

Example:

§6 Information Decay in Stochastic Processes

Topic 1

As an illustrative example we shall now apply the general definition of relative information to stochastic processes.⁸⁾ Suppose we have a stationary stochastic process with a finite number of states s_i , and that the process occurs at discrete (integral) times $1, \dots, n, \dots$, at which times the transition probability from the state s_i to the state s_j is T_{ij} . The probabilities T_{ij} then form what is called a stochastic matrix, i.e. the elements are between 0 and 1 and $\sum_j T_{ij} = 1$ for all i . If at any time k the probability distribution over the states is $\{P_i^k\}$ then at the next time the probabilities will be $P_j^{k+1} = \sum_i P_i^k T_{ij}$.

In the special case where the matrix is doubly-stochastic, which means that $\sum_i T_{ij}$, as well as $\sum_j T_{ij}$, equals unity, and which amounts to a principle of detailed balancing holding, it is known that the entropy of a probability distribution over the states, defined as $H = -\sum_i P_i \ln P_i$, is a monotone decreasing function of the time⁹⁾. This entropy is, however, simply the negative of the information relative to the uniform measure.

One can extend this result to more general stochastic processes only if one uses the more general definition of relative information. For an arbitrary (non-stationary) process the choice of an information measure which is stationary, i.e. for which

$$(6.1) \quad a_j = \sum_i q_i T_{ij} \quad (\text{all } j)$$

leads to the desired result. In this case the relative information

$$(6.2) \quad I = \sum_i P_i \ln \frac{P_i}{\alpha_i}$$

~~Doubtful as measure~~ is a monotone decreasing function of time and constitutes a suitable basis for the definition of the entropy $H = -I$.
Insert I Proof in App.

One can further drop the requirement that the stochastic process be stationary, and even allow that the states ~~can be different at different times~~ ^{completely} there are different sets of states at each time n , $\{S_i^n\}$, ~~where the numbers depend on n~~ so that the process is now given by a sequence of matrices T_{ij}^n representing transition prob. at time n from state S_i^n to state S_j^{n+1} . In this case probability distributions change according to:

$$(6.3) \quad P_j^{n+1} = \sum_i P_i^n T_{ij}^n$$

~~if we then choose any~~ choose any time-dependent information measure which satisfies the relations

$$(6.4) \quad \alpha_j^{n+1} = \sum_i \alpha_i^n T_{ij}^n \quad (\text{all } i, n)$$

~~they~~ the information of a probability distribution is ^{again} monotone decreasing with time. All of these results are easily extended to the continuous case, and we see that the concept of relative information allows us to define entropy for quite general stochastic processes.

Part 1

1

As an ~~illustrative~~ example illustrating the usefulness of the concept of relative information we shall consider briefly stochastic processes.

Part 2

Note that this definition leads to the previous result for doubly-stochastic processes, since the uniform measure, $\alpha_i = 1$ (all i), is obviously stationary in this case.