Isai Gordeev et Zhicheng Hui
Régressions, MAP361
École Polytechnique

T1

On dispose le modèle linéare

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Comme ε_i est une variable aléatoire gaussienne centrée de variance $\sigma^2 > 0$, on l'exprime avec la variable y_i et la variable fixe x_i

$$\varepsilon_i = y_i - (\beta_0 + \beta_1 x_i) \sim \mathcal{N}(0, \sigma^2)$$

D'où

$$\frac{\varepsilon_i}{\sigma} = \frac{y_i - (\beta_0 + \beta_1 x_i)}{\sigma} \sim \mathcal{N}(0, 1)$$

$(\epsilon_i)_{i\geq 1}$ étant indépendentes, $(\frac{\epsilon_i}{\sigma})_{i\geq 1}$ l'est aussi, donc on a

$$\frac{\varepsilon}{\sigma} = \frac{y - X\beta}{\sigma} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$$

On obtient ainsi la loi de y

$$y \sim \mathcal{N}(X\beta, \sigma^2 \boldsymbol{I})$$

Finalement

$$p(y) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{||y - X\beta||^2}{2\sigma^2}\right)$$

T2

L'estimateur du maximum de vraisemblance

$$\hat{\beta} \in \arg\min_{\beta \in R^2} ||y - X\beta||^2$$

Vérifions d'abord que la fonction $\beta \to ||y - X\beta||^2$ soit convexe et différentiable sur R^2 :
Soit $\beta_1, \beta_2 \in R^2$, soit $t \in [0, 1]$, alors on a selon l'inégalité triangulaire

$$||y - X(t\beta_1 + (1-t)\beta_2)||^2 \leq (|t| \cdot ||y - X\beta_1|| + |1-t| \cdot ||y - X\beta_2||)^2 \leq t^2 ||y - X\beta_1||^2 + (1-t)^2 ||y - X\beta_2||^2$$

Or, on a $t^2 \leq t$ et $(1-t)^2 \leq 1 - t$ pour tout $t \in [0, 1]$

$$||y - X(t\beta_1 + (1-t)\beta_2)||^2 \leq (|t| \cdot ||y - X\beta_1|| + |1-t| \cdot ||y - X\beta_2||)^2 \leq t ||y - X\beta_1||^2 + (1-t) ||y - X\beta_2||^2$$

Donc cette fonction est convexe.
Puis, on calcule

$$\frac{\partial ||y - X\beta||^2}{\partial \beta_1} = -X \begin{pmatrix} 1 \\ 0 \end{pmatrix} (y - X\beta)^{\mathrm{T}} - \begin{pmatrix} 0 & 1 \end{pmatrix} X^{\mathrm{T}} (y - X\beta)$$

$$\frac{\partial ||y - X\beta||^2}{\partial \beta_2} = -X \begin{pmatrix} 0 \\ 1 \end{pmatrix} (y - X\beta)^{\mathrm{T}} - \begin{pmatrix} 1 & 0 \end{pmatrix} X^{\mathrm{T}}(y - X\beta)$$

Les deux dérivées partielles soient continues sur R^2, on sait que cette fonction est différentiable sur R^2.

Si $\hat{\beta}$ satisfait (4.2), il suffit que

$$\nabla_\beta ||y - X\beta||^2(\hat{\beta}) = 0$$

$$\nabla_\beta \, (y - X\beta)^{\mathrm{T}} \, (y - X\beta)(\hat{\beta}) = 0$$

Ainsi

$$\nabla_\beta \left(y^{\mathrm{T}} y - 2y^{\mathrm{T}} X\beta + X^{\mathrm{T}} X \beta^{\mathrm{T}} \beta \right) (\hat{\beta}) = 0$$

$$-2y^{\mathrm{T}} X + 2X^{\mathrm{T}} X\hat{\beta} = 0$$

$$X^{\mathrm{T}} X\hat{\beta} = X^{\mathrm{T}} y$$

Ce qui implique que $\mathrm{Im}(X^{\mathrm{T}}X) = \mathrm{Im}(X^{\mathrm{T}})$, et l'ensemble des solutions $\arg\min_{\beta \in R^2} ||y - X\beta||^2$ est celui du système linéaire :

$$X^{\mathrm{T}} X\hat{\beta} = X^{\mathrm{T}} y$$

où $\hat{\beta} \in R^2$ est l'inconnu.

