Statistique Asymptotique

Université Hassiba Benbouali de Chlef

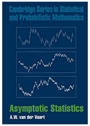
Plan du Cours

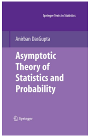
- Modes de convergence
- Méthode Delta
- M et Z-estimateurs
- Méthode Delta fonctionnelle
- U-statistiques
- Processus empirique

Références

- Van der Vaart, A. W. (2000). Asymptotic statistics. Cambridge university press.

- DasGupta, A. (2008). Asymptotic theory of statistics and probability. Springer.





Méthode Delta

Si au lieu d'estimer directement θ, on veut estimer une fonction de θ, on sait que $\phi(\widehat{\theta}_n)$ est l'estimateur de maximum de vraisemblance de $\phi(\theta)$.

- On souhaite déduire la normalité asymptotique pour un estimateur $\phi(\widehat{\theta}_n)$ fonction d'un estimateur $\widehat{\theta}_n$ de θ asymptotiquement normal.
- Pour cela, si ϕ est une fonction suffisamment régulière, une idée naturelle est d'exploiter un développement de Taylor pour ramener l'étude du comportement de $\phi(\widehat{\theta}_n)$ à celle de $\widehat{\theta}_n$.

Méthode Delta (suite)

Les propriétés de cet estimateur sont données par le théorème suivant. Il porte le nom de méthode Delta.

Théorème 1.1

Si :

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \sigma^2(\theta)\right)$$

alors

$$\sqrt{n}\left(\phi(\widehat{\theta}_n) - \phi(\theta)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \left[\phi'(\theta)\right]^2 \sigma^2(\theta)\right)$$

pourvu que la dérivée $\phi'(\theta)$ existe et soit non nulle.

Preuve

On sait que,

$$\widehat{\theta}_n \xrightarrow[n\to+\infty]{P} \theta.$$

Dire que ϕ est dérivable en θ signifie qu'on peut écrire

$$\phi(\widehat{\theta}_n) = \phi(\theta) + (\widehat{\theta}_n - \theta)\left(\phi'(\theta) + o_P(1)\right).$$

C'est-à-dire,

$$\sqrt{n}\left(\phi(\widehat{\theta}_n) - \phi(\theta)\right) = \sqrt{n}\left(\widehat{\theta}_n - \theta\right)\left(\phi'(\theta) + o_P(1)\right).$$

Finalement, on applique le Théorème de Slutsky.

Exemple

Le modèle de Poisson

L'estimateur du maximum de vraisemblance est $\widehat{\theta}_n = \overline{X}_n$, par le TCL

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

L'estimateur de $\mathrm{e}^{-\theta}$ est $\mathrm{e}^{-\widehat{\theta}_n}$. La méthode Delta permet d'écrire que

$$\sqrt{n}\left(\phi(\widehat{\theta}_n) - \phi(\theta)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \theta\mathrm{e}^{-2\theta}\right)$$

Développements à l'ordre supérieur

Le résultat présenté par le Théorème 1.1 repose sur un développement de Taylor à l'ordre 1. Cependant, lorsque $\phi'(\theta)$ est nulle, la loi limite est dégénérée en 0. Il est alors intéressant de pousser le développement à un ordre supérieur.

Développements à l'ordre supérieur (suite)

Théorème 1.2

Soit $\widehat{\theta}_n$ un estimateur tel que

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \sigma^2(\theta)\right).$$

Soit ϕ une application k-différentiable en θ ($k \geq 1$) avec $\phi^{(k)}(\theta) \neq 0$ et $\phi^{(j)}(\theta) = 0$ pour $j < k$. Alors

$$\left(\sqrt{n}\right)^k \left(\phi(\widehat{\theta}_n) - \phi(\theta)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \frac{1}{k!}\phi^{(k)}(\theta)\left[\mathcal{N}\left(0, \sigma^2(\theta)\right)\right]^k.$$

