
Act-7009 Gestion quantitative des risques

Chapitre 4 Données extrêmes et méthode du POT

Professeur : Etienne Marceau

12 novembre 2018

1 Introduction

On commence par considérer le cas d'un portefeuille de risques (contrats) homogènes.

Par la suite, on considère le cas de portefeuilles avec des risques hétérogènes.

2 Contexte

Considérons l'échantillon de montants de sinistres $\{x_1, \dots, x_n\}$.

On suppose que x_1, \dots, x_n sont des réalisations indépendantes de X_1, \dots, X_n .

On fait l'hypothèses que les montants de sinistres X_1, \dots, X_n sont i.d. et ont pour fonction de répartition F_X .

Cela est plausible si les sinistres proviennent d'une même portefeuille de risques que l'on suppose homogène.

On ne connaît pas F_X .

On cherche à identifier F_X .

3 Formats des données

Les données de sinistres peuvent se présenter sous forme complète, tronquée, groupée, ou censurée.

4 Étapes

Les étapes sont :

- Analyse exploratoire des données (moyenne, variance, qq-plot, histogramme, fonction d'excès moyen)
- Estimation des paramètres
- Adéquation et validation du modèle avec données (test d'adéquation du khi-deux, test de Kolmogorov, test d'Anderson-Darling, critère de Akaike)
- Choix de modèle (test du ratio de vraisemblance, etc.)

5 Fonction empirique

La fonction empirique est utile pour identifier la forme de F_X .

On définit la fonction empirique \hat{F}_n par

$$\hat{F}_n(x) = \frac{\sum_{i=1}^n \mathbf{1}_{(-\infty, x]}(x_i)}{n}$$

pour $x \in \mathbb{R}$.

La fonction empirique satisfait les propriétés d'une fonction de répartition.

Selon le théorème de Glivenko-Cantelli, on a

$$\sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F_X(x) \right| \rightarrow 0 \text{ (presque sûrement)}$$

lorsque $n \rightarrow \infty$.

Ainsi, \hat{F}_n peut offrir une bonne approximation de F_X lorsque n est grand.

On peut utiliser \hat{F}_n pour des fins de calculs actuariels.

Comme

$$n\hat{F}_n(x) = \sum_{i=1}^n 1_{(-\infty, x]}(x_i)$$

correspond à une somme de v.a. i.i.d. de loi Bernoulli avec probabilité $F_X(x)$, il en résulte

$$n\hat{F}_n(x) \sim \text{Bin}(n, F_X(x))$$

En vertu du théorème central limite, on a

$$\sqrt{n} \left(\hat{F}_n(x) - F_X \right) \xrightarrow{d} Norm(0, \sigma_x^2)$$

quand $n \rightarrow \infty$ où

$$\sigma_x^2 = Var \left(1_{(-\infty, x]}(X) \right) = F_X(x) (1 - F(x))$$

On utilise ce résultat pour construire un intervalle de confiance de niveau α (ex : $\alpha = 95\%$) pour l'estimateur de $\hat{F}_n(x)$ de $F_X(x)$

$$\hat{F}_n(x) \pm z_{1-\frac{(1-\alpha)}{2}} \frac{\sqrt{\hat{F}_n(x) (1 - \hat{F}_n(x))}}{\sqrt{n}}$$

où $z_{1-\frac{(1-\alpha)}{2}}$ est tel que

$$\Pr \left(Z \leq z_{1-\frac{(1-\alpha)}{2}} \right) = 1 - \frac{(1-\alpha)}{2}$$

avec $Z \sim \text{Norm}(0, 1)$.

6 Analyse graphique avec les q-q plot

Les *q-q plot* servent à faire une analyse graphique des données.

Définition - Fonction quantile. La fonction inverse généralisée d'une fonction de répartition F définie par

$$F^{\leftarrow}(\alpha) = \inf \{x \in \mathbb{R} : F(x) \geq \alpha\}$$

pour $0 < \alpha < 1$, est appelée la fonction quantile d'une fonction de répartition F .

On convient de définir $x_\alpha = F^{\leftarrow}(\alpha)$ comme étant le α -quantile de F .

Le théorème implique que

$$\hat{F}_n^{\leftarrow}(\alpha) \rightarrow F^{\leftarrow}(\alpha)$$

presque sûrement quand $n \rightarrow \infty$ en tout point de continuité de α de $F(\alpha)$.

Idée du qq-plot. Supposons que $\{x_1, \dots, x_n\}$ sont des réalisations i.i.d. provenant d'une v.a. de fonction de répartition continue. Alors, si on trace le graphique $\hat{F}_n^{\leftarrow}(\alpha)$ en fonction de $F^{\leftarrow}(\alpha)$ (pour $0 < \alpha < 1$) on doit s'attendre à une droite.

En pratique, on trace

$$\left\{ \left(x_{(k)}, F^{\leftarrow} \left(\frac{k}{n+1} \right), k = 1, 2, \dots, n \right) \right\}$$

pour une certaine fonction de répartition F .

7 Fonction empirique d'excès moyen

La fonction empirique est un outil d'adéquation graphique.

Considérons l'échantillon de montants de sinistres $\{x_1, \dots, x_n\}$.

On suppose que les éléments de l'échantillon sont déjà ordonnés de façon croissante, c'est-à-dire

$$x_1 < \dots < x_n.$$

La fonction d'excès moyen est estimée par la fonction empirique d'excès moyen

définie par

$$\begin{aligned}\hat{e}_n(y) &= \frac{\sum_{j=1}^n x_j \mathbf{1}_{(y, \infty)}(x_j)}{\sum_{j=1}^n \mathbf{1}_{(X_2, \infty)}(x_j)} - y \\ &= \frac{\sum_{j=1}^n x_j \mathbf{1}_{(y, \infty)}(x_j)}{\text{nombre de } x_j > y} - y.\end{aligned}$$

En traçant la courbe de la courbe de $\hat{e}_n(y)$ pour $y \in \{x_1, \dots, x_n\}$ avec

$$\hat{e}_n(x_k) = \frac{\sum_{j=k+1}^n x_j}{n - k} - x_k \quad (k = 1, 2, \dots, n - 1),$$

il est possible de faire un premier diagnostic quant au modèle à choisir pour la sévérité.

Si la fonction $\hat{e}_n(X_2)$ est croissante, alors la distribution possède une queue plus épaisse que la distribution exponentielle.

Si elle décroissante, alors la queue de la distribution est moins épaisse que celle de la distribution exponentielle.

Le *q-q plot* est aussi utilisé comme outil d'adéquation graphique.

Voir, e.g., Beirlant et al. [2] ou Klugman et al. [6].

8 Méthode du maximum de vraisemblance – données complètes

On dispose de l'échantillon complet des montants de sinistres $\{x_1, \dots, x_n\}$.

On considère un modèle paramétrique pour X avec fonction de densité $f_X(x; \underline{\theta})$ où $\underline{\theta} \in \Omega_{\underline{\theta}} \subseteq \mathbb{R}^h$ est le vecteur de h paramètres pour le modèle paramétrique de X .

On cherche à estimer $\underline{\theta}$.

La fonction de vraisemblance $L(\underline{\theta})$ est définie par

$$L(\underline{\theta}) = \prod_{j=1}^n f_X(x_j; \underline{\theta}) .$$

On cherche la valeur de $\underline{\theta}$ qui maximise la fonction de vraisemblance $L(\underline{\theta})$ i.e.

$$\hat{\underline{\theta}}_{MLE} = \arg \max_{\underline{\theta} \in \Omega_{\theta}} L(\underline{\theta})$$

Cela revient à maximiser le logarithme népérien de $L(\underline{\theta})$, où

$$\begin{aligned} l(\underline{\theta}) &= \ln \{L(\underline{\theta})\} \\ &= \sum_{j=1}^n \ln \{f_X(x_j; \underline{\theta})\}. \end{aligned}$$

Alors, on cherche

$$\hat{\underline{\theta}}_{MLE} = \arg \max_{\underline{\theta} \in \Omega_{\theta}} l(\underline{\theta})$$

L'estimateur $\hat{\underline{\theta}}_{MLE}$ correspond à la solution du système des h équations sui-

vantes

$$\begin{aligned}\frac{\partial l(\underline{\theta})}{\partial \theta_1} &= \sum_{j=1}^n \frac{\partial \ln \{f_X(x_j; \underline{\theta})\}}{\partial \theta_1} = 0 \\ &\vdots \\ \frac{\partial l(\underline{\theta})}{\partial \theta_h} &= \sum_{j=1}^n \frac{\partial \ln \{f_X(x_j; \underline{\theta})\}}{\partial \theta_h} = 0.\end{aligned}$$

Supposons que les vrais paramètres du modèle paramétrique sont $\underline{\theta}_0$.