T3

On a pout tout $1 \leq i \leq n$,

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i = (X\beta)_i + \varepsilon_i$$

Donc, on a bien

$$y = X\beta + \varepsilon$$

Si X et β sont fixés, d'après **T1**,

$$y \sim \mathcal{N}(X\beta, \sigma^2 \boldsymbol{I})$$

Comme on a $\hat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y$ et $y = X\beta + \varepsilon$,

$$E[\hat{\beta}] = E[(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\,(X\beta + \epsilon)] = E[(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}X\beta] + E[(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\epsilon]$$

Si X et β sont fixés, $E[X\beta] = X\beta$, par construction $E[\epsilon] = 0$,

$$E[\hat{\beta}] = \underbrace{(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}X}_{Id}\,\beta = \beta$$

On peut dire que l'estimateur $\hat{\beta}$ est non-biaisé.

T4

Remarquons que

$$\begin{aligned}
\hat{\beta} - \beta &= (X^T X)^{-1} X^T y - \beta \\
&= (X^T X)^{-1} X^T (X\beta + \epsilon) - \beta \\
&= (X^T X)^{-1} X^T X \beta + (X^T X)^{-1} X^T \epsilon - \beta \\
&= \underbrace{(X^T X)^{-1} X^T X}_{I} \beta + (X^T X)^{-1} X^T \epsilon - \beta \\
&= (X^T X)^{-1} X^T \epsilon
\end{aligned}$$

$$X(\hat{\beta} - \beta) = X(X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} \varepsilon$$

On note la matrice $H = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$, on vérifie

$$H^2 = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}} = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}} = H$$

$$H^{\mathrm{T}} = (X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}})^{\mathrm{T}} = (X^{\mathrm{T}})^{\mathrm{T}}((X^{\mathrm{T}}X)^{-1})^{\mathrm{T}}X^{\mathrm{T}} = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}} = H$$

Donc H est une projection orthogonale dans R^2 sur le sous-espace $\mathrm{Vect}(X)$, engendré par les vecteurs colonnes de X.

Notons $B = H\varepsilon$, comme ε est un vecteur gaussien dont les éléments sont i.i.d., centrées et de variance $\sigma > 0$, on obtient

$$\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$$

$$B = H\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 H \boldsymbol{I} H^{\mathrm{T}})$$

Comme $H\boldsymbol{I}H^{\mathrm{T}} = HH = H$,

$$B \sim \mathcal{N}(\mathbf{0}, \sigma^2 H)$$

$$\frac{B}{\sigma} \sim \mathcal{N}(\mathbf{0}, H)$$

Donc, on a

$$\sigma^{-2}(\hat{\beta} - \beta)^{\mathrm{T}} X^{\mathrm{T}} X (\hat{\beta} - \beta) = \sigma^{-2} B^{\mathrm{T}} B = ||\frac{B}{\sigma}||^2$$

On montre que

$$\sigma^{-2}(\hat{\beta} - \beta)^{\mathrm{T}} X^{\mathrm{T}} X (\hat{\beta} - \beta) \sim \chi^2(2)$$

Puis,

$$\begin{aligned}
P(\beta \in \mathcal{E}_\alpha) &= P(\sigma^{-2}(\beta - \hat{\beta})^{\mathrm{T}} X^{\mathrm{T}} X (\beta - \hat{\beta}) \leq q_{\chi^2(2)}(1 - \alpha)) \\
&= P(\sigma^{-2}(\hat{\beta} - \beta)^{\mathrm{T}} X^{\mathrm{T}} X (\hat{\beta} - \beta) \leq q_{\chi^2(2)}(1 - \alpha))
\end{aligned}$$

Selon la définition du quantile, en notant $F_{\chi^2(2)}$ la fonction de répartition d'une variable aléatoire suivant une loi du Chi^2 à n degré de liberté, on a

$$q_{\chi^2(2)}(1 - \alpha) = \inf\{x \in R, F_{\chi^2(2)}(x) \geq 1 - \alpha\}$$

Il est immédiat que

$$P(\beta \in \mathcal{E}_\alpha) = 1 - \alpha$$

T5

D'après le calcul on a

$$(X^n)^{\mathrm{T}}X^n = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}$$

et

$$(X^n)^{\mathrm{T}}\varepsilon^n = \begin{pmatrix} \sum_{i=1}^n \varepsilon_i \\ \sum_{i=1}^n x_i\varepsilon_i \end{pmatrix}$$

