Fiche résumée du cours de Statistiques, par G.Biau

1 Concepts fondamentaux de la statistique

Définition 1.0.1 (Modèle statistique) *Un **modèle statistique** est la donnée d'un espace mesuré $(E, \mathscr{E})$ et d'une famille $\{P_\theta,\ \theta \in \Theta\}$ de lois de probabilités. Le modèle associé est noté $(E, \mathscr{E}, P_\theta, \theta \in \Theta)$. Quand il existe un entier d tel que $\Theta \subset \mathbb{R}^d$, le modèle est dit paramétrique, et non paramétrique sinon.*

Définition 1.0.2 (Observation) *Une **observation** $\mathbb{X}$ est une variable aléatoire à valeurs dans E et dont la loi appartient à la famille $(P_\theta)_{\theta \in \Theta}$.*

Définition 1.0.3 (Echantillon) *Pour $n \in \mathbb{N}^*$, un n-échantillon sdde loi ν est la donnée de n variables aléatoires $X_1, ...X_n$ indépendantes et identiquement distribuées de loi ν.*

2 Estimation

Définition 2.0.4 (Estimateur) *Soit $g : \Theta \to \mathbb{R}^p$ une fonction de θ à estimer. Un estimateur $\widehat{g}$ de $g(\theta)$ est une application mesurable de l'observation $\mathbb{X}$ et indépendante de θ. Tout estimateur est donc de la forme $\widehat{g} = h(\mathbb{X})$, où h est une application mesurable $E \to \mathbb{R}^p$.*

2.1 construction de l'estimateur

Définition 2.1.1 (Estimateur consistant, fortement consistant) *Soit $\mathbb{X} = (X_1, ..., X_n)$ un n-échantillon, et $\widehat{g}_n = h_n(X_1, ..., X_n)$ un estimateur de $g(\theta)$.Cet estimateur est dit :*

*i) **Consistant** si, pour tout θ dans Θ, $\widehat{g}_n$ converge en $\mathbb{P}_\theta$-probabilités vers $g(\theta)$ lorsque $n \to \infty$.*

*ii) **Fortement consistant** si , pour tout θ dans Θ, $\widehat{g}_n$ converge $\mathbb{P}_\theta$-p.s. vers $g(\theta)$ lorsque $n \to \infty$.*

Par abus de language, on parle toujours de la consistance de l'estimateur là où l'on devrait parler de la consistance de la suite d'estimateurs.

2.2 Construction d'estimateurs

Methode des moments :
Cette méthode est utile lorsque les paramètres d'intérêt s'expriment en fonction des moments de la loi des observations : par exemple dans les cas où ces moments existent, on approchera $\mu_1 = \mathbb{E}_\theta(X_1)$ et $\mu_2 = \mathbb{E}_\theta(X_1^2)$ respectivement par

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \text{ , et } \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i^2 \text{ .}$$

Plus généralement, étant données q fonctions $\phi_1, ...\phi_q$, en posant $g_i(\theta) = \mathbb{E}_\theta(\phi_i(1))$, on estime $g(\theta) = \psi(g_1(\theta), ..., g_q(\theta))$ par

$$\widehat{g}_n = \psi(\sum_{i=1}^n \phi_1(X_i), ..., \sum_{i=1}^n \phi_q(X_i)) \text{ .}$$

Par exemple, la variance $\sigma^2 = \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2$, va être estimée par

$$\widehat{S}_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2 \text{ .}$$

Methode du maximum de vraisemblance :
Cette méthode ne s'applique que lorsque toutes les lois du modèle admettent une densité par rapport à une mesure commune, typiquement la mesure

de Lebesgue. Dans ce cas, on note f_θ cette densité. La méthode consiste à approcher $g(\theta) = \theta$ par l'élément de θ maximisant la **vraisemblance**

$$\begin{array}{rccc} V(\mathbb{X}) & : & \Theta & \rightarrow & \mathbb{R} \\ & & \theta & \mapsto & f_\theta(\mathbb{X}) = \prod f_\theta(X_i) \end{array}$$

