Statistiques

Remise à niveau

Bastien Le Chenadec

Table des matières

Première partie

Cours

1 Modèles statistiques

Définition 1.1. Une observation est un point d'un espace mesurable $(\mathcal{Z}, \mathcal{A})$. Un modèle statistique est la donnée de $(\Omega, \mathcal{F}, \mathcal{P}, Z)$.

- Ω univers
- $\mathcal{F} \subset \mathcal{P}(\Omega)$ tribu sur Ω
- $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$

Remarque 1.1.

- On suppose qu'il existe $\theta^* \in \Theta$ tel que $Z \sim P_{\theta^*}^Z$.
- Souvent $\mathcal{Z} = \mathbb{R}^{np}, Z = (X_1, \dots, X_n), X_i \in \mathbb{R}^p$.
- On suppose souvent les X_i i.i.d.
- Si $\Theta \subset \mathbb{R}^q$ (avec $q << n$) on parle de modèle paramétrique.

Exemple 1.1.

$\Omega = \mathbb{N}, \Theta = [0,1]$, n fixé. $\mathcal{P} = \{\mathcal{B}(n,\theta), \theta \in \Theta\}$

Modèle multinomial : $Z \sim \mathcal{M}(\pi_1, \dots, \pi_q, n)$.

$$\begin{aligned} P(Z = (n_1, \dots, n_q)) &= \binom{n}{n_1, \dots, n_q} \pi_1^{n_1} \dots \pi_q^{n_q} \mathbb{1}_{\{n_1 + \dots + n_q = n\}} \\ &= \frac{n!}{n_1! \dots n_q!} \pi_1^{n_1} \dots \pi_q^{n_q} \mathbb{1}_{\{n_1 + \dots + n_q = n\}} \end{aligned}$$

Modèle gaussien : $\Theta = \mathbb{R} \times \mathbb{R}_+^*$. $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2), (\mu, \sigma^2) \in \Theta\}$

Modèle gaussien multivarié : $\mu \in \mathbb{R}^p, \Sigma \in S_p^{++}(\mathbb{R})$. $\mathcal{P} = \{\mathcal{N}(\mu, \Sigma), (\mu, \Sigma) \in \Theta\}$

Définition 1.2. $\mathcal{M}$ est identifiable si $\theta \mapsto P_\theta$ est injective.

Définition 1.3. Une statistique (T) est une fonction mesurable sur $\mathcal{Z}$.

Exemple 1.2.

- Moyenne empirique $T(Z) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{Z}$.
- Moments empiriques.
- Variance empirique.
- Médiane...

On veut étudier la loi (asymptotique) de $T(Z)$.

Définition 1.4. Une procédure statistique est une fonction :

$$d = \phi \circ Z$$

où $\phi : \mathcal{Z} \mapsto D$ est un "traitement". (D est l'ensemble des décisions possibles).

Définition 1.5. On se donne une fonction de perte :

$$C : \begin{cases} \{\Omega \to D\} \times \Theta \to \mathbb{R}_+ \\ (d, \theta) \mapsto c(d, \theta) \end{cases}$$

Le risque est $R_\theta(d) = \mathbb{E}_\theta[C(d, \theta)]$.

2 Estimation

2.1 Non paramétrique

Définition 2.1 (Mesure empirique)**.** La mesure empirique est la statistique :

$$\hat{P} : \begin{cases} \mathcal{Z} \to \{\text{mesures de probas sur } \Omega\} \\ (X_i)_{1 \leq i \leq n} \mapsto \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i\}} \end{cases}$$

Remarque 2.1.

L'espérance et la variance empiriques sont $\mathbb{E}_{\hat{P}}$ et $\mathrm{Var}_{\hat{P}}$.

Définition 2.2 (Fonction de répartition empirique)**.**

$$\hat{F} : x \mapsto \hat{P}(]-\infty, x])$$

Théorème 2.1 (Glivenko-Cantelli)**.** $\hat{F}_n$ converge uniformément vers F presque sûrement.

$$\sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F(x) \right| \xrightarrow[n \to \infty]{p.s.} 0$$

Théorème 2.2 (Donsker)**.**

$$\sqrt{n} \left(\hat{F}_n - F \right) \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0, F)$$

où $F(x, x') = F(\min(x, x')) - F(x)F(x')$.

