Chapitre 11

Déconvolution. Méthode du maximum d'entropie.

1. La théorie de la surprise
2. Le principe du maximum d'entropie
3. Déconvolution par la méthode du maximum d'entropie

La théorie de la surprise

On désire une grandeur qui permette de quantifier la notion de ***surprise*** causée par la réalisation d'un événement (ou la confirmation d'une proposition)

Il est raisonnable de supposer que la surprise est une fonction de la probabilité de l'événement (ou du degré de confirmation de la proposition)

Soit

$$S = S(p) \qquad (0 < p \leq 1)$$

la quantité mesurant la surprise

Pour constituer une mesure raisonnable, cette quantité doit satisfaire à un certain nombre d'axiomes.

Axiome 1 :

$$S(1) = 0$$

(on n'est pas surpris d'apprendre qu'un événement certain s'est produit)

Axiome 2 :

$$p < q \quad \Rightarrow S(p) > S(q)$$

(plus l'événement est improbable, plus grande est notre surprise s'il se produit)

Axiome 3 :

$$S(p) \text{ est une fonction continue de } p$$

(un petit changement de p correspond à un petit changement de S)

Axiome 4 :

$$S(pq) = S(p) + S(q) \quad (0 < p \leq 1, \; 0 < q \leq 1)$$

(Soient E et F deux événements indépendants :

$$P(E) = p, \;\; P(F) = q, \;\; P(E \cdot F) = pq$$

On sait que E s'est produit
Si F se produit, la surprise vaut $S(pq) - S(p)$
mais aussi $S(q)$ car E et F indépendants)

Théorème :

Si $S(p)$ satisfait aux axiomes $A1$ à $A4$, alors

$$S(p) = -C \ln p$$

où C est un réel positif arbitraire

Démonstration :

$m, n =$ entiers positifs

$r =$ rationnel positif, $x =$ réel positif

$$(A4) \quad \Rightarrow \quad S(p^2) = 2\, S(p)$$

$$\Rightarrow S(p^m) = m\, S(p)$$

$$p = p^{1/n} \cdots p^{1/n} \Rightarrow S(p^{1/n}) = \frac{1}{n}\, S(p)$$

$$\Rightarrow S(p^{m/n}) = \frac{m}{n}\, S(p)$$

$$\Rightarrow S(p^r) = r\ S(p)$$

$$(A3) \quad \Rightarrow \quad S(p^x) = x\ S(p)$$

Posons

$$x = -\ln p \quad (0 < p \leq 1)$$

$$\Rightarrow p = (1/e)^x$$

$$\Rightarrow S(p) = S[(1/e)^x] = x\ S(1/e) = -C \ln p$$

$$\text{où } C =: S(1/e) > S(1) = 0 \qquad (A1,\ A2)$$

En général, on adopte $C = 1$

Soit x une variable aléatoire, pouvant prendre les valeurs $x_1, \ldots, x_N$, avec les probabilités $p_1, \ldots, p_N$

La surprise causée par l'événement « x prend la valeur x_i » est égale à :

$$S(p_i) = -\ln p_i$$

$\Rightarrow$ la quantité moyenne de surprise à laquelle on peut s'attendre en apprenant la valeur de x vaut :

$$H(x) = -\sum_{i=1}^{N} p_i \ \ln p_i$$

$H(x)$ = *entropie* de la variable aléatoire x

= quantité d'incertitude quant à la valeur de x

= quantité moyenne d'*information* reçue en observant la valeur de x

Le principe du maximum d'entropie

Soit x une variable aléatoire, pouvant prendre les valeurs x_i $(i = 1, \ldots, N)$

Soit p_i la probabilité (*inconnue* !) que x prenne la valeur x_i

Supposons que l'on connaisse la valeur moyenne d'une fonction $f(x)$:

$$< f(x) >= \sum_{i=1}^{N} p_i \, f(x_i) = F_0$$

Sur la base de cette information *uniquement*, on demande d'estimer la valeur moyenne d'une autre fonction $g(x)$

Données:

(1) $\Sigma_{i=1}^{N} p_i = 1$

(2) $\Sigma_{i=1}^{N} p_i \, f(x_i) = F_0$

$\Rightarrow$ 2 équations pour N inconnues !

Comment choisir les p_i sur la base d'informations incomplètes ?

(1) ***Bernoulli, Laplace***,...

En l'absence d'indications contraires, supposer tous les p_i égaux

(2) ***Jaynes*** (1957)

Choisir les p_i qui, tout en satisfaisant aux contraintes, maximisent l'entropie

$\Rightarrow$ sélection de la distribution « la moins informative », « la moins biaisée »

$\Rightarrow$ *Principe du maximum d'entropie*

Déconvolution par la méthode du maximum d'entropie

Supposons le modèle, les données et le profil instrumental normalisés :

$$\sum_{i=1}^{N} f_i = 1$$

où f_i est la fraction de la grandeur à mesurer contenue dans l'échantillon i

$\Rightarrow$ les f_i satisfont aux axiomes de la théorie des probabilités

$\Rightarrow$ ils peuvent être interprétés comme des probabilités

Exemple :

En imagerie, f_i = probabilité que le prochain photon soit émis dans la « zone » i

$\Rightarrow$ on définit l'entropie du modèle :

$$H = -\sum_{i=1}^{N} f_i \, \ln f_i$$

$\Rightarrow$ parmi tous les modèles qui sont compatibles avec les données, on choisit celui qui a l'entropie la plus grande

Statistique gaussienne :

Minimiser :

$$-H = \sum_{i=1}^{N} f_i \, \ln f_i$$

sous la contrainte de normalisation :

$$\sum_{i=1}^{N} f_i = 1$$

et celle de compatibilité avec les données :

$$\chi^2 \equiv \sum_{i=1}^{N} \left(\frac{d_i - \sum_{j=1}^{N} s_{ij} f_j}{\sigma_i} \right)^2 = \chi_0^2$$

avec $\chi_0^2 \sim N$

Solution par méthode des paramètres de Lagrange :

minimiser :

$$F = \sum_{i=1}^{N} f_i \, \ln f_i + \lambda \left(\sum_{i=1}^{N} f_i - 1 \right) + \mu \left(\chi^2 - \chi_0^2 \right)$$

λ et μ déterminés pour satisfaire aux deux contraintes

Déconvolution de l'exemple 5

Méthode du maximum d'entropie