

Analyse de données et exploitation de mesures

Jean-François ROCH
Préparation à l'agrégation de physique – ENS Cachan

January 3, 2005

Contenu

1	Mesurages et statistiques	3
1.1	Rappels sur la distribution normale de Gauss	3
1.1.1	Intérêt de la distribution normale de Gauss	3
1.1.2	Expression mathématique de la loi normale	5
1.1.3	Normalisation et fonction “erreur”	5
1.2	Interprétation statistique d’une opération de mesurage	7
1.2.1	Moyenne et incertitude d’une mesure	7
1.2.2	Propagation des erreurs	8
1.2.3	Erreurs aléatoires et erreurs systématiques dans un mesurage	9
2	Exploitation d’une série de mesure	11
2.1	Estimateurs de la distribution de probabilité associée à la mesure	11
2.1.1	Position du problème et maximum de vraisemblance	11
2.1.2	Estimateur de la moyenne μ	13
2.1.3	Précision sur la détermination de l’estimateur de la moyenne	14
2.1.4	Moyenne d’échantillon et loi des grands nombres	15
2.1.5	Estimateur de la variance de la distribution de probabilité	16
2.1.6	Précision sur la détermination de l’estimateur de la variance	18
2.1.7	Cas d’un petit nombre de mesures : fonction de Student	19
2.2	Exemple concret d’analyse d’une série de mesures	21
2.2.1	Position du problème	21
2.2.2	Premier niveau d’analyse	21
2.2.3	Deuxième niveau d’analyse : étude statistique	21
2.2.4	Comparaison avec la distribution gaussienne estimée	22
2.2.5	Vocabulaire de métrologie : incertitudes de type A et de type B	23
2.2.6	Budget d’incertitude de la mesure	24
3	Ajustement de courbes expérimentales	25
3.1	Données et moindres carrés	25
3.1.1	Position du problème	25
3.1.2	Hypothèse de départ	26
3.2	Ajustement par une fonction affine	27
3.2.1	Ajustement sans prendre en compte d’incertitudes	27
3.2.2	Droite de régression et coefficient de corrélation	28
3.2.3	Incertainces sur les paramètres de l’ajustement	29
3.3	Ajustement par une fonction polynômiale	30
3.4	Ajustement non-linéaire	31

4	Annexes	33
4.1	Annexe A : rappels élémentaires sur les lois de probabilité	33
4.1.1	Loi de probabilité et fonction caractéristique	33
4.1.2	Cas de la distribution normale	33
4.2	Quinconce de Galton ou la “planche à clous”	34
4.2.1	Description du dispositif	34
4.2.2	Modélisation mathématique de la “planche à clous”	35
4.2.3	Composition de deux dispersions consécutives	36
4.3	Annexe C : variance d’échantillon	36
4.3.1	Annexe D : méthode des moindres carrés et poids des mesures	37

1 Mesurages et statistiques

1.1 Rappels sur la distribution normale de Gauss

1.1.1 Intérêt de la distribution normale de Gauss

Commençons par nous intéresser à une fonction d'une grande importance en physique : la distribution normale, également connue sous le nom de "gaussienne" ¹. D'après le théorème de la limite centrale, cette distribution correspond à la distribution limite associée à la mesure d'une grandeur x sujette à de petites incertitudes aléatoires. En guise d'introduction, nous ne résistons pas au plaisir de "raconter" une petite fable ².

Lorsque l'Allemagne eut perdu la guerre en 1945, la situation économique continua à se détériorer rapidement dans le pays. Toute la nourriture resta rationnée, même le pain ; cette denrée était limitée à 200 grammes par personnes et par jour. Les boulangers étaient par conséquent obligés de se procurer des moules spéciaux pour fabriquer des pains pesant exactement 200 grammes et ils en vendaient un par jour à chaque consommateur.

Chaque matin, en se rendant à l'université, un vieux professeur de physique, Herr Pr. Dr. Herbert W., passait chez le boulanger pour prendre sa ration quotidienne. Un jour il lui dit :

— Vous êtes un profiteur, vous volez vos clients. Les moules dont vous vous servez sont de 5 pour cent plus petits qu'ils ne devraient l'être pour faire des pains de 200 grammes, et la farine que vous économisez, vous la vendez au marché noir.

— Mais, Monsieur le Professeur, s'écria le boulanger, personne ne peut donner à tous ses pains exactement le même poids. Certains sont quelques pour cents plus légers, d'autres quelques pour cent plus lourds.

— C'est bien cela, répliqua le professeur, depuis quelques mois, je pèse chaque jour le pain que vous me vendez sur la balance de précision de mon laboratoire. Son poids varie suivant la loi normale. Mais voici la courbe qui donne la répartition des pains de divers poids (figure 1a), comparée au poids de la ration. Vous voyez que certains pains ne pèsent que 185 grammes et que d'autres pèsent 205 grammes mais que le poids moyen obtenu dans ces mesures est de 195 grammes au lieu de 200. Vous allez vous procurer des moules de taille correcte, ou je vous dénonce au service du ravitaillement.

— Je le ferai dès demain, Monsieur le Professeur, dit le boulanger atterré, et soyez bien assuré que cette faute ne se renouvellera pas.

Quelques mois plus tard, le professeur revint trouver le boulanger.

— Je viens de vous dénoncer au service du ravitaillement, dit-il. Vous n'avez pas changé vos moules et vous continuez à voler vos clients.

— Mais, Monsieur le Professeur, s'exclama le boulanger, vous ne pouvez pas m'accuser de vol ! Vous ai-je jamais donné des pains trop légers ces derniers mois ?

— En effet, ils pesaient tous au moins 200 grammes. Mais ce n'est pas parce que vous avez utilisé des moules plus grands ; vous avez tout simplement choisi pour moi les plus gros pains, et gardé les plus petits pour les autres clients.

— Vous ne pouvez pas le prouver, dit le boulanger avec arrogance.

¹La découverte de l'importance de cette loi est attribuée aux pionniers du développement de la théorie des probabilités : d'abord DE MOIVRE (1667–1754), mathématicien d'origine française protestant qui émigra en Angleterre en 1685 après la révocation de l'Edit de Nantes et l'expulsion des Huguenots, puis le Marquis Pierre Simon de LAPLACE (1749–1827) et enfin évidemment, le "petit prince des mathématiques" Karl Fiedrich GAUSS (1777–1855).

²Cette histoire est tirée du merveilleux livre de G. GAMOW et M. STERN, "*Jeux Mathématiques*", Dunod (Paris, 1966), dédié à "Théodore Von Karman, qui aime bien les petits problèmes". Il est indiqué dans le livre que cette fable est rapportée comme véridique et se serait passée à Hambourg après la guerre.

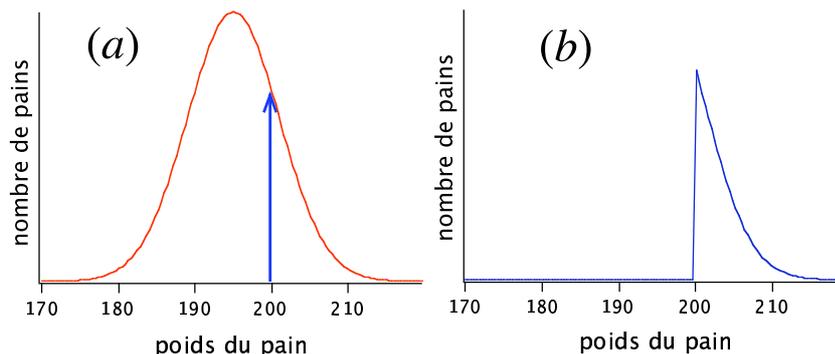


Figure 1: Distribution du poids des pains achetés chez le boulanger par le Pr. Dr. Herbert W.

— Justement si, déclara le professeur. Regardez la distribution statistique que j’ai obtenue en pesant vos pains au cours de ces derniers mois (figure 1b). Au lieu de la distribution normale démontrée par le grand mathématicien Karl Friedrich GAUSS, vous avez une courbe qui correspond à des pains trop lourds ; elle s’arrête brusquement à gauche, et descend progressivement à droite. Jamais des écarts statistiques autour de la moyenne ne pourraient conduire à une telle distribution, et il est clair que vous avez obtenu cette courbe artificiellement en choisissant des pains qui pesaient plus que le poids minimum autorisé. Cette courbe est tout simplement le pied de la courbe de Gauss (figure 1a) : elle représente la même distribution que celle que j’avais obtenue en pesant mes pains avant notre précédente conversation. Je suis certain que le service du rationnement me croira sur parole.

Et, tournant les talons, le Pr. Dr. Herbert W. sortit de la boutique.

On notera qu’en réalité, aucune mesure n’est jamais distribuée exactement suivant la loi gaussienne. Si on répète un grand nombre de fois la mesure d’un paramètre et qu’on trace l’histogramme de ces mesures en fonction de la valeur trouvée, la forme de l’histogramme tend vers une répartition de Gauss lorsqu’on fait tendre le nombre de mesures vers l’infini. Pourquoi cela ? On dit parfois que les expérimentateurs pensent qu’elle a été prouvée par les mathématiciens, puisque sa forme peut être justifiée lorsque les conditions de BOREL sont vérifiées, c’est-à-dire lorsqu’il existe des causes d’erreurs multiples d’importance comparable (*cf.* l’expérience du quinconce de Galton décrite dans l’Annexe ??). À l’inverse, les mathématiciens pensent qu’elle est un résultat de l’expérience, puisqu’elle est obtenue dans la quasi-totalité des cas où un expérimentateur soigneux a pris le temps de regrouper un très grand nombre de mesures indépendantes et de poids égaux d’une même grandeur.

Si la distribution normale est effectivement obtenue pour un très grand nombre d’expériences, certaines mesures effectuées dans des régimes de comptage ³ n’obéissent cependant plus à ce type de statistique lorsque le nombre de “coups” est très faible. Il faut alors utiliser des procédures d’analyse de données adaptées à la situation considérée. On notera également qu’il existe d’autres distributions de probabilité, appelées “vols de Lévy”, qui s’écartent de manière fondamentale de cette limite centrale puisqu’elles ne possèdent ni valeur moyenne ni variance [?]. Alors que celles-ci ont longtemps été considérées comme de simples curiosités mathématiques, on a découvert récemment qu’elles jouent un rôle crucial dans des domaines aussi variés que la biologie, les marchés financiers, ou encore les embouteillages de la circulation automobile ! Elles se rencontrent également dans différents phénomènes de physique dont l’évolution temporelle est dominée par l’occurrence d’événements rares, comme par exemple la diffusion de particules dans les milieux très fortement hétérogènes ou le refroidissement d’atomes par laser.

³On pensera par exemple au cas de la détection de désintégrations radioactives en physique nucléaire ou bien celui d’une mesure optique effectuée pour un très faible flux lumineux dans le régime de comptage de photons.

1.1.2 Expression mathématique de la loi normale

L'expression mathématique de la loi normale, fonction de deux paramètres μ et σ (avec $\sigma > 0$), est :

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

Nous savons évidemment que $G(x)$ a une forme de courbe “en cloche”, symétrique par rapport à μ qui correspond à la valeur de x pour laquelle la courbe atteint son maximum. Le paramètre σ caractérise la largeur de la courbe, tandis que le facteur $\sigma\sqrt{2\pi}$ assure la normalisation de la distribution de probabilité $d\mathcal{P} = G(x)dx$ associée à cette fonction, c'est-à-dire :

$$\int d\mathcal{P} = \int_{-\infty}^{+\infty} G(x)dx = 1. \quad (2)$$

Les paramètres μ et σ peuvent être interprétés comme liés aux premiers moments de la distribution de probabilité ⁴ donnée par la loi normale $G(x)$:

- Le paramètre μ correspond à la valeur moyenne $\langle x \rangle$ de la variable x :

$$\mu = \langle x \rangle = \int_{-\infty}^{+\infty} xG(x)dx. \quad (3)$$

Cette quantité porte également le nom d'*espérance mathématique* ⁵, notée $E[x]$.

- Le paramètre σ correspond à l'*écart-type* ou *écart quadratique moyen* de la variable x , également noté Δx et défini comme :

$$(\Delta x)^2 = \sigma^2 = \langle (x - \langle x \rangle)^2 \rangle \quad \text{soit} \quad \sigma^2 = \langle x^2 \rangle - \langle x \rangle^2. \quad (4)$$

Le paramètre σ^2 est encore appelé *variance* $V[x]$ de la loi de probabilité. Plus σ est petit, plus il est probable que l'on trouve une valeur de x proche de la moyenne. La quantité σ constitue par conséquent une mesure de l'écart à la moyenne, et elle quantifie les fluctuations de la variable aléatoire x autour de la valeur moyenne μ . Cette propriété apparaît également si nous évaluons la largeur totale à mi-hauteur de la gaussienne :

$$\Delta x_{1/2} = 2\sqrt{\ln 2} \sigma \text{ (FWHM)} \quad \text{soit} \quad \Delta x_{1/2} \simeq 2.355 \times \sigma. \quad (5)$$

De même, la hauteur de la gaussienne en $x = \mu \pm \sigma$ est égale à $1/\sqrt{e}$. Puisque $1/\sqrt{e} \simeq 0.607$, l'écart-type σ peut être considéré de manière approchée comme la demi-largeur à mi-hauteur de la gaussienne. Nous dirons ainsi que la “largeur” d'une distribution gaussienne est de l'ordre de son écart-type σ .

1.1.3 Normalisation et fonction “erreur”

En faisant les changements de variable

$$z = \frac{x - \mu}{\sigma} \quad \text{et} \quad dz = \frac{1}{\sigma} dx,$$

⁴On appelle moment d'ordre n d'une distribution de probabilité $\mathcal{P}(x)$ la valeur moyenne de x^n , calculée selon $\langle x^n \rangle = \int_{-\infty}^{+\infty} x^n \mathcal{P}(x) dx$. Quelques rudiments de statistique sont rappelés dans l'Annexe ??.

⁵Cette appellation tire son origine de l'analyse faite au XVIIème siècle du “problème des partis”, en lien avec l'application des probabilités aux jeux de hasard : quelle est la juste répartition des enjeux lorsqu'une partie d'un jeu de pur hasard est interrompue avant son terme. Des savants aussi illustres que PASCAL, FERMAT et HUYGENS se sont intéressés à ce problème et ont montré que la répartition doit se faire selon les probabilités de gagner des joueurs, calculées à partir de toutes les suites possibles de la partie. Chaque joueur recevra alors son “espérance de gain” au moment de l'interruption du jeu.

la fonction gaussienne définie précédemment peut être mise sous une forme normalisée, avec une valeur moyenne nulle et un écart-type égal à l'unité (figure 2-a) :

$$y = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}. \quad (6)$$

Cette expression donne en particulier la probabilité $F(a) = \Pr(z \leq a)$ qu'une mesure de la grandeur z soit plus petite que la valeur a :

$$F(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-u^2/2} du. \quad (7)$$

Cette intégrale, dont la variation en fonction de a est représentée sur la figure 2-a correspond à l'aire de la gaussienne située à gauche de la limite $z = a$. Elle permet de déterminer la probabilité que z soit contenu dans tout intervalle $b \leq z \leq a$:

$$\Pr(b \leq z \leq a) = F(a) - F(b). \quad (8)$$

Ainsi, nous pouvons calculer la probabilité que le tirage de la variable aléatoire tombe à plus ou moins un écart-type σ de la valeur moyenne μ :

$$\Pr(-1 \leq z \leq 1) = F(1) - F(-1) = 0.8413 - 0.1587 = 0.6826$$

c'est-à-dire une probabilité de 68 %. Quand on augmente la largeur de cet intervalle, cette probabilité va tendre rapidement vers l'unité. Ainsi, la probabilité qu'une mesure soit à $\pm 2\sigma$ (resp. $\pm 3\sigma$) est de 95.4 % (resp. 99.7 %).

On définit également la fonction "erreur" $\operatorname{erf}(z)$ comme :

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt \quad (9)$$

dont la variation est représentée sur la figure 2-b. Cette fonction, qui est intégrée dans les logiciels de calcul ou de traitement de données, permet de calculer ces mêmes probabilités, au moyen de l'expression :

$$\Pr(z \leq a \mid \text{gaussienne}, \mu, \sigma) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{a - \mu}{\sigma\sqrt{2}} \right) \right]. \quad (10)$$

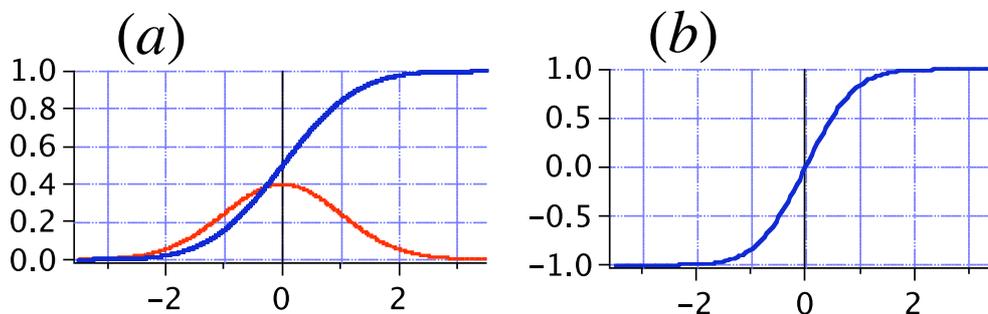


Figure 2: (a) – Tracé de la gaussienne normalisée et de son aire $F(z)$, obtenue par simple intégration numérique. (a) – Tracé de la fonction "erreur" $\operatorname{erf}(z)$, permettant de calculer au moyen de la formule (10) la probabilité $\Pr(\pm t \times \sigma)$ que le résultat d'un tirage de la variable aléatoire se trouve dans un intervalle de t fois son écart-type σ autour de la valeur moyenne μ .

