Entropies

Order / Disorder

from Time Series
Any method involving the notion of entropy, the very existence of which depends on the second law of thermodynamics, will doubtless seem to many far-fetched, and may repel beginners as obscure and difficult of comprehension.

Willard Gibbs
Graphical Methods in the Thermodynamics of Fluids (1906)
fundamental concept in thermodynamics and statistical mechanics (1850s – 1880s)

entropy → expression of the disorder, or randomness of a system

- macroscopically: \[ S = k_B \ln \Omega \quad [J/K] \]
  \( \Omega \) denotes number of microstates
  \( k_B \approx 1.38 \cdot 10^{-23} \quad [J/K] \)

- microscopically: \[ S = -k_B \sum_i p_i \ln p_i \]
  \( p_i = \frac{1}{\Omega} \) for microcanonical ensemble

phase transitions, entropy-driven order (Landau theory); adiabatic demagnetization; …
fundamental concept in information theory (1940-1950)

Entropy $\rightarrow$ amount of information needed to specify the full microstate of the system $X$ (Shannon entropy)

$$S(X) = -\sum_i p(x_i) \ln p(x_i)$$

Extensions and generalizations useful for time series analysis:

Rényi entropies $\rightarrow$ diversity, uncertainty, or randomness of a system

Kolmogorov-Sinai entropies $\rightarrow$ chaoticity of a system
entropy and Information

observing a system (measurement) is source of information

system with 2 states has maximum information content: 1 bit

system with 4 states has maximum information content: 2 bits

system with $M$ states has maximum information content:

$$I = \log_2 M$$
**entropy and Information**

measuring statistical events and average information gain

given a priori knowledge: 
$M$ events ($M$ system states) will appear (will be taken) with probabilities $\{p_i\}$, with $\sum_i p_i = 1$

measurement:
if you learn that event $j$ ($j \in M$) appeared (system state $j$ has been taken) then you will gain “average information” (through many measurement repetitions) as

$$I = -\sum_i p_i \log_2 p_i$$

(denoted as Shannon information)
entropy and Information

measuring statistical events and average information gain

eexample: coin flipping; head ($p_1$) or tail ($p_2$)?
equal probability for outcome: $p_1 = p_2 = 0.5$

measurement $\rightarrow$ head $\rightarrow$ information gain $I = 1$

and with probabilities:

$$I = - (0.5 \log_2 0.5 + 0.5 \log_2 0.5) = - (-0.5 - 0.5) = 1$$
linear methods for estimating entropies

recall: Fourier transform and Parseval’s theorem (see Linear Methods) with normalized power spectrum

$$\hat{P} = \sum_{k=1}^{N} |\hat{v}_k|^2 = 1$$

we can estimate the entropy $S$ of the relative spectral density as:

$$S = -\sum_{k=1}^{N} \hat{P}(k) \log_2 \hat{P}(k)$$

$S$ characterizes homogeneity of power spectrum:
- $S$ is minimum for line spectra (single Fourier component)
- $S$ is maximum for broad-band spectra (white noise)
- $S$ for chaotic dynamics? (looks like white noise)

need other methods to characterize entropy of chaotic dynamics
entropy and Information

**Given:**
- measured data follows some probability distribution
- transitions between successive data points occur with well-defined probabilities

**Qs:**
- if you have performed exactly one measurement, how much do you learn about the state of a system?
- if you have observed the entire past of a system, how much information do you have about future observations?

**As:**
can be found with generalized Rényi entropies
generalized entropies

order-$q$ Rényi entropies

... characterize the amount of information needed to specify the value of an observable with a certain precision if only the probability density is known that observable has value $x$.