Selon certaines conditions de régularités, l'estimateur $\hat{\underline{\theta}}_{MLE}$ possèdent les propriétés suivantes.

Propriétés de $\hat{\underline{\theta}}_{MLE}$.

- l'estimateur $\hat{\underline{\theta}}_{MLE}$ est consistant, i.e. pour tout $\varepsilon > 0$, on a

$$\lim_{n \rightarrow \infty} \Pr \left(\left| \hat{\underline{\theta}}_{MLE} - \underline{\theta}_0 \right| > \varepsilon \right) \rightarrow 0$$

ou

$$\hat{\underline{\theta}}_{MLE} \xrightarrow{d} \underline{\theta}_0$$

- l'estimateur $\hat{\underline{\theta}}_{MLE}$ est asymptotiquement sans biais

$$\lim_{n \rightarrow \infty} E \left[\hat{\underline{\theta}}_{MLE} \right] = \underline{\theta}_0$$

- l'estimateur $\hat{\underline{\theta}}_{MLE}$ est asymptotiquement normal

$$\sqrt{n} \left(\hat{\underline{\theta}}_{MLE} - \underline{\theta} \right) \xrightarrow{d} Norm \left(\underline{0}, I^{-1} \right)$$

où

$$I = E \left(\begin{array}{ccc|c} \frac{\partial^2 l(\underline{\theta})}{\partial \theta_1^2} & \cdots & \frac{\partial^2 l(\underline{\theta})}{\partial \theta_1 \partial \theta_h} & \\ \frac{\partial^2 l(\underline{\theta})}{\partial \theta_h \partial \theta_1} & & \frac{\partial^2 l(\underline{\theta})}{\partial \theta_h^2} & \end{array} \middle|_{\underline{\theta} = \hat{\underline{\theta}}_0} \right)$$

- l'estimateur $\hat{\underline{\theta}}_{MLE}$ est asymptotiquement efficient i.e. parmi les estimateurs dont la loi asymptotique est normale, il est celui ayant la plus petite variance. Cette variance correspond à la borne de Rao-Cramer.

Important. Ces propriétés sont satisfaites si le modèle est correctement spécifié.

On considère les exemples suivants.

Exemple (Loi exponentielle). La loi exponentielle compte un seul paramètre, λ ($h = 1$). Il en résulte que

$$\begin{aligned} L(\lambda) &= \prod_{j=1}^n \lambda \exp(-\lambda x_j) \\ &= \lambda^n \prod_{j=1}^n \exp(-\lambda x_j) \end{aligned}$$

puis

$$\begin{aligned} l(\lambda) &= \ln \{L(\lambda)\} \\ &= n \ln \{\lambda\} - \sum_{j=1}^n \lambda x_j. \end{aligned}$$

On détermine la dérivée de $l(\lambda)$ par rapport à λ et, en isolant λ , on obtient

$$\begin{aligned}\hat{\lambda}_{MLE} &= \frac{1}{\left(\frac{\sum_{j=1}^n x_j}{n}\right)} \\ &= \frac{1}{\text{moyenne empirique}}.\end{aligned}$$

□

Exemple : La loi lognormale compte deux paramètres, μ et σ^2 ($h = 2$).

Si X obéit à une loi lognormale de paramètres μ et σ^2 , alors $Y = \ln(X)$ est de loi normale avec les paramètres μ et σ^2 .

À partir de cette relation, on déduit que les estimateurs de μ et σ^2 du maximum de vraisemblance sont

$$\hat{\mu} = \frac{1}{n} \sum_{j=1}^n \ln(x_j)$$

et

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{j=1}^n \left\{ \ln(x_j) - \hat{\mu} \right\}^2.$$



En général, on doit avoir recours à des méthodes numériques telles que l'algorithme de Newton-Raphson (voir Klugman et collab. (1998) pour les détails).

Exemple : La loi Pareto compte deux paramètres, α et λ ($h = 2$).

La fonction de vraisemblance et le logarithme naturel de la fonction de vraisemblance sont

$$\begin{aligned} L(\alpha, \lambda) &= \prod_{j=1}^n \frac{\alpha \lambda^\alpha}{(\lambda + x_j)^{\alpha+1}} \\ &= (\alpha \lambda^\alpha)^n \prod_{j=1}^n \frac{1}{(\lambda + x_j)^{\alpha+1}} \end{aligned}$$

et

$$\begin{aligned} l(\alpha, \lambda) &= \ln \{L(\alpha, \lambda)\} \\ &= n \ln \{\alpha\} + n\alpha \ln \{\lambda\} - \sum_{j=1}^n (\alpha + 1) \ln \{\lambda + x_j\}. \end{aligned}$$

On détermine les dérivées partielles de $l(\alpha, \lambda)$ par rapport à α et à λ

$$\frac{\partial l(\alpha, \lambda)}{\partial \alpha} = \frac{n}{\alpha} + n \ln \{\lambda\} - \sum_{j=1}^n \ln \{\lambda + x_j\} = 0$$

$$\frac{\partial l(\alpha, \lambda)}{\partial \lambda} = \frac{n\alpha}{\lambda} - \sum_{j=1}^n \frac{\alpha + 1}{\lambda + x_j} = 0.$$

L'algorithme de Newton-Raphson (voir Klugman et collab. (1998) pour les détails) peut être utilisé pour isoler α et λ .

Approche alternative : Utiliser un outil d'optimisation.

- EXCEL : Solveur. On peut trouver $\hat{\underline{\theta}}_{MLE}$ avec EXCEL en utilisant la macro SOLVEUR. On maximise directement $\ln L(\underline{\theta})$ avec SOLVEUR

$$\begin{aligned} l(\underline{\theta}) &= \ln \{L(\underline{\theta})\} \\ &= \sum_{j=1}^n \ln \{f_X(x_j; \underline{\theta})\} \end{aligned}$$

Cela est facile à appliquer. Cela fonctionne généralement bien quand le nombre de paramètres n'est pas trop élevé.

- R : Fonction Optim. Facile d'application.

9 Méthode du maximum de vraisemblance – données groupées

On suppose que les montants de sinistres de l'échantillon sont disponibles sous forme de données groupées.

On définit les bornes c_0, c_1, \dots, c_r de telle sorte que

$$0 \leq c_0 < c_1 < \dots < c_r \leq \infty$$

et on compte n_i sinistres dont les valeurs sont incluses dans $[c_{i-1}, c_i[$ pour $i = 1, 2, \dots, r$. On a

$$n_1 + \dots + n_r = n.$$

On considère un modèle paramétrique pour X avec fonction de densité $f_X(x; \underline{\theta})$ où $\underline{\theta}$ est le vecteur avec h paramètres pour le modèle paramétrique de X . On cherche à estimer $\underline{\theta}$.

Compte tenu des observations disponibles, la fonction de vraisemblance $L(\underline{\theta})$ est définie par

$$L(\underline{\theta}) = \prod_{i=1}^r \{F_X(c_i; \underline{\theta}) - F_X(c_{i-1}; \underline{\theta})\}^{n_i}.$$

On cherche la valeur de $\underline{\theta}$ qui maximise la fonction de vraisemblance $L(\underline{\theta})$ ou celle du logarithme népérien de $L(\underline{\theta})$, avec

$$\begin{aligned} l(\underline{\theta}) &= \ln \{L(\underline{\theta})\} \\ &= \sum_{i=1}^r n_i \ln \{F_X(c_i; \underline{\theta}) - F_X(c_{i-1}; \underline{\theta})\}. \end{aligned}$$

Dans le cas de la loi exponentielle qui compte un seul paramètre, λ ($h = 1$), on a

$$L(\lambda) = \prod_{i=1}^r \{\exp(-\lambda c_{i-1}) - \exp(-\lambda c_i)\}^{n_i}$$

avec

$$\begin{aligned} l(\lambda) &= \ln \{L(\lambda)\} \\ &= \sum_{i=1}^r n_i \ln \{\exp(-\lambda c_{i-1}) - \exp(-\lambda c_i)\}. \end{aligned}$$

La dérivée de $l(\lambda)$ par rapport à λ est

$$\frac{\partial l(\lambda)}{\partial \lambda} = \sum_{i=1}^r n_i \frac{c_i \exp(-\lambda c_i) - c_{i-1} \exp(-\lambda c_{i-1})}{\exp(-\lambda c_{i-1}) - \exp(-\lambda c_i)}.$$

L'estimateur $\hat{\lambda}$ du maximum de vraisemblance de λ est la solution de

$$\sum_{i=1}^r n_i \frac{c_{i-1} \exp(-\lambda c_{i-1}) - c_i \exp(-\lambda c_i)}{\exp(-\lambda c_{i-1}) - \exp(-\lambda c_i)} = 0$$

qui ne peut être obtenue qu'avec l'aide de méthodes numériques telle que l'algorithme de Newton-Raphson (voir Klugman et collab. (1998) pour les détails).

