

## TP 3

### Avertissement

L'objectif de ce TP est de programmer manuellement une minimisation de cout par une méthode "brut force". L'idée est simplement ici de programmer la fonction qui calcule le risque a posteriori associé à une décision, puis de la minimiser en parcourant plusieurs décisions possibles. L'objectif pédagogique est de s'assurer de la bonne compréhension de tous les niveaux (perte prédictive, perte, risque a posteriori...) d'intégration.

Vous organisez une conférence, et vous devez effectuer les réservations. 100 personnes se sont inscrites (le maximum que vous acceptez<sup>1</sup>), mais d'expérience des 2 années précédentes, vous savez qu'il y a toujours des désistements de dernière minute. En effet, les années passées, sur 100 inscrit-e-s, seulement 79 et 85 étaient effectivement venues.

La politique de réservation de l'hôtel est la suivante :

- En cas de sur-réservation, vous devez payer 300 euros de dédommagement par chambre sur-réservée.
- En cas de sous-réservation, cela vous coûte 100 euros de plus par chambre non réservée à l'avance.

En bon mathématicien-ne, vous décidez de formaliser le problème afin de trouver le nombre de chambres à réserver.

1. Si on note  $X_1$  et  $X_2$  les variables aléatoires correspondant aux nombre de personnes effectivement venues les 2 années précédentes, et  $X_3$  celle correspondant au nombre qui vont venir cette année, par quelle loi peut-on modéliser ces variables ? (on notera  $N, p$  les paramètres).
2. En bon Bayésien-ne, vous allez considérer que  $X_1, X_2, X_3$  sont i.i.d. conditionnellement à  $p$ , et mettre un prior sur  $p$ . Quelle famille de prior peut-on choisir pour que les calculs soient simples ? Quels hyperparamètres  $a_0, b_0$  correspondent à un prior uniforme dans ce cadre ? À partir de maintenant, on travaille avec ce prior.
3. On note  $\mathcal{D} = (X_1, X_2)$  les données observées, et  $d(\mathcal{D})$  la décision correspondant au nombre de chambres réservées que vous devez prendre. À partir de maintenant, on travaille sur les machines.  
Écrire (i.e. programmer) la fonction Lp donnant un coût correspondant à la politique de l'hôtel, en fonction de  $X_3$  et de la décision  $d$ .
4. Écrire (i.e. programmer) la fonction L donnant, en fonction d'un paramètre  $p$ , et de la décision  $d$ , le coût prédictif.
5. Écrire (i.e. programmer) la fonction rho donnant, en fonction des hyperparamètres  $a, b$  du posterior, et de la décision  $d$ , le coût à posteriori.
6. Écrire (i.e. programmer) la fonction minimpost donnant, en fonction des hyperparamètres  $a, b$  du posterior, la décision optimale pour ce problème de décision.
7. En utilisant les données de l'énoncé, calculer les hyperparamètres du posterior, et donner (numériquement) la décision optimale. Comment feriez-vous si vous disposiez uniquement d'un échantillon tiré sous la loi a posteriori ?
8. Faire l'analyse théorique du problème : obtenir la décision optimale comme un quantile de la loi prédictive, en résolvant (théoriquement) le problème de minimisation.
9. Vérifier la valeur numérique de  $d(\mathcal{D})$  ainsi obtenue, en utilisant la fonction qbetabinom du package rmutil, qui donne les quantiles de la loi betabinomiale.
10. L'hôtel vous propose une nouvelle politique, plutôt que payer 300 euros par chambre en sur-réservation, il vous proposer de payer un forfait de 1000 euros en cas de sur-réservation (indépendamment du nombre de chambres). Que devient la décision optimale avec cette politique ? Quelle politique préféreriez-vous ?
11. Comment pourriez vous accélérer votre code ?

---

<sup>1</sup> votre conférence est très demandée