EXERCICE 4 Une entreprise de téléphonie a constaté que ses clients se répartissaient selon deux groupes : ceux qui téléphonent en moyenne aussi souvent vers un fixe que vers un portable et ceux qui téléphonent en moyenne deux fois plus souvent vers un portable que vers un fixe. Les clients du premier groupe sont environ deux fois plus nombreux que les seconds. L'entreprise souhaite adresser une offre promotionnelle aux clients du second groupe, pour éviter qu'ils ne changent de compagnie et ce, le plus tôt possible. Comme information sur sa clientèle, l'entreprise dispose des types d'appels effectués. On notera P (resp. F) un appel vers un portable (resp. vers un fixe).

Question 1 Peut-on décider sur la base des k premiers appels si un client appartient au premier groupe ou au second ? Supposons que $k = 4$ et qu'on observe la suite P, P, F, P. Quelle décision conduit à prendre la règle majoritaire ? la règle du maximum de vraisemblance ? la règle de Bayes ?

Correction : Il est évident que pour un k assez grand, une décision pourra être prise avec un risque probablement faible.

Modélisation du problème Les clients de l'entreprise peuvent être divisés en deux groupes : Soit G_1 le groupe des clients téléphonant aussi souvent sur les fixes que sur les portables et G_2 l'autre groupe. On note P (respectivement F) un appel vers une téléphone portable (resp. fixe). L'énoncé s'écrit alors :

$$\begin{array}{lll} P(G_1) = 2/3 & P(P|G_1) = 1/2 & P(F|G_1) = 1/2 \\ P(G_2) = 1/3 & P(P|G_2) = 2/3 & P(F|G_2) = 1/3 \end{array}$$

Règle majoritaire. La règle majoritaire classe tout exemple dans le groupe G_1.

Règle du maximum de vraissemblance. On a :

$$P(PPFP|G_1) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{16} \text{ et } P(PPFP|G_2) = \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{8}{81}$$

Comme $\frac{1}{16} < \frac{8}{81}$, la règle de maximum de vraissemblance classe tout nouvel exemple dans le groupe G_2.

Règle de Bayes. Comme dans l'exercice précédent, pour calculer $P(G_1|PPFP)$ et $P(G_2|PPFP)$ ils nous faut calculer $P(PPFP)$. D'après la loi de probabilité totale : $P(PPFP) = P(PPFP|G_1)P(G_1) + P(PPFP|G_2)P(G_2) = \frac{1}{16} \cdot \frac{2}{3} + \frac{8}{81} \cdot \frac{1}{3} = \frac{145}{3 \cdot 8 \cdot 81}$.

D'où : $P(G_1|PPFP) = \frac{P(PPFP|G_1)P(G_1)}{P(PPFP)} = \frac{81}{145} > 0.5$.

La règle de Bayes conduit à attribuer la classe G_1 à l'observation P, P, F, P.

Question 2 Supposons que le bénéfice de l'opération soit mesuré selon la matrice suivante :

Prédit/Réel	G_1	G_2
G_1	0	0
G_2	-1	2

Quelle décision l'observation de la suite P, P, F, P doit-elle entraîner ?

Correction : on s'intéresse ici à maximiser les bénéfices de l'opération. Cette matrice de coût est telle qu'un client prédit comme appartenant au groupe G_1 ne coûte rien à l'entreprise mais ne rapporte rien. Pour savoir quelle décision doit être prise suite à l'observation de P, P, F, P, on doit calculer l'espérance de gain pour chaque décision.

$*$ Pour la décision G_1, l'espérance du gain est nulle.

$*$ Pour la décision G_2, l'espérance du gain vaut $-1 \cdot P(G1|PPFP) + 2 \cdot P(G2|PPFP)$. Ces probabilités ont été calculées à la question précédente :

$$P(G1|PPFP) = \frac{81}{145} \text{ et } P(G2|PPFP) = 1 - P(G1|PPFP) = \frac{64}{145}$$

Le gain vaut donc $-\frac{81}{145} + \frac{128}{145} = \frac{47}{145}$. Il est positif, la décision à prendre est donc G_2 puisque l'espérance de gain est supérieure à celle de la décision G_1.

Laissé en exercice : mêmes questions pour les observations $FFPF$ et $PFPF$.