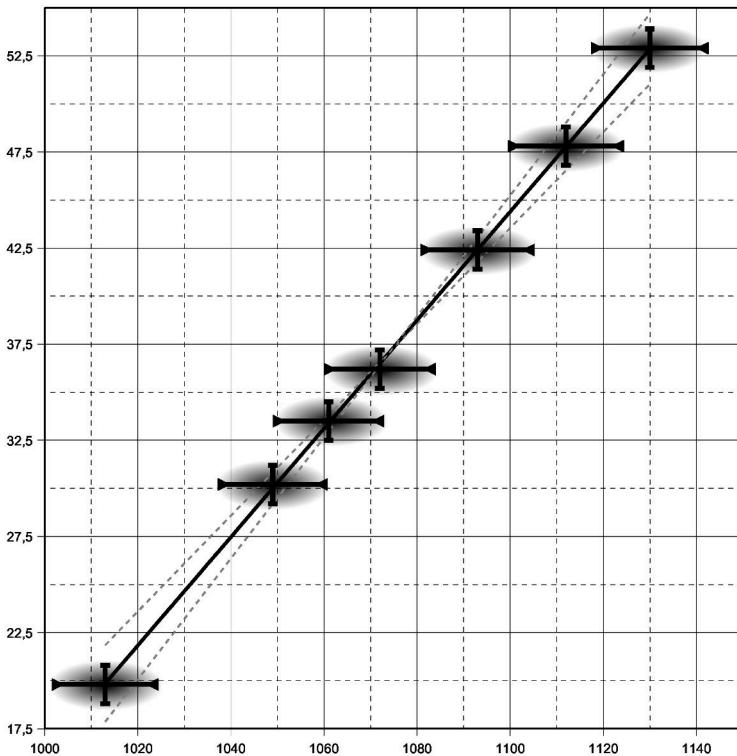
Les instruments de mesure ne sont pas parfaits et les notices indiquent leurs précisions. La précision du manomètre est à un pour cent près et celle du thermomètre au degré. Nous avons donc une incertitude sur x et y :

$$M_i(x_i \pm \Delta x_i, y_i \pm \Delta y_i)$$

Nous allons ajouter ce qu'on appelle des barres d'erreurs, nous n'avons plus des points mais des rectangles. Plus un point est précis, plus il joue un rôle important. La méthode des moindres carrés est modifiée, et la somme des carrés est pondérée par ce "poids" :

$$\sum_i w_i e_i^2 \quad \text{avec} \quad w_i = \frac{1}{(\Delta y_i)^2 + (\alpha \Delta x_i)^2}$$

C'est une méthode itérative, nous mettons initialement une valeur de α estimée, puis la valeur de α obtenue la remplace jusqu'à l'égalité des valeurs.



Nous obtenons alors : $\Theta_{0K} = -266 \pm 35^\circ C$ avec la même confiance que les incertitudes sur x_i et y_i . La valeur est, cette fois, satisfaisante. Les sources principales d'incertitudes semblent retenues.

Nous pourrions aussi considérer l'incertitude de modélisation induite par l'hypothèse de gaz parfait, mais, dans les conditions expérimentales de cette expérience, l'hypothèse est excellente. Cette source d'incertitude est négligeable devant les autres considérées ici. L'utilisation d'un modèle de gaz réel (Van der Waals par exemple) permettrait de le démontrer.

Formules [i] :

$$S^2 = \sum_i w_i [y_i - (ax_i + b)]^2$$

$$\frac{\partial S^2}{\partial b} = 0 \quad \text{et} \quad \frac{\partial S^2}{\partial a} = 0$$

conduit à :

$$b = \frac{\sum w_i y_i \sum w_i x_i^2 - \sum w_i x_i \sum w_i x_i y_i}{\Delta}$$

et

$$a = \frac{\sum w_i \sum w_i x_i y_i - \sum w_i x_i \sum w_i y_i}{\Delta}$$

avec

$$\Delta = \sum w_i \sum w_i x_i^2 - (\sum w_i x_i)^2$$

puis

$$\Delta b = \sqrt{\frac{\sum w_i x_i^2}{\Delta}}$$

et

$$\Delta a = \sqrt{\frac{\sum w_i}{\Delta}}$$

4) Linéarisation

Dans de nombreux cas nous pouvons nous ramener à un modèle linéaire (ici au sens $y=ax+b$). Nous donnons quelques exemples ci-dessous :

$y = \alpha x^\beta$ $x, y, \alpha > 0$	$y' = ax' + b$ avec $y' = \ln(y)$ $x' = \ln(x)$ et $\beta = a$ $\alpha = e^b$ et $\Delta \beta = \Delta a$ $\Delta \alpha = \alpha \Delta b$
$y = \alpha e^{\beta x}$ $x, \alpha > 0$	$y' = \ln(y)$ $x' = x$
$y = \frac{1}{\alpha + \beta x}$	$y' = \frac{1}{y}$
$y = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}$ (loi logistique)	$y' = \ln(\frac{y}{1-y})$ $y' = \alpha + \beta x$
$y = 1 - \left(\frac{\alpha}{x}\right)^\beta$ (loi de Pareto)	$y' = \ln(1-y)$ $x' = \ln(x)$ $\beta = -a$ $\alpha = e^{-\frac{b}{a}}$
$y = 1 - e^{-\left(\frac{x}{\alpha}\right)^\beta}$ (loi de Weibull)	$y' = \ln\left(\ln(\frac{1}{1-y})\right)$ $x' = \ln(x)$ $\beta = a$ $\alpha = e^{-\frac{b}{a}}$
$y = \alpha + e^{\beta x}$	non linéarisable
$y = \frac{\alpha x}{\beta + x}$	non linéarisable
$y = \alpha + \beta x + \gamma x^2$	non linéarisable selon $y(x)$

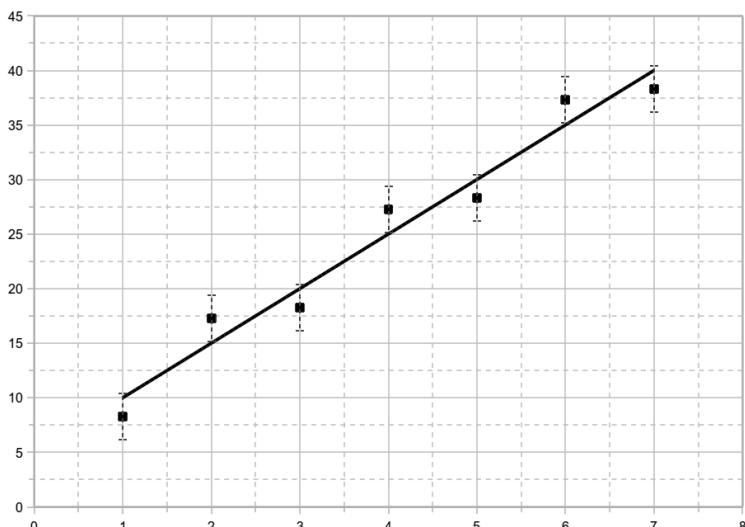
5) Comparaison des méthodes

Dans toutes les méthodes de régression que nous considérons nous nous intéressons à la variabilité de y à x fixé (la démarche inverse donnerait des résultats équivalents).

Pour un x_i donné nous avons un y_i et son écart-type σ_{y_i} .

a) Résumé

1- Régression simple



$$y_I = 8,3 \quad y_I = 17,3 \quad y_I = 18,3 \quad y_I = 27,3 \quad y_I = 28,3 \quad y_I = 37,3 \quad y_I = 38,3$$

$$\text{Cas 1 : } a=5 \quad b=5 \quad s_r=2,34 \quad s_a=0,443 \quad \text{et} \quad s_b=1,98$$

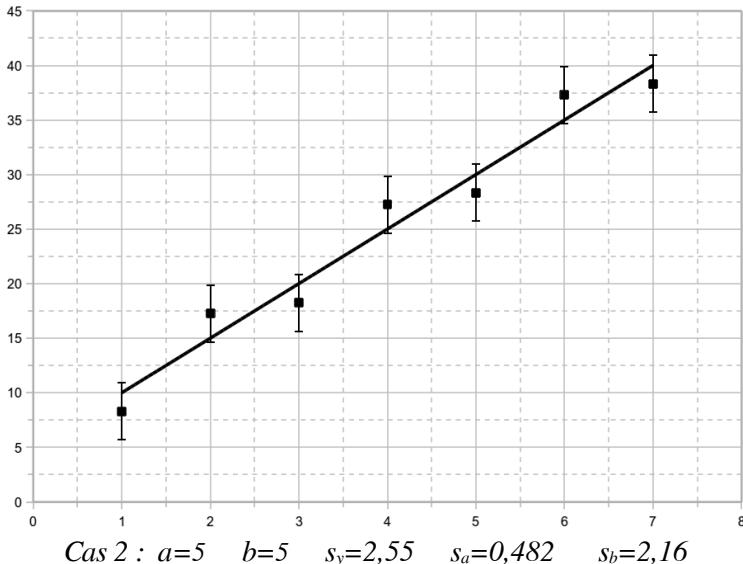
La régression simple ne correspond pas à des incertitudes sur les données nulles. Elles sont inconnues et nous les estimons à partir des données elles-mêmes. Les incertitudes

sont considérées constantes quelque soit y_i . L'incertitude correspond à l'écart-type des y_i par rapport à la droite estimée :

$$s_{y_i} = s_y = s_r = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n-2}}$$

$$s_a = \frac{s_r}{\sqrt{\sum_i (x_i - \bar{x})^2}} \quad s_b = \sqrt{\frac{\bar{x}^2}{\sum_i (x_i - \bar{x})^2}} s_r$$

2- Régression avec écarts-types constants selon y

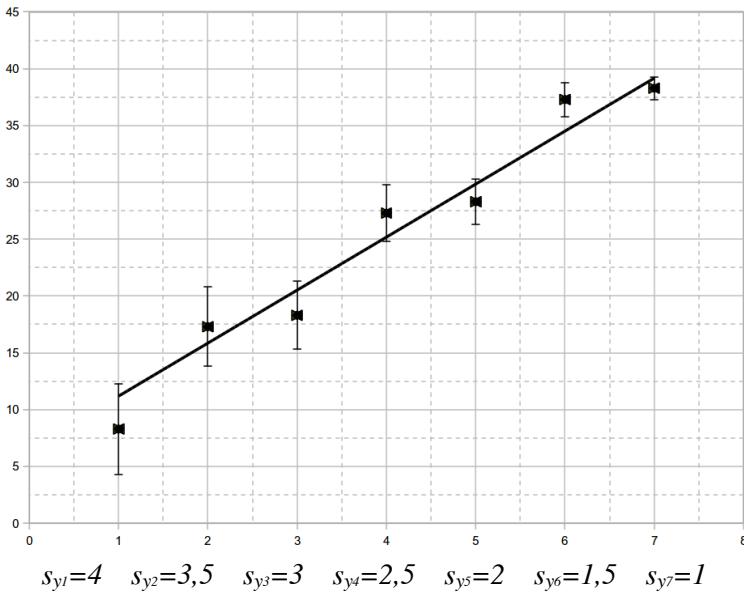


Dans ce cas les s_{y_i} sont égaux et connus : $s_{y_i} = s_y$ et

$$s_a = \frac{s_y}{\sqrt{\sum_i (x_i - \bar{x})^2}} \quad s_b = \sqrt{\frac{\bar{x}^2}{\sum_i (x_i - \bar{x})^2}} s_y$$

Si la droite ne passe pas par les barres d'erreurs, on peut penser que toutes les sources d'incertitudes n'ont pas été retenues. Il faut, soit les intégrer dans s_y , soit appliquer la méthode précédente.

3- Régression avec écart-types selon y



$$s_{y_1}=4 \quad s_{y_2}=3,5 \quad s_{y_3}=3 \quad s_{y_4}=2,5 \quad s_{y_5}=2 \quad s_{y_6}=1,5 \quad s_{y_7}=1$$

$$\text{Cas 3 : } a=4,67 \quad b=6,51 \quad s_a=0,482 \quad s_b=2,16$$

Les s_{y_i} sont connus. Nous pouvons appliquer la formule de propagation des écart-types :

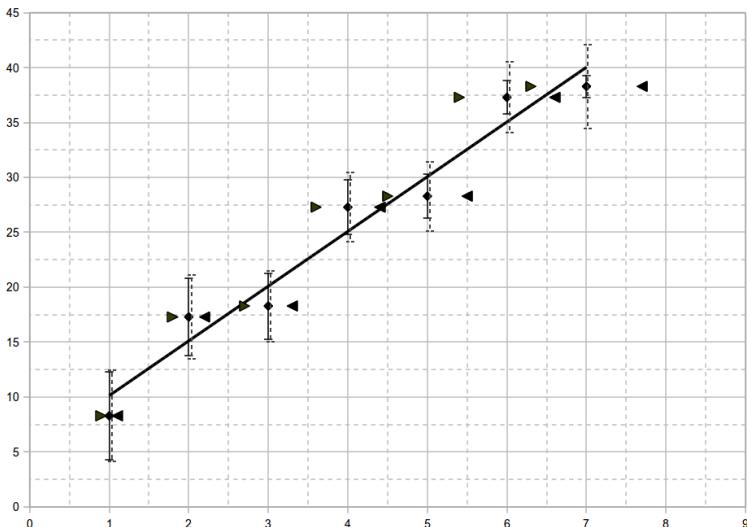
$$s_a^2 = \sum_i \left(\frac{\partial a}{\partial y_i} \right)^2 s_{y_i}^2 \quad \text{et} \quad s_b^2 = \sum_i \left(\frac{\partial b}{\partial y_i} \right)^2 s_{y_i}^2$$

Les formules sont exactes et permettent de retrouver les expressions des cas précédents. Dans le cas plus complexe

d'écart-types variables selon y_i nous avons les estimations suivantes :

$$\Delta a = \frac{1}{\sqrt{\sum w_i}} \frac{1}{\sqrt{x^2 - \bar{x}^2}} \quad \Delta b = \frac{1}{\sqrt{\sum w_i}} \sqrt{\frac{\bar{x}^2}{x^2 - \bar{x}^2}} \quad w_i = \frac{1}{\sigma_{y_i}^2}$$

4- Régression avec écart-types selon y et x



$$s_{y_1}=4 \quad s_{y_2}=3,5 \quad s_{y_3}=3 \quad s_{y_4}=2,5 \quad s_{y_5}=2 \quad s_{y_6}=1,5 \quad s_{y_7}=1$$

$$s_{x_1}=0,1 \quad s_{x_2}=0,2 \quad s_{x_3}=0,3 \quad s_{x_4}=0,4 \quad s_{x_5}=0,5 \quad s_{x_6}=0,6 \quad s_{x_7}=0,7$$

$$s_1=4,0 \quad s_2=3,6 \quad s_3=3,4 \quad s_4=3,2 \quad s_5=3,2 \quad s_6=3,4 \quad s_7=3,6$$

$$\text{Cas 4 : } a=4,98 \quad b=5,14 \quad s_a=0,695 \quad s_b=3,16$$

Comme x_i est fixé nous reportons la dispersion de x_i

$$\text{sur } y_i : \quad s_{y_i \text{Total}}^2 = s_i^2 = s_{y_i}^2 + a^2 s_{x_i}^2$$

À nouveau tout se passe comme si seuls les y_i avaient des

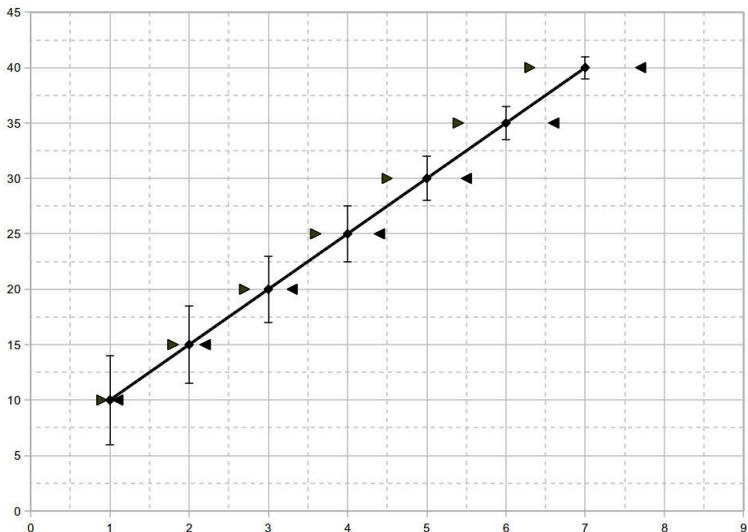
écart-types s_i . D'où les formules :

$$s_a^2 = \sum_i \left(\frac{\partial a}{\partial y_i} \right)^2 s_i^2 \quad \text{et} \quad s_b^2 = \sum_i \left(\frac{\partial b}{\partial y_i} \right)^2 s_i^2$$

Les dérivées sont complexes à calculer (les poids dépendent de a), nous pouvons par contre les évaluer numériquement. Aussi nous utilisons habituellement les estimations suivantes :

$$\Delta a = \sqrt{\frac{\sum w_i}{\Delta}} \quad \Delta b = \sqrt{\frac{\sum w_i x_i^2}{\Delta}} \quad w_i = \frac{1}{\sigma_{y_i}^2 + a^2 \sigma_{x_i}^2}$$

b) Discussion



$$\text{Cas 5 : } a=5 \quad b=5 \quad s_a=0,696 \quad s_b=3,16$$

Les incertitudes de ce cas peuvent avoir pour origine les instruments de mesure. La dispersion des valeurs du pre-

mier cas p68 peuvent provenir d'erreurs accidentelles liées à l'expérience ou d'un phénomène par essence aléatoire.

Nous aimerais que le quatrième cas p71 intègre l'ensemble les sources d'incertitudes des cas 1 et 5 :

$$s_{a_1} \simeq 0,443 \text{ et } s_{a_5} \simeq 0,696 \text{ donne } s_{a_4} \simeq 0,695$$

Nous avons utilisé les formules classiques. Nous avons l'impression que la dispersion autour de la droite de régression n'est pas incluse. Nous allons effectuer le calcul direct avec la formule de propagation et je propose la méthode des petites variations :

$$s_a^2 = \sum_i \left(\frac{\partial a}{\partial y_i} \right)^2 s_i^2 \quad \text{et} \quad \left(\frac{\partial a}{\partial y_j} \right) = \lim_{\Delta y_j \rightarrow 0} \left(\frac{\Delta a}{\Delta y_j} \right)$$

En gardant tous les $y_{i \neq j}$ constants : $y_j + \Delta y_j \rightarrow a + \Delta a$

Δy_j est une petite variation par rapport à y_j , si Δy_j devient chaque fois plus petit, $\frac{\Delta a}{\Delta y_j}$ reste constant (définition de la dérivée).

Essayons de retrouver le résultat pour s_{a_5} : remplaçons $y_1 = 10$ par $y_1 = 10,001$ alors de $a = 5$ nous avons après itérations $a = 4,999907$ d'où $\frac{\Delta a}{\Delta y_1} \simeq -0,093$.

Nous remettons y_1 à 10 et répétons le procédé pour y_2 en le remplaçant par 15,001. D'où les résultats suivants :

	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$
$\frac{\partial a}{\partial y_j}$	-0,093	-0,078	-0,049	-0,006	0,041	0,080	0,105

Nous trouvons alors $s_{a_5} \simeq 0,696$, le même résultat que l'autre méthode.

Procérons de même pour s_{a_4} :

	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$
$\frac{\partial a}{\partial y_j}$	-0,096	-0,080	-0,050	-0,007	0,081	0,122	0,147

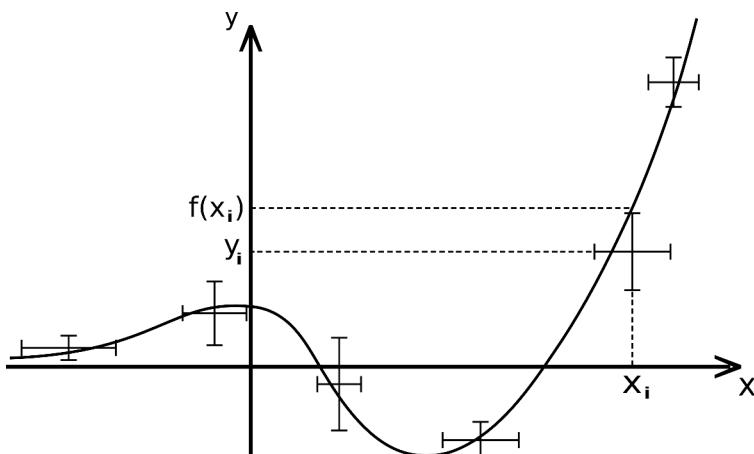
Nous trouvons alors $s_{a_4} \simeq 0,786$, résultat significativement différent de l'estimation classique qui semble maintenant intégrer l'ensemble des sources d'incertitudes.

Dans l'exercice *Écart-types proportionnel à y*, page 98, nous étudions un cas particulier où nous effectuons le calcul analytique direct. La comparaison peut ainsi être élargie.

D. Régression généralisée

Nous généralisons la méthode des moindres carrés pondérés pour une fonction et une dépendance des paramètres non linéaires. Bien que la régression multiple ait des développements analogues elle n'est pas traitée dans cet ouvrage [x].

1) Principe



Nous comparons y_i à la valeur $f(x_i)$ donnée par la fonction cherchée : $y_i - f(x_i)$. Le poids attribué à $(y_i - f(x_i))^2$ est inversement proportionnel à la variance du terme $y_i - f(x_i)$.

Les grandeurs x_i et y_i sont indépendantes¹⁴ d'où :

14 Parler d'indépendance entre deux variables alors que nous recherchons une relation fonctionnelle entre les deux peut sembler illogique. Nous faisons ici référence à la détermination expérimentale de chaque grandeur qui est indépendante au sens des incertitudes de l'autre (discuté dans le complément p200).

$$V(y_i - f(x_i)) = V(y_i) + V(f(x_i))$$

En appliquant à nouveau la formule de propagation de la variance :

$$V(f(x_i)) = f'(x_i)^2 V(x_i)$$

dans : $S^2 = \sum_i w_i (y_i - f(x_i))^2$

avec $w_i = \frac{1}{\sigma_{y_i}^2 + f'(x_i)^2 \sigma_{x_i}^2}$ ¹⁵

Ensuite $\frac{\partial S^2}{\partial a_k} = 0$ permet de déterminer les paramètres a_k de notre fonction (par un calcul analytique ou une résolution numérique).

À chaque fois nous pouvons nous ramener à un système d'équations linéaires de la forme $HA=B$, avec H une matrice carrée et A le vecteur associé aux paramètres. Ensuite $A=H^T B$, où H^T est la matrice inverse de H .

Les incertitudes sur les paramètres sont données par les termes diagonaux de la matrice H^T :

$\sigma_{a_k}^2 = (H^{-1})_{kk} \sigma_r^2$ où σ_r^2 correspond à la variance des résidus par rapport à la courbe estimée.

Quand il n'y a pas d'incertitudes sur les données explicitées, l'écart-type des résidus à p paramètres s'écrit :

15 S^2 est aussi appelé χ^2 . Si nous supposons les lois de probabilités normales, les écart-types peuvent être remplacés par les incertitudes en utilisant la loi de Student.

$$s_r = \sqrt{\frac{\sum (y_i - f(x_i))^2}{n-p}}$$

Quand w_i dépend de paramètres nous itérons la méthode jusqu'à ce que nous puissions considérer les poids constants. Si nous connaissons les écart-types des données, les écart-types des paramètres peuvent se calculer avec la formule de propagation, ou bien être estimés comme en régression linéaire avec barres d'erreurs p66.

2) Régression polynomiale

Dans ce cas :

$$f(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m = \sum_{i=0}^m a_i x^i$$

Cette fonction est non-linéaire par rapport à x et linéaire par rapport aux paramètres.

Illustrons avec l'exemple de la loi de Cauchy due au phénomène de dispersion de la lumière :

$$n(\lambda) = a_0 + \frac{a_1}{\lambda^2} + \frac{a_2}{\lambda^4}$$

avec les données suivantes :

$\lambda(\mu m)$	0,6157	0,5892	0,5685	0,5152	0,4981
n	1,71276	1,71578	1,71852	1,72716	1,73060

Les incertitudes sur λ et n sont dans un premier temps négligées. Que valent a_0, a_1, a_2 et leurs incertitudes ?

Nous avons une régression parabolique sous la forme :

$$f(x) = a_0 + a_1 x + a_2 x^2 \quad \text{avec} \quad f = n, \quad x = 1/\lambda^2.$$

$$S^2 = \sum (y_i - a_0 - a_1 x_i - a_2 x_i^2)^2 \quad \text{et} \quad \partial S^2 / \partial a_k = 0$$

$$\text{donne : } \begin{cases} \bar{y} - a_0 - a_1 \bar{x} - a_2 \bar{x}^2 = 0 \\ \bar{x}\bar{y} - a_0 \bar{x} - a_1 \bar{x}^2 - a_2 \bar{x}^3 = 0 \\ \bar{x}^2\bar{y} - a_0 \bar{x}^2 - a_1 \bar{x}^3 - a_2 \bar{x}^4 = 0 \end{cases}$$

d'où $HA=B$ avec :

$$H = \begin{pmatrix} 1 & \bar{x} & \bar{x}^2 \\ \bar{x} & \bar{x}^2 & \bar{x}^3 \\ \bar{x}^2 & \bar{x}^3 & \bar{x}^4 \end{pmatrix}, \quad A = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} \bar{y} \\ \bar{x}\bar{y} \\ \bar{x}^2\bar{y} \end{pmatrix}.$$

Avec un tableur nous pouvons facilement faire les calculs :

$$H \approx \begin{pmatrix} 1 & 3,3 & 11 \\ 3,3 & 11 & 38 \\ 11 & 38 & 135 \end{pmatrix}, \quad H^{-1} \approx \begin{pmatrix} 4150 & -2530 & 376 \\ -2530 & 1546 & -230 \\ 376 & -230 & 34,3 \end{pmatrix},$$

$$B \approx \begin{pmatrix} 1,7 \\ 5,7 \\ 19 \end{pmatrix} \quad \text{d'où} \quad A = H^{-1}B \approx \begin{pmatrix} 1,68129 \\ 0,01135 \\ 0,00022 \end{pmatrix}.$$

Pour les incertitudes :

$$\Delta a_k = \sqrt{(H^{-1})_{kk}} t_{n-3} s_r \quad \text{avec} \quad s_r = \sqrt{\frac{\sum (y_i - f(x_i))^2}{n-3}}$$

$s_r=1,87 \cdot 10^{-5}$ et à 95% de confiance $t_{n-3}=4,30$

D'où :

$$a_0=1,6813 \pm 0,0017, \quad a_1=(1,13 \pm 0,10) \cdot 10^{-2} \mu\text{m}^2$$

et $a_2=(2,2 \pm 1,5) \cdot 10^{-4} \mu\text{m}^4$

Prenons maintenant l'incertitude $\Delta n=0,00004$ à 95% de confiance, nous avons une incertitude sur y et pas sur x . Les poids sont donc constants : $w_i=w=1/\Delta n^2$. Nous parvenons au même système d'équations, et de même pour $H, H^T B$ et A .

Ici, en procédant de manière analogue à la régression linéaire avec barres d'erreurs nous pouvons en première approche estimer la dispersion des résidus par $1/\sqrt{w_i}$:

$$\Delta a_k = \sqrt{\frac{(H^{-1})_{kk}}{\sum w_i}}, \text{ d'où :}$$

$$a_0=1,6813 \pm 0,0012, \quad a_1=(1,14 \pm 0,07) \cdot 10^{-2} \mu\text{m}^2$$

et $a_2=(2,2 \pm 1,0) \cdot 10^{-4} \mu\text{m}^4$

Avec l'ensemble des incertitudes, $\Delta n=0,00004$, $\Delta \lambda=0,005 \mu\text{m}$ et $\Delta x=2\Delta \lambda/\lambda^3$, les poids dépendent du point. Nous parvenons néanmoins au même système d'équations en considérant les poids constants sur une itération, et nous calculons $H, H^T B$ et A avec des paramètres estimés.

Pour calculer les moyennes nous utilisons l'expression du poids suivante :

$$w_i = \frac{1}{\Delta n^2 + (a_1 + 2a_2 x_i)^2 \Delta x_i^2}$$

Nous obtenons ainsi une première expression des paramètres. Pour itérer nous remplaçons les paramètres estimés par ces nouveaux paramètres. Nous itérons autant que nécessaire. La convergence est souvent très rapide :

Itération	a_0	a_1	a_2
Estimés	1,5	0,005	0,0001
1	1,681282848208	0,011350379724	0,000219632215
2	1,681269875466	0,011358254795	0,000218466771
...			

Ici aussi nous prenons : $\Delta a_k = \sqrt{\frac{(H^{-1})_{kk}}{\sum w_i}}$, d'où :

$$a_0 = 1,6813 \pm 0,0013, \quad a_1 = (1,14 \pm 0,08) \cdot 10^{-2} \mu m^2 \\ \text{et } a_2 = (2,2 \pm 1,2) \cdot 10^{-4} \mu m^4$$

3) Régression non linéaire

Nous allons partir de p paramètres a_k estimés et utiliser une méthode d'itération. Sur chaque itération les poids seront considérés constants et la fonction sera linéarisée pour chacun des n points sur l'ensemble des paramètres.

La fonction dépend des x_i et des paramètres a_k .

Nous posons : $f_i = f(x_i; a_k)$.

$$S^2 = \sum_i w_i (y_i - f_i)^2 \quad \text{et} \quad \frac{\partial S^2}{\partial a_k} = -2 \sum_{i=1}^n w_i \frac{\partial f}{\partial a_k} (y_i - f_i) = 0$$

Les premiers paramètres estimés sont notés $a_{0,k}$. ensuite nous les noterons $a_{j,k}$ pour la $j^{\text{ème}}$ itération. Nous allons effectuer un développement limité d'ordre un pour une petite variation $\delta a_{0,k}$ autour de $a_{0,k}$ ¹⁶.

Notons $\vec{a} = (a_1, \dots, a_k, \dots, a_p)$ et $\vec{\delta a} = (\delta a_1, \dots, \delta a_k, \dots, \delta a_p)$:

$$f(x_i; \vec{a}_0 + \vec{\delta a}_0) = f(x_i; \vec{a}_0) + \sum_{l=1}^p \delta a_{0,l} \left(\frac{\partial f(x_i; \vec{a})}{\partial a_l} \right)_{a_0}$$

soit en allégeant les notations :

$$f_i = f_{0,i} + \sum_l \left(\frac{\partial f}{\partial a_l} \right)_{0,i} \delta a_{0,l}$$

$$\text{ainsi } \sum_i w_i \frac{\partial f}{\partial a_k} (y_i - f_{0,i} - \sum_l \left(\frac{\partial f}{\partial a_l} \right)_{0,i} \delta a_{0,l}) = 0$$

$$\text{et } \sum_i w_i \frac{\partial f}{\partial a_k} (y_i - f_{0,i}) = \sum_{i,l} w_i \frac{\partial f}{\partial a_k} \frac{\partial f}{\partial a_l} \delta a_{0,l}.$$

$$\text{Nous posons } H_{k,l} = \sum_i w_i \left(\frac{\partial f}{\partial a_k} \right)_i \left(\frac{\partial f}{\partial a_l} \right)_i = H_{l,k},$$

$$B_k = \sum_i w_i \frac{\partial f}{\partial a_k} (y_i - f_{0,i}) \quad \text{et} \quad A_l = \delta a_{0,l}.$$

d'où à nouveau $HA=B$ et $A=H^T B$. Nous itérons jusqu'à convergence et variations sur les paramètres négligeables.

16 MATH : nous généralisons la notion de dérivée en ajoutant les variations selon tous les paramètres :

$$f(x_0 + \epsilon) \approx f(x_0) + \epsilon (f'(x))_{x_0}$$

Ici aussi nous utilisons : $\sigma_{a_k}^2 = (H^{-1})_{kk} \sigma_r^2$.

Illustrons avec une expérience de biologie, où l'on étudie la relation entre la concentration du substrat $[S]$ et la vitesse de réaction dans une réaction enzymatique à partir de données reportées dans le tableau suivant [ix] :

i	1	2	3	4	5	6	7
$[S]$	0,038	0,194	0,425	0,626	1,253	2,500	3,740
v	0,050	0,127	0,094	0,2122	0,2729	0,2665	0,3317

Nous souhaitons ajuster sous la forme :

$$y = \frac{\alpha x}{\beta + x} \text{ en posant } v=y \text{ et } [S]=x.$$

Nous partons des estimations $\alpha_0=0,9$ et $\beta_0=0,2$.

$$\frac{\partial f}{\partial \alpha} = \frac{x}{\beta + x}, \quad \frac{\partial f}{\partial \beta} = -\frac{\alpha x}{(\beta + x)^2}, \quad H = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix},$$

$$A = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}.$$

En absence d'incertitudes fournies sur x_i et y_i :

$$H_{11} = \sum_{i=1}^7 \left(\frac{\partial f}{\partial \alpha} \right)_{\alpha_0, i} \left(\frac{\partial f}{\partial \alpha} \right)_{\alpha_0, i} = \sum_i \left(\frac{\partial f}{\partial \alpha} \right)_{\alpha_0, i}^2 = \sum_i \left(\frac{x_i}{\beta_0 + x_i} \right)^2,$$

$$H_{12}=H_{21}=\sum_i \frac{\alpha_0 x_i^2}{(\beta_0+x_i)^3} \text{ et } H_{22}=\sum_i \frac{\alpha_0^2 x_i^2}{(\beta_0+x_i)^4}.$$

$$B_1=\sum_i \frac{x_i}{\beta_0+x_i}(y_i-f_{0,i}) \text{ et } B_2=-\sum_i \frac{\alpha_0 x_i}{(\beta_0+x_i)^2}(y_i-f_{0,i})$$

$$\text{avec } f_{0,i}=\frac{\alpha_0 x_i}{\beta_0+x_i}$$

Nous pouvons maintenant tout mettre dans un tableur, nous obtenons :

$$H_{11}\simeq 3,81, \quad H_{12}=H_{21}\simeq -2,89, \quad H_{22}\simeq 3,70,$$

$$B_1\simeq -2,33 \quad \text{et} \quad B_2\simeq 1,86.$$

puis

$$H^{-1}\simeq \begin{pmatrix} 0,649 & 0,508 \\ 0,508 & 0,668 \end{pmatrix} \quad \text{d'où} \quad A=H^{-1}B\simeq \begin{pmatrix} -0,567 \\ 0,0602 \end{pmatrix}.$$

Ainsi de $\alpha_1=\alpha_0+\delta\alpha_0$ et $\beta_1=\beta_0+\delta\beta_0$ nous avons les nouveaux paramètres estimés :

$$\alpha_1\simeq 0,9-0,567\simeq 0,333 \quad \text{et} \quad \beta_1\simeq 0,2+0,06\simeq 0,260$$

Nous replaçons ces valeurs au début des calculs à la place de α_0 et β_0 pour itérer. Nous calculons de nouvelles matrices et vecteurs H , H^T , B et A , et nous obtenons α_2 et β_2 .

Nous donnons l'ensemble des résultats dans le tableau suivant :

Itérat°	α	β	$\delta\alpha$	$\delta\beta$	S^2
	0,9	0,2	-0,57	0,060	1,4454965815
1	0,333	0,26	0,0101	0,166	0,0150720753
2	0,343	0,43	0,0150	0,103	0,0084583228
3	0,358	0,53	0,0040	0,024	0,0078643240
4	0,3614	0,554	0,0004	0,0024	0,0078441826
5	0,36180	0,5561	0,00003	0,00018	0,0078440067
6	0,36183	0,5563	0,000002	0,000013	0,0078440057

Après suffisamment d'itérations : $H^{-1} \simeq \begin{pmatrix} 1,52 & 6,34 \\ 6,34 & 36,2 \end{pmatrix}$

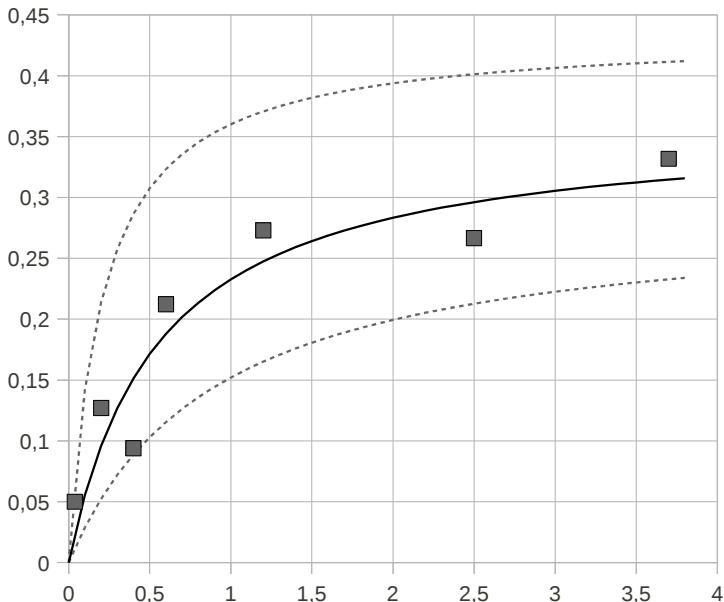
Calculons les incertitudes sur les paramètres :

$$s_r = \sqrt{\frac{S^2}{n-2}} \text{ et } \Delta \alpha = \sqrt{(H^{-1})_{11}} t_{n-2} s_r \simeq \sqrt{1,52} \cdot 1,48 \cdot \sqrt{\frac{0,00784}{5}}$$

d'où $\Delta \alpha \simeq 0,07$, $\Delta \beta \simeq 0,35$.

Finalement : $\alpha = 0,36 \pm 0,07$ et $\beta = 0,56 \pm 0,35$ à 80%.

Le graphique qui suit représente la vitesse de réaction en fonction de la concentration du substrat. Les carrés sont les points expérimentaux, la ligne continue la courbe optimale et en pointillés les deux courbes extrêmes $f_{\alpha_{max}, \beta_{min}}$ et $f_{\alpha_{min}, \beta_{max}}$.



Si nous voulons considérer les incertitudes pour chaque point nous devons ajouter les poids. Ceux-ci sont considérés constants sur chaque itération. Il faudra calculer la dérivée de f par rapport à x qui intervient dans l'expression du poids.

Pour des écart-types sur l'ensemble des données nous pouvons calculer les écart-types des paramètres en s'inspirant des méthodes de la régression linéaire p68.

E. Exercices

Exercice 1 : Corrélations corrigé p241

1- On réalise neuf expériences, nous obtenons à chaque fois trois réalisations des grandeurs X_1 , X_2 et X_3 :

$i =$	1	2	3	4	5	6	7	8	9
X_1	$x_1^1 = -1$	-1	-1	0	0	0	1	1	1
X_2	$x_2^1 = 1$	0	-1	1	0	-1	1	0	-1
X_3	$x_3^1 = -1$	0	1	-1	0	1	-1	0	1

a) Sur l'ensemble des expériences déterminez les moyennes arithmétiques et les écart-types de ces trois grandeurs.

b) Tracez X_2 en fonction de X_1 . Puis $X_3(X_1)$ et $X_3(X_2)$.

c) Calculez les coefficients de corrélation r_{12} , r_{13} et r_{23} . Commentez.

2- Même chose avec les données suivantes :

X_1	0	1	-1	2	0	-1	0	-2	1
X_2	1	2	-1	2	0	-2	-1	-2	1

3- Même chose avec les données suivantes :

X_1	-1	2	-2	0	-2	2	1
X_2	-1	0	2	-2	0	2	-1

Exercice 2 : Volumes corrigé p242

À l'aide d'une pipette jaugée nous remplissons quatre bêchers avec 100 mL d'eau chacun. Pour tester la pipette et connaître précisément la quantité d'eau, nous effectuons quatre pesées au décigramme et nous obtenons, ramené en mL, les résultats suivants pour les différents bêchers :

$$V_1 = \{100,1 ; 100,0 ; 99,9 ; 100,0\}$$

1- Calculez la moyenne et l'écart-type de V_1 . Estimez la précision de la pipette avec une confiance de 95%.

Nous remplissons maintenant deux bêchers et rassemblons le contenu des deux dans un seul :

$$V = V_1 + V_2.$$

Correspondant à V_1 , nous avons les mesures suivantes pour V_2 :

$$V_2 = \{100,0 ; 100,1 ; 100,0 ; 99,9\}$$

Par exemple pour la troisième mesure nous avons $V_1=99,9$ mL et $V_2=100,0$ mL.

2- Montrez que V_1 et V_2 sont des grandeurs indépendantes.

3- Calculez la moyenne de V , son écart-type et l'incertitude ΔV à 95%.

4- Pourriez-vous retrouver ce résultat avec la formule de propagation des incertitudes?

(Pour affiner le test il faudrait prendre plus de mesures, mais le principe reste le même, et les résultats restent valides car nous avons élargi avec le Student, considéré des données décorrélées et des paquets globalement gaussiens. Nous devrions aussi tenir compte des incertitudes sur les mesures -résolution- en plus de leur dispersion.)

Exercice 3 : Arbres corrigé p243

Nous voulons mesurer la distance d entre deux arbres. Pour cela nous disposons d'un bâton d'une longueur d'un mètre. D'un arbre à l'autre nous reportons le bâton cent fois. Nous estimons pour chaque report une incertitude de 1 cm.

Quelle est l'incertitude sur la valeur de d ?

Exercice 4 : Méthode de Bessel corrigé p244

C'est une méthode de focométrie qui permet de mesurer la distance focale f d'une lentille convergente. Pour cela on mesure la distance D entre un objet lumineux et son image sur un écran. Quand $D > 4f$ il existe deux positions où l'image est nette. La distance entre ces deux positions est notée d . Nous avons ensuite la distance focale de la lentille par la relation $f = (D^2 - d^2) / 4D$. Nous mesurons $D = 2000 \pm 10$ mm et $d = 536 \pm 20$ mm.

Quelle est alors l'incertitude sur f ?

Exercice 5 : Indice corrigé p246

Nous voulons déterminer expérimentalement l'indice n_2 d'un verre. Pour cela nous effectuons l'expérience de la réfraction d'un faisceau laser. Les lois de Descartes pour l'optique géométrique indique que $n_1 \cdot \sin(i_1) = n_2 \cdot \sin(i_2)$, avec n_i les indices lumineux et i_i les angles par rapport à la normale au dioptre. Nous avons $n_1 = n_{\text{air}} = 1$, $i_1 = 30 \pm 1^\circ$ et $i_2 = 20 \pm 2^\circ$.

Déterminez n_2 avec son incertitude.

Exercice 6 : Formule de Cauchy corrigé p247

Nous voulons mesurer la variation d'indice lumineux n en fonction de la longueur d'onde λ dans un milieu transparent (phénomène de dispersion). Pour cela nous utilisons un prisme, une lampe à sodium et un goniomètre. D'après la théorie la variation de $n(\lambda)$ doit suivre la formule de Cauchy dans le spectre visible :

$$n(\lambda) = A + \frac{B}{\lambda^2}$$

Le spectre de raie du sodium est connu. Pour chaque raie de longueur d'onde λ_i on calcule l'indice n_i correspondant en utilisant la formule du prisme suivante :

$$n_i = \frac{\sin\left(\frac{A+D_{m,i}}{2}\right)}{\sin\left(\frac{A}{2}\right)}$$

D_m est la déviation minimale. $A=60^\circ$ est l'angle au sommet du prisme. Ces deux angles sont mesurés à $2'$ près ($1'=$ une minute d'arc et $1^\circ=60'$).

Nous obtenons ainsi le tableau de données suivant :

$\lambda(nm)$	615,7	589,2	568,5	515,2	498,1
Couleur	rouge	jaune	vert-jaune	vert	bleu-vert
D_m	$57^\circ 49,5'$	$58^\circ 9'$	$58^\circ 28'$	$59^\circ 26,5'$	$59^\circ 50'$
n	1,71276	1,71568	1,71852	1,72716	1,73060

1- Déterminez l'incertitude sur n (celle-ci est globalement constante d'une valeur à l'autre).

2- Acceptant l'expression en $1/\lambda^2$ de la formule de Cauchy, en déduire A et B avec leurs incertitudes. Que vaut le coefficient de régression r ?

3- Peut-être qu'en traçant n en fonction de $1/\lambda$ ou $1/\lambda^3$ nous aurions un meilleur alignement des points ?

Nous voulons vérifier expérimentalement que la variation en $1/\lambda^2$ est bien la meilleure des relations polynomiales. Pour cela nous prenons la forme :

$$n(\lambda) = A + B \cdot \lambda^\alpha$$

Proposez une méthode pour déterminer A, B et α .

Nous pourrons vérifier le modèle car nous aurons α avec son incertitude.

Exercice 7 : Mur corrigé p249

Nous avons un mur d'une surface $S=72 \text{ m}^2$. La température extérieure est de 6°C et la température intérieure est maintenue à 18°C . Ce mur de 50 cm d'épaisseur est constitué de $e_p=40 \text{ cm}$ de paille compressée (conductivité thermique $\lambda_p=45 \text{ mW/K/m}$) et de $e_e=10 \text{ cm}$ d'enduit ($\lambda_e=200 \text{ mW/K/m}$). Les λ sont données à 10% près, les épaisseurs au cm et les température au demi degré.

1- Déterminez la résistance thermique, avec son incertitude, de la paille pour ce mur ($R \cdot \lambda \cdot S = e$)

2- Même chose pour l'enduit.

3- Sachant que les résistances thermiques s'associent comme les résistances électriques en série, déterminez la résistance thermique totale du mur avec son incertitude.

4- Quelle doit être la puissance minimale du chauffage de la maison rien que pour compenser les pertes par les murs ? ($\Delta T = R \cdot \Phi$)

Exercice 8 : Isolation et inertie corrigé p250

Nous avons une maison proche des préconisations pour une *maison énergie zéro*. Les résistances thermiques ramenées au m², soit e/λ , sont pour le toit de 8 m².K/W pour les murs et le sol de 4 m².K/W et pour les huisseries de 1 m².K/W (les résistances sont connues avec une précision de 10%). La maison a une surface au sol de 36 m², 54 m² de toit, 82 m² de mur et 8 m² d'huisseries.

1- Les résistances équivalentes du toit, du mur, du sol et des huisseries étant en parallèle, déterminez la résistance thermique totale (en K/W) avec sa précision.

Les températures extérieures et intérieures sont constantes.

2- Quelle doit être la puissance minimale du chauffage de la maison pour maintenir la température intérieure constante en compensant les pertes ?

3- Nous coupons le chauffage et nous mesurons la température au cours du temps :

t en heures	0	1	2	4	5	6	8	9	10
T en °C	18	16	14	12	11	10	9	9	8

Expliquez pourquoi la décroissance ne peut être linéaire. Nous considérons la décroissance de type exponentielle : $T(t)=a \cdot \exp(-t/\tau)+b$. Déterminez b, a et τ , avec leurs incertitudes.

4- La maison est isolée par l'extérieur. Le flux thermique perdu correspond à une diminution de l'énergie emmagasinée dans la maison. Cette inertie est due à la capacité thermique C des matériaux (J/K).

a) En raisonnant sur un intervalle de temps infinitésimal dt trouvez l'équation différentielle

vérifiée par $T(t)$ et retrouvez l'expression du 3.

b) Quelle est la relation entre τ , R et C ? Déterminez C et son incertitude.

Au 3. nous pouvons aussi tenir compte des incertitudes de mesure : le temps peut-être considéré comme parfaitement connu et la température est mesuré avec un thermomètre à dilatation d'un liquide avec des graduations tous les degrés Celsius.

Pour simplifier nous avons considéré que la température extérieure reste constante (pour tenir compte des variations jour/nuit nous envisagerions des variations sinusoïdales et une approche harmonique).

Exercice 9 : Rendement corrigé p253

Des quantités d'engrais déterminées sont répandues sur des champs et nous obtenons les rendements suivants :

Engrais (kg/ha)	100	200	300	400	500	600	700
Rendement (Q/ha)	41	44	53	63	66	65	78

1- Déterminez la droite de régression qui passe par ce nuage de points. Pente, ordonnée à l'origine et incertitudes avec une confiance de 95%.

2- Pour 550 kg d'engrais par ha estimez le rendement.

3- Même chose en absence d'engrais.

4- Si un agriculteur répand 250 kg d'engrais par hectare, quelle est la probabilité qu'il obtienne 40 à 60 quintaux de céréales ?

Exercice 10 : Étude d'une pile corrigé p255

Nous cherchons à déterminer la tension à vide E et la résistance interne r .

Pour cela nous mesurons pour la pile différentes valeurs de U et I avec un voltmètre et un ampèremètre ($U = E - r.I$) :

calibre pour U :	unité : V précision $\pm 0,05\% \pm 0,003$									
U (V)	4, 731	4, 731	4, 730	4, 728	4, 724	4, 724	4, 722	4, 721	4, 719	4, 716
calibres pour I :	unité : μA précision $\pm 0,2\% \pm 0,03$			unité : mA précision $\pm 0,2\% \pm 0,0003$						
I	92, 83	115, 45	152, 65	0, 2352	0, 4686	0, 5200	0, 5841	0, 6661	0, 7750	0, 9264

1- Sans les barres d'erreurs : déterminez $E \pm \Delta E$ et $r \pm \Delta r$.

2- Même chose en incluant cette fois les incertitudes de chaque mesure indiquées dans la notice du fabricant du multimètre.

Exercice 11 : Focométrie corrigé p257

Nous désirons déterminer la distance focale f d'une lentille convergente. La lentille est en O et fait l'image net en A' d'un point objet A . Nous mesurons OA , OA' et leurs incertitudes (sont incluses toute les sources d'incertitudes, géométriques, latitude de mise au point et modélisation). Nous considérons que la lentille vérifie la relation de conjugaison :

$$\frac{1}{OA'} + \frac{1}{OA} = \frac{1}{f}.$$

Déterminez f par l'intermédiaire d'une régression linéaire. Tableau page suivante.

Données expérimentales (mm) :			
OA	ΔOA	OA'	$\Delta OA'$
635	5	150	15
530	5	160	17
496	5	164	15
440	5	172	18
350	5	191	20
280	5	214	25
210	5	292	28
150	5	730	102

Théorie

Exercice 12 :

Pour la régression simple, montrez que :

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

Exercice 13 :

Pour la régression simple, montrez que nous pouvons aussi écrire $s_b = s_r \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}}$.

Exercice 14 :

Pour la régression simple, montrez que : $\Delta b = \sqrt{\bar{x}^2} \Delta a$.

Exercice 15 : Asymptotes corrigé p260

Pour la régression simple, montrez que au dessus et en dessous de la droite de régression les asymptotes des droites extrêmes, des courbes de confiance et de prédiction se confondent quand x_o devient grand.

Exercice 16 : Intervalle de confiance et de prédiction pour la régression linéaire avec barres d'erreurs corrigé p260

Régression avec barres d'erreurs

1- Donnez les expressions de \bar{x} et $\bar{x^2}$ en utilisant les poids w_i . Quel lien constatez-vous avec l'expression de Δ (p66) ?

Analogie

2- À partir des intervalles de confiance et de prédiction pour y_o en régression simple (p61 et suivante) déterminez par analogie les formules qui suivent :

estimation :

$$\Delta y_o = \frac{1}{\sqrt{n \sum w_i}} \sqrt{1 + \frac{(x_o - \bar{x})^2}{\bar{x}^2 - \bar{x}^2}}$$

prédiction :

$$\Delta y_o = \frac{1}{\sqrt{n \sum w_i}} \sqrt{1 + n + \frac{(x_o - \bar{x})^2}{\bar{x}^2 - \bar{x}^2}}$$

3- Déterminez les écarts selon y à la droite de régression pour les droites extrêmes, les courbes de confiance et les courbes de prédiction quand x_o devient grand.

Dans un exercice précédent nous avons montré que pour la régression linéaire les asymptotes se confondaient à l'infini. Par analogie que doit-on poser pour qu'il en soit de même en régression avec barres d'erreurs ?

