Stochastic Processes

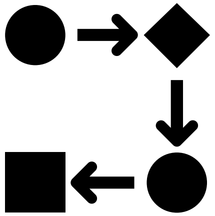
Deepthi Rajashekar

Outline

- Background
  - Motivation
  - Examples
- Definition
  - Terminologies
  - Examples
- Classification
  - Discrete time
    - Simple Random Walk
    - Markov Chains
  - Continuous time
    - Brownian Process
    - Poisson Process
    - Gaussian Process

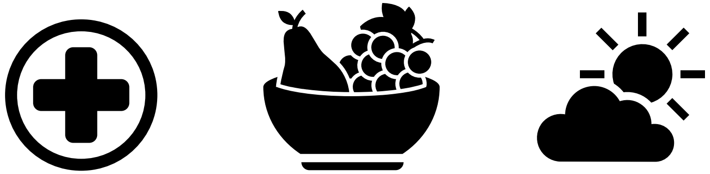
Motivation

Consumer behavior



Fidelity &Demand

Population growth



Uncertain factors

Queueing

Cost optimization

Congestion

Operational efficiency

Refresher

X =

1 

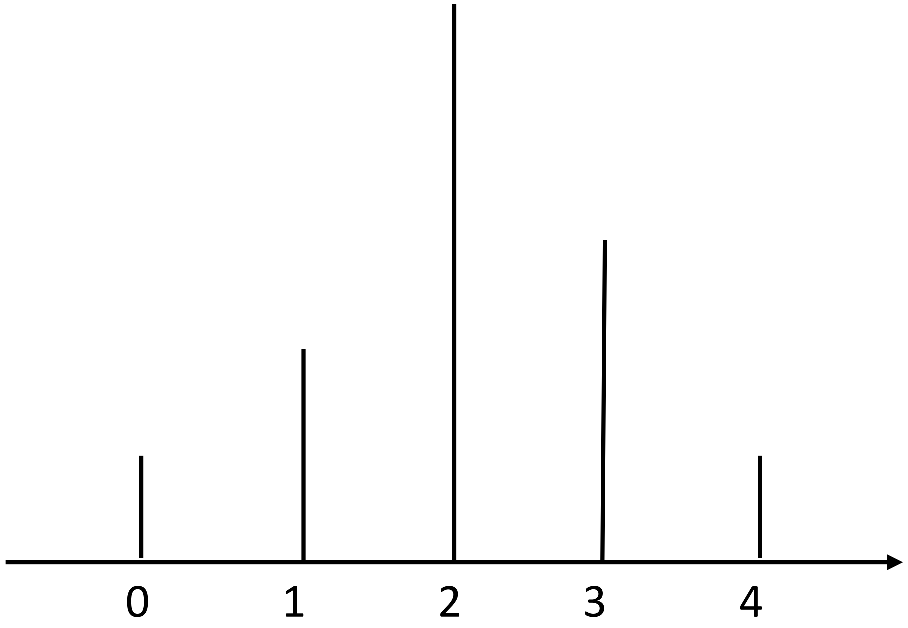
0 

Y = sum of money won in a coin toss game after time ‘t’.

Random variable := mapping *outcomes* of a random process to *numbers*

X := Number of workouts this week

X	P(x)
0	0.1
1	0.15
2	0.4
3	0.25
4	0.1



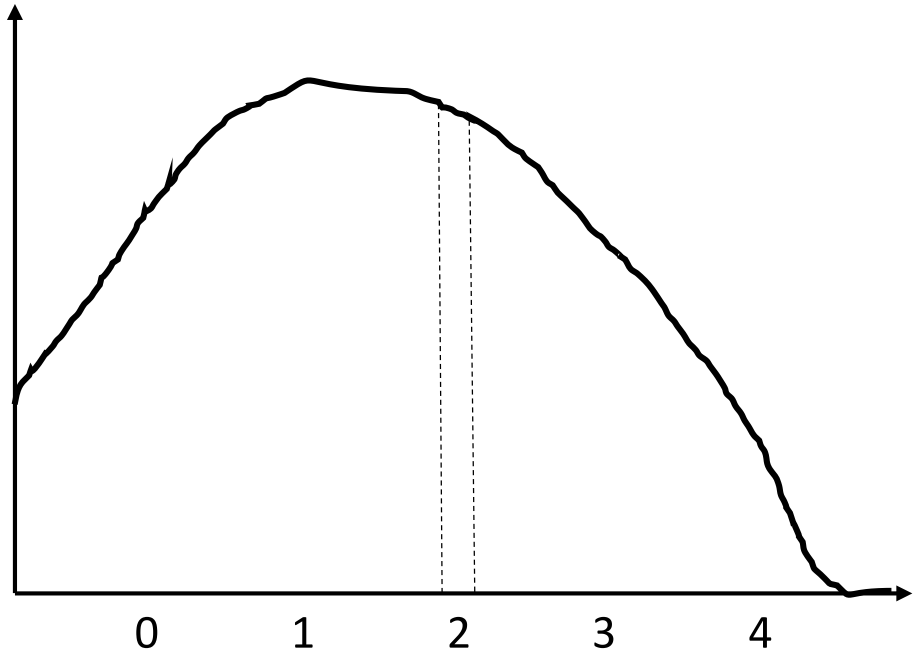
Discrete probability distribution

$E(x) = u_x$
$= 0*0.1 + 1*0.15 + 2*0.4 + 3*0.25 + 4*0.1$
$= 2.1$

$Var(x) = (0-2.1)^2*0.1 + (1-2.1)^2*0.15 + \ldots$
$+ (4-2.1)^2*0.1$

$Std(x) = \sqrt{var(x)} = 1.09$

Y := *exact* amount of snow this weekend



Area under curve

$$P(|Y-2| < 0.1) = \int_{1.9}^{2.1} f(y)\,dy$$

Informal definition

Brand Switching:

Consumer preference at time ‘t’