2.3 Normalité asymptotique

Définition 2.3.1 (Normalité asymptotique) *Soit $\mathbb{X} = (X_1, ..., X_n)$, et $\widehat{\theta}$ un estimateur de θ. Il est dit **asymptotiquement normal** si les trois conditions suivantes sont vérifiées :*

i) La vitesse de convergence de l'estimateur est en $\sqrt{n}$.

ii) La convergence a lieu en loi.

iii) La loi limite est normale.

Théorème 2.3.1 (Methode Δ) *On se donne une suite $(U_n)_{n\in\mathbb{N}}$ de vecteurs aléatoires de $\mathbb{R}^m$, une suite déterministe $(a_n)_{n\in\mathbb{N}}$ et une application $l : \mathbb{R}^m \to \mathbb{R}^p$ tels que :*

i) $a_n \to \infty$.

ii) $\exists U \in \mathbb{R}^m$, un vecteur déterministe et un vecteur aléatoire V tel que

$$a_n(U_n - U) \xrightarrow{\mathscr{L}} V \ .$$

iii) l est une fonction différentiable en U de différentielle notée $D_l(U) \in M_{p,n}(\mathbb{R})$.

Alors, on a

$$a_n(l(U_n - l(U)) \xrightarrow{\mathscr{L}} D_l(U)V \ .$$

Lemme 2.3.1 (Slutsky) *Soient $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de vecteurs aléatoires prenant leurs valeurs respectivement dans $\mathbb{R}^d$ et dans $\mathbb{R}^p$, où p et d sont dans $\mathbb{N}^*$. Si $X_n \to X$, et $Y_n \xrightarrow{\mathbb{P}} C$ pour une certaine variable aléatoire X et un vecteur déterministe C, alors*

$$(X_n, Y_n) \underset{\mathscr{L}}{\rightarrow} (X, C) \ .$$

Corollaire 2.3.1 *Soit $(X_1, ...X_n)$ un n-échantillon de vecteurs de $\mathbb{R}^n$, de loi commune admettant un moment d'ordre 2, d'espérance μ et de matrice de variance covariance Σ. Si l est une fonction différentiable en μ, de différentielle notée $D_l(\mu)$, alors*

$$\sqrt{n}(l(\overline{X}_n) - l(\mu)) \xrightarrow{\mathscr{L}} \mathscr{N}(0, D_l(\mu)\Sigma D_l(\mu)) \ .$$

2.4 Estimation sans biais

Définition 2.4.1 (biais) *Soit $\mathbb{X}$ une observation et $\widehat{g}(\theta) = h(\mathbb{X})$ un estimateur de $g(\theta)$. vérifiant $\mathbb{E}_\theta(\|\widehat{g}\|) < \infty$ pour tout θ.On appelle **biais** la fonction*

$$b : \theta \mapsto \mathbb{E}_\theta(\widehat{g}) - g(\theta) \ .$$

$\widehat{g}$ est dit sans biais lorsque son biais est identiquement nul.

Définition 2.4.2 (biais asympotique) *Soit $\mathbb{X} = (X_1, ..., X_n)$ une observation et $\widehat{g}_n(\theta) = h(\mathbb{X})$ un estimateur de $g(\theta)$. vérifiant $\mathbb{E}_\theta(\|\widehat{g}_n\|) < \infty$ pour tout n, θ. On appelle **biais** la suite de fonctions*

$$b_n : \theta \mapsto \mathbb{E}_\theta(\widehat{g}_n) - g(\theta) \ .$$

$\widehat{g}$ est dit asymptotiquement sans biais lorsque la limite $n \to \infty$ de son biais est identiquement nul.

