Bayesian model selection in meta-analysis

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

Bohua Chen Christian Röver and Tim Friede

University Medical Center Göttingen, Germany

Meta-regression model

Let y_i denote the observed value in the ith study and θ_i the corresponding true parameter. We assume that

$$y_i = \theta_i + \varepsilon_i, \tag{1}$$

where $\varepsilon_i \sim N(0, \nu_i)$ denotes the sampling error and ε_i the sampling variance of the ith estimate. According to the random-effects model, the true effect sizes are heterogenous and are given by

$$\theta_i = \mu + u_i, \tag{2}$$

where $u_i \sim N(0, \tau^2)$. Therefore, τ^2 denotes the heterogeneity in the true effects and μ the average true effect. A special case of the random-effects model arises when $\tau^2 = 0$, in which case the true effects are homogeneous. We can then set up a **mixed-effects meta-regression model** of the form with p potential moderator variables.

$$\theta_i = \beta + \beta_i x_{i1} + \cdots + \beta_p x_{ip} + \mu_i, \tag{3}$$

where x_{ij} denotes the **observed value** of the j**th moderator variable** in the ith study. $\beta_j (j = 1, \cdots, p)$ denotes how $E[\theta_i]$ changes for **a one-unit increase in** x_{ij}, and $\mu_i \sim N(0, \tau^2))$ as before, and τ^2 now denotes **residual heterogeneity**.

Model fitting and inference

Let $\boldsymbol{X}$ denote the $(k \times (p+1))$ model martix containing the values of p moderator variables. Next, let $\boldsymbol{y}$ denote the observed effect size estimates and $\boldsymbol{V}$ a diagonal matrix with the sampling variances along the diagonal. The random/mixed-effects model can be then be written as

$$\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \boldsymbol{M}), \tag{4}$$

where $\boldsymbol{M} = \boldsymbol{V} + \tau^2 \boldsymbol{I}$ and $\boldsymbol{I}$ denotes a identity martix. Letting $\boldsymbol{W} = \boldsymbol{M}^{-1}$, the log likelihood function is therefore given by

$$ll_{\mathrm{ML}}\left(\boldsymbol{\beta}, \tau^2\right) = -\frac{k}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{M}| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{X\beta}), \tag{5}$$

The maximun likelihhod estimate of $\boldsymbol{\beta}$ is given by:

$$\boldsymbol{b} = \left(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{y} \tag{6}$$

Finding the **maximum liklihood estimates of β and τ^2** is considerably simplified by maximizing the profiled log likelihood over τ^2

$$ll_{\mathrm{ML}}\left(\tau^2\right) = -\frac{k}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{M}| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{Xb})'\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{Xb}) \tag{7}$$

The restricted log likelihood function is given by

$$ll_{\mathrm{REML}}\left(\tau^2\right) = -\frac{k-p-1}{2}\ln(2\pi) + \frac{1}{2}\ln|\boldsymbol{X}'\boldsymbol{X}| - \frac{1}{2}\ln|\boldsymbol{M}| - \frac{1}{2}\ln|\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{Xb})'\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{Xb}) \tag{8}$$

Table 1: **Value of the AICc (based on llREMLr), WAIC and LOO for the 16 models fitted to the data examining the influence of mycorrhizal inoculation on plant biomass**

Model	Moderator(s)	AICc	WAIC(N)	WAIC(U)	LOO(N)	LOO(U)
01	FN + FP	203.21(01)	120.28(16)	118.02(13)	132.91(16)	126.45(11)
02	STER + FN + FP	205.16(02)	119.08(15)	119.23(15)	129.00(15)	130.71(15)
03	FUN + FN + FP	205.51(03)	117.22(13)	119.07(14)	128.63(14)	130.92(16)
04	FUN + STER + FP	207.54(04)	117.21(12)	116.19(11)	128.10(13)	126.12(10)
05	FUN+FP	207.57(05)	115.37(11)	113.64(07)	126.74(10)	122.18(04)
06	FUN + STER + FN + FP	207.61(06)	117.50(14)	120.06(16)	126.87(11)	128.22(14)
07	STER + FP	208.53(07)	114.56(07)	116.45(12)	125.45(08)	126.66(12)
08	FP	212.52(08)	112.88(04)	113.54(06)	123.98(06)	124.34(07)
09	FUN	213.05(09)	112.32(03)	112.06(02)	121.80(03)	120.57(01)
10	FUN + STER	214.16(10)	112.25(02)	112.87(03)	119.77(01)	121.95(03)
11	FUN + FN	215.74(11)	114.80(09)	113.34(05)	127.61(12)	123.40(06)
12	STER	216.28(12)	114.58(08)	111.83(01)	123.38(04)	121.12(02)
13	FUN + STER + FN	216.77(13)	114.15(05)	115.00(10)	125.36(07)	125.41(09)
14	FN	217.80(14)	114.29(06)	114.68(09)	125.89(09)	127.21(13)
15	STER + FN	218.33(15)	114.84(10)	113.17(04)	123.91(05)	122.35(05)
16	-	218.49(16)	110.98(01)	113.96(08)	120.86(02)	124.68(08)

(N): By using Normal(0,2.82) prior distribution for each moderator,
(U): By using Uniform(-2,2) prior distribution for each moderator,
(·): The rank of values in the corresponding information criteria,
AICc: AICc is ranked according to TABLE 2 from Cinar et al..

