Statistique (MA101) Cours 2

ENSTA 1ère année

Christine Keribin

christine.keribin@math.u-psud.fr

Laboratoire de Mathématiques
Université Paris-Sud

2017-2018

Propriétés asymptotiques
Consistance
Loi asymptotique
Construction d'estimateurs
Moments
EMV

Résumé Cours 1 : modèle stat. et estimateur

- Données observées : $x = (x_1, \ldots, x_n)$.
- Modèle : x est une réalisation de $X = (X_1, \ldots, X_n)$, n variables aléatoires dont on modélise la loi.
- Un n-échantillon X est i.i.d. ssi
  - $\hookrightarrow$ les X_i sont indépendantes
  - $\hookrightarrow$ les X_i ont la même loi marginale $\mathbb{P}$

  La loi de l'échantillon est la loi produit $\mathbb{P}^{\otimes n}$

$$X_i \sim \mathbb{P}; \quad X_1, \ldots, X_n \sim \mathbb{P}^{\otimes n}$$

- Modèle paramétrique : $\mathbb{P}$ appartient à une famille de lois $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}$ paramétrée par un paramètre θ de dimension finie
- Estimateur de $\nu_n(\theta)$, variable aléatoire $\widehat{\nu}(X)$ définie à partir de l'échantillon.
  Propriétés : biais, variance, risque

Sommaire

Propriétés asymptotiques
Consistance
Loi asymptotique

Construction d'estimateurs
Moments
EMV

Rappels Convergence (cours de probabilité)

Soit X_n une suite de v.a. et X une v.a.. On étudie son comportement quand $n \to \infty$

- $X_n \xrightarrow{\mathcal{L}} X$: la suite (X_n) converge en loi vers la v.a. X si
$$\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$$
pour tout x où la fct de répartition de X est continue.
- $X_n \xrightarrow{\mathcal{P}} X$: La suite X_n converge en probabilité vers la v.a. X si
$$\forall \varepsilon, \lim_{n \to \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$
- $X_n \xrightarrow{p.s.} X$: la suite X_n converge presque sûrement vers la v.a. X si
$$\mathbb{P}(\lim_{n \to \infty} |X_n - X| = 0) = 1$$
- $X_n \xrightarrow{L^2} X$: La suite X_n converge en moyenne quadratique vers la v.a. X si $\mathbb{E}[(X_n - X)^2] \to 0$

Propriétés

- cvg ps $\Rightarrow$ cvg en proba $\Rightarrow$ cvg en loi
- cvg L2 $\Rightarrow$ cvg en proba
- Soit c une constante réelle. $X_n \xrightarrow{\mathcal{L}} c \Leftrightarrow X_n \xrightarrow{\mathcal{P}} c$

Lemme (Lemme de Slutsky)

Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} c$ où c est une constante, alors $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, c)$. En appliquant cette convergence jointe à une fonction continue de (x, y), on a en particulier

(i) $X_n + Y_n \xrightarrow{\mathcal{L}} X + c$

(ii) $X_n Y_n \xrightarrow{\mathcal{L}} cX$

(iii) $X_n / Y_n \xrightarrow{\mathcal{L}} X / c$

Consistance d'un estimateur

Soit l'estimateur $\widehat{\nu}_n$ de $\nu(\theta)$

- $\widehat{\nu}_n$ est consistant ssi $\widehat{\nu}_n$ tend en probabilité vers $\nu(\theta)$ quand $n \to \infty$:

$$\forall \theta \in \Theta, \forall \epsilon, \lim_{n \to \infty} \mathbb{P}_\theta(|\widehat{\nu}_n - \nu| > \epsilon) = 0$$

- $\widehat{\nu}_n$ est fortement consistant ssi $\forall \theta \in \Theta$,

$$\mathbb{P}_\theta(\lim_{n \to \infty} |\widehat{\nu}_n - \nu| = 0) = 1$$

Outils pour montrer la convergence

Théorème (Inégalité de Bienaymé-Tchebychev)

Soit T une v.a. telle que $\mathbb{E}(T^2) < +\infty$. Alors,

$$\forall t > 0, \ \ \mathbb{P}(\{|T - \mathbb{E}(T)| > t\}) \leq \frac{\text{Var}(T)}{t^2}$$

Rem : si le risque quadratique tend vers 0, alors l'estimateur est consistant

Outils pour montrer la convergence

Théorème (Loi des grands nombres)

Soit $X_1, \ldots, X_n$ un n-échantillon i.i.d. de loi $\mathbb{P}_\theta$ intégrable, d'espérance μ, la moyenne empirique $\bar{X}$ satisfait les propriétés suivantes :

- *C'est un estimateur sans biais de μ :*

$$B_\theta(\bar{X}, \mu) = \mathbb{E}_\theta(\bar{X}) - \mu = 0$$

- *Si $\mathbb{P}_\theta$ a une variance σ^2 finie, la variance de $\bar{X}$ est*

$$Var_\theta(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^{n} Var_\theta(X_i) = \frac{\sigma^2}{n},$$

et $\bar{X}$ converge p.s., la suite $\bar{X}$ est fortement consistante.

Outils pour montrer la convergence

Théorème (Loi des grands nombres de Kolmogorov)

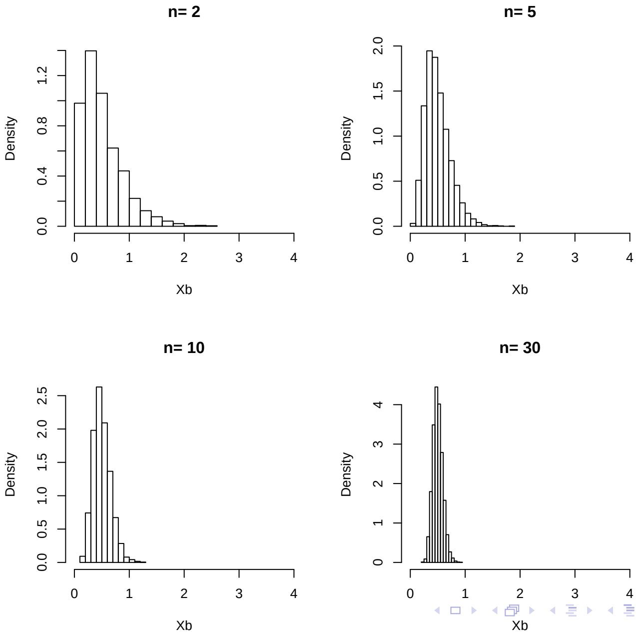
Soit $X_1, \ldots, X_n$ un n-échantillon i.i.d. de loi $\mathbb{P}_\theta$, tel que $\mathbb{E}(|g(X_1)|)$ soit fini. Alors l'estimateur $\widehat{\nu}_n$

$$\widehat{\nu}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i) \to \mathbb{E}_\theta[g(X_1)] = \nu(\theta) \ \ ps$$

est fortement consistant pour estimer $\nu(\theta)$.

La LGN est le premier théorème fondamental en statistique

Ex : loi exponentielle d'espérance 0.5



Outils pour montrer la convergence

Proposition (Continuité pour la convergence)

Soit g une fonction continue.
Si $\widehat{\nu}$ est un estimateur (fortement) consistant de $\nu(\theta)$, alors $g(\widehat{\nu})$ est un estimateur (fortement) consistant de $g(\nu(\theta))$.

Deuxième théorème fondamental

Théorème (de limite centrale)

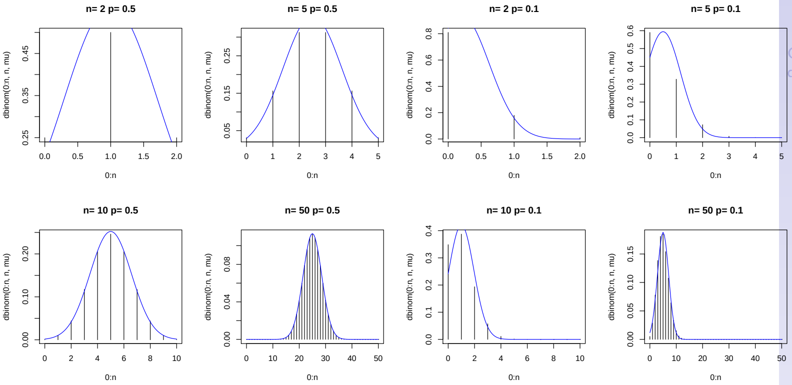
Soit $\{X_n\}$ une suite de variables aléatoires i.i.d. admettant une espérance μ et une variance $\sigma^2 > 0$ finie. Alors, la suite des variables $W_n = \sqrt{n}(\bar{X}_n - \mu)$ converge en loi vers la v.a. $\mathcal{N}(0, \sigma^2)$ quand $n \to \infty$

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

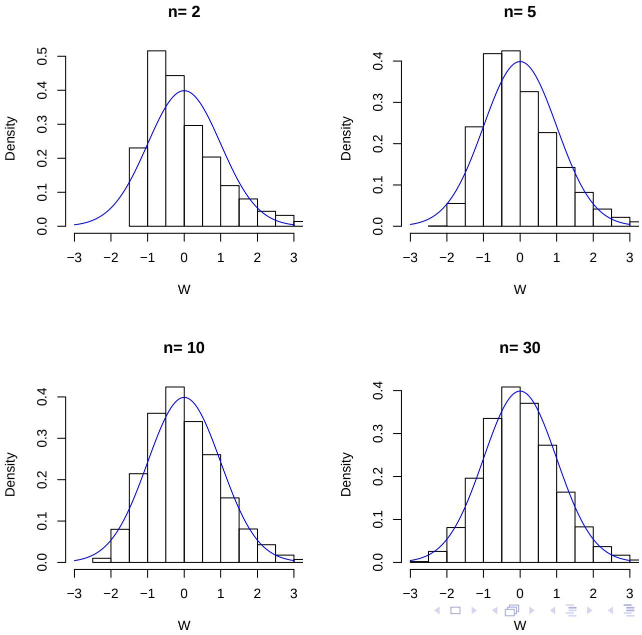
Rem :

- On a aussi : $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$
- TLC, TCL, CLT ...
- Lemme de l'application continue

Exemple : approx. loi Binomiale



Exemple : loi exponentielle



Eléments de preuve TCL

- On pose $Z_k = X_k - \mu$ de fonction caract. $\varphi_z(t) = \mathbb{E}(e^{itZ_k})$;
- $W_n = \sum_{k=1}^n Z_k/\sqrt{n}$, les Z_k sont iid, donc

$$\varphi_{w_n}(t) = \left[\varphi_{z_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

- Pour tt $x \in \mathbb{R}$, $\left|e^{ix} - 1 - ix + \frac{x^2}{2}\right| \leq \min\left(\frac{|x|^3}{6}, x^2\right)$, d'où

$$e^{i\frac{t}{\sqrt{n}}Z_1} = 1 + i\frac{t}{\sqrt{n}}Z_1 - \frac{t^2}{2n}Z_1^2 + h_n(Z_1)$$

avec $|h_n(Z_1)| \leq \frac{t^2}{n}\min\left(\frac{t|Z_1|^3}{6\sqrt{n}}, Z_1^2\right) \leq \frac{t^2}{n}Z_1^2$ donc $h_n(Z_1)$ intégrable et

$$\varphi_{y_1}\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2\sigma^2}{2n} + \mathbb{E}[h_n(Z_1)]$$

où $\lim_{n\to\infty} n\mathbb{E}[h_n(Z_1)] = 0$ (cvg dominée)

Eléments de preuve TCL

- Pour n assez grand tel que $t^2\sigma^2/n < 1$, on a

$$\left|\varphi_{Z_1}\left(\frac{t}{\sqrt{n}}\right)^n - \left(1 - \frac{t^2\sigma^2}{2n}\right)^n\right| \leq \sum_{k=1}^{n} \underbrace{\left|\varphi_{Z_1}\left(\frac{t}{\sqrt{n}}\right) - 1 + \frac{t^2\sigma^2}{2n}\right|}_{|\mathbb{E}(h_n(Z_1))|}$$

- $$\left|\varphi_{Z_1}\left(\frac{t}{\sqrt{n}}\right)^n - e^{-\frac{t^2\sigma^2}{2}}\right| \leq n|\mathbb{E}(h_n(Z_1))| + \left|\left(1 - \frac{t^2\sigma^2}{2n}\right)^n - e^{-\frac{t^2\sigma^2}{2}}\right|$$

- $\left(1 - \frac{t^2\sigma^2}{2n}\right)^n = e^{n\log(1-\frac{t^2\sigma^2}{n})} \to e^{-\frac{t^2\sigma^2}{2}}$
- pour tt $t \in \mathbb{R}$

$$\lim_{n\to\infty} \varphi_{W_n}(t) = e^{-\frac{t^2\sigma^2}{2}} = \varphi_{\mathcal{N}(0,\sigma^2)}(t)$$

Application

La loi de l'estimateur empirique de l'espérance d'une loi quelconque de moment d'ordre 2 fini peut être approximée par une loi gaussienne

$$\bar{X} \overset{appr}{\sim} \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

Exemple si $X \sim \mathcal{B}(\theta)$, si $n\theta > 5$ et $n(1-\theta) > 5$

- $\bar{X} \overset{appr}{\sim} \mathcal{N}\left(\theta, \frac{\theta(1-\theta)}{n}\right)$
- $n\bar{X} \overset{appr}{\sim} \mathcal{N}(n\theta, n\theta(1-\theta))$
- $\sqrt{n}\frac{\bar{X}-\theta}{\sqrt{\widehat{\theta}(1-\widehat{\theta})}} \overset{appr}{\sim} \mathcal{N}(0,1)$

Utilisation...

Normalité asymptotique

Définition

Si un estimateur $\widehat{\nu}_n$ de $\nu \in \mathbb{R}^p$ de variance $\mathrm{Var}(\widehat{\nu}_n) = V_n$ *a un comportement asymptotiquement normal si*

$$V_n^{-1/2}(\widehat{\nu}_n - \nu) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, Id_p)$$

Si $nV_n \to V_0$ où $V_0 > 0$ est finie, on dit que la vitesse de l'estimateur est en $\sqrt{n}$

- Un estimateur est d'autant meilleur que sa vitesse de convergence est rapide et sa loi limite concentrée autour de 0.
- Il existe des estimateurs non asymptotiquement normaux

Delta-méthode

Proposition

Si $\nu(\theta)$ est une fonction différentiable de $\theta \in \mathbb{R}^p$ et $\widehat{\theta}$ un estimateur asympt. normal

$$V_n^{-1/2}(\widehat{\theta} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, Id_p)$$

alors $\nu(\widehat{\theta})$ est un estimateur asympt. normal de $\nu(\theta)$

$$(D_\nu V_n D'_\nu)^{-1/2}(\widehat{\nu}_n - \nu(\theta)) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, Id_p)$$

avec $D_\theta = \begin{pmatrix} \partial\nu(\theta)/\partial\theta_1 & \dots & \partial\nu(\theta)/\partial\theta_p \end{pmatrix}$

Sommaire

Propriétés asymptotiques
Consistance
Loi asymptotique

Construction d'estimateurs
Moments
EMV

Méthode des moments

Soit $X_1, \ldots, X_n \sim \mathbb{P}_\theta$.

- Les moments de $\mathbb{P}_\theta$ dépendent de θ. Moment d'ordre k :
$$\mathbb{E}(X_1^k) = m_k(\theta)$$
- On définit le moment empirique d'ordre k
$$\widehat{m}_k(\theta) = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$
fortement consistant si $\mathbb{E}(|X^k|)$ existe.
- Un estimateur $\widehat{\theta}_n$ de θ obtenu par la méthode des moments est solution de
$$\begin{array}{ccc} m_1(\widehat{\theta}) & = & \widehat{m}_1 \\ \vdots & \vdots & \vdots \\ m_p(\widehat{\theta}) & = & \widehat{m}_p. \end{array}$$

Exemple : estimateur empirique de la variance

Méthode des moments

Extension :
Soit $X_1, \ldots, X_n \sim \mathbb{P}_\theta$ et $\nu(\theta) = \phi(m_1(\theta), \ldots, m_k(\theta))$ où $m_k(\theta) = \mathbb{E}_\theta(g_k(X_1))$.
Un estimateur des moments de $\nu(\theta)$ est

$$\widehat{\nu}_n = \phi\left(\frac{1}{n}\sum_{i=1}^{n} g_1(X_i), \ldots, \frac{1}{n}\sum_{i=1}^{n} g_k(X_i)\right)$$

On remplace les espérances par leur version empirique. Si $\mathbb{E}_\theta(|g_k(X_1)|)$ est fini pour tout k, et ϕ est continue, alors la méthode des moments est consistante.

Méthode du maximum de vraisemblance

Cadre du modèle paramétrique dominé : les lois $(\mathbb{P}_\theta)_{\theta\in\Theta}$ admettent une densité $f_\theta(x)$ par rapport à une mesure commune ξ (mesure de Lebesgue, mesure de comptage)

$$\forall A \in \mathcal{A}, \;\; \mathbb{P}_\theta(A) = \int_A f_\theta(x)\, d\xi(x)$$

Définition

Dans un modèle paramétrique dominé, on appelle vraisemblance d'une réalisation $(x_1, \ldots, x_n)$ du n-échantillon, la fonction de θ :

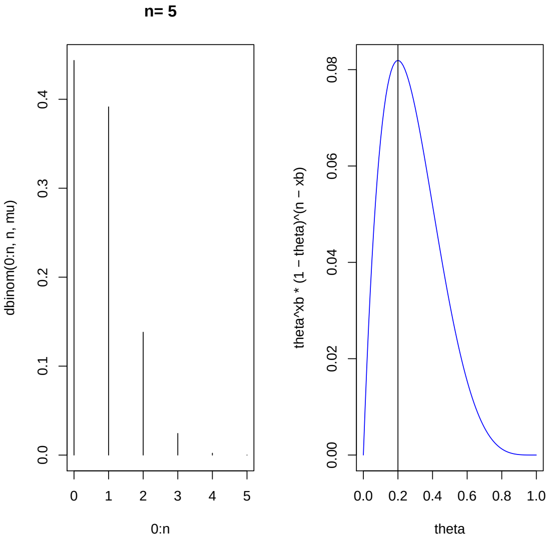
$$\theta \mapsto L(\theta; x_1, \ldots, x_n) = f_\theta(x_1, \ldots, x_n)$$

Pour un échantillon i.i.d. : $L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^n f_\theta(x_i)$

! : Vraisemblance $\neq$ densité

Exemple : loi gaussienne, loi de Bernoulli

Exemple : loi de Bernoulli



Densité (à gauche) et la vraisemblance (à droite)

Estimateur du maximum de vraisemblance

La valeur θ_1 de θ est plus vraisemblable que la valeur θ_2, si $L(\theta_1; x) > L(\theta_2; x)$; graphe !

Définition (EMV)

On appelle estimation du maximum de vraisemblance, une valeur $\widehat{\theta}_n$ maximisant la vraisemblance

$$\widehat{\theta}_n \in \operatorname{Arg}\max_{\theta\in\Theta} L(\theta; x).$$

$\widehat{\theta}_n = t(x_1, \ldots, x_n)$ *est une fonction des données, ce qui induit la statistique* $t(X_1, \ldots, X_n)$ *que l'on note (abusivement) avec la même notation :*
$\widehat{\theta}_n = t(X_1, \ldots, X_n)$ *est appelé Estimateur du Maximum de Vraisemblance*

Exemple : loi gaussienne, loi de Bernoulli

EMV

Remarque : quand l'échantillon est i.i.d. on utilise plutôt la log-vraisemblance

$$\ell_n(\theta; x) = \log L(\theta; x) = \sum_{i=1}^{n} \log f_\theta(x_i)$$

Propriétés

- Si $\theta \to \ell_n(\theta; x)$ est une fonction continue et que Θ est compact, l'existence de l'EMV est garantie.
- Si la vraisemblance est dérivable en θ,
  - $\hookrightarrow$ soit l'EMV annule le gradient,
  - $\hookrightarrow$ soit il est situé sur le bord de Θ

Calcul de l'EMV

Si le domaine de définition de f_θ ne dépend pas de θ, et si la vraisemblance est deux fois dérivable :

- Chercher $\widehat{\theta}_n$ annulant les équations de vraisemblance (ou équations normales ou équations du score)

$$U_n(\widehat{\theta}_n) := \nabla \ell_n(\widehat{\theta}_n; x) = \left(\frac{\partial}{\partial \theta_k} \ell_n(\widehat{\theta}_n; x) \right)_{k=1,\ldots,dim(\theta)} = 0,$$

- Vérifier que $\widehat{\theta}_n$ est bien un maximum : $H_n(\theta) = \nabla^2 \ell_n(\theta; x)$ est définie négative autour de $\widehat{\theta}_n$.

Exemple : loi gaussienne, loi de Bernoulli

Calcul numérique de la vraisemblance

Dans le cas où le calcul analytique de l'EMV est impossible
$\hookrightarrow$ schéma numérique, par ex. Newton-Raphson :

- Initialiser $\theta^{(0)}$
- Itérer jusqu'à converge ou à un nombre de pas fixés
  - $\hookrightarrow$ Linéarisation de U_n au point courant $(\theta^{(m)}, U_n(\theta^{(m)}))$

$$0 = U_n(\theta^{(m)}) + H_n(\theta^{(m)})(\theta^{(m+1)} - \theta^{(m)})$$

  - $\hookrightarrow$ si $H_n(\theta^{(m)})$ est inversible

$$\theta^{(m+1)} = \theta^{(m)} - [H_n(\theta^{(m)})]^{-1} U_n(\theta^{(m)})$$

Attention, il peut y avoir des maxima locaux

Propriétés de l'EMV

Propriété

Soit $X_1, \ldots, X_n \sim \mathbb{P}_\theta$. Sous des conditions de régularité du modèle, l'EMV $\widehat{\theta}_n$ de θ est consistant et asymptotiquement normal.
Il est donc asymptotiquement sans biais mais peut être biaisé à distance finie

Nous démontrerons ces propriétés au cas par cas en 1A, à suivre dans le cours de 2A !

Remarque : Quand l'échantillon est i.i.d. on utilise plutôt la log-vraisemblance

$$\ell_n(\theta; x) = \log L(\theta; x) = \sum_{i=1}^{n} \log f_\theta(x_i)$$

Propriétés asymptotiques
Consistance
Loi asymptotique
Construction d'estimateurs
Moments
EMV