The Abdus Salam International Centre for Theoretical Physics

United Nations Educational, Scientific and Cultural Organization

IAEA International Atomic Energy Agency

2060-8

Advanced School on Non-linear Dynamics and Earthquake Prediction

28 September - 10 October, 2009

On the Power Law Statistical Distribution of Observations

Antoni M. Correig

Universitat de Barcelona
Spain

ton.correig@am.ub.es

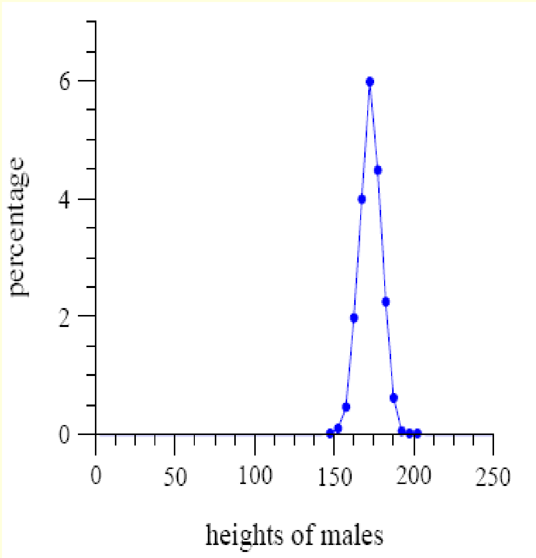
ON THE POWER LAW STATISTICAL DISTRIBUTION OF OBSERVATIONS

Antoni M. Correig (2009)

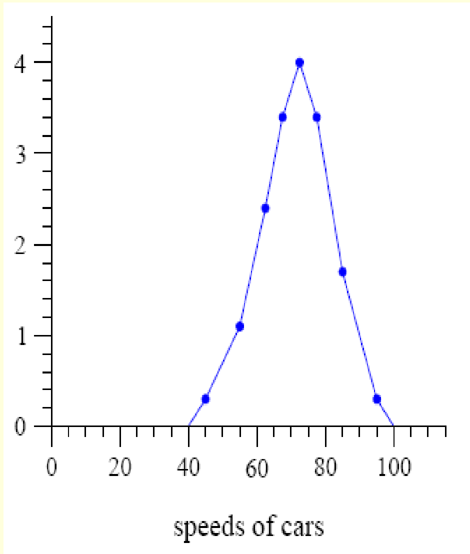
[based on Newman (2005)
and Claused *et al.* (2007)]

NORMAL DISTRIBUTION

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



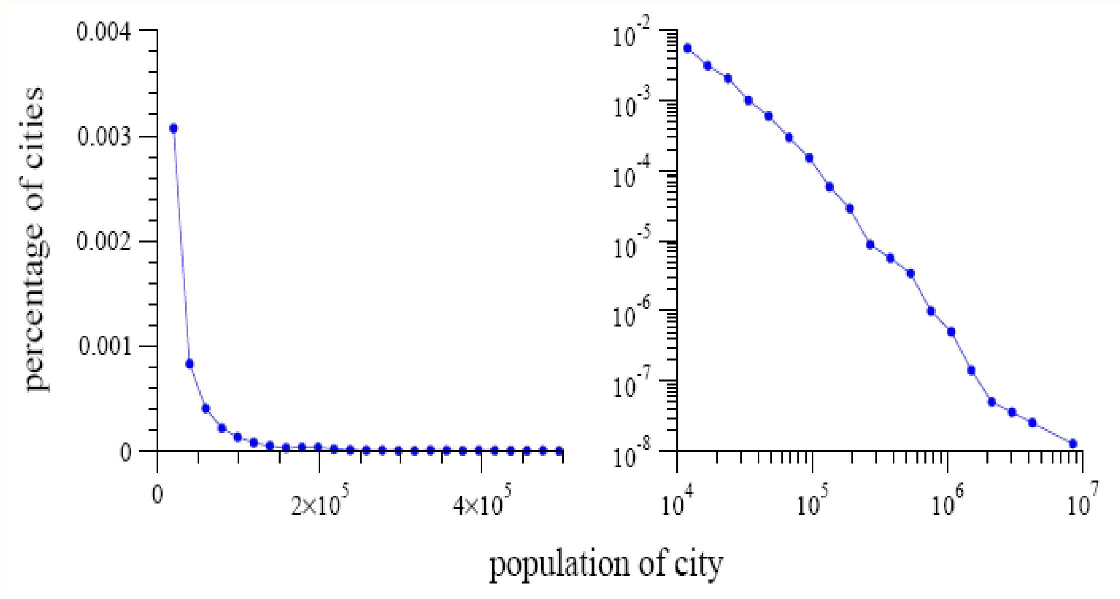
Histogram of heights in centimeters of American males.



Histogram of speeds in miles per hour of cars on UK motorways.

POWER LAW DISTRIBUTION

But not all things we measure are peaked around a typical value. Some vary over an enormous dynamic range, sometimes many orders of magnitude.



Left: histogram of the populations of all US cities with population of 10 000 or more. Right: another histogram of the same data, but plotted on logarithmic scales. The approximate straight-line form of the histogram in the right panel suggest that the distribution follows a power law.

Power law distribution

- Straight line on a log-log plot

$$\ln(p(x)) = c - \alpha \ln(x)$$

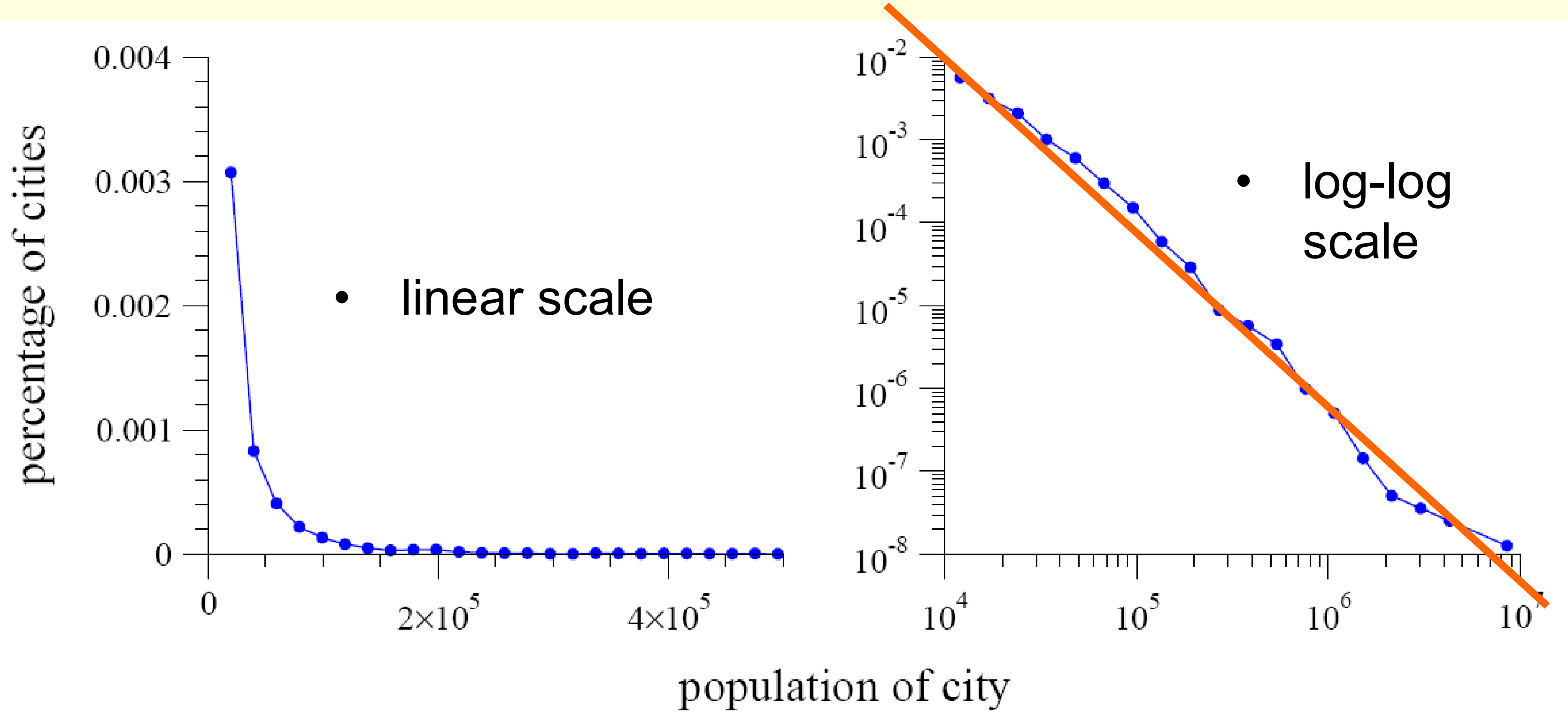
- Exponentiate both sides to get that *p(x)*, the probability of observing an item of size 'x' is given by

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

normalization constant (probabilities over all *x* must sum to 1)

power law exponent α

Power-law distribution



- high skew (asymmetry)
- straight line on a log-log plot

MEASURING POWER LAWS

Identifying power-law behavior in either natural or man-made systems can be tricky. The standard strategy makes use of a histogram of a quantity with a power-law distribution appears as a straight line when plotted on logarithmic scales. Just making a simple histogram, however, and plotting it on log scales to see if it looks straight is, in most cases, a poor way proceed.

