

Elicitation de loi a priori par dires d'experts. Application en risque alimentaire

I. Albert¹, S. Donnet², C. Guihenneuc³, S. Low Choy⁴, K.
Mengersen⁴ et J. Rousseau²

¹Unité meta@risk, INRA, Paris

²Université Paris-Dauphine, Cérémade

³Université Paris-Descartes, MAP5 et Inserm U754

⁴Queensland University of Technology, Brisbane, Australia

Journée de la recherche 2011 - mardi 22 mars

Listeria monocytogenes

- ▶ On s'intéresse au pouvoir pathogène de la Listeria monocytogenes
- ▶ Bactérie largement répandue dans la nature
- ▶ Peu pathogène pour l'homme mais capable d'infections sévères chez les individus fragilisés (immuno-déprimés , femmes enceintes, nouveau-nés)
- ▶ Maladie rare mais grave : la mortalité est estimée à 25 – 30% avec 40% de séquelles neurologiques.
- ▶ Contamination humaine est d'origine alimentaire.
- ▶ Aliments à risque :
 - ▶ charcuteries (pâté, rillettes, langue de porc, beefsteak haché) ,
 - ▶ certains produits laitiers (fromages au lait cru) ,
 - ▶ certains poissons fumés (saumon),
 - ▶ certains végétaux (soja).

Expérience in vivo et modélisation

Injection par intraveineuse d'une dose d de *Listeria monocytogenes* à n souris.

On note X_d le nombre de souris mortes après l'injection parmi les n

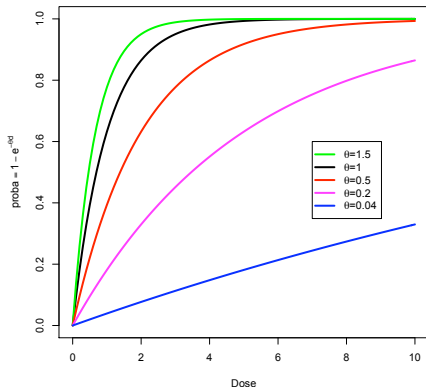
$$X_d \sim \text{Bin}(n, p) \quad \text{avec} \quad p = p(d, \theta) = 1 - e^{-\theta d}$$

- ▶ $p = p(d, \theta)$ la probabilité pour une souris de mourir avec une dose d de Listeria [Haas & al. (1999)] .

Objectif : estimer le paramètre θ

A propos du paramètre θ

- ▶ Caractérise le caractère pathogène.
- ▶ Propre à l'espèce de souris, à la souche de *Listeria*, etc.
- ▶ Paramètre de forme, sans interprétation concrète.



Estimation de θ

► **Estimation :**

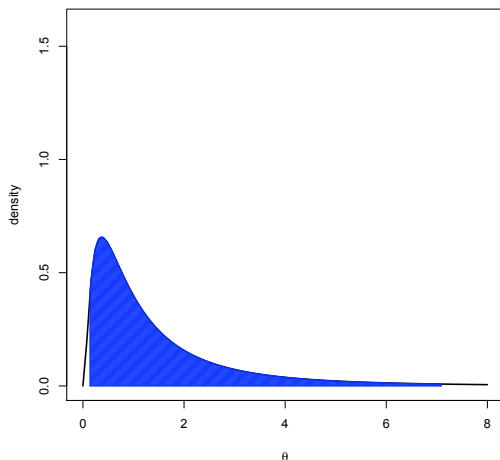
$$\hat{p} = \frac{X_d}{n} \quad \Rightarrow \quad \hat{\theta} = -\frac{1}{d} \log(1 - \hat{p})$$

- **Mais :** Puisqu'on a peu d'observations \Rightarrow Grande incertitude sur θ
- **Idée :** Utiliser des experts¹ pour améliorer notre connaissance de θ .

¹Biologistes, spécialistes du domaine

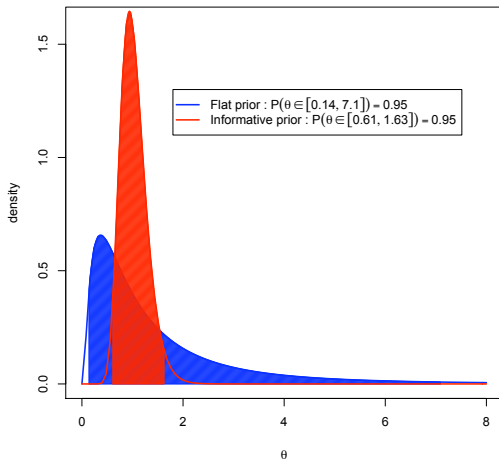
Cadre bayésien : loi a priori

- ▶ Prise en compte de l'incertitude sur θ : θ suit une certaine loi de probabilité π dite "loi a priori".
- ▶ Par exemple : $\log \theta \sim \mathcal{N}(\mu, \sigma^2)$



Cadre bayésien : loi a priori

- ▶ Prise en compte de l'incertitude sur θ : θ suit une certaine loi de probabilité π dite "loi a priori".
- ▶ Par exemple : $\log \theta \sim \mathcal{N}(\mu, \sigma^2)$



Cadre bayésien : loi a posteriori

Théorème de Bayes :

$$\pi(\theta|X_d) = \frac{f(X_d|\theta)\pi(\theta)}{\pi(X_d)}$$

Estimateur de θ : l'espérance ou le mode de la loi a posteriori

Elicitation

Problématique

Comment construire une loi a priori informative en utilisant le savoir d'experts du domaine ?

