

Chapitre 2 : Vraisemblance et Estimation ponctuelle

Organisation du chapitre 2

1 Vraisemblance et Estimation

- Vraisemblance, exhaustivité, liberté
- La famille exponentielle
- Comment Construire un Estimateur ?
- Information de Fisher
- Comportement asymptotique des estimateurs

2 Intervalle de Confiance

- Intervalle de Confiance : définition
- Construction pratique d'un Intervalle de Confiance
- Intervalle de Confiance Asymptotique

1. Vraisemblance et Estimation

1.1. Vraisemblance, exhaustivité, liberté

Vraisemblance d'un échantillon

- **Définition 2.1** : Soit $s_n = (x_1, \dots, x_n)$ réalisation d'un échantillon (i.i.d) de loi de densité $f(x; \theta)$ appartenant au modèle $\mathcal{M} = \{f(x; \theta) | \theta \in \Theta\}$. La vraisemblance de l'échantillon s_n est la fonction $L(\theta, x_1, \dots, x_n)$:

$$\begin{aligned} L : \Theta &\longrightarrow \mathbb{R}^+ \\ \theta &\longmapsto f(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) \end{aligned}$$

- La log-vraisemblance est

$$\begin{aligned} \mathcal{L}(\theta, x_1, \dots, x_n) &= \log L(\theta, x_1, \dots, x_n) \\ &= \sum_{i=1}^n \log(f(x_i; \theta)) \end{aligned}$$

- Exemples :

- ▶ Echantillon issue d'une loi normale.

- ▶ Echantillon issue d'une loi de Weibull

$\mathcal{M} = \left\{ f(x; \lambda, \alpha) = \frac{\alpha}{\lambda} \left(\frac{x}{\lambda}\right)^{\alpha-1} \exp\left(-\left(\frac{x}{\lambda}\right)^\alpha\right) 1_{x \geq 0} \mid \alpha, \lambda > 0 \right\}$ (α paramètre de forme, λ paramètre d'échelle).

Exhaustivité

- **Définition 2.2** : Soit $\mathcal{M} = \{f(x, \theta), x \in \mathcal{X}, \theta \in \Theta\}$ un modèle paramétrique et $S : \mathcal{X}^n \rightarrow \mathcal{Y}$ une statistique. La statistique $S(x_1, \dots, x_n)$ est dite **exhaustive** si la loi de X sachant $S = s$ est indépendante de θ .
- Interprétation : $S(x_1, \dots, x_n)$ contient toute l'information de l'échantillon sur θ , c'est un **résumé** qui préserve l'information.
- **Théorème de factorisation** : Une statistique $S(x_1, \dots, x_n)$ est exhaustive du modèle paramétrique si et seulement si la densité s'écrit

$$f(x_1, \dots, x_n; \theta) = g_\theta(S(x_1, \dots, x_n); \theta) h(x)$$

- Propriété : Si S_1, S_2 sont des statistiques telles que $S_1 = h(S_2)$. Alors si S_1 est exhaustive pour θ , alors S_2 est exhaustive pour θ .
- **Définition 2.3** : Une statistique S est **exhaustive minimale** si tout autre statistique exhaustive T s'écrit $S = h(T)$.
- **Définition 2.4** : Dans un modèle paramétrique \mathcal{M} , une statistique S est **libre** en θ si sa loi ne dépend pas de θ .

Exemples

- L'échantillon lui-même
- Le cas Bernoulli $\mathcal{B}(1, p)$
- Le cas gaussien $N(m, \Sigma)$
- Le cas uniforme $U([a, b])$

1.2. La famille exponentielle

La famille exponentielle

- **Définition 2.5** : Un modèle statistique $\{f(x; \theta) | \theta \in \Theta\}$ sur $\mathcal{X} = \mathbb{R}^d$ est une **famille exponentielle** si les densités $f(x; \theta)$ s'écrivent

$$f(x; \theta) = \exp \left(\sum_{i=1}^s \eta_i(\theta) T_i(x) - B(\theta) \right) h(x)$$

avec $h, B, \eta_i, T_i, i = 1, \dots, s$ à valeurs réels, et $h(x) \geq 0$.