Example on an artificially generated data set

- Take 1 million random numbers from a distribution with $\alpha = 2.5$
- Can be generated using the so-called 'transformation method'
- Generate random numbers r on the unit interval $0 \le r < 1$
- then $x = (1-r)^{-1/(\alpha-1)}$ is a random power law distributed real number in the range $1 \le x < \infty$

REPRESENTATION OF HISTOGRAMS

a) ***Normal histogram*** of the numbers: produced by binning them into bins of equal size 0.1. On linear scales used this produces a smooth curve.

b) ***Logarithmic scales.*** Same as a) but in ***logarithmic scales***: the characteristic straight-line form is revealed. The right-hand end of the distribution is noisy due to statistical errors.

c) ***Logarithmic binning.*** Same as b) but with variable width of the bins: each bin is a fixed multiple wider than the previous one. The bins of the tail of the distribution get more samples and reduces the statistical errors in the tail.

d) ***Cumulative distribution function***. Make a plot of the probability *P(x)* that *x* has a value greater than or equal to *x*: $P(x) = \int_x^\infty p(x')dx'$

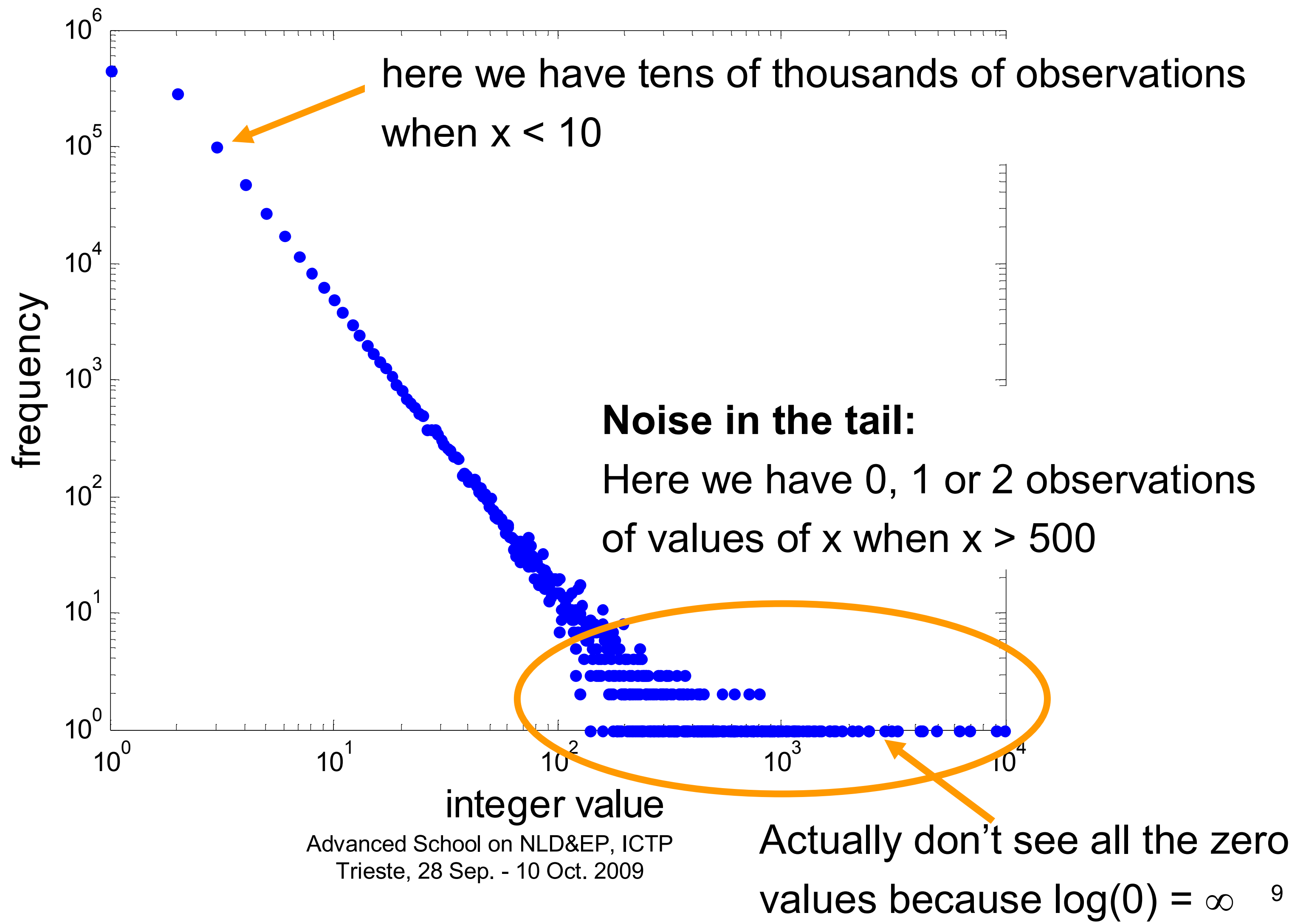
For a power law PDF, $p(x) = Cx^{-\alpha}$

$$\boxed{P(x) = \int_x^\infty C(x')^{-\alpha}\,dx' = \frac{C}{\alpha - 1}x^{-(\alpha-1)} \;\rightarrow\; \text{power law with exponent } \alpha\text{-1}}$$

There is no need to bin the data at all to calculate *P(x)*.

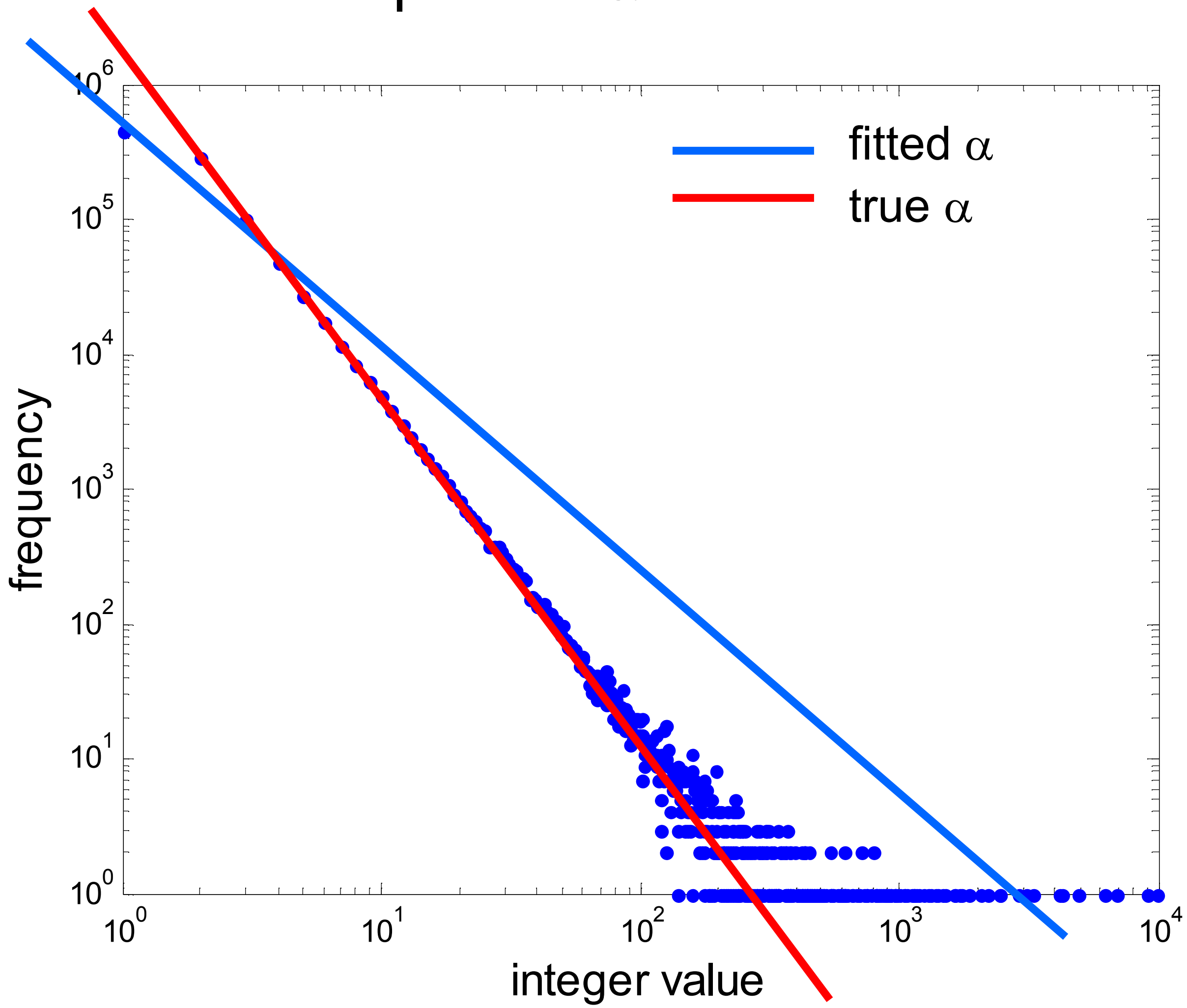
Log-log scale plot of straight binning of the data

