ECN option mathématiques Parcours S2D

Statistique Bayésienne.

Anne Philippe
Université de Nantes
Laboratoire de Mathématiques Jean Leray

Fiche 1. Information a priori

Exercice 1. Modèle exponentiel

L'objectif de cet exercice est de visualiser l'influence de la loi a priori en fonction de la qualité de l'information a priori et du nombre d'observations.

Résultat du cours .

On suppose que les observations sont iid suivant une loi exponentielle de paramètre $\theta > 0$

Si la loi a priori du paramètre θ est la loi gamma $\Gamma(a,b)$ alors la loi a posteriori est aussi une loi gamma de paramètres $n+a, b+\sum_{i=1}^n X_i$.

Rappel sur les lois .

La loi Gamma de paramètres $(a,b) \in {\mathbb{R}_+^*}^2$ (notée $\Gamma(a,b)$) est la loi qui admet pour densité (par rapport à la mesure de Lebesgue)

$$f(x) = b^a \frac{1}{\Gamma(a)} e^{-bx} x^{a-1} \mathbb{I}_{\mathbb{R}^+}(x).$$

Son espérance est égale à a/b et sa variance à a/b^2

1) Récupérer le fichier de données

`http://www.math.sciences.univ-nantes.fr/~philippe/data/duree-de-vie.txt`

qui contient des durées de fonctionnement de 1000 ampoules.

Commande R .

Avec la fonction `scan`, on peut importer des données à partir d'un fichier local ou d'un site web. Par exemple

```
data = scan("http://www.math.sciences.univ-nantes.fr/~philippe/data/duree-de-vie.txt")
```

On modélise ces données par des variables aléatoires $X_1, ..., X_n$ iid suivant la loi exponentielle de paramètre $\theta \in \mathbb{R}_*^+$. On veut estimer le paramètre θ à l'aide d'un modèle bayésien. On choisit une loi gamma comme loi a priori sur θ

2) L'information fournit a priori est θ devrait être proche de 1/2. On note τ la variance de la loi a priori. Proposer des paramètres pour la loi a priori.
3) On choisit $\tau = 1/2$. Superposer la densité de la loi a priori et les densités a posteriori pour différentes tailles d'échantillon n (par exemple `n in c(2,5,10,100,500,1000)`)
4) Reprendre la question précédente pour différentes valeurs de $\tau = 1/100, 1/10, 10, 100$.
5) Construire trois estimateurs de θ à partir de la loi a posteriori.
6) Pour les différentes valeurs de τ :
   6 - a) Calculer et représenter ces estimateurs en fonction de n le nombre d'observations.
   6 - b) Calculer et ajouter au graphique précédent l'estimateur du maximum de vraisemblance de θ
7) Interpréter les résultats obtenus.
8) L'information a priori sur θ est maintenant la suivante "θ est autour de 3". Reprendre les questions précédentes
9) Conclure

Exercice 2.

Rappel sur les lois .

La loi beta de paramètres $(a, b) \in {\mathbb{R}_+^*}^2$ (notée $\beta(a, b)$) est la loi qui admet pour densité (par rapport à la mesure de Lebesgue sur [0,1])

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}\, \mathbb{I}_{[0,1]}(x).$$

Son espérance est égale à $\frac{a}{a+b}$ et sa variance $\frac{ab}{(a+b)^2(a+b+1)}$

Trois personnes veulent estimer la proportion p d'étudiants qui ne résident pas sur le campus.
— A suppose que la loi a priori sur p est la loi discrète définie par

p_i	.1	.2	.3	.4	.5
$\pi(p_i)$	.5	.2	.2	.05	.05

— B suppose que la loi a priori sur p est la loi Beta de paramètres $(3, 12)$
— C suppose que la loi a priori sur p est Beta de paramètres $(1, 4)$

1) Calculer la moyenne et l'écart type de ces trois lois a priori.
2) Ont-ils la même information a priori sur la localisation du paramètre p ? Accordent ils la même confiance à l'information a priori obtenue ?
3) Soit y le nombre d'étudiants qui habitent hors du campus dans un échantillon de taille 12. Donner l'expression de la loi prédictive a priori $m(y)$ pour les trois lois a priori.
4) Superposer les courbes représentatives des densités des lois prédictives a priori.