Latent variable models for ecology and evolution

Presentation

Sophie Donnet. **INRAE**

Master 2 MathSV. November 26, 2024

université PARIS-SACLAY

AgroParisTech

Classical statistical models are quite limited in practice

- Linear models : $Y_i = x_i^T \beta + \varepsilon_i, \epsilon \sim \mathcal{N}(0, \sigma_2 I)$

- Generalized linear models :

$$\mathbb{E}[Y_i] = \phi(x_i^T \beta)$$

- $Y_i \sim \mathcal{F}_\theta(\cdot)$

Objectives

- **Purpose of this lecture**: present a series of statistical models involving hidden (also called latent) variables used in ecology and evolution:

$$
\begin{aligned}
Y|Z &\sim p_{\theta^{obs}}(Y|Z) \qquad \text{Observed} \\
Z &\sim p_{\theta^{lat}}(Z) \qquad \text{Hidden}
\end{aligned}
$$

- Estimation of the parameters

EM algorithm and extensions

$$\widehat{\theta} = \arg \max_{\theta} \log \ell(Y; \theta)$$

$$\log \ell(Y; \theta) = \log \int_Z p_{\theta^{obs}}(Y|Z) p_{\theta^{lat}}(Z) dZ$$

- Likelihood Expectation can be difficult to compute because of the integral form
- Resort to Expectation-Maximisation algorithm [Dempster et al., 1977] and extensions (Variational EM, Monte-Carlo EM, etc...)
- Model selection criterion

Examples

- **Mixture models** and EM
  - Gaussian mixture models: sizes of animals
  - Zero inflated Poisson : abundances for biodiversity
- **Hidden Markov models** and EM
  - Trajectories of animals (movement ecology)
  - Gaussian models for traits evolution
- **Stochastic Block models** and VEM
  - Parasitism network , pollination networks
  - Gaussian models for traits evolution
- **Poisson log-normal** and VEM
  - Dependent couting data : abundances of sereral interdependent species
- **Variational auto-encoder** and VEM
  - Deep learning

Dates (séance de 3h00)

14:00 - 17:00. Campus Agro Palaiseau, salle Pal-C2.1.33

- 08/01/2025
- 15/01/2025
- 22/01/2025
- 29/01/2025
- 12/02/2025
- 19/02/2025
- 05/03/2025
- 12/03/2025 (A confirmer)

Evaluation

Examen écrit de 3h le 19/03/2025 (sans documents)

References

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977).
Maximum likelihood from incomplete data via the EM algorithm.
Jr. R. Stat. Soc. B, 39:1–38.