Combinaison de dires d'experts en élicitation de lois a priori. Application à un modèle dose-réponse pour Listeria chez la souris.

Exposé AppliBugs

ISABELLE ALBERT

18 / 12 / 2013

INTRODUCTION

Cet exposé présente une partie du travail publié dans l'article :

I. Albert, S. Donnet, C. Guihenneuc-Jouyaux, S. Low Choy, K. Mengersen, J. Rousseau.
Combining expert opinions in prior elicitation
(with Discussion (French et Gosling)),
Bayesian analysis, 7(3): 503-532, 2012.

Définition :
« Éliciter » : est l'action d'aider un expert ou des experts à formaliser leurs connaissances pour permettre de les sauvegarder, les partager et/ou les utiliser.

Contexte : Statistique bayésienne
Objectif : Construire la loi a priori à partir de dires d'experts.
Littérature abondante sur le sujet d'un point de vue psychologique et statistique.
Proposer une approche cohérente avec la démarche bayésienne (les données élicitées vont mettre à jour un a priori vague) et pouvant prendre en compte d'éventuelles interactions entre experts.

SOMMAIRE

_01

Contexte bayésien et loi a priori

Contexte bayésien et loi a priori

Y = Données, Modèle paramétrique : $Y|\theta \sim M_\theta$
θ = paramètres

Loi a priori sur θ : $[\theta]$

Vraisemblance (données) : $[Y|\ \theta]$

Loi a posteriori sur θ : $[\theta|Y]$ par le théorème de Bayes : $[\theta|Y] \propto [Y|\ \theta]\ [\theta]$

- Choix de la loi a priori
  - **Pas d'information** sur θ (sauf sur le support) : loi a priori « plate » (gaussienne avec une large variance) ou loi a priori de Jeffreys (utilisant l'information de Fisher)
  - **Des études antérieures** : la loi a posteriori d'études antérieures peut servir de loi a priori
  - **Prise en compte de l'avis d'experts** pour construire la loi a priori

_02

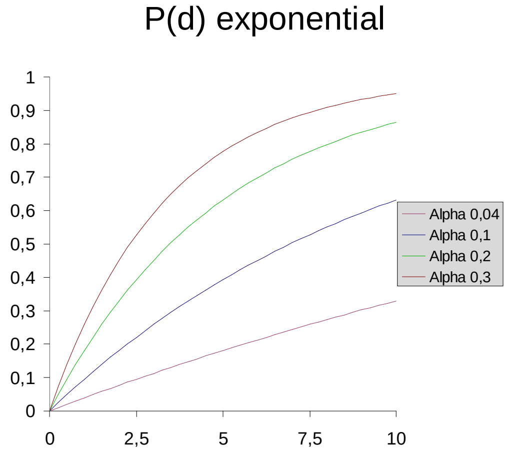
Exemple en risque alimentaire

Modèle dose-réponse à Listeria

- Choix d'un modèle exponentiel :

$$p(d,\alpha)=1-e^{-\alpha d},\ \alpha>0$$

où $p(d,\alpha)$ est la probabilité pour une souris de mourir avec la dose d.



- Des études antérieures ont montré que α est très mal connu après inférence (grande imprécision sur son estimation).
- Donc utiliser de l'avis d'experts biologistes réalisant ces expériences pour mieux estimer α.
- Utiliser la statistique bayésienne et construire une loi a priori informative $[\alpha]$ à partir des dires de ces experts.