$\{X(t), t \geq 0\}$

Maxwell

Lavazza

Nespresso

State space

Stochastic Process

Parameter space

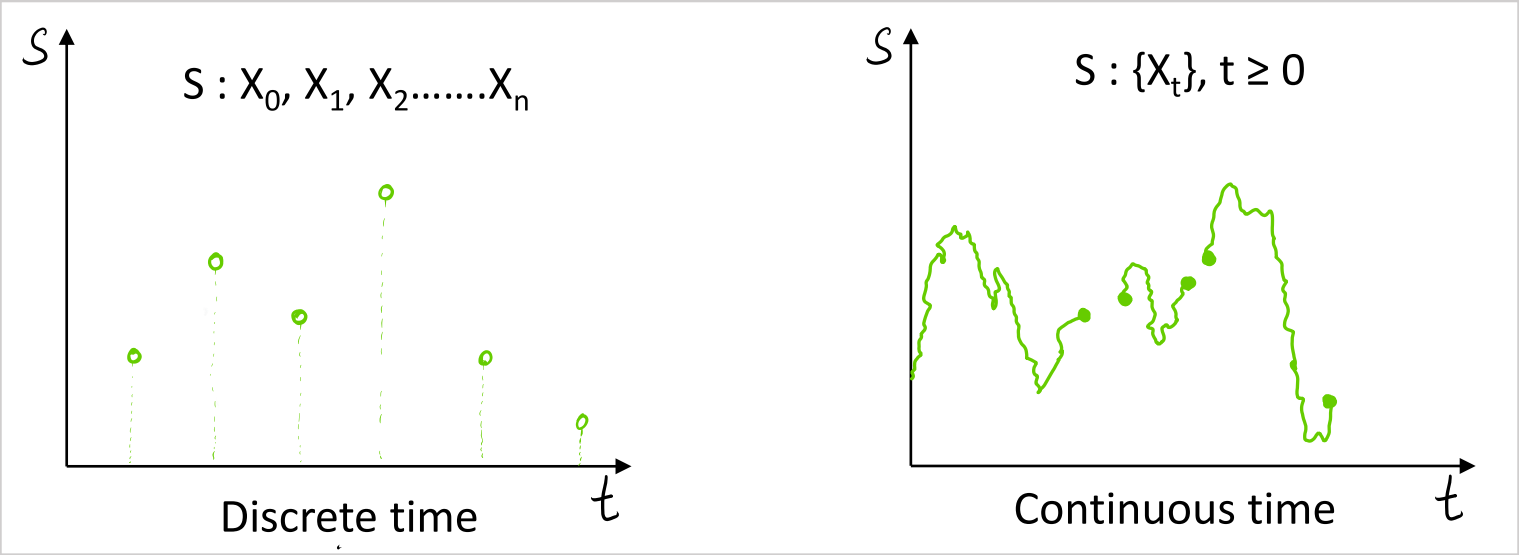
Stochastic process := collection of random variables indexed by time $_{\text{(usually)}}$

Probability distribution over a space of paths.

Classification

STATE SPACE \ PARAMETER SPACE	Discrete	Continuous
Discrete	Monthly consumer preference	# jobs waiting at *any time* of day
Continuous	*exact* amount of demand each day of the month	waiting time of an arriving job before it is serviced

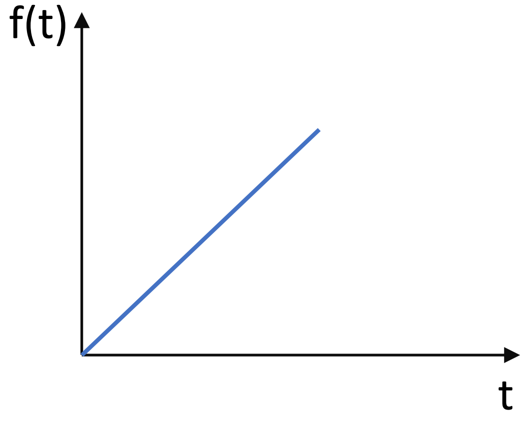
Classification

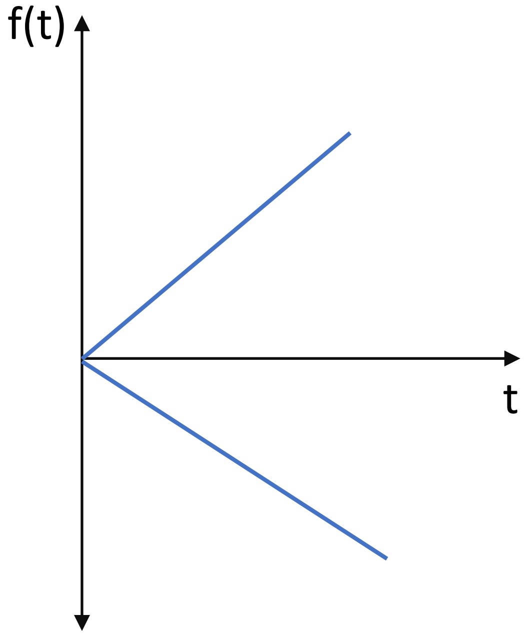


Discrete time (recall formal definition)

$f(t) = t, \qquad pr = 1$

$$f(t) = \begin{cases} -t, & pr = 0.5 \\ \ \ t, & pr = 0.5 \end{cases}$$



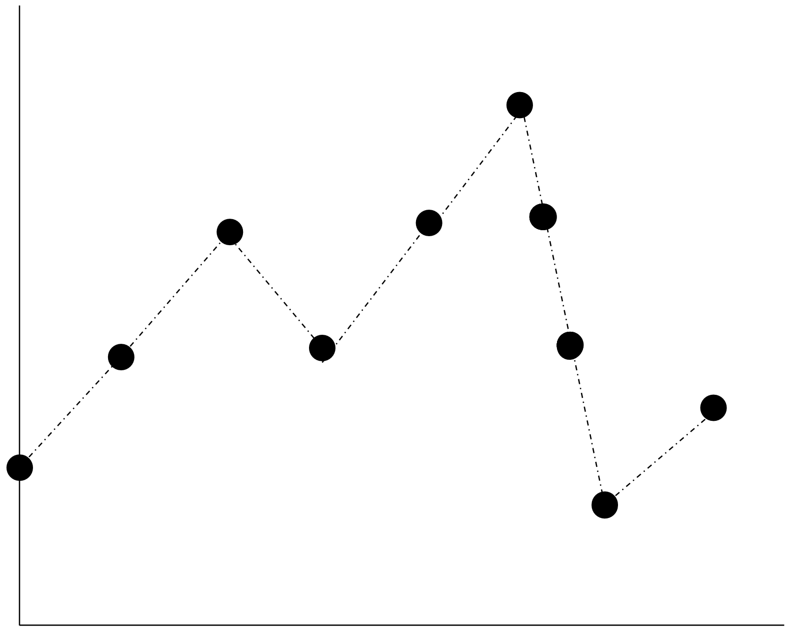


Random walk [theory]

$$Y_i = \begin{cases} 1 & pr = 0.5 \\ -1 & pr = 0.5 \end{cases}$$

$$X_t = \sum_{i=1}^{t} Y_i$$

$$S:\ X_0, X_1, X_2, X_3, \ldots\ldots$$



Random variable

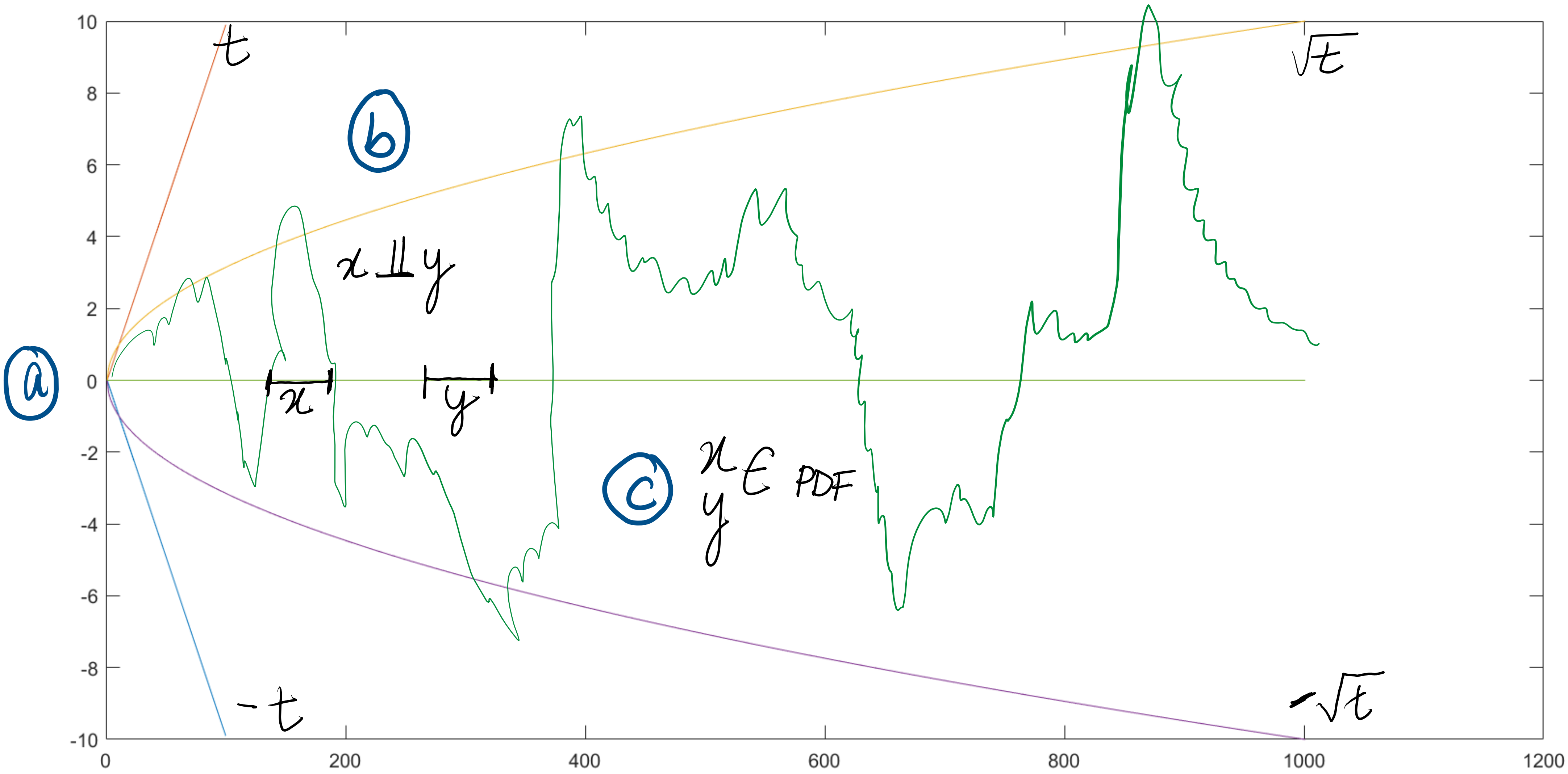
$E(Y_i)$ = 1*0.5 – 1*0.5 = 0

$Var(Y_i) = (1-0)^2 * 0.5 + (-1-0)^2 * 0.5$ = 1

Random walk

$$E(X_t) = \sum_{i=1}^{t} E(Y_i) = 0$$

$$Var(X_t) = \sum_{i=1}^{t} Var(Y_i) = t$$



10
8
6
4
2
0
-2
-4
-6
-8
-10
0
200
400
600
800
1000
1200
t
-t
√t
-√t
a
b
c
x ⫫ y
x
y
x, y ∈ PDF

Random walk [application]

Non-random walks in monkeys and humans

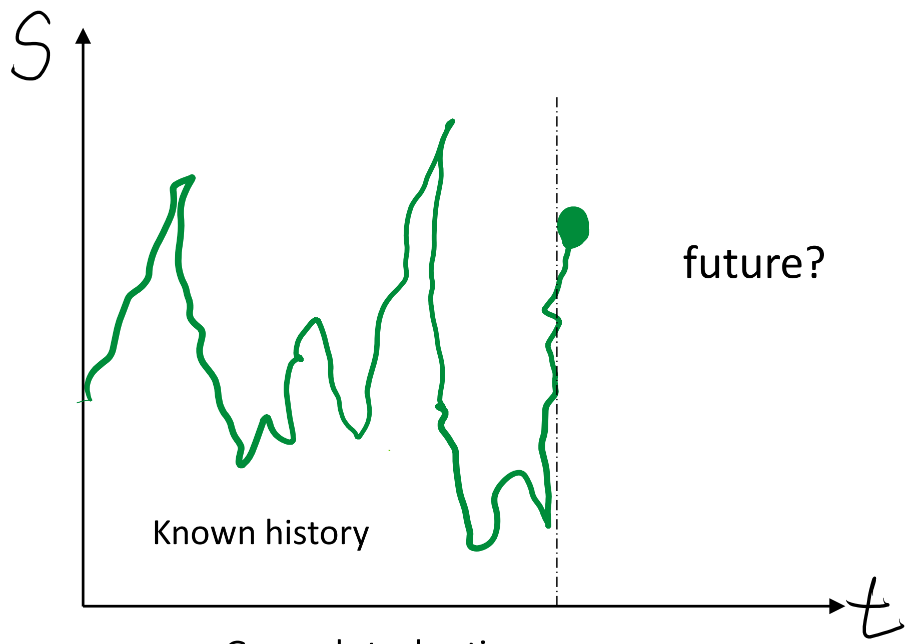
Denis Boyer, Margaret C. Crofoot and Peter D. Walsh

Published: **26 October 2011** | https://doi.org/10.1098/rsif.2011.0582

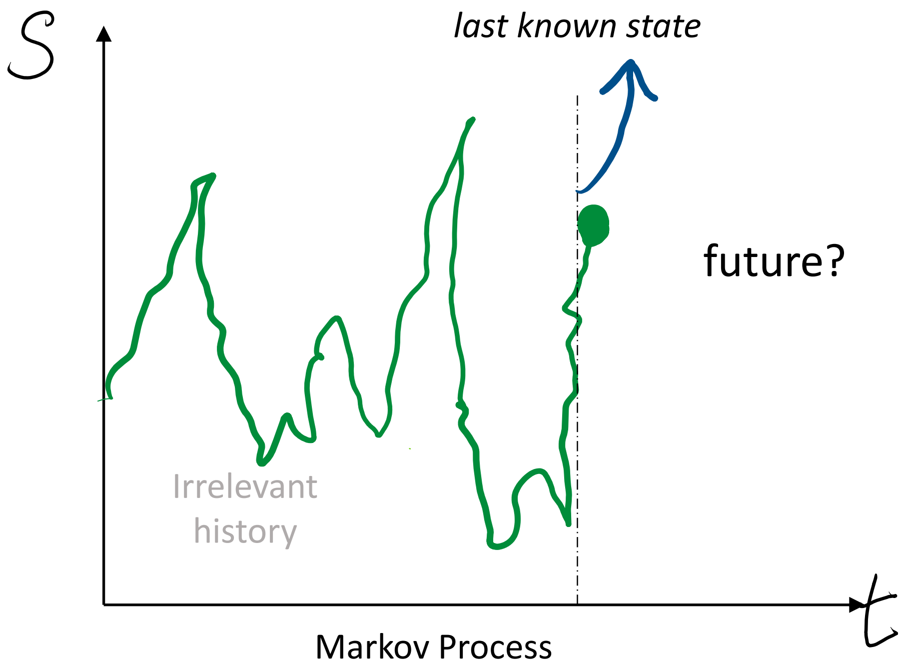
Abstract

Principles of self-organization play an increasingly central role in models of human activity. Notably, individual human displacements exhibit strongly recurrent patterns that are characterized by scaling laws and can be mechanistically modelled as self-attracting walks. Recurrence is not, however, unique to human displacements. Here we report that the mobility patterns of wild capuchin monkeys are not random walks, and they exhibit recurrence properties similar to those of cell phone users, suggesting spatial cognition mechanisms shared with humans. We also show that the highly uneven visitation patterns within monkey home ranges are not entirely self-generated but are forced by

Markov Chain [theory]



General stochastic process



Markov Process

$$P(X_{t+1} = j | X_1 = i_1, X_2 = i_2, \dots \dots X_t = i_t) = P(X_{t+1} = j | X_t = i_t)$$

Markov Chain [theory]

Transition probability matrix

$$P_{ij} = P(X_{t+1} = j |\, X_t = i)$$

$$A = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mn} \end{pmatrix}$$

$$\sum_{j \in S} P_{ij} = 1$$

$$Q_{ij} = P(X_{t+2} = j |\, X_t = i)$$

$$A^2 = \begin{pmatrix} q_{11} & \cdots & q_{1n} \\ \vdots & \ddots & \vdots \\ q_{m1} & \cdots & q_{mn} \end{pmatrix}$$

$$q_{ij} = \sum P\,(X_{t+1} = k\ |X_t = i)\ \mathrm{P}(X_{t+2} = j\ |X_{t+1} = k)$$

Finite markov chains have a transitional probability matrix.