On sait que $(x_i)_{i>=1}$ est une suite de variables i.i.d. et suivent la loi uniforme sur $(0,1)$, donc intégrables. On a

$$E[x_1] = \frac{1}{2}$$

D'après la loi des grands nombres,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n x_i = \frac{1}{2}\ p.s.$$

De même, on peut calculer les espérances :

$$E[x_1^2] = \int_R x^2 I_{[0,1]}\,\mathrm{d}x = \int_0^1 x^2\,\mathrm{d}x = \frac{1}{3}$$

$$E[\varepsilon_1] = 0$$

$$E[x_1\varepsilon_1] = E[x_1]E[\varepsilon_1] = 0$$

Puisque $(x_i^2)_{i>=1}$, $(\varepsilon_i)_{i>=1}$ et $(x_i\varepsilon_i)_{i>=1}$ sont aussi i.i.d. et intégrables, appliquons la loi des grands nombres sur chacune de ces suites.

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n x_i^2 = \frac{1}{3}\ p.s.$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \varepsilon_i = 0\ p.s.$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n x_i\varepsilon_i = 0\ p.s.$$

Donc,

$$\lim_{n\to\infty} \frac{1}{n}(X^n)^{\mathrm{T}}X^n = \begin{pmatrix} 1 & \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n x_i \\ \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n x_i & \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n x_i^2 \end{pmatrix} = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix}\ p.s.$$

$$\lim_{n\to\infty} \frac{1}{n}(X^n)^{\mathrm{T}}\varepsilon^n = \begin{pmatrix} \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \varepsilon_i \\ \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n x_i\varepsilon_i \end{pmatrix} = \mathbf{0}\ p.s.$$

Puis, comme la matrice $(X^n)^{\mathrm{T}}X^n$ est inversible, et on a la relation

$$\frac{1}{n}(X^n)^{\mathrm{T}}X^n n((X^n)^{\mathrm{T}}X^n)^{-1} = n((X^n)^{\mathrm{T}}X^n)^{-1}\frac{1}{n}(X^n)^{\mathrm{T}}X^n = \boldsymbol{I}$$

En passant à l'inverse, on obtient

$$\lim_{n\to\infty} (\frac{1}{n}(X^n)^{\mathrm{T}}X^n)^{-1} = \lim_{n\to\infty} n((X^n)^{\mathrm{T}}X^n)^{-1} = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix}^{-1}\ p.s.$$

En les multipliant, on obtient

$$\lim_{n\to\infty} n((X^n)^{\mathrm{T}}X^n)^{-1}\frac{1}{n}(X^n)^{\mathrm{T}}\varepsilon^n = \lim_{n\to\infty}((X^n)^{\mathrm{T}}X^n)^{-1}(X^n)^{\mathrm{T}}\varepsilon^n = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix}^{-1} \mathbf{0} = \mathbf{0}\ p.s.$$

Selon **T4**, on a

$$\hat{\beta}^n - \beta = ((X^n)^{\mathrm{T}}X^n)^{-1}(X^n)^{\mathrm{T}}\varepsilon^n$$

Donc

$$\lim_{n\to\infty}\hat{\beta}^n - \beta = \lim_{n\to\infty}((X^n)^{\mathrm{T}}X^n)^{-1}(X^n)^{\mathrm{T}}\varepsilon^n = \mathbf{0}\ p.s.$$

$$\lim_{n\to\infty}\hat{\beta}^n = \beta\ p.s.$$

T6

Pout tout $1 <= i <= n$, on a

$$y_i = \frac{1}{1+\exp(-\theta_0 - \theta_1 x_i)} + \varepsilon_i$$

$$\varepsilon_i = y_i - \frac{1}{1+\exp(-\theta_0 - \theta_1 x_i)} = y_i - f(\theta, x_i) \sim \mathcal{N}(0, \sigma^2)$$

Quand les données $(x_i, y_i)_{1\le i\le n}$ sont fixées, $y_i - f(\theta, x_i)$ est une fonction de θ, on peut définir la vraisemblance comme

$$p_i(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{|y_i - f(\theta, x_i)|^2}{2\sigma^2}\right)$$

qui correspond à la probabilité pour une variable aléatoire gaussienne centrée de variance $\sigma^2 > 0$ d'être égale à $y_i - f(\theta, x_i)$.