_03

Construction de la loi a priori à partir des dires d'experts

Construction de la loi a priori

Soit $\boldsymbol{D} = \{D_1,\ldots,D_N\}$ les données fournies par les experts. Ces données sont appelées ***données élicitées*.**

Difficultés pour définir l'a priori :

- les experts ont une connaissance limitée, ne connaissent pas la loi sur tout son support
- en un temps raisonnable, on ne peut leur poser qu'un nb de questions limité, donc ne fournissent souvent qu'une petite quantité de données

Donc raisonnable de supposer que [θ] appartienne à une famille paramétrique :

$$[\theta] \in \left\{ [\theta|\lambda], \lambda \in \Lambda \right\}$$

Donc **But** de l'élicitation **: trouver les valeurs de λ les plus appropriées pour représenter le savoir des experts**

- *Dans l'exemple :* $p(d,\alpha) = 1 - e^{-\alpha d}$
  - $\theta = \alpha$
  - $\log(\alpha) \sim N(\mu, \sigma^2)$ donc $\lambda = (\mu, \sigma^2)$

Construction de la loi a priori

Processus d'élicitation :

1. Préparation : sélection et motivation des experts, formation aux probabilités.
2. Définir les quantités à éliciter : interroger les experts sur des quantités tangibles pour eux.
3. Interroger les experts
4. Traitement statistique : estimation de λ à partir des données élicitées
5. Validation auprès des experts

Nature des données élicitées :

- encoder le savoir des experts en distributions de probabilités donc les interroger « **en probabilité** » ;
- aperçu plus cohérent de leur savoir en utilisant **différents types** de questions : des probabilités et des quantiles par exemple d'une quantité tangible X.

Dans l'exemple : modèle dose-réponse à Listeria

Problème **: α n'a pas d'interprétation concrète pour nos experts (paramètre de forme sur les courbes)**

- Élicitation de probabilités :
  - Question structurelle (directement sur $p(d,\alpha)$) :

    « Vous avez fait un grand nombre d'expériences, donnez le pourcentage d'expériences qui ont conduit à moins de q = 20% de souris mortes.» (avec aussi q = 60%)

    - Réponse théorique pour chaque expert i :

$$P_{iq}^{*} = P\left(p(d,\alpha) \leq q \middle| \mu_i, \sigma_i^2\right)$$

avec $\log(\alpha) \sim N\left(\mu_i, \sigma_i^{2}\right)$

    - Traitement statistique :

$$\begin{cases} \mu_i = \sigma_i \Phi^{-1}\left(P_{iq_1}\right) - \dfrac{\log(1-q_1)}{d} \\ \sigma_i = \dfrac{-\log(1-q_1)+\log(1-q_2)}{d\left(\Phi^{-1}\left(P_{iq_1}\right) - \Phi^{-1}\left(P_{iq_2}\right)\right)} \end{cases}$$

Mais questions difficiles pour les experts

Dans l'exemple : modèle dose-réponse à Listeria

Problème **: α n'a pas d'interprétation concrète pour nos experts (paramètre de forme sur les courbes)**

Idée : interroger sur X_n = **Nombre de souris mortes sur n souris ayant reçu une même dose d de Listeria**

$$X_n \sim Bin\left(n, p\left(d, \alpha\right)\right)$$

- Élicitation de probabilités :
  - Question **prédictive** :

« Une expérience consiste en l'injection de la dose ***d*** à 10 souris. Dans combien d'expériences parmi 100, obtiendriez-vous moins de ***q*** = 2 souris mortes parmi les 10 ? » (aussi avec ***q*** = 6)

- Réponse théorique pour chaque expert ***i*** :

$$P_{iq}^* = P\left(X_{10} \leq q \middle| \mu_i, \sigma_i^2\right)$$

où $X_{10} \sim Bin\left(10, p\left(d, \alpha\right)\right)$ avec $\log\left(\alpha\right) \sim N\left(\mu_i, \sigma_i^2\right)$

Quantités observables mais plus difficile de trainement

- Traitement statistique :

$$\left(\mu_i, \sigma_i\right) = \arg\min_\lambda \sum_q \left[P_{iq} - P_{iq}^*\right]^2$$

avec $P_{iq}^* = \sum_{r=0}^{q} C_{10}^r \int_0^\infty \left(1 - e^{-td}\right)^r e^{-(10-r)td} f\left(\left(\log t - \mu_i\right) / \sigma_i\right) \sigma_i^{-1} t^{-1} dt$ (réponse théorique)

Dans l'exemple : modèle dose-réponse à Listeria

- Élicitation de quantiles :
  - **Question prédictive** :