1.2 Interprétation statistique d'une opération de mesurage

1.2.1 Moyenne et incertitude d'une mesure

Les rappels que nous venons d'effectuer vont nous permettre de mieux comprendre l'écriture du résultat de la mesure d'une grandeur donnée x , sous la forme :

$$x = x_m \pm \Delta x \quad (11)$$

où x_m est la meilleure estimation de la grandeur mesurée x et Δx l'incertitude sur la mesure. Ne craignons pas d'insister sur l'importance de l'estimation de l'incertitude Δx . En son absence, il devient impossible de chercher à comparer entre eux des résultats. Comment savoir, sans connaître l'incertitude, si une grandeur a évolué, si tel procédé de mesure conduit au même résultat que tel autre, si la différence éventuellement observée entre les résultats obtenus par des méthodes ou des équipes différentes est réellement significative ou bien si elle est uniquement due à des phénomènes aléatoires mal maîtrisés ? Comment comparer le résultat qu'on a obtenu à la valeur qui a été publiée dans un article scientifique ? Comment le comparer à des valeurs de référence spécifiées par exemple dans une norme ou dans les objectifs d'un contrat, et garantir ainsi la conformité du produit ou du système réalisé ?

Dans son rapport sur le concours 1998, le jury de l'épreuve de montage à l'agrégation de physique avait insisté sur le fait que trop de candidats ne savent pas comment aborder le problème de la précision des résultats et l'évaluation d'une incertitude de mesure, représentée par le terme Δx dans l'équation (11). Il précise en particulier que la critique des résultats obtenus repose trop souvent sur un calcul d'incertitude uniquement basé sur l'utilisation de différentielles dont on a pris la valeur absolue. Malheureusement, cette méthode, qui a été longtemps enseignée et utilisée par les physiciens, n'est pas appropriée car le résultat qu'elle fournit est à la fois optimiste et pessimiste :

- Il est pessimiste car il suppose que toutes les causes d'erreur agissent dans le même sens. Il s'agit là d'un cas particulièrement défavorable qui n'a dans la réalité que peu de chances de se produire.
- Il est optimiste car il suggère que la valeur cherchée ne peut pas être à l'extérieur de la barre d'incertitude $[x_m - \Delta x, x_m + \Delta x]$ et que celle-ci représente par conséquent un domaine de valeurs qui contient à "coup sûr" la valeur "exacte" x_{exact} de la grandeur mesurée.

La définition de l'incertitude sur la mesure est en réalité liée à des considérations d'ordre statistique⁶ : l'intervalle Δx doit être présenté comme un domaine à l'intérieur duquel la valeur x_{exact} a de "fortes chances" d'être contenue. La dispersion des résultats de mesure observée lorsqu'on répète l'opération de mesure dans des conditions apparemment identiques, justifie l'adoption d'une description statistique du mesurage. Puisque la réalisation d'une mesure comporte nécessairement une part d'aléatoire, la grandeur physique x est alors caractérisée, non plus par une valeur exacte, mais par la probabilité de trouver telle ou telle valeur. Nous devons ainsi interpréter le résultat d'une mesure de cette grandeur comme une valeur particulière, issue du tirage au sort d'une **variable aléatoire** X dans l'ensemble, en général infini, des résultats possibles. Ce tirage au sort de X est associé à une **fonction de distribution de probabilité** $\Pr(X = x)$ de trouver la valeur x , pour laquelle nous ferons l'abus de la simple notation $\Pr(x)$.

Nous pouvons maintenant interpréter correctement les différents termes dans l'équation (11) :

⁶Il faut ainsi se débarrasser de certains a priori sur l'utilisation des outils d'analyse statistique, lesquels sont malheureusement souvent considérés comme une simple description austère et sans âme des faits économiques ou démographiques. Les analyses des statisticiens sont également parfois perçues avec un mélange de scepticisme ou de méfiance, à l'instar de Mark TWAIN qui déclarait qu'il existe trois formes de mensonges : *lies - damned lies - and statistics*. La statistique est en réalité une discipline importante, partie intégrante de tout processus de recherche élaboré à partir d'un ensemble de données brutes.

- La moyenne x_m correspond à la valeur moyenne $\langle x \rangle$, calculée avec la distribution de probabilité $\Pr(x)$:

$$\langle x \rangle = \int x \Pr(x) dx ;$$

- L’intervalle (ou la barre d’erreur) $[x_m - \Delta x, x_m + \Delta x]$ est défini comme un “intervalle de confiance”, associé à une probabilité donnée \mathcal{P} de contenir la valeur vraie x_{exact} de la grandeur mesurée [?]. Cette probabilité est appelée **niveau de confiance** et la valeur choisie en pratique est $\mathcal{P} = 0.95 = 95\%$ ou $\mathcal{P} = 0.99 = 99\%$ ⁷.

Supposons que les différentes valeurs pouvant être obtenues pour la mesure d’une grandeur physique donnée soient distribuées selon une probabilité gaussienne. Nous pouvons alors évaluer à partir du calcul effectué précédemment de la fonction erreur $\text{erf}(z)$ le risque encouru lorsqu’on donne pour la valeur de la grandeur un intervalle de confiance, appelé encore intervalle de “certitude maîtrisée”. En limitant cet intervalle à $\pm(2 \times \sigma)$ autour de la valeur estimée, le risque de trouver une valeur au-delà des bornes de l’intervalle est inférieur à 5 % ⁸.

Les considérations précédentes montrent qu’il convient de faire, pour chaque mesurage, un inventaire précis et rigoureux des causes d’erreur susceptibles de biaiser le résultat. Il faut ensuite effectuer les corrections nécessaires, évaluer l’incertitude associée à chacune de celles-ci, puis composer ces termes avec la dispersion liée à la non répétabilité du procédé de mesure. On pourra alors exprimer le résultat du mesurage sous la forme d’un intervalle de confiance, dont la valeur aura été déterminée à partir de l’écart-type σ associé au résultat obtenu.

1.2.2 Propagation des erreurs

La composition quadratique des différentes contributions à l’incertitude de mesure, chacune exprimée sous forme d’écart-type (même si l’appréciation de celui-ci ne résulte pas d’une multiplicité de mesures), correspond alors mieux à la prise en compte simultanée des causes d’erreur indépendantes. Ainsi, lors de la détermination d’une résistance R par voltmètre et ampèremètre, il n’est *a priori* pas justifié de corrélérer les composantes de l’incertitude associées à la mesure de la tension U et à celle de l’intensité I . Cette corrélation est implicite dans le calcul usuel par différentielle logarithmique

$$\frac{\Delta R}{R} = \frac{\Delta U}{U} + \frac{\Delta I}{I}$$

qui tend à quantifier un hypothétique “intervalle maximal” pour le résultat. Pour exprimer l’écart-type σ_R sur la mesure de la résistance R , il est plus approprié d’utiliser la relation de “propagation des erreurs”

$$\left(\frac{\sigma_R}{R}\right)^2 = \left(\frac{\sigma_U}{U}\right)^2 + \left(\frac{\sigma_I}{I}\right)^2 .$$

Si cette expression a un sens mathématique bien défini, on pourra se contenter dans la pratique de dire que l’incertitude relative $\Delta R/R$ est de l’ordre de :

$$\frac{\Delta R}{R} \approx \sqrt{\left(\frac{\Delta U}{U}\right)^2 + \left(\frac{\Delta I}{I}\right)^2} .$$

⁷Il est bon de savoir que l’utilisation de ces méthodes statistiques pour la donnée des incertitudes se sont imposées seulement au début des années 1990, avec l’installation des normes d’assurance qualité ISO 9000 dans le monde de l’entreprise. Ces méthodes avaient en réalité été mises au point dès le début du XXème siècle et adoptées par quelques industries, comme par exemple l’agriculture et la fabrication textile de fil et de tissu, dans le but de rationaliser et améliorer la production.

⁸On notera l’importance du choix du niveau de certitude, corrélé à la définition de l’incertitude de la mesure. Si nous avons choisi de définir l’incertitude Δx à $\pm(1 \times \sigma)$, un tiers des mesures effectuées tomberait alors statistiquement en dehors de l’intervalle $[x_m - \sigma, x_m + \sigma]$. Remarquons qu’en toute rigueur, le niveau de confiance $\mathcal{P} = 95\%$ correspond à $\pm(1.96 \times \sigma)$.

1.2.3 Erreurs aléatoires et erreurs systématiques dans un mesurage

Une première cause de variation des résultats de mesure réside dans les appréciations du manipulateur lui-même. On ne refait jamais la mesure exactement dans les mêmes conditions, parce que la lecture des instruments va changer d'une mesure à la suivante : erreur de parallaxe dans le repérage d'un trait ou de l'indication d'une aiguille, effets de ménisque pour le niveau de liquide dans une pipette, fatigue au bout de longues heures de travail sur le dispositif expérimental, etc.

Il existe également bien d'autres sources d'erreurs qui peuvent affecter le résultat brut d'un mesurage. Voici tout d'abord quelques exemples d'erreurs systématiques :

- La grandeur mesurée peut être parfois mal définie. Il faut alors se demander quelles sont les conditions de validité de la loi modèle prévue, savoir si la grandeur mesurée varie dans le temps ou dans l'espace. La grandeur peut également avoir une dispersion intrinsèque. C'est par exemple le cas pour la mesure de l'énergie d'un niveau instable d'un noyau ou d'un système atomique ou la mesure de la position d'une particule microscopique lorsque celle-ci est sujette au mouvement brownien créé par son environnement.
- La grandeur étudiée peut être affectée par le procédé de mesure lui-même, ce qui impose alors l'usage de méthodes d'échantillonnage. Le mode opératoire choisi peut également introduire des erreurs. Ainsi, une pesée effectuée dans l'air doit prendre en compte une correction due à la poussée d'Archimède ; une mesure de résistance effectuée à l'aide d'un voltmètre et d'un ampèremètre est biaisée aussi bien dans les configurations "courte" ou "longue" dérivation, etc.
- Les capteurs et les instruments utilisés pour la mesure, et avec lesquels le système est mis en interaction, peuvent présenter des défauts : temps de réponse fini et bande passante limitée, justesse et sensibilité de l'instrument de mesure utilisé⁹, réponse non-linéaire d'un capteur, effet d'hystérésis...
- De nombreuses grandeurs qui caractérisent les conditions d'ambiance dans lesquelles la mesure est réalisée vont influencer sur le résultat. Il s'agit selon les cas de la température, des conditions électriques ou magnétiques, de la présence ou non de lumière parasite...

De tels effets doivent être recherchés puis quantifiés, de manière à pouvoir éventuellement introduire des corrections permettant de compenser les erreurs systématiques ainsi introduites. Les constructeurs fournissent en particulier des indications concernant la précision de leurs appareils ou de leurs composants, sous la forme d'une incertitude à attribuer aux mesures effectuées. Il faut alors se reporter aux notices du constructeur pour connaître le sens précis... de la "précision" indiquée, sachant que celle-ci est également donnée suivant une norme industrielle. Ainsi, la norme ISO 9000 précise aux fabricants comment ils doivent procéder en pratique pour fournir les indications quantitatives sur les caractéristiques des produits qu'ils commercialisent. Prenons ainsi l'exemple d'une résistance donnée avec une précision de 1 %. Cette indication recouvre d'abord un aspect d'échantillonnage, puisque l'industriel va fabriquer des milliers – ou des millions – de produits, dont les résistances vont varier nécessairement un peu d'un exemplaire à l'autre. Il va également inclure un aspect de fonctionnement, puisque la résistance va changer avec la température du composant, qui est elle-même fonction de l'intensité du courant. En donnant R avec une incertitude de 1 %, le fabricant donne une limite à l'effet global de ces différents facteurs, la valeur correspondant dans la norme à deux fois l'écart-type σ sur

⁹On prendra soin de distinguer la *sensibilité*, qui indique la variation de l'indication de l'appareil en fonction de la variation de la grandeur à mesurer, et la *justesse* de l'appareil. S'il est mal calibré, un dispositif de mesure peut être sensible sans être pour autant juste. C'est par exemple le cas d'une mesure de température effectuée à l'aide d'une thermistance, par comparaison avec une mesure de cette même grandeur effectuée au moyen d'une sonde étalon à résistance de platine.

l'échantillon de mesures.

Lors des montages, il s'agit souvent de comparer le résultat de la mesure à ce qu'on appelle parfois abusivement la valeur "théorique" et qui, la plupart du temps, est issue de "valeurs nominales". Si celle-ci figure dans un recueil de données comme dans le cas d'une constante physique fondamentale ou de la longueur d'onde d'une radiation, elle est effectivement bien connue. Dans ce cas, l'incertitude sur la valeur tabulée dans le recueil peut être négligée et il suffit de vérifier que cette valeur tombe à l'intérieur de l'intervalle de confiance évalué de façon réaliste pour la mesure. Dans d'autres cas, la connaissance imparfaite des valeurs de référence utilisées, des courbes d'étalonnage des instruments induit également des erreurs sur la valeur "théorique" de comparaison. Il est alors nécessaire de lui attribuer également une incertitude afin de pouvoir parvenir à une comparaison pertinente entre le résultat de la mesure et la valeur de référence.

Pour donner un exemple d'une telle situation, considérons ainsi la détermination de la fréquence de résonance ω_0 d'un circuit RLC série et la vérification de la formule $LC\omega_0^2 = 1$. La mesure directe de la fréquence de résonance ω_0 doit être effectuée avec soin. Si la technique des courbes de LISSAJOUS à l'oscilloscope est utilisée, il faut se placer dans les meilleures conditions pour repérer la situation dans laquelle l'ellipse est réduite à un segment de droite. On doit donc veiller à faire les traces les plus fines possibles et à travailler avec des calibres qui correspondent à des valeurs maximales du gain. Si la fréquence ω_0 est mesurée à l'aide d'un fréquencemètre de qualité, la seule incertitude à prendre en compte est alors pratiquement celle du pointé. L'incertitude sur le calcul de la fréquence ω_0 peut ensuite être estimée à l'aide des formules habituelles. Si l'on cherche ensuite à vérifier la relation $LC\omega_0^2 = 1$, il faut savoir qu'une bobine réelle peut être assimilée à une inductance L qui n'est pas absolument constante lorsque la fréquence à laquelle elle est utilisée varie. Ainsi, il est loin d'être acquis que le calcul de ω_0 à partir des valeurs de l'inductance L et de la capacité C conduise à une précision meilleure que celle d'une mesure directe de ω_0 .

Quitte à nous répéter, disons-le encore une fois : **une valeur mesurée n'est pas une valeur certaine**. La dispersion des résultats est la signature de la variabilité de la mesure, laquelle est un phénomène objectif. De plus, il existe une certaine méconnaissance de la valeur pour chaque correction individuelle et par conséquent de la correction totale. Une fois prises en compte toutes les causes d'erreur, on appelle **incertitude de mesure** le paramètre associé au résultat qui caractérise la dispersion des valeurs numériques.

Nous retiendrons pour la suite :

Une mesure s'exprime toujours par la valeur mesurée et son incertitude.

On aura évidemment l'honnêteté de ne **jamais** évaluer l'erreur sur une quantité mesurée en comparant la valeur trouvée expérimentalement à un quelconque chiffre déniché dans un livre ou sur une page web !

2 Exploitation d'une série de mesure

2.1 Estimateurs de la distribution de probabilité associée à la mesure

2.1.1 Position du problème et maximum de vraisemblance

La réalisation d'une mesure d'une grandeur physique est schématisée sur la figure 3. La répétition (à l'identique, autant que faire se peut) du procédé de mesure direct conduit à associer à la grandeur mesurée, appelée *mesurande*, un ensemble de n valeurs numériques, qui correspondent aux résultats bruts des mesurages :

$$\{x_1, x_2, \dots, x_n\}_{i=1 \dots n} .$$

L'ensemble de ces valeurs constitue ce nous appellerons l'*échantillon de mesures*.

