

Université d'Orléans

U.F.R. Faculté des Sciences

Licence de Physique

Analyse Numérique pour les Sciences Physiques

Traitement de données expérimentales

Jean-Philippe Grivet, 2003

1 Introduction

Chaque fois que je cherche à mesurer pratiquement une grandeur, j'obtiens un résultat plus ou moins éloigné de la "vraie" valeur. En répétant l'expérience, en améliorant la technique, la méthode et les instruments de mesure, j'ai l'espoir (souvent vérifié) que la part d'erreur va diminuer et que les résultats vont s'approcher asymptotiquement d'une estimation raisonnable et sûre. Des considérations de ce genre s'appliquent à toute activité et en particulier aux mesures physiques, qui sont toujours entachées d'erreurs et d'incertitudes, lesquelles doivent être réduites par des perfectionnements de la technique et aussi estimées pour établir la validité des résultats.

Il est commode d'établir une classification des erreurs qui guettent la personne qui effectue des mesures. On distingue ainsi les erreurs systématiques, qui découlent par exemple d'un mauvais étalonnage d'un instrument ou d'un biais de l'observateur. Si, pour prévoir l'issue d'un prochain scrutin présidentiel, je n'interrogeais que des personnes de sexe masculin, d'âge supérieur à 50 ans et dont les revenus sont supérieurs à 300 000 F/an, j'aboutirais à une prédiction « biaisée », entachée d'une erreur de méthode. De même la mesure des dimensions d'une pièce avec une règle tordue, la mesure précise d'une masse sans tenir compte de la poussée d'Archimède de l'air...

D'autre part, malgré la bonne volonté de la personne chargée des mesures, il peut s'introduire des erreurs non reproductibles, aléatoires. C'est par exemple le cas lorsque la sensibilité de l'appareil est insuffisante et que le signal cherché est accompagné de « bruit ». Les erreurs aléatoires affectent la précision de la mesure : plus elles sont importantes, plus les mesures sont dispersées. Les erreurs systématiques affectent la justesse du résultat.

Si la dispersion se détecte aisément, le biais est moins visible ; en toute rigueur, il faudrait, pour le mettre en évidence, disposer d'une mesure juste et indépendante de la quantité en question. On augmente la précision d'une expérience en répétant celle-ci plusieurs fois et en retenant la moyenne des résultats individuels. On améliore la justesse en analysant avec soin la mesure et en cherchant toutes les causes d'erreurs possibles.

Dans les paragraphes qui suivent, je considérerais a priori que je dispose de mesures parfaitement justes, seulement affectées d'erreurs aléatoires. Cette situation permet de faire des prédictions assez simplement. Le résultat d'une mesure est considéré comme le résultat du tirage d'une variable aléatoire, superposition de la valeur "vraie" (non fluctuante ou certaine) et de l'erreur ; je supposerai que les erreurs sont additives, c'est-à-dire qu'un résultat Y est la somme de la valeur vraie inconnue s et d'une erreur ϵ : $Y = s + \epsilon$.

Comme ϵ est réputée aléatoire, je suis conduit à considérer les lois de probabilités les plus fréquentes qui peuvent caractériser une grandeur aléatoire. Une mesure physique est rarement utilisée seule ; je chercherai souvent à tirer parti d'une série de résultats, pour déterminer les paramètres qui définissent un modèle théorique. Quelle est l'influence des erreurs de mesures sur la précision de mes connaissances des paramètres ? Je donnerai quelques éléments de réponse à la fin du chapitre.

2 Probabilité

Je suppose que les lecteurs ont des connaissances élémentaires en théorie des probabilités et je ne ferai que rappeler quelques définitions. Je commence par le cas de grandeurs aléatoires discrètes, qui ne prennent que des valeurs entières ou équivalentes à des entiers (pile ou face, lancement d'un dé, tirage du loto). Les différents résultats possibles d'un tirage ou d'une épreuve, X_i , sont repérés par l'indice i , $1 \leq i \leq n$. Soit p_i la probabilité du résultat i . Ces grandeurs sont normalisées, en ce sens que

$$\sum_{i=1}^n p_i = 1.$$

La valeur moyenne d'une grandeur aléatoire X est notée $\langle X \rangle$ ou μ ; on a :

$$\langle X \rangle \equiv \sum_1^n p_i X_i. \quad (1)$$

Un autre paramètre important est l'écart type σ tel que :

$$\sigma^2 \equiv \langle (X - \langle X \rangle)^2 \rangle; \quad (2)$$

σ^2 est appelée la variance de X .

Il arrive souvent que l'on sache que tous les résultats d'une épreuve sont également probables (lancement d'un dé équilibré par exemple). La probabilité d'un événement complexe est alors proportionnelle au nombre d'événements élémentaires qui le compose (exemple : faire apparaître un nombre pair en lançant un dé a une probabilité de 3/6). On traduit souvent ce fait en disant que la probabilité d'un événement composé est le rapport du nombre de cas «favorables» au nombre de cas total.

Ces définitions prennent une forme un peu différente lorsque la quantité X est une grandeur réelle, continue, définie dans un intervalle fini ou non. Il est commode d'introduire la fonction de répartition de X , qui est la probabilité d'observer une valeur de X inférieure ou égale à x :

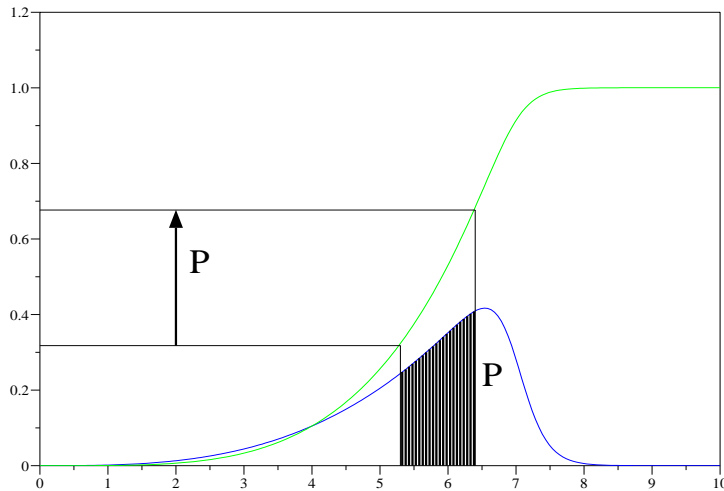
$$F(x) = \text{Proba}(X \leq x)$$

et la densité de probabilité de X qui est la dérivée de la fonction précédente :

$$f(x) = F'(x).$$

On l'interprète comme la probabilité de trouver X dans un petit intervalle :

$$f(x)dx = \text{Proba}(x \leq X \leq x + dx).$$



La figure montre les deux fonctions F et $F' = f$ et la représentation géométrique de la probabilité de trouver X dans l'intervalle $[5,3..6,4]$. La densité de probabilité est normalisée :

$$\int f(x)dx = 1,$$

ce qui implique que la fonction de répartition F varie de 0 à 1 d'une extrémité à l'autre de l'intervalle de définition. La valeur moyenne et l'écart type sont alors définis comme :

$$\langle x \rangle \equiv \int x f(x) dx, \quad (3)$$

$$\sigma^2 \equiv \int (x - \langle x \rangle)^2 f(x) dx, \quad (4)$$

les intégrales étant étendues à tout le domaine permis de x .

