Exercice 1

On s'intéresse au problème suivant : nous avons un test pour savoir si une personne est enceinte ou non. Sur un échantillon de 2500 personnes, nous avons trouver 80% des personnes enceintes. Par ailleurs, parmi les 2000 personnes que notre test donne non-enceinte, 75% n'étaient effectivement pas enceintes.

Question 1 : Dresser le tableau de contingence.

Le tableau de contingence est

400	500
100	1500

Question 2 : Calculer les paramètres de l'évaluation.

Nous allons calculer la sensibilité, spécificité, VPP et VPN.

$$\begin{aligned} \text{sensibilité} &= \frac{VP}{VP+FN} = \frac{400}{400+100} \approx 80\% \\ \text{spécificité} &= \frac{VN}{FP+VN} = \frac{1500}{1500+500} = 75\% \\ \text{VPP} &= \frac{VP}{VP+FP} = \frac{400}{400+500} = 44.4\% \\ \text{VPN} &= \frac{VN}{VN+FN} = \frac{1500}{1500+100} = 93.75\% \end{aligned}$$

Question 3 : Changez les valeurs de VP et FP de telle sorte à ce que la VPP soit égale à 90% et la FN soit à 0

Le tableau de contingence devient

900	100
0	1500

Exercice 2

Nous avons une maladie M avec une valence de 60%. Et notre modèle a une précision de 80%. Enfin, parmi les non-malades, nous trouvons 95% de tests négatifs

Question 1 : Traduire l'énoncé en termes de probabilités.

$$\begin{aligned} \mathbb{P}(M) &= 0.60 \\ \mathbb{P}((M \cap S) \cup (\bar{M} \cap \bar{S})) &= 0.80 \\ \mathbb{P}(\bar{S}|\bar{M}) &= 0.95 \end{aligned}$$

Question 2 : Calculer les les paramètres de l'évaluation.

Pour résoudre ce problème, on commence par rappeler les définitions des paramètres de l'évaluation en terme de probabilités conditionnelles.

$$\begin{aligned} \text{sensibilité} &= \mathbb{P}(S|M) \\ \text{spécificité} &= \mathbb{P}(\bar{S}|\bar{M}) \\ \text{VPP} &= \mathbb{P}(M|S) \\ \text{VPN} &= \mathbb{P}(\bar{M}|\bar{S}) \end{aligned}$$

Ensuite, nous allons essayer de voir ce que l'on peut déduire de l'énoncé.

$$\mathbb{P}(\bar{M} \cap \bar{S}) = \mathbb{P}(\bar{S}|\bar{M})(1 - \mathbb{P}(M)) = 0.95(1 - 0.6) = 0.38$$

En conséquence,

$$\mathbb{P}(M \cap S) = \mathbb{P}((M \cap S) \cup (\bar{M} \cap \bar{S})) - \mathbb{P}(\bar{M} \cap \bar{S}) = 0.80 - 0.38 = 0.42$$

On peut calculer la probabilité de S.

$$\mathbb{P}(S \cap \bar{M}) = \mathbb{P}(S|\bar{M})\mathbb{P}(\bar{M}) = (1 - \mathbb{P}(\bar{S}|\bar{M}))(1 - \mathbb{P}(M)) = (1 - 0.95)(1 - 0.60) = 0.05 \times 0.4 = 0.02$$

Ainsi,

$$\mathbb{P}(S) = \mathbb{P}(S \cap \bar{M}) + \mathbb{P}(S \cap M) = 0.42 + 0.02 = 0.44$$

Enfin, nous devrions avoir tous les éléments pour résoudre l'exercice

$$\begin{aligned}
\text{sensibilité} = \mathbb{P}(S|M) = \frac{\mathbb{P}(M \cap S)}{\mathbb{P}(M)} = \frac{0.42}{0.60} &= 0.7 \\
\text{spécificité} = \mathbb{P}(\bar{S}|\bar{M}) = \frac{\mathbb{P}(\bar{M} \cap \bar{S})}{\mathbb{P}(\bar{M})} = \frac{0.38}{0.40} &= 0.95 \\
\text{VPP} = \mathbb{P}(M|S) = \frac{\mathbb{P}(M \cap S)}{\mathbb{P}(S)} = \frac{0.42}{0.44} &\approx 0.9545 \\
\text{VPN} = \mathbb{P}(\bar{M}|\bar{S}) = \frac{\mathbb{P}(\bar{M} \cap \bar{S})}{\mathbb{P}(\bar{S})} = \frac{0.38}{0.56} &= 0.67857
\end{aligned}$$

Exercice 3

Le but de cet exercice est de s'entraîner à formuler correctement le problème pour appliquer le TCL.

Question 1 : On considère une chaîne de production de boule de billard. On constate qu'en moyenne la machine est précise moyenne à 1mm près avec une variance de 0.5mm sur l'échantillon. Ces valeurs sont obtenues sur un échantillon de 35 éléments.

- [x] X suit une loi gaussienne
- [ ] il faut étudier la variable moyenne de X
- [x] on peut appliquer le TCL

Question 2 : On considère des boules de billards produites par un humain. On constate que pour une boule, nous avons une précision moyenne de 3mm et une variance de 1mm.

- [ ] X suit une loi gaussienne
- [x] il faut étudier la variable moyenne de X
- [ ] on peut appliquer le TCL

Question 3 : On considère un modèle prédictif qui classe des chat et des chiens. Il trouve correctement la classe avec une probabilité de 0.99 **sur un échantillon de** 100 **exemples.**

- [ ] X suit une loi gaussienne
- [x] il faut étudier la variable moyenne de X
- [ ] on peut appliquer le TCL

Question 4 : En reprenant la question précédente mais sur un échantillon de 10000 **exemples :**

- [ ] X suit une loi gaussienne
- [x] il faut étudier la variable moyenne de X
- [x] on peut appliquer le TCL

Exercice 4

Pour la question 1 et 4, je vous demander trouver l'intervalle de confiance à 95%

Dans un cas, comme dans l'autre, nous allons résoudre le problème suivant

$$\mathbb{P}\left(M_N \in [\mu - a; \mu + a]\right) = \mathbb{P}\left(M_N - \mu \in [-a; +a]\right) = \mathbb{P}\left(\frac{M_N - \mu}{\sigma} \in \left[-\frac{a}{\sigma}; \frac{a}{\sigma}\right]\right)$$

Il faut donc trouver μ et σ. Pour la question 1, on a simplement $\mu = 1$ et $\sigma = 0.5$ puisqu'on s'intéresse directement à X qui suit une loi gaussienne. Pour la question 4, on a $\mu = 0.99$ (espérance d'une loi de bernoulli) et $\sigma = \frac{0.99}{10000}$.