Idea:
- partition phase space into $M$ disjoint hypercubes (boxes) of side length $\varepsilon$ (set of all these hypercubes is called a partition $P_\varepsilon$)
- estimate probability $p_j$ to find state $x$ in box $j$
- define order-$q$ Rényi entropy for partition $P_\varepsilon$ as:

$$\tilde{H}_q(P_\varepsilon) = \frac{1}{1-q} \ln \sum_{j=1}^{M(\varepsilon)} p_j^q$$
generalized entropies

order-\(q\) Rényi entropies

for \(q = 1\), we derive (L’Hôpital’s rule) the Shannon entropy:

\[
\tilde{H}_1(\mathcal{P}_\epsilon) = - \sum_j p_j \ln p_j
\]

which is the only Rényi entropy that is additive:

the Rényi entropy of a joint process is the sum of the entropies of the independent processes

(cf. mutual information)
generalized entropies  

example: Rényi entropy of a uniform distribution

given: probability density \( \mu(x) = \begin{cases} 1 & \text{if } x \in [0, 1] \\ 0 & \text{else} \end{cases} \)

partition the unit interval into \( N \) partitions of length \( \epsilon = \frac{1}{N} \)

we find: \( \tilde{H}_q(\epsilon) = \frac{1}{1-q} \ln \left( N \epsilon^q \right) = -\ln \epsilon = \ln N \)

- all order-\( q \) entropies are the same
  (due to the homogeneity of the uniform distribution)

- the better you resolve the real numbers by the partition, the more information you gain
generalized entropies and dimensions

relationship: order-\( q \) entropies and order-\( q \) dimensions

\[
\tilde{H}_q(\mathcal{P}_\epsilon) = \frac{\ln \sum_{j=1}^{M(\epsilon)} p_j^q}{1-q}
\]

\[
D_q := \lim_{\epsilon \to 0} \frac{\ln \left( \sum_{i=1}^{M(\epsilon)} p_i^q \right)}{(q-1) \ln(\epsilon)}
\]

- disjoint vs. non-disjoint partitioning

dimensions are the scaling exponents of the Rényi entropies computed for equally-sized partitions as functions of \( \epsilon \) and in the limit \( \epsilon \to 0 \).
generalized entropies

so far: entropies for static distributions

- can characterize attractor “as a whole”
- similar to dimension $\rightarrow$ no further gain of information
- no information about dynamics on the attractor

idea:

- consider entropies for transition probabilities
- characterize flow of information from small to large scales (typical for chaotic systems)
generalized entropies

- partition $m$-dimensional phase space into $M$ disjoint hypercubes (boxes) of side length $\varepsilon^m$

- let $p_{i_1,\ldots,i_m}$ denote the joint probability that state $X(t=1)$ is in box $i_1$, state $X(t=2)$ is in box $i_2$, etc., and that state $X(mt)$ is in box $i_m$

- define block-entropies of block-size $m$ as:

$$H_q(m, P_\varepsilon) = \ln \frac{\sum_{i_1,\ldots,i_m}^M p_{i_1,\ldots,i_m}^q}{1-q}$$
generalized entropies \hspace{2cm} \text{Kolmogorov-Sinai entropy}

for \( m \to \infty \), block-entropies are related to order-\( q \) entropies as:

\[
h_q = \sup_{\mathcal{P}} \lim_{m \to \infty} \frac{1}{m} H_q(m, \mathcal{P}_\varepsilon)
\]

\[
h_q = \lim_{m \to \infty} H_q(m + 1, \mathcal{P}_\varepsilon) - H_q(m, \mathcal{P}_\varepsilon)
\]

the supremum indicates: maximize over all possible partitions \( \mathcal{P} \), and

implies the limit \( \varepsilon \to 0 \)

\( h_0 \) is called \textit{topological entropy} (also abbreviated with \( K_0 \))

\( h_1 \) is called \textit{Kolmogorov-Sinai entropy} (also abbreviated with \( K_1 \))
generalized entropies

Kolmogorov-Sinai entropy

what do order-$q$ entropies and order-$q$ dimensions characterize?

topological entropy and Hausdorff dimension

- $h_0$ (or $K_0$) counts number of different orbits
- $D_0$ counts number of non-empty boxes

Kolmogorov-Sinai entropy and information dimension

- $h_1$ (or $K_1$) is a measure for the average rate of loss of information loss about a system state
- $D_1$ is a measure for a gain of information when findings a state in a given box
entropies from time series

entropies provide important information on topology of folding processes, disorder, chaoticity, and predictability

estimating order-q entropies from data is hard, particularly for high-dimensional systems (require more data than dimensions or Lyapunov exponents)

taking the limit $m \to \infty$ is difficult

box-counting (evaluate $m$-dimensional histograms) is most direct approach but turned out to be impractical

alternative ansatz: importance sampling
entropies from time series

idea:

- instead of using uniformly distributed partitions of phase space center partitions (boxes with fixed $\varepsilon$) on phase-space vectors

- use correlation sum (see Dimensions) to derive correlation entropy $K_2$
entropies from time series
with order-\(q\) correlation sum

\[
C_q(\epsilon) := \frac{1}{N} \sum_i \left( \frac{1}{N} \sum_j \Theta (\epsilon - |\vec{v}_i - \vec{v}_j|) \right)^{q-1}
\]

we find for \(q = 2\)

\[C_2(\epsilon) \propto \text{const.} \, \epsilon^{D_2}\]

in general, we have for \(q > 1\)

\[C_q(\epsilon) \propto \epsilon^{(q-1)D_q} \epsilon^{(1-q)H_q(m)}\]

if the systems exhibits a scaling region, we have \(\epsilon^{D_q} \approx \text{const.}\).

we can then find correlation entropy from

\[
h_q = \lim_{m \to \infty} H_q(m + 1, \epsilon) - H_q(m, \epsilon)
\]

\[= \lim_{m \to \infty} \ln \left( \frac{C(m, \epsilon)}{C(m + 1, \epsilon)} \right) =: K_2\]
entropies from time series
pros and cons of correlation entropy

- conceptually easy
- quickest to calculate

- requires existence of scaling region (independent on $\varepsilon$)
  (if you can’t find a scaling region do not apply this method!)
- needs lots of data
  (you lose $\varepsilon^{-h}$ neighbors when going from $m$ to $m+1$)

→
check robustness
constancy for a range of $\varepsilon$ values and embedding dimensions $m$
entropies from time series

\[ x_{n+1} = 1 - ax_n^2 + y_n \]
\[ y_{n+1} = bx_n \]

where
\[ a = 1.4; b = 0.3 \]

example: Hénon map

literature \((m \to \infty)\):
\[ K_2 \sim 0.33 \]
entropies from time series

difficult to identify scaling region

example: white noise

no constancy for range of $\varepsilon$ values
entropies from time series

example: EEG data

healthy subject

epilepsy patient seizure-free interval

epilepsy patient seizure
entropies from time series  
example: EEG data

Fundamentals of Analyzing Biomedical Signals  
Entropies
entropies

field applications

- number of data points \( \lim N \to \infty \) and \( m \to \infty \)
- data precision
  - adopt to requirement of small \( \varepsilon \)-neighborhood
- strong correlations in data (sampling interval)
  - use Theiler correction (see Dimensions)
- noise, filtering
  - similar impact as with Dimensions and Lyapunov exponents
- identifiable scaling region

what can go wrong?
entropies

- in general, we have: $K_{q'} \leq K_q$ for $q' > q$

- disorder, chaoticity of a system and type of the dynamics:
  - $K > 0$: chaos, unstable dynamics
  - $K = 0$: regular dynamics
  - $K = \infty$: noise

- average rate of loss of information due to action of nonlinearity

- prediction horizon:

$$T_p \approx \frac{-\ln(\rho)}{K}$$

where:

$\rho$ denotes accuracy of measurement (initial state)
Pesin’s identity

relationship between entropy and Lyapunov exponents

- entropy characterizes average rate of loss of information loss about a system state

- Lyapunov exponents characterize exponential divergence of initially close system states

Pesin’s identity:

\[ K_1 = \sum_{i, \lambda_i > 0} \lambda_i \]
Pesin’s identity

relationship between entropy and Lyapunov exponents

consistency checks for time-series analysis

estimate $K_1$ from sum over all positive Lyapunov exponents

note that $K_1 \leq \sum_i \lambda_i$

due to $K_{q'} \leq K_q$ for $q' > q$

we have $K_2 = \sum_{i, \lambda_i > 0} \lambda_i$

compare with $K_2$ estimate from correlation sum