3 Lois de probabilité

3.1 Loi Binomiale

Soit une épreuve aléatoire n'ayant que deux résultats possibles : A et B ; si A ne se réalise pas, il y aura réalisation de l'événement complémentaire B (événements mutuellement exclusifs). Soit p la probabilité de A, et donc $q = 1 - p$ celle de B. Procédons à n épreuves consécutives et indépendantes. Je peux définir une variable aléatoire à valeurs entières X égale au nombre d'apparitions du résultat A sur l'ensemble des n épreuves. La probabilité pour que A se réalise exactement k fois (et donc B $n - k$ fois), sans tenir compte de l'ordre des événements, s'écrit :

$$P(X = k) \equiv b(k; n, p) = C_n^k p^k q^{n-k}. \quad (5)$$

Cette loi de probabilité est appelée loi binomiale, ou de Bernouilli. Le maximum de $b(k; n, p)$ est atteint approximativement pour $k = np \equiv k_m$. On montre aussi que pour la loi binomiale :

$$\langle k \rangle = np \quad ; \quad \sigma = \sqrt{npq}. \quad (6)$$

3.2 Loi de Poisson

Lorsque la probabilité p de A est faible et que n est assez grand, de telle sorte que $\lambda \equiv np$ reste moyen, la loi binomiale tend vers une forme limite, connue sous le nom de loi de Poisson. J'essaie de rendre plausible le passage d'une loi à l'autre, par le raisonnement suivant. Pour $k = 0$, j'ai

$$b(0; n, p) = (1 - p)^n = \left(1 - \frac{\lambda}{n}\right)^n \simeq e^{-\lambda}.$$

De plus, je déduis de la définition de $b(k; n, p)$ la relation

$$\frac{b(k; n, p)}{b(k-1; n, p)} = \frac{\lambda - (k-1)p}{k(1-p)} \simeq \frac{\lambda}{k}.$$

Cette relation de récurrence me permet de calculer successivement

$$b(1; n, p) = \lambda e^{-\lambda}, \quad b(2; n, p) = \frac{\lambda^2}{2} e^{-\lambda}$$

et ainsi de suite. Je note $p(k; \lambda)$ cette valeur approchée de $b(k; n, p)$: c'est la loi de Poisson, laquelle s'écrit en général

$$p(k; \lambda) = e^{-\lambda} \frac{\lambda^k}{k!}. \quad (7)$$

On montre que $\langle k \rangle = \lambda$ et que l'écart-type obéit à $\sigma^2 = \langle k \rangle$. Cette loi de probabilité intervient dans tous les problèmes de comptage en radioactivité comme en microbiologie.

3.3 Loi normale ou de Gauss

Lorsque le nombre d'épreuves n et le nombre de «succès» k deviennent grands (en pratique supérieurs à 10), la loi binomiale tend vers la loi de Gauss. Pour le voir, il est commode de faire, dans $b(k; n, p)$, un changement de repère tel que le maximum de $b(k)$ soit à l'origine; j'ai dit que ce maximum avait lieu pour $k \simeq np = \langle k \rangle$. Il faut donc choisir comme nouvelle variable l'écart à la moyenne $x = k - \langle k \rangle$. De plus, lorsque n et k sont grands, je peux approcher les factorielles par la formule de Stirling : $n! = n^n e^{-n} \sqrt{2\pi n}$. Il vient alors

$$b(x; n, p) \simeq \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{x^2}{2npq}\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right),$$

en introduisant l'écart-type $\sigma = \sqrt{npq}$; J'ai supposé que n et k étaient grands; la quantité $\frac{x^2}{2\sigma^2}$ varie donc très peu lorsque k augmente d'une unité et je vais pouvoir la considérer comme une variable continue.

Je cherche la probabilité pour que la variable aléatoire X soit inférieure ou égale à x . Cet événement se produit si $X = x - 1$ ou si $X = x - 2, \dots$ si bien que la fonction de répartition de X s'écrit

$$F(x) = \sum_{x'} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x'^2}{2\sigma^2}\right).$$

Comme les valeurs de x' sont très serrées, je peux remplacer cette somme par une intégrale

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x'^2}{2\sigma^2}\right) dx'.$$

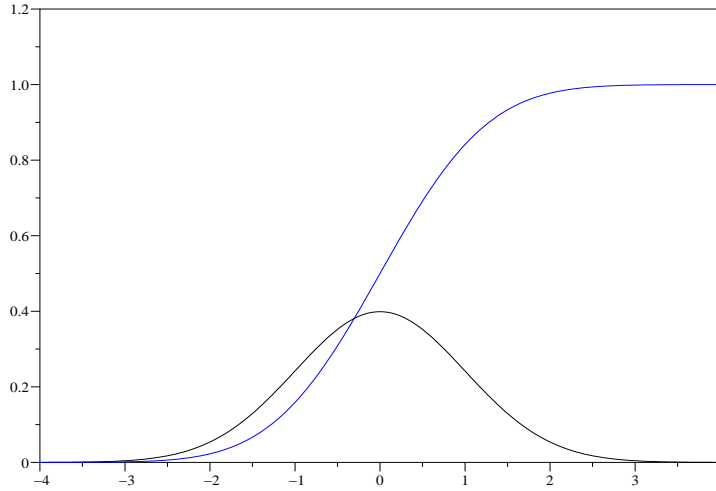
La fonction exponentielle décroît très rapidement lorsque x' devient supérieur à quelque fois σ ; je ne commet pratiquement pas d'erreur en calculant l'intégrale depuis $-\infty$. La fonction correspondante est considérée comme la fonction de répartition de la loi de Gauss-Laplace. Plus généralement, celle-ci est définie par

$$N(x; \mu, \sigma) \equiv \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left[-\frac{(x' - \mu)^2}{2\sigma^2}\right] dx'. \quad (8)$$

La probabilité pour que X soit compris entre x et $x + dx$ est proportionnelle à la densité de probabilité $n(x) = N'(x)$:

$$n(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right] dx. \quad (9)$$

Il est commode d'introduire la loi normale réduite (ou standard) pour laquelle $\sigma = 1$ et $\langle x \rangle = 0$, représentée ci-dessous.



