

1 Convergence de variables aléatoires

Calcul sur les événements

Prop. • Si $(A_n)_n$ est croissante, $\mathbf{P}(\bigcup_n A_n) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n)$.
 • Si $(A_n)_n$ est décroissante, $\mathbf{P}(\bigcap_n A_n) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n)$.
 • Si $\forall n, \mathbf{P}(A_n) = 0$ alors $\mathbf{P}(\bigcup_n A_n) = 0$.
 • Si $\forall n, \mathbf{P}(A_n) = 1$ alors $\mathbf{P}(\bigcap_n A_n) = 1$.

Def. $\limsup_{n \rightarrow \infty} A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k$, i.e. $\omega \in \limsup_{n \rightarrow \infty} A_n \iff \forall n, \exists k \geq n, \omega \in A_k$.
 Donc $\limsup_{n \rightarrow \infty} A_n$ est réalisé ssi une infinité de A_n est réalisé.

Lem (de Borel-Cantelli). Si $\sum_n \mathbf{P}(A_n) < \infty$, alors $\mathbf{P}(\limsup_{n \rightarrow \infty} A_n) = 0$.

Autrement dit, il y a une proba 1 pour que seulement un nombre fini de A_n soient réalisés.

Convergence p.s., en probabilité et dans L^p

Def. (i) On dit que $X_n \xrightarrow{\mathbf{P}} X$ (**converge en probabilité**) si $\forall \epsilon > 0, \mathbf{P}(\|X_n - X\| > \epsilon) \xrightarrow{n \rightarrow \infty} 0$.

(ii) On dit que $X_n \xrightarrow{\text{p.s.}} X$ (**converge presque sûrement**), si $\forall \omega$ \mathbf{P} -p.p, $X_n(\omega) \rightarrow X(\omega)$. Autrement dit il existe $A \in \mathcal{F}$ tel que $\mathbf{P}(A) = 1$ et $\forall \omega \in A, \lim_n X_n(\omega) = X(\omega)$.

(iii) On dit que $X_n \xrightarrow{L^p} X$ (**converge vers X dans $L^p(\Omega, \mathbf{R}^d)$**) si $X_n, X \in L^p$ et $\mathbf{E}(\|X_n - X\|^p) \xrightarrow{n \rightarrow \infty} 0$.

Prop. On note $X_n = (X_n^{(1)}, \dots, X_n^{(d)})$ sur $\mathcal{X} = \mathbf{R}^d$. Alors $X_n \xrightarrow{\text{p.s.}} X$ p.s. (resp. en probabilité, dans L^p) ssi $\forall k \in \llbracket 1; d \rrbracket, X_n^{(k)} \xrightarrow{\text{p.s.}} X^{(k)}$ (resp. en probabilité, dans L^p).

Prop. Si $X_n \xrightarrow{\text{p.s.}} X$ ou $X_n \xrightarrow{L^p} X$ alors $X_n \xrightarrow{\mathbf{P}} X$.

Prop. Si $\forall \epsilon > 0, \sum_n \mathbf{P}(\|X_n - X\| > \epsilon) < \infty$ alors $X_n \xrightarrow{\text{p.s.}} X$.

Prop. $X_n \xrightarrow{\mathbf{P}} X$ ssi de toute sous-suite $X_{\varphi(n)}$ on peut extraire une autre sous-suite $X_{\varphi \circ \psi(n)}$ telle que $X_{\varphi \circ \psi(n)} \xrightarrow{\text{p.s.}} X$.

Th (de continuité). Soit $h: \mathbf{R}^d \rightarrow \mathbf{R}^p$ mesurable et continue sur $C \in \mathcal{B}(\mathbf{R}^d)$ tel que $\mathbf{P}(X \in C) = 1$:

(i) Si $X_n \xrightarrow{\text{p.s.}} X$ alors $h \circ X_n \xrightarrow{\text{p.s.}} h \circ X$

(ii) Si $X_n \xrightarrow{\mathbf{P}} X$ alors $h \circ X_n \xrightarrow{\mathbf{P}} h \circ X$.

