PopularnośćRozkładPotęgowy

Szymon Talaga

09.12.2014

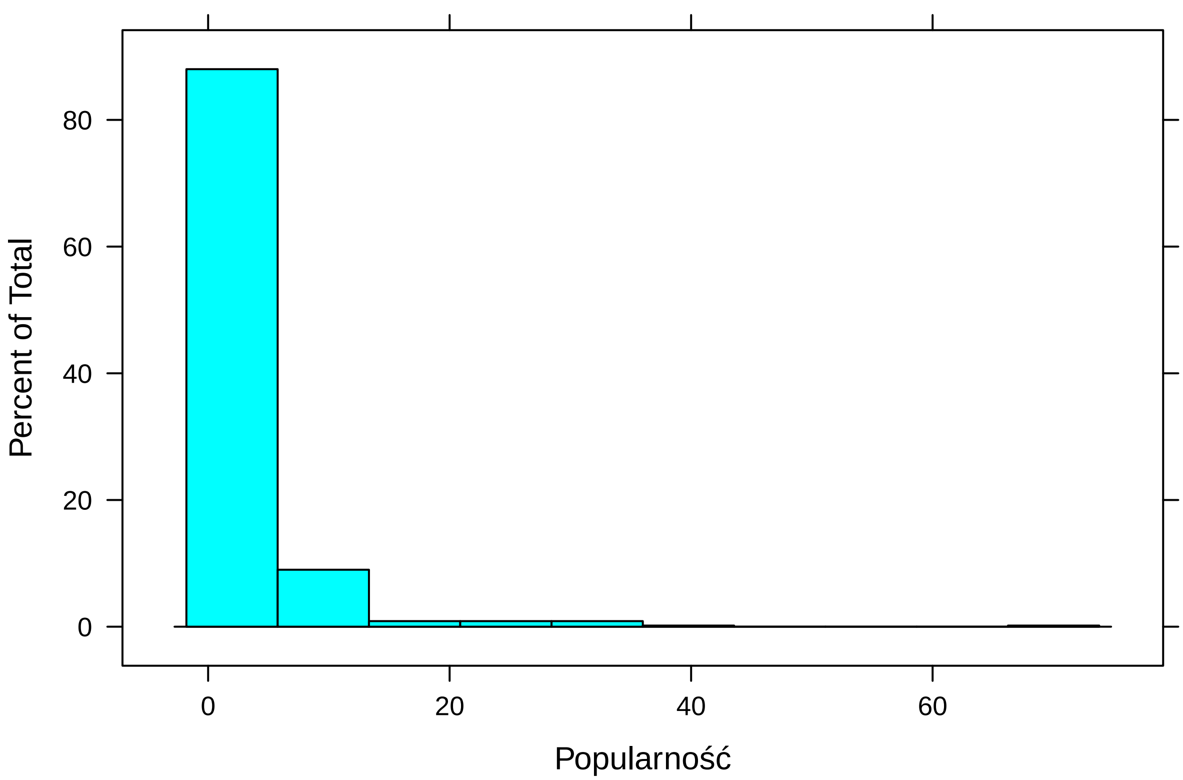
Popularność miejsc - rozkład potęgowy

Raport ten zawiera weryfikacje hipotezy potęgowego rozkładu popularności miejsc. Współczynniki rozkładu zostaną estymowane przy użyciu metody Największej Wiarygodności, jako że metoda Najmniejszych Kwadratów jest nieadekwatna w przypadku rozkładów potęgowych (Fronaczk i Fronczak, 2009). Ostateczna weryfikacja hipotezy zostanie oparta o test Kołmogorova-Smirnova.

Wczytanie niezbędnych pakietów i funkcji oraz danych:

```
library(igraph)
library(lattice)
library(latticeExtra)
source("../HelperFunctionsMisc/ComputingMisc.R")
source("../Networks/NetworkMethods.R")
load("../Places/PlaceData.RData") # Dane na poziomie miejsc
```

Rozkład popularności - jak widać wydaje się być zbliżony do potęgowego:



Estymacja (metodą Największej Wiarygodności) wykładnika α rozkładu potęgowego oraz jego błędu standardowego (wzór na błąd standardowy za: Fronczak i Fronczak, 2009)

```r
n = dim(Pdat)[1]
fit = power.law.fit(Pdat$popularity, xmin=1, implementation="plfit")
alpha.mle = fit$alpha
alpha.mle.se = (alpha.mle - 1) / sqrt(n)
```

Wykładnik: $\alpha = 2.0043986$ Błąd standardowy wykładnika: $\sigma_\alpha = 0.0421808$