    « Une expérience consiste en l'injection de la dose *d* à 10 souris. Vous faites un grand nombre d'expériences, quel serait le nombre *Q* de souris mortes parmi les 10 tel que *t**100% des expériences aboutissent à un nombre de souris mortes inférieur ou égal à *Q* ? (avec *t* = 0.1,0.5 et 0.9)

    - Réponse théorique pour chaque expert *i* :

$$P\left(X_{10} \le Q^{*}_{it} \middle| \mu_i, \sigma_i^2\right) = t$$

où $X_{10} \sim Bin(10, p(d,\alpha))$ avec $\log(\alpha) \sim N\left(\mu_i, \sigma_i^2\right)$

  - Question structurelle :

    « Quelle serait la proportion *Q* de souris mortes telle que *t**100% des expériences aboutissent à un nombre de souris mortes inférieur ou égal à *Q* ? (avec *t* = 0.1,0.5 et 0.9)

    - Réponse théorique pour chaque expert *i* :

$$P\left(p(d,\alpha) \le Q^{*}_{it} \middle| \mu_i, \sigma_i^2\right) = t$$

où $\log(\alpha) \sim N\left(\mu_i, \sigma_i^2\right)$

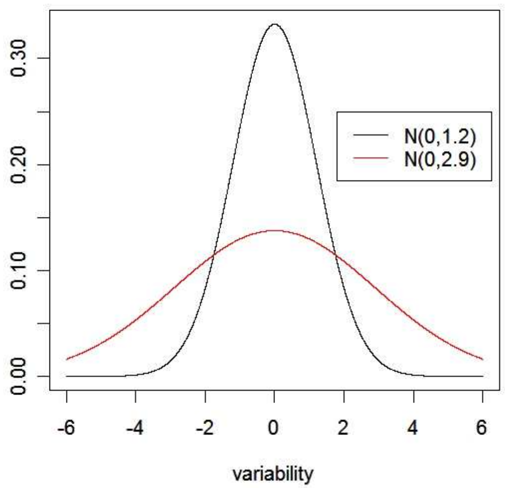
Dans l’exemple : modèle dose-réponse à Listeria

- Traitement statistique des quantiles :
  - Prise en compte de l’incertitude des experts sur les quantiles donnés :
    - À chaque quantile donné, les experts associent un degré de confiance c_{it} compris entre 1 et 10.
    - Donc on peut construire un modèle d’erreur de type Probit : (vraisemblance des données élicitées)

$$\begin{cases} \Phi^{-1}(Q_{it}) = \Phi^{-1}(Q^*_{it}) + \varepsilon_{it} \\ \varepsilon_i \sim N(0, v_{it}) \end{cases}$$

où $Q^*_{i,t}$ est le quantile théorique $P\left(p(d,\alpha) \le Q^*_{it} \middle| \mu_i, \sigma_i^2\right) = t$

(d’où $Q^*_{it} = 1 - e^{-d\exp\left(\mu_i + \sigma_i \Phi^{-1}(t)\right)}$) et v_{it} est reliée au degré de confiance : (q_i* précision fixée par le statisticien)



$P(|\ \Phi^{-1}(Q_{it}) - \Phi^{-1}(Q^*_{it})\ | < q_i^*) = c_{it} \Leftrightarrow P(-\ q_i^* \le \varepsilon_{it} \le q_i^*) = c_{it}$

$\Leftrightarrow \sqrt{v_{it}} = q_i^* / \Phi^{-1}((c_{it}+1)/2)$

Par ex. : $q_i^* = 1.96$, si $c_{it} = 90\% \rightarrow \sqrt{v_{it}} = 1.2$
et si $c_{it} = 50\% \rightarrow \sqrt{v_{it}} = 2.9$

_04

Combiner des experts

Combiner des experts

❖ Soient plusieurs experts interrogés. De chaque experts i, on déduit une loi a priori mais comment obtenir une loi unique ?

$$\forall i, \left[\theta \middle| \hat{\lambda}_i\right] \to [\theta] ?$$

1. Approche par mélange

$$[\theta] = \sum_i w_i \left[\theta \middle| \hat{\lambda}_i\right]$$

   modélise les différences entre experts

2. Approche par moyenne

$$\hat{\lambda} = \sum_i w_i \hat{\lambda}_i \Rightarrow [\theta] = \left[\theta \middle| \hat{\lambda}\right]$$

   modélise un consensus entre experts

Difficultés :

Comment choisir les poids w_i ? Comment prendre en compte des interactions entre experts (non indépendance des experts s'ils proviennent par exemple de mêmes instituts) ?