La loi Pareto compte deux paramètres, α et λ ($h = 2$). La fonction de vraisemblance et le logarithme naturel de la fonction de vraisemblance sont respectivement

$$L(\alpha, \lambda) = \prod_{i=1}^r \left\{ \left(\frac{\lambda}{\lambda + c_{i-1}} \right)^{\alpha} - \left(\frac{\lambda}{\lambda + c_i} \right)^{\alpha} \right\}^{n_i}$$

et

$$\begin{aligned} l(\alpha, \lambda) &= \ln \{L(\alpha, \lambda)\} \\ &= \sum_{i=1}^r n_i \ln \left\{ \left(\frac{\lambda}{\lambda + c_{i-1}} \right)^\alpha - \left(\frac{\lambda}{\lambda + c_i} \right)^\alpha \right\}. \end{aligned}$$

L'algorithme de Newton-Raphson (voir Klugman et collab. (1998) pour les détails) peut être utilisé pour isoler α et λ , une fois que les dérivées par rapport à chacun des paramètres ont été calculées.

Approche alternative : Utiliser un outil d'optimisation fourni dans un logiciel.

Propriétés de $\hat{\theta}_{MLE}$. Identiques à celles énoncées dans le cas des données complètes.

10 Méthode du maximum de vraisemblance – données tronquées à gauche

On suppose que les données se présentent en étant tronquées à gauche.

Pour $d \geq 0$, cela signifie que l'on dispose uniquement des données qui sont supérieures à d . On n'a pas les données qui sont inférieures à d .

On définit

$$X^{(d)} = \begin{cases} 0, & X < d \\ X, & X > d \end{cases}$$

ou $X^{(d)} = X \times \mathbf{1}_{(d, \infty)}(X)$. On déduit que

$$\begin{aligned} F_{X^{(d)}}(x) &= \Pr(X^{(d)} \leq x) = \Pr(X \leq x | X > d) \\ &= \frac{\Pr(d < X \leq x)}{\Pr(X > d)} = \frac{F_X(x) - F_X(d)}{1 - F_X(d)} \end{aligned}$$

De plus, on a

$$f_{X^{(d)}}(x) = f_{X|X>d}(x) = \frac{f_X(x)}{1 - F_X(d)}.$$

Il en résulte que la fonction de vraisemblance s'écrit sous la forme suivante :

$$L(\underline{\theta}) = \prod_{i=1}^r f_{X^{(d)}}(x_i; \underline{\theta}) = \prod_{i=1}^r \frac{f_X(x_i; \underline{\theta})}{1 - F_X(d; \underline{\theta})} = \frac{1}{(\bar{F}_X(d; \underline{\theta}))^r} \prod_{i=1}^r f_X(x_i; \underline{\theta}).$$

La fonction $l(\underline{\theta})$ devient

$$\begin{aligned} l(\underline{\theta}) &= \ln \{L(\underline{\theta})\} \\ &= \prod_{i=1}^r \ln \{f_X(x_i; \underline{\theta})\} - r \ln \left(\overline{F}_X(d; \underline{\theta}) \right). \end{aligned}$$

Exemple. On suppose que $X \sim \text{Exp}(\beta)$. On a

$$\begin{aligned} l(\underline{\theta}) &= \prod_{i=1}^r \ln \{f_X(x_i; \underline{\theta})\} - r \ln (\overline{F}_X(d; \underline{\theta})) . \\ &= \prod_{i=1}^r \ln (\beta e^{-\beta x_i}) - r \ln e^{-\beta d} \\ &= r \ln (\beta) - \beta \sum_{i=1}^r x_i + r\beta d. \end{aligned}$$

On obtient

$$\hat{\beta} = \frac{1}{\frac{1}{r} \sum_{i=1}^r (x_i - d)}.$$

□

11 Méthode du maximum de vraisemblance – données censurées à droite

On suppose que les données se présentent en étant censurées à droite

Pour $d \geq 0$, cela signifie que l'on dispose uniquement des données qui sont inférieures ou à égales à d . On n'a pas les données qui sont inférieures à d .

On définit $X^{(d)} = \min(X; d)$. Alors, $X^{(d)}$ est une v.a. mixte : continue sur $(0, d)$ avec une masse à d .

Supposons qu'il y ait r_0 ($r_0 < r$) données qui soient inférieures à d et que $x_i = d$ pour $i = r_0 + 1, \dots, r$

Il en résulte que la fonction de vraisemblance s'écrit sous la forme suivante :

$$L(\underline{\theta}) = \prod_{i=1}^{r_0} f_X(x_i; \underline{\theta}) \prod_{i=r_0+1}^r \overline{F}_X(d; \underline{\theta}).$$

La fonction $l(\underline{\theta})$ devient

$$\begin{aligned} l(\underline{\theta}) &= \ln \{L(\underline{\theta})\} \\ &= \prod_{i=1}^{r_0} \ln \{f_X(x_i; \underline{\theta})\} + (r - r_0) \ln \left(\overline{F}_X(d; \underline{\theta}) \right). \end{aligned}$$

Exemple. On suppose que $X \sim \text{Exp}(\beta)$. On a

$$\begin{aligned}
 l(\underline{\theta}) &= \prod_{i=1}^{r_0} \ln \{f_X(x_i; \underline{\theta})\} + (r - r_0) \ln (\overline{F}_X(d; \underline{\theta})) . \\
 &= \prod_{i=1}^r \ln (\beta e^{-\beta x_i}) + (r - r_0) \ln e^{-\beta d} \\
 &= r \ln (\beta) - \beta \sum_{i=1}^{r_0} x_i - (r - r_0) \beta d.
 \end{aligned}$$

Après les calculs habituels, on obtient

$$\hat{\beta} = \frac{1}{\frac{1}{r} \left(\sum_{i=1}^{r_0} x_i + (r - r_0) d \right)}.$$

□

12 Tests d'adéquation

Il est important de vérifier si le modèle offre une description appropriée du comportement des données.

Approches complémentaires :

- Outils d'analyse graphique
- Tests d'adéquation

Outils d'analyse graphique :

- *qq-plot*
- fonction d'excès moyen empirique

Tests d'adéquation :

- Test de Kolmogorov-Smirnov
- Test de Anderson-Darling
- Test d'adéquation du khi-deux

Les tests de Kolmogorov-Smirnov et Anderson-Darling s'appliquent sur des données continues.

Les données doivent être groupées pour utiliser le test d'adéquation du khi-deux, qui s'applique aussi pour les données discrètes.

12.1 Test de Kolmogorov-Smirnov.

On suppose que la fonction de la répartition, notée F_X , de X est continue

On a recours à la statistique de Kolmogorov-Smirnov

$$D_n = \max\left(\sup_{x_i} \left| \hat{F}_n(x_i) - F_X(x_i) \right|, \sup_{x_i} \left| \hat{F}_n(x_i-) - F_X(x_i) \right| \right)$$

où $\hat{F}_n(x_i-)$ est la valeur de la fonction empirique juste avant le saut à x_i .

Cela a nécessité que l'on teste $2n$ valeurs.

Sous l'hypothèse H_0 , on ne connaît pas la nature exacte de la distribution de la statistique D_n .

Soit c_α tel que

$$\Pr(D_n > c_\alpha) = \alpha.$$

Les valeurs critiques c_α pour certaines valeurs de niveau de confiance α sont fournies dans le tableau ci-dessous :

α	20%	10%	5%	1%
c_α	$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Les approximation sont bonnes si $n > 15$.

