

Titre

Nom

2 mars 2025

1 Introduction

Le théorème central limite, formulé par Pierre-Simon de Laplace en 1809, garantit que, sous des conditions raisonnables, la somme normalisée de ces variables suit asymptotiquement une loi normale. Cette convergence est utile pour étudier le comportement **global** des observations, mais elle ne renseigne pas sur le comportement des valeurs extrêmes.

Il est donc naturel de se demander quelle peut être la convergence en loi de ses valeurs extrêmes. Autrement dit, pour $X = (X_1, \dots, X_n)$ un échantillon de variables aléatoires i.i.d, on pose :

$$M_n = \max\{X_i \mid i \in \{1, \dots, n\}\}$$

et on s'intéresse à la convergence de M_n , ainsi qu'aux hypothèses sous lesquelles cette convergence a lieu.

Remarque : Etudier le minimum est totalement analogue dans ce qui suit.

2 Lois de M_n

2.1 Quelque notation

On commence par faire une remarque sur la fonction de repartition de M_n en utilisant le fait que les X_i sont i.i.d :

En effet,

$$F_{M_n}(t) = \mathbb{P}(M_n < t) = \mathbb{P}(X_1 < t, \dots, X_n < t) = \mathbb{P}(X_1 < t)^n = F_{X_1}^n(t)$$

Dans la suite, on notera $F(t)$, la fonction de repartition des X_i .

Mais on rencontre un problème ici, puisque si $n \rightarrow +\infty$, $F(t)^n$ converge vers 0 (ou 1 si t est la borne sup du support des X_i).

L'idée est donc d'introduire 2 suites (b_n) et (a_n) (avec $a_n > 0$ pour tout n) afin de pouvoir contrôler M_n .

Puis étudier la loi de la limite de $\frac{M_n - b_n}{a_n}$. Comme la fonction de repartition caractérise la loi, il nous suffit d'étudier la fonction G définie pour tout t dans le support des X_i comme :

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} < t\right) \xrightarrow{n \rightarrow +\infty} G(t)$$

à ce stade là, il nous faut donc trouver toutes les distributions G qui peuvent apparaître comme limite dans l'équation ci-dessus.

Pour ce faire, nous allons utiliser le théorème suivant :

Théorème : Soit Y_n une variable aléatoire de fonction de répartition F_n , et soit Y une variable aléatoire de fonction de répartition F . Alors $Y_n \xrightarrow{\mathcal{L}} Y$ si et seulement si pour toute fonction z réelle, bornée et continue :

$$\mathbb{E}[z(Y_n)] \rightarrow \mathbb{E}[z(Y)].$$

En prenant ici $Y_n = \frac{M_n - b_n}{a_n}$, on obtient :

$$\mathbb{E}\left[z\left(\frac{M_n - b_n}{a_n}\right)\right] = \int_{-\infty}^{\infty} z\left(\frac{x - b_n}{a_n}\right) n F^{n-1}(x) dF(x)$$

L'astuce ici va être de faire un changement de variable astucieux. On va poser :

$$x = Q\left(1 - \frac{1}{y}\right) = K(y) \quad \text{avec } Q \text{ la fonction quantile}$$

$$\text{donc} \quad \int_{-\infty}^{\infty} z\left(\frac{x - b_n}{a_n}\right) n F^{n-1}(x) dF(x) = \int_0^n z\left(\frac{K\left(\frac{n}{v}\right) - b_n}{a_n}\right) \left(1 - \frac{v}{n}\right)^{n-1} dv$$

Or, on a $\lim_{n \rightarrow \infty} (1 - \frac{v}{n})^{n-1} = e^{-v}$, et on a $\lim_{n \rightarrow \infty} \int_0^n = \int_0^{+\infty}$.

2.2 Paramètre b_n

On en déduit une bonne valeur pour b_n . En effet,

$$\begin{aligned} \mathbb{P}\left(\frac{M_n - b_n}{a_n} < t\right) &\xrightarrow{n \rightarrow +\infty} G(t) \in]0 : 1[\\ \iff F^n(a_n t + b_n) &\xrightarrow{n \rightarrow +\infty} G(t) \\ \iff n \ln(F(a_n t + b_n)) &\xrightarrow{n \rightarrow +\infty} \ln(G(t)) \\ \iff n(-F(a_n t + b_n) + 1) &\xrightarrow{n \rightarrow +\infty} \ln(G(t)) \\ \iff n \mathbb{P}(X_1 > a_n t + b_n) &\xrightarrow{n \rightarrow +\infty} -\ln(G(t)) \end{aligned}$$

On obtient alors pour paramètre d'échelle :

$$\begin{aligned} n\mathbb{P}(X_1 > b_n) = 1 &\iff \mathbb{P}(X_1 < b_n) = 1 - \frac{1}{n} \\ \iff F(b_n) &= 1 - \frac{1}{n} \\ b_n = Q(1 - \frac{1}{n}) &= K(n) \end{aligned}$$

Dans la dernière équivalence, on a composé par la fonction quantile (qui est bien définie pour chaque distribution)

2.3 Paramètre a_n

Avec le paramètre b_n définie au dessus, on obtient alors une condition (C), il faut qu'il existe une fonction a tel que $\lim_{x \rightarrow \infty} \frac{K(\frac{x}{a_n}) - K(x)}{a(x)}$ converge (vers une fonction $h(n)$).

Proposition :

Les limites possibles sont données par :

$$c h_\gamma(u) = c \int_1^u v^{-\gamma-1} dv = c \frac{u^\gamma - 1}{\gamma}.$$

Nous interprétons $h_0(u) = \log(u)$ lorsque $\gamma = 0$.

On ne veut pas que $c = 0$, car il conduit à une limite dégénérée pour $\frac{X_{n,n} - b_n}{a_n}$. Ensuite, le cas $c > 0$ peut être ramené au cas $c = 1$ en incorporant c dans la fonction a .

Nous remplaçons la condition (C) par une condition plus informative, appelée "condition extrême du domaine d'attraction" :

$$\lim_{x \rightarrow \infty} \frac{U(xu) - U(x)}{a(x)} = h_\gamma(u), \quad \text{pour } u > 0.$$

Ceci indique que les limites possibles sont essentiellement décrites par la famille à un paramètre h_γ . Si nécessaire, nous ferons référence à la fonction auxiliaire a en notant $\mathcal{C}_\gamma(a)$. Nous dirons que la distribution F satisfait la condition de valeurs extrêmes $\mathcal{C}_\gamma(a)$ avec la fonction auxiliaire a .

Preuve de la Proposition

Soient $u, v > 0$. Alors :

$$\frac{U(xuv) - U(x)}{a(x)} = \frac{U(xuv) - U(xu)}{a(xu)} \frac{a(xu)}{a(x)} + \frac{U(xu) - U(x)}{a(x)}. \quad (2.3)$$

Si la limite dans (C) est satisfaite, alors le rapport

$$\frac{a(ux)}{a(x)}$$

doit converger. Nous notons cette limite par $g(u)$. En effet, pour $u, v > 0$,

$$\frac{a(xuv)}{a(x)} = \frac{a(xuv)}{a(xv)} \frac{a(xv)}{a(x)}.$$

Par passage à la limite, la fonction g satisfait l'équation fonctionnelle de Cauchy :

$$g(uv) = g(u)g(v).$$

3 Estimation du paramètre gamma

Soient X_1, \dots, X_n i.i.d. On définit la fonction de répartition :

$$F(x) = P(X_1 \leq x), \quad x \in \mathbb{R}$$

Nous allons définir la fonction de répartition empirique.

3.1 Fonction de répartition empirique

Définition : Pour tout $x \in \mathbb{R}$, la fonction de répartition empirique est donnée par :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{(-\infty, x]}(X_i)$$

où $\mathbb{I}_{(-\infty, x]}(X_i)$ est l'indicatrice de l'événement $\{X_i \leq x\}$. La fonction de répartition empirique utilise la statistique d'ordre.

3.2 Quantiles et inverse généralisée

Définition : Soit $p \in (0, 1)$, on appelle p -quantile, noté x_p , de la loi F toute quantité satisfaisant :

$$F(x_p) = p$$

Pour définir un quantile de manière plus générale, on considère l'inverse généralisée de F .

Définition : Pour tout $u \in [0, 1]$, on appelle *inverse généralisée* de F , notée F^- , la fonction définie par :

$$F^-(u) = \inf\{x \in \mathbb{R} \mid F(x) \geq u\}$$

3.3 Quantiles empiriques

Définition : Soit $p \in (0, 1)$, on appelle p -quantile empirique, noté $x_p(n)$, la valeur :

$$x_p(n) = F_n^-(p) = \inf\{x \in \mathbb{R} \mid F_n(x) \geq p\}$$

3.4 Estimation de γ par la méthode des quantiles

Pour déterminer les estimateurs de γ pour ces trois distributions, nous utilisons l'estimation par la méthode des quantiles. Nous allons commencer par estimer le paramètre γ pour la distribution de Fréchet.

3.4.1 Rappel de la fonction de répartition

Rappelons que la fonction de répartition de la loi de Fréchet s'écrit comme :

$$F(x) = \exp(-x^{-\gamma}), \quad x > 0, \quad \gamma > 0.$$

On cherche x tel que :

$$\exp(-x^{-\gamma}) = p.$$

En prenant le logarithme, on obtient :

$$\ln(p) = -x^{-\gamma}$$

$$x^\gamma = -\frac{1}{\ln(p)}$$

3.4.2 Estimation de γ à partir du quantile médian

On utilise l'estimateur du quantile avec la médiane de la distribution ($x_{1/2}$) :

$$x_{1/2} = (\ln 2)^{-1/\gamma} = f(\gamma),$$

où $f(\gamma)$ est une fonction bijective. Si l'on sait estimer $x_{1/2}$, alors on peut facilement estimer γ en inversant f :

$$\gamma = f^{-1}(x_{1/2}).$$

On résout :

$$x = f(\gamma) \iff x = (\ln 2)^{-1/\gamma}$$

$$\iff (\ln 2)^{1/\gamma} = \frac{1}{x}$$

$$\iff \frac{1}{\gamma} \ln(\ln 2) = \ln\left(\frac{1}{x}\right) = -\ln x.$$

D'où :

$$\gamma = -\frac{\ln(\ln 2)}{\ln x} = f^{-1}(x).$$

3.4.3 Convergence et estimation empirique

Rappelons que le quantile empirique converge en probabilité vers le quantile théorique :

$$x_p(n) \xrightarrow{\mathbb{P}} x_p = F^{-1}(p).$$

En particulier, pour $p = \frac{1}{2}$:

$$x_{1/2}(n) \xrightarrow{\mathbb{P}} x_{1/2},$$

où $x_{1/2}(n)$ est le quantile empirique d'ordre 1/2 (médiane empirique).

Or, la convergence en probabilité est stable par transformation continue, d'où :

$$f^{-1}(x_{1/2}(n)) \xrightarrow{\mathbb{P}} f^{-1}(x_{1/2}) = \gamma.$$

Ainsi,

$$\gamma = f^{-1}(x_{1/2}(n))$$

est un estimateur convergent de γ . Finalement, l'estimateur de γ est donné par :

$$\hat{\gamma} = -\frac{\ln(\ln 2)}{\ln x_{1/2}(n)}.$$

3.5 Estimation de γ pour la distribution de Gumbel

La fonction de répartition de Gumbel est donnée par :

$$F(x) = \exp(-\exp(-x/\gamma)).$$

Ainsi, pour $p = \frac{1}{2}$, on a :

$$F(x_{1/2}) = \frac{1}{2}$$

$$\iff \exp(-\exp(-x_{1/2}/\gamma)) = \frac{1}{2}$$

$$\iff -\exp(-x_{1/2}/\gamma) = -\ln(2)$$

$$\iff \exp(-x_{1/2}/\gamma) = \ln(2)$$

$$\iff -\frac{x_{1/2}}{\gamma} = \ln(\ln(2))$$

$$\iff \gamma = -\frac{x_{1/2}}{\ln(\ln(2))} = f(x_{1/2}),$$

où f est une fonction continue.

3.5.1 Convergence et estimation empirique

En utilisant le quantile empirique, on a :

$$x_{1/2}(n) \xrightarrow{\mathbb{P}} x_{1/2}.$$

Puisque la convergence en probabilité est stable par transformation continue, il en résulte :

$$f(x_{1/2}(n)) \xrightarrow{\mathbb{P}} \gamma = f(x_{1/2}).$$

Ainsi, un estimateur convergent de γ est donné par :

$$\hat{\gamma} = f(x_{1/2}(n)) = -\frac{x_{1/2}(n)}{\ln(\ln(2))}.$$

3.6 Estimation de γ pour la distribution de Weibull

Ainsi, l'estimateur $\hat{\gamma}$ est un estimateur convergent de γ .

La fonction de répartition de Weibull est donnée par :

$$F(x) = 1 - \exp(-x^\gamma).$$

Pour $p = \frac{1}{2}$, nous avons :

$$F(x_{1/2}) = \frac{1}{2}$$

$$\iff 1 - \exp(-x_{1/2}^\gamma) = \frac{1}{2}$$

$$\iff \frac{1}{2} = \exp(-x_{1/2}^\gamma)$$

$$\iff -\ln(2) = -x_{1/2}^\gamma$$

$$\iff \ln(2) = x_{1/2}^\gamma.$$

En appliquant le logarithme népérien :

$$\gamma \ln(x_{1/2}) = \ln(\ln 2)$$

$$\iff \gamma = \frac{\ln(\ln 2)}{\ln(x_{1/2})} = \varphi(x_{1/2}),$$

où φ est une fonction continue.

3.6.1 Convergence et estimation empirique

Puisque $x_{1/2}(n)$ converge en probabilité vers $x_{1/2}$, on a :

$$\varphi(x_{1/2}(n)) \xrightarrow{\mathbb{P}} \varphi(x_{1/2}) = \gamma.$$

Finalement, un estimateur convergent de γ est donné par :

$$\hat{\gamma} = \frac{\ln(\ln 2)}{\ln(x_{1/2}(n))}.$$

4 Méthodes d'estimation de l'indice de valeurs extrêmes

Dans cette section, nous nous intéressons aux différentes méthodes d'estimation du paramètre γ , intervenant dans la distribution des valeurs extrêmes généralisée. D'une part, des approches non paramétriques sont dédiées à l'estimation de l'indice de queue, notamment les estimateurs de Hill et de Pickands. D'autre part, des méthodes paramétriques ont été développées, parmi lesquelles la méthode du maximum de vraisemblance, la méthode des moments et les approches bayésiennes.

Définition : On appelle *statistique d'ordre* la permutation aléatoire de l'échantillon X_1, \dots, X_n , qui ordonne les valeurs de l'échantillon par ordre croissant :

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

En particulier, $X_{(1)} = \min X_i$ et $X_{(n)} = \max X_i$.

Attention !!! Même si les X_i sont i.i.d., les statistiques d'ordre $X_{(i)}$ ne le sont pas.

Définition : On dit qu'une suite $(k_n)_{n \geq 0}$ d'entiers est intermédiaire si :

$$\lim_{n \rightarrow \infty} k_n = \infty \quad \text{et} \quad \lim_{n \rightarrow \infty} \frac{k_n}{n} = 0$$

4.1 Estimateur de Pickands

L'estimateur de Pickands est défini par la statistique

$$\hat{\gamma}_{k,n} = \frac{1}{\ln(2)} \ln \left(\frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}} \right)$$

Cet estimateur présente l'avantage d'être applicable quelle que soit la distribution des extrêmes. Cependant, la représentation graphique de cet estimateur en fonction du nombre k d'observations considérées révèle généralement un comportement volatil au départ, ce qui peut nuire à la lisibilité du graphique. De plus, cet estimateur est particulièrement sensible à la taille de l'échantillon sélectionné, ce qui le rend peu robuste.

Cet estimateur est asymptotiquement normal, avec :

$$\sqrt{k} \frac{\hat{\gamma}_{k,n} - \gamma}{\sigma(\gamma)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Lorsque $k \rightarrow +\infty$, la variance asymptotique est donnée par :

$$\sigma(\gamma) = \frac{\gamma \sqrt{2^{(2\gamma+1)} + 1}}{2(2^\gamma - 1) \ln(2)}$$

Une généralisation de l'estimateur de Pickands a été introduite comme suit :

$$\hat{\gamma}_{(k,u,v)} = \frac{1}{\ln(v)} \ln \left(\frac{X_{n-k+1,n} - X_{n-[uk]+1,n}}{X_{n-[vk]+1,n} - X_{n-[uvk]+1,n}} \right)$$

où u et v sont des réels positifs différents de 1, de sorte que les indices $[vk]$, $[uk]$ et $[uvk]$ ne dépassent pas n . Lorsque $u = v = 2$, on retrouve l'estimateur de Hill $\hat{\gamma}_{k,n}$.

4.2 Estimateur de Hill

Tout d'abord, l'estimateur de Hill est applicable uniquement aux distributions de Fréchet ($\gamma > 0$), où il permet d'obtenir un estimateur de l'indice de queue plus efficace que celui de Pickands. Cet estimateur est défini par la statistique suivante :

$$\hat{\gamma}_{k,n} = \frac{1}{k} \sum_{i=1}^k \ln\left(\frac{X_{n-i+1,n}}{X_{n-k,n}}\right)$$

pour $k \in \{1, \dots, n-1\}$. Si l'on choisit $k, n \rightarrow +\infty$, de sorte que $\frac{k}{n} \rightarrow 0$, alors on peut montrer que $\lim_{k \rightarrow \infty} \hat{\gamma}_{k,n} = \gamma$. Cet estimateur possède la propriété d'être asymptotiquement normal, ce qui signifie que :

$$\sqrt{k} \frac{\hat{\gamma}_{k,n} - \gamma}{\gamma} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Il existe plusieurs approches pour construire l'estimateur de Hill. Une approche possible consiste à utiliser la méthode du maximum de vraisemblance.

Tout d'abord, on considère une suite de variables aléatoires X_1, \dots, X_n i.i.d. suivant une loi de Pareto de paramètre $\lambda > 0$, dont la fonction de répartition est donnée par :

$$F(x) = 1 - x^{-\lambda}, \quad \text{pour } x \geq 1.$$

La densité de probabilité associée est alors :

$$f(x) = \lambda x^{-\lambda-1}, \quad \text{pour } x \geq 1.$$

La fonction de vraisemblance est donnée par :

$$L(x_1, \dots, x_n, \lambda) = \prod_{i=1}^n f(x_i) = \lambda^n \prod_{i=1}^n x_i^{-\lambda-1}.$$

En prenant le logarithme, on obtient la log-vraisemblance :

$$\log L(x_1, \dots, x_n, \lambda) = n \log \lambda - (\lambda + 1) \sum_{i=1}^n \log x_i.$$

En dérivant cette expression par rapport à λ , on obtient :

$$\frac{d \log L}{d \lambda} = \frac{n}{\lambda} - \sum_{i=1}^n \log x_i.$$

En dérivant une seconde fois, nous obtenons :

$$\frac{d^2 \log L}{d \lambda^2} = -\frac{n}{\lambda^2} < 0,$$

ce qui confirme qu'il s'agit bien d'un maximum.

Ainsi, l'estimateur du maximum de vraisemblance de $\frac{1}{\lambda}$ est donné par :

$$\hat{\lambda}^{-1} = \frac{1}{n} \sum_{i=1}^n \log X_i.$$

Cela implique que l'estimateur du paramètre λ est :

$$\hat{\lambda} = \left(\frac{1}{n} \sum_{i=1}^n \log X_i \right)^{-1}.$$

4.3 Estimateur des moments

L'estimateur des moments a été proposé par Dekkers, Einmahl et Haan en 1989 comme une alternative aux estimateurs de Hill et de Pickands pour l'estimation de l'indice de queue γ dans les modèles de valeurs extrêmes. Ce type d'estimateur est particulièrement adapté lorsque les données ne suivent pas exactement les hypothèses des autres estimateurs. Cet estimateur est défini comme suit