Définition 2.3 (Densité empirique)**.** Soient $a_0 < \cdots < a_N \in \mathbb{R}$.

$$\hat{f}_n = \frac{1}{n} \sum_{i=1}^N \left(\frac{1}{a_i - a_{i-1}} \sum_{j=1}^N \mathbb{1}_{\{]a_{i-1}, a_i]\}}(x_j) \right) \mathbb{1}_{\{]a_{i-1}, a_i]\}}$$

Remarque 2.2.

Les segments sont constants, $\hat{f}_n$ est un mauvais estimateur. Voir les méthodes à noyau.

2.2 Méthode des moments

$\Theta \subset \mathbb{R}^q, Z = (X_i)_{1 \leq i \leq n}$ iid.

$\forall \theta \in \Theta$, le moment d'ordre k est :

$$\mu_{\theta,k} = \mathbb{E}_\theta(X_1^k)$$

Le moment empirique d'ordre k est :

$$\frac{1}{n} \sum_{i=1}^{n} X_i^k$$

On résout :

$$\begin{cases} \mu_{0,1} & = \sum_{i=1}^{n} X_i \\ \mu_{0,q} & = \sum_{i=1}^{q} X_i^q \end{cases}$$

(q équations, q inconnues).

Exemple 2.1.

$\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+^*\}$
$\mathbb{E}_{(\mu,\sigma^2)}(X_1) = \mu$
$\mathbb{E}_{(\mu,\sigma^2)}(X_1^2) = \sigma^2 + \mu^2$
On observe $(x_1, \ldots, x_n)$.

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\hat{\mu}^2 + \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \hat{\mu}^2 = \text{variance asymptotique}$$

2.3 Maximum de vraisemblance

On représente $\mathcal{P}$ par ses densités :

$$\mathcal{P} = \{P_\theta, \theta \in \Theta\}$$

Définition 2.4. La fonction de vraisemblance est définie par :

$$L : \begin{cases} \mathcal{Z} \times \Theta \to \mathbb{R}^+ \\ (Z, \theta) \mapsto P_\theta(Z) \end{cases}$$

Si $Z = (X_i)_{1 \leq i \leq n}$ iid, $L(Z, \theta) = \prod_{i=1}^{n} P_\theta(X_i)$.

Définition 2.5 (MLE)**.** L'estimateur du maximum de vraisemblance est :

$$\text{MLE} = \arg\max_{\theta \in \Theta} L(z, \theta)$$

pour une réalisation z de Z.

Remarque 2.3.

Dans le cas iid, on regarde :

$$\ln(L(z,\theta)) = \sum_{i=1}^{n} \ln(P_\theta(x_i))$$

3 Propriétés des estimateurs

3.1 Risque, biais, variance

$\Theta \subset \mathbb{R}^q$, $\mathcal{P}$ représenté par les densités, $Z = (X_i)_{1\leq i\leq n}$ iid de loi $P_{\Theta^*}, \Theta^* \in \Theta$. Pour tout n, notre estimateur est une statistique $T_n(Z)$.

Définition 3.1 (Risque quadratique)**.**

$$R(T_n, \theta^*) = \mathbb{E}_{\theta^*}(||T_n(Z) - \theta^*||^2)$$

Définition 3.2. T est dit meilleur que T' si :

- $\forall\theta \in \Theta, R(T,\theta) \leq R(T',\theta)$
- $\exists\theta_0 \in \Theta, R(T,\theta_0) < R(T',\theta_0)$

Définition 3.3. Si $\mathbb{E}(||T||) < +\infty$, le biais de T est :

$$b_{\theta^*} = \mathbb{E}_{\theta^*}(T) - \theta^*$$

Si $b_\theta(T) = 0$ pour tout $\theta \in \Theta$, T est dit non-biaisé (ou sans biais).

