The Scaling Laws Of Human Travel

Verena Zuber

Department of Statistics
University of Munich

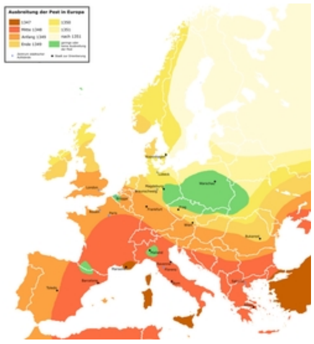
7. July 2006

Institut für Statistik München

Introduction: the plague

- highly infectious disease
- several pandemics have influenced the human history
- the severest pandemic has taken place in the 14th century
- about one third of the whole population died
- the beginning goes back to the 30ies in central Asia
- 1347 it was first observed on the European continent near the Krim (at the Black Sea)
- 1348 it reached south Europe
- 1349 it had spread almost all over Europe

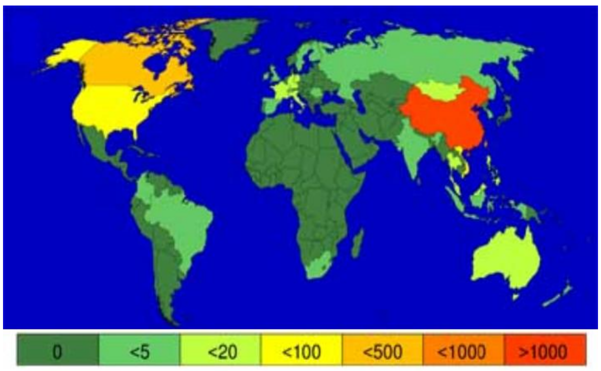
The spreading of the plague in Europe since 1347



Introduction: SARS (severe acute respiratory syndrom)

- infectious disease
- discovered in November 2002 in the chinese province Guangdong
- in February 2003 SARS has reached Vietnam and Hong-Kong
- in March about 2 000 cases were reported in Asia
- it also had jumped over to another continent: over 200 cases were registered in Canada
- soon SARS had spread worldwide: there have been about 8 500 cases in over 30 countries (about 900 lethal)
- there were 14 cases in Germany, too
- in summer 2003 SARS declined completely

The spreading of SARS in 2003



Result:

Today pandemics spread differently than in the Middle Ages: it took the plague two years to cover a distance of 2 000 km; SARS traversed the Pacific Ocean in less then a month!

- Surprisingly, until recently models to simulate the course of pandemics were used, that were appropriate for pandemics before the 20th century.
- As human mobility has undergone a significant change, these models failed to predict the course of recent pandemics.
- In this talk I am going to present a new approach (2006) by D. Brockmann, L. Hufnagel and T. Geisel

I. How to asses human travel?

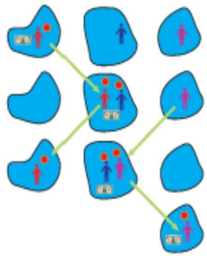
How to asses human travel?

No appropriate data exists to describe human travel:
thus the relationship between human travel and the movement of banknotes was analysed.
The banknote data is from the online bill-tracking website:

At the moment over 85 million bills are registered. This data includes 464 670 dollar bills with 1 033 095 reports. From these data we can obtain the distance traveled and the time elapsed.

The theoretical approach

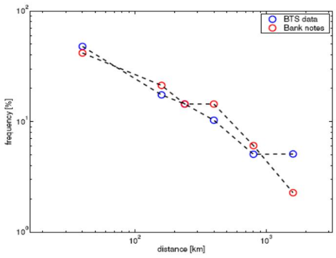
Home range:
Geographical patch in which a person resides most of the time.
People interact and consequently either share home ranges or entrance other peoples' home ranges. Thus an individuum infects others with a desease.



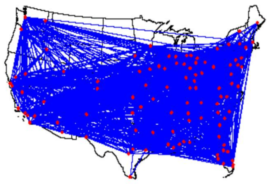
1.) US Bureau of Transportation Statistics Data

There are over 2.5 million round trip distances recorded,
but the spatial resolution is low and no data exists for the time the movement has taken.

