Exercice 10. Estimation d'un proportion

On veut estimer la proportion p d'étudiants qui dorment plus de 8 heures par nuit. Les observations sur un échantillon de 27 étudiants sont :

s= 11 étudiants dorment plus de 8 heures
f=16 étudiants dorment moins de 8 heures.

On note S la variable aléatoire qui représente le nombre d'étudiants qui dorment plus de 8 heures dans un échantillon de taille $n = 27$.

On envisage deux lois a priori sur le paramètre $p \in]0, 1[$:

Modèle A- la loi discrète définie par

i	b_i	$P(p = b_i)$
1	0.05	0.03
2	0.15	0.18
3	0.25	0.28
4	0.35	0.25
5	0.45	0.16
6	0.55	0.07
7	0.65	0.03

Modèle B- la loi Beta de paramètres $a = 3.4$ et $b = 7.4$.

Modèle C- la loi non informative de Jeffreys

Comparaison des trois modèles bayésiens

I-1) Représenter graphiquement les trois lois a priori.

I-2) Calculer la moyenne et la variance des trois lois a priori

I-3) Commenter les résultats obtenus. Les lois apportent-elles la même information a priori ?

I-4) Donner l'expression des trois lois marginales $m(x) = P(X = x)$

I-5) En déduire la valeur du facteur de Bayes pour comparer
- les modèles A et B.
- les modèles B et C.
- les modèles A et C.

Commenter les résultats obtenus.

lois a posteriori

II-1) Calculer les trois lois a posteriori.

II-2) Représenter sur un même graphique les lois a priori et les lois a posteriori

II-3) Calculer la moyenne et la variance des lois a posteriori.

II-4) Commenter les résultats obtenus.

Régions de confiance bayésiennes pour le paramètre p

Modèle a priori A

III-1) Construire une région HPD de niveau 95% (ou au moins 95%) ? S'il est différent de 95% quel est le niveau exact de la région HPD.

Modèle a priori B

III-2) Calculer le plus court intervalle de crédibilité au niveau 95% en utilisant la fonction `qbeta`.

III-3) Quelle est la région HPD de niveau 95% ?

Modèle a priori C

III-4) Reprendre les questions précédentes pour le modèle C

Conclusion

III-5) Comparer et commenter les résultats obtenus

Prévision

On veut prévoir S^* le nombre d'étudiants qui dorment plus de 8 heures dans un groupe de taille 20.

Modèle a priori B

VI-1) Trouver les fonctions $a(S), b(S)$ pour que la valeurs s^* simulée à partir de l'algorithme ci-dessous soit une réalisation de la loi prédictive de S^*.
Quelle expression théorique de la loi prédictive permet de justifier cet algorithme ?

1. `Simuler` p `suivant la loi beta de paramètre` $(a(S), b(S))$
2. `Simuler` s^* `suivant la loi binomiale de paramètre` $(20, p)$

VI-2) Simuler un échantillon de longueur M suivant la loi predictive de S^*

VI-3) A partir de l'échantillons simulé,

(a) calculer et représenter graphiquement une approximation de la loi prédictive,

(b) calculer une approximation de la région HPD de la loi prédictive au niveau 95 % (ou au moins 95%),

(c) calculer une approximation du meilleur prédicteur ponctuel au sens de l'erreur L^2

VI-4) Reprendre les questions précédentes pour le modèle a priori C

VI-5) Adapter l'algorithme précédent pour simuler un échantillon suivant la loi prédictive associée au modèle a priori A, puis reprendre les mêmes questions.

Comparaison.

VI-6) Représenter sur un même graphique les lois prédictives associées aux trois modèles.

VI-7) Comparer les régions HPD des lois prédictives de niveau 95 % (ou au moins 95%) et les prédicteurs ponctuels.

VI-8) Commenter les résultats