Compatible imputation models for meta-regression

Jacob M. Schauer

14 Oct 2019

Overview

In this memo, we discuss the idea compatible imputation models, and demonstrate ways to generate compatible imputations for simple meta-regression models.

Meta-regression model and MI

Consider a simple meta-regression model with effect estimates T_i, estimation error variances v_i, and a single covariate X_i for $i = 1, \ldots k$. Assume that some of the X_i are missing at random, and denote X_{obs} as the vector of observed X_i and X_{mis} the missing X_i. Note that the parameter of interest can be written $\eta = (\beta_0, \beta_1, \tau^2)$.

- θ_i effect(s) in study i
- T_i estimate of θ_i
- v_i conditional standard error of T_i (estimation error variance)
- X_i the covariate

$$T_i = \beta_0 + \beta_1 X_i + u_i + e_i$$
$$\theta_i = \beta_0 + \beta_1 X_i + u_i$$
$$X_i, \beta_j \in \mathbb{R}$$
$$V[e_i] = v_i, V[u_i] = \tau^2, u_i \perp e_i$$
$$X_{obs}, X_{mis}$$
$$\eta = (\beta_0, \beta_1, \tau^2)$$

The original derivation of MI assumed a Bayesian approach to analysis (Rubin, 1987) and is explained nicely by Murray (2018). The idea behind the Bayesian approach to MI is that we are after the posterior distribution

$$p(\eta | T, X_{obs}, v) = \int p(\eta | T, X_{obs}, X_{mis}, v) p(X_{mis} | T, X_{obs}, v) dX_{mis} \quad (1)$$

More specifically, we may wish to obtain the posterior mean and variance:

$$E[\eta | T, X_{obs}, v] = E[E[\eta | T, X_{obs}, X_{mis}, v] | T, X_{obs}, v]$$
$$V[\eta | T, X_{obs}, v] = E[V[\eta | T, X_{obs}, X_{mis}, v] | T, X_{obs}, v]$$
$$+ V[E[\eta | T, X_{obs}, X_{mis}, v] | T, X_{obs}, v]$$

When we impute the missing values X_{mis} via Monte Carlo methods and follow the standard combining rules for MI, this results in Monte Carlo estimates of the posterior mean and variance.

An important point here is that any draws of X_{mis} in MI are drawn from the posterior predictive distribution $p(X_{mis} | T, X_{obs}, v)$. Note that we can decompose this into

$$p(X_{mis} | T, X_{obs}, v) = \int p(X_{mis} | \gamma, T, v) p(\gamma | T, X_{obs}, v) d\gamma \quad (2)$$

where γ is the parameter that describes the distribution of $X | T, v$:

$$p(\gamma | T, X_{obs}, v) \propto p(X_{obs} | T, v, \gamma) p(\gamma)$$

Thus, generating imputations involves specifying some model of $p(X | T, v, \gamma)$ and doing the following:

1. Draw γ^* from the posterior distribution $p(\gamma|X_{obs}, T, v)$.
2. Draw X^*_{mis} from a model with probability proportional to $p(X|T, v, \gamma)$.

Compatibility

Various authors have pointed out that estimates based on MI are only accurate insofar as they involve (a) a correctly specified analytic model $p_A(\eta|T, X_{obs}, X_{mis}, v)$ and (b) a correctly specified imputation model that is congenial/compatible with the analytic model $p_A(X_{mis}|T, X_{obs}, v)$.

$$p_A(\eta|T, X_{obs}, v) = \int p_A(\eta|T, X_{obs}, X_{mis}, v) p_A(X_{mis}|T, X_{obs}, v) dX_{mis}$$

Here, we use $p_A(X_{mis}|T, X_{obs}, v)$ to denote that the imputation model is compatible/congenial with the analytic model (discussed further below).

If we use an imputation model $p_I(X_{mis}|T, X_{obs}, v)$ that is not correctly specified, then the moments estimated via MI will not be of the posterior we're interested in.

$$p_H(\eta|T, X_{obs}, v) = \int p_A(\eta|T, X_{obs}, X_{mis}, v) p_I(X_{mis}|T, X_{obs}, v) dX_{mis}$$
$$\neq p_A(\eta|T, X_{obs}, v)$$

- Analytic model:
  $p_A(\eta|T, X_{obs}, X_{mis}, v)$
- Imputation model:
  $p_I(\eta|T, X_{obs}, X_{mis}, v)$
- Accurate inferences depend on using the correctly specified imputation model so that $I = A$. Note that this only means that the imputation model is related to the analytic model in specific ways, and not that they are identical.

