Loi des valeurs extrêmes

October 10, 2017

Contents

1 Simulations **1**
1.1 Domaine d'attraction 1
1.2 Estimation dans le modèle GEV 2

2 Données Réelles **2**
2.1 Estimation statistique de la crue centennale à Bar sur Seine (modèle POT) . 2
2.2 Estimation statistique sur un modèle de maximum par paquets 3

1 Simulations

1.1 Domaine d'attraction

On considère les lois suivantes:

- Loi uniforme sur $[0,1]$,
- Loi exponentielle d'espérance 1.5 (de fonction de répartition $F(z) = 1 - \exp(-\frac{z}{1.5})$, pour $z > 0$),
- Loi de Pareto (de fonction de répartition $F(z) = 1 - z^{-0.5}$ pour $z > 1$).

Question 1 *Pour chacune des lois ci-dessus: Trouver par le calcul les suites renormalisantes* μ_n *et* σ_n *qui assure la convergence du maximum. En déduire l'indice de forme des extrêmes* ξ*. Donner la loi généralisée des extrêmes qui approche la loi du maximum renormalisé.*

Question 2 *Simuler* $N = 100$ *échantillons du maximum de* $n = 12, 365, 1\ 000$ *et* $10\ 000$ *valeurs pour les trois lois ci-dessus, puis tracer les histogrammes des maximums correspondants et la densité de la loi des extrêmes théoriques. On utilisera les fonctions unif.sce, expo.sce, pareto.sce. Discuter de la qualité d'ajustement. (Pour créer et afficher les graphiques dans la fenêtre graphique* `i`*, utiliser la commande* `xset(``window'',i)`*.)*

Le vecteur `unif(:,2)` contient l'échantillon de N maxima de $n = 365$ simulations de la loi uniforme.

1.2 Estimation dans le modèle GEV

On reprend les simulations du paragraphe précédent.

Question 3 *Écrire la vraisemblance et réaliser l'inférence par maximum de vraisemblance. Donner un intervalle de confiance pour μ, σ et ξ. On utilisera la fonction* `gev_fit` *définie dans le fichier gev.sce.*

La fonction `gev_fit (y, init)` calcule les estimateurs du maximum de vraisemblance de μ, σ et ξ de l'échantillon y ainsi que leur intervalle de confiance. Le paramètre optionnel `init` permet de spécifier les valeurs initiales de μ, σ et ξ utilisées pour l'optimisation de la log-vraisemblance.

Question 4 *Relancer* 10 *fois les simulations, et comparer les estimations avec les valeurs théoriques.*

2 Données Réelles

2.1 Estimation statistique de la crue centennale à Bar sur Seine (modèle POT)

A Bar sur Seine, sur 26 ans, 74 valeurs de débits de la Seine ont dépassé 250 (dixièmes de mm équivalent lame d'eau sur le bassin versant). Voici les valeurs de dépassements et le nombre de dépassement sur chaque année :

```
// valeurs de depassement
x=[110, 21, 8, 34, 134, 156, 252, 126, 141, 259, 1, 913, 333, 32, 19, 300, 97, 259, 329,
   23, 100, 178, 392, 119, 34, 4, 137, 248, 41, 303, 1, 156, 12, 170, 163, 263, 230, ...
   122, 311, 80, 69, 447, 32, 69, 55, 126, 518, 49, 75, 6, 196, 6, 58, 28, 28, 97, 113,
   21, 182, 163, 47, 385, 425, 178, 374, 303, 156, 63, 104, 156, 52, 337, 25, 41];

// nbre de depassements par an
f=[2,3,5,1,3,1,1,6,1,3,3,4,3,3,2,3,1,4,2,5,3,1,2,3,5,4];

// 15 premieres annees
x_15=x(1:sum(f(1:15)));
```

On ne considère dans un premier temps que les 15 premières années de données.

Question 5 *Réaliser l'inférence du modèle POT par maximum de vraisemblance et donner un intervalle de confiance pour λ, σ et ξ. Pour l'estimation et l'intervalle de confiance de σ, ξ, utiliser fonction* `pot_fit` *définie dans le fichier pot.sce.*

```
[sigma, xi, cov] = pot_fit(x_15);
```

Question 6 *Peut on accepter l'hypothèse d'un dépassement de forme exponentielle $\xi = 0$?*

Question 7 *Donner un intervalle de confiance pour le débit centennal à Bar sur Seine. On pourra utiliser la fonction pot_quantile(y, f, p).*

```
pot_quantile(x_15, f(1:15), 0.01)
```

Question 8 *Même travail avec toutes les données disponibles. Que constatez vous?*

2.2 Estimation statistique sur un modèle de maximum par paquets

Les maxima annuels sur les 31 années de 1950 à 1980 de la Seine à Bar sur Seine sont

```
maxima=[147, 159, 162, 184, 240, 251, 269, 305, 330, 347, 360, 369, 384, 413, 432, ...
        446, 498, 509, 513, 550, 553, 553, 561, 579, 587, 635, 642, 675, 697, 768, 1163
```

Question 9 *Écrire la vraisemblance et réaliser l'inférence par maximum de vraisemblance. Donner un intervalle de confiance pour pour μ, σ et ξ. De nouveau, on utilisera la fonction* `gev_fit`.

Question 10 *Donner un intervalle de confiance pour le débit maximal annuel centennal à Bar sur Seine.*

On utilisera la fonction gev_quantile qui prend en argument `y, p, mu, sigma ,xi` où `y` est l'échantillon des maxima annuels, `p=0.01` dans notre cas et `mu, sigma ,xi` sont les valeurs des estimateurs trouvés à la question précédente.

Question 11 *On suppose μ et σ sont connus, donnés par l'évaluation à la question 9.*

Le coût unitaire de construction (surélévation) est de 100. On suppose que le coût des dommages est égal à la puissance 1.2 du débit non contenu par la digue et que la durée du projet est de 50 ans. En utilisant un calcul par simulation de Monte-Carlo, quelle hauteur recommandez vous si $\xi = 0$, si $\xi = 0.1$, si $\xi = 0.3$. Est-il plus pertinent de répondre à cette question avec le modèle GEV ou POT?

Pour cette question, on utilisera la fonction simul_gev(mu, sigma, xi, z) qui calcule par une méthode de Monte Carlo (avec 10000 simulations) l'espérance de ${(D - z)_+}^{1.2}$ lorsque la variable aléatoire D (représentant le débit) suit respectivement un modèle GEV de paramètres `mu, sigma, xi` La fonction fournit également la précision de la méthode de Monte Carlo à 95%.

On pourra aussi utiliser la boucle:

```
xbasc();
z=[100:10:600];

for i=1:length(z)
y(i)=50*simul_gev(mu, sigma, xi, z(i))+ 100*z(i);
end
plot(z,y);
```