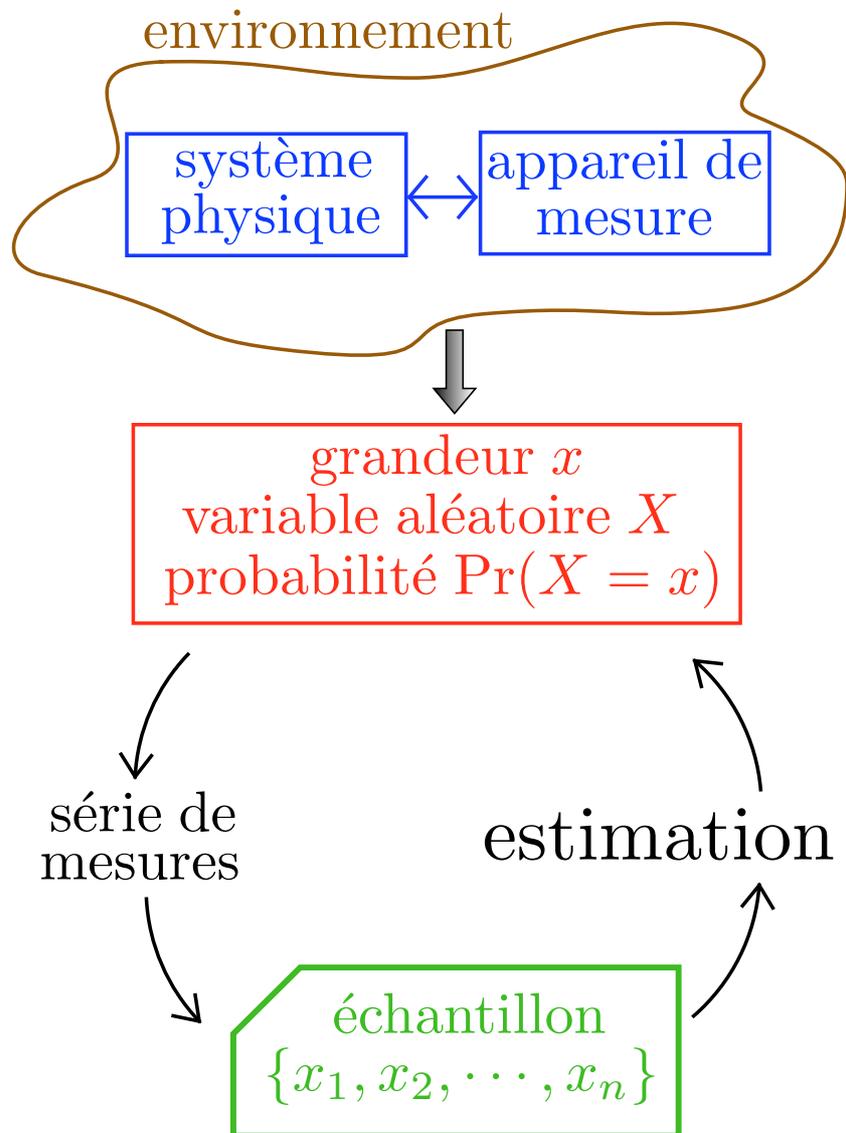


Figure 3: Mesure de la grandeur x , représentée par une variable aléatoire X et associée à une distribution de probabilité inconnue $\Pr(X = x)$. La figure montre la relation entre l'échantillonnage, qui correspond à n tirages au sort de la variable aléatoire X , et l'estimation, procédure d'exploitation de l'échantillon de mesure pour obtenir des informations sur la distribution de probabilité $\Pr(x)$. Les paramètres de cette distribution sont estimés à partir d'une statistique calculée sur la base de l'échantillon des n valeurs de la variable aléatoire X .

En pratique, on ne fait jamais assez de mesures pour connaître la distribution limite associée à la mesure de la grandeur x , de sorte que les paramètres μ et σ qui interviennent dans l'expression de la distribution de probabilité restent inconnus. Le problème va donc se ramener à une exploitation optimale de l'échantillon de mesures afin d'obtenir les valeurs *estimées* des paramètres de la distribution de probabilité $\Pr(x)$ associée à la mesure de la grandeur x . Cette estimation, dont nous attendons qu'elle soit d'autant meilleure que le nombre n d'éléments dans l'échantillon est grand, nous permettra alors de déterminer de manière statistique le résultat de mesures ultérieures de la grandeur x . Afin de bien faire ressortir la différence entre paramètres de la distribution de probabilité et grandeurs statistiques, nous utiliserons des symboles différents. Ainsi, les caractéristiques de la distribution de probabilité seront notées par des lettres grecques tandis que les caractéristiques statistiques de l'échantillon seront notées par des lettres romaines.

Il existe plusieurs méthodes pour déterminer les “estimateurs” des paramètres inconnus. Nous pouvons en particulier utiliser une méthode d'estimation du “maximum de vraisemblance” qui fut introduite par R. A. FISHER (figure 4).



Figure 4: Sir Ronald Aylmer FISHER (1890–1962). Ce scientifique anglais est reconnu comme un des fondateurs de la statistique moderne. Il fut en particulier un des premiers à introduire les méthodes statistiques en agronomie, en partant de schémas d'expériences afin de rechercher et de contrôler les sources de variations.

Pour appliquer cette méthode, nous devons connaître les distributions de probabilité normalisées des n variables x_i qui forment notre ensemble de données,

$$\Pr(x_i | \theta)$$

où θ correspond de manière générique au paramètre à estimer duquel dépend la distribution de probabilité. Nous pouvons alors considérer le produit de ces fonctions pour l'ensemble des mesures effectuées :

$$\mathcal{L}(x_1, x_2, \dots, x_n; \theta) = \Pr(x_1 | \theta) \times \Pr(x_2 | \theta) \times \dots \times \Pr(x_n | \theta), \quad (12)$$

que nous appellerons “fonction de vraisemblance” pour le paramètre θ . Le principe de “vraisemblance maximale” nous dit que pour l'ensemble d'observations, la valeur θ^* qui rend maximal \mathcal{L} correspond au meilleur estimateur du paramètre θ :

$$\left(\frac{\partial \mathcal{L}(x_1, x_2, \dots, x_n; \theta)}{\partial \theta} \right)_{\theta=\theta^*} = 0. \quad (13)$$

Pour chercher cet estimateur, il est plus commode, et cependant équivalent du point de vue mathématique puisque le logarithmique est une fonction toujours croissante, de chercher le maximum de $\ln \mathcal{L}$. Supposons en particulier que l'ensemble des n données $\{x_1 \cdots x_n\}$ soit tiré d'une distribution de probabilité sous-jacente correspondant à une répartition gaussienne de valeur moyenne μ et d'écart-type σ :

$$\Pr(x_i | \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\mu - x_i}{\sigma} \right)^2 \right]. \quad (14)$$

En supposant que les n mesures sont indépendantes, nous avons alors :

$$\mathcal{L} = \prod_{i=1}^n \Pr(x_i | \mu, \sigma) \quad \text{et} \quad W = \ln \mathcal{L} = -n \ln(\sigma\sqrt{2\pi}) - \frac{1}{2} \sum_{i=1}^n \left(\frac{\mu - x_i}{\sigma} \right)^2. \quad (15)$$

2.1.2 Estimateur de la moyenne μ

Pour déterminer l'estimateur de la valeur moyenne μ , nous pouvons ainsi calculer

$$\frac{\partial W}{\partial \mu} = - \sum_{i=1}^n \frac{\mu - x_i}{\sigma^2}. \quad (16)$$

Cette expression est nulle pour :

$$\sum_{i=1}^n \frac{\mu^* - x_i}{\sigma^2} = 0$$

soit :

$$\mu^* = m \quad \text{avec} \quad m = \frac{1}{n} \sum_{i=1}^n x_i. \quad (17)$$

Par conséquent, si un ensemble de mesures obéit à une distribution de probabilité gaussienne, la "valeur la plus vraisemblable" de la grandeur mesurée x vis-à-vis des valeurs mesurées $\{x_i\}_{i=1 \dots n}$ est simplement la **moyenne arithmétique** m de ces valeurs. Nous appellerons cette grandeur la *moyenne d'échantillon* ou la *moyenne expérimentale*.

Pris dans le sens inverse, c'est en réalité en partant de ce résultat que Karl Friedrich GAUSS, qui dirigeait l'Observatoire de Göttingen, fut conduit à introduire la loi normale après s'être intéressé au problème des erreurs dans les mesures astronomiques.¹⁰ Son argumentation part du postulat intuitif de la moyenne : "étant données plusieurs valeurs observées d'une quantité inconnue, la valeur la plus probable de cette quantité est la moyenne des valeurs observées". En supposant que les diverses observations sont représentées par des variables aléatoires indépendantes ayant une même distribution de probabilité, on montre que la fonction correspondante est une loi normale.

Remarquons que cette correspondance exacte entre la moyenne m de l'échantillon de mesure et l'estimateur du maximum de vraisemblance n'est pas nécessairement valable pour des distributions autres que la gaussienne. Considérons par exemple le cas d'une distribution rectangulaire sur un intervalle (a, b) . L'estimateur μ^* est égal à la valeur médiane $\frac{1}{2} [x_{(1)} + x_{(2)}]$, où $x_{(1)}$ et $x_{(2)}$ sont respectivement la mesure la plus faible et la mesure la plus élevée dans l'échantillon. On constate que dans ce cas, l'estimateur est différent de la moyenne d'échantillon m .

¹⁰Ce résultat fut publié dans l'ouvrage *Theoria motus corporum caelestium* (1809). GAUSS (1777-1855) fut un des derniers scientifiques à utiliser le latin pour la diffusion de ses travaux scientifiques.

2.1.3 Précision sur la détermination de l'estimateur de la moyenne

Si nous reproduisons l'ensemble des n mesures dans des conditions identiques, la dispersion des résultats conduira à des valeurs différentes $\{x'_1, \dots, x'_n\}$ dans l'échantillon. La moyenne d'échantillon m est par conséquent elle-même sujette à fluctuations d'une série de mesures à l'autre. Dans l'ensemble d'une infinité virtuelle d'échantillons obtenus indépendamment les uns des autres et dans des conditions identiques, on aura une moyenne (ou espérance mathématique), une variance et une distribution. Cette distribution de probabilité est la distribution d'échantillonnage de la moyenne.

L'espérance $E(m)$ et la variance $\text{Var}(m) = (\sigma_m)^2$ associées à la moyenne d'échantillon m peuvent être exprimées en fonction des paramètres μ et σ . En effet, si $\{x_1, \dots, x_n\}$ sont des copies indépendantes d'une grandeur x de moyenne μ et de variance σ , alors :

$$E(m) = E\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n} \sum_{i=1}^n E(x_i) = \frac{1}{n} \times (n\mu) \quad (18)$$

soit :

$$E(m) = \mu . \quad (19)$$

De même :

$$(\sigma_m)^2 = \text{Var}(m) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(x_i) = \frac{1}{n^2} \times (n\sigma^2) \quad (20)$$

soit :

$$(\sigma_m)^2 = \frac{\sigma^2}{n} \quad \text{et} \quad \sigma_m = \frac{\sigma}{\sqrt{n}} . \quad (21)$$

Une autre méthode d'évaluation des paramètres de la distribution d'échantillonnage de la moyenne consiste à utiliser le principe du maximum de vraisemblance afin de déterminer l'écart-type σ_θ associé à l'estimateur θ^* , lorsque celui-ci est mesuré sur un grand nombre d'échantillons. Il n'est pas difficile de se convaincre que :

$$\frac{1}{(\sigma_\theta)^2} = - \left(\frac{\partial^2 W}{\partial \theta^2} \right)_{\theta=\theta^*} \quad (22)$$

En dérivant une deuxième fois l'expression (16) par rapport au paramètre μ , nous obtenons ainsi

$$\frac{1}{(\sigma_m)^2} = - \left(\frac{\partial^2 W}{\partial \mu^2} \right)_{\mu=\langle m \rangle} = \sum_{i=1}^n \frac{1}{\sigma^2} \quad (23)$$

soit :

$$\sigma_m = \frac{\sigma}{\sqrt{n}} . \quad (24)$$

Cette dernière expression, que nous avons ainsi établie de deux manières différentes, correspond à un résultat que nous connaissons bien. Nous avons pu constater expérimentalement que la valeur moyenne d'un ensemble de n mesures tend à se stabiliser lorsque la taille n de l'échantillon augmente. De manière plus précise, nous venons d'établir que l'écart-type sur la moyenne de n observations est plus petit que l'écart-type associé à une mesure unique, dans un rapport de $1/\sqrt{n}$.

Puisqu'une augmentation du nombre de points de mesure conduit à une meilleure localisation de la valeur moyenne, on pourrait alors penser qu'en partant d'une mesure peu précise, on puisse très facilement augmenter la précision de celle-ci grâce à une simple répétition de la mesure. En réalité, il n'en est rien puisque le gain dans la précision ne varie qu'en \sqrt{n} . Ainsi, réduire

l'incertitude d'un facteur 10 requiert d'effectuer 100 fois plus de mesures, et une amélioration d'un facteur 1000 de cette même précision obligerait à reproduire un million de fois la mesure. Il faut également être conscient du fait que si un ensemble répété de mesures permet effectivement de réduire l'erreur statistique, l'augmentation du temps nécessaire pour effectuer cette répétition des mesures nécessite une très grande stabilité du dispositif ou de l'appareil de mesure. En effet, si celui-ci dérive au cours du temps, la déviation standard cesse d'être un bon indicateur de la précision de la mesure ¹¹.

2.1.4 Moyenne d'échantillon et loi des grands nombres

Nous avons supposé que les mesures individuelles $\{x_1, \dots, x_N\}$ de la grandeur x suivaient une distribution normale, centrée sur μ et d'écart-type σ . Si le même appareil permet d'effectuer plusieurs déterminations indépendantes de la moyenne de N mesures, et que ces mesures ne sont pas affectées par des erreurs systématiques, la moyenne d'échantillon m suivra une distribution normale centrée sur μ et d'écart-type σ/\sqrt{N} . Ce résultat est en réalité tout-à-fait général et est indépendant de la forme initiale de la distribution de probabilité, comme l'indique la loi des grands nombres.

Afin de comprendre pourquoi la forme initiale de la distribution de probabilité $\Pr(x)$ associée à une mesure individuelle de la grandeur x se trouve lissée pour aboutir à une loi gaussienne pour la moyenne d'échantillon, commençons par considérer la distribution de probabilité associée à la somme de deux tirages. Celle-ci va s'écrire :

$$\Pr(\text{somme de deux tirages} = x) = \int \Pr(y)\Pr(x - y) dy, \quad (25)$$

et ainsi de suite pour la composition de plusieurs tirages. En d'autres mots, la probabilité d'une somme est le produit de convolution des probabilités individuelles.

Pour continuer le calcul et simplifier cette expression, introduisons la fonction caractéristique $\varphi_x(t)$ associée à la distribution de probabilité normalisée $\Pr(x)$ (cf. Annexe A) :

$$\int_{-\infty}^{+\infty} \Pr(x) dx = 1 \quad \text{et} \quad \varphi_x(t) = \int_{-\infty}^{+\infty} e^{itx} \Pr(x) dx \quad (26)$$

telle que :

$$\begin{aligned} \varphi_x(t) &= \langle e^{itx} \rangle \\ &= 1 + it\langle x \rangle - \frac{t^2}{2}\langle x^2 \rangle + \dots \end{aligned} \quad (27)$$

À la condition qu'elle soit n -fois dérivable, cette fonction $\varphi_x(t)$, qui porte également le nom de "fonction génératrice des moments", permet de calculer les moments d'ordre 1 à n de la distribution de probabilité. Remarquons que puisque $\Pr(x)$ est normalisée, la fonction $\varphi_x(t)$ existe dans tous les cas mais ce n'est pas forcément le cas pour ses dérivées. On pensera ainsi à une distribution lorentzienne (ou de Cauchy), pour laquelle

$$\Pr(x) = \frac{1}{\pi(1+x^2)} \quad \text{et} \quad \varphi_x(t) = e^{-|t|}.$$

¹¹Ce problème se pose en particulier lorsqu'on cherche à évaluer la stabilité à long terme d'un oscillateur ou d'une horloge. Ainsi, depuis le début des années soixante et les travaux sur les horloges atomiques, de nombreuses études théoriques et expérimentales ont été effectuées pour caractériser d'une part les fluctuations d'oscillation des résonateurs, qui conduisent à une *instabilité en fréquence*, et d'autre part à ses instabilités de fonctionnement au cours du temps. Nous renvoyons le lecteur intéressé par ce problème à des références plus spécialisées sur la métrologie des temps et des fréquences (par exemple [?]).

Ce type de distribution possède ni valeur moyenne ni moments d'ordre supérieur. De façon équivalente, la fonction caractéristique qui lui est associée possède une singularité en $t = 0$ et ne permet donc pas de générer les moments de la distribution.

