extreme value statistics

statistics and data analysis (chapter 11)

Björn Malte Schäfer

Graduate School for Fundamental Physics
Fakultät für Physik und Astronomie, Universität Heidelberg

June 25, 2016

outline

1. Markov-chains
2. statistical physics
3. extreme value statistics
4. copulas
5. summary

Markov-chains

- Markov-chains are random processes, where the outcome of the random experiment depends on previous outcomes
- example: chain bivariate Gaussian, where one of the variables is free and the other is set to the previous outcome
- only sensible **if** there's a nonzero correlation and off-diagonal parts in the covariance
- distribution $p(x, y)$

$$p(x, y) = \frac{1}{\sqrt{(2\pi)^2 \det C}} \exp\left(-\frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix}^t C^{-1} \begin{pmatrix} x \\ y \end{pmatrix}\right) \tag{1}$$

but set $y_n = x_{-1}$ and draw new x_n

Metropolis-Hastings as a Markovian process

- Metropolis-Hasting is an example of a Markov-chain
- the position of the chain (i.e. the current configuration) is a memory of all previous samples

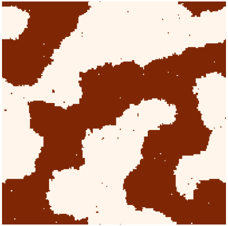
Metropolis-Hastings and the Ising model

- select a lattice site
- switch its configuration from up→down and from down→up
- compute change in energy associated with the switch of configuration
  - keep the flip if you go to a lower energy state
  - carry out a flip which needs energy ϵ with the probability $\exp(-\epsilon/(kT))$
- repeat

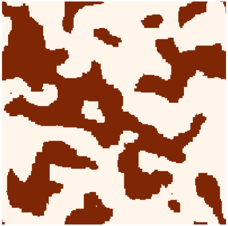
Metropolis-Hastings sampling

generates samples for the configurations at thermodynamic equilibrium and establishes a distribution lattice sites. there's a competition between randomness by thermal fluctuations and magnetisation.

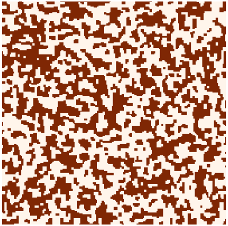
Ising-model in 2d: configurations at different T



low T



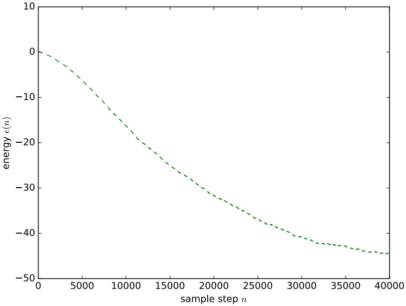
medium T



high T

- more random patterns at high T
- formation of large zones of equal magnetisation at low T

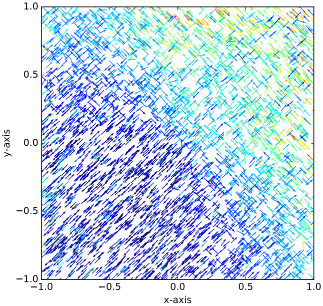
Ising-model in 2d: equilibration



equilibration

- the system finds a lowest energy state by equilibration

Ising-model in 2d: compass needles



compass needles

- configuration of compass needles with a dipolar interaction

statistical model of a polymer: analytic solution

- polymers contract under increasing temperatures: what's the reason for this?
- imagine a simple model of a polymer: it consists of n monomers
- each monomer has a long axis a and a short axis b
- total length:
$$l(i) = i \times a + (n - i) \times b \tag{2}$$
- if a tension σ is applied to the string of monomers, there's an energy change $\sigma \Delta l$ associated to a change of configuration
- a shortening of the chain of monomers is suppressed with the Boltzmann-factor
$$p \propto \exp\left(-\frac{\sigma \Delta l}{kT}\right) \tag{3}$$
at which the system can "borrow" thermal energy at temperature T
- $\Delta l = b - a$ for an exchange $b \rightarrow a$ somewhere in the chain

partition sums and thermodynamic potentials

- in fact, the polymer generates an **entropic** force
- change in internal energy $\mathrm{d}U$:

$$\mathrm{d}U = \underbrace{\frac{\partial U}{\partial S}}_{T}\mathrm{d}S + \underbrace{\frac{\partial U}{\partial l}}_{-\sigma}\mathrm{d}l \quad (4)$$