Proposition : **Modélisation hiérarchique**

Permet de prendre en compte leurs connaissances communes
mais permet une variabilité entre et à l'interieur des groupes d'experts

Formulation du modèle hiérarchique

- Soient J groupes de N_j experts indicés par j (i l'expert, et au total N experts) :
  - **Modèle hiérarchique** :

$$\lambda_{ij} \overset{i.i.d}{\sim} g\left(.\middle|\lambda_j, b_j\right) \forall\, i = 1,\ldots,N_j,$$

$$\lambda_j \overset{i.i.d}{\sim} g\left(.\middle|\lambda, b\right), \forall\, j = 1,\ldots,J,$$

$$\lambda \sim \pi_0.$$

    - Les experts regroupés en classes homogènes avec une même distribution (première ligne)
    - La connaissance des groupes liée à une distribution commune (ligne 2)
    - π_0 **loi peu informative** représentant l'incertitude globale sur λ. λ peut être vu comme le paramètre de consensus entre experts.
  - Dans l'exemple : $\lambda_j = \left(\mu_j, \sigma_j^2\right), b_j = \left(\tau_j, \xi_j\right), \lambda = \left(\mu, \sigma^2\right)$ et $b = \left(\tau, \xi\right)$

$$\mu_{ij} \left|\mu_j, \tau_j \overset{i.i.d}{\sim} N\left(\mu_j, \tau_j\right) \text{ et } \frac{\sigma_{ij}^2}{\sigma_j^2}\right|\sigma_j, \xi_j \overset{i.i.d}{\sim} \Gamma\left(\xi_j, \xi_j\right) \forall\, j = 1,\ldots,J,\; i = 1,\ldots,N_j$$

$$\mu_j \left|\mu, \tau \overset{i.i.d}{\sim} N\left(\mu, \tau\right) \text{ et } \frac{\sigma_j^2}{\sigma^2}\right|\sigma^2, \xi \overset{i.i.d}{\sim} \Gamma\left(\xi, \xi\right) \forall\, j = 1,\ldots,J$$

$$\mu \sim N\left(\mu_0, V\right) \text{ et } \sigma^2 \sim \sigma_0^2 \Gamma\left(a, a\right)$$

Formulation du modèle hiérarchique

- Traitement statistique (proche du bayesien empirique) :
  - **Idée : loi [α] = loi a posteriori [α|D]** (approche supra-bayésienne) **de la forme :**

$$\left[\alpha|D\right] \propto \int \left[\alpha|\lambda\right] \prod_{ijt} \left[D_{ijt}|\lambda_{ij}\right] \prod_{ij} \left[\lambda_{ij}|\lambda_j, b_j\right] \prod_j \left[\lambda_j|\lambda, b\right] \pi_0(\lambda)\, d\lambda_{ij}\, d\lambda_j\, d\lambda$$

  - Dans l'exemple (5 experts) : **2** gps d'experts : gp « Pasteur » avec **3** experts et gp « INRA-Tours » avec **2** experts – **Deux étages dans l'estimation**
    1) Estimation des hyper-paramètres τ_j, ξ_j, τ, ξ par la méthode des moments en utilisant les **probabilités élicitées** : $\left(\hat{\mu}_{ij}, \hat{\sigma}_{ij}\right) = \arg\min_{\lambda} \sum_q \left[P_{ijq} - P^*_{ijq}\right]^2$

– $\hat{\tau}_j = \dfrac{1}{N_j - 1} \sum_{i=1}^{N_j} \left(\hat{\mu}_{ij} - \hat{\mu}_j\right)^2$ avec $\hat{\mu}_j$ la moyenne des $\hat{\mu}_{ij}$ dans chaque gp j.