Donc la vraisemblance totale vaut

$$p(\theta) = \prod_{i=1}^{n} p_i(\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{\sum_{i=1}^{n}|y_i - f(\theta, x_i)|^2}{2\sigma^2}\right)$$

C'est une fonction décroissante de $\sum_{i=1}^{n}|y_i - f(\theta, x_i)|^2$, donc on a

$$\arg\max_{\theta\in R^2} p(\theta) = \arg\min_{\theta\in R^2}\sum_{i=1}^{n}|y_i - f(\theta, x_i)|^2 = \arg\min_{\theta\in R^2}\frac{1}{n}\sum_{i=1}^{n}|y_i - f(\theta, x_i)|^2$$

Finalement on obtient

$$\hat{\theta} \in \arg\min_{\theta\in R^2} L(\theta)$$

T7

- Tirage avec remplacement

On note m_i le i-ième indice dans I_m avec $1 \le i \le m$.

Comme I_m est constitué de m indices tirés de manière indépendente et uniforme dans $\{1, ..., n\}$, alors pour tout $1 \le i \le m$, m_i suit une loi uniforme sur $\{1, ..., n\}$.

On obtient

$$\forall i \in N^*, i <= m, E(G_{m_i}) = \frac{1}{n}\sum_{i=1}^{n} G_i = g(\theta)$$

Donc, on a

$$E(g_m(\theta)) = E(\frac{1}{m}\sum_{j\in I_m} G_j) = E(\frac{1}{m}\sum_{i=1}^{m} G_{m_i}) = \frac{1}{m}\sum_{i=1}^{m} E(G_{m_i}) = g(\theta)$$

- Tirage sans remplacement

Supposons m donné, pour une tirage avec remplacement, chaque partie de m éléments de l'ensemble $\{1, ..., n\}$ a une chance de $1/\binom{n}{m}$ d'être tirée.

Notons T_m l'ensemble de tirages possibles, par un dénombrement, on peut déduire que

$$\text{Card}\,(T_m) = \binom{n}{m}$$

et

$$\forall k \in N^*, k \leq n, \sum_{T\in T_m} I_{k\in T} = \binom{n-1}{m-1}$$

Donc, on a

$$E(g_m(\theta)) = \frac{1}{\binom{n}{m}} \sum_{T\in T_m} \left(\frac{1}{m}\sum_{t\in T} G_t\right)$$

En changeant l'ordre des deux sommes, on obtient

$$E(g_m(\theta)) = \frac{1}{\binom{n}{m} m}\sum_{t=1}^{n}\sum_{T\in T_m} I_{t\in T} G_t = \frac{1}{\binom{n}{m} m}\sum_{t=1}^{n}\binom{n-1}{m-1} G_t = \frac{1}{n}\sum_{t=1}^{n} G_t = g(\theta)$$

T8

On note $g(\theta)_0, g(\theta)_1$ les deux éléments du vecteur colonne $g(\theta)$, puisque l'on a montré dans **T7** que

$$\forall i \in N^*, i \leq n, E(G_i) = g(\theta)$$

donc on a

$$E(2(f(\theta, x_i) - y_i)\frac{\partial f(\theta, x_i)}{\partial \theta_0}) = g(\theta)_0$$

$$E(2(f(\theta, x_i) - y_i)\frac{\partial f(\theta, x_i)}{\partial \theta_1}) = g(\theta)_1$$

On écrit $g_m(\theta)$ et $\Sigma(\theta)$ sous forme d'une matrice

$$g_m(\theta) = \frac{1}{m}\sum_{j\in I_m} G_j = \frac{1}{m}\sum_{j\in I_m} 2(f(\theta, x_j) - y_j)\nabla_\theta f(\theta, x_j) = \begin{pmatrix} \frac{1}{m}\sum_{j\in I_m} 2(f(\theta, x_j) - y_j)\frac{\partial f(\theta, x_j)}{\partial \theta_0} \\ \frac{1}{m}\sum_{j\in I_m} 2(f(\theta, x_j) - y_j)\frac{\partial f(\theta, x_j)}{\partial \theta_1} \end{pmatrix}$$

$$\Sigma(\theta) = \frac{1}{n}\sum_{i=1}^{n} \overline{G}_i \overline{G}_i^{\mathrm{T}} = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} 2(f(\theta, x_i) - y_i)\frac{\partial f(\theta, x_i)}{\partial \theta_0} - g(\theta)_0 \\ 2(f(\theta, x_i) - y_i)\frac{\partial f(\theta, x_i)}{\partial \theta_1} - g(\theta)_1 \end{pmatrix} \cdot$$
$$\cdot \begin{pmatrix} 2(f(\theta, x_i) - y_i)\frac{\partial f(\theta, x_i)}{\partial \theta_0} - g(\theta)_0 & 2(f(\theta, x_i) - y_i)\frac{\partial f(\theta, x_i)}{\partial \theta_1} - g(\theta)_1 \end{pmatrix}$$

On utilise les notations du problème précédent.

