Three imputation methods for multiple $\overset{\swarrow \text{MLR}}{\text{linear}}$ regression

[van Buuren 2018, ch 3.2.1+3.2.2]

NA = missing value (shorthand)

Notation: now we work with the imputation model

and assume that $\overset{N\times1}{Y} = \overset{N\times q}{X}\overset{q\times1}{\phi} + \varepsilon$, $\varepsilon \sim N(0, \sigma^2 I)$

but NA's are only in Y, and write

$$\underset{N\times1}{y} = \begin{bmatrix} y_o \\ y_m \end{bmatrix} \begin{matrix} \leftarrow n_1\times1 \\ \leftarrow n_0\times1 \end{matrix} \quad \text{and} \quad \underset{N\times q}{X} = \begin{bmatrix} X_o \\ X_m \end{bmatrix} \begin{matrix} n_1\times q \\ n_0\times q \end{matrix}$$

↑ no missing here - only labelled based on NA's in y

NB: book use β, but this is imp and not analysis model, so we use ϕ to emphasize this!

Aim: Construct imputation(s) for y_m.

This will use ϕ as a "helper".

$\dot{y}_m$ = imputed value

In L5 we looked at two solution using the MLR

1) Predict $\dot{y}_m = X_m\hat{\phi}$ $\quad \hat{\phi} = (X_o^T X_o)^{-1} X_o^T y_o$ [one value]

2) Predict + noise $\dot{y}_m = X_m\hat{\phi} + \dot{\varepsilon}$, $\dot{\varepsilon} \sim N(0, \hat{\sigma}^2)$
↑ drawn from

where $\hat{\sigma}^2 = \frac{1}{N-q}(y_{obs} - X_{obs}\hat{\phi})^T(y_{obs} - X_{obs}\hat{\phi})$

With method 2, we may potentially draw many imputed datasets (e.g. m)

Now we add a Bayesian solution!

↓

Bayesian (multiple) imputation in MLR

Likelihood: $p(y \mid X, \phi, \sigma^2) \propto (\sigma^2)^{-N/2} \exp\left(-\frac{1}{2\sigma^2}(y - X\phi)^T(y - X\phi)\right)$

"then posterior derived analytically"
↓
Conjugate prior: $p(\phi, \sigma^2) = p(\phi \mid \sigma^2) \cdot p(\sigma^2)$

$p(\phi \mid \sigma^2) \sim N(\mu_0, \sigma^2 \Lambda_0^{-1})$ $\quad$ Λ_0 is precision matrix ← Bayesian like this!
inverse of covariance mat

$p(\sigma^2) \sim$ inverse-gamma

Posterior: $p(\phi, \sigma^2 \mid y, X) \propto \overset{\text{likelihood}}{p(y \mid X, \phi, \sigma^2)} \cdot \overbrace{p(\phi \mid \sigma^2)\, p(\sigma^2)}^{\text{prior}}$

Some rearrangement:

$E(\phi, \sigma^2 \mid y, X)$: the ϕ' part: $\mu_n = (\underbrace{X^T X + \Lambda_0}_{\Lambda_n})^{-1}(X^T X \underset{\uparrow \text{ LS est.}}{\hat{\phi}} + \Lambda_0 \mu_0)$

$E(\phi, \sigma^2 \mid y, X) = \begin{bmatrix} \mu_n \\ E\sigma^2 \end{bmatrix}$

$Var(\phi, \sigma^2 \mid y, X) = \sigma^2 \Lambda_n^{-1}$

How to use this in Bayesian imputation?

Similar to 2) above, but replace $\hat{\phi}$ and $\hat{\sigma}^2$ with draws from the posterior $\Rightarrow \dot{\phi}, \dot{\sigma}^2$ so that

$\dot{y}_m = X_m \dot{\phi} + \dot{\varepsilon}$ and $\dot{\varepsilon} \sim N(0, \dot{\sigma}^2)$

The frequentist interpretation is that the parameter uncertainty is now also taken into account. As for 2) we may draw many (m) imputed data sets.

See van Buuren (2018) Algo 3.1 for how this is implemented in mice R package (function norm)