- Same bins, but plotted on a log-log scale



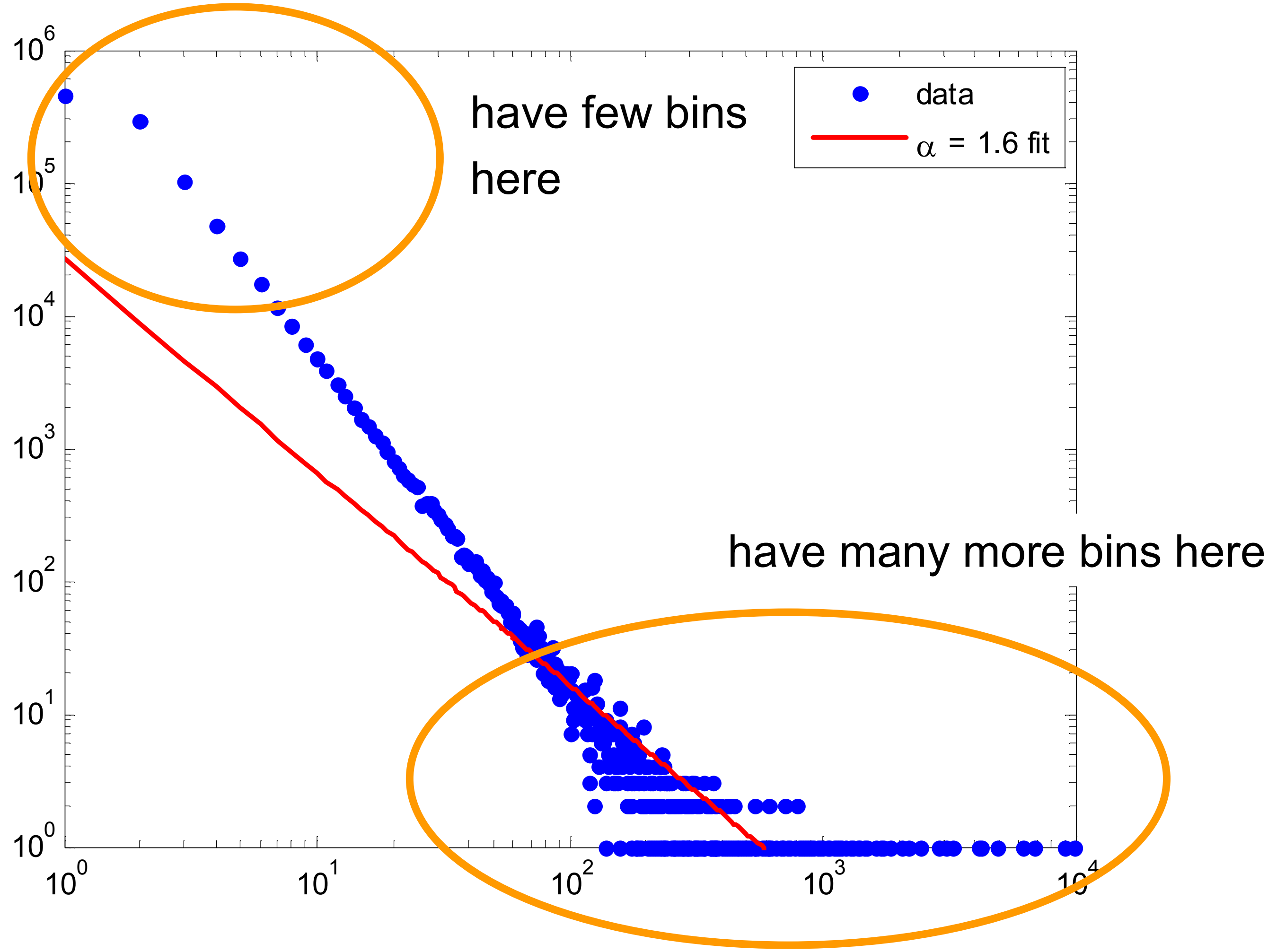
Log-log scale plot of straight binning of the data

- Fitting a straight line to it via least squares regression will give values of the exponent α that are too low



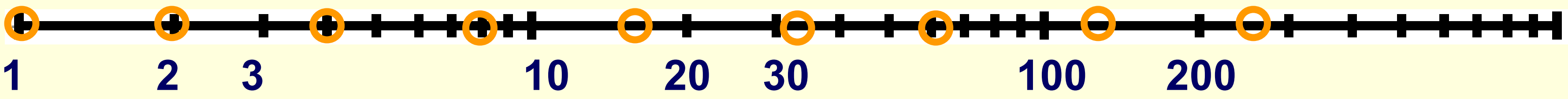
What goes wrong with straightforward binning

- Noise in the tail skews the regression result



Logarithmic axes

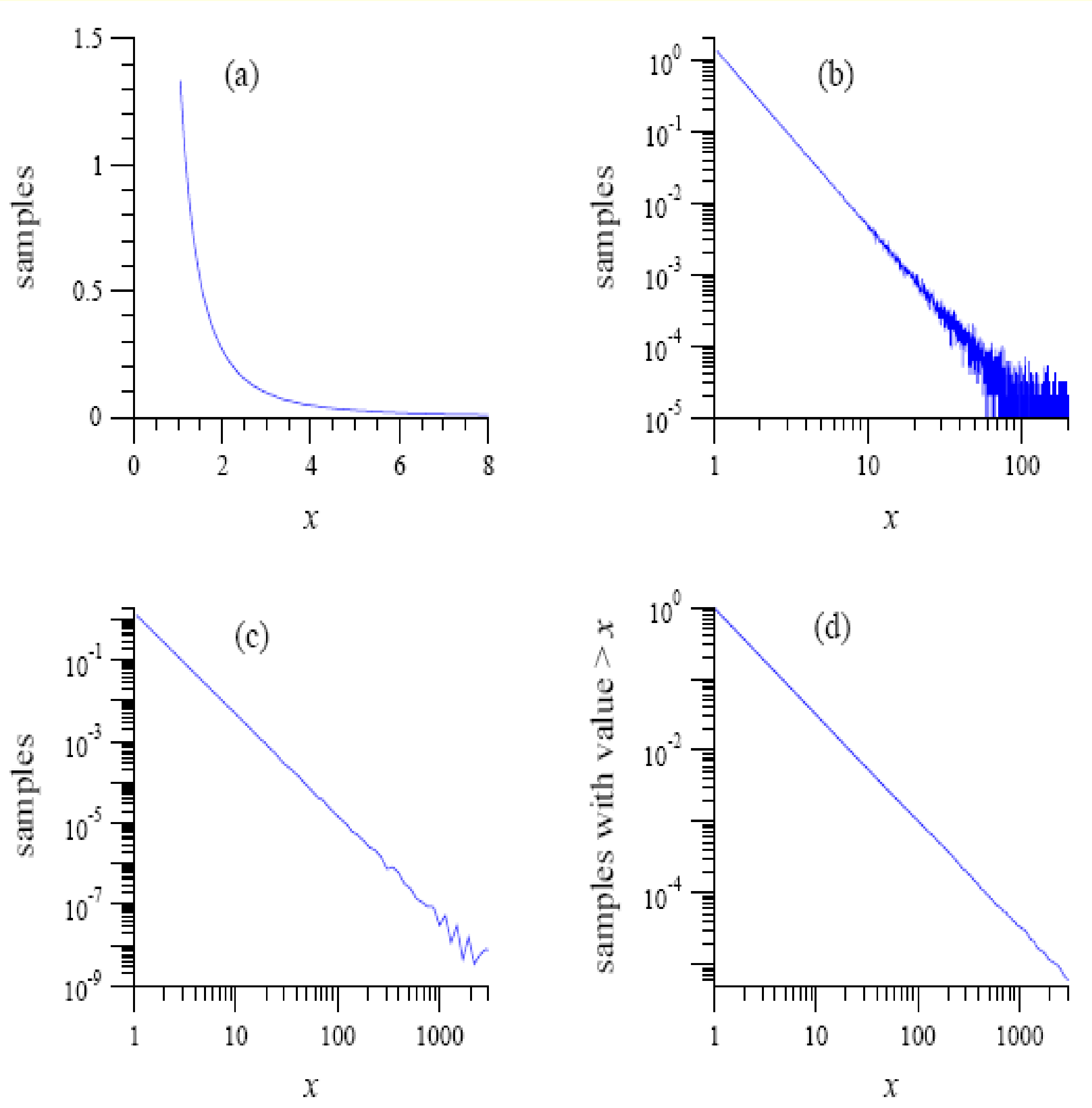
- powers of a number will be uniformly spaced



- $2^0=1$, $2^1=2$, $2^2=4$, $2^3=8$, $2^4=16$, $2^5=32$, $2^6=64$,….

