CMU SCS

15-826: Multimedia Databases and Data Mining

Lecture #13: Power laws - equivalences
Addendum
C. Faloutsos

CMU SCS

Jumping to the conclusion:

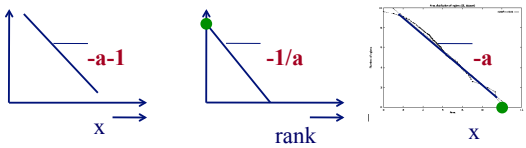
CMU SCS

3 versions of P.L.

PDF = frequency-count plot | Zipf plot = Rank-frequency | NCDF = CCDF

IF ONE PLOT IS P.L., SO ARE THE OTHER TWO

Prob(area = x) | area | Prob(area >= x)



CMU SCS

Details, and proof sketches:

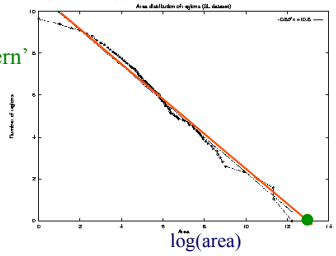
CMU SCS

Reminder

More power laws: areas – Korcak's law

log(count(>= area))

'Vaenern'



log(area)

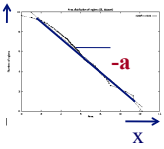
Scandinavian lakes area vs complementary cumulative count (log-log axes)

CMU SCS

3 versions of P.L.

NCDF = CCDF

Prob(area >= x)

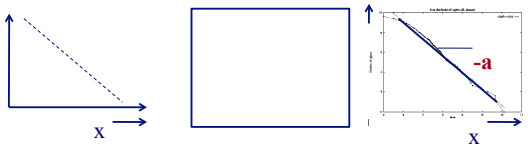


CMU SCS

3 versions of P.L.

PDF NCDF = CCDF

Prob(area = x) Prob(area >= x)

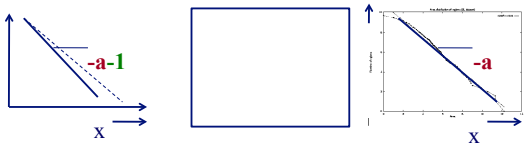


CMU SCS

3 versions of P.L.

PDF NCDF = CCDF

Prob(area = x) Prob(area >= x)

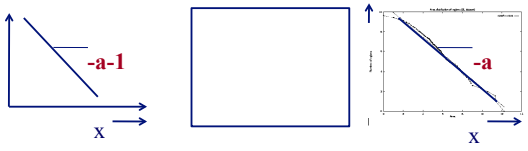


CMU SCS

3 versions of P.L.

PDF NCDF = CCDF

Prob(area = x) Prob(area >= x)

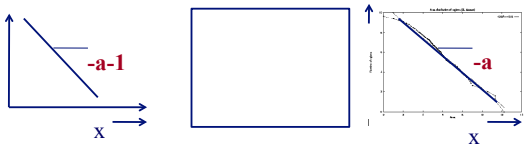


CMU SCS

3 versions of P.L.

PDF Zipf plot = Rank-frequency NCDF = CCDF

Prob(area = x) Prob(area >= x)

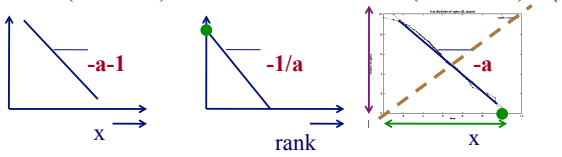


CMU SCS

3 versions of P.L.

PDF Zipf plot = Rank-frequency NCDF = CCDF

Prob(area = x) area Prob(area >= x)

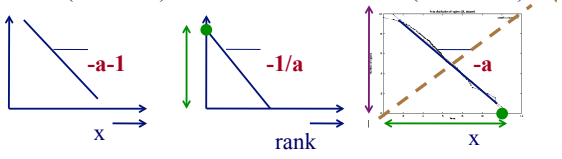


CMU SCS

3 versions of P.L.