- internal energy $U(S,l)$ is a function of entropy and volume, but we can change the dependencies by Legendre-transform to e.g. T, l with the free energy $F(T,l)$ or to T, σ with the Gibbs enthalpy $G(T,\sigma)$
- partition sum counts all states weighted with the Boltzmann-factor

$$G(T,\sigma) = -kT\ln Z \quad \text{with} \quad Z = \sum_{\text{states}} \exp\left(-\frac{\sigma l(s)}{kT}\right) \quad (5)$$

with the string tension σ

partition sums and thermodynamic potentials

- binomial coefficient $\binom{n}{k}$ gives the number of possibilities to disperse i short polymers among $n - i$ long ones
- all states with the same number i of short polymers and $n - i$ long polymers have the same energy
- partition sum

$$Z = \sum_i \binom{n}{i} \exp\left(-\frac{\sigma l(i)}{kT}\right) \quad (6)$$

- substitution of $l(i)$ yields:

$$Z = \left(\exp\left(-\frac{\sigma a}{kT}\right) + \exp\left(-\frac{\sigma b}{kT}\right)\right)^n \quad (7)$$

  such that the n-particle partition function is the nth power of the 1-particle partition
- recover Hooke's law including a temperature dependence: $l = \mathrm{d}G/\mathrm{d}\sigma$ is a function of string tension σ and temperature T

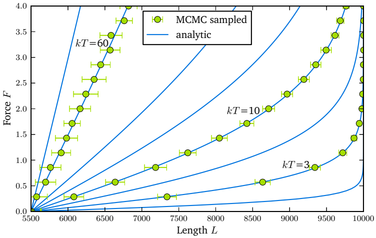
Metropolis-Hastings sampling

- select a monomer from the string
- switch its configuration from $a \rightarrow b$ or from $b \rightarrow a$
- compute change in energy associated with the switch of configuration
  - replace always a short by a long element, release of energy
  - replace a long element by a short one with a probability $\exp(-\sigma\Delta l/(kT))$
- repeat

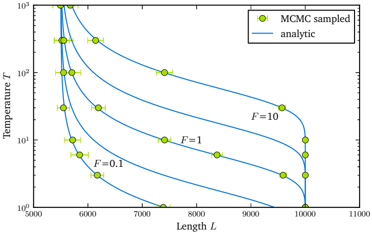
Metropolis-Hastings sampling

generates samples for the configurations at thermodynamic equilibrium and establishes a distribution of the chain length. there's a competition between randomness by thermal fluctuations and string tension.

relations between tension, length and temperature



force-length relation



temperature-length relations

- system shows the correct behaviour: increased restoring force for increasing length, and contraction under temperature increase
- credit: M. Kretschmer

marginalisation conditions and stat. dependence

- back to the Gaussian chain: the nonzero covariance between successive draws links the random events to each other
- depending on the dimensionality n of the Guassian one has constructed a Markovian proces with length n
- only if the covariance is diagonal, the process separates into independent random processes, then, the length of the Markovian process is 1 and the Gaussians are 1-dimensional
- there might be very unusual ways of linking the individual events
- for instance, one could draw n random numbers, order them by magnitude and keep the largest (or the smallest):

$$x_n \geq x_{n-1} \geq \ldots \geq x_2 \geq x_1 \tag{8}$$

  with the largest value x_+ and the smallest value x_-
- the events are not independent, because the choice of $x\pm$ depends on all other samples

maximum distribution

- distribution of the largest value x_+ in n draws from $p(x)\mathrm{d}x$
- cumulative distribution $P(x) = \int_{-\infty}^{x} \mathrm{d}x\, p(x)$: probability for a sample to be $< x$
- probability that n samples are $< x$: $P(x)^n$
- probability that at least one sample is $> x$ in n trials: $1 - P(x)^n$
- differentiate for the probability density

$$p_+(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(1 - P(x)^n\right) = nP(x)^{n-1}p(x) \tag{9}$$

minimum distribution

- distribution of the smalles value x_- in n draws from $p(x)\mathrm{d}x$
- complementary cumulative distribution $1 - P(x) = \int_x^\infty \mathrm{d}x\, p(x)$: probability for a sample to be $> x$
- probability that n samples are $> x$: $(1 - P(x))^n$
- probability that at least one sample is $< x$ in n trials: $1 - (1 - P(x))^n$
- differentiate for the probability density

$$p_-(x) = \frac{\mathrm{d}}{\mathrm{d}x}(1 - (1 - P(x))^n) = n(1 - P(x))^{n-1} p(x) \tag{10}$$

alternative derivation

- it is possible to derive the extreme value distribution in a marginalisation process
- (but using the cumulative distribution is much easier!)
- probability distribution for the largest samples requires that the other $n-1$ samples were each larger than the previous (if ordered)
- look at two samples x_1 and x_2 and construct

$$p_+(x_2) = \int_{x_2>x_1} \mathrm{d}x_1\ p(x_1)p(x_2) \tag{11}$$

  which is a marginalisation over the condition $x_2 > x_1$
- in analogy: look at 3 samples x_1, x_2 and x_3:

$$p_+(x_3) = \int_{x_3>x_2} \mathrm{d}x_2\ p(x_3) \int_{x_2>x_1} \mathrm{d}x_1\ p(x_2)p(x_1) \tag{12}$$