2.5 Estimation optimale

Définition 2.5.1 (Risque quadratique d'un estimateur) *A tout estimateur $\widehat{g}$ de $g(\theta)$, on associe son risque (quadratique) défini comme l'application*

$$\begin{array}{rccc} R & : & \Theta & \rightarrow & [0, +\infty] \\ & & \theta & \mapsto & \mathbb{E}_\theta((\widehat{g} - g(\theta))^2) \end{array} .$$

Un estimateur est d'autant meilleur que son rique est faible sur Θ. On a la relation fondamentale suivante :

$$R(\theta) = b(\theta)^2 + var(\widehat{g})$$

2.6 Bornes de Cramér-Rao

On suppose $\Theta =]a, b[\subset \mathbb{R}$, et on suppose le modèle dominé par μ. On note f_θ les densités respectives, et on se place sous les hypothèses de régularité suivantes : pour $\mu-$presque tout x, $\theta \mapsto f_\theta(x)$ est continue et continument différentiable sauf éventuellement en un nombre fini de points, et la quantité $f'_\theta(x)^2/f_\theta(x)\mathbb{1}_{ftheta(x)>0}$ est intégrable en μ d'intégrale

$$I(\theta) = \int_{ftheta(x)>0} \frac{f'_\theta(x)^2}{f_\theta(x)} d\mu(x) \text{ continue.}$$

Cette intégrale est apellée **information de Fisher** du modèle.

Proposition 2.6.1 *Dans ces conditions, considérons une statistique $T(\mathbb{X})$ telle que $\mathbb{E}_\theta(T^2(\mathbb{X}))$ est bornée sur Θ.Alors, l'application*

$$\theta \mapsto \mathbb{E}_\theta(T(\mathbb{X}))$$

est continument différentiable sur Θ, de dérivée

$$\int_{f_\theta(x)>0} T(x) f'_\theta(x) d\mu(x) = \mathbb{E}_\theta\left(T(\mathbb{X})\frac{\partial}{\partial\theta}\log(f_\theta(\mathbb{X}))\right) .$$

Corollaire 2.6.1 *Si la famille (P_θ) est régulière, alors*

$$\int_{f_\theta(x)>0} f'\theta(x) d\mu(x) = 0 .$$

Cela implique en particulier que

$$I(\theta) = var_\theta\left(\frac{\partial}{\partial\theta}\log f_\theta(\mathbb{X})\right) .$$

Théorème 2.6.1 *Soit $\widehat{g} = h(\mathbb{X})$ un estimateur sans biais de $g(\theta)$, où g est une fonction dérivable. Si le modèle est régulier, sur $\Theta =]a, b[\subset \mathbb{R}$ d'information strictement positive, et si $\mathbb{E}_\theta(h^2(\mathbb{X}))$ est bornée, on a $\forall\theta \in \Theta$*

$$var_\theta(\widehat{g}) \geq \frac{(g'(\theta)^2)}{I(\theta)} .$$

Le membre de droite est apellé borne de Cramér-Rao, et donne une minoration du risque de l'estimateur $\widehat{g}$.

Corollaire 2.6.2 *Supposons que l'on observe un n-échantillon $\mathbb{X} = (X_1, ..., X_n)$, et que la loi commune admette I_1 comme information de Fisher. Alors, sous les notations et hypothèses du théorème précédent , pour tout $\theta \in \Theta$, on a*

$$I(\theta) = nI_1(\theta), \text{ et } var_\theta(h(X_1, ..., X_n)) \geq \frac{(g'(\theta))^2}{nI_1(\theta)} .$$

3 Intervalles et régions de confiance

Définition 3.0.1 (Région de confiance) *Pour tout $\alpha \in]0, 1[$, une **région de confiance** de $g(\theta)$ de niveau $1-\alpha$ est un ensemble $\widehat{C}$ construit mesurablement par rapport à $\mathbb{X}$ et tel que, pour tout $\theta \in \Theta$, on ait*

$$\mathbb{P}_\theta(g(\theta) \in \widehat{C}) \geq 1 - \alpha .$$

Lorsque cette inégalité est une égalité, on dit que le niveau de confiance est exactement égal à $1-\alpha$.