On rejette H_0 (le modèle) si

$$D_n > c_\alpha$$

On peut construire un intervalle (bande) de confiance autour de $F(x)$.

On a

$$\Pr \left(\max(\hat{F}_n(x) - F(x)) \leq c_\alpha \right) = 1 - \alpha$$

ce qui est équivalent à

$$\Pr \left(\hat{F}_n(x) - c_\alpha \leq F(x) \leq \hat{F}_n(x) + c_\alpha \right) = 1 - \alpha$$

On déduit les bornes

$$\begin{aligned} F_L(x) &= \max \left(0; \hat{F}_n(x) - c_\alpha \right) \\ F_U(x) &= \min \left(1; \hat{F}_n(x) + c_\alpha \right) \end{aligned}$$

pour $F(x)$.

Le test de Kolmogorov-Smirnov ne s'applique pas sur des données groupées.

12.2 Test de Anderson-Darling

Le test de Kolmogorov-Smirnov a l'avantage d'être non paramétrique. Néanmoins, ce test n'est pas bon pour identifier les problèmes éventuels dans la queue de la distribution. A cet égard, le test d'Anderson-Darling est une modification du test de Kolmogorov-Smirnov qui vise à corriger ce défaut.

(...)

13 Test d'adéquation du khi-deux

Pour appliquer le test du khi-deux, les données doivent être groupées. Si on dispose de données complètes, il suffit de les grouper.

Objectif : On veut vérifier si le modèle choisi est adéquat pour les données (fit to the data)

Statistique du test :

$$Q = \sum_{i=1}^r \frac{(n_i - E_i)^2}{E_i}$$

On définit E_j = nombre espérés d'observations

$$\begin{aligned} E_i &= n \{F_X(c_i; \underline{\theta}) - F_X(c_{i-1}; \underline{\theta})\} \\ &= n \Pr(c_{i-1} < X \leq c_i; \underline{\theta}) \end{aligned}$$

Convention habituelle :

$$E_i \geq 5$$

On définit une v.a. Z qui obéit à une loi du chi-deux

$$d.f. = (nbclasses) - (nbparam) - 1$$

On définit $Z_{\alpha,df}$ (appelée valeur critique) tel que

$$\Pr(Z > Z_{\alpha,df}) = \alpha \quad (e.g. \alpha = 5\%)$$

On rejette le modèle (H_0) si

$$Q > Z_{\alpha,df}$$

On définit la p_{value}

$$p_{value} = \Pr(Z > Q)$$

De façon équivalente, on rejette le modèle (H_0) si

$$p_{value} < \alpha$$

14 Sélection d'un modèle

Principe de base. Il est important de choisir le modèle le plus simple possible. Il est recommandé de respecter le principe de parsimonie : le modèle le plus simple est toujours préféré au modèle plus complexe.

On doit se servir de son jugement lors de la sélection d'un modèle.

On peut utiliser les outils d'analyse graphique afin d'aider à la sélection du modèle.

Souvent, dans la pratique, on doit aussi tenir compte des contraintes de temps et de la mise en place du modèle.

On a aussi recours à des méthodologies formelles et informelles. Un exemple de méthodologie formelle est le test de ratio de vraisemblance. Le test du ratio de vraisemblance permet de comparer des modèles provenant d'une même famille. Exemple : exponentielle et gamma. Toutefois, très souvent, on doit comparer des lois qui ne sont pas de la même familles. Exemple : gamma, lognormale, Pareto.

14.1 Test du ratio de vraisemblance

Le test de ratio de vraisemblance permet de choisir entre deux modèles, où l'un est le cas particulier de l'autre. On dit que les modèles sont imbriqués.

Exemples de modèles imbriqués :

- Exponentielle vs Gamma
- Mélanges de $K - 1$ expon vs mélanges de K expon ($K = 2, 3, \dots$)

etc.

Exemples de modèles non-imbriqués :

- Exponentielle vs lognormale
- Lognormale vs Pareto
- Lognormale vs Gamma
- Gamma vs Pareto.

Idée du test. On considère deux modèles imbriqués avec paramètres $\underline{\theta}_0$ et $\underline{\theta}_1$ en supposant

$$nbparam(\underline{\theta}_0) < nbparam(\underline{\theta}_1)$$

Statistic du test

$$R = 2 \{ \ln L(\underline{\theta}_1) - \ln L(\underline{\theta}_0) \}$$

Soit W une v.a qui obéit à un loi du chi-deux

$$d.f. = nbparam(\underline{\theta}_1) - nbparam(\underline{\theta}_0)$$

Soit $w_{\alpha, df}$ (appelée valeur critique) tel que

$$\Pr(W > w_{\alpha, df}) = \alpha \quad (e.g. \alpha = 5\%)$$

On rejette H_0 (modèle avec $\underline{\theta}_0$) si

$$R > w_{\alpha, df}.$$

14.2 Méthodologies informelles

On a recours à des méthodologies informelles, qui sont basées sur des *scores*.

- Comparer les valeurs des statistiques du test de Kolmogorov-Smirnov. On choisit le modèle avec la plus petite valeur.
- Comparer les valeurs des statistiques du test de Anderson-Darling. On choisit le modèle avec la plus petite valeur.
- Comparer les p – *value* produites par le test d'adéquation du khi-deux. On choisit le modèle avec la plus grande valeur.
- Comparer les valeurs du critère de l'information de Akaike. On choisit le modèle avec la plus petite valeur.
- Comparer les valeurs du critère bayésien de Schwartz. On choisit le modèle avec la plus petite valeur.

Le critère de l'information de Akaike et le critère bayésien de Schwartz reposent sur la même idée. On compare les valeurs des log des fonctions de vraisemblance et choisit le modèle avec la valeur la plus élevée. Toutefois, afin de tenir compte du nombre de caractères les corrections ont été suggérées :

- Critère de l'information de Akaike :

$$AIC = 2h - 2l(\underline{\theta}_h)$$

où h = nombre de paramètres.

- Critère bayésien de Schwartz :

$$BSC = h \ln(r) - 2l(\underline{\theta}_h)$$

où h = nombre de paramètres et r = nb de données.

Comme on le constate, les deux critères sont de la même forme. Il vise à pénaliser pour le nombre de paramètres. Le critère bayésien de Schwartz pénalise davantage.

15 Intervalles de confiance

Parmi ses propriétés, on sait que l'estimateur $\hat{\underline{\theta}}_{MLE}$ est asymptotiquement normal

$$\sqrt{n} \left(\hat{\underline{\theta}}_{MLE} - \underline{\theta} \right) \xrightarrow{d} Norm \left(\underline{0}, I^{-1} \right)$$

où

$$I = E \left(\begin{array}{ccc} \frac{\partial^2 l(\underline{\theta})}{\partial \theta_1^2} & \cdots & \frac{\partial^2 l(\underline{\theta})}{\partial \theta_1 \partial \theta_h} \\ \frac{\partial^2 l(\underline{\theta})}{\partial \theta_h \partial \theta_1} & & \frac{\partial^2 l(\underline{\theta})}{\partial \theta_h^2} \end{array} \middle| \underline{\theta} = \hat{\underline{\theta}}_0 \right)$$

On utilise les composantes de I^{-1} pour construire des intervalles de confiance

pour les paramètres

$$\hat{\underline{\theta}}_{MLE} = (\hat{\theta}_{1,MLE}, \dots, \hat{\theta}_{h,MLE})$$

de $(\theta_1, \dots, \theta_h)$ où

$$I^{-1} = \begin{pmatrix} Var(\hat{\theta}_{1,MLE}) & Cov(\hat{\theta}_{1,MLE}, \hat{\theta}_{2,MLE}) & Cov(\hat{\theta}_{1,MLE}, \hat{\theta}_{h,MLE}) \\ Cov(\hat{\theta}_{2,MLE}, \hat{\theta}_{1,MLE}) & Var(\hat{\theta}_{2,MLE}) & Cov(\hat{\theta}_{2,MLE}, \hat{\theta}_{h,MLE}) \\ Cov(\hat{\theta}_{h,MLE}, \hat{\theta}_{1,MLE}) & Cov(\hat{\theta}_{h,MLE}, \hat{\theta}_{2,MLE}) & Var(\hat{\theta}_{h,MLE}) \end{pmatrix}$$

L'intervalle de confiance de niveau α est

$$\hat{\theta}_{i,MLE} \pm t_{n-1, \frac{\alpha}{2}} \sqrt{Var(\hat{\theta}_{i,MLE})}$$

avec

$$\Pr\left(Z > t_{n-1, \frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

où Z obéit à une loi de Student de $n - 1$ degrés de liberté n (ex : $\alpha = 1\%, 5\%$ ou 10%).