Montrer que nous avons alors les formules suivantes :

Estimation :

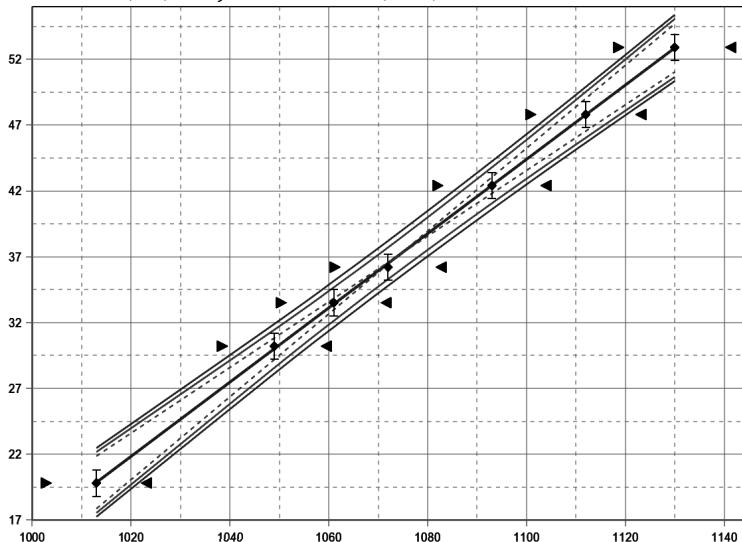
$$\Delta y_o = \frac{1}{\sqrt{\sum w_i}} \sqrt{1 + \frac{(x_o - \bar{x})^2}{x^2 - \bar{x}^2}}$$

Prédiction :

$$\Delta y_o = \frac{1}{\sqrt{\sum w_i}} \sqrt{2 + \frac{(x_o - \bar{x})^2}{x^2 - \bar{x}^2}}$$

Les formules ainsi obtenues par analogie sont empiriques, elles semblent expérimentalement cohérentes, et demandent confirmation par une démonstration théorique.

Courbes de confiance et prédiction pour l'expérience du zéro absolu : $T(^{\circ}\text{C})$ en fonction de $P(\text{hPa})$



Exercice 17 : Autres expressions corrigé p262

Pour la régression avec barres d'erreurs donnez l'expression de a , b , Δa et Δb en fonction de \bar{x} , \bar{y} , xy , \bar{x}^2 et y^2 . Comparez avec la régression simple.

Exercice 18 : Méthode des moindres carrés

corrigé p263

Démontrez par la méthode des moindres carrés les expressions de a et b :

1- pour la régression linéaire simple.

2- pour la régression linéaire avec barres d'erreurs.
Les Δx_i et Δy_i sont considérés petits par rapport à x_i et y_i .

Démonstration des expressions de Δa et Δb pour la régression simple :

Méthode 1 :

1- Montrez que $a = \sum_i p_i y_i$ avec $p_i = \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2}$.

2- Déduire de cette formule de a sa variance $V(a)$ ¹⁷.

Méthode 2 :

Utilisez la formule de propagation des écart-types.

Pour la régression linéaire simple, pourriez-vous retrouver a , b , Δa et Δb en utilisant la méthode matricielle de la régression généralisée ?

Exercice 19 : Espérance de a

corrigé p265

Pour la régression linéaire, nous notons α et β les paramètres de la population : $E(y_i) = \alpha x_i + \beta$.

a et b sont les paramètres estimés à partir d'un échantillon : $\hat{y}_i = a x_i + b$

Montrez que nous avons un estimateur non biaisé pour α , soit $E(a) = \alpha$.

17 MATH : $E(X+Y) = E(X) + E(Y)$. Si X et Y sont deux variables indépendantes : $V(X+Y) = V(X) + V(Y)$.

Exercice 20 : Écart-types proportionnels à y

corrigé p265

Nous nous plaçons dans le cas théorique particulier où les écart-types de la régression linéaire sont proportionnels à y : $\sigma_{y_i} = k y_i$. Ce cas, où les incertitudes relatives sont constantes, est expérimentalement courant. Nous nous intéressons ici au coefficient directeur a .

1- Montrez que :
$$a = \frac{\sum \frac{1}{y^2} \sum \frac{x}{y} - \sum \frac{x}{y^2} \sum \frac{1}{y}}{\sum \frac{1}{y^2} \sum \frac{x^2}{y^2} - \left(\sum \frac{x}{y} \right)^2}$$

2- Exprimez $\frac{\partial a}{\partial y_j}$ (calcul long).

3- Calculez s_a en utilisant les expressions trouvées pour les deux jeux de données suivants :

x_i	1	2	3	4	5	6	7
1 : y_i	10	15	20	25	30	35	40
2 : y_i	8,286	17,286	18,286	27,286	28,286	37,286	38,286

(Nous prenons $k=0,1$).

Retrouver ces résultats avec la méthode numérique (évaluation des dérivées par petites variations).

Comparer avec les valeurs obtenues par la méthode classique.

Exercice 21 : Interprétation de l'expression de w_i

corrigé p267

Justifiez graphiquement la position de a dans l'expression de w_i .

Régression non-linéaire

Exercice 22 : Décomposition en gaussiennes *corrigé p268*

Une fabrique usine des clous. Nous mesurons avec un pied à coulisse la taille de 48 clous :

Taille	59,97	59,98	59,99	60,00	60,01	60,02	60,03	60,04	60,05
Nombre	2	4	6	5	5	10	9	6	1

Deux machines fabriquent les clous. Les clous fabriqués par une machine n'ont pas exactement la même taille, nous supposons que la taille est distribuée selon un profil gaussien. Déterminer les caractéristiques des profils gaussiens des clous fabriqués par chacune des deux machines (maximum, moyenne et écart-type). Combien de clous ont fabriqués chaque machine ?

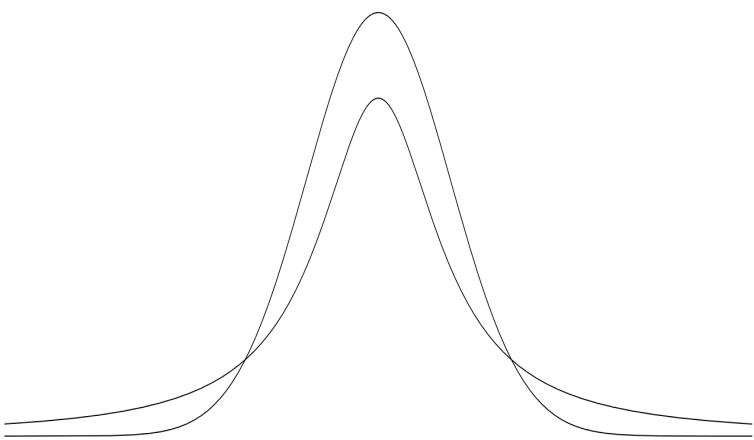
Pour vérifier vos calculs de corrélation vous pouvez utiliser la fonction du tableur OOo :

COEFFICIENT.CORRELATION(*;**).

Courbes : Insertion>Diagramme...>XY(Dispersion)>etc.

Pour les calculs matriciels, inversion de matrices : INVERSEMAT(*), produit de matrices : PRODUITMAT(*;**).

Vous pouvez utiliser le fichier IncertitudesLibresOOo32.ods sur le site www.incertitudes.fr pour réaliser les régressions. Feuille 2, régression simple et feuille 3 avec barres d'erreurs.



100

III. LOIS DE PROBABILITÉS

Nous allons lister différentes lois de probabilités discrètes puis continues. Nous montrerons qu'il s'agit bien de lois de probabilité et nous déterminerons leur espérance $\mu=E(X)$ et variance $V=E[(X-\mu)^2]$. La variance pourra aussi se calculer avec la formule $V=E[X^2]-E[X]^2$ et nous aurons ainsi l'écart-type $\sigma=\sqrt{V}$.

Nous pourrons calculer d'autres moments qui permettent de déterminer par exemple la symétrie et l'aplatissement de la distribution de probabilité.

Les moments ordinaires : $\mu_k=E(X^k)$

Les moments centrés réduits sont instructifs car caractéristiques de la forme de la distribution :

$$\beta_{k-2}=E\left[\left(\frac{X-\mu}{\sigma}\right)^k\right] \text{ ou } \beta_{k-2}=\frac{\mu_k}{\sigma^k}$$

β_1 : coefficient d'asymétrie (moment centré réduit d'ordre trois)

β_2 : coefficient d'aplatissement

Nous nous intéresserons aussi à la somme de variables aléatoires indépendantes : $Z=X+Y$.

A. Lois discrètes

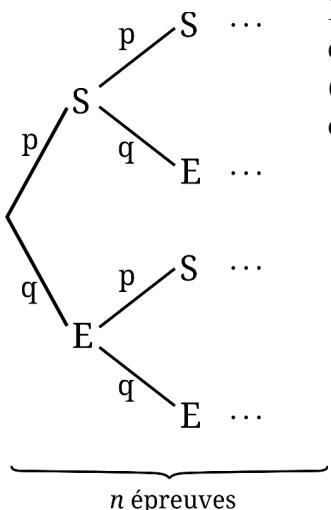
1) Loi binomiale

Nous considérons une succession de n épreuves identiques et indépendantes. Chaque épreuve a deux issues que nous nommons succès et échec. La probabilité du succès est notée $p=P(S)$. Les deux paramètres de la loi sont n et p et la loi peut s'écrire $\mathcal{B}(n,p)$.

Nous voulons déterminer la probabilité d'avoir k succès suite aux n épreuves. Un chemin comportant k succès comporte nécessairement $n-k$ échecs et sa probabilité vaut $p^k q^{n-k}$, avec $q=1-p$, probabilité de l'échec.

Il suffit ensuite de dénombrer le nombre de chemins qui comprennent k succès, manières de choisir k objets parmi n . n choix pour la position du premier succès, $n-1$ pour le deuxième et $n+1-k$ pour le k^{em} succès :

$$n(n-1)\dots(n+1-k)=n!/(n-k)! \text{ possibilités}^{18}.$$



Nous divisons ensuite par $k!$ pour enlever les comptages multiples (par exemple S_1S_2E et S_2S_1E correspondent au même chemin).

D'où :

$$P(X=k) = \binom{n}{k} p^k q^{n-k}$$

$$\text{avec } \binom{n}{k} = \frac{n!}{(n-k)!k!} \quad (\text{se dit « k parmi n »})$$

$$\text{et } \sum_{k=0}^n P(X=k) = 1$$

18 $n!$, se lit « n factorielle » avec $n!=1\times 2\times 3\dots\times n$.

On montre que $E(X)=np$ et $V(X)=npq$.

La somme de deux lois binomiales indépendantes de même p et aussi une loi binomiale. La somme d'une $\mathcal{B}_x(n_1, p)$ et d'une $\mathcal{B}_y(n_2, p)$ est une $\mathcal{B}_z(n_1+n_2, p)$.

Pour déterminer la loi de Z somme de X et Y nous utilisons, pour les lois discrètes, la propriété suivante :

$$P(Z=k) = \sum_i P([X=i] \cap [Y=k-i])$$

2) Loi géométrique

Nous considérons une épreuve à deux issues dont la probabilité du succès est p et celle de l'échec est notée $q=1-p$. Nous répétons cette épreuve de manière indépendante jusqu'au premier succès. Nous définissons la variable aléatoire X qui donne le rang du premier succès. Cette loi est notée $\mathcal{G}(p)$.

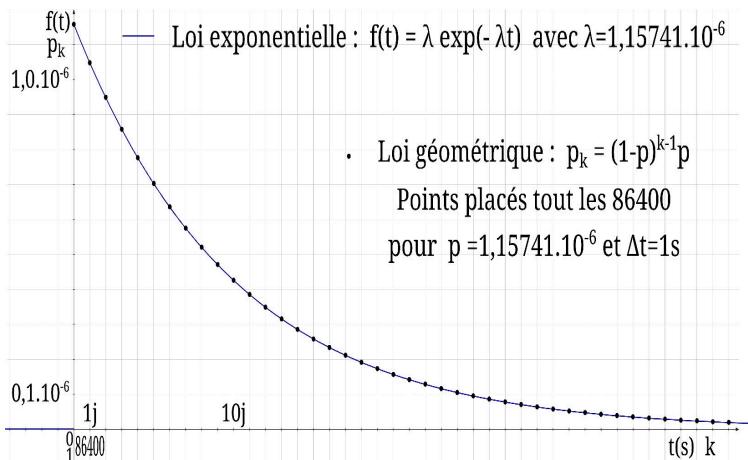
$$P(X=k) = q^{k-1} p \quad , \quad \sum_{k=1}^{\infty} P(X=k) = 1$$

$$E(X) = 1/p \quad \text{et} \quad V(X) = q/p^2$$

- Exemple : Nous avons une urne comportant 2 boules blanches et une boule noire. Les boules sont indiscernables au toucher. Quelle est la probabilité de tirer la première boule noire au troisième tirage avec remise ?

Réponse : $P(X=3) = (2/3)^2 \cdot 1/3 = 4/27$.

- Application : nous avons ici un modèle discret de durée de vie. Prenons par exemple une cellule qui chaque jour a un chance sur dix de mourir. Si elle n'est pas morte le lendemain, sa probabilité de mourir le jour suivant n'a pas changée : il n'y a pas de vieillissement. Chaque jour elle peut échapper à la mort et elle peut vivre éternellement. De manière équivalente, plutôt que de prendre une décision chaque jour, nous pouvons décider de sa mort toutes les heures avec une probabilité q' de survivre telle que $q'^{24} = q$ soit $q' \approx 0,996$. Ou même un choix toutes les secondes, nous avons alors des lois géométriques qui modélisent une même réalité avec $p \rightarrow 0$ et $q \rightarrow 1$.



Nous passons d'une loi discrète à une loi continue, le temps écoulé depuis le début de l'expérience est $t = k \Delta t$.

Regardons la fonction de répartition :

$$P(X \leq k) = \sum q^{k-1} p = \sum e^{\ln(1-p) \cdot (k-1)} p \approx \sum p e^{-pk}$$

Nous avons utilisé un développement limité, le fait que $t \gg \Delta t$ et $k \gg 1$.

Dans le cas limite :

$P(X \leq k) \rightarrow \int f(t) dt$, $\sum p e^{-\frac{p}{\Delta t}t} \rightarrow \int \lambda e^{-\lambda t} dt$ ainsi la loi exponentielle est la loi continue équivalente à la loi géométrique en prenant $\lambda = p/\Delta t$, ici pour notre cellule λ vaut environ $1,16 \cdot 10^{-6}$ par seconde.

La somme de deux variables aléatoires géométriques indépendantes $\mathcal{G}(p)$ est une loi binomiale négative $\mathcal{BN}(2,p)$, rang pour l'obtention de deux succès.

Pour une loi binomiale le nombre d'épreuves est fixé et nous regardons le nombre de succès, dans le cas de la loi binomiale négative c'est l'inverse nous nous intéressons au nombre d'épreuves nécessaires pour réaliser un nombre de succès fixé à l'avance. $\mathcal{BN}(r,p)$ est la loi de probabilité du rang du r^{em} succès¹⁹. Ainsi $\mathcal{G}(p) = \mathcal{BN}(1,p)$.

3) Loi de Poisson

Soient des événements qui se produisent avec une fréquence connue indépendante du temps écoulé depuis les événements précédents. Si λ est le nombre de fois que l'événement se produit en moyenne sur un intervalle donné alors la probabilité que l'événement se produise k fois sur

¹⁹ Il existe aussi une autre définition : nombre d'échecs précédents le r^{em} succès. Les relations restent vraies en redéfinissant de même la loi géométrique.

cet intervalle de temps est :

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad , \quad \sum_{k=0}^{\infty} P(X=k) = 1$$

$$E(X) = \lambda \quad \text{et} \quad V(X) = \lambda$$

- Exemple : Lors des Perséides d'août 1911 le taux horaire d'étoiles filantes était de 50, qu'elle était alors la probabilité de voir exactement 7 météores en 12 minutes ?

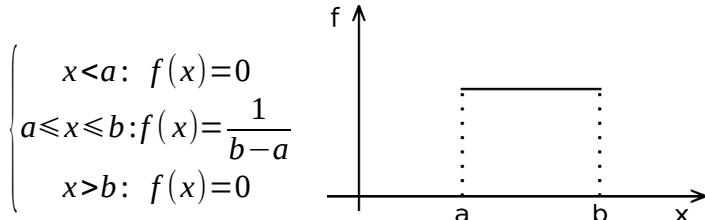
Réponse :

$$\lambda = \frac{12}{60} \cdot 50 = 10 \quad \text{et} \quad P(X=7) = \frac{10^7}{7!} e^{-10} \simeq 9\%$$

La somme de deux variables aléatoires de Poisson indépendantes $\mathcal{P}_X(\lambda_1)$ et $\mathcal{P}_Y(\lambda_2)$ est une loi de Poisson $\mathcal{P}_Z(\lambda_1 + \lambda_2)$.

B. Lois continues

1) Loi uniforme



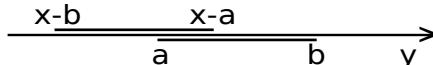
$$E(X) = \frac{a+b}{2} \quad \text{et} \quad V(X) = \frac{(b-a)^2}{12}$$

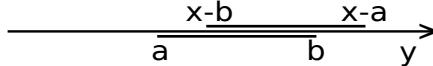
Pour déterminer la loi de Z somme de X et Y variables aléatoires indépendantes continues nous utilisons un produit de convolution :

$$f_Z(x) = \int_{-\infty}^{+\infty} f_X(y) f_Y(x-y) dy$$

Considérons la somme de deux lois uniformes continues indépendantes $\mathcal{U}(a,b)$. Ici l'intégrant est non nul si :

$$a < y < b \quad \text{et} \quad a < x - y < b \quad \text{soit} \quad x - b < y < x - a$$

Si $2a < x < a+b$: 

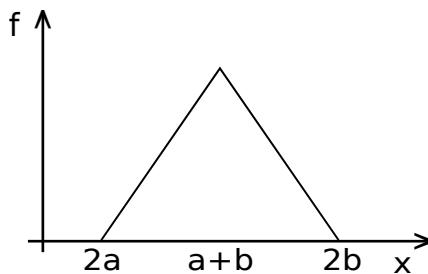
Si $a+b < x < 2b$: 

Si $2a < x < a+b$

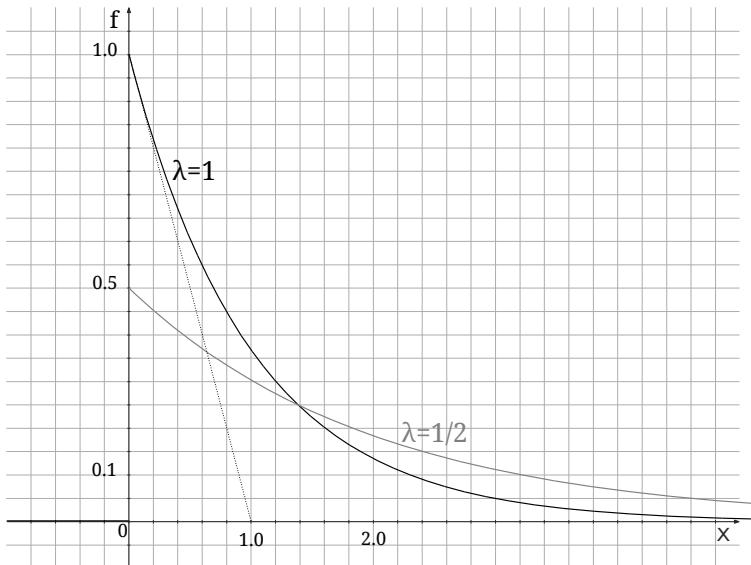
$$\text{alors } f_z(x) = \int_a^{x-a} \frac{1}{(b-a)^2} dy = \frac{x-2a}{(b-a)^2}$$

$$\begin{cases} x < 2a : f_z(x) = 0 \\ 2a \leq x < a+b : f_z(x) = \frac{x-2a}{(b-a)^2} \\ a+b \leq x \leq 2b : f_z(x) = \frac{2b-x}{(b-a)^2} \\ x > 2b : f_z(x) = 0 \end{cases}$$

Nous obtenons une loi triangulaire :



2) Loi exponentielle



$$\begin{cases} t \geq 0: f(t) = \lambda e^{-\lambda t} \\ t < 0: f(t) = 0 \end{cases} \quad E(T) = \frac{1}{\lambda} \quad \text{et} \quad V(T) = \frac{1}{\lambda^2}$$

Loi de durée de vie sans vieillissement, en effet, on montre pour cette loi que : $P(T>b) = P_{T>a}(T>a+b)$.

Bien différencier la vie moyenne $E(T)$ avec la demi-vie $t_{1/2}$ telle que $P(T>t_{1/2})=0,5$.

La loi de la somme de deux lois exponentielles indépendantes n'est pas une loi exponentielle.

3) Loi normale

La loi normale, encore nommée gaussienne, a précédemment été décrite page 20.

La somme de deux variables aléatoires normales indépendantes $\mathcal{N}_X(\mu_1, \sigma_1^2)$ et $\mathcal{N}_Y(\mu_2, \sigma_2^2)$ est une loi normale $\mathcal{N}_Z(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

4) Loi de Student

La loi de Student dépend du nombre de degrés de liberté k et utilise la fonction Gamma, fonction particulière décrite dans les outils mathématiques.

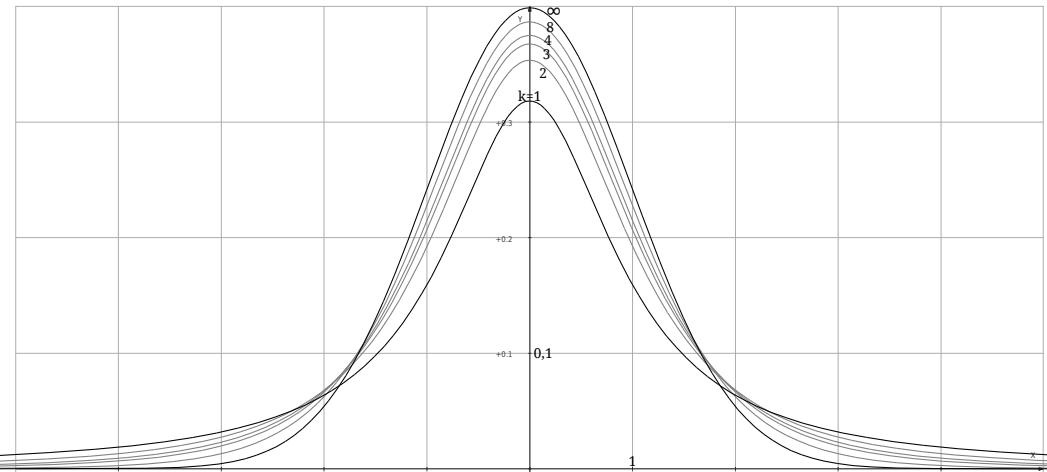
$$\text{Pour } k \geq 1, \quad f_k(x) = \frac{1}{\sqrt{k}\pi} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{k}\right)^{\frac{k+1}{2}}}$$

La loi de Student présentée tend vers la loi normale centrée réduite $\mathcal{N}(0,1)$ quand k tend vers l'infini :

$$\lim_{k \rightarrow +\infty} f_k(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$\text{Variance : } V_k = \frac{k}{k-2}, \text{ pour } k \geq 3.$$

$$\text{Coefficient d'aplatissement : } \beta_k = 3 \frac{k-2}{k-4}, \text{ pour } k \geq 5.$$



En exercices les expressions des premières Students sont exprimées, nous montrons que $\int_{-\infty}^{+\infty} f_k(x) dx = 1$, nous calculons l'expression de la variance et finalement nous montrons sur des exemples que la somme de deux Students indépendantes n'est pas une Student.

5) Loi du Khi-Deux

Soit k lois normales $\mathcal{N}(0,1)$ T_1, T_2, \dots et T_k indépendantes. La somme $X_k = T_1^2 + T_2^2 + \dots + T_k^2$ obéit à la loi du χ^2 . La loi du Khi² dépend du nombre de degrés de liberté k

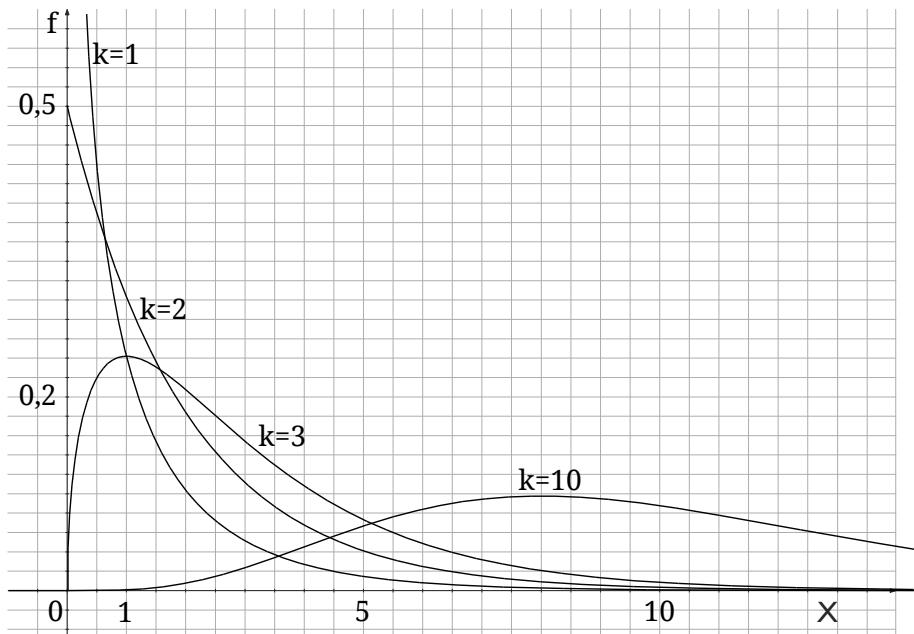
$$E(X_k) = \sum_{i=1}^k E(T_i^2) = k(V(T) + E(T)^2) = k$$

Pour $k \geq 1$ et $x \geq 0$, $f_k(x) = \frac{1}{2^{k/2} \Gamma(k/2)} x^{k/2-1} e^{-x/2}$

Espérance : $E_k = k$ Variance : $V_k = 2k$

Coefficient d'asymétrie : $\beta_{1,k} = \sqrt{8/k}$

Coefficient d'aplatissement : $\beta_{2,k} = 3 + 12/k$



La loi du Khi-deux tend vers la loi normale d'espérance k et de variance $2k$ quand k est grand.

C. Fonctions de variables à densité

Nous avons une variable aléatoire Y définie comme une fonction d'une variable aléatoire X continue : $Y=\varphi(X)$.

Nous connaissons la loi de X et nous voulons déterminer celle de Y. Nous raisonnons sur les fonctions de répartition F et nous dérivons ensuite pour avoir la densité f.

Fonction de répartition de X :

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(x) dx$$

Nous considérons le cas où $\varphi(x)$ est une fonction strictement monotone. La loi de Y est dans ce cas continue.
Fonction de répartition de Y :

$$F_Y(y) = P(Y \leq y) \text{ puis } f_Y(y) = \frac{dF_Y(y)}{dy}$$

Nous cherchons à exprimer $F_Y(y)$ en fonction de F_X .

1) Cas où $\varphi(x)=\ln(x)$: $x>0$ et $Y=\ln X$

$$F_Y(y) = P(Y \leq y) = P(\ln X \leq y) = P(e^{\ln X} \leq e^y) = P(X \leq e^y)$$

Nous avons appliqué la fonction réciproque $\varphi^{-1}(x)=e^x$.

La fonction exponentielle est strictement croissante, le sens de l'inégalité a donc été conservé.

Ainsi : $F_Y(y) = F_X(e^y)$ et $f_Y(y) = e^y f_X(e^y)$

- Exemple : la loi de X est la loi uniforme $\mathcal{U}(1;2)$. Quelle est la loi de Y ?

Nous avons : $f_X(x) = \begin{cases} 0 & x \leq 1 \\ 1 & 1 < x < 2, \text{ ainsi si } e^y \leq 1, \\ 0 & x \geq 2 \end{cases}$

$y \leq 0$ et $f_Y(y) = 0$. Nous continuons ainsi pour les deux autres cas et nous avons la loi de Y :

$$f_Y(y) = \begin{cases} 0 & y \leq 0 \\ e^y & 0 < y < \ln 2 \\ 0 & y \geq \ln 2 \end{cases}$$

2) Cas où $q(x)=ax+b$: $a \neq 0$ et $Y=aX+b$

Si $a > 0$:

$$F_Y(y) = P(Y \leq y) = P(ax + b \leq y) = P\left(X \leq \frac{y-b}{a}\right)$$

La fonction affine est strictement croissante pour a strictement positif, le sens de l'inégalité est conservé.

Ainsi : $F_Y(y) = F_X\left(\frac{y-b}{a}\right)$ et $f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$

Si $a < 0$:

$$F_Y(y) = P\left(X \geq \frac{y-b}{a}\right) = 1 - P\left(X < \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right)$$

et $f_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right)$ d'où $f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$

- Exemple 1 : la loi de X est la loi uniforme $\mathcal{U}(0;1)$. Quelle est la loi de Y ?

Nous avons : $f_X(x) = \begin{cases} 0 & x \leq 0 \\ 1 & 0 < x < 1, \text{ ainsi si } \frac{y-b}{a} \leq 0 \\ 0 & x \geq 1 \end{cases}$

$a > 0$ alors $y \leq b$ et $f_Y(y) = 0$. Nous continuons ainsi pour les deux autres cas et nous avons la loi de Y :

$$f_Y(y) = \begin{cases} 0 & y \leq b & U(0,1) \rightarrow U(b, a+b) \\ \frac{1}{a} & b < y < a+b & \text{Si } \varphi(x) = (b-a)x + a \text{ et } a < b: \\ 0 & y \geq a+b & U(0,1) \rightarrow U(a, b) \end{cases}$$

- Exemple 2 : la loi de X est la loi normale $\mathcal{N}(0;1)$. Nous pouvons retrouver la loi d'une $\mathcal{N}(\mu, \sigma)$ avec $Y = \sigma X + \mu$. De manière générale par application d'une fonction affine nous obtenons une loi de même type.

3) Cas où $\varphi(x) = x^2$: $Y = X^2$ et $y > 0$

$$F_Y(y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_Y(\sqrt{y}) - F_Y(-\sqrt{y})$$

et $f_Y(y) = \frac{1}{2\sqrt{y}}(f_X(\sqrt{y}) + f_X(-\sqrt{y}))$ si $y > 0$ et zéro sinon.

4) Cas où $g(x)=e^x$: $Y=e^X$ et $y>0$

$$F_Y(y)=P(e^X \leq y)=P(X \leq \ln y)=F_X(\ln y)$$

$$\text{et } f_Y(y)=\frac{1}{y}f_X(\ln y) \text{ si } y>0 \text{ et zéro sinon.}$$

D. Simulation numérique

Nous simulons à l'aide d'ordinateurs des lois de probabilités continues et discrètes. Pour cela nous utilisons des lois uniformes créées par des algorithmes de génération de nombres aléatoires :

Lois uniformes continues $U(0,1)$: Ran# sur calculatrice, ALEA() sur LibreOffice, etc.

Lois uniformes discrètes $U(i,j)$: par exemple, rand(i,j) en langage PHP. $\text{rand}(0,999)/1000$ simule une loi uniforme continue discrétisée au millième.

- Méthode d'inversion : détermination de la loi de X en fonction de celle de U par inversion de la fonction de répartition. Avec F strictement croissante : $X=F^{-1}(U)$.

Cas d'une loi exponentielle : $f_X(x)=\lambda e^{-\lambda x}$ pour $x>0$ et zéro sinon, d'où $F_X(x)=\int_{-\infty}^x f_X(x)dx=1-e^{-\lambda x}=y$ pour

$x > 0$ et zéro sinon. Ainsi $x = -\ln \frac{(1-y)}{\lambda}$ et finalement nous simplifions, sachant que $1-U$ et U ont la même loi.

Simulation d'une loi exponentielle : $X = -\frac{\ln U}{\lambda}$

- Simulation de deux lois normales $\mathcal{N}(0;1)$ indépendantes X_1 et X_2 à partir de U_1 et U_2 indépendantes :

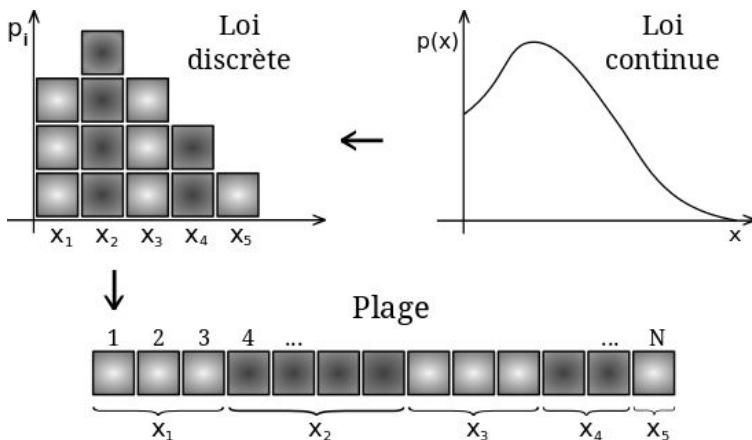
$$X_1 = \sqrt{-2 \ln U_1} \cos(2\pi U_2)$$

$$X_2 = \sqrt{-2 \ln U_1} \sin(2\pi U_2)$$

C'est la méthode de Box-Müller. Pour une gaussienne il n'y a pas de formulation directe de la fonction de répartition F . Ainsi F^{-1} n'a pas d'expression analytique simple.

- Méthode pour une loi X quelconque :

Si nous avons une loi continue nous discrétisons x par classes. Nous avons alors un histogramme et chaque barre de hauteur p_i est aussi discrétisée en unités. Une réalisation de la loi revient alors à tirer au hasard et de manière équitable une unité de l'histogramme. Nous obtenons en mettant les barres bout à bout une plage complète des unités. Nous utilisons une loi uniforme discrète $U(1,N)$ où N correspond au nombre total d'unités de l'histogramme. La valeur générée est comparée à sa position sur la plage et nous en déduisons une valeur pour la réalisation x_i .



- Il existe de nombreuses autres méthodes qui utilisent les différentes propriétés des lois de probabilité. Par exemple, en simulant des lois de Bernoulli²⁰, nous obtenons, par somme, une loi binomiale qui elle-même permet de simuler une loi normale.

20 Loi à deux issues : $p=P(S)$ et $q=1-p=P(E)$.

E. Exercices

Exercice 1 : Loi binomiale corrigé p272

Vérifier qu'il s'agit d'une loi de probabilité. Déterminer l'espérance et la variance d'une loi binomiale en fonction de n et p .

Exercice 2 : Somme de binomiales corrigé p273

Montrer que la somme de loi binomiales indépendantes de même paramètres p est elle-même une loi binomiale dont on déterminera les paramètres.

Exercice 3 : Loi géométrique corrigé p273

Montrer qu'il s'agit bien d'une loi de probabilité.
Déterminer l'espérance et la variance d'une loi géométrique en fonction de p .

Montrer que la somme de deux variables aléatoires géométriques indépendantes de paramètre p est une loi binomiale négative $\mathcal{NB}(2,p)$.

Exercice 4 : Premiers succès corrigé p274

1- Soit une pièce équilibrée. Qu'elle est la probabilité que le premier Pile apparaisse au cinquième lancer ? Sachant que le premier Pile n'aït toujours pas apparu au troisième lancer, qu'elle est la probabilité qu'il apparaissent pour la première fois au huitième lancer ?

2- Soit un dé équilibré. En moyenne après combien de lancers apparaît le premier six ? Qu'elle est la probabilité que le premier six apparaisse au cours des six premiers lancers ?

Exercice 5 : Loi de Poisson corrigé p274

Montrer qu'il s'agit bien d'une loi de probabilité.
Déterminer l'espérance et la variance d'une loi Poisson
en fonction du paramètre λ .
Montrer la loi de la somme.

Exercice 6 : Loi uniforme corrigé p275

Retrouver l'expression de la variance pour une loi
uniforme continue $\mathcal{U}(a,b)$.

Exercice 7 : Loi exponentielle corrigé p276

Montrer qu'il s'agit bien d'une loi de probabilité.
Déterminer l'espérance et la variance d'une loi
exponentielle en fonction du paramètre λ .
Trouver la loi de la somme.

Exercice 8 : Somme de gaussiennes corrigé p277

- 1- Trouver la loi de la somme de deux $\mathcal{N}(0,1)$.
- 2- Trouver la loi de la somme dans le cas général.

Exercice 9 : Premières Students corrigé p278

- 1- Donner les expressions des premières fonctions de Student.
- 2- Donner le degré du polynôme au dénominateur, la variance, le coefficient d'aplatissement et la valeur en zéro.
- 3- Donner l'expression de la Student pour $k=9$, centrée et avec une variance de $14/5$.

Exercice 10 : Loi de Student corrigé p278

Quelque soit k , montrer que la loi de Student correspond bien a une loi de probabilité.

On pourra poser l'intégrale $I_k = \int_{-\infty}^{+\infty} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}} dx$ et

effectuer le changement de variable $u = \frac{x}{\sqrt{k}}$.

Exercice 11 : Variance d'une Student corrigé p280

Déterminer la variance d'une la loi de Student.

Exercice 12 : Somme de Students corrigé p281

Sur un exemple montrer que la somme de deux variables aléatoires de Student indépendantes n'est pas une variable aléatoire de Student.

Exercice 13 : Lois du Khi-Deux corrigé p283

Donner les expressions des premières fonctions du Khi-Deux pour $k=1$, $k=2$, $k=3$ puis $k=10$.

Exercice 14 : Loi du produit corrigé p283

Soient deux lois X et Y indépendantes.

1. Proposer une méthode générale permettant de déterminer la loi de probabilité de $Z=XY$.
2. Nous considérons le cas où X et Y sont des lois uniformes continues $\mathcal{U}(1;2)$ indépendantes.
 - a. Déterminer l'expression analytique de la loi de Z.
 - b. Retrouver l'allure de la loi de probabilité de Z avec une simulation du produit sur tableur pour $n=10\,000$.

Exercice 15 : Loi de la somme de lois exponentielles corrigé p284

Soient n lois exponentielles X_i de même paramètre λ et indépendantes. Nous appelons S_n la loi de la somme : $S_n = X_1 + X_2 + \dots + X_n$. Nous avons aussi la loi : $M_n = S_n/n$.

1. Déterminer la loi de probabilité de S_2 .
2. Déterminer la loi de probabilité de S_n .
3. Déterminer la loi de probabilité de M_n .

Exercice 16 : Loi de l'inverse corrigé p285

Soit X une loi de probabilité définie strictement positive.

1. Déterminer la loi de $Y=1/X$.
2. a. Quelle la loi de T_n inverse de M_n défini à l'exercice précédent ?
b. Quelle la loi de l'inverse de la loi de Cauchy :

$$\text{loi de } X : f_x = \frac{1}{\pi} \frac{1}{1+x^2} \text{ pour tout } x.$$



123

IV. ESTIMATEURS

Nous avons une loi de probabilité X qui dépend d'un certain nombre de paramètres.

À l'aide d'un échantillon (X_1, X_2, \dots, X_n) de variables aléatoires mutuellement indépendantes et qui suivent la loi de X , nous voulons disposer d'une méthode pour estimer au mieux les différents paramètres qui définissent notre loi (statistique inférentielle). Notre échantillon est de taille n . Nos paramètres sont de manière générique désignés par la lettre θ et nous nommons T_n notre estimateur de θ .

Nous effectuons dans un premier temps une estimation ponctuelle, valeur la plus probable du paramètre, dans un second temps nous déterminerons une estimation par intervalle.

A. Qualité d'un estimateur

1) Biais

Nous appelons biais de T_n l'espérance $E(T_n - \theta)$ d'où, par linéarité de l'espérance :

$$b_{T_n}(\theta) = E(T_n) - \theta$$

L'idéal est d'avoir un estimateur sans biais ou sinon un biais qui tend vers zéro quand la taille de l'échantillon tend vers l'infini.

2) Risque

Nous appelons risque quadratique de T_n l'espérance $E[(T_n - \theta)^2]$ et on montre que :

$$r_{T_n}(\theta) = V(T_n) + b^2$$

En effet $E[(T_n - \theta)^2] = E(T_n^2) - 2\theta E(T_n) + \theta^2$ (par linéarité de l'espérance), et finalement après simplifications de $r = V(T_n) + E(T_n)^2 - 2\theta(b + \theta) + \theta^2$ nous retrouvons l'expression du risque. À moins d'un estimateur non biaisé le risque n'est pas simplement égal à la variance de l'estimateur. On a intérêt à choisir un estimateur dont le risque tend vers zéro avec la taille de l'échantillon et plus la convergence est rapide, meilleur est l'estimateur.

- Exemple 1 : Nous appelons \bar{X}_n la moyenne empirique associée à l'échantillon (X_1, X_2, \dots, X_n) définie par :

$$T_n = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Montrez que \bar{X}_n est un estimateur de l'espérance $\theta = m = E(X)$. Déterminer la qualité de cette estimateur.

$E(T_n) = 1/n \sum E(X_i)$ par linéarité de l'espérance, d'où $E(T_n) = 1/n \cdot n m = m$ et $b = 0$.

$V(T_n) = 1/n^2 \sum V(X_i)$ (propriété de la variance et indépendance des variables), d'où $r = 1/n^2 \cdot n \sigma^2$ et $r = \sigma^2/n$.

La moyenne empirique est un bon estimateur de l'espérance d'une variable aléatoire. L'estimateur est de

biais nul et le risque tend vers zéro quand n tend vers l'infini.

- Exemple 2 : Nous cherchons un estimateur de la variance d'une variable aléatoire, nous proposons les deux estimateurs suivants :

$$S_n^2 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n} \quad \text{et} \quad R_n^2 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n-1}$$

Lequel, selon vous, est le meilleur estimateur de σ^2 ?

$$S_n^2 = 1/n \sum (X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2) \quad \text{et}$$

$$S_n^2 = 1/n \sum X_i^2 - 2T_n \sum X_i/n + T_n^2 = 1/n \sum X_i^2 - T_n^2$$

$$\text{D'où } E(S_n^2) = E(X^2) - E(T_n^2) = \sigma^2 + m^2 - V(T_n) - E(T_n)^2$$

$$\text{Et } E(S_n^2) = \sigma^2 - \frac{\sigma^2}{n} \text{ soit } E(S_n^2) = \frac{n-1}{n} \sigma^2 \text{ et } b_{S_n^2} = -\frac{\sigma^2}{n}.$$

Par un calcul tout à fait analogue nous trouvons $b_{R_n^2} = 0$.

Les deux quantités sont des estimateurs de la variance car les biais sont, soit nul, soit asymptotiquement non biaisé.

R_n^2 est le meilleur estimateur car il n'a pas de biais.

B. Construction d'estimateurs

Nous voulons disposer d'une méthode générale permettant de trouver les estimateurs adéquats pour les paramètres d'une loi.

1) Méthode des moments

Nous identifions les moments de la population avec ceux de l'échantillon. Nous considérons autant de moments que nous avons de paramètres en commençant par le moment d'ordre 1. Comme nous le verrons sur des exemples cette méthode nous fournit des estimateurs mais ne nous garantit pas que ce sont les meilleurs en terme de biais et de risque quadratique.

Moments de la loi : $m_k = E(X^k)$

Moments de l'échantillon : $\overline{X_n^k} = \frac{1}{n} \sum_{i=1}^n X_i^k$

Nous pouvons aussi considérer les moments centrés ou même centrées réduits, l'approche reste la même. Le moment du n-échantillon d'ordre 1 est la moyenne empirique et correspond bien à un excellent estimateur comme montré précédemment. Par contre le moment centré d'ordre 2 a un biais, comme nous l'avons montré avant il faudrait diviser par $n-1$ au lieu de n pour avoir un biais nul.

- Exemple 1 : Nous disposons d'un damier de 100 cases et de 200 graines. Chaque seconde nous plaçons aléatoirement une graine dans une des cases. À la fin de l'expérience nous comptons le nombre de graines dans chaque case et nous dénombrons le nombre de cases qui ne contiennent aucune graine, puis une graine, puis deux et ainsi de suite. Soit la loi

X du nombre de graines par case. Nous obtenons la distribution suivante :

k	0	1	2	3	4	5	6	7	8
n	12	28	34	10	8	7	0	1	0

Nous supposons que cette distribution suit une loi de Poisson, en déduire la valeur du paramètre λ de la loi.

Le paramètre de cette loi est égal à l'espérance donc d'après le théorème de moments, λ est estimé par la moyenne

$$\text{empirique : } \lambda = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{k=1}^p n_k x_k \text{ soit } \lambda = 2^{21}.$$

Résultat cohérent avec notre modèle, en effet nommons n le nombre de cases et N le nombre de graines. Nous avons une distribution aléatoire uniforme. Nous pourrions par exemple utiliser deux dés équilibrés dix faces de couleurs différentes pour la position horizontale et verticale. La

21 En utilisant le moment d'ordre 2 nous trouverions $\lambda = 2,06$.

fréquence de l'événement par seconde pour chaque case vaut $1/n$. Pour une loi de Poisson à tout instant l'événement peut se produire, ici le temps est discrétilisé, néanmoins l'approximation d'un temps continu est correct car une seconde est une petite durée au regard de celle de 200 secondes de l'expérience. Nous pouvons donc prendre $\lambda = N/n$. Nous tendons vers une loi de Poisson quand le nombre des cases et de graines tendent vers l'infini.

- Exemple 2 : Nous supposons que la durée de vie d'un verre de table suit une loi exponentielle. Nous observons les différentes durées de vie en mois suivantes : 7, 5, 37, 35, 17, 9, 6, 13, 4 et 8 mois. Déterminez le paramètre λ de la loi.

Pour un loi exponentielle $E(T) = \frac{1}{\lambda}$ d'où en utilisant le moment d'ordre 1 : $\lambda = \frac{n}{\sum_{i=1}^n t_i}$ et $\lambda = \frac{10}{141} \simeq 0,071$.

Nous avons ainsi une estimation ponctuelle du paramètre λ et l'espérance de vie vaut environ 14 mois.

Nous démontrerons en exercice que cet estimateur T_n possède un biais : $E(T_n) = \frac{n}{n-1}\lambda$ et $b_{T_n} = \frac{\lambda}{n-1}$.

Ce biais est asymptotiquement nul et son risque tend vers zéro : $r_{T_n} = \lambda^2 \frac{(n+2)}{(n-1)(n-2)}$.

Nous construisons à partir de cet estimateur un nouvel estimateur W_n sans biais : $W_n = \frac{n-1}{n} T_n = \frac{n-1}{\sum X_i}$.

On montre : $E(W_n) = \lambda$ et $r_{W_n} = V(W_n) = \lambda^2/(n-2)$.

Nous avons un meilleur estimateur : biais nul et risque plus faible. Une nouvelle estimation du paramètre donne $\lambda \approx 0,064$ et l'espérance de vie vaut environ 16 mois.

- Exemple 3 : Nous considérons que la masse des pommes produites par un arbre suivent une loi normale. Nous prélevons au hasard et mesurons les masses suivantes en grammes :

158	131	146	158	125	153	166	121
127	123	195	149	124	153	123	129

Déterminer la moyenne et l'écart-type de la loi.

$$\mu = \frac{1}{n} \sum_{i=1}^n m_i \approx 142,5, \quad \sigma^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^n m_i^2 \text{ et } \sigma \approx 20,2.$$

- Exemple 4 : Les planches produites par une scierie ont une longueur qui suivent une loi uniforme $\mathcal{U}(a, b)$.

Nous mesurons les longueurs en millimètres de 8 planches tirées au hasard : 2017, 1987, 2018, 2014, 2003, 1985, 2013 et 1981. Estimez a et b.

$$\text{Nous posons } E(X) = \frac{a+b}{2} = \overline{X}_n^1$$

$$\text{et } E(X^2) = \frac{(b-a)^2}{12} + \frac{(a+b)^2}{4} = \frac{a^2 + ab + b^2}{3} = \overline{X}_n^2$$

$$\text{d'où } a+b = 2\overline{X}_n^1 \text{ et } ab = 4\overline{X}_n^1 - 3\overline{X}_n^2$$

$$\text{Finalement : } a \approx 1977 \text{ et } b \approx 2028$$

2) Méthode du maximum de vraisemblance

Soit un variable aléatoire X discrète ou continue dont nous voulons estimer un paramètre θ à partir d'un ensemble d'observations $\{x_i\}$.

Nous définissons une fonction $f(x, \theta)$ telle que :

$$f(x, \theta) = \begin{cases} P_\theta(X=x) & \text{pour une variable discrète} \\ & \text{ou} \\ f_\theta(x) & \text{pour une variable continue} \end{cases}$$

La vraisemblance V est ainsi définie :

$$V(x_1, x_2, \dots, x_i, \dots, x_n, \theta) = f(x_1, \theta) \times f(x_2, \theta) \times \dots \times f(x_n, \theta)$$

notée aussi :

$$V(x_1, x_2, \dots, x_i, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta)$$

La valeur du paramètre pour laquelle le produit des probabilités, ou densités de probabilités, prises aux différents points de l'échantillon est maximale, est considérée comme la valeur la plus vraisemblable.

Le principe est simple et la méthode est facile à mettre en œuvre :

$$\frac{\partial V(x_i, \theta)}{\partial \theta} = 0 \quad \text{et} \quad \frac{\partial^2 V(x_i, \theta)}{\partial \theta^2} < 0$$

La nullité de la dérivée permet d'obtenir notre estimateur : relation entre θ et les x_i .

La fonction logarithme étant strictement croissante nous pouvons aussi travailler avec $\ln V$ pour des calculs plus simples.

- Exemple 1 : Reprenons le cas d'une loi de Poisson et déterminons l'estimateur de λ par la méthode du maximum de vraisemblance.

$$P_\lambda(X=x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{et} \quad V = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = e^{-n\lambda} \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!}$$

$$\ln V = -n\lambda + \sum_{i=1}^n \ln \frac{\lambda^{x_i}}{x_i!} \quad \text{et} \quad \frac{\partial \ln V}{\partial \lambda} = -n + \sum_{i=1}^n \frac{x_i}{\lambda} = 0$$

$$\text{D'où} \quad \lambda = \frac{1}{n} \sum_{i=1}^n x_i$$

Nous trouvons le même estimateur que par la méthode précédente.

- Exemple 2 : Reprenons le cas d'une loi exponentielle et déterminons l'estimateur de λ par la méthode du maximum de vraisemblance.

$$f_\lambda(t) = \lambda e^{-\lambda t} \quad \text{et} \quad V = \prod_{i=1}^n \lambda e^{-\lambda t_i} = \lambda^n e^{-\lambda \sum_{i=1}^n t_i}$$

$$\ln V = n \ln \lambda - \lambda \sum_{i=1}^n t_i \quad \text{et} \quad \frac{\partial \ln V}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^n t_i$$

$$\text{D'où} \quad \lambda = \frac{n}{\sum_{i=1}^n t_i}$$

Le même estimateur que par la méthode précédente.

- Exemple 3 : Reprenons le cas d'une loi normale et déterminons les estimateurs de μ et σ par la méthode du maximum de vraisemblance.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{et} \quad V = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

$$\text{d'où} \quad \ln V = -n \ln \sigma - \frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^n \left(\frac{x_i-\mu}{\sigma}\right)^2$$

$$\frac{\partial \ln V}{\partial \mu} = \sum_{i=1}^n \frac{x_i - \mu}{\sigma^2} = 0 \quad \text{et donc} \quad \mu = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\frac{\partial \ln V}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2 = 0 \quad \text{et} \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

Nous obtenons les mêmes estimateurs que par le théorème des moments et nous avons à nouveau un estimateur de la variance biaisé.

- Exemple 4 : Reprenons le cas de la loi uniforme.

$$V(x_i, a, b) = \prod_{i=1}^n f(x_i, a, b) = \begin{cases} \frac{1}{(b-a)^n} & \text{si } \{x_i\} \subset [a, b] \\ 0 & \text{sinon} \end{cases}$$

Pour a fixé, plus b est petit plus la vraisemblance est grande donc $b = \max(\{x_i\})$.

Pour b fixé, plus a est grand plus la vraisemblance est grande donc $a = \min(\{x_i\})$.

Nous avons ici des estimateurs bien différents que par la méthode des moments, pour l'exemple des planches nous avons les estimations suivantes : $a = 1981$ et $b = 2018$.

Ces estimateurs sont biaisés, par exemple b est nécessairement inférieur à la valeur théorique, mais contrairement à la méthode des moments nous sommes assuré que les x_i appartiennent à $[a, b]$.

C. Estimation par intervalle

Nous avons maintenant des outils efficaces pour déterminer la valeur des paramètres d'une loi. Nous avons déterminé des valeurs ponctuelles et à présent nous voulons déterminer un intervalle de confiance.

Le théorème central limite permet à partir d'un échantillon de grande taille d'estimer la moyenne d'une loi de probabilité avec un intervalle de confiance.

Mais qu'en est-il si nous voulons déterminer d'autres paramètres différents de la moyenne ? Par exemple, quelles sont les incertitudes sur le paramètre λ d'une loi exponentielle, ou les bornes a et b pour une loi uniforme ?

Nous considérons un estimateur non biaisé et si l'estimateur est biaisé nous en créons un nouveau en retirant le biais.

Nous utilisons trois méthodes différentes. La méthode par intégration qui exige de déterminer la loi de probabilité complète de l'estimateur. Une seconde méthode par encadrement, plus simple, qui surestime notre incertitude, mais ne requiert que la connaissance de la variance de l'estimateur. Et finalement une troisième par simulation numérique.