Proposition 3.1 (Décomposition biais-variance)**.** On suppose $\mathbb{E}(||T||^2) < +\infty$.

$$R(T,\theta^*) = ||b_{\theta^*}(T)||^2 + \mathrm{Tr}(V_{\theta^*}(T))$$

où $V_{\theta^*}(T)$ est la matrice de covariance de T.

Définition 3.4. Soient T, T' deux estimateurs de même biais. T est dit plus efficace que T' si :

$$\begin{cases} \forall\theta \in \Theta, V_\theta(T) & \leq V_\theta(T') \\ \exists\theta_0 \in \Theta, V_{\theta_0}(T) & < V_{\theta_0}(T') \end{cases}$$

où :

$$\begin{cases} A \leq b \Leftrightarrow B - A \in S_n^+(\mathbb{R}) \\ A < b \Leftrightarrow B - A \in S_n^{++}(\mathbb{R}) \end{cases}$$

3.2 Information de Fisher et efficacité

Définition 3.5. Sous certaines hypothèses de régularité, l'information de Fisher est la matrice $q \times q$ telle que :

$$I(\theta) = -\mathbb{E}_\theta\left(\frac{\partial \ln(p_\theta(Z))}{\partial\theta_i\partial\theta_j}\right)$$

Proposition 3.2. Soient $\mathcal{M}$, $\mathcal{M}'$ deux modèles statistiques avec le même Θ. Soient $X \sim P_\theta \in \mathcal{P}$ et $Y \sim P'_\theta \in \mathcal{P}'$ tels que $X \perp\!\!\!\perp Y$.

$$I_{(X,Y)}(\theta = I_X(\theta)) + I_Y(\theta)$$

Corollaire 3.1. Si $Z = (X_i)_{1 \leq i \leq n}$ iid,

$$I_Z(\theta) = nI_{X_1}(\theta)$$

On a toujours $I(\theta) \in S_n^+(\mathbb{R})$ et on aura souvent $I(\theta) \in S_n^{++}(\mathbb{R})$.

Théorème 3.1 (Borne de Cramer-Rao)**.** Soit T un estimateur sans biais de paramètre $\theta \in \Theta$. Sous certanies hypothèses (dont $I(\theta) \in S_n^{++}(\mathbb{R}) \forall \theta \in \Theta$), on a pour tout $\theta \in \Theta$ tel que $\mathbb{E}(||T||^2) < +\infty$:

$$V_\theta(T) \geq I^{-1}(\theta)$$

Définition 3.6. S'il y a égalité, T est dit efficace.

Corollaire 3.2. Si $Z = (X_i)_{1 \leq i \leq n}$ respecte les hypothèses, alors :

$$V_\theta(T) \geq \frac{1}{n} I_{X_1}^{-1}(\theta)$$

3.3 Propriétés asymptotiques

Définition 3.7. T_n est asymptotiquement sans biais si $\forall \theta \in \Theta, b_\theta(T_n) \xrightarrow[n \to +\infty]{} 0$.

Définition 3.8. T_n est consistant (convergent) si $\forall \theta \in \Theta, T_n \xrightarrow[n \to +\infty]{proba} \theta$. S'il y a convergence presque sûre, (T_n) est fortement consistant.

Définition 3.9. Soit T un estimateur consistant tel qu'il existe une variable aléatoire Y et une suite $(a_n)_{n \geq 0}$ strictement positive tels que :

$$a_n(T_n - \theta) \xrightarrow{loi} Y, \forall \theta$$

Alors (a_n) est la vitesse de convergence de T.

Exemple 3.1.

- En pratique, on a souvent $\sqrt{n} \leq a_n \leq n$.
- Lois $\mathcal{N}(\mu, 1)$ moyenne empirique. $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ consistant par LGN. TCL : $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{loi} \mathcal{N}(0,1)$.

