



Université Gaston Berger de Saint-Louis  
U.F.R des Sciences Appliquées et de Technologie  
Section Mathématiques Appliquées.



UNIVERSITE  
GASTON BERGER

*L'excellence au service du développement*

---

Sujet 11 : Comparaison des estimateurs Hill et Hill adapté  
sur la distribution de Fréchet et sur la distribution de Paréto

---

Présenté par :  
Safiétou NDIAYE  
Sokhar SAMB

Professeur  
Aliou DIOP

Année Scolaire : 2015-2016

# Table des matières

Introduction . . . . .	2
0.1 Simulations . . . . .	2
0.1.1 Distribution de Fréchet . . . . .	2
0.1.2 Distribution de Paréto . . . . .	3
0.2 Évaluation des Estimateurs . . . . .	3
0.2.1 Estimateur de Hill . . . . .	3
0.2.2 Estimateur de Hill adapté . . . . .	6
0.3 Erreur quadratique moyenne . . . . .	7
0.3.1 Erreur quadratique moyenne de l' estimateur de Hill . . . . .	7
0.3.2 Graphe de l'erreur quadratique moyenne de l' estimateur de Hill . . . . .	8
0.3.3 Erreur quadratique moyenne de l' estimateur de Hill adapté . . . . .	9
0.3.4 Graphe de l'erreur quadratique moyenne de l' estimateur de Hill adpté . . . . .	9
0.4 Nombre optimal d'extrêmes au point $k_0 = k_{opt}$ . . . . .	10
0.4.1 Nombre optimal d'extrêmes avec l' estimateur de Hill . . . . .	10
0.4.2 Nombre optimal d'extrêmes avec l' estimateur de Hill adapté . . . . .	10
0.5 Valeur de l' estimateur moyen au point $k_0 = k_{opt}$ . . . . .	10
0.5.1 Pour l' estimateur de Hill . . . . .	11
0.5.2 Pour Hill adapté . . . . .	11
0.6 Erreur absolue au point $k_0 = k_{opt}$ . . . . .	11
0.6.1 Erreur absolue au point $k_0$ pour l' estimateur de Hill . . . . .	12
0.6.2 Erreur absolue au point $k_0$ pour l' estimateur de Hill adapté . . . . .	12
0.7 Erreur quadratique moyenne au point $k_0$ . . . . .	12
0.7.1 Erreur quadratique moyenne de Hill avec le nombre optimal d'extrêmes $MSE(k_0)$ . . . . .	12
0.7.2 Erreur quadratique moyenne de Hill adapté avec le nombre optimal d'extrêmes $MSE(k_0)$ . . . . .	13
0.8 Déviation standard . . . . .	13
0.8.1 Déviation standard de l' estimateur de Hill . . . . .	13
0.8.2 Déviation standard de l' estimateur de Hill adapté . . . . .	14
0.9 Étude comparative des différents estimateurs . . . . .	14
0.9.1 Pour la distribution de Fréchet . . . . .	14
0.9.2 Pour la distribution de Paréto . . . . .	15
Conclusion . . . . .	15

# Introduction

Au cours des dernières décennies, dans le monde scientifique, nous avons pu observé un développement de la modélisation et de l'analyse de la statistique des événements rares qualifiés d'événements extrêmes. La théorie des valeurs extrêmes est basée sur l'approximation asymptotique des lois des maxima convenablement normalisés de vecteur aléatoire dont les composantes sont des variables supposées indépendantes et identiquement distribuées. Cette théorie basée sur l'étude des valeurs extrêmes d'un échantillon, est vu comme étant la contrepartie de la statistique classique qui est essentiellement basée sur l'étude de la moyenne d'un échantillon. L'étude de ces valeurs, revient à l'étude des queues de distribution, à l'analyse de la plus grande observation d'un échantillon. Le comportement de la queue de distribution d'une fonction est caractérisé par un unique paramètre noté  $\gamma$  appelé indice des valeurs extrêmes. Ainsi dans ce projet, nous tenterons de faire une étude comparative de deux estimateurs usuels de l'indice des valeurs extrêmes  $\gamma$  : l'estimateur de Hill et l'estimateur de Hill adapté. Nous appliquerons cette comparaison tout d'abord sur la distribution de Fréchet et puis sur la distribution de Paréto généralisée. Les algorithmes utilisés sont explicités en annexe.

## 0.1 Simulations

### 0.1.1 Distribution de Fréchet

En théorie des probabilité et en statistique, la loi de Fréchet est un cas spécial de loi d'extrémum généralisée au même titre que la loi de Gumbel et de la loi de Weibull. Le nom de cette loi est issu de Maurice Fréchet.

Supposons qu'on ait la fonction de répartition  $\mathbb{F}_X^{-1}$  de la variable aléatoire  $X$  et que l'on veuille simuler des observations ayant la même loi que celle de  $X$ . Cela peut facilement se faire à partir d'un générateur  $U$  de la loi  $\mathcal{U}([0, 1])$  et au moyen de la formule suivante :

$$X = \mathbb{F}_X^{-1}$$

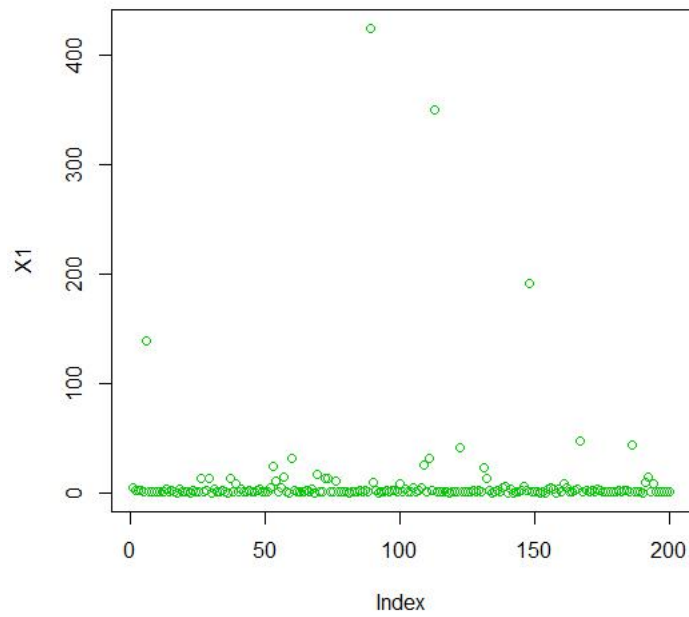
En effet, il se trouve que la variable aléatoire a pour fonction de répartition  $\mathbb{F}_X$ . Cette propriété est connue sous le nom de transformation intégrale de probabilité et a été découvert par R.A.Fischer.

#### **Représentation graphique pour n=200**

```
X1=frechet(200)
```

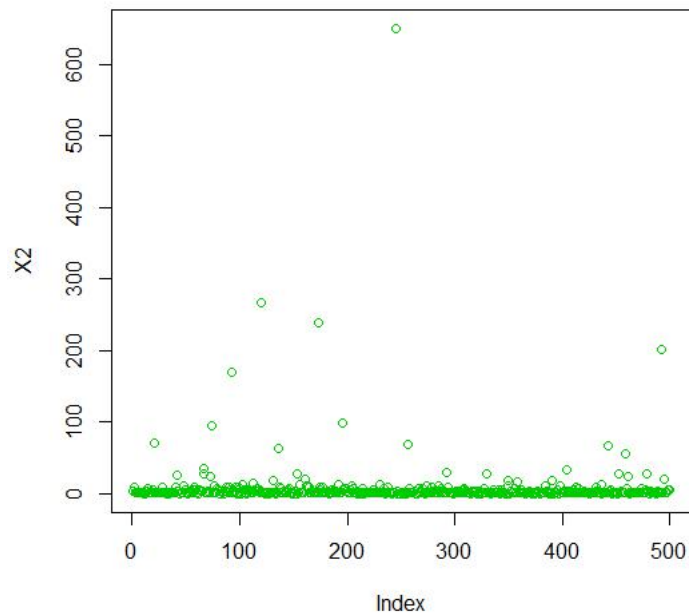
```
plot(X1,col=11)
```

Après exécution de ces deux commandes, nous obtenons le graphe suivant :



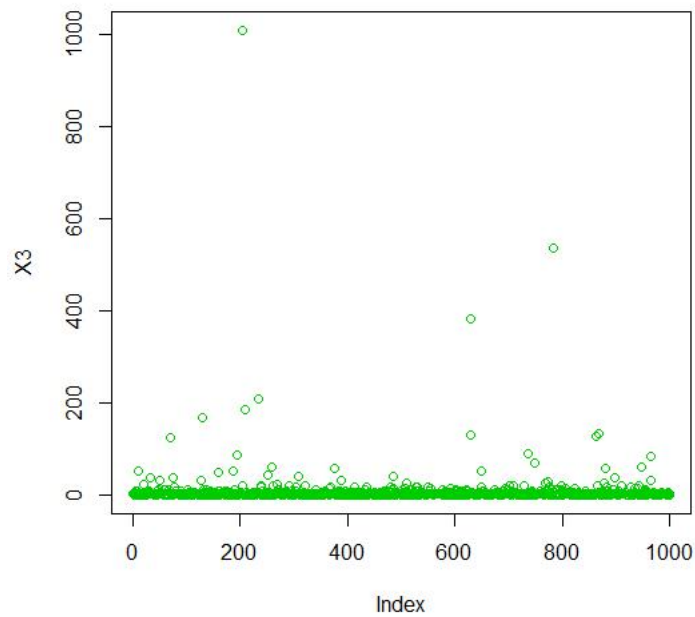
### Représentation graphique pour n=500

```
X2=frechet(500)
plot(X2,col=11)
```