L'inégalité de Bienaymé-Tchebychev fournit l'encadrement de la deuxième méthode. Elle est universelle car elle s'applique à toute loi X possédant une espérance et une variance. Nous devons cependant connaître la variance de notre estimateur et elle fournit un majorant assez grossier en utilisant la probabilité suivante :

$$\forall \epsilon > 0, \quad P(|X - E(X)| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2}$$

- Exemple 1 : Considérons à nouveau le damier et la loi de Poisson de paramètre λ . Ce paramètre est estimé par la moyenne et nous pouvons donc, dans le cas des grands nombres, utiliser le théorème central limite :

$$\lambda = \lambda_m \pm t_\infty \sigma / \sqrt{n}$$

Estimons la variance :

k	0	1	2	3	4	5	6	7	8
$k - \lambda_m$	-2	-1	0	1	2	3	4	5	6
$(k - \lambda_m)^2$	4	1	0	1	4	9	16	25	36
n	12	28	34	10	8	7	0	1	0

$$\sigma_\lambda^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad \sigma_\lambda \approx 1,44 \text{ et } \lambda = 2,0 \pm 0,3 \text{ à 95%}.$$

Ici la taille de l'échantillon n'est pas suffisante pour cette méthode mais nous avons illustré la méthode générale pour de grands échantillons. Nous avons probablement sous-estimé la largeur de l'intervalle.

- Deuxième exemple : Le temps d'attente d'un train suit une loi uniforme $\mathcal{U}(0, a)$. Nous observons les temps d'attente suivants 3, 12, 7, 17, 8, 14, 2, 5, 10, 14, 15 et 11 minutes. Que vaut a ?

Mais quel est le biais, la variance de l'estimateur T_n et l'incertitude sur a à 90% de confiance ?

a) Utilisons tout d'abord la méthode du maximum de vraisemblance qui nous fournit la loi du max pour connaître la loi de probabilité de T_n : $T_n = \max(\{X_i\})$.

a est donc ici estimé à 17 minutes.

Déterminons la loi de l'estimateur, loi du max :

$$P(T_n \leq x) = P([X_1 \leq x] \cap [X_2 \leq x] \cap \dots \cap [X_n \leq x])$$

or les variables sont indépendantes :

$$\text{d'où } P(T_n \leq x) = \prod_{i=1}^n P(X_i \leq x)$$

Avec les fonctions de répartition : $F_X(x) = P(X \leq x)$

$$F_{T_n}(x) = \prod_{i=1}^n F_{X_i}(x) = \left(\frac{x}{a}\right)^n \text{ pour } 0 \leq x \leq a$$

$$F_{T_n}(x) = 0 \text{ si } x < 0, \quad F_{T_n}(x) = 1 \text{ si } x > a$$

$$\text{d'où la loi de probabilité de } T_n : f_{T_n}(x) = \frac{dF_{T_n}(x)}{dx} = n \frac{x^{n-1}}{a^n}$$

$$\text{pour } 0 \leq x \leq a \text{ et } f_{T_n}(x) = 0 \text{ sinon.}$$

$$\text{Espérance : } E(T_n) = \int x f_{T_n}(x) dx = \frac{n}{a^n} \int_0^a x^n dx = \frac{n}{n+1} a$$

$$\text{l'estimateur } T_n \text{ est biaisé : } b_{T_n} = \frac{n}{n+1} a - a = -\frac{a}{n+1}$$

a est sous-estimé et le biais est asymptotiquement nul.

$$\text{Variance : } E(X^2) = \int x^2 f_{T_n}(x) dx = \frac{n}{a^n} \int_0^a x^{n+1} dx = \frac{n}{n+2} a^2$$

$$V(T_n) = E(X^2) - E(X)^2 = \frac{n}{n+2}a^2 - \frac{n^2}{(n+1)^2}a^2 = \frac{na^2}{(n+2)(n+1)^2}$$

Il est préférable de prendre un estimateur non biaisé, pour cela nous retirons le biais et nous obtenons W_n :

$$W_n = \frac{n+1}{n}T_n \quad \text{avec } b_{W_n} = 0 \quad \text{et} \quad V(W_n) = \frac{a^2}{n(n+2)}$$

Nouvelle estimation de a avec W_n : $a \approx 18,4$ minutes.

Pour l'incertitude Δa nous allons comparer trois méthodes.

-> Déterminons un intervalle de confiance grâce à l'inégalité de Bienaymé-Tchebychev :

Nous appliquons l'inégalité à W_n :

$$P(|W_n - a| \geq \epsilon) \leq \frac{V(W_n)}{\epsilon^2}$$

$$\text{or } P(|W_n - a| \leq \epsilon) = P(-\epsilon \leq W_n - a \leq \epsilon)$$

$$\text{et } P(\epsilon \geq a - W_n \geq -\epsilon) = P(U_n + \epsilon \geq a \geq W_n - \epsilon)$$

Puis en posant $\epsilon = \sqrt{\frac{V(W_n)}{\alpha}}$ nous avons :

$$1 - P(W_n - \epsilon \leq a \leq W_n + \epsilon) \leq \alpha$$

Et finalement :

$$P(W_n - \sqrt{\frac{V(W_n)}{\alpha}} \leq a \leq W_n + \sqrt{\frac{V(W_n)}{\alpha}}) \geq 1 - \alpha$$

Pour un confiance de 90%, $\alpha=0,1$ et pour notre échantillon

$$V(W_{12}) = \frac{18,4^2}{12 \times 14} \simeq 2,0 \text{ et } \Delta a = \sqrt{\frac{V(W_n)}{\alpha}} \simeq 4,5.$$

D'où $13,9 \leq a \leq 22,9$ et $a \simeq 18,4 \pm 4,5$ minutes.

-> Déterminons un intervalle de confiance avec un calcul d'intégrales sur la densité de probabilité de l'estimateur.

Nous déterminons la densité de probabilité de W_n :

$$P(W_n \leq x) = P\left(\frac{n}{n+1} W_n \leq \frac{n}{n+1} x\right) = P\left(T_n \leq \frac{n}{n+1} x\right)$$

$$F_{W_n}(x) = F_{T_n}\left(\frac{n}{n+1} x\right) = \left(\frac{n}{n+1} \frac{x}{a}\right)^n \text{ pour } 0 \leq x \leq \frac{n+1}{n} a.$$

$$f_{W_n}(x) = \left(\frac{n}{a(n+1)}\right)^n n x^{n-1} \text{ pour } 0 \leq x \leq \frac{n+1}{n} a$$

$$\text{et } f_{W_n}(x) = 0 \text{ sinon.}$$

Nous avons une loi de probabilité dissymétrique dont le maximum correspond à la borne supérieure. Pour une confiance de 90% nous retirons les 10% de la queue de distribution gauche de l'intervalle de confiance. Ainsi nous définissons a_{max} et a_{min} tels que :

$$\text{Borne supérieure : } a_{max} = \frac{n+1}{n} a$$

soit $a_{max} = \left(\frac{13}{12}\right)^2 \times 17 \simeq 19,95$ et $a_{max} \simeq 20,0$ minutes.

Borne inférieure : $\int_{a_{min}}^{a_{max}} f_{W_n}(x) dx = 1 - \alpha$

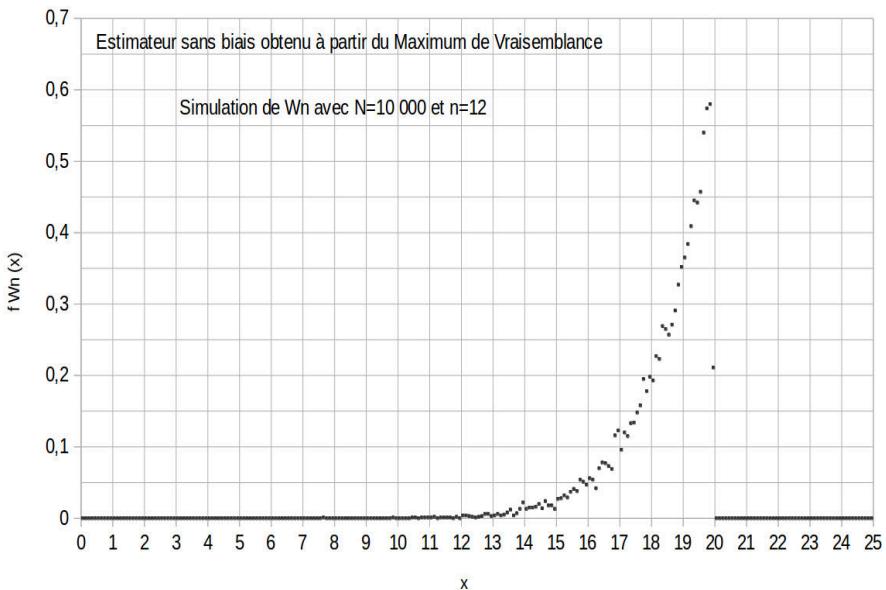
par calcul numérique avec $a \simeq 18,4$ et $n=12$,

$\int_{a_{min}}^{20} f_{W_{12}}(x) dx = 0,9$ pour $a_{min} \simeq 16,4$ minutes.

En conclusion $a \simeq 18,4 \begin{matrix} +1,6 \\ -2,0 \end{matrix}$ minutes à 90%.

L'intervalle trouvé est dissymétrique. Nous avons un encadrement inclus dans celui de l'inégalité de Bienaymé-Tchebychev et nous avons ici une estimation plus précise de a .

-> Effectuons maintenant une simulation numérique sur un tableur. La fonction ALEA() du tableur fournit des réels aléatoirement et uniformément répartis entre 0 et 1. La loi des $X_i = U_i(0, a)$ est alors obtenue en multipliant par l'estimation ponctuelle de a . Nous générerons 10 000 échantillons de taille 12. Nous plaçons le maximum de chacun de ces échantillons sur un graphe et nous avons ainsi la distribution d'échantillonnage de T_n :



Nous trouvons ainsi les mêmes résultats que par la méthode précédente (fichier : www.incertitudes.fr/livre/Train.ods).

a) Utilisons maintenant l'estimateur du théorème des moments : $T_n = 2\bar{X}_n$. Ici, pour n grand, nous pouvons utiliser le théorème central limite, car une fonction affine d'une loi donne une nouvelle loi de même forme. Nous estimons la moyenne avec la loi de Gauss :

$$\bar{x} = \frac{3+12+\dots+11}{12} \approx 9,83 \text{ et } s = \sqrt{\frac{(3-\bar{x})^2 + \dots}{11}} \approx 4,88$$

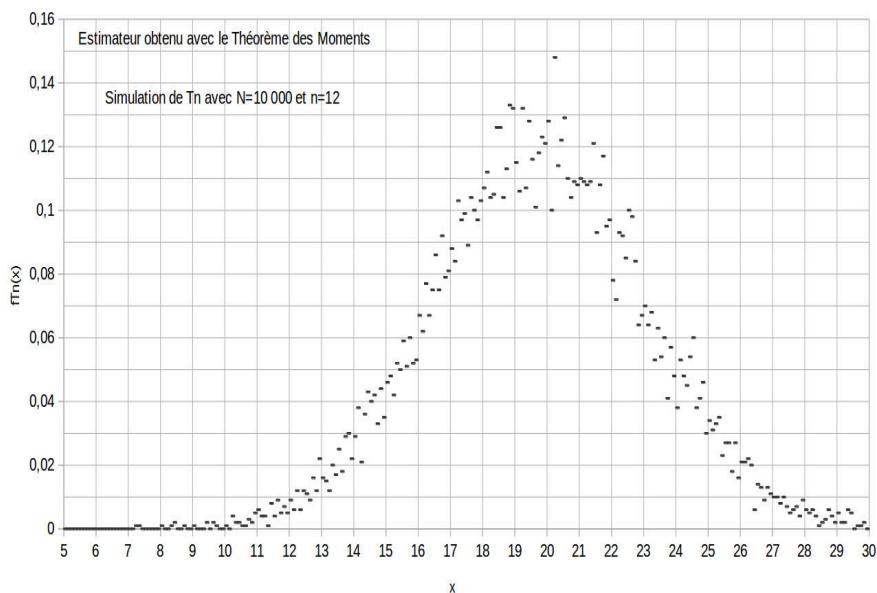
$$m = \bar{x} \pm t_{\alpha/2} s / \sqrt{n} = 9,8 \pm 2,3 \text{ à } 90\%$$

$$\text{D'où : } a = 19,7 \pm 4,6 \text{ minutes à } 90\%.$$

Dans ce cas n vaut seulement 12 et la loi de la population n'est pas normale, nous voulons néanmoins montrer

comment on procéderait pour n grand. Comme nous le constaterons avec la simulation numérique, l'intervalle est ainsi sous-estimé.

En effet en réalisant une simulation numérique avec cet estimateur, $a = 19,7 \pm 5,5$ minutes à 90% :

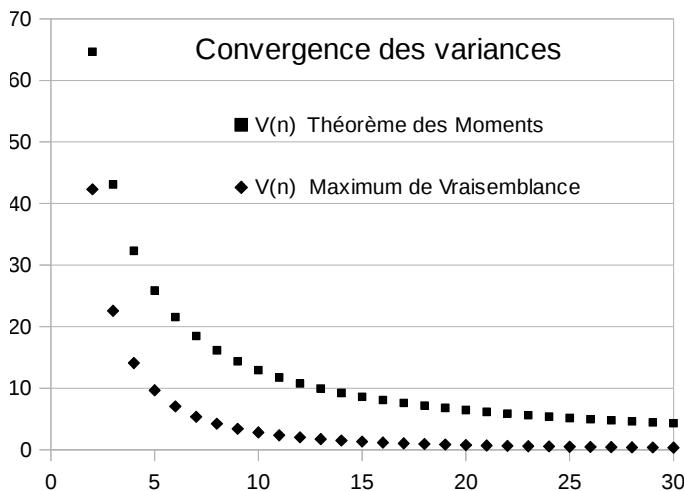


Comparaison des intervalles à 90% suivant les méthodes :

Maximum de Vraisemblance	Bienaymé-Tchebychev	$13,9$	$18,4$	$22,9$
	Calcul d'intégrales et Simulation	$16,4$	$18,4$	$20,0$
Théorème des Moments	<i>Théorème Central Limite (échantillon trop petit)</i>	$15,1$	$19,7$	$24,3$
	Simulation	$14,2$	$19,7$	$25,2$

En conclusion, l'estimateur du maximum de vraisemblance converge bien plus rapidement que celui du théorème des moments et nous préférions cette première méthode. La variance converge en $1/n^2$ au lieu de $1/n$:

$$[V(W_n)]_{MV} = \frac{a^2}{n(n+2)} \quad \text{et} \quad [V(T_n)]_{TM} = \frac{a^2}{3n}$$



D. Exercices

Exercice 1 : Estimateurs de la moyenne

corrigé p286

Soit l'échantillon (X_1, X_2, X_3). X_1, X_2 et X_3 sont trois variables indépendantes, de même loi, d'espérance m et de variance σ^2 .

Comparez les trois estimateurs suivants proposés pour estimer la moyenne m de l'échantillon :

$$A_3 = (X_1 + X_2 + X_3)/3, B_3 = (X_1 + 2X_2 + 3X_3)/6 \text{ et}$$
$$C_3 = (X_1 + 2X_2 + X_3)/3.$$

Exercice 2 : Faisceau homocinétique *corrigé p286*

Soit un faisceau homocinétique d'atomes de carbones ionisés C^+ . Nous mesurons la quantité de mouvement et l'énergie cinétique de 100 atomes du faisceau. Le faisceau est considéré parfaitement unidirectionnel.

La norme de la quantité de mouvement et l'énergie cinétique totale s'écrivent et valent :

$$p = \sum_{i=1}^{100} m v_i = 2,418 \cdot 10^{-21} \text{ kg.m/s et}$$

$$E_c = \sum_{i=1}^{100} \frac{1}{2} m v_i^2 = 1,518 \cdot 10^{-18} \text{ J avec } m = 1,993 \cdot 10^{-26} \text{ kg.}$$

Soit V la loi de probabilité de la vitesse des ions. Déterminer, grâce à l'échantillon prélevé et les estimateurs adéquats, la vitesse moyenne v_m , sa variance σ_v^2 et l'incertitude Δv à 95%.

Exercice 3 : Deux estimateurs corrigé p287

Soit X la variable aléatoire discrète suivante :

Valeurs	0	1	2
Probabilités	3θ	θ	$1 - 4\theta$

1. Quelles valeurs de θ définissent une loi de probabilité ?
2. Calculer $E(X)$ et $V(X)$.
3. Soit un échantillon (X_1, X_2, \dots, X_n) de X.

Nous avons $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ et $T_n = a \bar{X}_n + b$.

Déterminer a et b pour que T_n soit un estimateur sans biais de θ .

Déterminer $V(T_n)$.

4. Soit la variable aléatoire Y_i définie pour tout $i \in \llbracket 0 ; n \rrbracket$ par $Y_i = 1$ si $X_i = 1$ et $Y_i = 0$ sinon.

Soit $Z_n = \sum_{i=1}^n Y_i$. Déterminer $E(Z_n)$.

Montrer que $U_n = \frac{Z_n}{n}$ est un estimateur non biaisé de θ . Déterminer $V(U_n)$.

5. Nous effectuons des estimations de θ avec les réalisations suivantes :

Valeurs	0	1	2
Effectifs	31	12	57

Estimer θ . Quel estimateur préférez-vous ?

Exercice 4 : Urnes corrigé p287

Soient deux urnes identiques qui contiennent la même proportion p de boules noires. Dans la première urne nous tirons avec remise un échantillon de taille n_1 et nous notons P_1 la proportion de boules noires de cet échantillon. Nous réalisons la même expérience pour la deuxième urne.

$$\text{Nous définissons } T = \frac{P_1 + P_2}{2} \text{ et } U = x P_1 + (1-x) P_2 \\ \text{avec } x \in]0;1[$$

Montrer que T et U sont deux estimateurs de p .

Lequel est le meilleur ?

Déterminer la valeur de x permettant d'avoir l'estimateur optimum.

Exercice 5 : Variable à densité corrigé p288

Soit la variable aléatoire X définie par la densité de probabilité suivante :

Où a est le paramètre que nous voulons estimer ($a > 1$).

$$f(x) = \begin{cases} \frac{a}{x^{a+1}} & \text{si } x > 1 \\ 0 & \text{sinon} \end{cases}$$

1. Montrer que f définit une loi de probabilité.
2. Calculer $E(X)$.
3. Déterminer des estimateurs de a par la méthode du maximum de vraisemblance et par la méthode des moments.
4. Donner des estimations ponctuelles de a pour les observations suivantes :
1,16 / 1,80 / 1,04 / 3,40 / 1,22 / 1,06 / 1,35 / 1,10.
5. Peut-on effectuer le calcul de $V(X)$?
6. Effectuer une simulation numérique de la loi de X . Que pouvons nous conjecturer quant aux biais et convergences des estimateurs trouvés ?

Exercice 6 : Densité linéaire corrigé p291

Soit la variable aléatoire X définie par la densité de

probabilité suivante : $f(x) = \begin{cases} ax+b & si \ 0 \leq x \leq 1 \\ 0 & sinon \end{cases}$

1. Exprimer b en fonction de a de manière à ce que f définisse une loi de probabilité.
2. Calculer $E(X)$ et $V(X)$.
3. Déterminer un estimateur T_n de a par la méthode des moments. Discuter les propriétés de cet estimateur.
4. Soient les données suivantes :

0,81	0,67	0,72	0,41	0,93	0,55	0,28	0,09	0,89
------	------	------	------	------	------	------	------	------

Déterminer une estimation ponctuelle de a . Comment pourrions-nous obtenir une estimation par intervalle ?

Exercice 7 : Estimateurs pour la loi exponentielle corrigé p292

L'exercice du chapitre précédent page 122 fournit l'expression de la densité de probabilité de l'estimateur T_n de λ obtenue dans le cours :

$$f_{T_n}(x) = \frac{n^n \lambda^n}{(n-1)! x^{n+1}} e^{-\frac{n\lambda}{x}} \text{ si } x > 0 \text{ et zéro sinon.}$$

1. Déterminer l'espérance, le biais, la variance et le risque quadratique de T_n .
2. Soit $W_n = \frac{n-1}{n} T_n$. Déterminer l'espérance, le biais, la variance et le risque quadratique de W_n .
3. Quel estimateur conseillerez-vous pour λ ?

Exercice 8 : Désintégrations corrigé p293

La loi de probabilité X de désintégration d'une particule au cours du temps suit une loi exponentielle de paramètre λ . Nous mesurons la durée de vie en 2,20 3,32 1,93 0,88 4,36 microsecondes d'un échantillon de dix particules et 7,10 0,65 0,53 4,69 0,31 nous voulons en déduire une estimation ponctuelle et par intervalle de λ . Nous allons utiliser différentes méthodes et les commenter.

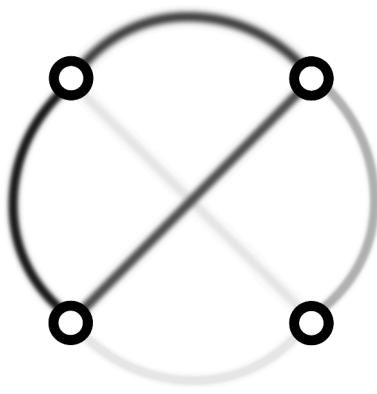
1. En utilisant le théorème central limite, sauriez-vous estimer l'espérance m de X avec son incertitude à 90% de confiance (m : durée de vie moyenne) ? En utilisant la formule de propagation des incertitudes nous pourrions ensuite en déduire λ avec son incertitude. Que pensez-vous de cette estimation de λ ?

2. Soit T_n l'estimateur de λ trouvé en cours. Comme montré précédemment celui-ci est biaisé et nous

$$\text{utilisons donc } W_n = \frac{n-1}{n} T_n .$$

Déterminer par un calcul d'intégrale l'incertitude à 90% sur λ .

3. Retrouver ce résultat par simulation numérique.



V. APPROCHES MULTICRITÈRES

Les approches multicritères correspondent à un ensemble d'outils d'aide à l'analyse et à la décision en présence d'un nombre important de paramètres. Nous verrons deux méthodes. La méthode par pondération, rapide à comprendre et à mettre en œuvre, mais d'un pouvoir d'analyse sommaire. Puis nous étudierons l'analyse en composantes principales, plus difficile à maîtriser, mais qui permet une étude fine et approfondie.

Pour illustrer, nous nous appuierons dans les deux cas, sur l'exemple suivant du choix d'un isolant :

Données :

	Densité kg/m ³	Épais. cm	Prix €/m ²	Confort sur 20
1. Ouate de cellulose	55	25	14	16
2. Laine de roche en vrac	25	36	18	20
3. Perlite expansée	90	25	30	15
4. Polystyrène expansé	18	35	28	20
5. Laine de verre ou roche en rouleau	18	42	34	20
6. Laine de verre ou roche semi-lourde	70	21	36	18
7. Laine de chanvre en rouleau	25	32	41	20
8. Polystyrène extrudé	35	25	49	20
9. Laine de bois	150	18	50	13
10. Laine roche pour sol	130	19	54	15
11. Polyurethane	35	26	106	20

A. Méthode par pondération

Soit un ensemble d'individus qui ont un ensemble de caractéristiques. En statistique, nous comprenons par individus toutes entités semblables pourvues de propriétés multiples. Les caractéristiques ont des unités différentes. De plus les données pour chaque caractéristique sont différemment centrées et dispersées. Pour normaliser les différences, nous allons attribuer des notes :

- caractéristique positive : $note = \frac{V - V_{min}}{E} \times 20$
- caractéristique négative : $note = \frac{V_{max} - V}{E} \times 20$

V : valeur

V_{min} : valeur minimale

$E = V_{max} - V_{min}$: étendue

V_{max} : valeur maximale

Pour la note finale nous pondérons les différentes caractéristiques :

$$note_{finale} = \sum_j p_j n_j$$

avec $\sum_{j=1}^p p_j = 1$ et n_j : note de la $j^{\text{ème}}$ caractéristique.

Pour notre exemple, concentrons-nous sur trois caractéristiques des isolants : l'épaisseur, le prix et le confort.

Nous prenons deux cas, l'un, où le plus important c'est le prix (prix 50%, confort 30% et épaisseur 20%), et un second cas, où, avant tout le confort et l'encombrement du matériaux importent sans trop regarder au prix (confort 50%, épaisseur 40% et prix 10%). Le choix final dépendra de ces priorités.

Priorités 1 :	Épaisseur			Note		Prix			Note		Confort			Note		Note totale
	cm	Note	pondérée				€/m ²	Note	pondérée			sur 20	Note	pondérée		
Pondération :	0,2			0,5						0,3						
Ouate de cellulose	25	14	2,8	14	20	10,0	16	9	2,6	15						
Laine de roche en vrac	36	5	1,0	18	19	9,6	20	20	6,0	17						
Perlite expansée	25	14	2,8	30	17	8,3	15	6	1,7	13						
Polystyrène expansé	35	6	1,2	28	17	8,5	20	20	6,0	16						
Laine de verre ou roche en rouleau	42	0	0,0	34	16	7,8	20	20	6,0	14						
Laine de verre ou roche semi-lourde	21	18	3,5	36	15	7,6	18	14	4,3	15						
Laine de chanvre en rouleau	32	8	1,7	41	14	7,1	20	20	6,0	15						
Polystyrène extrudé	25	14	2,8	49	12	6,2	20	20	6,0	15						
Laine de bois	18	20	4,0	50	12	6,1	13	0	0,0	10						
Laine roche pour sol	19	19	3,8	54	11	5,7	15	6	1,7	11						
Polyurethane	26	13	2,7	106	0	0,0	20	20	6,0	9						

moyennes :	28	12	42	14	18	14	14
étendues :	24	20	92	20	7	20	8
écart-types :	8	6	25	5	3	8	3

Priorités 2 :	Épaisseur			Note		Prix			Note		Confort		Note		Note totale	
	cm	Note	pondérée	€/m²	Note	pondérée	sur 20	Note	pondérée	sur 20	Note	pondérée	sur 20	Note	pondérée	
Pondération :	0,4			0,1						0,5						
Ouate de cellulose	25	14	5,7	14	20	2,0	16	9	4,3	12						
Laine de roche en vrac	36	5	2,0	18	19	1,9	20	20	10,0	14						
Perlite expansée	25	14	5,7	30	17	1,7	15	6	2,9	10						
Polystyrène expansé	35	6	2,3	28	17	1,7	20	20	10,0	14						
Laine de verre ou roche en rouleau	42	0	0,0	34	16	1,6	20	20	10,0	12						
Laine de verre ou roche semi-lourde	21	18	7,0	36	15	1,5	18	14	7,1	16						
Laine de chanvre en rouleau	32	8	3,3	41	14	1,4	20	20	10,0	15						
Polystyrène extrudé	25	14	5,7	49	12	1,2	20	20	10,0	17						
Laine de bois	18	20	8,0	50	12	1,2	13	0	0,0	9						
Laine roche pour sol	19	19	7,7	54	11	1,1	15	6	2,9	12						
Polyurethane	26	13	5,3	106	0	0,0	20	20	10,0	15						

Dans la première situation la laine de roche en vrac sera la mieux adaptée, dans la deuxième, ce sera le polystyrène extrudé.

Bien sûr, si nous ajoutons d'autres critères comme la densité ou le bilan écologique (traité dans un devoir page 212), la conclusion pourra évoluer.

B. Analyse en Composantes Principales

1) Principles

Soit X une caractéristique. Celle-ci prend des valeurs différentes suivant n individus. Ces réalisations, encore appelées mesures ou observations, sont notées :

$$\{x_1, x_2, \dots, x_i, \dots, x_n\}$$

ou plus simplement $\{x_i\}$ avec $1 \leq i \leq n$

Prenons maintenant p caractéristiques, nous avons donc pas seulement une, mais p variables aléatoires:

$$\{X_1, X_2, \dots, X_j, \dots, X_p\}$$

, et les réalisations de X_j se notent :

$$\{x_{1j}, x_{2j}, \dots, x_{ij}, \dots, x_{nj}\}^{22}$$

La matrice des données regroupe l'ensemble des informations :

$$M = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{ip} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np} \end{pmatrix}$$

22 Les indices sont ici inversés par rapport à la notation page 47.

Comme $n \neq p$ cette matrice n'est pas, en général, carrée.

Une colonne représente la viabilité d'une caractéristique selon les individus :

$$X_j = \begin{pmatrix} x_{1j} \\ \vdots \\ x_{ij} \\ \vdots \\ x_{nj} \end{pmatrix}$$

Une ligne de la matrice représente les différentes caractéristiques d'un individu. Mais un individu peut aussi être représenté par un vecteur colonne :

$$\vec{I}_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ij} \\ \vdots \\ x_{ip} \end{pmatrix}$$

Dans l'espace des individus, les individus sont un ensemble de points. Cet espace a pour axes les p caractéristiques et la base $\{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_j, \dots, \vec{e}_p\}$.

Nous allons commencer par calculer la matrice des données normalisée m (centrage et réduction) :

$$\hat{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}$$

Valeur

moyenne : $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$

Écart-

type : $\sigma_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}}$

Les caractéristiques fluctuent autour d'une valeur moyenne. Les valeurs moyennes et les dispersions dépendent de chaque caractéristique. Or l'analyse en composantes principales (ACP) est basée sur les coefficients de corrélations linéaires entre les différentes caractéristiques, et ces coefficients ne dépendent pas de la moyenne et de l'écart-type :

$$r_{jk} = \frac{\sum_{i=1}^n \hat{x}_{ij} \hat{x}_{ik}}{n-1}$$

De ce fait il est plus clair de travailler directement avec la matrice m normale. Si nous gardions M les résultats obtenus seraient les mêmes. Nous trouverons au final, dans un nouvel espace, une nouvelle matrice m' des données que nous pourrions aussi normaliser.

Il faut donc bien garder à l'esprit que pour l'ACP seules les corrélations linéaires comptent. Qu'une grandeur varie sur un millimètre, ou un mètre, autour d'un kilomètre, ou un micromètre, peu où prou, il importe seulement de savoir si cette grandeur varie, ou pas, dans le même sens que les autres. Il convient donc de réfléchir avant de commencer une étude afin de savoir si c'est bien ce type d'analyse qui convient à votre problématique.

Pour les isolants nous obtenons pour M et m :

	X_1	X_2	X_3	X_4
I_1	$X_1^1 = 55$	$X_2^1 = 25$	$X_3^1 = 14$	$X_4^1 = 16$
I_2	$X_1^2 = 25$	$X_2^2 = 36$	$X_3^2 = 18$	$X_4^2 = 20$
I_3	$X_1^3 = 90$	$X_2^3 = 25$	$X_3^3 = 30$	$X_4^3 = 15$
I_4	$X_1^4 = 18$	$X_2^4 = 35$	$X_3^4 = 28$	$X_4^4 = 20$
I_5	$X_1^5 = 18$	$X_2^5 = 42$	$X_3^5 = 34$	$X_4^5 = 20$
I_6	$X_1^6 = 70$	$X_2^6 = 21$	$X_3^6 = 36$	$X_4^6 = 18$
I_7	$X_1^7 = 25$	$X_2^7 = 32$	$X_3^7 = 41$	$X_4^7 = 20$
I_8	$X_1^8 = 35$	$X_2^8 = 25$	$X_3^8 = 49$	$X_4^8 = 20$
I_9	$X_1^9 = 150$	$X_2^9 = 18$	$X_3^9 = 50$	$X_4^9 = 13$
I_{10}	$X_1^{10} = 130$	$X_2^{10} = 19$	$X_3^{10} = 54$	$X_4^{10} = 15$
I_{11}	$X_1^{11} = 35$	$X_2^{11} = 26$	$X_3^{11} = 106$	$X_4^{11} = 20$

Moyennes : 59,18 27,64 41,82 17,91
 Écart-types : 46,09 7,65 24,81 2,66

	X_1	X_2	X_3	X_4
I ₁	$X_1^1 = -0,09$	$X_2^1 = -0,34$	$X_3^1 = -1,12$	$X_4^1 = -0,72$
I ₂	$X_1^2 = -0,74$	$X_2^2 = 1,09$	$X_3^2 = -0,96$	$X_4^2 = 0,79$
I ₃	$X_1^3 = 0,67$	$X_2^3 = -0,34$	$X_3^3 = -0,48$	$X_4^3 = -1,09$
I ₄	$X_1^4 = -0,89$	$X_2^4 = 0,96$	$X_3^4 = -0,56$	$X_4^4 = 0,79$
I ₅	$X_1^5 = -0,89$	$X_2^5 = 1,88$	$X_3^5 = -0,32$	$X_4^5 = 0,79$
I ₆	$X_1^6 = 0,23$	$X_2^6 = -0,87$	$X_3^6 = -0,23$	$X_4^6 = 0,03$
I ₇	$X_1^7 = -0,74$	$X_2^7 = 0,57$	$X_3^7 = -0,03$	$X_4^7 = 0,79$
I ₈	$X_1^8 = -0,52$	$X_2^8 = -0,34$	$X_3^8 = 0,29$	$X_4^8 = 0,79$
I ₉	$X_1^9 = 1,97$	$X_2^9 = -1,26$	$X_3^9 = 0,33$	$X_4^9 = -1,84$
I ₁₀	$X_1^{10} = 1,54$	$X_2^{10} = -1,13$	$X_3^{10} = 0,49$	$X_4^{10} = -1,09$
I ₁₁	$X_1^{11} = -0,52$	$X_2^{11} = -0,21$	$X_3^{11} = 2,59$	$X_4^{11} = 0,79$

Moyennes : 0 0 0 0
Écart-types : 1 1 1 1

L'idée fondamentale de cette étude est basée sur la matrice de corrélation. Nous cherchons les corrélations entre les différentes caractéristiques. Si un ensemble de caractéristiques sont très corrélées entre elles, cela signifie qu'une seule suffirait à décrire notre système sans perte importante d'information.

Comme nous sommes mieux habitué à visualiser les choses en deux dimensions, nous cherchons deux caractéristiques, combinaisons de toutes les autres, qui contiendraient la majorité de l'information. Ainsi nous pourrions nous ramener à l'essentiel d'un problème en peu de paramètres et simplifier la vision avec des caractéristiques mieux adaptées auxquelles nous n'aurions pas pensées à priori.

Coefficients de corrélations linéaires :

$$r_{jk} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{(n-1)\sigma_j\sigma_k}$$

$$\text{avec } r_{jj} = 1 \text{ et } r_{kj} = r_{kj}$$

nous avons alors une matrice carrée appelée la matrice des corrélations :

$$r = \begin{pmatrix} 1 & r_{12} & r_{13} & \dots & r_{1p} \\ r_{12} & 1 & \dots & \dots & \dots \\ r_{13} & \dots & 1 & \dots & \dots \\ \dots & \dots & \dots & 1 & \dots \\ r_{1p} & \dots & \dots & \dots & 1 \end{pmatrix} \quad -1 \leq r_{ij} \leq 1$$

Nous diagonalisons ensuite la matrice r :

$$r' = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_j & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_p \end{pmatrix}$$

$$\text{avec } \sum_{j=1}^p \lambda_j = p = \text{tr}(r) = \text{tr}(r')$$

La somme des valeurs propres λ_j est p , le nombre de caractéristiques. La somme des valeurs sur la diagonale d'une matrice carré est appelée trace de la matrice, celle-ci est invariante par changement de base. Nous les plaçons tel que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$.

Nous avons alors une nouvelle base, appelée base des vecteurs propres $\{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_j, \dots, \vec{e}_p\}$, pour la projection de nos individus sur les nouvelles caractéristiques $\{X'_1, X'_2, \dots, X'_j, \dots, X'_p\}$.

Dans ce nouvelle espace toutes les nouvelles caractéristiques sont décorrélées entre elles. C'est une grande avancée pour l'analyse, nous savons maintenant qu'elles sont les "vraies" caractéristiques. Nous avons le point de vue sur le problème qui donne la vision la plus simple et la plus claire.

De plus avec les valeurs propres nous avons un ordre d'importance des caractéristiques. Le pourcentage d'information contenu dans une caractéristique X'_j est :

$$\%_j = \frac{\lambda_j}{p}$$

Il n'est pas rare que deux caractéristiques sur une dizaine contiennent à elles seules 90% de l'information totale d'un système de données. C'est un outil descriptif très puissant.

Les valeurs propres se calculent en résolvant l'équation $\det(r - \lambda I) = 0$. Nous calculons le déterminant de la matrice des corrélations à laquelle nous avons soustrait la matrice identité multipliée par la valeur propre recherchée. Il s'agit d'un problème classique de mathématiques abordé en début d'enseignement supérieur scientifique. Les calculs sont effectués sur des exemples dans le deuxième exercice. Nous avons ensuite l'habitude de classer les valeurs propres en ordre décroissant dans la matrice r' .

Les vecteurs propres sont déterminés en résolvant l'équation :

$$r \vec{e'}_j = \lambda_j \vec{e'}_j$$

Pour obtenir la nouvelle matrice réduite des données avec les individus en fonction des nouvelles caractéristiques nous avons besoin de la matrice de passage P . Celle-ci est une matrice carrée dont les colonnes sont les composantes des vecteurs propres. Ensuite $mP = m'$.

Ce travail effectué pour nos isolants donne :

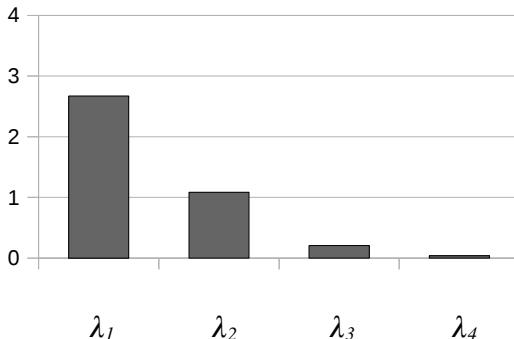
$$r = \begin{vmatrix} 1,000 & -0,810 & 0,114 & -0,936 \\ -0,810 & 1,000 & -0,307 & 0,725 \\ 0,114 & -0,307 & 1,000 & 0,097 \\ -0,936 & 0,725 & 0,097 & 1,000 \end{vmatrix}$$

$$r' = \begin{vmatrix} 2,671 & 0,000 & 0,000 & 0,000 \\ 0,000 & 1,083 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,206 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,040 \end{vmatrix}$$

Cumul des valeurs propres :

Valeurs propres	%	cumul	%
2,67	67	2,67	67
1,08	27	3,75	94
0,21	5	3,96	99
0,04	1	4	100

Éboulis des valeurs propres :



La matrice de passage permet d'avoir la nouvelle base en fonction de l'ancienne, par exemple :

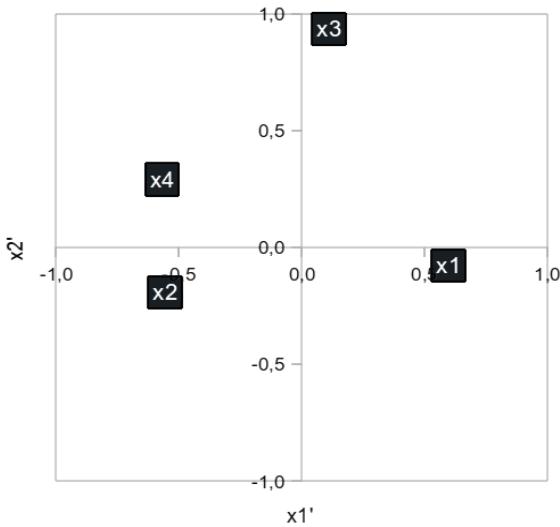
$$\vec{e}_1 \approx 0,60 \vec{e}_1 - 0,56 \vec{e}_2 + 0,11 \vec{e}_3 - 0,57 \vec{e}_4$$

$$P = \begin{vmatrix} 0,60 & -0,08 & -0,34 & -0,72 \\ -0,56 & -0,19 & -0,81 & -0,06 \\ 0,11 & 0,94 & -0,31 & 0,14 \\ -0,57 & 0,29 & 0,37 & -0,68 \end{vmatrix}$$

$$\text{norme : } \quad \begin{matrix} 1 & 1 & 1 & 1 \end{matrix}$$

Or X_1' et X_2' contiennent 94% de l'information du système.

Et tracer X_1 , X_2 , X_3 et X_4 en fonction de X_1' et X_2' donne une bonne vision globale de l'information.

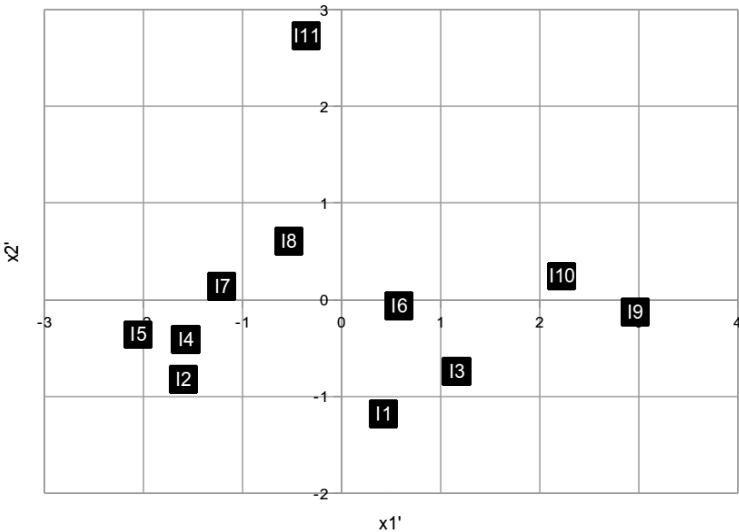


La perte d'information est de seulement 6%.

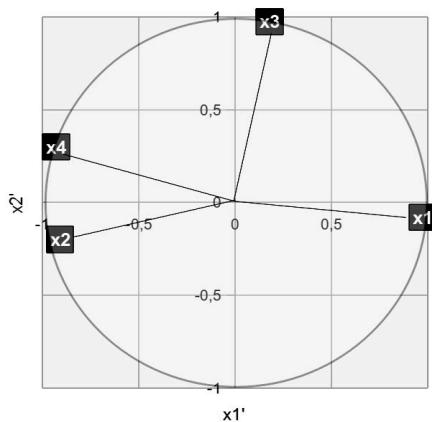
Nous obtenons ensuite la matrice des données en fonction des composantes principales :

	x_1'	x_2'	x_3'	x_4'
I_1	0,42	-1,18	0,39	0,42
I_2	-1,60	-0,83	-0,04	-0,19
I_3	1,16	-0,74	-0,21	0,21
I_4	-1,58	-0,41	-0,01	-0,01
I_5	-2,06	-0,36	-0,82	-0,03
I_6	0,58	-0,06	0,70	-0,18
I_7	-1,21	0,14	0,09	-0,03
I_8	-0,54	0,60	0,66	-0,09
I_9	2,96	-0,13	-0,44	-0,06
I_{10}	2,22	0,24	-0,17	-0,24
I_{11}	-0,36	2,73	-0,15	0,21

Nous représentons enfin ces données en fonctions des composantes principales :



Pour trouver rapidement le sens de ces deux nouvelles composantes principales il est pratique de tracer un **cercle des corrélations** :



Pour cela nous avons calculé la matrice des corrélations entre anciennes et nouvelles variables (composantes principales) :

	X_1'	X_2'	X_3'	X_4'
X_1	$r_{11'} = 0,97$	$r_{12'} = -0,08$	$r_{13'} = -0,16$	$r_{14'} = -0,14$
X_2	$r_{21'} = -0,91$	$r_{22'} = -0,2$	$r_{23'} = -0,37$	$r_{24'} = -0,01$
X_3	$r_{31'} = 0,18$	$r_{32'} = 0,97$	$r_{33'} = -0,14$	$r_{34'} = 0,03$
X_4	$r_{41'} = -0,93$	$r_{42'} = 0,3$	$r_{43'} = 0,17$	$r_{44'} = -0,14$

L'interprétation peut maintenant commencer :

Tout d'abord les quatre caractéristiques sont proches des bords du cercle de rayon 1. Elles vont, donc, toutes jouer un rôle important. Si nous avions des caractéristiques proches du centre elles seraient peu ou pas prises en considération dans la description.

Nous voyons que X_1 , X_2 et X_4 sont très corrélés avec la première composante principale X_1' . X_1 positivement, X_2 et X_4 négativement. Tout d'abord, quand l'épaisseur augmente, le confort aussi (toujours en terme probabiliste, bien sûr). De plus, plus la densité d'un matériau est grande, moins il est confortable (on sait que les bons isolants sont souvent remplis d'air, donc peu denses). Aussi plus la densité d'un matériau est grande et moins il est épais (et volumineux, ce qui est logique aussi). Apparaît donc une caractéristique nouvelle qui combine ces trois paramètres et qui contient 67% de l'information du problème.

Sur le cercle des corrélation X_3 est selon X_2' (il est très corrélés positivement avec lui $r_{32'}$ vaut environ 0,97). La deuxième composante principale correspond donc simplement au prix (poids 27%).

Le prix des matériaux est sans lien avec leur confort. Des grandeurs décorrélées sur le cercle des corrélations sont à angle droit avec le centre. Plus elles sont corrélées dans le même sens, plus l'angle est faible ; et si elles sont corrélées en sens contraire, l'angle vaut 180°.

Regardons maintenant les individus et essayons de les regrouper en catégories :

Tout d'abord un isolé, l'individu 11, qui à un confort/épaisseur moyen et un prix élevé.

Le groupe le plus intéressant en bas à gauche (2, 5 et 4), qui correspond à des isolants à la fois pas cher et confortable (avec une bonne épaisseur).

Et puis ceux qui partent vers la droite, d'un prix moyen mais pas confortable, donc à éviter, à moins que le volume disponible pour l'isolation soit faible.

A vous de prendre votre décision ; cet outil permet de choisir son isolant en connaissance de cause.

Dans le deuxième devoir sont étudiées d'autres isolants avec l'éco-bilan, l'isolation et le confort d'été.

2) Une deuxième illustration

Nous considérons maintenant un exemple plus mathématique qui permet de visualiser et d'élargir notre compréhension de la méthode.

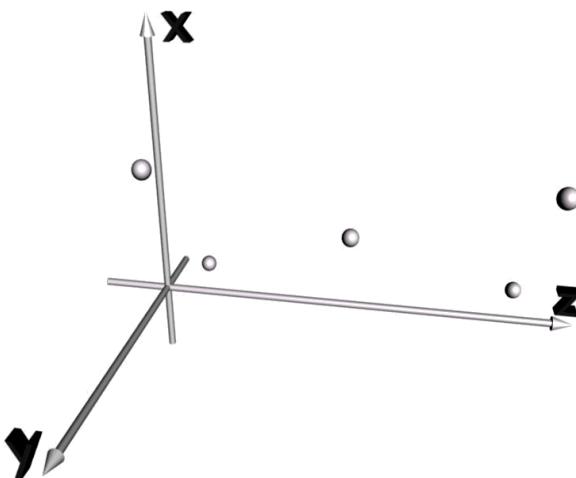
Nous prenons des points dans l'espace. Les caractéristiques prises sont les positions x, y et z. Nous avons cinq points : $M_1(1,1,1)$, $M_2(1,1,5)$, $M_3(2,2,3)$, $M_4(3,3,1)$ et $M_5(3,3,5)$.

Matrice de données :

$$M = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 5 \\ 2 & 2 & 3 \\ 3 & 3 & 1 \\ 3 & 3 & 5 \end{pmatrix}$$

Matrice réduite :

$$m = \begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix}$$



$$\text{Matrice des corrélations : } r = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

d'où : $\lambda_1=2$, $\lambda_2=1$ et $\lambda_3=0$.

$$\text{Matrice de passage : } P = \begin{pmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{pmatrix}$$

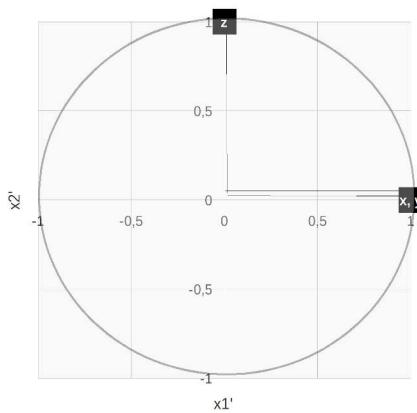
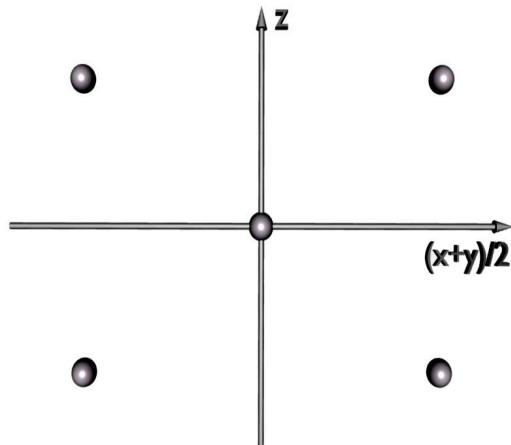
d'où : $\vec{e}'_1 = (\vec{e}_1 + \vec{e}_2)/\sqrt{2}$, poids $w_1=2/3$, $\vec{e}'_2 = \vec{e}_3$, poids $w_2=1/3$ et $\vec{e}'_3 = (\vec{e}_2 - \vec{e}_1)/\sqrt{2}$, poids $w_3=0$.

Matrice des individus dans la nouvelle base :

$$m' = \begin{pmatrix} -\sqrt{2} & -1 & 0 \\ -\sqrt{2} & 1 & 0 \\ 0 & 0 & 0 \\ \sqrt{2} & -1 & 0 \\ \sqrt{2} & 1 & 0 \end{pmatrix}$$

$$\text{Corrélations entre nouvelles et anciennes caractéristiques : } \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

D'où le graphe avec les composantes principales $X'_1=(x+y)/2$ et $X'_2=z$ et le cercle des corrélations suivants :



Ressort ainsi la figure suivante : un carré avec un point au milieu. À partir d'un rectangle incliné dans l'espace, on a une figure simple dans un plan qui contient toutes les informations relatives.

3) Résumé

Matrice des données
(individus selon les lignes et
grandeurs selon les colonnes)

Matrice des données
réduite

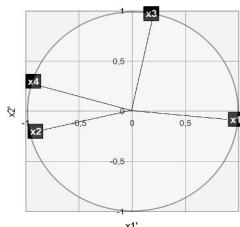
Matrice des corrélations

Valeurs propres (ordonnées
dans l'ordre décroissant)

Vecteurs propres
correspondants

Matrice de passage (les
colonnes sont les vecteurs
propres)

Matrice des données dans
la nouvelle base des
vecteurs propres
(possibilité de normaliser m')



$$x_{ij}$$



$$\hat{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}$$



$$r_{jk} = \frac{\sum_{i=1}^n \hat{x}_{ij} \hat{x}_{ik}}{n-1}$$



$$\det(r - \lambda I) = 0$$



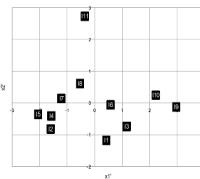
$$w_j = \frac{\lambda_j}{p}$$



$$r \vec{e}_j = \lambda_j \vec{e}_j$$



$$m' = m P$$



Représentation
des individus
dans
le nouvel espace

↓
Matrice
des corrélations
entre nouvelles
et anciennes
grandeurs :
**Cercle
des
corrélations**

	X'_1	X'_2	...	X'_p
X_1	$r_{11'}$	$r_{12'}$		$r_{1p'}$
X_2	$r_{21'}$	$r_{22'}$		$r_{2p'}$
...
X_p	$r_{p1'}$	$r_{p2'}$...	$r_{pp'}$

C. Exercices

Exercice 1 : Fournisseurs corrigé p295

Nous voulons choisir entre deux fournisseurs. Ceux-ci sont notés selon trois critères : les tarifs, la qualité et la solidité financière.

Pour le fournisseur 1 les notes sont, sur dix : 4, 8 et 5.
Pour le deuxième : 6, 3 et 6.

Quel fournisseur choisir si l'on pondère 50% pour les tarifs, 30% pour la qualité et le reste pour la solidité financière ?

Exercice 2 : Principe de l'ACP corrigé p295

Réalisez l'analyse mathématique pour les deux matrices de données suivantes (la première sans l'aide d'un ordinateur et pour la seconde vous pourrez utiliser le logiciel Scilab) :

	X_1	X_2	X_3	X_4
I_1	$X_1^1 = -2$	$X_2^1 = 4$	$X_3^1 = 0$	$X_4^1 = 8$
I_2	$X_1^2 = 0$	$X_2^2 = 3$	$X_3^2 = -3$	$X_4^2 = 3$
I_3	$X_1^3 = 0$	$X_2^3 = 3$	$X_3^3 = 3$	$X_4^3 = 13$
I_4	$X_1^4 = 2$	$X_2^4 = 2$	$X_3^4 = 0$	$X_4^4 = 8$

	X_1	X_2	X_3
I_1	$X_1^1 = 1$	$X_2^1 = 3$	$X_3^1 = 2$
I_2	$X_1^2 = 2$	$X_2^2 = 1$	$X_3^2 = 3$
I_3	$X_1^3 = 3$	$X_2^3 = 2$	$X_3^3 = 1$

Pour vous exercer davantage sur l'ACP vous avez deux cas concrets traités en devoir.