Exemple

Soit $X_1, \dots, X_n$ n copies indépendantes de la loi $\mathcal{B}(p)$ $(0 < p < 1)$. Par le TCL

$$\sqrt{n}(\overline{X}_n - p) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}\left(0, p(1-p)\right).$$

Un estimateur de la variance $p(1-p)$ est $\overline{X}_n(1-\overline{X}_n)$. Il s'écrit comme $\phi(\overline{X}_n)$ avec $\phi(x) = x(1-x)$.
On a $\phi'(x) = 1 - 2x$. Si $p \neq \frac{1}{2}$, alors $\phi'(p) \neq 0$. D'après le Théorème 1.1, la distribution asymptotique de cet estimateur est

$$\sqrt{n}\left(\phi(\overline{X}_n) - \phi(p)\right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}\left(0, p(1-p)(1-2p)^2\right)$$

Exemple (suite)

En revanche $\phi'\left(\frac{1}{2}\right) = 0$. Cependant $\phi''\left(\frac{1}{2}\right) = -2$, d'où d'après le Théorème 1.2

$$\begin{aligned} n\left(\phi(\overline{X}_n) - \left(\frac{1}{2}\right)\right) & \xrightarrow[n\to+\infty]{\mathcal{L}} \frac{1}{2!}(-2)\left[\mathcal{N}\left(0, \frac{1}{4}\right)\right]^2 \\ & \xrightarrow[n\to+\infty]{\mathcal{L}} -\frac{1}{4}\chi_1^2 \end{aligned}$$

Stabilisation de la variance

- Un choix judicieux de ϕ nous donnera pour la loi Normale limite une variance qui ne dépendra plus d'un paramètre.
- Il suffit de trouver ϕ telle que $\left[\phi'(\theta)\right]^2 \sigma^2(\theta) \equiv 1$. C'est-à-dire $\phi = \int \frac{1}{\sigma}$.

Remarque

Remarque

Puisque $\sigma(\theta) > 0$, ceci implique que ϕ, en plus d'être continue, est strictement croissante, donc inversible. Il vient

$$\mathbb{P}\left(\left|\phi\left(\widehat{\theta}_n\right) - \phi(\theta)\right| \leq \frac{z_{1-\alpha/2}}{\sqrt{n}}\right) =$$
$$\mathbb{P}\left(\phi^{-1}\left(\phi\left(\widehat{\theta}_n\right) - \frac{z_{1-\alpha/2}}{\sqrt{n}}\right) \leq \theta \leq \phi^{-1}\left(\phi\left(\widehat{\theta}_n\right) + \frac{z_{1-\alpha/2}}{\sqrt{n}}\right)\right).$$

qui correspond donc à un intervalle de confiance asymptotique de niveau $(1-\alpha)$.

Le modèle de Poisson

L'estimateur du maximum de vraisemblance est $\widehat{\theta}_n = \overline{X}_n$, par le TCL

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

Le principe du plug-in donne l'intervalle de confiance suivant :

$$\left[\overline{X}_n - \frac{\sqrt{\overline{X}_n}}{\sqrt{n}} z_{1-\alpha/2} \; ; \; \overline{X}_n + \frac{\sqrt{\overline{X}_n}}{\sqrt{n}} z_{1-\alpha/2}\right]$$

Appliquons la méthode de stabilisation de la variance : on cherche ϕ telle que

$$\phi'(\theta) = \frac{1}{\theta} \Longleftrightarrow \phi(\theta) = 2\sqrt{\theta}$$

Le modèle de Poisson (suite)

ce qui donne donc

$$\sqrt{n}\left(2\sqrt{\overline{X}_n} - 2\sqrt{\theta}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0,1\right).$$

