Preparació de les eines: Estimació d'un quantil elevat

Esther Amores, Anna Costa i Oscar Ortiz

2/6/2021

Abans de començar amb la preparació de les eines, carreguem els paquets necessaris per tal d'executar les instruccions que es troben a continuació, així com també les dades `danish` del paquet `evir()`. Aquestes dades ens serviran com a referència per tal d'estimar els paràmetres d'una distribució power-law, així com també la distància de Kolmogorov-Smirnov.

```
# Paquets
library(poweRlaw)

# Dades
data("danish", package="evir")
```

Programem en `R` la funció que dóna l'estimació màxim versemblant dels paràmetres d'una power-law.

Creem una funció anomenada `ePL` que retorna $\hat{\mu}$, que és el valor mínim de les dades, el paràmetre estimat α i el valor de la log-likelihood.

Donada una mostra $\{x_1, x_2, \ldots, x_m\}$ busquem la màxima versemblança d'una distribució power-law.

- La funció de densitat és:

$$f(x; \alpha, \mu) = \frac{\alpha}{\mu} \left(\frac{\mu}{x}\right)^{\alpha+1}$$

- La funció de versemblança és:

$$L(\alpha, \mu) = \alpha^n \mu^{n\alpha} \left(\prod_{i=1}^{n} x_i\right)^{-(\alpha+1)}, \qquad (\mu \leq x_{min})$$

- La funció log-versemblança és:

$$l(\alpha, \mu) = logL(x) = n \cdot log(\alpha) + n \cdot \alpha \cdot log(\mu) - (\alpha + 1) \sum_{i=1}^{n} log(x_i)$$

Fixat α el màxim és $\mu = x_{min}$. Així doncs, els paràmetres estimats d'una power-law són:

$$\hat{\mu} = x_{1,n}$$

$$\hat{\alpha} = \left[\frac{1}{n} \sum_{x=1}^{n} log\left(\frac{x_i}{x_{1,n}}\right)\right]^{-1} = \frac{1}{\xi}$$

que anomenarem `xm` i `alpha`, respectivament.

```r
ePL <- function(xdt){
  xm <- min(xdt)
  xi <- mean(log(xdt/xm))
  n <- length(xdt)
  al <- 1/xi
  lpl <- n*log(al)+n*al*log(xm)-(al+1)*sum(log(xdt))
  list(min=xm, alpha=1/xi, lPL=lpl)
}

ePL(danish)
```

```
## $min
## [1] 1
##
## $alpha
## [1] 1.270729
##
## $lPL
## [1] -3353.128
```

Construïm un generador de nombres aleatoris per a la distribució power-law composant el generador d'uniformes, U(0,1), amb la funció quantil.

La funció de distribució d'una power-law és:

$$F(x) = 1 - \left(\frac{x}{\mu}\right)^{-\alpha}$$

La funció quantil d'una distribució power-law es calcula fent $F^{-1}(x) = y$ i aillant la x d'aquesta equació. És a dir:

$$\begin{aligned}
F^{-1}(x) = 1 - \left(\frac{x}{\mu}\right)^{-1/\alpha} &= y \\
\Leftrightarrow y - 1 &= -\left(\frac{x}{\mu}\right)^{-1/\alpha} \\
\Leftrightarrow (y-1)^{-1/\alpha} &= -\frac{x}{\mu} \\
\Leftrightarrow x &= \mu(1-y)^{-1/\alpha}
\end{aligned}$$

$$Q(y) = F^{-1}(x) = \mu(1-y)^{-1/\alpha}$$

```r
rgpl <- function(mu, alpha, n, seed){
  set.seed(seed)
  y <- runif(n)
  x <- mu*((1-y)^(-1/alpha))
  return(x)
}
```

Validem les dades amb el test de Kolmogorov-Smirnov

```
mu <- 1
al <- 1

# Simulació d'un vector de dades aleatòries
x <- rgpl(mu=mu, alpha=al, n=10000, seed=1)

# Estadístic de contrast
X <- sort(x)
n <- length(X)
F <- 1-(X/mu)^(-al)
E <- seq(1:n)/n
D <- ks.test(x=F, y=E)$statistic[[1]]
est.con <- sqrt(n)*D

print(sprintf("est.con=%.3f, D=%.3f", est.con, D))
```

```
## [1] "est.con=0.900, D=0.009"
```

Veiem que per qualsevol dels valors d'α el resultat de l'estadístic de Kolmogorov-Smirnov, D, és el mateix (és a dir: 0.009).

Simulem dades d'una distribució power-law amb la funció de l'apartat anterior i verifiquem que la funció dóna l'estimació correcta, utilitzant el test de Kolmogorov-Smirnov.

$$D_n = \sup_{x_m < x < \infty} |F_n(x) - F_\alpha(x)|$$

En primer lloc, generem valors d'una distribució power-law amb paràmetres determinats α i μ. Després, calculem l'estadístic de contrast D. Aleshores, si X és un vector aleatori de dades $\{x_1, x_2, \ldots, x_n\}$ i $X \sim PL(\alpha, \mu)$, el test d'hipòtesis que avaluarem és el següent:

$$\begin{cases} H_0 & = X \sim PL(\alpha, \mu) \\ H_1 & = H_0 \text{ falsa} \end{cases}$$

Rebutjarem l'hipòtesi nul · la si $\sqrt{n} D_n > k_\alpha$.

```
mu <- 1
al <- 3

# Simulació d'un vector de dades aleatòries
x <- rgpl(mu=mu, alpha=al, n=10000, seed=1)

# Estadístic de contrast
X <- sort(x)
n <- length(X)
F <- 1-(X/mu)^(-al)
E <- seq(1:n)/n
D <- max(abs(E-F))
est.con <- sqrt(n)*D
print(sprintf("est.con=%.8f, D=%.8f", est.con, D))
```

```
## [1] "est.con=0.89300049, D=0.00893000"
```

α	1
Punt crític	1.22
$\sqrt{n}D_n$	0.8930005

No tenim evidències per rebutjar la hipòtesi nul · la quan $\alpha = 0.01, 0.05$ o 0.10 perquè cap dels estadístics de contrast obtinguts superen el llindar k_α. Per tant, les dades generades segueixen una distribució power-law amb parametres $\mu = 1$, $\alpha = 1$.