Multilevel Process Monitoring: A Case Study to Predict Student Success or Failure

July 13, 2020

Supplementary Material

This supplementary material gives the full conditional distributions and the JAGS code of the three-level model discussed used to model grades for high-school students.

Full Conditional Distributions

The Gibbs sampling procedure uses the full conditional distributions of the unknown parameter space. Although they are not necessary when using rJAGS (Plummer (2018)), we report them below to be used in a Gibbs sampler or similar Markov Chain Monte Carlo (MCMC) sampling methods.

The likelihood function of the n_0 observed grades is the joint density of the data conditional on the parameters.

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n_0} f(y_i|\boldsymbol{\theta}) = (2\pi)^{-n_0/2}\sigma^{-n_0}\exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{n_0}(y_i - \boldsymbol{X}_i^{(L_0)}\boldsymbol{\beta}_{j[i]}^{(L_0)})^2\right).$$

Let vector $\boldsymbol{Y}$ of size n_0 contain the observed values y_i. The full conditional distributions of the individual parameters are each proportional to $p(\boldsymbol{Y},\boldsymbol{\theta})$:

$$f(\boldsymbol{Y},\boldsymbol{\theta}) \propto f(\boldsymbol{\beta}^{(\boldsymbol{L_2})})f(\sigma^2)f(\boldsymbol{\Sigma}^{(\boldsymbol{L_1})})f(\boldsymbol{\Sigma}^{(\boldsymbol{L_2})})\prod_{h=1}^{n_2} f(\boldsymbol{\beta}_{\boldsymbol{h}}^{(\boldsymbol{L_1})}|\boldsymbol{\Sigma}^{(\boldsymbol{L_2})},\boldsymbol{\beta}^{(\boldsymbol{L_2})})\prod_{j=1}^{n_1} f(\boldsymbol{\beta_{jh}}|\boldsymbol{\Sigma}^{(\boldsymbol{L_1})},\boldsymbol{\beta}_{\boldsymbol{h}}^{(\boldsymbol{L_1})})\prod_{i=1}^{n_0} f(y_{ijh}|\boldsymbol{\beta_{jh}},\boldsymbol{\beta}_{\boldsymbol{h}}^{(\boldsymbol{L_1})},\sigma^2).$$

We then calculate the full conditional distributions by multiplying the prior by the likelihood and simplifying.

Calculation of the full conditional distribution of $\boldsymbol{\beta}_j^{(L_0)}$

We calculate the full conditional distribution of $\boldsymbol{\beta}_j^{(L_0)}$ by multiplying the prior by the likelihood and simplifying

$$f(\boldsymbol{\beta}_j^{(L_0)},|\boldsymbol{Y},\sigma^2,\boldsymbol{\beta}_h^{(L_1)},\boldsymbol{\beta}^{(L_2)},\boldsymbol{\Sigma}^{(L_1)},\boldsymbol{\Sigma}^{(L_2)})\propto f(\boldsymbol{\beta_j^{(L_0)}}|\boldsymbol{\Sigma_j},\boldsymbol{\beta_h^{(L_1)}})\times\prod_{i=1}^{N_j}f(y_{ij}|\boldsymbol{\beta_j^{(L_0)}},\boldsymbol{\beta_h^{(L_1)}},\sigma^2)\propto$$

$$N(\boldsymbol{V^{-1}M},\boldsymbol{V})$$

with $\boldsymbol{V}=(\boldsymbol{\Sigma}^{(L_1)-1}+\sigma^{-2}\boldsymbol{X}_{i\in j}^{(L_0)\prime}\boldsymbol{X}_{i\in j}^{(L_0)})^{-1}$ and $\boldsymbol{M}=(\boldsymbol{\Sigma}^{(L_1)-1}\boldsymbol{\beta}_h^{(L_1)}\boldsymbol{X}_j^{(L_1)\prime}+\sigma^{-2}\boldsymbol{X}_{i\in j}^{(L_0)\prime}\boldsymbol{Y}_{i\in j})$.