Vous pouvez maintenant oublier l'origine discontinue de la variable x et la considérer comme une variable continue définie sur tout l'axe réel. Vous pourrez montrer directement (à l'aide des formules 3) que, comme on pouvait s'y attendre, la moyenne de X est μ et que son écart type est σ .

3.4 Loi du chi-deux ou de Pearson

Si X_1, X_2, \dots, X_n sont n variables aléatoires indépendantes, chacune distribuée selon la même loi normale de moyenne μ et d'écart type σ , alors la quantité $\sum (X_i - \mu)^2 / \sigma^2$ est distribuée selon une loi dite « du χ^2 (prononcez qui-deux) à n degrés de liberté ». Si $\langle X \rangle = \sum X_i / n$, la variable $\sum (X_i - \langle X \rangle)^2 / \sigma^2$ est distribuée comme le χ^2 à $n - 1$ degrés de liberté.

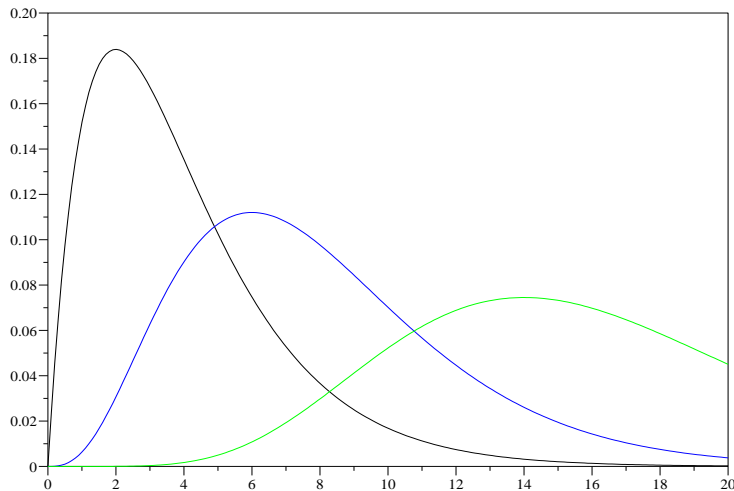
Une variable aléatoire X est distribuée comme le chi-carré (ou comme le chi-deux) à n degrés de liberté si sa densité de probabilité est :

$$f(x, n) = \frac{x^{n/2-1} e^{-x/2}}{2^{n/2} \Gamma(n/2)}. \quad (10)$$

La « fonction gamma », $\Gamma(x)$, est l'intégrale :

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$$

Elle obéit à la relation de récurrence $\Gamma(x + 1) = x\Gamma(x)$ et, de plus, $\Gamma(1/2) = \sqrt{\pi}$.



La figure montre trois versions de la loi du χ^2 , respectivement pour $n = 4, 8, 16$. On utilise souvent la fonction de répartition $P(X \leq \chi^2)$ ou la fonction complémentaire $Q(X \geq \chi^2)$, que l'on trouve tabulées dans les ouvrages de statistique mais que l'on peut aussi calculer facilement par intégration numérique.

3.5 Paramètres de la loi de probabilité et paramètres de l'échantillon

Dans la pratique, je dispose d'un certain nombre n d'observations d'une grandeur Y , réparties au hasard. L'ensemble des y_i constitue un échantillon extrait de la «population» de toutes les valeurs possibles de Y . La variable aléatoire Y obéit à une certaine loi de probabilité $p(y)$, laquelle est caractérisée par une moyenne μ et un écart-type σ . Je n'aurai jamais connaissance de ces paramètres, mais je pourrai essayer de les estimer à partir de propriétés de l'échantillon. Il importe de ne pas confondre paramètres de la population et paramètres de l'échantillon. En fait, on a les résultats suivants :

* La meilleure estimation de la moyenne μ de la population parente est la moyenne m de l'échantillon ($m = (1/n) \sum y_i$).

* La meilleure estimation de l'écart-type σ de la population parente est égale à l'écart-type s de l'échantillon, défini par :

$$s^2 = \frac{1}{n-1} \sum_1^n (y_i - m)^2. \quad (11)$$

Le facteur $n - 1$ (au lieu de n que l'on pourrait attendre) tient compte de ce qu'une partie de l'information contenue dans l'échantillon a été «consommée» pour calculer m . On dit que les nombres m et s sont respectivement des estimateurs de μ et σ . Ce sont des variables aléatoires (un autre échantillon donnerait d'autres valeurs de m et s). On montre que ces estimateurs sont sans biais, c'est-à-dire que $\langle m \rangle = \mu$ et que $\langle s^2 \rangle = \sigma^2$.

3.6 Vérification d'une loi de probabilité

Je dispose d'un échantillon d'une variable aléatoire Y : suis-je capable de déterminer la loi de probabilité à laquelle obéit Y ? Non, pas plus que je ne pourrais trouver rigoureusement la moyenne ou l'écart-type de Y . La seule approche possible est la suivante. Je fais le choix d'une loi de probabilité $f(y)$ et j'essaie de vérifier l'hypothèse : « les résultats expérimentaux sont compatibles

avec la loi f , avec une probabilité de $\alpha\%$. » Je ne considère ici qu'un exemple élémentaire. Soit une épreuve conduisant à l'un ou l'autre de r résultats mutuellement exclusifs dont les probabilités (inconnues) sont p_1, p_2, \dots, p_r . Par ailleurs, je note q_i mon estimation de la probabilité du résultat i . Je vais essayer de vérifier l'hypothèse $q_i = p_i, 1 \leq i \leq r$. Je répète n fois l'épreuve et je désigne par Y_i le nombre d'apparitions du résultat i . La quantité Y_i/n est proche de p_i (loi des grands nombres) et, si l'hypothèse est valable, elle est aussi voisine de q_i . Je dispose alors du théorème suivant. La variable S , définie comme

$$S = \sum_1^r \frac{(Y_i - np_i)^2}{np_i}$$

obéit à la loi de Pearson à $r - 1$ degrés de liberté, à condition que les nombres np_i soient assez grands ($np_i \geq 10$ en pratique). Comme j'ignore les p_i , je considère, à la place de S , l'estimateur

$$T = \sum_1^r \frac{(Y_i - nq_i)^2}{nq_i}, \quad (12)$$

qui obéit à peu près à la même loi de répartition. Il est clair que T sera petit si mon hypothèse est vérifiée. Je choisis une probabilité ϵ ; T' étant une variable distribuée selon le chi-deux à $r - 1$ degrés de liberté, je détermine un nombre χ^2 vérifiant

$$\text{Proba}(T' \geq \chi^2) \leq \epsilon.$$

Je rejette l'hypothèse si $T \geq \chi^2$. En d'autres termes, si l'écart entre fréquences observées et fréquences calculées me paraît trop grand pour être dû au hasard, je dois conclure que mon hypothèse est fautive.