Pour faire vos calculs pour l'ACP vous avez Scilab un très bon logiciel de calcul numérique libre, gratuit et multiplateforme. Vous pouvez ensuite présenter vos résultats avec un tableur.

Par exemple pour le devoir Suricate : ouvrir le logiciel Scilab (prononcez : " saille lab", contraction de *Scientific Laboratory* en anglais) puis *Fichier > exécuter...* le fichier acpsuricate.sce :

```
function cov=covariance(A)
    moyenne = mean(A,"r");
    n = size(A,"r")
    Acent = A - ones(n,1)*moyenne;
    cov=(Acent'*Acent)/(n-1);
endfunction;

function c=correlation(A)
    cov=covariance(A);
    ectinv=(1)./sqrt(diag(cov));
    c=diag(ectinv)*cov*diag(ectinv);
endfunction;

function [c]=val_prop(A);
    [Diag,Base]=bdiag(A);
    c=qsort(diag(Diag));
endfunction;

function c=vect_prop(A);
    [Diag,Base]=bdiag(A);
    if Base(:,1)==-abs(Base(:,1)) then Base=-1*Base; end;
    [Valpr,k]=qsort(diag(Diag));
    c = Base(:,k);
endfunction;

function c=acp_indiv(A);
    mat_cor=correlation(A);
```

```

X=centrer_reduire(A);
VectP=vect_prop(mat_cor);
c = X*VectP;
endfunction;

function c=centrer_reduire(A)
moyenne = mean(A,"r");
cov=covariance(A);
n = size(A,"r");
Acent = A - ones(n,1)*moyenne;
c=Acent*(diag((1)./sqrt(diag(cov))));;
endfunction;

function c=acp_var(A);
mat_cor=correlation(A);
X=centrer_reduire(A);
VectP=vect_prop(mat_cor);
ValP=val_prop(mat_cor);
c = VectP*diag(sqrt(ValP));
endfunction;

M=[ 78 357 24 4.5 10 1 5 7.2 ; 190 1300 85 44
38 10 165 6 ; 486 403 84 10 1187 3.2 9.1 7.2 ;
555 403 110 11 1479 4 14 7.2 ; 90 4368 11 322
174 132 1708 6.6 ; 202 402 43 8 336 2 4.7
7.5 ; 11.5 71 8 13.5 8.1 6.2 11.6 6.95 ; 103
2989 10 235 138 66 1172 6.8 ; 253 820 11 4 25
3 7 6]

m=centrer_reduire(M)
r=correlation(m)
val_prop(r)
vect_prop(r)
acp_indiv(m)
acp_var(m)

```

VI. COMPLÉMENTS

A. Mesure avec une règle

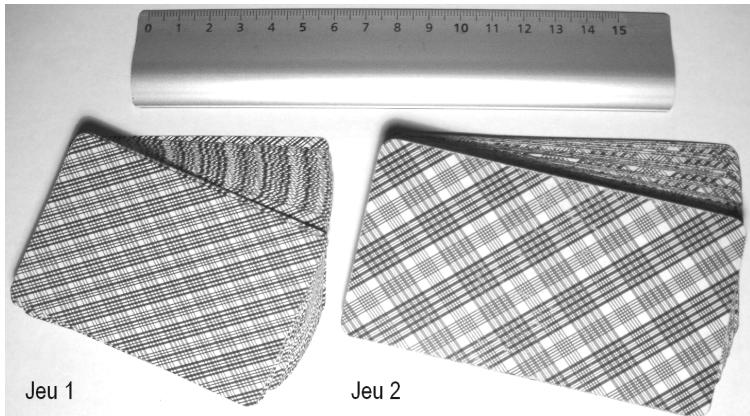
Article publié dans le BUP [ii].

RÉSUMÉ

La mesure d'une grandeur par un système d'acquisition induit de part sa résolution une erreur de discréétisation. Nous nous attachons ici à la mesure d'une longueur avec une règle graduée. Ce type de mesure nous amène à considérer une loi de probabilité continue uniforme. Nous utilisons ensuite un produit de convolution pour déterminer l'incertitude avec sa confiance d'une somme de longueurs. Nous généralisons finalement au cas général du calcul d'incertitudes pour des variables aléatoires indépendantes en utilisant la formule de propagation des erreurs.

INTRODUCTION

Nous voulons mesurer des longueurs et évaluer les incertitudes le plus précisément possible. Incertitudes sur les valeurs mesurées et leurs sommes. Nous disposons d'une règle de 15cm graduée au millimètre et de deux jeux de cartes. La règle est supposée parfaite et les cartes de chaque jeu à l'identique.



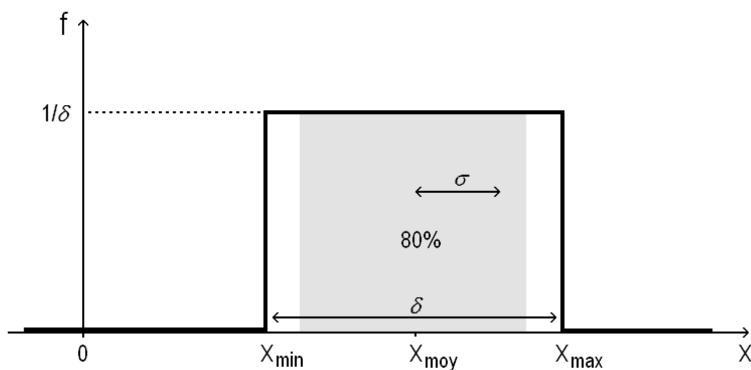
1. MESURE DE LA LONGUEUR D'UNE CARTE

Nous plaçons la graduation du zéro sur le bord gauche de la carte. Sur le bord droit nous considérons la graduation la plus proche du bord. L'expérimentateur ne lit pas entre les graduations. L'épaisseur des traits qui délimitent une graduation est considérée comme négligeable devant la largeur de cette graduation. Nous obtenons ainsi pour le jeu 1:

$$x_1 = 8,4 \pm 0,05 \text{ cm}.$$

Pour le jeu 2:

$$x_2 = 11,2 \pm 0,05 \text{ cm}.$$



Nous acceptons une perte d'information due à la résolution $\delta = 1 \text{ mm}$ de la règle. Lorsque ultérieurement nous exploitons ces données toutes les valeurs entre $x_{\min} = x_{\text{moy}} - \delta/2$ et $x_{\max} = x_{\text{moy}} + \delta/2$ sont équiprobables. La loi de probabilité de la variable continue aléatoire X est uniforme. x est une réalisation de X. Cette distribution de probabilité a une étendue $E = x_{\max} - x_{\min}$ et nous vérifions pour la densité de probabilité $f(x)$:

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

La probabilité pour que la valeur de X soit comprise entre x et $x+dx$ est de $f(x)dx$. Le résultat sera compris avec certitude entre x_{\min} et x_{\max} : par exemple $x_1 = 8,4 \pm 0,05 \text{ cm}$ à 100% de confiance, mais $x_1 = 8,4 \pm 0,04 \text{ cm}$ avec une probabilité de 80%.

Pour caractériser l'étalement d'une distribution considérons l'étendue E et l'écart-type σ dont la définition pour une loi continue est:

$$V = \sigma^2 = \int (x - x_{moy})^2 f(x) dx ,$$

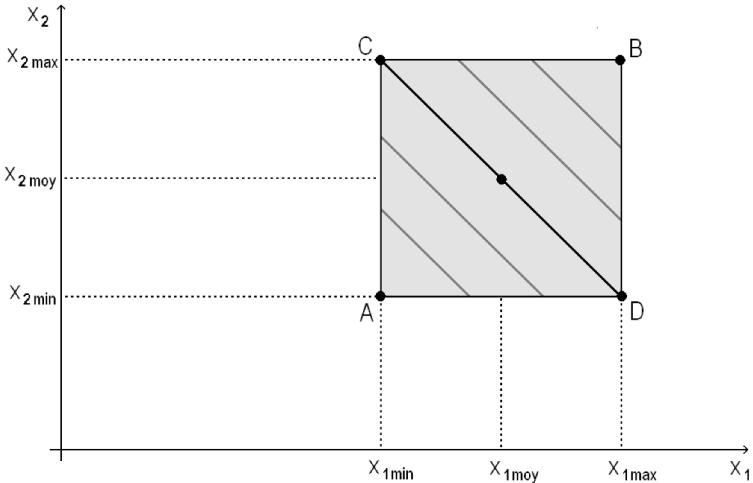
V est appelée la variance. Pour une loi uniforme:

$$\sigma = \delta / \sqrt{12} \approx 0,29 \delta ,$$

et nous avons $x = x_{moy} \pm \sigma$ avec une confiance de 58%. L'écart-type est une grandeur adéquate pour caractériser la largeur d'une distribution. L'étendue quant à elle est définie par les valeurs extrêmes qui peuvent être peu représentatives ou pire des valeurs aberrantes.

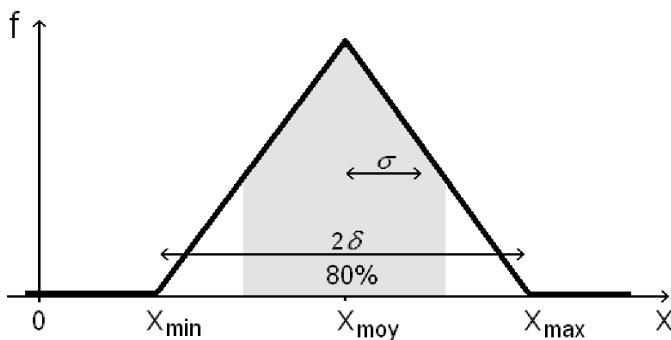
2. LONGUEUR DES DEUX CARTES MISES BOUT À BOUT

Nous souhaitons déterminer l'incertitude sur x avec $x = x_1 + x_2$. Si nous traçons x_2 en fonction de x_1 l'ensemble des points possibles forme un domaine carré. L'ensemble des points tel que x soit constant est une portion de droite de pente -1 et d'ordonnée à l'origine x : $x_2 = -x_1 + x$. Il n'y a qu'un cas qui réalise $x = x_{\min}$ soit $\{x_1 = x_{1\min}; x_2 = x_{2\min}\}$ au point A sur la figure. Par contre sur l'ensemble du segment [CD] $x = x_{moy}$. Nous comprenons que toutes les valeurs de x ne sont pas équiprobables.



La loi de probabilité f de X se calcule à partir de celle f_1 de X_1 et f_2 de X_2 . Pour une somme de variables aléatoires indépendantes le résultat est donné par un produit de convolution [iii] :

$$f(x) = \int f_1(y)f_2(x-y)dy \Rightarrow \begin{cases} x < x_{\min} \Rightarrow f(x) = 0 \\ x_{\min} < x < x_{\text{moy}} \Rightarrow f(x) = \frac{(x-x_{\min})}{\delta^2} \\ x_{\text{moy}} < x < x_{\max} \Rightarrow f(x) = \frac{(x_{\max}-x)}{\delta^2} \\ x > x_{\max} \Rightarrow f(x) = 0 \end{cases}$$

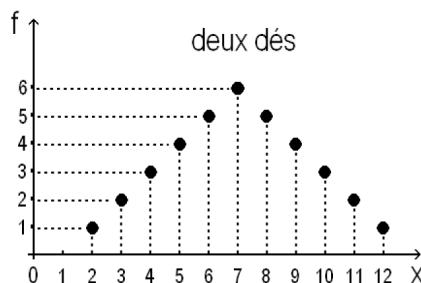
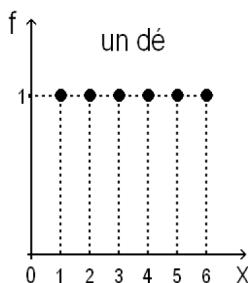


Nous avons alors une loi de probabilité triangulaire.

Nous obtenons $x=19,6 \pm 0,1\text{ cm}$ avec 100% de confiance,
et $x=19,6 \pm 0,055\text{ cm}$ avec 80% de confiance.

3. ANALOGIE AVEC LE LANCER DE DEUX DÉS

Pour chaque dé les six valeurs sont équiprobables. Ici la loi de probabilité n'est plus continue mais discrète. Pour le lancer simultané de deux dés, la somme des valeurs obtenues est comprise entre deux et 12. Dans ce cas il n'y a plus équiprobabilité, une manière de faire deux avec un double un, deux manières de faire trois avec un et deux ou deux et un... Pour faire sept nous obtenons le maximum de possibilités. Nous retrouvons ainsi une loi triangulaire.



4. LONGUEUR DE DEUX CARTES D'UN MÊME JEU MISSES BOUT À BOUT

Les cartes d'un jeu étant supposées identiques si la longueur de l'une d'elle est surestimée, il en sera de même pour la deuxième. Dans ce cas les erreurs s'ajoutent et ne peuvent pas se compenser. Pour deux cartes différentes, la première mesure pouvait être sous-estimée et la deuxième surestimée, une compensation pouvant alors se produire. Ici ce n'est plus le cas et pour $X = X_i + X_i'$ nous obtenons une loi de probabilité à nouveau uniforme de largeur 2δ . Nos variables aléatoires ne sont plus indépendantes.

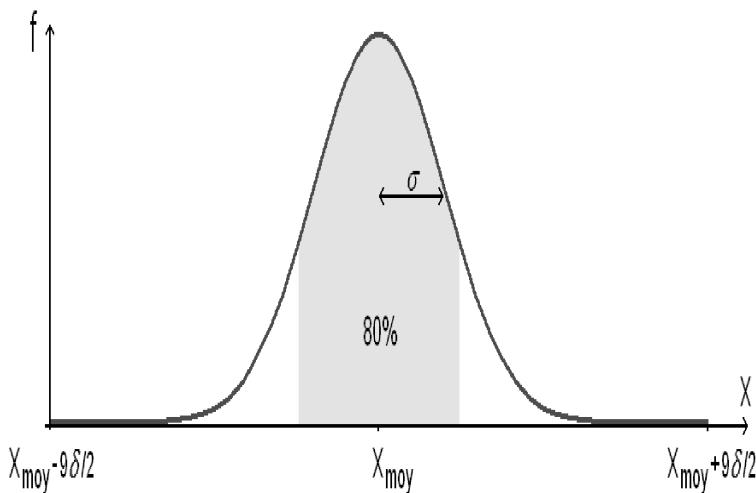
Pour le jeu 1:

$$x_1 = 8,4 \pm 0,04 \text{ cm} \Rightarrow x = 2x_1 = 16,8 \pm 0,08 \text{ cm}$$

à 80% de confiance.

5. SOMME DE N LONGUEURS INDÉPENDANTES

Nous avons $X = \sum_{i=1}^N X_i$. Chaque longueur X_i suit une loi uniforme de largeur δ . Pour la somme de neuf variables aléatoires indépendantes après itération du calcul nous obtenons la courbe suivante :



Nous avons dans ce cas $x = x_{moy} \pm 0,11\text{ cm}$ à 80%. A 100% de confiance $x = x_{moy} \pm 0,45\text{ cm}$ ce qui amène à considérer des domaines où la probabilité de présence de X est vraiment négligeable. Une incertitude de 0,45cm semble inutile alors que 99% des cas étaient déjà présents avec une incertitude de 0,22cm.

Raisonner avec une confiance de 100% revient à considérer l'étendue, celle-ci est additive pour une somme de variables. L'étendue est proportionnelle à N.

	<i>80%</i>	<i>95%</i>	<i>99%</i>
N=1	<i>0,40 δ</i>	<i>0,48 δ</i>	<i>0,50 δ</i>
2	<i>0,55 δ</i>	<i>0,78 δ</i>	<i>0,90 δ</i>
3	<i>0,66 δ</i>	<i>0,97 δ</i>	<i>1,19 δ</i>
4	<i>0,75 δ</i>	<i>1,12 δ</i>	<i>1,41 δ</i>
5	<i>0,84 δ</i>	<i>1,25 δ</i>	<i>1,60 δ</i>
6	<i>0,92 δ</i>	<i>1,38 δ</i>	<i>1,76 δ</i>
7	<i>0,99 δ</i>	<i>1,49 δ</i>	<i>1,91 δ</i>
8	<i>1,06 δ</i>	<i>1,59 δ</i>	<i>2,05 δ</i>
9	<i>1,12 δ</i>	<i>1,69 δ</i>	<i>2,18 δ</i>
10	<i>1,2 δ</i>	<i>1,8 δ</i>	<i>2,3 δ</i>
20	<i>1,7 δ</i>	<i>2,5 δ</i>	<i>3,3 δ</i>
50	<i>2,6 δ</i>	<i>4,0 δ</i>	<i>5,2 δ</i>
100	<i>3,7 δ</i>	<i>5,7 δ</i>	<i>7,4 δ</i>

Mais cette approche ne tient pas compte d'une chose: la courbe se resserre autour de la moyenne quand N augmente. Il existe une autre grandeur additive: la variance. L'écart-type racine de la variance est proportionnel à \sqrt{N} et tient compte des compensations d'erreurs.

La courbe obtenue est ce qu'on appelle une courbe en cloche. Un théorème statistique, appelé théorème central limite, indique que pour N grand la courbe tend vers une gaussienne. L'étendue d'une gaussienne est infinie pour un écart-type fini.

Nous pouvons résumer l'évolution de l'incertitude sur la somme de N longueurs indépendantes mesurées avec une même résolution δ dans un tableau. En italique, à partir de N=10, il s'agit de simulations numériques réalisées sur

ordinateur par génération de nombres aléatoires.

Les résultats des mesures sont souvent donnés avec une confiance de 95%, ce qui correspond pour une gaussienne à une incertitude d'environ 2σ .

6. AUTRES APPLICATIONS

Un coureur souhaite mesurer son temps de parcours. Sa montre à affichage numérique indique qu'il part à $10 h 52 \text{ min}$ et qu'il arrive à $11 h 11 \text{ min}$. L'affichage est à la minute, il est donc parti entre $10 h 52 \text{ min } 00 \text{ s}$ et $10 h 52 \text{ min } 59 \text{ s}$. D'où la date de départ dans l'intervalle $t_1 = 10 h 52 \text{ min } 30 \text{ s} \pm 30 \text{ s}$. La résolution est d'une minute. La durée du parcours est $\Delta t = t_2 - t_1$. Les résultats restent vrais pour une différence. Nous avons $N=2$ et $\Delta t = 19 \text{ min} \pm 47 \text{ s}$ avec 95% de confiance.

Même démarche si des étudiants mesurent une différence d'angles sur un goniomètre. Chaque mesure étant à la minute d'arc près l'incertitude du résultat est de 47 secondes d'arc à 95%.

Sept personnes veulent rentrer en même temps dans un ascenseur. Sa charge maximale est de 500kg. Leurs masses individuelles sont mesurées avec un pèse personne d'une résolution de un kilogramme. La masse totale est de 499kg. Quelle est la probabilité d'être en surcharge?

Pour N=7 l'incertitude atteint un kilogramme avec une confiance de 80%. Il y a donc une chance sur dix pour que l'ascenseur soit en surcharge.

Au laboratoire de nombreux appareils de mesure disposent d'affichages numériques. La résolution est au dernier digit près. Mais l'incertitude globale est bien supérieure. Il faut consulter la notice de chaque appareil.

CONCLUSION

La démarche générale consiste à combiner des lois de probabilités. L'outil mathématique utilisé est un changement de variables, puis une ou plusieurs intégrations. Pour la mesure avec une règle décrite dans cet article, il s'agissait d'une somme de deux variables aléatoires indépendantes et nous avons obtenu un produit de convolution.

Si l'on veut faire un calcul plus rapide une analyse de variance peut suffire. Nous avons une variable aléatoire X qui dépend de N variables aléatoires indépendantes X_i : $X = f(X_1, X_2, \dots, X_i, \dots, X_N)$. Nous appelons σ_i l'écart-type de X_i et σ celui de X . Pour des σ_i finis et de petites variations, nous avons la formule de propagation des écart-types [iv] :

$$\sigma^2 = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2$$

Et, indépendamment des lois de probabilités, cette relation entre les variances reste vraie. On pourra ainsi donner son résultat avec une incertitude à 2σ ou 3σ .

Existe-t-il une formule analogue en terme de confiance? Oui, mais elle est approximative, c'est la formule de propagation des incertitudes:

$$\Delta f^2 = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 \Delta x_i^2 ,$$

avec $x_i = x_{i\text{moy}} \pm \Delta x_i$, $f = f_{\text{moy}} \pm \Delta f$ et une confiance constante. Cette formule est très pratique et permet un calcul rapide et raisonnable des incertitudes combinées. Qui plus est, elle est exacte si la forme des distributions est la même pour X et les X_i . Par exemple si les X_i sont à distribution gaussienne toute combinaison linéaire l'est aussi.

Nous tenons ainsi compte des compensations et nous évitons d'utiliser la formule $\Delta f = \sum_{i=1}^n |\partial f / \partial x_i| \Delta x_i$ qui

surestime les incertitudes, parfois même avec un tel excès que l'on en perd son sens physique. Cette dernière formule ne tient compte d'aucunes compensations, on a la pire des situations, statistiquement improbable. Ici, par exemple pour $N = 100$, on aurait une incertitude de 50δ , au lieu de $5,7 \delta$ dans la pratique (confiance de 95%).

Dans cet article nous nous sommes concentré sur la résolution d'un système d'acquisition qui donne une erreur de discréétisation. Mais on peut aussi être amené à considérer des erreurs systématiques et des erreurs aléatoires. Ici la règle était supposée parfaite, c'est à dire juste et fidèle.

B. Métrologie

1. Exemples

Intéressons-nous à deux instruments de mesure particuliers : un chronomètre et un ampèremètre.

Mesure d'une durée avec un chronomètre :

Nous mesurons la durée d d'un phénomène. Nous supposons que l'expérimentateur appuie aux bons moments pour mesurer ce que l'on appelle les dates initiale t_1 et finale t_2 . Pour être un bon expérimentateur il faut avoir une connaissance minimale de ses instruments de mesure. Un chronomètre, du type utilisé ici, fonctionne avec un quartz très précis. La dérive est d'une quinzaine de secondes par an, soit environ 0,00004 secondes sur notre expérience. Ici, nous négligeons la dérive. Nous pouvons supposer le chronomètre parfait (juste et fidèle). Seule la résolution va intervenir ($\delta = 1/10$ ème de seconde).

003'32.8

004'55.7

$$3min32,75s < t_1 < 3min32,85s$$

$$4min55,65s < t_2 < 4min55,75s$$

$$\text{et } d = t_2 - t_1$$

Les lois de probabilité de t_1 et t_2 sont uniformes et celle de d triangulaire (voir article précédent).

D'où²³ $\sigma_d = \delta/\sqrt{6} \approx 0,41\delta$, $\Delta d \approx 0,78\delta$ (tableau page 182),

23 Calculons l'écart-type d'une distribution triangulaire (après recentrage) : $\sigma^2 = \int_{-\delta}^0 x^2 \frac{(x+\delta)}{\delta^2} dx + \int_0^\delta x^2 \frac{(\delta-x)}{\delta^2} dx = \frac{\delta^2}{6}$. Ou, si nous utilisons la formule de propagation des écart-types on trouve que pour une somme, ou différence, de n grandeurs de même écart-

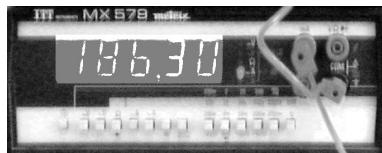
$d = 82,900 \pm 0,078$ s et $\Delta d/d \approx 0,1\%$ avec 95% de confiance.

Maintenant, si le phénomène durait plusieurs mois, l'incertitude la plus grande serait la dérive, le chronomètre ne serait plus juste. Nous voyons que l'ensemble de la situation expérimentale est à prendre en compte et pas seulement l'instrument de mesure. Nous pourrons incorporer cette incertitude systématique de dérive par la connaissance de la valeur du biais.

Si la dérive n'est pas connue, nous pouvons prendre 100 chronomètres identiques que nous lançons simultanément. Nous aurons au final une dispersion aléatoire des valeurs qui donne une incertitude liée à la dérive.

Mesure d'un courant avec un ampèremètre :

L'ampèremètre, placé en série dans un circuit, mesure le courant de la branche où il est situé.



Nous choisissons le calibre adapté, ici pour des courants de 20 à 200 mA, et nous mesurons : $I = 186,30$ mA.

Quelle est l'incertitude sur I ?

Nous pourrions penser, comme pour le chronomètre, qu'elle est sur le dernier digit, en fait la notice indique une plus grande incertitude :

types nous avons $\sigma_n = \sqrt{n} \sigma_1$. Ici $\sigma_2 = \sqrt{2} \sigma_1 = \sqrt{2} \delta / \sqrt{12} = \delta / \sqrt{6}$.

COURANTS CONTINUS:

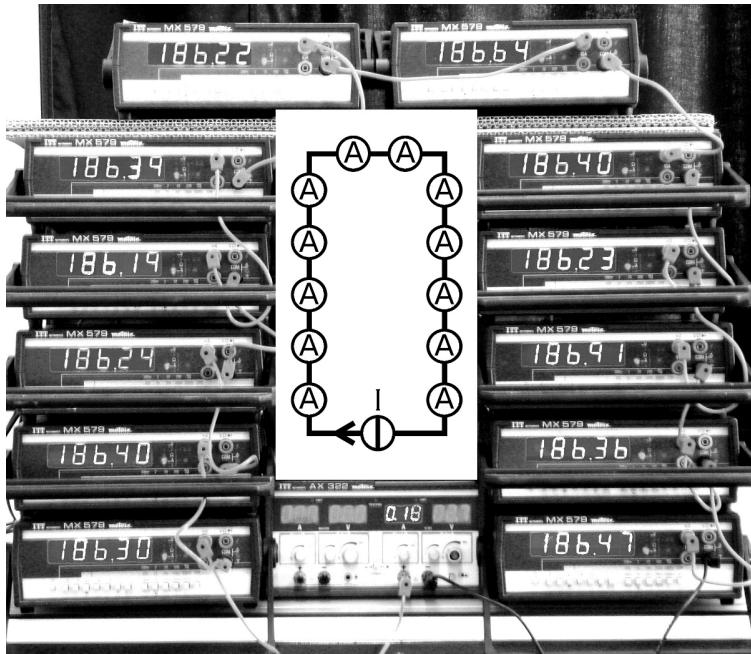
Calibre	Résolution	Précision $\pm(n\%L + mUR^*)$	Coeff. de Ture.	Chute de tension	Protec- tion
200µA	10nA	0,2 % + 3 UR	2E-4/K	<300mV	fusible 2A 10A
2mA	100nA	"	"	"	
20mA	1µA	"	"	"	
200mA	10µA	0,7 % + 3 UR	7E-4/K	"	
2A	100µA	"	"	<700mV	
10A	1mA	"	"	"	

$$\text{D'où } \Delta I = 0,7\% \times 186,30 \text{ mA} + 3 \times 10 \mu\text{A}.$$

Pour simplifier nous pouvons considérer que les indications des fabricants sont données avec **une confiance de 95% et un profil gaussien.**

Alors $I = 186,30 \pm 1,34 \text{ mA}$
 avec un niveau de confiance de 95%
 $(184,96 \text{ mA} < I < 187,64 \text{ mA})$

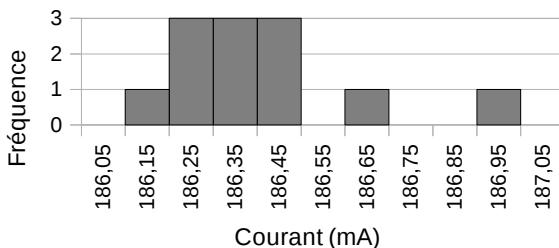
Essayons de retrouver l'erreur systématique donnée par le fabricant en regardant l'erreur aléatoire donnée par 12 ampèremètres de même modèle. Nous les plaçons en série et ainsi ils mesurent tous la même grandeur :



Le calcul de l'écart-type donne : $s_I = 0,21 \text{ mA}$, le coefficient de Student vaut $t = 2,2$, d'où $\Delta I = t \cdot s = 0,45 \text{ mA}$.

Soit $I = 186,40 \pm 0,45 \text{ mA}$ avec un niveau de confiance de 95% ($185,94 \text{ mA} < I < 186,85 \text{ mA}$).

L'incertitude est ici, à peu près, 3 fois plus faible, ce qui est normal car nous avons une meilleure connaissance de la moyenne ($\Delta \bar{I} = t \cdot s / \sqrt{n} = 0,13$



I (mA)
186,30
186,40
186,24
186,19
186,39
186,22
186,64
186,40
186,23
186,91
186,36
186,47

mA). Si nous considérons l'ampèremètre pour lequel $I=186,91\text{mA}$, seule l'incertitude systématique plus grande fournie par le constructeur garantie un bon encadrement.

Par ailleurs l'incertitude du constructeur concerne l'ensemble des ampèremètres de ce modèle qu'il produit et pas seulement ceux livrés au lycée. Aussi, nous devons considérer, pour la comparaison des résultats, que les 12 ampèremètres n'ont pas été recalibrés avec un étalon depuis leur achat d'il y a plus de 15 ans.

2. Normes

Pour déterminer rigoureusement l'incertitude, il faudrait que tous les fabricants donnent la confiance de leurs incertitudes et les profils des lois de probabilité.

Pour contribuer à une information complète sur l'expression de l'incertitude et fournir une base pour la comparaison internationale des résultats de mesure *Le comité international des poids et mesures* et *l'Organisation internationale de normalisation* ont développés un guide [vii].

Il s'agit de mettre en place des méthodes d'évaluation et d'expression des incertitudes harmonisées pour tous les pays et tous les domaines (de la santé, de la sécurité, industriel, commercial, etc.).

3. Vocabulaire

Que l'on parle avec le vocabulaire de la statistique ou avec celui de la métrologie les notions sont les mêmes. Pour aider à faire la jonction, nous avons synthétisé certaines expressions courantes dans un tableau. Sur une même ligne vous trouverez les termes équivalents.

statistique [vi] [iv]		métrologie [v] [vii] [viii]		
grandeur à mesurer X		mesurande M		
mesure	action de mesurer		mesurage	
ensemble des mesures $\{x_i \pm \Delta x_i\}$		résultat du mesurage		
x_i	une mesure		m_i	
$X = \bar{x} \pm \Delta x$	résultat d'une mesure		$M = m \pm \Delta M$	
dispersion	fidélité (répétabilité et reproductibilité)	erreurs aléatoires (accidentielles)	erreurs de type A	
justesse	biais	erreurs systématiques	erreurs de type B	
résolution (sensibilité)	(erreur de) discrétisation			
μ (espérance)	moyenne de la population (souvent inconnue)	la valeur vraie M_{vrai}		
\bar{x}	moyenne de l'échantillon	\bar{m} ou m		
σ : écart-type de la population (souvent inconnu)				
s	s_x	écart-type de l'échantillon	écart-type expérimental	
s / \sqrt{n}	$s_{\bar{x}}$	écart-type estimé de la distribution d'échantillonnage	incertitude- type	
coefficient de Student t $\Delta x = t \cdot s / \sqrt{n}$		facteur d'élargissement k $\Delta M = k \cdot s$		
Δx	incertitude (incertitude absolue)	incertitude-type élargie	ΔM	
$\Delta x / \bar{x}$	précision (incertitude relative)		$\Delta M / m$	

Remarques :

- Plus le biais est grand, moins la mesure est juste.
- Plus les mesures sont dispersées, moins la fidélité est bonne.
- L'écart-type de l'échantillon est un estimateur biaisé de l'écart-type de la population.
- Dans ce livre nous avons évité de parler d'*erreur*, nous avons préféré le terme *incertitude*. En métrologie, nous parlons de différentes erreurs (erreur de mesure, erreur aléatoire et erreur systématique).
- Nous utilisons le terme écart pour la différence entre la moyenne de la mesure et la valeur attendue (une valeur donnée par le fabricant ou dans les tables).

4. Méthode simplifiée d'évaluation de l'incertitude de type B

Si le fabricant donne des limites et que l'on ne dispose d'aucune autre information, il est raisonnable de supposer la probabilité égale sur l'intervalle et nulle en dehors. Nous utilisons alors les caractéristiques d'une loi uniforme ([vii] p23 et [v] p7) :

Cas :	Incertitude-type :	Incertitude élargie :
Appareil à cadran, lecture d'un régllet, digit, etc. Résolution : $\delta = x_{max} - x_{min}$, $x = \bar{x} \pm \delta/2$	$s_{\bar{x}} = \frac{\delta}{\sqrt{12}}$	$\Delta x = k s_{\bar{x}}$
Le constructeur fournit la classe : $\pm a$ a correspond à $\delta/2$	$s_{\bar{x}} = \frac{a}{\sqrt{3}}$	($k=2$ à 95%)

Remarques :

- Pour le premier cas nous ne disposons pas d'information du constructeur, dans le second, il fournit ce que l'on appelle la classe *a* de l'appareil.
- Nous prenons k=2 avec une confiance de 95% comme si nous avions une gaussienne (approximation).
- Si la notice est conçue selon cette norme, nous aurons pour l'ampèremètre (étudié page 188), $s_i = 1,34 \text{ mA}$ et $I = 186,30 \pm 2,68 \text{ mA}$ à 95%.
- Sur le matériel de verrerie jaugé, la classe est toujours indiquée.

25ml

Par exemple, sur une pipette jaugée de 25 mL, nous avons les indications suivantes : Ex 20°C ± 0,06ml

La classe est 0,06 mL, d'où $s_v = 0,035 \text{ mL}$ et $V = 25 \pm 0,07 \text{ mL}$ avec une confiance de 95%. La précision sur le volume est de 0,28%.

- Pour une résistance R de 1000Ω et d'une précision de 5% (brun, noir, rouge, or), nous avons :
 $\Delta R = 2 \cdot 1000 \cdot (5/100) / \sqrt{3} \approx 58\Omega$ à 95% de confiance.
 $R = 1000 \pm 58\Omega$ et $\Delta R/R \approx 5,8\%$.
- Pesée : une balance numérique, de résolution 0,1g, indique une masse m de 73,7g.
 $\Delta m = 2 \cdot 0,1 / \sqrt{12} \approx 0,06 \text{ g}$, $m = 73,70 \pm 0,06 \text{ g}$ à 95%. $\Delta m/m \approx 0,08\%$
- Par la méthode d'autocollimation, nous mesurons la distance focale f' d'une lentille convergente mince. L'image est nette (latitude de mise au point) de $f'_{min} = 195 \text{ mm}$ à $f'_{max} = 203 \text{ mm}$. D'où $s_{optique} = 2,31 \text{ mm}$.
 $s_{géométrique} = 0,29 \text{ mm}$ (δ de 1mm du banc d'optique) et $s_{modélisation} = 0,58 \text{ mm}$ (épaisseur de 2mm de la lentille).

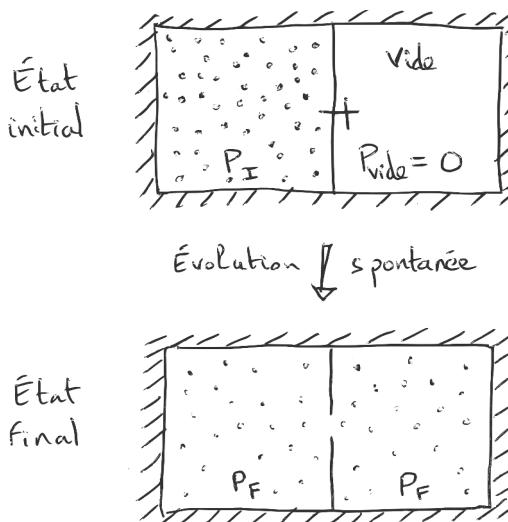
$$s = \sqrt{s_o^2 + s_g^2 + s_m^2} \quad f' = 199,0 \pm 4,8 \text{ mm} (95\%)$$

C. Thermodynamique

Nous allons étudier la détente de Joule et Gay-Lussac. Pour cette détente, nous allons utiliser deux méthodes pour déterminer les fluctuations de la pression dans l'état final : l'une classique en thermodynamique, l'autre en utilisant la formule de propagation. Dans ce cadre, nous montrerons que cette formule est en accord avec les lois de la physique.

1. Description de la détente de Joule Gay-Lussac

Nous avons deux enceintes, aux parois indéformables et calorifugées, pouvant communiquer au moyen d'un robinet. Initialement un fluide est enfermé dans l'un des compartiments, l'autre compartiment est vide. Nous ouvrons le robinet et une détente dans le vide se produit. La pression s'égalise.



Nous considérerons le cas où les deux enceintes ont le même volume.

De plus, le fluide étudié sera un gaz parfait : le premier principe de la thermodynamique permet alors de montrer que la température finale est égale à la température initiale : $T_F=T_I$; et pour les volumes $V_F=2V_I$.

2. Hypothèses du gaz parfait

Les molécules sont considérées ponctuelles, identiques et sans interactions entre elles. Elles évoluent donc, de manière tout à fait indépendante, elles s'ignorent mutuellement, et les seuls contacts des particules sont avec les parois des récipients. Leurs chocs sur les parois exercent une pression, celle-ci dépend de la densité des particules et de la température :

$$PV=nRT \Rightarrow P=n^*k_B T$$

$$\text{avec } N=n \cdot N_A \text{ et } n^*=N/V$$

n : nombre de moles de gaz dans la boîte

N : nombre de molécules de gaz dans la boîte

n^* : densité de particules

(nombre de particules par unité de volume)

$k_B=R/N_A$: constante de Boltzmann

R : constante des gaz parfaits N_A : nombre d'Avogadro

Nous avons, ici, les mêmes températures initiale et finale, ainsi, $P_F=P_I/2$, et la pression est directement proportionnelle à la densité de particules :

$$P \propto n^*$$

3. Hypothèses de la physique statistique

Nous distinguons l'état macroscopique du gaz, appelé macroétat (décrit par les variables d'états de la thermodynamique, telles la pression et la température) de tous les états microscopiques correspondants, appelés microétats (décrit par les positions et les vitesses individuelles de toutes les molécules du gaz). Un expérimentateur (macroscopique) qui observe la détente ne peut mesurer que le nombre total de particules (et donc la pression) dans chaque compartiment. Il n'a pas accès à l'ensemble des informations microscopiques du système : par exemple, il ne sait pas si une particule donnée A est dans le compartiment de gauche ou de droite, il peut seulement en connaître la probabilité.

Nous noterons k le nombre total de molécules dans le compartiment de gauche.

Imaginons notre gaz constitué de cette seule particule, deux possibilités :

$$k=0 \quad \boxed{\text{A}} \quad p=1/2 \quad \text{ou} \quad k=1 \quad \boxed{\text{A}} \quad p=1/2$$

Les deux microétats possibles ont la même probabilité p , et correspondent chacun à un macroétat.

Notre gaz a maintenant deux particules, $N=2$:

macroétats	$k=0$	$k=1$	$k=2$
microétats	$\boxed{\text{A}, \text{B}}$	$\boxed{\begin{matrix} \text{A} & \text{B} \\ \hline \text{B} & \text{A} \end{matrix}}$	$\boxed{\text{A}, \text{B},}$
probabilités	$p=1/4$	$p=1/2$	$p=1/4$

Le macroétat $k=1$ est représenté par deux microétats.

Regardons aussi le cas $N=3$:

macroétats	$k=0$	$k=1$	$k=2$	$k=3$
microétats				
probas	$p=1/8$	$p=3/8$	$p=3/8$	$p=1/8$

Allons encore plus loin pour $N=4$:

macroétats	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
microétats	(,ABCD)	(A,BCD) (B,ACD) (C,ABD) (D,ABC)	(AB,CD) (AC,BD) (AD,BC) (BC,AD) (BD,AC) (CD,AB)	(BCD,A) (ACD,B) (ABD,C) (ABC,D)	(ABCD,)
probas	$p=1/16$	$p=1/4$	$p=3/8$	$p=1/4$	$p=1/16$

Nous voyons que, plus le nombre de particules est grand, plus les macroétats autour de $k \simeq N/2$ sont probables.

Nous avons deux principes physiques :

- les microétats sont équiprobables.
- hypothèse ergodique : au cours du temps tous les microétats sont indifféremment balayés et le système reste un temps équivalent dans chacun deux.

4. État le plus probable et fluctuations

Nous comprenons maintenant que le système sera, la plupart du temps, dans le macroétat $k=N/2$ (prenant N pair, ce qui ne restreint pas notre étude car N est très grand et nous ne sommes donc pas à une particule près).

Nombre de microétats Ω et probabilité p

pour un macroétat k donné :

$$\Omega = C_N^k = \frac{N!}{k!(N-k)!} \quad \text{et} \quad p(k) = \frac{C_N^k}{2^N}$$

Finalement, cela va se passer autour de $N/2$, alors symétrisons, et posons $k=N/2+i$, d'où : $\Omega = N!/(N/2+i)!(N/2-i)!$

Nous cherchons l'expression de $p(i)$ pour N grand et i petit, utilisons la formule de Stirling $\ln(n!) \approx n \ln(n) - n$ avec n grand et le développement limité $\ln(1+\varepsilon) \approx \varepsilon$ avec ε petit ; nous avons alors : $\ln[p(i)] \approx -4i^2/N$ et $p(i) \propto e^{-4i^2/N}$.

La probabilité obtenue obéit à une loi gaussienne.

Déterminons l'écart-type : $\frac{p(\pm\sigma)}{p_{max}} = \frac{1}{\sqrt{e}}$ d'où $\sigma \propto \frac{\sqrt{N}}{2}$.

Nous avons l'incertitude relative à un sigma pour le nombre de particules et donc aussi pour la pression :

$$\frac{\sigma_{N/2}}{N/2} = \frac{1}{\sqrt{N}} \quad \text{et} \quad \frac{\sigma_P}{P} = \frac{1}{\sqrt{N}}$$

N est grand, ainsi la pression fluctue très peu et est parfaitement définie.

5. Utilisation de la formule de propagation

Nous avons N particules de pression p_j qui contribuent, ou pas, à la pression du compartiment de droite : $P = \sum_{j=i}^N p_j$.

Si la particule est dans le compartiment de gauche $p_j=0$, si elle est à droite $p_j=P_F/(N/2)$. Donc $\sigma_p = P_F/N$ (même principe que les pièces page 8). En utilisant la formule de propagation des écart-types nous obtenons : $\sigma_p = \sqrt{N} \sigma_p$

$$\text{et } \frac{\sigma_p}{P} = \frac{1}{\sqrt{N}}$$

6. Conclusion

La formule de propagation permet de retrouver rapidement le résultat de la physique statistique.

Si nous faisions l'erreur de considérer que les incertitudes s'ajoutent nous aurions $P_F = P_F \pm P_F$. Ainsi la pression et toutes les variables d'état de la thermodynamique ne seraient pas définies. De plus, cette détente, démontrée irréversible en vertu du second principe, deviendrait réversible.

Nous voyons à nouveau, sur cet exemple, que la formule de propagation n'est pas seulement un outil rigoureux pour le calcul des incertitudes mais participe, aussi, aux fondements des lois de la nature.

D. Indépendance des variables

Considérons un système. Nous mesurons deux grandeurs associées à celui-ci. Normalement les deux processus de mesure s'effectuent indépendamment. L'action de mesurer la première grandeur fournit un ensemble d'observations relatives à cette grandeur. Ces observations ont un centre et une dispersion. Dans un second temps, nous mesurons le deuxième grandeur et il n'y a aucune raison que cette deuxième collecte d'informations soit perturbée par la première. Les fluctuations de la première ne sont pas corrélées avec celles de la deuxième.

En effet, la plupart du temps, le mesurage ne modifie pas le système et les différents processus de mesure n'interagissent pas entre eux.

Exemple : nous avons un système électrique en régime stationnaire constitué d'un générateur de tension continue U qui alimente une résistance R parcourue par un courant I . Nous avons la relation théorique $U=RI$ et nous souhaitons estimer R .

Nous effectuons 100 observations de la tension U (nous pouvons prendre 100 voltmètres différents) et nous en déduisons \bar{U} et ΔU . Dans un second temps, nous procédons de même pour I , d'où \bar{I} et ΔI . Nécessairement, le coefficient de corrélation r_{UI} entre $\{U_1, \dots, U_{100}\}$ et $\{I_1, \dots, I_{100}\}$ est nul. Les grandeurs U et I sont indépendantes. D'où $\bar{R}=\bar{U}/\bar{I}$ et $\Delta R/R=\sqrt{(\Delta U/U^2+\Delta I/I^2)}$.

Deuxième exemple : nous lançons quatre dés équilibrés six faces de couleurs différentes. Nous calculons la somme : $S=d_1+d_2+d_3+d_4$. Si les lancers des trois premiers dés donnent à

chaque fois 6 cela ne préjuge en rien la probabilité d'obtenir 6 avec le quatrième dé, c'est toujours une chance sur six. Nos quatre grandeurs sont indépendantes. Conclusion : quelque soit $i,j=1,2,3,4$ avec $i \neq j$ $r_{ij}=0$ et $\sigma_s^2=\sigma_1^2+\sigma_2^2+\sigma_3^2+\sigma_4^2$.

Il existe une généralisation de la formule de propagation des écart-types dans le cas de grandeurs corrélées mais, dans la plupart des cas, nous n'avons pas à l'utiliser :

$$\sigma_f^2 = \sum_j \left(\frac{\partial f}{\partial x_j} \right)^2 \sigma_j^2 + 2 \sum_{i < j} \left(\frac{\partial f}{\partial x_i} \right) \left(\frac{\partial f}{\partial x_j} \right) r_{ij} \sigma_i \sigma_j$$

Remarques :

- $\sigma_f \leq \sum |\partial f / \partial x_i| \sigma_i$.
- si $f = x_1 + x_2$, $\sigma_1 = \sigma_2$ et $r_{12} = -1$ alors $\sigma_f = 0$.

VII. DEVOIRS

Durée de composition d'une heure pour chaque devoir. Calculatrice autorisée. Documents, ordinateurs et téléphones interdits. Toutes les réponses doivent être justifiées et clairement explicitées. Les tables de Student et de k_{hi} deux sont fournies, ainsi qu'un formulaire avec la formule de l'écart-type et celle du coefficient de corrélation.

Devoirs donnés en deuxième année d'école d'ingénieur (École Hubert Curien / Maîtrise de l'efficacité énergétique).

A. *Devoir Suricate*

corrigé p298

Exercice 1 (3 pts) :

Nous lançons quatre pièces équilibrées.

Quelle est la probabilité d'en avoir quatre qui tombent du côté pile?

Même question pour avoir trois faces et un pile suite au lancer des quatre pièces.

Nous allons maintenant nous intéresser à la composition de différentes eaux minérales :

	Calcium mg/L	Bicarbonates mg/L	Magnésium mg/L	Chlorures mg/L	Sulfates mg/L	Potassium mg/L	Sodium mg/L	pH
Evian	78	357	24	4,5	10	1	5	7,2
Badoit	190	1300	85	44	38	10	165	6
Contrex	486	403	84	10	1187	3,2	9,1	7,2
Hépar	555	403	110	11	1479	4	14	7,2
St-Yorre	90	4368	11	322	174	132	1708	6,6
Vittel	202	402	43	8	336	2	4,7	7,5
Volvic	11,5	71	8	13,5	8,1	6,2	11,6	6,95
Vichy célestin	103	2989	10	235	138	66	1172	6,8
Salvetat	253	820	11	4	25	3	7	6

Exercice 2 (7 pts) :

Nous allons tout d'abord considérer la quantité B de bicarbonates en mg/L :

- a) Déterminez la valeur moyenne \bar{B} de B.
- b) Calculez l'écart-type σ_B sur B.
- c) Avec une confiance de 95%, quelle est l'incertitude ΔB ?

Nous considérons maintenant la concentration S en sulfates.

- d) Les grandeurs B et S sont-elles corrélées ? (si la valeur absolue du coefficient de corrélation r est inférieure à la valeur critique r_c ici égale à 0,52 les grandeurs peuvent être considérées comme indépendantes).
- e) Déterminez la teneur moyenne totale $\{B \oplus S\}$ en mg/L des bicarbonates et sulfates avec son incertitude.

Exercice 3 (3 pts) :

Nous considérons une population d'un pays constituée de pauvres et de non-pauvres. Les pays est découpé en quatre régions. Dans chaque région nous tirons 1000 personnes au hasard et nous obtenons la répartition suivante :

Zone géographique :	Nord - Est	Centre - Nord	Sud	Ouest
Pauvres :	112	105	154	113
Non pauvres :	888	895	846	887

Nous faisons l'hypothèse que la pauvreté est uniformément répartie sur tout le pays.

L'hypothèse est-elle soutenable avec une probabilité critique de 5% ?

Exercice 4 (3 pts) :

Nous traçons maintenant pour les différentes eaux minérales la quantité de sodium en fonction de celle de chlorure. Une régression nous donne la pente et l'ordonnée à l'origine avec leurs incertitudes à 95% de confiance. Pour l'ordonnée à l'origine nous trouvons : $b = -43 \pm 25$ mg/L. Nous vous proposons des possibilités pour la pente dont une est la bonne, expliquez votre choix et pourquoi les autres ne peuvent convenir.

Pour ce faire vous devez vous aider d'une matrice de l'exercice 5. D'autres données qui peuvent vous aider : $M_{Na}=23,0\text{g/mol}$, $M_{Cl}=35,5\text{g/mol}$. Vous pourrez conclure sur l'information que l'on en tire sur les eaux minérales en général.

Cas :	1)	2)	3)	4)	5)	6)	7)	8)
pente a :	-5,34	5,34	-0,65	0,65	5,34	-5,34	0,65	-0,65
Δa :	1,19	0,19	0,19	1,19	1,19	0,19	0,19	1,19

Exercice 5 (8 pts dont 4 en bonus) :

Nous voulons réaliser une analyse en composantes principales avec les eaux minérales mentionnées sur la première page du devoir. Montrez sur l'exemple de x_7^2 comment on passe de la matrice de données \mathbf{M} à la matrice de données réduite \mathbf{m} suivante :

-0,753	-0,597	-0,475	-0,569	-0,660	-0,539	-0,532	0,693
-0,154	0,044	1,059	-0,238	-0,609	-0,339	-0,281	-1,542
1,430	-0,566	1,034	-0,523	1,454	-0,490	-0,525	0,693
1,799	-0,566	1,687	-0,514	1,979	-0,472	-0,518	0,693
-0,689	2,132	-0,802	2,089	-0,365	2,371	2,139	-0,424
-0,089	-0,567	0,003	-0,540	-0,074	-0,517	-0,532	1,252
-1,109	-0,792	-0,877	-0,494	-0,663	-0,424	-0,521	0,228
-0,619	1,194	-0,827	1,361	-0,430	0,905	1,298	-0,052
0,183	-0,282	-0,802	-0,573	-0,633	-0,495	-0,529	-1,542

Quel est l'intérêt d'une telle transformation ?

Nous construisons ensuite la matrice des corrélations \mathbf{r} . Complétez les valeurs manquantes :

?	-0,339	0,825	-0,375	0,922	-0,358	-0,373	0,204
?	?	-0,373	0,982	?	0,969	0,984	-0,336
?	?	?	-0,413	0,781	-0,411	-0,421	0,168
?	?	?	?	-0,234	0,980	0,999	-0,201
?	?	?	?	?	-0,213	-0,236	0,471
?	?	?	?	?	?	0,986	-0,194
?	?	?	?	?	?	?	-0,200
?	?	?	?	?	?	?	?

Nous diagonalisons ensuite la matrice **r** et nous donnons les quatre premières valeurs propres :

$$\lambda_1 = 4,76; \lambda_2 = 2,05; \lambda_3 = 0,95 \text{ et } \lambda_4 = 0,20.$$

Que pensez-vous des autres valeurs propres en terme d'information ?

Expliquez le principe qui permet d'obtenir les valeurs propres.

Nous obtenons ensuite la matrice de passage **P** (en dessous), les colonnes sont les vecteurs propres en fonction des vecteurs de la base initiale.

-0,305	0,472	-0,223	0,481	-0,457	-0,283	-0,278	0,192
0,422	0,258	-0,108	-0,063	-0,442	-0,173	0,502	-0,512
-0,312	0,402	-0,286	-0,806	0,006	-0,050	-0,051	0,068
0,422	0,265	0,050	-0,059	-0,104	0,493	-0,657	-0,244
-0,267	0,553	0,089	0,289	0,492	0,366	0,316	-0,236
0,417	0,272	0,055	0,015	0,531	-0,652	-0,203	0,030
0,423	0,264	0,053	-0,022	-0,085	0,256	0,307	0,762
-0,169	0,171	0,917	-0,166	-0,231	-0,143	-0,025	-0,002

Expliquez le principe qui permet d'obtenir la matrice de passage.