Ainsi la variance de la loi Normale limite ne dépend plus du paramètre inconnu θ.
D'où les intervalles de confiance asymptotiques :

$$\mathbb{P}\left(\left|\sqrt{n}\left(2\sqrt{\overline{X}_n} - 2\sqrt{\theta}\right)\right| \leq z_{1-\alpha/2}\right) \underset{n\to\infty}{\longrightarrow} 1-\alpha$$

$$\mathbb{P}\left(2\sqrt{\overline{X}_n} - \frac{z_{1-\alpha/2}}{\sqrt{n}} \leq 2\sqrt{\theta} \leq 2\sqrt{\overline{X}_n} + \frac{z_{1-\alpha/2}}{\sqrt{n}}\right) \underset{n\to\infty}{\longrightarrow} 1-\alpha$$

Le modèle de Poisson (suite)

$$\mathbb{P}\left(\left(\sqrt{\overline{X}_n} - \frac{z_{1-\alpha/2}}{2\sqrt{n}}\right)^2 \leq \theta \leq \left(\sqrt{\overline{X}_n} + \frac{z_{1-\alpha/2}}{2\sqrt{n}}\right)^2\right) \underset{n\to\infty}{\longrightarrow} 1-\alpha$$

C'est-à-dire :

$$\left[\overline{X}_n - \frac{\sqrt{\overline{X}_n}}{\sqrt{n}} z_{1-\alpha/2} + \frac{z_{1-\alpha/2}^2}{4n} \ ; \ \overline{X}_n + \frac{\sqrt{\overline{X}_n}}{\sqrt{n}} z_{1-\alpha/2} + \frac{z_{1-\alpha/2}^2}{4n}\right]$$

La méthode de stabilisation de la variance a donc eu pour effet de translater les intervalles de confiance de $\frac{z_{1-\alpha/2}^2}{4n}$.

Vecteurs aléatoires

Soit $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} \in \mathbb{R}^d$ un vecteur aléatoire dont les composantes sont de carré intégrable.

- Le vecteur moyenne de X est défini par : $\mathbb{E}[X] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_d] \end{pmatrix}$
- et sa matrice de covariance par :

$$\begin{aligned}
\mathrm{Var}[X] &= \mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top\right] \\
&= \begin{pmatrix} \mathrm{Var}\left[X_1\right] & \mathrm{Cov}\left(X_1, X_2\right) & \cdots & \mathrm{Cov}\left(X_1, X_d\right) \\ \mathrm{Cov}\left(X_2, X_1\right) & \mathrm{Var}\left[X_2\right] & \cdots & \mathrm{Cov}\left(X_2, X_d\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}\left(X_d, X_1\right) & \mathrm{Cov}\left(X_d, X_2\right) & \cdots & \mathrm{Var}\left[X_d\right] \end{pmatrix}
\end{aligned}$$

Remarque

Remarque

- Pour tout $1 \leq i, j \leq d$, la covariance entre X_i et X_j est donnée par

$$\begin{aligned}\mathrm{Cov}\left(X_i, X_j\right) &= \mathbb{E}\left[\left(X_i - \mathbb{E}\left[X_i\right]\right)\left(X_j - \mathbb{E}\left[X_j\right]\right)\right] \\ &= \mathbb{E}\left[X_i X_j\right] - \mathbb{E}\left[X_i\right]\mathbb{E}\left[X_j\right]\end{aligned}$$

- Il est facile de voir que l'on a $\mathrm{Cov}\left(X_i, X_i\right) = \mathrm{Var}\left[X_i\right]$.
- De plus, si les variables X_i et X_j sont indépendantes on a $\mathrm{Cov}\left(X_i, X_j\right) = 0$.
- En général, la réciproque est fausse, sauf pour des vecteurs gaussiens, comme nous le verrons plus loin.