Markov Chain [application]

"The PageRank value of a page reflects the chance that the random surfer will land on that page by clicking on a link. It can be understood as a Markov chain in which the states are pages, and the transitions, which are all equally probable, are the links between pages."

src:https://en.wikipedia.org/wiki/PageRank

MUSICAL MARKOV CHAINS

DIMA VOLCHENKOV*,‡ and JEAN RENÉ DAWIN†

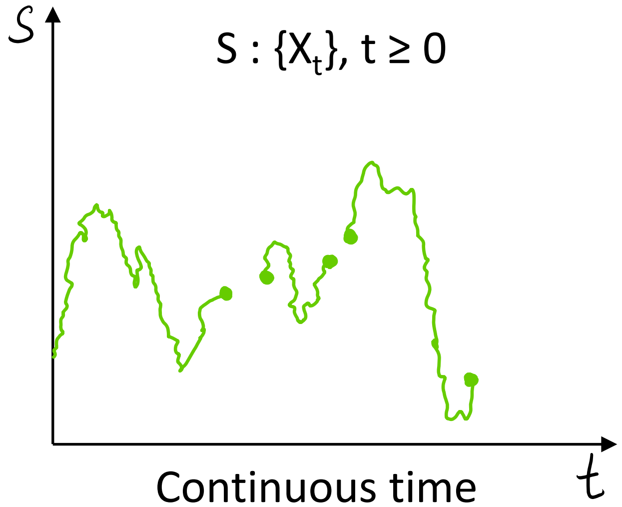
*Center of Excellence Cognitive Interaction Technology,
Universität Bielefeld, Postfach 10 01 31,
33501 Bielefeld, Germany

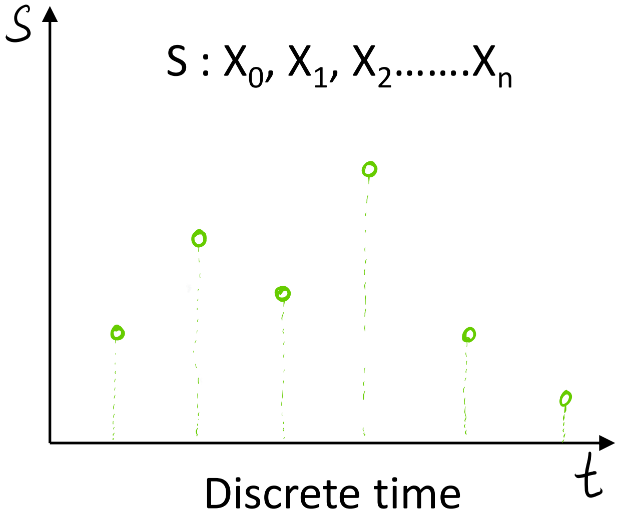
†Department of Physics, Universität Bielefeld,
Postfach 10 01 31, 33501 Bielefeld, Germany
‡volchenk@physik.uni-bielefeld.de

A system for using dice to compose music randomly is known as the musical dice game. The discrete time MIDI models of 804 pieces of classical music written by 29 composers have been encoded into the transition matrices and studied by Markov chains. Contrary to human languages, entropy dominates over redundancy, in the musical dice games based on the compositions of classical music. The maximum complexity is achieved on the blocks consisting of just a few notes (8 notes, for the musical dice games generated over Bach's compositions). First passage times to notes can be used to resolve tonality and feature a composer.

Keywords: Markov chains; entropy and complexity of music; first passage times.

Continuous time

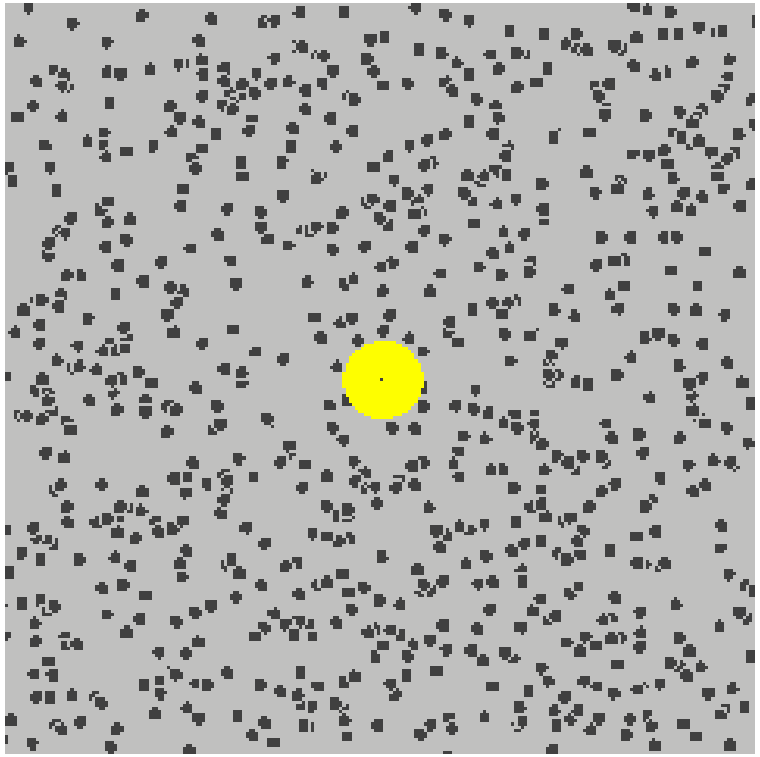




Define PDF?

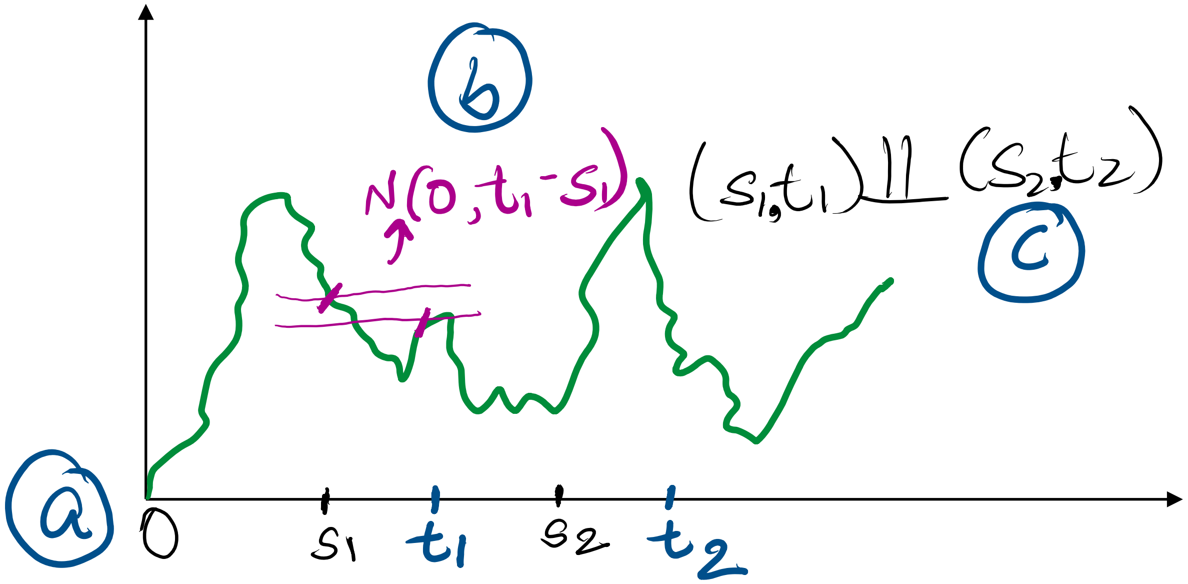
$$X_t - X_{t-1} = \begin{cases} 1, & pr = 0.5 \\ -1, & pr = 0.5 \end{cases}$$