Th (Loi forte des grands nombres). Soit (X_n) i.i.d. telle que $\mathbf{E}(\|X_1\|) < \infty$. Alors $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{p.s.}} \mathbf{E}(X_1)$.

Th (Loi faible des grands nombres). Soit (X_n) i.i.d. telle que $\mathbf{E}(\|X_1\|^2) < \infty$. On a $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbf{P}} \mathbf{E}(X_1)$.

2 Convergence en loi

Définitions et propriétés

Def. Une mesure de proba μ sur $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ est caractérisée par sa **fonction de répartition**

$$F_\mu: (x_1, \dots, x_d) \mapsto \mu \left(\prod_i]-\infty; x_i] \right).$$

Prop. • F_μ croissante.

• $F_\mu(-\infty) = 0, F_\mu(+\infty) = 1$

• F_μ est continue à droite et $\mu(x_0) = F_\mu(x_0) - F_\mu(x_0^-)$

Th. Deux mesures distinctes ne peuvent pas avoir la même fonction de répartition.

Soit $X: \Omega \rightarrow \mathbf{R}^d$ une v.a. On note $P_X = \mathbf{P} \circ X^{-1}$ la loi de X . P_X est une mesure de proba sur \mathbf{R}^d . On note F_X sa fonction de répartition. Pour $d = 1, F_X(x) = \mathbf{P}(X \leq x)$.

Def. Soit $(\mu_n)_n, \mu$ des mesures de proba sur \mathbf{R}^d . On dit que μ_n **converge faiblement (ou étroitement)** vers μ si $\lim_n F_{\mu_n}(x) = F_\mu(x)$ en tout x point de continuité de F_μ . On note $\mu_n \Rightarrow \mu$.

Def. $(X_n)_n, X$ v.a. sur \mathbf{R}^d . On dit que X_n **converge en loi** vers X (noté $X_n \xrightarrow{\mathcal{L}} X$) si $P_{X_n} \Rightarrow P_X$.

Prop. Si $X_n \xrightarrow{\mathbf{P}} X$ alors $X_n \xrightarrow{\mathcal{L}} X$.

Th (de représentation de Skorohod). Soit $(\mu_n)_n, \mu$ des mesures de proba sur \mathbf{R}^d telles que $\mu_n \Rightarrow \mu$. Il existe un espace de proba et des v.a. $(Y_n), Y$ sur cet espace à valeurs dans $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ telles que $Y \sim \mu, \forall n, Y_n \sim \mu_n$ et Y est limite simple des Y_n , i.e. $\forall \omega, Y_n(\omega) \rightarrow Y(\omega)$.

Th (de continuité). Soit $h: \mathbf{R}^d \rightarrow \mathbf{R}^p$ mesurable et continue sur $C \in \mathcal{B}(\mathbf{R}^d)$ tel que $\mathbf{P}(X \in C) = 1$. Si $X_n \xrightarrow{\mathcal{L}} X$ alors $h \circ X_n \xrightarrow{\mathcal{L}} h \circ X$.

Th. Les affirmations suivantes sont équivalentes :

- (i) $\mu_n \Rightarrow \mu$,
- (ii) pour toute fonction f continue, $\mu_n(f) \rightarrow \mu(f)$,
- (iii) pour toute fonction f lipschitzienne, $\mu_n(f) \rightarrow \mu(f)$,
- (iv) pour tout $A \in \mathcal{B}(\mathbf{R}^d)$ tel que $\mu(\delta A) = 0$ (frontière de A), $\mu_n(A) \rightarrow \mu(A)$.

Th (de Portmanteau). On a équivalence entre :

- (i) $X_n \xrightarrow{\mathcal{L}} X$,
- (ii) $\forall f: \mathbf{R}^d \rightarrow \mathbf{R}$ continue bornée, $\mathbf{E}(f(X_n)) \rightarrow \mathbf{E}(f(X))$,
- (iii) $\forall A \subset \mathbf{R}^d$ tel que $\mathbf{P}(X \in \delta A) = 0$, on a $\mathbf{P}(X_n \in A) \rightarrow \mathbf{P}(X \in A)$ où $\delta A = \bar{A} \setminus \overset{\circ}{A}$,
- (iv) $\forall t \in \mathbf{R}^d$, $\lim_n \phi_{X_n}(t) = \phi_X(t)$.