Supposons maintenant que la distribution de probabilité $\Pr(x)$ soit telle que $\langle x \rangle$ et $\langle x^2 \rangle$ existent. Sans perdre de généralité, nous pouvons également supposer que la moyenne est nulle et que la variance est égale à l'unité. Dans ce cas :

$$\varphi_x(t) = 1 - \frac{1}{2}t^2 + t^2 o(t) \quad \text{avec} \quad o(t) \rightarrow 0 \quad \text{pour} \quad t \rightarrow 0. \quad (28)$$

Pour $N \gg 1$, cette équation (28) conduit à la relation approchée :

$$[\varphi_x(t)]^n \simeq \exp\left(-n \frac{t^2}{2}\right). \quad (29)$$

En utilisant la relation de transformée de Fourier :

$$\begin{aligned} f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-itx) F(t) dt &\leftrightarrow F(t) = \int_{-\infty}^{+\infty} \exp(itx) f(x) dx \\ f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{x^2}{2\sigma^2}\right) &\leftrightarrow F(t) = \exp\left(-\frac{1}{2}t^2\sigma^2\right) \end{aligned} \quad (30)$$

nous obtenons ainsi :

$$\Pr(\text{somme de } n \text{ tirages} = x) = \frac{1}{\sqrt{2\pi N}} \exp\left(-\frac{x^2}{2N}\right). \quad (31)$$

Nous avons ainsi donné une démonstration ¹² du théorème de la limite centrale pour une somme de variables aléatoires indépendantes. Nous en rappelons ici l'énoncé. Soit

$$X_N = \frac{1}{N} \sum_{i=1}^N x_i \quad (32)$$

la moyenne d'une somme de variables aléatoires indépendantes, décrites toutes par la même loi de probabilité $\Pr(x)$ pour laquelle la moyenne $\langle x \rangle$ et la variance σ^2 sont tous deux finis. Le théorème de la limite centrale indique alors que :

$$\lim_{N \rightarrow \infty} \Pr(a \leq X_N \leq b) = \frac{1}{\sqrt{2\pi} \frac{\sigma}{\sqrt{N}}} \int_a^b du \exp\left[-\frac{(u - \langle x \rangle)^2}{2 \frac{\sigma^2}{N}}\right]. \quad (33)$$

2.1.5 Estimateur de la variance de la distribution de probabilité

Le résultat démontré ci-dessus est remarquable car il nous dispense de connaître la forme exacte de la distribution de probabilité associée aux variables aléatoires initiales. Son application pose cependant deux problèmes :

- Le résultat dépend de la dimension N de la somme ;
- Pour l'appliquer, il faut connaître d'une part la moyenne et d'autre part la variance de la distribution de probabilité initiale.

Nous avons vu comment la donnée d'un échantillon de mesures $\{x_1, \dots, x_N\}$ nous permettait d'estimer la moyenne $\langle x \rangle$ grâce à la *moyenne d'échantillon* m . Nous pouvons chercher de la même manière quel est le meilleur estimateur σ^* de l'écart-type σ . Supposons à nouveau que

¹²Les "puristes" auront certainement trouvé cette "démonstration" insuffisamment rigoureuse..

l'échantillon de mesures soit issu d'une distribution gaussienne sous-jacente. Pour déterminer l'estimateur σ^* , nous allons dériver l'équation (15) par rapport au paramètre σ :

$$\frac{\partial W}{\partial \sigma} = -\frac{N}{\sigma} + \sum_{i=1}^N \left[\left(\frac{\mu - x_i}{\sigma} \right) \left(\frac{\mu - x_i}{\sigma^2} \right) \right],$$

de sorte que :

$$\frac{\partial W}{\partial \sigma} = 0 \quad \longrightarrow \quad (\sigma^*)^2 = \frac{1}{N} \sum_{i=1}^N (\mu - x_i)^2. \quad (34)$$

Comme nous ne connaissons pas la valeur moyenne "vraie" μ , le résultat de l'équation (34) est cependant inutilisable ! En utilisant la moyenne d'échantillon m , nous pouvons penser le remplacer par l'estimateur :

$$S^2 = \frac{1}{N} \sum_{i=1}^N (m - x_i)^2. \quad (35)$$

Nous devons cependant avoir une exigence dans le choix de l'estimateur. Puisque la valeur de l'estimateur peut fluctuer d'une réalisation de l'échantillon à l'autre, une qualité que nous devons imposer est que l'ensemble de toutes les estimations soit en moyenne égal à la valeur exacte du paramètre correspondant de la distribution de probabilité. Ainsi, la relation $E(m) = \mu$ nous indique que la moyenne d'échantillon m constitue bien un estimateur sans biais de la moyenne μ de la distribution de probabilité. À l'inverse, le choix de S^2 comme estimateur de la variance σ^2 introduit un biais, puisque $E(S^2)$ n'est pas exactement égal à σ^2 . On montre en effet (cf. Annexe ??) que :

$$E(S^2) = \frac{N-1}{N} \sigma^2.$$

Par conséquent, pour que l'estimateur s^2 de la variance σ^2 soit sans biais, il faut ajuster S^2 par le facteur $(N-1)/N$, soit :

$$\frac{N}{N-1} S^2 = \frac{N}{N-1} \times \frac{1}{N} \sum_{i=1}^N (m - x_i)^2.$$

L'estimateur *sans biais* de la variance σ^2 s'écrit ainsi :

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (m - x_i)^2. \quad (36)$$

La grandeur s définie par l'équation (36) est appelée *écart-type d'échantillon*. Elle correspond à une caractérisation intrinsèque de la dispersion sur notre échantillon $\{x_1, \dots, x_N\}$ de mesures. Insistons sur le fait qu'elle est de nature fondamentalement différente de l'écart-type σ , puisque ce paramètre se réfère au moment d'ordre deux de la loi de probabilité associée à la mesure de la grandeur x .

Notons que l'apparition du facteur $N-1$ dans l'expression (36) est cohérente avec le fait qu'une seule mesure est insuffisante pour espérer pouvoir estimer l'écart-type de la distribution de probabilité associée à la mesure. Nous voyons ainsi apparaître la notion de "degrés de liberté" pour l'obtention des estimateurs des moments de la distribution de probabilité sous-jacente à la mesure :

- L'estimateur de la valeur moyenne μ est fondée sur les N points de mesure, c'est-à-dire N degrés de liberté ;
- L'estimateur de la variance σ^2 est obtenu avec seulement $N-1$ degrés de liberté, puisque nous avons dû dans un premier temps estimer la valeur moyenne de la distribution.

Nous pouvons également remarquer que les deux expressions

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (m - x_i)^2 \quad \text{et} \quad (\sigma^*)^2 = \frac{1}{N} \sum_{i=1}^N (\mu - x_i)^2$$

présentent aucune différence notable pour $N \gg 1$, puisque $\sqrt{N-1} \simeq \sqrt{N}$ et que la moyenne d'échantillon m va converger vers μ pour cette limite des grands nombres.

2.1.6 Précision sur la détermination de l'estimateur de la variance

Nous devons également être conscients du fait que la valeur de l'écart-type d'échantillon s , évaluée selon l'expression (36) va fluctuer d'un échantillon de mesures $\{x_1, \dots, x_N\}$ à l'autre. Pour évaluer l'écart-type associé à cette dispersion, nous pouvons partir de l'équation (22), avec :

$$\frac{\partial^2 W}{\partial \sigma^2} = \frac{N}{\sigma^2} - \frac{3}{\sigma^4} \sum_{i=1}^N (\mu - x_i)^2 = \frac{N}{\sigma^2} \left[1 - 3 \left(\frac{\sigma^*}{\sigma} \right)^2 \right] \rightarrow - \left(\frac{\partial^2 W}{\partial \sigma^2} \right)_{\sigma=\sigma^*} = \frac{2N}{(\sigma^*)^2}. \quad (37)$$

En remplaçant N par $N-1$, nous en déduisons immédiatement la variance $V(s)$ associée à la détermination de l'écart-type d'échantillon s :

$$V(s) = \frac{s^2}{2(N-1)} \quad (38)$$

Comme représenté sur la figure 5, ce résultat montre que l'écart-type d'échantillon s est connu avec une précision relative

$$\frac{\Delta s}{s} \simeq \frac{1}{\sqrt{2(N-1)}}. \quad (39)$$

Si la forme de la dépendance par rapport à N ne nous étonne pas, nous voyons que pour $N \simeq 10$ mesures, l'estimation s de l'écart-type σ de la distribution de probabilité est effectuée avec une incertitude de l'ordre de 25%. Pour un échantillon de telle "petite" taille, il est par conséquent inutile de chercher à attribuer une valeur parfaitement définie à l'écart-type d'échantillon, tel qu'on peut le calculer à l'aide de la formule (36).

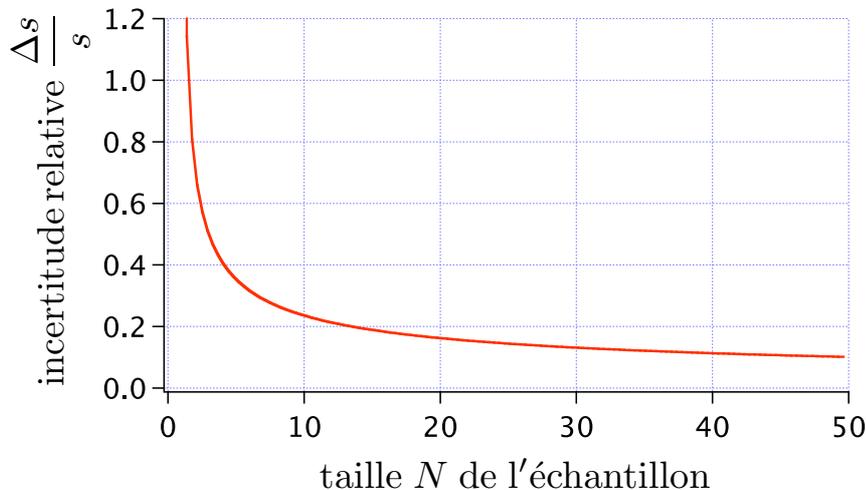


Figure 5: Précision relative sur l'écart-type d'échantillon s en fonction de la taille N de l'échantillon considéré. Pour 5 à 6 mesures, cette précision est de l'ordre de 30%. Il est nécessaire d'effectuer plus de 30 mesures pour que celle-ci devienne inférieure à 20%.

2.1.7 Cas d'un petit nombre de mesures : fonction de Student

Nous venons d'établir que l'écart-type d'échantillon s est un estimateur sans biais de l'écart-type σ de la distribution de probabilité associée à la mesure. Cette approximation est d'autant meilleure que le nombre N de points de mesure dans l'échantillon est grand et dès lors que $N \geq 20$, il n'y a plus de différence significative entre s et σ .

Dans le cas d'un échantillon de petite taille pour lequel N est par exemple inférieur à 10, l'analyse est cependant plus compliquée. Supposons en effet que l'une des valeurs x_k dans l'ensemble des mesures $\{x_1, \dots, x_N\}$ soit anormalement élevée, sa valeur étant a priori relativement improbable compte tenu de la distribution de probabilité. Une telle mesure "anormale" va alors fortement influencer d'une part la moyenne d'échantillon m , prise comme estimateur de la mesure "vraie" x_{vraie} , et d'autre part la valeur de l'écart-type d'échantillon s . Dans ces conditions, la probabilité $\mathcal{P}_{0.95}$ que l'intervalle de confiance $[m - 2s, m + 2s]$ contienne la valeur x_{vraie} sera inférieure à 95 %. Il est par conséquent nécessaire de revoir la procédure d'exploitation de l'échantillon de mesures, en cherchant une distribution dont la forme soit effectivement liée à la loi de Gauss mais qui soit cependant mieux adaptée à la situation considérée d'un échantillon de petite taille.

Notre objectif consiste plus précisément à obtenir un intervalle de confiance pour le résultat de la mesure, en utilisant les valeurs m et s des estimateurs de la moyenne et de l'écart-type. Nous cherchons donc la quantité e telle que :

$$\Pr(|m - \mu| \leq e) = 1 - \alpha$$

soit :

$$\Pr\left(\frac{|m - \mu|}{\frac{s}{\sqrt{N}}} \leq e'\right) = 1 - \alpha \quad \text{avec} \quad e' = \frac{e}{\frac{s}{\sqrt{N}}}.$$

Étant donné que le numérateur $m - \mu$ et le dénominateur $\frac{s}{\sqrt{N}}$ sont tous deux aléatoires, la distribution de probabilité de la variable réduite

$$t = \frac{x - m}{\frac{s}{\sqrt{N}}} \tag{40}$$

est celle d'un ratio entre deux variables aléatoires. Cette distribution de probabilité n'est plus une fonction normale comme cela est le cas quand la variance σ^2 – le dénominateur du ratio – est un paramètre fixé et connu. On montre que la variable réduite t suit une distribution appelée "loi de Student" à $N - 1$ degrés de liberté et définie par :

$$f_{N-1}(t) = \frac{C}{\left(1 + \frac{t^2}{N-1}\right)^{\frac{N}{2}}} \quad \text{avec} \quad C = \frac{\Gamma(\frac{N}{2})}{\sqrt{\pi(N-1)} \Gamma(\frac{N-1}{2})} \tag{41}$$

où la fonction $\Gamma(x)$ est l'intégrale d'Euler de deuxième espèce ¹³.

Comme représenté sur la figure 6, cette fonction a une structure relativement simple. Pour $N = 2$, c'est-à-dire l'exploitation de seulement deux points de mesure, la fonction de Student se confond avec une lorentzienne $1/\pi(1 + x^2)$. Lorsque N devient très grand, la distribution de Student tend asymptotiquement vers une distribution de Gauss normalisée $1/\sqrt{2\pi} \times \exp\left(-\frac{x^2}{2}\right)$.

Dans l'hypothèse où toute erreur systématique a été écartée et où les N mesures de l'échantillon $\{x_1, \dots, x_N\}$ sont tirées d'une distribution gaussienne plus vaste, l'intervalle de confiance de la

¹³Cette fonction spéciale constitue une généralisation de la fonction factorielle au cas d'un argument non-entier, avec $\Gamma(N + 1) = N!$ pour N entier.

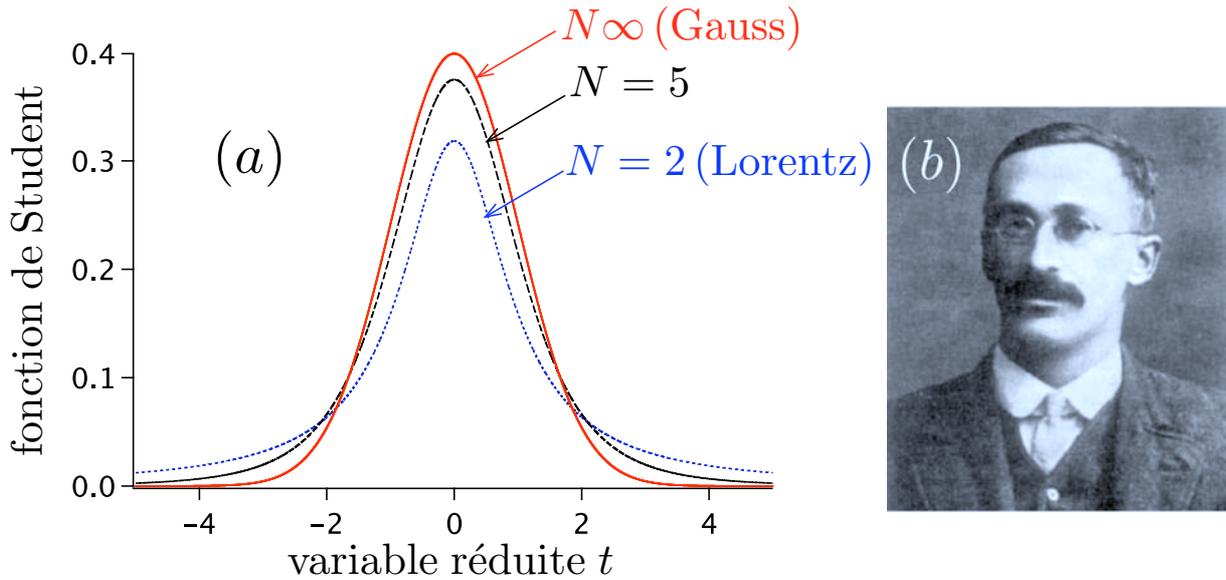


Figure 6: (a) – Comparaison de la distribution t de Student pour différentes valeurs du nombre N dans l'échantillon de mesures. Lorsque $N \gg 1$, ces courbes tendent asymptotiquement vers une distribution de Gauss. On notera que pour un échantillon contenant un nombre fini de valeurs, l'utilisation de la distribution de Gauss conduit à *sous-évaluer* la probabilité d'obtenir des écarts importants, qui seraient par exemple supérieurs à 3σ , par rapport à la valeur moyenne. (b) – On doit ces lois à W. S. GOSSET (1876–1937) qui était à l'origine employé par les célèbres brasseries Guinness afin de mettre au point des techniques de contrôle de qualité, lesquelles sont aujourd'hui d'une utilisation courante dans tout le monde industriel. Il publia ses travaux en 1908 sous le pseudonyme de STUDENT, car la société GUINNESS souhaitait préserver le secret sur les travaux d'exploitation statistique, pensant qu'ils étaient la clé d'un avantage industriel par rapport à des sociétés concurrentes !

grandeur x qui correspond à un niveau de confiance α donné (par exemple 95 %) est alors de la forme :

$$\left[\mu - t_\alpha \times \frac{s}{\sqrt{N}} ; \mu + t_\alpha \times \frac{s}{\sqrt{N}} \right], \quad (42)$$

Cette expression fait intervenir le coefficient de Student t_α . Ce coefficient est obtenu par intégration de la distribution de Student pour un échantillon de taille N , correspondant à $N_d = N - 1$ degrés de liberté.