Histogram of the set of 1 million random numbers

Histogram constructed using "logarithmic binning"



same histogram on logarithmic scales.

l.s.f.:
α = 2.26 +- 0.02

cumulative histogram or rank/frequency plot

Numerical simulation generated by a power law distribution $p(x) = Cx^{-\alpha}$, wit **α = 2.5**

RECOVERING THE EXPONENT α

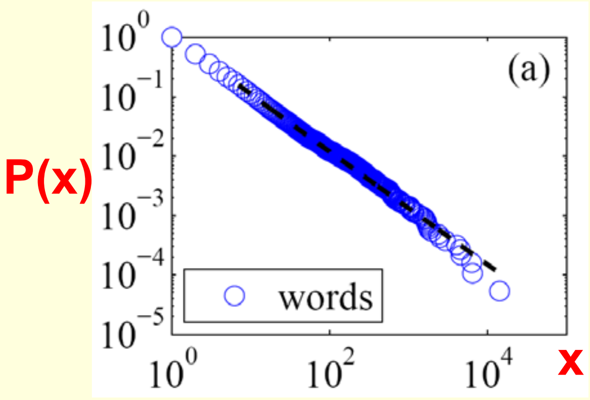
Very often: fit the slope of the line in plots like Figs. b, c or d. Unfortunately, it is known to introduce systematic biases into the value of the exponent. For example, a least-squares fit of a straight line to Fig. b gives α = 2.26 +- 0.02, which is clearly incompatible with the known value of α = 2.5.

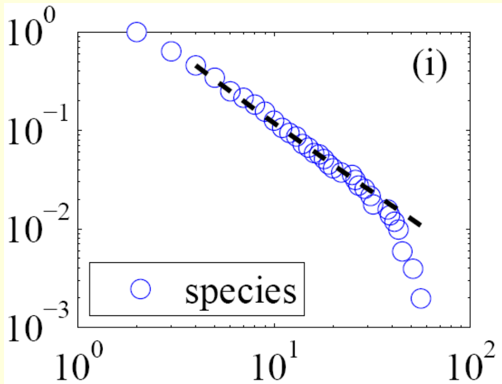
Alternative method: **Maximum likelihood fitting**

$$\alpha = 1 + n\left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}}\right]^{-1}, \quad \sigma = \sqrt{n}\left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}}\right]^{-1} = \frac{\alpha - 1}{\sqrt{n}}$$

The quantities x_i, i = 1 . . . n are the measured values of x; x_{min} corresponds to the smallest value of *x* for which the power-law behavior holds (the measured series is complete).
For the generated data set the recovered value is α = 2.503, very close to the true value.

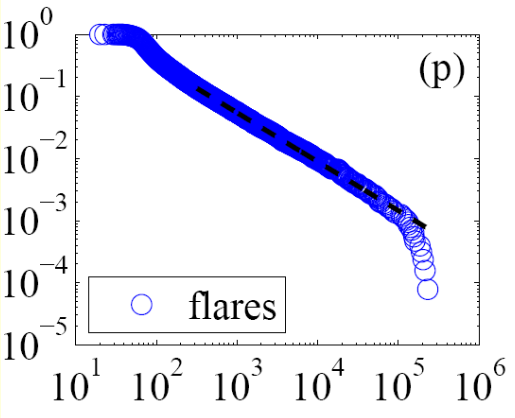
EXAMPLES

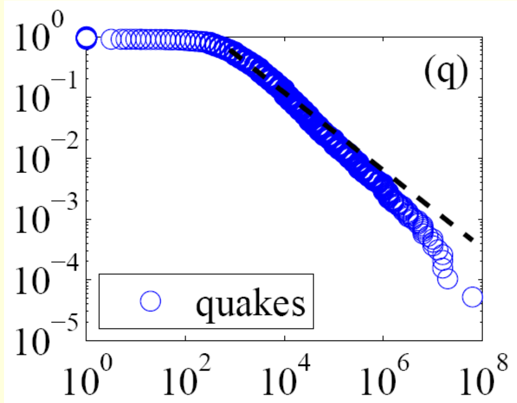




The frequency of occurrence of unique words in the novel Moby Dick by Herman Melville.

The number of species per genus of mammals. This data set, is composed primarily of species alive today but also includes a subset of recently extinct species.





Peak gamma-ray intensity of solar flares between 1980 and 1989.

Intensity of earthquakes occurring in California between 1910 and 1992, measured as the maximum amplitude of motion during the quake.

DISTRIBUTIONS THAT DO NOT FOLLOW A POWER LAW

Power-law distributions are impressively ubiquitous, but they are not the only form of broad distribution. A few examples:

a) The abundance of bird species spans over five orders of magnitude, but is probably distributed according to a log-normal (normal distribution of the logarithm of the quantity)

b) The number of entries in people's email address books, which spans over three orders of magnitude, seems to follow a stretched exponential (a curve of the form $\exp\left(-ax^{b}\right)$ for some constants *a* and *b*.

c) The distribution of the sizes of the forest fires, which spans six orders of magnitude, could follow a power law but with an exponential cutoff.

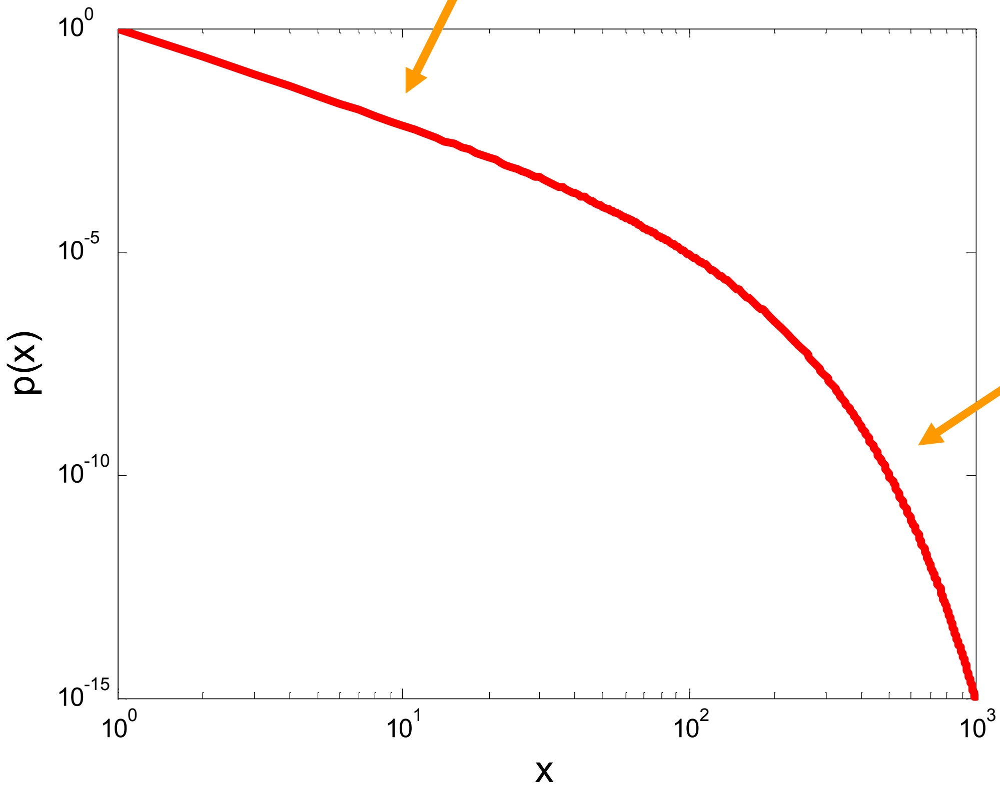
d) It could be also the case for earthquakes.

In analyzing a new set of data having a broad dynamical range and a highly skewed distribution, we should bean in mind that **a power law model is only one of several possibilities for fitting it**.