Définition 3.10. Un estimateur T est asymptotiquement normal s'il existe une matrice $\Sigma(\theta) \in S_n^{++}(\mathbb{R})$ telle que :

$$\sqrt{n}(T_n - \theta) \xrightarrow{loi} \mathcal{N}(0, \Sigma(\theta)), \forall \theta$$

Proposition 3.3. Asymptotiquement normal $\implies$ consistant.

Définition 3.11. Soient T, T' deux estimateurs asymptotiquement normaux. T est asymptotiquement plus efficace que T' si :

$$\begin{cases} \forall\theta, \Sigma(\theta) & \leq \Sigma'(\theta) \\ \exists\theta_0, \Sigma(\theta_0) & < \Sigma'(\theta_0) \end{cases}$$

Définition 3.12. T est asymptotiquement efficace si :

$$\forall\theta, I^{-1}(\theta) = \Sigma(\theta)$$

Sous de bonnes hypothèses, le MLE est consistant, asymptotiquement normal et asymptotiquement efficace.

3.4 Régions de confiance

Définition 3.13. Soit $\alpha \in [0,1]$ un niveau de risque. Une région de confiance de $I_\alpha(Z)$ de niveau $1-\alpha$ est une statistique à valeurs dans $\mathcal{P}(\Theta)$ telle que :

$$P_\theta(\theta \in I_\alpha(Z)) \geq 1-\alpha$$

S'il y a égalité, on dit que $I_\alpha(Z)$ est de taille $1-\alpha$.

Remarque 3.1.

Si $\Theta \subset \mathbb{R}$, on prend des intervalles de confiance.

Définition 3.14. $Z = (X_i)_{1\leq i\leq n}$ iid. $I_\alpha(Z)$ est région de confiance asymptotique de niveau $1-\alpha$ si :

$$\lim_{n\to+\infty} P_\theta(\theta \in I_\alpha(Z_n)) \geq 1-\alpha$$

Construction des régions de confiance : On veut une statistique $S(Z,\theta)$ dont la loi (connue) ne dépend pas de θ. Pour α fixé, on utilise les quantiles q_γ et $q_{\gamma+1-\alpha}$ pour γ tel que γ et $\gamma+1-\alpha \in [0,1]$. On a alors :

$$P_\theta(q_\gamma \leq S(Z,\theta) \leq q_{\gamma+1-\alpha}) \geq 1-\alpha$$

Sinon, on cherche $S(Z,\theta)$ de loi asymptotique indépendante de θ.

Définition 3.15. Le quantile d'odre γ d'une loi dont on a la fonction de répartition F est :

$$q_\gamma = \inf\{x \in \mathbb{R} : F(x) \geq \gamma\}$$

Exemple 3.2.

- Loi normale $\mathcal{N}(0,1)$: $q_{0.975} = 1.96$. Par symmétrie : $q_{0.025} = -1.96$.
- Pour $X \sim \mathcal{N}(\mu,\sigma^2), S = \frac{X-\mu}{\sigma^2} \sim \mathcal{N}(0,1)$.
  $P(S \in \{-1.96, 1.96\}) = 0.95$.
  $P(\mu - 1.96\sigma^2 \leq X \leq \mu + 1.96\sigma^2) = 0.95$.
- On utilise souvent des lois χ^2 et de Student.

Théorème 3.2. $\Theta \subset \mathbb{R}$, soit T un estimateur asymptotiquement normal de variance asymptotique $V(\theta)$. On a donc :

$$\sqrt{n}(T_n - \theta) \xrightarrow{loi} \mathcal{N}(0, V(\theta)), \forall\theta$$

On suppose $V(\theta) > 0 \forall\theta$ et V continue en θ. Alors :

$$\sqrt{n}\frac{T_n - \theta}{\sqrt{V(\theta)}} \xrightarrow{loi} \mathcal{N}(0,1)$$

Et :

$$\sqrt{n}\frac{T_n - \theta}{\sqrt{V(T_n)}} \xrightarrow{loi} \mathcal{N}(0,1)$$

Corollaire 3.3.

$$I_\alpha = \left[T_n - q_{1-\frac{\alpha}{2}}\sqrt{\frac{V(T_n)}{n}}, T_n + q_{1-\frac{\alpha}{2}}\sqrt{\frac{V(T_n)}{n}}\right]$$

est un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour θ.