Exemple. J'ai lancé 120 fois un dé, pour obtenir les résultats consignés ci-dessous. Je fais l'hypothèse que le dé est équilibré, ou encore que la probabilité de chaque face est $1/6$.

Rang i	nombres observés	nombres calculés	$(y_i - nq_i)^2/nq_i$
1	25	20	1,25
2	21	20	0,05
3	16	20	0,8
4	28	20	3,2
5	14	20	1,8
6	16	20	0,8
	total : 120	total : 120	$T = 7,9$

Je constate, à la lecture d'une table du χ^2 , que la probabilité pour que T (avec 5 degrés de liberté) soit au moins égal à 7,9 est comprise entre 0,1 et 0,2. L'hypothèse est considérée comme plausible et elle est acceptée. Le tableur « EXCEL » (entre autres) permet d'obtenir ce résultat en quelques pressions de touches. La fonction « loi.khi-deux(a2;n) » renvoie la probabilité d'apparition d'une valeur de T' au moins égale au contenu de la cellule «a2» pour un nombre de degrés de liberté égal à n ; le résultat est ici 0,1618. La fonction « test.khi-deux » effectue automatiquement l'ensemble des calculs, à condition de fournir en arguments la plage de cellules contenant les résultats observés et la plage destinée à contenir les résultats attendus.

4 Propagation des erreurs

Je m'intéresse maintenant à une grandeur physique G qui est déterminée indirectement à partir des mesures de deux grandeurs U et V . Sachant que les valeurs de U et de V sont entachées d'erreurs aléatoires, je me demande quelle peut être la dispersion de $G = g(U, V)$. La probabilité d'obtenir, lors d'une mesure particulière, le résultat (u, v) dépend de la «densité de probabilité conjointe»

$f(u, v)$. Plus précisément, la probabilité pour que U soit dans l'intervalle $[u, u + du]$ et V dans $[v, v + dv]$ est :

$$\text{Proba}(u \leq U \leq u + du; v \leq V \leq v + dv) = f(u, v)dudv.$$

Il faut encore généraliser la notion de valeur moyenne :

$$\langle g(u, v) \rangle = \int \int g(u, v)f(u, v)dudv.$$

La probabilité pour que U soit dans l'intervalle $[u, u + du]$, cela **quelque soit** V , s'écrit :

$$\text{Proba}(u \leq U \leq u + du; \forall V) = f_1(u) = \int f(u, v)dv.$$

À partir de la densité f_1 , je peux calculer la valeur moyenne $\langle u \rangle$ et l'écart-type σ_u de U , quelque soit V ; des formules symétriques existent pour la variable V . J'introduis ensuite les «moments» de la distribution f :

$$m_{p,q} \equiv \langle u^p v^q \rangle = \int \int f(u, v)u^p v^q dudv,$$

puis les "moments centrés" $\mu_{p,q} = \langle (u - \langle u \rangle)^p (v - \langle v \rangle)^q \rangle$. Je trouve en particulier :

$$\mu_{2,0} = \sigma_u^2 \quad ; \quad \mu_{0,2} = \sigma_v^2.$$

Le moment centré $\mu_{1,1} \equiv \langle (u - \langle u \rangle)(v - \langle v \rangle) \rangle$ est souvent appelé la covariance de U et V . Les variables aléatoires U et V sont dites corrélées si $\mu_{1,1} \neq 0$. Bien entendu, je ne dispose pas des « vraies » valeurs de U et V ; je ne connais que des estimateurs U^* et V^* (qui pourraient être les moyennes de quelques mesures de U et V). Ces estimateurs sont supposés sans biais : $\langle U^* \rangle = U$ et $\langle V^* \rangle = V$. Il est raisonnable de choisir $G^* = g(U^*, V^*)$ comme estimateur de G . Je vais calculer la moyenne et la variance de G^* . Je fais un développement limité de g autour du point U, V :

$$g(U^*, V^*) = g(U, V) + (U^* - U)g_u + (V^* - V)g_v + \dots$$

Dans cette expression, les dérivées ($g_U = \partial g / \partial U$, $g_V = \partial g / \partial V$) sont calculées en U, V et sont certaines (non aléatoires). La valeur moyenne de cette expression vaut, sachant que U^* , et V^* sont sans biais,

$$\langle g^* \rangle \equiv \langle g(U^*, V^*) \rangle = g(U, V)$$

en négligeant les termes d'ordre supérieur. g^* est donc lui-même, à cette approximation, un estimateur non biaisé. Je calcule maintenant la variance de g

$$\sigma_g^2 = \langle [G(U^*, V^*) - G(U, V)]^2 \rangle = \langle [(U^* - U)g_U + (V^* - V)g_V]^2 \rangle$$

et, en développant la valeur moyenne du carré :

$$\sigma_g^2 = \mu_{2,0}g_U^2 + 2\mu_{1,1}g_U g_V + \mu_{0,2}g_V^2.$$

Cette formule donne l'écart type sur g connaissant les moments centrés de U et V ; on l'écrit en général en fonction des écarts types :

$$\sigma_g^2 = \sigma_u^2 g_U^2 + 2\sigma_{u,v} g_U g_V + \sigma_v^2 g_V^2, \quad (13)$$

où j'ai introduit la notation habituelle $\sigma_{uv} = \mu_{1,1}$ pour la covariance. En pratique, on ne connaît pas les $\mu_{i,j}$ mais seulement des estimateurs construits à partir de résultats expérimentaux. Si les observations de U et V ne sont pas corrélées, alors $\sigma_{uv} = 0$ et la formule précédente se simplifie :

$$\sigma_g^2 = \sigma_u^2 \left(\frac{\partial g}{\partial U} \right)^2 + \sigma_v^2 \left(\frac{\partial g}{\partial V} \right)^2. \quad (14)$$

Exemple. Je considère les fonctions $x = uv$ et $y = u/v$; les variances correspondantes sont :

$$\sigma_x^2 = \sigma_u^2 v^2 + \sigma_v^2 u^2 + 2\sigma_{uv} uv \quad ; \quad \sigma_y^2 = \sigma_u^2 / v^2 + \sigma_v^2 u^2 / v^4 - 2\sigma_{uv} u / v^3$$

ce que l'on écrit en général de façon plus symétrique ($z = uv$ ou $z = u/v$) :

$$\left(\frac{\sigma_z}{z}\right)^2 = \left(\frac{\sigma_u}{u}\right)^2 + \left(\frac{\sigma_v}{v}\right)^2 \pm 2\left(\frac{\sigma_{uv}}{uv}\right)^2 \quad (15)$$

Dans le cas de variables non corrélées, on retrouve un résultat qui évoque l'addition des erreurs relatives que l'on enseigne dans les cours élémentaires.