Brownian Motion [theory]

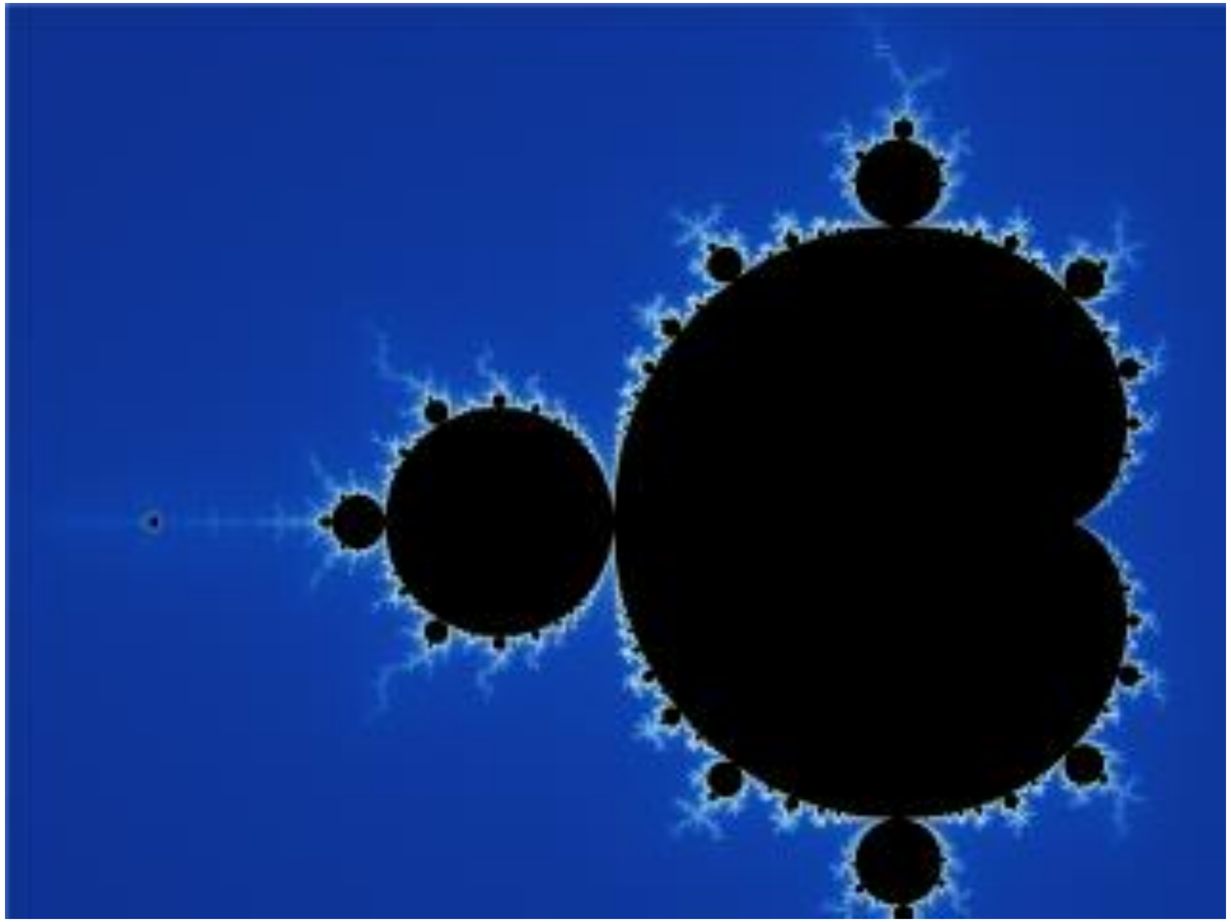


src:https://en.wikipedia.org/wiki/Brownian_motion

There is a probability distribution over set of continuous functions that..



Brownian Motion [application]



src:https://en.wikipedia.org/wiki/Mandelbrot_set

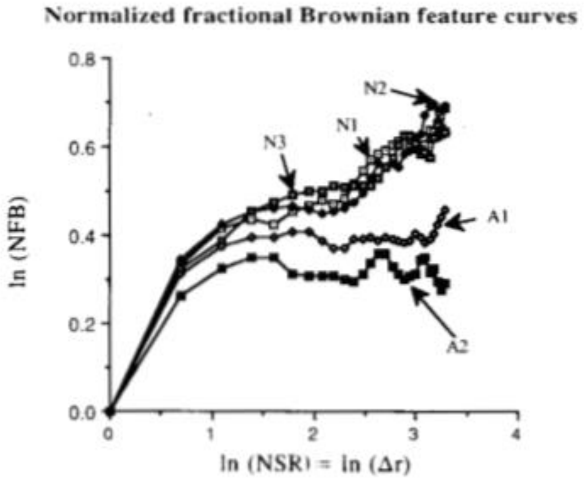
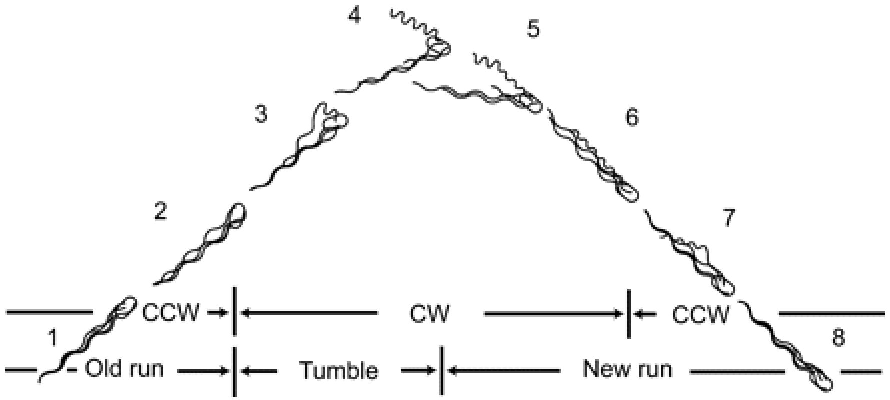


