

UNCERTAINTY, ENTROPY, VARIANCE AND THE EFFECT OF PARTIAL INFORMATION

JAMES V. ZIDEK AND CONSTANCE VAN EEDEN
University of British Columbia

Uncertainty about the value of an unmeasured real random variable Y is commonly represented by either the entropy or variance of its distribution. If it becomes known that Y lies in a subset A of the support of Y 's distribution, one might expect uncertainty about Y to decrease. In other words, one might expect the entropy and variance of Y 's conditional distribution given $Y \in A$ to be less than their counterparts for the unconditional distribution. Going further it might be conjectured that the uncertainty about Y would be greater given the knowledge that $Y \in B$ as compared with $Y \in A \subset B$.

We do not know whether these conjectures are correct. However, we give sufficient conditions in certain cases where they are true. In particular, when Y is normally distributed we can make considerable progress. For example, we show in the case that $A = [a, b]$ and Y normally distributed with mean η and variance 1, that the variance of the conditional distribution of Y given that $a \leq Y \leq b$ is less than that of the unconditional distribution, thereby confirming our intuitive reasoning in this case. This last example also shows that for this exponential family the variance is less than 1 for all $a < b$ and all η —a result that is not known among the experts on exponential families we consulted.

The only relevant thing is uncertainty—the extent of our own knowledge and ignorance. de Finetti (1970, Preface, pp. xi–xii)

1. Introduction

A search of the Science Citation Index for the years since 1989 lends support to de Finetti's contention. The keyword "uncertainty" yielded 29,386 documents. However, a quick scan of the abstracts for just the most cited documents, as well as standard statistical references reveal a variety of interpretations of this concept, leading us to paraphrase Basu (1975) on "information":

But, what is uncertainty? No other concept in statistics is more elusive in its meaning and less amenable to a generally agreed definition.

As with "information," an agreed on, operationally useful meaning of the term would be desirable. However, such a meaning seems elusive.

Bernardo and Smith (1994) describe uncertainty as "incomplete knowledge in relation to a specified objective." Frey and Rhodes (1996) echo that definition, saying that "uncertainty arises due to a lack of knowledge regarding an unknown quantity." In achieving their comprehensiveness, these definitions suffer from a lack of specificity.