Si l'on trouve dans tous les traités de statistique des tableaux donnant les valeurs numériques de ces coefficients en fonction du niveau de confiance α et du nombre N de mesures, le logiciel IGOR propose deux fonctions utiles qui sont directement liées à ces distributions de Student :

- **StudentA(t, DegFree)** donne l'aire de la distribution de Student correspondant à un nombre $N_d = \text{DegFree}$ de degrés de liberté, comprise entre les valeurs $-t$ et $+t$;
- **StudentT(Prob, DegFree)** retourne la valeur de t qui correspond à une aire **Prob** en-dessous de la distribution de Student à $N_d = \text{DegFree}$ degrés de liberté.

2.2 Exemple concret d'analyse d'une série de mesures

2.2.1 Position du problème

Considérons 10 mesures successives d'une tension, effectuées à l'aide d'un même voltmètre¹⁴.

N° de la mesure	1	2	3	4	5	6	7	8	9	10
Valeur lue (en V)	5.11	5.10	5.08	5.08	5.07	5.10	5.10	5.10	5.10	5.11

Si nous examinons les résultats obtenus, dont les valeurs sont indiquées dans le tableau ci-dessus, nous constatons tout d'abord qu'apparemment, aucune d'entre elles ne s'écarte de manière significative des autres. Par conséquent, nous ne pouvons déceler aucune erreur de manipulation dans la prise de ces données. Quel est alors le résultat de la mesure et comment l'exprimer ?

2.2.2 Premier niveau d'analyse

Il peut sembler logique de supposer que la vraie valeur de la tension se trouve comprise entre la valeur maximale $U_{\max} = 5.11 \text{ V}$ et la valeur minimale $U_{\min} = 5.07 \text{ V}$ de la série de mesure. On trouvera parfois écrit, ou affirmé lors de la présentation d'un montage, que l'écart entre ces deux valeurs conduit à une estimation de l'erreur ΔU , leur moyenne donnant une valeur "probable" de la grandeur mesurée :

$$U_{\text{probable}} = \frac{U_{\max} + U_{\min}}{2} = \quad \text{et} \quad \Delta U = \frac{U_{\max} - U_{\min}}{2},$$

ce qui revient à donner le résultat de la mesure sous la forme $U = (5.09 \pm 0.02) \text{ V}$.

Cette analyse est évidemment simple et rapide. Nous espérons cependant avoir convaincu le lecteur qu'elle est sans fondement, et que le résultat ainsi énoncé n'a aucune valeur scientifique. En particulier, si on venait à effectuer une nouvelle mesure de cette tension, rien n'interdirait d'obtenir un résultat dont la valeur soit située nettement en dehors de l'intervalle écrit ci-dessus. Même si ce type d'analyse donne pratiquement des résultats "acceptables", il faut une fois pour toutes y renoncer et lui préférer l'exploitation statistique de la série de mesures.

2.2.3 Deuxième niveau d'analyse : étude statistique

Nous allons maintenant voir comment une analyse statistique de cet échantillon de mesures peut être effectuée simplement à l'aide du logiciel IGOR. Après avoir créé la wave `tension` contenant les dix valeurs mesurées, nous allons tout d'abord nous faire une idée sur la répartition de ces valeurs. Allons pour cela dans le menu `Analysis → Wave Stats...`, ou tapons directement la commande correspondante `WaveStats Tension` dans la ligne de commande. Nous voyons alors s'imprimer dans la fenêtre d'histoire les renseignements suivants :

```
V_npnts= 10; V_numNaNs= 0; V_numINFs= 0; V_avg= 5.095;
V_sdev= 0.0135401; V_rms= 5.09502; V_adev= 0.011;
V_skew= -0.604267; V_kurt= -1.22603; V_minloc= 4; V_min= 5.07;
V_maxloc= 0; V_max= 5.11;
```

Parmi ces différentes variables dont on trouvera la signification dans la documentation du logiciel¹⁵, `V_avg` et `V_sdev` donnent respectivement la valeur moyenne d'échantillon $m = 5.095$ et l'écart-type d'échantillon $s = 0.0135401$.

¹⁴Cet exemple est tiré du polycopié de TP d'André GALAIS sur le traitement et l'analyse des données.

¹⁵IGOR PRO Version 4.0, Vol. II *User's Guide* et Vol. III *Programming and Reference Manual* (Wavemetrics, 2000).

Nous allons ainsi pouvoir exprimer le résultat de ces mesures sous la forme d'un intervalle de confiance, en prenant en compte les corrections de Student. Pour les $N - 1 = 9$ degrés de liberté de l'échantillon, la valeur de t_α correspondant à un niveau de confiance de $\alpha = 0.95$ peut être obtenue grâce aux tables. Nous pouvons également utiliser à nouveau le logiciel IGOR et sa fonction pré-définie `StudentT` :

```
print StudentT(0.95,9)
2.26214
```

de sorte que :

$$t_\alpha \times \frac{s}{\sqrt{N}} = 2.26214 \times \frac{0.0135401}{\sqrt{10}} = 0.009686.$$

Nous avons maintenant tous les éléments pour exprimer le résultat sous la forme :

$$U = m \pm \Delta U \quad \text{soit} \quad U = (5.095 \pm 0.010) \text{ V}$$

Compte tenu de l'arrondi, cet intervalle correspond à un niveau de confiance légèrement supérieur à 95 % pour les 10 mesures effectuées. Le résultat précédent peut également être écrit sous la forme

$$U = 5.095(10) \text{ V}.$$

Cette convention de notation est en particulier utilisée pour l'écriture des valeurs numériques des constantes fondamentales, car elle conduit à une lecture plus agréable lorsque la valeur numérique comporte un grand nombre de chiffres significatifs.

Remarquons que si nous n'avions pas tenu compte du facteur correctif introduit par la distribution de Student, nous nous serions contentés de donner l'incertitude sur la moyenne des 10 mesures pour un intervalle de probabilité de 95 % comme

$$2 \times \frac{s}{\sqrt{N}} \simeq 0.0086 \text{ V}.$$

Comme nous avons montré précédemment que l'évaluation de l'incertitude est effectuée à environ 25 % près, nous pouvons tout aussi bien lui donner une valeur comprise entre 0.0064 V et 0.0107 V. L'intervalle de confiance déterminé à partir du coefficient de Student recouvre ces valeurs et c'est par conséquent celui que nous conserverons pour la suite. Enfin, nous pouvons noter que le premier niveau d'analyse, qui consiste à prendre les valeurs extrêmes de la série de mesures comme bornes d'un intervalle fictif qui contiendrait la "vraie" valeur conduit à une surestimation de l'incertitude sur la mesure, nettement supérieure à l'incertitude effective Δs sur l'écart-type.

2.2.4 Comparaison avec la distribution gaussienne estimée

Si cette analyse est rigoureuse et donne un résultat final quantitatif, on peut cependant lui reprocher de supposer implicitement que les mesures de la tension U sont distribuées suivant une loi normale de Gauss. Nous pouvons par conséquent nous interroger sur la justesse de cette hypothèse, et chercher des critères quantitatifs qui nous permettraient de la tester. Une première méthode qualitative peut être de simplement tracer l'histogramme des points de mesure en choisissant un pas ¹⁶ approprié, par exemple $\delta V_{\text{bin}} = 0.01 \text{ V}$).

Commençons pour cela par créer une wave `tension_histo` dans le logiciel IGOR PRO, puis sélectionnons r le menu `Analysis > Histogram`. Dans la fenêtre qui apparaît alors, précisons r la wave de départ (`tension`), la wave de destination (`tension_histo`), le bin start (5), la largeur du pas (0.01) ainsi que le nombre de pas (20). La taille et l'échelle de la wave (`tension_histo`) sont alors automatiquement calculées. Nous pouvons représenter la wave `tension_histo` sous la

¹⁶En anglais, *bin*.

forme d'un histogramme, en sélectionnant le mode `Bars to next` à la place de `Line between points` pour l'affichage de la courbe (figure 7).

Nous pouvons ensuite superposer à cet histogramme la distribution de probabilité déterminée à partir de l'échantillon de mesures. La probabilité d'obtenir une valeur U_{lue} de la tension, avec le pas δV_{bin} choisi, va alors s'écrire :

$$\begin{aligned} \Pr(U_{\text{lue}}) &= \frac{1}{s\sqrt{2}\pi} \exp\left[-\frac{(U_{\text{lue}} - m)^2}{2s^2}\right] \times \delta V_{\text{bin}} \\ &= \frac{1}{0.0135 \times \sqrt{2}\pi} \exp\left[-\frac{(U_{\text{lue}}(\text{V}) - 5.095)^2}{2 \times (0.0135)^2}\right] \times 0.01. \end{aligned} \quad (43)$$

L'histogramme des valeurs de l'échantillon et la loi de Gauss correspondante $N \times \Pr(U_{\text{lue}})$, où le facteur N correspond au nombre total de points de mesure dans l'échantillon, sont représentés sur la figure 7. Compte tenu de la faible statistique, l'échantillon contenant seulement 10 mesures, il est évidemment difficile d'exploiter cette simple représentation pour se convaincre ou non de la validité de l'hypothèse d'une répartition des mesures effectuées selon la loi normale de Gauss. Le test du χ^2 , dont on trouvera la description dans tout ouvrage de statistique, permet de rendre ces considérations plus quantitatives.

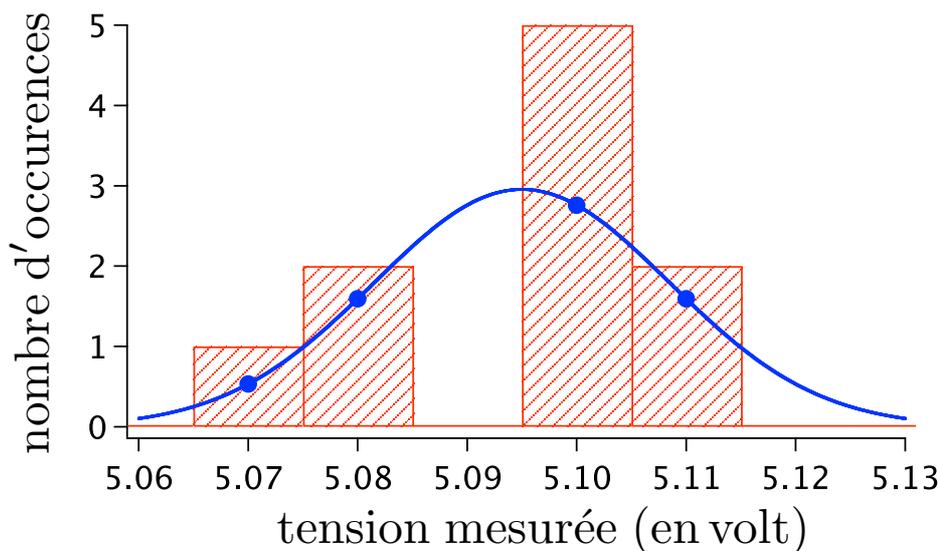


Figure 7: Histogramme des mesures de la tension, tracé pour un pas d'échantillonnage de 0.01 V, et représentation en pointillé de la distribution de probabilité associée. Cette courbe correspond à une gaussienne avec comme paramètres les valeurs des estimateurs m et s obtenus à partir de l'exploitation de l'échantillon de mesures. Les points (●) correspondent aux occurrences pour chaque valeur de la tension obtenue dans l'échantillon considéré.

2.2.5 Vocabulaire de métrologie : incertitudes de type A et de type B

L'incertitude ΔU que nous venons d'évaluer est celle sur la **moyenne** m des dix mesures. Elle ne correspond pas à celle qui affecte une mesure **unique** de la tension, laquelle est caractérisée par l'écart-type σ dont s est un estimateur sans biais statistique. Ainsi, l'intervalle de confiance de 95 % pour une mesure unique de la tension U sera donné comme $[m - 2s; m + 2s]$, soit $[5.081 \text{ V}; 5.109 \text{ V}]$.

Dans le langage des métrologues, nous venons d'utiliser une *méthode d'évaluation de type A* de l'incertitude¹⁷. Cette dénomination correspond par définition à une technique d'évaluation de l'incertitude à partir d'un procédé statistique qui prend en compte la dispersion des résultats observée lorsqu'on répète à l'identique une mesure.

Les métrologues définissent également les *méthodes d'évaluation de type B*, lesquelles correspondent a contrario aux techniques d'évaluation de l'incertitude qui ne sont pas fondées sur une procédure d'exploitation statistique d'un échantillon de mesures. L'évaluation de ce type d'incertitude s'appuie en général sur un jugement scientifique élaboré à partir d'une analyse physique de l'expérience. Il faut en particulier cerner la part d'incertitude qui est due à chacune des corrections associées aux différentes causes d'erreur ayant été identifiées. On applique le même formalisme que pour les méthodes de type A, à savoir évaluer l'incertitude par une valeur ΔU_k qu'on peut traiter comme un écart-type. Cette valeur est déterminée à partir d'éléments qui peuvent provenir d'une bibliographie sur les matériaux et les capteurs utilisés dans l'expérience, de notices des constructeurs, d'éventuels certificats d'étalonnage, etc. mais également à partir des connaissances et du savoir-faire de l'expérimentateur. Pour obtenir ensuite l'incertitude $\Delta U^{(B)}$ sur la correction totale, on applique la loi de composition des variances pour la somme de variables aléatoires indépendantes. Dans l'hypothèse où causes d'erreur et corrections ne sont pas corrélées, celle-ci s'écrira :

$$\left(\Delta U^{(B)}\right)^2 = \sum_k (\Delta U_k)^2 . \quad (44)$$

On peut également convenir de donner une incertitude globale en introduisant un facteur multiplicatif conventionnel $\times 2$, lequel correspond pour une loi normale de Gauss à la donnée d'un intervalle à risque d'erreur inférieur à 5 %.

L'incertitude composée ΔU_{tot} sur la mesure va tenir compte des incertitudes $\Delta U^{(A)}$ et $\Delta U^{(B)}$ évaluées respectivement par les méthodes de type A et B. Elle est obtenue par composition des variances :

$$\left(\Delta U_{\text{tot}}\right)^2 = \left(\Delta U^{(A)}\right)^2 + \left(\Delta U^{(B)}\right)^2 \quad \text{soit} \quad \left(\Delta U_{\text{tot}}\right)^2 = t_\alpha^2 \times \frac{s^2}{N} + \left(\Delta U^{(B)}\right)^2 . \quad (45)$$

L'incertitude finale sur le résultat sera ensuite donnée par la racine carrée de cette variance. Cette expression montre également que l'incertitude de type B affecte de la même manière une mesure unique et la valeur moyenne d'un échantillon de mesures. Contrairement à l'incertitude de type A, ce type d'incertitude ne peut être réduit par le moyennage d'une série de mesures répétées. Tout l'art du physicien "métrologue" sera alors de chercher à identifier le mieux possible les différentes causes d'erreur qui affectent la mesure qu'il réalise.

2.2.6 Budget d'incertitude de la mesure

Les métrologues ont pour coutume de présenter les deux types de contributions sous la forme d'un *budget d'incertitude*. Nous allons illustrer ce concept à travers l'exemple des dix mesures de la tension. Celles-ci ont été effectuées au moyen d'un multimètre numérique TEKTRONIX DMM 912 dont nous disposons au laboratoire. Nous avons évalué la dispersion associée à la répétabilité des mesures, représentée par une incertitude

$$\Delta U^{(A)} = 9.2 \text{ mV} .$$

¹⁷B. N. TAYLOR et C. E. KUYATT, *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, NIST Technical Note 1297 (1994). Ce document peut être téléchargé directement à l'adresse [web du nist](http://www.nist.gov).

L'incertitude de type B sera quant-à elle due à l'étalonnage du voltmètre et à la limitation imposée par la résolution de l'appareil numérique, soit :

$$\Delta U^{(B)} = 2 \times (0.2\% \text{ VL} + 1 \text{ UL})$$

où VL correspond à la valeur lue (5.10 V) et UL à une unité du dernier chiffre exprimant la mesure. Le multimètre ayant 2000 points et étant utilisé sur le calibre 10 V, UL correspond à 5 mV. On obtient ainsi un budget d'incertitude qu'on peut représenter par le tableau ci-dessous :

<i>Composantes évaluées par l'analyse statistique</i>	<i>en V</i>
Exploitation de 10 mesures répétées de la tension	
$\Delta U^{(A)}$ (incertitude de type A sur la moyenne des 10 mesures)	0.009 7
<i>Composantes évaluées par d'autres moyens</i>	<i>en V</i>
Étalonnage du multimètre (0.2 % de la valeur lue)	0.010 2
Résolution du multimètre (1 unité de lecture)	0.005 0
$\Delta U^{(B)}$ (incertitude de type B élargie à 2σ)	0.030 4
<i>Incertitude totale</i>	<i>en V</i>
$\Delta U_{\text{tot}} = \sqrt{(\Delta U^{(A)})^2 + (\Delta U^{(B)})^2}$ (valeur élargie à 2σ)	0.032 1

Le résultat final de la mesure, qui prend en compte les deux types d'incertitudes, sera ainsi donné comme :

$$U = (5.095 \pm 0.032) \text{ V} .$$

À la lumière du budget d'incertitude, nous avons montré que la précision de la mesure était essentiellement fixée par la calibration du multimètre... On se convaincra du fait qu'avec un peu d'entraînement et l'habitude d'un travail méthodique, une telle analyse peut être effectuée très rapidement.