A frequency distribution comparing the distance traveled:



2.) US aviation network data

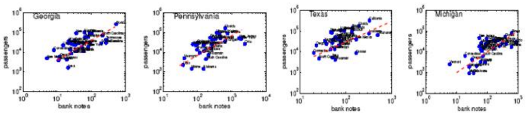


The passenger flux on the US aviation network as well as the movement of bills between different states can be represented as a matrix W or M, respectively, where:

- $W_{m,n}$: number of passengers travelling from state m to n
- $M_{m,n}$: number of banknotes travelling from state m to n

2.) Results:

Scatter plots for Georgia, Pennsylvania, Texas and Michigan:



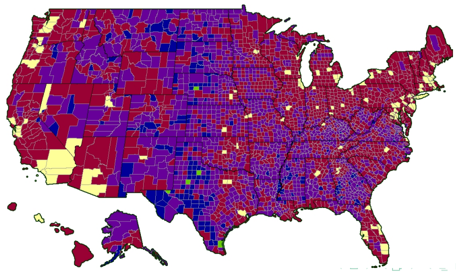
The degree of correlation can be estimated by the correlation coefficient R for matrices:

$$R = \frac{\langle W_{m,n} M_{m,n} \rangle - \langle W_{m,n} \rangle \langle M_{m,n} \rangle}{\sqrt{(\langle W_{m,n}^2 \rangle \langle W_{m,n} \rangle)^2 (\langle M_{m,n}^2 \rangle \langle M_{m,n} \rangle)^2}}$$

- the single correlation for almost all states is significantly positive
- overall correlation coefficient is approximately 0.5

Data set

The data are based on the online bill-tracking system:
www.wheresgeorge.com
At the moment more then 6% of the total U.S. Currency in circulation is registered. This data set includes 464 670 dollar bills with 1 033 095 reports.



Data set

Variables:

- **geographical displacement**: $\mid x_2 - x_1 \mid$
  x_1 first and x_2 secondary report of a banknote;
  boundarys:
  - short range radius (10km)
  - approximate average East-West extension of the US ($\approx$ 3 200km)
- **elapsed time** T between first and secondary report
- **population size** of the initial entry location:
  - highly populated metropolitan areas ($N_{loc} \geq$ 120 000; about 35.7% of the entire population)
  - cities of intermediate size (120 000 $\geq N_{loc} \geq$ 22 000; about 29.1% of the entire population)
  - small towns ($N_{loc} \leq$ 22 000; about 25.2% of the entire population)

II. The theory of stochastic processes

Overview

state space / **parameter space**

parameter space	discrete (state space)	continuous (state space)
discrete	Simple Random Walk, Levy Flight, Markov-chain	Gaussian Random Walk
continuous	Discrete Markov-chain, Poisson Process, Counting Process	Wiener Process, Brownian Motion, Continuous Time Random Walk (CTRW)

The following theory is presented for only **one** dimension; it can be easily extended to more dimensions.

Diffusion (Brownian Motion)

"We say a particle is *diffusing* about in space whenever it experiences erratic and disordered motion through the space. For example, we may speak of radioactive particles diffusing through the atmosphere, or even of a rumour diffusing through a population."

stochastic model: the **Wiener Process** $W = \{W(t) : t \geq 0\}$:

- $W(t)$ has independent and stationary increments, that means the distribution is only dependent on δt
- $W(s+t) - W(s)$ is distributed as $N(0, \sigma^2 t)$ for all $s, t \geq 0$, where σ^2 is a positive constant
- the sample paths of $W(t)$ are continous, but nowhere differentiable functions of t
- self-similarity (fractals)

Simple Random Walk (SRW)

The position of a SRW is frequently defined as a sum of N i.i.d. displacements ΔX_n:

$$X_N = \sum_{n=1}^{N} \Delta X_n$$

Each displacement is drawn from the same probability density function (pdf) $p(\Delta x)$. The scaled position is defined as:

$$Y_N = \frac{X_N}{\sqrt{N}}$$

And thus we get the universal scaling relation with the coefficient $\beta = 2$:

$$X_N \sim \sqrt{N} = N^{\frac{1}{\beta}} = N^{\frac{1}{2}}$$

Lévy Flight (LF)

LF can be described by analogy to the Simple Random Walk by a sum of i.i.d. random increments. But the single step pdf $p(\Delta x)$ shows algebraic tails, so that its variance is infinite:

$$p(\Delta x) \sim \frac{1}{\Delta x^{1+\beta}}$$

where: $0 \leq \beta \leq 2$
The scaled position is defined as:

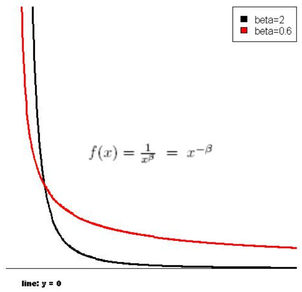
$$Y_N = \frac{X_N}{N^{\frac{1}{\beta}}}$$

And thus we get the universal scaling relation:

$$X_N \sim N^{\frac{1}{\beta}}$$

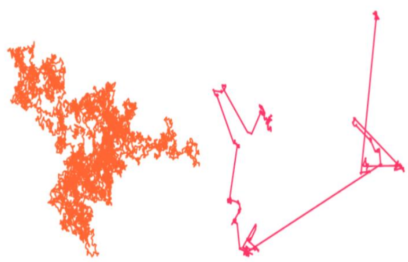
Comparing Simple Random Walk and Lévy Flight

In contrast:
the universal scaling relation
for SRW (in black: $\beta = 2$) and LF (in red $\beta = 0.6$) :



Comparing Simple Random Walk and Lévy Flight

In contrast:
two-dimensional paths of a SRW (in red; on the left) and LF (in pink; on the right):



Continuous Time Random Walk (CTRW)

A simple version of a CTRW is defined by two pdfs:

- for spacial displacements $f(\Delta x)$
- for temporal increments $\phi(\Delta t)$

A CTRW consists of pairwise random and stochastically independent events:

- a spatial displacement Δx
- a temporal increment Δt

Both are drawn from the combined pdf:

$$p(\Delta x, \Delta t) = f(\Delta x)\phi(\Delta t)$$

Continuous Time Random Walk (CTRW)

After N iterations:
the position of the walker is given by $X_N = \sum_{n=1}^{N} \Delta X_n$
and the time elapsed is $T_N = \sum_{n=1}^{N} \Delta t_n$.

The pdf $W(x,t)$, indicating the position $X(t)$ after time t, can be obtained using the Laplace- and Fourier transformation.

From this generalized model we can derive more special ones by changing the characteristics of the two pdfs $f(\Delta x)$ and $\phi(\Delta t)$.

Continuous Time Random Walks (CTRW)

- **Simple Random Walk**
  When both, the variance of the spatial steps σ^2 and the expectation value of the temporal increments τ exist, then CTRW are equivalent to Simple Random Walks.
- **Lévy Flight**
  A CTRW with algebraically distributed spatial steps of infinite variance is equivalent to an ordinary Lévy Flight.
- **Fractional Brownian Motion**
  The complementary scenario occurs, when ordinary spatial steps with finite variance are combined with a pdf of the temporal increments with infinite expectation value τ.
  In this case, the time between successive spatial increments can be very long, effectively slowing down the random walk.

Ambivalent Processes

The last and most interesting combination of waiting times and spatial steps is the one, in which long waiting times compete and interfere with long range spatial steps. Thus we get the following pdfs:

$$f(\Delta x) \sim \frac{1}{\Delta x^{1+\beta}} \qquad \text{where} \qquad 0 \leq \beta \leq 2$$

$$\phi(\Delta t) \sim \frac{1}{\Delta t^{1+\alpha}} \qquad \text{where} \qquad 0 \leq \alpha \leq 1$$

The scaling relation is defined as:

$$X(t) \sim t^{\frac{\alpha}{\beta}} = t^{\frac{1}{\mu}}$$

Ambivalent Processes

The ratio of the exponents $\frac{\beta}{\alpha} = \mu$ resembles the interplay between sub- and superdiffusion:

- superdiffusive $\beta \leq 2\alpha$ or $\mu < 2$
- quasidiffusive $\beta = 2\alpha$ or $\mu = 2$
- subdiffusive $\beta \geq 2\alpha$or $\mu > 2$

III. Application

Dispersal

What is **dispersal** here?
"*geographical displacements of individuals in a short time*"