### Représentation graphique pour n=1000

```
X3=frechet(1000)
plot(X3,col=11)
```



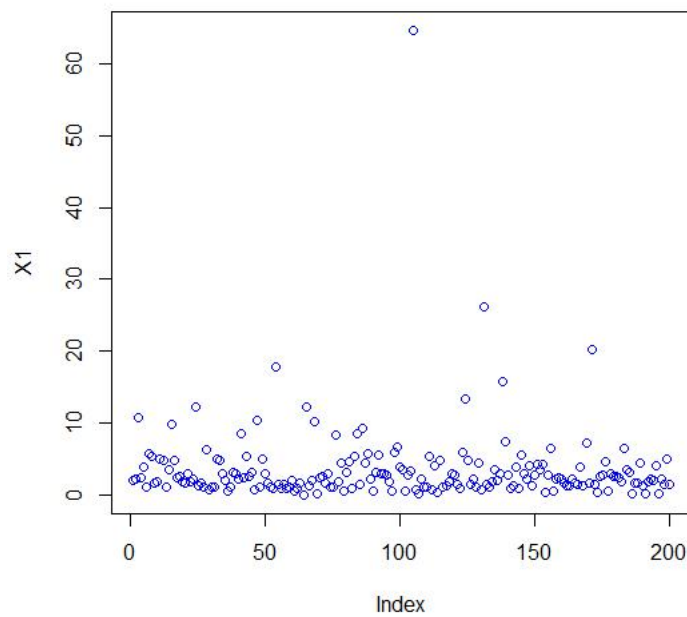
La fonction **frechet(n)** , nous a permis d’observer graphiquement les suites de variables aléatoires générées pour un échantillon de taille n dans les graphes ci dessus. Donc pour simuler m échantillons de taille n de variable aléatoires issues de la loi de Fréchet, nous utilisons la commande suivante :

$$X = m\_frechet(m,n)$$

### 0.1.2 Distribution de Paréto

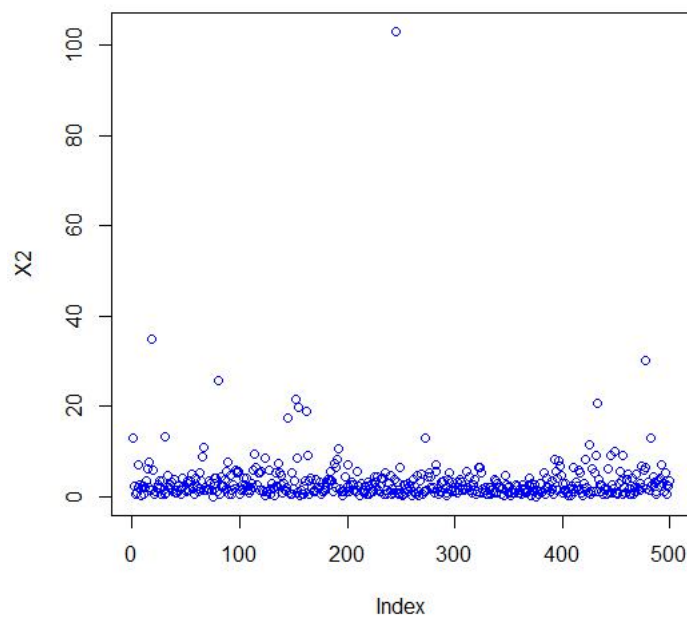
**Représentation graphique pour n=200**

```
X1=pareto(200)
plot(X1,col="blue")
```



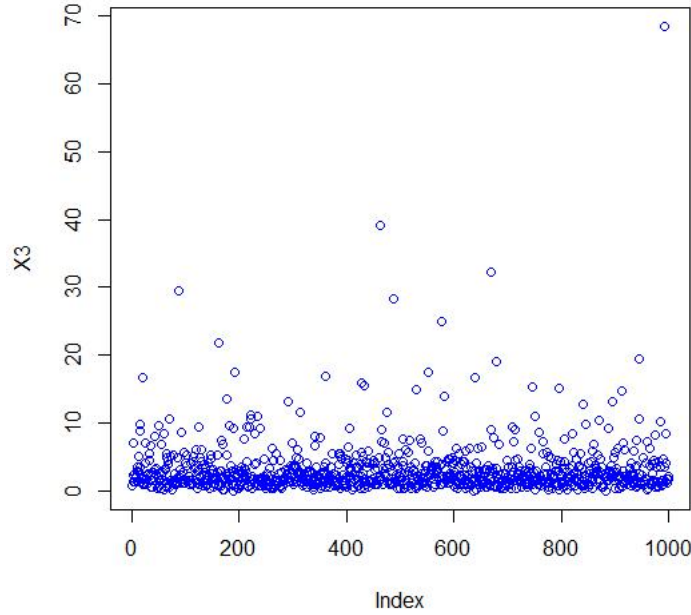
### Représentation graphique pour n=500

```
X2=pareto(500)
plot(X2,col="blue")
```



### Représentation graphique pour n=1000

```
X3=pareto(1000)
plot(X3,col="blue")
```



Par analogie à la fonction **frechet(n)**, la fonction **pareto(n)** , nous a permis d'observer graphiquement les suites de variables aléatoires générées pour un échantillon de taille n dans les graphes ci dessus. Donc pour simuler m échantillons de taille n de variable aléatoires issues de la loi de Paréto généralisé, nous utilisons la commande suivante :

$$X = m\_pareto(m,n)$$

## 0.2 Évaluation des Estimateurs

Soit  $(X_n)$  une suite de copies indépendantes d'une variable aléatoire  $X$  et de fonction de distribution  $\mathbb{F}(x) = \mathbb{P}(X \leq x)$ . On note  $X_{1,n} \leq \dots \leq X_{n,n}$  l'échantillon ordonné pour tout  $i=1,\dots,n$  , la variable aléatoire  $X_{i,n}$  s'appelle la ième statistique d'ordre de l'échantillon. Dans toute la suite  $k$  est une séquence d'entiers représentant le nombre de statistiques d'ordre supérieur( nombres d'extrêmes)utilisé pour l'estimation de  $\gamma$ .

Parmi les estimateurs les plus répandus , nous trouvons l' estimateur de Hil et l' estimateur de Hill adapté.

### 0.2.1 Estimateur de Hill

Cet estimateur est proposé par Hill en 1975. Il est seulement applicable pour les  $\gamma > 0$  et est défini par :

$$\hat{\gamma}_{k,n}^H = \frac{1}{k} \sum_{j=1}^k \log X_{n-j+1,n} - \log X_{n-k,n} = \frac{1}{k} \sum_{j=1}^k j(\log X_{n-j+1,n} - \log X_{n-j,n})$$

La fonction *estimateur\_hill*( $X, n, m, k$ ), avec  $X$  la matrice de  $m$  échantillon de même taille ,  $n$  la taille de l'échantillon, renvoie un vecteur de taille  $m$  dont chaque composante représente la valeur de l' estimateur de Hill pour chaque échantillon.

## Pour la distribution de Fréchet

Pour la loi de Fréchet, les estimateurs de hill sont obtenus par la fonction **estimateur\_hill** et son résultat est gardé dans la variable **estimateur\_frechet** définie par :

$$estimateur\_frechet = estimateur\_hill(X, n, m, k)estimateur\_frechet$$

### Exemple :

Après avoir simuler les m échantillons générés par la loi de Fréchet de taille n=200 et en posant k=30, l'évaluation des estimateurs de Hill est donné par le graphe suivant :



```

> k=30

> estimateur_Frechet = estimateur_hill(X,n,m,k)

> estimateur_Frechet

[1] 1.0125167 1.2375560 0.9161263 0.9265521 0.8771979 1.0040730 1.0928704
[8] 0.8125300 1.1281891 1.0131866 0.8737390 1.0052362 1.3310865 1.4270586
[15] 1.2461758 1.1641311 1.0434810 0.8840382 0.8059782 0.9314410 1.2521594
[22] 0.8384197 0.9291358 0.7925070 1.4353072 0.9532482 0.8858554 0.9513087
[29] 1.1083225 1.0037083 0.9031830 1.3271343 0.9992740 1.0373034 0.9228243
[36] 1.2777292 0.9750006 1.0659633 0.9940728 1.0918712 1.1944926 0.5666221
[43] 1.2924884 0.9083992 1.2587309 1.3572000 1.1851747 1.1501892 1.2318711
[50] 0.9808916 0.6946805 0.9669151 0.9967556 0.8887740 0.9718316 1.1087723
[57] 0.8203926 0.9185903 1.2423345 0.8884698 0.9900045 1.0471811 1.0991093
[64] 1.0032694 0.8494080 1.4310276 1.0956727 0.8289192 1.1196298 0.7040237
[71] 1.3075827 1.4525393 0.8933324 1.2541825 0.9366448 1.0869934 1.0903863
[78] 1.0612904 0.9619760 0.9223130 0.9939100 1.2742596 1.0419298 0.9586180
[85] 1.2324142 1.1872509 1.0409626 1.0752224 0.7566882 1.0106341 0.9498847
[92] 1.0675712 0.9628985 1.1449858 0.9025723 0.8543882 1.0127345 1.3113283
[99] 0.6709934 0.9422104 1.0094573 1.2651114 1.4575252 1.0795991 1.0790108
[106] 0.7805381 1.1402021 1.0261505 0.7962714 1.2513338 0.9627781 0.6890429
[113] 1.2666530 0.7907517 1.1765709 0.8052124 0.8281111 0.7799337 1.0369880
[120] 0.8100729 0.8653585 1.3257803 0.9657748 1.0728599 0.7545780 0.6038018
[127] 0.9730351 0.9459041 0.8421318 1.3700486 1.0066474 0.9580834 0.9883812
[134] 1.3137777 1.4001956 0.8050680 1.3709362 1.1385425 0.8384109 1.1046499
[141] 1.0032475 0.7016648 1.0871943 0.8026489 1.0351851 0.7111800 1.0156046
[148] 1.2548813 1.2056290 0.9972318 0.7938395 1.1783550 1.1454552 1.2141138
[155] 1.0259884 0.8263474 0.7527548 1.1741779 0.8076669 1.1131290 1.3473443
[162] 0.6969776 1.4049245 0.7922204 0.8345625 0.8832969 1.2295219 1.2201149
[169] 0.9151313 0.7997395 0.7931496 1.1226843 1.1290933 1.0566546 1.2901484
[176] 1.0328472 1.3198778 1.2616384 0.9585140 1.0564211 0.9140546 0.9761980
[183] 0.9099171 0.8790828 0.8793479 0.7908106 0.8645623 1.2383979 1.3122934
[190] 1.0226004 1.3359423 1.2238553 0.8101288 0.8199179 1.1858293 0.6280127
[197] 1.0097768 0.7817456 0.9581183 0.8510940

```

## Pour la distribution de Paréto

Pour la loi de Paréto, les estimateurs de Hill sont obtenus par la fonction `estimateur_hill` et son résultat est gardé dans la variable `estimateur_pareto` définie par :

$$estimateur\_pareto = estimateur\_hill(X, n, m, k)estimateur\_pareto$$

Après simulation de m échantillons générés par loi de Paréto généralisé de taille n=200 et en posant k=30, l'évaluation des estimateurs de Hill est donné par le graphe suivant :

