1 Eléments de probabilités

Dans un sac, nous disposons de trois pièces : l'une d'elles est équilibrée, et les deux autres sont pipées (informations de distribution de probabilités : 1/3 de `pile`, et 2/3 de `face`).

1. On tire une pièce dans le sac et on la jette. Quelle est la probabilité d'observer `pile` ?
2. On tire une pièce dans le sac et on la jette deux fois : on observe alors `pile-pile`. On la jette une troisième fois. Quelle est la probabilité d'observer `pile` ?
3. On tire une pièce dans le sac et on la jette deux fois. Que peut-on décider sur la nature de la pièce (équilibrée ou pipée), en considérant chacune des règles de décision, et dans le cas de chacune des observations suivantes : `pile-pile`, `face-pile`, `pile-face`, et `face-face` ? Indiquer, dans chaque cas, le taux d'erreur.

2 Détection de fraude

L'apprentissage automatique peut être utilisé pour détecter les fraudes : l'exercice ci-après en est une illustration très simple.
On dispose d'un dé à 6 faces, parfaitement équilibré. On confie ce dé à des individus en leur demandant de procéder à un certain nombre de lancers et de faire part de leurs résultats. La population est composée de personnes honnêtes (H) qui font exactement ce qu'on leur demande, mais aussi d'un certain nombre de tricheurs (T) qui, chaque fois qu'on leur demande de lancer une fois le dé, le lancent en réalité deux fois et annoncent le plus grand des nombres obtenus. Ainsi, si l'on demande à un tricheur de lancer 5 fois le dé, il pourra obtenir la suite de résultats (2,2), (5,2), (4,1), (5,4), (6,3), et annoncer (2,5, 4, 5, 6).

1. Calculer $p(i|H)$ et $p(i|T)$ pour $i = 1$ à 6.
2. Calculer $p(25456|H)$ et $p(25456|T)$.
3. On suppose que la population contient 10% de tricheurs. Que doit-on décider sur l'honnêteté d'un individu qui annonce 25456 si l'ont suit respectivement :
   (a) la règle majoritaire,
   (b) la règle du maximum de vraisemblance,
   (c) la règle de décision de Bayes ?
4. TD2 : On souhaite maintenant fonder la décision sur le moyenne des résultats obtenus. Quelle est l'espérance du résultat annoncé par :
   (a) une personne honnête,
   (b) un tricheur ?
5. TD2 : Combien faut-il faire faire de tirages à un individu pour être sûr, avec une confiance de 95%, de différencier une personne honnête d'un tricheur ?

3 Patients malades et patients sains

La population consiste en un ensemble de patients. Ces patients doivent être répartis en deux classes : la classe M des malades, et la classe S des sains.
Chaque individu est décrit à l'aide de deux attributs logiques T (indiquant une tension artérielle anormale) et C (indiquant un taux de cholesterol anormal).
On suppose que la population est un espace probabilisé et on note P la loi de probabilité. Nous allons utiliser l'approche bayésienne classique pour classer les patients au vu de leurs descriptions.
Les probabilités suivantes sont connues :

classe k	S (sains)	M (malade)
$P(k)$	0.7	0.3
$P(T/k)$	0.25	0.7
$P(C/k)$	0.4	0.7

Pour pouvoir calculer les probabilités nécessaires à l'application de la règle de Bayes, nous allons faire l'hypothèse supplémentaire suivante : les deux attributs sont indépendants. Cette hypothèse permet, par exemple, d'écrire :

$$P(C \cap \bar{T}/S) = P(C/S) \times P(\bar{T}/S)$$

1. Déterminer la procédure de classification C_{Bayes} en utilisant la règle de décision de Bayes.
2. Soient les procédures de classification C_1 et C_2 : $C_1(x) = S$ et $C_2(x) =$ si T anormale alors M sinon S. Calculer les erreurs au sens de la probabilité d'erreur pour C_1, C_2 et C_{Bayes}.
3. Plutôt que de chercher à minimiser l'erreur au sens de la probabilité d'erreur, on peut introduire des coûts pour les mauvaises classifications. On définit un coût pour tout couple de classes (k, i) noté $cout(k, i)$. On définit alors le coût moyen de l'affectation à la classe k d'une description d de D par

$$coutMoyen(k/d) = \sum_{i=1,\cdot,c} cout(k, i) \times P(i/d)$$

La règle de décision du coût minimum est : *Choisir* C_{coutMin} *qui à toute description d associe la classe k qui minimise* coutMoyen(k/d). On définit, sur notre exemple, les coûts suivants :

$$cout(S, S) = cout(M, M) = 0, cout(S, M) = 2, cout(M, S) = 1$$

Déterminer $C_{coutMin}$.