Méthodes de Monte Carlo et inférence bayésienne

Introduction

Pierre Gloaguen

pierre.gloaguen@agroparistech.fr

Déroulé du cours

- 7 séances de 3h;
- Chaque début de séance consacré aux questions et à la présentation du cours;
- TD en autonomie ensuite (avec réponse aux questions);
- Evaluation en contrôle continu (exercices à rendre);

Objectifs de cette 1ère séance

- Introduction et motivation du sujet du cours;
- Principe des méthodes de Monte Carlo;
- Premiers exercices et implémentation sous `R`.

Introduction et motivations

Modèlisation statistique

- Formulation probabiliste d'un problème:
  - *On suppose que les données sont les réalisations de variables aléatoires telles que ...*;
- Quantification de l'incertitude:
  - *d'après le modèle posé, la probabilité que tel événement arrive est de ...*;
- Recours permanent au **calcul de probabilités**;

Exemple immédiat

- Test d'hypothèse: On veut tester une hypothèse H_0 contre une hypothèse H_1
  - *Ex: Comparaison de moyennes de deux échantillons Gaussiens*;
- On construit une statistique de test T;
- Décision binaire, pour un risque de première espèce α, on définit un zone de rejet $\mathcal{R}$, telle que

$$\mathbb{P}_{H_0}(T \in \mathcal{R}) = \alpha$$

Exemple immédiat

- Test d'hypothèse: On veut tester une hypothèse H_0 contre une hypothèse H_1
  - *Ex: Comparaison de moyennes de deux échantillons Gaussiens*;
- On construit une statistique de test T;
- Décision binaire, pour un risque de première espèce α, on définit un zone de rejet $\mathcal{R}$, telle que

$$\mathbb{P}_{H_0}(T \in \mathcal{R}) = \alpha$$

Rappel:

$$\mathbb{P}_{H_0}(T \in \mathcal{R}) = \mathbb{E}_{H_0}[\mathbf{1}_{T \in \mathcal{R}}]$$

- On doit donc être capable d'évaluer une espérance (ici, la probabilité d'évènements) sous H_0 pour construire notre test.

Exemple de modèle statistique: Biomasse d'une population de poisson

Ce qu'on connaît

Captures C

Observations scientifiques Y

Quantité d'intérêt

Biomasse de poisson X

Modèle probabiliste d'observation de la dynamique de population

$$X_{t+1} = \left(X_t + rX_t\left(1 - \frac{X_t}{K}\right) - C_t\right)\exp(\varepsilon_{t+1}), \text{ Biomasse cachée}$$

$$Y_t | X_t = qX_t \exp(\nu_t), \; C_t \text{ Observations}$$

$$\varepsilon_t \overset{i.i.d}{\sim} \mathcal{N}\left(-\sigma^2/2, \sigma^2\right), \quad \nu_t \overset{i.i.d}{\sim} \mathcal{N}\left(-\sigma^2_{\text{obs}}/2, \sigma^2_{\text{obs}}\right).$$

- X_t: Biomasse à l'année t (non observée);
- Y_t: Abundance observée à l'année t;
- C_t: Captures à l'année t;
- K: Capacité d'accueil du milieu (paramètre);
- r: Taux de croissance de la population (paramètre);
- $\sigma, \sigma_{\text{obs}}$: Paramètres de l'aléa;
- q: Détectabilité (paramètre);

Questions classiques d'inférence

Questions classiques d'inférence

- Pour un tel modèle, étant donnée une population initiale X_0, quelle est la moyenne attendue du nombre de poissons au bout de 10 ans si on se fixe une quantité de captures?

Questions classiques d'inférence

- Pour un tel modèle, étant donnée une population initiale X_0, quelle est la moyenne attendue du nombre de poissons au bout de 10 ans si on se fixe une quantité de captures?
  - $\mathbb{E}\left[X_{10}|X_0\right]$?

