Différentes contributions à l'estimation de quantiles extrêmes

PAR

Jonathan EL METHNI

sous la direction de

Stéphane GIRARD & Laurent GARDES

Journée de rentrée des doctorants en Statistique du LJK
20 Novembre 2012

Motivations

Exemple

La hauteur d'une rivière est modélisée par une variable aléatoire Y ayant pour fonction de survie $\overline{F} = 1 - F$.
On dispose de $Y_{1,n} \leq \cdots \leq Y_{n,n}$ un échantillon ordonné de hauteurs d'eau annuelles.

Motivations

Exemple

La hauteur d'une rivière est modélisée par une variable aléatoire Y ayant pour fonction de survie $\overline{F} = 1 - F$.
On dispose de $Y_{1,n} \leq \cdots \leq Y_{n,n}$ un échantillon ordonné de hauteurs d'eau annuelles.

Problème

- On veut estimer le niveau d'eau h qui est atteint ou dépassé en moyenne une fois sur T années avec $T > n$ (taille de l'échantillon) *i.e.* on veut estimer h tel que

$$1/T = \mathbb{P}(Y \geq h) = \overline{F}(h)$$

Motivations

Exemple

La hauteur d'une rivière est modélisée par une variable aléatoire Y ayant pour fonction de survie $\overline{F} = 1 - F$.
On dispose de $Y_{1,n} \leq \cdots \leq Y_{n,n}$ un échantillon ordonné de hauteurs d'eau annuelles.

Problème

- On veut estimer le niveau d'eau h qui est atteint ou dépassé en moyenne une fois sur T années avec $T > n$ (taille de l'échantillon) *i.e.* on veut estimer h tel que

$$1/T = \mathbb{P}(Y \geq h) = \overline{F}(h)$$

Autrement dit on veut estimer

$$h = \overline{F}^{-1}(1/T)$$

Motivations

Exemple

La hauteur d'une rivière est modélisée par une variable aléatoire Y ayant pour fonction de survie $\overline{F} = 1 - F$.
On dispose de $Y_{1,n} \leq \cdots \leq Y_{n,n}$ un échantillon ordonné de hauteurs d'eau annuelles.

Problème

- On veut estimer le niveau d'eau h qui est atteint ou dépassé en moyenne une fois sur T années avec $T > n$ (taille de l'échantillon) *i.e.* on veut estimer h tel que
$$1/T = \mathbb{P}(Y \geq h) = \overline{F}(h)$$
Autrement dit on veut estimer
$$h = \overline{F}^{-1}(1/T)$$

Définition : Quantile

Le quantile d'ordre $\alpha \in]0, 1[$ noté $q(\alpha)$ est définie par $q(\alpha) := \overline{F}^{-1}(\alpha)$.
On cherche à estimer $h = \overline{F}^{-1}(1/T)$ qui est un quantile d'ordre $\alpha = 1/T$.

Motivations

Définition : Quantile extrême

Un quantile est dit extrême si son ordre $\alpha = \alpha_n \longrightarrow 0$ lorsque $n \to \infty$.

Motivations

Définition : Quantile extrême

Un quantile est dit extrême si son ordre $\alpha = \alpha_n \longrightarrow 0$ lorsque $n \to \infty$.

On veut estimer $h = \overline{F}^{-1}(1/T)$, l'ordre du quantile vaut $\alpha_n = 1/T$ avec $T > n$ on a donc

$$\alpha_n = 1/T < 1/n \underset{n \to \infty}{\longrightarrow} 0.$$

On cherche donc à estimer un quantile extrême.

Motivations

Définition : Quantile extrême

Un quantile est dit extrême si son ordre $\alpha = \alpha_n \longrightarrow 0$ lorsque $n \to \infty$.

On veut estimer $h = \overline{F}^{-1}(1/T)$, l'ordre du quantile vaut $\alpha_n = 1/T$ avec $T > n$ on a donc

$$\alpha_n = 1/T < 1/n \underset{n \to \infty}{\longrightarrow} 0.$$

On cherche donc à estimer un quantile extrême.