5 Méthode du maximum de vraisemblance

Un noyau radioactif a une certaine probabilité de désintégration par unité de temps, notée $1/\tau$; en conséquence, pour un grand nombre de noyaux, le nombre de désintégrations par unité de temps, dN/dt , (la fréquence de l'événement) est proportionnel au nombre de noyaux :

$$dN/dt = -N(t)/\tau,$$

équation différentielle dont la solution s'écrit $N(t) = N_0 e^{-t/\tau}$. La probabilité de désintégration par seconde d'un noyau est $p(t) = (1/\tau)e^{-t/\tau}$. Comment déterminer la « durée de vie » ? Je peux (en principe au moins) mesurer les dates t_1, t_2, \dots, t_N auxquelles chacun de ces N noyaux disparaît. Si je recommençais l'expérience avec une nouvelle famille de N noyaux, j'obtiendrais des valeurs des t_i différentes des précédentes : ces fluctuations sont dues à la nature aléatoire du processus radioactif et existent même si les erreurs de mesure sont négligeables. La probabilité $V(\tau)$ d'observer effectivement les désintégrations aux époques t_1, t_2, \dots s'écrit (événements indépendants) :

$$V(\tau) = \prod_i^N p(t_i) = \exp \left\{ -\frac{1}{\tau} \sum_1^N t_i - N \ln \tau \right\}.$$

Le meilleur choix de τ est celui qui maximise V , c'est à dire celui qui rend le plus vraisemblable possible le résultat effectivement observé. Le maximum de V est atteint lorsque $-\ln V$ est minimal, soit pour la valeur :

$$\tau^* = \frac{1}{N} \sum_1^N t_i.$$

Autrement dit, la valeur de τ égale à la moyenne des t_i calculée sur l'échantillon rend maximale la probabilité d'apparition de cet ensemble de valeurs. On dit que V est une fonction de vraisemblance et que τ a été choisi selon un critère de maximum de vraisemblance.

L'exemple qui précède est particulier : si les mesures ne sont pas reproductibles, c'est à cause de la nature aléatoire du phénomène. De plus, le nombre de désintégrations par seconde est donné par une loi exponentielle (c'est en fait un cas particulier de loi de Poisson). Dans beaucoup d'autres cas, les mesures vont être entachées d'erreurs expérimentales. Je donne encore un exemple simple d'application du principe de maximum de vraisemblance dans ce cas.

Je suppose que j'ai réalisé n déterminations de la grandeur X , que les conditions sont un peu différentes d'une mesure à l'autre, si bien qu'à chaque mesure est associée une incertitude (écart-type) σ_i particulière. La probabilité d'apparition de la valeur x_i est, comme très souvent pour des erreurs expérimentales, donnée par une loi de Gauss, de valeur moyenne μ et de variance σ_i^2 :

$$p(x_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left(-\frac{(x_i - \mu)^2}{2\sigma_i^2} \right)$$

Je me demande quelle est la « meilleure » valeur de μ , en supposant, pour simplifier le raisonnement, que les σ_i sont connues. Pour répondre, je construis une fonction de vraisemblance $V(\mu)$, égale au

produit des $p(x_i)$, et qui représente la probabilité d'observer l'ensemble des résultats x_i . Comme au paragraphe précédent, je vais chercher la valeur de μ qui rend maximale la probabilité V , ce qui revient à chercher le minimum de $\ln V$ ou de l'argument de l'exponentielle. Il vient :

$$V(\mu) = \left[\prod_1^n \frac{1}{\sigma_i \sqrt{2\pi}} \right] \exp \left[-\frac{1}{2} \sum_1^n \left(\frac{x_i - \mu}{\sigma_i} \right)^2 \right].$$

En annulant la dérivée par rapport à μ de l'exposant, je trouve :

$$\mu^* = \frac{\sum_1^n \frac{x_i}{\sigma_i^2}}{\sum_1^n \frac{1}{\sigma_i^2}}$$

La valeur la plus vraisemblable de μ est la moyenne pondérée des valeurs observées x_i , les poids étant les inverses des variances relatives à chaque mesure. Je peux même déterminer l'incertitude sur μ : cette quantité dépend en effet des observations x_i et obéit à la loi de propagation des erreurs (nous supposons les mesures indépendantes, donc les corrélations nulles) :

$$\sigma_\mu^2 = \sum_1^n \left(\frac{\partial \mu}{\partial x_i} \right)^2 \sigma_i^2.$$

Un calcul simple montre que

$$\sigma_\mu^2 = \frac{1}{\sum_1^n (1/\sigma_i^2)}. \quad (16)$$

Lorsque toutes les incertitudes sont égales à σ , cette relation se réduit à

$$\sigma_\mu = \frac{\sigma}{\sqrt{n}}. \quad (17)$$

L'incertitude sur la valeur moyenne d'une série de mesures équivalentes décroît comme l'inverse de la racine carrée du nombre de mesures indépendantes.

6 Méthode des moindres carrés

Je vais maintenant appliquer la méthode du maximum de vraisemblance à un problème plus général. Il me faut, cependant, faire une remarque préliminaire. Bien des gens considèrent la méthode (on dit souvent le principe) du maximum de vraisemblance comme peu convaincante. Ils préfèrent se référer plus directement à un «principe des moindres carrés». Sauf cas particuliers, les deux démarches sont équivalentes.

J'imagine que j'ai mesuré la variation d'une grandeur physique y en fonction d'une autre grandeur indépendante x : par exemple la susceptibilité magnétique d'un matériau en fonction de la température. Je dispose donc d'une série de résultats expérimentaux $\{x_i, y_i\}, i = 1, 2, \dots, n$. Les valeurs $\{x_i\}$ de la variable indépendante sont supposées parfaitement exactes, alors que les mesures de y sont entachées d'erreurs aléatoires. D'autre part, j'ai des raisons de penser que y est lié à x par une loi physique de la forme :

$$y = f(x, a_1, a_2, \dots, a_m)$$

où les a_i sont des paramètres constants. Je suppose que les erreurs qui affectent chaque mesure y_i sont additives, indépendantes et réparties selon une loi normale. Je vais donc utiliser le modèle

$$Y = f(x, a_1, a_2, \dots, a_m) + \epsilon.$$

Je supposons encore que ce modèle est "juste" ou non biaisé, c'est-à-dire que $\langle \epsilon \rangle = 0$ ou encore que la moyenne de la distribution $p(y_i)$ est $f(x_i, a_1, \dots) = f_i = \langle Y_i \rangle$. Je souhaite trouver les valeurs

des a_i telles que la loi précédente représente "au mieux" l'ensemble des résultats expérimentaux $\{x_k, y_k\}, k = 1 \dots n$.