Quand $n \rightarrow \infty$, la loi de Z tend vers une loi $N(0, 1)$.

On veut maintenant construire un intervalle de confiance pour une fonction de h paramètres $g(\underline{\theta})$.

On a

$$g(\hat{\theta}_{MLE}) \pm t_{n-1, \frac{\alpha}{2}} \sqrt{g'(\underline{\theta})^t I^{-1} g'(\underline{\theta})}$$

où

$$g'(\underline{\theta})^t = \left(\frac{\partial g(\underline{\theta})}{\partial \theta_1}, \dots, \frac{\partial g(\underline{\theta})}{\partial \theta_h} \right).$$

Cette méthode est appelée la méthode delta.

16 Estimation des distributions avec montants de sinistres élevés

On examine l'estimation de sinistres avec montants élevés.

On introduit aussi la méthode du POT.

Le contenu de cette section est basé sur, e.g., McNeil [7], Embrechts et al. [5], Cizek et al. [3], Beirlant et al. [2], McNeil et al. [5], Albrecher et al. [1].

Pour une illustration récente des présentées dans cette section, voir, e.g., Cizek et al. [3] et Prettenthaler et al. [9].

17 Rappels

17.1 Outils d'analyse

Quantile :

$$q_X(\kappa) = F_X^{-1}(\kappa) = \inf \{x \in \mathbb{R}, F_X(x) \geq \kappa\}$$

Fonction d'excès moyen ou espérance résiduelle :

$$\begin{aligned}e_X(d) &= E[X - d | X > d] \\&= E[X | X > d] - d \\&= \frac{E[\max(X - d; 0)]}{1 - F(d)} \\&= \frac{\int_d^\infty \{1 - F_X(x)\} dx}{1 - F(d)} \\&= \frac{E[X] - E[\min(X; d)]}{1 - F(d)}.\end{aligned}$$

Échantillon aléatoire ordonné : $\{x_1, \dots, x_n\}$ i.i.d. provenant d'une même loi

Fonction de répartition empirique \hat{F}_n :

$$\hat{F}_n(x) = \frac{\sum_{i=1}^n \mathbf{1}_{\{x_i \leq x\}}}{n}$$

pour $x \in \mathbb{R}$.

Quantile empirique :

$$\hat{Q}_n(\kappa) = F_n^{-1}(\kappa) = \inf \left\{ x \in \mathbb{R}, \hat{F}_n(x) \geq \kappa \right\}$$

Fonction d'excès moyen empirique :

$$\begin{aligned} \hat{e}_n(d) &= \frac{\sum_{j=1}^n x_j \mathbf{1}_{\{x_j > d\}}}{\sum_{j=1}^n \mathbf{1}_{\{x_j > d\}}} - d \\ &= \frac{\sum_{j=1}^n x_j \mathbf{1}_{\{x_j > d\}}}{\text{nombre de } x_j > d} - d. \end{aligned}$$

Q-q plot :

$$\left\{ \left(x_{(k)}, F_X^{-1} \left(\frac{k}{n+1} \right), k = 1, 2, \dots, n \right) \right\}$$

17.2 Lois continues

Loi Pareto (paramètres $\alpha > 1$ pour le chapitre et $\lambda > 0$)

– Quantile :

$$q_X(\kappa) = \frac{\lambda}{(1 - \gamma)^{\frac{1}{\alpha}}} - \lambda$$

– Fonction d'excès moyen :

$$e_X(d) = \frac{\lambda}{(\alpha - 1)} + \frac{d}{(\alpha - 1)} = E[X] + \frac{d}{(\alpha - 1)}$$

Loi Exponentielle (paramètre $\beta > 0$)

– Quantile :

$$q_X(\kappa) = -\frac{1}{\beta} \ln(1 - \kappa)$$

– Fonction d'excès moyen :

$$e_X(d) = \frac{1}{\beta} = E[X]$$

18 Comportement de la valeur extrême

Soit un nombre n de v.a. i.i.d. X_1, \dots, X_n avec fonction de répartition F_X

On définit

$$M_n = \max(X_1, \dots, X_n)$$

La fonction de répartition de M_n est donnée par

$$\begin{aligned} F_{M_n}(x) &= \Pr(M_n \leq x) = \Pr(\max(X_1, \dots, X_n) \leq x) \\ &= \Pr(X_1 \leq x, \dots, X_n \leq x) \\ &= \Pr(X_1 \leq x) \times \dots \times \Pr(X_n \leq x) \text{ (indépendance)} \\ &= (F_X(x))^n \text{ (i.d.)} \end{aligned}$$

On veut examiner le comportement asymptotique de M_n (i.e. quand $n \rightarrow \infty$)

A cette fin, on définit des distributions extrêmes.

19 Distributions de la valeur extrême

Il y a trois distributions de la valeur extrême.

Distribution Gumbel. Fonction de répartition :

$$F(x) = \exp(-\exp(x))$$

pour $-\infty < x < \infty$.

Avec des paramètres de location et d'échelle $\mu \in \mathbb{R}$ et $\theta > 0$, on a

$$F(x) = \exp\left(-\exp\left(\frac{x - \mu}{\sigma}\right)\right)$$

pour $-\infty < x < \infty$.

Distribution Fréchet. Fonction de répartition :

$$F(x) = \exp\left(-x^{-\alpha}\right),$$

pour $x \geq 0$ et $\alpha > 0$.

Avec des paramètres de location $\mu \in \mathbb{R}$ et d'échelle $\sigma > 0$, on a

$$F(x) = \exp\left(-\left(\frac{x - \mu}{\sigma}\right)^{-\alpha}\right), \quad x \geq \mu.$$

Distribution Weibull. Fonction de répartition :

$$F(x) = \exp\left(-(-x)^{-\alpha}\right),$$

pour $x \leq 0$ et $\alpha < 0$.

Avec des paramètres de location $\mu \in \mathbb{R}$ et d'échelle $\sigma > 0$, on a

$$F(x) = \exp \left(- \left(-\frac{x - \mu}{\sigma} \right)^{-\alpha} \right),$$

pour $x \leq \mu$ et $\alpha < 0$. Cette forme n'est pas à confondre avec la forme présentée en annexe.

Elle n'est pas appliquée en actuariat.

Il est possible de regrouper ces trois sous la forme d'une seule loi générale, appelée la distribution généralisée de la valeur extrême.

20 Distribution généralisée de la valeur extrême

La distribution généralisée de la valeur extrême est utilisée pour décrire le comportement du maximum d'un échantillon de v.a. i.i.d.

On pose $\xi = \frac{1}{\alpha}$ dans les distributions de Gumbel, Fréchet et Weibull.

On conserve le paramètre d'échelle σ .

La forme standard de la fonction de répartition de cette distribution est donnée

par

$$H_{\xi, \sigma}(x) = \begin{cases} \exp \left\{ - \left(\frac{1}{1 + \frac{\xi x}{\sigma}} \right)^{\frac{1}{\xi}} \right\}, & \xi < 0, 0 < x < -\frac{\sigma}{\xi} \quad (Weibull) \\ e^{\{-\exp(-\frac{x}{\sigma})\}}, & \xi = 0, x > 0 \quad (Gumbel) \\ \exp \left\{ - \left(\frac{1}{1 + \frac{\xi x}{\sigma}} \right)^{\frac{1}{\xi}} \right\}, & \xi > 0, x > 0 \quad (Fréchet) \end{cases} .$$

Lorsque $\xi = 0$, on retrouve le cas particulier de la distribution Gumbel alors que pour $\xi > 0$ et $\xi < 0$, on obtient respectivement la distribution Fréchet et la distribution Weibull.

Le paramètre σ est un paramètre d'échelle.

