Ordinary meeting contribution

Paper under discussion:
D. J. Spiegelhalter, N. G. Best, B. P. Carlin and A. van der Linde: Bayesian measures of model complexity and fit

Contribution by:
Dr.Tech. Aki Vehtari, Helsinki University of Technology

Full address:
Laboratory of Computational Engineering
Helsinki University of Technology
P.O.Box 9400, FIN-02015, HUT, Finland
Aki.Vehtari@hut.fi

The authors mention that DIC estimates the expected loss, with deviance as loss function. I think this connection should be more emphasized. It should be remembered that the estimation of the expected deviance was Akaike's motivation for deriving the very first information criterion (AIC) (Akaike, 1973). In prediction and decision problems, it is natural to assess the predictive ability of the model by estimating the expected utilities, as the principle of rational decisions is based on maximizing the expected utility (Good, 1952) and the maximization of expected likelihood maximizes the information gained (Bernardo, 1979). It is often useful to use other than likelihood based utilities. For example, in classification problems it is much more meaningful for the application expert to know the expected classification accuracy, than just the expected deviance value (Vehtari, 2001). Given arbitrary utility function u, it is possible to use Monte Carlo samples to estimate $E_\theta[\bar{u}(\theta)]$ and $\bar{u}(E_\theta[\theta])$, and then compute an expected utility estimate as

$$\bar{u}_{\mathrm{DIC}} = \bar{u}(E_\theta[\theta]) + 2\big(E_\theta[\bar{u}(\theta)] - \bar{u}(E_\theta[\theta])\big),$$

which is a generalization of DIC (Vehtari, 2001).

The authors also mention the known asymptotic relationship of AIC to cross-validation. Equally important is to note that the same asymptotic relationship holds also for NIC (Stone, 1977, equation 4.5). The asymptotic relationship is not surprising, as it is known that cross-validation can also be used to estimate expected utilities with Bayesian justification (Bernardo & Smith, 1994, chap. 6; Vehtari, 2001; Vehtari & Lampinen, 2002a). Below some main differences between the CV and DIC are listed. See Refs. (Vehtari, 2001; Vehtari & Lampinen, 2002b) for full discussion and empirical comparisons. CV can use full predictive distributions. In the CV approach, there are no parametrization problems, as it deals directly with predictive distributions. CV estimates the expected utility directly, but it can also be used to estimate the effective number of parameters if desired. In the CV approach, it is easy to estimate the distributions of the expected utility estimates, which can for example be used to automatically determine if the difference between two models is "important". The importance-sampling leave-one-out CV (Gelfand et al., 1992; Gelfand, 1996) is computationally as light as DIC, but it seems to be numerically more unstable. k-fold-CV is very stable and reliable, but on the other hand it requires k times more computation time to use. k-fold-CV can also handle finite range dependencies in the data. For example, in the Six-cities study, the wheezing statuses of a single child at different ages are not independent. DIC, which assumes independency, under-estimates the expected deviance. In k-fold-CV it is possible to group the dependent data and handle independent groups and thus get better estimates (Vehtari, 2001; Vehtari & Lampinen, 2002b).

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov, & F. Csaki (eds.), *Second International Symposium on Information Theory*, (pp. 267–281). Budapest: Academiai Kiado. Reprinted in Kotz, S. and Johnson, N. L., editors, (1992). *Breakthroughs in Statistics Volume I: Foundations and Basic Theory*, pp. 610–624. Springer-Verlag.

Bernardo, J. M. (1979). Expected information as expected utility. *Annals of Statistics*, 7(3), 686–690.

Bernardo, J. M., & Smith, A. F. M. (1994). *Bayesian Theory*. John Wiley & Sons.

Gelfand, A. E. (1996). Model determination using sampling-based methods. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (eds.), *Markov Chain Monte Carlo in Practice*, (pp. 145–162). Chapman & Hall.

Gelfand, A. E., Dey, D. K., & Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (eds.), *Bayesian Statistics 4*, (pp. 147–167). Oxford University Press.

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 14(1), 107–114.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 44–47.

Vehtari, A. (2001). *Bayesian Model Assessment and Selection Using Expected Utilities*. Dissertation for the degree of Doctor of Science in Technology, Helsinki University of Technology. Available also online at `http://lib.hut.fi/Diss/2001/isbn9512257653/`.

Vehtari, A., & Lampinen, J. (2002a). Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Computation*, 14(10), 2439–2468.

Vehtari, A., & Lampinen, J. (2002b). Cross-validation, information criteria, expected utilities and the effective number of parameters. In preparation.