Incertitudes dans Regressi

Jean-Michel Millet

24 juin 2019

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.

La propagation des incertitudes (incertitudes composées) se fait par addition des variances.

Exemple : $U = R \cdot I$, $\frac{u(U)^2}{U^2} = \frac{u(R)^2}{R^2} + \frac{u(I)^2}{I^2}$

Rappel : expression de l'incertitude type

— Pour une mesure avec des graduations de longueur pas, l'incertitude-type est : $pas/\sqrt{12}$
— Pour un instrument avec une $erreur$ de justesse maximale donnée : $erreur/\sqrt{3}$
— Pour un appareil de précision p, l'incertitude-type est $p/\sqrt{3}$
— Pour un appareil type voltmètre avec une précision de (pc % de lecture + $N\times$ chiffre le moins significatif), l'incertitude sur x est donnée par $\left(x \cdot pc + N \cdot ms\right)/\sqrt{3}$ 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 115% (sic) et $3u$ à 173% (resic)

2 Méthode des moindre carrés

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

On peut alors considérer que $\left(y_i - Ax_i - B\right)$ représente l'écart entre la valeur mesurée y_i et la valeur « vraie » $A \cdot x_i + B$, et donc évaluer l'incertitude sur y par $\sigma_y^2 = \dfrac{1}{N-2} \sum_{i=1}^{N} \left(y_i - A \cdot x_i - B\right)^2$, 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 $\sigma_B = \sigma_y \sqrt{\dfrac{\sum x^2}{\Delta}}$ et $\sigma_A = \sigma_y \sqrt{\dfrac{N}{\Delta}}$.

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 $y - A \cdot x - B$, 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 $\hat{y} = y + \delta y + A \cdot \delta x$, $\hat{y}$ 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 $\sum_{i=1}^{N} \dfrac{\left(y_i - A - B \cdot x_i\right)^2}{\sigma_y^2 + (A\sigma_x)^2}$, c'est la méthode du χ^2. Il *suffit* alors de faire le même calcul qu'au §**??**, pour déterminer A et B, on fait apparaitre le coefficient de pondération $k_i = \dfrac{1}{\sigma_y^2 + (A \cdot \sigma_x)^2}$ 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.

4.1 Prise en compte des incertitudes

Cette fois on minimise l'écart quadratique $y - \mathrm{f(x)}$. On a $\hat{y} = y + \delta y + \frac{\mathrm{d}y}{\mathrm{d}x}\delta x$, $\hat{y}$ valeur mesurée, y valeur « vraie » et δx, δy variables aléatoires centrées, la variance associée à y est donc $\sigma_y^2 + \left(\frac{\mathrm{df}}{\mathrm{d}x}\sigma_x\right)^2$, 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 $\frac{1}{u(y)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 u(x)^2}$ 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 χ^2 » 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).

5 Ancienne version

5.1 Algorithmes de modélisation

On suppose que l'on modélise $y = f(x, a)$ sur N points avec p paramètres (a). On pose $m = N - p$.

5.1.1 Moindres carrés

Lorsqu'on ne définit pas d'incertitude, la méthode utilisée est la méthode des moindres carrés : la grandeur que le logiciel cherche à minimiser est $\Sigma\left(y_{\text{mesure}} - y_{\text{modele}}\right)^2$. La précision relative sur la fonction est définie par racine de $\frac{\Sigma\left(y_{\text{mesure}} - y_{\text{modele}}\right)^2}{\Sigma y_{\text{mesure}}^2}$, c'est donc le rapport entre la moyenne quadratique des écarts fonction théorique / expérimentale et la moyenne quadratique de la fonction expérimentale. On rappelle que la méthode des moindres carrés suppose une incertitude constante sur y et nulle sur x. En supposant que l'écart mesure modèle vient de cette incertitude, on voit que le numérateur est un estimateur de l'écart type, et donc que l'écart relatif vaut, sous cette hypothèse, $\frac{\sigma_y}{y_m}$ si y_m est la moyenne quadratique de y. On a donc ici un moyen de vérifier la cohérence des résultats. Cette même hypothèse permet de calculer l'incertitude sur les paramètres expérimentaux.

5.1.2 Remarque sur les incertitudes

La valeur donnée considère que l'écart théorie expérience provient de l'incertitude des mesures qui est donc ainsi estimée (cf. méthode du χ^2). On peut alors en déduire l'incertitude sur les paramètres. Le tableau ci-dessous a été réalisé en faisant 16 fois la charge d'un condensateur : τ est la constante de temps, $\delta\tau$ est l'incertitude obtenue par modélisation.

