1 Introduction

Regressi effectue la régression par deux méthodes

1. par défaut, la méthode des moindres carrés classique et donc sans prise en compte des incertitudes. On suppose que l'incertitude sur l'abscisse est négligeable et que l'incertitude sur l'ordonnée est la même pour tous les points. La technique de détermination des incertitudes sur les paramètres repose sur l'assimilation entre l'écart modèle-données et l'incertitude sur l'ordonnée.
2. si les incertitudes sont définies et que l'utilisateur a coché « méthode du χ^2 », la dite méthode, les incertitudes sur x et y étant prises en compte.

Le test du χ^2 suppose que les incertitudes soient des incertitudes-type, et on élargit l'incertitude par un coefficient de Student de paramètre (nombre de mesures - nombre de paramètres), ce qui suppose que la statistique sur les paramètres est bien gaussienne.

Note modélisation s'entend au sens d'ajustement des paramètres d'une fonction donnée (fit en anglais). J'utilise modélisation car, pour moi, le choix de la fonction n'est pas arbitraire mais résulte de la modélisation du dispositif étudié.

1.1 Incertitudes des données

Les incertitudes définies pour les grandeurs expérimentales devraient donc être des incertitudes-type, notées u, de manière à ce que la loi de propagation puisse s'appliquer. Les options de tracé des ellipses le supposent également.

Rappel : expression de l'incertitude type

— Pour une mesure avec des graduations de longueur pas, l'incertitude-type est : $\dfrac{\text{pas}}{\sqrt{12}}$

— Pour un instrument avec une erreur de justesse maximale donnée : $\dfrac{\text{erreur}}{\sqrt{3}}$

— Pour un appareil de précision p, l'incertitude-type est $\dfrac{p}{\sqrt{3}}$

— Pour un appareil type voltmètre avec une précision de $\pm(pc$ % de lecture$+N\times$chiffre le moins significatif), l'incertitude sur x est donnée par $\dfrac{x\cdot pc+N\cdot ms}{\sqrt{3}}$ avec ms valeur correspondant à l'unité du dernier chiffre.

1.2 Affichage

L'affichage des ellipses d'incertitude suppose qu'il n'y a pas corrélation entre les deux grandeurs. Si on a entré des incertitudes-type

— avec une loi normale, une ellipse de demi-axe u correspondra à un intervalle de confiance de 68%, $2u$ de 95% et $3u$ de 99,7%.

— pour une loi rectangulaire u correspondra à 58%, $2u$ à 10% (sic) et $3u$ à 173% (resic)

2 Méthode des moindre carrés

Soit une grandeur y fonction affine d'une autre grandeur x : $y = Ax + B$. On a N couples de mesures $\{y_i, x_i\}$. Pour trouver A et B, on cherche à minimiser l'écart quadratique $\sum_{i=1}^{N}\left(y_i - A - Bx_i\right)^2$. Cela conduit, en notant $\Delta = N\sum x^2 - \left(\sum x\right)^2$, à $A = \dfrac{\sum x^2\sum y - \sum x\sum(xy)}{\Delta}$, et $B = \dfrac{N\sum(xy) - \sum x\sum y}{\Delta}$

On peut alors considérer que $\left(y_i - A - Bx_i\right)$ représente l'écart entre la valeur mesurée y_i et la valeur « vraie » $A + Bx_i$, et donc évaluer l'incertitude sur y par $\sigma_y^2 = \dfrac{1}{N-2}\sum_{i=1}^{N}\left(y_i - A - Bx_i\right)^2$, le 2 venant des deux paramètres A et B. L'incertitude sur y étant maintenant estimée, on peut évaluer l'incertitude sur A par propagation des incertitudes dans l'expression précédente de A. On trouve $\sigma_A = \sigma_y\sqrt{\dfrac{\sum x^2}{\Delta}}$ et $\sigma_B = \sigma_y\sqrt{\dfrac{N}{\Delta}}$.

On suppose que l'incertitude sur l'abscisse est négligeable et que l'incertitude sur l'ordonnée est la même pour tous les points. La technique de détermination des incertitudes sur les paramètres repose sur l'assimilation entre écart modèle-données et l'incertitude sur l'ordonnée.