Définition 3.0.2 (Région de confiance asymptotique) *Pour tout $\alpha \in]0, 1[$, une **région de confiance asymptotique** de $g(\theta)$ de niveau $1-\alpha$ est une suite d'ensembles $\left(\widehat{C}_n\right)_{n\in\mathbb{N}}$ chacun construit mesurablement par rapport à $\mathbb{X} = (X_1, ..., X_n)$ et tel que, pour tout $\theta \in \Theta$, on ait*

$$\liminf_{n\to\infty} \mathbb{P}_\theta(g(\theta) \in \widehat{C}_n) \geq 1 - \alpha .$$

3.1 Premières constructions

Définition 3.1.1 (quantile) *Pout tout $\beta \in [0, 1]$, on apelle quantile d'ordre β d'une loi de proba P à support inclus dans $\mathbb{R}$ la quantité*

$$q_\beta = \inf\{x \in \mathbb{R}, P(]-\infty, x]) \geq \beta\} .$$

3.2 Intervalles de confiance exacts obtenus par des inégalités de probabilités

3.2.1 Cas des variances uniformément bornées

Proposition 3.2.1 (Inégalité de Bienaymé-Tchebitchev) *Soit Y une variable aléatoire admettant un moment d'ordre 2. Alors, pour tout $\varepsilon > 0$,*

$$\mathbb{P}(|Y - \mathbb{E}(Y)| > \varepsilon \leq \frac{varY}{\varepsilon^2} \ .$$

3.2.2 Cas des lois à supports tous inclus dans un compact borné.

Lemme 3.2.1 (Inégalité de Hoeffding) *Soient $Y_1, ..., Y_n$ des variables aléatoires réelles indépendantes et $(\alpha_1, ..., \alpha_n)$ et $(\beta_1, ..., \beta_n)$ deux n-uplets de réels tels que, pour tout i, on ait*

$$\mathbb{E}(Y_i) = 0 \text{ et } \alpha_i \leq Y_i \leq \beta_i \text{ p.s.}$$

Alors, pour tout $t > 0$,

$$\mathbb{P}\left(\sum_{i=1}^{n} Y_i \geq t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(\beta_i - \alpha_i)^2}\right) \ .$$

3.3 Intervalle de confiance asymptotique

4 Test d'hypothèses

4.1 Formalisme et démarche expérimentale

Dans le cadre du modèle statistique $(E, \mathscr{E}, P_\theta, \theta \in \Theta)$. On se donne deux sous ensembles disjoints Θ_0 et Θ_1 inclus dans Θ, et on souhaite déterminer si θ est dans Θ_0 ou Θ_1. On définit donc les deux hypothèses, l'hypothèse nulle $H_0 \ : \ \theta \in \Theta_0$, et l'hypothèse alternative $H_1 \ : \ \theta \in \Theta_1$. L'hypothèse est dite simple si l'ensemble correspondant est un singleton, et composite sinon.

Définition 4.1.1 (Test) *On appelle **test** de l'hypothèse H_0 contre H_1 toutee fonction $\phi(\mathbb{X})$, à valeurs dans $\{0, 1\}$, où ϕ est mesurable, et peut dépendre de Θ_0 et Θ_1. On accepte H_i si $\phi(\mathbb{X}) = i$. On peut toujours écrire $\phi(\mathbb{X}) = \mathbb{1}_{h(\mathbb{X}) \in R}$, où R est appellé région de rejet.*

