Série 9

Exercice 1. Vous voulez connaître le pourcentage de femmes parmi les étudiant-e-s de l'EPFL, mais vous n'avez pas accès à la liste d'étudiant-e-s. Vous décidez donc d'utiliser ce que vous avez appris dans le cours afin d'estimer ce pourcentage.

(a) Quel sera votre modèle statistique ?

(b) Quel sera le paramètre d'intérêt ?

(c) Comment choisissez-vous votre échantillon ?

(d) Quel sera votre estimateur (un choix intuitif est suffisant ici) ?

(e) L'estimateur est une variable aléatoire. Comment voit-on cela dans votre situation ?

(f) Etant une variable aléatoire, votre estimateur a une espérance et une variance. Supposez que le vrai pourcentage est de 40 % (valeur fictive !), et que vous avez un échantillon de taille 100. Calculez l'espérance et la variance de votre estimateur. Comment ces valeurs changent-elles en fonction de la taille de l'échantillon ? Est-ce que votre estimateur est biaisé ?

(g) Si le vrai pourcentage est de 40 %, quelle est la taille de l'échantillon dont vous avez besoin afin d'être certain à 95 % que votre estimateur sera plus petit que 50 % ? Indication : utilisez une loi approximative.

Exercice 2. Soient $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{B}(p)$ avec $p \in (0, 1)$.

(a) Soit $x_1, x_2, \ldots, x_n$ une réalisation de cet échantillon. Ecrivez la fonction de vraisemblance de p basée sur cette réalisation.
Rappel : pour une variable $X \sim \mathcal{B}(p)$ on a $P(X = x) = p^x(1 - p)^{1-x}$ pour $x = 0, 1$.

(b) Trouvez $\hat{p}_{MC}$, l'estimateur des moindres carrés de p.

(c) Trouvez $\hat{p}_{ML}$, l'estimateur du maximum de vraisemblance de p.

(d) Les estimateurs définis ci-dessus sont-ils biaisés ? Trouvez leur variance.

Exercice 3. Soit X_1 une variable aléatoire dont la fonction de densité est $f(x) = cx^{\theta-1}$ pour $x \in (0, 1)$ et $f(x) = 0$ autrement. Ici c est une constante et $\theta > 0$ est un paramètre inconnu.

(a) Calculez c.

(b) Calculez l'espérance de X_1.

Maintenant, considérons un échantillon $X_1, X_2, \ldots, X_n$ de variables de cette distribution.

(c) Trouvez $\hat{\theta}_{ML}$, l'estimateur du maximum de vraisemblance de θ.

Exercice 4. On aimerait connaître le numéro de plaque d'immatriculation le plus élevé dans le canton de Vaud. On fait un tour au stationnement de l'EPFL et on relève les numéros de quelques plaques vaudoises (données fictives) :

58285, 166933, 277869, 170764, 205779, 289401, 190749, 132604, 102773, 91313, 286130, 111962, 274737, 298158, 140726, 286010, 86987, 139899, 43212, 262779.

Intuitivement, estimez le numéro de plaque le plus élevé dans le canton à l'aide de cet échantillon (une solution intuitive est suffisante ici : l'exercice suivant propose des solutions rigoureuses).

Exercice 5. Soit $X_1, X_2, \ldots, X_n$ un échantillon de la loi uniforme $\mathrm{U}[0, \theta]$ avec $\theta > 0$.

(a) Donnez la fonction de vraisemblance de θ basée sur une réalisation $x_1, \ldots, x_n$.

(b) Dessinez cette fonction (de θ).

(c) Montrez que $M_n = \max(X_1, X_2, \ldots, X_n)$ est l'estimateur du maximum de vraisemblance.

(d) Trouvez le biais de M_n, et déduisez-en un estimateur non-biaisé de θ. On appellera ce nouvel estimateur $\hat{\theta}_{NB}$. Quelle est sa variance ?

(e) Parmi les estimateurs $\hat{\theta}_{ML}$, et $\hat{\theta}_{NB}$ lequel choisiriez-vous et pourquoi ?

(f) Estimez le numéro de plaque le plus élevé dans le canton à l'aide de l'échantillon de l'exercice précédent en utilisant les deux méthodes d'estimation définies plus haut, et comparez les résultats avec celui que vous avez suggéré dans l'exercice précédent.