> estimateur_Pareto					
[1]	0.120376351	0.120388216	0.940576012	0.329191278	0.661677578
[7]	0.544133459	0.134543179	0.691888526	0.179512022	0.290099631
[13]	0.166424949	0.342379969	0.507960363	0.858158646	0.654511389
[19]	0.428469608	0.301137687	0.662062684	0.290968650	0.355467563
[25]	0.482251380	1.266471436	0.599440768	0.813807371	0.381736112
[31]	0.580533221	0.576310555	2.257938033	0.686116380	0.279643327
[37]	0.167696932	0.060610115	0.905553935	0.284260766	0.697852208
[43]	0.229256568	0.666564934	0.117269416	0.006501023	0.526237611
[49]	0.359120370	0.573432298	0.279373240	1.654110329	0.023479526
[55]	0.514850478	0.196105638	0.058915282	1.011453390	0.175223158
[61]	1.714732337	0.617233403	0.338012510	0.308626321	1.081115134
[67]	0.009207422	0.013915414	0.040426990	0.711744530	1.859510462
[73]	0.346271392	0.718369061	0.866883173	1.383250096	0.101961130
[79]	0.697171446	0.030795235	0.364025090	0.423729104	0.274174033
[85]	0.287010715	0.344579709	0.047073812	0.915013922	0.268929591
[91]	0.563928696	0.102041716	0.274033797	0.140637633	0.143443161
[97]	0.089260153	1.004659968	0.361503775	0.953836900	0.997796891
[103]	0.046751625	0.053338029	0.079583636	0.229624288	0.396407644
[109]	0.427035868	0.401063316	0.131583974	1.052308691	1.112699231
[115]	0.409431807	0.292195939	0.754857084	0.175179245	0.540760845
[121]	0.168848553	0.067218902	1.990852649	0.864105656	0.003382547
[127]	0.040110684	0.050773581	0.142340133	0.296587101	0.003815563
[133]	0.951573383	0.264914568	0.782222360	0.075832601	0.286516994
[139]	0.616563706	0.706344765	0.130924149	0.108722269	0.306505036
[145]	0.533453323	0.087183203	0.205043945	0.043881380	0.843072451
[151]	0.517897975	0.966928023	0.280312010	0.362860696	0.727762698
[157]	0.693233942	0.299640498	0.285084419	0.057408045	0.774466173
[163]	0.350496033	0.273113894	1.643792110	1.477916735	3.341317863
[169]	0.060743117	0.289132595	0.251130481	0.653807858	0.420505367
[175]	0.751835804	0.218729167	0.423362284	0.061137008	0.253916501
[181]	0.412510224	0.498121807	0.720447101	0.830045695	0.149249811
[187]	1.403186144	1.631365823	1.162526576	0.432692105	1.636999497
[193]	0.578717449	0.162538613	0.115513871	0.483271428	0.778843318
[199]	0.372143941	0.986820790			

## Estimateur moyen Hill

L' estimateur moyen est défini par :

$$\bar{\gamma}_{k,n}^H = \frac{1}{m} \sum_{j=1}^m \hat{\gamma}_{k,n,i}^H$$

Pour l'évaluation de cet estimateur, nous avons la fonction **estimateur\_moyen\_hill()** qui renvoie la moyenne des estimateurs de Hill des m échantillons.

## Estimateur moyen pour la distribution de Fréchet

Pour n=200 et k=30

```
> estimateur_moyen_hill(X,n,m,k)
[1] 1.023431
```

$$\bar{\gamma}_{k,n}^H = 1.023431$$

Pour n=500 et k=70

```
> estimateur_moyen_hill(X,n,m,k)
[1] 1.040598
```

$$\bar{\gamma}_{k,n}^H = 1.040598$$

Pour n=1000 et k=150

```
> estimateur_moyen_hill(X,n,m,k)
[1] 0.9919213
```

$$\bar{\gamma}_{k,n}^H = 0.9919213$$

## Estimateurs moyen pour la distribution de Paréto

Pour n=200 et k=30

```
> estimateur_moyen_hill(X, n, m,k)
[1] 0.5408905
```

$$\bar{\gamma}_{k,n}^H = 0.5408905$$

Pour n=500 et k=70

```
> estimateur_moyen_hill(X,n,m,k)
[1] 0.5395884
```

$$\bar{\gamma}_{k,n}^H = 0.5395884$$

Pour n=1000 et k=150

```
> estimateur_moyen_hill(X, n, m, k)
[1] 0.5406236
```

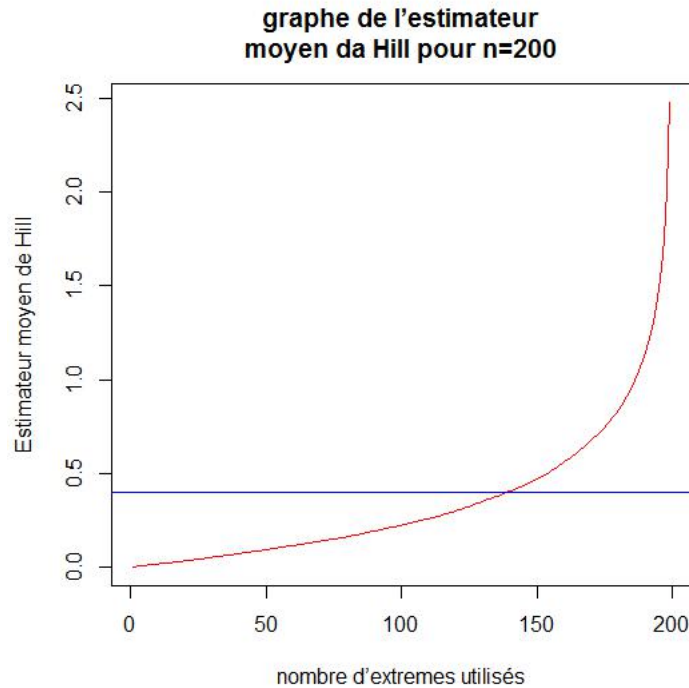
$$\bar{\gamma}_{k,n}^H = 0.5406236$$



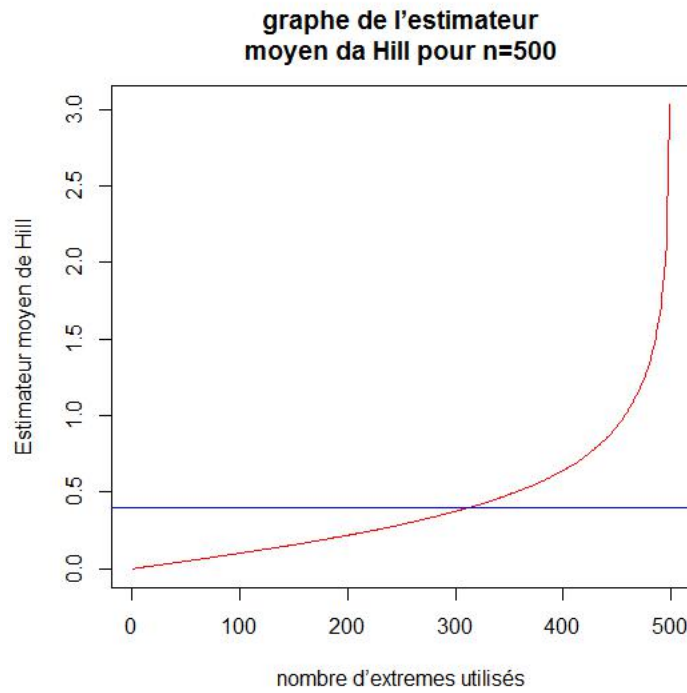
## Graphe des Estimateurs moyen de Hill

Dans cette partie , pour chaque valeur de  $n$  nous allons tracer le graphe  $(k, \bar{\gamma}_{k,n}^H), 1 \leq k < n$  de l'estimateur  $\gamma$  puis nous spécifierons la droite correspondante à la vraie valeur de  $\gamma$ . De ce fait, nous faisons appel à la fonction **vect\_estimateur\_moyen\_hill()** qui renvoie un vecteur de taille  $n - 1$  dont chaque composante représente la valeur de l'estimateur moyen de Hill pour les  $m$  échantillons suivant la valeur de  $k=1,...,n-1$ . **Graphes de l'estimateur moyen pour la distribution de Fréchet**