Selection via information criteria

Based on the p moderator variables, a total of $R = 2^p$ models can be fitted to the given data. Information criteria, which penalize the maximized likelihoods for model complexity, can be used for model selection. Let ll deonte eithr ll_{ML} or ll_{REML}

- **AIC (Akaike Information Criterion)**

$$AIC = -2ll + 2(s+2) \tag{9}$$

- **BIC (Bayesian Information Criterion)**

$$BIC = -2ll + (s+2)\ln(k^*) \tag{10}$$

where $k^* = k$ for ML estimate and $k^* = k - s - 1$ for REML estimation. When $k^* \geq 8$, the BIC penalizes the model fit more heavily than the AIC and therefore should tend to select models with fewer fixed effects.

- **AICc (a finite sample size (second-order bias) corrected version of the AIC)**

$$\mathrm{AICc} = -2ll + 2(s+2)\left(\frac{k^*}{k^* - (s+2) - 1}\right) \tag{11}$$

where $k^* = max(k, s+4)$ for ML estimation and $k* = max(k - s - 1, s+4)$ for REML estimation. If k is small, AICc will favor models with fewer parameters [1]. If k is large, AIC will favour models with more parameters.

- **WAIC (Watanabe-Akaike information criterion)**

$$LPPD = \sum_{i=1}^{n}\log\int p\left(y_i \mid \theta\right)p_{\mathrm{post}}(\theta)d\theta$$
$$p_{WAIC} = \sum_{i=1}^{n}\mathrm{var}_{\mathrm{post}}\left(\log p\left(y_i \mid \theta\right)\right)$$
$$WAIC = -2LPPD + 2p_{WAIC} \tag{12}$$

Where $p_{\mathrm{post}}(\theta)$ for posterior distribution, and LPPD for log pointwise predictive density which summarize the predictive accuracy of the fitted model to data. And the penalty term p_{WAIC} is "the variance of individual terms in the log predictive density summed over the n data points".

- **LOO (leave-one-out cross validation)**

$$LOO = -2LPPD_{loo} = -2\sum_{i=1}^{n}\log\int p\left(y_i \mid \theta\right)p_{\mathrm{pos}(-i)}(\theta)d\theta \tag{13}$$

Based on the data minus the data point i, $p_{\mathrm{pos}(-i)}$ is the posterior distribution. Data point i is only used for prediction, not computation, in contrast to $LPPD$, which uses it for both prediction and posterior distribution computation.

Conclusions

In Table 1, we consider two different Prior settings for our binary value moderator, $Normal(0, 2.82)$[2, 3] and $Uniform(-2, 2)$. And $HalfNormal(0.5)$ for the heterogeneity. A large variance makes the Normal prior less informative, spreading the probability mass over a wider range. A Normal prior becomes less specific as variance increases because it allows almost any value of the parameter within that range, which makes it more like a Uniform prior. It is evident that the models are ranked very differently based on different information criteria. In contrast to the frequentist information criterion, the Bayesian information criterion provides completely different ranking results. These differences may be due to the fact that AIC does not work in settings with strong prior information, and WAIC uses a data partition that would make it difficult to use structured models like spatial or network data. However, cross-validation is computationally expensive as well as not always well defined in dependent data settings. Considering that the Bayesian Information Criterion is computationally very time-consuming in model selection, especially when using MCMC, we will also examine whether the Laplace approximation in "INLA" [4] can provide better computational assistance.

References

[1] Ozan Cinar, James Umbanhowar, Jason D Hoeksema, and Wolfgang Viechtbauer. Using information-theoretic approaches for model selection in meta-analysis. *Research Synthesis Methods*, 12(4):537–556, 2021.

[2] Burak Kürsad Günhan, Christian Röver, and Tim Friede. Random-effects meta-analysis of few studies involving rare events. *Research Synthesis Methods*, 11(1):74–90, 2020.

[3] Christian Röver and Tim Friede. Using the bayesmeta r package for bayesian random-effects meta-regression. *Computer Methods and Programs in Biomedicine*, 229:107303, 2023.

[4] Shuangshuang Xu, Marco AR Ferreira, Erica M Porter, and Christopher T Franck. Bayesian model selection for generalized linear mixed models. *Biometrics*, 79(4):3266–3278, 2023.