- $T(x) = (T_1(x) \dots T_s(x)) \in \mathbb{R}^s$ est la statistique privilégiée du modèle (les T_i sont linéairement libres = de rang plein)
- Il est plus commode d'utiliser directement les η_i comme paramètre : la densité s'écrit alors sous **forme canonique**

$$f(x; \eta) = \exp \left(\sum_{i=1}^s \eta_i T_i(x) - A(\eta) \right) h(x)$$

et $\eta = (\eta_1, \dots, \eta_s)$ est appelé **paramètre naturel**.

Propriétés

- $C(\eta) = \exp(-A(\eta))$ est la constante de normalisation.
- L'ensemble $H = \{\eta \in \mathbb{R}^s \mid \int_{\mathcal{X}} \exp(\sum_{i=1}^s \eta_i T_i(x)) < \infty\}$ est un convexe.
- Pour une famille exponentielle, le modèle d'échantillonnage est encore une famille exponentielle avec la statistique privilégiée :

$$T(x_1, \dots, x_n) = \sum_{i=1}^n T(x_i).$$

- On peut écrire $\sum_{i=1}^s \eta_i T_i(x)$ sous la forme d'un produit scalaire adapté : $\langle \eta, T(x) \rangle_H$ dans un bon espace de paramètre.
- La fonction $\eta \mapsto A(\eta)$ est indéfiniment dérivable et
 - 1 $E_{\eta} [T(X)] = \nabla_{\eta} A(\eta)$
 - 2 $\text{cov}_{\eta}(T_i, T_j) = \frac{\partial^2}{\partial \eta_i \partial \eta_j} A(\eta)$
- $\theta \mapsto \eta(\theta)$ est le plus souvent un difféomorphisme.

Beaucoup de modèles usuels sont exponentiels...

- Le modèle Gamma est

$$\mathcal{M} = \left\{ f(x; a, b) = \frac{x^{a-1} b^a}{\Gamma(a)} \exp(-bx) 1_{x \geq 0} \mid a, b > 0 \right\}$$

la densité se réécrit :

$$f(x; a, b) = \exp(-bx + a \log(x) + a \log b - \log \Gamma(a)) \frac{1}{x} 1_{x \geq 0}.$$

- Soit $T_1(x) = -x$ et $T_2(x) = \log x$; $\eta_1 = b$ et $\eta_2 = a$, avec $B(a, b) = a \log b - \log \Gamma(a)$ que l'on ré-écrit $A(\eta_1, \eta_2) = -\eta_1 \log \eta_2 + \log \Gamma(\eta_2)$, $h(x) = \frac{1}{x} 1_{x \geq 0}$

- Le modèle Gaussien multivarié

$$\mathcal{M} = \left\{ f(x; \mu, \Sigma) = \frac{|\det \Sigma|^{-1/2}}{(2\pi)^{p/2}} \exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\right) \mid \mu \in \mathbb{R}^p, \Sigma \succ 0 \right\}$$

- Soit $T_1(x) = x$, $T_2(x) = -\frac{1}{2}xx^\top$; $\eta_1 = \Sigma^{-1}\mu$, $\eta_2 = \Sigma^{-1}$, $A(\eta_1, \eta_2) = -\frac{1}{2} \log \det \eta_2 + \frac{1}{2} \eta_1^\top \eta_2^{-1} \eta_1$.

Exhaustivité dans les familles exponentielles

- Propriété : Dans une famille exponentielle \mathcal{M} , si le modèle est de rang plein (égale à s), la **statistique privilégiée est exhaustive minimale**.

Réciproquement,

- **Théorème de Pitman-Koopman** : Si le modèle est homogène (le support ne dépend pas du paramètre) et est tel que pour une taille d'échantillon n assez grand, il existe une statistique exhaustive de taille fixe s , alors la famille est exponentielle.
- Exemple : La loi de Fréchet $f(x; s, \alpha) = \frac{\alpha}{s} \left(\frac{x}{s}\right)^{-(1+\alpha)} \exp\left(-\left(\frac{x}{s}\right)^{-\alpha}\right)$.

