

Robert Owen

From: "Lana Hollady" <lhollady@charter.net>
To: "Lana Hollady" <lhollady@charter.net>
Sent: Wednesday, June 28, 2006 2:55 PM
Attach: Bayes Theorem Laver 06-11-06.doc; Donaldson-Sex-Offender-Risk.doc; Mossman-2006-Another-look-at-interpreting-risk.pdf
Subject: Bayes Theorem document

Evaluators,

I have asked Dr. Gary Laver (Statistics, Cal Poly) to help interpret the attached Mossman paper addressing Bayes Theorem that Lana distributed to you recently. His comments are attached. Dr. Donaldson also wrote a paper on Bayes Theorem and Lana has attached that as well (it has also been distributed previously).

What follows are comments from Karl Hanson about Dr. Laver's comments on the Mossman paper, the Donaldson document and Bayes Theorem. This e-mail is intended to be read after you read Dr. Laver's document.

I hope this helps you in understanding these statistical issues as they may come up on court testimony, especially Sacramento County.

Amy,

It is generally clear and should be helpful to the evaluators. I have read Mossman's and Doren's articles, but not Donaldson's. I have a copy of Donaldson's paper, but I have not been motivated to read it. Laver's summary of Mossman's article appears reasonable.

A couple of additional points.

There are two very different ways that base rate change between samples. One way concerns relatively uninformative features such as follow-up time, detection rates, or artificial oversampling of recidivists (e.g., MnSOST developmental sample, Dempster's thesis). With this type of variation, the observed recidivism rates per category should vary, but the likelihood ratios should be consistent across samples. It is plausible in these situations to use Bayes' theorem to "adjust" the expected recidivism rates to other samples with other base rates (e.g., shorter follow-up periods). The defining condition is that the factors influencing the recidivism rate are unrelated to the pre-existing risk level (and risk scores) of the sample under investigation.

Another way in which the base rates can vary is when the offenders are selected based on factors related to risk. When the variation in base rate is related to risk, then Bayes' theorem is not longer a valid way of estimating the estimated recidivism rates in new samples with different base rates. For example, it makes no sense to use likelihood ratios from the Static-99 developmental sample and the different base rate for older offenders because age is related to Static-99 scores.

When the risk measure captures all the variability in risk levels, the likelihood ratios will be consistent even with the risk levels in the samples are restricted (e.g., all high risk or all low risk). When the variation in recidivism rates is not fully captured by the risk measure, then the likelihood ratios will change across samples.

When the variability in the base rate is attributable to variation in the underlying risk scores, it is quite possible for the base rates to vary while the recidivism rates per category are consistent. In fact this pattern is expected if the risk instrument is valid, and it is this pattern that Doren believed he had found. What Mossman points out is what Doren actually found was a happy coincidence in which the risk scales did not fully capture the variation in

recidivism rates (likelihood ratios differed), but the nature of the differences were such that the average values were close to those predicted.

In short, Mossman suggests a way of describing the accuracy of risk prediction tools that is most detailed than the simple rank order indices we typically use (e.g., ROC areas). Specifically, he argues that we should not only look at the relative ranking (ROC area), but also look at how the predictive accuracy varies across risk levels (i.e., the shape of the ROC curve). The analyses he present demonstrate how the shape of the ROC curve can differ across studies, even when the total ROC area is consistent. Given that there is even variability in the size of the ROC areas across studies (significant Q value in Hanson & Morton-Bourgon, 2004 for Static-99), I am not surprised that the shape of the curves also changes across studies.

My bottom line: The commonly used actuarial measures (e.g., Static-99) are useful indicators of group patterns, but they do not capture all variability due to risk. Individuals (and groups) may have characteristics not measured by the instrument that increase or decrease their probability of recidivism.

I hope that is helpful.

Karl.

[Hanson, Karl]