Définition : "Eliciter" est l'action d'aider un expert à formaliser ses connaissances pour permettre de les sauvegarder et/ou les partager.

But : encoder le savoir des experts en distributions de probabilité.

- ▶ Difficile pour l'esprit humain de quantifier un savoir qualitatif.
- ▶ Largement étudié d'un point de vue mathématique ET psychologique [O'Hagan & al. (2006)]

Qu'attendre d'un expert ?

- ▶ $\log \theta \sim \mathcal{N}(\mu, \sigma^2)$. On cherche μ et σ^2 .
- ▶ On ne peut pas poser de questions directement sur θ
- ▶ On va poser des questions sur $p(d, \theta) = 1 - e^{-\theta d}$ ou sur X_d

Questions sur $p(d, \theta)$: questions structurelles

Probabilités

Pour une dose d^1 , dans combien d'expériences parmi 100 obtiendrez-vous moins de 20% (ou moins de 60%) de souris mortes ?

réponse = $\mathbb{P}(p(d, \theta) \leq 0.2 | \mu, \sigma^2)$ où $p(d, \theta) = 1 - e^{-\theta d}$ et $\log \theta \sim \mathcal{N}(\mu, \sigma^2)$

Quantiles

Quelle serait la proportion q de souris mortes telle que 10% (ou 50% ou 90%) des expériences aboutissent à une proportion de souris mortes inférieure ou égale à q ?

réponse = q tel que $\mathbb{P}(p(d, \theta) \leq q | \mu, \sigma^2) = 20\%$

- ▶ Quantités non observables donc difficile pour les experts mais facile pour les statisticiens.

Questions sur X_d : questions prédictives

Probabilités

Une expérience consiste en l'injection de la dose d^1 à 10 souris.
Dans combien d'expériences parmi 100 obtiendrez-vous moins de 2
(ou 6) souris mortes parmi les 10 ?

réponse = $\mathbb{P}(X_d \leq 2 | \mu, \sigma^2)$ où $X_d \sim \text{Bin}(10, p(d, \theta))$ avec $p(d, \theta) =$
 $1 - e^{-\theta d}$ et $\log \theta \sim \mathcal{N}(\mu, \sigma^2)$

Quantiles

Une expérience consiste en l'injection d'une dose d^1 à 10 souris.
Quel serait le nombre q de souris mortes parmi les 10 tel que 10%
(ou 50% ou 90%) des expériences aboutissent à un nombre de
souris mortes inférieur ou égal à q ?

- ▶ Quantités observables donc facile pour les experts mais difficile pour les statisticiens.

¹[choisie par l'expert]

Remarques

- ▶ Quantiles et probabilités sont plus fiables. On peut aussi demander la moyenne, la médiane...
- ▶ Questions probabilistes parfois complexes pour les experts \Rightarrow ils subissent une session d'entraînement qui peut aussi permettre de les évaluer.
- ▶ Les experts fournissent un indice de confiance dans leur réponse (modèle d'erreur sur les données fournies par les experts)

Collecte des dires d'experts

1. **Préparation** : Sélection et motivation des experts
2. **Définition** des quantités à éliciter
3. **Interrogation** des experts : veiller à la cohérence des réponses.
4. **Traitement statistique** : construire la loi a priori à partir des données élicitées
5. **Validation** auprès des experts

Notre travail : combinaison de dires d'experts

Problématique

Comment fournir une seule loi a priori élicitée si on a plusieurs experts ?

- ▶ **Méthode 1** : rassembler les experts et les laisser trouver un consensus. Statisticien = modérateur
- ▶ **Méthode 2** : écrire un modèle mathématique permettant de prendre en compte la dépendance entre experts :
 - ▶ Les experts peuvent avoir des sources d'information similaires (mêmes expérimentations, mêmes instituts) \Rightarrow devraient fournir des valeurs similaires.
 - ▶ Si deux experts sont du même institut, les valeurs qu'ils fournissent ont-elles deux fois plus de chances d'être correctes ?
 - ▶ Sont-elles aussi informatives que celles d'experts indépendants ?
- ▶ Proposition d'une méthode systématique : modèle hiérarchique

Bibliographie

-  Cooke R. M. and Goossens, L. H. J. (2004). Expert judgement elicitation for risk assessments of critical infrastructures. *Journal of Risk Research* **7**, 643-656.
-  Haas C.N., Rose J. B. and Gerba C.,P. (1999). *Quantitative microbial risk assessment*. New York : Wiley.
-  O'Hagan A., Buck C. E., Daneshkhah A., Eiser R., Garthwaite P., Jenkinson D., Oakley J. and Rakow T. (2006). *Uncertain Judgements : Eliciting Experts' Probabilities*. New York : Wiley.
-  Albert I., Donnet S., Guihenneux C., Low Choy S., Mengersen K. and Rousseau J. (2011) *Combining expert opinions in prior elicitation*. In preparation