Démonstration.
- 1er cas : TCL
- V est $\mathcal{C}^0$ et > 0 donc $\frac{\sqrt{V(\theta)}}{\sqrt{V(T_n)}} \xrightarrow{proba} 1$. T_n est consistant.
- Théorème de Slutsky : $X_n \xrightarrow{loi} X$ et $Y_n \xrightarrow{proba} c \in \mathbb{R}$ alors $(X_n, Y_n) \xrightarrow{loi} (X, c)$.
  Ici, $\sqrt{n}\frac{T_n-\theta}{\sqrt{V(\theta)}}\frac{\sqrt{V(\theta)}}{\sqrt{V(T_n)}} \xrightarrow{loi} \mathcal{N}(0,1)$.

□

3.5 Estimation bayesienne

Idée : Mettre une loi de probabilités sur Θ.

Définition 3.16. Un modèle statistique bayesien est un modèle statistique $\mathcal{M}$ muni d'une loi de probabilités Π sur $(\Theta, \mathcal{F}_\Theta)$ avec $\mathcal{F}_\Theta$ une tribu sur Θ. Π est la loi *a priori* (ou *prior*).

On prend $\Theta \subset \mathbb{R}^q$, $\mathcal{F}_\Theta = \mathcal{B}(\Theta)$ et on suppose que les P_θ ont des densités p_θ. On suppose aussi que Π a une densité π.

On note :

$$p(z|\theta) = p_\theta(z) = L(z, \theta)$$

Et :

$$P_Z(z) = \int_\Theta p(z|\theta)\pi(\theta)d\theta$$

Formule de Bayes : Si $P_Z(z) > 0$ alors :

$$p(\theta|z) = \frac{p(z|\theta)\pi(\theta)}{P_Z(z)} \propto p(z|\theta)\pi(\theta)$$

On veut estimer $p(\theta|z)$: loi a posteriori.

Remarque 3.2.
Le choix de la loi a posteriori est important. Par exemple, utiliser les estimations faites sur une population pour définir le priori pour un nouvel individu $\rightarrow$ modèles à effets mixtes.

4 Tests d'hypothèses

4.1 Généralités

Soit $\mathcal{M}$ un modèle statistique de loi inconnue $P^* \in \mathcal{P}$. On cherche à valider / infirmer une hypothèse du type : $P^* \in \mathcal{P}_0 \subset \mathcal{P}$. Soient $\mathcal{P}_0, \mathcal{P}_1 \subset \mathcal{P}$, $\mathcal{P}_0 \cap \mathcal{P}_1 = \emptyset$. Les hypothèses sont :

- $H_0 : P^* \in \mathcal{P}_0$
- $H_1 : P^* \in \mathcal{P}_1$

Remarque 4.1.

En général, $\mathcal{P}_0 \cup \mathcal{P}_1 = \mathcal{P}$ quitte à se restreindre à $\mathcal{P}_0 \cup \mathcal{P}_1$.

Définition 4.1. Un test statistique est une statistique δ à valeurs dans $\{0, 1\}$ telle que :

- si $\forall z \in \mathcal{Z}, \delta(z) = 1$, on accepte H_0.
- si $\exists z \in \mathcal{Z}, \delta(z) = 0$, on rejette H_0.

Remarque 4.2.

H_0 est l'hypothèse conservative.

Définition 4.2. La région de rejet est définie par $\mathcal{R} = \delta^{-1}(\{0\})$. On peut définir un test par $(\mathcal{P}_0, \mathcal{P}_1, \mathcal{R})$.

Pour construire les régions de rejet, on utilise une statistique de test $T : \mathcal{Z} \to \mathbb{R}$. $\mathcal{R}$ est du type : $\mathcal{R} = \{z \in \mathcal{Z} : T(z) \in W\}$ où $W \subset \mathbb{R}$.

Définition 4.3.

