Plan

1. La loi de Pareto généralisée.

2. Méthode du maximum de vraisemblance.

3. Test lisse de Neyman appliqué à la loi de Pareto généralisée.

4. Résultats de simulations.

5. Conclusion.

La loi de Pareto généralisée

La loi de Pareto généralisée (GPD) est une vaste famille, englobant entre autre la famille exponentielle, la loi uniforme et la loi de Pareto. Sa fonction de répartion est :

$$F(x) = \begin{cases} 1 - (1 - kx/\sigma)^{1/k} & \text{sur} \quad [0, \sigma/k] \quad \text{si} \quad k > 0 \\ 1 - \exp(-x/\sigma) & \text{sur} \quad \mathbb{R}^+ \quad \text{si} \quad k = 0 \\ 1 - (1 - kx/\sigma)^{1/k} & \text{sur} \quad \mathbb{R}^+ \quad \text{si} \quad k < 0 \end{cases}$$

$k = 0$ loi exponentielle.
$k = 1$ loi uniforme.
$k < 0$ loi de Pareto.

Estimateurs MV

Deux méthodes de calcul des EMV de la GPD: Grimshaw(1993), Davison et Smith(1984).
On a utilisé la deuxième qui consiste à faire le changement suivant : $(\sigma, k) \longrightarrow (\tau = \frac{k}{\sigma}, k)$
$\widehat{\tau}$ calculé itérativement par la méthode de Newton-Raphson, $\widehat{k}$ puis $\widehat{\sigma}$ s'en déduit aisément.

pour $k < \frac{1}{2}$:

$$\sqrt{n}(\widehat{k} - k, \widehat{\sigma} - \sigma) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

$$\Sigma = \begin{pmatrix} 2(1-k)\sigma^2 & (1-k)\sigma \\ (1-k)\sigma & (1-k)^2 \end{pmatrix}$$

Tests lisses de Neyman

Nous voulons tester l'hypothèse nulle que les variables $X_1, ..., X_n$ sont issues de la GPD de densité $f(x; \beta)$, où $\beta = (\sigma, k)$.

On plonge cette densité dans une famille exponentielle de densités :

$$g_r(x; \theta, \beta) = f(x; \beta) \exp\{\sum_{i=1}^{r} \theta_i \overline{F}^i(x; \beta) - K(\theta)\}$$

$$\overline{F}(x; \beta) = 1 - F(x; \beta)$$

Tester l'adéquation de $f(x; \beta)$ revient à effectuer le test d'hypothèse :

$$H_0 : \theta = 0 \text{ contre } H_1 : \theta \neq 0.$$

β paramètre de nuisance.

Tests lisses pour la GPD

La statistique de test est :

$$\widehat{\Psi}_r^2 = \frac{1}{n}\,\widehat{U}^t \widehat{A}^{-1} \widehat{U}$$

$$\widehat{A} = I_{\theta\theta} - I_{\theta\widehat{\beta}} I_{\widehat{\beta}\widehat{\beta}}^{-1} I_{\widehat{\beta}\theta}$$

$\widehat{U}$ est le score par rapport à θ sous H_0 en remplaçant β par $\widehat{\beta}$.

Sous l'hypothèse nulle $\widehat{\Psi}_r^2$ converge en distribution vers une loi du chi-deux à r degrés de liberté.

$$
\begin{aligned}
I_{\theta\theta}(i,j) &= -\mathbb{E}\left[\frac{\partial^2 L(\theta,\beta)}{\partial\theta_i\partial\theta_j}\right]|_{\theta=0} \\
I_{\theta\beta}(i,j) &= -\mathbb{E}\left[\frac{\partial^2 L(\theta,\beta)}{\partial\theta_i\partial\beta_j}\right]|_{\theta=0}
\end{aligned}
$$

