Email follow-up 1: Here is one more piece of input from the client: "As far as a proof, my follow up thought is that conventional linear integer scoring has a known error associated with it while the error associated with probabilistic scoring should be based on sample size. It seems like you should be able to estimate probabilistic error base on sample size and translate it to linear scored error, establishing how large a sample is need to beat conventional scoring."

Email follow-up 2: To make the argument that the accuracy of probabilistic scoring is a function of the accuracy of the P(E|C) and P(E) evidence probability terms, it will probably need to be based on information theory concepts. I did a quick search and found the abstract below. A bit Greek to me but I think I'm reading that Bayes error only differs by the weights imposed. Can this be flushed out to show that the information in the IRT items is preserved during the scoring process and only affected by the error in the Bayes evidence probability terms.

The second thought is to get our hands on a large dataset (I've already started making inquiries), subsample it at various samples sizes, compute the probabilistic evidence terms and measure the convergence of the error (assuming it does) to get a handle on sample size and magnitude of error in probabilistic evidence terms.

On Divergences and Informations in Statistics and Information Theory

Abstract:
The paper deals with the f-divergences of Csiszar generalizing the discrimination information of Kullback, the total variation distance, the Hellinger divergence, and the Pearson divergence. All basic properties of f-divergences including relations to the decision errors are proved in a new manner replacing the classical Jensen inequality by a new generalized Taylor expansion of convex functions. Some new properties are proved too, e.g., relations to the statistical sufficiency and deficiency. The generalized Taylor expansion also shows very easily that all f-divergences are average statistical informations (differences between prior and posterior Bayes errors) mutually differing only in the weights imposed on various prior distributions. The statistical information introduced by De Groot and the classical information of Shannon are shown to be extremal cases corresponding to alpha=0 and alpha=1 in the class of the so-called Arimoto alpha-informations introduced in this paper for 0

Published in: IEEE Transactions on Information Theory (Volume: 52 , Issue: 10 , Oct. 2006)

Page(s): 4394 - 4412

Date of Publication: 25 September 2006

INSPEC Accession Number: 9116315

DOI: 10.1109/TIT.2006.881731

Publisher: IEEE

ISSN Information: