Master Program in *Data Science and Business Informatics*

Statistics for Data Science

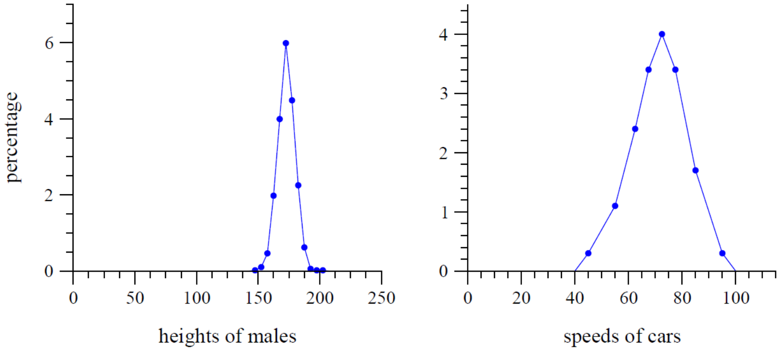
Lesson 13 - Power laws and Zipf's law

Salvatore Ruggieri

Department of Computer Science
University of Pisa, Italy
salvatore.ruggieri@unipi.it

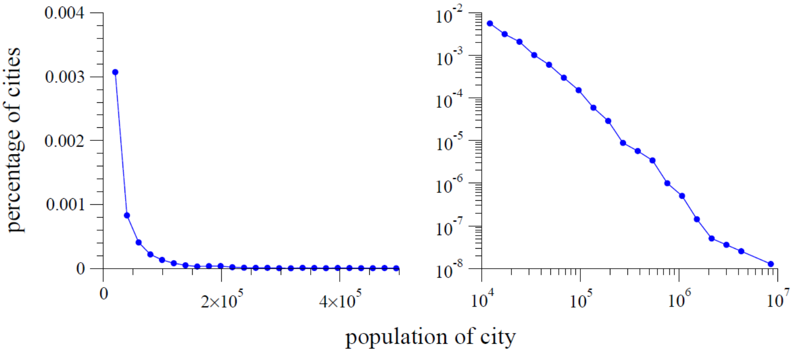
Scaled distributions

- Many of the things that scientists measure have a typical size or “scale” — a typical value around which individual measurements are centered



Scale-free distributions

- But not all things we measure are peaked around a typical value. Some vary over an enormous dynamic range.



Look at Figure 4 of [Newman2005]

Continuous power-law

Power-law

A continuous random variable X has the *power-law distribution*, if for some $\alpha > 1$ its density function is given by$^{(\star)}$

$$p(x) = C \cdot x^{-\alpha} \quad \text{for } x \geq x_{min}$$

We denote this distribution by $Pow(x_{min}, \alpha)$.

- C is called the **intercept**, and α the **exponent**.
- Passing to the logs:

$$\log p(x) = -\alpha \cdot \log(x) + \log C$$

linearity in log-log scale plots!

See R script

$(^\star)$ We use $p(x)$ for the density function (instead of $f(x)$) to be consistent with [Newman2005].

Scale-free distributions

$$p(bx) = g(b)p(x)$$

- Measuring in cm, inches, Km, or miles does not change the form of the distribution (up to some constant)!
- For a power-law $p(x) = Cx^{-\alpha}$

  $$p(bx) = b^{-\alpha}Cx^{-\alpha}$$

  hence, $g(b) = b^{-\alpha}$
- Actually, power-laws are the only scale-free distributions!
  - see Eq. 30-34 of [Newman2005] for a proof

Intercept

- What is the constant C?

$$1 = \int_{x_{min}}^{\infty} C \cdot x^{-\alpha} dx = \frac{C}{-\alpha + 1} \left[x^{-\alpha+1} \right]_{x_{min}}^{\infty} \overset{(\star)}{=} \frac{C}{-\alpha + 1} \left(0 - x_{min}^{-\alpha+1} \right) = \frac{C}{\alpha - 1} x_{min}^{-\alpha+1}$$

($\star$) finite only for $\alpha > 1$, because:
  - for $\alpha < 1$: $\lim_{x \to \infty} x^{-\alpha+1} = \infty$
  - for $\alpha = 1$: denominator $-\alpha + 1$ is 0

- Therefore:

$$C = (\alpha - 1)/x_{min}^{-\alpha+1} \tag{1}$$

and, in summary:

$$p(x) = \frac{(\alpha - 1)}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha}$$

CCDF

- Let's compute:

$$P(X > x) = \int_x^\infty p(y)dy = C\int_x^\infty y^{-\alpha}dy = \frac{C}{-\alpha+1}\left[y^{-\alpha+1}\right]_x^\infty = \frac{C}{\alpha-1}x^{-\alpha+1}$$

and since $C = (\alpha-1)/x_{min}^{-\alpha+1}$:

$$P(X > x) = \left(\frac{x}{x_{min}}\right)^{-\alpha+1} = \left(\frac{x}{x_{min}}\right)^{-(\alpha-1)}$$

- Same form as df (see Eq. 1) but with exponent $(\alpha - 1)$

See R script

Pareto distribution

- **Vilfredo Pareto** noticed that the number of people whose income exceeded level x (i.e., CCDF) is well approximated by C/x^{β} for some constants C and $\beta > 0$
  - It appears that for all countries $\beta \approx 1.5$.

Pareto distribution

A continuous random variable X has the *Pareto distribution*, if for some $\beta > 0$ its density function is given by

$$p(x) = C \cdot x^{-(\beta+1)} \quad \text{for } x \geq x_{min}$$

We denote this distribution by $Par(x_{min}, \beta)$.

- $Par(x_{min}, \beta) = Pow(x_{min}, \beta + 1)$ or $Pow(x_{min}, \alpha) = Par(x_{min}, \alpha - 1)$
- CCDF of $Par(x_{min}, \beta)$ is $(\frac{x}{x_{min}})^{-((\beta+1)-1)} = (\frac{x}{x_{min}})^{-\beta}$

See R script

Expectation and variance of a power-law

- What is the expectation of $X \sim Pow(x_{min}, \alpha)$?

$$E[X] = \int_{x_{min}}^{\infty} x \cdot p(x) dx = C \int_{x_{min}}^{\infty} x^{-\alpha+1} dx = \frac{C}{-\alpha+2} \left[x^{-\alpha+2}\right]_{x_{min}}^{\infty} \overset{(\star)}{=} \frac{C}{\alpha-2} x_{min}^{-\alpha+2}$$

$(\star)$ Finite only for $\alpha > 2$, because:

$$\lim_{x \to \infty} x^{-\alpha+2} = \infty \text{ for } \alpha \leq 2$$

and since $C = (\alpha - 1)/x_{min}^{-\alpha+1}$:

$$E[X] = \frac{\alpha - 1}{\alpha - 2} x_{min}$$

  - For $1 < \alpha \leq 2$, there is no expectation: the mean of a dataset has no reliable value!

- $Var(X)$ finite only for $\alpha > 3$
  - For $2 < \alpha \leq 3$, there is no variance: the empirical variance of a dataset has no reliable value!

Discrete power-law

Discrete power-law

A discrete random variable X has the *power-law distribution*, if for some $\alpha > 1$ its p.m.f. function is given by

$$p(k) = C \cdot k^{-\alpha} \quad \text{for } k = k_{min}, k_{min}+1, \dots$$

We denote this distribution by $Pow(k_{min}, \alpha)$.

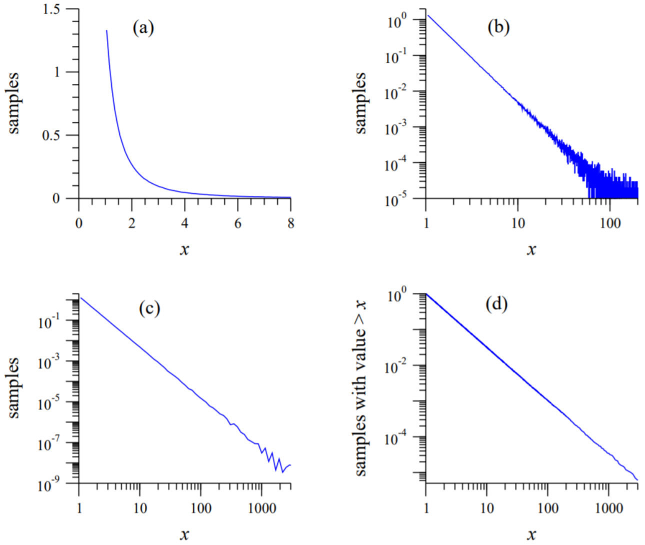
- Population of cities, number of books sold, number of citations, etc.
- Since $1 = \sum_{k=k_{min}}^{\infty} C \cdot k^{-\alpha}$, we have

$$C = \frac{1}{\sum_{k=k_{min}}^{\infty} k^{-\alpha}} = \frac{1}{\zeta(\alpha, k_{min})}$$

  where $\zeta(\alpha, k_{min}) = \sum_{k=k_{min}}^{\infty} k^{-\alpha}$ *[Hurwitz zeta-function]*
- Special case: $\zeta(\alpha) = \zeta(\alpha, 1) = \sum_{k=1}^{\infty} k^{-\alpha}$ *[Riemann zeta-function]*