Another common distribution: power-law with an exponential cutoff

- $p(x) \sim x^{-a} e^{-k/\kappa}$

starts out as a power law



ends up as an exponential

but could also be a lognormal or double exponential…

	name	distribution $p(x) = Cf(x)$ $f(x)$	C
continuous	power law	$x^{-\alpha}$	$(\alpha-1)x_{\min}^{\alpha-1}$
	power law with cutoff	$x^{-\alpha}e^{-\lambda x}$	$\frac{\lambda^{\alpha-1}}{\Gamma(1-\alpha,\lambda x_{\min})}$
	exponential	$e^{-\lambda x}$	$\lambda e^{\lambda x_{\min}}$
	stretched exponential	$x^{\beta-1}e^{-\lambda x^{\beta}}$	$\beta\lambda e^{\lambda x_{\min}^{\beta}}$
	log-normal	$\frac{1}{x}\exp\left[-\frac{(\ln x-\mu)^2}{2\sigma^2}\right]$	$\sqrt{\frac{2}{\pi\sigma^2}}\left[\operatorname{erfc}\left(\frac{\ln x_{\min}-\mu}{\sqrt{2}\sigma}\right)\right]^{-1}$

Definition of the power-law distribution and several other common statistical distributions. For each distribution we give the basic functional form f(x) and the appropriate normalization constant C such that $\int_{x_{\min}}^{\infty} C f(x)dx = 1.$

NORMALIZATION

The constant C is given by the normalization requirement that

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

This only makes sense if α > 1, since otherwise the right-hand side of the equation would diverge: power laws with exponents less than unity cannot be normalized and don't normally occur in nature. If α > 1 then $C = (\alpha-1)x_{\min}^{\alpha-1}$ and the correct normalized expression for the power law itself is

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

Some distributions follow a power law for part of their range but are cut off at high values of *x*. That is, above some value they deviate from the power law and fall off quickly towards zero. If this happens, then the distribution may be normalized no matter what the value of the exponent .

MOMENTS

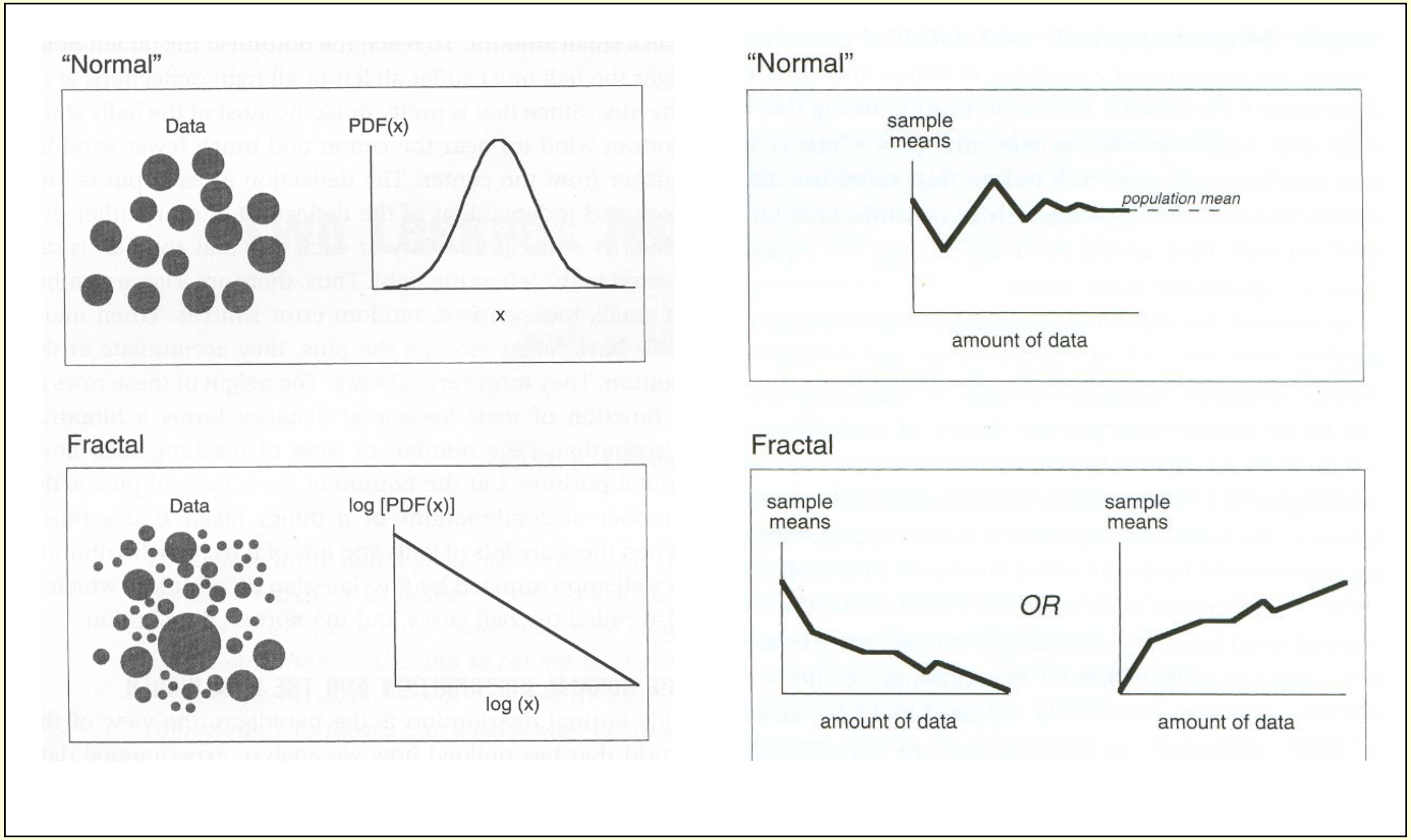
The mean value of the power-law distributed quantity *x* is given by

$$\langle x \rangle = \int_{x_{\min}}^{\infty} x\, p(x)\, dx = C \int_{x_{\min}}^{\infty} x^{-\alpha+1} dx$$

Note that this expression becomes infinite if *α* ≤ 2. Power laws with such low values of α have no finite mean: if we were to repeat our finite experiment many times and calculate the mean for each repetition, then the mean of those many means is itself also formally divergent, since it is simply equal to the mean we would calculate if all the repetitions were combined into one large experiment. This implies that, while the mean may take a relatively small value on any particular repetition of the experiment, it must occasionally take a huge value, in order that the overall mean diverge as the number of repetitions does. Thus there must be very large fluctuations in the value of the mean, and this is what the divergence really implies.

We can also calculate higher moments of the distribution *p(x).*

$$\langle x^m \rangle = \frac{\alpha - 1}{\alpha - 1 - m} x_{\min}^m$$


"Normal"
Data
PDF(x)
x
Fractal
Data
log [PDF(x)]
log (x)
"Normal"
sample means
population mean
amount of data
Fractal
sample means
OR
sample means
amount of data
amount of data

SCALE-FREE DISTRIBUTIONS

A power-law distribution is also sometimes called a scale-free distribution: a power law is the only distribution that is the same whatever scale we look at it on. Suppose we have some probability distribution *p(x)* for a quantity *x*, and satisfies the property that, for any *b*

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

If we increase the scale or units by which we measure x by a factor of b, the shape of the distribution p(x) is unchanged, except for an overall multiplicative constant.

There are some systems that become scale-free for certain special values of their governing parameters. The point defined by such a special value is called a “**continuous phase transition**” and the argument given above implies that at such a point the observable quantities in the system should adopt a power-law distribution.

MANIFESTATIONS OF POWER LAWS

Power laws can be viewed from two different points of view:

- Power-law decay of correlations
- Power-law size distributions

Correlations are related to systems in thermodynamic equilibrium, that should have correlations which decay **exponentially** over space and time, and how big a typical fluctuation should be.

Power law correlations are related to far from equilibrium systems or in equilibrium but very close to a critical point. Phase transitions have fluctuations which decay like power law, and many non-equilibrium systems do too.

Autocorrelation functions are defined in time domain, but very often they are transformed into the Fourier spectrum (**power spectrum**). A power-law decay for the correlations as a function of time translates into a power-law decay of the spectrum as a function of frequency. This is also called "**$1/f^{\alpha}$ noise**", whiz $\alpha \sim 1$.

MANIFESTATIONS OF POWER LAWS

Power law distributions say there is no typical scale or size for the variable, whereas the exponential and he Gaussian cases both have natural scale parameters.

$1/f^{\alpha}$ noise. Many different kinds of stochastic process, with no connection to critical phenomena, have power-law correlations. From the point of view of time series analysis, $1/f^{\alpha}$ noise can be considered as **long memory** processes, that can be obtained as a superposition of Gaussian autoregressive processes, also known as **fractional Brownian motion**, a particular case of self-affine time series.