1.3. Comment Construire un estimateur ?

La méthode des moments

- Estimation de θ dans un modèle $\mathcal{M} = \{f(x, \theta) | \theta \in \Theta\}$. On suppose que l'on ait capable de calculer théoriquement $E_{\theta}[X] = \mu_1(\theta)$, $E_{\theta}[X^2] = \mu_2(\theta), \dots, E_{\theta}[X^K] = \mu_K(\theta)$.
- Les espérances sont estimables par leur version empirique : $E_{\theta}[X^k] \approx \frac{1}{n} \sum_{i=1}^n x_i^k \triangleq \overline{X^k}$ (estimateurs consistants).
- Détermination de $\hat{\theta}(X_1, \dots, X_n)$ par moindres carrés

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{i=1}^K \left(\mu_k(\theta) - \overline{X^k} \right)^2$$

- Exemple : Loi Gamma $E(X) = \frac{a}{b}$, $V(X) = \frac{a}{b^2}$. Loi de Poisson $P(X = k) = \frac{\exp(-\lambda)\lambda^k}{k!}$, et $E(X) = \lambda$, $V(X) = \lambda$

La méthode des moindres carrés

- Le modèle de régression est définie par $Y = a + bx + \varepsilon$, avec $E[\varepsilon] = 0$. Nous avons les observations $Y_i = a + bx_i + \varepsilon_i$ avec ε_i i.i.d.
- Paramètre à estimer $\theta = (a, b)$. Nous avons pour tout $i = 1 \dots n$,
$$E[Y_i] = a + bx_i = [1 \ x_i] \begin{bmatrix} a \\ b \end{bmatrix}.$$
- L'estimateur par moindres carrés est

$$\begin{aligned} R(\theta) &= \sum_{i=1}^n (y_i - (a + bx_i))^2 \\ &= \|Y - X\theta\|^2 \end{aligned}$$

Estimateur du Maximum de Vraisemblance

- **Définition 2.6** : Soit \mathcal{M} un modèle paramétrique et x_1, \dots, x_n un échantillon de vraisemblance $L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta)$. Lorsqu'il existe, l'estimateur du maximum de vraisemblance est définie par

$$\begin{aligned}\hat{\theta}_n^{EMV} &= \arg \max_{\theta \in \Theta} L(\theta; x_1, \dots, x_n) \\ &= \arg \max_{\theta \in \Theta} \mathcal{L}(\theta; x_1, \dots, x_n)\end{aligned}$$

Remarque

Les conditions suffisantes d'optimalité pour déterminer $\hat{\theta}_n^{EMV}$ sont :

$$\left\{ \begin{array}{l} \frac{\partial \mathcal{L}}{\partial \theta}(x, \theta)|_{\theta = \hat{\theta}_n^{EMV}} = 0 \\ \frac{\partial^2 \mathcal{L}}{\partial \theta^2}(x_1, \dots, x_n, \hat{\theta}_n^{EMV}) \leq 0 \\ \mathcal{L}(x, \hat{\theta}_n^{EMV}) \geq \mathcal{L}(x_1, \dots, x_n, \theta), \forall \theta \in \Theta. \end{array} \right.$$

Exemples d'estimateurs d'EMV

- Le cas Bernoulli $L(p; x_1, \dots, x_n) = p^{\sum_{i=1}^n x_i} (1-p)^{\sum_{i=1}^n (1-x_i)}$, soit $\mathcal{L}(p; x_1, \dots, x_n) = \sum_{i=1}^n x_i \log\left(\frac{p}{1-p}\right) + n \log(1-p)$.
- Le cas Poisson $L(p; x_1, \dots, x_n) = \exp(-n\lambda) \frac{\lambda^{(x_1+\dots+x_n)}}{x_1! \dots x_n!}, \dots$

Exemples d'estimateurs d'EMV

- Le cas de la régression logistique $Y_i \sim B(1, p_i = p(x_i))$, et l'échantillon est $S_n = ((x_1, y_1), \dots, (x_n, y_n))$. On définit $\varphi(p) = \log \frac{p}{1-p}$ et on suppose que $\varphi(p_i) = ax_i + b$, avec $\theta = a, b \in \mathbb{R}$ alors

$$\mathcal{L}(\theta; x_i, y_i) = \sum_{i=1}^n y_i(ax_i + b) + \sum_{i=1}^n \log(1 - \varphi(ax_i + b))$$