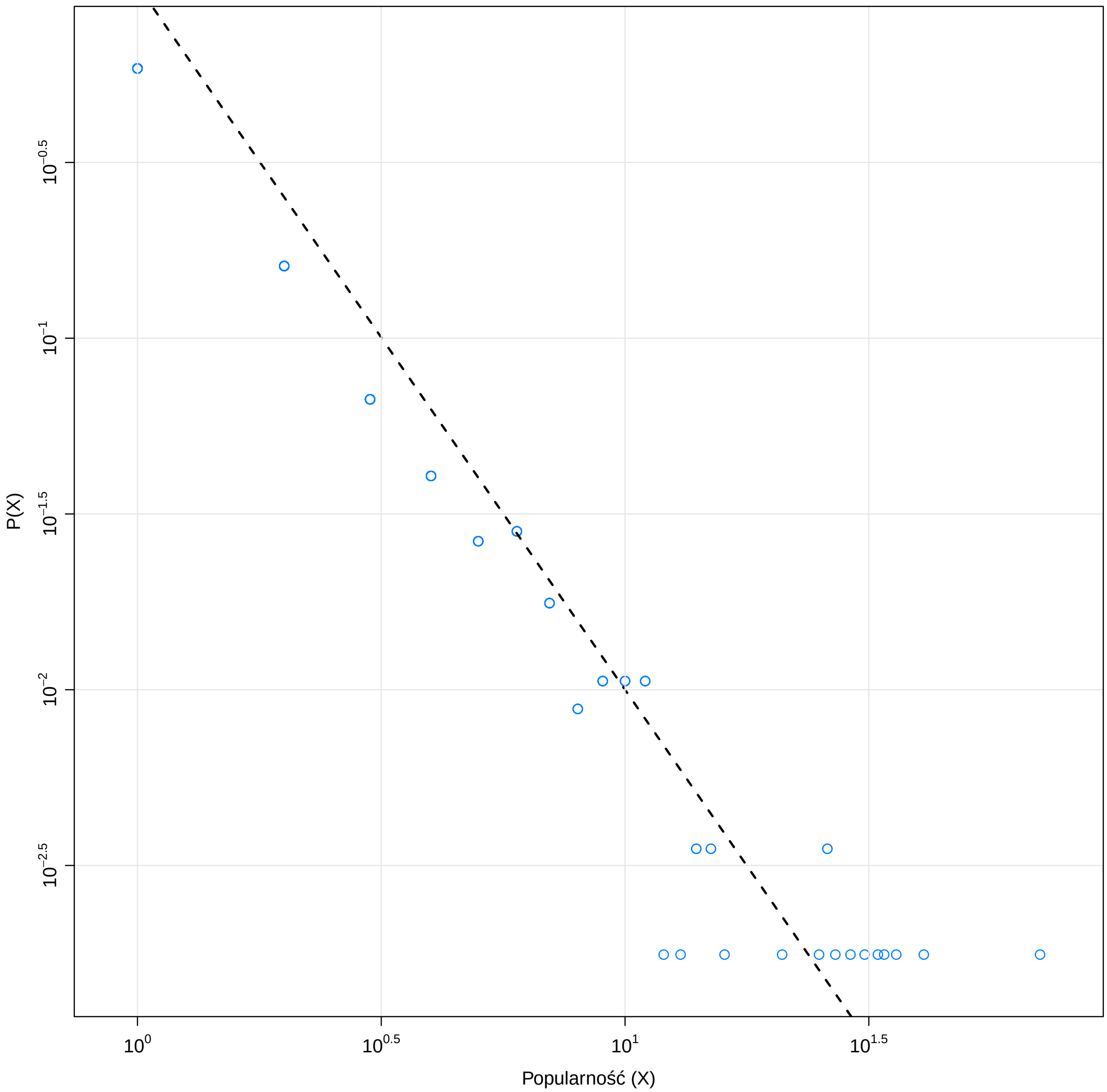
Widać, że estymacja jest obciążona bardzo małym błędem.

Na tej podstawie można obliczyć wartość stałej normującej C dla funckji gęstości rozkładu. Wyprowadzenie wzoru opiera się na prostej obserwacji, że całka funkcji gęstości prawdopodobieństwa po całej dziedzinie powinna równać się 1. Co w przypadku, gdy minialna popularność wynosi 1, daje wzór postaci: $C = \alpha - 1$. Wówczas C wynosi 1.0043986.

```r
C = alpha.mle - 1
```

Na podstawie powyższej analizy można przeprowadzić wstępną graficzną weryfikację hipotezy w oparciu o wykres zależności częstości od popularności na skali podwójnie logarytmicznej (podstawa dziesiętna). Jeżeli rozkład popularności jest naprawdę potęgowy to równanie $lnP(popularno) = -\alpha ln(popularno) + C$ powinno dać dobre liniowe dopasowanie względem korelogramu.

Wykres:



Wykres zdaje się świadczyć na korzyść hipotezy. Ostatecznym sprawdzianem będzie test Kołmogorova-Smirnova dopasowania danych empirycznych do teoretycznego rozkładu:

```
p = fit$KS.p
D = fit$KS.stat
```

Przetestowane zostały następujące hipotezy:

H_0 : dane pochodzą z rozkładu potęgowego o wyznaczniku $\alpha = 2.0043986$

H_1 : dane pochodzą z jakiegoś innego rozkładu

Test przyniósł wynik negatywny: $D = 0.024$; $p = 0.902$. Oznacza to, że nie udało się odrzucić hipotezy zerowej, co oznacza, że rozkład popularności miejsc jest potęgowy. Ma on zatem własności bezskalowe - np. samopodobieństwo. Najważniejszą tego konsekwencją jest fakt, że ma on nieokreśloną wartość charakterystyczną, czyli nie można mówić o "przeciętnie" popularnym miejscu. Innymi słowy, pomimo że większość miejsc charakteryzuje się małą popularnością, to obok nich mogą się pojawiać super popularne miejsca (Fronczak i Fronczak, 2009)

Koniec raportu