See R script

Logarithmic binning vs CCDF



See R script

Zipf's law

Zipf's law distribution

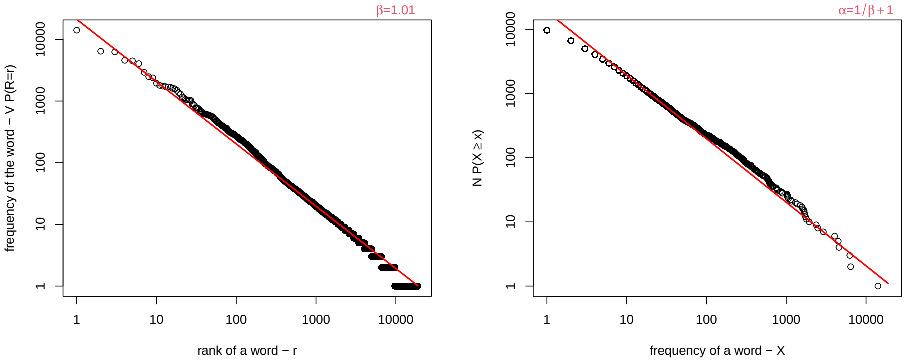
A discrete random variable R has the *Zipf's law distribution*, if for some $\alpha > 1$ its p.m.f. function is given by

$$p(r) = C \cdot r^{-\alpha} \quad \text{for } r = 1, 2, \ldots, N$$

We denote this distribution by $Zipf(\alpha)$.

- Since $\sum_{r=1}^{N} C \cdot r^{-\alpha} = 1$, we have:
$$C = \frac{1}{\sum_{r=1}^{N} r^{-\alpha}} = \frac{1}{\zeta(\alpha) - \zeta(\alpha, N+1)}$$
- Read $p(r)$ as the probability of an event based its rank $r = 1, \ldots, N$ (finitely many!)
- E.g., prob. of occurrence of a word in a book given the word rank, prob. of occurrence of an inhabitant of a city given the city rank
  - *Contrast to discrete power laws*: prob. of words with a given number of occurrences, prob. of cities with a given number of inhabitants

Zipf's law



Left: frequency of words based on rank *[Zipf's law]*

Right: number of words with a given minimum frequency *[CCDF of a Power-law]*

Main point: a word with frequency f has rank r iff there are r words with frequency $\geq f$

From power-law to Zipf's law and vice-versa

- $\Omega = \{\omega_1, \dots, \omega_N\}$, ω_i is a city with n_i inhabitants, for a total of N cities and $V = \sum_{i=1}^{N} n_i$ inhab.
  - e.g., $X(\omega_{Tokyo}) = 37M$ for the city of **Tokyo** (world's most populated city)
- $X(\omega_i) = n_i$ is the population of the city ω_i
  - $P(X = k) = |\{\omega \in \Omega \mid X(\omega) = k\}|$, fraction of cities with population $= k$
  - Assume $X \sim Pow(1, \alpha)$ for some α
  - $P(X > k) \propto k^{-(\alpha-1)}$, fraction of cities with more than k inhabitants

    [$\propto$ reads "proportional to" up to multip./additive constants]
  - $N \cdot P(X > k) \propto N \cdot k^{-(\alpha-1)}$, numbr of cities with more than k inhabitants
- $R(\omega_i)$ = rank of the city ω_i w.r.t. city population
  - $P(R = r) = X(\omega_{r\text{-}th})/V$, fraction of world's population in the r^{th} largest city
  - e.g., $R(\omega_{Tokyo}) = 1$ for the city of Tokyo, and $P(R = 1) = 37M/8B \approx 0.46\%$
  - $R(\omega_i)$ = numbr of cities with $X(\omega_i)$ or more inhab. $\propto N \cdot X(\omega_i)^{-(\alpha-1)} + 1 \propto X(\omega_i)^{-(\alpha-1)}$
- In summary $R(\omega) \propto X(\omega)^{-(\alpha-1)}$, or, by inverting, $X(\omega) \propto R(\omega)^{-\frac{1}{\alpha-1}}$, and then:

$$P(R = r) = \frac{X(\omega_{r\text{-}th})}{V} \propto X(\omega_{r\text{-}th}) \propto r^{-\beta} \quad \text{where } \beta = \frac{1}{\alpha - 1}$$

  i.e., $R \sim Zipf(\beta)$ – the r^{th} most populated city has population proportional to $r^{-\beta}$

See R script

Mandatory reference

Sections I, II, III(A,B,E,F) of the following paper are mandatory teaching material for this lesson.

M. E. J. Newman (2005)
Power laws, Pareto distributions and Zipf's law
Contemporary Physics 46 (5), 323–351.