Propriétés de l'EMV

- Sous les hypothèses du théorème de factorisation, $\hat{\theta}_n^{EMV}$ est fonction de toute statistique exhaustive.
- Pour l'EMV, il y a des problèmes :
 - ▶ Existence d'une solution au problème de maximisation
 - ▶ Existence de plusieurs solutions et de minima locaux
 - ▶ Numérique (pour l'optimisation)
- Pour un modèle exponentiel de plein rang (avec la paramétrisation naturelle) et $\mathcal{I}(\eta) > 0$ pour tout η , alors $\mathcal{L}(\eta; x_1, \dots, x_n)$ est strictement concave et l'EMV est l'unique solution de

$$\frac{1}{n} \sum_{i=1}^n T(x_i) = \nabla_{\eta} A(\hat{\eta}^{EMV})$$

1.3. Information de Fisher

Score et Information de Fisher

- On suppose que $\mathcal{M} = \{f(x; \theta), \theta \in \Theta\}$ est **régulier**. Le **score** est le vecteur aléatoire

$$S(X; \theta) = \nabla_{\theta} \log f(X; \theta)$$

Dans ce cas, $\forall \theta \in \Theta$,

$$E_{\theta} [\nabla_{\theta} \log f(X; \theta)] = 0$$

Attention au support!

- La matrice d'information de Fisher** $I(\theta)$, ou information apportée par X sur θ :

$$\mathcal{I}(X; \theta) = \text{cov}_{\theta} (\nabla_{\theta} \log f(X; \theta))$$

- Additivité de l'Information de Fisher** pour un échantillon i.i.d :

$$\mathcal{I}(X_1, \dots, X_n; \theta) \triangleq \mathcal{I}_n(\theta) = n\mathcal{I}_1(\theta)$$

Calcul de l'Information de Fisher

- Théorème : Si $\frac{\partial^2}{\partial \theta_i \partial \theta_j} f(x; \theta)$ existe pour tout x et tout θ , et si on peut dériver 2 fois sous le signe intégral $\int f(x; \theta) \mu(dx)$ alors

$$\mathcal{I}(X; \theta) = - \left(E_{\theta} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(X; \theta) \right] \right)_{1 \leq i, j \leq p}$$

- Pour une famille exponentielle, si on utilise la paramétrisation canonique :

$$\begin{cases} S(X; \eta) &= T(X) - \nabla_{\eta} A(\eta) \\ \mathcal{I}(X; \eta) &= \nabla_{\eta}^2 A(\eta) = \text{cov}_{\eta}(T(X)) \end{cases}$$

Inégalité d'information de Cramér-Rao

Théorème

Soit \mathcal{M} modèle statistique **régulier**, g fonction dérivable et $T_n = T(X_1, \dots, X_n)$ un estimateur sans biais de $g(\theta)$.

Cas θ scalaire : la borne de Cramer-Rao est définie par $BCR(\theta) = \frac{g'(\theta)^2}{\mathcal{I}_n(\theta)}$ et

$$V_{\theta}(T_n) \geq BCR(\theta) = \frac{g'(\theta)^2}{\mathcal{I}_n(\theta)}$$

Cas θ vectoriel :

$$V_{\theta}(T_n) \geq J(g)(\theta)^{\top} \mathcal{I}_n(\theta)^{-1} J(g)(\theta)$$

avec $J(g)(\theta) =$ Jacobien de g évalué au point θ .

Efficacité d'un estimateur

- La comparaison entre les matrices de covariances est au sens de l'ordre (partiel) matrices s.d.p : A, B s.d.p. $A \leq B \iff B - A$ s.d.p
- Inégalité d'information = $BCR(\theta)$ est la meilleure variance possible pour un estimateur sans biais.
- Un estimateur T_n sans biais de $g(\theta)$ est un estimateur **efficace** si $V_{\theta}(g(\theta)) = BCR(\theta)$.