Difficultée

La fonction de survie $\overline{F}$ est inconnue et difficile à estimer au-delà du maximum. En effet on a :

$$\text{Si } n\alpha_n \underset{n \to \infty}{\longrightarrow} 0 \text{ alors } \mathbb{P}(q(\alpha_n) > Y_{n,n}) \underset{n \to \infty}{\longrightarrow} 1.$$

Résultat fondamental de la Théorie des valeurs extrêmes

Théorème [Fisher et Tippett (1928) et Gnedenko (1943)]

Sous certaines conditions de régularité sur la fonction de répartition, il existe un paramètre réel γ et deux suites $(a_n)_{n\geq 1} > 0$ et $(b_n)_{n\geq 1} \in \mathbb{R}$ tels que pour tout $y \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{Y_{n,n} - b_n}{a_n} \leq y\right) = \mathcal{H}_\gamma(y),$$

avec

$$\mathcal{H}_\gamma(y) = \begin{cases} \exp\left(-(1+\gamma y)_+^{-1/\gamma}\right) & \text{if} \quad \gamma \neq 0, \\ \exp\left(-e^{-y}\right) & \text{if} \quad \gamma = 0, \end{cases}$$

où $z_+ = \max(0, z)$.

- $\mathcal{H}_\gamma$ est la loi des valeurs extrêmes.
- a_n et b_n sont des paramètres de normalisation.
- γ est l'indice des valeurs extrêmes.

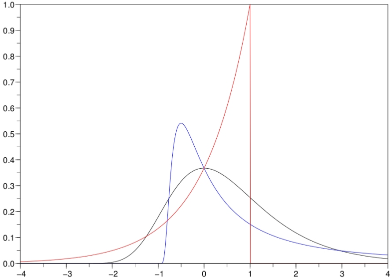
3 domaines d'attraction

- Si F vérifie le Théorème [Fisher et Tippet (1928) et Gnedenko (1943)], on dit alors que F appartient au domaine d'attraction de $\mathcal{H}_\gamma$.

Fréchet ($\gamma > 0$)	Gumbel ($\gamma = 0$)	Weibull ($\gamma < 0$)
Pareto	Normale	Uniforme
Student	Exponentielle	Beta
Burr	Log-normale	
Fréchet	Gamma	
	Weibull	

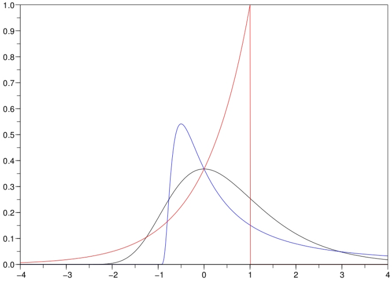
- Si $\gamma > 0$ on a une loi dite à queue lourde.
- Si $\gamma = 0$ on a une loi dite à queue légère.
- Si $\gamma < 0$ on a une loi dite à queue finie.

3 domaines d'attraction



Densité de la loi des valeurs extrêmes pour $\gamma = -1$, $\gamma = 0$ et $\gamma = 1$

3 domaines d'attraction

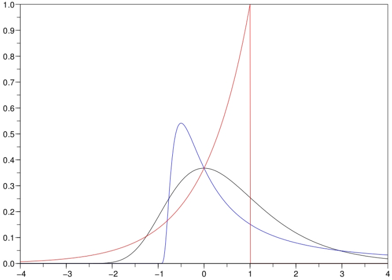


Densité de la loi des valeurs extrêmes pour $\gamma = -1$, $\gamma = 0$ et $\gamma = 1$

Première partie de ma thèse

- Proposer un modèle unificateur des domaines d'attraction de Fréchet ($\gamma > 0$) et de Gumbel ($\gamma = 0$).

