Information transfer and entropy in large-dimensional systems

Peter Jan van Leeuwen
Data Assimilation Research Centre (DARC)
University of Reading

Entropy as a measure of uncertainty

- How much information is received when we observe a specific value of random variable x?
- Information is ‘degree of surprise’.
- A highly probable observation contains less information than a highly improbable one.
- So the information content h(x) depends on the pdf of x, and information has to be monotonically dependent on this pdf.
- If two events x and y are unrelated their information content should by additive: h(x,y) = h(x) + h(y) when p(x,y)=p(x)p(y).
- This forces us to assume h(x) = - log (p(x)).
- The average amount of information is given by

$$E[p(x)] = -\sum_x p(x) \log p(x)$$

Examples

- For a uniform pdf over the unit interval with probability intervals 1/N we find

$$E[p(x)] = -\sum_x p(x)\log p(x) = -\sum \frac{1}{N}\log\left[\frac{1}{N}\right] = \log N$$

- For a peaked pdf with probability 1 at one of the intervals 1/N and zero elsewhere we find

$$E[p(x)] = -\sum_x p(x)\log p(x) = 0$$

Prediction and relative entropy

- Before an actual prediction we know the system by its climatological or equilibrium pdf.
- So we need a measure of how far the prediction pdf is away from climatology.
- A useful measure is the relative entropy:

$$E[p|\mu] = -\int p(\psi)\log\left(\frac{p(\psi)}{\mu(\psi)}\right)d\psi$$

- Positive definite and temporal monotonic:
  $E\left[p(t_1)|\mu(t_1)\right] \geq E\left[p(t_2)|\mu(t_2)\right]$ for $t_2 > t_1$
- Jaynes (1963) has shown that this information entropy is equal to the thermodynamic entropy and as such is a natural extension for non-equilibrium thermodynamics.

Information flow I

- Consider the change in relative entropy for a variable x due to information from y, the Mutual Information of x and y:

$$MI[x,y] = E[x] - E[x|y] = \int p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y))}\right)dxdy$$

- Advantage is that the *MI* can be obtained from ensemble integrations
- Disadvantage is that when x and y are both related to z the *MI* can be high, while x and y are not physically coupled.

  An example is uncoupled processes that both react to the seasonal cycle. (Compare correlation and causality.)
- Note that the *MI* is symmetric in x and y.

Information flow II

- Define information flow from y to x as the relative entropy evolution of x for full system evolution minus the relative entropy evolution of x when y is kept constant:

$$T_{y\rightarrow x} = \frac{dE[p(x)]}{dt} - \left.\frac{dE[p(x)]}{dt}\right|_{y=constant}$$

Hence compare the the relative entropy evolution of this system:

$$\begin{aligned} dx &= f(x,y)dt \\ dy &= g(x,y)dt + d\beta \end{aligned}$$

to that of this system:

$$\begin{aligned} dx &= f(x,y)dt \\ dy &= 0 \end{aligned}$$

Information flow III

- Expressions for the relative entropy evolution can be obtained from the Liouville (or Fokker-Planck, Kolmogorov, equation):

$$\frac{\partial p}{\partial t} + \nabla_x \cdot (fp) + \nabla_y \cdot (gp) = \frac{1}{2}\nabla_y \cdot (Q\nabla_y p)$$

as

$$\begin{aligned} \frac{dE[p(x)]}{dt} &= -\frac{d}{dt}\int p(x,y)\log p(x)\ dxdy = \ldots \\ &= \int p(y|x)(f \cdot \nabla p(x))\ dxdy \end{aligned}$$

and

$$\left.\frac{dE[p(x)]}{dt}\right|_{y=constant} = \ldots = \int p(x,y)(\nabla_x \cdot f)\ dxdy$$

Information flow IV

- For the climate prediction problem, when the equilibrium pdf is Gaussian interesting relations can be derived, like

$$\frac{d}{dt}E[p(x)|\mu(x)] = -T_{y\to x} + \frac{1}{\sigma^2}K_{y\to x}$$

in which the energy transfer from y to x is equal to the time-rate-of-change of the energy in x:

$$K_{y\to x} = \frac{dK(p(x))}{dt} = \frac{d}{dt}\left(\frac{1}{2}\int x\cdot x\; p(x)\; dx\right)$$

(Work by Kleeman, Madja, Liang, Harlim)

Observation information I

- Kleeman (2007) applied ‘this’ to find the information impact of observations on predictions. However, the update of the model pdf by the observations before prediction was not taken into account. Using *MI on the full model state at observation time*:

$$
\begin{aligned}
MI[x,y] &= E[x] - E[x|y] = \int p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y))}\right) dxdy \\
&= \int p(y|x)p(x)\log\left(\frac{p(y|x)}{p(y)}\right)\, dxdy
\end{aligned}
$$

Observation information II

- The Mutual information is also equal to the mean of the entropy of p(x|y) relative to p(x):

$$
\begin{aligned}
MI[x,y] &= \int p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y))}\right) dxdy \\
&= \int p(y) \int p(x|y) \log\left(\frac{p(x|y)}{p(x)}\right)\ dxdy \\
&= \int p(y) E[p(x|y)|p(x)]\ dy
\end{aligned}
$$

So the value of the observation does not matter, as in the linear methods.

Observation information III

An ensemble representation gives:

$$p(x) = \frac{1}{N}\sum_i \delta(x - x_i)$$

$$\begin{aligned} MI[x,y] &\approx \frac{1}{N}\sum_i \int p(y|x_i) \log\left(\frac{Np(y|x_i)}{\sum_j p(y|x_j)}\right) dy \\ &\approx \frac{1}{2} + \log N + \frac{1}{NN_{obs}}\sum_i\sum_k \log\left(\sum_j p(y_{ik}|x_j)\right) \end{aligned}$$

Where we used also an ensemble of observations drawn from each $p(y|x_i)$ as:

$$p(y|x_i) = \frac{1}{N_{obs}}\sum_k \delta(y - y_{ik})$$

The point of all this is to show that we can extend the common linear ‘observation information measures’ to fully nonlinear ones.

Observation information IV

- The previous slides got us the *MI* for the full model state, so the full model pdf. This work can be extended to *local MI*, but that becomes more complicated.

- The problem is that a certain variable has no direct physical relation to a specific observation, so we need to evaluate the joint pdf of that model variable and the model equivalent of the observation.

- It might be doable, however…

And prediction…?

- Mutual information is the natural tool for data assimilation.
- But not so for prediction, where the evolution equations have to be exploited and the ‘information flow’ is used.
- How to combine them? What is the exact relation between Mutual Information and Information Flow?