Théorème 1.3

Si X est un vecteur (colonne) aléatoire de $\mathbb{R}^d$ de vecteur moyenne $\boldsymbol{m}$ et de matrice de covariance $\boldsymbol{\Sigma}$ Alors si A est une matrice réelle $k \times d$, le vecteur aléatoire AX de $\mathbb{R}^k$ a pour vecteur moyenne $A\boldsymbol{m}$ et pour matrice de covariance $A\boldsymbol{\Sigma}A^\top$.

Preuve

C'est une simple conséquence de la linéarité de l'espérance. Pour la moyenne on a :

$$\mathbb{E}[AX] = A\mathbb{E}[X] = A\boldsymbol{m}$$

et pour la matrice de covariance

$$\begin{aligned}
\mathrm{Var}[AX] &= \mathbb{E}\left[(AX - \mathbb{E}[AX])(AX - \mathbb{E}[AX])^\top\right] \\
&= \mathbb{E}\left[A(X - \mathbb{E}[X])\,(A(X - \mathbb{E}[X]))^\top\right] \\
&= A\mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top\right] A^\top = A\boldsymbol{\Sigma}A^\top
\end{aligned}$$

Densité gaussienne en dimension d

Définition 1.4

Un vecteur aléatoire de $\mathbb{R}^d$ est un vecteur aléatoire gaussien si et seulement si toute combinaison linéaire de ses composantes est une variable aléatoire réelle gaussienne, i.e. :

$$\forall a \in \mathbb{R}^d \qquad a^\top X \sim \mathcal{N}(m, \sigma^2).$$

Soit $X \sim \mathcal{N}_d(\boldsymbol{m}, \boldsymbol{\Sigma})$ un vecteur gaussien en dimension d. Si $\boldsymbol{\Sigma}$ est inversible, la densité de X est donc donnée par

$$f(\mathbf{x}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}(2\pi)^{\frac{d}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{m})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{m})\right),$$

Théorème 1.5

Pour tout vecteur gaussien X de $\mathbb{R}^d$, les trois propriétés suivantes sont équivalentes :

1. *Les composantes $X_1, \ldots, X_d$ sont mutuellement indépendantes.*
2. *Les composantes $X_1, \ldots, X_d$ sont deux à deux indépendantes.*
3. *La matrice de covariance $\mathbf{\Sigma}$ de X est diagonale.*

Théorème central limite multidimensionnel

Théorème 1.6

Soient $\mathbf{X_1}, \ldots, \mathbf{X_n}$ *des vecteurs aléatoires i.i.d. de* $\mathbb{R}^d$ *d'espérance* $\boldsymbol{m}$ *et de matrice de covariance* $\boldsymbol{\Sigma} = \mathbb{E}\left[(\mathbf{X} - \boldsymbol{m})(\mathbf{X} - \boldsymbol{m})^\top\right]$. *Alors*

$$\sqrt{n}\left(\overline{\mathbf{X}}_\mathbf{n} - \boldsymbol{m}\right) \xrightarrow{\mathcal{L}} \mathcal{N}_d\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$$

Méthode Delta multivariée

La matrice de gradient

$$\nabla \boldsymbol{\phi}(\boldsymbol{a}) = \left. \begin{pmatrix} \frac{\partial \phi_1(\boldsymbol{x})}{\partial x_1} & \cdots & \frac{\partial \phi_k(\boldsymbol{x})}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial \phi_1(\boldsymbol{x})}{\partial x_d} & \cdots & \frac{\partial \phi_k(\boldsymbol{x})}{\partial x_d} \end{pmatrix} \right|_{\boldsymbol{x}=\boldsymbol{a}}$$

Par définition, la matrice de gradient satisfait la formule de Taylor du premier ordre

$$\boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{\phi}(\boldsymbol{a}) + \nabla \boldsymbol{\phi}(\boldsymbol{a})^\top (\boldsymbol{x} - \boldsymbol{a}) + \boldsymbol{r}(\boldsymbol{x}, \boldsymbol{a})$$

avec $\frac{\boldsymbol{r}(\boldsymbol{x},\boldsymbol{a})}{\|\boldsymbol{x}-\boldsymbol{a}\|} \longrightarrow 0$ quand $\boldsymbol{x} \longrightarrow \boldsymbol{a}$.