- Tirage avec remplacement

On calcule

$$
\begin{aligned}
\mathrm{Var}\left(\frac{1}{m}\sum_{j\in I_m} 2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_0}\right) &= \mathrm{Var}\left(\frac{1}{m}\sum_{i=1}^{m} 2(f(\theta,x_{m_i})-y_{m_i})\frac{\partial f(\theta,x_{m_i})}{\partial\theta_0}\right) \\
&= \frac{1}{m^2}\sum_{i=1}^{m} E\left((2(f(\theta,x_{m_i})-y_{m_i})\frac{\partial f(\theta,x_{m_i})}{\partial\theta_0}-g(\theta)_0)^2\right)
\end{aligned}
$$

Pour tout $1 \leq i \leq m$, m_i suit une loi uniforme sur $\{1, ..., n\}$, donc

$$
E((2(f(\theta,x_{m_i})-y_{m_i})\frac{\partial f(\theta,x_{m_i})}{\partial\theta_0}-g(\theta)_0)^2) = \sum_{i=1}^{n}\frac{1}{n}(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}-g(\theta)_0)^2
$$

Donc, on a

$$
\begin{aligned}
(\mathrm{Cov}\,(g_m(\theta)))_{1,1} &= \mathrm{Var}\left(\frac{1}{m}\sum_{j\in I_m} 2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_0}\right) \\
&= \frac{1}{mn}\sum_{i=1}^{n}(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}-g(\theta)_0)^2 \\
&= \frac{1}{m}\,(\Sigma(\theta))_{1,1}
\end{aligned}
$$

Par un raisonnement analogue, on a

$$
(\mathrm{Cov}\,(g_m(\theta)))_{1,2} = \frac{1}{mn}\sum_{i=1}^{n}(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}-g(\theta)_0)(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_1}-g(\theta)_1) = \frac{1}{m}\,(\Sigma(\theta))_{1,2}
$$

$$
(\mathrm{Cov}\,(g_m(\theta)))_{2,1} = \frac{1}{mn}\sum_{i=1}^{n}(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}-g(\theta)_0)(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_1}-g(\theta)_1) = \frac{1}{m}\,(\Sigma(\theta))_{2,1}
$$

$$
\begin{aligned}
(\mathrm{Cov}\,(g_m(\theta)))_{2,2} &= \mathrm{Var}\left(\frac{1}{m}\sum_{j\in I_m} 2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_1}\right) \\
&= \frac{1}{mn}\sum_{i=1}^{n}(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_1}-g(\theta)_1)^2 \\
&= \frac{1}{m}\,(\Sigma(\theta))_{2,2}
\end{aligned}
$$

Finalement, on montre que

$$
\mathrm{Cov}\,(g_m(\theta)) = \frac{1}{m}\Sigma(\theta)
$$

- Tirage sans remplacement
On calcule

$$
\begin{aligned}
(\Sigma(\theta))_{1,1} &= \frac{1}{n}\sum_{i=1}^{n}(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}-g(\theta)_0)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}\right)^2 - \frac{2}{n}g(\theta)_0\sum_{i=1}^{n}\left(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}\right) + g^2(\theta)_0
\end{aligned}
$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}\right)^2 - g^2(\theta)_0$$

$$= \left(\frac{1}{n}-\frac{1}{n^2}\right)\sum_{i=1}^{n}\left(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}\right)^2 - \frac{2}{n^2}\sum_{1\leq ij\leq n}\left(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}\right)$$
$$\left(2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_0}\right)$$

Et

$$\left(\mathrm{Cov}\left(g_m(\theta)\right)\right)_{1,1} = \mathrm{Var}\left(\frac{1}{m}\sum_{j\in I_m}2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_0}\right) = E\left(\left(\frac{1}{m}\sum_{j\in I_m}2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_0}-g^2(\theta)_0\right)\right)$$

$$= E\left(g^2(\theta)_0 - \frac{2}{m}g(\theta)_0\sum_{j\in I_m}2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_0} + \frac{1}{m^2}\left(\sum_{j\in I_m}2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_0}\right)^2\right)$$

$$= g^2(\theta)_0 - \frac{2}{m}g(\theta)_0 E\left(\sum_{j\in I_m}2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_0}\right) + \frac{1}{m^2}E\left(\left(\sum_{j\in I_m}2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_0}\right)^2\right)$$