GENERATION OF POWER LAW DISTRIBUTIONS

We can deal with the generation of power law distributions from two different points of view:

- Phenomenological (purely mathematical without any underlying physics): **algebraic models**
- Based on the **underlying physics**: complex approach, notably the physics of critical phenomena and the tools of the renormalization group that are used to analyze it. Two approaches are in order:
    - ❖ By developing physical models
    - ❖ Through numerical simulations

ALGEBRAIC MODELS

- Combination of exponentials
- Inverse quantities
- Random walks
- The Yule process

[See Newman (2005)]

PHYSICAL MODELS

- Phase transitions and critical phenomena
- Self-organized criticality
- Highly optimized tolerance
- Coherent noise

[See: Newman (2005),
Sornette (2000)]

NOISE 1/f

[Dutta & Horn (1981)]

1/f noise refers to fluctuations which have spectral densities varying approximately as 1/f over a large range of frequency, f. Fluctuations with such spectra have been observed in a tremendous variety of dissimilar physical systems. Is there a universality in the underlying equations which leads to 1/f noise in many apparently unrelated systems? The shape of the power spectrum uniquely characterizes the process only if it is stationary and Gaussian (all higher-order correlations are zero).

In a generic way, 1/f noise can be contemplated as an **activated random process**. A random process with a **characteristic time** τ has a Debye-Lorentzian spectrum

$$S(\omega) \propto \frac{\tau}{\omega^2\tau^2 + 1}$$

Any spectrum may be generated by postulating an appropriate **distribution *D(τ)* of the characteristic times** within the sample. This could arise if, for example, the sample was inhomogeneous. Then

$$S(\omega) \propto \int \frac{\tau}{\omega^2\tau^2 + 1} D(\tau)\, d\tau .$$

NOISE 1/f

In particular, if

$$D(\tau) \propto \tau^{-1} \qquad for \qquad \tau_1 \le \tau \le \tau_2,$$

then

$$S(\omega) \propto \omega^{-1} \qquad for \qquad \tau_2^{-1} << \omega << \tau_1^{-1}$$

If, as in many physical processes, τ is thermally activated,

$$\tau = \tau_0 \exp(E / kT),$$

then the required energy distribution is

$$D(E) = const \quad for \qquad kT \ln(\tau_1 / \tau_0) \le E \le kT \ln(\tau_2 / \tau_1)$$

The problem of justifying a 1/f spectrum has now been shifted to one of motivating the **required energy distribution**. Very often, specially in natural phenomena, several intervals can be found behaving as $1/f^{\alpha}$, each one with a different value of α.

NOISE 1/f

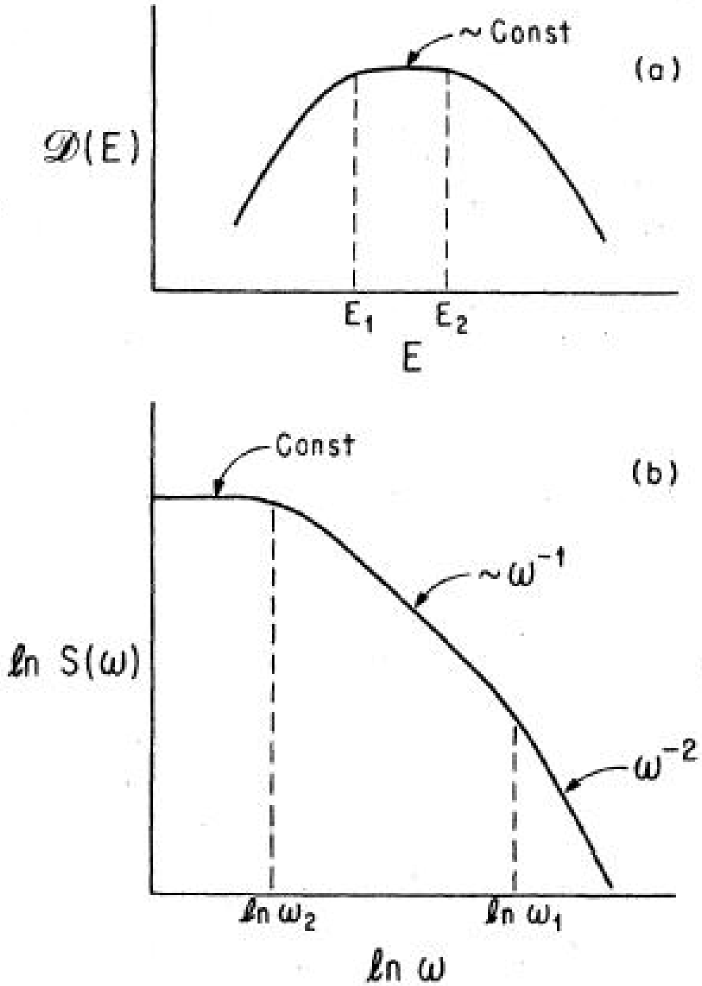
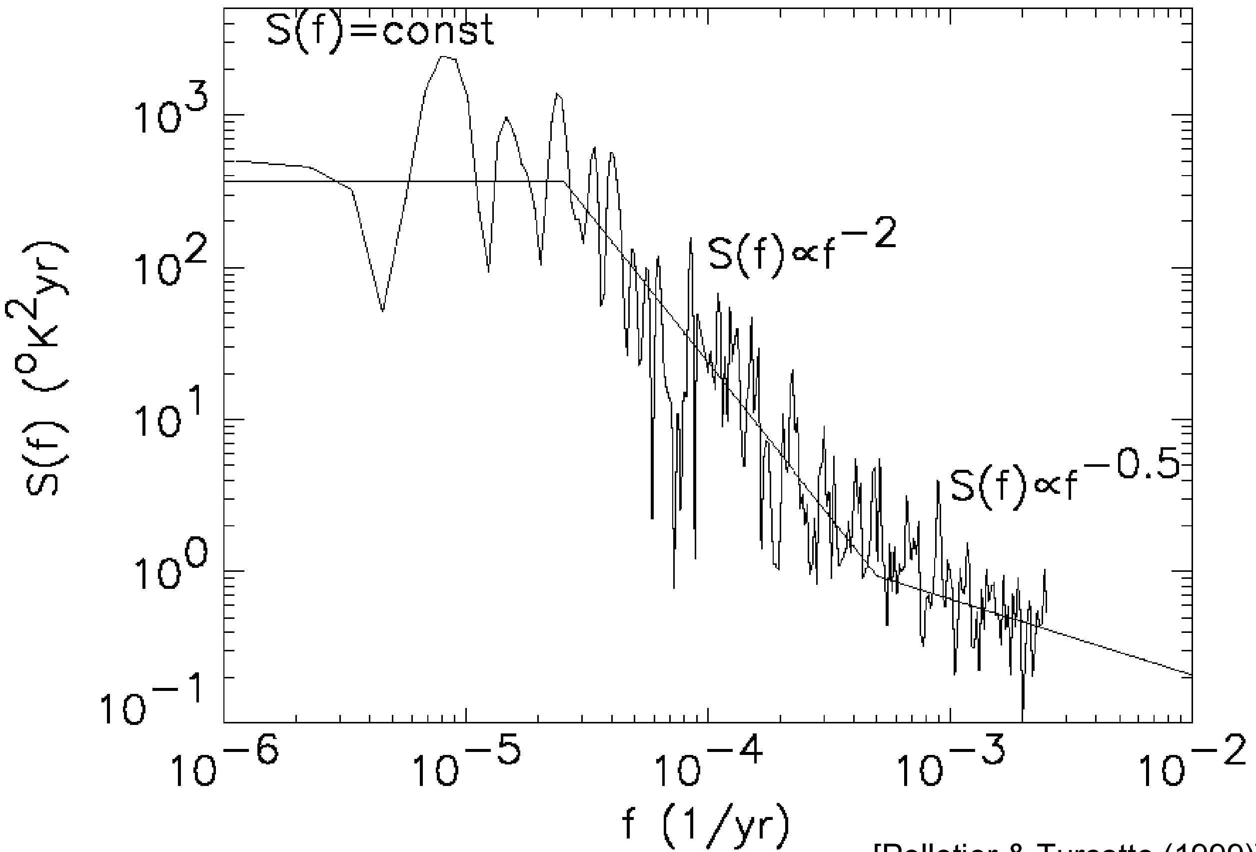


FIG. 5. (a) "flat" energy distribution. (b) frequency spectrum resulting from this distribution

EXAMPLES



[Pelletier & Turcotte (1999)]

Figure 2.2: Power-spectral density estimated with the Lomb periodogram of the **temperature** inferred from the Deuterium concentrations in the Vostok (East Antarctica) ice core. The power-spectral density S is given as a function of frequency for time scales of 500 yr to 200 kyr.