Méthode Delta multivariée

Théorème 1.7

Soit $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}$ une suite de vecteurs aléatoires dans $\mathbb{R}^d$ satisfaisant $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} \mathcal{N}_d\left(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right)$. Soit $\phi \ : \ \mathbb{R}^d \longrightarrow \mathbb{R}^k$ différentiable en $\boldsymbol{\theta}$. Alors

$$\sqrt{n}\left(\phi(\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}) - \phi(\boldsymbol{\theta})\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}_k\left(0, \left[\nabla\phi(\boldsymbol{\theta})\right]^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})\nabla\phi(\boldsymbol{\theta})\right)$$

Exemple : La variance empirique

Soit $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} = \begin{pmatrix} \overline{X}_n \\ \overline{X^2}_n \end{pmatrix}$ et $\boldsymbol{\theta} = \begin{pmatrix} \mathbb{E}[X] \\ \mathbb{E}[X^2] \end{pmatrix}$. Par le TCL multivarié, on a

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} \mathcal{N}_2\left(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right)$$

avec

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \mathrm{Var}[X] & \mathrm{Cov}(X, X^2) \\ \mathrm{Cov}(X^2, X) & \mathrm{Var}[X^2] \end{pmatrix}.$$

La variance empirique $S_n^2 = \overline{X^2}_n - \overline{X}_n^2$ s'écrit comme $\boldsymbol{\phi}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}})$ avec $\boldsymbol{\phi}(\boldsymbol{x}) = x_2 - x_1^2$.

On a $\nabla\boldsymbol{\phi}(\boldsymbol{x}) = \begin{pmatrix} -2x_1 \\ 1 \end{pmatrix}$. Donc

$$[\nabla\boldsymbol{\phi}(\boldsymbol{\theta})]^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})\nabla\boldsymbol{\phi}(\boldsymbol{\theta}) = \begin{pmatrix} -2\mathbb{E}[X] & 1 \end{pmatrix} \begin{pmatrix} \mathrm{Var}[X] & \mathrm{Cov}(X, X^2) \\ \mathrm{Cov}(X^2, X) & \mathrm{Var}[X^2] \end{pmatrix} \begin{pmatrix} -2\mathbb{E}[X] \\ 1 \end{pmatrix}$$

Exemple : La variance empirique (suite)

Si on suppose que $\mathbb{E}[X] = 0$ et $\mathrm{Var}[X] = \sigma^2$, alors

$$\sqrt{n}\left(S_n^2 - \sigma^2\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \mathbb{E}[X^4] - \sigma^4\right).$$

Méthode des moments

Soit $(X_1, \ldots, X_n)$ un échantillon de variables aléatoires i.i.d. de loi $\mathbb{P}_\theta$, avec $\theta \in \Theta$ intervalle de $\mathbb{R}$, et ϕ un $\mathcal{C}^1$ -difféomorphisme de Θ dans $\phi(\Theta)$. Si $\widehat{\phi}_n = \widehat{\phi}_n\left(X_1, \ldots, X_n\right)$ est un estimateur convergent de $\phi(\theta)$ et θ un point intérieur à Θ, alors $\widehat{\theta}_n = \phi^{-1}\left(\widehat{\phi}_n\right)$ est défini avec une probabilité qui tend vers 1 lorsque $n \to \infty$ et

$$\widehat{\theta}_n \xrightarrow[n\to+\infty]{P} \theta$$

De plus, si

$$\sqrt{n}\left(\widehat{\phi}_n - \phi(\theta)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \sigma^2(\theta)\right),$$

Méthode des moments (suite)

alors

$$\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2(\theta)}{\left[\phi'(\theta)\right]^2}\right).$$

Exemple