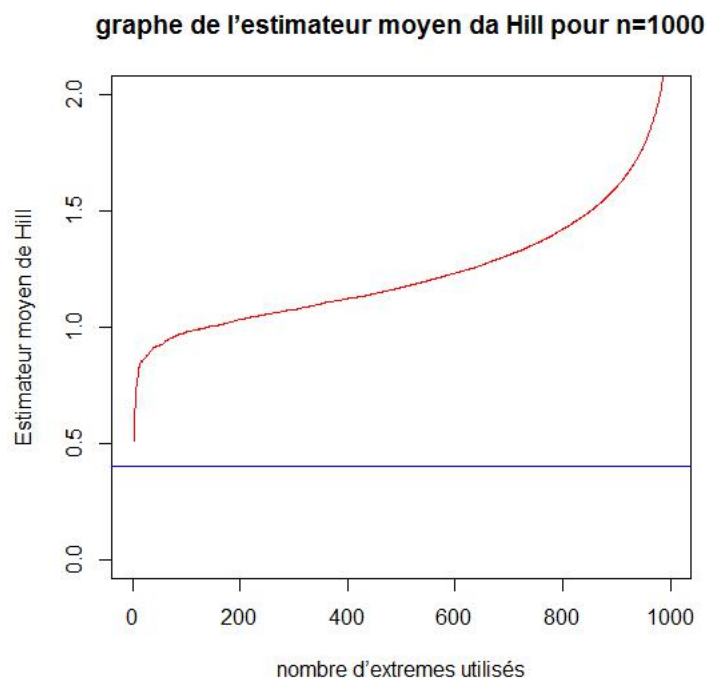
**Pour  $n=200$**



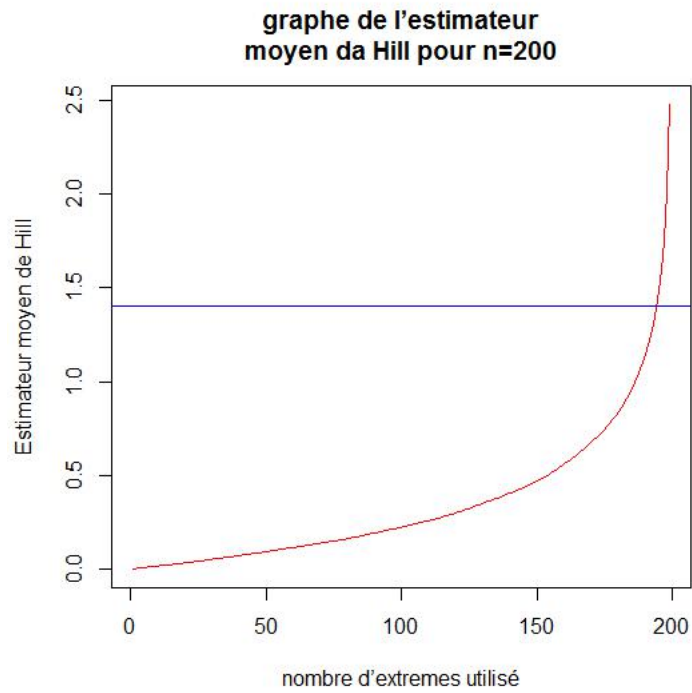
**Pour  $n=500$**



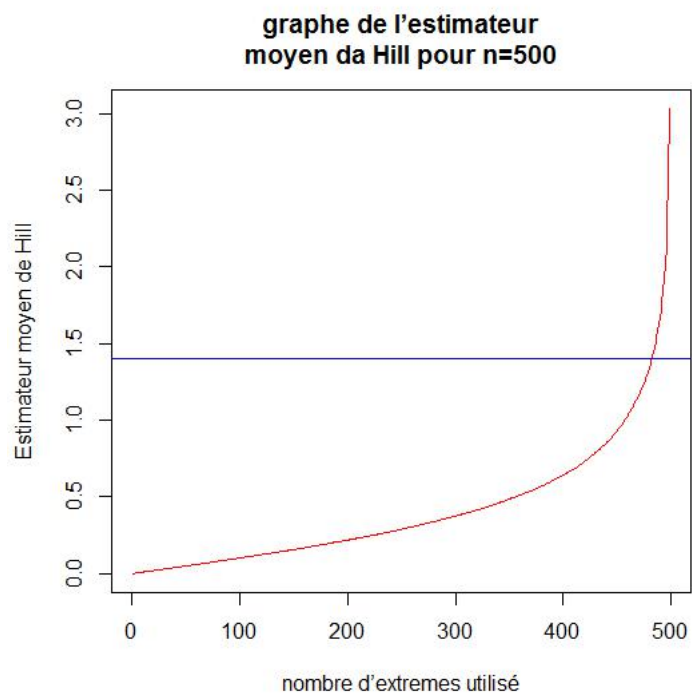
Pour  $n=1000$



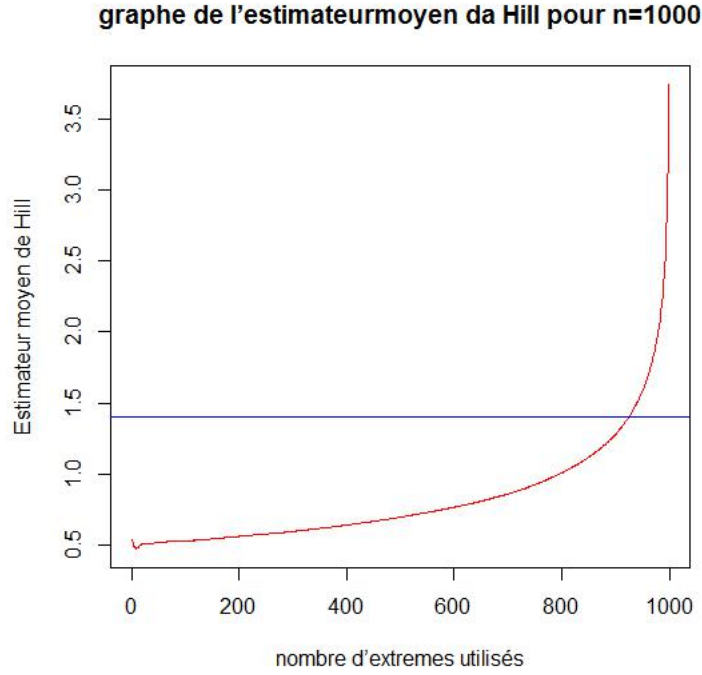
Graphes de l' estimateur moyen pour la distribution de Paréto  
Pour  $n=200$



Pour  $n=500$



Pour  $n=1000$



### 0.2.2 Estimateur de Hill adapté

Cet estimateur a été proposé par Beirlant et al en 1996. Il est applicable pour tout  $\gamma \in \mathbb{R}$  et est défini par :

$$\hat{\gamma}_{k,n}^{UH} = \frac{1}{k} \sum_{j=1}^k \log UH_{j,n} - \log UH_{k+1,n}, \quad 1 \leq k < n$$

où

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

#### Pour la distribution de Fréchet

Pour la loi de Fréchet, les estimateurs de Hill adapté sont obtenus par la fonction **estimateur\_hill\_adapte()** et son résultat est gardé dans la variable **estimateur\_frechet** définie par :

$$estimateur\_frechet = estimateur\_hill\_adapte(X, n, m, k) estimateur\_frechet$$

#### Exemple :

Après avoir simulé les  $m$  échantillons générés par la loi de Fréchet de taille  $n=200$  et en posant  $k=30$ , l'évaluation des estimateurs de Hill adapté est donné par le graphe suivant :

#### Pour la distribution de Paréto

Pour la loi de Paréto, les estimateurs de Hill adapté sont obtenus par la fonction **estimateur\_hill\_adapte** et son résultat est gardé dans la variable **estimateur\_pareto** définie par :

$$estimateur\_pareto = estimateur\_hill\_adapte(X, n, m, k) estimateur\_pareto$$

Après simulation de  $m$  échantillons générés par loi de Paréto généralisé de taille  $n=200$  et en posant  $k=30$ , l'évaluation des estimateurs de Hill adapté est donné par le graphe suivant :

## Estimateur moyen de Hill Adapté

L' estimateur moyen est défini par :

$$\bar{\gamma}_{k,n}^{UH} = \frac{1}{m} \sum_{j=1}^m \hat{\gamma}_{k,n,i}^{UH}$$

Pour l'évaluation de cet estimateur, nous avons la fonction **estimateur\_moyen\_hill()** qui renvoie la moyenne des estimateurs de Hill adapté des m échantillons.

### Pour la distribution de Fréchet

Pour n=200 et k=30

```
> estimateur_moyen_hill_adapte(X,n,m,k)
[1] 0.8773189
```

$$\bar{\gamma}_{k,n}^{UH} = 0.8773189$$

### Pour la distribution de Paréto

Pour n=200 et k=30

```
> estimateur_moyen_hill_adapte(X,n,m,k)
[1] 0.407259
```

$$\bar{\gamma}_{k,n}^{UH} = 0.407259$$

Pour n=500 et k=70

```
> estimateur_moyen_hill(X,n,m,k)
[1] 0.5393595
```

$$\bar{\gamma}_{k,n}^{UH} = 0.5393595$$

Pour n=1000 et k=150

```
> estimateur_moyen_hill_adapte(X,n,m,k)
[1] 0.4669065
```

$$\bar{\gamma}_{k,n}^{UH} = 0.4669065$$

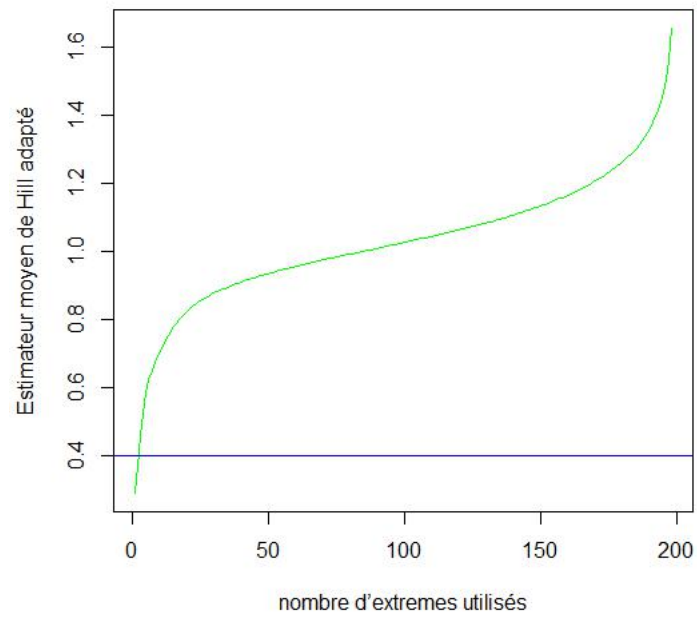
## Graphe des Estimateurs moyen de Hill adapté

### Graphes de l' estimateur moyen pour la distribution de Fréchet

Pour n=200

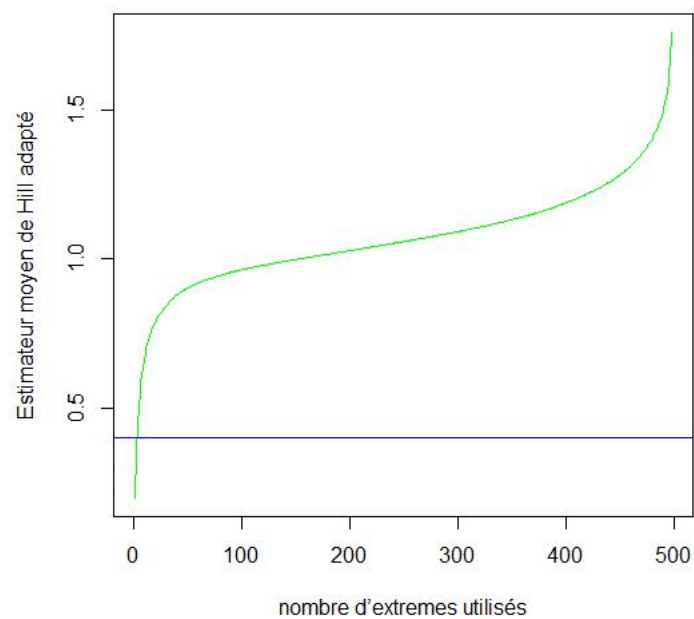


**graphe de l'estimateur moyen de Hill adapté n=200**



Pour  $n=500$

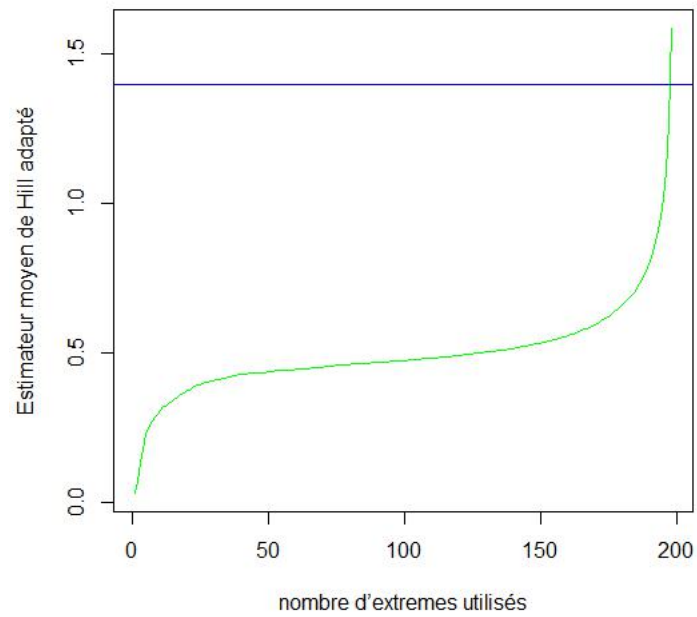
**graphe de l'estimateur moyen de Hill adapté n=500**