On peut aussi ajouter un paramètre de localisation μ conduisant à la fonction de répartition suivante :

$$\begin{aligned}
 H_{\xi, \sigma, \mu}(x) &= H_{\xi, \sigma}(x - \mu) \\
 &= \begin{cases} \exp \left\{ - \left(\frac{1}{1 + \frac{\xi(x-\mu)}{\sigma}} \right)^{\frac{1}{\xi}} \right\}, & \xi < 0, 0 < x < -\frac{\sigma}{\xi} \quad (Weibull) \\ e^{\left\{ -\exp\left(-\frac{x-\mu}{\sigma}\right) \right\}}, & \xi = 0, x > 0 \quad (Gumbel) \\ \exp \left\{ - \left(\frac{1}{1 + \frac{\xi(x-\mu)}{\sigma}} \right)^{\frac{1}{\xi}} \right\}, & \xi > 0, x > 0 \quad (Fréchet) \end{cases} .
 \end{aligned}$$

On examine l'utilisation de cette distribution dans la prochaine sous-section. On dit que $H_{\xi, \sigma, \mu}$ et $H_{\xi, \sigma}$ sont de type H_{ξ} où

$$H_{\xi} = H_{\xi, 1}.$$

On passe au théorème Fisher-Tippett.

21 Théorème Fisher-Tippett

On considère une suite de v.a. i.i.d. X_1, X_2, \dots dont la fonction de répartition commune F_X est inconnue.

On définit le maximum des n premières v.a. par

$$M_n = \max(X_1, \dots, X_n).$$

On suppose qu'il existe une suite de nombres réels $a_n > 0$ et b_n telle que

$$\begin{aligned} \Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) &= \Pr(M_n \leq a_n x + b_n) \\ &= \{\Pr(X \leq a_n x + b_n)\}^n \\ &= F_X(a_n x + b_n)^n \\ &\rightarrow H(x), \text{ quand } n \rightarrow \infty, \end{aligned}$$

où $H(x)$ est une fonction de répartition non dégénérée.

Si cette condition est respectée, on dit que F_X est élément du domaine d'attraction du maximum de la distribution H , notée par

$$F_X \in DAM(H).$$

Selon le théorème Fisher-Tippett, la seule fonction de répartition H non dégénérée est la fonction de répartition de la distribution généralisée de la valeur extrême H_ξ en supposant que les suites de nombres réels $a_n > 0$ et b_n est choisie de façon appropriée.

Il y a trois classes de distributions $F_X \in DAM(H_\xi)$ selon que $\xi < 0$, $\xi = 0$ et $\xi > 0$.

- La classe de distributions appartenant au domaine d'attraction de la distribution Fréchet

$$F_X \in DAM(H_\xi), \quad \xi > 0,$$

correspondent aux distributions *heavy-tailed* (à queue épaisse) et inclut les distributions Pareto, Burr, loggamma, Cauchy, Student ainsi qu'une variété de mélanges.

- Les distributions appartenant au domaine d'attraction de la distribution Gumbel

$$F_X \in DAM(H_\xi), \quad \xi = 0,$$

correspondent aux distributions *medium-tailed* (à queue moyenne) et comprend, par exemple, les distributions normale, exponentielle, gamma et lognormale.

- Les distributions appartenant au domaine d'attraction de la distribution Weibull

$$F_X \in DAM(H_\xi), \quad \xi < 0,$$

correspondent aux distributions *light-tailed* (à queue mince) et contient, notamment, les distributions uniforme et bêta.

En actuariat, pour les données d'assurance, on utilise principalement les distributions *heavy-* et *medium-tailed*.

Exemple – Loi exponentielle avec $\beta = 1$.

On fixe

$$\begin{aligned}a_n &= 1 \\b_n &= \ln n\end{aligned}$$

On a

$$\begin{aligned}\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) &= \Pr(M_n \leq xa_n + b_n) \\&= F_X(xa_n + b_n)^n \\&= (1 - \exp(-(x + \ln n)))^n \\&= \left(1 - \frac{e^{-x}}{n}\right)^n\end{aligned}$$

On déduit que

$$\lim \Pr \left(\frac{M_n - b_n}{a_n} \leq x \right) = \exp(-\exp(-x))$$

qui correspond, comme prévu par le théorème à la fonction de répartition de la loi Gumbel. \square

Exemple – Loi Pareto avec $\alpha, \lambda > 0$.

On fixe

$$\begin{aligned}a_n &= \frac{\lambda n^{\frac{1}{\alpha}}}{\alpha} \\b_n &= \lambda n^{\frac{1}{\alpha}} - \lambda\end{aligned}$$

On a

$$\begin{aligned}\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) &= \Pr(M_n \leq xa_n + b_n) \\&= F_X(xa_n + b_n)^n \\&= \left(1 - \left(\frac{1}{1 + \frac{xa_n - b_n}{\lambda}}\right)^\alpha\right)^n \\&= \left(1 - \left(\frac{1}{1 + \frac{x\lambda n^{\frac{1}{\alpha}} + \lambda n^{\frac{1}{\alpha}} - \lambda}{\alpha}}\right)^\alpha\right)^n \\&= \left(1 - \frac{1}{n} \left(\frac{1}{1 + \frac{x}{\alpha}}\right)^\alpha\right)^n\end{aligned}$$

On déduit que

$$\lim \Pr \left(\frac{M_n - b_n}{a_n} \leq x \right) = \exp \left(- \left(1 + \frac{x}{\alpha} \right)^{-\alpha} \right)$$

qui correspond, comme prévu par le théorème à la fonction de répartition de la loi Fréchet avec $\mu = -\alpha$, $\sigma = \alpha$ et $\xi = \frac{1}{\alpha}$ \square

22 Distribution de l'excédent

Soit la v.a. X avec fonction de répartition F_X et la v.a. W_u représentant l'excédant d'un sinistre par rapport à un seuil u

$$W_u = (X - u \mid X > u).$$

On sait que la fonction de répartition W_u est

$$\begin{aligned} F_{W_u}(x) &= \Pr(W_u \leq x) \\ &= \Pr(X - u \leq x \mid X > u) \\ &= \frac{F_X(x + u) - F_X(u)}{1 - F_X(u)}, \quad x > 0. \end{aligned}$$

Par conséquent, on a, pour $x > u$,

$$F_X(x) = \{1 - F_X(u)\} F_{W_u}(x - u) + F_X(u).$$

23 Distribution Pareto généralisée

La distribution Pareto généralisée résulte de la reparamétrisation de la distribution Pareto et d'une extension des valeurs possibles de son paramètre de forme.

La fonction de répartition de la loi Pareto avec les paramètres α et λ est

$$\begin{aligned} F_X(x) &= 1 - \left(\frac{\lambda}{\lambda + x} \right)^\alpha \\ &= 1 - \left(\frac{1}{1 + \frac{x}{\lambda}} \right)^\alpha. \end{aligned}$$

En fixant

$$\xi = \frac{1}{\alpha}$$

et

$$\begin{aligned}\lambda &= \sigma\alpha \\ &= \frac{\sigma}{\xi},\end{aligned}$$

on obtient une forme reparamétrisée de la loi Pareto dont la fonction de répartition est notée par

$$GP_{\xi,\sigma}(x) = 1 - \left(\frac{1}{1 + \frac{\xi x}{\sigma}} \right)^{\frac{1}{\xi}},$$

où ξ est un paramètre de forme et σ est un paramètre d'échelle.

En fait, cette distribution correspond à la distribution Pareto généralisée au sens entendu par les références sur les distributions extrêmes. Cette classe peut

être élargie en considérant les cas où $\xi = 0$ et $\xi < 0$

$$GP_{\xi, \sigma}(x) = \begin{cases} 1 - \left(\frac{1}{1 + \frac{\xi x}{\sigma}} \right)^{\frac{1}{\xi}}, & \xi < 0, 0 < x < -\frac{\sigma}{\xi} \quad (\text{bêta}) \\ 1 - \exp\left(-\frac{x}{\sigma}\right), & \xi = 0, x > 0 \quad (\text{exponentielle}) \\ 1 - \left(\frac{1}{1 + \frac{\xi x}{\sigma}} \right)^{\frac{1}{\xi}}, & \xi > 0, x > 0 \quad (\text{Pareto reparamétrisée}) \end{cases}.$$

Quand $\xi = 0$, la distribution correspond à la loi exponentielle.

Quand $\xi < 0$, la distribution correspond à une distribution de type bêta (voir Reiss et Thomas (1997) pour les détails) et quand $\xi > 0$, la distribution correspond à la version de la distribution Pareto à deux paramètres comme on vient de le montrer ci-dessus.