Fig. 5. *NFB* versus ln (*NSR*) for each area of interest in images *N*1, *N*2, *N*3, *A*1, and *A*2. Vertical axis is ln (*NFB*). Horizontal axis is log (*NSR*). The *NFB* is the normalized fractional Brownian motion feature vector. The *NSR* is the normalized scale range vector.

Chen, C-C., John S. DaPonte, and Martin D. Fox. "Fractal feature analysis and classification in medical imaging." *IEEE transactions on medical imaging* 8.2 (1989): 133-142.

Brownian Motion [application]



Darnton, Nicholas C., et al. "On torque and tumbling in swimming Escherichia coli." *Journal of bacteriology* 189.5 (2007): 1756-1764.

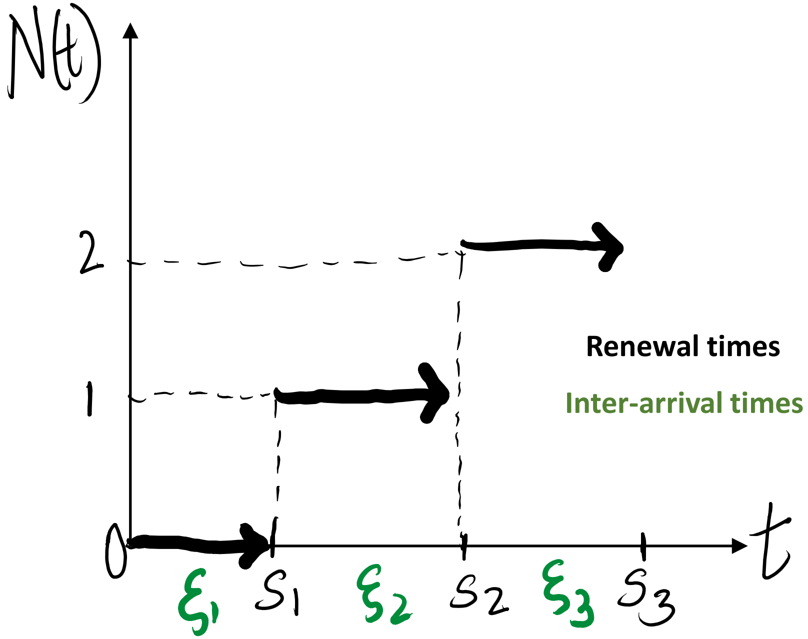
Molecular diffusion

Enzyme-substrate encounters

Suggested reading:http://www.doc.ic.ac.uk/~nd/surprise_95/journal/vol2/ykl/article2.html

Poisson Process [theory]

{X_t} stochastic process with *{S:0,1,2,3,….}* state space and *{t ≥ 0}* parameter space



Renewal times

Inter-arrival times

Poisson distribution is the limit on Binomial distr. :

$$P(X_t = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}, k = 0,1,2, \dots$$

for all (0,t]

$$pmf = \binom{n}{k} p^k (1-p)^{n-k}$$

Independence

Suggested tutorial:https://www.khanacademy.org/math/statistics-probability/random-variables-stats-library/poisson-distribution/v/poisson-process-2

Poisson Process [application]

© Journal of the American Statistical Association
June 1971, Volume 66, Number 334
Applications Section

Accident Rate Potential: An Application of Multiple Regression Analysis of a Poisson Process

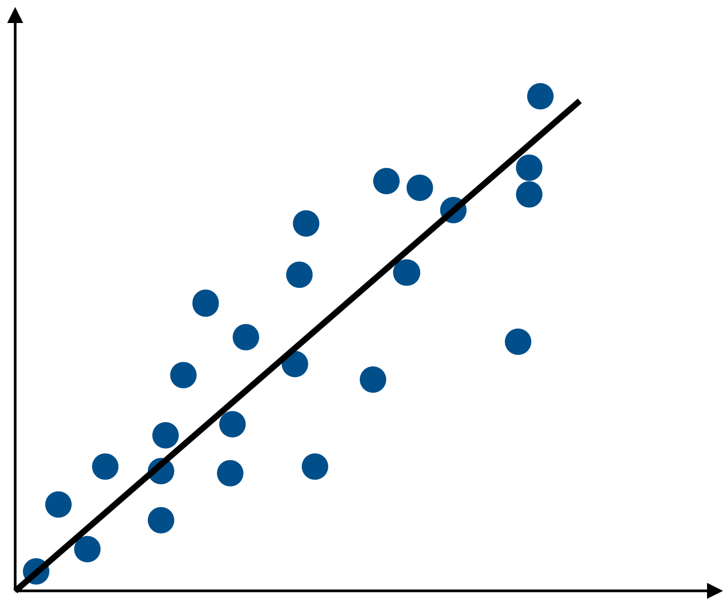
DONALD C. WEBER*

Various accident frequency models have appeared in the literature which predict the distribution of future accidents based on the number of past accidents. This article presents a method for deriving such distributions using several predictive criteria. It is assumed that an individual's accident experience is a Poisson process with the parameter a linear function of criterion variables. An iterative weighted least-squares procedure is used to solve the system of maximum likelihood equations required for estimating this parameter and a large sample test procedure is illustrated. The tenability of the model is viewed in the light of actual data.

driven and conviction history. Accepting the proposition that an individual driver's accident frequency over a short period of time follows (1.1), the purpose of this article is to demonstrate a method for estimating the parameter λ, the "accident rate potential" associated with the individual, as a function of k criteria.