Comment à partir de cette matrice déterminons-nous la nouvelle matrice \mathbf{m}' des individus avec les nouvelles caractéristiques ?

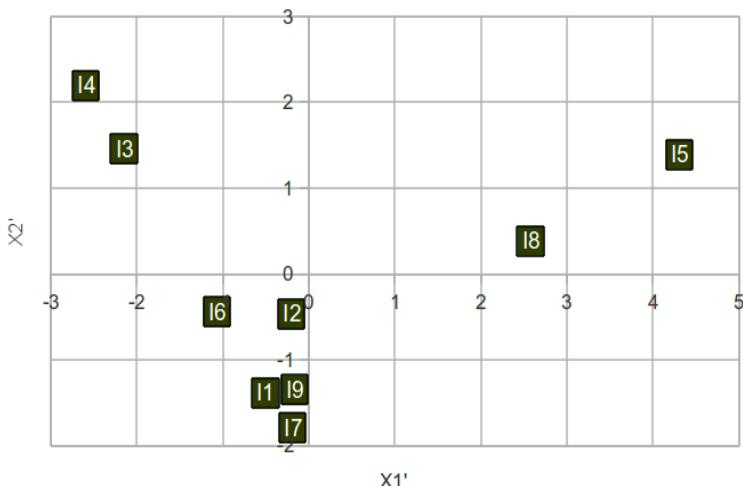
Illustrez en calculant l'élément x'_2 .

Nous avons restitué les trois premières colonnes de \mathbf{m}' :

-0,50	-1,38	0,86	...
-0,20	-0,47	-1,79	...
-2,15	1,46	0,13	...
-2,59	2,19	-0,09	...
4,31	1,39	0,08	...
-1,07	-0,44	1,14	...
-0,19	-1,79	0,66	...
2,58	0,39	0,35	...
-0,17	-1,35	-1,33	...
-0,50	-1,38	0,86	...
-0,20	-0,47	-1,79	...
-2,15	1,46	0,13	...
-2,59	2,19	-0,09	...
4,31	1,39	0,08	...
-1,07	-0,44	1,14	...
-0,19	-1,79	0,66	...
2,58	0,39	0,35	...
-0,17	-1,35	-1,33	...

Finalement nous obtenons la distribution des individus :

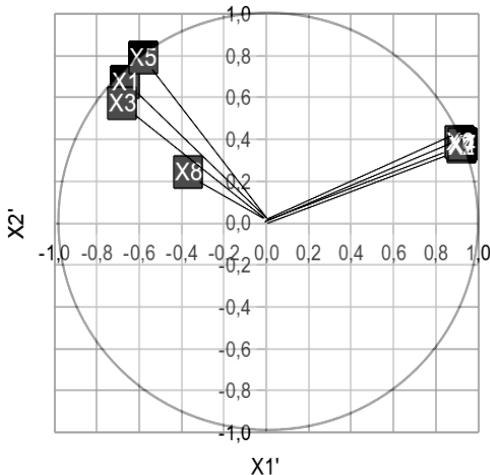
Eaux selon les deux composantes principales



Quel sens donneriez-vous à X_1' et X_2' ?

Vous pouvez aussi vous aider des corrélations entre nouvelles et anciennes variables, et du cercle des corrélations.

	X_1'	X_2'	X_3'
x_1	-0,666	0,675	-0,218 ...
x_2	0,920	0,369	-0,105 ...
x_3	-0,680	0,575	-0,279 ...
x_4	0,921	0,379	0,049 ...
x_5	-0,583	0,791	0,087 ...
x_6	0,910	0,389	0,054 ...
x_7	0,923	0,378	0,051 ...
x_8	-0,368	0,245	0,893 ...



Quels enseignements tirez-vous concernant les différents types d'eaux ?

Apparaît-il une classification particulière ? Par exemple selon la minéralisation, eau gazeuse, etc.

Que pensez-vous de X3' ?

B. Devoir Narval

corrigé p302

Exercice 1 (3 pts) :

Nous lançons deux dés équilibrés à six faces.

Quelle est la probabilité que la somme des deux dés fasse 12 ?

Même question pour une somme égale à 7.

Exercice 2 (7 pts) :

Pour huit maisons du même modèle nous mesurons les résistances thermiques. Pour les murs nous mesurons, de la première à la huitième maison, les valeurs R_1 suivantes en mK/W :

39	41	41	43	41	37	39	39
----	----	----	----	----	----	----	----

- Déterminez la valeur moyenne \bar{R}_1 de R_1 .
- Calculez l'écart-type σ_1 sur R_1 .
- Avec une confiance de 95%, quelle est l'incertitude ΔR_1 ?

Nous mesurons maintenant les résistances thermiques R_2 des huisseries. Pour les mêmes maisons, dans le même ordre, les valeurs mesurées en mK/W :

104	102	98	102	98	102	96	98
-----	-----	----	-----	----	-----	----	----

- Les grandeurs R_1 et R_2 sont-elles corrélées ?
- Déterminez la résistance thermique R de l'ensemble {murs \oplus huisseries} avec son incertitude.

Exercice 3 (3 pts) :

Nous faisons l'hypothèse d'égalité des salaires hommes-femmes au sein de l'entreprise suivante :

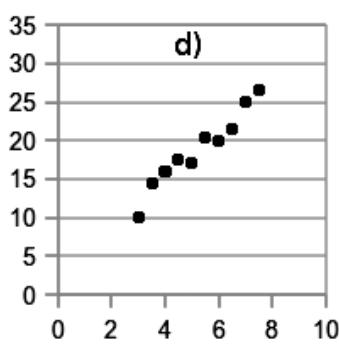
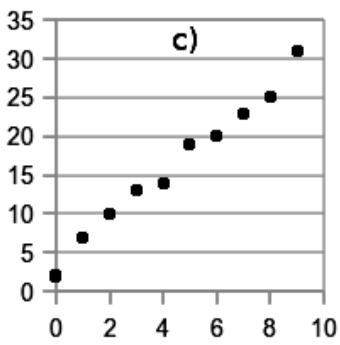
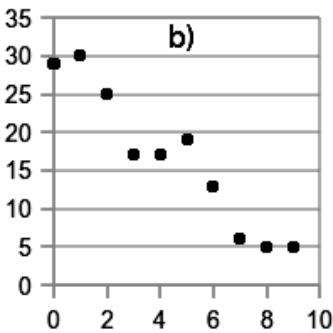
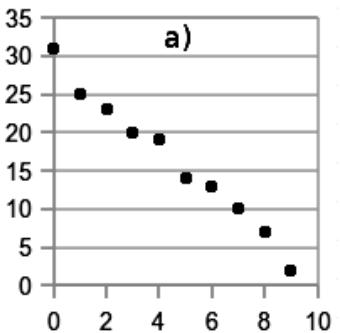
Salaires en € :	1000 à 1999	2000 à 2999	3000 à 3999	4000 à 4999
Hommes (effectifs):	139	124	68	29
Femmes (effectifs) :	261	66	22	11

L'hypothèse est-elle soutenable avec une probabilité critique de 5% ?

Exercice 4 (3 pts) :

Pour les courbes suivantes de température T ($^{\circ}\text{C}$) dans une habitation en fonction du temps t(h), indiquez si des cas du tableau correspondent à des graphes. Dans les cas positifs donnez la correspondance, sinon, expliquez pourquoi cela n'est pas possible. N'oubliez pas que vos réponses doivent être scientifiquement justifiées.

Cas :	1)	2)	3)	4)
pente a :	3,16	-5,80	2,90	-3,01
incertitude sur a :	0,63	0,31	0,31	0,69
ordonnée à l'origine b :	2,27	20,44	3,36	30,13
incertitude sur b :	3,40	1,70	1,70	3,70
coeff. de corrélation :	0,9712	-0,9914	0,9914	-0,9627



Exercice 5 (8 pts dont 4 en bonus) : Isolants

Nous voulons réaliser une analyse en composantes principales avec les huit isolants du tableau :

Données :	éco-bilan		iso-lation	confort d'été
	constr- uction	démo- lition		
1. Vermiculite	-	.	-	.
2. Polyuréthane	--	--	++	.
3. Liège expansé	.	+	+	++
4. Laine de roche	-	-	+	.
5. Laine de mouton	.	+	+	+
6. Isolant mince	--	--	--	--
7. Béton cellulaire	.	.	+	++
8. Botte de paille	+	+	+	++

Pour chiffrer, nous transformons les appréciations (-- , - , . , + , ++) en notes (-2, -1, 0 , 1 , 2).

Montrez sur l'exemple de x_1^1 comment on passe de la matrice de données \mathbf{M} à la matrice de données réduite \mathbf{m} suivante :

	X_1	X_2	X_3	X_4
I_1	$X_1^1 = ?$	$X_2^1 = 0,195$	$X_3^1 = -1,146$	$X_4^1 = -0,444$
I_2	$X_1^2 = -1,296$	$X_2^2 = -1,365$	$X_3^2 = 1,146$	$X_4^2 = -0,444$
I_3	$X_1^3 = 0,589$	$X_2^3 = 0,975$	$X_3^3 = 0,382$	$X_4^3 = 0,977$
I_4	$X_1^4 = -0,354$	$X_2^4 = -0,585$	$X_3^4 = 0,382$	$X_4^4 = -0,444$
I_5	$X_1^5 = 0,589$	$X_2^5 = 0,975$	$X_3^5 = 0,382$	$X_4^5 = 0,266$
I_6	$X_1^6 = -1,296$	$X_2^6 = -1,365$	$X_3^6 = -1,909$	$X_4^6 = -1,864$
I_7	$X_1^7 = 0,589$	$X_2^7 = 0,195$	$X_3^7 = 0,382$	$X_4^7 = 0,977$
I_8	$X_1^8 = 1,532$	$X_2^8 = 0,975$	$X_3^8 = 0,382$	$X_4^8 = 0,977$

Quel est l'intérêt d'une telle transformation ?

Nous construisons ensuite la matrice des corrélations \mathbf{r} .
Complétez les valeurs manquantes :

?	?	0,360	0,873
?	?	0,255	0,811
?	?	?	0,659
?	?	?	?

Nous diagonalisons ensuite la matrice **r** et nous obtenons les valeurs propres :

$$\lambda_2 = 0,8552; \lambda_3 = 0,0779 \text{ et } \lambda_4 = 0,0588.$$

Que vaut λ_1 ?

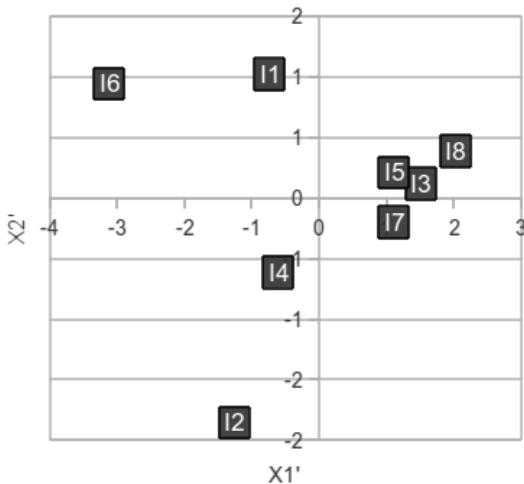
Expliquez le principe qui permet d'obtenir les valeurs propres. Nous obtenons ensuite la matrice de passage **P**, les colonnes sont les vecteurs propres en fonction des vecteurs de la base initiale :

0,544	0,285	0,488	0,620
0,520	0,411	-0,747	-0,056
0,347	-0,856	-0,251	0,287
0,560	-0,127	0,375	-0,728

Expliquez le principe qui permet d'obtenir la matrice de passage. Comment à partir de cette matrice nous déterminons la nouvelle matrice **m'** des individus avec les nouvelles caractéristiques ? Illustrez en calculant l'élément x'_2 .

	X_1'	X_2'	X_3'	X_4'
I_1	-0,738	1,017	-0,197	-0,236
I_2	-1,265	-1,856	-0,068	-0,075
I_3	1,507	0,118	-0,171	-0,291
I_4	-0,612	-0,612	0,002	0,247
I_5	1,109	0,208	-0,437	0,226
I_6	-3,122	0,941	0,168	0,081
I_7	1,101	-0,203	0,412	-0,247
I_8	2,019	0,387	0,290	0,294

Finalement nous obtenons la distribution suivante des individus:

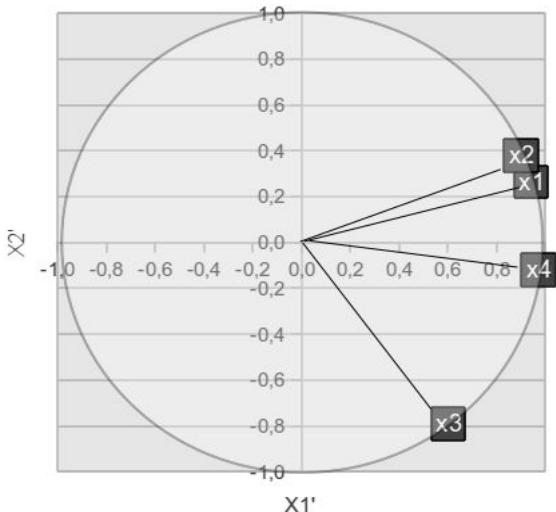


Quel sens donneriez-vous à X_1' et X_2' ?

En omettant X_3' et X_4' avons-nous perdu de l'information ?

Vous pouvez aussi vous aider des corrélations entre nouvelles et anciennes variables, et du cercle des corrélations :

	X_1'	X_2'	X_3'	X_4'
X_1	$r_{11'} = 0,943$	$r_{12'} = 0,264$	$r_{13'} = 0,136$	$r_{14'} = 0,150$
X_2	$r_{21'} = 0,901$	$r_{22'} = 0,380$	$r_{23'} = -0,208$	$r_{24'} = -0,014$
X_3	$r_{31'} = 0,602$	$r_{32'} = -0,792$	$r_{33'} = -0,070$	$r_{34'} = 0,070$
X_4	$r_{41'} = 0,972$	$r_{42'} = -0,117$	$r_{43'} = 0,105$	$r_{44'} = -0,177$



Quels enseignements tirez-vous concernant les différents types d'isolants ?

Apparaît-il une classification particulière selon que nous ayons une isolation répartie, un isolant synthétique, une laine minérale ou un isolant sain (biosourcé) ?

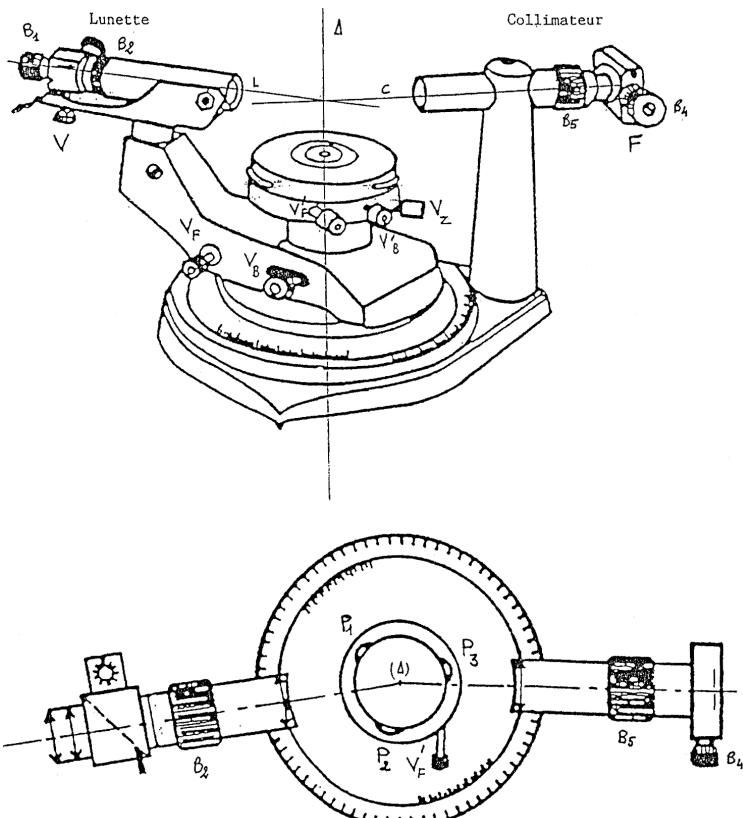
VIII. TRAVAUX PRATIQUES

A. Mesure d'un indice lumineux

(TP réalisé au Lycée Alain-Fournier de Bourges)

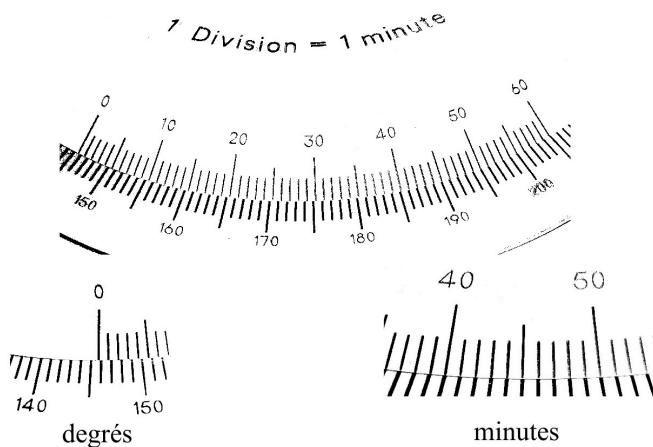
Utilisation du goniomètre et étude du prisme

Un goniomètre permet de mesurer des angles dans un plan horizontal. Il est constitué de trois parties principales mobiles autour d'un axe central vertical : le collimateur, la lunette et la plateforme.



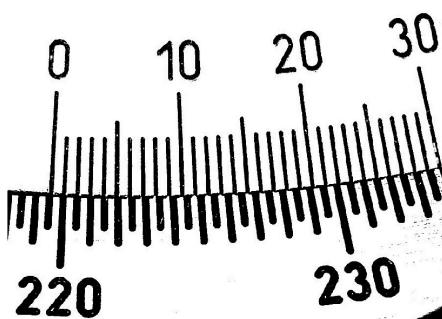
Lecture du vernier

Vernier sur 60' :



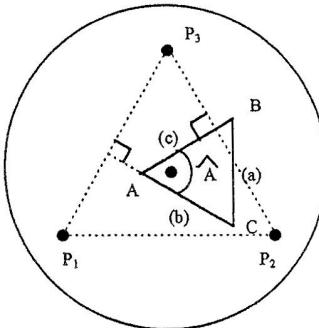
Vernier sur 30' :

Déterminer les deux
abscisses angulaires en
degré et minutes (...° ..')



Mesure de l'angle au sommet du prisme par réflexion :

Mise en place du prisme :



Méthode par auto-collimation :

Placer la face (a) approximativement perpendiculaire à l'axe optique C du collimateur. Bloquer la plate-forme avec la vis de blocage V_a' .

Avec la lunette, viser la face B du prisme de façon à ce quelle soit perpendiculaire à (b). Déterminer avec précision la position angulaire de la lunette pour laquelle le fil vertical du réticule et son image par réflexion sur (b) coïncident : pour cela, on bloquera la rotation manuelle de la lunette avec la vis de blocage V_B au voisinage de la bonne position et on ajustera avec la vis de réglage fin V_F . Noter alors l'abscisse angulaire β_1 .

Faire de même avec la face (c). Noter β_2 correspondant.

Étude théorique :

Montrer que l'angle entre les deux positions de la lunette est égal à $\pi - A$.

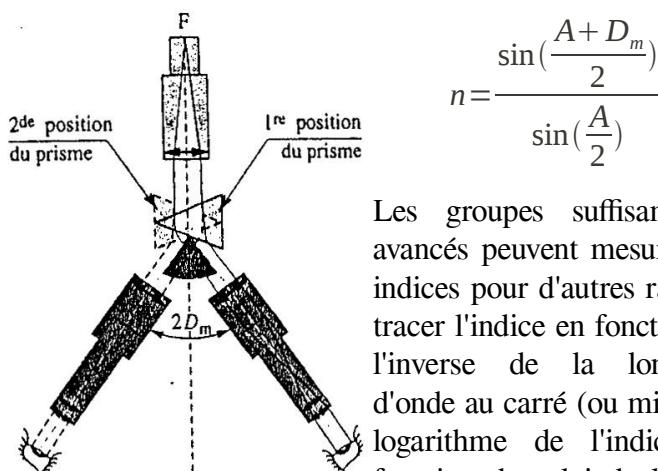
Exploitation des mesures : Déduire de ce qui précède A mesuré avec son incertitude ΔA .

Minimum de déviation et indice du prisme :

Nous plaçons la lampe à sodium derrière le collimateur. Nous obtenons un spectre de raies à la sortie du prisme. De la mesure de la déviation minimale, connaissant A, nous en déduisons l'indice via la formule du prisme donnée plus bas. Déterminez l'indice pour les raies rouge, jaune et vert-jaune. Garder le nombre de chiffres significatifs qui vous semble pertinent.

Comment évolue l'indice avec la longueur d'onde ?

Comment s'appelle ce phénomène ?



Les groupes suffisamment avancés peuvent mesurer les indices pour d'autres raies et tracer l'indice en fonction de l'inverse de la longueur d'onde au carré (ou mieux le logarithme de l'indice en fonction de celui de la longueur d'onde). Conclusion ?

raies	1	2	3	4	5	6	7
	rouge	jaune	vert-jaune	vert	bleu-vert	bleu-violet	violet
$\lambda(\text{nm})$	615,7	589,3	568,5	515,2	498,1	475,0	466,7

B. Le miroir sphérique

Nous considérons deux miroirs concaves M_1 et M_2 .

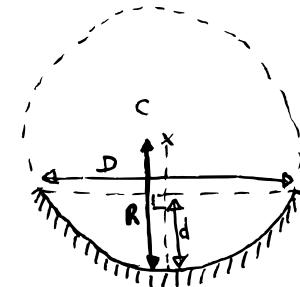
Mesure géométrique du rayon

Théorie :

Exprimez R en fonction de d et D .

Expérience : Mesurez d et D .

Estimez le rayon R avec son incertitude ΔR .



Mesure mécanique du rayon

La pulsation de petites oscillations d'une bille, sans amortissement, dans une cuvette circulaire, est $\omega_0 = \sqrt{g/R}$. Estimez R .

Mesure optique du rayon

Construction géométrique et relation de conjugaison :

Montrez que $A=A'$ avec $\gamma=-1$ pour $A=A'=C$.

Sachant cela, en déduire un protocole permettant la mesure de R .

Expérience : Mesure de $R \pm \Delta R$.

Comparez avec les autres méthodes.

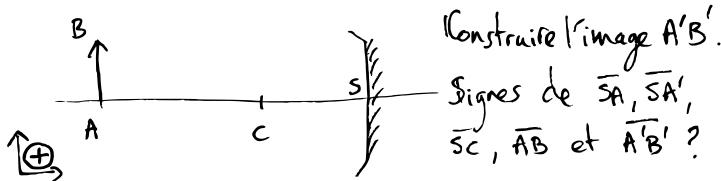
Mesure optique de la focale

On utilisera une lentille convergente telle que $A=F$ pour disposer d'un faisceau parallèle ($A'=\infty$).

En déduire une méthode de mesure de la focale f . Faire la construction.

Estimez $f \pm \Delta f$, et comparez aux résultats obtenus pour R .

Relation de conjugaison



$-\bar{SA}$	(min)	(max)	(C)	(Points Intermédiaires)
$-\bar{SA}'$	(max)	(min)	(C)	
$\Delta\bar{SA}$
$\Delta\bar{SA}'$.			
$x = -1/\bar{SA}$.			
$y = -1/\bar{SA}'$				
Δx				
Δy				

Montrez que, d'après la théorie, nous avons une relation linéaire entre x et y . Si l'on note $y=ax+b$, à quoi doivent correspondre a et b ?

Tracer les points sur une feuille de papier millimétrée (minimum 5).

Comment estimer les incertitudes sur x et y ?

Ceci fait les rectangles d'incertitudes seront placés sur le graphique, et les droites extrêmes passant au mieux par les rectangles permettront de déterminer $a \pm \Delta a$ et $b \pm \Delta b$.

Comparaison aux valeurs attendues.

Les groupes suffisamment avancés pourront ensuite rentrer les points sur ordinateur dans le tableur prévu à cet effet et procéder par itération.

Conclusion.

C. Relation de conjugaison d'une lentille

Soit une lentille mince sphérique d'une focale de +200mm (valeur annoncée par le fabricant). Nous voulons vérifier la relation de conjugaison, et calculer la focale à partir de 8 points expérimentaux mesurés sur un banc optique de 2m. Nous plaçons successivement la lampe avec l'objet lumineux, la lentille puis l'écran (nous avons ainsi un objet et une image réels).

L'axe optique est orienté suivant la direction du rayon incident. Montrez que, d'après la théorie, nous avons une relation linéaire entre $x = -1/\overline{OA}$ et $y = 1/\overline{OA'}$. Si l'on note $y = ax + b$, à quoi doivent correspondre a et b ?

Comment estimer les incertitudes sur x et y ?

Pour chaque point mesuré on prendra le plus grand soin dans l'estimation des incertitudes (incertitudes optiques dues à la latitude de mise au point, incertitudes géométriques dues à la résolution du banc, épaisseur de la lentille, etc.)

$-\overline{OA}$	(min)	(max)	(Points Intermédiaires)
$\overline{OA'}$	(max)	(min)	
$\Delta \overline{OA}$.	.	.
$\Delta \overline{OA'}$.	.	
$x = -1 / \overline{OA}$.	.	
$y = 1 / \overline{OA'}$			

Δx			
Δy			

Rentrer les points expérimentaux sur ordinateur dans le tableau prévu à cet effet et procéder par itération. Déterminez ainsi $a \pm \Delta a$ et $b \pm \Delta b$. En déduire $f' \pm \Delta f'$.

Comparaison aux valeurs attendues.

Liste du matériel pour chaque binôme :

- un banc d'optique 2m
- 3 pieds
- une lampe source avec la flèche
- une alimentation électrique
- un écran blanc
- une lampe torche
- 1 porte lentille
- un ordinateur avec le tableau *régression avec barres d'erreurs*

Résultats obtenus par 7 groupes le 22 novembre 2007 au Lycée Alain-Fournier :

	a	Δa	f' (cm)	$\Delta f'$ (cm)
1	-0,99	0,036	19,9	0,28
2	-0,98	0,055	20,5	0,52
3	-1,07	0,036	18,9	0,29
4	-0,96	0,1	20,6	0,74
5	-1,03	0,049	20,6	0,44
6	-1,03	0,041	20,13	0,31
7	-1,02	0,059	19,58	0,42

D. Dioptrès et lentilles minces sphériques

But du TP :

- Nous désirons mesurer les rayons de courbure d'une lentille mince sphérique.
- Mesurer sa distance focale par une méthode classique.
- Déduire de ces mesures l'indice de la lentille.

Théorie :

- Reprendre le théorème des vergences pour un système constitué d'une lentille et d'un miroir sphérique accolés dans les conditions de Gauss.
- Démontrez la relation de conjugaison au sommet d'un dioptre sphérique dans les conditions de Gauss. Associez deux tels dioptrès accolés pour trouver l'expression de f' , focale image de la lentille, en fonction de n , R_1 et R_2 .

Proposez deux méthodes expérimentales permettant de mesurer les rayons de courbures d'une lentille mince sphérique (l'une optique, l'autre par mesure directe).

Données pour la méthode optique : Pour un rayon incident sur un dioptre, seule environ 96% de l'énergie lumineuse est transmise (réfractée), l'autre partie est réfléchie. Expli-

quez comment l'étude de cette portion réfléchie permet de remonter au rayon de courbure d'un des deux dioptrés de la lentille (est-ce toujours possible ?), puis à l'indice de la lentille.

Expérience : Mesurez les rayons de courbures d'au moins deux lentilles différentes par les deux méthodes.

Proposition : celle de 200mm de grand diamètre et celle de 50cm de distance focale.

Exploitation : Une grande importance est donnée aux calculs d'incertitudes.

Mesures optiques :

	L ₁	L ₂
f _{mesuré}		
Δf _{mesuré}		
R _{mesuré}		
ΔR _{mesuré}		
R _{1 calculé}		
ΔR _{1 calculé}		
R' _{mesuré}		
ΔR' _{mesuré}		
R _{2 calculé}		
ΔR _{2 calculé}		
n _{calculé}		
Δn _{calculé}		

Mesure directes : Établir le même type d'exploitation.

Commentaires ? Précision sur l'indice comparée à celle obtenue avec le goniomètre ?

IX. OUTILS MATHÉMATIQUES

Les outils mathématiques utilisés au chapitre 1 sont du niveau lycée. Les dérivées partielles utilisées au chapitre 2 sont de la première année d'enseignement supérieur mais nous comprenons très vite le lien avec les dérivées du lycée. C'est à la fin du second chapitre, avec l'utilisation des matrices pour la régression généralisée, que nous plongeons dans l'enseignement universitaire. Mon but n'est pas ici de revoir ou d'introduire toutes ces notions, seulement un petit formulaire qui pourra être utile pour résoudre les exercices proposés.

A - Dérivées

1- Définition : $f'(x) = \lim_{\epsilon \rightarrow 0} \left(\frac{f(x+\epsilon) - f(x)}{\epsilon} \right)$

Par exemple si $f(x)=x^2$:

$$(x+\epsilon)^2 - x^2 = x^2 + 2x\epsilon + \epsilon^2 - x^2 \simeq 2x\epsilon \quad \text{et } f'(x) = 2x.$$

Graphiquement la dérivée correspond à la pente de la courbe en un point.

2- Principales dérivées utilisées :

fonction f	dérivée f'	Δf
$a x$	a (constante)	$\Delta(ax) = a \Delta x$
x^α	$\alpha x^{\alpha-1}$	$\Delta(x^\alpha) = \alpha x^{\alpha-1} \Delta x$
$\sin(x)$	$\cos(x)$	$\Delta(\sin(x)) = \cos(x) \Delta x$

$\cos(x)$	$-\sin(x)$	$\Delta(\cos(x)) = \sin(x) \Delta x$
e^x	e^x	$\Delta(e^x) = e^x \Delta x$
$\ln(x)$	$1/x$	$\Delta(\ln(x)) = \Delta x / x $
$u+v$	$u'+v'$	(u et v fonctions de x)
uv	$u'v + v'u$	
$\frac{u}{v}$	$\frac{u'v - v'u}{v^2}$	
$f(g(x))$	$g'(x)f'(g(x))$	(fonctions composées)

- $\frac{1}{x} = x^{-1}$ donc $\left(\frac{1}{x}\right)' = (-1)x^{-1-1} = -\frac{1}{x^2}$.
- $\sqrt{x} = x^{\frac{1}{2}}$ donc $(\sqrt{x})' = \frac{1}{2}x^{\frac{1}{2}-1} = \frac{1}{2\sqrt{x}}$.
- $(\sin(x^2))' = (x^2)' \cos(x^2) = 2x \cos(x^2)$

B - Dérivées partielles

Celles-ci concernent les fonctions à plusieurs variables. Par exemple considérons la fonction à trois variables :

$$f(x,y,z) = x^2 - 2z + xy$$

Nous pouvons regarder les variations de cette fonction par rapport à une variable tout en considérant les autres constantes. On procède alors comme pour une dérivée. On a donc :

$$\left(\frac{\partial f}{\partial x}\right)_{y,z} = 2x + y, \quad \left(\frac{\partial f}{\partial y}\right)_{x,z} = x \quad \text{et} \quad \left(\frac{\partial f}{\partial z}\right)_{x,y} = -2$$

On parle de dérivée partielle et par exemple pour la première on dit : "dé rond" f sur "dé rond" x , à (avec) y et z constants.

C - Développement limité

Avec la notion de dérivée nous avions étudié le comportement au premier ordre d'une fonction autour d'un point, nous affinons ici aux ordres supérieurs.

Pour toute fonction infiniment dérivable et pour $\epsilon \ll 1$ nous avons le développement suivant au voisinage d'un point :

$$f(x_0 + \epsilon) = f(x_0) + \epsilon f'(x_0) + \frac{\epsilon^2}{2} f''(x_0) + \frac{\epsilon^3}{3!} f^{(3)}(x_0) + \dots + \frac{\epsilon^n}{n!} f^{(n)}(x_0) + \dots$$

Plus nous prenons de termes de haut degré plus l'approximation est précise. Par exemple pour $f(x) = \sin(x)$ et $x_0 = 0$ nous trouvons : $\sin(\epsilon) \simeq \epsilon - \frac{\epsilon^3}{3!}$, de même $\cos(\epsilon) \simeq 1 - \frac{\epsilon^2}{2}$.

$$\text{Aussi : } \exp(\epsilon) \simeq 1 + \epsilon \quad \ln(1 + \epsilon) \simeq \epsilon \quad (1 + \epsilon)^\alpha \simeq 1 + \alpha \epsilon$$

D - Intégrales

On montre en math que : $\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$.

Pour la loi normale standard, nous pouvons vérifier que la moyenne est nulle et que l'écart-type est égal à 1 :

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \text{ et } \mu = \int_{-\infty}^{+\infty} x \cdot p(x) dx = 0 \text{ car l'intégrale sur}$$

un intervalle symétrique d'une fonction impaire est nulle.

$$\sigma^2 = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (-x)(-x e^{-\frac{x^2}{2}}) dx \text{ d'où}$$

$$\sigma^2 \sqrt{2\pi} = [-x e^{-\frac{x^2}{2}}]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} (-1) e^{-\frac{x^2}{2}} dx = 0 + \sqrt{2} \int_{-\infty}^{+\infty} e^{-x^2} dx$$

d'où l'écart-type égal à l'unité.

Nous avons utilisé une **intégration par parties** :

$$\int_a^b u(x)v'(x)dx = [u(x)v(x)]_a^b - \int_a^b u'(x)v(x)dx$$

puis un changement de variable $x' = x/\sqrt{2}$.

Nous pouvons aller plus loin, en calculant, le *coefficient de dissymétrie* $\mu_3 = \int_{-\infty}^{+\infty} x^3 \cdot p(x)dx$, le *coefficient d'aplatissement* $\mu_4 = \int_{-\infty}^{+\infty} x^4 \cdot p(x)dx$, et, de manière générale, les *moments d'ordre n* $\mu_n = \int_{-\infty}^{+\infty} x^n \cdot p(x)dx$. L'ensemble de ces moments permettent de caractériser une distribution.

Pour une gaussienne $\mu_3=0$ (symétrique). Si ce coefficient est négatif la courbe s'étale vers la gauche, s'il est positif la courbe s'étale vers la droite. Pour une gaussienne $\beta_2=\mu_4/\sigma^4=3$. Si ce coefficient est inférieur à 3 la courbe est plus aplatie qu'une gaussienne. Pour une loi binomiale : $\beta_2=3-8/n$.

Changement de variable :

Soit une nouvelle variable $u=g^{-1}(x)$ avec la fonction g^{-1} continue et strictement monotone sur $[a,b]$ et g la fonction réciproque, alors :

$$\int_a^b f(x)dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(u))g'(u)du$$

E – Séries

Formule du binôme de Newton : $(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$

Séries géométriques dérivées : si $|q| < 1$ alors $\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$

nous dérivons ensuite par rapport à q : $\sum_{k=1}^{\infty} k q^{k-1} = \frac{1}{(1-q)^2}$

puis $\sum_{k=2}^{\infty} k(k-1)q^{k-2} = \frac{2}{(1-q)^3} \dots$

et finalement nous trouvons la formule du binôme négatif :

$\sum_{k=r}^{\infty} k(k-1)\dots(k-r+1)q^{k-r} = \frac{r!}{(1-q)^{r+1}}$

soit $\frac{1}{(1-q)^{r+1}} = \sum_{k=r}^{\infty} \binom{k}{r} q^{k-r}$

Une définition de la fonction exponentielle : $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$

F - Fonction Gamma

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$$

Cette fonction permet de généraliser la notion de factorielle à \mathbb{C} privé de \mathbb{Z} ²⁴. Nous l'utiliserons pour les nombres demi-entiers.

On montre par intégration par parties que : $\Gamma(x+1) = x \Gamma(x)$

$\Gamma(1) = 1$ d'où, pour n entier, $\Gamma(n+1) = n!$.

De plus $\Gamma(1/2) = \sqrt{\pi}$ permet le calcul de la fonction pour les demi-entiers.

24 Par exemple $\pi! \approx 7,2$.

X. CORRECTIONS

Corrections complètes ou sommaire suivant les cas.

Chapitre I : Variable Aléatoire

E1 : Âges énoncé p37

n=15; mode=18; médiane=18; moyenne≈18,333; moy. géométrique≈18,303; étendue=4; écart-type≈1,113; écart quadratique moyen≈1,075; écart moyen≈0,844.

E2 : Cartes énoncé p37

Nb de tirages possibles : $32 \times 31 \times 30 \times 29 \times 28$ divisé par toutes les permutations possibles des 5 cartes $5! = 120$, soit **201 376** jeux.

* Nb de tirages correspondant aux carrés d'As : une seule manière d'avoir les 4 As, et 28 possibilités pour la carte qui les accompagne, donc **28** jeux possibles (un jeu = 120 tirages possibles).

D'où la probabilité $p = 28/201376 = 139$ chances sur un million=1 chance sur $7192 = \mathbf{0,014\%}$.

* Pour une couleur : chaque couleur a 8 cartes. Nb de manières de choisir 5 cartes parmi 8 = $\frac{8!}{5!(8-5)!} = \binom{8}{5} = C_8^5$ (nombre de combinaisons) = 56.

Comme il y a 4 couleurs : $4 \times 56 = 224$.

D'où la probabilité $p = 224 / 201 376 = 1$ chance sur $899 = \mathbf{0,11\%}$.

E3 : Champ de pesanteur énoncé p37

a) Très dispersé. On pourrait indiquer moins de chiffres significatifs.

b) $\bar{g} = 9,32 \text{ m/s}^2$. $\sigma_g = 1,994 \text{ m/s}^2 \approx 2 \text{ m/s}^2$

c) Nous supposons une distribution normale des données ainsi nous avons une Student pour la distribution d'échantillonnage :

$\Delta g = 2,36 \cdot 1,994 / \sqrt{8} = 1,66 \text{ m/s}^2$. D'où $g=9,32 \pm 1,66 \text{ m/s}^2$ et la valeur attendue appartient à l'intervalle de confiance (estimation) :

$7,66 \text{ m/s}^2 < g = 9,81 \text{ m/s}^2 < 10,98 \text{ m/s}^2$ à 95% de confiance.

d) On est environ à σ de $g \approx 10 \text{ m/s}^2$ d'où 68% de chance (intervalle de fluctuation : gaussienne).

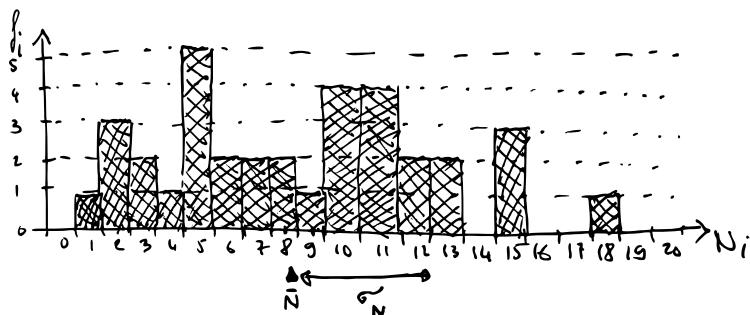
E4 : Ascenseur énoncé p38 25 chances sur mille.

E5 : Devoir énoncé p38

a) $n=35$, $\bar{N} \approx 8,29$ et $\sigma_N \approx 4,40$.

b) Notes avec leur fréquences :

Notes	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Fréq.	0	1	3	2	1	5	2	2	2	1	4	4	2	2	0	3	0	0	1	0	0

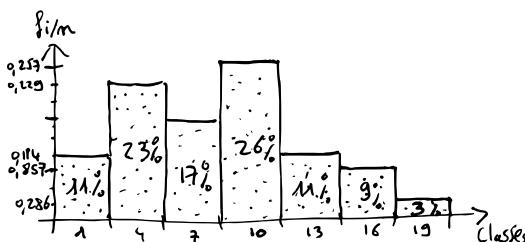


c) Classes, fréquences et fréquences relatives :

Classes	[0,1,2]	[3,4,5]	[6,7,8]	[9,10,11]	[12,13,14]	[15,16,17]	[18,19,20]
Fréqu.	4	8	6	9	4	3	1
F. rel.	11,4%	22,9%	17,1%	25,7%	11,4%	8,57%	2,86%

En regroupant les données par classes il y a une perte d'information. La perte est ici faible. Mais par ailleurs nous avons un gain de visibilité.

Nous pouvons plus facilement faire ressortir un comportement global et des catégories d'étudiants. Les fréquences relatives indiquent directement le pourcentage d'étudiant dans chaque classe, par exemple 34% des étudiants semblent en grande difficulté, ce qui est plus difficile d'interprétation sur le premier diagramme.



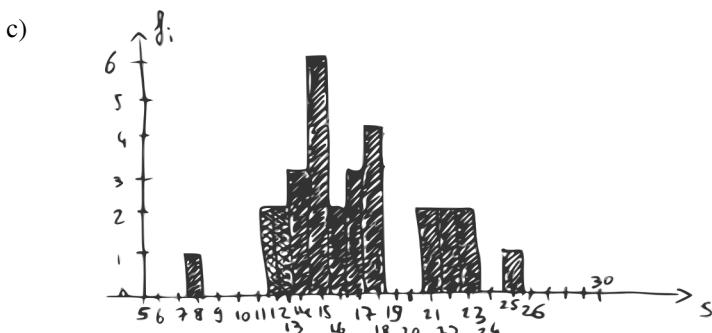
E6 : Yams énoncé p38

1) $1/1296$

2) $(1+5+15+35+70)/6^5 = 1,65\%$

3) a) $n=30$, $\bar{S} \approx 16,73$ et $\sigma_S \approx 3,88$.

b) Pour la somme de cinq dés nous avons une distribution de la population que nous pouvons supposer déjà proche d'une gaussienne (voir p18 pour quatre dés) et nous utilisons le coefficient d'élargissement de Student : $\Delta\bar{S} \approx 1,45$ et $\bar{S} = 16,73 \pm 1,45$ avec une confiance de 95%. Ici la population est infinie et connue. On trouve $\mu_{\text{pop}} = 17,5$. C'est cohérent. Il n'y a pas de raison de penser que les dés sont pipés.



d) La somme est le résultat de cinq lancers indépendants, nous pouvons

penser d'après le théorème central limite que nous nous approchons d'une gaussienne : $(24-\mu)/\sigma_s \approx 1,68$. On est proche de $t_{\infty,90\%}$, donc la moitié 5% sur la queue de gaussienne de droite. Donc 1 chance sur vingt (sur nos lancers ici : 1/30, cohérent).



E7 : Élastiques énoncé p39

H_0 : 10 élastiques sur 1000 sont endommagés.

Nous considérons la variable aléatoire discrète X_0 qui vaut 1 si l'élastique est défectueux et 0 s'il est en bonne état. Probabilité d'avoir un élastique défectueux : $p=0,01$.

$$\mu = 0,01 \times 1 + 0,99 \times 0 = 0,01 \text{ et } \sigma^2 = 0,01 \times 0,99^2 + 0,99 \times 0,01^2 \text{ soit } \sigma \approx 0,0995.$$

$$\text{Nous avons } \bar{x} = \mu + t_\infty \sigma / \sqrt{n} \text{ or } \bar{x} = n_0 / n \text{ d'où } n_0 = n(\mu + t_\infty \sigma / \sqrt{n}).$$

*Pour $n=1000$: n est grand et d'après le théorème central limite nous avons une loi normale. Nous avons 1 % en queue de distribution droite.

D'où : $n_0 = 1000(0,01 + 2,33 \cdot 0,099 / \sqrt{1000}) \approx 17,3$ et un refus de la livraison à partir de $n_0=18$ élastiques endommagés sur les 1000 tirés au hasard. En utilisant la loi binomiale $n_0=19$, proche de 18 (voir ci-dessous pour la justification).

*Pour $n=200$: en suivant le même raisonnement nous avons $n_0=6$, en effet $n_0 \approx 200(0,01 + 2,33 \cdot 0,099 / \sqrt{200}) \approx 5,3$. Mais l'hypothèse des grands nombres est-elle encore ici justifiée ? n doit être grand, par exemple supérieur à 30, mais aussi np et $n(1-p)$ doivent aussi être suffisamment grand, par exemple supérieur à 5. Comme ici p n'est pas autour d'une chance sur deux mais au contraire proche d'un bord, peu probable ou quasi sûr, nous allons vérifier la validité en faisant un calcul exact.

Nous avons une succession d'épreuves identiques et indépendantes (le grand nombre d'élastiques présents dans les lots livrés permet de considérer un tirage avec remise). Épreuves à deux issues succès (S) de probabilité p ou échec (E) de probabilité $q=1-p$. Nous reconnaissions une loi binomiale de paramètres n et p (voir p102 pour plus de détails). Nous nommons X la loi qui correspond au nombre k de succès, élastiques endommagés sur n épreuves. Nous cherchons la valeur de $k=n_0$ à

partir de laquelle nous rejeterions H_0 avec moins d'une chance sur 100 de se tromper : $P(X \geq k) \leq 1\%$.

$$P(X \geq 6) = 1 - P(X \leq 5) \approx 1,6\% \quad \text{et} \quad P(X \geq 7) \approx 0,4\%$$

(Calcul fait sur tableur : LOI.BINOMIALE(k;n;p;1)). D'où : n_o=7.

*Pour n=50 : avec la loi binomiale n_o=4 (et non n_o>2,1 soit n_o=3).

E8 : Test d'un isolant énoncé p39

$x = 39,28 \pm 0,49$ mW/m/K à 95%, soit une incertitude relative de 1,2% (la distribution de la population a été supposée normale et celle de la distribution d'échantillonnage de Student). Résultats conformes. D'après ces résultats le constructeur pourrait annoncer une marge plus faible. Selon un test de Student bilatéral classique la valeur annoncée par le constructeur est bien dans l'intervalle de confiance. Le constructeur a indiqué une moyenne mais n'a pas donné d'écart-type, nous l'avons estimé avec celui de l'échantillon : $\sigma \approx s \approx 0,69$. La probabilité critique α vaut 23% (avec $t = (\bar{x} - \mu) / \sqrt{n} / \sigma \approx 1,29$ et =TDIST(1,29,9,2) sur LibreOffice), si l'on rejettait l'hypothèse de conformité il y aurait 23% de chance de la rejeter à tort ce qui est un risque bien trop important.

E9 : Pièces énoncé p40

1) $\chi^2 \approx 2,56 < 3,84$. Il n'y a pas lieu de remettre en cause l'équilibre des pièces. La différence par rapport à l'équiprobabilité s'explique par les fluctuations statistiques.

Ou par un test bilatéral classique (intervalle de fluctuation) :

$$|(\bar{x} = 42/100) - (\mu = 0,5)| \leq t_{\infty 95\%} (\sigma = 0,5) / \sqrt{n = 100}$$

2) $\chi^2 \approx 0,4 < 3,84$. Idem. p.c. (probabilité critique) $\approx 20,8\%$, il y a une chance sur 5 de rejeter l'hypothèse alors quelle est juste, on ne peut pas la rejeter.

3) $\chi^2 \approx 25,6 >> 3,84$! La pièce n'est sûrement pas équilibrée. Une chance sur deux millions de rejeter l'hypothèse alors quelle est juste, on la rejette !

E10 : Parité énoncé p40

Valeurs espérées $\chi^2 = \frac{(470-288,5)^2}{288,5} + \frac{(107-288,5)^2}{288,5} + \dots$ et

288,5	288,5	$\chi^2 \approx 344$, avec ddl=2x1=2 la probabilité de
174	174	rejeter l'hypothèse alors quelle est vraie est très	inférieure à une chance sur 1000 (13,8 << 344
6	6	pour 0,1%).	<u>La parité n'est pas respectée.</u>

Comprenant ces résultats il faudrait être d'une mauvaise foi extrême pour prétendre qu'il n'y a aucune raison de penser que la parité n'est pas respectée en justifiant les écarts par les fluctuations statistiques.

E11 : Naissances énoncé p41

$\chi^2 \approx 6,16 < 7,8$. L'hypothèse ne peut être abandonnée, on a besoin de plus d'effectif pour trancher.

$\chi^2 \approx 128 >> 7,8$. Hypothèse totalement rejetée.

E12 : Gaussiennes dans le plan et dans l'espace énoncé p42

1D :

1- $\bar{x} = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0$ car la fonction intégrée est définie et impaire sur un intervalle d'intégration symétrique (propriété de l'intégrale). $|\bar{x}| = \int_{-\infty}^{+\infty} |x| \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 2 \int_0^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ car cette fois la fonction est paire. $|\bar{x}| = \frac{-2}{\sqrt{2\pi}} [e^{-\frac{x^2}{2}}] = \sqrt{\frac{2}{\pi}}$. La distance correspond à une valeur absolue, sa moyenne est donc non nulle et positive. \bar{x} correspond à la moyenne de l'abscisse, grandeur algébrique.

$\sigma_x = \sigma_{|x|} = 1$, car $x^2 = |x|^2$ et que c'est une loi normale réduite.

2- $P(|x| \leq 1) = P(-1 \leq x \leq 1) = \int_{-1}^{+1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ nous calculons numéri-

quement (à la calculatrice, sur un ordinateur...) ces intégrales et retrouvons les résultats du chapitre sur la distribution de Gauss à σ , deux σ et trois σ .

2D :

1- $p(x,y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$ $p(x,y)$ est bien une fonction positive, ce qui est conforme à une loi de probabilité.

Sur tout le plan nous avons bien une probabilité de 100% :

$$\begin{aligned} \iint p(x,y) dx dy &= \int \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \int \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1.1 = 1 \\ 2- \iint p(x,y) dx dy &= \int \frac{1}{2\pi} e^{-\frac{\rho^2}{2}} 2\pi \rho d\rho = 1 \quad p(\rho) = \rho e^{-\frac{\rho^2}{2}} \\ \text{et } \int_0^{+\infty} p(\rho) d\rho &= -[e^{-\frac{\rho^2}{2}}] = 1. \end{aligned}$$

$$\bar{\rho} = \int_0^{+\infty} \rho p(\rho) d\rho = \int \rho^2 e^{-\frac{\rho^2}{2}} d\rho = [\rho^2(-e^{-\frac{\rho^2}{2}})] - \int 2\rho e^{-\frac{\rho^2}{2}} d\rho$$

$\bar{\rho} = 0 - 0 + \sqrt{2} \int_0^{+\infty} e^{-t^2} dt = \sqrt{2} \frac{\sqrt{\pi}}{2} = \sqrt{\frac{\pi}{2}}$ après une intégration par parties ($u=\rho$) et un changement de variable ($\rho = \sqrt{2} t$ et on reconnaît une intégrale connue).

$$\sigma_\rho^2 = \bar{\rho}^2 = \int_0^{+\infty} \rho^2 p(\rho) d\rho = \int \rho^3 e^{-\frac{\rho^2}{2}} d\rho = [\rho^2(-e^{-\frac{\rho^2}{2}})] + 2 \int \rho e^{-\frac{\rho^2}{2}} d\rho$$

on retrouve une intégrale calculée en 1D : $\sigma_\rho^2 = 0 - 0 + 2 \cdot 1 \quad \sigma_\rho = \sqrt{2}$

$$3- P(\rho \leq \sigma_\rho) = \int_0^{\sqrt{2}} \rho e^{-\frac{\rho^2}{2}} d\rho = 1 - \frac{1}{e} \quad P(\rho \leq 2\sigma_\rho) = 1 - \frac{1}{e^4}$$

3D :

$$1- \quad p(x, y, z) = \frac{1}{(\sqrt{2\pi})^3} e^{-\frac{r^2}{2}} \quad p(r) = \alpha r^2 e^{-\frac{r^2}{2}} \quad \alpha = \frac{4\pi}{(\sqrt{2\pi})^3}$$

$$2- \quad \int_0^{+\infty} p(r) dr = \alpha \times \sqrt{\frac{\pi}{2}} = 1 \quad (\text{intégrale déjà calculée en 2D})$$

$$3- \quad \bar{r} = \int_0^{+\infty} r p(r) dr = \alpha \int_0^{+\infty} r^3 e^{-\frac{r^2}{2}} dr = \alpha \times 2 = \frac{4}{\sqrt{2\pi}}$$

$$\sigma_r^2 = \int_0^{+\infty} r^2 p(r) dr = \alpha \int_0^{+\infty} r^4 e^{-\frac{r^2}{2}} dr = 3 \quad \text{par intégration par parties.}$$

$$\text{d'où : } \sigma_r = \sqrt{3}$$

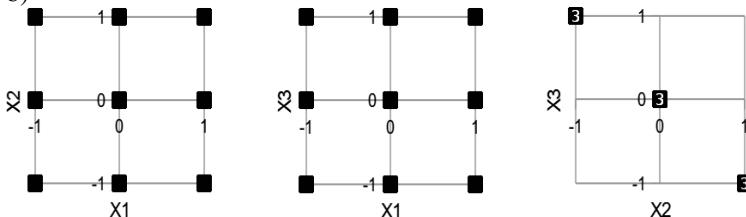
$$4- \quad P(r \leq \sigma_r) = \int_0^{\sqrt{3}} \rho e^{-\frac{r^2}{2}} dr \simeq 0,608 \quad \text{et les deux autres aussi par calcul numérique.}$$

Chapitre II : Corrélations et indépendances

E1 : Corrélations énoncé p86

- 1- a) $\bar{x}_1=(-1-1-1+0+0+0+1+1+1)/9=0$ de même $\bar{x}_2=0$ et $\bar{x}_3=0$.
 $\sigma_1=\sqrt{\sum_{i=1}^9 (x_{i1}-\bar{x}_1)^2/(9-1)}=\sqrt{6}\times 1/8$ d'où
 $\sigma_1=\sigma_2=\sigma_3=\sqrt{3}/2 \approx 0,87$.

b)



c) $r_{12}=0$, $r_{13}=0$, $r_{23}=-1$. X_1 et X_2 ne sont pas corrélées. De même pour X_1 et X_3 . X_2 et X_3 sont dépendantes et totalement corrélées.