3 Prise en compte des incertitudes

On voit que dans la technique usuelle, on fait une évaluation a posteriori de l'incertitude sur y. On choisit maintenant de minimiser toujours l'écart quadratique $y - Ax - B$, mais on pondère chacun des éléments par l'inverse de la variance et on en profite pour prendre en compte l'incertitude sur x. On a $\hat{y} = y + \delta y + A\delta x$, $\hat{y}$ valeur mesurée, y valeur « vraie » en x et δx, δy variables aléatoires centrées, la variance associée à y est donc $\sigma_y^2 + (A\sigma_x)^2$, par addition des variances. On minimise donc $\sum_{i=1}^{N}\dfrac{\left(y_i - A - Bx_i\right)^2}{\sigma_y^2 + (A\sigma_x)^2}$, c'est la méthode du χ^2. Il *suffit* alors de faire le même calcul qu'au §2, pour déterminer A et B, on fait apparaitre le coefficient de pondération $k_i = \dfrac{1}{\sigma_y^2 + (A\sigma_x)^2}$ dans les sommes.

$\Delta = \sum k \sum(kx^2) - \left(\sum kx\right)^2$; $A = \dfrac{\sum(kx^2)\sum(ky) - \sum(kx)\sum(kxy)}{\Delta}$; $B = \dfrac{\sum k \sum(kxy) - \sum(kx)\sum(ky)}{\Delta}$

Et de même pour les incertitudes

$\sigma_A = \sqrt{\dfrac{\sum(kx^2)}{\Delta}}$; $\sigma_B = \sqrt{\dfrac{\sum k}{\Delta}}$.

Le test du χ^2 affiche un χ^2 réduit $\dfrac{\chi^2}{n-p}$, avec n nombre de données et p nombre de paramètres. Ce χ^2 réduit devrait être proche de 1, si le modèle est correct et si les incertitudes données pour les grandeurs sont des incertitudes-type.

Remarque dans le cas simple où l'incertitude sur x est quasi-nulle et celle sur y constante, on devrait retrouve le même incertitude sur les paramètres que dans le cas précédent, sinon cela signifierait un χ^2 réduit relativement différent de 1, de mauvais augure.

4 Fonction quelconque

Ce qui est important dans les relations précédentes est que $y = A + B \cdot \mathrm{f(x)}$, avec dans le cas usuel $f(x) = x$, autrement dit la linéarité qui compte dans le calcul est celle relative à A et B. Dans le cas d'une fonction quelconque $y = \mathrm{f(x{,}A{,}B)}$, on suppose donc que, localement en A et B, f est bien une fonction linéaire de A et B et on applique les relations précédentes.

5 Prise en compte des incertitudes

Cette fois on minimise l'écart quadratique $y - \mathrm{f(x)}$. On a $\hat{y} = y + \delta y + \dfrac{\mathrm{d}y}{\mathrm{d}x}\delta x$, $\hat{y}$ valeur mesurée, y valeur « vraie » et δx, δy variables aléatoires centrées, la variance associée à y est donc $\sigma_y^2 + \left(\dfrac{\mathrm{df}}{\mathrm{d}x}\sigma_x\right)^2$, par addition des variances.

Les points de la courbe $y(x)$ sont pondérés par leur incertitude à la fois en x et en y : le coefficient de pondération est de $\dfrac{1}{u(y)^2 + \left(\dfrac{\mathrm{d}y}{\mathrm{d}x}\right)^2 u(x)^2}$ si $u(y)$ et $u(x)$ sont les incertitudes données dans l'onglet correspondant. Pour que les incertitudes soient actives, il faut qu'elles soient définies pour toutes les grandeurs expérimentales (double clic sur l'en-tête du tableau de valeurs pour l'éditer) et que vous ayez coché la case « méthode du chi2 » dans la boite de dialogue options obtenue à l'aide du menu local de la modélisation accessible par le clic droit. Une incertitude peut éventuellement être à zéro (cas du temps fréquemment).