2. THE DATA

Gaussian Process [theory]



$$f(x) = w^T x + \varepsilon$$

$D = \{(x_1,y_1), (x_2,y_2), (x_3,y_3)\ldots\ldots (x_n,y_n)$

Learn ‘w’

Maximum likelihood estimation

$P(D|w)$

Maximum a posteriori

$P(w|D)$

Once predictions done, ‘w’ is not useful

Why not model the predictions directly?

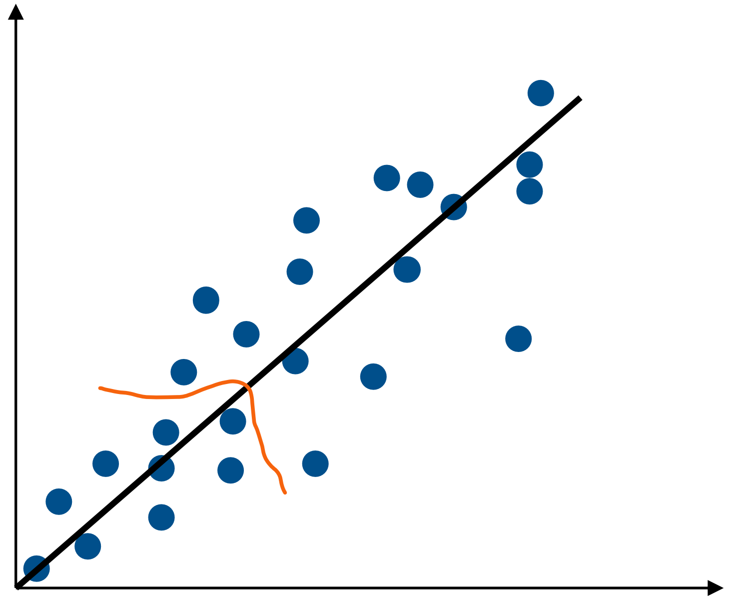
Gaussian Process [theory]

$$P(y|x,D) = \int P(y|x,w)\, P(w|D)\, dw$$

$$\frac{P(D|w)P(w)}{Z}$$

(Baye's rule)

$$P(y_i|x_i,w) = N(w^T x, \varepsilon^2 I)$$



But P(y|x,D) is Gaussian!!

Gaussian Process [theory]

But P(y|x,D) is Gaussian implies

$$P\begin{pmatrix} y1 \\ y2, [x1 \quad x2 \quad xn] \\ yn \end{pmatrix} \sim N(\mu, \Sigma)$$

Gaussian process is a probability distribution over infinite functions

$$\sum \cong \begin{pmatrix} 1 & \cdots & tr_{1n} \\ \vdots & \ddots & \vdots \\ tr_{n1} & \cdots & 1 \end{pmatrix}$$

Is the covariances of training data

GP regression : which training data most similar to test data?

Gaussian Process [theory]

Covariance matrix is the kernel i.e. positive semi-definite

Need a kernel function such that similar points get more weight

Kernel that mimics nearest-neighbor

Example: Radial basis function!

Feature vector

Kernel function

Gaussian distribution

Assume gaussian priors

Demo applet:http://www.tmpl.fi/gp/

Gaussian Process [application]

Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013

*GBD 2013 Risk Factors Collaborators**

Summary

Background The Global Burden of Disease, Injuries, and Risk Factor study 2013 (GBD 2013) is the first of a series of annual updates of the GBD. Risk factor quantification, particularly of modifiable risk factors, can help to identify emerging threats to population health and opportunities for prevention. The GBD 2013 provides a timely opportunity to update the comparative risk assessment with new data for exposure, relative risks, and evidence on the appropriate counterfactual risk distribution.

Methods Attributable deaths, years of life lost, years lived with disability, and disability-adjusted life-years (DALYs) have been estimated for 79 risks or clusters of risks using the GBD 2010 methods. Risk–outcome pairs meeting explicit evidence criteria were assessed for 188 countries for the period 1990–2013 by age and sex using three inputs: risk exposure, relative risks, and the theoretical minimum risk exposure level (TMREL). Risks are organised into a hierarchy with blocks of behavioural, environmental and occupational, and metabolic risks at the first level of the hierarchy. The next level in the hierarchy includes nine clusters of related risks and two individual risks, with more detail provided at levels 3 and 4 of the hierarchy. Compared with GBD 2010, six new risk factors have been added: handwashing practices, occupational exposure to trichloroethylene, childhood wasting, childhood stunting, unsafe sex, and low glomerular filtration rate. For most risks, data for exposure were synthesised with a Bayesian meta-regression method, DisMod-MR 2.0, or spatial-temporal Gaussian process regression. Relative risks were based on meta-regressions of published cohort and intervention studies. Attributable burden for clusters of risks and all risks combined took into account evidence on the mediation of some risks such as high body-mass index (BMI) through other risks such as high systolic blood pressure and high cholesterol.

Lancet 2015; 386: 2287–323

Published Online
September 11, 2015
http://dx.doi.org/10.1016/S0140-6736(15)00128-2

See Comment page 2235

*Collaborators listed at the end of the Article

Correspondence to:
Prof Christopher J L Murray, Institute for Health Metrics and Evaluation, 2301 5th Avenue, Suite 600, Seattle, WA 98121, USA
cjlm@uw.edu

Resources

Bhat, U. Narayan, and Gregory K. Miller. *Elements of applied stochastic processes*. Vol. 3. Hoboken^ eN. JNJ: Wiley-Interscience, 2002.

Rasmussen, Carl Edward. "Gaussian processes in machine learning." *Summer School on Machine Learning*. Springer, Berlin, Heidelberg, 2003.

MIT open courseware (18.S096 Fall 2013); Cornell (Cornell CS4780 SP17)