3 Ajustement de courbes expérimentales

3.1 Données et moindres carrés

3.1.1 Position du problème

Nous avons vu dans la partie précédente comment il est possible d'exploiter une série de mesures correspondant à la détermination *directe* d'une grandeur comme la tension, l'intensité, la position d'un objet, etc. Dans son travail, le physicien expérimentateur est très souvent confronté à un problème de nature un peu différente¹⁸. Lorsqu'on réalise une expérience, on est amené à mesurer une grandeur y en fonction d'une autre grandeur x , soit qu'il s'agisse d'établir une loi empirique, soit qu'il s'agisse d'utiliser une loi connue pour déterminer une grandeur inconnue. Ainsi, on peut chercher à établir la relation entre la pression et le volume d'un gaz à température constante, déterminer comment varie le gain électronique d'un amplificateur en fonction de la fréquence du signal d'entrée, etc.

On peut également posséder une théorie prédisant la loi $y(x)$ et chercher à déterminer si cette théorie est juste. C'est bien là l'essence même d'une modélisation d'expérience, où on cherche à faire entrer les mesures dans le cadre théorique et prédictif d'un modèle, et à valider les hypothèses physiques sur lesquelles il repose. En général, la loi issue de la théorie va également dépendre de paramètres dits "ajustables", dont on ne connaît pas la valeur a priori et qu'on va tenter d'ajuster afin d'obtenir le meilleur accord possible avec les données expérimentales. On est ainsi amené à poser *deux questions distinctes* :

¹⁸Cette partie est très largement inspirée du polycopié de cours de Philippe DEPONDT *Physique numérique* (Licence Phytém, 2004), dont nous recommandons vivement la lecture.

- (1) Est-ce que la théorie marche ?
- (2) Quelles sont les valeurs des paramètres ajustables, et les incertitudes associées ?

Compte tenu de l'importance de ce problème, nul ne doutera de l'intérêt de regarder d'un peu plus près comment un logiciel comme IGOR Pro peut être utilisé pour répondre à ces questions, et quels sont les algorithmes qu'il utilise. Nous pouvons également nous demander quelles sont les hypothèses sous-jacentes, et surtout quelles sont les éventuelles limites de ce type d'exploitation.

3.1.2 Hypothèse de départ

En pratique, l'expérimentateur relève, pour N points de mesure $\{x_1, \dots, x_N\}$, les N valeurs $\{y_1, \dots, y_N\}$ correspondantes. Supposons que l'on dispose d'une loi théorique qui donne l'évolution de la grandeur y en fonction de x . Les valeurs correspondantes, que nous noterons \hat{y} sont supposées être fonction de la variable x et d'un certain nombre de paramètres $\{\alpha_1, \dots, \alpha_m\}$. Nous écrirons par conséquent :

$$\hat{y}_i = \hat{y}(x | \alpha_1, \dots, \alpha_m) = \hat{y}(x | \{\alpha_j\}) \quad (46)$$

Nous supposons par ailleurs que chaque résultat mesure \bar{y}_i est affecté d'une incertitude caractérisée par un écart-type σ_i , comme représenté sur la figure 8

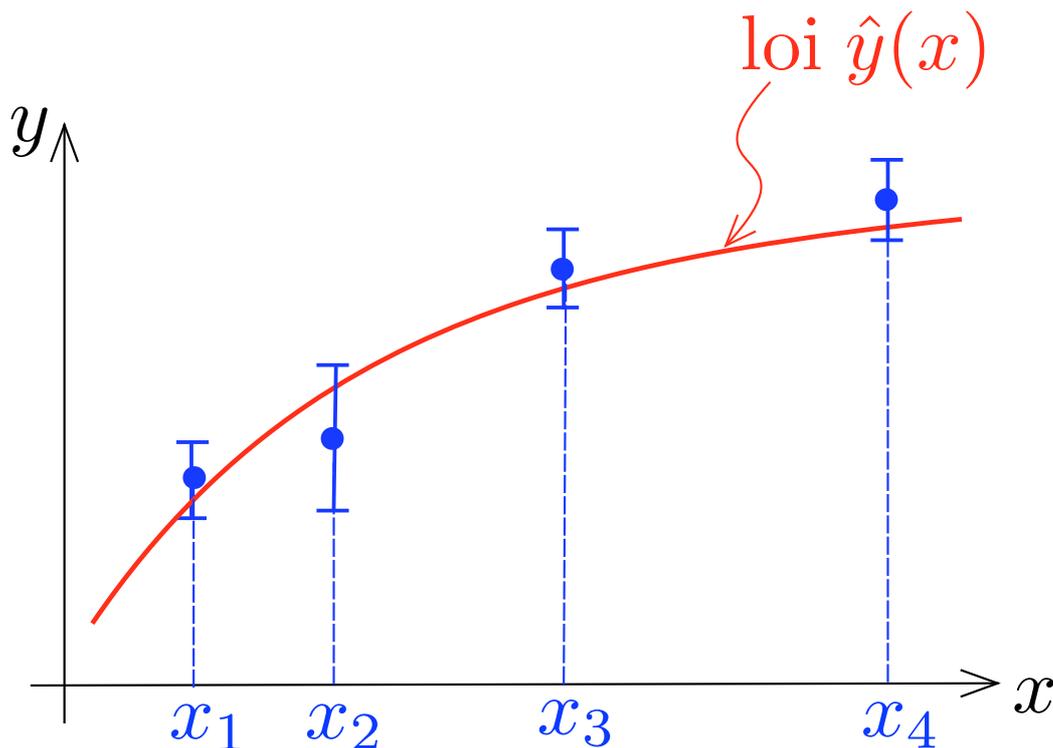


Figure 8: Principe de l'ajustement d'une loi théorique $\hat{y}(x)$ à un ensemble de données de mesure $\{(x_1, y_1), (x_2, y_2), \dots\}$. Pour chaque valeur x_i , la mesure correspondante y_i est effectuée avec une incertitude Δy_i . Cette précision sur les mesures est représentée par la barre d'erreur sur le graphique $y(x)$.

Ce qui nous intéresse, ce sont les écarts entre les y_i mesurés et les valeurs \hat{y}_i calculées à l'aide de la loi théorique pour les différentes valeurs x_i . On définit ainsi la quantité "chi-deux" comme :

$$\chi^2(\alpha_1, \dots, \alpha_m) = \sum_{i=1}^N \frac{1}{\sigma_i^2} \times [y_i - \hat{y}(x_i | \{\alpha_1, \dots, \alpha_m\})]^2 . \quad (47)$$

Nous cherchons alors l'ensemble des paramètres $\{\alpha_1, \dots, \alpha_m\}$ qui rendra le coefficient χ^2 le plus petit possible. C'est pour cette raison que la méthode porte le nom d'“ajustement par moindres carrés” puisque χ^2 est défini comme la somme (pondérée par l'incertitude) des carrés des écarts entre les mesures et les valeurs théoriques calculées.

Remarquons que nous aurions pu choisir autre chose que la somme des carrés des écarts, par exemple la somme des valeurs absolues ou la borne supérieure des carrés. La justification du choix que nous avons fait repose en fait sur l'hypothèse que chaque mesure individuelle y_i est tirée d'une distribution gaussienne sous-jacente dont la valeur moyenne est donnée par la valeur vraie de y_i , $\hat{y}(x_i | \{\alpha_1, \dots, \alpha_m\})$. Pour l'écart-type de cette loi de Gauss, nous prendrons naturellement l'écart-type σ_i associé à l'incertitude de mesure sur chacun des y_i . La distribution de probabilité correspondante va alors s'écrire

$$\Pr(y_i | x_i | \{\alpha_1, \dots, \alpha_m\}) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \frac{[y_i - \hat{y}(x_i | \{\alpha_1, \dots, \alpha_m\})]^2}{\sigma_i^2} \right\}. \quad (48)$$

Par analogie avec l'équation (12), la fonction de vraisemblance va s'écrire

$$\mathcal{L}(y_1 \cdots y_N | x_1 \cdots x_N | \{\alpha_j\}) = \prod_{i=1}^N \Pr(y_i | x_i | \{\alpha_1, \dots, \alpha_m\}) \quad (49)$$

et nous cherchons les estimateurs $\{\alpha_1^*, \dots, \alpha_m^*\}$ qui rendent maximale cette fonction de vraisemblance, ou plutôt de manière équivalente son logarithme $W = \ln \mathcal{L}$. Compte tenu de la définition de \mathcal{L} , celui-ci va s'écrire :

$$W = -\sum_{i=1}^N \ln(\sigma_i \sqrt{2\pi}) - \frac{1}{2} \sum_{i=1}^N \frac{[\hat{y}(x_i | \{\alpha_1, \dots, \alpha_m\}) - y_i]^2}{\sigma_i^2}. \quad (50)$$

Puisque les valeurs des σ_i sont fixées par la méthode de mesure, les estimateurs $\{\alpha_1^*, \dots, \alpha_m^*\}$ vont correspondre aux valeurs des paramètres $\{\alpha_1, \dots, \alpha_m\}$ qui rendent minimale la somme des carrés des écarts $\chi^2(\alpha_1, \dots, \alpha_m)$ définie par l'équation (47).

3.2 Ajustement par une fonction affine

3.2.1 Ajustement sans prendre en compte d'incertitudes

Considérons pour commencer le cas le plus simple d'une dépendance affine $\hat{y} = ax + b$, où les deux paramètres ajustables sont les coefficients a et b . Nous supposons également que chaque mesure a le même poids statistique, de sorte que nous cherchons a et b tels que

$$\chi^2(a, b) = \sum_{i=1}^N (ax_i + b - y_i)^2 \quad (51)$$

prenne une valeur minimale. Il faut pour cela que les dérivées de χ^2 par rapport aux deux paramètres a et b soient nulles :

$$\begin{cases} \frac{\partial}{\partial a} [\chi^2(a, b)] &= \sum_{i=1}^N 2x_i(ax_i + b - y_i) &= 0 \\ \frac{\partial}{\partial b} [\chi^2(a, b)] &= \sum_{i=1}^N 2(ax_i + b - y_i) &= 0 \end{cases} \quad (52)$$

de sorte que les valeurs de a et b sont obtenues par la résolution d'un système linéaire de deux équations à deux inconnues :

$$\begin{cases} a \sum_{i=1}^N x_i^2 + b \sum_{i=1}^N x_i &= \sum_{i=1}^N x_i y_i \\ a \sum_{i=1}^N x_i + bN &= \sum_{i=1}^N y_i \end{cases}. \quad (53)$$

Notons que la dernière équation montre que la fonction obtenue par cet ajustement passe nécessairement par le point dont les coordonnées suivant x et y correspondent respectivement aux moyennes \bar{x} et \bar{y} des échantillons $\{x_1 \cdots x_N\}$ et $\{y_1 \cdots y_N\}$:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{et} \quad \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i . \quad (54)$$

Quelques lignes de calcul suffisent ensuite à déterminer les valeurs de a et b :

$$a = \frac{\sum_{i=1}^N x_i (y_i - \bar{y})}{\sum_{i=1}^N x_i (x_i - \bar{x})} \quad \text{et} \quad b = \bar{y} - a\bar{x} . \quad (55)$$

Lire dans un fichier ou dans une mémoire les N valeurs de x et y , puis effectuer ces calculs pour obtenir la droite d'ajustement par moindres carrés est donc extrêmement simple. On comprend pourquoi cette fonctionnalité est disponible sur toutes les calculatrices scientifiques, et qu'elle est également proposée dans tous les tableurs.

3.2.2 Droite de régression et coefficient de corrélation

La droite obtenue par la procédure d'ajustement par moindres carrés porte également le nom de “droite de régression”. Ce nom provient des travaux effectués au début du XX^e siècle par le généticien Francis GALTON, dont on pourra lire une biographie sommaire dans l'Annexe B. Il découvrit en particulier un phénomène qu'il qualifia de “régression vers la moyenne”. À partir de l'étude de la taille des populations, il tenta d'expliquer pourquoi la taille des fils tend à régresser vers la taille moyenne de la population, par comparaison avec la taille du père. Ainsi, un père de grande taille tend à avoir un fils de plus petite taille et vice versa. De manière plus générale, l'analyse par régression permet de s'interroger sur la relation qui peut exister e cause à effet entre deux variables x et y , en identifiant cette relation et en mesurant cette importance.

Afin de montrer comment la droite d'ajustement par moindres carrés permet de quantifier la régression entre une grandeur “explicative” x et une grandeur “expliquée” y (figure 9), introduisons les notations suivantes :

$$\begin{cases} s_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \bar{x}^2 \\ s_y^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2 = \frac{1}{N} \sum_{i=1}^N y_i^2 - \bar{y}^2 \\ s_{xy} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{N} \sum_{i=1}^N x_i y_i - \bar{x}\bar{y} \end{cases} \quad (56)$$

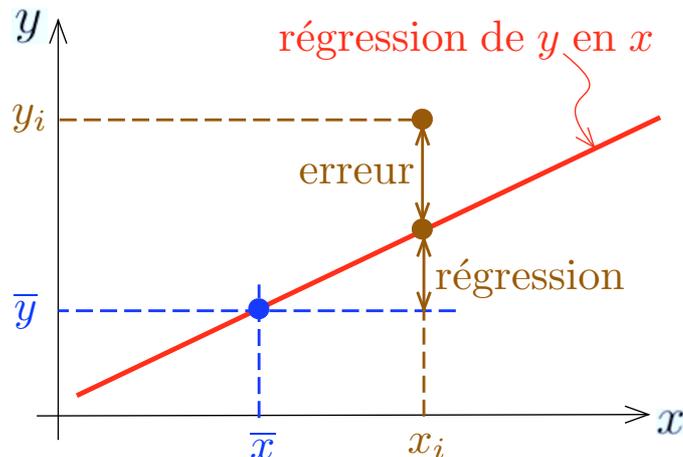


Figure 9: Droite de régression de y en x , ayant pour but de quantifier le degré de corrélation entre ces deux grandeurs.

D'après l'équation (55), la pente de la droite de régression de y en fonction de x est donnée par :

$$a = r_{xy} \frac{s_y}{s_x} \quad \text{avec} \quad r_{xy} = \frac{s_{xy}}{s_x s_y} . \quad (57)$$

La quantité r_{xy} est appelée "coefficient de corrélation linéaire". Nous pouvons également définir une variance résiduelle s_R^2 par la relation

$$s_R^2 = \frac{1}{N} \chi^2 \quad (58)$$

soit

$$s_R^2 = \frac{1}{N} \sum_{i=1}^N [y_i - (ax_i + b)]^2 = s_y^2 (1 - r_{xy}^2) . \quad (59)$$

Cette grandeur correspond à la variance non expliquée par la droite de régression de y en x . D'autre part, la variance s_E expliquée par la droite de régression est donnée par :

$$s_E = \frac{1}{N} \sum_{i=1}^N (ax_i + b - \bar{y})^2 . \quad (60)$$

Un peu d'algèbre permet d'établir la relation suivante :

$$s_E = r_{xy}^2 s_y^2 = s_y^2 - s_R^2 . \quad (61)$$

Ces résultats permettent d'établir les propriétés les plus importantes du coefficient de corrélation, à savoir :

- (a) Le coefficient de corrélation r_{xy} est une quantité comprise entre -1 et $+1$;
- (b) Le coefficient de corrélation en valeur absolue $|r_{xy}|$ en valeur absolue mesure, sur une échelle allant de 0 à 1 , le degré d'alignement des points correspondant aux observations ;
- (c) Un coefficient de corrélation r_{xy} égal à -1 ou $+1$ indique un alignement parfait, soit une corrélation parfaite entre les deux grandeurs, tandis qu'un coefficient de corrélation égal à 0 indique une absence de corrélation ;
- (d) Le signe du coefficient de corrélation indique si l'alignement est de pente positive, correspondant à une corrélation positive, ou de pente négative, correspondant à une corrélation négative.

3.2.3 Incertitudes sur les paramètres de l'ajustement

Il y a évidemment une faiblesse criante dans ce que nous avons fait jusqu'à présent, puisque nulle part ne figure d'incertitude sur cet ajustement. Plus précisément, nous aimerions connaître l'incertitude sur les paramètres a et b de la droite d'ajustement qui résulte de l'écart-type σ sur les mesures de la grandeur y . Imaginons pour cela la réalisation de nouvelles mesures de la grandeur y correspondant aux mêmes N valeurs $\{x_1, \dots, x_N\}$ de la grandeur x . Nous obtiendrions alors un nouvel échantillon de valeurs $\{y'_1, \dots, y'_N\}$, de sorte que l'ajustement par moindres carrés des nouveaux couples de points $\{(x_1, y'_1), \dots, (x_N, y'_N)\}$ conduirait à de nouvelles valeurs a' et b' de la droite d'ajustement par moindres carrés.