Graphes de l' estimateur moyen pour la distribution de Paréto

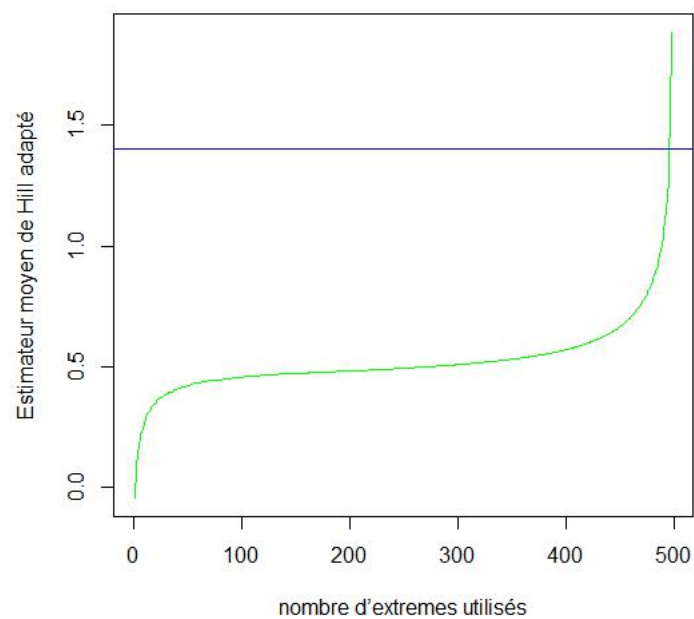
Pour  $n=200$

**graphe de l'estimateur moyen de Hill adapté n=200**

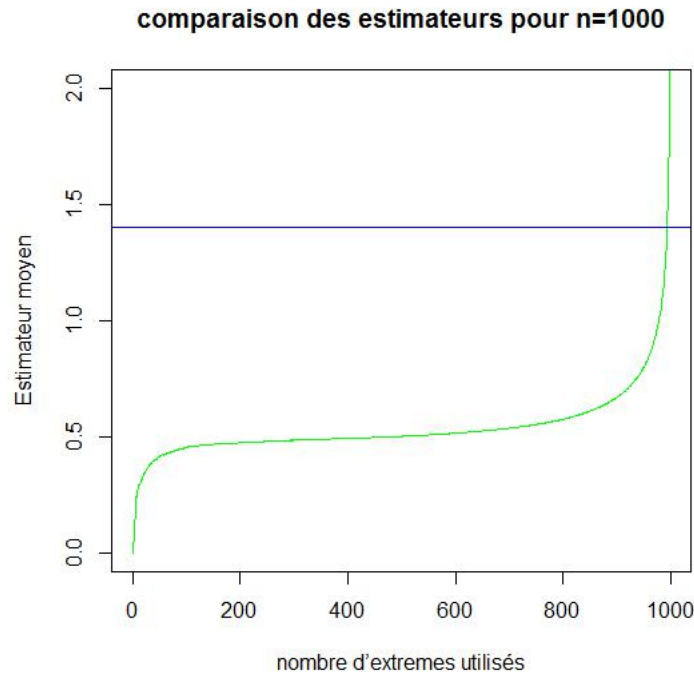


Pour  $n=500$

**graphe de l'estimateur moyen de Hill adapté n=500**



Pour  $n=1000$



## 0.3 Erreur quadratique moyenne

L'erreur quadratique moyenne d'un estimateur  $\hat{\gamma}_{k,n}$  d'un paramètre  $\gamma$  est une mesure caractérisant la "précision" de cet estimateur et est définie par

$$MSE(k) = \frac{1}{m} \sum_{i=1}^m (\hat{\gamma}_{k,n,i} - \gamma)^2, 1 \leq k < n$$

### 0.3.1 Erreur quadratique moyenne de l' estimateur de Hill

L'évaluation de l'erreur quadratique moyenne est obtenue grâce à la fonction **MSE()** et son résultat est gardé dans la variable **mse**.

**Pour la distribution de Fréchet**

**Pour n=200**

```
> mse=MSE(X,n,m,k,gamma)
> mse
[1] 0.4478109
```

**MSE(30) =**

**Pour n=500**

```
> k=70
> mse=MSE(X,n,m,k,gamma)
> mse
[1] 0.4221256
```

**MSE(70)= 0.4221256**

## Pour la distribution de Paréto

Pour  $n=200$

```
> mse=MSE(X,n,m,k,gamma)
> mse
[1] 0.7465444
```

$MSE(30) = 0.7465444$

Pour  $n=500$

```
> mse=MSE(X,n,m,k,gamma)
> mse
[1] 0.7439456
```

$MSE(70) = 0.7439456$

Pour  $n=1000$

```
>k=150
> mse=MSE(X, n, m, k, gamma)
> mse
[1] 0.7400603
```

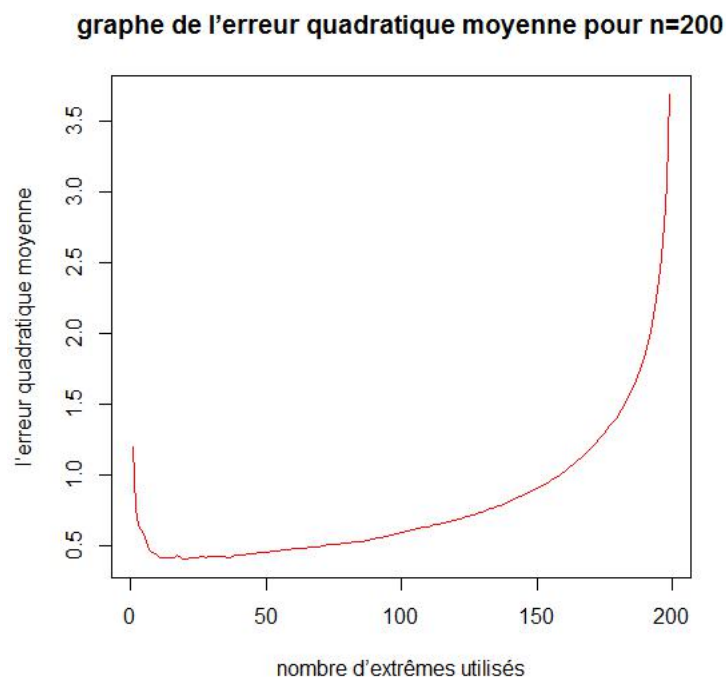
$MSE(150) = 0.7400603$

### 0.3.2 Graphe de l'erreur quadratique moyenne de l'estimateur de Hill

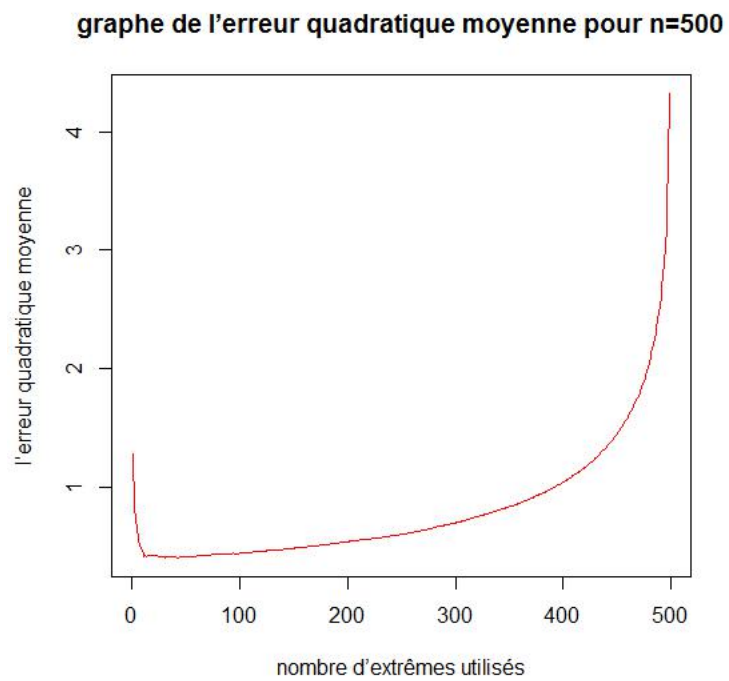
Dans cette partie, il s'agira de tracer le graphe  $(k, MSE(k))$ ,  $1 \leq k < n$  de l'erreur quadratique moyenne en fonction du nombre d'extrêmes  $k$  utilisés. De ce fait, nous faisons appel à la fonction **msev\_vec()** qui renvoie un vecteur de taille  $n - 1$  dont chaque composante représente la valeur de l'erreur quadratique moyenne de Hill pour les  $m$  échantillons suivant la valeur de  $k=1, \dots, n-1$ .