On examine une propriété intéressante de la loi Pareto généralisée.

Supposons que la distribution de la v.a. X est Pareto généralisée.

La fonction de répartition de la v.a. $W_u = (X - u \mid X > u)$ est

$$\begin{aligned} F_{W_u}(x) &= \Pr(X - u \mid X > u) \\ &= \frac{F_X(x + u) - F_X(u)}{1 - F_X(u)} \\ &= \frac{GP_{\xi, \sigma}(x + u) - GP_{\xi, \sigma}(u)}{1 - GP_{\xi, \sigma}(u)} \end{aligned}$$

qui devient

$$\begin{aligned}
 F_{W_u}(x) &= \frac{\left(\frac{1}{1+\frac{\xi u}{\sigma}}\right)^{\frac{1}{\xi}} - \left(\frac{1}{1+\frac{\xi(x+u)}{\sigma}}\right)^{\frac{1}{\xi}}}{\left(\frac{1}{1+\frac{\xi u}{\sigma}}\right)^{\frac{1}{\xi}}} \\
 &= 1 - \left(\frac{1}{\left(1 + \frac{\xi u}{\sigma} + \frac{\xi x}{\sigma}\right) \times \left(\frac{1}{1+\frac{\xi u}{\sigma}}\right)} \right)^{\frac{1}{\xi}} \\
 &= 1 - \left(\frac{1}{1 + \xi \frac{x}{\sigma + \xi u}} \right)^{\frac{1}{\xi}}, \quad x > 0.
 \end{aligned}$$

Par conséquent, la v.a. W_u obéit à une loi Pareto généralisée avec paramètres ξ et $\sigma + \xi u$

$$F_{W_u} = GP_{\xi, \sigma + \xi u}.$$

24 Liens entre la distribution Pareto généralisée et la distribution généralisée de la valeur extrême

On considère un seuil élevé u et une v.a. X avec distribution $F_X \in DAM(H_\xi)$. Selon le théorème de Pickand-Balkema-de Haan, il existe une suite de nombre réels $\sigma(u)$ (fonction de u) telle que

$$\lim_{u \rightarrow \infty} \sup_{x \geq 0} |F_{W_u}(x) - G_{\xi, \sigma(u)}(x)| = 0$$

si et seulement si $F_X \in DAM(H_\xi)$. En d'autres termes, H_ξ gouverne le comportement de M_n si et seulement si G_ξ gouverne le comportement de W_u .

D'après ce résultat, pour des seuils u suffisamment élevés, on peut approximer F_{W_u} par $G_{\xi,\sigma}$.

Ce résultat donne une base théorique suffisante suggérant de modéliser les excédents dépassant des seuils u élevés par une distribution Pareto généralisée. Quand l'actuaire croit que F_X est une distribution *heavy-tailed* ($F_X \in DAM(H_\xi)$, $\xi > 0$), alors il est suggéré d'approximer F_{W_u} par $G_{\xi,\sigma}$ avec $\xi > 0$. Il en est de même pour les distributions *medium-* et *light-tailed* avec $\xi = 0$ et $\xi < 0$.

25 Domaine d'attraction de la loi Fréchet

Les lois Pareto, Burr, loggamma, Cauchy, Student, ... sont membres du domaine d'attraction de la loi Fréchet.

On dit qu'elle ont un comportement semblable à la loi Pareto – lois cousines de la loi Pareto

Loi Pareto (paramètres $\alpha > 1$ pour le chapitre et $\lambda > 0$)

– Fonction de répartition :

$$F_X(x) = 1 - \left(\frac{\lambda}{\lambda + x} \right)^\alpha$$

– Quantile :

$$q_X(\kappa) = \frac{\lambda}{(1 - \gamma)^{\frac{1}{\alpha}}} - \lambda$$

Loi Burr (paramètres $\alpha > 1$ pour le chapitre et $\lambda > 0$)

– Fonction de répartition :

$$F_X(x) = 1 - \left(\frac{\lambda}{\lambda + x^\tau} \right)^\alpha$$

– Quantile :

$$q_X(\kappa) = \left(\frac{\lambda}{(1 - \gamma)^{\frac{1}{\alpha}}} - \lambda \right)^{\frac{1}{\tau}}$$

Pour la Pareto et ses cousines, on peut montrer

$$q_X(1 - \kappa) = (\kappa)^{-\gamma} l\left(\frac{1}{\kappa}\right)$$

pour $\kappa \in]0, 1[$. Les formes de $l(x)$ et γ diffèrent selon la loi. Le paramètre γ correspond à l'indice Pareto. Pour la loi Pareto, $\gamma = \frac{1}{\alpha}$ et $l(x) = 1$

Alors, on a

$$\ln q_X(1 - \kappa) = -\gamma \ln \kappa + \ln l\left(\frac{1}{\kappa}\right)$$

On peut montrer que

$$\lim_{\kappa \rightarrow 0} \frac{\ln q_X(1 - \kappa)}{-\ln \kappa} \rightarrow \gamma.$$

Pour la loi *Burr*, $\gamma = \frac{1}{\alpha\lambda}$ et $l(x) = \left(\frac{\lambda}{1+\frac{\lambda}{x^\tau}}\right)^\alpha$.

On peut utiliser l'estimateur de Hill pour estimer l'indice Pareto γ .

26 Estimateur de Hill

L'estimateur permet d'estimer l'indice de Pareto γ .

Il se fonde sur la relation

$$\lim_{\kappa \rightarrow 0} \frac{\ln q_X(1 - \kappa)}{-\ln \kappa} \rightarrow \gamma.$$

On suppose un échantillon

$$\{X_{1,n}, \dots, X_{n,n}\}$$

L'estimateur de Hill est construit à partir des données transformées par la fonction \log .

Estimateur de Hill :

$$H_{k,n} = \frac{1}{k} \sum_{j=1}^k \ln X_{n-j+1,n} - \ln X_{n-k,n}$$

Interprétation : $H_{k,n}$ = estimateur de la fonction d'excès moyen construit à partir des données transformées par la fonction log.

Exemples :

– $k = 1$:

$$H_{1,n} = \ln X_{n,n} - \ln X_{n-1,n} = \ln \left(\frac{X_{n,n}}{X_{n-1,n}} \right)$$

– $k = 2$:

$$H_{1,n} = \frac{\ln X_{n-1,n} + \ln X_{n,n}}{2} - \ln X_{n-2,n}$$

– $k = 10$:

$$H_{10,n} = \frac{\ln X_{n-9,n} + \dots + \ln X_{n,n}}{2} - \ln X_{n-10,n}$$

27 Méthode POT pour sinistres élevés

27.1 Introduction

Les sinistres en assurance dommage peuvent prendre des valeurs très élevées.

La santé financière d'une société d'assurance est influencée par le nombre et les valeurs des sinistres élevés. Une bonne appréciation du risque d'assurance dépend grandement de la justesse du modèle utilisé pour les sinistres élevés.

L'assureur doit choisir avec soin le modèle approprié pour la tarification de la prime de ce contrat. Comme le choix du modèle est crucial, on traite séparément la modélisation (et l'estimation) de la distribution pour les sinistres avec montants élevés.

On considère un contexte où une compagnie d'assurance veut tarifier un contrat avec protection en excédent de sinistres.

On suppose que les limites par sinistre offertes dans les contrats de la compagnie sont supérieures à un *seuil* (*threshold*) u d'une valeur élevée.

Comme on le précise plus loin, le choix du seuil u est assez arbitraire et découle de l'expérience de l'analyste.

La compagnie dispose d'un échantillon avec un nombre (assez ou très) important de données.

On suppose que les données sont complètes.

En général, on compte plusieurs observations de valeurs moyennes ou basses mais très peu d'observations de valeurs élevées, c'est à dire supérieures à u .

Habituellement, la compagnie d'assurance peut avoir recours à deux stratégies. Selon la première stratégie, elle cherche comme il est expliqué précédemment

à trouver le modèle le plus approprié à partir de l'ensemble des données en utilisant, par exemple, la méthode du maximum de vraisemblance. Ce modèle « approprié » offrira une bonne adéquation sur l'ensemble des données à l'exception généralement des données élevées (ou extrêmes) supérieures à u . La tarification du contrat avec protection en excédent de sinistres (*excess-of-loss*) requiert un modèle adéquat pour les données très élevées, c.-à-d. supérieures à u .