$P(r)$ is the pdf for finding a displacement r in less then δt days.
It quantifies the frequency of travel distances for individuals as a function of geographical distance.
Dispersal equation:

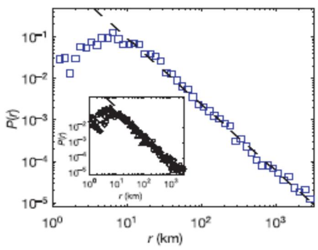
$$P(r) \sim r^{-(1+\beta)}$$

Visualization: dispersal curve

Dispersal curve

$P(r)$ of traversing a distance r in less then $t = 4$ days ($N = 20\,540$)
black line: dispersal kernel $\beta = 0.59 \pm 0.02$

→ **Model 1: Lévy Flight with** $\beta = 0.59$

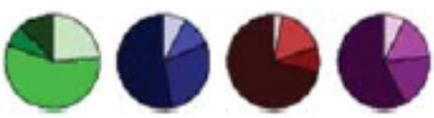


Model 1: Lévy Flight

Model 1: Lévy Flight with $\beta = 0.59$

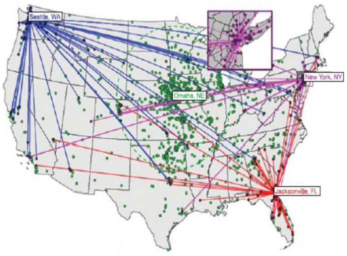
It is typical for Lévy Flights, that after a certain time T_{eq} all subjects are spread equally over the whole space.
In this case: If you look at a number of bills with the same initial entry location, after T_{eq} they should be spread all over the US and only a tiny fraction is to remain in the initial area.
T_{eq} can be estimated Lévy Flight with $\beta = 0.59$: $\hat{T}_{eq} \approx 68$ days

Pie Charts indicating the proportion of bills to be found inside the radius of 50 km around the initial area (dark section) for Omaha, Seattle, Jacksonville and New York after 100 days:



Model 1: Lévy flight with $\beta = 0.59$

Omaha: about every fifth bill stays in a radius about 50km!
Short-time trajectories ($t < 14$ days) for these four states are represented by lines
and only for Omaha are visualized the long-time trajetories ($t > 100$ days) by green dots.



Model 1: Lévy flight with $\beta = 0.59$

$P_0^i(t)$: relative proportion of banknotes which are reported in a small radius (20 km) of the initial entry location i at time t:

$$P_0(t) \sim t^{-\eta}$$

From the data we get: $\eta = 0.6 \pm 0.03$ universally for the different initial entry locations ($N = 25\,375$).
For a LF the relative proportion is defined as:

$$P_0(t) \sim t^{-\frac{2}{\beta}}$$

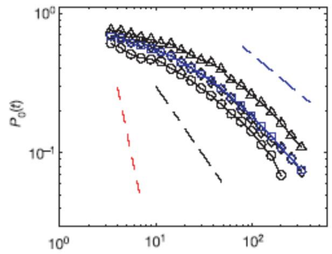
with $\beta = 0.59$ we get: $\eta \approx 3.33 \neq 0.6$

Model 1: Lévy flight with $\beta = 0.59$

$P_0^i(t)$ is computed for the different initial entry locations:

- highly populated metropolitan areas: black triangles
- cities of intermediate size: blue diamonds
- small towns: black circles

A Lévy flight with $\beta = 0.59$ is presented as a red line.



Result

A pure Lévy Flight can NOT explain the motion of the bills!
There must be some kind of attenuation owing to:

- spatial inhomogenities
- long periods of rest (attenuation of superdiffusion)

Possible solution:
Creating a tail in the probability density $\phi(t)$ for the time t between successive spatial displacements (subdiffusion).