Th. Soit $m \in \mathbf{N}^*$ et $h: \mathbf{R}^d \rightarrow \mathbf{R}^p$ mesurable et continue sur $C \in \mathcal{B}(\mathbf{R}^d)$ tel que $\mathbf{P}(X \in C) = 1$. Si $\mu_n \Rightarrow \mu$ alors $\mu_n h^{-1} \Rightarrow \mu h^{-1}$.

Th (Procédé de Cramer-Wold). Soit X_n, X des v.a. sur \mathbf{R}^d . On a $X_n \xrightarrow{\mathcal{L}} X \iff \forall t, \langle t | X_n \rangle \xrightarrow{\mathcal{L}} \langle t | X \rangle$.

Mesures tendues

Lem (d'Helly). Soit $(F_n)_n$ une suite de fonctions de répartition. Il existe une sous-suite $(\varphi(n))_n$ et $F: \mathbf{R} \rightarrow [0; 1]$ croissante, continue à droite, telle que $\lim_n F_{\varphi(n)}(x) = F(x)$ en tout x point de continuité de F .

On ajoute une condition pour que la limite vérifie $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.

Def. $(\mu_n)_n$ est dite **tendue** si $\forall \varepsilon > 0, \exists \mathcal{K}$ compact, $\forall n, \mu_n(\mathcal{K}) \geq 1 - \varepsilon$.

Dans le cas $d = 1$ on peut prendre $\mathcal{K} = [-K; K]$.

Def. $(X_n)_n$ est **tendue** si $\forall \varepsilon > 0, \exists \mathcal{K}$ compact, $\forall n, \mathbf{P}(X_n \in \mathcal{K}) \geq 1 - \varepsilon$.

Prop. (i) Toute famille finie de mesures de probabilité est tendue.

(ii) Une suite de mesures qui converge faiblement est tendue.

Th (de Prokhorov). Soit une famille \mathcal{M} de mesures de probabilité. Alors \mathcal{M} est tendue si et seulement si elle est relativement séquentiellement compacte pour la topologie de la convergence faible, i.e. de toute suite $(\mu_n)_n$ de \mathcal{M} on peut extraire une sous suite $(\mu_{\varphi(n)})_n$ telle que $\mu_{\varphi(n)} \Rightarrow \mu$ avec μ une mesure de probabilité.

Prop. Soit $(\mu_n)_n$ tendue. Si toute sous-suite faiblement convergente de $(\mu_n)_n$ converge vers μ , alors $\mu_n \Rightarrow \mu$.

Fonction caractéristique

Def. La fonction caractéristique d'une mesure de proba μ sur \mathbf{R}^d est $\phi_\mu: \begin{matrix} \mathbf{R}^d & \rightarrow & \mathbf{C} \\ t & \mapsto & \int e^{i\langle t|x \rangle} d\mu(x) \end{matrix}$.

Th. $\varphi_\mu = \varphi_\nu \implies \mu = \nu$.

Prop. Soit μ une mesure sur \mathbf{R} et $p \in \mathbf{N}$ tel que $\int |x|^p d\mu(x) < \infty$. Alors ϕ_μ est p fois continument dérivable et $\forall t \in \mathbf{R}$, sa p^e dérivée satisfait $\forall t \in \mathbf{R}, \phi_\mu^{(p)} = \int i^p x^p e^{itx} d\mu(x)$.

Ex. $\forall t \in \mathbf{R}, \phi_{\mathcal{N}(0,1)}(t) = e^{-t^2/2}$.

Prop. Pour $Y = AX + b$ on a $\forall t, \phi_Y(t) = e^{i\langle t|b \rangle} \phi_X(A^*t)$.

Prop. ϕ_μ est continue en zéro.

Th (de Lévy). Soit $(\mu_n)_n, \mu$ des mesures de probabilité sur \mathbf{R}^d . $\mu_n \Rightarrow \mu$ ssi $\forall t \in \mathbf{R}^d, \phi_{\mu_n}(t) \rightarrow \phi_\mu(t)$.