- Risque de première espèce : rejeter H_0 alors que H_0 est vraie $\sim P^*(Z \in \mathcal{R})$ si $P^* \in \mathcal{P}_0$.
- Risque de deuxième espèce : accepter H_0 alors que H_0 est fausse $\sim P^*(Z \notin \mathcal{R})$ si $P^* \in \mathcal{P}_1$.
- Puissance du test : probabilité de rejeter H_0 quand elle est fausse $\sim 1 - P^*(Z \notin \mathcal{R})$ si $P^* \in \mathcal{P}_1$.

Remarque 4.3.

- Compromis entre risque de première espèce et puissance du test.
- P^* inconnue, on calcule des majorants / minorants.
- Généralement, on fixe niveau de risque α et on cherche à maximiser la puissance du test parmi les tests de niveau α.

Définition 4.4. On définit la taille du test par :

$$\sup_{P^* \in \mathcal{P}_0} P^*(Z \in \mathcal{R}) = \alpha$$

Un majorant de la taille du test est appelé un niveau du test.

On prend $\mathcal{R} = \{z \in \mathcal{Z} : T(z) \in W\}$. Par exemple W peut être un intervalle de $\mathbb{R}$. On fait grossir W pour augmenter la puissance du test, sans d"passer la taille α.

Autre méthode : On regarde la taille des zones de rejet qui permettent de rejetter H_0.

Définition 4.5. Soit $(\mathcal{R}_t)_{t\in T}$ une famille de zones de rejet de tailles $(\alpha_t)_{t\in T}$. La **p-valeur** du test est la statistique :

$$\alpha^* = \inf_{t\in T}\{\alpha_t : Z \in \mathcal{R}_t\}$$

En pratique :

- Si $\alpha^* < 0.01$: très forte évidence contre H_0.
- Si $\alpha^* \in [0.01, 0.05]$: forte évidence contre H_0.
- Si $\alpha^* > 0.05$: faible évidence contre H_0, voire aucune.

4.2 Tests paramétriques

Définition 4.6. $\mathcal{M}$ modèle paramétrique. Un test paramétrique est un test du type :

$$\begin{cases} H_0 : \theta^* \in \Theta_0 \subset \Theta \\ H_1 : \theta^* \in \Theta_1 \subset \Theta \end{cases}$$

avec $\Theta_0 \cap \Theta_1 = \emptyset$ et $P^* = P_{\theta^*}$.

Exemple 4.1.

Test simple : $\Theta_0 = \{\theta_0\}$ et $\Theta_1 = \{\theta_1\}$.

Définition 4.7. Fonction puissance d'un test paramétrique :

$$\pi : \begin{cases} \Theta_0 \cup \Theta_1 \to \mathbb{R}_+ \\ \theta \mapsto \pi(\theta) = P_\theta(Z \in \mathcal{R}) \end{cases}$$

Exemple 4.2.

Cas du test simple :

- $\pi(\theta_0)$: risque de première espèce.
- $\pi(\theta_1)$: puissance.
- $1 - \pi(\theta_1)$: risque de deuxième espèce.

Remarque 4.4.

La taille du test est alors $\alpha = \sup_{\theta\in\Theta_0} \pi(\theta)$.

Définition 4.8. Soient T, T' tests paramétriques.
Si $\Theta_1 = \{\theta_1\}$, on dit que T est plus puissant que T' si $\pi(\theta_1) > \pi'(\theta_1)$.
Si Θ_1 est quelconque, T est uniformément plus puissant que T' si $\forall\theta \in \Theta_1, \pi(\theta) > \pi'(\theta)$.

4.3 Test du rapport de vraisemblance

Modèle à densités, $Z = (X_i)_{1\le i\le n}$ iid. Test simple.

Définition 4.9. Le test du rapport de vraisemblance a pour statistique de test :

$$T(Z) = \frac{L(Z, \theta_1)}{L(Z, \theta_0)}$$

avec $\mathcal{R}_\alpha = \{z \in \mathcal{Z}, T(z) > c_\alpha\}$, avec c_α tel que $P_{\theta_0}(Z \in \mathcal{R}_\alpha) \le \alpha$.