$$\hat{\gamma}_n = 1 + \hat{\gamma}_n^{(1)} - \frac{1}{2} \left[1 - \frac{(\hat{\gamma}_n^{(1)})^2}{\hat{\gamma}_n^{(2)}} \right]^{-1}$$

avec

$$\hat{\gamma}_n^{(i)} = \left[\frac{1}{k} \sum_{j=1}^k \log \left(\frac{X_{j,n}}{X_{k-1,n}} \right) \right]^i, \quad \text{où } i = 1, 2$$

5 Convergence des estimateurs

Définition : On dit qu'un estimateur $\hat{\gamma}_n$ est convergent s'il converge en probabilité vers γ , soit :

$$\lim_{n \rightarrow \infty} P(|\hat{\gamma}_n - \gamma| > \epsilon) = 0 \quad \forall \epsilon > 0$$

5.1 Estimateur de Hill

Si on choisit $k, n \rightarrow +\infty$ de sorte que $\frac{k}{n} \rightarrow 0$ alors on peut montrer que $\lim_{k \rightarrow \infty} \hat{\gamma}_{k,n} = \gamma$.

Démonstration :

5.2 Estimateur de Pickands

Lemme : Soit X_1, \dots, X_n des variables aléatoires indépendantes et de fonction de répartition F . Soit U_1, \dots, U_n des variables aléatoires indépendantes de loi uniforme sur $[0, 1]$. Alors :

$$(F(U_{(1,n)}), \dots, F(U_{(n,n)}))$$

a la meme loi que :

$$(F(X_{(1,n)}), \dots, F(X_{(n,n)}))$$

Pour la suite, nous utiliserons cette notation :

$$F \in D(H(\gamma)) \iff n \lim_{n \rightarrow \infty} n \bar{F}(xan + bn) = -\log(H(\gamma(x)))$$

pour une certaine suite $(an, bn)_{n \geq 1}$ où $an > 0$ et $bn \in \mathbb{R}$. On a alors la convergence en loi de $(an^{-1}(Mn - bn))_{n \geq 1}$ vers une variable aléatoire de la fonction de répartition $H(\gamma)$.

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes de même fonction de répartition $F \in D(H(\gamma))$, où $\gamma \in \mathbb{R}$. Si $\lim_{n \rightarrow \infty} k(n) = \infty$ et $\lim_{n \rightarrow \infty} \frac{k(n)}{n} = 0$, alors l'estimateur de Pickands converge en probabilité vers γ .

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes de même fonction de répartition $F \in D(H(\gamma))$, où $\gamma \in \mathbb{R}$. Si $\lim_{n \rightarrow \infty} k(n) = \infty$ et $\lim_{n \rightarrow \infty} \frac{k(n)}{n} = 0$, alors l'estimateur de Pickands converge en probabilité vers γ .

5.3 Estimateur des moments

Supposons que $F \in D(H_\gamma)$, $\gamma \in \mathbb{R}$, $k_n \rightarrow \infty$ et $\frac{k}{n} \rightarrow 0$ quand $n \rightarrow \infty$:

$$\hat{\gamma}_n \xrightarrow{\mathbb{P}} \gamma \quad \text{quand } n \rightarrow \infty$$

6 Sélection des estimateurs de l'indice de valeurs extrêmes

Le choix de l'estimateur dépend du type de distribution sous-jacente. L'estimateur de Hill est spécifiquement adapté aux distributions de Fréchet ($\gamma > 0$), caractérisées par des queues lourdes. Il est donc plus efficace dans ce cas et sera préféré à l'estimateur de Pickands.

Cependant, pour les distributions de Weibull ($\gamma < 0$) et Gumbel ($\gamma = 0$), l'estimateur de Hill n'est pas applicable. Dans ces cas, on utilise l'estimateur de Pickands, qui est valide quel que soit le signe de γ .

Ainsi, la sélection de l'estimateur repose sur la nature de la distribution des valeurs extrêmes.