En supposant que chaque y_i est tiré au sort à partir d'une loi normale centrée sur la valeur vraie \bar{y}_i correspondant à x_i avec un écart-type σ (figure 10), les paramètres x_i et \bar{y}_i sont reliés par :

$$\bar{y}_i = \bar{a} x_i + \bar{b} \quad (62)$$

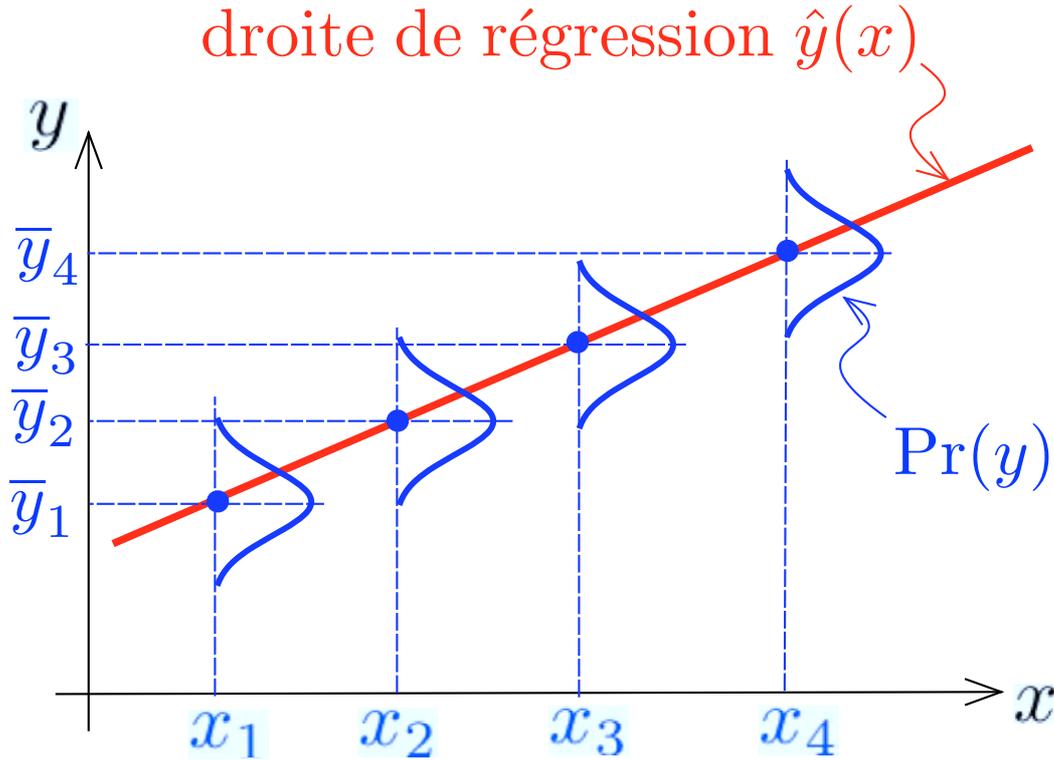


Figure 10: Mesures répétées pour un ensemble fixé de valeurs $\{x_1, x_2, \dots, x_N\}$ de la grandeur x . Pour chaque valeur x_i , le résultat y_i de la mesure forme une distribution normale centrée sur la valeur vraie \bar{y}_i . Les mesures étant supposées avoir le même poids statistique, chaque distribution gaussienne possède le même écart-type σ . Chaque nouvel ensemble de mesures $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ correspondra à un nouveau jeu de valeurs des paramètres a et b de la droite $y = ax + b$ d'ajustement par moindres carrés. L'ajustement vrai correspond quant-à lui à la droite d'équation $y = \bar{a}x + \bar{b}$, passant par les points $\{(x_1, \bar{y}_1), (x_2, \bar{y}_2), \dots, (x_N, \bar{y}_N)\}$.

où \bar{a} et \bar{b} correspondent aux valeurs *vraies* des paramètres de l'ajustement par moindres carrés.

Si nous ne sommes pas capables de connaître ces valeurs puisque nous disposons d'un seul ensemble de mesures, nous voulons pouvoir estimer les erreurs standard sur les valeurs de a et b données par l'équation (55) :

$$(\Delta a)^2 = \langle (a - \bar{a})^2 \rangle \quad \text{et} \quad (\Delta b)^2 = \langle (b - \bar{b})^2 \rangle. \quad (63)$$

À partir de la dispersion des valeurs de y par rapport à celles estimées par le modèle, nous montrerons en cours qu'il est possible de faire une estimation des incertitudes Δa et Δb .

3.3 Ajustement par une fonction polynômiale

On peut généraliser facilement la démarche précédente au cas de l'ajustement des points de mesure par un polynôme de degré p :

$$P(x) = c_0 + c_1x + c_2x^2 + \dots + c_px^p = \sum_{j=0}^p c_j x^j. \quad (64)$$

Les coefficients c_j du polynôme d'ajustement sont à nouveau déterminés par la méthode des moindres carrés, soit :

$$\frac{\partial \chi^2}{\partial c_j} = \frac{\partial}{\partial c_j} \left[\sum_{i=1}^N \left\{ \sum_{k=0}^p (c_k x_i^k - y_i) \right\}^2 \right] = 0. \quad (65)$$

On peut se convaincre aisément que le problème va en fait se ramener à un système linéaire de p équations à p inconnues correspondant aux coefficients c_k du polynôme. On obtiendra par conséquent dans tous les cas une solution pour les paramètres de l'ajustement.

Si une telle procédure est commode afin de par exemple pouvoir modéliser une courbe d'étalonnage d'un capteur ou d'un composant sans avoir besoin de connaître un quelconque modèle physique, nous pouvons nous demander s'il s'agit de la méthode générique d'ajustement. Poser la question revient évidemment à lui donner un début de réponse. Il suffit de considérer les équations de ce chapitre pour constater que cette méthode donnera **toujours** un résultat, aussi absurde qu'il puisse être. La figure 11 montre ainsi un exemple d'un ajustement de données expérimentales par une loi quadratique $y = a + bx + cx^2$, dont on peut douter de l'à-propos...

Moralité : ce n'est pas parce qu'un programme d'ajustement par moindres carrés donne un résultat que (1) le modèle est validé et que (2) les paramètres donnés par le calcul algorithmique sont effectivement pertinents vis-à-vis du problème physique considéré.

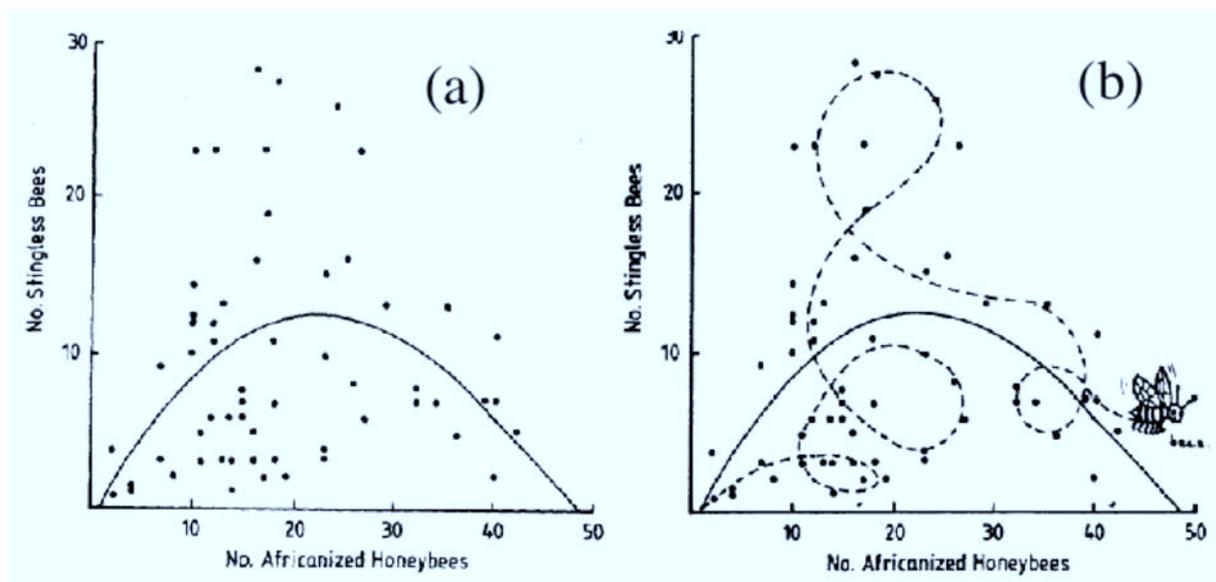


Figure 11: Un exemple d'ajustement. La figure (a) représente l'abondance relative d'abeilles "africanisées" (axe horizontal) et d'abeilles sans dard (axe vertical), et l'ajustement de ces données par un polynôme du second degré. Cette courbe est tirée d'un article publié en 1978 dans la revue prestigieuse *Nature*. La figure (b) propose un autre type d'ajustement...

3.4 Ajustement non-linéaire

Ce qui rend l'ajustement de lois affines, et par extension de polynômes, si simple, c'est que la dépendance d'un polynôme par rapport aux paramètres qui le composent est linéaire. Ainsi, la fonction $y = a + bx + cx^2$ est linéaire en a , b et c , même si ce n'est pas une fonction linéaire de la variable x . Si la fonction d'ajustement ne dépend plus de façon linéaire des coefficients de l'ajustement, le principe de la minimisation des moindres carrés peut continuer à être appliqué, tout au moins dans son principe. On utilise alors un algorithme d'ajustement non-linéaire qui fonctionne selon la procédure suivante. Imaginons qu'on cherche à trouver le fond d'une vallée la nuit et dans le brouillard. À partir d'un point de départ donné, on suit naturellement la direction de plus grande pente pour descendre, et on continue à le faire jusqu'à aboutir à un minimum "local". Bien évidemment, si on fait un pas "de géant", il se peut qu'on passe par-dessus le fond de la vallée. À l'inverse, plus on arrive vers le fond de la vallée et plus il est nécessaire de faire des petits pas pour être sûr de bien localiser le minimum. Dans la recherche du minimum

du χ^2 de l'ajustement, IGOR utilise l'algorithme de LEVENBERG–MARQUARDT¹⁹ qui est une méthode numérique astucieuse et efficace pour définir la taille du pas au fur et à mesure qu'on se rapproche du minimum .

Attention cependant ! Ce n'est pas parce que cette méthode est intégrée dans le logiciel IGOR Pro ou un autre, qu'elle est infaillible. Si le point de départ de la recherche, c'est-à-dire les valeurs proposées initialement pour les paramètres de l'ajustement²⁰, est trop éloigné de la solution, le résultat obtenu risque d'être absurde ou bien l'algorithme risque de ne jamais converger. Un autre danger est d'utiliser un trop grand nombre de paramètres ajustables, chaque paramètre ajoutant une dimension à l'espace de recherche. Dans ce cas, la signification physique du résultat mathématique obtenu risque d'être assez discutable.

Indépendamment de l'utilisation d'un algorithme d'ajustement non-linéaire, il existe au reste une première méthode efficace pour tester, tout au moins de façon semi-quantitative, la validité d'un modèle. Une bonne habitude est de tracer sur un même graphe la courbe théorique et les résultats expérimentaux pour voir "à l'œil" si ça ressemble²¹ ! On peut pousser un peu plus loin l'analyse en traçant en parallèle la courbe des écarts entre fonction théorique et points de mesure. L'ajustement des valeurs des paramètres du "fit" se fait alors par une procédure essai-erreur, qui converge plus ou moins rapidement...

Résumons-nous pour conclure.

Lorsqu'on ajuste un modèle théorique à des données expérimentales, on est en pratique confronté à deux types de difficultés:

- La recherche du minimum de χ^2 , sachant qu'il est parfois difficile de prouver que le minimum trouvé par le logiciel est bien le véritable minimum de la fonction ;
- Le fait que d'avoir trouvé le minimum de χ^2 ne valide en aucun cas le modèle utilisé.

C'est en fait le bon sens du physicien, et surtout son analyse critique, qui permet d'avoir confiance dans l'ajustement. On doit considérer les paramètres de l'ajustement comme des grandeurs physiques, voir si l'ordre de grandeur obtenu au moyen de l'ajustement est effectivement correct, vérifier si une modification d'un paramètre donné entraîne bien une variation dans le sens qu'on observe expérimentalement, etc. On pourra lire dans l'Annexe D ce que le physicien français du début du XX^e siècle Henri BOASSE pensait, avec son outrance habituelle, de la méthode d'ajustement par moindres carrés.

¹⁹W. H. PRESS, S. A. TEUKOLSKY, W. T. VETTERLING et B. P. FLANNERY, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press (2^eme édition, 1992). Le contenu de cet ouvrage de base sur les méthodes numériques peut être consulté *online*, à l'adresse <http://www.library.cornell.edu/nr/bookcpdf.html>.

²⁰C'est là une différence fondamentale par rapport à l'ajustement d'un polynôme, où les valeurs des paramètres sont déterminées directement par la résolution d'un système linéaire d'équations.

²¹Le logiciel IGOR Pro offre cette possibilité une fois la fonction de "fit" définie.

4 Annexes

4.1 Annexe A : rappels élémentaires sur les lois de probabilité

4.1.1 Loi de probabilité et fonction caractéristique

Considérons une variable aléatoire X associée à une distribution de probabilité $\Pr(X, x)$, telle que $d\Pr = \Pr(X, x)dx$ est la probabilité élémentaire pour que le tirage de la variable X se fasse entre les valeurs x et $x + dx$. Nous ferons l'abus de langage traditionnel du physicien, qui consiste à confondre le nom de la variable aléatoire X avec celui de son échantillon x réel. La fonction, maintenant notée $\Pr(x)$, peut être caractérisée par la fonction caractéristique $\varphi_x(t)$, avec t réel, définie par :

$$\varphi_x(t) = \langle e^{itx} \rangle \quad \text{soit} \quad \varphi_x(t) = \int_{-\infty}^{+\infty} \Pr(x) e^{itx} dx . \quad (66)$$

En développant e^{itx} en série entière, on obtient :

$$\varphi_x(t) = \sum_{n=0}^{\infty} (it)^n \frac{\langle x^n \rangle}{n} \quad (67)$$

de sorte que la connaissance de $\varphi_x(t)$ permet d'avoir directement les moments $\langle x^n \rangle$ de la variable aléatoire, sans qu'il soit pour autant nécessaire de disposer d'une expression explicite de la fonction de distribution $\Pr(x)$.

Nous pouvons également noter que :

$$\begin{aligned} \left(\frac{d\varphi_x(t)}{dt} \right) (t=0) &= i \langle x \rangle \\ \left(\frac{d^2\varphi_x(t)}{dt^2} \right) (t=0) &= -\langle x^2 \rangle \\ \left(\frac{d^n\varphi_x(t)}{dt^n} \right) (t=0) &= i^n \langle x^n \rangle \end{aligned} \quad (68)$$

offrant ainsi une manière simple pour calculer les moments de la loi de probabilité $\Pr(x)$:

$$\langle x^n \rangle = \frac{1}{i^n} \left(\frac{d^n\varphi_x(t)}{dt^n} \right) (t=0) . \quad (69)$$

4.1.2 Cas de la distribution normale

Considérons la distribution de probabilité gaussienne centrée sur $\mu = 0$ et d'écart-type σ . La fonction caractéristique est alors donnée par :

$$\varphi_x(t) = \int_{-\infty}^{+\infty} e^{itx'} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x'^2}{2\sigma^2}} dx' \quad \text{soit} \quad \varphi_x(t) = e^{-\frac{\sigma^2 t^2}{2}} \quad (70)$$

par le calcul habituel de la transformée de Fourier d'une fonction gaussienne. Par développement en série, on obtient alors :

$$\varphi_{x'}(t) = \sum_{k=0}^{\infty} (-1)^k \frac{\sigma^k t^{2k}}{2^k k} \quad (71)$$

On voit ainsi que tous les moments d'ordre impair sont nuls

$$\langle x'^{2p+1} \rangle = 0$$

tandis que ceux d'ordre pair valent :

$$\langle x'^{2p} \rangle = \sigma^{2p} C_{2p} \quad \text{avec} \quad C_{2p} = \frac{(2p)}{2^p p} .$$

Grâce à cette expression, on retrouve la variance de la distribution $\langle x'^2 \rangle = \sigma^2$. On a de même $\langle x'^4 \rangle = 3\sigma^4$ de sorte que :

$$\frac{\langle x'^4 \rangle}{\langle x'^2 \rangle^2} = 3.$$

Cette dernière relation est souvent considérée comme la signature d'une distribution gaussienne, lorsqu'on considère une distribution de probabilité $\text{Pr}(x)$ qui est au départ inconnue.