Calculation of the full conditional distribution of $\boldsymbol{\beta}_h^{(L_1)}$

We calculate the full conditional distribution of $vec(\boldsymbol{\beta}_h^{(L_1)})$ by multiplying the prior by the likelihood and simplifying

$$f(vec(\boldsymbol{\beta}_h^{(L_1)})|\boldsymbol{Y},\sigma^2,\boldsymbol{\beta}_j^{(L_0)},\boldsymbol{\beta}^{(L_2)},\boldsymbol{\Sigma}^{(L_1)},\boldsymbol{\Sigma}^{(L_2)})\propto$$

$$f(vec(\boldsymbol{\beta}_h^{(L_1)})|\boldsymbol{\Sigma}^{(L_2)},\boldsymbol{\beta}^{(L_2)})\prod_{j=1}^{N_{j\in h}}f(\boldsymbol{\beta}_j^{(L_0)}|\boldsymbol{\Sigma}^{(L_1)},\boldsymbol{\beta}_h^{(L_1)})\prod_{i=1}^{N_{i\in h}}f(y_{ijh}|\boldsymbol{\beta}_j^{(L_0)},\boldsymbol{\beta}_h^{(L_1)},\sigma^2)\propto$$

$$exp\Bigg[-(vec(\boldsymbol{\beta}_h^{(L_1)})-\boldsymbol{\beta}^{(L_2)}\boldsymbol{X}_h^{(L_2)\prime})'\boldsymbol{\Sigma}^{(L_2)-1}(vec(\boldsymbol{\beta}_h^{(L_1)})-\boldsymbol{\beta}^{(L_2)}\boldsymbol{X}_h^{(L_2)\prime})-$$

$$\sum_{j\in h}\left((\boldsymbol{\beta}_j^{(L_0)}-\boldsymbol{\beta}_h^{(L_1)}\boldsymbol{X}_j^{(L_1)\prime})'\boldsymbol{\Sigma}^{(L_1)-1}(\boldsymbol{\beta}_j^{(L_0)}-\boldsymbol{\beta}_h^{(L_1)}\boldsymbol{X}_j^{(L_1)\prime})\right)\Bigg]$$

For further calculations, we define $A=vec(\boldsymbol{\beta}_h^{(L_1)})$, $B=\boldsymbol{\beta}^{(L_2)}\boldsymbol{X}_h^{(L_2)\prime}$, $C=\boldsymbol{\Sigma}^{(L_2)}$, $D_j=\boldsymbol{\beta}_j^{(L_0)}$, E=$\boldsymbol{\Sigma}^{(L_1)}$, $X_j=\boldsymbol{X}_j^{(L_1)}$, this gives

$$f(vec(\boldsymbol{\beta}_h^{(L_1)})|\boldsymbol{Y},\sigma^2,\boldsymbol{\beta}_j^{(L_0)},\boldsymbol{\beta}^{(L_2)},\boldsymbol{\Sigma}^{(L_1)},\boldsymbol{\Sigma}^{(L_2)})\propto P(A|B,C,D_j,E,X_j)\propto$$

$$exp\Bigg[-(A-B)'C^{-1}(A-B)-\sum_{j\in h}\Big[D_j'E^{-1}D_j-2(X_j\otimes(D_j'E^{-1}))A+A'(X_j\otimes(X_j'\otimes E))A\Big]\Bigg]\propto$$

$$exp\Bigg[-A'\left(C^{-1}+\sum_{j\in h}\left(X_j\otimes(X_j'\otimes E^{-1})\right)\right)A+A'(C^{-1}B)+\left(B'C^{-1}+2\sum_{j\in h}\left(X_j\otimes(D_j'E^{-1})\right)\right)A\Bigg]$$

which shows that the full conditional distribution of $vec(\boldsymbol{\beta}_h^{(L_1)})$ is a multivariate normal distribution with covariance matrix

$$\left(\boldsymbol{\Sigma}^{(L_2)-1}+\sum_{j\in h}\left(\boldsymbol{X}_j^{(L_1)}\otimes(\boldsymbol{X}_j^{(L_1)\prime}\otimes\boldsymbol{\Sigma}^{(L_1)-1})\right)\right)^{-1}$$

and mean

$$\left(\boldsymbol{\Sigma}^{(L_2)-1}+\sum_{j\in h}\left(\boldsymbol{X}_j^{(L_1)}\otimes(\boldsymbol{X}_j^{(L_1)\prime}\otimes\boldsymbol{\Sigma}^{(L_1)-1})\right)\right)(\boldsymbol{\Sigma}^{(L_2)-1}\boldsymbol{\beta}^{(L_2)}\boldsymbol{X}_h^{(L_2)\prime}).$$

