

# Incertitudes dans Regressi

Jean-Michel Millet

6 avril 2020

## 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  $\chi^2$  », la dite méthode, les incertitudes sur  $x$  et  $y$  étant prises en compte.

L'option est accessible par un clic sur le bouton « options », ou par un clic droit dans le panneau modélisation.

Une incertitude peut éventuellement être à zéro (cas du temps fréquemment).

Le test du  $\chi^2$  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.

## 2 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  $(x \cdot pc + N \cdot ms) / \sqrt{3}$  avec  $ms$  valeur correspondant à l'unité du dernier chiffre.

**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)

### 3 Méthode des moindres 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  $\sum_{i=1}^N (y_i - Ax_i - B)^2$ . Cela conduit, en notant  $\Delta = N \sum x^2 - (\sum x)^2$ , à  $B = \frac{\sum x^2 \sum y - \sum x \sum (xy)}{\Delta}$ , et  $A = \frac{N \sum (xy) - \sum x \sum y}{\Delta}$

On peut alors considérer que  $(y_i - Ax_i - B)$  représente l'écart entre la valeur mesurée  $y_i$  et la valeur « vraie »  $Ax_i + B$ , et donc évaluer l'incertitude sur  $y$  par  $\sigma_y^2 = \frac{1}{N-2} \sum_{i=1}^N (y_i - Ax_i - B)^2$ , 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

$$\sigma_B = \sigma_y \sqrt{\frac{\sum x^2}{\Delta}} \text{ et } \sigma_A = \sigma_y \sqrt{\frac{N}{\Delta}}.$$

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.

La précision relative sur la fonction est définie par racine de  $\frac{\sum (y_{\text{mesure}} - y_{\text{modele}})^2}{\sum y_{\text{mesure}}^2}$ , c'est donc le rapport entre la moyenne quadratique des écarts fonction théorique / expérimentale et la moyenne quadratique de la fonction expérimentale.

### 4 Méthode du $\chi^2$

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  $\delta x, \delta y$  variables aléatoires centrées, la variance associée à  $y$  est donc  $\sigma_y^2 + (A\sigma_x)^2$ , par addition des variances. On minimise donc

$\sum_{i=1}^N \frac{(y_i - A - Bx_i)^2}{\sigma_y^2 + (A\sigma_x)^2}$ , c'est la méthode du  $\chi^2$ . Il suffit alors de faire le même calcul qu'au §2, pour déterminer  $A$  et  $B$ , on fait apparaître le coefficient de pondération  $k_i = \frac{1}{\sigma_y^2 + (A\sigma_x)^2}$  dans les sommes.

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

Et de même pour les incertitudes

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

Le test du  $\chi^2$  affiche un  $\chi^2$  réduit  $\frac{\chi^2}{n-p}$ , avec  $n$  nombre de données et  $p$  nombre de paramètres.

Ce  $\chi^2$  réduit devrait être proche de 1, si le modèle est correct et si les incertitudes données pour les grandeurs sont des incertitudes-type.

## 4.1 Comparaison

Dans le cas simple où l'incertitude sur  $x$  est quasi-nulle et celle sur  $y$  constante, on devrait retrouver le même incertitude sur les paramètres que dans le cas précédent, sinon cela signifierait un  $\chi^2$  réduit relativement différent de 1, de mauvais augure. Cela signifie en particulier, que l'incertitude-type sur  $y$  doit être du même ordre de grandeur que l'écart-type données-modèle de la méthode des moindres carrés.

## 4.2 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  $\chi^2/m$  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  $\chi^2/m$  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,  $\chi^2/m$  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  $\chi^2$  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  $\chi^2$ .

## 5 Fonction quelconque

Ce qui est important dans les relations précédentes est que  $y = A + B \cdot f(x)$ , 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  $y = f(x, A, B)$ , 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.

**Méthode du  $\chi^2$**  Cette fois on minimise l'écart quadratique  $y - f(x)$ . On a  $\hat{y} = y + \delta y + \frac{dy}{dx} \delta x$ ,  $\hat{y}$  valeur mesurée,  $y$  valeur « vraie » et  $\delta x, \delta y$  variables aléatoires centrées, la variance associée à  $y$  est donc  $\sigma_y^2 + \left(\frac{df}{dx} \sigma_x\right)^2$ , 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  $\frac{1}{u(y)^2 + \left(\frac{dy}{dx}\right)^2 u(x)^2}$  si  $u(y)$  et  $u(x)$  sont les incertitudes données dans l'onglet correspondant.

## 6 Remarque sur les incertitudes sans $\chi^2$

La valeur donnée considère que l'écart théorie expérience provient de l'incertitude des mesures qui est donc ainsi estimée. 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 :  $\tau$  est la constante de temps,  $\delta \tau$  est l'incertitude obtenue par modélisation.

| Page<br>n° | $\tau$<br>ms | $\delta\tau$<br>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  $\tau$  moyen  $1,126 \cdot 10^{-3}$  s ; un écart type sur tau  $2,7 \cdot 10^{-6}$  s. L'incertitude indiquée par Regressi sur  $\tau$  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.

## 7 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>