# Travaux Dirigés

**• Méthode du maximum de vraisemblance MV (maximum likelihood ML) :**

○ Cette méthode permet de calculer, à partir d'un échantillon observé, la (les) meilleure(s) valeur(s) d'un paramètre d'une loi de probabilité.

○ *Le principe de la méthode MV :* Si un phénomène $X$ a été l'objet de n observations indépendantes $x_1, x_2, ..., x_n$ les unes des autres, sa loi de probabilité $P(X = x)$(dans le cas discret : loi binomiale, loi de Poisson) ou sa densité (en cas de loi continue, comme la loi normale) est une fonction $f(x; \theta_1, ..., \theta_n)$ où $\theta_1, ..., \theta_n$ sont les paramètres de la loi.

· *Dans le cas discret :* on définit la fonction de **MV** est la probabilité de l'échatillon observé en fonction des paramètres $\theta = (\theta_1, ..., \theta_n)$ :

$$\begin{aligned} L(x_1, x_2, ..., x_n; \theta) &= P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n; \theta) \\ &= P_X\ (x_1, x_2, ..., x_n; \theta), \end{aligned}$$

avec $X = (X_1, X_2, ..., X_n)$.

· *Dans le cas continu :* on définit la fonction de **MV** est définie par

$$L(x_1, x_2, ..., x_n; \theta) = P_X\ (x_1, x_2, ..., x_n; \theta).$$

· On maximise la fonction $L(x_1, x_2, ..., x_n; \theta)$ sur l'ensemble des paramètres $\theta$ pour trouver $\widehat{\theta}$ l'estimateur de $MV$,

$$\widehat{\theta} = \arg\max_{\theta} L(x_1, x_2, ..., x_n; \theta).$$

**Exercice 1.** Soit la variable aléatoire $X \rightsquigarrow \mathcal{B}(n, p)$ et $(X_1, X_2, ..., X_n)$ un échatillon $i.i.d$ de même loi que $X$.

a) Calculer la fonction de **MV** de cette échatillon.

b) Estimer le paramètre $p$ de cette loi.

**Exercice 2.** Soit la variable aléatoire $X \rightsquigarrow \mathcal{N}(m, \sigma)$ et $(X_1, X_2, ..., X_n)$ un échatillon $i.i.d$ de même loi que $X$.

a) Calculer la fonction de **MV** de cette échatillon.

b) Estimer les paramètres de cette loi.

**Exercice 3.**

a) Montrer les propriétés suivantes :

•
$$OR(x, \widetilde{x}) > 1 \Longleftrightarrow P(Y = 1|X = x) > P(Y = 1|X = \widetilde{x}).$$

•
$$OR(x, \widetilde{x}) = 1 \Longleftrightarrow P(Y = 1|X = x) = P(Y = 1|X = \widetilde{x}).$$

•
$$OR(x, \widetilde{x}) < 1 \Longleftrightarrow P(Y = 1|X = x) < P(Y = 1|X = \widetilde{x}).$$

• si $P(Y = 1|X = x)$ et $P(Y = 1|X = \widetilde{x})$ sont très petits par rapport à 1, on peut faire l'approximation
$$OR(x, \widetilde{x}) = \frac{P(Y = 1|X = x)}{P(Y = 1|X = \widetilde{x})}.$$

b) Quel est le meilleurs modèles parmi les trois modèles suivant :

· Modèle 1 : AIC = 51.09.

· Modèle 2 : AIC = 42.87.

· Modèle 3 : AIC = 43.10.

$$odds(x) = \frac{P(Y = 1|X = x)}{1 - P(Y = 1|X = x)}.$$

• L'odds ratio $OR$ (**rapport des odds**) (**rapport des chances**) entre deux individus $x$ et $\widetilde{x}$ est
$$OR(x, \widetilde{x}) = \frac{odds(x)}{odds(\widetilde{x})}.$$

**Exercice 4.** Données sur maladie coronarienne :

AGE : age.

CHD : diagnostic de maladie coronarienne.

| $CHD \setminus Age$ | $x = 1$ $(Age \geq 55)$ | $x = 0$ $(Age < 55)$ | $Total$ |
|---|---|---|---|
| $y = 1$ $(Yes)$ | 21 | 22 | 43 |
| $y = 0$ $(No)$ | 6 | 51 | 57 |
| $Total$ | 27 | 73 | 100 |

a) Quelle est la loi de la variable $y =$"CHD" ?

b) Déterminer les paramètres $y$.

c) Quelle est l'espérance de $y$ ?

d) Déterminer l'expression de la fonction reliant l'espérace de la variable $y =$"CHD" et la variable explicative $x =$"Age".

e) Calculer les "*odds*" de $x = 1$ et de $x = 0$. Déduire "Odds Ratio" $OR\,(x = 1, x = 0)$ .