4.2 Mesure de la qualité d'un test

Définition 4.2.1 (Risques de première et seconde espèce) *Les risques de première espèce et de seconde espèce du test $\phi(\mathbb{X})$ sont définies respectivement comme les fonctions $\underline{\alpha}$ sur Θ_0 et $\underline{\beta}$ sur Θ_1 par*

$$\begin{array}{cccc} \underline{\alpha} & : & \Theta_0 & \to & [0,1] \\ & & \theta & \mapsto & \mathbb{P}_\theta(\phi(\mathbb{X}) = 1) \end{array} \quad \text{et} \quad \begin{array}{cccc} \underline{\beta} & : & \theta_1 & \to & [0,1] \\ & & \theta & \mapsto & \mathbb{P}(\phi(\mathbb{X}) = 0) \end{array} \ .$$

On appelle puissance d'un test la fonction $(1 - \underline{\beta})$ la probabilité de rejeter à raison. Sa taille est le réel

$$\alpha^* = \sup_{\theta \in \Theta_0} \underline{\alpha} = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\phi(\mathbb{X}) = 1) \ .$$

On dit enfin que le test est de niveau α si sa taille α^* est inférieure à α.

4.3 propriétés éventuelles d'un test

Un test est dit sans biais quand sa fonction de puissance vérifie $1 - \underline{\beta} > \alpha$ sur Θ_1. Une suite de tests est consistante s'ils sont tous de niveau α et si, pour tout $\theta \in \Theta_1$, $\underline{\beta}_n \to 0$ quand $n \to \infty$. La robustesse exprime l'absence de sensibilité d'un test éventuellement asymptotique au modèle. (i.e à la loi ou la forme de l'observation).

Définition 4.3.1 *Pour $\phi(\mathbb{X})$ et $\phi'(\mathbb{X})$ deux tests de niveau α, on dit que $\phi(\mathbb{X})$ est uniformément plus puissant que $\phi'(\mathbb{X})$ si les fonctions puissance vérifient*

$$1 - \underline{\beta}' \geq 1 - \underline{\beta} \ , i.e. \forall \theta \in \Theta_1, \ \mathbb{P}_\theta(\phi(\mathbb{X}) = 1) \geq \mathbb{P}_\theta(\phi'(\mathbb{X}) = 1)$$

4.4 Un outil : la p-valeur

Définition 4.4.1 (p-valeur) *Supposons avoir construit une famille de tests $\phi_\alpha(\mathbb{X})$ de niveau $\alpha \in]0,1[$. La p-valeur associée à cette famille et à l'observation $\mathbb{X}$ est le réel défini par $\widehat{\alpha}(\mathbb{X}) = \sup_{\alpha \in]0,1[}\{\phi_\alpha(\mathbb{X}) = 0\}$. En d'autres termes, c'est le plus grand niveau autorisant l'acceptation de H_0, et forme donc, en un certain sens, l'indice de crédibilité du test.*

4.5 Botanique des tests

Théorème 4.5.1 (lemme de Neymann-Pearson) *On fixe θ_0 et θ_1 deux points distincts de Θ, et on s'intéresse au test de $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$. Pour $\alpha \in]0,1[$ fixé, donné, lorsqu'il existe un seuil k_α tel que le test du rapport de vraisemblance*

$$\phi(\mathbb{X}) = \mathbb{1}_{h(\mathbb{X},\theta_0,\theta_1) > k_\alpha},\ \text{où } h(\mathbb{X},\theta_0,\theta_1) = \frac{V_{\mathbb{X}}(\theta_1)}{V_{\mathbb{X}}(\theta_0)} = \frac{f_{\theta_1}(\mathbb{X})}{f_{\theta_0}(\mathbb{X})}$$

vérifie $\mathbb{P}(\phi(\mathbb{X}) = 1) = \alpha$, alors ce test est upp($\alpha$).