Avec les hypothèses faites, je peux écrire la probabilité d'apparition conjointe des événements y_1, y_2, \dots, y_n :

$$V(a_1, a_2, \dots, a_m) = \prod_1^m \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left[-\frac{(y_i - f_i)^2}{2\sigma_i^2} \right]$$

ou encore

$$V = \prod_1^m \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left(-\frac{\mathcal{S}}{2} \right) \quad ; \quad \mathcal{S} \equiv \sum_1^m \frac{(y_i - f_i)^2}{\sigma_i^2}$$

La quantité \mathcal{S} joue un grand rôle dans ce formalisme : c'est la somme des carrés des écarts entre valeurs observées (y_i) et valeurs calculées (f_i), chaque terme étant pondéré par l'inverse du carré de l'écart-type. Le maximum de V (la vraisemblance) est atteint lorsque l'exposant est minimal. On rejoint ici le «principe des moindres carrés» qui stipule que les meilleures valeurs des paramètres sont celles qui minimisent la somme (pondérée) des carrés des écarts entre valeurs expérimentales et valeurs théoriques.

Le maximum de V et le minimum de \mathcal{S} sont atteints quand

$$\frac{\partial \mathcal{S}}{\partial a_k} = -2 \sum_{i=1}^m \frac{y_i - f_i}{\sigma_i^2} \frac{\partial f_i}{\partial a_k} = 0.$$

6.1 Ajustement sur une fonction affine

Pour commencer, je considère un modèle dépendant linéairement de deux paramètres (et linéairement de x , bien que cette hypothèse ne soit pas nécessaire). Je dit que je vais ajuster les paramètres d'un modèle linéaire ou encore lisser les données à l'aide d'une droite. Le modèle s'écrit

$$Y = f(x, a, b) + \epsilon = ax + b + \epsilon$$

Comme $\partial f / \partial a = x$ et que $\partial f / \partial b = 1$, les conditions du maximum de vraisemblance sont alors

$$\sum_{i=1}^n \frac{y_i - f_i}{\sigma_i^2} = 0 \quad ; \quad \sum_{i=1}^n x_i \frac{y_i - f_i}{\sigma_i^2} = 0.$$

Ces relations constituent en fait un système de deux équations linéaires à deux inconnues (a et b). Il est commode de poser

$$S = \sum_{i=1}^n \frac{1}{\sigma_i^2} \quad ; \quad S_x = \sum_{i=1}^n \frac{x_i}{\sigma_i^2} \quad ; \quad S_{xx} = \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \quad ; \quad S_y = \sum_{i=1}^n \frac{y_i}{\sigma_i^2} \quad ; \quad S_{xy} = \sum_{i=1}^n \frac{x_i y_i}{\sigma_i^2}.$$

Le système mentionné ci-dessus s'écrit explicitement :

$$\begin{cases} aS_{xx} + bS_x &= S_{xy}, \\ aS_x + bS &= S_y. \end{cases} \quad (18)$$

Ces équations sont appelées «équations normales» ou «équations de Gauss». Le tableau des coefficients est symétrique et les valeurs expérimentales (y_i) n'apparaissent qu'au second membre. La solution est élémentaire ; je pose $\Delta = SS_{xx} - S_x^2$ et je trouve

$$a^* = \frac{1}{\Delta} (SS_{xy} - S_x S_y) \quad ; \quad b^* = \frac{1}{\Delta} (S_{xx} S_y - S_x S_{xy}). \quad (19)$$

Les coefficients a^* et b^* sont des fonctions des variables aléatoires y_i : ce sont donc elles-mêmes des variables aléatoires qui nous servent à estimer les « vraies » a et b . Pour apprécier la précision

(ou l'incertitude) de a^* et de b^* , il faut calculer l'écart-type de ces paramètres, au moyen de la formule de propagation des erreurs (14). Pour $u = a^*$ ou b^* , je sais que

$$\sigma_u^2 = \sum_1^n \left(\frac{\partial u}{\partial y_i} \right)^2 \sigma_i^2.$$

Je calcule les dérivées partielles à l'aide des formules (19) :

$$\frac{\partial a}{\partial y_i} = \frac{x_i S - S_x}{\sigma_i^2 \Delta}; \quad \frac{\partial b}{\partial y_i} = \frac{S_{xx} - x_i S_x}{\sigma_i^2 \Delta}.$$

Je calcule ensuite

$$\sigma_{a^*}^2 = \frac{S}{\Delta} \quad ; \quad \sigma_{b^*}^2 = \frac{S_{xx}}{\Delta}. \quad (20)$$

Cette méthode élémentaire ne permet pas de trouver la covariance de a^* et b^* ; je citerai simplement le résultat : $\sigma_{a^*b^*} = -S_x/\Delta$.

Les formules (19) et (20) se trouvent assez couramment programmées sur les calculettes. D'autre part, le système linéaire des équations normales définissant a^* et b^* fait intervenir la matrice :

$$\mathbf{M} = \begin{bmatrix} S_{xx} & S_x \\ S_x & S \end{bmatrix},$$

Je sais déjà que $\det \mathbf{M} = \Delta$ et je calcule \mathbf{M}^{-1} :

$$\mathbf{M}^{-1} = \frac{1}{\Delta} \begin{bmatrix} S & -S_x \\ -S_x & S_{xx} \end{bmatrix} = \begin{bmatrix} \sigma_{a^*}^2 & \sigma_{a^*b^*} \\ \sigma_{a^*b^*} & \sigma_{b^*}^2 \end{bmatrix}.$$

Vous voyez que l'inverse de la matrice des équations normales est la matrice dite des « variances-covariances ». C'est un cas où le calcul de l'inverse d'une matrice s'avère fructueux.

