UNSW FORENSIC SPEECH SCIENCE CONFERENCE 2012

The Likelihood Ratio goes to Monte Carlo: the effect of reference sample size on likelihood-ratio estimates.

phil rose

Assuming that the characteristics of an appropriate reference sample can be specified according to the defence hypothesis - for example "middle-aged Australian males with a broad accent", a question that is frequently raised, its uncertainty often implied as an argument against using a LR-based approach, is: how big should the reference sample be? The answer is the same as with all such statistical size-of-sample questions: it depends on your desired precision. The aim of this paper is to use a Monte Carlo approach to do some preliminary investigations into how the likelihood ratio varies with the size of the reference sample used to estimate it. To preserve a little reality in the face of the Monte Carlo approach, data is taken from a real case involving comparison of suspect and offender acoustic-phonetic F-pattern, where, unusually, the prior odds were also known.

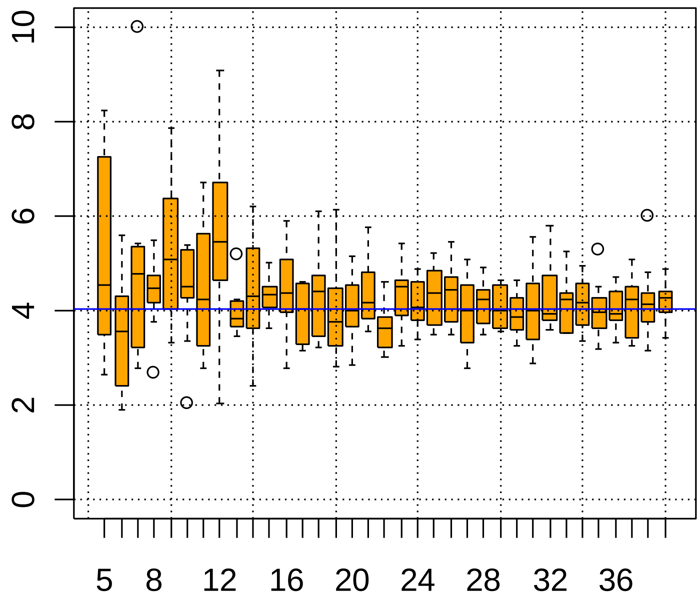

Figure 1. Effect on LR magnitude of number of speakers in reference sample. Vertical axis = univariate LR from /a:/ F2 values; horizontal axis = number of speakers in reference sample.

The test data are mean F1 thru F3 measurements from /a:/ in *car* (n = 13). A normal reference population is generated in R with parameters estimated from various sources on /a:/, and it is assumed that the LR obtained from this represents the true value. The size of the population is set suitably high, say 10,000. Samples are taken with replacement from the reference population increasing in size from 5 to 40, and a LR estimated. The LR from ten iterations using the same reference sample size are estimated and plotted against the number of reference sample speakers. Figure 1 is a box-and-whiskers plot showing the results from a comparison of /a:/ F2 values using Lindley's univariate normal LR, with a reference population = $N(\mu = 1400 \text{ Hz}, \sigma = 100 \text{ Hz}, n = 10000)$. (Each box extends from the lower quartile to the upper with the median dividing the box. The whiskers have a length not exceeding 1.5 * the interquartile range; outliers are then shown separately by a circle.)

It can be seen that the variance of the LRs gradually decreases with increasing reference sample speaker *n* and that its mean is very close (easily within +/- 2 LRs) to the 'true' LR (of about 4) by about 30 odd speakers. This is not surprising, of course, since it all has to do with sampling from a normal population. The mean and sd LR do not change much after about $n = 30$.

In my talk, I look at the results from the individual formants and from them combined with the Aitken & Lucy MVLR, and on the effect on the posterior.