Master professionnel II : Ingénierie mathématique : Option Statistique

Statistique Bayésienne.

Anne Philippe
Université de Nantes
Laboratoire de Mathématiques Jean Leray

Fiche 6. **Prévision bayésienne**

Exercice 1. *(Suite de l'exercice 1 de la fiche 3)*

On veut prévoir S^* le nombre d'étudiants qui dorment plus de 8 heures dans un groupe de taille N^*.

Données : on a observé S le nombre d'étudiants qui dorment plus de 8 heures dans un échantillon de taille $N = 28$. La valeur observée est $s = 11$.

On suppose que conditionnellement à p (la proportion des étudiants qui dorment plus de 8 heures par nuit), les deux échantillons sont indépendants

On suppose que la loi a priori sur le paramètre p est la loi beta de paramètres $a = 3.4$ et $b = 7.4$.

1) Quelle est la loi a posteriori de p ?
2) Quelle est la loi de S^* conditionnellement à (S, p)
3) En déduire la loi prédictive de S^* (autrement dit la loi conditionnelle de S^* sachant S).
4) Donner une prévision ponctuelle de S^* (optimale au sens L^2)
5) Donner pour S^* un intervalle de prévision de niveau de confiance 95%.
6) Comparaison avec l'approche non bayésienne :
   - a - Quelle est la prévision ponctuelle optimale au sens L^2 de S^* dans un contexte non bayesien ?
   - b - Quel prédicteur proposez-vous en pratique ?
   - c - Comparer avec le prédicteur obtenu à la question 4.
7) Représenter graphiquement
   -a- le prédicteur bayesien de S^* en fonction de N^*
   -b- ajouter les bornes du plus court l'intervalle de prévision de niveau de confiance 80%, 90%, 95% et 99%

Exercice 2.

Soit $X_1....X_n$ n variables aléatoires définies par

$$\begin{aligned} & X_0 = 0 \\ \text{pour tout } i = 1, ..., n, \qquad & X_i = aX_{i-1} + \varepsilon_i \end{aligned} \tag{1}$$

où $\varepsilon_1, ..., \varepsilon_n$ sont des variables aléatoires iid suivant la loi gaussienne $\mathcal{N}(0, 1)$ et a est un paramètre réel inconnu.

- Calcul de la vraisemblance :

1) Soit U et V deux variables aléatoires indépendantes. On suppose que la loi de U (respectivement de V) admet une densité f_U (respectivement f_V). Montrer que la loi conditionnelle de $W = g(U) + V$ sachant U admet pour densité
$$f_{W|U}(w|u) = f_V(w - g(u)).$$
2) Justifier que les variables aléatoires X_i et ε_{i+1} sont indépendantes.
3) En déduire que la loi de X_i sachant X_{i-1} est la loi normale de moyenne aX_{i-1} et de variance 1.
4) En déduire la densité $f_a^{(n)}$ de la loi des observations $X_1, ..., X_n$.

- Modèle bayésien. On suppose que la loi a priori sur a est la loi gaussienne $\mathcal{N}(0, 1)$.

5) Calculer la loi a posteriori de a.
6) Quelle est la loi conditionnelle de X_{n+1} sachant $(X_1, ..., X_n, a)$?
7) Récupérer le fichier de données *ARmodel.txt*. Il contient $n = 55$ observations.

Résultat du cours .

Mélange continu de lois Soit X une variable aléatoire dont la loi admet une densité f par rapport à la mesure de Lebesgue. On suppose que
$$f(x) = \int_{\mathbb{R}^p} h(x|y)g(y)\,\mathrm{d}y$$
où g et $h(\cdot|y)$, pour tout $y \in \mathbb{R}^p$, sont des densités de probabilité.
Pour simuler un nombre aléatoire x suivant la loi de densité f :

1. **on simule** y **suivant la loi de densité** g
2. **on simule** x **suivant la loi de densité** $h(\cdot|y)$
3. **on retourne** x

8) Programmer une fonction qui retourne un échantillon suivant la loi prédictive de X_{n+1} sachant $X_1, ...X_n$.
9) Simuler un échantillon suivant la loi prédictive pour $n = 50$
10) A partir de l'échantillon simulé, donner
    - -a- une approximation de la densité de la loi prédictive,
    - -b- le plus court intervalle de prévision de niveau de confiance 95 %,
    - -c- un prédicteur ponctuel.
11) Comparer avec le vraie valeurs de X_{51}.
12) Reprendre les questions précédentes pour $n = 51, ..., 54$