EXAMPLES

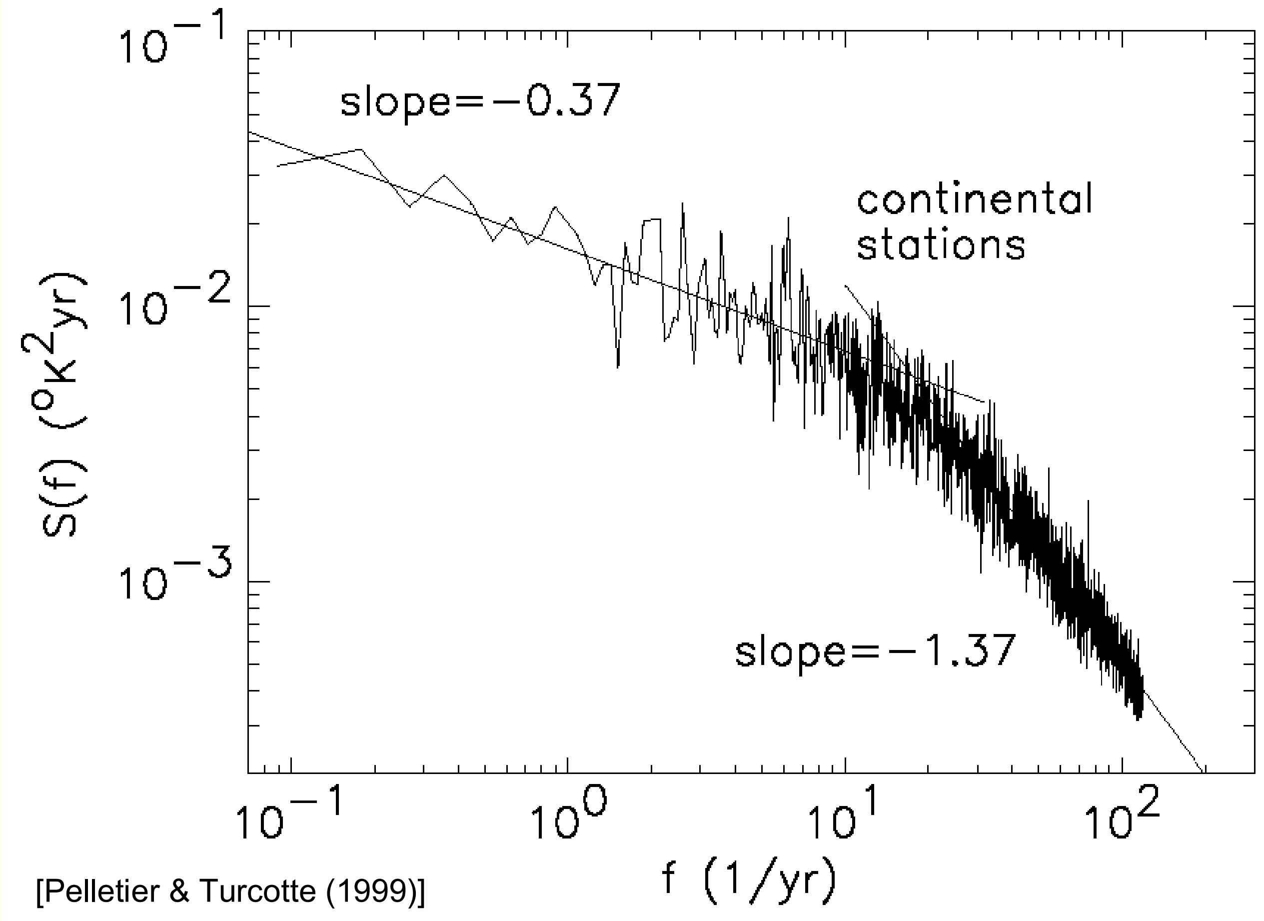
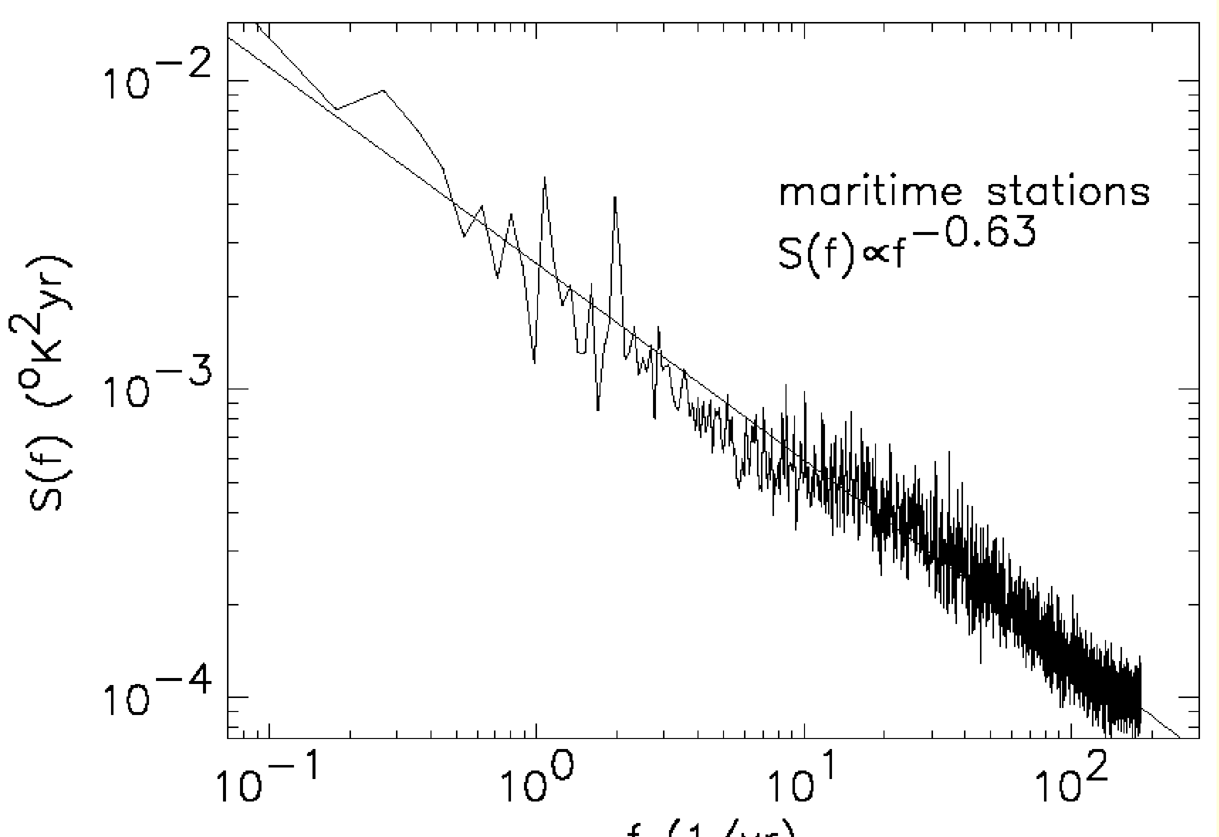


Figure 2.6: Average power-spectral density of 50 continental **daily temperature** time series from the data set of the National Climatic Data Center [1994] as a function of frequency in yr^{-1}. The power-spectral density S is given as a function of frequency for time scales of 2 days to 10 yr.

