Nonparametric bayesian modelling of individual co-exposures to various pesticides to determine cocktails

Amélie Crépet and Jessica Tressou

ANSES, French agency for food, environmental and occupational health safety

INRA UR-1204, Mét@risk, Paris, France

The Hong Kong University of Science and Technology, Hong Kong

Applibugs, Paris, 26 November, 2010
For statisticians
What are the cocktails of pesticides simultaneously present in the diet?

- Co-exposure assessment
- Co-exposure clustering

For toxicologists
What are the possible combined effects of multiple residues of pesticides?
Data

- **Contamination levels**: \((\text{DGCCRF, DGAL, SISE-EAUX})\)
  - \(p = 1, \ldots, P = 79\) pesticides
  - \(a = 1, \ldots, 121\) commodities
  - \(H: 0\) if the value < \(\text{LOD}\)

- **Consumption**: the National French survey (INCA 2, 2006)
  - consumed quantities \(c_{ia}\) of \(n = 3,337\) individuals
    - \(n = 1,439\) children (3-17 years) and \(n = 1,898\) adults
  - 7 days detailed
Empirical exposure distribution $x_{imp}$

$$x_{imp} = \log_{10}\left(\sum_{a=1}^{A_p} c_{ia} \times q_{pam} / w_i\right)$$

$M=100$ values computed by

- randomly sampling for each individual $i$ and pesticide $p$
  - consumptions of food $a$ observed on a day: $c_{ia}$
  - pesticide residue levels $q_{pam}$ of the different commodities $A_p$

- dividing by the associated individual body weight $w_i$

- normalized the logarithm of exposures
Exposure Structure

Co-exposures, \( \{x_{imp}, i : \text{individuals}; m : \text{residues levels}; p : \text{pesticides}\} \)

- \( M = 1, 95^{th} \) percentile of exposure to each pesticide \( P \)
- \( M = 100 \), account for contamination variability

\[ M = 1, 95^{th} \text{ percentile of exposure to each pesticide } P \]
\[ M = 100, \text{ account for contamination variability} \]
Co-exposures of 15 children to 79 pesticides

Density

Exposure log mg/kg body weight/day
Children exposure to a single pesticide

![Graph showing exposure log mg/kg body weight/day]

Exposure log mg/kg body weight/day
Infinite Mixture Models
Non-parametric Bayesian Models

- A way of getting very flexible models
- No assumption on the number of mixture components
- No prior parametric assumption
- Infer an adequate model (size/complexity) without doing Bayesian model comparison (AIC, DIC...)
- Derived from finite parametric model with number of parameters going to infinity
Mixture Models

Let \( x = (x_1, \ldots, x_n) \), with \( x_i \) a \( P \) dimensional vector \( x_i = (x_{i1}, \ldots, x_{iP}) \) distributed with a density of probability :

\[
f(x) = \int_\Theta k(x|\theta)G(d\theta)
\]

where

- \( k(\cdot|\theta) \) the known density of the mixture components,
- \( \theta \in \Theta \) is a latent variable,
- \( G \) the unknown mixture distribution, infinite dim. parameter

How to place an appropriate prior on \( G \) ?
Define an a priori on an unknown probability distribution

- \( G|\gamma, H \sim \mathcal{D}(dG|\gamma, H) \) is a Dirichlet process with parameters
  - a real \( \gamma > 0 \)
  - a probability measure \( H \)

- if and only if for any partition \( B_1, \ldots, B_K \) of \( \Omega \),
  \[
  (G(B_1), \ldots, G(B_K)) \sim \text{Dir}(\gamma H(B_1), \ldots, \gamma H(B_K))
  \]

- \( \mathbb{E}[G(B)] = H(B) \)
- \( \mathbb{V}[G(B)] = H(B)(1 - H(B))/(1 + \gamma) \)
Dirichlet Process Mixture Models - DPM

\[ G \sim DP(\gamma, H) \]
\[ \theta_i | G \sim G \]
\[ x_i | \theta_i \sim k(., \theta_i) \]
Stick-Breaking Representation - SB

Sample $G \sim DP(\gamma, H) \iff G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$

- $\phi_k \sim H$
- Infinite mixing proportions, $\sum_{k=1}^{\infty} \pi_k = 1 : \pi_k = \beta_k \prod_{l=1}^{k-1} (1 - \beta_l)$
- Infinite sequence of Beta r.v. : $\beta_k \sim Beta(1, \gamma)$
- Good quality of approximation for reasonable $K < \infty$

$G = \sum_{k=1}^{K} \pi_k \delta_{\phi_k}$ (Ishwaran and James 2001)
Stick-Breaking Representation - SB
Modeling co-exposure to pesticides $M = 1$

$x = (x_1, \ldots, x_n)$, with $x_i$ a $P$ dimensional vector $x_i = (x_{i1}, \ldots, x_{iP})$

$x_i|\theta_i \sim k(.|\theta_i)$ \hfill (1)

$\theta_i|G \sim G$ \hfill (2)

$G \sim DP(\gamma, H)$ \hfill (3)

- kernel $k$: a MultivariateNormal $\mathcal{N}(\mu, \tau)$, $\phi = (\mu, \tau)$
- base probability measure $H$: the conjugate Normal-Wishart
One level of hierarchy

\[ x_{im} = (x_{pim}, p = 1, \ldots, P) \] with \( i = 1, \ldots, n \) and \( m = 1, \ldots, M \)

\[
\begin{align*}
    x_{im} | \theta_{im} & \sim k(., \theta_{im}) \\
    \theta_{im} | G_i & \sim G_i \\
    G_i & \sim DP(\alpha_i, G_0) \\
    G_0 & \sim DP(\gamma, H)
\end{align*}
\]
Algorithm and Random Block

- Gibbs sampler based on the Stick-Breaking priors
- Random Block to reduce the heavy computational burden
  - at each Gibbs cycle select $d < P$ pesticides
  - a subset of random observations $x_i = (x_{il_1}, \ldots, x_{il_d})$ is used instead of $x_i = (x_{i1}, \ldots, x_{iP})$
Example of convergence checking

![Graph 1: Number of clusters vs iterations](image)

![Graph 2: Posterior Likelihood vs iterations](image)
Validation data

Comparison between Stick-breaking (SB) and “Random-Block Stick-Breaking” (RB-SB) algorithms ($N = 30$ atomes, 30000 iterations)

Q-criterion of the first 1,000 iterations performed with the SB and the SB-RB algorithms

Optimal partition obtained with the SB algorithm: mixture of 3 gaussian distributions
Children co-exposure to pesticides

Three main clusters of children with similar co-exposures
Box-plot of the cluster 1: 699 children
Box-plot of the cluster 2 : 238 children
Exposures correlations of cluster 1: 699 children

Cocktail 1
Cocktail 2
Cocktail 3
Cocktail 4
Cocktail 5
Cocktail 6
Cocktail 7
Exposures correlations of cluster 2: 238 children

Amélie Crépet (ANSES)
Extensions

- Prior distribution on $\gamma$: $\text{Gamma}(a_\gamma, b_\gamma)$
- Use Poisson-Dirichlet process instead of DP: slow convergence
  - $a_\gamma$: $\text{Uniform}[0, 1]$
  - $b_\gamma$: $\text{Gamma}(0.01, 0.01)$

- Hierarchical model not really adapted

- Pareto kernel for $k$

- DP for censored data

- 3 of the 7 defined cocktails have toxicological effects
Thanks

- F. Héraud, JC Leblanc and JL Volatier, AFSSA
- ORP (Pesticide Residues Observatory) for financial support
- ANR (National Research Agency) for financial support
- Thank you for your attention