New model:

Continuous Time Random Walk: Ambivalent Process

Model 2: Ambivalent Process

Model 2: Ambivalent Process

Remember: a CTRW can be defined by two pdfs:

- for spacial displacements $f(\Delta x)$ with variance σ^2
- for temporal increments $\phi(\Delta t)$ with mean τ

If σ^2 and τ are **finite**, we obtain Ordinary Diffusion. The diffusion equation for the pdf $W(x,t)$, indicating the position $X(t)$ after time t, states:

$$\delta_t W(x,t) = D\delta_x^2 W(x,t)$$

with diffusion coefficient $D = \frac{\sigma^2}{\tau}$

Model 2: Ambivalent Process

BUT here:
$P(\delta x_n) \sim\mid \delta x_n \mid^{-(1+\beta)}$ and
$\phi(\delta t_n) \sim\mid \delta t_n \mid^{-(1+\alpha)}$ have heavy tails; thus σ^2 and τ are **infinite**.

The diffusion equation (core dynamical equation) is given by:

$$\delta_t^\alpha W(x,t) = D_{\alpha,\beta}\delta_{|x|}^\beta W(x,t)$$

with:

- the fractional derivatives: δ_t^α and $\delta_{|x|}^\beta$, depending only on α and β
- a constant, generalized diffusion coeffizient: $D_{\alpha,\beta}$

Model 2: Ambivalent Process

With non-trivial methods (fractional calculus) this leads to an equation discribing the probability of having traversed a distance r at time t:

$$W_r(r,t) = t^{-\frac{\alpha}{\beta}} L_{\alpha,\beta}\left(\frac{r}{t^{-\frac{\alpha}{\beta}}}\right) \quad (1)$$

$L_{\alpha,\beta} = W_r(r,t)t^{\frac{\alpha}{\beta}}$ is the universal scaling function, that represents the characteristics of the process.
So a typical distance travelled scales:

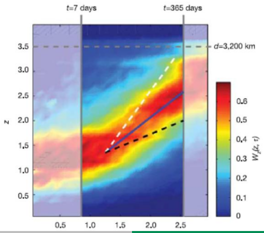
$$r(t) \sim t^{-\frac{\alpha}{\beta}} = t^{-\frac{1}{\mu}} \qquad \text{with } \mu = \frac{\beta}{\alpha} \quad (2)$$

Using the former results ($\beta = 0.59 \pm 0.02$ and $\alpha = 0.60 \pm 0.03$ and thus $\beta < 2\alpha \rightarrow$ superdiffusiv): we get $\mu = 0.98 \pm 0.08$.

Testing Validity

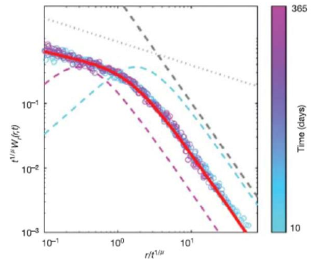
$W_r(r,t)$ can be estimated from our data: $\mu = 1.05 \pm 0.02$
The probablity of $W_r(r,t)$ is represented by the colouring (red areas idicate the highest probability).

- blue line: the prediction: $\mu = 1.05 \pm 0.02$
- white line: pure Lévy Process with $\beta = 0.6$
- black line: ordinary diffusion



Testing Validity

If the scaling function $L_{\alpha,\beta} = W_r(r,t)t^{\frac{\alpha}{\beta}}$ is universal, then the courses of the rescaled pdf ($W_r(r,t)t^{\frac{\alpha}{\beta}}$; y-axis) as functions of the rescaled distance ($rt^{\frac{\alpha}{\beta}}$; x-axis) for different elapse times t, are to collapse to one line.



Summary

- Different statistical models exist to describe the motion of almost anything:
  - Simple Random Walk
  - Lévy Flight
  - Fractional Brownian Motion
  - Ambivalent Process
- An Ambivalent Process is the first choice to fit the motion of banknotes.
- As the motion of banknotes is firmly linked to human travel we can conclude, that human travel can be described by an universal, superdiffusive Ambivalent Process.
- This is the first empirical evidence for such a process in nature.

Sources

- Introduction: Wikepedia
- The theory of stochastic processes:
  "Probability And Random Process" by G. Grimmet and D. Stirzaker; Oxford University Press: Oxford, 2001
  "The Random Walks Guide To Anomalous Diffusion: A Fractional Dynamics Approach" by R. Metzler and J. Klafter in Phys. Rep.; Vol 339, 2000
- Paper:
  "The Scaling Laws Of Human Travel" by D. Brockmann, L. Hufnagel and T. Geisel in nature; Vol 439, 26. January 2006

Thanks for your interest!