Validation and Expert Judgement

Validating expert probabilistic assessments is an essential aspect of the expert judgment process, which has too often been neglected. The overall goal of rendering expert uncertainty a form of scientific input in support of evidenced decision making is poorly served by ignoring validation.

We may distinguish two types of validation, namely *in-sample* and *out-of-sample.* In case experts are combined using differential weights derived from performance on calibration variables, comparing performance on these same variables is in-sample validation. In sample validation is important as it drives the definition of performance metrics. The classical model involves two generic performance metrics, calibration (statistical accuracy) and informativeness. Other performance metrics could be contemplated, such as distance of the median to the true values, likelihood of the true values (Cooke et al, 2008), percent of realizations falling with the 90% central confidence region (Lin and Bier, 2008) and the expected relative frequency of observed outcomes (ERF, Flandoli et al, 2008). Inter comparisons of these and other weighting schemes in-sample has not received much attention and would be of great value.

Out-of-sample validation involves either (a) using performance on calibration variables to predict variables of interest, or if these cannot be observed, (b) splitting the calibration variables into a training test set, initializing the model on a training set and evaluating performance on the test set. The best way of performing this split is unknown and deserves high research priority.

Other forms of validation based on other performance metrics can be contemplated. The TU Delft SEJ database is available to try out ideas.

Cooke, R.M., ElSaadany, S., Xinzheng Huang, X. (2008) On the Performance of Social Network and Likelihood Based Expert Weighting Schemes, Special issue on expert judgment Reliability Engineering & System Safety, 93, 745-756, Available online 12 March 2007, Volume 93, Issue 5, May 2008.

Flandoli, F., Giorgi, E., Aspinall W. P., and Neri, A., (2011) Comparison of a new expert elicitation model with the Classical Model, equal weights and single experts, using a cross-validation technique. Reliability Engineering and System Safety, 96, 1292-1310. doi:10.1016/j.ress.2011.05.012.

Lin, Shi-Woei, and Bier, V.M. (2008) "A Study of Expert Overconfidence" Reliability Engineering & System Safety, 93, 775-777, Available online 12 March 2007. Volume 93, Issue 5.

Eggstaff, J.W., Mazzuchi,T.A. Sarkani, S. (2014) The Effect of the Number of Seed Variables on the Performance of Cooke's Classical Model, Reliability Engineering and System Safety 121 (2014) 72–82. DOI: 10.1016/j.ress.2013.07.015

Flandoli, F., Giorgi, E., Aspinall W. P., and Neri, A., (2011) Comparison of a new expert elicitation model with the Classical Model, equal weights and single experts, using a cross-validation technique. Reliability Engineering and System Safety, 96, 1292-1310. doi:10.1016/j.ress.2011.05.012.

Lin, Shi-Woei, and Bier, V.M. (2008) "A Study of Expert Overconfidence" Reliability Engineering & System Safety, 93, 775-777, Available online 12 March 2007. Volume 93, Issue 5.

Lin, Shi-Woei, Cheng, Chih-Hsing (2008) "Can Cooke's Model Sift Out Better Experts and Produce Well-Calibrated Aggregated Probabilities?" Department of Business Administration, Yuan Ze University, Chung-Li, Taiwan Proceedings of the 2008 IEEE IEEM

Lin, Shi-Woei, Cheng, Chih-Hsing (2009) "The reliability of aggregated probability judgments obtained through Cooke's classical model", Journal of Modelling in Management, Vol. 4 Iss: 2, pp.149 – 161

Shi-Woei Lin, Ssu-Wei Huang, (2012) "Effects of overconfidence and dependence on aggregated probability judgments", Journal of Modelling in Management, Vol. 7 Iss: 1, pp.6 – 22.

If interested in this topic please contact Roger Cooke. cooke AT rff.org