3 domaines d'attraction



Densité de la loi des valeurs extrêmes pour $\gamma = -1$, $\gamma = 0$ et $\gamma = 1$

Première partie de ma thèse

- Proposer un modèle unificateur des domaines d'attraction de Fréchet $(\gamma > 0)$ et de Gumbel $(\gamma = 0)$.

$\Longrightarrow$ Estimer des quantiles extrêmes dans les deux domaines d'attraction.

Mesures de risque

Definition

Soit Y une variable aléatoire désignant un montant de perte. On appelle mesure de risque une fonction $\mathcal{R}$ associant à Y une valeur positive ou nulle telle que

$$\mathcal{R} : Y \longrightarrow \mathbb{R}^+$$

$\mathcal{R}(Y)$ représente le capital à détenir pour faire face aux pertes Y.
De grandes valeurs de $\mathcal{R}(Y)$ indiqueront que Y est "dangereux".

Exemples de mesures de risque $\mathcal{R}$:

- Value-at-Risk,
- Conditional Tail Expectation.

Value-at-Risk et Conditional Tail Expectation

Definition

La plus utilisée des mesures de risque est la Value-at-Risk (1993). La Value-at-Risk au niveau de confiance $\alpha \in]0,1[$ notée $VaR(\alpha)$ est définie par

$$VaR(\alpha) = \overline{F}^{-1}(\alpha) = q(\alpha).$$

Value-at-Risk et Conditional Tail Expectation

Definition

La plus utilisée des mesures de risque est la Value-at-Risk (1993). La Value-at-Risk au niveau de confiance $\alpha \in]0,1[$ notée $VaR(\alpha)$ est définie par

$$VaR(\alpha) = \overline{F}^{-1}(\alpha) = q(\alpha).$$

L'idée derrière le concept de la *VaR*

On fixe un seuil probabiliste α et on définit une valeur $VaR(\alpha)$ qui sera acceptable ssi la probabilité que la catastrophe survienne est plus petite que α.

Value-at-Risk et Conditional Tail Expectation

Definition

La plus utilisée des mesures de risque est la Value-at-Risk (1993). La Value-at-Risk au niveau de confiance $\alpha \in]0,1[$ notée $VaR(\alpha)$ est définie par

$$VaR(\alpha) = \overline{F}^{-1}(\alpha) = q(\alpha).$$

L'idée derrière le concept de la *VaR*

On fixe un seuil probabiliste α et on définit une valeur $VaR(\alpha)$ qui sera acceptable ssi la probabilité que la catastrophe survienne est plus petite que α.

Un des principaux reproches fait à la *VaR* et que des v.a à queues légères et à queues lourdes peuvent avoir la même $VaR(\alpha)$ (Embrechts et al. (1997)).

Value-at-Risk et Conditional Tail Expectation

Definition

La plus utilisée des mesures de risque est la Value-at-Risk (1993). La Value-at-Risk au niveau de confiance $\alpha \in]0,1[$ notée $VaR(\alpha)$ est définie par

$$VaR(\alpha) = \overline{F}^{-1}(\alpha) = q(\alpha).$$

L'idée derrière le concept de la *VaR*

On fixe un seuil probabiliste α et on définit une valeur $VaR(\alpha)$ qui sera acceptable ssi la probabilité que la catastrophe survienne est plus petite que α.

Un des principaux reproches fait à la *VaR* et que des v.a à queues légères et à queues lourdes peuvent avoir la même $VaR(\alpha)$ (Embrechts et al. (1997)).