– $\hat{\tau} = \dfrac{1}{J - 1} \sum_{j=1}^{J} \left(\hat{\mu}_j - \hat{\mu}\right)^2$ avec $\hat{\mu}$ la moyenne des $\hat{\mu}_j$

– $\hat{\xi}_j^{-1} = \dfrac{1}{N_j - 1} \sum_{i=1}^{N_j} \left(\dfrac{\hat{\sigma}_{ij}^2}{\hat{\sigma}_j^2} - 1\right)^2$ avec $\hat{\sigma}_j$ la moyenne des $\hat{\sigma}_{ij}$ dans chaque gp j

– $\xi^{-1} = \dfrac{1}{J - 1} \sum_{j=1}^{J} \left(\dfrac{\hat{\sigma}_j^2}{\hat{\sigma}^2} - 1\right)^2$ avec $\hat{\sigma}^2$ la moyenne des $\hat{\sigma}_j^2$

– $\hat{V} = \dfrac{1}{J} \sum_{j=1}^{J} \hat{\tau}_j + \hat{\tau},\ \hat{a}^{-1} = \dfrac{1}{J} \sum_{j=1}^{J} \xi^{-1}{}_j + \hat{\xi}^{-1},\quad \mu_0 = \hat{\mu},\quad \sigma_0^2 = \hat{\sigma}^2.$

Formulation du modèle hiérarchique

- Traitement statistique :
  - **Dans l'exemple** :

    2) **Loi [α|D] approchée par un algorithme MCMC** (programmable sous jags) en utilisant **les quantiles élicités** et **les hyper-paramètres obtenus précédemment « plugés »** (mis directement dans la vraisemblance)

  - Comparaison avec la méthode par « moyenne » où λ est remplacé par les estimateurs de la méthode des moments sur toutes les données élicitées :

