

Re: Shannon's information theory

Source: <http://coding.derkeiler.com/Archive/General/comp.theory/2004-12/0641.html>

From: Ajoy K Thamattoor (ajoyk_at_cs.stanford.edu)

Date: 12/24/04

Date: Fri, 24 Dec 2004 13:51:37 -0800
To: Amir massoud Farahmand <sologen@yahoo.com>

> *Thomas B. wrote:*

>

>>*Hello.*

>>*I have a question about calculating the entropy of an integer*

>>*value (32 bit).*

>>

>>*Let's call the value x . x 's range is $0-(2^{32}-1)$.*

>>*I make a measure of x . I got 5 samples and x is*

>>*100 everytime.*

>>*I repeat this experiment to verify my results,*

>>*everytime x is 5 times 100.*

>>

>>*Therefore my mind tells me: no entropy.*

Basically, that is a determination you need to make. Is the sample set of 5 you have representative enough?

>>*But what about the formula? How should I set the probability?*

>>

>>*If I set p to $5/5 = 1$ then entropy is 0.*

>>*(5 times occurrence, 5 samples)*

>>*But this looks wrong, because x can be*

>>*every value from 0 to $2^{32}-1$.*

Well, your experiments seem to suggest otherwise. You seem to be winding up with a deterministic value of 100. A deterministic random variable has entropy 0.

Entropy, in spite of its elegant mathematical setting, is hard to use in practical situations precisely because of this problem. The probability distribution associated with a random variable is almost never known in advance, and needs to be estimated empirically. There are many ways to do this, two of the most commonly used mechanisms are maximum likelihood estimation (which just means using the proportions in the observed data as the probabilities) and bayesian estimation. ML estimation runs into problems when certain values just don't appear in

comp.theory: Re: Shannon's information theory

the observed data – the standard method assigns an arbitrary small probability (a smoothing probability) to such values instead of the value 0. There are very many different ways to smooth, and they are all ad-hoc.

Bayesian estimation assumes a prior distribution (ie., it just assumes X has some distribution based on prior global/local knowledge), and then it adjusts this prior based on the observed data. Here the obvious problem is choosing the prior, and, unfortunately, way too often the prior is chosen based on mathematical convenience rather than any real-world criteria.

And, finally, the very question of whether the random variable X obeys a probability distribution (or whether it is truly random) is hard to settle for many cases. Mostly, again for mathematical convenience, the assumption of non-randomness is imposed ad-hoc and a-priori.

Ajoy.