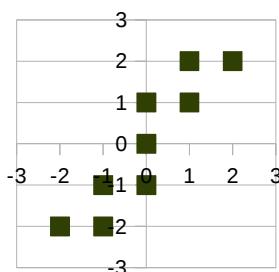
2- a) $\bar{x}_1=0$ et $\bar{x}_2=0$. $\sigma_1 \approx 1,22$ et $\sigma_2 \approx 1,58$.

c) $r_{12}=0,904$. Grandeur globalement corrélées positivement.

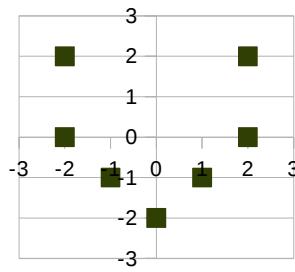
3- a) $\bar{x}_1=0$ et $\bar{x}_2=0$. $\sigma_1 \approx 1,73$ et $\sigma_2 \approx 1,53$.

c) $r_{12}=0$. Grandeur totalement décorrélées, ne pas oublier que les corrélations recherchées sont ici **linéaires**. Il y a une corrélation en V.

2- b)



3- c)



E2 : Volumes énoncé p87

1-

$$\bar{V}_1 = (100,1 + 100,0 + 99,9 + 100,0) / 4 \text{ d'où } \bar{V}_1 = 100,0 \text{ mL.}$$

$$\sigma_1 = \sqrt{\frac{0,1^2 + 0^2 + (-0,1)^2 + 0^2}{4-1}} = \sqrt{\frac{2}{3}} \cdot 0,1 \text{ mL} \quad \text{d'où} \quad \sigma_1 \approx 0,082 \text{ mL}$$

D'après le théorème central limite et la distribution de la population normale : $\Delta V = t \cdot \sigma / \sqrt{n} = 3,18 \times 0,082 / \sqrt{4} = 3,18 \times 0,0816 / 2$

$t_{ddl=3; 95\%} = 3,18$ soit $\Delta \bar{V}_1 \approx 0,13 \text{ mL}$ et $\Delta \bar{V}_1 / \bar{V}_1 \approx 0,13 / 100$

La pipette, avec une confiance de 95%, est à 0,13 mL près, soit pour 100 mL à 0,13% de précision.

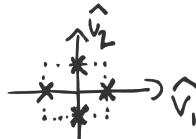
2-

$$\hat{V}^i = V^i - \bar{V} \quad \text{et}$$

$$\sum_i [(\hat{V}_1^i - \bar{V}_1)(\hat{V}_2^i - \bar{V}_2)] = \sum_i [\hat{V}_1^i \hat{V}_2^i] = 0,1 \cdot 0 + 0,0 \cdot 1 + (-0,1) \cdot 0 + 0,0 \cdot 1 = 0$$

$$\text{or } r_{12} = \frac{\sum_i [(\hat{V}_1^i - \bar{V}_1)(\hat{V}_2^i - \bar{V}_2)]}{\sqrt{\sum_i (\hat{V}_1^i - \bar{V}_1)^2} \sqrt{\sum_i (\hat{V}_2^i - \bar{V}_2)^2}}$$

soit $r_{12}=0$, les grandeurs sont totalement décorrélées et donc indépendantes.



3-

$$V = \{200,1 ; 200,1 ; 199,9 ; 199,9\} \text{ mL d'où } \bar{V} = 200 \text{ mL.}$$

$$\sigma_V = \sqrt{\frac{0,1^2 + 0,1^2 + (-0,1)^2 + (-0,1)^2}{4-1}} = \frac{2}{\sqrt{3}} \cdot 0,1 \text{ mL}$$

d'où $\sigma_V \approx 0,115 \text{ mL}$ et $\Delta \bar{V} \approx 0,183 \text{ mL}$ et $\Delta \bar{V} / \bar{V} \approx 0,09\%$

4-

$V(V_1, V_2)$ d'où :

$$\Delta V^2 = \left(\frac{\partial V}{\partial V_1} \right)_{V_2}^2 \Delta V_1^2 + \left(\frac{\partial V}{\partial V_2} \right)_{V_1}^2 \Delta V_2^2 = \Delta V_1^2 + \Delta V_2^2$$

et $\Delta V = \sqrt{2} \cdot \Delta V_1 \approx 0,18 \text{ mL}$ même résultat qu'au 3-.

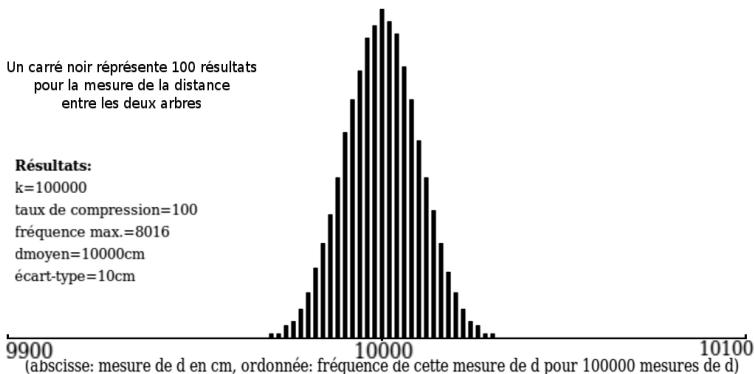
E3 : Arbres

énoncé p88

Nous avons $d = \sum x_i$ (soit $d = x_1 + x_2 + \dots + x_i + \dots + x_n$), où les x_i sont les mesures de longueur pour chaque report de bâton. Δx_i est l'incertitude sur chaque mesure x_i . Ici $\Delta x_i = 1\text{cm}$ et $x_i = 100 \pm 1\text{cm}$.

Quelle est alors l'incertitude Δd ?

Spontanément nous pourrions penser que $\Delta d = \sum \Delta x_i$, d'où ici $\Delta d=100cm$ soit $1m$ pour $100m$ mesurés d'un arbre à l'autre. Mais problème, ce n'est absolument pas ce qu'indique la simulation ci-dessous !



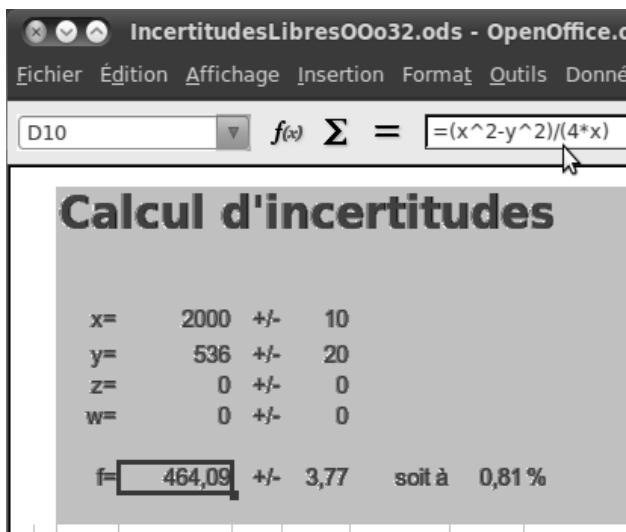
Elle indique plutôt $\Delta d=0,1m$. En effet un simple calcul de probabilités montre qu'il est extrêmement peu probable d'obtenir pour d , $99m$ ou $101m$, alors que l'on a une chance sur deux, pour x_i , d'avoir $0,99m$ ou $1,01m$. Or nous désirons avoir Δd à probabilité égale de Δx_i .

Ici $\partial d / \partial x_i = 1$ et $\Delta x_i = \Delta x$, quelquesoient $i=1\dots n$, ainsi $\Delta d = \sqrt{n} \cdot \Delta x$ soit $\Delta d = 0.1m$, ce qui correspond bien au résultat attendu !

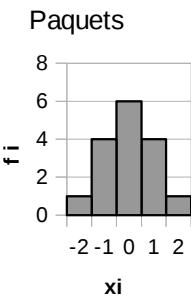
E4 : Méthode de Bessel énoncé p88

$(\partial f' / \partial D)_d = (D^2 + d^2) / 4d^2$ et $(\partial f' \partial / \partial D)_D = -d / 2D$, d'où
 $\Delta f' = \sqrt{[(D^2 + d^2) / 4d^2]^2 (\Delta D)^2 + (d / 2D)^2 (\Delta d)^2}$
puis $f' = 464,1 \pm 3,8 \text{ mm}$
et $\Delta f'/f' = 0,8\%$.

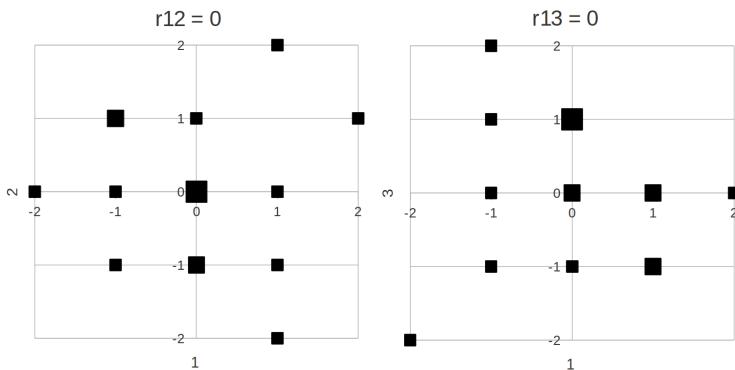
Soit par un calcul numérique avec un tableur et des macros (pas besoin alors d'effectuer des calculs de dérivées partielles) :

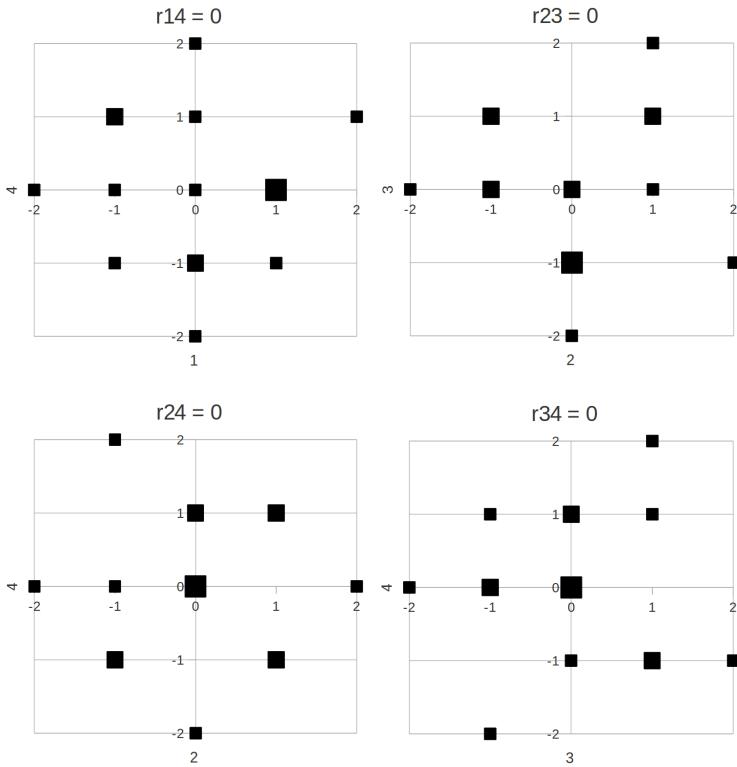


Le tableau utilise ici 2 des 4 paquets globalement gaussiens parfaitement décorrélés disponibles :



1	2	3	4	x	y	f
-2	0	-2	0	1973,71	536	457,04
-1	-1	0	0	1986,86	509,71	464,02
-1	0	-1	1	1986,86	536	460,56
-1	1	1	1	1986,86	562,29	456,93
-1	1	2	-1	1986,86	562,29	456,93
0	0	0	0	2000	536	464,09
0	0	-1	-2	2000	536	464,09
0	-1	1	-1	2000	509,71	467,52
0	1	1	-1	2000	562,29	460,48
0	-1	1	2	2000	509,71	467,52
0	0	0	1	2000	536	464,09
1	-2	0	0	2013,14	483,43	474,26
1	0	-1	0	2013,14	536	467,61
1	-1	0	-1	2013,14	509,71	471,02
1	2	-1	0	2013,14	588,57	460,27
2	1	0	1	2026,29	562,29	467,56





E5 : Indice énoncé p88

D'après la loi de Descartes pour la réfraction : $n_2 = n_1 \frac{\sin(i_1)}{\sin(i_2)}$

d'où $n_2 \approx 1,462$.

$n_2(i_1, i_2)$ alors

$$\Delta n_2^2 = \left(\frac{\partial n_2}{\partial i_1} \right)_{i_2}^2 \Delta i_1^2 + \left(\frac{\partial n_2}{\partial i_2} \right)_{i_1}^2 \Delta i_2^2$$

(la variation en fonction de l'indice n_1 n'est pas incluse car il est supposé connu avec une grande précision)

$$\text{On a } \left(\frac{\partial n_2}{\partial i_1} \right)_{i_2} = \frac{n_1}{\sin(i_2)} \cos(i_1) \quad \text{et} \quad \left(\frac{\partial n_2}{\partial i_2} \right)_{i_1} = n_1 \sin(i_1) \frac{-\cos(i_2)}{\sin^2(i_2)} .$$

$$\text{soit } \Delta n_2 = \left[\left(\frac{\cos(30^\circ)}{\sin(20^\circ)} \frac{\pi}{180} \right)^2 + \left(\frac{\sin(30^\circ) \cos(20^\circ)}{\sin^2(20^\circ)} \frac{2\pi}{180} \right)^2 \right]^{1/2}$$

(les angles sont mis dans leur unité naturelle, les radians, sans unité : $\pi \text{ rad} = 180^\circ$)

Finalement $\Delta n_2 = 0,1470$ avec une confiance de 95%, $n_2 = 1,46 \pm 0,15$ et $\Delta n_2/n_2 = 10\%$.

Avec le tableur : $n_2 = 1,46 \pm 0,16$ et $\Delta n_2/n_2 = 11\%$. La méthode numérique est approximative mais nous vérifions ainsi que nous n'avons pas fait d'erreur de calcul.

Par ailleurs avec un goniomètre nous pouvons faire des mesures d'angles bien plus précises à quelques minutes d'arc près (un minute d'arc = $1' = 1^\circ/60$).

E6 : Formule de Cauchy énoncé p89

1 -

$$\Delta n^2 = \left(\frac{\partial n}{\partial D_m} \right)_A^2 \Delta D_m^2 + \left(\frac{\partial n}{\partial A} \right)_{D_m}^2 \Delta A^2$$

$$\text{avec } \left(\frac{\partial n}{\partial D_m} \right)_A = \frac{\cos[(A+D_m)/2]}{2 \sin(A/2)} \quad \text{et}$$

$$\left(\frac{\partial n}{\partial A} \right)_{D_m} = \frac{\cos[(A+D_m)/2] \sin(A/2) - \sin[(A+D_m)/2] \cos(A/2)}{2 \sin^2(A/2)}$$

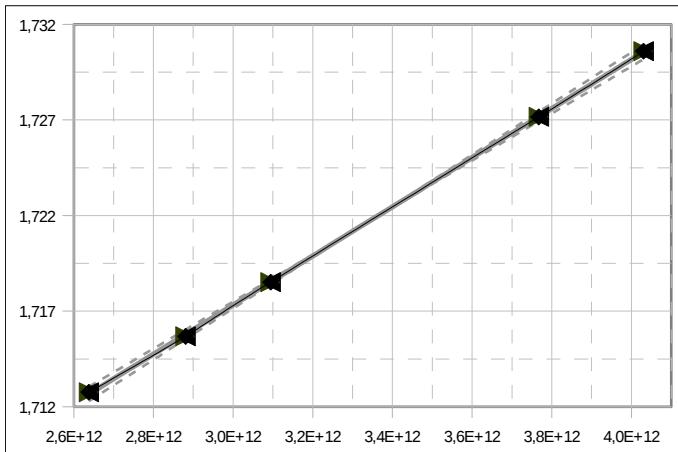
$$\text{avec } A=60^\circ \simeq D_m \quad : \quad \left(\frac{\partial n}{\partial D_m} \right)_A \simeq \frac{1}{2} \quad \text{et} \quad \left(\frac{\partial n}{\partial A} \right)_{D_m} \simeq -1$$

d'où $\frac{\Delta n \simeq 0,00065}{\text{(même chose en calcul numérique)}}$ et $\Delta n/n \simeq 0,04\%$

2-

Avec les barres d'erreurs, $\Delta n=0,00065$ et $\Delta(1/\lambda^2)=2\Delta\lambda/\lambda^3$ avec $\Delta\lambda=0,05\text{nm}$, on trouve **A=1,67877±0,00182**; $\Delta A/A=0,109\%$; **B=(1,285 ± 0,055).10⁻¹⁴ m²** et $\Delta B/B=4,26\%$

n en fonction de 1/λ²(1/m²) :



La régression simple donne **r=0,99995...**

(Si on oublie les barres d'erreurs $A=1,67877\pm 0,00075$; $\Delta A/A=0,044\%$; $B=(1,285 \pm 0,022).10^{-14} \text{m}^2$ et $\Delta B/B=1,75\%$)

3- En régression simple :

- $r=0,99947\dots$ pour $n(1/\lambda)$
- $r=0,99995\dots$ pour $n(1/\lambda^2)$
- $r=0,99982\dots$ pour $n(1/\lambda^3)$

La loi en $1/\lambda^2$ est donc meilleure,
mais qu'elle valeur de α permet d'optimiser la régression ?

Pour cela nous allons tracer $y = \ln(n - A) = \alpha x + \ln(B)$ avec $x = \ln(\lambda)$. α et $\ln(B)$ seront obtenus avec leur incertitudes et A sera choisi de manière à avoir r_{max} .

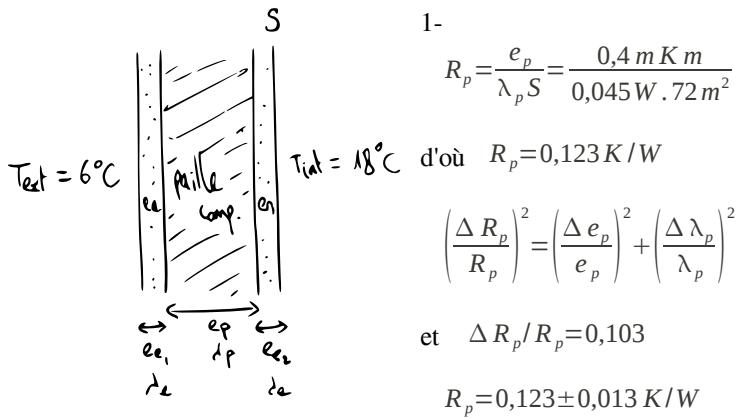
Les résultats sont les suivants $r_{max} = -0,99997024\ldots$ pour $A = 1,6843744$ et nous trouvons : $\alpha = -2,306 \pm 0,033$. Puis avec les barres d'erreurs : $\alpha = -2,3 \pm 0,1$. Conclusion $\alpha = -2$ n'est pas dans l'intervalle le modèle en $1/\lambda^2$ n'est pas validé.

Ceci dit il fonctionne correctement. La différence peut s'expliquer simplement car le modèle de l'électron élastiquement lié, utilisé ici, comporte d'autres termes (il s'agit d'un développement limité d'une fonction plus complexe) : $n = A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4}$. Une étude plus avancée pourrait le vérifier en incluant le nouveau terme en $1/\lambda^4$.

Il faudrait aussi vérifier nos incertitudes expérimentales, il est possible que nous ayons négligé, ou sous-estimé, des sources d'incertitudes (par exemple en triplant les incertitudes sur les angles à $6'$: $\alpha = -2,3 \pm 0,3 \ldots$).

E7 : Mur

énoncé p90



2- $R_e = 0,0069 \pm 0,0010 \text{ K/W}$

3- Les résistances sont en série : $R=R_p+R_e$
 et $\Delta R = \sqrt{(\Delta R_p)^2 + (\Delta R_e)^2}$ d'où $R=0,130 \pm 0,015 \text{ K/W}$

4- $\Phi = \frac{\Delta T}{R} = \frac{12}{0,130} = 92,3 \text{ W}$

$$(\Delta \Phi / \Phi)^2 = (\Delta (\Delta T) / \Delta T)^2 + (\Delta R / R)^2 \quad \Delta T = T_{\text{intérieur}} - T_{\text{extérieur}}$$

d'où $\Delta (\Delta T) = \sqrt{2} \cdot 0,5^\circ \text{C}$, $\Delta \Phi / \Phi = 12,9\%$ et
 $80 \text{ W} \leq \Phi \leq 104,3 \text{ W}$ d'où une puissance minimale de chauffage :

$$\mathbf{P_{\min} = 105 \text{ W}}$$

E8 : Isolation et inertie énoncé p91

1- En parallèle : $\frac{1}{R_{eq}} = \sum_i \left(\frac{1}{R_i} \right) = \sum_i \left(\frac{\lambda_i S_i}{e_i} \right)$

ou en terme de conductances thermiques $G=1/R$: $G_{eq} = \sum_i G_i$
 donc dans notre cas :

$$\frac{1}{R} = \frac{S_s}{(e_s/\lambda_s)} + \frac{S_t}{(e_t/\lambda_t)} + \frac{S_m}{(e_m/\lambda_m)} + \frac{S_h}{(e_h/\lambda_h)} = \frac{36}{4} + \frac{54}{8} + \frac{82}{4} + \frac{8}{1}$$

pour les incertitudes :

$$\Delta G = \sqrt{\sum_i \Delta G_i^2}, \quad \Delta G_i = S_i \frac{\Delta(e_i/\lambda_i)}{(e_i/\lambda_i)^2}, \quad \Delta G \approx 2,471 \text{ W/K}$$

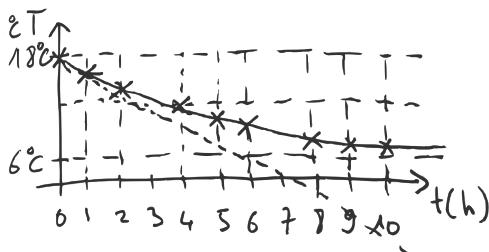
et $\Delta R = R^2 \Delta G \approx 1,26 \text{ mK/W}$

finalement $R = 22,6 \pm 1,3 \text{ mK/W}$ et $\Delta R / R = 5,6\%$

2- $\Phi = \frac{\Delta T}{R} = \frac{12}{22,6 \cdot 10^{-3}} \approx 531 \text{ W} \pm 8,1\%$ (comme à l'exercice *Mur*)
 d'où une puissance minimale de chauffage :

$$\mathbf{P_{min} = 575 \text{ W}}$$

3-

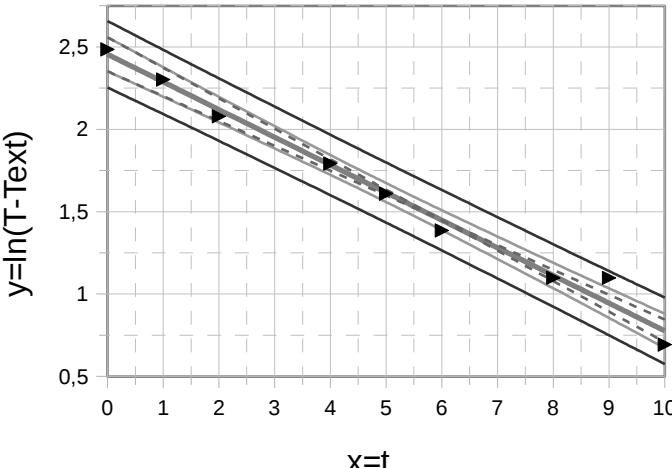


Nous avons nécessairement une asymptote à T_{ext} : la température ne peut descendre spontanément en dessous de T_{ext} .

Si linéaire la température tendrait vers moins l'infini avec le temps.

$$T - b = a e^{-t/\tau} \quad \text{d'où} \quad \ln(T - b) = \ln a - \frac{1}{\tau} t \quad ,$$

on trouve $b = T_{ext} = 5,7^\circ\text{C}$ (coefficient de corrélation r maximal),
 puis on fait une régression sur un tableur avec $y = \ln(T - b)$ et
 $x = t$ (page 2 d'*IncertitudesLibres*) :



coefficient directeur = $-1/\tau = -0,156 \pm 0,0157 \text{ h}^{-1}$
 ordonnée à l'origine = $\ln a = \ln(T_{int} - T_{ext}) = 2,474 \pm 0,0949$
 $r_{max} = -0,9937170474$ (meilleur alignement des points)

Nous avons : $\frac{\Delta \tau}{\tau} = \frac{\Delta(1/\tau)}{(1/\tau)}$ et $\Delta x = x \cdot \Delta(\ln x)$

D'où : $\tau = 6 \text{ h } 25 \text{ min} \text{ à } 10,1\%$ et $T_{int} = 17,5 \pm 1,2 \text{ °C}$.

4-

a)

flux sortant d'énergie pendant dt

= énergie calorifique perdue pendant ce même intervalle de temps

$$\text{soit } \Phi dt = \frac{\Delta T}{R} dt = \frac{T(t) - T_{ext}}{R} dt = C dT$$

$$\text{et } \frac{dT}{dt} + \frac{1}{\tau} T = \frac{1}{\tau} T_{ext} \text{ avec } \tau = RC$$

où l'expression du 3- est la solution générale.

$$\text{b) } C = \tau / R \text{ et } (\Delta C/C)^2 = (\Delta \tau / \tau)^2 + (\Delta R / R)^2$$

d'où $C = 1034 \pm 119 \text{ kJ/K}$

E9 : Rendement

énoncé p92

1-

Sur un tableur (droite de régression sur le graphique) :

n = 7

Conf.: 95 %

t = 2,57

x	y	(xi-xm)	(xi-xm) ²	yim	(yi-yim) ²	xi ²	(xi-xm) ²	(yi-yim) ²	(xi-xm)*(yi-yim)
100	41	5271,43	90000	40,79	0,05	10000	90000	308,76	5271,4
200	44	2914,29	40000	46,71	7,37	40000	40000	212,33	2914,3
300	53	557,14	10000	52,64	0,13	90000	10000	31,04	557,14
400	63	0	0	58,57	19,61	160000	0	19,61	0
500	66	742,86	10000	64,5	2,25	250000	10000	55,18	742,86
600	65	1285,71	40000	70,43	29,47	360000	40000	41,33	1285,7
700	78	5828,57	90000	76,36	2,70	490000	90000	377,47	5828,6

$$xm = ym = \text{Cov}(x,y) = \text{Var}(x) = 400 \quad 58,57 \quad 16600 \quad 280000$$

$$sy/x = sx^2 = 3,51 \quad 1400000$$

$$sx = sy = s^2xy = 216,02 \quad 13,20 \quad 2766,7$$

$$da = 0,017$$

$$da/|a| = 29\%$$

$$db = 8$$

$$db/|b| = 22\%$$

$$dy =$$

$$9,021$$

$$r = 0,9701$$

r: coefficient de corrélation

linéaire algébrique.

$$a = 0,059 + ou - 0,017$$

$$b = 35 + ou - 8$$

$$0,042 < a < 0,076$$

$$27 < b < 42$$

$$sa = 0,00663$$

$$sb = 2,96579$$

Coefficient de corrélation

linéaire critique:

$$rc = 0,669$$

Test de corrélation (linéaire):

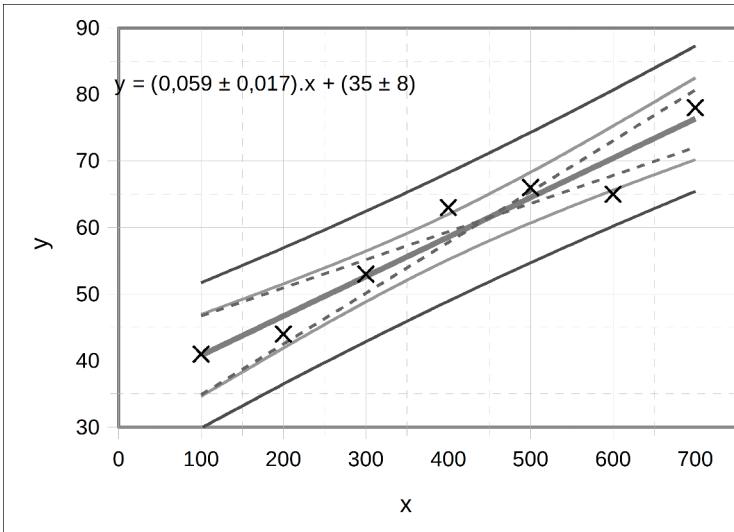
$$R = 0,941$$

si $|r| < rc$ alors grandeurs non corréléesR: coefficient de détermination ($R=r^2$).si $|r| > rc$ alors grandeurs corrélées94,1% de la variation de y est

Ici les grandeurs sont corrélées.

expliquée par x

Rendement en fonction de la quantité d'engrais :



2- Le rendement serait alors de 67 Q/ha à 4 Q/ha près avec une confiance de 95 pour cent.

Pour $x_0=550$

Estimation de la moyenne de y_0 :

$$\begin{array}{ll} 63 & < y_0 < 72 \\ \text{yom=} & 67 \\ \text{dyom=} & 4 \\ \text{dyom}/|\text{yom}|= & 6\% \end{array}$$

3- Le rendement serait alors de 35 Q/ha à 8 Q/ha près avec une confiance de 95 pour cent (ordonnée à l'origine).

Pour $x_0=0$

Estimation de la moyenne de y_0 :

$$\begin{array}{ll} 27 & < y_0 < 42 \\ \text{yom=} & 35 \\ \text{dyom=} & 8 \\ \text{dyom}/|\text{yom}|= & 22\% \end{array}$$

4- La probabilité est de 95% (intervalle de prédiction à 2 sigmas).

Pour $x_0=250$

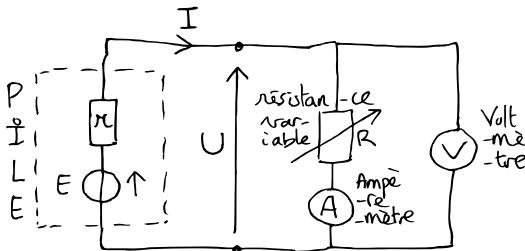
Prédition pour y_0 :

$$\begin{array}{ll} 40 & < y_0 < 60 \\ \text{yom=} & 50 \\ \text{dyom=} & 10 \\ \text{dyom}/|\text{yom}|= & 20\% \end{array}$$

E10 : Étude d'une pile

énoncé p93

Nous pouvons comme pour le tracé de la caractéristique de tout dipôle utiliser le montage suivant :



En faisant varier R le montage parcourt différents points de fonctionnement du dipôle. Nous avons choisi le montage longue dérivation ($R_V \rightarrow \infty$ et $R_A \rightarrow 0$).

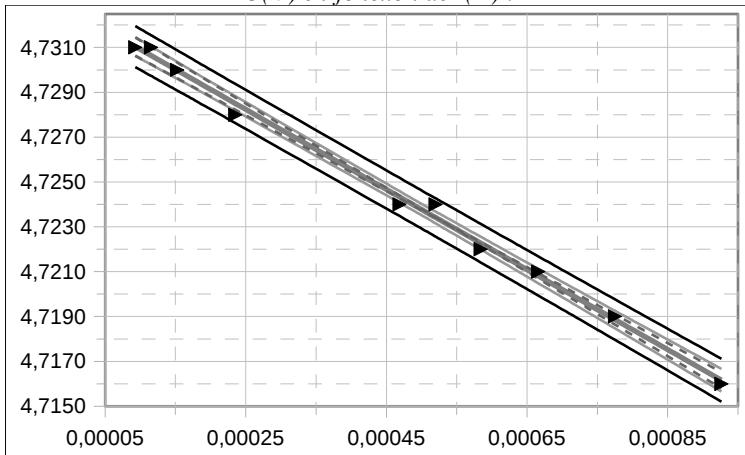
1- Le tableau donne :

$$\begin{array}{lll} a = -17,85 & + \text{ ou } - & 0,92 \\ b = 4,7327 & + \text{ ou } - & 0,00049 \\ r = -0,9980 & \text{Confiance:} & 95\% \end{array}$$

d'où :

$$\begin{aligned} E &= 4,7327 \pm 0,0005 \text{ V} \\ \text{précision :} & 0,010 \% \\ r &= 17,85 \pm 0,92 \Omega \\ \text{précision :} & 5,16 \% \end{aligned}$$

$U(V)$ en fonction de $I(A)$:



2-

 $n = 10$ $a \text{ estimé} = -17,851$

I(A)	dI(A)	U(V)	dU(V)	poids wi	wi/S
0,000093	2,16E-07	4,731	0,00537	347,4E+2	10%
0,000115	2,61E-07	4,731	0,00537	347,4E+2	10%
0,000153	3,35E-07	4,730	0,00537	347,4E+2	10%
0,000235	7,70E-07	4,728	0,00536	347,6E+2	10%
0,000469	1,24E-06	4,724	0,00536	347,8E+2	10%
0,000520	1,34E-06	4,724	0,00536	347,8E+2	10%
0,000584	1,47E-06	4,722	0,00536	347,9E+2	10%
0,000666	1,63E-06	4,721	0,00536	348,0E+2	10%
0,000775	1,85E-06	4,719	0,00536	348,1E+2	10%
0,000926	2,15E-06	4,716	0,00536	348,3E+2	10%

S=

347,8E+3

$da = 6,09$

$db = 0,00324$

$da/|a| = 34,1\%$

$db/|b| = 0,0685\%$

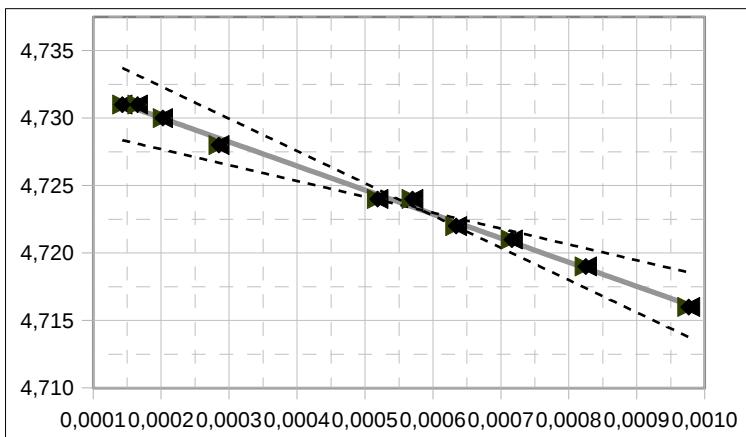
$sa = 6,09$

$sb = 0,00324$

$a = -17,85 + ou - 6,1$

$b = 4,7327 + ou - 0,0032$

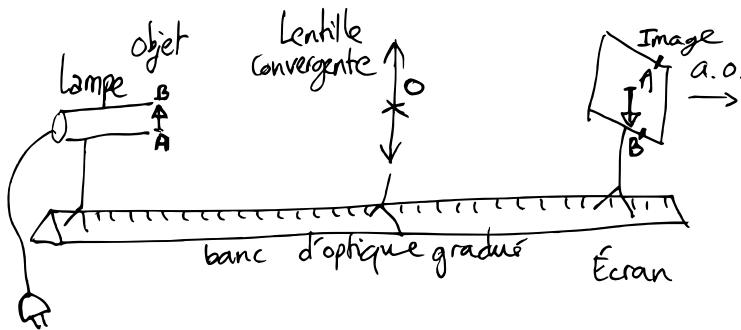
d'où : $E = 4,7327 \pm 0,0032 \text{ V}$, précision : 0,07 %
et $r = 17,85 \pm 6,1 \Omega$, précision : 34 %

 $U(V) \text{ en fonction de } I(A) :$ 

Ce sont bien sûr les résultats du 2- qui font foi, ceux du 1- n'ont pas tenu compte de sources d'incertitudes importantes.

E11 : Focométrie

énoncé p93



Dans les conditions de Gauss, la relation de conjugaison au centre de la lentille mince sphérique s'écrit :

$$\frac{1}{OA'} - \frac{1}{OA} = \frac{1}{f}, \quad , f': \text{distance focale image.}$$

Nous avons ici un objet et une image réels et suivant le choix d'axe optique $\overline{OA} < 0$ et $\overline{OA'} > 0$. Ainsi pour travailler avec des grandeurs positives nous posons : $\overline{OA} = -OA$ et $\overline{OA'} = OA'$.

La méthode la plus rapide pour mesurer la focale d'une lentille convergente est la méthode par auto-collimation. Ensuite nous avons la méthode de Bessel, plus précise mais plus longue à mettre en œuvre. L'avantage avec l'autre méthode de cet exercice, certes longue, est que nous ne travaillons pas avec une seule mesure mais avec n mesures. Plus le nombre mesure est grand plus le résultat est précis.

ΔOA contient les incertitudes géométriques (mesures de longueurs) et de modélisation (épaisseur de la lentille, conditions de Gauss). $\Delta OA'$ contient en plus les incertitudes optiques (latitudes de mise au point) :

$$\Delta OA' = \sqrt{(\Delta OA'_{\text{geo}})^2 + (\Delta OA'_{\text{mod}})^2 + (\Delta OA'_{\text{opt}})^2}$$

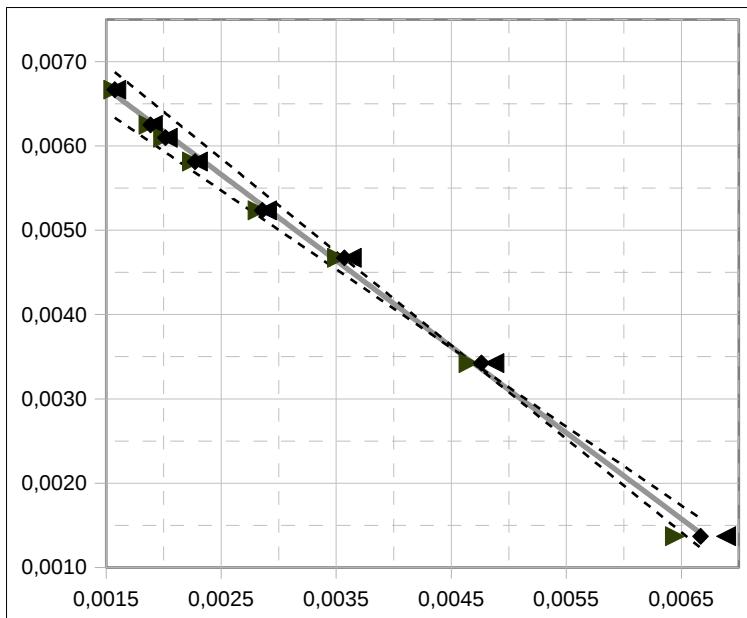
La relation entre OA et OA' est hyperbolique, nous considérons donc $x=1/OA$ et $y=1/OA'$. Nous avons alors une relations linéaire $y=ax+b$ où le coefficient directeur a doit être égal à -1 et l'ordonnée à l'origine b à $1/f'$. $\Delta x=\Delta(1/OA)=\Delta OA/OA^2$, de même pour Δy .

En utilisant un tableur (ici www.incertitudes.fr/livre/TPlentille.ods) nous obtenons :

$a = -1,02 \pm 0,09$ et $b = 0,0082 \pm 0,0004$ soit $f' = 122 \pm 6 \text{ mm}$ (cohérent avec l'indication du fabricant $f'=125 \text{ mm}$).

$a = -1$ est bien dans l'intervalle, nous vérifions aussi la validité de cette relation linéaire entre $1/OA$ et $1/OA'$. Pour préciser davantage la validité de la relation de conjugaison il faudrait aussi montrer que cette relation linéaire est celle qui donne la meilleure corrélation.

y en fonction de x :



n = 8

a estimé= -1,0210

a= -1,0210

x	dx	y	dy	poids wi	wi/S
0,00157	0,0000124	0,00667	0,000667	2249189	6%
0,00189	0,0000178	0,00625	0,000664	2265985	6%
0,00202	0,0000203	0,00610	0,000558	3210643	9%
0,00227	0,0000258	0,00581	0,000608	2696211	7%
0,00286	0,0000408	0,00524	0,000548	3308045	9%
0,00357	0,0000638	0,00467	0,000546	3308566	9%
0,00476	0,0001134	0,00342	0,000328	8248034	23%
0,00667	0,0002222	0,00137	0,000191	11349080	31%

S=
36635752

$$\begin{array}{lll} da = 0,0871 & da/|a| = 8,5\% & sa = 0,0871 \\ db = 0,000408 & db/|b| = 5,0\% & sb = 0,0004 \end{array}$$

$$\begin{array}{lll} a = & -1,021 & + ou - 0,087 \\ b = & 0,00821 & + ou - 0,00041 \\ & -1,11 & < a < -0,93 \end{array}$$

$f=1/b=$	122
$df=d(1/b)=$	6
$df/f=d(1/b)/ 1/b =$	5%

Théorie

E12 :

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n x_i^2 - 2 \sum_{i=1}^n x_i \bar{x} + \sum_{i=1}^n \bar{x}^2 = n \bar{x}^2 - 2 \bar{x} \sum_{i=1}^n x_i + n \bar{x}^2 \\ n \bar{x}^2 - 2 \bar{x} \sum_{i=1}^n x_i + n \bar{x}^2 &= n \bar{x}^2 - 2 \bar{x} n \bar{x} + n \bar{x}^2 = n (\bar{x}^2 - \bar{x}^2) \end{aligned}$$

E13 :

$$s_b = s_r \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}} = s_r \sqrt{\frac{\sum (x_i - \bar{x})^2 + n \bar{x}^2}{n \sum (x_i - \bar{x})^2}} = s_r \sqrt{\frac{n(\bar{x}^2 - \bar{x}^2) + n \bar{x}^2}{n \sum (x_i - \bar{x})^2}}$$

$$\text{d'où} \quad s_b = s_r \sqrt{\frac{n \overline{x^2}}{n \sum (x_i - \bar{x})^2}} = s_r \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}}$$

E14 :

$$s_a = \frac{s_r}{\sqrt{\sum (x_i - \bar{x})^2}} \quad \text{et} \quad s_b = s_r \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}} \quad \text{d'où} \quad s_b = s_a \sqrt{\frac{\sum x_i^2}{n}}$$

ce qui nous amène immédiatement au résultat recherché.

E15 : Asymptotes énoncé p95

Pour la droite extrême de plus grande pente :

$$\Delta y_{o+} = y_{o+} - y_o = (a_+ x_o + b_-) - (a_- x_o + b_+) = \Delta a_+ x_o - \Delta b_-$$

$$\lim_{x_o \rightarrow \infty} \Delta y_{o+} = \lim_{x_o \rightarrow \infty} (\Delta a \cdot x_o - \Delta b) = \Delta a \cdot x_o$$

Pour la courbe de prédiction :

$$\Delta y_{conf} = t_{n-2} s_r \sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = t_{n-2} s_r \sqrt{\frac{1}{n x_o^2} + \frac{\left(1 - \frac{\bar{x}}{x_o}\right)^2}{\sum (x_i - \bar{x})^2} x_o}$$

$$\text{ainsi } \lim_{x_o \rightarrow \infty} \Delta y_{conf} = t_{n-2} s_r \sqrt{0 + \frac{1}{\sum_i (x_i - \bar{x})^2}} x_o = \Delta a \cdot x_o$$

Pour la prédiction nous arrivons simplement à la même conclusion. Comme nous le pressentions expérimentalement toutes les asymptotes se rejoignent bien à l'infini.

E16 : Intervalles de confiance et prédition énoncé p95

$$1 - \bar{x} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad \text{et} \quad \bar{x}^2 = \frac{\sum w_i x_i^2}{\sum w_i}$$

$$\Delta = \sum w_i \sum w_i x_i^2 - (\sum w_i x_i)^2 = \left(\sum w_i\right)^2 (\bar{x}^2 - \bar{x}^2)$$

$$2-\text{En r\'egression simple : } \Delta y_o = t_{n-2} s_r \sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

$$s_a = \frac{s_r}{\sqrt{\sum_i (x_i - \bar{x})^2}} \text{ et } \Delta a = t_{n-2} s_a \text{ d'où } t_{n-2} s_r = \Delta a \sqrt{\sum (x_i - \bar{x})^2}$$

ainsi $\Delta y_o = \Delta a \sqrt{n(\bar{x}^2 - \bar{x}^2)} \sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{n(\bar{x}^2 - \bar{x}^2)}}$

En régression avec barres : $\Delta a = \sqrt{\frac{\sum w_i}{\Delta}}$

et $\Delta = (\sum w_i)^2 (\bar{x}^2 - \bar{x}^2)$ d'où par analogie nous avons :

$$\Delta y_o = \sqrt{\frac{\sum w_i}{\Delta}} \sqrt{n} \sqrt{\frac{\Delta}{(\sum w_i)^2}} \sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{n(\bar{x}^2 - \bar{x}^2)}}$$

soit $\Delta y_o = \frac{1}{\sqrt{n \sum w_i}} \sqrt{1 + \frac{(x_o - \bar{x})^2}{\bar{x}^2 - \bar{x}^2}}$ pour la confiance,

et $\Delta y_o = \frac{1}{\sqrt{n \sum w_i}} \sqrt{1 + n + \frac{(x_o - \bar{x})^2}{\bar{x}^2 - \bar{x}^2}}$ pour la prédition,

3- Par le même calcul qu'en régression simple $\lim_{x_o \rightarrow \infty} \Delta y_o = \Delta a \cdot x_o$

$$\Delta y_{conf} = \frac{1}{\sqrt{n \sum w_i}} \sqrt{1 + \frac{(x_o - \bar{x})^2}{\bar{x}^2 - \bar{x}^2}} = \frac{1}{\sqrt{n \sum w_i}} \sqrt{\frac{1}{x_o^2} + \frac{\left(1 - \frac{\bar{x}}{x_o}\right)^2}{\bar{x}^2 - \bar{x}^2}} x_o$$

$$\lim_{x_o \rightarrow \infty} \Delta y_{conf} = \frac{1}{\sqrt{n \sum w_i}} \sqrt{\frac{1}{\bar{x}^2 - \bar{x}^2}} x_o = \frac{1}{\sqrt{n \sum w_i}} \sqrt{\frac{(\sum w_i)^2}{\Delta}} x_o$$

$$\lim_{x_o \rightarrow \infty} \Delta y_{conf} = \frac{1}{\sqrt{n}} \sqrt{\frac{\sum w_i}{\Delta}} x_o = \frac{1}{\sqrt{n}} \Delta a \cdot x_o$$

Par analogie nous devons prendre dans les formules **n=1** pour la régression avec barres d'erreur. En fait nous devions nous attendre à un facteur de différence, en effet les poids w_i sont définis à un facteur près.

Par exemple nous aurions aussi pu prendre $w_i' = \frac{1/a^2}{(\Delta y_i/a)^2 + (\Delta x_i)^2}$.

Aussi en régression simple le nombre de points est clairement défini, alors qu'avec les barres d'erreurs certains points comptent plus que d'autres, $n=1$ signifie que tous les points correspondent à 100% des données. Pour trouver les formules de l'énoncé il suffit de remplacer n par 1.

E17 Autres expressions énoncé p96

$$\bar{x} = \frac{\sum w_i x_i}{\sum w_i} \quad \bar{x^2} = \frac{\sum w_i x_i^2}{\sum w_i} \quad \bar{xy} = \frac{\sum w_i x_i y_i}{\sum w_i} \quad \frac{\Delta}{(\sum w_i)^2} = \bar{x^2} - \bar{x}^2$$

$$a = \frac{\sum w_i \sum w_i x_i y_i - \sum w_i x_i \sum w_i y_i}{\Delta} = \frac{\bar{y} \bar{x^2} - \bar{x} \bar{xy}}{\bar{x^2} - \bar{x}^2}$$

$$b = \frac{\sum w_i y_i \sum w_i x_i^2 - \sum w_i x_i \sum w_i x_i y_i}{\Delta} = \frac{\bar{y} \bar{x^2} - \bar{x} \bar{xy}}{\bar{x^2} - \bar{x}^2}$$

$$\Delta a = \sqrt{\frac{\sum w_i}{\Delta}} = \frac{1}{\sqrt{\sum w_i}} \frac{1}{\sqrt{\bar{x^2} - \bar{x}^2}}$$

$$\Delta b = \sqrt{\frac{\sum w_i x_i^2}{\Delta}} = \frac{1}{\sqrt{\sum w_i}} \sqrt{\frac{\bar{x^2}}{\bar{x^2} - \bar{x}^2}}$$

Régression simple : $a = \frac{\bar{y} \bar{x} - \bar{x} \bar{y}}{\bar{x^2} - \bar{x}^2}$ (même formule avec les barres)

et $\bar{y} = a \bar{x} + b$ qui implique $b = \frac{\bar{y} \bar{x^2} - \bar{x} \bar{xy}}{\bar{x^2} - \bar{x}^2}$ (même formule)

$$\Delta a = t_{n-2} s_r \frac{1}{\sqrt{\sum_i (x_i - \bar{x})^2}} = \frac{t_{n-2} s_r}{\sqrt{n}} \frac{1}{\sqrt{\bar{x^2} - \bar{x}^2}} \text{ et}$$

$$\Delta b = t_{n-2} s_r \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}} = \frac{t_{n-2} s_r}{\sqrt{n}} \sqrt{\frac{\bar{x^2}}{\bar{x^2} - \bar{x}^2}}$$

Les formules sont semblables. Comme $n \sim 1$ nous avons

$\frac{1}{\sqrt{\sum w_i}} \sim t_{n-2} S_r$ ainsi $\frac{1}{\sqrt{\sum w_i}}$ pour la régression avec barres est équivalent à l'incertitude sur les résidus en régression simple.

E18 : Méthode des moindres carrés énoncé p97

❶ Régression simple : $\sum d^2 = \sum_i (y_i - a x_i - b)^2$

$$\frac{\partial(\sum d^2)}{\partial b} = \sum_i 2 \frac{\partial(y_i - a x_i - b)}{\partial b} (y_i - a x_i - b) = -2 \sum_i (y_i - a x_i - b)$$

d'où : $\sum (y_i - a x_i - b) = 0$, $\sum y_i - a \sum x_i - nb = 0$

et finalement en divisant par n : $\bar{y} = a \bar{x} + b$

$$\frac{\partial(\sum d^2)}{\partial a} = \sum_i 2 \frac{\partial(y_i - a x_i - b)}{\partial a} (y_i - a x_i - b) = -2 \sum_i x_i (y_i - a x_i - b)$$

d'où : $\sum (y_i - a x_i - b) x_i = 0$, $\sum y_i x_i - a \sum x_i^2 - b \sum x_i = 0$

et finalement en divisant par n : $\bar{x} \bar{y} - a \bar{x}^2 - b \bar{x} = 0$

$$\text{soit } \bar{x} \bar{y} - a \bar{x}^2 - (\bar{y} - a \bar{x}) \bar{x} = 0 \quad \text{et finalement : } a = \frac{\bar{x} \bar{y} - \bar{x} \bar{y}}{\bar{x}^2 - \bar{x}^2}$$

❷ Régression avec barres : $S^2 = \sum_i w_i (y_i - a x_i - b)^2$

$$\frac{\partial(\sum S^2)}{\partial b} = -2 \sum w_i (y_i - a x_i - b) \quad \text{car } w_i \text{ ne dépend pas de } b.$$

Ainsi nous obtenons aussi $\bar{y} = a \bar{x} + b$.

$$\frac{\partial w_i}{\partial a} = \frac{\partial}{\partial a} \left(\frac{1}{\Delta y_i^2 + a^2 \Delta x_i^2} \right) = -\frac{2a \Delta x_i^2}{(\Delta y_i^2 + a^2 \Delta x_i^2)^2} = -2a \Delta x_i^2 w_i^2$$

$$\frac{\partial(\sum S^2)}{\partial a} = -2a \sum_i \Delta x_i^2 w_i^2 (y_i - a x_i - b)^2 + \sum_i 2w_i (-x_i) (y_i - a x_i - b)$$

or Δx_i est petit devant x_i et y_i , nous pouvons donc négliger le premier terme devant le second et nous trouvons ainsi le même résultat qu'en régression simple.