$$\log \alpha \sim N\left(\hat{\mu}, \hat{\sigma}^2\right)$$

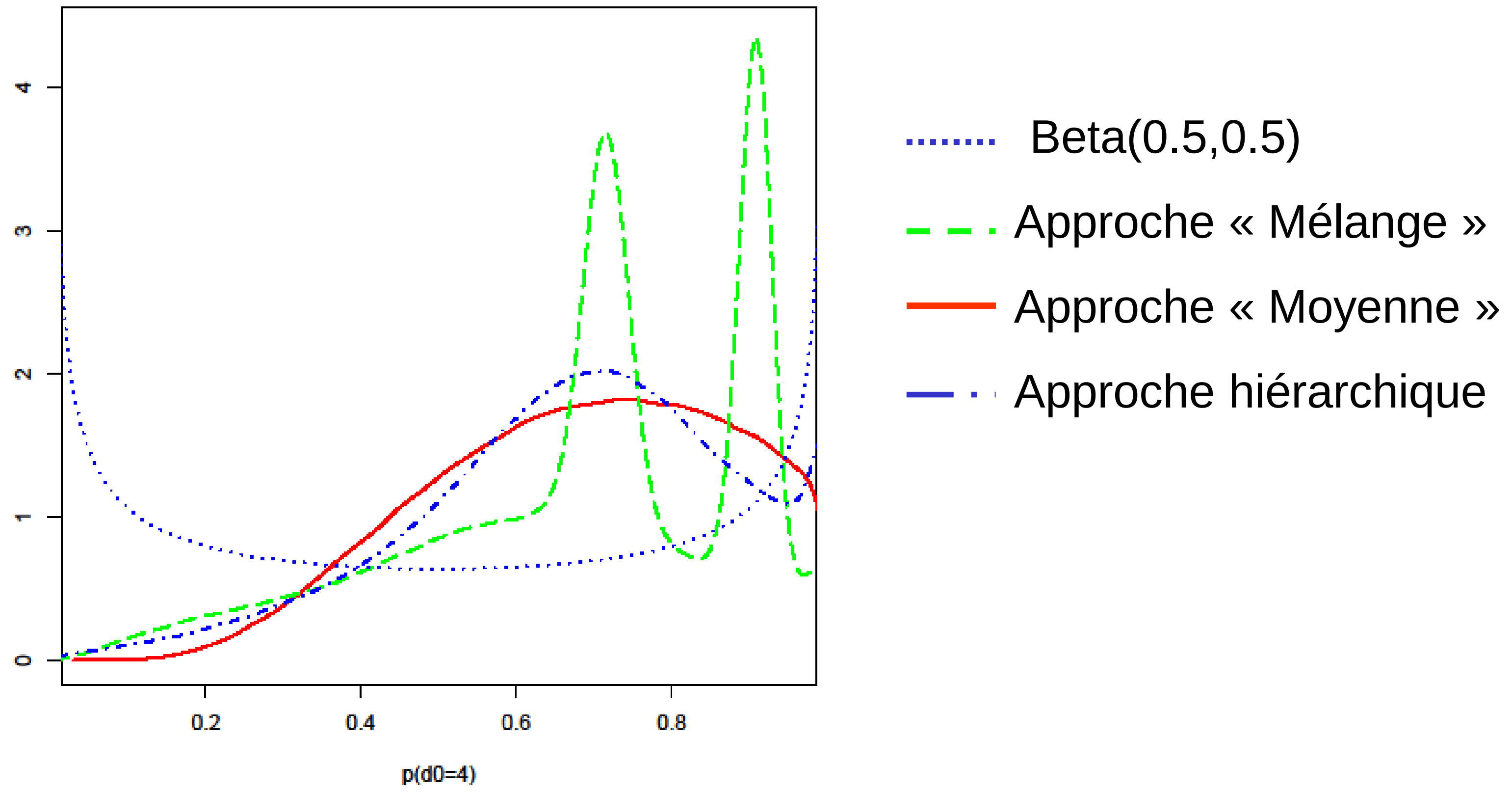
  - Comparaison avec la méthode par mélange (N experts) avec les λ_{ij} obtenus sur toutes les données élicitées par la minimisation des moindres carrés :

$$\log \alpha \sim \frac{1}{N} \sum_{ij} N\left(\hat{\mu}_{ij}, \hat{\sigma}_{ij}{}^2\right)$$

_05

Résultats

Résultats : Listeria/souris



_06 Conclusions et Discussion

Conclusions et discussion

- Approche qui garde la cohérence de l'approche d'apprentissage via la mise à jour bayésienne de la loi a priori par les données élicitées
- Approche qui utilise le cadre de la modélisation hiérarchique pour combiner les dires d'experts (*vers un consensus*) et tenir compte des interactions éventuelles entre experts
- Combinaison d'élicitations indirectes (sur des observables) de différents experts et de différentes natures (proba. et quantiles)
- Prise en compte de l'incertitude des experts en leur demandant leur confiance en leur répons (*envisager un autre type de calibration*)
- Chaque expert peut choisir la dose où il est le plus à l'aise

Conclusions et discussion

- Dans l'article, étude de simulation qui fait varier le nombre d'experts par groupe (même nombre, fort déséquilibre), la composition des groupes (deux groupes alors qu'il y en a qu'un)
- Autre exemple, modélisation de la durée d'une thèse en mathématiques dans une université australienne (3 groupes, modèle log-normal, 1 étape d'estimation)
- Feedback : consensus ?
- Modélisation plus fine : ex. : Déséquilibrer les gps : l'importance d'un groupe vs un autre ou un gp qui a un biais systématique : (connaissance autre) $\lambda_1 \overset{i.i.d}{\sim} g\left(.\middle|\lambda+\Delta,b\right), \Delta>0$ ou $\lambda_1 \overset{i.i.d}{\sim} g\left(.\middle|\lambda,bb'\right), bb'>b$

Merci de votre attention

ISABELLE ALBERT