1.5. Comportement asymptotique des estimateurs

Convergence de l'EMV

- Si le modèle paramétrique \mathcal{M} est régulier et si $\theta^* \in \overset{\circ}{\Theta}$ est le vrai paramètre alors

$$\hat{\theta}_n^{EMV} \longrightarrow \theta^*$$

en probabilité P_{θ^*} .

Remarque

La régularité suppose que l'information de Fisher $\mathcal{I}(\theta)$ existe et est définie positive. Une conséquence de $\mathcal{I}(\theta) > 0$ est que le modèle est localement **identifiable**, i.e $f(\cdot; \theta) = f(\cdot; \theta') \implies \theta = \theta'$, for all θ, θ' dans un voisinage de θ .

Normalité asymptotique de l'EMV

- Si le modèle paramétrique \mathcal{M} est régulier et si $\theta^* \in \overset{\circ}{\Theta}$ est le vrai paramètre alors

$$\sqrt{n} \left(\hat{\theta}_n^{EMV} - \theta^* \right) \rightsquigarrow N(0, \mathcal{I}_1^{-1}(\theta))$$

Remarque

- \sqrt{n} est la vitesse de convergence de l'estimateur. De manière générale, la convergence des estimateurs paramétriques est "en $\frac{1}{\sqrt{n}}$ ".
- Un estimateur consistant $\hat{\theta}_n$ est dit **asymptotiquement normal** si il existe $r_n > 0$ tel que

$$r_n \left(\hat{\theta}_n - \theta^* \right) \rightsquigarrow N(0, \mathcal{J}(\theta)).$$

$\mathcal{J}(\theta)$ est la **variance asymptotique**.

Delta-Method

- Soit $\hat{\theta}_n$, un estimateur asymptotiquement normal $r_n(\hat{\theta}_n - \theta^*) \rightsquigarrow N(0, \mathcal{J}(\theta))$, et ψ une fonction différentiable autour de θ^* , alors $\psi(\hat{\theta}_n)$ est asymptotiquement normal et

$$r_n\left(\psi\left(\hat{\theta}_n\right) - \psi\left(\theta^*\right)\right) \rightsquigarrow N\left(0, \psi'(\theta)^\top \mathcal{J}(\theta) \psi'(\theta)\right)$$

- Lemme de Slutsky : Si $Y_n \rightsquigarrow Y$ et $A_n \xrightarrow{\text{Proba}} A$ et $B_n \xrightarrow{\text{Proba}} B$, alors

$$A_n + B_n Y_n \rightsquigarrow A + BY$$

- Ces résultats sont utilisées pour obtenir la normalité asymptotique de statistique complexe (mais régulière) de l'échantillon.

2. Intervalle de confiance

2.1. Intervalle de Confiance : définition

Intervalle de Confiance

On considère un modèle d'échantillonnage paramétrique X_1, \dots, X_n de taille n . Les variables aléatoires X_1, \dots, X_n sont indépendantes et de même loi inconnue P , et elles sont à valeurs dans \mathcal{X} (\mathbb{R} ou \mathbb{R}^d).

On suppose que la loi P appartient à une famille de probabilité $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$.

Définition 3.1 (Région de Confiance)

Soit $x \mapsto \Lambda(x)$ une application de \mathcal{X}^n à valeurs dans l'ensemble des boréliens de $g(\Theta)$ telle que pour tout $\theta \in \Theta$, on a

$$P_\theta(g(\theta) \in \Lambda(X_1, \dots, X_n)) = 1 - \alpha.$$

On dit que l'ensemble aléatoire Λ est une région de confiance de niveau $(1 - \alpha)$.

Dans le cas où $g(\Theta) \subset \mathbb{R}$ et Λ est un intervalle, on parle d'un intervalle de confiance.

Intervalle de Confiance

Définition 3.4 (Statistique asymptotiquement libre)

On appelle statistique asymptotiquement libre pour θ , toute statistique dont sa loi asymptotique ne dépend pas de θ .