5 Vecteurs gaussiens

5.1 Propriétés des vecteurs gaussiens et théorème de la limite centrale

5.1.1 Loi normale multidimensionnelle

Définition 5.1.1 (Vecteur gaussien) *Une variable aléatoire réelle Z est dite gaussienne centrée réduite, ou de loi normale standard, noté $Z \sim \mathcal{N}(0,1)$ si sa loi admet pour densité sur $\mathbb{R}$*

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Une variable aléatoire réelle X est dite gaussienne, ou de loi normale de paramètres $\mu \in \mathbb{R}$ et $\sigma^2 \in \mathbb{R}_+$, ce que l'on note $X \sim \mathcal{N}(\mu,\sigma^2)$ s'il existe une variable aléatoire Z normale standard, et tel que $X = \mu + \sigma Z$. Lorsque $\sigma = 0$, on dit que X est dégénéré.

Théorème 5.1.1 *La fonction caractéristique de la loi $\mathcal{N}(\mu,\sigma^2)$ vaut*

$$\phi_{\mu,\sigma^2}(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}\ \forall t \in \mathbb{R}.$$

5.1.2 Loi normale multidimensionnelle

Définition 5.1.2 (Loi normale multidimensionnelle) *Un vecteur aléatoire X à valeurs dans $\mathbb{R}^d$ est dit gaussien si toute combinaison linéaire de ses composantes suit une loi normale. Si $X = (X_1, ..., X_n)$ est un vecteur gaussien, on définit son vecteur moyenne par $\mathbb{E}(X) = (\mathbb{E}(X_1), ..., \mathbb{E}(X_2))$, et sa matrice de variance covariance par*

$$var(X) = (X - \mathbb{E}(X))(X - \mathbb{E}(X))^T = (cov(X_i, X_j))_{i,j}.$$

5.1.3 Propriétés des vecteurs gaussiens

Théorème 5.1.2 *Soit $X = (X_1, ..., X_d)^T$ un vecteur gaussien d'espérance M et de matrice de variance covariance Σ. Alors, X admet pour fonction caractéristique*

$$\varphi_X(t) = \mathbb{E}\left(e^{it^T X}\right) = \exp\left(it^T M - \frac{t^T \Sigma t}{2}\right),\ \forall t \in \mathbb{R}^d.$$

Proposition 5.1.1 *Soit $X = (X_1, ..., X_d)^T$ un vecteur gaussien d'espérance M et de matrice de variance covariance Σ. Alors, pour toute matrice A $d' \times d$ et tout vecteur b dans $\mathbb{R}^{d'}$, on a $AX+b$ est un vecteur gaussien de moyenne $AM+b$ et de matice de variance covariance $A\Sigma A^T$. Pour tout couple i, j, on a de plus*

$$X_i \amalg X_j \Leftrightarrow cov(X_i, X_j) = 0 \ .$$

Enfin, X admet une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ si et seulement si Σ est inversible, i.e. définie positive. Sa densité est alors donnée par

$$f(x) = \frac{1}{\sqrt{2\pi}^d \sqrt{det(\Sigma)}} \exp\left(-\frac{1}{2}(x-m)^T \Sigma^{-1}(x-m)\right) \ .$$

Si elle n'est pas inversible, alors la loi de $X - m$ est p.s. portée par l'espace engendré par les vecteurs propres pour des valeurs propres non nulles de Σ.

Proposition 5.1.2 (Espérance conditionnelle) *Quand $(Y, X_1, ..., X_d)^T$ est un vecteur gaussien, $\mathbb{E}(Y \mid X_1, ..., X_d)$ est une fonction affine des composantes de $(X_1, ..., X_d)^T$.*

5.1.4 version multidimensionnelle du TCL

Théorème 5.1.3 *Soit $(V_1, ..., V_n)$ un n-échantillon de vecteurs de $\mathbb{R}^d$ admettant un moment d'ordre 2, dont on note m l'espérance et Σ la variance-covariance commune, alors*