Méthode 1 :

$$1- \quad a = \frac{\bar{xy} - \bar{x}\bar{y}}{\bar{x}^2 - \bar{x}^2} = \frac{\sum (x_i y_i - \bar{x} y_i)}{n(\bar{x}^2 - \bar{x}^2)} = \dots$$

$$2- \quad V(a) = \sum p_i^2 V(y_i) = \frac{\sum (x_i - \bar{x})^2}{\left(\sum (x_i - \bar{x})^2\right)^2} \sigma_r^2 = \dots$$

Car les variables y_i sont indépendantes et sont supposées avoir la même variance.

$$3- \quad b = \frac{\bar{y} \bar{x}^2 - \bar{x} \bar{xy}}{\bar{x}^2 - \bar{x}^2} = \frac{\sum (y_i \bar{x}^2 - \bar{x} x_i y_i)}{n(\bar{x}^2 - \bar{x}^2)} = \dots$$

$$V(b) = \frac{\sum (\bar{x}^2 - \bar{x} x_i)^2}{\left(\sum (x_i - \bar{x})^2\right)^2} \sigma_r^2 = \frac{n(\bar{x}^2)^2 - 2\bar{x}^2 \bar{x}^2 + \bar{x}^2 \bar{x}^2}{\left(\sum (x_i - \bar{x})^2\right)^2} \sigma_r^2 = \dots$$

Méthode 2 :

$$a = \frac{\sum x_i y_i - \bar{x} \sum y_i}{n(\bar{x}^2 - \bar{x}^2)} \quad \text{or}$$

$$\frac{\partial \sum y_i}{\partial y_j} = \frac{\partial (y_1 + \dots + y_j + \dots + y_n)}{\partial y_j} = 0 + \dots + 1 + \dots + 0 = 1$$

$$\text{et } \frac{\partial \sum x_i y_i}{\partial y_j} = \frac{\partial (x_1 y_1 + \dots + x_j y_j + \dots + x_n y_n)}{\partial y_j} = 0 + \dots + x_j + \dots + 0 = x_j$$

$$\text{ainsi } \frac{\partial a}{\partial y_j} = \frac{x_j - \bar{x}}{n(\bar{x}^2 - \bar{x}^2)} \quad \text{et} \quad s_a^2 = \sum \left(\frac{x_j - \bar{x}}{n(\bar{x}^2 - \bar{x}^2)} \right)^2 s_{y_i}^2 = \dots$$

Démarche analogue pour b .

Approche matricielle : le système d'équation p263 est équivalent à

$$\begin{cases} \bar{y} - b - a\bar{x} = 0 \\ \bar{x}\bar{y} - b\bar{x} - a\bar{x}^2 = 0 \end{cases}, \quad H = \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \bar{x}^2 \end{pmatrix}, \quad A = \begin{pmatrix} b \\ a \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} \bar{y} \\ \bar{x}\bar{y} \end{pmatrix}.$$

$$HA=B, \quad A=H'B, \quad H^{-1}=\frac{1}{\bar{x}^2-\bar{x}^2}\begin{pmatrix} \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \quad \text{et} \quad \sigma_b^2 = (H^{-1})_{11} \sigma_r^2$$

Nous avons une matrice de déterminant non nul, et nous avons appliqué la formule d'inversion pour une matrice 2x2 (on vérifie bien que $HH' = H'H = I$). Nous avons alors rapidement les formules recherchées.

E19 Espérance de a énoncé p97

$$E(a) = E(\sum_i p_i y_i) = \sum_i p_i E(y_i) = \sum_i p_i (\alpha x_i + \beta)$$

$$E(a) = \alpha \sum_i p_i x_i + \beta \sum_i p_i \quad \sum p_i = \frac{1}{\sum (x_i - \bar{x})^2} \sum (x_i - \bar{x}) = 0$$

$$\sum p_i x_i = \sum \frac{x_i(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \frac{1}{\sum (x_i - \bar{x})^2} \sum ((x_i - \bar{x})^2 + \bar{x}(x_i - \bar{x})) = 1 + 0$$

d'où $E(a) = \alpha$

E20 Écart-types proportionnels à y énoncé p98

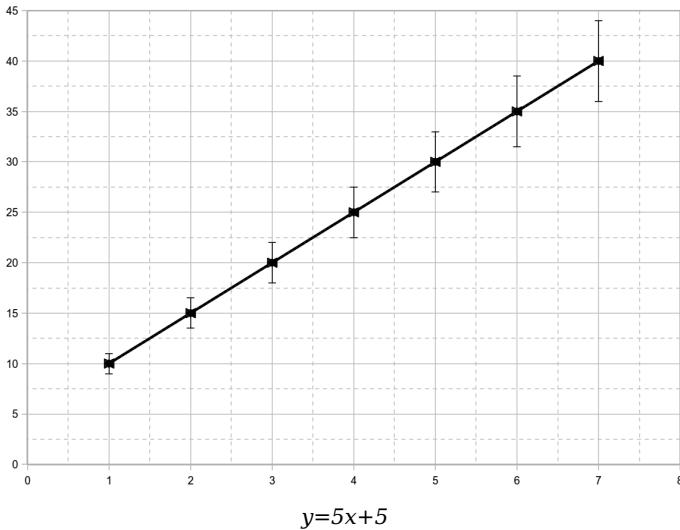
$$1- \quad a = \frac{\sum w_i \sum w_i x_i y_i - \sum w_i x_i \sum w_i y_i}{\sum w_i \sum w_i x_i^2 - (\sum w_i x_i)^2} \quad \text{avec} \quad w_i = \frac{1}{k^2 y_i^2}.$$

$$2- \quad a = \frac{u}{v} \quad \frac{\partial a}{\partial y_j} = \frac{u' v - v' u}{v^2}$$

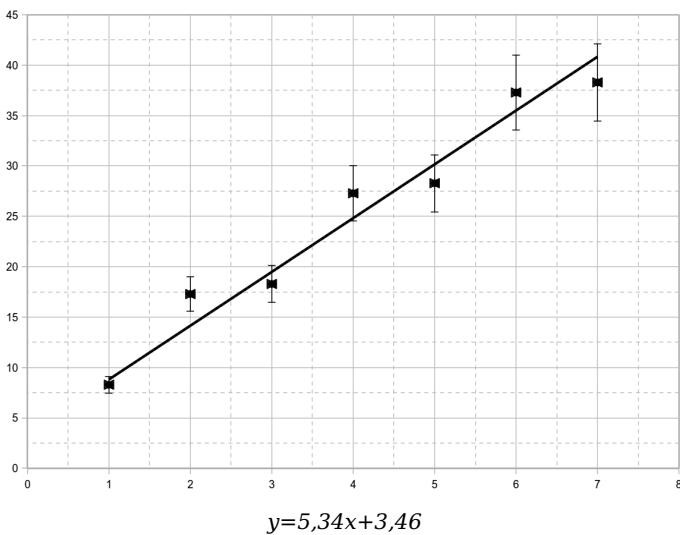
$$u' = \frac{\partial u}{\partial y_j} = -\frac{2}{y_j^3} \sum \frac{x}{y} + \sum \frac{1}{y^2} \frac{-x_j}{y_j^2} - \frac{-2x_j}{y_j^3} \sum \frac{1}{y} - \sum \frac{x}{y^2} \frac{-1}{y_j^2}$$

$$v' = \frac{\partial v}{\partial y_j} = \frac{-2}{y_j^3} \sum \frac{x^2}{y^2} + \sum \frac{1}{y^2} \frac{-2x_i^2}{y_j^3} - 2 \sum \frac{x}{y^2} \frac{-2x_j}{y_j^3}$$

Nous avons ainsi le calcul de la dérivée partielle, nous les ajoutons au carré selon toutes les valeurs de j sur un tableur afin d'obtenir s_a :



$$y = 5x + 5$$



Jeu de données 1 pour $k=0,1$: $s_a \approx 0,422$

Jeu de données 2 pour $k=0,1$: $s_a \approx 0,427$

Nous trouvons exactement les mêmes résultats avec les petites variations :

	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$
$\frac{\partial a}{\partial y_j}$	-0,2237	-0,0203	0,0331	0,0496	0,0542	0,0543	0,0528

d'où $s_a \approx 0,422$

	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$
$\frac{\partial a}{\partial y_j}$	-0,2232	-0,0016	0,0514	0,0343	0,0667	0,0410	0,0611

d'où $s_a \approx 0,427$

Nous avions avec la méthode classique p66 :

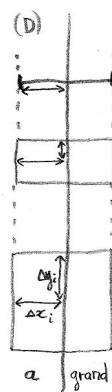
Jeu de données 1 pour $k=0,1$: $s_a \approx 0,422$

Jeu de données 2 pour $k=0,1$: $s_a \approx 0,399$

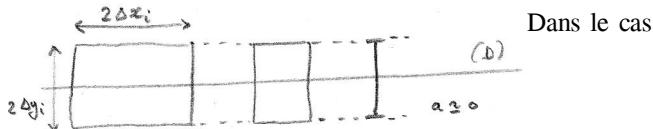
Encore une fois la méthode avec la formule de propagation et plus réaliste que la méthode classique.

E21 Interprétation de w_i énoncé p98

1- Nous voulons trouver la droite qui passe au mieux par tous les rectangles d'incertitudes, et si possible au plus près de leur centre. Si la pente de la droite est faible, la droite est proche de l'horizontale, $a \approx 0$ et seul Δy_i sera contraint pour ajuster la droite et $w_i \approx \frac{1}{(\Delta y_i)^2}$. Par exemple, un rectangle tel que $\Delta x_i = \Delta y_i$ encadrera autant la droite qu'un autre pour lequel $\Delta x_i = 0$. Au contraire si la



droite est verticale, $a \approx \infty$ et seul Δx_i sera contraint pour ajuster la droite et $w_i \approx \frac{1}{(a \Delta x_i)^2}$



intermédiaire où $a=1$, Δy_i et Δx_i jouent des rôles symétriques ce qui est satisfait par la formule $w_i = \frac{1}{(\Delta y_i)^2 + (\Delta x_i)^2}$.

E22 Décomposition en Gaussiennes énoncé p99

Nous considérons une décomposition de la forme :

$$f(x) = a_1 e^{-\frac{1}{2} \left(\frac{x-a_2}{a_3} \right)^2} + a_4 e^{-\frac{1}{2} \left(\frac{x-a_5}{a_6} \right)^2}$$

Nous avons six paramètres :

$$\frac{\partial f}{\partial a_1} = e^{-\frac{1}{2} \left(\frac{x-a_2}{a_3} \right)^2}, \quad \frac{\partial f}{\partial a_2} = \frac{a_1}{a_3^2} (x - a_2) e^{-\frac{1}{2} \left(\frac{x-a_2}{a_3} \right)^2}, \text{ etc.}$$

Nous partons des paramètres estimés suivants :

a_1	a_2	a_3	a_4	a_5	a_6
5,8	59,99	0,012	10	60,03	0,012

Si vous partez de paramètres quelconques, l'itération risque de diverger. Vous regardez d'abord si l'adéquation est graphiquement correcte et, surtout, vous essayez manuellement, à tâtons, de minimiser S^2 .

Nous obtenons les expressions suivantes pour la première itération :

$$H = \begin{vmatrix} 2,13 & 2,36 & 508 & 0,132 & -184 & 361 \\ 2,36 & 245572 & 7176 & 107 & -121329 & 202215 \\ 508 & 7176 & 354716 & 210 & -202215 & 303007 \\ 0,132 & 107 & 210 & 2,13 & -4,06 & 876 \\ -184 & -121329 & -202215 & -4,06 & 730001 & -21333 \\ 361 & 202215 & 303007 & 876 & -21333 & 1054447 \end{vmatrix}$$

$$B = \begin{vmatrix} 0,933 \\ 397 \\ 959 \\ 0,029 \\ -2767 \\ 520 \end{vmatrix} \quad \text{et } A = H^{-1}B = \begin{vmatrix} -0,0128 \\ -0,0009 \\ -0,0001 \\ -0,2927 \\ -0,0039 \\ 0,0008 \end{vmatrix}$$

Connaissant les variations des paramètres nous obtenons un nouveau jeu de paramètres pour la deuxième itération. Nous itérons autant que nécessaire pour convergence et stabilité des grandeurs :

	a_1	a_2	a_3	a_4	a_5	a_6	S^2
It. 0	5,8	59,99	0,012	10	60,03	0,012	13,2711
It. 1	5,787	59,9891	0,01194	9,707	60,0261	0,01285	2,9499
It. 2	5,784	59,9892	0,01210	10,283	60,0261	0,01197	2,3998
It. 3	5,764	59,9890	0,01191	10,271	60,0260	0,01218	2,3848
It. 4	5,762	59,9891	0,01200	10,279	60,0260	0,01212	2,3839
It. 5	5,762	59,9890	0,01197	10,276	60,0260	0,01214	2,3838
It. 6	5,762	59,9890	0,01198	10,277	60,0260	0,01214	2,3838

D'où les paramètres et matrice inverse finaux suivants :

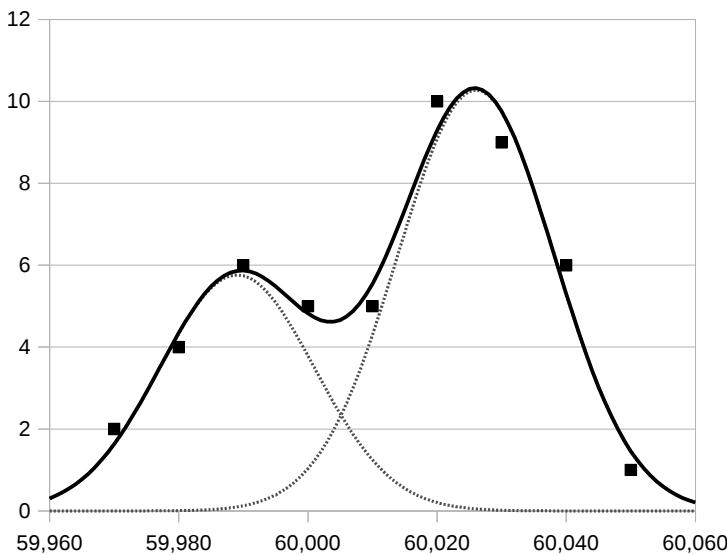
a_1	a_2	a_3	a_4	a_5	a_6
5,76	59,989	0,0120	10,28	60,026	0,0121

$$H^{-1} = \begin{vmatrix} 0,74315202 & -0,00004341 & -0,00118765 & 0,11299386 & -0,00011583 & -0,00006533 \\ -0,00004341 & 0,00000970 & 0,00000561 & 0,00017946 & 0,00000361 & -0,00000371 \\ -0,00118765 & 0,00000561 & 0,00001058 & -0,00014445 & 0,00000392 & -0,00000353 \\ 0,11299386 & 0,00017946 & -0,00014445 & 0,72254732 & 0,00002639 & -0,00062087 \\ -0,00011583 & 0,00000361 & 0,00000392 & 0,00002639 & 0,00000316 & -0,00000182 \\ -0,00006533 & -0,00000371 & -0,00000353 & -0,00062087 & -0,00000182 & 0,00000317 \end{vmatrix}$$

Nous pouvons ainsi déterminer les écart-types sur les paramètres :

$$s_r = \sqrt{\frac{\sum (y_i - f_i)^2}{n-6}} \approx 0,891 \quad \text{et} \quad s_{a_k}^2 = (H^{-1})_{kk} s_r^2 \quad \text{d'où}$$

σ_{a1}	σ_{a2}	σ_{a3}	σ_{a4}	σ_{a5}	σ_{a6}
0,77	0,003	0,003	0,76	0,002	0,0016



Un clou correspond à une aire de $0,01 \times 1$, donc pour une unité d'aire :
 $u.a.=100$ clous.

$$A = \int_{-\infty}^{+\infty} \alpha e^{-\frac{1}{2}\left(\frac{x-\beta}{\delta}\right)^2} dx = \sqrt{2\pi} \alpha \delta$$

suite au changement de variable $x' = \frac{x-\beta}{\delta}$

Pour le pic de gauche : $N_1 = \sqrt{2\pi} a_1 a_3 = 17,33 \pm 4,91$

Pour le pic de droite : $N_2 = \sqrt{2\pi} a_4 a_6 = 31,18 \pm 4,72$

Au total : $N = N_1 + N_2 = 48,5 \pm 6,8$ (cohérent avec 48 clous)

Chapitre III : Lois de probabilités

E1 : Loi binomiale énoncé p119

Probabilités positives. De plus, d'après la formule du binôme de Newton :

$$\sum_{k=0}^n P(X=k) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1$$

Probabilités comprises entre 0 et 1.

Espérance : $E(X) = \sum_{k=0}^n k P(X=k) = \sum_{k=1}^n k \frac{n!}{(n-k)!k!} p^k q^{n-k}$

d'où $E(X) = \sum_{k=1}^n \frac{n(n-1)!}{(n-1-k+1)!(k-1)!} p^k p^{k-1} q^{n-1-k+1}$

Nous sortons de la somme les grandeurs indépendantes de k :

$$E(X) = n p \sum_{k=1}^n \frac{(n-1)!}{(n-1-k+1)!(k-1)!} p^{k-1} q^{n-1-k+1}$$

Nous effectuons le changement d'indice $j=k-1$:

$$E(X) = n p \sum_{j=0}^{j=n-1} \frac{(n-1)!}{(n-1-j)!(j+1)!} p^j q^{n-1-j} = np(p+q)^{n-1} = np$$

Variance :

$$E(X^2) = \sum_{k=0}^n k^2 P(X=k) = \sum_{k=1}^n [k(k-1)+k] \frac{n!}{(n-k)!k!} p^k q^{n-k}$$

d'où $E(X^2) = \sum_{k=2}^n \frac{n!}{(n-k)!(k-2)!} p^k q^{n-k} + E(X)$ et

$$E(X^2) = \sum_{k=2}^n \frac{n(n-1)(n-2)!}{(n-2-k+2)!(k-2)!} p^2 p^{k-2} q^{n-2-k+2} + E(X)$$

$$E(X^2) = n(n-1)p^2 \times 1 + E(X) = n(n-1)p^2 + np = n^2 p^2 - np^2 + np$$

d'où $V(X) = E(X^2) - E(X)^2 = np(1-p) = npq$

E2 : Somme de binomiales énoncé p119

$$P(Z=k) = \sum_i P([X=i] \cap [Y=k-i]) = \sum_i P(X=i)P_{X=i}(Y=k-i)$$

Si les variables aléatoires sont indépendantes :

$$P(Z=k) = \sum_i P(X=i)P(Y=k-i)$$

Pour des lois binomiales $\mathcal{B}_x(n_1, p)$ et $\mathcal{B}_y(n_2, p)$:

$$P(Z=k) = \sum_{i=0}^{n_1+n_2} \binom{n_1}{i} p^i q^{n_1-i} \binom{n}{k-i} p^{k-i} q^{n_2-k+i}$$

$$\text{d'où } P(Z=k) = \left[\sum_{i=0}^{n_1+n_2} \binom{n_1}{i} \binom{n}{k-i} \right] p^k q^{n_1+n_2-k}$$

et finalement en utilisant l'identité de Vandermonde pour les coefficients binomiaux nous obtenons :

$$P(Z=k) = \binom{n_1+n_2}{k} p^k q^{n_1+n_2-k} \text{ soit une } \mathcal{B}_z(n_1+n_2, p).$$

E3 : Loi géométrique énoncé p119

Probabilités positives.

$$\sum_{k=1}^{\infty} P(X=k) = \sum_1^{\infty} q^{k-1} p = p \lim_{k \rightarrow \infty} \frac{1-q^k}{1-q} = 1$$

Probabilités comprises entre 0 et 1.

$$\text{Espérance : } \sum_{k=1}^{\infty} k P(X=k) = p \sum_1^{\infty} k q^{k-1} = p \frac{1}{(1-q)^2} = \frac{1}{p}$$

$$\text{Variance : } E(X^2) = \sum_{k=1}^{\infty} k^2 P(X=k) = p \sum_1^{\infty} [k(k-1)+k] q^{k-1}$$

$$\text{et } E(X^2) = p q \sum_2^\infty k(k-1)q^{k-2} + E(X) = p q \frac{2}{(1-q)^3} + \frac{1}{p}$$

$$\text{d'où } V(X) = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q+p-1}{p^2} = \frac{q}{p^2}$$

Somme de variables aléatoires géométriques de paramètre p indépendantes :

$$P(Z=k) = \sum_i P(X=i)P(Y=k-i) = \sum_{i=1}^{k-1} q^{i-1} p q^{k-i-1} p$$

$$\text{soit } P(Z=k) = (k-1)q^{k-2} p^2$$

Probabilité d'avoir un deuxième succès au rang k . Le premier succès a $k-1$ possibilités de la position 1 à $k-1$.

E4 : Premiers succès énoncé p119

1- $P(X=5) = (1/2)^4 \cdot 1/2 = 1/32 \approx 3\%.$ $P_{X>3}(X=8) = P(X=8-3) = P(X=5).$

2- $E(X)=1/p=6$, en moyenne après six lancers.

$$P(X \leq 6) = \sum_{k=1}^6 q^{k-1} p = p \sum_{i=0}^5 q^i = p \frac{1-q^6}{1-q} = 1-q^6 \approx 66\%$$

E5 : Loi de Poisson énoncé p120

Probabilités positives.

$$\sum_{k=0}^{\infty} P(X=k) = \sum_0^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_0^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

(définition de la fonction exponentielle en terme de série)

Probabilités comprises entre 0 et 1.

Espérance :

$$\sum_{k=0}^{\infty} k P(X=k) = \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \sum_{j=0}^{\infty} \frac{\lambda^{j+1}}{j!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda$$

Variance : $E(X^2) = \sum_{k=0}^{\infty} k^2 P(X=k) = \sum_{k=1}^{\infty} [k(k-1)+k] \frac{\lambda^k}{k!} e^{-\lambda}$

$$E(X^2) = \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} e^{-\lambda} + E(X) = e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} + \lambda = \lambda^2 + \lambda$$

d'où $V(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$

Somme de variables aléatoires de Poisson de paramètres λ_1 et λ_2 indépendantes :

$$P(Z=k) = \sum_i P(X=i) P(Y=k-i) = \sum_{i=0}^k \frac{\lambda_1^i}{i!} e^{-\lambda_1} \frac{\lambda_2^{k-i}}{(k-i)!} e^{-\lambda_2}$$

$$\text{et } P(Z=k) = \sum_{i=0}^k \frac{\lambda^k}{i!(k-i)!} e^{-\lambda_1-\lambda_2} = \left[\sum_{i=0}^k \binom{k}{i} \lambda_1^i \lambda_2^{k-i} \right] \frac{e^{-\lambda_1-\lambda_2}}{k!}$$

D'après la formule du binôme de Newton :

$$P(Z=k) = \frac{(\lambda_1 + \lambda_2)^k}{k!} e^{-(\lambda_1 + \lambda_2)}$$

L'expression obtenue correspond à une loi de Poisson de paramètre $\lambda_1 + \lambda_2$.

E6 : Loi uniforme énoncé p120

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{3} \frac{b^3 - a^3}{b-a} = \frac{1}{3} \frac{(b-a)(b^2 + ab + a^2)}{b-a}$$

d'où $V(X) = E(X^2) - E(X)^2 = \frac{1}{3} (b^2 + ab + a^2) - \frac{(a+b)^2}{4}$

$$\text{et } V(X) = \frac{(b-a)^2}{12}$$

E7 : Loi exponentielle énoncé p120

Densité de probabilité positive.

$$\int_{-\infty}^{+\infty} f(t) dt = \int_0^{+\infty} \lambda e^{-\lambda t} dt \text{ (intégrale généralisée)}$$

$$\text{or } I_A = \int_0^A \lambda e^{-\lambda t} dt = [-e^{-\lambda t}]_0^A = 1 - e^{-\lambda A} \text{ d'où } \lim_{A \rightarrow +\infty} I_A = 1$$

$$\text{Espérance : } E(T) = \int_{-\infty}^{+\infty} t f(t) dt = \int_0^{+\infty} \lambda t e^{-\lambda t} dt$$

$$J_A = \int_0^A \lambda t e^{-\lambda t} dt = [-t e^{-\lambda t}] + \int e^{-\lambda t} dt = -A e^{-\lambda A} - \frac{1}{\lambda} e^{-\lambda A} + \frac{1}{\lambda}$$

(intégration par parties)

$$\text{et } E(T) = \lim_{A \rightarrow +\infty} J_A = \frac{1}{\lambda}$$

$$\text{Variance : } E(T^2) = \int_{-\infty}^{+\infty} t^2 f(t) dt = \int_0^{+\infty} \lambda t^2 e^{-\lambda t} dt$$

Deux intégrations par parties successives :

$$K_A = \int_0^A \lambda t^2 e^{-\lambda t} dt = [-t^2 e^{-\lambda t}] + 2 \int t e^{-\lambda t} dt$$

$$\text{d'où } E(T^2) = \lim_{A \rightarrow +\infty} K_A = \frac{2}{\lambda^2} \text{ et } V(T) = \frac{1}{\lambda^2}$$

Somme de lois exponentielles indépendantes de même paramètre :

$$f_Z(x) = \int_{-\infty}^{+\infty} f_X(y) f_Y(x-y) dy$$

- si $x < 0$ alors $f_Z(x) = 0$

- si $x > 0$ alors $f_Z(x) = \int_0^x \lambda e^{-\lambda y} \lambda e^{-\lambda(x-y)} dy$

et $f_Z(x) = \lambda^2 e^{-\lambda x} \int_0^x dy = \lambda^2 x e^{-\lambda x}$, il ne s'agit pas d'une loi exponentielle, il s'agit d'une loi Gamma de paramètres 2 et $1/\lambda$.

E8 : Somme de gaussiennes énoncé p120

$$1- f_Z(x) = \int_{-\infty}^{+\infty} f_X(y) f_Y(x-y) dy = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} e^{-\frac{(x-y)^2}{2}} dy$$

$$f_Z(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}[y^2 + (x-y)^2]} dy = \frac{1}{2\pi} e^{-\frac{x^2}{4}} \int_{-\infty}^{+\infty} e^{-\left(y - \frac{x}{2}\right)^2} dy$$

d'où $f_Z(x) = \frac{1}{2\sqrt{\pi}} e^{-\frac{x^2}{4}} = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{1}{2}\frac{x^2}{\sqrt{2}}}$ nous reconnaissons bien

une loi normale. Loi centrée et comme attendue d'écart-type $\sqrt{2}$, en effet $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 = 1+1=2$.

$$2- f_Z(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{y-\mu_1}{\sigma_1}\right)^2} e^{-\frac{1}{2}\left(\frac{x-y-\mu_2}{\sigma_2}\right)^2} dy$$

nous avons ensuite un

calcul un peu long : $f_Z(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{\sigma_2^2(y-\mu_1)^2 + \sigma_1^2(x-y-\mu_2)^2}{2\sigma_1^2\sigma_2^2}} dy = \dots$ qui

une fois mené, fournit bien une loi normale de moyenne $\mu_Z = \mu_1 + \mu_2$ et de variance $\sigma_Z^2 = \sigma_1^2 + \sigma_2^2$.

E9 : Premières Students énoncé p120

$$f_1(x) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(1)}{\Gamma\left(\frac{1}{2}\right)(1+x^2)} \quad \text{d'où} \quad f_1(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

$$\text{de même } f_2(x) = \frac{1}{2\sqrt{2}} \frac{1}{\left(1+\frac{x^2}{2}\right)^{\frac{3}{2}}} , \quad f_3(x) = \frac{2}{\sqrt{3}\pi} \frac{1}{\left(1+\frac{x^2}{3}\right)^2}$$

$$\text{et } f_4(x) = \frac{1}{\sqrt{4\pi}} \frac{\Gamma\left(\frac{5}{2}\right)}{\Gamma(2)\left(1+\frac{x^2}{4}\right)^{\frac{5}{2}}} = \frac{1}{2\sqrt{\pi}} \frac{\frac{3}{2} \frac{1}{2} \Gamma\left(\frac{1}{2}\right)}{1! \left(1+\frac{x^2}{4}\right)^{\frac{5}{2}}}$$

$$\text{soit } f_4(x) = \frac{3}{8} \frac{1}{\left(1+\frac{x^2}{4}\right)^{\frac{5}{2}}}$$

E10 : Loi de Student énoncé p121

Densité de probabilité positive.

Montrons que quelque soit $k \geq 1$, $\int_{-\infty}^{+\infty} f_k(x) dx = 1$.

Nous supposons les intégrales convergentes.

$$dx = \sqrt{k} du \quad \text{d'où, avec le changement de variable puis en utilisant la parité :} \quad I_k = \int_{-\infty}^{+\infty} (1+u^2)^{-\frac{k+1}{2}} \sqrt{k} du = 2\sqrt{k} \int_0^{+\infty} (1+u^2)^{-\frac{k+1}{2}} du$$

Posons l'intégrale J_k telle que : $I_k = 2\sqrt{k} J_k$.

Considérons tout d'abord le cas où k est pair : $k=2n$

$$\text{soit } J_n = \int_0^{+\infty} (1+u^2)^{-n-\frac{1}{2}} du = \int_0^{+\infty} \frac{1}{\sqrt{1+u^2}} \frac{1}{(1+u^2)^n} du$$

$$\text{or } J_n = \int_0^{+\infty} \frac{1}{\sqrt{1+u^2}} \frac{1+u^2-u^2}{(1+u^2)^n} du \quad \text{d'où}$$

$$J_n = J_{n-1} - \int_0^{+\infty} \frac{1}{\sqrt{1+u^2}} \frac{u^2}{(1+u^2)^n} du, \text{ soit } K_n = \int_0^{+\infty} \frac{1}{\sqrt{1+u^2}} \frac{u^2}{(1+u^2)^n} du$$

Nous effectuons une intégration par parties :

$$K_n = 0 + \frac{1}{2\left(n-\frac{1}{2}\right)} J_{n-1} = \frac{J_{n-1}}{2n-1} \quad \text{et} \quad J_n = \frac{2n-2}{2n-1} J_{n-1}$$

$$\text{d'où } J_n = \frac{2n-2}{2n-1} \frac{2n-4}{2n-3} J_{n-2} = \dots = \frac{2n-2}{2n-1} \frac{2n-4}{2n-3} \cdots \frac{2}{3} J_1$$

Calcul de J_1 : $J_1 = \int_0^{+\infty} \frac{1}{(1+u^2)^{3/2}} du$ nous effectuons le changement de variable $u = sh x$ ²⁵ d'où $du = ch x dx$ et $J_1 = \int_0^{+\infty} \frac{1}{ch^2 x} dx$

$$\text{soit } J_1 = [th x]_0^{+\infty} = 1$$

Calcul de J_n : $J_n = \frac{n-1}{n-1/2} \frac{n-2}{n-3/2} \cdots \frac{1}{3/2} = \frac{\sqrt{\pi} \Gamma(n)}{2 \Gamma(n+1/2)}$

$$\text{d'où : } J_k = \frac{\sqrt{\pi} \Gamma\left(\frac{k}{2}\right)}{2 \Gamma\left(\frac{k+1}{2}\right)}$$

En considérant le cas k impair nous obtenons la même expression.

25 $sh x$ « s » « h » « x » : sinus hyperbolique avec $sh x = \frac{e^x - e^{-x}}{2}$,
 $ch x$: cosinus hyperbolique avec $ch x = \frac{e^x + e^{-x}}{2}$ et $th x = \frac{sh x}{ch x}$.

Finalement :

$$\int_{-\infty}^{+\infty} f_k(x) dx = \frac{1}{\sqrt{k\pi}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} I_k = \frac{1}{\sqrt{k\pi}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} 2\sqrt{k} \frac{\sqrt{\pi} \Gamma\left(\frac{k}{2}\right)}{2\Gamma\left(\frac{k+1}{2}\right)} = 1$$

E11 : Variance d'une Student énoncé p121

$$V(X) = E(X^2) - E(X)^2 = E(X^2) = \int_{-\infty}^{+\infty} x^2 f_k(x) dx$$

Cette intégrale converge si $k \geq 3$.

$$V(X) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \int_{-\infty}^{+\infty} \frac{x^2}{\left(1 + \frac{x^2}{k}\right)^{\frac{k+1}{2}}} dx = \frac{2}{\sqrt{k\pi}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \int_0^{+\infty} \frac{ku^2 \sqrt{k}}{\left(1 + u^2\right)^{\frac{k+1}{2}}} du$$

Tout d'abord si k est pair, en reprenant les notations et résultats de l'exercice précédent, $k=2n$:

$$V(X) = \frac{2k}{\sqrt{\pi}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \int_0^{+\infty} \frac{u^2}{\left(1 + u^2\right)^{\frac{k+1}{2}}} du = \frac{4n}{\sqrt{\pi}} \frac{\Gamma\left(n + \frac{1}{2}\right)}{\Gamma(n)} K_n$$

$$\text{Nous avions trouvé : } K_n = \frac{J_{n-1}}{2n-1} = \frac{\sqrt{\pi} \Gamma(n-1)}{2(2n-1)\Gamma(n-1/2)}$$

$$K_n = \frac{\sqrt{\pi} \Gamma(n)(n-1/2)}{2(2n-1)(n-1)\Gamma(n+1/2)} = \frac{\sqrt{\pi} \Gamma(n)}{4(n-1)\Gamma(n+1/2)}$$

$$\text{d'où : } V(X) = \frac{n}{n-1} = \frac{k/2}{k/2-1} = \frac{k}{k-2}$$

Pour k impair nous obtenons la même expression.

E12 : Somme de Students énoncé p121

- Pour $k=1$:

$$f_z(x) = \int_{-\infty}^{+\infty} f_1(y) f_1(x-y) dy = \frac{1}{\pi^2} \int_{-\infty}^{+\infty} \frac{1}{1+y^2} \frac{1}{1+(x-y)^2} dy$$

Puis en utilisant un logiciel de calcul symbolique nous trouvons :

$$f_z(x) = \frac{1}{2\pi} \frac{1}{1+x^2/4}$$

Dans ce cas nous retrouvons bien la forme générale d'une Student pour laquelle $k=1$.

- Pour $k=3$:

$$f_{f_3*f_3}(x) = \int_{-\infty}^{+\infty} f_3(y) f_3(x-y) dy = \frac{5\sqrt{3}}{12\pi} \frac{1+x^2/60}{(1+x^2/12)^3}$$

Nous ne reconnaissons pas la forme générale d'une Student.

- Pour $k=5$:

$$f_{f_5*f_5}(x) = \frac{400\sqrt{5}}{3\pi} \frac{x^4+120x^2+8400}{(x^2+20)^5}$$

Nous ne reconnaissons pas la forme générale d'une Student. Le coefficient d'aplatissement β_2 vaut 6 alors que le degré du polynôme au dénominateur moins celui du numérateur vaut 6. Ces données ne correspondent pas à une Student. Un tableau présenté à la suite résume les différentes propriétés des premières Students.

- Pour $k=7$:

$$f_{f_7*f_7}(x) = \frac{10976\sqrt{7}}{25\pi} \frac{5x^6+1092x^4+112112x^2+9417408}{(x^2+28)^7}$$

Nous ne reconnaissons pas la forme générale d'une Student. Nous pouvons à nouveau étudier les facteurs de forme : le coefficient d'aplatissement β_2 vaut 4, β_4 vaut 50 alors que le degré du polynôme au dénominateur moins celui du numérateur vaut 8.

Ces données ne correspondent pas à une Student.

k	d°	V	β_2	β_4	$f_k(x)$	$f_k(0)$
1	2	-	-	-	$\frac{1}{\pi} \frac{1}{1+x^2}$	$\approx 0,318$
2	3	-	-	-	$\frac{1}{2\sqrt{2}} 1/\left(1+\frac{x^2}{2}\right)^{\frac{3}{2}}$	$\approx 0,354$
3	4	3	-	-	$\frac{2}{\sqrt{3}\pi} 1/\left(1+\frac{x^2}{3}\right)^2$	$\approx 0,366$
4	5	2	-	-	$\frac{3}{8} 1/\left(1+\frac{x^2}{4}\right)^{\frac{5}{2}}$	$\approx 0,375$
5	6	$5/3$ $\approx 1,67$	9	-	$\frac{8}{3\sqrt{5}\pi} 1/\left(1+\frac{x^2}{5}\right)^3$	$\approx 0,380$
6	7	$3/2$ $= 1,5$	6	-	$\frac{15}{16\sqrt{6}} 1/\left(1+\frac{x^2}{6}\right)^{\frac{7}{2}}$	$\approx 0,383$
7	8	$7/5$ $= 1,4$	5	125	$\frac{16}{5\sqrt{7}\pi} 1/\left(1+\frac{x^2}{7}\right)^4$	$\approx 0,385$
8	9	$4/3$ $\approx 1,33$	$9/2$ $= 4,5$	67,5	$\frac{35}{64\sqrt{2}} 1/\left(1+\frac{x^2}{8}\right)^{\frac{9}{2}}$	$\approx 0,387$
9	10	$9/7$ $\approx 1,29$	$21/5$ $= 4,2$	49	$\frac{128}{105\pi} 1/\left(1+\frac{x^2}{9}\right)^5$...
10	11	$5/4$ $= 1,25$	4	
11	12	$11/9$ $\approx 1,22$	$27/7$ $\approx 3,9$			
12	13	$6/5$ $= 1,2$	$15/4$ $= 3,75$			
...						
∞		1	3	15	$1/\sqrt{2\pi} e^{-x^2/2}$	$\approx 0,399$

E13 : Lois du Khi-Deux énoncé p121

$$f_1(x) = \frac{1}{\sqrt{2\pi}x} e^{-\frac{x}{2}} \quad f_2(x) = \frac{1}{2} e^{-\frac{x}{2}}$$

$$f_3(x) = \sqrt{\frac{x}{2\pi}} e^{-\frac{x}{2}} \quad f_{10}(x) = \frac{x^4}{768} e^{-\frac{x}{2}}$$

E14 : Loi du produit énoncé p121

1. $W = \ln Z = \ln X + \ln Y = U + V$. Nous cherchons les lois de U et V , nous utilisons la loi de la somme et nous cherchons la loi de $Z = e^W$.

2. a. Loi de U et de V : $f_{U,V}(x) = \begin{cases} 0 & x \leq 0 \\ e^x & 0 < x < \ln 2 \\ 0 & x \geq \ln 2 \end{cases}$

Loi de W : $f_W(x) = \int_{-\infty}^{+\infty} f_U(y) f_V(x-y) dy$

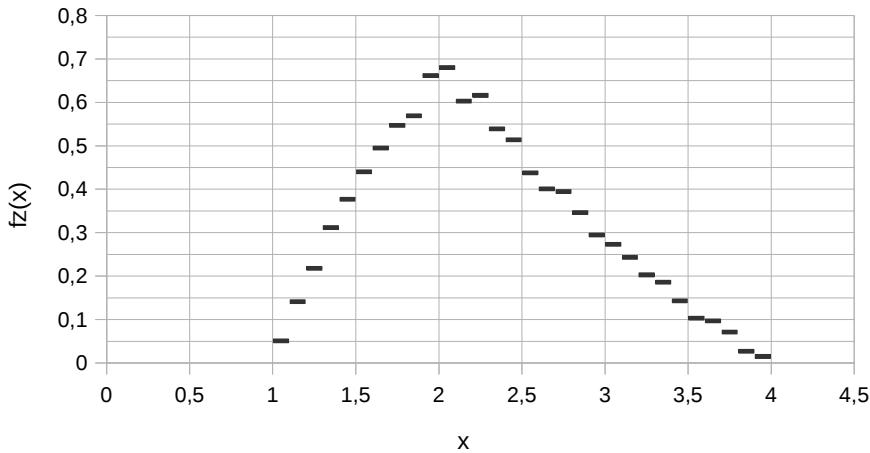
Après calcul : $f_W(x) = \begin{cases} 0 & x \leq 0 \\ xe^x & 0 < x \leq \ln 2 \\ (2\ln 2 - x)e^x & \ln 2 < x < 2\ln 2 \\ 0 & x \geq 2\ln 2 \end{cases}$

Loi de Z : $f_Z(x) = \frac{1}{x} f_W(\ln x)$ si $x > 0$ et zéro sinon.

Après calcul : $f_Z(x) = \begin{cases} 0 & x \leq 1 \\ \ln x & 1 < x \leq 2 \\ 2\ln 2 - \ln x & 2 < x < 4 \\ 0 & x \geq 4 \end{cases}$

2. b. Fichier pour la simulation sur tableur sous Libre Office : www.incertitudes.fr/livre/ProduitXY.ods

Simulation de la loi de densité de $Z=XY$ avec $n=10\ 000$



E15 : Loi de la somme de loi exponentielles énoncé p122

1. Loi exponentielle : $f_x(x)=\begin{cases} 0 & \text{si } x \leq 0 \\ \lambda e^{-\lambda x} & \text{si } x > 0 \end{cases}$

Somme de lois exponentielles : $f_{S_2}(x)=\int_{-\infty}^{+\infty} f_{X_1}(y) f_{X_2}(x-y) dy$

Intégrale non nulle si $y>0$ et $x-y>0$ et donc $0<y<x$:

$$f_{S_2}(x)=\lambda^2 e^{-\lambda x} \int_0^x dy \quad \text{et} \quad f_{S_2}(x)=\begin{cases} 0 & \text{si } x \leq 0 \\ \lambda^2 x e^{-\lambda x} & \text{si } x > 0 \end{cases}$$

2. Si nous continuons, nous obtenons la loi de $S_3=S_2+X_3$ par le même

méthode : $f_{S_3}(x)=\int_{-\infty}^{+\infty} f_{S_2}(y) f_{X_3}(x-y) dy$, non nulle si $0<y<x$.

$$f_{S_3}(x) = \lambda^3 e^{-\lambda x} \int_0^x y dy \quad \text{et} \quad f_{S_3}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \lambda^3 \frac{x^2}{2} e^{-\lambda x} & \text{si } x > 0 \end{cases}$$

Nous démontrons ainsi par récurrence la loi de S_n :

$$f_{S_n}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \lambda^n \frac{x^{n-1}}{(n-1)!} e^{-\lambda x} & \text{si } x > 0 \end{cases}$$

3. La loi de M_n est une fonction affine S_n : $f_{M_n}(x) = n f_{S_n}(nx)$

$$\text{D'où : } f_{M_n}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ n^n \lambda^n \frac{x^{n-1}}{(n-1)!} e^{-n\lambda x} & \text{si } x > 0 \end{cases}$$

E16 : Loi de l'inverse énoncé p122

$$1. F_Y(y) = P(Y \leq y) = P(X \geq 1/y) = 1 - P(X < 1/y)$$

$$\text{d'où } F_Y(y) = 1 - F_X(1/y) \text{ et } f_Y(y) = \frac{1}{y^2} f_X\left(\frac{1}{y}\right).$$

2. a.

$$f_{T_n}(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{n^n \lambda^n}{(n-1)! x^{n+1}} e^{-\frac{n\lambda}{x}} & \text{si } x > 0 \end{cases}$$

$$\text{b. } f_Y(y) = \frac{1}{y^2} \frac{1}{\pi} \frac{1}{1 + \frac{1}{y^2}} = \frac{1}{\pi} \frac{1}{1 + y^2}$$

L'inverse d'une loi de Cauchy est aussi une loi de Cauchy.

Chapitre IV : Estimateurs

E1 : Estimateurs de la moyenne énoncé p144

$E(A_3) = (E(X_1) + E(X_2) + E(X_3))/3$ par linéarité de l'espérance.
D'où : $E(A_3) = 3xm/3$ et $E(A_3) = m$. Aussi : $E(B_3) = m$.

Par contre : $E(C_3) = 4/3 m$.

Conclusion : A_3 et B_3 sont des estimateurs de m . C_3 n'est pas un estimateur de m . A_3 et B_3 sont des estimateurs non biaisés.

$V(A_3) = 1/3^2 (V(X_1) + V(X_2) + V(X_3))$, variables indépendantes.

D'où : $r_{A_3}(m) = V(A_3) = \sigma^2/3$.

$V(B_3) = 1/6^2 V(X_1) + 2^2/6^2 V(X_2) + 3^2/6^2 V(X_3)$

et $r_{B_3}(m) = V(B_3) = 7\sigma^2/18$.

Conclusion : $r_{A_3}(m) < r_{B_3}(m)$ et A_3 est un meilleur estimateur de la moyenne que B_3 .

E2 : Faisceau homocinétique énoncé p144

$$\sum_{i=1}^{100} v_i = 121,3 \cdot 10^3 \quad \text{et} \quad \sum_{i=1}^{100} v_i^2 = 152,3 \cdot 10^6$$

D'après le cours nous avons l'estimateur suivant, sans biais, de la

$$\text{moyenne } v_m : T_n = \bar{V}_n = \frac{1}{n} \sum_{i=1}^n V_i .$$

$$\text{D'où : } v_m = \frac{1}{100} \sum_{i=1}^{100} v_i \quad \text{et} \quad v_m = 1213 \text{ m/s.}$$

D'après le cours nous avons l'estimateur suivant, sans biais, de la variance σ_v^2 :

$$R_n^2 = \frac{\sum_{i=1}^n (V_i - \bar{V}_n)^2}{n-1} = \frac{n}{n-1} \frac{\sum_{i=1}^n (V_i - \bar{V}_n)^2}{n} = \frac{n}{n-1} \left(\frac{\sum_{i=1}^n V_i^2}{n} - \bar{V}_n^2 \right) .$$

$$\text{D'où : } \sigma_v^2 = \frac{100}{99} \left[\frac{1}{100} \sum_{i=1}^{100} v_i^2 - \left(\frac{1}{100} \sum_{i=1}^{100} v_i \right)^2 \right] \quad \text{et} \quad \sigma_v^2 = 52,15 \cdot 10^3 .$$

Or conformément au théorème central limite nous avons montré que $V(T_n) = \frac{\sigma_v^2}{n}$. La taille de l'échantillon est suffisante et la distribution de l'estimateur a un profil gaussien :

$$v = v_m \pm t \sigma_{T_n} = v_m \pm t_{\infty, 95\%} \sigma_v / \sqrt{n} = 1213 \pm 1,96 \cdot 228 / 10$$

et $v = 1213 \pm 45$ m/s avec 95 % de confiance.

E3 : Deux estimateurs énoncé p145

1. Pour tout θ $\sum p_i = 1$, de plus $0 < p_i < 1$ d'où $\theta \in]0; 1/4[$.
2. $E(X) = 0x30 + 1x0 + 2x(1-40)$ soit $E(X) = 2-70$. $E(X^2) = 4-150$.
 $V(X) = 4-150 - (2-70)^2$ et $V(X) = \theta(13-49\theta)$.
3. $E(T_n) = aE(\bar{X}_n) + b$ par linéarité de l'espérance.
 or $E(\bar{X}_n) = 1/n \cdot \sum E(X_i) = E(X)$ d'où $E(T_n) = a(2-70) + b = \theta$ (sans biais)
 et $a = -1/7$ et $b = 2/7$. Soit $T_n = (2-\bar{X}_n)/7$.
 $V(T_n) = 1/49 \cdot V(\bar{X}_n)$ ou $V(\bar{X}_n) = V(X)/n$ donc $V(T_n) = \theta(13-49\theta)/49n$
4. $E(Y_i) = \theta$ d'où $E(Z_n) = n\theta$ et $E(U_n) = \theta$, estimateur sans biais.

Valeurs de Y	0	1
Probabilités	1-θ	θ

$V(U_n) = 1/n^2 \cdot V(Z_n)$ or $V(Z_n) = nV(Y)$ et $V(Y) = \theta - \theta^2$ et $V(U_n) = \theta(1-\theta)/n$.
 5. $\bar{X}_{100} = (0x31 + 1x12 + 2x57)/100 = 1,26$ soit $T_{100} = (2-1,26)/7$ et $\underline{\theta_T \simeq 0,106}$
 $r_T(\theta) = V(T_{100}) \simeq 0,106(13-49,0,106)/4900$ et $r_U(\theta) = V(U_{100}) \simeq 1,7 \cdot 10^{-4}$.
 $U_{100} = 12/100$ soit $\underline{\theta_U = 0,12}$. $V(U_{100}) = 0,12 \cdot 0,88/100$ et
 $r_U(\theta) = V(U_{100}) \simeq 10,6 \cdot 10^{-4}$. $r_T(\theta) < r_U(\theta)$: Nous préférons T_{100} .
 Et T_n en général car $r_T - r_U = -360/49n < 0$ pour tout θ et n .

E4 : Urnes énoncé p146

$E(P_i) = p$ avec $i = 1$ ou 2 (espérance d'une loi binomiale).
 $E(T) = (E(P_1) + E(P_2))/2 = 2p/2$ et $E(T) = p$.
 $E(U) = xp + (1-x)p$ et $E(U) = p$.
 Nous avons deux estimateurs non biaisés de p .
 $V(P_i) = n_i pq$ avec $q = 1-p$ (variance d'une loi binomiale).
 $V(T) = 1/4(V(P_1) + V(P_2))$ car P_1 et P_2 sont des variables indépendantes.

D'où $V(T) = (n_1 + n_2)pq/4$. Aussi $V(U) = (x^2 n_1 + (1-x)^2 n_2)pq$.

$dV(U)/dx = (2xn_1 - 2(1-x)n_2)pq = 0$ soit $x_{\text{opt}} = n_2/(n_1 + n_2)$.

$V(U) = n_1 n_2 pq / (n_1 + n_2)$ et $V(U) - V(T) = -(n_1 - n_2)^2 / 4(n_1 + n_2)$.

U a un risque quadratique inférieur à V, donc U est un meilleur estimateur que V pour x_{opt} (U=V si $n_1=n_2$).

E5 : Variable à densité énoncé p146

1. Nous avons une densité positive.

Évaluons l'intégrale $\int_{-\infty}^{\infty} f(x) dx$ et montrons qu'elle vaut 1 :

$$I_A = \int_1^A \frac{a}{x^{a+1}} dx = a[-x^{-a}/a]_1^A = 1 - 1/A^a \text{ et } \lim_{A \rightarrow +\infty} I_A = 1.$$

$$2. J_A = \int_1^A \frac{ax}{x^{a+1}} dx = a[x^{-a+1}/(-a+1)]_1^A \text{ et } E(X) = \frac{a}{a-1}$$

$$3. V(x_i; a) = \prod_{i=1}^n f(x_i; a) = \prod_{i=1}^n \frac{a}{x_i^{a+1}} = \frac{a^n}{(\prod x_i)^{a+1}}$$

$$\text{d'où } \ln V = n \ln a - (a+1) \sum_{i=1}^n \ln x_i \text{ et } \frac{d \ln V}{da} = \frac{n}{a} - \sum_{i=1}^n \ln x_i$$

$$\text{et } \frac{d \ln V}{da} = 0 \text{ donne } a = \frac{n}{\sum \ln x_i} = \frac{n}{\ln(\prod x_i)}.$$

$$\text{Méthode des moments : } \bar{x} = \frac{a}{a-1} \text{ et } a = \frac{\bar{x}}{\bar{x}-1}.$$

4.

- Maximum de vraisemblance : $\hat{a} \simeq \frac{8}{\ln(14,18)}$ et $\hat{a} \simeq 3,02$.

- Théorème des moments : $\bar{x} \simeq 1,52$ et $\hat{a} \simeq 2,94$.