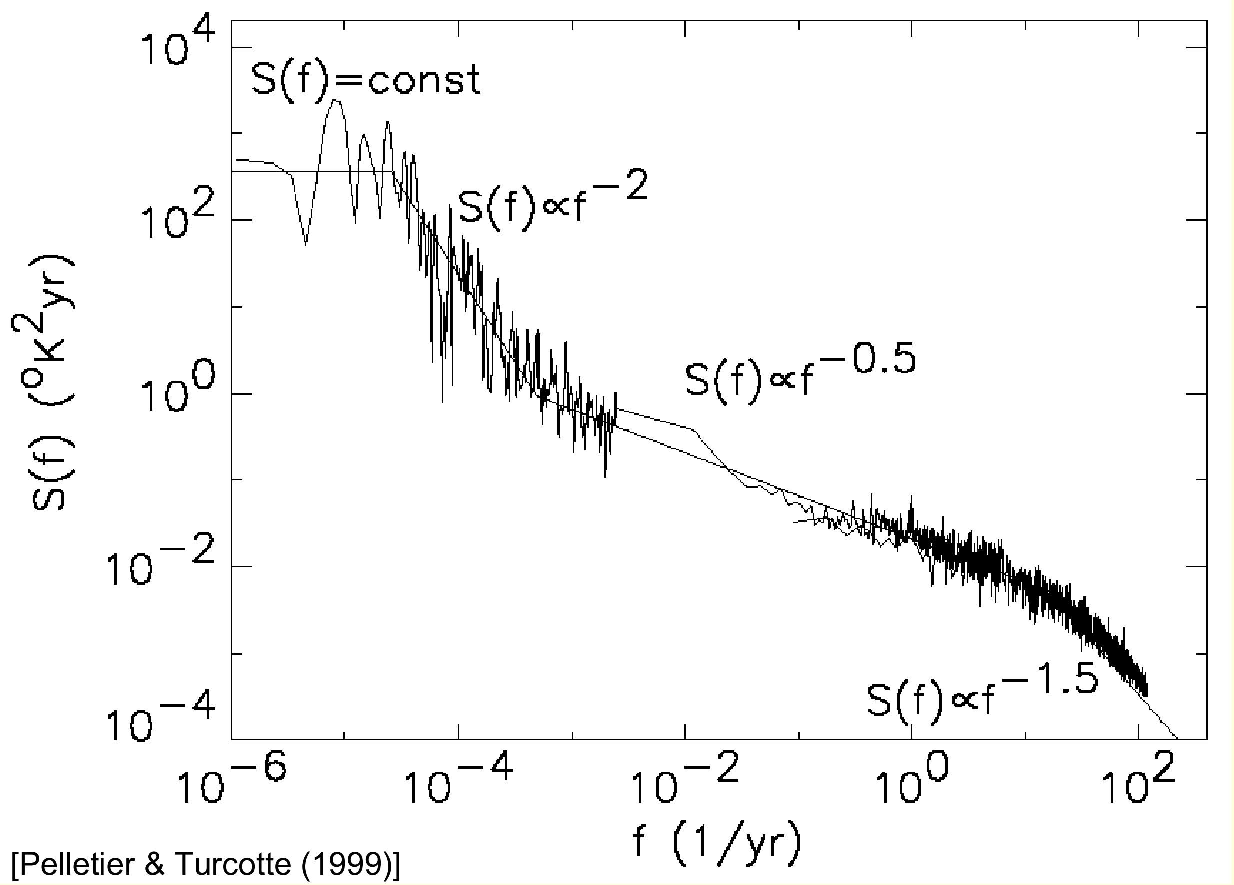
EXAMPLES



[Pelletier & Turcotte (1999)]

Figure 2.7: Average power-spectral density of 50 maritime **daily temperature** time series from the data set of the National Climatic Data Center [1994] as a function of frequency in yr-1. The power-spectral density S is given as a function of frequency for time scales of 2 days to 10 yr.

EXAMPLES



[Pelletier & Turcotte (1999)]

Figure 2.8: Power-spectral density of **local atmospheric temperature** from instrumental data and inferred from ice cores from time scales of 200 kyr to 2 days. The high frequency data are for continental stations. Piecewise power-law trends are indicated.

POWER LAW DISTRIBUTIONS IN EMPIRICAL DATA

Power-law distributions occur in many situations of scientific interest and have significant consequences for our understanding of natural and man-made phenomena.

Unfortunately, the empirical detection and characterization of power laws is made difficult by the large fluctuations that occur in the tail of the distribution.

In particular, standard methods such as least-squares fitting are known to produce systematically biased estimates of parameters for power-law distributions and should not be used in most circumstances.

In practice, we rarely, if ever, know for certain that an observed quantity is drawn from a power-law distribution. Instead, the best we can typically do is to say that our observations are consistent with a model of the world in which x is drawn from a power law distribution.

P-L DISTRIBUTION IN EMPIRICAL DATA

The maximum likelihood estimators are only guaranteed to be unbiased in the asymptotic limit of large sample size, $n \to \infty$. For finite data sets, biases are present but decay as $O(n^{-1})$ for any choice of x_{min}.

For very small data sets, such biases can be significant but in most practical situations they can be ignored because they are much smaller than the statistical error on the estimator, which decays as $O(n^{-1/2})$. Numerical simulations suggests that $n \geq 50$ is a reasonable rule of thumb for extracting reliable parameter estimates. Data sets smaller than this should be treated with caution.

Note, however, that there is another reason to treat small data sets with caution, which is typically more important, namely that it is difficult with such small set of data to rule out alternative forms for the distribution.

That is, for small data sets the power-law form may appear to be a good fit even when the data are drawn from a non-power-law distribution.

ESTIMATING THE LOWER BOUND ON P-L BEHAVIOR

The estimation of the lower limit x_{min} on the scaling behavior from data is an important issue.

It is important in the typical case where there is some non-power-law behavior at the lower end of the distribution of x. In such cases, we need a reliable method for estimating where power-law behavior starts: without it, we cannot make a reliable estimate of the scaling parameter.

If we choose too low a value for x_{min} we will get a biased estimate of the scaling parameter since we will be attempting to fit a power-law model to non-power-law data. On the other hand, if we choose too high a value for x_{min} we are effectively throwing away legitimate data points $x_i < x_{min}$, which increases both the statistical error on the scaling parameter and the bias from finite size effects.

Traditionally, x_{min} has been chosen either by visually identifying a point beyond which the PDF or CDF of the distribution becomes roughly straight on a log-log plot, or by plotting α as a function of x_{min} and identifying a point beyond which α appears relatively stable. These approaches are clearly subjective and can be sensitive to noise or fluctuations in the tail of the distribution.

TESTING THE POWER-LAW HYPOTHESIS

Assume we have performed a power-law fit to a given data set and provide good estimates of the parameters α and x_{min}. They tell us nothing, however, about whether the data are well fitted by the power law: obtaining roughly a straight line on a log-log plot is a necessary but not sufficient condition for power-law behaviour.

Given an observed data set and a power-law distribution from which the data are drawn, we want to know whether that hypothesis is a likely one given the data.

Questions of this type can be answered using goodness-of-fit tests that compare the observed data to the hypothesized distribution. Many such tests have been proposed, but one of the simplest, and more efficient, is based on the Kolmogorov-Smirnov statistic, the KS statistic.

THE KOLMOGOROV-SMIRNOV STATISTIC

The KS statistic is simply the maximum distance between the CDFs of the data and the fitted model:

$$D = \max_{x \geq x_{\min}} \left| S(x) - P(x) \right| .$$

Here S(x) is the CDF of the data for the observations with value at least x_{min}, and P(x) is the CDF for the power-law model that best fits the data in the region $x \geq x_{min}$.

The calculation returns a single number that is smaller for hypothesized distributions that are a better fit to the data. If this value is suitably small we can say that the power law is a plausible fit to the data; if the value is too large the power-law model can be ruled out. The crucial question we need to answer is, how large is too large?

The standard way to answer this question is to compute a ***p-value*** which quantifies the probability that our data were drawn from the hypothesized distribution, based on the observed goodness of fit.

THE P-VALUE

The p-value is defined to be the probability that a data set of the same size that is truly drawn from the hypothesized distribution would have goodness of fit D as bad or worse than the observed value. In essence, it tells you how likely it is that you saw results as bad as you did if the data really are power-law distributed.

If the p-value is much less than 1, then it is unlikely that the data are drawn from a power law. If it is closer to 1 then the data may be drawn from a power law, but it cannot be guaranteed.

This last point is an important one: the goodness-of-fit test and the accompanying p-value are a tool only for ruling out models, not for ruling them in. They can tell us when a model such as the power law is probably wrong, but they cannot tell us when it is right.

The best we can do by way of confirming the power-law model, in a strictly statistical sense, is to say that it is not ruled out by the observed data.

There is no known formula for calculating the p-value, but we can calculate it numerically by a Monte Carlo procedure.

REFERENCES

- Clauset, A., Rohilla Shalizi, C. And Newman, M.E.J. (2007). Power law distribution in empirical data, arXiv:0706.1062v1 [physics.data-an] (7 Jun 2007).
- Dutta & Horn (1981). Low-frequency fluctuations: 1/f noise, *Rev. Mod. Phys.,* 53, 497.
- Goldstein, M.L., Morris, S.A. And Yen, G.G. (2004). Problems with fitting to the power-law distribution, Eur. Phys.J.B, **41**, 255-258.
- Newman, M.E.J. (2005) Power laws, Pareto distributions and Zipf's law, Contemporary Physics, **46**, 323–351. / arXiv:cond-mat/0412004v3 [cond-mat.stat-mech] (29 May 2006).
- Pelletier, J.D. & Turcotte, D.L. (1999). Self affine time series: II. Applications and Models, in Advances in Geophysics, vol. 40, *Long-Range Persistence in Geophysical Time Series*. Academic Press, London.
- Sornette, D. (2000). Critical phenomena in Natural Sciences. Springer, Heidelberg.

School of Information
University of Michigan

Zipf's law & fat tails
Plotting and fitting distributions

Lecture 6

Instructor: Lada Adamic

Reading:

Lada Adamic, Zipf, Power-laws, and Pareto - a ranking tutorial,
http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html

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