$$\sqrt{n}(\overline{V}_n - m) \to \mathcal{N}(0, \Sigma) \text{ en loi.}$$

5.2 Loi du χ^2 et de Student

Définition 5.2.1 (Loi du χ^2) *Soit $X = (X_1, ..., X_d)^T$ un vecteur gaussien d'espérance m et de matrice de variance covariance $\Sigma = Id$. Alors, la loi de la variablealatoire $\|X\|_2^2 = X_1^2, ..., X_d^2$ ne dépend que de d et de la norme 2 de m. On dit que cette loi est une loi du χ^2 à d degrés de liberté, et de paramètre de décentrage $\|m\|_2^2$, ce que l'on note*

$$\|X\|_2^2 \sim \chi^2(d, \|m\|_2^2) \ .$$

Cette loi est dite décentrée lorsque le paramètre de décentrage est nul, on ne note alors pas le second paramètre.

Définition 5.2.2 (Loi de Student) *Lorsque X et Y sont deux variable aléatoire indépendantes, telles que $X \sim \mathcal{N}(\mu, 1)$, et $Y \sim \chi^2(d)$. Alors, la loi de la variable aléatoire $Z = X/\sqrt{Y/d}$ est appellée **loi de Student** à d degrés de liberté, notée $\mathcal{T}(d, \mu)$. Le nombre réel μ est le paramètre de décentrage, et on l'omet lorsqu'il est nul.*

5.3 Conséquences statistiques : intervalles de confiance et tests

5.3.1 Un outil fondamental : le théorème de Cochran

Théorème 5.3.1 (Cochran) *On considère un vecteur gaussien X de loi $\mathcal{N}(m, Id)$, et une décomposition $E_1 \oplus ... \oplus E_r$ de $\mathbb{R}^d$ en sous espaces orthogonaux de dimensions respectives $d_1, ..., d_r$. Alors, les projections orthogonales $\Pi_{E_1}X, ..., \Pi_{E_r}X$ forment des vecteurs gaussiens indépendants. Par ailleurs, pour $j \in \{1, ..., r\}$,*

$$\|\Pi_{E_j}X\|_2^2 \sim \chi^2\left(d_j, \|\Pi_{E_j}m\|_2^2\right) \ .$$

5.3.2 Intervalles de confiance et tests sur les paramètres d'une loi normale

Proposition 5.3.1 *Soit $(X_1, ..., X_n)$ un n-échantillon de loi $\mathcal{N}(\mu, \sigma^2)$. Alors, $\overline{X}_n$ et $\widehat{S}_n^2$ sont indépendants, et*

$$\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \text{ et } \frac{n\widehat{S}_n^2}{\sigma^2} \sim \chi^2(n-1) \ .$$

Enfin,

$$\sqrt{n}\frac{(\overline{X}_n-\mu)}{\sqrt{\widehat{S}_n^2}} \sim \mathcal{T}(n-1).$$

6 Le modèle linéaire gaussien

6.1 Definitions

Définition 6.1.1 (Modèle linéaire gaussien) *Soit Y un vecteur d'observations de $\mathbb{R}^n$ On dit que Y suit un modèle linéaire gaussien lorsque*

$$(*)\ Y = m+\varepsilon, \quad \text{où } \varepsilon \sim \mathcal{N}(0,\sigma^2 I_n)\ ,$$

pour deux paramètres inconnus $m \in \mathbb{R}^n$, et $\sigma^2 \in \mathbb{R}_+^*$. On fait l'hypothèse que m appartient à un sous-espace vectoriel V de $\mathbb{R}^n$. Le vecteur ε est appellé le "vecteurs des erreurs". On adapte souvent l'écriture matricielle pour reformuler un modèle linéaire gaussien, selon

$$Y = X\beta+\varepsilon,\ \varepsilon \sim \mathcal{N}(0,\sigma^2 I_n)\ ,$$

pour les paramètres inconnus $\beta \in \mathbb{R}^p$ et σ^2. La matrice X est connue. On notera Π_V le projecteur orthogonale sur le sous-espace V.

