M2-Mathématiques et Applications 2022-2023, Paris-Saclay
Remise à niveau Statistiques
christine.keribin@universite-paris-saclay.fr

Introduction à l'inférence bayésienne

1 Inférence bayésienne pour le modèle binomial

1. On rappelle que la loi Beta $\mathcal{B}e(\alpha, \beta)$ est une loi conjuguée pour la vraisemblance de Bernoulli $\mathcal{B}(1, \theta)$, d'espérance $\alpha/(\alpha+\beta)$ et de mode $(\alpha-1)/(\alpha+\beta-2)$. Quels sont les paramètres de la loi posteriori $\pi(\theta|x)$ où x désigne un n-échantillon?
2. Quelle est la loi a priori de Jeffreys et la loi a posteriori correspondante?
3. Représenter sur un même graphique les courbes des lois $\mathcal{B}e(3, 15)$, $\mathcal{B}e(15, 3)$, $\mathcal{B}e(1, 1)$, et de la loi de Jeffrey avec des attributs graphiques différents. Ajoutez une légende. Commentez les courbes obtenues
4. Simuler trois échantillons de taille $n = 20, 100, 1000$ suivant une loi de Bernoulli de paramètre $\theta = 0.1$.
5. Ecrire une fonction `plotposterior(tab,title)` qui trace pour un échantillon `tab` les quatre lois a posteriori sur un même graphique, affiche une légende et le titre `title`. Dans une fenêtre partagée en quatre, tracer les lois a priori et les lois a posteriori pour les trois échantillons, mettre un titre général.

   Discuter l'influence de la loi a priori et du nombre d'observations.
6. Reprendre les questions avec $\theta = 0.5$ et commenter.

2 Comparaison de deux échantillons

Soient $(X_1, ..., X_n)$ et $(Y_1, ..., Y_n)$ deux n-échantillons de lois respectives $\mathcal{N}(\theta_x, \sigma^2)$ et $\mathcal{N}(\theta_y, \sigma^2)$. On cherche à choisir entre le modèle $M1 : \theta_x = \theta_y$ et le modèle $M2 : \theta_x \neq \theta_y$.

On reparamètre $\theta_x = \theta - \delta$ et $\theta_y = \theta + \delta$.

1. Ecrire la définition du facteur de Bayes B_{21}. Expliquer pourquoi il est possible d'utiliser pour les deux modèles M1 et M2 la même loi a priori impropre $\pi(\sigma^2) = 1/\sigma^2$ sur σ^2, la même loi apriori $\pi(\theta) = 1$ impropre sur μ. La loi a priori sur δ dans M2 sera prise gaussienne $\pi(\delta) \sim \mathcal{N}(0, \tau^2)$.
2. Montrer que
$$B_{21} = \frac{\int [(2\delta + \bar{x} - \bar{y})^2 + 2s^2]^{-n+1/2} \, e^{-\delta^2/(2\tau^2)} d\delta / (\tau\sqrt{2\pi})}{[(\bar{x} - \bar{y})^2 + 2s^2]^{-n+1/2}}$$
avec $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$ et $\bar{x}$ et $\bar{y}$ sont les moyennes des deux échantillons.
3. Programmer le calcul du facteur de Bayes B_{21} en intégrant le numérateur par une méthode de Monte Carlo utilisant un $m = 1000$- échantillon d'une loi gaussienne.

   Application: Soit $\tau = 0.75$. Calculer B_{21} dans le cas de la comparaison des échantillons 5 et 6 du jeu `normaldata` du package `bayess`. Que conclure?
4. Les résultats changent-ils si $\tau = 0.1$?
5. Illustrer la variabilité de l'estimation de B_{21} en simulant 500 réplications de l'expérience précédente. Représenter l'histogramme des facteurs de Bayes calculés et superposer la courbe d'une densité gaussienne. Estimer la variance.