Re: Independence Further Examined in PI

Source: http://sci.tech−archive.net/Archive/sci.stat.math/2005−01/0317.html

---

From: Osher Doctorow (*mdoctorow_at_comcast.net*)
Date: 01/13/05

```
Date: Thu, 13 Jan 2005 15:40:25 +0000 (UTC)
```

On 13 Jan 05 02:37:07 −0500 (EST), Osher Doctorow wrote:

>1) P(A−−>B)_I = P(AB) − P(A) + 1 = P(A)P(B) − P(A) + 1
>2) DEP(A,B) or DEPENDENCE(A, B) = P(AB) − P(A)P(B)
>Theorem. P(A−−>B) = P(A−−>B)_I + DEP(A,B)
>Proof. P(A−−>B)_I = P(A)P(B) − P(A) + 1, and DEP(A,B) = P(AB) −
>P(A)P(B), so their sum is P(AB) − P(A) + 1 = P(A−−>B). Q.E.D.
>Remark. The Theorem shows that PI (Probable Influence) decomposes
>into a dependent part DEP(A,B) which is plausibly "(positive)

In any branch of mathematics or science, an anomaly, paradox, or enormous difficulty is often signalled by the absence of a definition for the "opposite" of a quantity or quality. This is especially the case for "independence" in statistics, since although everybody wants to study "dependence", the places where "dependence" appears with qualifiers in statistics are almost entirely in inequalities rather than equations that define particular types of dependence just as they define statistical independence already.

Positive quadrant dependence of Lehmann in the late 1960s is arguably the best definition of a type of dependence, although negative quadrant dependence is also of considerable interest. The definition of positive quadrant dependence of random variables X and Y is:

1) F(x,y) > FX(x)FY(y)

although this is sometimes postulated of the pdfs rather than the cdfs. For general random sets, it generalizes to:

2) P(AB) > P(A)P(B)

Why doesn't anybody propose that since F(x,y) − FX(x)FY(y) > 0 on (1), then F(x,y) − FX(x)FY(y) can be taken as a "measure" or even variable measuring positive statistical dependence as having a magnitude rather than just an inequality? I have in fact done just than in my last posting. Why others don't do it may not have one answer only, but in my opinion the deepest Bayesian thinkers would be worried by assigning importance to subtraction rather than division even though one of the terms involves multiplication. The most common excuse is "lack of motivation", which really doesn't carry much logical weight.

One could argue that since = instead of > in (2) gives statistical independence, therefore "any deviation from = " reflects "dependence" and that this doesn't require any equation since "not equals" is the opposite of "equals". However, almost everybody uses the word "dependence" with the intuitive understanding that there are degrees of dependence, while there are no degrees of "independence". Otherwise, the word even modified would be very differently defined from other words in the academic and applied academic worlds, where an attempt is made to retain at least some intuitive properties.

Osher Doctorow