Théorème centrale limite

Not. $\mathcal{N}(m, \sigma^2)$ désigne la loi de densité $\rho(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$, et si $\sigma^2 = 0$ c'est la loi δ_m .

Pour X gaussien, sa fonction caractéristique vérifie $\phi_X(t) = e^{i\langle t|m \rangle} e^{-\frac{1}{2}\langle t|\Sigma t \rangle}$ où $m = \mathbf{E}(X)$ et $\Sigma = \text{Cov}(X)$.

Th (central limite). Soit $(X_n)_n$ iid tel que $\mathbf{E}(\|X_1\|^2) < \infty$. Alors

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mathbf{E}(X_1)) \Rightarrow \mathcal{N}(0, \text{Cov}(X_1)).$$

Th (de Linderbergh). Soit un tableau de v.a. $(X_{i,n})_{1 \leq i \leq n}$ tel que

- $\forall n$, les v.a. $X_{1,n}, \dots, X_{n,n}$ sont indépendantes,
- $\forall n, \forall i \leq n, \mathbf{E}(X_{i,n}) = 0$,
- $\exists \Sigma, \lim_n \sum_{i=1}^n \text{Cov}(X_{i,n}) = \Sigma$
- $\forall \varepsilon > 0, \lim_n \sum_{i=1}^n \mathbf{E}(\|X_{i,1}\|^2 \mathbf{1}_{\|X_{i,n}\| > \varepsilon}) = 0$ (condition de Lindebergh).

Alors $\sum_{i=1}^n X_{i,n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma)$.

Lem. Soit $\mathcal{U} := \{z \in \mathbf{C} \mid |z| \leq 1\}$, $n \in \mathbf{N}^*$ et deux familles $(z_i)_{1 \leq i \leq n}$ et $(w_i)_{1 \leq i \leq n}$ de \mathcal{U} . Alors $|\prod_{i=1}^n z_i - \prod_{i=1}^n w_i| \leq \sum_{i=1}^n |z_i - w_i|$.

3 Manipulation des convergences

Quelques règles

Prop. (i) $\forall a \in \mathbf{R}^d, X_n \xrightarrow{\mathcal{L}} a \iff X_n \xrightarrow{\mathbf{P}} a$

(ii) $\left. \begin{array}{l} X_n \xrightarrow{\mathcal{L}} X \\ Y_n - X_n \xrightarrow{\mathcal{L}} 0 \end{array} \right\} \implies Y_n \xrightarrow{\mathcal{L}} X$

(iii) $\forall a \in \mathcal{Y}, \left. \begin{array}{l} X_n \xrightarrow{\mathcal{L}} X \\ Y_n \xrightarrow{\mathcal{L}} a \end{array} \right\} \implies (X_n, Y_n) \xrightarrow{\mathcal{L}} (X, a)$

(iv) $\left. \begin{array}{l} X_n \xrightarrow{\mathbf{P}} X \\ Y_n \xrightarrow{\mathbf{P}} c \end{array} \right\} \implies (X_n, Y_n) \xrightarrow{\mathbf{P}} (X, c)$

Notation o_P, O_P

Not. • On dit que $X_n = o_P(1)$ si $X_n \xrightarrow{\mathbf{P}} 0$.

• On dit que $X_n = o_P(Y_n)$ si $\exists (Z_n) \xrightarrow{\mathbf{P}} 0, X_n = Z_n Y_n$.

• On dit que $X_n = O_P(1)$ si (X_n) est borné en probabilité, i.e. (X_n) est tendue.

• On dit que $X_n = O_P(Y_n)$ si $\exists (Z_n) = O_P(1), X_n = Z_n Y_n$.

Prop. Si $X_n \xrightarrow{\mathcal{L}} X$ alors $X_n = O_P(1)$.

Prop. (i) $o_P(1) + O_P(1) = O_P(1)$,

(ii) $o_P(1) \cdot O_P(1) = o_P(1)$,

(iii) $o_P(1) + o_P(1) = o_P(1)$,

(iv) $\frac{1}{1+o_P(1)} = O_P(1)$.

