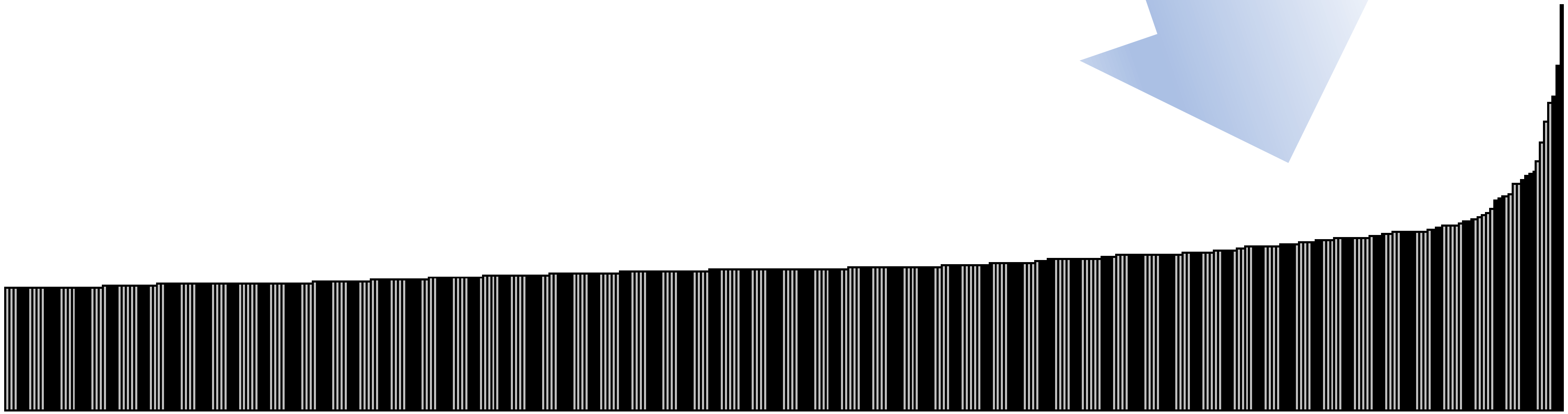
Power Laws and Popularity

Richard Vale

Weave on Boundary in partnership with Xccelerate, Hong Kong,
November 11, 2019

If we took the 500 tallest buildings in Hong Kong and lined them up, they would look like this



A few very tall buildings

A long tail of shorter buildings

Magic Formula

What if I told you that the heights of buildings in Hong Kong are described by a mysterious **magic formula**, and the **magic formula** is

$$f(x) = \frac{1{,}313{,}560{,}000{,}000}{x^{5.65}}$$

Would you be **amazed**?

Magic formula:

$$f(x) = \frac{1{,}313{,}560{,}000{,}000}{x^{5.65}}$$

What does the magic formula do?

If we want to know how many of the 500 buildings are taller than 200m, the magic formula gives

$\frac{1313560000000}{200^{5.65}} = 0.131$

The actual number of buildings which are taller than 200m is 65, which is 13% of 500, or 0.13.

This works for any other height in place of 200

$$\frac{1313560000000}{250^{5.65}} = 0.037 = 3.7\%$$

The actual number of buildings which are taller than 250m is 18, which is 3.6% of buildings

$$\frac{1313560000000}{300^{5.65}} = 0.013 = 1.3\%$$

Actually, 1.2% of buildings are greater than 300m

$$\frac{1313560000000}{170^{5.65}} = 0.33 = 33\%$$

Actually, 36% of buildings are greater than 170m

If you put a number x into the **magic formula**, it tells you (fairly accurately) what percentage of buildings in Hong Kong are taller than x metres.

It is surprising that the magic formula exists. The heights of skyscrapers depend on what and where people decide to build, what building laws apply, how the economy is doing… how is it possible that they can be described by a simple pattern if nobody decided that they should?

In this talk, you will learn the secrets…

- **Why** is there a **magic formula**?
- How would I have known in advance that there **is** a **magic formula**?
- How can I **make my own** **magic formula**?

What if we looked at the height of people in Hong Kong instead of buildings?

Is there a **magic formula** for this?

If you lined everybody up, their heights would look a bit like this

There is a **magic formula** for heights of people as well!

The percentage of people taller than x cm is roughly

$$\int_{x}^{\infty} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \times 1/\sigma\sqrt{2\pi} \; dt$$

where μ and σ are certain numbers. I hope you agree that, while this other **magic formula** is impressive, it is not the same kind of nice simple **magic formula** as the one we had for the heights of buildings.

Power Laws

A quantity follows a **power law** if the proportion of items (people; buildings; etc.) which are greater than x is given by a formula like

$$\frac{C}{x^a}$$

where *C* and *a* could be any positive numbers.

(Technical note: usually a quantity is also said to follow a power law if the law only applies for large values of x. I don't want to get into this complication here but will mention it again later).

Power Laws

Power laws tend to have

- A “long tail” of very small values
- A few very large values (called “outliers” in statistics)

"...consider a world where the heights of Americans were distributed as a power law. In this case, we would expect nearly 60,000 individuals to be as tall as the tallest adult male on record, at 2.72 meters. Further, we would expect ridiculous facts such as 10,000 individuals being as tall as an adult male giraffe, one individual as tall as the Empire State Building (381 meters), and 180 million diminutive individuals standing a mere 17 cm tall.“

- Aaron Clauset (power law expert) in a CSCI lecture, 2011

Examples of power laws

- Sizes of towns (many villages; a few huge cities)
- Number of species in families of biological organisms
- Incomes (Pareto’s Law)
- Number of friends on a social network
- Frequency of word use in text (Zipf’s Law)
- Number of citations of scientific papers
- Number of casualties in wars
- Earthquake sizes
- Upvotes in Reddit threads
- Popularity of books

But caution! Not everything with a long tail + outliers is always a power law

How can you recognize a power law?

The secret to making your own **magic formula** is to take logs of the power law equation

$$\boldsymbol{P(X \geq x) = \frac{C}{x^a}}$$

becomes

$$\boldsymbol{log\, P(X \geq x) = log(C) - a\, log(x)}$$

[Remember, log(a number) is defined by the rule: “10 to the power of log(number) gives you the number back again”. For example, log(10) = 1, log(100) = 2, etc. It’s also roughly the number of digits minus one.]

For the buildings, we see that the plot of log($P(X \geq x)$) versus log(x) looks like a straight line, which suggests that a power law is a good fit.

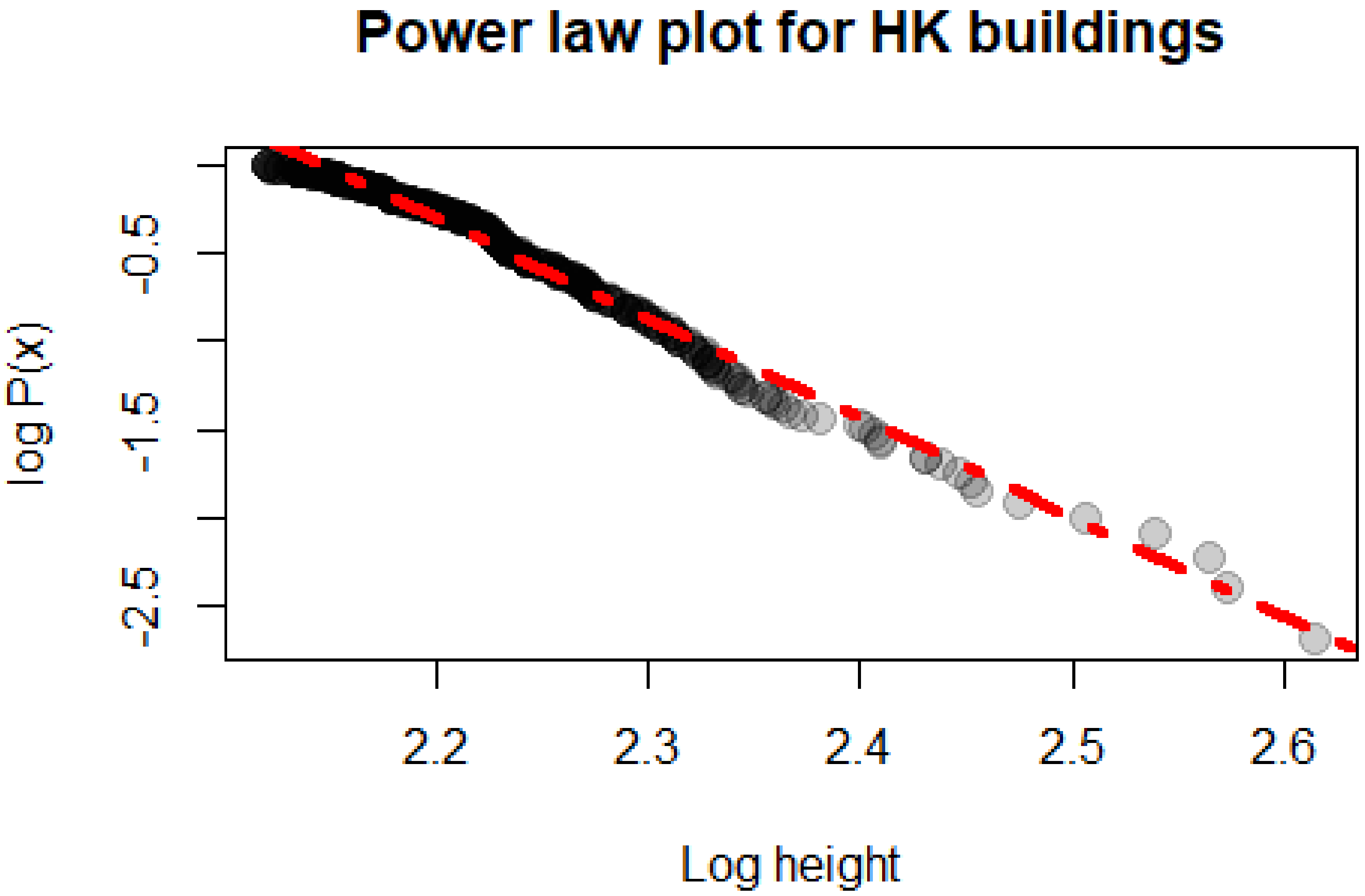
We can find the dotted red line using linear regression.

This gives the values

log(C) = 12.118
a = 5.65

That's where the **magic formula** came from!

[Footnote: technically, this technique is not the best way of getting the magic formula. See Clauset et al, *Power Law Distributions in Empirical Data, SIAM Review, vol.51 no.4, 2009* for the details.]



Answers to Questions

- **Why** is there a **magic formula**?

Answer: because the plot of log $\boldsymbol{P(X \geq x)}$ versus log $(\boldsymbol{x})$ is a straight line.

- How can I **make** my own **magic formula**?

Answer: fit a straight line to the log-log plot and read off the numbers.

One question remains:

- *How would I have known in advance that there* ***is*** *a* ***magic formula****?*

Where do power laws come from?

We are left with the question of why power laws are so common. How could we know in advance that some quantity is likely to follow a power law, before we have even looked at the data?

One way in which power laws arise is as the result of a **Yule process**, named after G. U. Yule. Also known as a **rich-get-richer** process.

G. Yule, 1871-1951, from Wikipedia

Yule Process

Suppose we start with 10 people in a row.

웃 웃 웃 웃 웃 웃 웃 웃 웃 웃

A new person arrives and randomly makes friends with someone

웃 웃 웃 웃 웃웃 웃 웃 웃 웃 웃

Now suppose that each subsequent person chooses to join an existing friendship group, **based on how large it is**. This means that the next person is twice as likely to join the group with two people as one of the others.

As more people join, some friendship groups become large, whereas others stay small.

웃 웃 웃 웃 웃웃 웃웃 웃웃 웃 웃
웃웃웃

The larger groups are more likely to grow even larger, and the smallest groups are likely to stay small.

Group sizes after 100 new people have joined:

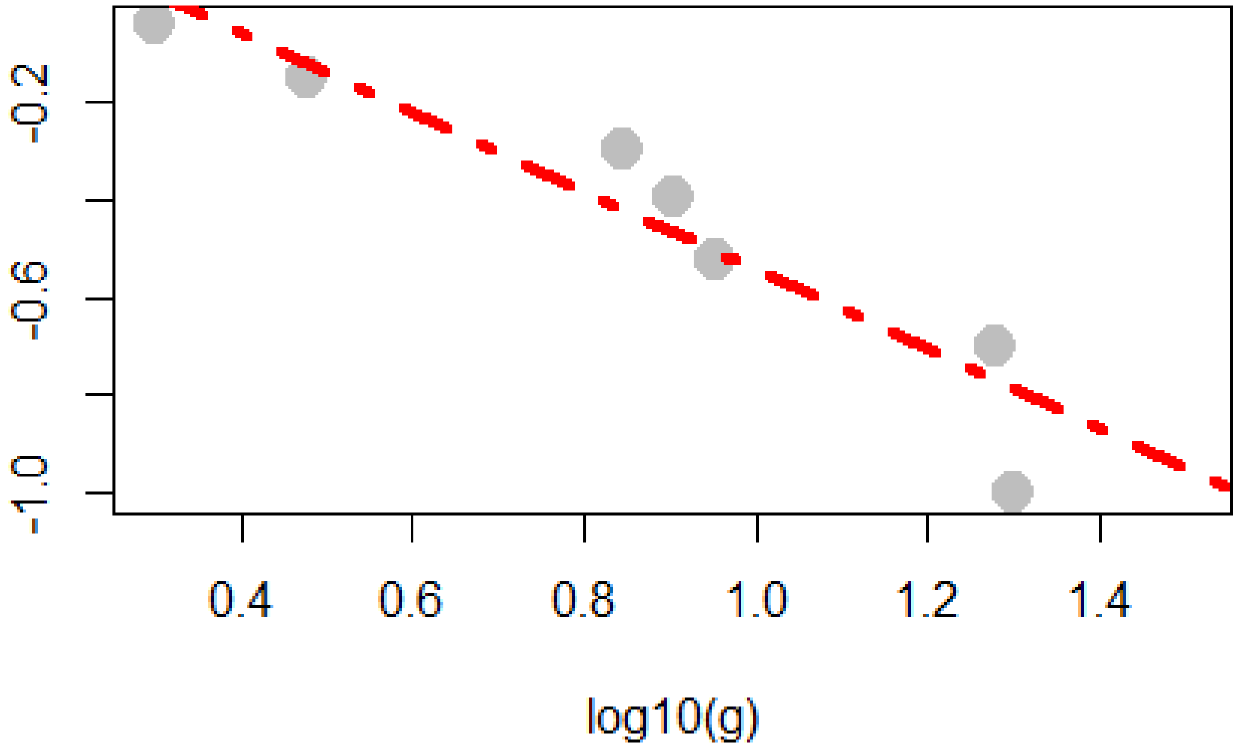
20 19 7 7 9 2 3 3 8 32

A power law plot for the numbers 20 19 7 7 9 2 3 3 8 32 looks like a straight line.

This always happens; in the long run, the group sizes you get from a Yule process will satisfy a power law.

In this case the power law is

$$P(X \geq x) = \frac{1.85}{x^{0.81}}$$



You can use this fact to predict that something will follow a power law. If it's likely to follow a rich-get-richer process, or if things grow in a way that is proportional to their size, then those sizes will follow a power law! **Pop**ulation and **Pop**ularity are **Po**wer Laws.

Examples of power laws, explained

- Sizes of towns (many villages; a few huge cities)

Explanation : people tend to join the largest towns to look for work, just like in the rich-get-richer process

- Incomes (Pareto’s Law)

Explanation : the rich get richer, literally

- Number of friends on a social network

Explanation : people with a lot of friends are more likely to make new friends via mutual contacts (popularity of something generally follows a power law)

- Number of citations of scientific papers

Explanation : highly-cited papers are more famous and more likely to be cited

- Number of casualties in wars

Explanation : wars tend to escalate; the more people killed already, the more casualties there are likely to be.

Examples of power laws, not really explained

- Earthquake magnitudes

Maybe: the more energy has already been released, the more will be released in total? Some kind of resonance?

- Building heights

Maybe: as the population grows exponentially, the heights of buildings grow in proportion, since people need more space?

- Word frequencies

Maybe: the more a word is used, the more likely it is to be used next?

Rich-get-richer processes are not the only source of power laws!

You would definitely expect that fighter pilot victories would follow a power law, because the more planes a pilot has shot down, the more experienced they are, and the more capable of shooting down even more.

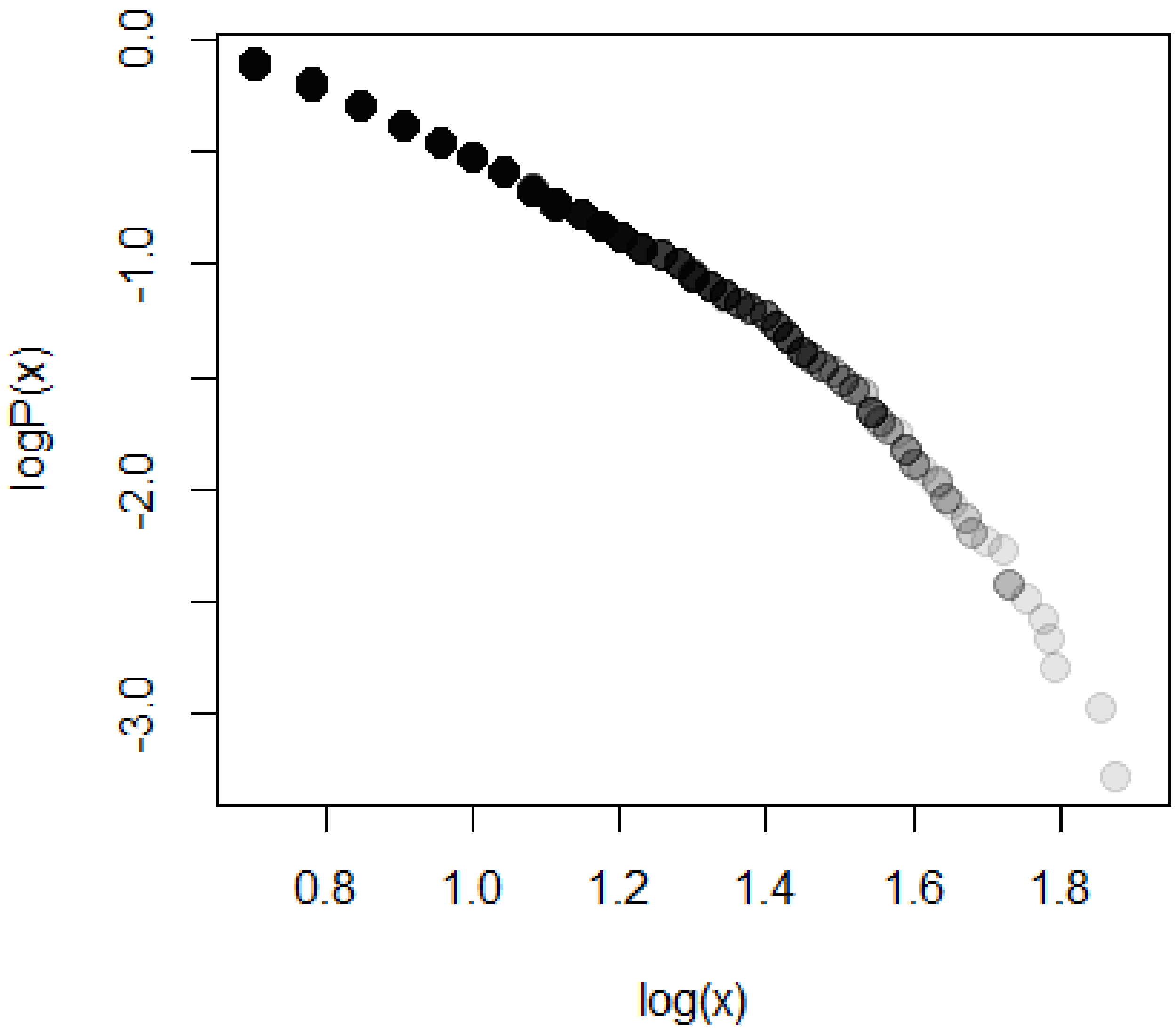
Example

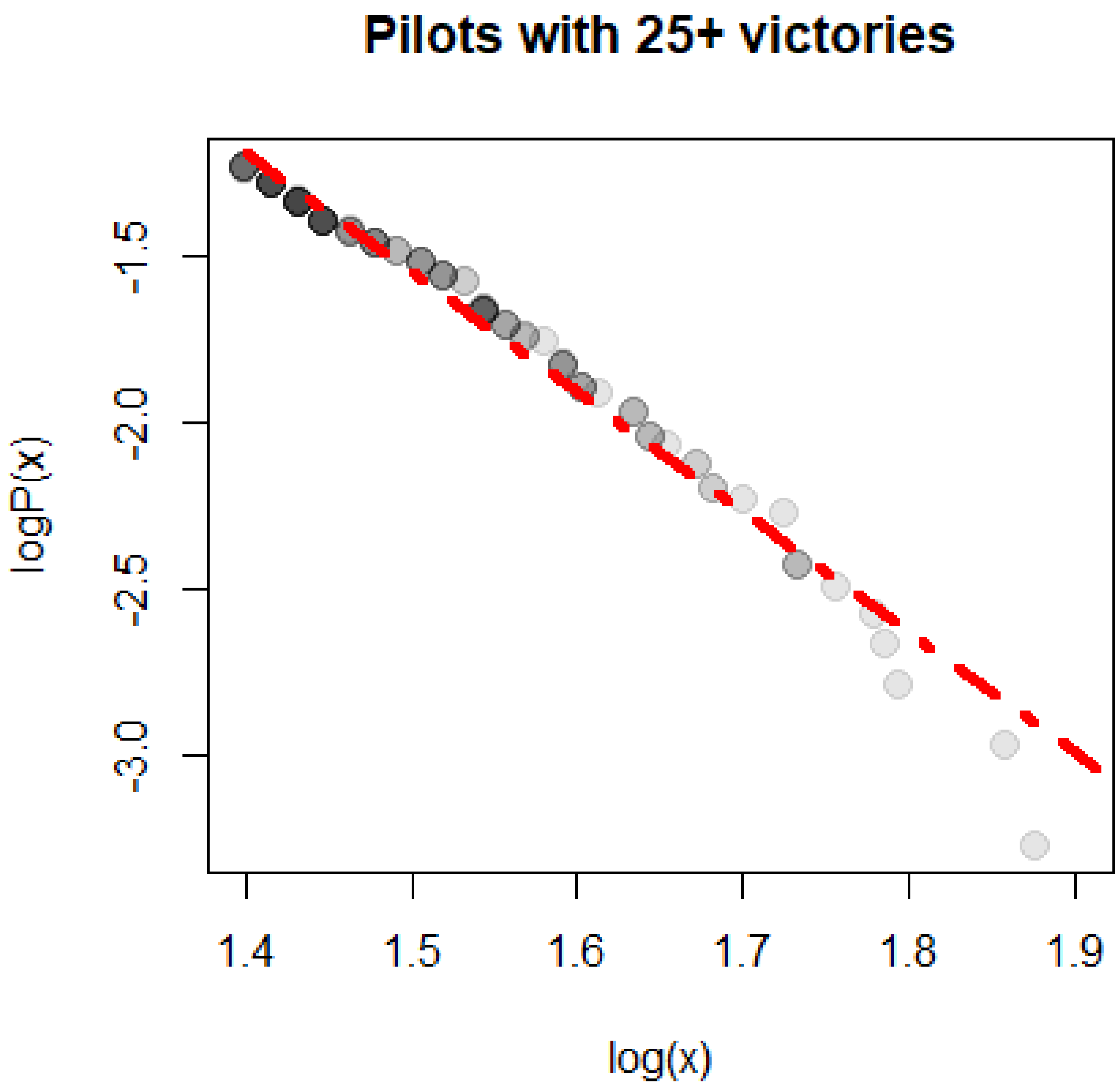
Analysis of all 1865 pilots (on all sides) who shot down 5 or more planes in the First World War.

Not straight!

Power law tail?

Multiple power laws at different scales?





Just like with buildings (we looked at the 500 tallest buildings) here we look at the 118 top pilots, and find that the distribution of victories for these pilots does look very much like a power law.

In this case, $a = 3.6$

Example 2 (interactive)

From the book Networks, Crowds, and Markets: Reasoning about a Highly Connected World. By David Easley and Jon Kleinberg. Cambridge University Press, 2010. Complete preprint on-line at http://www.cs.cornell.edu/home/kleinber/networks-book/

Suppose that some researchers studying educational institutions decide to collect data to address the following two questions.

(a) As a function of x, what fraction of Cornell classes have x students enrolled?

(b) As a function of x, what fraction of 3rd-grade elementary school classrooms in New York State have x pupils?

Which one of these would you expect to more closely follow a power-law distribution as a function of x? Give a brief explanation for your answer, using some of the ideas about power-law distributions developed in Chapter 18.

Things to ponder

If a rich-get-richer process is in operation, you only need to get a tiny bit lucky in the beginning in order to end up very rich indeed. Does this explain why some people/startups/sports teams/publishers/search engines/politicians… are highly successful, while many others are not?

Thank you for your attention!

多謝晒