Calculation of the full conditional distribution of $\mathbf{\Sigma}^{(L_1)}$

$$p(\mathbf{\Sigma}^{(L_1)}|\boldsymbol{Y},\sigma^2,\boldsymbol{\beta_j^{(L_0)}},\boldsymbol{\beta_h^{(L_1)}},\boldsymbol{\beta^{(L_2)}},\mathbf{\Sigma^{(L_2)}})\propto f(\mathbf{\Sigma}^{(L_1)})\prod_{j=1}^{n_1} f(\boldsymbol{\beta_j^{(L_0)}}|\mathbf{\Sigma^{(L_1)}},\boldsymbol{\beta_h^{(L_1)}})\propto$$
$$|\mathbf{\Sigma}^{(L_1)}|^{-(n_1+\nu+p_0+2)/2}exp(-tr((\boldsymbol{S^{L_1}}+\boldsymbol{C})\mathbf{\Sigma^{(L_1)-1}}/2)$$

with $\boldsymbol{S^{L_1}} = \sum_{j=1}^{n_1}(\boldsymbol{\beta_j^{L_0}} - \boldsymbol{\beta_{h[j]}^{(L_1)}}\boldsymbol{X_j^{(L_1)\prime}})'(\boldsymbol{\beta_j^{L_0}} - \boldsymbol{\beta_{h[j]}^{L_1}}\boldsymbol{X_j^{(L_1)\prime}})$, which shows that the full conditional distribution of $\mathbf{\Sigma}^{(L_1)}$ is $W^{-1}(\boldsymbol{S^{L_1}}+\boldsymbol{C}, n_1+\nu)$.

Calculation of the full conditional distribution of $\mathbf{\Sigma}^{(L_2)}$

$$p(\mathbf{\Sigma}^{(L_2)}|\boldsymbol{Y},\sigma^2,\boldsymbol{\beta_j^{(L_0)}},\boldsymbol{\beta_h^{(L_1)}},\boldsymbol{\beta^{(L_2)}},\mathbf{\Sigma^{(L_1)}})\propto f(\mathbf{\Sigma}^{(L_2)})\prod_{h=1}^{n_2} f(vec(\boldsymbol{\beta_h^{(L_1)}})|\mathbf{\Sigma^{(L_2)}},\boldsymbol{\beta^{(L_2)}})\propto$$
$$|\mathbf{\Sigma}^{(L_2)}|^{-(n_2+\nu+2p_0p_1+2)/2}exp(-tr((\boldsymbol{S^{L_2}}+\boldsymbol{D})\mathbf{\Sigma^{(L_2)-1}}/2)$$

with $\boldsymbol{S^{L_2}} = \sum_{h=1}^{n_2}(vec(\boldsymbol{\beta_h^{L_1}}) - \boldsymbol{\beta^{L_2}}\boldsymbol{X_h^{(L_2)\prime}})'(vec(\boldsymbol{\beta_h^{L_1}}) - \boldsymbol{\beta^{L_2}}\boldsymbol{X_h^{(L_2)\prime}})$, which shows that the full conditional distribution of $\mathbf{\Sigma}^{(L_2)}$ is $W^{-1}(\boldsymbol{S^{L_2}}+\boldsymbol{D}, n_2+\nu)$.

