A Goodness Criterion for Probabilistic Estimates

Jeremy Thorpe, Robert McEliece

December 3, 2001

Contents

Abstract

In this paper, we examine the need for a criterion with which to judge an oracle which gives estimated probability distributions. Giving A small number of motivated axioms, we arrive at a single suitable measure. We also give an interpretation of this measure in terms of "specificity" and "accuracy". Finally, we give several applications of this criterion including identifying the better weatherman.

1 Introduction

There are a tremendous number of applications in the human world which require accurate probabilistic information. For example, the price of wheat is governed to a large extent on how probable a shortage is. Our plans for next Sunday may depend on the probability that it will snow in the mountains. It is important to be able to judge the accuracy of the source of such estimates.

In many cases our estimates are generated by automated algorithms which take measurable input such as temperature and pressure readings and use probabilistic models perhaps generated by historical patterns. In any case, we will assume that the inner workings of the algorithm are unknown. We will assume, however, that we can eventually know the outcome.

2 Criterion

Consider a random variable X with outcome space $\mathbb{X}$ and (possibly hidden) probability mass function p and estimated mass function q. Further, consider a penalty function D on the estimate q which satisfies the following axioms:

Axiom 2.1 *D can be evaluated in terms of the estimate q and (the outcome of) the random variable X. In other words, $D = D(q, X)$*

Axiom 2.2

$$q_1(X) = q_2(X) \Rightarrow D(q_1, X) = D(q_2, X) \tag{1}$$

Axiom 2.3

$$q* \triangleq \text{argmax}_q E_p(D(q, X)) = p \tag{2}$$

Axiom 2.4

$$q(X) = 1 \Rightarrow D(q, X) = 0 \tag{3}$$

We will show that the following theorem holds:

Theorem 2.5 $D(q, X) = c\log(q(X))$

Axiom 1 can be motivated by such situations as when the probabilty mass function p may not be known, or possibly even well-defined. It may not be well-defined simply because an experiment is in principle not repeataple (i.e. what will the weather be on December 18, 2001) or "unknown" because it cannot be efficiently calculated (i.e. what is the probabilty that a number selected randomly from 1 to $10^2 0$ will have fewer than 6 prime factors).

Axiom 2 is currently motivated by the need to simplify the problem by reducing the solution space to essentially a function of one variable. (However, this axiom may be replaced with a weaker axiom in the future–I will explain this.)

Axiom 3 says that if there is a true distribution p, we give the estimate $q* = p$ the smallest possible expected penalty.

Axiom 4 says that if you predicted the correct outcome with certainty, the penalty is 0, an uncontroversial axiom.

(For the sake of simplicity, we will take c to be 1 throughout this paper. We could also write an axiom that enforces this, but in general, the base of the log in other information theoretic quantiantities is considered arbitrary, and an arbitrary c in our formula is equivalent.)

2.1 Consequences

The quantity $E_p D(q, X)$ has a nice interpretation in terms of information theory. In particular, we will see how it can apparently be broken into terms of well-known probabilistic quantities.

$$E_p D(q, X) = E_p \log(q(X)) \tag{4}$$

$$= E_p(\log(p(X))) + E_p(\log(q(X)) - \log(p(X))) \tag{5}$$

$$= H(p) - D(p||q) \tag{6}$$

Where $D(p||q)$ is the "kullbach-leibler" distance from p to q.

3 applications

There are many immediate applications for this theory. In principle, it can be used to evaluate the accuracy of any source of probabilistic information. For example, we can compare the weathermen of different TV stations to see who is the the most accurate.

Another possibility is to use this measure to compare probabilistic algorithms which return estimates of some true posterior probability distribution, estimates which cannot be guaranteed to be accurate.