Exemple 3.1 Soit X_1, \dots, X_n un échantillon de loi $\mathcal{N}(\mu, \sigma^2)$. Notons $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ et

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

- Si σ est connu, $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ est une statistique libre pour μ , de loi $\mathcal{N}(0, 1)$.
- Si σ est inconnu $\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n}$ est une statistique libre pour μ , de loi T_{n-1} .
- Si μ est connu $\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2$ est une statistique libre pour σ , de loi χ_n^2 .
- Si μ est inconnu $(n-1) \frac{S_n^2}{\sigma^2}$ est une statistique libre pour σ , de loi χ_{n-1}^2 .

2.2. Construction pratique d'un Intervalle de Confiance

Construction pratique d'un IC

Exemple 3.2 X_1, \dots, X_n un échantillon de loi $P(\frac{1}{\lambda})$. $T_n = \frac{\sqrt{n}(\bar{X}_n - \lambda)}{\lambda}$ est une statistique asymptotiquement libre pour λ . En effet, par le théorème central-limite $T_n \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0,1)$.

Construction pratique d'un IC

Soit T_n un estimateur du paramètre θ . On cherche $\varphi_1(T_n)$ et $\varphi_2(T_n)$ de sorte que leurs lois ne dépendent pas de θ et

$$P(\varphi_1(T_n) \leq \theta \leq \varphi_2(T_n)) = 1 - \alpha$$

Remarque Le choix d'une statistique libre revient à éviter d'avoir les bornes de l'IC dépendantes de θ

IC - Espérance d'une variable normale : σ connu

connu Soit X_1, \dots, X_n un échantillon de loi $\mathcal{N}(\mu, \sigma^2)$.

- Prenons $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ comme estimateur de μ .

- On utilise le fait que $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ est une statistique libre pour μ , de loi $\mathcal{N}(0, 1)$.

- L'intervalle de probabilité de \bar{X}_n à $1 - \alpha$ est :

$$\begin{aligned} -u_{\frac{\alpha}{2}} &\leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq u_{\frac{\alpha}{2}} \\ \mu - u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} &\leq \bar{X}_n \leq \mu + u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \end{aligned}$$

où $u_{\frac{\alpha}{2}}$ est telle que $\phi(u_{\frac{\alpha}{2}}) = P(N(0, 1) \leq u_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$.

d'où l'intervalle de confiance :

$$\bar{x}_n - u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x}_n + u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

- Application numérique : si $(1 - \alpha) = 0,95$, on a $u_{\frac{\alpha}{2}} = 1,96$. L'intervalle de confiance de niveau 95% pour μ est

IC - Espérance d'une variable normale : σ inconnu

On utilise le fait que $T_{n-1} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n}$ suit une loi de Student à $(n-1)$ degrés de liberté.

L'intervalle de probabilité pour T_{n-1} :

$$-t_{(n-1), \frac{\alpha}{2}} \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \leq t_{(n-1), \frac{\alpha}{2}}$$

d'où l'intervalle de confiance :

$$\bar{x}_n - t_{(n-1), \frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x}_n + t_{(n-1), \frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

Remarque Si la définition $S_n^{2*} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ alors l'intervalle de confiance devient :

$$\bar{x}_n - t_{(n-1), \frac{\alpha}{2}} \frac{s^*}{\sqrt{n-1}} \leq \mu \leq \bar{x}_n + t_{(n-1), \frac{\alpha}{2}} \frac{s^*}{\sqrt{n-1}}$$

Exemple 3.3 On prend, par exemple $\alpha = 0,05$, et $n = 25$, on a $t_{(n-1), \frac{\alpha}{2}} = 2,064$ et l'intervalle de confiance de niveau de confiance 95% pour μ est :

$$\left[\bar{x}_n - \frac{2,064}{\sqrt{25}} s_{25}, \bar{x}_n + \frac{2,064}{\sqrt{25}} s_{25} \right]$$

Remarque Le TCL a pour conséquence que les intervalles précédents sont valables pour

IC - Variance d'une variable normale

connu soit $\hat{U}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$. on utilise le fait que $\frac{n\hat{U}_n^2}{\sigma^2}$ suit une loi de χ_n^2 comme somme de n carrés de $\mathcal{N}(0,1)$ indépendantes.