Page n°	τ ms	$\delta\tau$ ms
1	1.128	0.0064
2	1.129	0.0055
3	1.124	0.0069
4	1.125	0.006
5	1.125	0.0058
6	1.131	0.0069
7	1.124	0.0055
8	1.128	0.0061
9	1.129	0.0069
10	1.124	0.0062
11	1.125	0.0055
12	1.126	0.0057
13	1.123	0.0063
14	1.12	0.0063
15	1.127	0.0058
16	1.123	0.0057

On obtient un τ moyen $1{,}126 \cdot 10^{-3}$ s; un écart type sur tau $2{,}7 \cdot 10^{-6}$ s. L'incertitude indiquée par Regressi sur τ vaut en moyenne $6 \cdot 10^{-6}$ s correspond à un intervalle de Student à 95 % qui dans ce cas vaut 2,12 fois l'écart type, or $2,7 \times 2,12 = 5,72$. Cela parait relativement cohérent.

5.1.3 Méthode du χ^2

La prise en compte des points est pondéré par l'inverse de leurs incertitudes au carré (variance) la grandeur que le logiciel cherche à minimiser est $\Sigma\left(\frac{y_{\text{mesure}} - y_{\text{modele}}}{\sigma_y}\right)^2$. Lorsqu'il y a également une incertitude sur x, la pondération se fait non par σ_y^2 mais $\sigma_y^2 + f'(x)\sigma_x^2$ (propagation d'erreur) et donc utilisation de la méthode des ellipses. Lors d'une modélisation par équation différentielle, l'incertitude sur x n'est pas prise en compte. Le résultat final est donné sous forme du critère χ^2/m qui devrait être proche de 1 ce qui se comprend aisément : l'écart mesure modélisation doit être du même ordre de grandeur que l'incertitude de mesure. On voit que la valeur du critère χ^2/m être très sensible aux valeurs des incertitudes et qu'une erreur d'un facteur 2 sur celle-ci modifie le critère d'un facteur 4. Il est donc important que l'incertitude soit donnée sous forme de son écart type. Si vous avez, par exemple, des erreurs aléatoires réparties de manière uniforme entre $x - \Delta x$ et $x + \Delta x$, vous devez entrer comme incertitude $\Delta x/\sqrt{3}$ (par définition même de l'écart type), et donc si vous entrez Δx comme incertitude l'optimum de χ^2/m sera aux alentours de $1/3$.

L'incertitude sur les paramètres est tirée de la valeur du χ^2, on montre que pour $\chi^2 = \frac{\chi^2_{\min}}{1 + (n/N)2}$ on obtient un intervalle de confiance de $\pm n\sigma$ sur les paramètres. La recherche de la valeur des paramètres conduisant à cette valeur de χ^2 s'effectue par le même méthode que les moindres carrés. On voit donc que la méthode du χ^2 n'est pas fondamentalement différente, elle apporte essentiellement la pondération des points et la prise en compte de l'incertitude sur x par l'utilisation de la méthode de l'ellipse. Pour des points normalement dispersés, les valeurs des paramètres et l'écart relatif ne devrait donc pas beaucoup évoluer par rapport à la méthode des moindres carrés, et la comparaison incertitude sur y / l'écart mesure modèle revient à comparer χ^2/m à 1.

Pour ce qui est de la valeur du critère, on a vu qu'il y avait une valeur critique de 1, cette valeur étant très sensible aux incertitudes. Si on se rappelle que la variance de $\chi^2(m)$ est $2m$, on voit que le critère χ^2/m devrait être égal à 1 à quelques $1/\sqrt{m}$ près. Pour prendre un exemple avec $m = 9$ (mesures manuelles) le critère se situe entre 0,23 et 2,41 à un taux de confiance de 2% (probabilité de 1% d'avoir $\chi^2/m < 0,23$ et de 99% d'avoir $\chi^2/m < 2,41$). Pour $m = 128$ (mesures automatiques) on obtient 0,733 et 1,315. Dans ce dernier cas, une *bonne* modélisation (*bonne* voulant dire un χ^2/m réel proche de 1 et non pas de 0) et une surestimation de l'incertitude d'un facteur 1,2 conduit à 0,694 donc rejet de l'hypothèse au niveau de test précédent, et de même pour une sous-estimation d'un facteur 1,2 qui conduit à 1,44. Et de même une "mauvaise" modélisation, χ^2/m réel proche de 4, avec une sur estimation d'un facteur 2 conduit à 1 et est donc acceptée. Le test d'un modèle par la valeur du χ^2 suppose donc des hypothèses drastiques sur l'évaluation des incertitudes. Il s'ensuit qu'il est très difficile de donner un taux de confiance sur la modélisation à partir de la valeur du χ^2.

5.2 Bibliographie

Les méthodes de calcul de modélisation proviennent de
Trigeassou, Recherche de modèles expérimentaux, Tec&Doc.
Références BUP :

— BUP 691 Le calcul des incertitudes
— BUP 752 La régression linéaire et ses conditions d'applications
— BUP 725 A propos de la méthode des moindres carrés
— BUP 702 L'anamorphose

et aussi
http ://www.sciences.univ-nantes.fr/physique/perso/cortial/optim