Calculation of the full conditional distribution of $\boldsymbol{\beta}^{(L_2)}$

$$p(vec(\boldsymbol{\beta^{(L_2)}})|\boldsymbol{Y},\sigma^2,\boldsymbol{\beta_j^{(L_0)}},\boldsymbol{\beta_h^{(L_1)}},\mathbf{\Sigma^{(L_1)}},\mathbf{\Sigma^{(L_2)}})\propto f(vec(\boldsymbol{\beta^{(L_2)}}))\prod_{h=1}^{n_2} f(vec(\boldsymbol{\beta_h^{(L_1)}})|\mathbf{\Sigma^{(L_2)}},\boldsymbol{\beta^{(L_2)}})$$

which similarly to $\boldsymbol{\beta_h^{(L_1)}}$ has a multivariate normal distribution with covariance matrix

$$\left(\boldsymbol{B}^{-1}+\sum_h\left(\boldsymbol{X}_h^{(L_2)\prime}\otimes(\boldsymbol{X}_h^{(L_2)}\otimes\mathbf{\Sigma}^{(L_2)-1})\right)\right)^{-1}$$

and mean

$$\left(\boldsymbol{B}^{-1}+\sum_h\left(\boldsymbol{X}_h^{(L_2)\prime}\otimes(\boldsymbol{X}_h^{(L_2)}\otimes\mathbf{\Sigma}^{(L_2)-1})\right)\right)(\boldsymbol{B}^{-1}\boldsymbol{a})$$

Calculation of the full conditional distribution of σ^2

$$p(\sigma^2|\boldsymbol{Y},\boldsymbol{\beta_j^{(L_0)}},\boldsymbol{\beta^{(L_2)}},\boldsymbol{\beta_h^{(L_1)}},\mathbf{\Sigma^{(L_1)}},\mathbf{\Sigma^{(L_2)}})\propto f(\sigma^2)\prod_{i=1}^{n_0} f(y_{ijh}|\boldsymbol{\beta_{jh}},\boldsymbol{\beta_h^{(L_1)}},\sigma^2)\propto$$
$$\sigma^{-(a+n_0/2+1)}exp(-\sigma^{-2}(b+\sum_{i=1}^{n_0}(y_i-\boldsymbol{X_i^{(L_0)}}\boldsymbol{\beta_{j[i]}^{(L_0)}})^2/2),$$

which shows that the full conditional distribution of σ^2 is proportional to an $IG(a + \frac{n_0}{2}, b + \frac{\sum_{i=1}^{n_0}(y_i - \boldsymbol{X_i^{(L_0)}}\boldsymbol{\beta_{j[i]}^{(L_0)}})^2}{2})$ distribution.

JAGS Code

```
### Dimensions and sizes
n0 <- dim(df)[1]
n1 <- length(unique(student))
n2 <- length(unique(class))
p0 <- dim(x0)[2]
p1 <- dim(x1)[2]
p2 <- dim(x2)[2]

### Hyper parameters
Sigma1.init <- diag(0.001,p0)
Sigma2.init <- diag(0.001,p0*p1)
Hypervar.init.Sigma <- diag(0.001,p0*p1*p2)
Hypervar.init.mu <- rep(0,p0*p1*p2)

### The JAGS model specification
model_string <- model{
    for(i in 1:n0) {
        y[i] ~ dnorm(mu[i],sigma)
        mu[i] <- b0[j_s[i],1:p0]%*%x0[i,1:p0]
    }
    sigma ~ dgamma(1000,1)
    for (j in 1:n1){
        for(p in 1:p0){
            mu1[j,p] <- b1[k_s[j],((p-1)*p1+1):(p*p1)]%*%x1[j,1:p1]
        }
        b0[j,1:p0] ~ dmnorm (mu1[j,1:p0],Sigma1[1:p0,1:p0])
    }
    Sigma1 ~ dwish(Sigma1.init,p0+1)
    for(h in 1:n2){
        for(p in 1:(p1*p0)){
            mu2[h,p] <- b2[((p-1)*p2+1):(p*p2)] %*% x2[h,1:p2]
```

```
        }
        b1[h,1:(p1*p0)]~ dmnorm (mu2[h,1:(p0*p1)],Sigma2)
    }
    Sigma2 ~ dwish(Sigma2.init,p0*p1+1)
    b2 ~ dmnorm(Hypervar.init.mu,Hypervar.init.Sigma)
}
```

References

Plummer, M. (2018). *rjags: Bayesian graphical models using MCMC*. R package version 4-8.
URL https://CRAN.R-project.org/package=rjags