## Pour la distribution de Fréchet

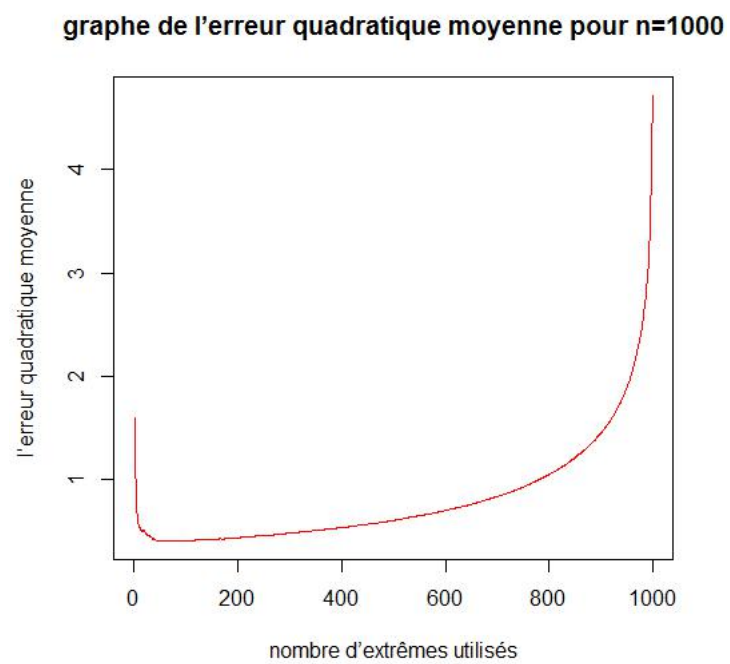
Pour  $n=200$



Pour  $n=500$



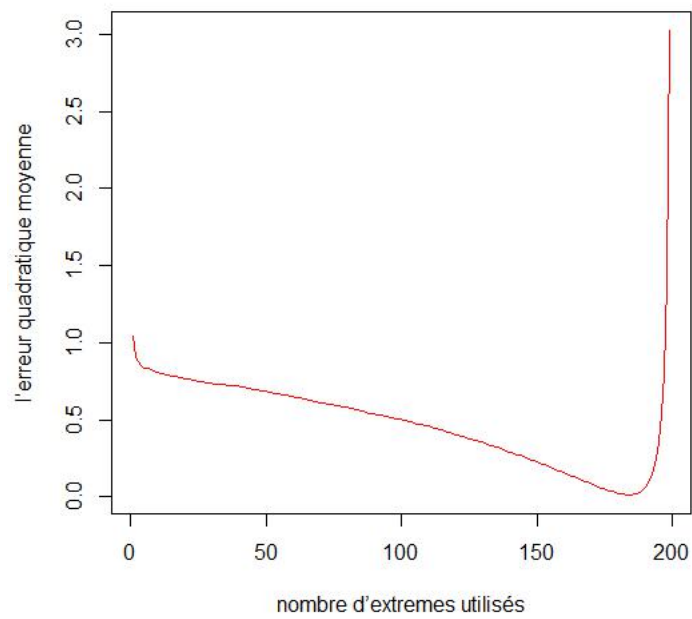
Pour  $n=1000$



Pour la distribution de Paréto

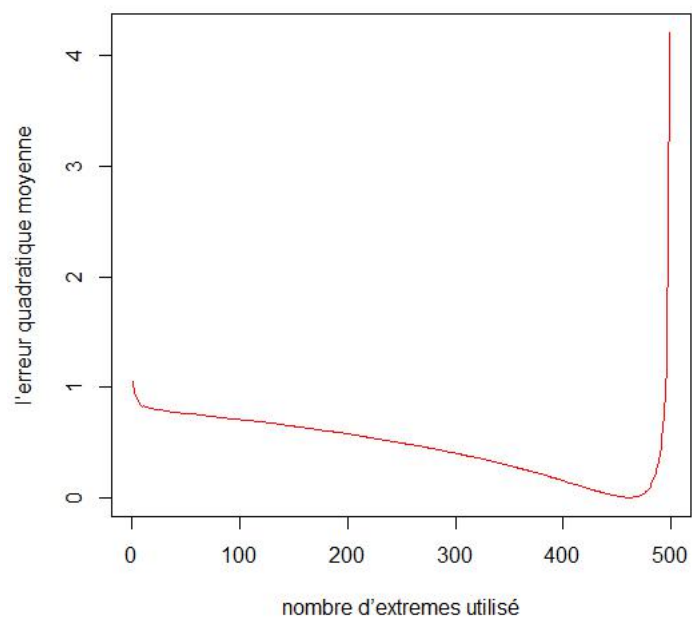
Pour  $n=200$

**graphe de l'erreur quadratique moyenne pour n=200**

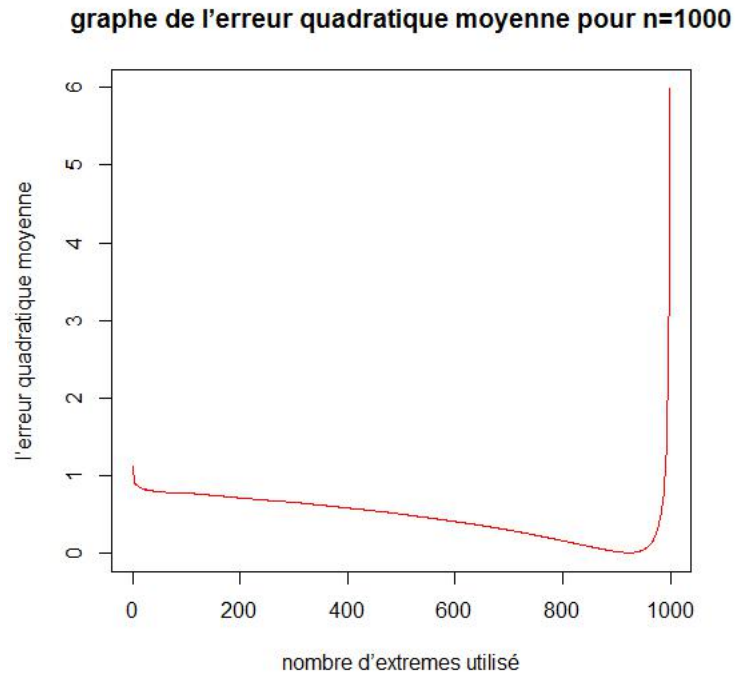


Pour n= 500

**graphe de l'erreur quadratique moyenne pour n=500**



Pour n=1000



### 0.3.3 Erreur quadratique moyenne de l' estimateur de Hill adapté

L'évaluation de l'erreur quadratique moyenne est obtenue grâce à la fonction **MSE()** et son résultat est gardé dans la variable **mse**.

**Pour la distribution de Fréchet**

**Pour n=200**

```
> mse=MSE(X,n,m,k,gamma)
> mse
[1] 0.2682075
```

$$\text{MSE}(30) = 0.26682075$$

**Pour la distribution de Paréto**

```
> mse=MSE(X,n,m,k,gamma)
> mse
[1] 1.015872
```

$$\text{MSE}(30) = 1.015872$$

```
> mse=MSE(X,n,m,k,gamma)
> mse
[1] 0.931997
```

$$\text{MSE}(70) = 0.931997$$

```
> mse=MSE(X,n,m,k,gamma)
> mse
[1] 0.8794401
```

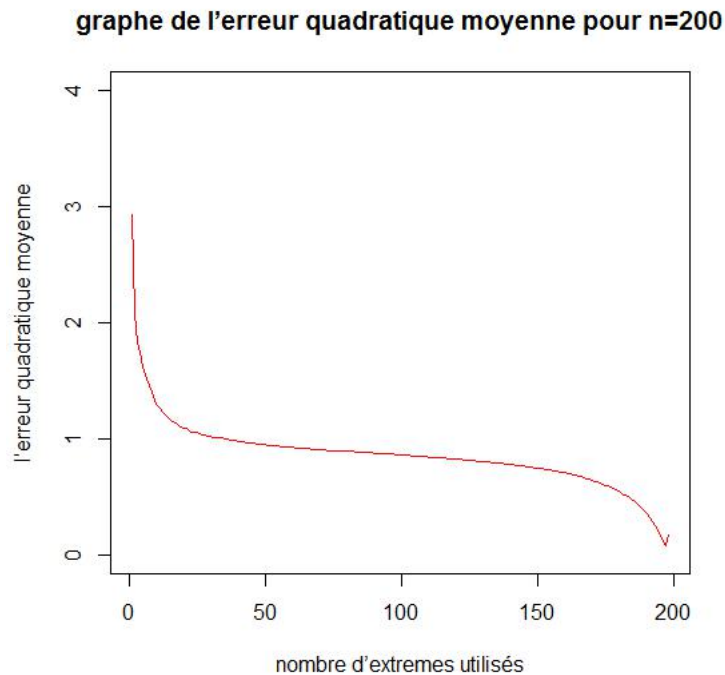
$$\text{MSE}(150) = 0.8794401$$

### 0.3.4 Graphe de l'erreur quadratique moyenne de l' estimateur de Hill adpté

Dans cette partie, il s'agira de tracer le graphe  $(k, \text{MSE}(k))$ ,  $1 \leq k < n$  de l'erreur quadratique moyenne en fonction du nombre d'extrêmes  $k$  utilisés. De ce fait, nous faisons appel à la fonction `msev_vec()` qui renvoie un vecteur de taille  $n - 2$  dont chaque composante représente la valeur de l' erreur quadratique moyenne de Hill adapté pour les  $m$  échantillons suivant la valeur de  $k=1, \dots, n-2$ .

Pour la distribution de Paréto

Pour  $n=200$



## 0.4 Nombre optimal d'extrêmes au point $k_0 = k_{opt}$

Suivant chaque valeur de  $n$ , nous allons trouver le nombre optimal d'extrêmes noté  $k_{opt}$  correspondant à la valeur de  $k$  qui minimise l'erreur quadratique moyenne  $\text{MSE}(k)$ ,  $1 \leq k < n$  et qui est défini par :

$$k_{opt} = \underset{k}{\operatorname{argmin}} \text{MSE}(k)$$

### 0.4.1 Nombre optimal d'extrêmes avec l' estimateur de Hill

Pour la distribution de Fréchet

Pour  $n=200$

```
> k_opt=which.min(msev)
> k_opt
[1] 21
```



$k_{opt} = 21$   
Pour n=500

```
> k_opt=which.min(msev)
> k_opt
[1] 30
```

$k_{opt} = 30$

Pour la distribution de Paréto

Pour n=200

```
> k_opt=which.min(msev)
> k_opt
[1] 184
```

$k_{opt} = 184$   
Pour n=500

```
> k_opt=which.min(msev)
> k_opt
[1] 461
```

$k_{opt} = 461$   
Pour n=1000

```
> k_opt=which.min(msev)
> k_opt
[1] 924
```

$k_{opt} = 924$

## 0.4.2 Nombre optimal d'extrêmes avec l' estimateur de Hill adapté

Pour la distribution de Paréto

Pour n=200

```
> k_opt=which.min(msev)
> k_opt
[1] 197
```

$k_{opt} = 197$

## 0.5 Valeur de l' estimateur moyen au point $k_0 = k_{opt}$

La valeur de l' estimateur moyen au point  $k_0$  est noté  $\bar{\hat{\gamma}}_{k,n}$ .