While this makes some intuitive sense (if we use the wrong imputation model, our inferences will be inaccurate), there are at least two ways to describe what it means for an imputation model to be correctly specified. Meng (1994) describes the idea of *congeniality*, and lays out conditions for what it means for an imputation model to be congenial with an analytic model. This largely amounts to a requirement that the imputation model be at least as complex/general as the analytic model.

A related concept is that of *compatibility*, which says that a series of conditional distributions imply a specific joint distribution that exists. Why does this matter? Consider our imputation problem. We want to impute values of X_{mis} given T, v, and then we turn around and model T given values of X_{mis} we just imputed. What we're really saying here his that X, T have some joint distribution from which our imputation and analytic models are resulting conditional distributions.

In the context of (meta-) regression, the ideas of compatibility and congeniality are related. If we use incompatible imputation models, then they are in some way uncongenial.

Compatibility: Let $g(X, T|\lambda)$ be a joint distribution, and $f_T(T|X, \lambda_T)$, $f_X(X|T, \lambda_X)$ be conditional distributions with parameter spaces Λ, Λ_j for $j = T, X$, respectively. f_j are compatible with g if there exist surjective maps from $\Lambda \mapsto \Lambda_j$ such that $f_X(X|T, \lambda_X) = g(X|T, \lambda)$ and $f_T(T|X, \lambda_T) = g(T|X, \lambda)$.

$p(T, X|y, \eta, \gamma) \leftrightarrow p(X|T, v, \gamma), p(T|X, v, \eta)$

Ensuring compatible imputations for covariates

According to the integral in (1), the inference we are interested in requires Monte Carlo simulations of the posterior predictive distribution $p(X_{mis}|T, X_{obs}, v)$, which in turn requires we specify an appropriate model $p(X|T, v, \gamma)$.

If we let $\gamma = (\eta, \psi) = (\beta_0, \beta_1, \tau^2, \psi)$, then Bartlett et al. (2015) show that a natural approach to ensuring that the imputation model is compatible with the analytic model is to set

$$p(X|T, v, \gamma) \propto p(T|X, v, \eta)p(X|\psi)$$

The first term above is just the analytic model likelihood, and the second term is analogous to a prior distribution on the missing X.

Assuming an agnostic prior $p(X|\psi) \propto a$, then

$$p(X|T, v, \gamma) \propto p(T|X, v, \eta) \tag{3}$$

In this case, the main parameter of interest for imputation η is the same parameter of interest for our analytic model.

$$p(X|T, v, \gamma) = \frac{p(X, T|v, \eta, \psi)}{p(T|v, \eta, \psi)}$$
$$= \frac{p(T|X, v, \eta)p(X|\psi)}{p(T|v, \eta, \psi)}$$
$$\propto p(T|V, v, \eta)p(X|\psi)$$

$p(X|\psi)$ lets us put a more refined model on X. It could also be used to help generate a proposal distribution if we use importance sampling to obtain imputations.

Generating compatible imputations

Recall that step 1 in our algorithm for imputing X_{mis} involves drawing $\gamma|X_{obs}, T, v$. As a consequence of the result from the previous section, one approach to generating compatible imputations would involve drawing $\eta|T, X_{obs}, v$. But that is just the posterior distribution of η given our complete cases.

Thus, we can generate draws from $p(X_{mis}|T, X_{obs}, v)$ in (2) by first generating draws of η from $p(\eta|T, X_{obs}, v)$. So long as we can fit our meta-regression model, we can actually draw X_{mis} from a model that is compatible with our meta-regression model.

The general algorithm looks something like this:

1. Draw η^* from $p(\eta|T, X_{obs}, v)$.
2. Draw X_{mis}^* from a model with probability proportional to $p_A(T|X_{mis}, v, \eta^*)$.

Categorical variables

When X is categorical with $j = 0, \dots, c$ categories (i.e., $c + 1$ categories), then the (saturated) regression model can be written:

$$T_i = \sum_{j=0}^{c} \beta_j \mathbf{1}\{X_i = j\} + u_i + e_i$$

Thus, $\gamma \equiv \eta = (\beta_0, \dots \beta_c, \tau^2)$. Note that since X is categorical, then we may model it as multinomial random variable when we impute it.