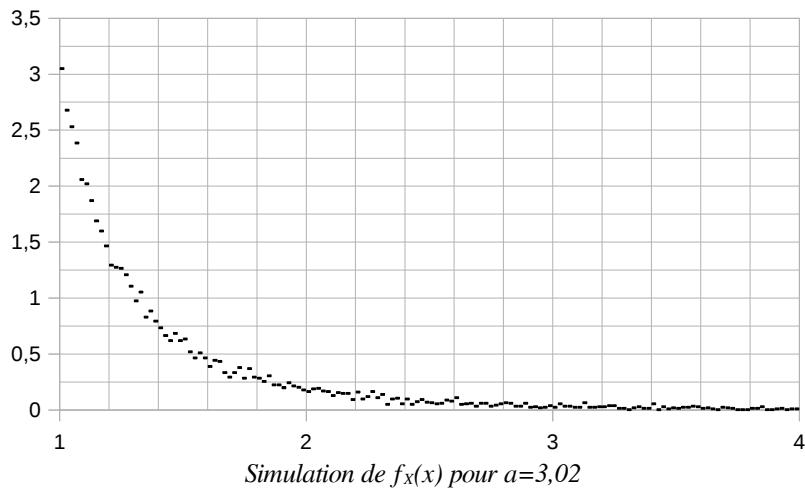
5. $K_A = \int_1^A \frac{a}{x^{a-1}} dx = \frac{a}{2-a} [1/x^{a-2}]_1^A$ cette intégrale ne tend vers une limite finie que si $a > 2$. Sinon l'intégrale est divergente et la variance n'est pas définie.

$$\text{Si } a > 2 \quad E(X^2) = \frac{a}{a-2} \quad \text{et} \quad V(X) = \frac{a}{(a-2)(a-1)^2}.$$

6. Nous voulons simuler X par une loi uniforme $U(0,1)$. Nous utilisons la méthode d'inversion. Le calcul de fonction de répartition donne :

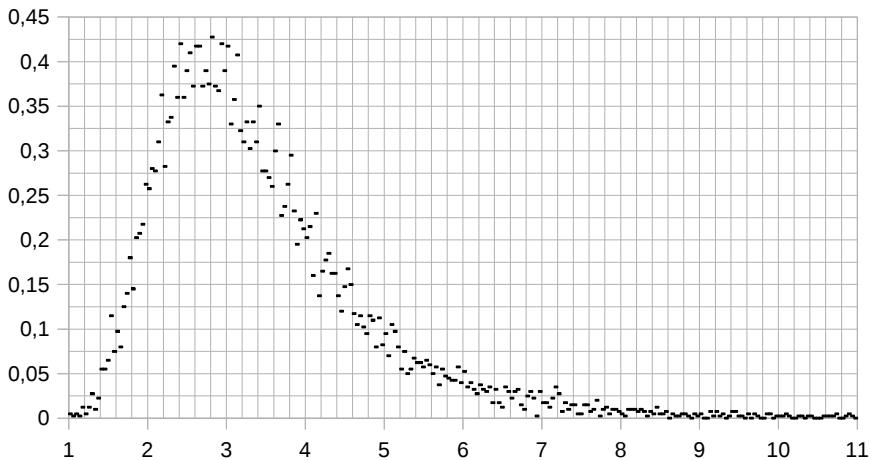
$$F(x) = 1 - \frac{1}{x^a} = y \quad \text{et} \quad x = \frac{1}{(1-y)^{1/a}} \quad \text{ainsi} \quad X = \frac{1}{U^{1/a}}.$$

Nous obtenons pour notre échantillon de taille $n=8$ et $N=10000$ tirages la distribution suivante :



Nous trouvons par la simulation : $E(X) \approx 1,49$ et $\sigma_X \approx 0,85$, valeurs qui correspondent aux valeurs théoriques.

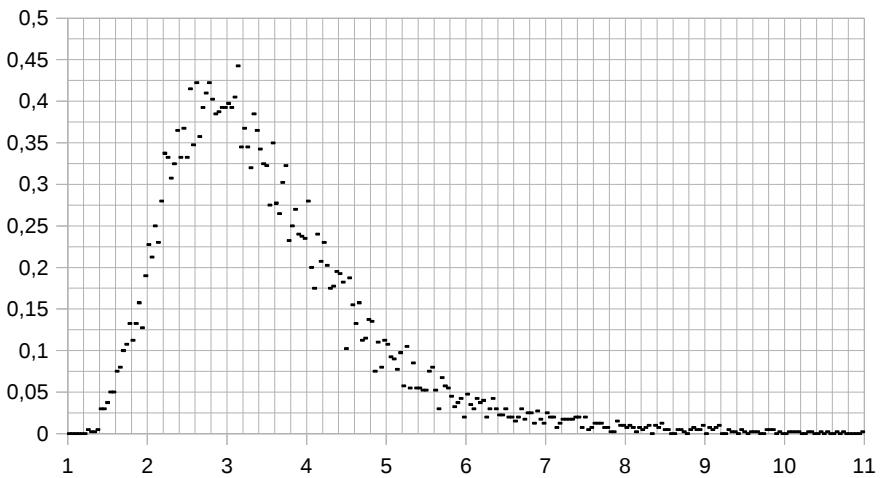
Pour l'estimateur du maximum de vraisemblance nous avons la distribution suivante pour $N=10000$:



Simulation de l'estimateur du maximum de vraisemblance pour $a=3,02$

Nous trouvons par la simulation une moyenne de 3,44 et un écart-type de 1,4.

Pour l'estimateur du théorème des moments nous avons la distribution suivante pour N=10000 :



Simulation de l'estimateur du théorème des moments pour $a=2,94$

Nous trouvons par la simulation une moyenne de 3,54 et un écart-type de 1,4.

Dans les deux cas nous avons des biais importants. La moyenne, espérance de l'estimateur, est bien au dessus de l'estimation ponctuelle de a . La variance est aussi importante, par exemple à 90% avec la méthode du maximum de vraisemblance, l'intervalle de a se situe entre 1,8 et 6,1 environ.

À ce stade de l'étude nous sommes ni assuré d'un biais asymptotiquement nul, ni d'une convergence. Il faudrait effectuer des simulations pour différentes tailles d'échantillons afin de conjecturer l'évolution en fonction de n du biais et du risque.

Nous pouvons aussi entreprendre un calcul analytique de la distribution des estimateurs par produits de convolutions et composition de fonctions.

Simulations sur le tableur de LibreOffice :

www.incertitudes.fr/livre/Simul5.ods

E6 : Densité linéaire énoncé p147

$$1. \int_0^1 (ax+b)dx = \frac{a}{2} + b = 1 \text{ et } b = 1 - \frac{a}{2}$$

De plus $f(0)$ et $f(1)$ doivent être positives d'où : $-2 \leq a \leq 2$.

$$2. E(X) = \int_0^1 (ax^2 + bx)dx = \frac{a}{3} + \frac{b}{2} \text{ et } E(X) = \frac{1}{2} + \frac{a}{12}$$

$$E(X^2) = \frac{a}{4} + \frac{b}{3} = \frac{1}{3} + \frac{a}{12} \text{ et } V(X) = \frac{1}{3} + \frac{a}{12} - \left(\frac{1}{2} + \frac{a}{12}\right)^2$$

$$\text{soit : } V(X) = \frac{1}{12} - \frac{a^2}{144} = \frac{12-a^2}{144}$$

$$3. \overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{2} + \frac{\hat{a}}{12} \text{ et } T_n = \hat{a} = 12 \left(\overline{X}_n - \frac{1}{2} \right)$$

$E(T_n) = a$, estimateur sans biais.

$$V(T_n) = 144 \times \frac{1}{n^2} \times n V(X) \text{ et } V(T_n) = \frac{12 - a^2}{n}$$

4. Avec l'estimateur de la question 3. : $a_{\text{moy}} \approx 0,59$.

Nous avons une relation linéaire entre l'estimateur de a et celui de la moyenne, ainsi, comme la moyenne, l'estimateur de a obéit au théorème central limite et dans le cas des grands nombres : $a = a_{\text{moy}} \pm t.s/\sqrt{n}$. Or ici $n=9$, l'échantillon est trop petit pour cette méthode. Nous ne pouvons plus utiliser la loi de Student car la loi de la population est linéaire et non gaussienne. Nous pourrions faire un calcul d'intégrales ou une simulation numérique.

E7 : Estimateurs pour la loi exponentielle énoncé p147

$$1. E(T_n) = \frac{n^n \lambda^n}{(n-1)!} \int_0^{+\infty} \frac{e^{-\frac{n\lambda}{x}}}{x^n} dx, \text{ nous posons : } u = \frac{n\lambda}{x} \text{ et}$$

$$E(T_n) = \frac{n\lambda}{(n-1)!} \int_0^{+\infty} u^{n-2} e^{-u} dx = \frac{n\lambda}{(n-1)!} I_n. \text{ Par intégration par}$$

$$\text{parties : } I_n = (n-2)I_{n-1} \text{ puis } I_n = (n-2)! \text{ car } I_3 = 1.$$

$$\text{Ainsi : } E(T_n) = \frac{n}{n-1} \lambda \text{ et } b_{T_n}(\lambda) = \frac{\lambda}{n-1}.$$

$$E(T_n^2) = \frac{n^2 \lambda^2}{(n-1)!} I_{n-1} = \frac{n^2 \lambda^2}{(n-1)(n-2)}, \quad V(T_n) = \frac{n^2 \lambda^2}{(n-2)(n-1)^2}$$

$$\text{et } r_{T_n}(\lambda) = \lambda^2 \frac{(n+2)}{(n-2)(n-1)}.$$

$$2. E(W_n) = \lambda \text{ et } V(W_n) = r_{W_n}(\lambda) = \frac{\lambda^2}{n-2} \quad (\text{en utilisant les propriétés de l'espérance et de la variance}).$$

3. Nous préférions W_n pour estimer λ car contrairement à T_n son biais est nul et de plus son risque est plus faible.

E8 : Désintégrations énoncé p148

1. Effectuons un faux raisonnement :

Moyenne des dix durées de vie mesurées : $m \approx 2,60 \mu s$.

Écart-type de l'échantillon : $s \approx 2,24 \mu s$.

Coefficient de Student : $t \approx 1,83$ (9 degrés de liberté et 90%).

Théorème central limite : $m = 2,6 \pm 1,3 \mu s$ à 90%.

Ce résultat constitue une bonne estimation pour la moyenne (biais nul, risque contrôlé).

Pour λ , l'inverse de la moyenne, nous utilisons maintenant la formule de propagation des incertitudes : $\lambda = 1/m$ d'où $\Delta \lambda = \Delta m / m^2$

Soit : $\lambda = 0,39 \pm 0,19 / \mu s$ à 90%.

Commentaires :

1. Ici le théorème centrale limite n'est pas applicable car nous avons clairement un petit échantillon ($n=10$). De plus, bien que la loi de Student peut s'utiliser pour les petits nombres, elle ne s'utilise que quand la population est normale, ce qui n'est absolument pas le cas ici car elle est de loi exponentielle. Si nous avions pris t_∞ c'était aussi faux car n petit. On peut imaginer qu'élargir avec la Student permettrait de contrebalancer le petit effectif, mais c'est du bidouillage et c'est faux !

2. La formule de propagation des incertitudes est une approximation, elle est exacte si toutes les distributions de probabilités ont la même forme, ici ce n'est pas le cas. Néanmoins l'approximation est correcte et le calcul est rapide. Mais rien nous assure que nous sommes à bien à 90%, c'est peut-être un peu moins ou un peu plus. Nous ne contrôlons pas ici précisément notre estimation.

2. En appliquant une fonction linéaire et d'après l'exercice page 147 :

$$f_{W_n}(x) = \frac{n}{n-1} f_{T_n}\left(\frac{nx}{n-1}\right) = \frac{(n-1)^{n-1} \lambda^n}{(n-2)! x^{n+1}} e^{-\frac{(n-1)\lambda}{x}}$$

Nous intégrons ensuite sur les 5 premiers centiles puis le 95 premiers centiles pour obtenir les limites de l'intervalle de confiance :

$$\int_0^{\lambda_{\min}} f_{W_{10}}(x) dx = 0,05 \quad \text{et} \quad \int_0^{\lambda_{\max}} f_{W_{10}}(x) dx = 0,95$$

Par calcul numérique des intégrales : $\lambda_{\min} \approx 0,221$ et $\lambda_{\max} \approx 0,639$.

Et finalement : $\lambda = 0,385^{+0,254}_{-0,164} / \mu s$ à 90%.

3. Nous générerons 10 lois exponentielles X_i identiques et indépendantes à partir de dix lois uniformes continues $U_i(0,1)$ indépendantes :

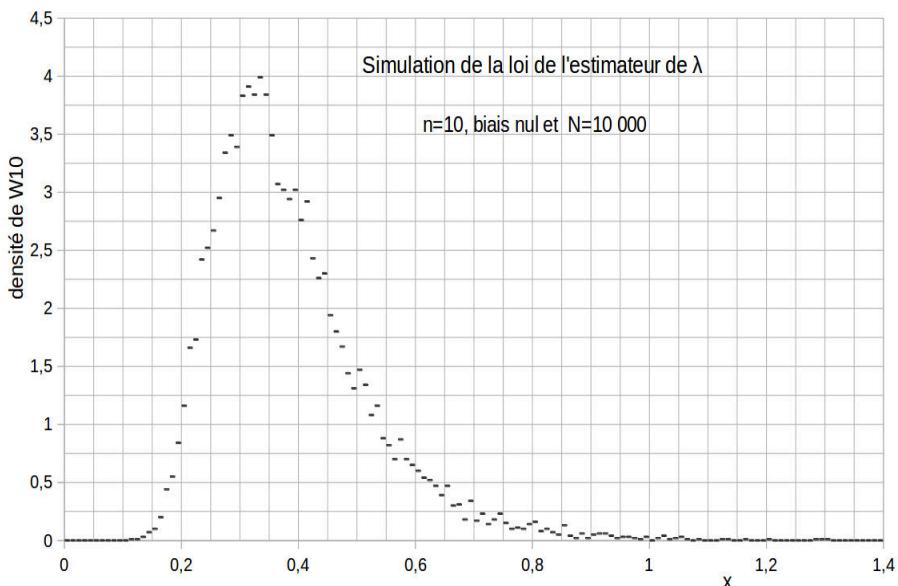
$$X_i = -\ln(U_i) / \lambda$$

Nous prenons pour λ l'estimation ponctuelle obtenue avec l'échantillon.

Ici $n=10$, d'où : $W_{10}=9/(X_1+X_2+\dots+X_{10})$. L'ordinateur réalise ce calcul $N=10000$ fois et les 10000 résultats pour W_{10} sont triés par classe et placés sur un graphe pour créer la distribution d'échantillonnage de W_{10} . Avec la fonction de répartition discrétisée nous évaluons les positions des 5em et 95em centiles : $x_{5\%} \approx 0,22$ et $x_{95\%} \approx 0,64$.

D'où : $\lambda = 0,39^{+0,25}_{-0,17} / \mu \text{ à } 90\%$.

Résultat qui corrobore pleinement le calcul exact.



Fichier créé avec le tableur de LibreOffice :
www.incertitudes.fr/livre/Expos.ods

Chapitre V : Approches multicritères

E1 : Fournisseurs énoncé p171

Nous avons déjà des notes, il n'est pas ici utile de normaliser, surtout qu'il n'y a que deux individus.

Fournisseur 1 : $4/10 \times 0,5 + 8/10 \times 0,3 + 5/10 \times 0,2 = 5,4/10$.

Fournisseur 2 : $6/10 \times 0,5 + 3/10 \times 0,3 + 6/10 \times 0,2 = 5,1/10$.

Choix du premier fournisseur.

E2 : Principe de l'ACP énoncé p171

$$\text{Cas 1 : } M = \begin{pmatrix} -2 & 4 & 0 & 8 \\ 0 & 3 & -3 & 3 \\ 0 & 3 & 3 & 13 \\ 2 & 2 & 0 & 8 \end{pmatrix} \quad m = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 \end{pmatrix}$$

$$r = \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

$$\det(r - \lambda I) = \det \begin{pmatrix} 1-\lambda & -1 & 0 & 0 \\ -1 & 1-\lambda & 0 & 0 \\ 0 & 0 & 1-\lambda & 1 \\ 0 & 0 & 1 & 1-\lambda \end{pmatrix}$$

$$\text{soit } \det(r - \lambda I) = \det \begin{pmatrix} 1-\lambda & -1 \\ -1 & 1-\lambda \end{pmatrix} \det \begin{pmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{pmatrix}$$

$$\text{d'où } ((1-\lambda)^2 - 1)(1-\lambda)((1-\lambda)^2 - 1)(1-\lambda) = 0 \quad \text{et} \quad \lambda_1 = \lambda_2 = 2$$

$$\lambda_3 = \lambda_4 = 0 \quad \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_1 - x_2 \\ -x_1 + x_2 \\ x_3 + x_4 \\ x_3 + x_4 \end{pmatrix} = \lambda_j \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \quad \text{d'où}$$

$$\vec{e}'_1 = (\vec{e}_1 - \vec{e}_2)/\sqrt{2} \quad , \quad \vec{e}'_2 = (\vec{e}_3 + \vec{e}_4)/\sqrt{2} \quad , \quad \vec{e}'_3 = (\vec{e}_1 + \vec{e}_2)/\sqrt{2} \quad \text{et}$$

$$\vec{e}'_4 = (\vec{e}_3 - \vec{e}_4)/\sqrt{2} \quad . \quad \text{Puis avec } P: \quad m' = \begin{pmatrix} -\sqrt{2} & 0 & 0 & 0 \\ 0 & -\sqrt{2} & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \\ \sqrt{2} & 0 & 0 & 0 \end{pmatrix}$$

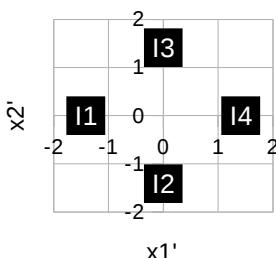
$$\vec{I}'_1 = \begin{pmatrix} -\sqrt{2} \\ 0 \end{pmatrix} \quad , \quad \vec{I}'_2 = \begin{pmatrix} 0 \\ -\sqrt{2} \end{pmatrix} \quad , \quad \vec{I}'_3 = \begin{pmatrix} 0 \\ \sqrt{2} \end{pmatrix} \quad \text{et} \quad \vec{I}'_4 = \begin{pmatrix} \sqrt{2} \\ 0 \end{pmatrix} \quad .$$

Pour le cercle des corrélations, nous calculons les corrélations entre les

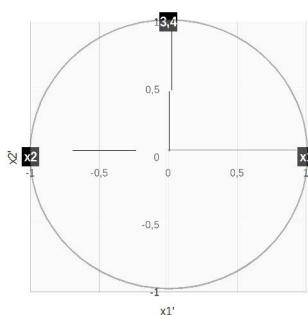
nouvelles et les anciennes valeurs $\begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$

(nous l'obtenons en calculant les corrélations entre les colonnes de m et m').

Individus selon X_1' et X_2' :



Cercle des corrélations :



Cas 2 : $\bar{x}_1=2$, $\bar{x}_2=2$ et $\bar{x}_3=2$.

$$M = \begin{pmatrix} 1 & 3 & 2 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \end{pmatrix} \quad m = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix} .$$

Matrice des corrélations : $r = \begin{pmatrix} 1 & -0,5 & -0,5 \\ -0,5 & 1 & -0,5 \\ -0,5 & -0,5 & 1 \end{pmatrix} .$

Valeurs propres : $\lambda_1=\lambda_2=1,5$ et $\lambda_3=0$.

100% de l'information sera représentée par X'_1 et X'_2 .

Vecteurs propres : $\vec{e}'_1 \approx -0,118\vec{e}_1 - 0,648\vec{e}_2 + 0,766\vec{e}_3$,
 $\vec{e}'_2 \approx 0,816\vec{e}_1 - 0,408\vec{e}_2 - 0,408\vec{e}_3$ et $\vec{e}'_3 \approx -0,577(\vec{e}_1 + \vec{e}_2 + \vec{e}_3)$.

Matrice de passage : $P \approx \begin{pmatrix} -0,118 & 0,816 & -0,577 \\ -0,648 & -0,408 & -0,577 \\ 0,766 & -0,408 & -0,577 \end{pmatrix}$ et

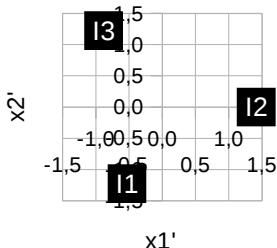
$$m' \approx \begin{pmatrix} -0,530 & -1,225 & 0 \\ 1,414 & 0 & 0 \\ -0,884 & 1,225 & 0 \end{pmatrix} .$$

$$\vec{I}'_1 \approx \begin{pmatrix} -0,530 \\ -1,225 \end{pmatrix} , \quad \vec{I}'_2 \approx \begin{pmatrix} 1,414 \\ 0 \end{pmatrix} \text{ et } \vec{I}'_3 \approx \begin{pmatrix} -0,884 \\ 1,225 \end{pmatrix} .$$

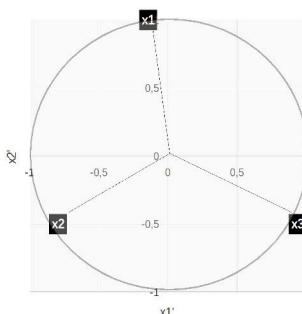
Corrélations entre nouvelles et anciennes caractéristiques :

$$\begin{pmatrix} -0,144 & 1 & 0 \\ -0,794 & -0,5 & 0 \\ 0,938 & -0,5 & 0 \end{pmatrix}$$

Individus selon X_1' et X_2' :



Cercle des corrélations :



Devoir Suricate : énoncé p202

E1 : Pièces

Nb de lancers possibles : $2 \times 2 \times 2 \times 2 = 16$.

Une seule possibilité correspond à 4 piles d'où une probabilité de **une chance sur 16** soit 6,25%.

Quatre possibilités pour un pile : **PFFF, FPFF, FFPF, FFFP**. D'où $p=4/16$ soit **une chance sur quatre** ($p=0,25=25\%$).

E2 : Eaux minérales (bicarbonates et sulfates)

a) **$\bar{B}=1234,8 \text{ mg/L}$** .

b) **$\sigma_B=1469 \text{ mg/L}$** .

c) $\Delta B = t_{\alpha/2} \cdot \sigma_B / \sqrt{n}$. $t_{\alpha/2}(df=8; 95\%) = 2,31$. **$\Delta B=1131 \text{ mg/L}$** . Grande dispersion, la teneur en bicarbonate est très variable d'une eau à l'autre.

d) $r_{BS} = -0,265 < 0,52$. Les grandeurs peuvent être considérées comme

non corrélées.

e) D'après la formule de propagation des incertitudes : si $T=B+S$ on a $\Delta T=\sqrt{(\Delta B^2+\Delta S^2)}$.

D'où $\Delta T=1209,5 \text{ mg/L}$, $\bar{T}=\bar{B}+\bar{S}$, $T=1612 \pm 1209 \text{ mg/L à 95\%}$, très dispersé.

E3 : Population (pauvres / non-pauvres)

Valeurs observées :

$O_{11}=112$	$O_{12}=105$	$O_{13}=154$	$O_{14}=113$
$O_{21}=888$	895	...	
1000	1000	1000	1000

Pauvres sur le pays : $112+105+154+113=484$ sur 4000 personnes au total, d'où 12,1% de pauvres. Nous en déduisons le tableau espéré pour l'hypothèse de répartition uniforme:

$E_{11}=121$	$E_{12}=121$	$E_{13}=121$	$E_{14}=121$
$E_{21}=879$	$E_{22}=879$	$E_{23}=879$	$E_{24}=879$
1000	1000	1000	1000

$$\chi^2 = \frac{(112-121)^2}{121} + \frac{(105-121)^2}{121} + \frac{(154-121)^2}{121} + \frac{(113-121)^2}{121} \approx 14,0$$

Le nombre de degrés de liberté : $ddl=(4-1)(2-1)$ d'où $\chi^2 = 7,815$ au maximum pour que la probabilité que l'hypothèse soit rejetée alors quelle est vraie soit inférieure à 5%. Ici on est bien au dessus, **l'hypothèse est donc rejetée** (pour être plus précis, la probabilité que l'hypothèse soit rejetée alors quelle est vraie n'est que de 0,29%).

E4 : Eaux minérales (sodium et chlorures) 2)

E5 : Eaux minérales (analyse en composantes principales)

$$\tilde{x}_7^2 (x_7^2 - \bar{x}_2)/\sigma_2 \simeq (71-1234,8)/1469 \simeq -0,792.$$

Nous avons réalisé une renormalisation qui permet de comparer à égalité les différentes caractéristiques (grandeur sans dimension, même centre et même dispersion).

$r_{25}=-0,266$ (corrélation entre sulfates et bicarbonates).

Ensuite la matrice est symétrique : $r_{ij}=r_{ji}$. Et $r_{ii}=1$.

$\sum_{j=1}^8 \lambda_j = 8 = \text{tr}(r)$. Le pourcentage d'information contenu dans les caractéristiques X'_5 , X'_6 , X'_7 et X'_8 est $\%_{5,6,7,8} = \frac{8-\lambda_1-\lambda_2-\lambda_3-\lambda_4}{8} = \frac{8-7,96}{8} = 0,5\%$ d'information perdue si on les ignore.

Pour obtenir les valeurs propres on résout $\det(r - \lambda I) = 0$.

Les colonnes de la matrice de passage sont les vecteurs propres déterminés en résolvant :

$$r \vec{e}_j = \lambda_j \vec{e}_j$$

$$m' = m P \cdot x'_2 = -0,154x(-0,305) + 0,044x0,422 + \dots = -0,202.$$

X'_1 et X'_2 sont les deux premières composantes principales. Elles contiennent $\%_{1,2} = (4,76+2,05)/8 = 85,1\%$ de l'information du système.

Depuis le cercle des corrélations nous constatons deux groupes d'ions très corrélés :

X'_1 représente principalement X_2 , X_4 , X_6 et X_7 (Bicarbonates / Cl et K / Na -ions alcalins).

X'_2 représente principalement X_1 , X_3 et X_5 (Ca / Mg -dureté de l'eau - ions alcalino-terreux- et Sulfates).

X_8 se rapproche du centre du cercle des corrélations, ainsi il est peu corrélé à X_1' et X_2' , et joue un rôle faible.

Aussi X_3' est très décorrélé des autres caractéristiques sauf de X_8 (le pH). Comportement intéressant, le pH évolue à part et pourrait être sortit de l'étude. Une interprétation précise peut être faite par un spécialiste en eaux minérales.

Deux remarques :

- $\text{pH} = -\log [\text{H}^+]$, donc le pH n'a pas une relation linéaire avec $[\text{H}^+]$. Il serait donc mieux adapté de prendre $[\text{H}^+]$ et non le pH pour X_8 . Toutes les autres grandeurs étant aussi des ions. Que donnerait l'ACP avec $[\text{H}^+]$ et aurions-nous le même résultat (X_8 décorrélée des autres grandeurs) ? L'étude faite, nous constatons cette déconnexion du pH, ou de H^+ , confirmée.
- la minéralisation (résidu sec) manque dans l'étude. En effet en renormalisant les caractéristiques la quantité totale en ions a disparue. La minéralisation de l'eau correspond globalement à la somme de toutes les concentrations. Après étude, en remplaçant le pH par la minéralisation, nous constatons que cette nouvelle grandeur joue un rôle important et que l'analyse est à reprendre dans son ensemble (par exemple, dans ce cas, les deux composantes principales contiennent à elles seules 96% de l'information et la minéralité est très corrélée aux bicarbonates, chlorures, potassium et sodium).

Devoir Narval : énoncé p209

E1 : Dès

36 possibilités. 12 : $p=1/36=2,8\%$. 7 : $p=6/36=1/6=16,7\%$.

E2 : Maisons

- a) b) c) d) $\bar{R}_1=40 \text{ mK/W}$; $\sigma_1=1,85 \text{ mK/W}$; $\Delta R_1=1,55 \text{ mK/W}$; $r_{12}=0$
- e) $R=28,6 \pm 0,8 \text{ mK/W}$

E3 : Salaires

$\chi^2=86,5 >> 7,8$ hypothèse non soutenable.

E4 : Courbes

- 1) d); 3) c) ; 4) b); 2) aucunes correspondances.

E5 : Isolants

$\bar{x}_1=-0,625$; $\sigma_1=1,06$; $x_{11}=-0,354$; $r_{12}=0,92$; $\lambda_1=3,0081$

sans X'_3 et X'_4 , 3,4% d'information perdue.

XI. Bibliographie / Sources / Logiciels / Illustrations

Livres

[iv] PROTASSOV Konstantin. *Probabilités et incertitudes dans l'analyse des données expérimentales.* Presses Universitaires de Grenoble, 1999. 128 p.

[vi] WONNACOTT. *Statistique Économie - Gestion - Sciences - Médecine.* Economica, 1972. 919 p.

[x] JOURNEAUX Roger. *Traitemet des mesures.* Ellipses, 2009. 377 p.

[iii] SAPORTA Gilbert. *Probabilités, analyse des données et statistique.* Technip, 2006. 622 p.

SHELDON M. ROSS. *Initiation aux probabilités.* Presses polytechniques et universitaires romandes, 2006. 585 p.

RONDY SYLVAIN. *Maths - Économique et Commerciale option Scientifique - 2^e année.* Collection Prépas Sciences, ellipse, 2014. 810 p.

ROUAUD Mathieu *Probabilités, statistiques et analyses multicritères.*, www.incertitudes.fr/proba-stat-acp/livre.pdf, 2012. 290 p, lulu.com.

Web

Placez *http://* devant le nom des sites. La plupart des fichiers sont en copie dans le dossier <www.incertitudes.fr/livre/>.

ROUAUD Mathieu. *Calcul d'incertitudes.* <www.incertitudes.fr/Incertitudes.html>

Calcul d'intégrales en ligne : <www.integral-calculator.com>

[viii] MARCADET Françoise. *Estimer une incertitude et Mesures et incertitudes (accompagnement personnalisé en 1^{ère}S)*. Professeur de Sciences Physiques appartenant au groupe académique. Lycée Alain Fournier de Bourges. Sur le site <physique.ac-orleans-tours.fr/accompagnement_personnalise/>

[v] *Nombres, mesures et incertitudes*. mai 2010.
<www.educnet.education.fr/rnchimie/recom/mesures_incertitudes.pdf>

[vii] *Évaluation des données de mesure — Guide pour l'expression de l'incertitude de mesure*. <www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_F.pdf>, 2008. 132p.

BREUIL P, DI BENEDETTO D. *Incertitude et étalonnage*. 2000. 16p.

[ix] *Algorithme de Gauss-Newton* - Wikipédia
15 juin 2013 à 17:32

Articles

[i] *Regression line analysis* K.S.KRANE et L.SCHECTER
American J. of Physics **50**,82 (1982).

[ii] ROUAUD Mathieu. *Mesure avec une règle*. Bulletin de l'Union des Physiciens n°913, avril 2009.

Logiciels

Tous les logiciels utilisés sont libres et gratuits.

- Traitement de texte et tableur *OpenOffice* et *LibreOffice*
- Graphisme : *Gimp* (points), *Inkscape* (vectoriel) et *Blender* (3D).
- Calcul : *Scilab* (numérique), *XCAS* (symbolique) et programmes en *PHP* (sur serveur).
- Tracé de courbes : *KmPlot*, *TeXgraph*.
- Système d'exploitation : *Ubuntu*.

Illustrations

Représentations 3D situées avant la table des matières et en troisième vignette de la couverture arrière.

Un marcheur aléatoire se déplace dans un plan. Il lance deux pièces et regarde les résultats : deux piles, un pile et un face, un face et un pile ou deux faces. La première pièce lui indique s'il doit faire un premier vers l'Est ou l'Ouest, la deuxième s'il doit faire un second pas vers le Sud ou le Nord. Ainsi à chaque intervalle de temps Δt il se déplace dans le plan de deux pas. Pendant Δt la distance parcourue par le marcheur est $\Delta d = 2 p$ (p longueur d'un pas).

À quelle distance du point de départ se trouve
le marcheur à l'instant t ? (après n intervalles de temps : $t = n \Delta t$)

Pour $n=1$, nous traçons les tableaux suivants :

0,5	1	1
-0,5	1	1
-0,5	0,5	

25%	25%
25%	25%

Le centre du tableau est son point de départ. En abscisse x (direction Est-Ouest) le déplacement est de plus ou moins un pas ($x=\pm p$, $p=\Delta d / 2$ et on a pris $\Delta d=1$). De même en ordonnée : $y=\pm p$. Dans chaque est indiqué le nombre de possibilités de se retrouver à cet endroit. Le second tableau indique les probabilités ($1/4$, $4=2 \times 2$).

Dans les quatre possibilités il est à une distance $\sqrt{2}/2$ du point d'origine. Et, en terme d'écart-type, la distance caractéristique est $s=\sqrt{(4.1/\sqrt{2})^2/(4-1)}\approx 0,816$.

Pour $n=2$, nous traçons les tableaux suivants :

1	1	2	1
0	2	4	2
-1	1	2	1
-1	0	1	

6%	13%	6%
13%	25%	13%
6%	13%	6%

Par exemple pour être en ($x=0$; $y=-1$), il y a deux chemins possibles : (PF,PP) et (FP,PP). Écart-type $s_2 \approx 1,033$.

Le marcheur a une chance sur quatre d'être revenu au point de départ.

Pour $n=3$, nous traçons les tableaux suivants :

1,5	1	3	3	1
0,5	3	9	9	3
-0,5	3	9	9	3
-1,5	1	3	3	1

-1,5 -0,5 0,5 1,5

2%	5%	5%	2%
5%	14%	14%	5%
5%	14%	14%	5%
2%	5%	5%	2%

$$s_3 \approx 1,234$$

Pour $n=4$, nous traçons les tableaux suivants :

2	1	4	6	4	1
1	4	16	24	16	4
0	6	24	36	24	6
-1	4	16	24	16	4
-2	1	4	6	4	1

-2 -1 0 1 2

0,4%	1,6%	2,3%	1,6%	0,4%
1,6%	6,3%	9,4%	6,3%	1,6%
2,3%	9,4%	14%	9,4%	2,3%
1,6%	6,3%	9,4%	6,3%	1,6%
0,4%	1,6%	2,3%	1,6%	0,4%

Pour $n=5$, nous traçons les tableaux suivants :

2,5	1	5	10	10	5	1
1,5	5	25	50	50	25	5
0,5	10	50	100	100	50	10
-0,5	10	50	100	100	50	10
-1,5	5	25	50	50	25	5
-2,5	1	5	10	10	5	1

-2,5 -1,5 -0,5 0,5 1,5 2,5

0,1%	0,5%	1,0%	1,0%	0,5%	0,1%
0,5%	2,4%	4,9%	4,9%	2,4%	0,5%
1,0%	4,9%	9,8%	9,8%	4,9%	1,0%
1,0%	4,9%	9,8%	9,8%	4,9%	1,0%
0,5%	2,4%	4,9%	4,9%	2,4%	0,5%
0,1%	0,5%	1,0%	1,0%	0,5%	0,1%

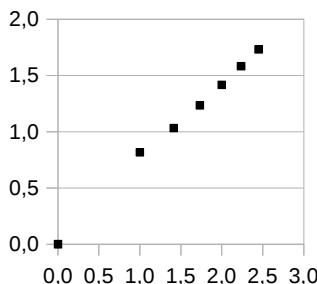
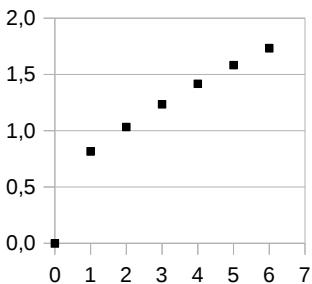
Pour $n=6$, nous traçons les tableaux suivants :

3	1	6	15	20	15	6	1	0,0%	0,1%	0,4%	0,5%	0,4%	0,1%	0,0%
2	6	36	90	120	90	36	6	0,1%	0,9%	2,2%	2,9%	2,2%	0,9%	0,1%
1	15	90	225	300	225	90	15	0,4%	2,2%	5,5%	7,3%	5,5%	2,2%	0,4%
0	20	120	300	400	300	120	20	0,5%	2,9%	7,3%	9,8%	7,3%	2,9%	0,5%
-1	15	90	225	300	225	90	15	0,4%	2,2%	5,5%	7,3%	5,5%	2,2%	0,4%
-2	6	36	90	120	90	36	6	0,1%	0,9%	2,2%	2,9%	2,2%	0,9%	0,1%
-3	1	6	15	20	15	6	1	0,0%	0,1%	0,4%	0,5%	0,4%	0,1%	0,0%
	-3	-2	-1	0	1	2	3							

D'où l'évolution de la distance quadratique moyenne du point d'origine en fonction du temps :

t (Δt)	0	1	2	3	4	5	6
\sqrt{t}	0,00	1,00	1,41	1,73	2,00	2,24	2,45
$s(2p)$	0,000	0,816	1,033	1,234	1,417	1,582	1,732

Nous traçons la courbe les points pour s en fonction de t , puis celle pour s en fonction de \sqrt{t} :



Nous constatons une bien meilleur corrélation en \sqrt{t} . En effet nous voyions directement que la distance à l'origine n'évoluait pas proportionnellement au temps, pour $n=2$ nous sommes à environ une unité de l'origine, nous devrions donc être vers 3 pour $n=6$.

Cette variation en \sqrt{t} est caractéristique des phénomènes de diffusion et trouve ici son analogie avec les compensations d'erreurs en \sqrt{n} .

XII. TABLES / Index

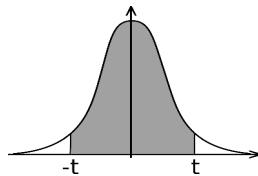
$$A. \text{ Loi normale centrée réduite}$$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

$$F(z) = \int_{-\infty}^z f(z) dz = P(Z \leq z) \quad P(Z > z) = 1 - P(Z \leq z)$$

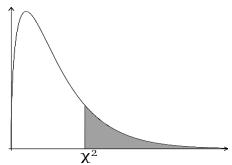
$$P(Z \leq -z) = P(Z > z) \quad \text{Exemple : } P(Z \leq 1,67) \simeq 0,95254$$

z	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
4,0	0,99997	0,99997	0,99997	0,99997	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998



B. Coefficients de Student

Degrés de liberté (taille de l'échantillon moins le nombre de paramètres)	Coefficient de Student t	Confiance (%)								
		50	80	90	95	98	99	99,5	99,8	99,9
	1	1,00	3,08	6,31	12,7	31,8	63,7	127	318	637
	2	0,82	1,89	2,92	4,30	6,96	9,92	14,1	22,3	31,6
	3	0,76	1,64	2,35	3,18	4,54	5,84	7,45	10,2	12,9
	4	0,74	1,53	2,13	2,78	3,75	4,60	5,60	7,17	8,61
	5	0,73	1,48	2,02	2,57	3,36	4,03	4,77	5,89	6,87
	6	0,72	1,44	1,94	2,45	3,14	3,71	4,32	5,21	5,96
	7	0,71	1,41	1,89	2,36	3,00	3,50	4,03	4,79	5,41
	8	0,71	1,40	1,86	2,31	2,90	3,36	3,83	4,50	5,04
	9	0,70	1,38	1,83	2,26	2,82	3,25	3,69	4,30	4,78
	10	0,70	1,37	1,81	2,23	2,76	3,17	3,58	4,14	4,59
	11	0,70	1,36	1,80	2,20	2,72	3,11	3,50	4,02	4,44
	12	0,70	1,36	1,78	2,18	2,68	3,05	3,43	3,93	4,32
	13	0,69	1,35	1,77	2,16	2,65	3,01	3,37	3,85	4,22
	14	0,69	1,35	1,76	2,14	2,62	2,98	3,33	3,79	4,14
	15	0,69	1,34	1,75	2,13	2,60	2,95	3,29	3,73	4,07
	16	0,69	1,34	1,75	2,12	2,58	2,92	3,25	3,69	4,01
	17	0,69	1,33	1,74	2,11	2,57	2,90	3,22	3,65	3,97
	18	0,69	1,33	1,73	2,10	2,55	2,88	3,20	3,61	3,92
	19	0,69	1,33	1,73	2,09	2,54	2,86	3,17	3,58	3,88
	20	0,69	1,33	1,72	2,09	2,53	2,85	3,15	3,55	3,85
	22	0,69	1,32	1,72	2,07	2,51	2,82	3,12	3,50	3,79
	24	0,68	1,32	1,71	2,06	2,49	2,80	3,09	3,47	3,75
	26	0,68	1,31	1,71	2,06	2,48	2,78	3,07	3,43	3,71
	28	0,68	1,31	1,70	2,05	2,47	2,76	3,05	3,41	3,67
	30	0,68	1,31	1,70	2,04	2,46	2,75	3,03	3,39	3,65
	40	0,68	1,30	1,68	2,02	2,42	2,70	2,97	3,31	3,55
	50	0,68	1,30	1,68	2,01	2,40	2,68	2,94	3,26	3,50
	60	0,68	1,30	1,67	2,00	2,39	2,66	2,91	3,23	3,46
	70	0,68	1,29	1,67	1,99	2,38	2,65	2,90	3,21	3,44
	80	0,68	1,29	1,66	1,99	2,37	2,64	2,89	3,20	3,42
	90	0,68	1,29	1,66	1,99	2,37	2,63	2,88	3,18	3,40
	100	0,68	1,29	1,66	1,98	2,36	2,63	2,87	3,17	3,39
	200	0,68	1,29	1,65	1,97	2,35	2,60	2,84	3,13	3,34
	300	0,68	1,28	1,65	1,97	2,34	2,59	2,83	3,12	3,32
	500	0,67	1,28	1,65	1,96	2,33	2,59	2,82	3,11	3,31
	1000	0,67	1,28	1,65	1,96	2,33	2,58	2,81	3,10	3,30
	∞	0,67	1,28	1,64	1,96	2,33	2,58	2,81	3,09	3,29



C. Valeurs critiques de Khi-deux

χ^2	Probabilité α de rejeter l'hypothèse alors qu'elle est vraie (%)										
	99	90	75	50	30	20	10	5	2	1	0,1
1	0,000157	0,0158	0,102	0,455	1,07	1,64	2,71	3,84	5,41	6,63	10,8
2	0,0201	0,211	0,575	1,39	2,41	3,22	4,61	5,99	7,82	9,21	13,8
3	0,115	0,584	1,21	2,37	3,66	4,64	6,25	7,81	9,84	11,3	16,3
4	0,297	1,06	1,92	3,36	4,88	5,99	7,78	9,49	11,7	13,3	18,5
5	0,554	1,61	2,67	4,35	6,06	7,29	9,24	11,1	13,4	15,1	20,5
6	0,872	2,20	3,45	5,35	7,23	8,56	10,6	12,6	15,0	16,8	22,5
7	1,24	2,83	4,25	6,35	8,38	9,80	12,0	14,1	16,6	18,5	24,3
8	1,65	3,49	5,07	7,34	9,52	11,0	13,4	15,5	18,2	20,1	26,1
9	2,09	4,17	5,90	8,34	10,7	12,2	14,7	16,9	19,7	21,7	27,9
10	2,56	4,87	6,74	9,34	11,8	13,4	16,0	18,3	21,2	23,2	29,6
11	3,05	5,58	7,58	10,3	12,9	14,6	17,3	19,7	22,6	24,7	31,3
12	3,57	6,30	8,44	11,3	14,0	15,8	18,5	21,0	24,1	26,2	32,9
13	4,11	7,04	9,30	12,3	15,1	17,0	19,8	22,4	25,5	27,7	34,5
14	4,66	7,79	10,2	13,3	16,2	18,2	21,1	23,7	26,9	29,1	36,1
15	5,23	8,55	11,0	14,3	17,3	19,3	22,3	25,0	28,3	30,6	37,7
16	5,81	9,31	11,9	15,3	18,4	20,5	23,5	26,3	29,6	32,0	39,3
17	6,41	10,1	12,8	16,3	19,5	21,6	24,8	27,6	31,0	33,4	40,8
18	7,01	10,9	13,7	17,3	20,6	22,8	26,0	28,9	32,3	34,8	42,3
19	7,63	11,7	14,6	18,3	21,7	23,9	27,2	30,1	33,7	36,2	43,8
20	8,26	12,4	15,5	19,3	22,8	25,0	28,4	31,4	35,0	37,6	45,3
21	8,90	13,2	16,3	20,3	23,9	26,2	29,6	32,7	36,3	38,9	46,8
22	9,54	14,0	17,2	21,3	24,9	27,3	30,8	33,9	37,7	40,3	48,3
23	10,2	14,8	18,1	22,3	26,0	28,4	32,0	35,2	39,0	41,6	49,7
24	10,9	15,7	19,0	23,3	27,1	29,6	33,2	36,4	40,3	43,0	51,2
25	11,5	16,5	19,9	24,3	28,2	30,7	34,4	37,7	41,6	44,3	52,6
26	12,2	17,3	20,8	25,3	29,2	31,8	35,6	38,9	42,9	45,6	54,1
27	12,9	18,1	21,7	26,3	30,3	32,9	36,7	40,1	44,1	47,0	55,5
28	13,6	18,9	22,7	27,3	31,4	34,0	37,9	41,3	45,4	48,3	56,9
29	14,3	19,8	23,6	28,3	32,5	35,1	39,1	42,6	46,7	49,6	58,3
30	15,0	20,6	24,5	29,3	33,5	36,3	40,3	43,8	48,0	50,9	59,7
31	15,7	21,4	25,4	30,3	34,6	37,4	41,4	45,0	49,2	52,2	61,1
32	16,4	22,3	26,3	31,3	35,7	38,5	42,6	46,2	50,5	53,5	62,5
33	17,1	23,1	27,2	32,3	36,7	39,6	43,7	47,4	51,7	54,8	63,9
34	17,8	24,0	28,1	33,3	37,8	40,7	44,9	48,6	53,0	56,1	65,2
35	18,5	24,8	29,1	34,3	38,9	41,8	46,1	49,8	54,2	57,3	66,6
36	19,2	25,6	30,0	35,3	39,9	42,9	47,2	51,0	55,5	58,6	68,0
37	20,0	26,5	30,9	36,3	41,0	44,0	48,4	52,2	56,7	59,9	69,3
38	20,7	27,3	31,8	37,3	42,0	45,1	49,5	53,4	58,0	61,2	70,7
39	21,4	28,2	32,7	38,3	43,1	46,2	50,7	54,6	59,2	62,4	72,1
40	22,2	29,1	33,7	39,3	44,2	47,3	51,8	55,8	60,4	63,7	73,4

nombre de degrés de liberté

Index

- Analyse en composantes principales.....150
Approches multicritères..150
Asymptotes.....95
Biais.....124
Calcul d'incertitude.....53
Capacité thermique.....1
Cercle des corrélations....164
Changement de variable. 231
Chi 2.....76
Classe.....**6**, 38
Clous.....99
Coefficient d'aplatissement101, 231
Coefficient de corrélation..47
Coefficient de dissymétrie101, 231
Coefficient de Student **12**, 60, 309
Coefficients binomiaux...102
Compensations.....54, 182
Conductivité thermique....90
Confiance.....12
Décomposition en gaussiennes.....99
Densité de probabilité.....**19**, 176
Densité linéaire.....147
Dérivée partielle.....59, 229
Dés.....16, 38, 179
Désintégrations.....148
Détente de Joule Gay-Lussac194
Développement limité....81, **230**
Développement limité,...105
Diffusion.....307
Dispersion.....191
Distribution
d'échantillonnage.....10, 24
Distribution de Gauss. **10**, **19**
Droites extrêmes.....61
Droites extrêmes.....223
Eaux minérales.....202
Écart.....192
Écart moyen.....3, 37
Écart quadratique moyen....4
Écart-type.....3, 191
Équation différentielle.....91
Erreur accidentelle....33, 191
Erreur aléatoire.....185, 191, 192
Erreur de discrétisation...33, 174, 191
Erreur de mesure.....192
Erreur de type A.....191
Erreur de type B.....191

Erreurs systématiques	33, 185, 191, 192	Fréquence.....	6
Espace des individus	155	Gaussienne 3D.....	42
Espérance	20, 191	Gaz parfait.....	65, 195
Estimateur	124	Goniomètre.....	217
Estimation	61	Hypothèse ergodique.....	197
Estimation par intervalle	124, 135	Incertitude.....	13
Estimation ponctuelle	124	Incertitude absolue.....	13, 55
Étendue	3, 176	Incertitude relative.....	13, 55
Étoiles filantes	106	Incertitude-type.....	191
Facteur d'élargissement	191	Indice.....	88, 217
Faisceau homocinétique	144	Inégalité de Bienaymé-Tchebychev.....	135, 138
Fidélité	33	Intégrale.....	20, 230
Fluctuations statistiques	12	Intégrale multiple.....	42
Focométrie	88, 93, 193	Intervalle de confiance.....	12, 61, 95
Fonction de répartition	104, 113	Intervalle de fluctuation.....	12
Fonction Gamma	232	Intervalle de prédiction.....	62, 95
Fonction réciproque	113	Intervalle dissymétrique ..	140
Fonctions de variables à densité	113	Isolant.....	39, 150, 212
Formule de Cauchy	77, 89	Justesse.....	33, 191
Formule de propagation des écarts-types	52, 184	Linéarisation.....	67, 80
Formule de propagation des incertitudes	52, 185	Loi binomiale. 102 , 118, 119	
Formule du binôme de Newton	232	Loi binomiale négative ..	105, 119
Formule du binôme négatif	232	Loi de Bernouilli.....	14

Loi de Poisson	105, 120, 128
Loi de probabilité triangulaire.....	108, 179, 186
Loi de probabilité uniforme	107, 176
Loi de Student.....	110, 120
Loi du Khi-Deux.....	111
Loi du produit.....	121
Loi exponentielle. .	109, 120, 129
Loi géométrique.....	103 , 119
Loi normale.....	110, 130
Loi uniforme.....	120, 130
Lois continues.....	107
Lois discrètes.....	102
Lois du Khi-Deux.....	121
Macroétat.....	196
Matrice de corrélation....	158
Matrice de passage.....	161
Matrice des données.....	154
Matrice des données normalisée.....	155
Matrice réduite.....	167
Médiane.....	2
Mesurage.....	191
Mesurande.....	191
Mesure optique.....	222
Méthode d'inversion.....	116
Méthode de Box-Müller.	117
Méthode des moments....	127
Méthode des petites variations.....	73
Méthode du maximum de vraisemblance.....	131
Méthode itérative..	64, 77, 80
Méthode par pondération	150
Métrologie.....	186
Microétat.....	196
Miroir sphérique.....	221
Mode.....	1
Moindres carrés....	58, 64, 75
Moment.....	101, 231
Montage longue dérivation	255
Moyenne.....	2
Moyenne arithmétique.....	2
Moyenne géométrique.....	2
Nuages de points.....	49
Ordonnée à l'origine.....	59
Parfaite.....	185
Pente.....	59
Physique statistique.....	196
Pile électrique.....	93
Prédiction.....	62
Prisme.....	89, 217
Produit de convolution..	174, 178
Proposition contraposée....	25
Puissance du test.....	29
Régression avec barres d'erreurs.....	64

Régression généralisée.....	75	Somme de binomiales.....	119
Régression linéaire.....	58	Somme de gaussiennes...	120
Régression multiple.....	75	Somme de Students.....	121
Régression non linéaire....	80	Somme de variables aléatoires.....	101
Régression parabolique....	78	Spectre de raie.....	89
Régression polynomiale....	77	Temps d'attente.....	136
Répétabilité.....	33, 191	Test d'hypothèse.....	24
Reproductibilité.....	33, 191	Test du Khi-deux.....	30
Résidu.....	59, 76	Théorème central limite..	10 , 24, 136, 182
Résistance thermique	90, 210	Train.....	136
Résolution.....	33, 174, 191	Urnes.....	146
Risque de deuxième espèce	26	Valeur vraie.....	191
Risque de première espèce	26	Valeurs propres.....	159
Risque quadratique.....	125	Variance...20, 101, 177, 182	
Scilab.....	172	Vecteurs propres.....	159
Séries géométriques dérivées	232	Vernier.....	218
Simulation numérique...116, 140		Vieillissement.....	104, 109
Somme d'exponentielles.	122	Vraisemblance.....	131
		Zéro absolu.....	62

Ce livre s'adresse à tout public intéressé par les sciences expérimentales et les mathématiques. Le champ d'application déborde largement sur des thèmes très présents dans notre société : les sondages, les modélisations économiques et l'aide à la décision. Par exemple l'analyse en composantes principales est un outil très utilisé.

Le fil directeur suit l'enchaînement d'un cours donné en école d'ingénieur, ainsi ce livre pourra être très utile à l'étudiant tout comme à l'enseignant. Aussi, je souhaite que l'expérimentateur trouve tous les outils nécessaires à des calculs d'incertitudes simples et rigoureux.

En effet, la science s'efforce de lier les phénomènes naturels à une logique mathématique. Cette quête d'une vérité et d'une cohérence globale nécessite d'exercer un esprit critique qui s'appuie sur des mesures accompagnées de leurs incertitudes.

3 1115926535897932384626133832795028811971693
99375 105820974944592307816406286208998628034
8253421 1706798214808651328230664709384460955
058223172535940812848 11174502841027019385211

$$\nabla^2 = \sum_{(i,j)} \frac{(O_{ij} - E_{ij})}{E_{ij}} = 0$$

$$\Delta a = t^{n-2} \sum_{j=1}^p \left(\frac{\partial f}{\partial x_j} \right)^2 \Delta x_j^2$$

$$r e_j = \lambda_j e_j$$

$$w_i = \frac{1}{(\Delta y_i)^2 + (a \Delta x_i)^2}$$

$$p(x) = \frac{1}{\sqrt{2\pi \cdot \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\bar{x} = \frac{s}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\mu = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$