6.2 Estimateurs des paramètres du modèle

Théorème 6.2.1 *On se place dans le modèle $(*)$, où les paramètres sont $\sigma^2>0$, et $m\in V$ pour un sous-espace vectorielV connu de $\mathbb{R}^n$ de dimension $p \geq 1$. Alors :*

i) L'estimateur du maximum de vraisemblance de $\theta=(m,\sigma^2)$ est $\widehat{\theta}_n = (\widehat{m},\widehat{S}_n^2)$ où $\widehat{m}=\Pi_V Y$, et $\widehat{S}_n^2 = 1/n\|Y-\Pi_V Y\|_2^2$.

ii) Les estimateurs $\widehat{m}$ et $\widehat{S}_n^2$ sont indépendants de lois respectives

$$\widehat{m} \sim \mathcal{N}(m,\sigma^2\Pi_V), \text{ et } \frac{n\widehat{S}_n^2}{\sigma^2} \sim \chi(n-p)\ .$$

Théorème 6.2.2 *On se place dans le modèle $(**)$, où les paramètres sont $\sigma^2>0$, et $\beta\in\mathbb{R}^p$ Alors :*

i) L'estimateur du maximum de vraisemblance de $\theta=(\beta,\sigma^2)$ est $\widehat{\theta}_n = (\widehat{\beta}_n,\widehat{S}_n^2)$ où $\widehat{\beta}_n=(X^TX)^{-1}X^TY$, et $\widehat{S}_n^2 = 1/n\|Y-X\widehat{\beta}_n\|_2^2$.

ii) Les estimateurs $\widehat{\beta}_n$ et $\widehat{S}_n^2$ sont indépendants de lois respectives

$$\widehat{\beta}_n \sim \mathcal{N}(m,\sigma^2(X^TX)^{-1}), \text{ et } \frac{n\widehat{S}_n^2}{\sigma^2} \sim \chi(n-p)\ .$$

6.3 Regions de cofiance et tests fondamentaux

6.3.1 Tests et intervalles de confiance por la variance

6.3.2 Test de Student d'une relation affine

On se place dans le modèle (**).On souhaite tester, étant donné $C\in\mathbb{R}^p$ et $a\in\mathbb{R}$, l'hypothèse $H_0 : C^T\beta = a$ contre $H_1 : C^T\beta \neq a$. La statistique de test naturelle est donnée par

$$T_n = \frac{C^T\beta_n - a}{\widehat{\sigma}_n\sqrt{C^T(X^TX)^{-1}C}}\ .$$

Le test $\phi(Y) = \mathbb{1}_{|T_n|>t_{1-\alpha/2}(n-p}$ est alors de taille α.

6.3.3 Test de Fisher d'un sous modèle

Définition 6.3.1 *Le loi de fisher à k et l degrés de liberté, notée $F(k,l)$ est la loi du quotient de deux χ^2 à k et l degrés de liberté.*

On se donne W un sous-espace de V, et on souhaite tester $H_0 : m\in W$ contre $H_1 : m\in V-W$.

Théorème 6.3.1 *lorsque $m \in W$, alors la statistique*

$$\frac{(\|Y - \Pi_W Y\|_2^2 - \|Y - \Pi_V Y\|_2^2)/(p-q)}{\|Y - \Pi_V(Y)\|_2^2/(n-p)}$$

suit une loi de fisher à $p-q$ et $n-p$ degrés de liberté.

6.3.4 Test de Wall de plusieurs relations affines

On se place dans le modèle (**). On souhaite tester, étant donné une matrice de transition C et $a \in \mathbb{R}^n$, l'hypothèse $H_0 : C\beta = a$ contre $H_1 : C\beta \neq a$.

Théorème 6.3.2 *Sous H_0 la statistique*

$$W = \frac{[C\widehat{\beta}_n - a]^T (C(X^T X)^{-1} C)(C\widehat{\beta}_n - a)}{\|Y - X\widehat{\beta}_n\|_2^2/(n-p)} \sim \mathscr{F}(k - np) .$$

6.3.5 Application à la regréssion linéaire