$$
I_{\theta\theta}(i,j) = \frac{ij}{(i+j+1)(i+1)(j+1)}
$$

$$
I_{\theta\beta}(i,1) = \frac{-i}{(i+1-k)(i+1)\sigma}
$$

$$
I_{\theta\beta}(i,2) = \frac{i}{(i+1-k)(i+1)^2}
$$

et

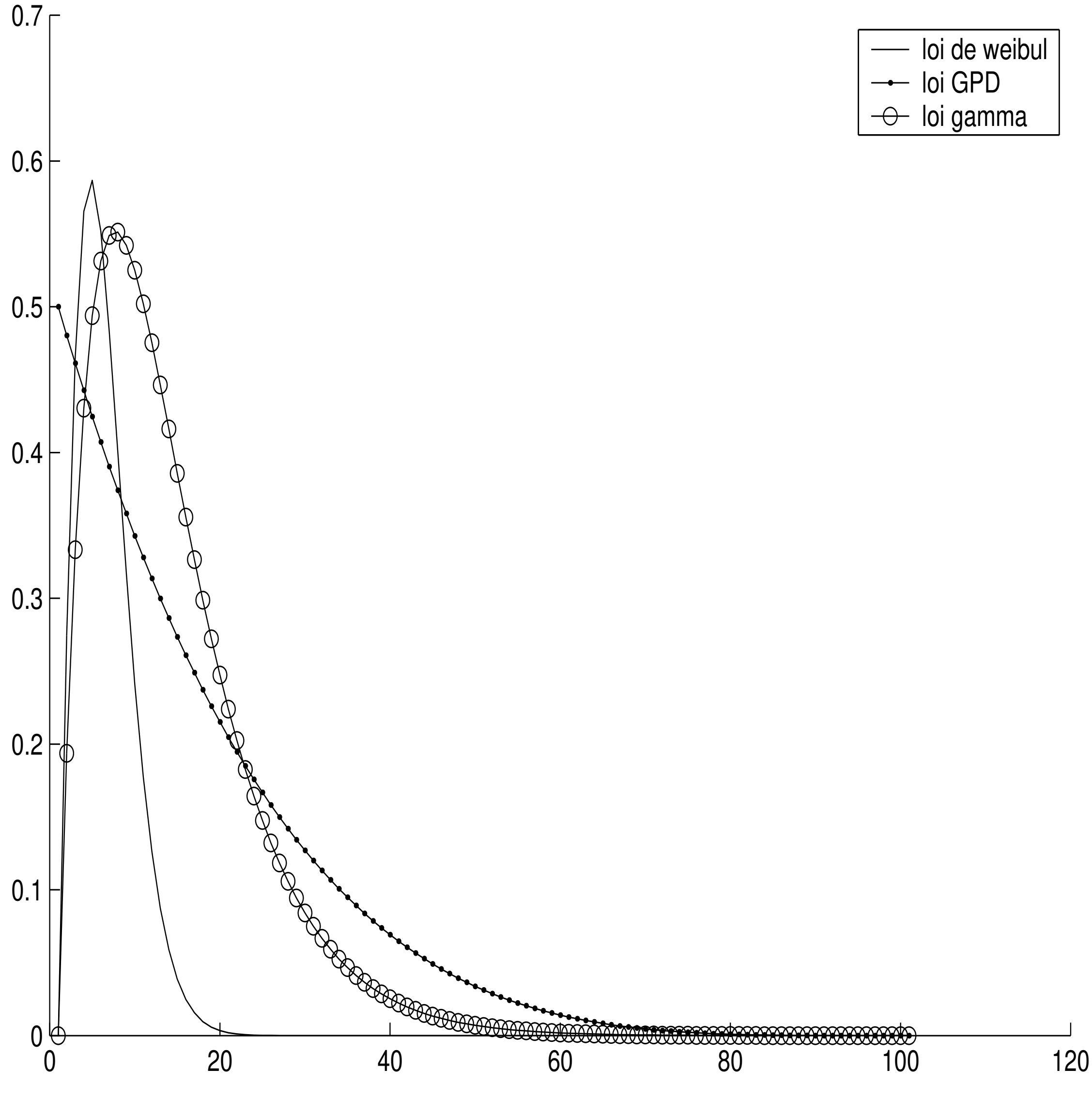
$$
I_{\beta\beta} = \begin{pmatrix} \frac{1}{(1-2k)\sigma^2} & \frac{-1}{\sigma(1-k)(1-2k)} \\ \frac{-1}{\sigma(1-k)(1-2k)} & \frac{2}{(1-k)(1-2k)} \end{pmatrix}
$$

Simulation GPD(2;0.4) et test à la GPD.