Questions classiques d'inférence

- Pour un tel modèle, étant donnée une population initiale X_0, quelle est la moyenne attendue du nombre de poissons au bout de 10 ans si on se fixe une quantité de captures?
  - $\mathbb{E}[X_{10}|X_0]$?
- Etant données des observations sur 10 années, et en supposant tous les paramètres connus, que puis je dire sur la quantité de poissons qu'il y avait durant ces 10 ans?

Questions classiques d'inférence

- Pour un tel modèle, étant donnée une population initiale X_0, quelle est la moyenne attendue du nombre de poissons au bout de 10 ans si on se fixe une quantité de captures?
  - $\mathbb{E}\left[X_{10}|X_0\right]$?
- Etant données des observations sur 10 années, et en supposant tous les paramètres connus, que puis je dire sur la quantité de poissons qu'il y avait durant ces 10 ans?
  - $\mathbb{E}\left[X_{0:10}|Y_{0:10}\right]$?

Questions classiques d'inférence

- Pour un tel modèle, étant donnée une population initiale X_0, quelle est la moyenne attendue du nombre de poissons au bout de 10 ans si on se fixe une quantité de captures?
  - $\mathbb{E}[X_{10}|X_0]$?
- Etant données des observations sur 10 années, et en supposant tous les paramètres connus, que puis je dire sur la quantité de poissons qu'il y avait durant ces 10 ans?
  - $\mathbb{E}[X_{0:10}|Y_{0:10}]$?
- Etant données des observations, que puis je dire sur la valeur des paramètres de dynamique de population?

Questions classiques d'inférence

- Pour un tel modèle, étant donnée une population initiale X_0, quelle est la moyenne attendue du nombre de poissons au bout de 10 ans si on se fixe une quantité de captures?
  - $\mathbb{E}[X_{10}|X_0]$?
- Etant données des observations sur 10 années, et en supposant tous les paramètres connus, que puis je dire sur la quantité de poissons qu'il y avait durant ces 10 ans?
  - $\mathbb{E}[X_{0:10}|Y_{0:10}]$?
- Etant données des observations, que puis je dire sur la valeur des paramètres de dynamique de population?
  - Inférence des paramètres:
  - Méthode des moments (nécessite un calcul d'espérance);

Questions classiques d'inférence

- Pour un tel modèle, étant donnée une population initiale X_0, quelle est la moyenne attendue du nombre de poissons au bout de 10 ans si on se fixe une quantité de captures?
  - $\mathbb{E}[X_{10}|X_0]$?
- Etant données des observations sur 10 années, et en supposant tous les paramètres connus, que puis je dire sur la quantité de poissons qu'il y avait durant ces 10 ans?
  - $\mathbb{E}[X_{0:10}|Y_{0:10}]$?
- Etant données des observations, que puis je dire sur la valeur des paramètres de dynamique de population?
  - Inférence des paramètres:
  - Méthode des moments (nécessite un calcul d'espérance);
  - Méthode du maximum de vraisemblance (nécessite ici un calcul d'espérance);

Questions classiques d'inférence

- Pour un tel modèle, étant donnée une population initiale X_0, quelle est la moyenne attendue du nombre de poissons au bout de 10 ans si on se fixe une quantité de captures?
  - $\mathbb{E}\left[X_{10}|X_0\right]$?
- Etant données des observations sur 10 années, et en supposant tous les paramètres connus, que puis je dire sur la quantité de poissons qu'il y avait durant ces 10 ans?
  - $\mathbb{E}\left[X_{0:10}|Y_{0:10}\right]$?
- Etant données des observations, que puis je dire sur la valeur des paramètres de dynamique de population?
  - Inférence des paramètres:
  - Méthode des moments (nécessite un calcul d'espérance);
  - Méthode du maximum de vraisemblance (nécessite ici un calcul d'espérance);
  - Estimateur Bayésien: Nécessite un calcul d'espérance.