Selon la deuxième stratégie, pour choisir et estimer les paramètres du modèle (méthode du maximum de vraisemblance) on utilise les données supérieures à u en sachant qu'elles sont supérieures à u . La fonction de vraisemblance $L(\underline{\theta})$

est donnée par

$$\begin{aligned} L(\underline{\theta}) &= \prod_{i=1}^{m(u)} f_B(x_i \mid B > u; \underline{\theta}) \\ &= \prod_{i=1}^{m(u)} \frac{f_B(x_i; \underline{\theta})}{1 - F_B(u)}, \end{aligned}$$

où $m(u)$ est le nombre de sinistres dont le montant excède le seuil u .

Cette seconde stratégie a pour résultat de produire des estimés qui sont peu fiables.

En se basant sur la théorie des valeurs extrêmes et en utilisant la distribution Pareto généralisée, Davison et Smith (1990) ont proposé d'utiliser une troisième stratégie basée sur la méthode *Peak-over-Threshold* (*POT*).

Cette méthode a l'avantage de fournir une très bonne adéquation aux données très élevées, supérieures au seuil u .

L'objectif de la présente section est de présenter cette méthode.

La méthode POT a été proposée par Davison et Smith (1990).

Une application de cette méthode avec des données provenant de l'assurance maladie collective est réalisée dans Cedrian et collab. (2001).

La définition utilisée dans cette section pour la Pareto généralisée correspond à celle donnée dans les références traitant le sujet (mentionnées précédemment) et non à celle qu'on retrouve dans Klugman et collab. (1998) ou dans Hogg et Klugman (1984).

27.2 Idée de la méthode Peak-over-Threshold (POT)

L'idée de la méthode *Peak-over-threshold* (POT) est de « fitter » la queue de la distribution de X par une loi Pareto généralisée.

Soit la v.a X avec fonction de répartition F_X et la v.a. W_u définie par

$$W_u = (X - u \mid X > u).$$

On sait que la fonction de répartition F_{W_u} est

$$F_{W_u}(x) = \frac{F_X(x + u) - F_X(u)}{1 - F_X(u)}, \quad x > 0.$$

Ainsi, on a

$$F_X(x) = \{1 - F_X(u)\} F_{W_u}(x - u) + F_X(u), \quad x > u.$$

Pour simplifier la présentation, on considère seulement le cas qui nous pré-occupe des distributions *heavy-tailed* ($\xi > 0$). L'idée de la méthode POT est d'approximer $F_X(u)$ par $F_n(u)$ et $F_{W_u}(x - u)$ par $G_{\xi, \sigma}(x - u)$ pour $x > u$ où u est le seuil fixé à une valeur élevée

$$\hat{F}_X(x) = \{1 - F_n(u)\} GP_{\xi, \sigma}(x - u) + F_n(u)$$

et F_n est la fonction de répartition empirique

$$F_n(x) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{(-\infty, x]}(x_j).$$

On peut montrer que

$$\hat{F}_X(x) = GP_{\xi, \sigma^*}(x - u^*),$$

où

$$\sigma^* = \sigma (1 - F_n(u))^\xi$$

et

$$u^* = u - \sigma \frac{\left\{ (1 - F_n(u))^{-\xi} - 1 \right\}}{\xi},$$

car, pour $x > u$,

$$\begin{aligned}
 \hat{F}_X(x) &= \{1 - F_n(u)\} GP_{\xi, \sigma}(x - u) + F_n(u) \\
 &= GP_{\xi, \sigma}(x - u) + F_n(u) \{1 - GP_{\xi, \sigma}(x - u)\} \\
 &= 1 - \left(\frac{1}{1 + \xi \frac{x-u}{\sigma}} \right)^{\frac{1}{\xi}} + F_n(u) \left(\frac{1}{1 + \xi \frac{x}{\sigma + \xi u}} \right)^{\frac{1}{\xi}} \\
 &= 1 - \left(\frac{1}{1 + \xi \frac{x-u}{\sigma}} \right)^{\frac{1}{\xi}} (1 - F_n(u))^{\frac{\xi}{\xi}} \\
 &= 1 - \left(\frac{1}{1 + \xi \frac{x-u}{\sigma(1-F_n(u))^\xi}} \right)^{\frac{1}{\xi}} \\
 &= 1 - \left(\frac{1}{1 + \xi \frac{x-u}{\sigma^*}} \right)^{\frac{1}{\xi}}.
 \end{aligned}$$

On estime les paramètres ξ et σ à partir de la méthode du maximum de vraisemblance. Quand $\xi > -0.5$ (on s'intéresse au cas $\xi > 0$), on peut montrer, pour un seuil u fixé, que

$$n(u)^{\frac{1}{2}} \begin{pmatrix} \hat{\xi}_{n(u)} \\ \hat{\sigma}_{n(u)} \end{pmatrix} \xrightarrow{u} norm \left(\begin{pmatrix} \xi \\ \sigma \end{pmatrix}, \begin{pmatrix} (1+\xi)^2 & \sigma(1+\xi) \\ \sigma(1+\xi) & 2\sigma(1+\xi) \end{pmatrix} \right),$$

pour $n(u) \rightarrow \infty$.

27.3 Choix du seuil u

Il n'existe pas de méthode pour établir le seuil u optimal. Le choix du seuil u pose un problème : en prenant un seuil trop élevé, on dispose alors de peu de données pour estimer les paramètres et on risque d'avoir des résultats imprécis et biaisés. En fixant un seuil u trop bas, l'estimation se base sur des données où il n'est pas approprié d'appliquer la distribution Pareto généralisée.

Une façon adéquate de remédier à ce problème est de tracer la fonction empirique $\hat{e}_n(x)$ de l'excès moyen et de choisir le seuil u de telle sorte que $\hat{e}_n(x)$ est approximativement linéaire pour les valeurs de $x \geq u$.

Travaux de recherche en cours sur ce sujet..

28 Distributions composites

Distributions composites = "splicing"

La méthode est semblable à la méthode du POT : on remplace la portion "non-paramétrique" par une portion paramétrique : lognormale, mélange d'Erlang.

Pour un exposé sur les distributions composites, voir, e.g., Albrecher et al. [1].

29 Distributions composites

Distributions composites = "splicing"

La méthode est semblable à la méthode du POT : on remplace la portion "non-paramétrique" par une portion paramétrique : lognormale, mélange d'Erlang.

Pour un exposé sur les distributions composites, voir, e.g., Albrecher et al. [1].

30 Sources de données et packages R utiles

Le site et les packages R ci-dessous sont recommandés :

1. Site (plusieurs sources de données) : <https://lstat.kuleuven.be/Wiley/>
2. Packages R sur Cran :
 - <https://cran.r-project.org/web/packages/ReIns/ReIns.pdf>
 - <https://cran.r-project.org/web/packages/insuranceData/insuranceData.pdf>
 - <https://cran.r-project.org/web/packages/evir/evir.pdf>
3. Autres packages R :
 - <http://cas.uqam.ca/pub/R/web/CASdatasets-manual.pdf>

Références

- [1] Albrecher, H., Beirlant, J. & Teugels, J. (2017). Reinsurance : Actuarial and Statistical Aspect. John Wiley & Sons, Ltd, Chichester, UK.
- [2] Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. L. (2006). Statistics of extremes : theory and applications. John Wiley & Sons.
- [3] Cizek, P., Härdle, W. K., & Weron, R. (Eds.). (2005). Statistical tools for finance and insurance. Springer Science & Business Media.
- [4] Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. Journal of the Royal Statistical Society. Series B (Methodological), 393-442.
- [5] Embrechts, P., Klüppelberg, C., & Mikosch, T. (2013). Modelling extremal events : for insurance and finance (Vol. 33). Springer Science & Business Media.

- [6] Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2012). Loss models : from data to decisions (Vol. 715). John Wiley & Sons.
- [7] McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin : The Journal of the IAA*, 27(1), 117-137.
- [8] McNeil, A. J., Frey, R., & Embrechts, P. (2015). Quantitative risk management : Concepts, techniques and tools. Princeton university press.
- [9] Prettenthaler, F., Albrecher, H., Köberl, J., & Kortschak, D. (2012). Risk and insurability of storm damages to residential buildings in Austria. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 37(2), 340-364.