On a

$$E\left(\sum_{j\in I_m}2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_0}\right) = \sum_{T\in T_m}\frac{1}{\binom{n}{m}}\sum_{t\in T}2(f(\theta,x_t)-y_t)\frac{\partial f(\theta,x_t)}{\partial\theta_0} = n\frac{\binom{n-1}{m-1}}{\binom{n}{m}}g(\theta)_0 = mg(\theta)_0$$

$$E\left(\left(\sum_{j\in I_m}2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_0}\right)^2\right) = \sum_{T\in T_m}\frac{1}{\binom{n}{m}}\left(\sum_{t\in T}2(f(\theta,x_t)-y_t)\frac{\partial f(\theta,x_t)}{\partial\theta_0}\right)^2$$

$$= \frac{\binom{n-1}{m-1}}{\binom{n}{m}}\sum_{i=1}^{n}\left(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}\right)^2$$
$$+ \frac{2\binom{n-2}{m-2}}{\binom{n}{m}}\sum_{1\leq i<j\leq n}\left(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}\right)\left(2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_0}\right)$$

$$= \frac{m}{n}\sum_{i=1}^{n}\left(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}\right)^2$$
$$+ \frac{2m(m-1)}{n(n-1)}\sum_{1\leq i<j\leq n}\left(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}\right)\left(2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_0}\right)$$

Donc

$$
\begin{aligned}
(\mathrm{Cov}\,(g_m(\theta)))_{1,1} = -g^2(\theta)_0 + \frac{1}{mn}\sum_{i=1}^{n}\left(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}\right)^2 \\
+ \frac{2(m-1)}{mn(n-1)}\sum_{1\le i<j\le n}\left(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}\right)\left(2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_0}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \left(\frac{1}{mn}-\frac{1}{n^2}\right)\sum_{i=1}^{n}\left(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}\right)^2 \\
&+ \left(\frac{2(m-1)}{mn(n-1)}-\frac{2}{n^2}\right)\sum_{1\le i<j\le n}\left(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}\right)\left(2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_0}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{n-m}{n^2m}\sum_{i=1}^{n}\left(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}\right)^2 \\
&+ \frac{2(m-n)}{n^2(n-1)m}\sum_{1\le i<j\le n}\left(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}\right)\left(2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_0}\right)
\end{aligned}
$$

Par l'équivalence des coefficients, on a

$$
\begin{aligned}
&\frac{1}{m}\frac{n-m}{n-1}\left(\Sigma(\theta)\right)_{1,1} \\
&= \frac{n-m}{n^2m}\sum_{i=1}^{n}\left(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}\right)^2 \\
&- \frac{2(n-m)}{n^2(n-1)m}\sum_{1\le i<j\le n}\left(2(f(\theta,x_i)-y_i)\frac{\partial f(\theta,x_i)}{\partial\theta_0}\right)\left(2(f(\theta,x_j)-y_j)\frac{\partial f(\theta,x_j)}{\partial\theta_0}\right) \\
&= (\mathrm{Cov}\,(g_m(\theta)))_{1,1}
\end{aligned}
$$

Par un raisonnement similaire, on montre que

$$\frac{1}{m}\frac{n-m}{n-1}\left(\Sigma(\theta)\right)_{1,2} = (\mathrm{Cov}\,(g_m(\theta)))_{1,2}$$

$$\frac{1}{m}\frac{n-m}{n-1}\left(\Sigma(\theta)\right)_{2,1} = (\mathrm{Cov}\,(g_m(\theta)))_{2,1}$$

$$\frac{1}{m}\frac{n-m}{n-1}\left(\Sigma(\theta)\right)_{2,2} = (\mathrm{Cov}\,(g_m(\theta)))_{2,2}$$

Finalement, on obtient

$$\mathrm{Cov}\,(g_m(\theta)) = \frac{1}{m}\frac{n-m}{n-1}\Sigma(\theta)$$

Puisque $\frac{n-m}{n-1} \le 1$, la matrice de covariance dans le cas sans remplacement est plus petit en termes, donc il est plus préférable.