

Introduction à la Modélisation Statistique Bayésienne

Organizer(s) [Thierry Phénix, PhD](#)

1 ECTS

Summary

L'approche Bayésienne est caractérisée par le fait de traiter chaque entité (e.g., variables observées, paramètres du modèle, données manquantes) comme des variables aléatoires caractérisées par des distributions. Dans une analyse Bayésienne, chaque entité inconnue se voit assignée une distribution a priori qui représente un état de connaissance avant d'observer les données. Une fois que les données ont été observées, le théorème de Bayes est utilisé pour mettre à jour la distribution a priori, permettant d'obtenir une distribution a posteriori. Les distributions a posteriori sont le but d'une telle analyse et peuvent être résumées par des valeurs ponctuelles ou des intervalles, et interprétées directement par un raisonnement probabiliste.

Cette approche Bayésienne se différencie –à la fois philosophiquement et en pratique– de l'approche traditionnelle fréquentiste, qui constitue actuellement la majorité des formations proposées. Un des avantages de l'approche Bayésienne est qu'elle permet à l'analyste de résoudre des problèmes difficiles voire impossibles pour l'approche fréquentiste traditionnelle.

Au fil des exemples proposés, nous réaliserons que même dans des situations de modélisation simples, l'approche Bayésienne permet un raisonnement probabiliste qui est plus naturel et plus flexible que la machinerie inférentielle de l'approche fréquentiste. La modélisation statistique Bayésienne et l'inférence Bayésienne représentent une alternative attirante aux approches fréquentiste en ce qu'elles offrent un cadre cohérent à la modélisation statistique, permettant de construire des modèles complexes, interpréter les résultats, faire des inférences et représenter l'incertitude.

Objectives

L'objectif de cette formation est de vous faire découvrir l'approche Bayésienne. Les conceptions et outils qui seront introduits tout au long de la formation seront illustrés par des cas concrets d'analyse de données.

Pre-requirements

- Être familier avec l'approche fréquentiste (e.g., test d'hypothèse, intervalles de confiance, régression linéaire, ANOVA).
- Connaissances élémentaires en manipulation de données avec Python et les paquets Pandas, Numpy et Matplotlib.

Course dates

Always on Thursdays from 16 – 18h on the following days :
October 14, 21 and 28
November 4, 11, 18 and 25
December 2, 9 and 16

Location If the sanitary situation allows, the course will take place in person in Geneva at the CMU, Room C07.17.32.

Course sessions Le cours comprend dix séances de deux heures durant lesquelles seront dispensées des connaissances théoriques et travaux pratiques en langage Python, dans l'environnement Conda / Jupyterlab.

Session 1: (2h)

- Introduction aux concepts probabilistes
- Modèles et théories, théorème de Bayes
- Configuration de l'environnement de travail

Session 2: (2h)

- Présentation de la démarche Bayésienne
- Modèle beta-binomial

Session 3: (2h)

- Présentation du modèle Gaussien
- Régression linéaire sans prédicteur
- Régression linéaire avec un prédicteur continu
- Régression polynomiale
- Régression robuste

Session 4: (2h)

- Régression linéaire avec deux prédicteurs continus ou plus
- Colinéarité, variables confondues
- Méthodes des résidus
- Relations masquées

Session 5: (2h)

- Régression linéaire et relation de causalité
- DAG (Direct Acyclic Graph)
- Schémas de base
- Cas pratiques

Session 6: (2h)

- Régression linéaire avec un prédicteur catégoriel
- Régression linéaire avec plusieurs prédicteurs catégoriels
- Régression linéaire avec un prédicteur ordinal
- taille d'effet, contrastes

Session 7: (2h)

- Modèle d'interactions
- Entre deux prédicteurs continus
- Entre deux prédicteurs catégoriels

Session 8: (2h)

- Markov Chain Monte Carlo
- Algorithmes d'échantillonnage (Metropolis, Hamilton...)
- Outils de diagnostique

Session 9: (2h)

- Modèle linéaire généralisé (GLM)
- Régression logistique, fonction de lien
- Régression de poisson

Session 10: (2h)

- Data hackaton
- Questions, recommendations

Course materials Course materials and papers will be stored under "[Github](#)". The repository will be available around one month before course start.

Evaluation Active participation in at least 80% of the course sessions is required; the data hackathons are mandatory course sessions aiming to assess the active involvement of students and their application of learned concepts to data.

Registration The course is limited to 14 participants. Register before September 20, 2021, by writing a mail to Ulrike.toepel@unil.ch.