Soit k_1 et k_2 , les bornes de l'intervalle de probabilité d'un χ_n^2 :

$$P(k_1 \leq \frac{n\hat{U}_n^2}{\sigma^2} \leq k_2) = 1 - \alpha$$

d'où l'intervalle de confiance :

$$\frac{n\hat{u}_n^2}{k_2} \leq \sigma^2 \leq \frac{n\hat{u}_n^2}{k_1}$$

connu On utilise $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$, et on sait que $\frac{(n-1)S_n^2}{\sigma^2}$ suit une loi χ_{n-1}^2 ,

soit b_1 et b_2 les bornes de l'intervalle de probabilité :

$$P(b_1 \leq \frac{(n-1)S_n^2}{\sigma^2} \leq b_2) = 1 - \alpha$$

d'où l'intervalle de confiance :

$$\frac{(n-1)s_n^2}{b_2} \leq \sigma^2 \leq \frac{(n-1)\hat{s}_n^2}{b_1}$$

Intervalle de Confiance : Exemple

Exemple 3.4 Soit X_1, \dots, X_n un échantillon de loi uniforme $\mathcal{U}([0, \theta])$.

La vraisemblance associée à l'échantillon est définie par

$$L_n(x, \theta) = \prod_{i=1}^n \frac{1}{\theta} \mathbf{1}_{[0, \theta]}(x_i)$$

Déterminons l'EMV : $L_n(x, \theta)$ est maximum en θ ssi $\prod_{i=1}^n \mathbf{1}_{[0, \theta]}(x_i) = 1$.

Autrement dit $\forall i = 1, \dots, n \theta \geq x_i$, donc, $\sup_{1 \leq i \leq n} x_i \leq \theta$. D'où l'EMV :

$$\hat{\theta}_n = \sup_{1 \leq i \leq n} X_i$$

Cherchons une statistique libre pour θ (à partir de $\hat{\theta}_n$). D'abord calculons la loi de $\hat{\theta}_n$, $\forall x \in [0, \theta]$,

$$\begin{aligned} F_{\hat{\theta}_n}(x) &= P(\hat{\theta}_n \leq x) \\ &= \left(\frac{x}{\theta}\right)^n \end{aligned}$$

Vérifions que $T_n = \frac{\hat{\theta}_n}{\theta}$ est une statistique libre pour θ . En effet, $\forall x \in [0, 1]$, on a

Intervalle de Confiance : Exemple

IC Méthode 1 : Déterminons un $\varphi_1(\hat{\theta}_n)$ et $\varphi_2(\hat{\theta}_n)$ telle que

$$P\left(\varphi_1(\hat{\theta}_n) \leq \theta \leq \varphi_2(\hat{\theta}_n)\right) = 1 - \alpha \quad (1)$$

A partir de (1), il faut construire autour de θ , l'estimateur T_n

$$P\left(\frac{1}{\varphi_2(\hat{\theta}_n)} \leq \frac{1}{\theta} \leq \frac{1}{\varphi_1(\hat{\theta}_n)}\right) = 1 - \alpha$$
$$P\left(\frac{\hat{\theta}_n}{\varphi_2(\hat{\theta}_n)} \leq T_n \leq \frac{\hat{\theta}_n}{\varphi_1(\hat{\theta}_n)}\right) = 1 - \alpha$$

Comme on connaît la loi de la statistique libre T_n , on cherche donc un $a := \frac{\hat{\theta}_n}{\varphi_2(\hat{\theta}_n)}$ et $b := \frac{\hat{\theta}_n}{\varphi_1(\hat{\theta}_n)}$ tels que

$$P(a \leq T_n \leq b) = 1 - \alpha$$
$$F_{T_n}(b) - F_{T_n}(a) = 1 - \alpha$$
$$b^n - a^n = 1 - \alpha$$

Il y a une infinité de a et b possibles. On choisit a et b tels que l'intervalle $[a, b]$ soit le plus court possible. Soit $b = 1$ et donc $a = \alpha^{\frac{1}{n}}$.