### 0.5.1 Pour l' estimateur de Hill

Pour la distribution de Fréchet

Pour n=200

```
> estimateur_moyen_k_o= estimateur_moyen_hill(X,200,m,k_opt)
> estimateur_moyen_k_o
[1] 0.9996577
```

$$\bar{\hat{\gamma}}_{k_0,n}^H = 0.9996577$$

Pour n=500

```
> estimateur_moyen_k_o= estimateur_moyen_hill(X,n,m,k_opt)
> estimateur_moyen_k_o
[1] 1.010668
```

$$\bar{\hat{\gamma}}_{k_0,n}^H = 1.010668$$

Pour la distribution de Paréto

Pour n=200

```
> estimateur_moyen_k_o=estimateur_moyen_hill(X,n,m,k_opt)
> estimateur_moyen_k_o
[1] 1.382868
```

$$\bar{\hat{\gamma}}_{k_0,n}^H = 1.382868$$

Pour n=500

```
> estimateur_moyen_k_o= estimateur_moyen_hill(X,n,m,k_opt)
> estimateur_moyen_k_o
[1] 1.387603
```

$$\bar{\hat{\gamma}}_{k_0,n}^H = 1.387603$$

Pour n=1000

```
> estimateur_moyen_k_o= estimateur_moyen_hill(X, n, m,k_opt)
> estimateur_moyen_k_o
[1] 1.397688
```

$$\bar{\hat{\gamma}}_{k_0,n}^H = 1.397688$$

## 0.5.2 Pour Hill adapté

Pour n=200

```
> estimateur_moyen_ko=estimateur_moyen_hill_adapte(X,n,m,k_opt)
> estimateur_moyen_ko
[1] 1.252602
```

$$\bar{\hat{\gamma}}_{k_0,n}^H = 1.382868$$

## 0.6 Erreur absolue au point $k_0 = k_{opt}$

L'erreur absolue représente la différence entre la valeur réelle et la valeur mesurée. En évaluant l'erreur absolue, nous pouvons avoir une idée de la précision d'une valeur. Calculer l'erreur absolue au point  $k_{opt}$ , revient à calculer la valeur suivante :

$$|\bar{\hat{\gamma}}_{k_0,n}^H - \gamma|$$

La fonction `erreur_abs()` nous permettra d'obtenir la valeur ci dessus.

### 0.6.1 Erreur absolue au point $k_0$ pour l' estimateur de Hill

Pour la distribution de Fréchet

Pour n=200

```
> erreur_abs=abs(estimateur_moyen_k_o - gamma)
> erreur_abs
[1] 0.5996577
```

$$|\bar{\hat{\gamma}}_{k_0,200}^H - \gamma| = 0.5996577$$

Pour n=500

```
> erreur_abs=abs(estimateur_moyen_k_o - gamma)
> erreur_abs
[1] 0.6106681
```

$$|\bar{\hat{\gamma}}_{k_0,500}^H - \gamma| = 0.6106681$$

Pour la distribution de Paréto

Pour n=200

```
> erreur_abs=abs(estimateur_moyen_k_o - gamma)
> erreur_abs
[1] 0.01713226
```

$$|\hat{\gamma}_{k_0,200}^H - \gamma| = 0.01713226$$

Pour n=500

```
> erreur_abs=abs(estimateur_moyen_k_o - gamma)
> erreur_abs
[1] 0.01239732
```

$$|\hat{\gamma}_{k_0,500}^H - \gamma| = 0.01239732$$

Pour n=1000

```
> erreur_abs=abs(estimateur_moyen_k_o - gamma)
> erreur_abs
[1] 0.00231228
```

$$|\hat{\gamma}_{k_0,1000}^H - \gamma| = 0.0231228$$

## 0.6.2 Erreur absolue au point $k_0$ pour l' estimateur de Hill adapté

Pour la distribution de Paréto

Pour n=200

```
> erreur_abs=abs(estimateur_moyen_ko - gamma)
> erreur_abs
[1] 0.1473979
```

$$MSE(k_0) = 0.1473979$$

## 0.7 Erreur quadratique moyenne au point $k_0$

### 0.7.1 Erreur quadratique moyenne de Hill avec le nombre optimal d'extrêmes $MSE(k_0)$

Pour la distribution de Fréchet

Pour n=200

```
> mse_k_o=MSE(X,n,m,k_opt,gamma)
> mse_k_o
[1] 0.407003
```

$$MSE(k_0) = 0.407003$$

Pour n=500

```
> mse_k_o=MSE(X,n,m,k_opt,gamma)
> mse_k_o
[1] 0.4037678
```

$$MSE(k_0) = 0.4037678$$

**Pour la distribution de Paréto**

**Pour n=200**

```
> mse_k_o=MSE(X,n,m,k_opt,gamma)
> mse_k_o
[1] 0.01550283
```

$$MSE(k_0) = 0.01550283$$

**Pour n=500**

```
> mse_ko=MSE(X,n,m,k_opt)
> mse_ko
> 0.005863853
```

$$MSE(k_0) = 0.005863853$$

**Pour n=1000**

```
> mse_k_o=MSE(X, n, m, k_o, gamma)
> mse_k_o
[1] 0.002775768
```

$$MSE(k_0) = 0.002775768$$

## 0.7.2 Erreur quadratique moyenne de Hill adapté avec le nombre optimal d'extrêmes $MSE(k_0)$

**Pour la distribution de Paréto**

**Pour n=200**

```
> mse_k_o=MSE(X,n,m,k_opt,gamma)
> mse_k_o
[1] 0.08018812
```

$$MSE(k_0) = 0.08018812$$

## 0.8 Déviation standard

La déviation standard(écart type) est une notion mathématique définie en probabilité et appliquée à la statistique. en probabilité, l'écart type est une mesure de la dispersion d'une variable aléatoire ; en statistique elle est une mesure de dispersion de données Elle est définie comme suit :

$$\sqrt{\frac{1}{m} \sum_{i=1}^m (\hat{\gamma}_{k_0,n,i} - \bar{\hat{\gamma}}_{k_0,n})^2}$$

### 0.8.1 Déviation standard de l' estimateur de Hill

Pour la distribution de Fréchet

Pour n=200

```
> D=deviation(X,n,m,k_opt)
> D
[1] 0.4554269
```

D=0.4554269

Pour n=500

```
> D=deviation(X,n,m,k_opt)
> D
[1] 0.4368664
```

D=0.4368664

Pour la distribution de Paréto

Pour n=200

```
> D=deviation(X,n,m,k_opt)
> D
[1] 0.5075166
```

D=0.5075166

Pour n=500

```
> D=deviation(X,n,m,k_opt)
> D
[1] 0.5141685
```

D=0.5141685

Pour n=1000

```
> D=deviation(X,n,m,k_opt)
> D
> 0.5041231
```

D=0.5041231

## 0.8.2 Déviation standard de l' estimateur de Hill adapté

Pour la distribution de Paréto

Pour n=200

```
> D=deviation(X,n,m,k_opt)
> D
[1] 1.131344
```

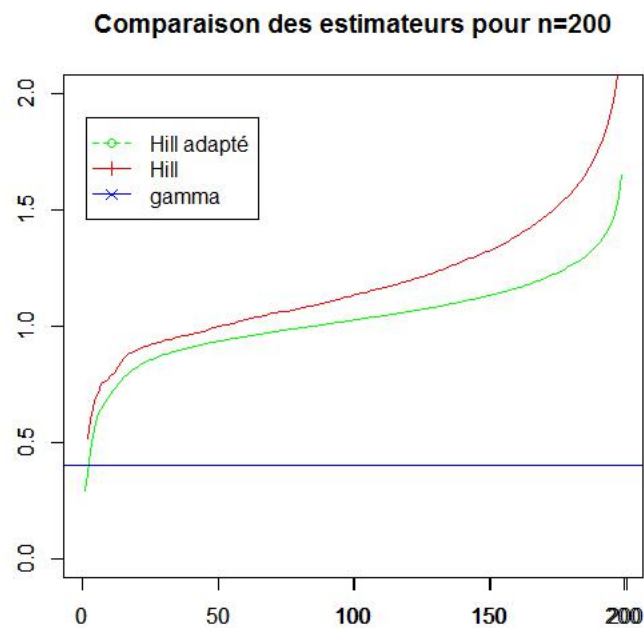
D=0.4554269

## 0.9 Étude comparative des différents estimateurs

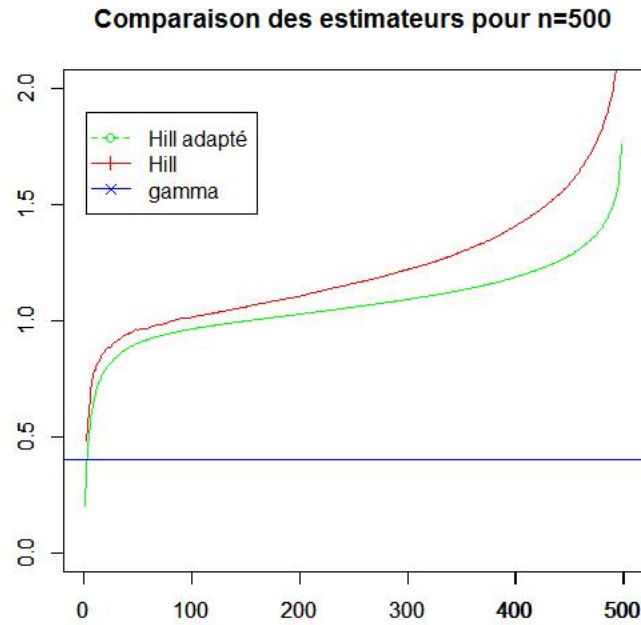
### 0.9.1 Pour la distribution de Fréchet

Étude graphique

Pour n=200



Pour n=500



Pour les graphes ci dessus, nous constatons que l'estimateur Hill adapté est plus proche de la vraie valeur de  $\gamma$  que l'estimateur de Hill. Également, une légère amélioration est observée lorsque la taille de l'échantillon augmente.

### Étude numérique

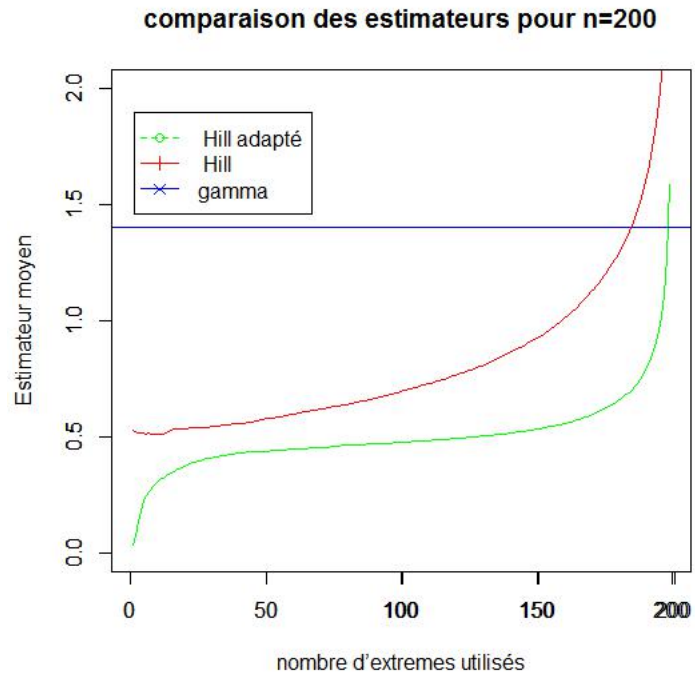
	n	$\gamma$	$k_o$	$\hat{\gamma}_o$	$Erreur_o$	$MSE_o$	$Deviation_o$
estimateur de Hill	200	0.4	21	0.9996577	0.59696577	0.407003	0.4554269
	500	0.4	30	1.010668	0.6106681	0.4037678	0.4368664
	1000	0.4					
estimateur de Hill adapté	200	0.4					
	500	0.4					
	1000	0.4					