n	20	40	60	80	100
$r = 1$	0.0135	0.0332	0.0301	0.0400	0.0440
$r = 2$	0.0310	0.0504	0.0511	0.0380	0.0660
$r = 3$	0.0512	0.0393	0.0511	0.0631	0.0550
$r = 4$	0.0561	0.0634	0.0611	0.0550	0.0620
$r = 5$	0.0458	0.0547	0.0612	0.0600	0.0440
$r = 6$	0.0554	0.0727	0.0561	0.0490	0.0670
$r = 7$	0.0578	0.0535	0.0671	0.0470	0.0700
$r = 8$	0.0466	0.0594	0.0680	0.0540	0.0540
$r = 9$	0.0608	0.0658	0.0650	0.0640	0.0630
$r = 10$	0.0636	0.0584	0.0610	0.0550	0.0510

Simulation GPD(1;-0.25) et test à la GPD.

n	20	40	60	80	100
$r = 1$	0.0350	0.0420	0.0320	0.0440	0.0460
$r = 2$	0.0320	0.0510	0.0520	0.0480	0.0560
$r = 3$	0.0520	0.0490	0.0620	0.0520	0.0550
$r = 4$	0.0554	0.0610	0.0600	0.0610	0.0460
$r = 5$	0.0450	0.0724	0.0561	0.0480	0.0460
$r = 6$	0.0560	0.0520	0.0450	0.0620	0.0670
$r = 7$	0.0500	0.0540	0.0655	0.0460	0.0720
$r = 8$	0.0455	0.0524	0.0620	0.0520	0.0600
$r = 9$	0.0602	0.0658	0.0610	0.0610	0.0620
$r = 10$	0.0620	0.0580	0.0600	0.0540	0.0610



0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
20
40
60
80
100
120
loi de weibul
loi GPD
loi gamma

Simulation de la loi gama(2;1.5) vs GPD

n	20	40	60	80	100
$r = 1$	0.9540	0.9760	0.9820	0.9860	0.9970
$r = 2$	0.9790	0.9870	0.9820	0.9930	0.9980
$r = 3$	0.8550	0.8920	0.8980	0.9000	0.9970
$r = 4$	0.8480	0.8560	0.8580	0.8600	0.9960
$r = 5$	0.8520	0.8540	0.8560	0.8580	0.9950
$r = 6$	0.8510	0.8540	0.8550	0.8560	0.9610
$r = 7$	0.8500	0.8510	0.8530	0.8550	0.9550
$r = 8$	0.8500	0.8500	0.8520	0.8550	0.9450
$r = 9$	0.8490	0.8510	0.8530	0.8540	0.9450
$r = 10$	0.8450	0.8500	0.8520	0.8540	0.9420

Simulation loi Weibull(1.5;2) vs GPD

n	20	40	60	80	100
$r = 1$	0.9645	1.0000	1.0000	1.0000	1.0000
$r = 2$	0.8555	0.9975	1.0000	1.0000	1.0000
$r = 3$	0.7155	0.9965	1.0000	1.0000	1.0000
$r = 4$	0.6380	0.9875	1.0000	1.0000	1.0000
$r = 5$	0.5495	0.9775	1.0000	1.0000	1.0000
$r = 6$	0.5535	0.9595	1.0000	1.0000	1.0000
$r = 7$	0.6965	0.9330	0.9995	1.0000	1.0000
$r = 8$	0.5435	0.9155	0.9985	1.0000	1.0000
$r = 9$	0.5990	0.8860	0.9970	1.0000	1.0000
$r = 10$	0.6630	0.8535	0.9955	0.9995	1.0000

Simulation de la loi exp(3) et test à la GPD

n	20	40	60	80	100
$r = 1$	0.2940	0.3760	0.4680	0.5420	0.5850
$r = 2$	0.3960	0.4730	0.5600	0.6140	0.6550
$r = 3$	0.4330	0.5200	0.6120	0.6650	0.6960
$r = 4$	0.4510	0.5530	0.6510	0.7010	0.7220
$r = 5$	0.4550	0.5620	0.6750	0.7120	0.7390
$r = 6$	0.4530	0.5770	0.6860	0.7290	0.7560
$r = 7$	0.4570	0.5910	0.6920	0.7370	0.7690
$r = 8$	0.4700	0.6000	0.7080	0.7390	0.7800
$r = 9$	0.4850	0.6050	0.7190	0.7490	0.7810
$r = 10$	0.4920	0.6110	0.7190	0.7500	0.7910

Conclusion

Le test lisse de Neyman pour la GPD peut être relativement aisément mis en pratique si k est nettement $\neq 0$. Pour $k \approx 0$ on peut tester le modèle exponentiel.

Donc il faudra envisager d'autres estimateurs : moments, moments pondérés de Hosking et Wallis (1987), ce qui entrainera une modification de la statistique du test.