Intervalle de Confiance : Exemple

Soit $\varphi_1(\hat{\theta}_n) = \hat{\theta}_n$ et $\varphi_2(\hat{\theta}_n) = \alpha^{-\frac{1}{n}} \hat{\theta}_n$ Nous avons donc obtenu un intervalle de confiance :

$$\left[\hat{\theta}_n ; \alpha^{-\frac{1}{n}} \hat{\theta}_n \right]$$

IC **Méthode 2** : On peut partir directement de T_n , et l'on cherche un a et b tel que

$$\begin{aligned} P(a \leq T_n \leq b) &= 1 - \alpha \\ F_{T_n}(b) - F_{T_n}(a) &= 1 - \alpha \\ b^n - a^n &= 1 - \alpha \end{aligned}$$

On choisit les a et b tels que $[a, b]$ soit l'intervalle le plus court possible.

Puis on remonte le raisonnement pour arriver à un intervalle de confiance (voir exemple sur les espérances/variances normales).

2.3. Intervalle de Confiance Asymptotique

Intervalle de Confiance Asymptotique

éthode Modèle régulier.

- On construit l'intervalle asymptotique en utilisant l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ .
- On sait que $\hat{\theta}_n$ est asymptotiquement normal, i.e. :

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}\left(0, \frac{1}{\mathcal{I}_1(\theta)}\right).$$

- Autrement dit :

$$\sqrt{n \mathcal{I}_1(\theta)}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, 1).$$

On obtient :

$$P\left(-u_{\frac{\alpha}{2}} \leq \sqrt{n \mathcal{I}_1(\hat{\theta}_n)}(\hat{\theta}_n - \theta) \leq u_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

d'où l'intervalle de confiance asymptotique :

$$\left[\hat{\theta}_n - \frac{u_{\frac{\alpha}{2}}}{\sqrt{n \mathcal{I}_1(\hat{\theta}_n)}}, \hat{\theta}_n + \frac{u_{\frac{\alpha}{2}}}{\sqrt{n \mathcal{I}_1(\hat{\theta}_n)}} \right].$$

Intervalle de Confiance Asymptotique : Exemple

Remarque On a mentionné l'EMV car il est asymptotiquement normal, donc il converge en loi vers une loi connue (ici la loi normale).

Exemple 3.5 Soit X_1, \dots, X_n un échantillon de loi uniforme $U([0, \theta])$.

L'Estimateur de Maximum de Vraisemblance : $\hat{\theta}_n = \sup_{1 \leq i \leq n} X_i$.

Soit $T_n := n \left(1 - \frac{\hat{\theta}_n}{\theta}\right)$, T_n est une statistique asymptotiquement libre pour θ , en effet ;

$$\begin{aligned} F_{T_n}(x) &= P(T_n \leq x) \\ &= P\left(\hat{\theta}_n \geq \theta \left(1 - \frac{x}{n}\right)\right)^n \\ &= 1 - \left(1 - \frac{x}{n}\right)^n \end{aligned}$$

et $\lim_{n \rightarrow \infty} F_{T_n}(x) = 1 - e^{-x}$. Donc, $T_n \rightarrow T$, où T suit une loi exponentielle $\mathcal{E}xp(1)$.

Intervalle de Confiance Asymptotique : Exemple

On cherche un intervalle de la forme $[\hat{\theta}_n ; b]$ tel que pour un α donné on ait

$$P(\hat{\theta}_n \leq \theta \leq b) = 1 - \alpha.$$

$$\begin{aligned} P(\hat{\theta}_n \leq \theta \leq b) &= P\left(0 \leq T_n \leq n\left(1 - \frac{\hat{\theta}_n}{b}\right)\right) \\ &\approx_{n \rightarrow \infty} F_T\left(n\left(1 - \frac{\hat{\theta}_n}{b}\right)\right) = 1 - \alpha \end{aligned}$$

d'où $b = \frac{\hat{\theta}_n}{1 - \frac{1}{n} F_T^{-1}(1 - \alpha)} = \frac{\hat{\theta}_n}{1 - \frac{1}{n} \log \alpha}$ On a donc

$$IC = \left[\hat{\theta}_n ; \frac{\hat{\theta}_n}{1 - \frac{1}{n} \log \alpha} \right]$$