STA721

TakeHome

The paper by Berry et al (2009) Clinical Trials 6:28-41 presents an approach to Meta-Analysis using Bayesian Model Averaging. The data consist of summaries from 22 studies (a row in the data file) that explore the relationship between vitamin E doses and all-cause mortality. The paper focuses on 4 functions for how vitamine E dose enters the mean: 1) no effect; 2) a quadratic-linear spline; 3) a linear effect; and 4) a quadratic does effect. While several other covariates are entertained in addition to the dose of vitamen E, their main results are based on the 4 models in vitamin E with uniform prior probabilities over the 4 models.

This BMA approach has been critiqued by Greenland (2009) Clinical Trials 6:50-51 who argues that "point null hypotheses" of no effect should not be considered and that results are driven by the prior probability (0.25) placed on the null model.

The response for the meta-analysis is the empirical log-odds ratio:

$$d_i = log\left(\frac{\hat{p}(D=1 \mid T_i = E_i)/(1-\hat{p}(D=1 \mid T_i = E_i))}{\hat{p}(D=1 \mid T_i = C_i)/(1-\hat{p}(D=1 \mid T_i = C_i))}\right)$$

where $\hat{p}(D=1 \mid T_i = E_i) = n_{iDE}/n_{iE}$, n_{iDE} is the number of deaths in the treated group, n_{iE} is the total number in the treated group, $\hat{p}(D=1 \mid T_i = U_i) = n_{iDU}/n_{iU}$, n_{iDU} is the number of deaths in the untreated group, n_{iU} is the total number in the untreated group (controls)

Asymptotically

$$d_i \overset{\text{ind}}{\sim} \mathsf{N}(\mu_i, \sigma^2 s_i^2) \tag{1}$$

$$s_i^2 = \frac{1}{n_{iDE}} + \frac{1}{n_{iE} - n_{iDE}} + \frac{1}{n_{iDU}} + \frac{1}{n_{iU} - n_{iDU}} \tag{2}$$

where if the model is correct, $\sigma^2 = 1$ (i.e. no over or under dispersion relative to the binomial model). This would require Weighted Regression to find MLEs or the likelihood, $\boldsymbol{\mu} \in C(\mathbf{X})$.

Read the paper and discussion. The objective of this TakeHomw is to reanalyze the Vitamin E data from a Bayesian perspective using both model averaging and from a single "rich" model that captures important uncertainties described in the paper and comment. Create credible/confidence intervals (ideally graphs) to show the effect of vitamin E on all-cause mortality after adjusting for (any/all) covariates.

Your write-up should provide: a (1) (brief) introduction to the problem; (2) the methods/models that you have used in enough detail that someone with the statistical background of this course could reproduce your model without having the code (i.e models, priors, discussion of hyperparameters; (3) results (discuss similarities/differences of frequentist intervals, BMA and other models, sensitivity to error/model/prior assumptions, does vitamin E supplementation appear to be harmful or beneficial (at all or at specific doses); and (4) overall discussion/summary and recommendations. Do you think that BMA is appropriate here or

do you recommend a different model? Quantitatively how sensetive are results to modelling assumptions at all stages?

Prepare a report using R/Rmarkdown or related of up to two pages. Where possible illustrate main conclusions with figures (up to 2 in the main body)

You do not need to describe all of the exploratory work that you do or wrong turns, but include the most important points in the body. The JAGS model or other R code may be included in a supplemental appendix.

A few things to consider (but you do not need to address point-by-point):

- To start, can you replicate any of the figures in the paper?
- Are there any studies that are outliers or violations of the normality assumption? What do you think is the best way to address this if this is an issue?
- Is there evidence of over/under dispersion? How sensitive is this to the prior? Is a "default" reference prior approriate here?
- Which assumptions affect the results the most? (using weights, model for likelihood, form for how vit E enters the mean function, priors on coefficients, priors on models)?
- What is the justification for your prior(s)? If you simulate data based on generating from the prior/model do the outcomes seem plausible? (what does the prior say about probabilities instead of log-odds ratios?)
- How could validate the model? How would you create validation/test data? Is out-of-sample prediction relevant to effect quantification?
- for energetic students, could you fit this with logistic regression and if so do conclusions change?
- In writing-up results, be sure to not only address questions qualitative, but quantify estimates with point estimates and uncertainty intervals, with interpretations.