Questions classiques d'inférence

- Pour un tel modèle, étant donnée une population initiale X_0, quelle est la moyenne attendue du nombre de poissons au bout de 10 ans si on se fixe une quantité de captures?
  - $\mathbb{E}\left[X_{10}|X_0\right]$?
- Etant données des observations sur 10 années, et en supposant tous les paramètres connus, que puis je dire sur la quantité de poissons qu'il y avait durant ces 10 ans?
  - $\mathbb{E}\left[X_{0:10}|Y_{0:10}\right]$?
- Etant données des observations, que puis je dire sur la valeur des paramètres de dynamique de population?
  - Inférence des paramètres:
  - Méthode des moments (nécessite un calcul d'espérance);
  - Méthode du maximum de vraisemblance (nécessite ici un calcul d'espérance);
  - Estimateur Bayésien: Nécessite un calcul d'espérance.

Ces espérances n'ont, en général, pas d'expressions directes!

Méthodes de Monte Carlo

Méthodes de Monte Carlo

- **But:** Approcher des espérances (intégrales) en utilisant des simulations probabilistes;

Méthodes de Monte Carlo

- **But:** Approcher des espérances (intégrales) en utilisant des simulations probabilistes;
- **Idée:** La loi des grands nombres! La moyenne empirique d'une variable aléatoire va tendre, si on répète l'expérience, vers la moyenne théorique.

Exemple

On dispose d'un dé à 6 faces et seulement de ce dé. Comment peut on essayer de savoir s'il est biaisé?

- Si on lance le dé suffisament de fois, on obtient une information;

Exemple

On dispose d'un dé à 6 faces et seulement de ce dé. Comment peut on essayer de savoir s'il est biaisé?

- Si on lance le dé suffisament de fois, on obtient une information;
- Combien de fois faut il lancer le dé pour avoir une idée précise?

Exemple

On dispose d'un dé à 6 faces et seulement de ce dé. Comment peut on essayer de savoir s'il est biaisé?

- Si on lance le dé suffisament de fois, on obtient une information;
- Combien de fois faut il lancer le dé pour avoir une idée précise?
- À quel point peut être confiant en notre réponse?

Exemple

On dispose d'un dé à 6 faces et seulement de ce dé. Comment peut on essayer de savoir s'il est biaisé?

- Si on lance le dé suffisament de fois, on obtient une information;
- Combien de fois faut il lancer le dé pour avoir une idée précise?
- À quel point peut être confiant en notre réponse?
- Encore faut il savoir lancer le dé!

Programme du cours

- Présentation formelle des méthodes de Monte Carlo pour le calcul d'intégrales;
- Application directe en statistique classique (évaluation d'une probabilité, aide à la décision);

Programme du cours

- Présentation formelle des méthodes de Monte Carlo pour le calcul d'intégrales;
- Application directe en statistique classique (évaluation d'une probabilité, aide à la décision);
- Comment peut on simuler des variables aléatoires génériques (avec un ordinateur)?

Programme du cours

- Présentation formelle des méthodes de Monte Carlo pour le calcul d'intégrales;
- Application directe en statistique classique (évaluation d'une probabilité, aide à la décision);
- Comment peut on simuler des variables aléatoires génériques (avec un ordinateur)?
- Une méthode d'inférence dépendante de la simulation: l'inférence bayésienne;

Programme du cours

- Présentation formelle des méthodes de Monte Carlo pour le calcul d'intégrales;
- Application directe en statistique classique (évaluation d'une probabilité, aide à la décision);
- Comment peut on simuler des variables aléatoires génériques (avec un ordinateur)?
- Une méthode d'inférence dépendante de la simulation: l'inférence bayésienne;
- Une extension nécessaire, les Méthodes de Monte Carlo par chaîne de Markov (MCMC).

Prérequis

- Résulats statistiques asymptotiques:
  - Loi des grands nombres, théorème central limite, lemme de Slutsky, Delta méthode.
- Chaînes de Markov:
  - Loi de transition, irréductibilité, périodicité, mesure invariante ...
- Logiciel `R`
  - Logiciel `R` installé ainsi que l'IDE `Rstudio`.
  - Connaissance minimale du langage (boucles, fonctions, graphiques de base...).