PDF Zipf plot = Rank-frequency NCDF = CCDF

Prob(area = x) area Prob(area >= x)

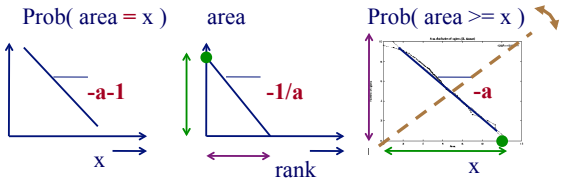


CMU SCS

3 versions of P.L.

PDF | Zipf plot = Rank-frequency | NCDF = CCDF

Prob(area = x) | area | Prob(area >= x)

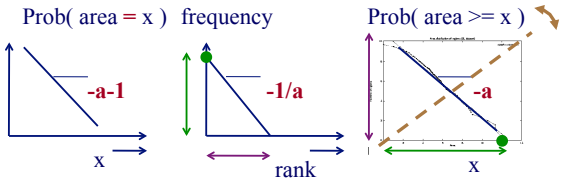


CMU SCS

3 versions of P.L.

PDF | Zipf plot = Rank-frequency | NCDF = CCDF

Prob(area = x) | frequency | Prob(area >= x)

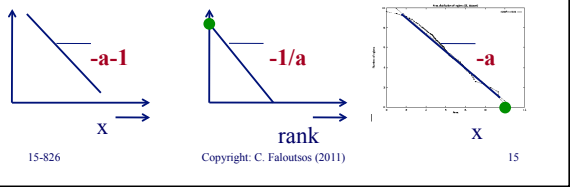


CMU SCS

3 versions of P.L.

PDF = frequency-count plot | Zipf plot = Rank-frequency | NCDF = CCDF

Prob(area = x) | area | Prob(area >= x)

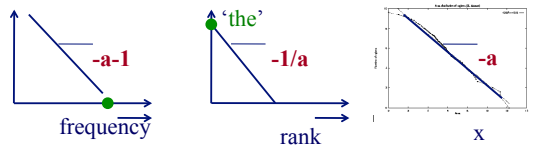


CMU SCS

3 versions of P.L.

PDF = frequency-count plot | Zipf plot = Rank-frequency | NCDF = CCDF

count | frequency | Prob(area >= x)



CMU SCS

Sanity check:

- Zipf showed that if
  - Slope of rank-frequency is -1
  - Then slope of freq-count is -2
- Check it!

CMU SCS

3 versions of P.L.

PDF = frequency-count plot | Zipf plot = Rank-frequency | NCDF = CCDF

slope = -2 $\Leftrightarrow$ slope = -1

count | frequency | Prob(area >= x)

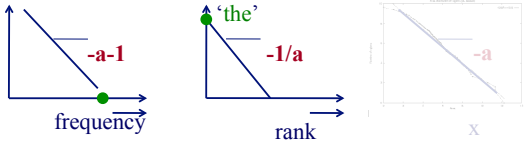
CMU SCS

3 versions of P.L.

PDF = frequency-count plot | Zipf plot = Rank-frequency | NCDF = CCDF

✓ slope = -2 ⟺ slope = -1

count | frequency | Prob(area >= x)



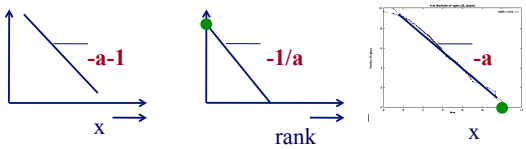
CMU SCS

3 versions of P.L.

PDF = frequency-count plot | Zipf plot = Rank-frequency | NCDF = CCDF

IF ONE PLOT IS P.L., SO ARE THE OTHER TWO

Prob(area = x) | area | Prob(area >= x)



CMU SCS

References

- *Zipf, Power-laws, and Pareto - a ranking tutorial*, Lada A. Adamic www.hpl.hp.com/research/idl/papers/ranking/ranking.html
- L.A. Adamic and B.A. Huberman, *'Zipf's law and the Internet'*, *Glottometrics* 3, 2002,143-150
- *Human Behavior and Principle of Least Effort*, G.K. Zipf, Addison Wesley (1949)