

Inférence bayésienne

Exercice 1. Le virus de l'immunodéficience humaine (VIH) est responsable du SIDA. Il existe aujourd'hui des tests rapides appelés « Tests rapides d'orientation diagnostique » (TROD) qui ont l'avantage de pouvoir être réalisés à partir d'un échantillon de salive ou d'une goutte de sang prélevée sur le bout du doigt.

1. On a testé avec les deux types de TROD (salivaire et sanguin) deux populations : une première composée de 10 000 personnes infectées par le VIH et une seconde composée de 100 000 personnes non infectées. On a obtenu les résultats suivants.

	Personnes infectées	Personnes non infectées
Tests salivaires positifs	9 803	260
Tests sanguins positifs	9 968	90

Calculer la sensibilité et la spécificité de chacun des deux tests.

2. a. En 2017, la population mondiale exposée au VIH était estimée à 6 milliards et, dans cette population, le nombre de personnes infectées par le VIH était estimé à 37 millions.

Calculer la valeur prédictive positive de chacun des deux tests pour la population mondiale exposée.

- b. En 2017, la population française exposée était estimée à 50 millions et, dans cette population, le nombre de personnes infectées par le VIH était estimé à 150 000.

Calculer la valeur prédictive positive de chacun des deux tests pour la population française exposée.

- c. En 2017, la population sud-africaine exposée était estimée à 35 millions et, dans cette population, le nombre de personnes infectées par le VIH était estimé à 7 millions.

Calculer la valeur prédictive positive de chacun des deux tests pour la population sud-africaine exposée.

- d. Que mettent en évidence les trois exemples précédents ?

Exercice 2. Parmi les femmes de 40 ans ayant effectué une mammographie, 1% a un cancer du sein. À la suite de mammographies sur un échantillon, on a établi que :

- pour 82% des femmes ayant un cancer du sein, la mammographie détecte une anomalie ;
- pour 9% des femmes n'ayant pas de cancer du sein, la mammographie détecte une anomalie.

On suppose que 10 000 femmes de 40 ans ont effectué une mammographie.

1. Déterminer la sensibilité et la spécificité d'une mammographie.
2. Compléter le tableau suivant.

	Anomalie détectée	Pas d'anomalie détectée	Total
Personnes malades			
Personnes non malades			
Total			10 000

3. Une femme de 40 ans a effectué une mammographie qui a permis de détecter une anomalie. Quelle est la probabilité qu'elle soit atteinte d'un cancer du sein ?
4. Calculer les valeurs prédictives positive et négative d'une mammographie chez les femmes de 40 ans.

Exercice 3. On suppose que la somme de la sensibilité et de la spécificité d'un test est égale à 1 (c'est-à-dire, avec les notations du cours, $Se + Sp = 1$).

Expliquer pourquoi le test est inutile dans ce cas.

Exercice 4. Détection de spams

Après une phase d'apprentissage sur un nombre important de courriels, il est estimé que 12% des courriels sont des courriels indésirables (ou spams). 95% des courriels indésirables contiennent l'expression E alors que 99% des courriels légitimes ne la contiennent pas.

1. Pour 10000 courriels reçus, déterminer le nombre de courriels qui :

- Sont indésirables ;
- Sont indésirables et contiennent l'expression E ;
- Sont légitimes et ne contiennent pas l'expression E.

2. Recopier et compléter le tableau suivant :

	Contient l'expression E	Ne contient pas l'expression E	Total
Indésirable			
Légitime			
Total			10000

3. Déterminer la proportion de courriels qui :

- Sont indésirables parmi ceux qui contiennent l'expression E ;
- Sont indésirables parmi ceux qui ne contiennent pas l'expression E.

4. Un détecteur de spams déplace les courriels qui contiennent l'expression E dans le dossier « Spams » et il accepte les autres.

En septembre, il en a déplacé 81 et il en a accepté 459 sur la boîte électronique de Mei. Estimer le nombre de courriels indésirables que Mei a reçus en septembre.