Lem (de Slutsky). Soit $(X_n)_n$ et $(Y_n)_n$ tels que $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} a$:

(i) Pour $a \in \mathcal{X}$ quelconque, $X_n + Y_n \xrightarrow{\mathcal{L}} X + a$.

(ii) Dans le cas réel $a \in \mathbf{R}$, $X_n Y_n \xrightarrow{\mathcal{L}} aX$ et $\frac{X_n}{Y_n} \xrightarrow{\mathcal{L}} \frac{X}{a}$ si $a \neq 0$.

Delta-méthode

Soit $g: \mathbf{R}^d \rightarrow \mathbf{R}^m$ dérivable en un point $\nu \in \mathbf{R}^d$ de matrice jacobienne $\nabla g(\nu)$.

Rappel : $\lim_{h \rightarrow 0} \frac{\|g(\nu+h) - g(\nu) - \nabla g(\nu) \cdot h\|}{\|h\|} = 0, \nabla g(\nu) = \left(\frac{\partial g_i(\nu)}{\partial \nu_j} \right)_{i \in [1;d], j \in [1;m]}$.

Th. Soit $g: \mathbf{R}^d \rightarrow \mathbf{R}^m$ dérivable en ν . Soient T_n, T des v.a. sur \mathbf{R}^d et (r_n) une suite réelle telle que $\lim_n r_n = +\infty$ et $r_n(T_n - \nu) \xrightarrow{\mathcal{L}} T$. Alors $r_n(g(T_n) - g(\nu)) \xrightarrow{\mathcal{L}} \nabla g(\nu) \cdot T$.

4 Statistique asymptotique

Not. • $(\Omega, \mathcal{F}, \mathbf{P})$ espace de proba,

• $(X_i)_{i \in \mathbf{N}}$ v.a. iid

• $\forall i \in \mathbf{N}, X_i = (X_i^{(1)}, \dots, X_i^{(d)})^\top$,

• $\|\cdot\|$ norme euclidienne.

Introduction

Def. Estimateur $\hat{\theta}_n$ à valeurs dans $\Theta \subset \mathbf{R}^q$: transformation mesurable de (X_1, \dots, X_n) . $\hat{\theta}_n$ est faiblement consistant si $\hat{\theta}_n \xrightarrow{\mathbf{P}} \theta_0$. $\hat{\theta}_n$ est fortement consistant si $\hat{\theta}_n \xrightarrow{\mathbf{P}^{\text{s}}} \theta_0$. $\hat{\theta}_n$ est asymptotiquement normal si $\sqrt{n}(\hat{\theta}_n - \theta_0) \implies \mathcal{N}(0, \sigma_0^2)$.

Rem. La consistance est différente du biais. En effet, soit $\bar{X}^n = \frac{1}{n} \sum_{i=1}^n X_i, \hat{\theta}_n = \bar{X}^n + \frac{1}{n}$ est fortement consistant (si $\mathbf{E}(X_1) < \infty$) et biaisé car $\mathbf{E}(\hat{\theta}_n) - \mathbf{E}(X_1) = \frac{1}{n}$.

À l'inverse $\hat{\theta}_n = X_1$ est sans biais mais non consistant.

Def. $\hat{\theta}_n$ est un M-estimateur si $\hat{\theta}_n \in \arg \min_{\theta \in \Theta} M(\theta)$. $\hat{\theta}_n$ est un Z-estimateur si $\Psi_n(\hat{\theta}_n) = 0$.

Ex. • Moindres carrés : $\hat{\beta}_n$ est défini par $\hat{\beta}_n = \arg \min_{\beta \in \mathbf{R}^d} \sum_{i=1}^n (Y_i - X_i^\top \beta)^2$.

- Maximum de vraisemblance : soit la famille paramétrique $\mathcal{P} = \{f_\theta \mid \theta \in \Theta\}$ selon laquelle est distribuée les données (X_1, \dots, X_n) .