Using the result above, this means that

$$P[X = k|T, v, \gamma] = \frac{p(T|X = k, v, \eta)}{\sum_{j=0}^{c} p(T|X = j, v, \eta)} \qquad (4)$$

Under the normal analytic meta-regression model, this reduces to:

$$P[X = k|T, v, \gamma] \propto \frac{\exp\left\{-\frac{(T-\beta_k)^2}{2(\tau^2+v)}\right\}}{\sum_{j=0}^{c} \exp\left\{-\frac{(T-\beta_j)^2}{2(\tau^2+v)}\right\}} \qquad (5)$$

Thus, we can generate imputations of X_{mis} that are compatible with $p(T|X, v, \eta)$ by the following algorithm:

1. Draw η^* from the posterior distribution $p(\eta|T, X_{obs}, v)$

2. Draw X_{mis} from a multinomial distribution with probabilities given in (5), substituting $\eta^* = (\beta_0^*, \ldots, \beta_c^*, \tau^{2*})$

Continuous variables

When X is continuous, then the regression model is

$$T_i = \beta_0 + \beta_1 X_i + u_i + e_i$$

The imputation model is given by

$$p(X|T, v, \gamma) \propto p(T|X, v, \eta)p(X|\psi)$$

We can specify a probability model $p(X|\psi)$ based on prior knowledge about X (e.g., if we know it's distribution should be skewed). However, if we assume $p(X|\psi) \propto a$ is flat, then, the imputation model is

$$p(X|T, v, \gamma) \propto p(T|X, v, \eta) = \frac{1}{\sqrt{2\pi(v + \tau^2)}} \exp\left\{-\frac{(T - \beta_0 - \beta_1 X)^2}{2(\tau^2 + v)}\right\}$$

This implies that $\beta_1 X$ is normally distributed with mean $\beta_0 - T$ and variance $v + \tau^2$. Therefore, we can generate compatible imputations (under the assumption that $p(X|\psi) \propto a$) by the following algorithm:

1. Draw η^* from $p(\eta|T, v, X_{obs})$.
2. Draw $Z^* \sim N(\beta_0^* - T, v + \tau^{2*})$.
3. Set $X^* = Z^*/\beta_1^*$.

Incorporating information beyond the analytic model

We ignored $p(X|\psi)$ above, which is equivalent to saying all we need to know to impute X is contained in the analytic model $p(T|X, v, \eta)$.

However, it is possible to specify $p(X|\psi)$ that incorporates informa-
tion about X not contained in the analytic model. This can complicate
things in the sense that direct sampling may not be possible. Above,
we can simply draw η and then draw X directly from it's posterior
predictive distribution. However, depending on the nature of $p(X|\psi)$,
generating imputations of X may require more computationally in-
tensive methods, such as importance or rejection sampling.

Some questions to answer, then, include:

1. Are there "conjugate" families of $p(X|\psi)$ that result in $p(T|X,v,\eta)p(X|\psi)$
   to have a likelihood that resembles a normal distribution (or some
   other known distribution like a t distribution).

2. Are there models $p(X|\psi)$ that can put boundary restrictions on the
   support of X such that $p(X|T,v,\gamma)$ is easy to sample for? For in-
   stance, if we know X must be positive, how can we specify $p(X|\psi)$
   to do that such that sampling is easy?

3. If sampling is not easy, is there a "general" importance sampling
   approach?

   - We could draw $X \sim g(\lambda)$, and then re-sample according to
     weights $p(X|T,v,\gamma)/g(\lambda)$.

4. If there is no good importance sampling approach, what is the
   best rejection sampling approach?

Other questions

The approach outlined here assumes that we can fit and sample from
the posterior distribution of the analytic model. This is easy to do
directly when $\tau^2 = 0$ and we fit a model where $\eta = (\beta_0, \beta_1)$. Draws
from the posterior are just draws from a normal distribution with a
known mean and variance.

However, for the random effects model, sampling from the pos-
terior of τ^2 is not straightforward. Most approaches use rejection
sampling (Metropolis-Hastings). This will slow down imputations. Is
there a useful approximation to the posterior that might be of use?