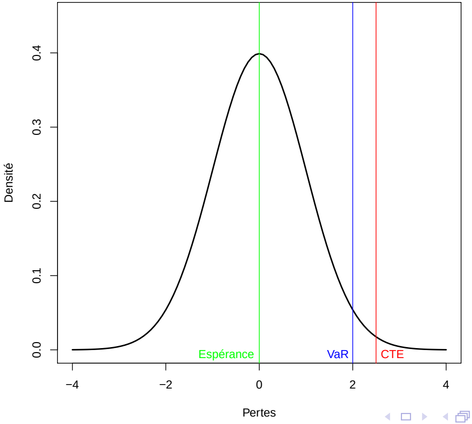
Definition

La Conditional Tail Expectation au niveau de confiance $\alpha \in]0,1[$ notée $CTE(\alpha)$ est une mesure de risque définie par

$$CTE(\alpha) = \mathbb{E}(Y | Y > VaR(\alpha)).$$

VaR et CTE

La $CTE(\alpha)$ donne des informations sur la distribution de Y au delà de la $VaR(\alpha)$ et donc contrairement à la $VaR(\alpha)$, sur l'épaisseur de la queue de distribution.



Contributions

Deuxième partie de ma thèse

L'apport nouveau de ce travail consiste en l'ajout de deux difficultés supplémentaires dans le cadre de l'estimation de mesures de risque (VaR et CTE) de pertes pour des lois à queues lourdes.

Contributions

Deuxième partie de ma thèse

L'apport nouveau de ce travail consiste en l'ajout de deux difficultés supplémentaires dans le cadre de l'estimation de mesures de risque (*VaR* et *CTE*) de pertes pour des lois à queues lourdes.

- On ajoute la présence d'une covariable $X \in \mathbb{R}^p$.

Contributions

Deuxième partie de ma thèse

L'apport nouveau de ce travail consiste en l'ajout de deux difficultés supplémentaires dans le cadre de l'estimation de mesures de risque (*VaR* et *CTE*) de pertes pour des lois à queues lourdes.

- On ajoute la présence d'une covariable $X \in \mathbb{R}^p$.

- On s'intéresse à l'estimation de mesures de risque (*VaR* et *CTE*) de pertes extrêmes pour des lois à queues lourdes.

$\Longrightarrow$ Pour cela on remplace α par une suite $\alpha_n \to 0$ lorsque $n \to \infty$.

Contributions

Deuxième partie de ma thèse

L'apport nouveau de ce travail consiste en l'ajout de deux difficultés supplémentaires dans le cadre de l'estimation de mesures de risque (VaR et CTE) de pertes pour des lois à queues lourdes.

- On ajoute la présence d'une covariable $X \in \mathbb{R}^p$.

- On s'intéresse à l'estimation de mesures de risque (VaR et CTE) de pertes extrêmes pour des lois à queues lourdes.

$\Longrightarrow$ Pour cela on remplace α par une suite $\alpha_n \to 0$ lorsque $n \to \infty$.

- $VaR(\alpha_n|x) = q(\alpha_n|x) = \overline{F}^{-1}(\alpha_n|x)$.
- $CTE(\alpha_n|x) = \mathbb{E}(Y|Y > VaR(\alpha_n|X), X = x)$.

Estimation du quantile conditionnel extrême

Pour estimer la fonction de survie conditionnelle de Y sachant $X = x$, on propose d'utiliser un estimateur à noyau introduit par Collomb (1980). Il est défini pour $(x, y) \in \mathbb{R}^p \times \mathbb{R}$ par

$$\widehat{\overline{F}}_n(y|x) = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \mathbb{I}\{Y_i > y\} \bigg/ \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right).$$

$\Longrightarrow$ Estimation non-paramétrique.

Estimation du quantile conditionnel extrême

Pour estimer la fonction de survie conditionnelle de Y sachant $X = x$, on propose d'utiliser un estimateur à noyau introduit par Collomb (1980). Il est défini pour $(x, y) \in \mathbb{R}^p \times \mathbb{R}$ par

$$\widehat{\overline{F}}_n(y|x) = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \mathbb{I}\{Y_i > y\} \Bigg/ \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right).$$

$\Longrightarrow$ Estimation non-paramétrique.