Théorème 4.1 (lemme de Neymann-Pearson)**.** Le test du rapport de vraisemblance de taille α est le plus puissant parmi les tests de niveau α.

À voir : test du χ^2, de Pearson, ANOVA, de Student, d'ajustement, Wilcoxson.

Deuxième partie

Exercices

Exercice 3 PCA

$$\begin{aligned} Var((v^T x_i)_{1\leq i\leq n}) &= \frac{1}{n}\sum_{i=1}^{n}(v^T x_i)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(v^T x_i x_i^T v) \\ &= v^T\frac{1}{n}\sum_{i=1}^{n}(x_i x_i^T v) \\ &= v^T\sum_{i=1}^{n}(v) \end{aligned}$$

Soient $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$ les valeurs propres de $\hat{\Sigma}$, et $e_1, \ldots, e_p$ les vecteurs propres associés.

$$e_1^T\hat{\Sigma}e_1 = \lambda_1$$

$$v = \sum_{i=1}^{p}\alpha_i e_i$$

$$||v||^2 = 1 = \sum_{i=1}^{n}\alpha_i^2$$

$$v^T\hat{\Sigma}v = v^T\sum_{i=1}^{p}\lambda_i\alpha_i e_i$$

$$v^T\hat{\Sigma}v = \sum_{i=1}^{p}\lambda_i\alpha_i^2 \leq \lambda_1\sum\alpha_i^2 = \lambda_1$$

Exercice 4 Method of moments vs maximum likelihood estimation

1. $\mathbb{E}_\theta(X_1) = \theta/2$. On résout $\frac{\hat{\theta_{MO}}}{2} = \frac{1}{n}\sum_{i=1}^{n}X_i \implies \hat{\theta_{MO}} = \frac{2}{n}\sum_{i=1}^{n}X_i$
2. Soit $(x_1, \ldots, x_n) \in \mathbb{R}^n$

$$L(x_1, \ldots, x_n, \theta) = \prod_{i=1}^{n}\frac{1}{\theta}\mathbb{1}_{\{[0n\theta]\}}(x_i) = \frac{1}{\theta^n}\prod_{i=1}^{n}\mathbb{1}_{\{[0,\theta]\}}(x_i)$$

$$\arg\max_{\theta>0} L(x_1,\ldots,x_n,\theta) = \max(x_1,\ldots,x_n) = \hat{\theta_{MLE}}$$

3. Soit $x \in \mathbb{R}$,

$$\begin{aligned} P(\hat{\theta}_{MLE} \leq x) &= P(\cap_{i=1}^n X_i \leq x) \\ &= \prod_{i=1}^n P(X_i \leq x) \\ &= \left(\frac{x}{\theta}\right)^n \end{aligned}$$

$$P_{\hat{\theta}_{MLE}}(x) = n\frac{x^{n-1}}{\theta^n}$$
$$\hat{\theta}_{MLE} \sim \text{Beta}(n,1)$$

4.

$$\mathbb{E}_\theta(\hat{\theta}_{MLE}) = \frac{n\theta}{n+1}$$
$$b_\theta(\hat{\theta}_{MLE}) = \frac{-\theta}{n+1}$$
$$\text{Var}(\hat{\theta}_{MLE}) = \frac{\theta^2 n}{(n+1)^2(n+2)}$$

5.

$$\mathbb{E}_\theta(\hat{\theta}_{MO}) = \frac{2}{n}\sum\frac{\theta}{2} = \theta$$
$$\text{Var}_\theta(\hat{\theta}_{MO}) = \frac{4}{n^2}\sum_{i=1}^n \frac{\theta^2}{12} = \frac{\theta^2}{3n}$$

6.

$$MSE(\hat{\theta}_{MO}) = \frac{\theta^2}{3n}$$

$$\begin{aligned} MSE(\hat{\theta}_{MLE}) &= \frac{\theta^2}{(n+1)^2} + \frac{\theta^2 n}{(n+1)^2(n+2)} \\ &= \frac{\theta^2(2n+2)}{(n+1)^2(n+2)} \\ &= \frac{2\theta^2}{(n+1)(n+2)} \end{aligned}$$