6.2 Linéarisation

Bien des fonctions peuvent, par un changement de variable approprié, se ramener au modèle linéaire du paragraphe précédent. Je vais encore traiter un cas particulier. L'étude théorique de la cinétique d'une réaction chimique montre que la concentration d'un réactif dépend du temps selon le modèle :

$$y = y_0 e^{-\alpha t}.$$

L'expérience m'a fourni des valeurs y_i relevées aux instants t_i . Pour déterminer y_0 et α , je linéarise ce modèle en posant $z = \ln y$:

$$z = \ln y_0 - \alpha t \equiv \beta - \alpha t.$$

Dans un calcul de moindres carrés habituel, chaque terme $y_i - f_i$ est pondéré par l'inverse de l'écart-type, σ_i . Quel est l'écart-type sur $\ln y_i$? Il se déduit de la loi de propagation des erreurs :

$$\sigma_z = \frac{\partial z}{\partial y} \sigma_y = \frac{\sigma_y}{y}$$

en supposant que chaque mesure souffre de la même incertitude. L'algorithme des moindres carrés va fournir la meilleure valeur de β , avec son écart-type, σ_β , mais c'est y_0 qui a un sens physique ; j'utilise encore la relation (14) pour trouver cette fois : $\sigma_{y_0} = y_0 \sigma_\beta$.

7 Qualité de l'ajustement

Après avoir déterminé les meilleures valeurs des paramètres a et b caractérisant un modèle linéaire, je dois me demander si l'expérience vérifie ce modèle. On dispose pour cela de divers tests statistiques. Je vais décrire le plus courant, le test dit du chi-deux. J'ai montré comment la méthode du maximum de vraisemblance conduisait aux valeurs les plus probables de a et b en rendant minimale la quantité

$$S^2 = \sum \left(\frac{y_i - f_i}{\sigma_i} \right)^2$$

où y_i est un résultat de mesure et f_i la valeur de y calculée par le modèle pour la même valeur de la variable indépendante x_i . Si le modèle était exact et les fluctuations absentes, S^2 serait nul. Ce ne sera pas le cas en pratique, à cause des erreurs de mesure d'une part, et des défauts du modèle d'autre part. J'espère néanmoins que S^2 est « petit ». Comme il s'agit d'une somme sur toutes les mesures (n en tout), S^2 augmente avec n . On peut raisonnablement s'attendre à ce que $y_i - f_i$ soit de l'ordre de σ_i et S^2 soit de l'ordre de n . Pour m'affranchir de l'effet du nombre de mesures, je pourrais être tenté de considérer la quantité S^2/n . Ce n'est pas tout à fait le bon choix, comme le montre l'exemple de $n = 2$. Je peux toujours faire passer une droite exactement par deux points, mais la perfection de cette ajustement n'est pas très convaincante. Il faut en fait rapporter S^2 au « nombre de degrés de liberté » de l'expérience (ν), défini comme le nombre de résultats indépendants (n) diminué du nombre de contraintes ou du nombre de paramètres ajustables (p en général, deux pour l'exemple d'une fonction affine), $\nu = n - p$. Une meilleure mesure de la qualité de l'ajustement est donc fournie par le « chi-deux réduit »

$$\chi_\nu^2 \equiv \frac{S^2}{\nu} = \frac{1}{N - p} \sum_i \left(\frac{y_i - f_i}{\sigma_i} \right)^2.$$

D'après les considérations précédentes, χ_ν^2 est une fonction aléatoire dont la valeur moyenne est proche de 1. Aussi, une valeur de χ^2/ν voisine de 1 signale-t-elle un bon modèle. De façon un peu plus précise, je remarque que l'une ou l'autre des quantités χ^2 ou S^2 est une somme de carrés de variables aléatoires gaussiennes ; elles obéissent donc à la loi du chi-deux. Je vais devoir décider si, et avec quelle probabilité, la valeur de χ_ν^2 déduite de mes mesures peut être le fait du hasard. Je fais donc l'hypothèse que mon modèle est juste, je fais le choix d'un seuil ϵ et, muni d'une table de la fonction de répartition de la loi de Pearson, je cherche un nombre u tel que pour une variable T' (distribuée selon la loi du chi-carré), $Proba(T' > u) = \epsilon$. Si $\chi^2/\nu > u$, c'est que l'écart théorie-expérience est très improbable, ou que mon hypothèse est fausse.

Il est malheureusement assez fréquent que je ne connaisse pas de l'incertitude affectant les y_i : c'est le cas lorsque je n'ai pas les moyens de répéter les expériences nécessaires. Je suis alors contraint de faire le raisonnement simplifié suivant. Je suppose que chaque mesure a la même dispersion : $\sigma_i = \sigma$ pour l'instant inconnue, je détermine les meilleures valeurs de a et b , puis je recalcule σ comme

$$\sigma^2 = \frac{1}{n - 2} \sum (y_i - a^* x_i - b^*)^2$$

Les écarts-types sur a^* et b^* peuvent alors être calculés par les formules (20) multipliées par le facteur $\sqrt{S/(n-2)}$.

Exemple. J'ai relevé les couples de valeurs suivantes :

x	1	2	4	5	8	9	10	12	15
y	-0,1	1,7	3,3	4,9	7,6	8,4	9,1	10,7	12,3

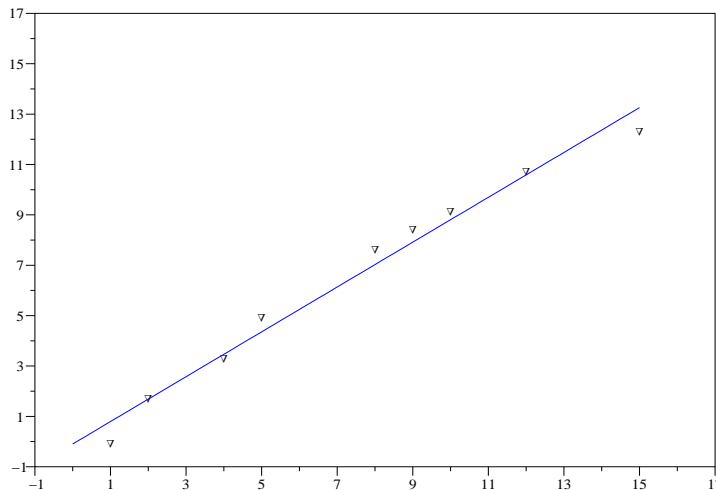
et je veux lisser les valeurs de y à l'aide du modèle $y = ax + b$, autrement dit, ajuster au sens des moindres carrés les paramètres a et b pour que l'équation proposée représente au mieux les résultats expérimentaux. Vous pouvez vous prêter au jeu suivant : reporter ces données sur papier

millimétré et déterminer « à l'oeil » la meilleure droite et les valeurs correspondantes de a et b , avant de faire les calculs.