$\widehat{\overline{F}}_n(.|x)$ est une fonction décroissante on donc peut définir un estimateur de $\overline{F}^{-1}(.|x)$ par $\widehat{\overline{F}}_n^{-1}(.|x)$.

Estimation du quantile conditionnel extrême

Pour estimer la fonction de survie conditionnelle de Y sachant $X = x$, on propose d'utiliser un estimateur à noyau introduit par Collomb (1980). Il est défini pour $(x, y) \in \mathbb{R}^p \times \mathbb{R}$ par

$$\widehat{\overline{F}}_n(y|x) = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \mathbb{I}\{Y_i > y\} \Big/ \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right).$$

$\Longrightarrow$ Estimation non-paramétrique.

$\widehat{\overline{F}}_n(.|x)$ est une fonction décroissante on donc peut définir un estimateur de $\overline{F}^{-1}(.|x)$ par $\widehat{\overline{F}}_n^{-1}(.|x)$.

Estimateur de $q(\alpha_n|x)$

On peut donc estimer le quantile conditionnel extrême par

$$q(\alpha_n|x) = \widehat{\overline{F}}_n^{-1}(\alpha_n|x)$$

Estimation du quantile conditionnel extrême

Pour estimer la fonction de survie conditionnelle de Y sachant $X = x$, on propose d'utiliser un estimateur à noyau introduit par Collomb (1980). Il est défini pour $(x, y) \in \mathbb{R}^p \times \mathbb{R}$ par

$$\widehat{\overline{F}}_n(y|x) = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \mathbb{I}\{Y_i > y\} \Big/ \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right).$$

$\Longrightarrow$ Estimation non-paramétrique.

$\widehat{\overline{F}}_n(.|x)$ est une fonction décroissante on donc peut définir un estimateur de $\overline{F}^{-1}(.|x)$ par $\widehat{\overline{F}}_n^{-1}(.|x)$.

Estimateur de $q(\alpha_n|x)$

On peut donc estimer le quantile conditionnel extrême par

$$q(\alpha_n|x) = \widehat{\overline{F}}_n^{-1}(\alpha_n|x) = VaR(\alpha_n|x).$$

Conclusion

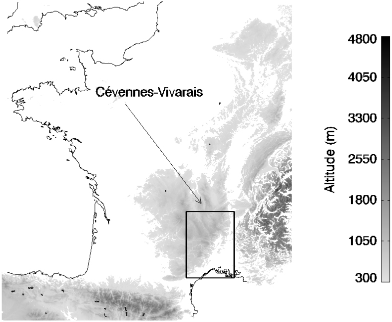
Conclusion

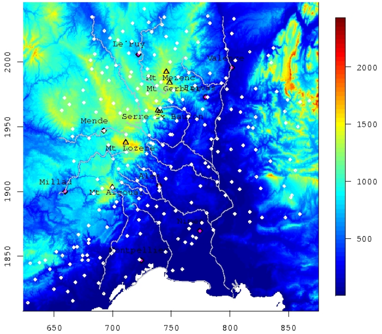
- Première partie terminée avec application à un jeu de données réelles qui a conduit à l'estimation de crues extrêmes.
- Il me reste à terminer la deuxième partie et à l'appliquer à un jeu de données réelles issu de la pluviométrie.

Conclusion

Conclusion

- Première partie terminée avec application à un jeu de données réelles qui a conduit à l'estimation de crues extrêmes.
- Il me reste à terminer la deuxième partie et à l'appliquer à un jeu de données réelles issu de la pluviométrie.





Principales références

- Gnedenko, B., (1943). Sur la distribution limite du terme maximum d'une série aléatoire, *The annals of Mathematics*, **44**, 423–453.
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., De Waal, D. and Ferro, C., (2004). Statistics of Extremes : Theory and Applications, *John Wiley & Sons*
- Embrechts, P., Kluppelberg, C. et Mikosh, T. (1997). Modelling Extremal Events, *Springer editions*.

Merci de votre attention.