Exercice 5 Maximum likelihood estimators

1. $X_i \sim \mathcal{B}(p)$

$$\begin{aligned} L(X_1,\ldots,X_n,p) &= \prod_{i=1}^n p^{X_i}(1-p)^{1-X_i} \\ &= p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum_{i=1}^n X_i} \end{aligned}$$

$$\ln(L(X_1,\ldots,X_n,p)) = \ln(p)\sum_{i=1}^n X_i + \ln(1-p)(n - \sum_{i=1}^n X_i)$$

$$\frac{\partial \ln(L(X_1,\ldots,X_n,p))}{\partial p} = \frac{\sum_{i=1}^n X_i}{p} - \frac{n-\sum_{i=1}^n X_i}{1-p}$$

$$\begin{aligned}\frac{\partial \ln(L(X_1,\ldots,X_n,p))}{\partial p} = 0 &\Leftrightarrow \frac{\sum_{i=1}^n X_i}{p} = \frac{n-\sum_{i=1}^n X_i}{1-p}\\ &\Leftrightarrow \sum X_i = n\hat{p}\\ &\Leftrightarrow \boxed{\hat{p} = \tfrac{1}{n}\sum X_i}\end{aligned}$$

De plus par le TCL on a :

$$\sqrt{n}(\hat{p}_n - p) \xrightarrow{loi} \mathcal{N}(0, p(1-p))$$

2. $X_i \sim \mathcal{N}(\mu, \sigma^2)$

$$\begin{aligned}L(X_1,\ldots,X_n,\mu,\sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(X_i-\mu)^2)\\ &= \frac{1}{(\sqrt{2\pi\sigma^2})^n}\exp(-\frac{1}{2\sigma^2}(\sum_{i=1}^n (X_i-\mu)^2))\end{aligned}$$

On passe au log :

$$\ln(L((X_1,\ldots,X_n,\mu,\sigma^2))) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu)^2$$

$$\frac{\partial \ln(L(X_1,\ldots,X_n,\mu,\sigma^2))}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i-\mu)$$

$$\frac{\partial \ln(L(X_1,\ldots,X_n,\mu,\sigma^2))}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^n (X_i-\mu)^2$$

En annulant ces deux dérivées partielles, on obtient :

$$\begin{cases}\hat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i\\ \hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (X_i-\hat{\mu})^2\end{cases}$$

Exercice 6 Linear regression

1.
$$||Y - X\beta||^2 = ||Y||^2 + \beta^T X^T X\beta - 2Y^T X\beta$$
$$\nabla||Y - X\beta||^2 = 2\beta^T X^T X - 2Y^T X = 0$$
$$\Leftrightarrow \beta^T X^T X = Y^T X$$
$$\Leftrightarrow X^T X\beta = X^T Y$$
$$\Leftrightarrow \beta = (X^T X)^{-1} X^T Y$$

2.
$$\mathbb{E}(\hat{\beta}) = (X^T X)^{-1} X^T \mathbb{E}(X\beta + \epsilon) = (X^T X)^{-1} X^T X\beta = \beta$$

Donc $b(\hat{\beta}) = 0$.

$$\mathrm{Var}(\hat{\beta}) = V((X^T X)^{-1} X^T (X\beta + \epsilon)) = (X^T X)^{-1} X^T \mathrm{Var}(\epsilon) X (X^T X)^{-1^T} = \sigma^2 (X^T X)^{-1^T}$$

Exercice 7 Bayesian estimation

Soient $(x_1, \ldots, x_n) \in \{0,1\}^n$.

$$\begin{aligned} p(\theta|x) &\propto p(x|\theta)\pi(\theta) \\ &= \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\mathbb{1}_{[0,1]}(\theta) \\ &\sim \mathrm{Beta}(\sum x_i + 1, n + 1 \sum x_i) \end{aligned}$$