4.2 Quinconce de Galton ou la “planche à clous”

4.2.1 Description du dispositif

Une expérience élémentaire permet de visualiser la limite gaussienne d'un phénomène aléatoire. Sur un plan incliné, on plante en quinconce des clous de sorte qu'ils soient régulièrement espacés, en les disposant sur des lignes horizontales parallèles (figure 12). À la base du plan incliné, se trouvent des compartiments identiques, séparés par des cloisons qui prolongent les clous situés à la base du quinconce. On place sur la ligne supérieure une bille dont le diamètre est sensiblement égal à la distance entre deux clous voisins. Depuis cette position initiale, la bille va descendre en passant à gauche ou à droite de chaque clou rencontré avec la probabilité $\frac{1}{2}$, jusqu'à ce qu'elle tombe dans l'un des compartiments à la base du quinconce. Lorsqu'on répète cette petite expérience avec un grand nombre de billes, les billes accumulées dans les différents compartiments forment un histogramme dont l'enveloppe se rapproche de plus en plus d'une courbe en cloche. On trouvera sur le web ²² de très nombreuses simulations de cette expérience, dont le physicien n'aura pas manqué de remarquer le lien avec le problème de la marche au hasard à une dimension.

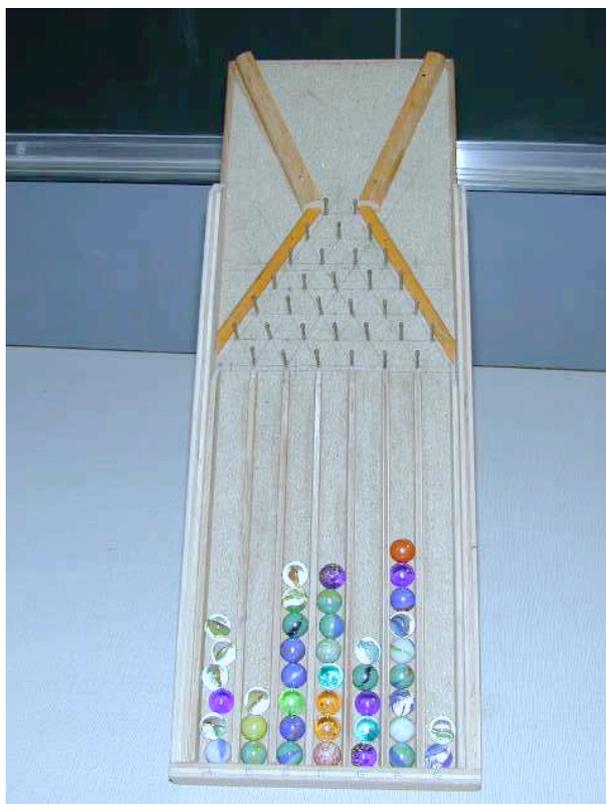


Figure 12: Expérience du “quinconce” ou “planche” de GALTON. On peut voir ce même dispositif, réalisé à plus grande échelle, à la Cité des Sciences de Paris.

²²Par exemple <http://www-sop.inria.fr/mefisto/java/tutorial1/node11.html>.

Pour la petite histoire, ce dispositif fut imaginé par Sir Francis GALTON (1822–1911), qui vécut à l’époque victorienne et fut l’un des pionniers des études statistiques. Ce scientifique, qui commença sa carrière par l’exploration en 1844 du Nil blanc et du sud-ouest de l’Afrique, aborda tout au long de sa vie de nombreux sujets, avec à chaque fois la volonté de quantifier l’objet de ses études. Séduit par la théorie de l’évolution de son cousin Charles DARWIN, il voulut justifier la transmission des capacités intellectuelles par l’hérédité afin de permettre l’amélioration de l’espèce humaine. Son point de départ était le paradoxe suivant. Comment expliquer qu’à chaque génération, on observe une dispersion des tailles, alors qu’à celle des parents devra s’ajouter celle des enfants, et qu’en même temps la taille moyenne des individus d’une population et la dispersion par rapport à cette moyenne restent à peu près constantes quand les générations se succèdent. Les concepts et méthodes qu’il introduisit pour ses études, comme la corrélation et les analyses de régression, sont devenus des outils de base de la statistique moderne.

4.2.2 Modélisation mathématique de la “planche à clous”

On trouvera sur le web ²³ de très nombreuses simulations de cette expérience, dont le physicien n’aura pas manqué de remarquer le lien avec le problème de la marche au hasard à une dimension. Ce dispositif illustre également la convergence de la binomiale $B(\frac{1}{2}, n)$, correspondant à n tirages indépendants d’une pièce de monnaie non biaisée, vers une loi normale, montrant ainsi comment une succession de processus élémentaires, indépendants les uns des autres, donne un effet final dont l’histogramme de probabilité tend vers une distribution normale.

Nous pouvons également justifier ce résultat de manière élémentaire. À chaque choc de la bille contre un clou, elle a une probabilité $\frac{1}{2}$ d’aller à gauche ($\Delta x = -a$) ou à droite ($\Delta x = +a$). Si nous notons x la variable aléatoire qui représente la position horizontale de la bille, sa fonction de distribution est :

$$f(x) = \begin{cases} \frac{1}{2} & \text{si } x = +a \\ \frac{1}{2} & \text{si } x = -a \\ 0 & \text{si } x \neq +a - a \end{cases}$$

en choisissant de prendre la position du clou comme origine des axes. Par conséquent :

$$\begin{aligned} E(x) &= \frac{1}{2} \times (-a) + \frac{1}{2} \times (+a) = 0 \\ V(x) &= E([x - E(x)]^2) = \frac{1}{2} \times (-a - 0)^2 + \frac{1}{2} \times (+a - 0)^2 = a^2. \end{aligned}$$

Si nous nous intéressons maintenant à la position de la boule après N chocs, son abscisse est une variable aléatoire

$$X = \sum_{i=1}^N x_i$$

dont la fonction de distribution $f_N(x)$ est celle de la somme de N variables aléatoires identiques x_i d’espérance mathématique $E(x) = 0$ et de variance $V(x) = a^2$. Le théorème de la limite centrale nous dit alors que cette fonction de distribution se comporte pour $N \gg 1$ comme une loi normale :

$$f_N(X) \simeq \frac{1}{\sigma\sqrt{2\pi}} e^{-X^2/2\sigma^2} \quad \text{avec} \quad \sigma^2 = \sum_{i=1}^N V(x_i) = Na^2.$$

La répartition que l’on matérialise en envoyant un grand nombre de billes qui vont venir s’accumuler dans les différentes colonnes a donc pour limite la fonction

$$f_\infty(X) = \frac{1}{Na\sqrt{2\pi}} e^{-X^2/(2Na^2)}.$$

²³Par exemple <http://www-sop.inria.fr/mefisto/java/tutorial1/node11.html>.

Là encore, le lien avec la marche au hasard à une dimension est transparent. Le lecteur intéressé peut également s'amuser à retrouver ce même résultat en considérant les chemins parcourus par une bille et en évaluant les probabilités que la bille, partant de $X = 0$ arrive à une position X donnée.

4.2.3 Composition de deux dispersions consécutives

Le quinconce permet également à GALTON de comprendre l'effet de la combinaison de deux dispersions successives. Les billes sont alors recueillies dans une première série de bacs, où elles dessinent une distribution normale d'écart-type σ' . On rouvre ensuite les bacs séparément et on peut constater que d'une part, chacun des bacs engendre une distribution normale, de même écart-type σ'' et que d'autre part, la réunion de toutes ces petites distributions conduit à une grande distribution normale d'écart-type $\sigma > \sigma'$. On vérifie ainsi la propriété de **stabilité** des variables aléatoires gaussiennes : si l'on compose deux variables gaussiennes stochastiquement indépendantes, on obtient de nouveau une variable gaussienne.

Pour démontrer ce résultat, considérons deux variables gaussiennes x_1 et x_2 indépendantes, caractérisées par leurs valeurs moyennes μ_1 et μ_2 , ainsi que par leurs écarts-types σ_1 et σ_2 . Posons $X = x_1 + x_2$; la fonction caractéristique de cette nouvelle variable aléatoire va alors s'écrire :

$$\varphi_X(t) = \langle e^{itX} \rangle = \langle e^{it(x_1+x_2)} \rangle = \langle e^{itx_1} \rangle \langle e^{itx_2} \rangle$$

de sorte que

$$\varphi_X(t) = \varphi_{x_1}(t)\varphi_{x_2}(t) = e^{it(\langle x_1 \rangle + \langle x_2 \rangle)} e^{-\frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}}.$$

Nous voyons ainsi que X correspond à une variable aléatoire gaussienne avec comme paramètres :

$$\langle X \rangle = \langle x_1 \rangle + \langle x_2 \rangle \quad \text{et} \quad \sigma^2 = \sigma_1^2 + \sigma_2^2.$$

On notera que ce résultat peut être immédiatement généralisé à la somme de plusieurs variables aléatoires gaussiennes.

4.3 Annexe C : variance d'échantillon

Dans cette annexe, nous allons montrer que la variance d'échantillon s^2 , définie par l'équation (36), correspond à un paramètre qui, dans la limite d'un grand nombre d'échantillons de mesures $\{x_i\}$, donne en moyenne la vraie variance σ^2 de la distribution de probabilité $\text{Pr}(x)$ associée à la mesure de la grandeur x .

Calculons ainsi :

$$\begin{aligned} \langle s^2 \rangle &= \frac{1}{N-1} \sum_{i=1}^N \langle (x_i - m)^2 \rangle \\ &= \frac{1}{N-1} \sum_{i=1}^N \langle (x_i - \mu + \mu - m)^2 \rangle \\ &= \frac{1}{N-1} \left[\sum_{i=1}^N \langle (x_i - \mu)^2 \rangle - 2 \sum_{i=1}^N \langle (x_i - \mu)(m - \mu) \rangle + \sum_{i=1}^N \langle (m - \mu)^2 \rangle \right]. \end{aligned} \quad (72)$$

Nous reconnaissons sans difficulté le premier et le troisième terme dans cette somme :

$$\sum_{i=1}^N \langle (x_i - \mu)^2 \rangle = N\sigma^2 \quad \text{et} \quad \langle (m - \mu)^2 \rangle = \frac{\sigma^2}{N}. \quad (73)$$

Nous devons calculer le deuxième terme, que nous pouvons écrire à partir de la définition de la valeur moyenne d'échantillon m :

$$\sum_{i=1}^N \langle (x_i - \mu)(m - \mu) \rangle = \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^N \langle (x_i - \mu)(x_k - \mu) \rangle.$$

Dans cette expression, la seule contribution non nulle correspond à $i = k$, d'où :

$$\sum_{i=1}^N \langle (x_i - \mu)(m - \mu) \rangle = \frac{1}{N} \sum_{i=1}^N \langle (x_i - \mu)^2 \rangle = \sigma^2 \quad (74)$$

En reportant les expressions (73) et (74) dans l'équation (72), nous obtenons finalement la valeur moyenne de la variance d'échantillon s^2 :

$$\langle s^2 \rangle = \frac{1}{N-1} \left(N\sigma^2 - 2\sigma^2 + N\frac{\sigma^2}{N} \right) \quad \text{soit} \quad \langle s^2 \rangle = \sigma^2 \quad \text{CQFD.} \quad (75)$$

4.3.1 Annexe D : méthode des moindres carrés et poids des mesures

Dans l'exploitation d'une série de mesures, il est possible d'affecter chacun des points par un "poids" inversement proportionnel à la précision de la mesure correspondante. Nous ne résistons pas au plaisir de reproduire un extrait d'un livre de Henri BOUASSE²⁴ et nous laisserons au lecteur le soin de se forger sa propre opinion sur l'emploi de telles pondérations dans l'exploitation des résultats expérimentaux.

"Nous entrons ici dans le domaine du tripatouillage systématique et du bafouillage corrélatif. Ce que nous allons exposer est l'art de faire rendre aux nombres bruts le résultat qu'on désire, pour lécher les bottes d'un pontife, faire cadrer avec une théorie absurde dont on est l'auteur, ou tout simplement pour montrer qu'on est un imbécile : il y a des gens qui pourraient se taire, mais qui n'ont de cesse qu'on les ait classés parmi les idiots.

"Le *poids* est un coefficient dont on affecte le résultat d'une mesure ; il est censé déterminer sa précision. Comme je ne veux pas avoir l'air de dénigrer systématiquement, je cite l'exemple suivant emprunté à feu l'astronome HATT (membre de l'Institut, il va sans dire) : il est bien caractéristique de la sottise de certains savants.

On a mesuré une même longueur successivement avec deux règles, l'une en millimètres, l'autre en centimètres, en estimant chaque fois la fraction de division. La première mesure, dix fois supérieure en précision a donné 0.1853 m ; la deuxième a donné 0.186 m. En multipliant par 10 l'équation de condition correspondant à la première mesure, on obtiendra deux seconds membres ayant la même précision :

$$10x = 1.853 \text{ m} \quad \text{et} \quad x = 0.186 \text{ m}$$

La règle des moindres carrés nous conduira à écrire pour l'équation normale :

$$101x = 18.530 + 0.186 = 18.716 \text{ m} \quad \text{soit} \quad x = 0.185307 \text{ m}$$

"Ce texte est-il assez bête ! Quel crétin, ayant mesuré avec le mètre plus précis, tiendra compte de la mesure avec l'autre mètre ? Quel crétin encore plus gâteux s'imaginera obtenir une précision meilleure parce qu'il aura joint à une mesure d'une certaine précision une mesure

²⁴H. BOUASSE, "*Construction des instruments de mesure*", p. 536, Delagrave (Paris, 1935). Les ouvrages de cet auteur sont connus pour la vigueur avec laquelle certaines thèses sont défendues, en particulier dans les préfaces. Il ne faudrait cependant pas oublier que ce monumental traité, constitué de plus d'une trentaine de volumes, regorge d'explications qu'on ne trouve nulle part ailleurs.

évidemment moins précise ? Quel abruti admettra qu'une mesure, estimée dix fois plus précise, doit être affectée, personne n'a jamais su pourquoi, du coefficient 100 ?

“Là-dessus vous direz que l'exemple est schématique, qu'il ne faut pas attacher d'importance à une caricature ! Malheureusement, sous des fioritures, chaque fois qu'un auteur met des poids, l'opération de diffère en rien comme stupidité de celle que suppose notre astronome hydrographe !

“L'homme intelligent raisonne à l'inverse. Quand une mesure vous semble mauvaise, laissez-là de côté : il n'y a pas de méthode rationnelle pour utiliser l'erreur. Cent mille mauvaises mesures n'en font pas une bonne.

“Encore une fois, les exemples sont infiniment nombreux des résultats déplorables que ces méthodes de trompe-l'œil ont toujours donnés.

“En définitive, mettre des poids, c'est supposer qu'une opération compte pour plusieurs ; on applique donc la méthode ordinaire, en répétant plusieurs fois l'équation de condition que fournit la mesure en question.

“Comme le poids est arbitraire, il n'y a pas à discuter si l'opération de poids p doit être considérée comme p , p^2 ou p^n opérations, l'arbitraire du choix de p pouvant évidemment se reporter sur le choix de n . Aussi bien on ne discute pas des sottises ; j'en ai assez dit.

“HATT parle du *poids des observateurs*. Il consent à écrire : ‘

On apprécie d'une manière plutôt vague le poids des observateurs. Quant aux instruments, on en détermine l'erreur moyenne par l'ensemble des résultats antérieurs.

“J'avais pour collègue pendant de longues années un homme dont les étrangers considèrent les expériences comme bonnes, et qui cependant était notre joie pour la façon grotesque dont il expérimentait. Ses préparateurs se fichaient de lui à tel point qu'ils faussaient systématiquement les expériences : jamais notre imbécile ne s'en est aperçu. Voilà à quels résultats d'autres imbéciles mettront des poids ! (...)

“L'appareil ne vaut que par celui qui l'emploie ; mais il est impossible de juger de l'habileté de l'observateur sans recommencer ses expériences.

“Voilà ce que dit le bon sens... et tout le reste n'est que fichaise !

“Mettez des poids et des moindres carrés si ça vous amuse. Ça tient de la place, ça fait du papier noirci. Vous trouverez toujours dans les académies des gens, certes non pour vous lire, mais, sur des recommandations bien agencées, pour vous donner un prix, voire un fauteuil ! Vous ne demandez rien de plus ? Mettez des poids et des moindres carrés !”

Et pour finir, voici la conclusion d'Henri BOUASSE, si on en doutait encore :

“Ma conclusion est très simple. Vous voulez déterminer un cercle : déterminez-en trois points avec toute la précision possible ; ça vaut beaucoup mieux que d'en déterminer cent mille par-dessous la jambe. Quand vos trois points seront connus *au mieux de vos appareils et de votre habileté*, faites passer le cercle. Si vous avez quatre points qui refusent de se placer sur le cercle à *la précision estimée de vos appareils*, menez un cercle à l'œil dans leur voisinage : tous les calculs que vous ajouriez, perdraient votre temps sans aucun bénéfice.

“C'est ainsi qu'opèrent les meilleurs physiciens : seuls les ânes conservent l'espoir de remplacer l'habileté manuelle et les appareils bien étudiés par une manivelle mathématique. mais comme ils sont le nombre, ce n'est pas demain qu'on cessera d'appliquer la méthode des moindres carrés.”