### 0.9.2 Pour la distribution de Paréto

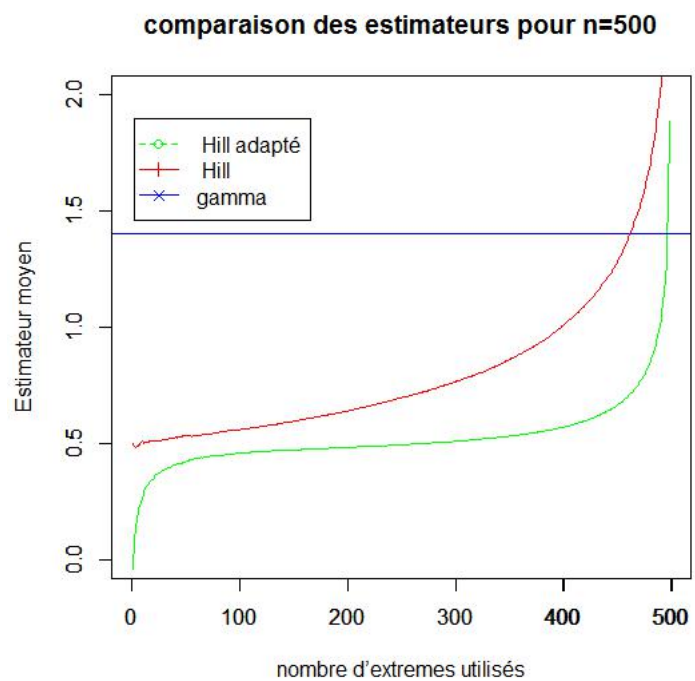
#### Étude graphique

Pour n=200

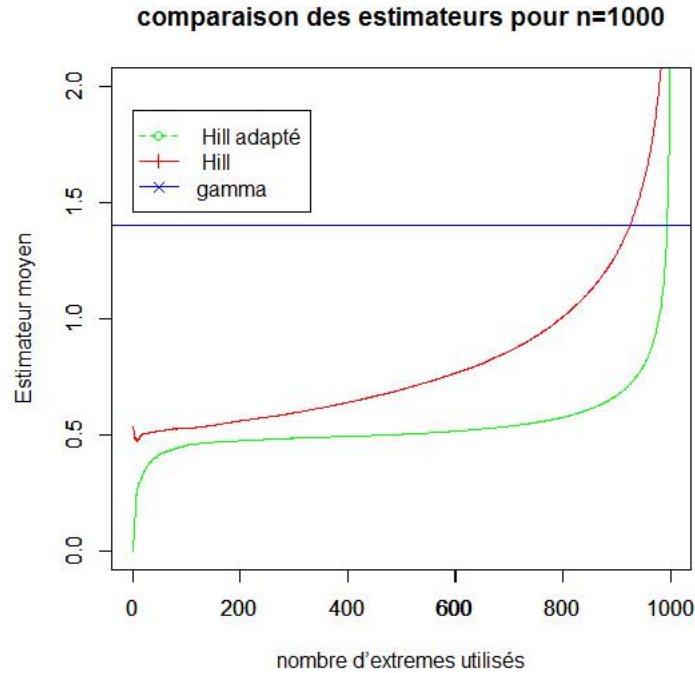




Pour  $n=500$



Pour  $n=1000$



Concernant ces trois graphes comparatifs pour la distribution de Paréto, l' estimateur de Hill est plus proche de la vraie valeur de  $\gamma$  contrairement à la distribution de Fréchet. aussi, plus la taille de l'échantillon augmente plus une légère amélioration est constatée pour les deux estimateurs.

### Étude numérique

	n	$\gamma$	$k_o$	$\hat{\gamma}_o$	$Erreur_o$	$MSE_o$	$Deviation_o$
estimateur de Hill	200	1.4	184	1.382868	0.01713226	0.01550283	0.5075166
	500	1.4	461	1.38794	0.01206038	0.005863853	0.5142409
	1000	1.4	924	1.397688	0.00231228	0.002775768	0.5041231
estimateur de Hill adapté	200	1.4	197	1.252602	0.1473979	0.08018812	1.131344
	500	1.4					
	1000	1.4					

Les résultats observés sur ce tableau confirment l'étude graphique : l' estimateur de Hill est meilleur que l' estimateur de Hill adapté. En effet, la valeur de l' estimateur de Hill au point optimal est très proche de la vraie valeur de gamma. Aussi l'erreur absolue et l'erreur quadratique moyenne tendent vers zéro au niveau de ce point.

### Conclusion

Dans ce projet, nous avons eu à étudier les estimateurs de Hill et Hill adapté sur les distributions de Fréchet et de Paréto avec des échantillon de tailles différentes. Après une étude comparative des différents estimateurs, nous avons constaté que l' estimateur de Hill est meilleur que l' estimateur de Hill adapté sur la distribution de Fréchet. Tandis que l' estimateur de Hill adapté est meilleur que celui de Hill sur la loi de Paréto. aussi, lorsque la taille de l'échantillon augmente, les estimations deviennent plus précises.

## Annexe

```
m=200 #####Fréchet#####
gamma= 0.4
#####Paréto#####
gamma= 1.4
#####PREMIÈRE PARTIE : SIMULATION#####
##### Simuler m = 200 échantillons basés sur la distribution de Frechet #####
set.seed(123)
frechet=function(n)
{
X=numeric(n)
for (i in 1 :n)
X[i]=-1/(log(runif(1)))
X
}
m_frechet=function(m,n)
{
X=matrix(nrow=n,ncol=m)
for(j in 1 :m)
X[,j]=sort(frechet(n))
X
}
X = m_frechet(m,n)
X
##### Simuler m = 200 échantillons basés sur la distribution de Paréto #####
set.seed(123)
pareto=function(n)
{
Y=1 :n
for (i in 1 :n)
Y[i]=2*sqrt(1/(1-runif(1))-1)
Y
}
m_pareto=function(m,n)
{
Y=matrix(nrow=n,ncol=m)
for(j in 1 :m)
Y[,j]=sort(pareto(n))
return(Y)
}
Y=m_pareto(m,n)
Y
#####DEUXIEME PARTIE : EVALUATION DES ESTIMATEURS#####
#####A.ESTIMATEUR DE HILL#####
somme=function(i,n,k,X)
{
som=0
```

```

for(j in 1 :k)
{
som=som + log(X[(n-j+1),i])
}
return(som)
}
estimateur_hill = function(X,n,m,k)
{
EH = numeric(m)
for (i in 1 :m)
{
EH[i]=((1/k)*somme(i,n,k,X))- log(X[n-k,i])
}
return(EH)
}

#####A. Estimateur de Hill adapté #####
j=numeric
UH=function(i,j,n,X)
{
uh=numeric
som=0
for (l in 1 :j)
som=som + (log(X[(n-l+1),i])- log(X[(n-j),i]))
uh=X[(n-j),i]*(1/j)*som
return(uh)
}
somme_hill_adapte=function(i,X,n,k)
{
som=0
for(j in 1 :k)
som=som+log(UH(i,j,n,X))
return(som)
}
estimateur_hill_adapte=function(X,n,m,k)
{
eha=numeric(m)
for (i in 1 :m)
eha[i]=((1/k)*somme_hill_adapte(i,X,n,k))- log(UH(i,(k+1),n,X))
return(eha)
}
#####ESTIMATEUR MOYEN de HILL #####
estimateur_moye_hill=function(X,n,m,k)
{
M = mean(estimateur_hill(X,n,m,k))
return(M)
}

```

```

#####Estimateur moyen Hill adapté #####
estimateur_moyen_hill_adapte=function(X,n,m,k)
{ M=mean(estimateur_hill_adapte(X,n,m,k))
return(M)
}

#####TROISIEME PARTIE#####
##### GRAPHE(k,estimateurmoyen_(k,n))#####
vect_estimateur_moyen_hill=function(X,n,m,k)
{
em=numeric(n-1)
j=1
for(k in 1 : n-1)
{
em[j]=estimateur_moyen_hill(X,n,m,k)
j=j+1
}
return(em)
}
k= numeric
k_estimateur=vect_estimateur_moyen_hill(X,n-1,m,k)
k_estimateur
k=1 :(n-1)
X11()
plot(k, k_estimateur, main="graphe de l' estimateur moyen da Hill pour n=500",
xlab = "nombre d' extrêmes utilisés", ylab = "Estimateur moyen de Hill", type="l", col="red")
abline(h=0.4, col="blue")
savePlot(filename="emf-fh200",type="jpeg")
##### QUATRIEME PARTIE#####
##### EVALUATION de l'ERREUR QUADRATIQUE MOYENNE #####
MSE=function(X,n,m,k,gamma)
{
mse=numeric
som=0
for ( i in 1 :m)
{
som=som +( estimateur_hill(X,n,m,k)[i]- gamma)^2
}
mse=(1/m)*som
return(mse)
}
mse=MSE(X,n,m,k,gamma)
mse

##### Traçons le graphe de l'erreur quadratique moyenne#####

mse_vec=function(X,n,m,k,gamma)
{ msev=numeric(n-1)
j=1

```

```

for(k in 1 :(n-1))
{
msev[j]=MSE(X,n,m,k,gamma)
j=j+
}
return(msev)
}
k=numeric

msev=mse_vec(X,n,m,k,gamma)
msev
k=1 :(n-1)
X11()
plot(k , msev, main="graphe de l'erreur quadratique moyenne pour n=500" , xlab="nombre d'extrêmes
utilisés" , ylab="l'erreur quadratique moyenne" , type="l",col="red")
savePlot(filename="mse-fh500",type="jpeg")
##### CINQUIEME PARTIE#####
#####NOMBRE OPTIMAL d'EXTREME#####
k_opt=which.min(msev)
k_opt

##### SIXIEME PARTIE #####
##### Pour chaque valeur de n, calculons au point k_0= k_opt#####
A.L'estimateur de Hill#####
estimateur_moyen_k_o= estimateur_moyen_hill(X,n,m,k_opt)
estimateur_moyen_k_o
#####B.L'erreur absolue#####
erreur_abs=abs(estimateur_moyen_k_o - gamma)
erreur_abs

#####C.L'erreur moyenne quadratique#####
mse_k_o=MSE(x,n,m,k_opt,gamma)
mse_k_o

#####D.Déviati on Standard#####
deviation=function(X,n,m,k_opt)
{
som=0
for (i in 1 :m)
som=som + (estimateur_hill(X,n,m,k_opt)[i] - erreur_abs)2
dev=sqrt(1/m*som)
return(dev)
}
D=deviation(X,n,m,k_opt)
D

```