Including Model Selection Uncertainty in Estimates of Precision

or

Making Inferences From More Than a Single Model

Unconditional Estimates of Variances and Standard Errors

The precision of an estimator should ideally have 2 variance components: (1) the conditional sampling variance, given a model $\left(\widehat{\text{var}}(\hat{\theta}_i \mid M_i) \right)$ and (2) variation associated with model selection uncertainty. Buckland et al. (1997) provide an effective method to estimate an estimate of precision that is not conditional upon a particular model. Assume that the scalar parameter θ is in common to all models considered (e.g., ϕ_4, or N, or S_2). This will often be the case for our full set of *a priori* models, and is always the case if the objective is prediction. [If our focus is on a model structural parameter that appears only in a subset of our full set of models, then we must restrict ourselves to that subset in order to make the sort of inferences considered here about the parameter of interest.]

From Buckland et al. (1997) we will take the estimated unconditional var($\hat{\theta}$) as

$$\widehat{\text{var}}(\hat{\theta}) = \left[\sum_{i=1}^{R} \hat{w}_i \sqrt{\widehat{\text{var}}(\hat{\theta}_i \mid M_i) + (\hat{\theta}_i - \hat{\theta}_a)^2} \right]^2,$$

where,

$$\hat{\theta}_a = \sum_{i=1}^{R} \hat{w}_i \hat{\theta}_i$$

and the w_i are the Akaike weights (Δ_i) scaled to sum to 1. The subscript *i* refers to the i^{th} model. θ_a is a weighted average of the estimated parameter over *R* models (*i* = 1, 2, ..., *R*). This estimator of the *unconditional* variance is clearly the sum of 2 components: the conditional sampling variance var($\hat{\theta}_i \mid M_i$) and a term for the variation in the estimates across the *R* models $(\hat{\theta}_i - \hat{\theta}_a)^2$. The square root of these terms is then merely weighted by the w_i. This approach is very useful. Obviously, the estimated conditional

$$\widehat{\text{se}}(\hat{\theta}) = \sqrt{\widehat{\text{var}}(\hat{\theta})}.$$

We expect even better approaches to be discovered over the next few years.

The approach of Buckland et al. (1997) entails an assumption of perfect pairwise correlation, ρ_{ih}, of $\hat{\theta}_i - \theta_a$ and $\hat{\theta}_h - \theta_a$ for all $i \neq h$ (both i and h index models). Such pairwise correlation of $\rho_{ih} = 1$ is unlikely, however, it will be high. The choice of a value of $\rho_{ih} = 1$ is conservative in that $\text{var}(\hat{\theta}_a)$ will tend to be too large if this assumption is in error.

Model Averaging

Sometimes there are several models that seem plausible, based on the QAIC_c values. In this case, there is a formal way to base inference on more than a single model. This entails a weighted average of the estimates of a parameter for R models. Akaike weights are a natural to use (alternative weights can come from estimates of model selection frequencies, based on the bootstrap). Again, we assume that the parameter θ is the same across the models (or that only a subset of models containing the parameter of interest is considered). Again, define the estimator

$$\hat{\theta}_a = \sum_{i=1}^{R} \hat{w}_i \hat{\theta}_i$$

and in this case, $\hat{\theta}_a$ is the parameter of interest. The estimator of the unconditional variance is the same as that given above.

Unconditional Confidence Intervals

The matter of a $(1 - \alpha)100\%$ unconditional confidence interval is now considered. The simplest such interval is given by the end points

$$\hat{\theta}_i \pm z_{1-\alpha/2}\, \widehat{\text{se}}(\hat{\theta}_i),$$

where $\widehat{\text{se}}(\hat{\theta}_i) = \sqrt{\widehat{\text{var}}(\hat{\theta}_i)}$.

Here, the confidence interval is set around a single $\hat{\theta}$ or a model averaged estimate $\hat{\theta}_a$. When there is no model selection then an interval, conditional on Model i is

$$\hat{\theta}_i \pm t_{df,1-\alpha/2}\, \hat{\text{se}}(\hat{\theta}_i \mid M_i),$$

where it is clear what the degrees of freedom (*df*) are for the *t*-distribution.

When model selection is done, one needs an unconditional confidence interval. If the degrees of freedom df_i for the estimator $\hat{\text{var}}(\hat{\theta}_i \mid M_i)$, then for generally small degrees of freedom use the interval

$$\hat{\boldsymbol{\theta}}_i \pm \mathbf{z}_{1-\alpha/2}\, \hat{\mathbf{a}}\mathbf{se}(\hat{\boldsymbol{\theta}}_i)$$

where the adjusted standard error estimator is

$$\hat{\mathbf{a}}\mathbf{se}(\hat{\boldsymbol{\theta}}_a) = \sum_{i=1}^{R} \hat{w}_i \sqrt{\left(\frac{t_{df_i,1-\alpha/2}}{z_{1-\alpha/2}}\right)^2 \hat{\mathbf{v}}\mathbf{ar}(\hat{\boldsymbol{\theta}}_i \mid M_i) + (\hat{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_a)^2}\ .$$

In cases where $\hat{\theta}_i \pm z_{1-\alpha/2}\, \hat{\text{se}}(\hat{\theta}_i)$ is not justified by a normal sampling distribution (as judged by the conditional distribution of $\hat{\theta}_i$), intervals with improved coverage can be based on a transformation of $\hat{\theta}_i$ if a suitable transformation is known (e.g., the log and logit transforms).

Profile Likelihood Intervals

A general alternative when there is no model selection is the profile likelihood interval approach. We suggest here an adaptation of that approach that widens the likelihood interval to account for model selection uncertainty. Let the vector parameter $\underline{\theta}$ be partitioned into the component of interest, θ, and the rest of the parameters, denoted here as $\underline{\gamma}$. Then the profile likelihood, as a function of θ_i (the subscript denotes the model used) for model M_i is given by

$$\mathcal{PL}(\theta_i \mid \underline{x}\ , M_i) = \max_{\underline{\gamma}_i \mid \theta_i} \left[\mathcal{L}(\theta_i, \underline{\gamma}_i \mid \underline{x}\ , M_i)\right];$$

almost always $\mathcal{PL}(\theta_i \mid \underline{x}\ , M_i)$ has to be computed numerically. We define a profile deviance as

$$\mathcal{PD}(\theta_i) = 2\left[\mathcal{PL}(\hat{\theta}_i \mid \underline{x}\ , M_i) - \mathcal{PL}(\theta_i \mid \underline{x}\ , M_i)\right] .$$

The large sample profile likelihood interval ignoring model selection uncertainty is the set of θ_i that satisfy the condition $\mathcal{PD}(\theta_i) \leq \chi^2_{1,1\text{-}\alpha}$. Here, $\chi^2_{1,1\text{-}\alpha}$ is the upper $1 - \alpha$ percentile of the central chi-squared distribution on 1 *df*. This interval is approximately a $(1 - \alpha)100\%$ confidence interval.

The interval can be adjusted (widened) for model selection uncertainty: the set of all θ_i that satisfies

$$\boldsymbol{\mathcal{PD}(\theta_i) \leq \left[\frac{\hat{\text{v}}\text{ar}(\hat{\theta})}{\hat{\text{v}}\text{ar}(\hat{\theta}_i|M_i)}\right] \chi^2_{1,1\text{-}\alpha}}.$$

Literature Cited

Buckland, S. T., K. P. Burnham, and N. H. Augustin. 1997. Model selection: an integral part of inference. Biometrics 53:603-618.