Rapport - Statistique bayésienne

Philippe Real

01 mars, 2020

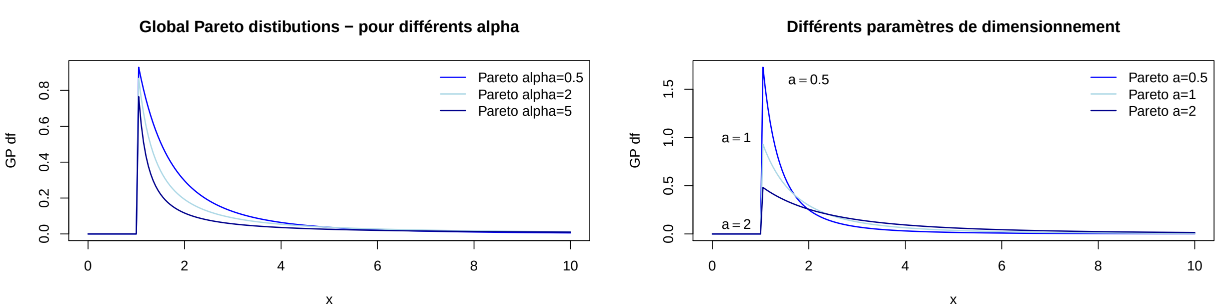
Contents

1 Loi de Pareto

On ignore maintenant les covariables, et on s'intéresse uniquement à la loi du nombre de points nécessaire (colonne Barre). La loi gaussienne peut paraître peu pertinente pour ces données : on va plutôt proposer une loi de Pareto. Pour $m > 0$ et $\alpha > 0$, on dit que $ZPareto(m; \alpha)$ si Z est à valeurs dans $[m; +1[$ de densité: $f(z \mid \alpha, m) = \alpha \frac{m^\alpha}{z^{\alpha+1}} \nVdash_{[>,+\infty[}$

1.1 Package R pour générer des réalisation d'une loi de Paréto

On peut utiliser le package *extRemes* et la fonction *devd*



1.2 Choix d'une loi à priori pour α

- Loi de paréto :

$$f(z \mid \alpha, m) = \alpha \frac{m^\alpha}{z^{\alpha+1}} \nVdash_{[>,+\infty[}$$

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    21.0   111.0   196.0   321.9   292.0  2056.0
```

Au vu des données on prend : m=21

A une constante multiplicative près et après transformation en log, on reconnaît une loi exponentielle de paramètre α.

$$f(z \mid \alpha, m) \propto \alpha e^{\alpha log(m/z)}$$

En applicant la transformation : $z \rightarrow ln(\frac{z}{m})$ a notre échantillon (Z_i), on a que $ln(\frac{Z}{m}) \sim Exp(\alpha)$

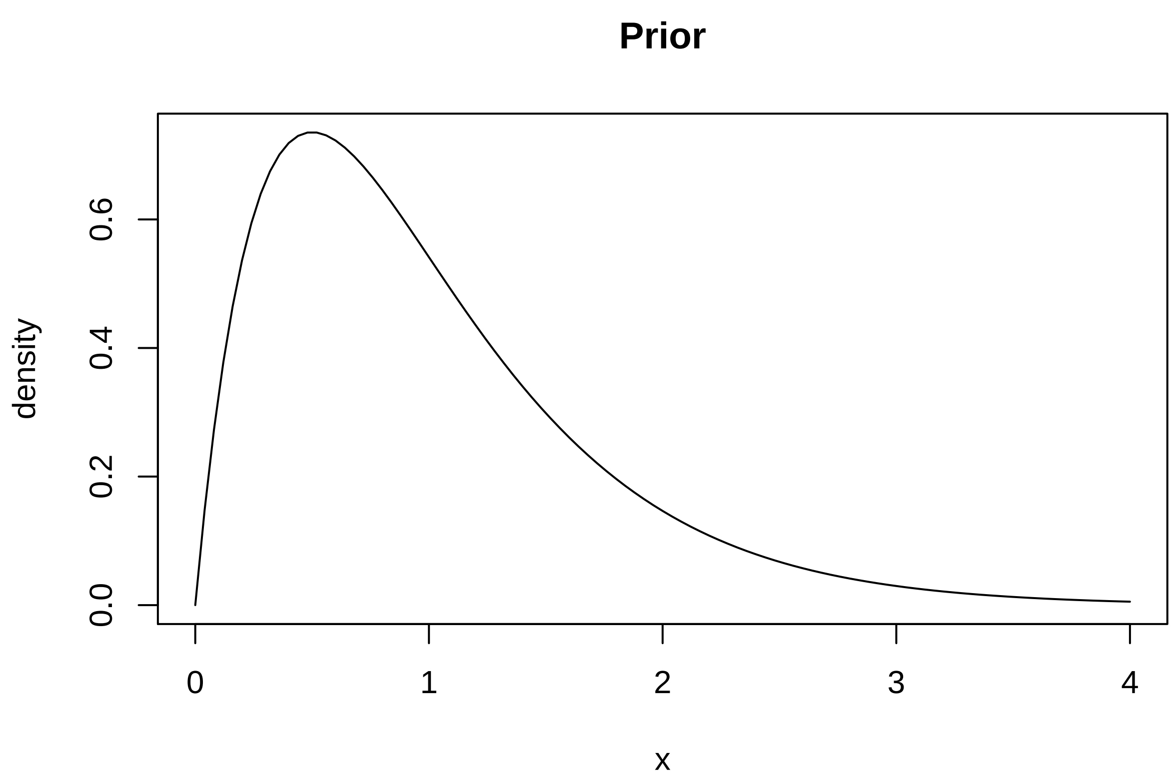
On peut alors estimer le paramètre α par mle à partir de la fonction R: *fitdist* du package *fitdistrplus*.

```
m=21
y.exp<-log(y.tot/m)
fit.exp <- fitdist(y.exp, "exp", method="mle")
fit.exp
```

```
## Fitting of the distribution ' exp ' by maximum likelihood
## Parameters:
##       estimate Std. Error
## rate 0.4502063 0.01981913
```

On peut prendre pour loi à priori la loi $\Gamma(a, b)$ de manière à avoir une loi conjuguée.

```
gam.prior<-dgamma(3, 2)
curve(dgamma(x, 2, 2), xlim=c(0, 4), main="Prior", ylab="density")
```



- EMV de *alpha*

Llog = nlog alpha + alpha n log m-(alpha+1)Somme des Xi EMV(alpha) = n/(Somme (log (Zi) + nlog m)

```
m = 21
n=length(y.tot)

EMV_alpha = n/(sum(log(y.tot)) + n*log(m))
EMV_alpha
```

```
## [1] 0.1203333
```

1.3 Loi à postériori de α

La loi à postériori correspondante est la loi : $\Gamma(a+n, b+\sum_{i=1}^{n} ln(\frac{Z_i}{m}))$

1.4 Echantillon de la loi à postériori de α

Par la méthode de votre choix, tirer un échantillon de la loi a posteriori de α. Donner un intervalle de crédibilité à 95%.

```
y.exp=log(y.tot/m)
a=2
b=1
sy=sum(y.exp)
n=length(y.tot)
alpha_mc10=rgamma(10,a+sy,b+n)
alpha_mc100=rgamma(100,a+sy,b+n)
alpha_mc1000=rgamma(1000,a+sy,b+n)
```

```
mean(alpha_mc10)
```

```
## [1] 2.240455
```

```
mean(alpha_mc100)
```

```
## [1] 2.219847
```

```
mean(alpha_mc1000)
```

```
## [1] 2.222055
```

```
mean(alpha_mc10<2.3)
```

```
## [1] 0.8
```

```r
mean(alpha_mc100<2.3)
```

```
## [1] 0.86
```

```r
mean(alpha_mc1000<2.3)
```

```
## [1] 0.883
```

```r
pgamma(2.3,a+sy,b+n)
```

```
## [1] 0.8858041
```

Quantiles: A 95% a partir de qgamma

```
## [1] 2.094162 2.351054
```

Intervalle de confiance à 95% peut aussi être obtenu à partir des path Monte Carlo:

```
##     2.5%    97.5%
## 2.123656 2.342398
```

```
##     2.5%    97.5%
## 2.095044 2.342522
```

```
##     2.5%    97.5%
## 2.080005 2.339875
```

1.5 On se concentre uniquement sur les mutations en mathématiques et en anglais. Répéter l'analyse pour chacune de ces deux catégories. Que penser de l'hypotèse d'égalité des *alpha*

```
## [1] 0.4977
```