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} \frac{1}{n} \log(f_\theta(X_i))$$

- Estimateur des moments et estimateur des moments généralisés : $\hat{\theta}_n \in \arg \min_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^n g(X_i) - \int g d\mathbf{P}_\theta \right\|$.

Rem. Un Z-estimateur est toujours un M-estimateur car $\forall \theta \in \Theta, 0 = \left\| \Psi_n(\hat{\theta}_n) \right\| \leq \left\| \Psi_n(\theta) \right\|$. Un M-estimateur est un Z-estimateur si M_n est continuellement dérivable sur Θ et $\hat{\theta}_n$ est un point intérieur à Θ . Alors $\nabla M_n(\hat{\theta}_n) = 0$.

Prop (Consistance).

Prop (Consistance Z-estimateur). Si $\hat{\theta}_n$ est un Z-estimateur et

- $\sup_{\theta \in \Theta} \left\| \Psi_n(\theta) - \Psi(\theta) \right\| \xrightarrow{\mathbf{P} \text{ resp. p.s.}} 0,$
- $\forall \varepsilon > 0, \inf_{\theta \in \Theta \setminus B(\theta_0, \varepsilon)} \left\| \Psi(\theta) \right\| > \left\| \Psi(\theta_0) \right\|,$

alors $\hat{\theta}_n \xrightarrow{\mathbf{P} \text{ resp. p.s.}} \theta_0$.

Lem. Supposons

- Θ compact,
- $\forall \theta \in \Theta, \mathbf{E}(|\rho(X_1, \theta)|) < \infty,$
- $\exists r: \mathcal{X} \rightarrow \mathbf{R}_+, \mathbf{E}(r(X_1)) < \infty$ où ...

Lem (Vérification de la condition d'identifiabilité). Soit $\Theta \subset \mathbf{R}^q$ compact et $M \in \mathcal{C}^0(\Theta)$ telle que θ_0 en est l'unique maximum. Alors $\forall \varepsilon > 0, \inf_{\theta \in \Theta \setminus B(\theta_0, \varepsilon)} M(\theta) > M(\theta_0)$.

Prop. Soit Θ compact. Supposons $\mathbf{E}_{\theta_0} \|g\| < +\infty$ et $\hat{\theta}_n \in \arg \min_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^n g(X_i) - \mathbf{E}_\theta g \right\|$. Si, de plus, $\theta \mapsto \mathbf{E}_\theta g$ est injective et continue alors $\hat{\theta}_n \xrightarrow{\text{p.s.}} \theta_0$.

Normalité asymptotique

On considère ici uniquement les Z-estimateurs : $\frac{1}{n} \sum_{i=1}^n \psi(X_i, \hat{\theta}_n) = \Psi_n(\hat{\theta}_n) = 0$.

Th. Supposons que :

- $\hat{\theta}_n \xrightarrow{\mathbf{P}} \theta,$
- il existe un voisinage $v(\theta_0)$ tel que $\forall x \in \mathcal{X}, \theta \mapsto \psi(x, \theta)$ est \mathcal{C}^2 sur $v(\theta_0)$ et $\forall k \in \llbracket 1; q \rrbracket, \mathbf{E} \left[\sum_{\theta \in v(\theta_0)} \|\psi_k(X_1, \theta)\| \right] < +\infty,$
- soit $\Phi: (x, \theta) \mapsto \begin{pmatrix} \nabla_\theta \psi_1(x, \theta) \\ \vdots \\ \nabla_\theta \psi_q(x, \theta) \end{pmatrix} \in \mathbf{R}^{q \times q}$ tel que $\mathbf{E} [\|\Phi(X_1, \theta_0)\|] < +\infty$ et $\mathbf{E}[\Phi(X_1, \theta_0)] = Q(\theta_0)$ est inversible,
- $\mathbf{E}[\|\psi(X_1, \theta_0)\|^2] < +\infty.$

Alors $\sqrt{n}(\hat{\theta}_n - \theta_0) = -Q(\theta_0)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(X_i, \theta_0) \right) + o_P(1)$. En particulier

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, Q(\theta_0)^{-1} \text{Var}(\psi(X_1, \theta_0)) (Q(\theta_0)^{-1})^\top \right).$$