1er cas (fréquent) : j'ignore les valeurs des erreurs commises sur les y , mais je pense qu'elles sont toutes aléatoires et distribuées de façon identique. Je pose alors $\sigma_i = \sigma = 1$. Comme toutes les formules qui donnent a^ et b^* sont homogènes en σ , la valeur exacte importe peu. Le calcul se fait facilement avec une calculette ou avec Scilab. Dans ce logiciel, je défini deux vecteurs x et y contenant les données. J'ai alors les instructions

```
S = length(x); Sx = sum(x); Sxx = sum(x.*x); Sy = sum(y); Sxy = sum(x.*y);
Delta = S*Sxx-Sx*Sx;
a = (S*Sxy - Sx*Sy)/Delta, b = (Sxx*Sy - Sx*Sxy)/Delta;
siga = sqrt(S/Delta); sigb = sqrt(Sxx/Delta)
```

Les résultats sont $a^* = 0.8903$, $\sigma_a = 0.0754$, $b^* = -0.0958$, $\sigma_b = 0.645$; les points expérimentaux et la droite sont représentés sur la figure. Si la pente est bien déterminée, l'ordonnée à l'origine est imprécise.



Avec ces valeurs de a^* et b^* , je recalcule σ : j'admet que tout écart entre valeur expérimentale et valeur calculée est due à une erreur aléatoire; je retiens la moyenne du carré de ces écarts comme valeur de σ^2 (comme indiqué plus haut, je considère qu'il n'y a que $n - 2 = 7$ mesures). Je trouve que $\sigma = 0.62$, d'où je déduis les «bonnes» valeurs $\sigma_a = 0.047$, $\sigma_b = 0.401$. Les amateurs d'EXCEL verront que ce logiciel, grâce à la fonction « DROITEREG », peut faire automatiquement les calculs précédents.

*2ème cas (idéal) : Je sais que les valeurs de y sont entachées d'erreurs telles que l'écart-type sur les 4 premières valeurs de y est de 0,4 et qu'il est de 0,6 sur les 5 dernières. J'obtiens alors :

$$S = 38,89 \quad S_x = 225,0 \quad S_y = 194,86 \quad S_{xx} = 1993,06 \quad S_{xy} = 1757,08$$

d'où je tire $a = 0,911$, $\sigma_a = 0,038$, $b = -0,259$, $\sigma_b = 0,272$. La quantité S^2 vaut 11,04 soit un χ^2_ν réduit (7 degrés de liberté) de 1,577. Les tables donnent une probabilité d'environ 0,14 d'obtenir par hasard une valeur au moins aussi élevée. Le modèle est donc accepté.

8 Coefficient de corrélation

Une suite d'observations m'a fourni des couples de valeurs (x_i, y_i) , comme par exemple le nombre de taches solaires observées pendant l'année i et le cours moyen du bourgogne à la vente

des Hospices de Beaune la même année. Je me pose la question de savoir s'il existe une relation causale entre ces deux variables, en d'autres termes, je cherche à savoir si x est corrélé à y . Je vais examiner le cas le plus simple, celui d'une corrélation linéaire. Si y dépend de x selon la loi linéaire $y = ax + b$, les résultats du paragraphe précédent me permette d'estimer le coefficient a . Si x et y sont indépendants, y ne doit, en moyenne, ni croître ni décroître quand x augmente, donc $a = 0$. Il est aussi permis d'estimer les paramètres de la loi $x = a'y + b'$. Ils sont différents des précédents, mais a et a' sont liés si x et y sont corrélées. Si la corrélation entre x et y est parfaite, les deux lois sont décrites par des fonctions inverses l'une de l'autre et $a = 1/a'$ ou $aa' = 1$. Je définis le coefficient empirique de corrélation par la relation $r = \sqrt{aa'}$. En supposant tous les écarts types égaux (une hypothèse qu'il est facile de lever), on trouve la formule ci-dessous.

$$r \equiv \frac{\sum_1^n (x_i - \langle x \rangle)(y_i - \langle y \rangle)}{\sqrt{\sum_1^n (x_i - \langle x \rangle)^2 \sum_1^n (y_i - \langle y \rangle)^2}}. \quad (21)$$

$|r|$ est compris entre 0 (pas de corrélation) et 1 (corrélation complète) : on considère que x et y sont fortement corrélés lorsque que $|r|$ est voisin de 1, pratiquement indépendants si r est proche de 0, mais il est difficile de donner un critère vraiment quantitatif en l'absence d'hypothèses plus précises sur la nature des erreurs entachant x et y .

9 Ajustement sur une fonction linéaire de m paramètres

Je généralise hardiment le raisonnement des paragraphes précédents, pour obtenir l'expression :

$$y(x) = \sum_1^m a_k \varphi_k(x) \quad (22)$$

qui dépend linéairement des m paramètres a_k , alors que les fonctions φ_k qui définissent la loi en x sont quelconques. Je souhaite de nouveau rendre minimale la somme des carrés des écarts :

$$\mathcal{S}^2 = \sum_1^n \frac{[y_i^{obs} - y_i^{calc}]^2}{\sigma_i^2}$$

avec, pour chacune des n mesures, $y_i^{calc} = y(x_i)$, la valeur prédite par le modèle. Il est commode d'introduire les notations suivantes. \mathbf{A} est la matrice à n lignes et m colonnes d'éléments $A_{ij} = \varphi_j(x_i)/\sigma_i$ et \mathbf{b} est le vecteur à n coordonnées $b_i = y_i^{obs}/\sigma_i$. De plus, les paramètres a_k sont considérés comme les coordonnées d'un vecteur \mathbf{a} . Les conditions pour que \mathcal{S}^2 soit minimale sont :

$$\sum_{i=1}^n \frac{1}{\sigma_i^2} [y_i^{calc} - y_i^{obs}] \varphi_k(x_i) = 0 \quad ; \quad k = 1, 2, \dots, m.$$

ce qui s'écrit :

$$\begin{aligned} \mathbf{M}\mathbf{a} &= \beta \quad ; \quad M_{ik} = \sum_{i=1}^n \frac{1}{\sigma_i^2} \varphi_k(x_i) \varphi_j(x_i), \\ \beta_j &= \sum_{i=1}^n \frac{1}{\sigma_i^2} y_i \varphi_k(x_i). \end{aligned}$$

Vous pourrez vérifier les relations

$$\mathbf{M} = \mathbf{A}^T \mathbf{A} \quad ; \quad \beta = \mathbf{A}^T \mathbf{b}$$

où \mathbf{A}^T désigne la matrice transposée de \mathbf{A} . La matrice \mathbf{M} est symétrique définie positive et le système linéaire se résout facilement par la méthode de Cholesky. La solution formelle est

$$\mathbf{a} = \mathbf{M}^{-1} \beta = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}.$$

Cependant, comme dans le cas à 2 inconnues, l'inverse de \mathbf{M} présente un intérêt : l'élément diagonal $[\mathbf{M}^{-1}]_{ii}$ est la variance (carré de l'incertitude) du paramètre a_i , alors que les éléments non diagonaux $[\mathbf{M}^{-1}]_{ik}$ sont les covariances entre a_i et a_k .