Fiche de TD – Statistiques Inférentielles

Linda MATSING FOGANG

Exercice 1 : Population, échantillon et statistiques descriptives

1.1 Rappel de cours

- Une population est l'ensemble des individus sur lesquels porte une étude.
- Un échantillon est une sous-partie de la population utilisée pour faire des inférences.
- Une statistique est une fonction calculée à partir de l'échantillon (ex : moyenne, variance).

Objectif : décrire l'échantillon pour comprendre la population.

1.2 Énoncé

Une entreprise souhaite étudier les salaires de ses employés. Elle sélectionne aléatoirement un échantillon de 30 employés et relève leurs salaires mensuels (en euros) :

2200	2500	2400	2300	2700	2100
2600	2500	2350	2450	2550	2400
2500	2750	2650	2200	2300	2500
2550	2600	2450	2350	2400	2550
2500	2650	2700	2400	2300	2250

(a) Calculez la moyenne, la médiane, la variance et l'écart-type de l'échantillon.

(b) Commentez la distribution : est-elle symétrique ? Présente-t-elle des outliers ?

(c) En quoi cet échantillon permet-il de dire quelque chose sur la population ?

Exercice 2 : Estimateur, biais et consistance

2.1 Rappel de cours

Un estimateur est une variable aléatoire utilisée pour estimer un paramètre inconnu. Il est sans biais si son espérance est égale au paramètre. Il est consistant si sa probabilité de s'éloigner du paramètre tend vers 0 quand $n \to \infty$.

2.2 Énoncé

On observe un échantillon $X_1, ..., X_n \sim \mathcal{U}[0, \theta]$. On propose deux estimateurs de θ : $T_1 = 2\bar{X}_n$ et $T_2 = \max(X_1, ..., X_n)$.

(a) Calculez le biais de chacun.

(b) Étudiez leur consistance.

Exercice 3 : Propriétés d'un bon estimateur (biais, variance, EQM)

3.1 Rappel de cours

L'erreur quadratique moyenne (EQM) d'un estimateur est :

$$\mathrm{EQM}(T) = \mathrm{Var}(T) + \mathrm{Biais}^2(T)$$

3.2 Énoncé

On veut estimer la moyenne μ d'une loi normale $\mathcal{N}(\mu, \sigma^2)$. On compare deux estimateurs : $T_1 = \bar{X}_n$, $T_2 = \bar{X}_n + 1$.

(a) Calculez leur biais, variance et EQM.

(b) Lequel est préférable et pourquoi ?

Exercice 4 : Méthode des moments et estimation du maximum de vraisemblance

4.1 Rappel de cours

— Méthode des moments : on égalise les moments empiriques et théoriques.

— EMV : on maximise la fonction de vraisemblance pour estimer les paramètres.

4.2 Énoncé

Soit $X_1, ..., X_n \sim \mathrm{Exp}(\lambda)$.

(a) Estimez λ par la méthode des moments.

(b) Estimez λ par EMV.

(c) Comparez les deux estimateurs.

Exercice 5 : Intervalle de confiance pour une moyenne

5.1 Rappel de cours

Si $X_i \sim \mathcal{N}(\mu, \sigma^2)$, l'intervalle de confiance de niveau $1 - \alpha$ pour μ est :

$$\left[\bar{X}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

5.2 Énoncé

Une machine remplit des bouteilles. Un échantillon de 36 bouteilles donne une moyenne de 498 ml et un écart-type de 4 ml.

(a) Construisez un IC à 95% pour le volume moyen.

(b) Que conclure sur la conformité de la machine (objectif : 500 ml) ?

Exercice 6 : Test d'hypothèse classique

6.1 Rappel de cours

H_0 : hypothèse nulle, H_1 : hypothèse alternative. On rejette H_0 si la statistique dépasse un certain seuil. Risques d'erreur : α (faux positif), β (faux négatif).

6.2 Énoncé

On veut tester si un médicament a un effet. $H_0 : \mu = 0$, $H_1 : \mu > 0$ Échantillon de 50 patients, moyenne $= 1.1$, $\sigma = 2$.

(a) Réalisez le test à 5%.

(b) Calculez la p-valeur.

(c) Que signifie rejeter H_0 ici ?

Exercice 7 : Théorème Central Limite (TCL)

7.1 Rappel de cours

TCL : sous certaines conditions,

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1)$$

7.2 Énoncé

On observe $X_1, ..., X_{100} \sim \text{Poisson}(\lambda = 3)$.

(a) Approximez $\mathbb{P}(2.8 < \bar{X}_{100} < 3.2)$.

(b) Vérifiez la validité de l'approximation normale.

(c) Simulez cette probabilité et comparez.

Exercice 8 : Lemme de Slutsky et convergence en loi

8.1 Rappel de cours

Si $X_n \xrightarrow{d} X$ et $Y_n \xrightarrow{P} c$, alors $X_n Y_n \xrightarrow{d} cX$.

8.2 Énoncé

Soient $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, et soit :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad \text{et} \quad Y_n = 1 + \frac{1}{n}.$$

(a) Montrez n utilisant le lemme de Slutsky que $\sqrt{n}\,\bar{X}_n\,Y_n \xrightarrow{d} \mathcal{N}(0,1)$.

(b) Expliquez intuitivement pourquoi ce résultat est vrai.

Exercice 9 : Loi forte des grands nombres (LFGN)

9.1 Rappel de cours

LFGN : $\bar{X}_n \xrightarrow{a.s.} \mu$ si X_i i.i.d. avec $\mathbb{E}[X_i] = \mu$.

9.2 Énoncé

Simulez 1000 réalisations de $X_i \sim \mathcal{U}[0,1]$, calculez $\bar{X}_n$ pour n allant de 1 à 1000.

(a) Tracez $\bar{X}_n$ en fonction de n.

(b) Que constatez-vous ?

(c) En quoi cela illustre-t-il la LFGN ?

Exercice 10 : Bootstrap, jackknife et estimation empirique de l'EQM

10.1 Rappel de cours

— Bootstrap : on rééchantillonne avec remise pour estimer la variance ou l'EQM.

— Jackknife : on laisse de côté une observation à chaque fois.

— Ces méthodes sont non paramétriques.

10.2 Énoncé

On observe $x = [2, 3, 5, 7, 11]$.

(a) Estimez la moyenne et son écart-type par bootstrap ($B = 1000$).

(b) Estimez la même chose avec le jackknife.

(c) Comparez les deux méthodes.