Chapitre 2

Caractéristique de la loi a posteriori

Nous avons vu que dans l'approche bayésienne la loi *a posteriori* $p(\boldsymbol{\theta}|\boldsymbol{x})$ contient toute information concernant les paramètres $\boldsymbol{\theta}$ aprés l'observations des données. On peut alors décider de transmettre cette loi à l'utilisateur comme telle qui en fera ce que lui semble bon. Mais une telle attitude n'est pas trés responsable, surtout pour un ingénieur qui cherche à fournir des réponses concrètes aux problèmes qui a à résoudre. Souvent on cherche à caractériser cette loi *a posteriori* par un nombre trés limité (souvent un ou deux) de paramètres.

Quand il s'agit d'un seule paramètre on peut caractériser la loi *a posteriori* $p(\theta|\boldsymbol{x})$ par des quantités suivantes :

– la mode

$$\mathrm{M}\left[\theta\right] = \sup_{\theta\in\Theta} p(\theta|\boldsymbol{x}),$$

– la moyenne

$$\mathrm{E}\left[\theta\right] = \int_{\theta\in\Theta} \theta\, p(\theta|\boldsymbol{x})\,\mathrm{d}\theta,$$

– la variance

$$\mathrm{Var}\left[\theta\right] = \int_{\theta\in\Theta} \left[\theta - \mathrm{E}\left[\theta\right]\right] p(\theta|\boldsymbol{x})\,\mathrm{d}\theta,$$

– les α–quantiles

$$Q_\alpha(\theta) : \Pr\left\{\theta \leq Q_\alpha(\theta)\right\} = \alpha,$$

– la médiane

$$\text{Méd}\left[\theta\right] = Q_{\frac{1}{2}}(\theta) : \Pr\left\{\theta \leq \text{Méd}\left[\theta\right]\right\} = \frac{1}{2},$$

– la région de la plus heute probabilité ou le support α–interquantile

$$[a,b] = \left[Q_{(1-\alpha)/2}(\theta), Q_{(1+\alpha)/2}(\theta)\right] : \Pr\left\{a \leq \theta \leq b\right\} = 1-\alpha,$$

Dans le cas d'un vecteur de paramètres $\boldsymbol{\theta}$ on peut bien entendu définir les lois marginales pour chaque paramètres θ_k

$$p(\theta_k|\boldsymbol{x}) = \int p(\boldsymbol{\theta}|\boldsymbol{x})\,\mathrm{d}\theta_1 \cdots \mathrm{d}\theta_{k-1}\,\mathrm{d}\theta_{k+1} \cdots \mathrm{d}\theta_K, \quad k = 1, \cdots, K$$

et les caractériser par les grandeurs que nous avons définis précédement :

– la mode :

$$\mathrm{M}\left[\theta_i\right] = \sup_{\theta_i\in\Theta_i} p(\theta_i|\boldsymbol{x}),$$

– la moyenne :
$$\mathrm{E}\left[\theta_i\right] = \int_{\theta_i \in \Theta_i} \theta_i \, p(\theta_i|\boldsymbol{x}) \, \mathrm{d}\theta_i,$$
– la variance :
$$\mathrm{Var}\left[\theta_i\right] = \int_{\theta_i \in \Theta_i} \left[\theta_i - \mathrm{E}\left[\theta_i\right]\right] p(\theta_i|\boldsymbol{x}) \, \mathrm{d}\theta_i,$$
Mais cela ne suffit pas car il faut aussi caractériser la corrélations entre les différents paramètres :
  – la covariance :
$$\mathrm{Cov}\left[\theta_i, \theta_j\right] = \iint_{\boldsymbol{\theta} \in \Theta} \left[\theta_i - \mathrm{E}\left[\theta_i\right]\right] \left[\theta_j - \mathrm{E}\left[\theta_j\right]\right] p(\boldsymbol{\theta}|\boldsymbol{x}) \, \mathrm{d}\boldsymbol{\theta}.$$
  – la corrélation :
$$\mathrm{Cov}\left[\theta_i, \theta_j\right] = \iint_{\boldsymbol{\theta} \in \Theta} \theta_i \, \theta_j \, p(\boldsymbol{\theta}|\boldsymbol{x}) \, \mathrm{d}\boldsymbol{\theta}.$$
Notons aussi que la mode $\mathrm{M}[\theta_i]$ n'est pas forcément égale à $\mathrm{M}[\boldsymbol{\theta}]_i$ où:
$$\mathrm{M}\left[\boldsymbol{\theta}\right] = \sup_{\boldsymbol{\theta} \in \Theta} p(\boldsymbol{\theta}|\boldsymbol{x})$$
Dans le cas de deux paramètres on peut encore dessiner la fonction $p(\theta_1, \theta_2|\boldsymbol{x})$ sous formes de nappes $z = p(\theta_1, \theta_2|\boldsymbol{x})$ ou d'une image avec des niveaux de couleurs ou encore sous forme de tracé de contours (iso densités). Ceci n'est plus facile dans le cas où on a plus de deux paramètres.

A illustrer avec des exemples

En pratique l'utilisateur se contente d'une valeur pour ses paramètres (une estimée) et éventuellement d'une valeur qui lui donnerais une estimation de la précision de cette estimation.

Quand il s'agit d'un seule paramètre, souvent on se contente d'une fonction dépendant des observations (un estimateur) qui fournit une valeur (une estimée) pour une observation donnée.

Si on pouvait calculer analytiquement l'expression exacte de la loi *a posteriori* alors on peut

Quand il s'agit d'un seule paramètre il peut tout simplement tracer la courbe. Il faut alors fournir des outils pour condenser