Mutual information between successive reorientations

Subhaneil Lahiri

Harvard University,

May 3, 2011

Abstract

We show how mutual information can be used to describe the independence of successive reorientations

1 Reorientation sequences

sec:reoseq

As a worm navigates, it performs a sequence of turns. When turns occur sufficiently close to each other, they are grouped into a reorientation event. These reorientations have several characteristics, e.g. the types of turn of which it is composed, the difference in heading direction before and after, the duration of the run leading into it. We wish to know if the characteristics of one reorientation are independent of the characteristics of previous reorientations.

Consider a sequence of r successive reorientations. The values of a particular characteristic of these reorientations is an r-tuple of random variables: $(X_1, \ldots, X_r)$. We are asking whether or not $P(X_1, \ldots, X_r) = P(X_1) \ldots P(X_r)$.

2 Entropy and mutual information

ec:entropy

The **entropy** of a probability distribution is a measure of the lack of information we have about a random variable:

$$H(X) = \langle -\log P(X) \rangle . \tag{1}$$ eq:ent

It takes its minimum value of 0 when X can only take one value. It takes its maximum value of $\log n$ when X has a uniform distribution over n possibilities.

With several random variables, we can define a joint entropy from their joint probability distribution:

$$H(X_1, \ldots, X_r) = \langle -\log P(X1, \ldots, X_r) \rangle . \tag{2}$$ eq:jointent

It satisfies the bounds

$$\max_i H(X_i) \leq H(X_1, \ldots, X_r) \leq \sum_{i=1}^{r} H(X_i). \tag{3}$$ eq:entbounds

The lower bound is saturated when one of the variables is enough to determine the others. The upper bound is saturated when the X_i are independent:

$$P(X1,\dots,X_r) = \prod_{i=1}^{r} P(X_i) \qquad \Longrightarrow \qquad H(X_1,\dots,X_r) = \sum_{i=1}^{r} H(X_i). \tag{4}$$

eq:indent

We can define the following measure of (lack of) independence:

$$I(X_1,\dots,X_r) = \sum_{i=1}^{r} H(X_i) - H(X_1,\dots,X_r). \tag{5}$$

eq:mutinf

In the case $r = 2$, this is the **mutual information** between X_1 and X_2. For $r > 2$, there are many different generalisations of mutual information. This one is called the **total correlation** [1], or multiinformation. It has the properties:

- it vanishes if and only if the random variables are independent
- otherwise, it is positive.
- it is bounded from above by $\sum_{i=1}^{r} H(X_i) - \max_i H(X_i)$.

In our cases, the random variables, X_i, all have the same distribution, so the total correlation satisfies the bounds

$$0 \leq I_r(X_1,\dots,X_r) \leq (r-1)H(X). \tag{6}$$

eq:mutinfbou

We can define a normalised total correlation:

$$C_r = \frac{I_r}{(r-1)H_1}, \qquad 0 \leq C_r \leq 1. \tag{7}$$

eq:normmutin

The lower bound corresponds to complete independence. The upper bound corresponds to complete redundancy.

A Bias and standard error

sec:stderr

We will follow the approach of [2]. Our situation is slightly different from that one. As all the X_i have the same distribution, we will estimate $P(X)$ from the pooled data, rather that estimating the $P(X_i)$ separately. This means that our estimates may not satisfy the bounds, such as (6).

Let $p_{i_1 \dots i_r}$ denote the probability $P(X_1 = x_{i_1},\dots,X_r = x_{i_r})$ and $n_{i_1 \dots i_r}$ denote the number of corresponding r-tuples in the sample. We can estimate $p_{i_1 \dots i_r}$ with

$$q_{i_1 \dots i_r} = \frac{n_{i_1 \dots i_r}}{N}, \qquad N = \sum_{j_1 \dots j i_r} n_{j_1 \dots j_r}. \tag{8}$$

eq:tupprob

We can then estimate $p_j = P(X = x_j)$ with

$$q_j = \sum_{i_1 \dots i_r} \left(\frac{q_{i_1 \dots i_r}}{r} \sum_{a=1}^{r} \delta_{j,i_a} \right). \tag{9}$$ eq:singprob

From now on, we will use A to denote the estimate of $A(p)$ with p replaced by q and $\widehat{A} = A - \text{Bias}(A)$.

Our bias estimates are essentially the same as those of [2], except that the number of samples for H_1 is rN instead of N:

$$B_1 = \text{Bias}(H_1) = -\frac{\#b_1}{2rN}, \qquad B_r = \text{Bias}(H_r) = -\frac{\#b_r}{2N}, \tag{10}$$ eq:biasH

where $\#b_r$ is the number of non-empty bins. The bias estimate for I and C follow in the same way.

We can estimate the standard errors with

$$\text{Var}(A) = \sum_{i_1 \dots i_r} \left(\frac{\partial A}{\partial n_{i_1 \dots i_r}} \right)^2 \text{Var}(n_{i_1 \dots i_r}), \qquad \text{Var}(n_{i_1 \dots i_r}) \approx N q_{i_1 \dots i_r} (1 - q_{i_1 \dots i_r}). \tag{11}$$ eq:stderr

where the corrections to the last formula are lower order in N.

We find that

$$\begin{aligned} \frac{\partial q_{i_1 \dots i_r}}{\partial n_{j_1 \dots j_r}} &= \frac{\left(\prod_a \delta_{i_a, j_a}\right) - q_{i_1 \dots i_r}}{N}, & \frac{\partial B_r}{\partial n_{j_1 \dots j_r}} &= -\frac{B_r}{N}, \\ \frac{\partial q_i}{\partial n_{j_1 \dots j_r}} &= \frac{\frac{1}{r}\left(\sum_a \delta_{i, j_a}\right) - q_i}{N}, & \frac{\partial B_1}{\partial n_{j_1 \dots j_r}} &= -\frac{B_1}{N}, \end{aligned} \tag{12}$$ eq:dqbydn

which leads to

$$\begin{aligned} \frac{\partial H_r}{\partial n_{j_1 \dots j_r}} &= -\frac{\log q_{i_1 \dots i_r} + H_r}{N}, & \frac{\partial I_r}{\partial n_{j_1 \dots j_r}} &= -\frac{\left(\sum_a \log q_{j_a}\right) - \log q_{i_1 \dots i_r} + I_r}{N}, \\ \frac{\partial H_1}{\partial n_{j_1 \dots j_r}} &= -\frac{\frac{1}{r}\left(\sum_a \log q_{j_a}\right) + H_1}{N}, & \frac{\partial C_r}{\partial n_{j_1 \dots j_r}} &= \frac{\log q_{i_1 \dots i_r} H_1 - \frac{1}{r}\left(\sum_a \log q_{j_a}\right) H_r}{(r-1)N(H_1)^2}, \end{aligned} \tag{13}$$ eq:dhbydn

all of these formulae apply if you put hats on every capital letter.

References

58.1661265 [1] S. Watanabe, "Information theoretical analysis of multivariate correlation," *IBM J. Res. Dev.* **4** (January, 1960) 66–82.

.125..285R [2] M. S. Roulston, "Estimating the errors on measured entropy and mutual information," *Physica D Nonlinear Phenomena* **125** (Jan., 1999) 285–294, `SAO/NASA ADS:1999PhyD..125..285R`.