- ordering cuts off a part of the distribution and **correlates** the otherwise independent distribution

example: Gumbel-distribution

- start with the exponential probability density $p(x) = \exp(-x)$ with the cumulative distribution $P(x) = 1 - \exp(-x)$
- use a centralise variable $u = x - \ln(u)$ such that the expectation value is zero

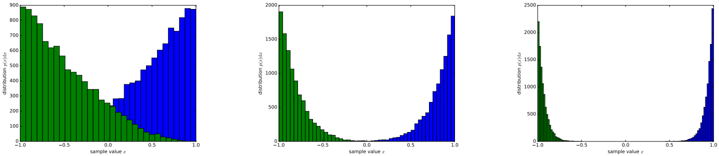
$$G(u) = P(x_+ - \ln u < u) = P(u + \ln u) = (1 - \exp(-u)/u)^n = \exp(-\exp(-u)) \tag{13}$$

- extreme value distribution

$$p_+(x) = \frac{\mathrm{d}}{\mathrm{d}x} \exp(-\exp(-x)) = \exp(-\exp(-x)) \exp(-x) \tag{14}$$

- heuristically, extreme value distributions look very similar
- there are three major types of extreme value distributions: Weibull, Gumbel and Frechet

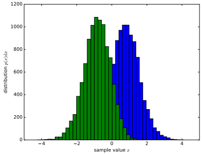
extreme value statistics from a uniform distribution



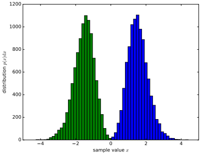
$n = 3$ $n = 10$ $n = 30$

- maximum (blue) and minimum (green) distributions based on the uniform distribution

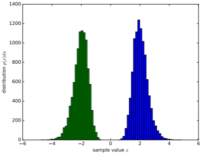
extreme value statistics from a uniform distribution



$n = 3$



$n = 10$



$n = 30$

- maximum (blue) and minimum (green) distributions based on the Gaussian distribution

bivariate processes and copulas

- a copula of a multivariate distribution is a generalisation of covariance, which is only exhaustive for Gaussian distributions (as a second moment, but of course there might be some nonzero, higher-order mixed moment)
- example: Gumbel's bivariate cumulative distribution

$$P(x,y) = \frac{1}{1+\exp(-x)+\exp(-y)} \tag{15}$$

- look at the edges $x \to \infty$ or $y \to \infty$

$$P_x(x) = P(x,\infty) = \frac{1}{1+\exp(-x)} \quad \text{and} \quad P_y(y) = P(\infty,y) = \frac{1}{1+\exp(-y)} \tag{16}$$

- use identities $x = P_x^{-1}(P_x(x))$ and $y = P_y^{-1}(P_y(y))$:

$$P(x,y) = \frac{1}{1+\exp(-P_x^{-1}P_x(x))+\exp(-P_y^{-1}(P_y(y))} \tag{17}$$

copula density

- define $u = P_x(x)$ und $v = P_y(y)$:

$$P(P_x^{-1}(x), P_y^{-1}(y)) = \frac{uv}{u + v - uv} = C(u, v) \tag{18}$$

- $C(u, v)$ is the cupola and lets you extrapolate the distribution $P(x, y)$ from the edges $P_x(x)$ and $P_y(y)$:

$$P(x, y) = C(P_x(x), P_y(y)) \tag{19}$$

- cupola density is the derivative of the cupola (which is a cumulative distribution)

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v) \tag{20}$$

- for Gumbel's bivariate distribution:

$$c(u, v) = 2\frac{uv}{(u + v - uv)^3} \tag{21}$$

- if two distributions are identical, $C(u, v) = uv$, so independent distributions have $c(u, v) = 0$

summary

- random processes with dependence:
    - Markov-chains
    - statistical systems
    - extreme values
- Markov-chains show correlations along a sequence of random numbers
- statistical systems are based on randomness: Ising-model for magnetisation, polymer-model for a string
- extreme value statistics introduce correlations through ordering