A Multinomial Regression Approach to Model Outcome Heterogeneity

BaoLuo Sun, Tyler VanderWeele and Eric J. Tchetgen Tchetgen

Web Appendix

Web Appendix 1. Implementation of Constrained Bayesian Estimation in Simulation

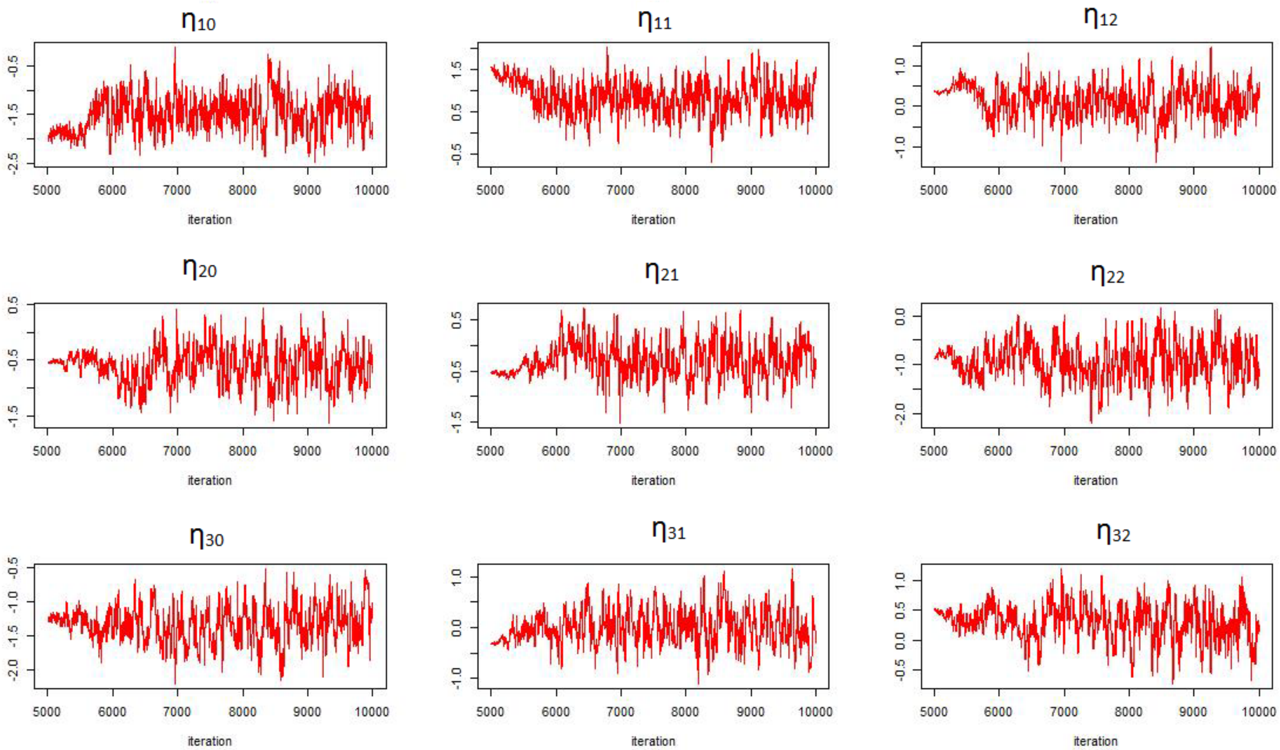
The OpenBUGS code for posterior computation in the simulation study is shown below. For the i^{th} individual with outcome $Y = k$, the encoding follows that $Y_k[i] = 1$ and $Y_j[i] = 0$ for $j \neq k$. The first part of the code describes each individual's contribution to the observed data likelihood, where the $N \times 1$ dummy vector Z of all 1's is assumed to be the results of Bernoulli trials with probabilities $p[i]$. By making each p[i] equal to the probability of observing the outcome for the i^{th} individual, the required likelihood term is provided. The input covariates are the $N \times 1$ vectors X_1 and X_2.

Prior distributions for the parameters in the model are specified as independent $N(0, 10^2)$. The final part of the code imposes constraints on the sampling space so that $0 < \Pr\{Y = 0|X_i; \hat{\eta}\} < 1$ for all individuals $i = 1, 2, ..., n$. Posterior mean, median and 95% credible intervals can be obtained directly from Markov-chain Monte-Carlo sampling in R through BRugs, an interface to the OpenBUGS software.

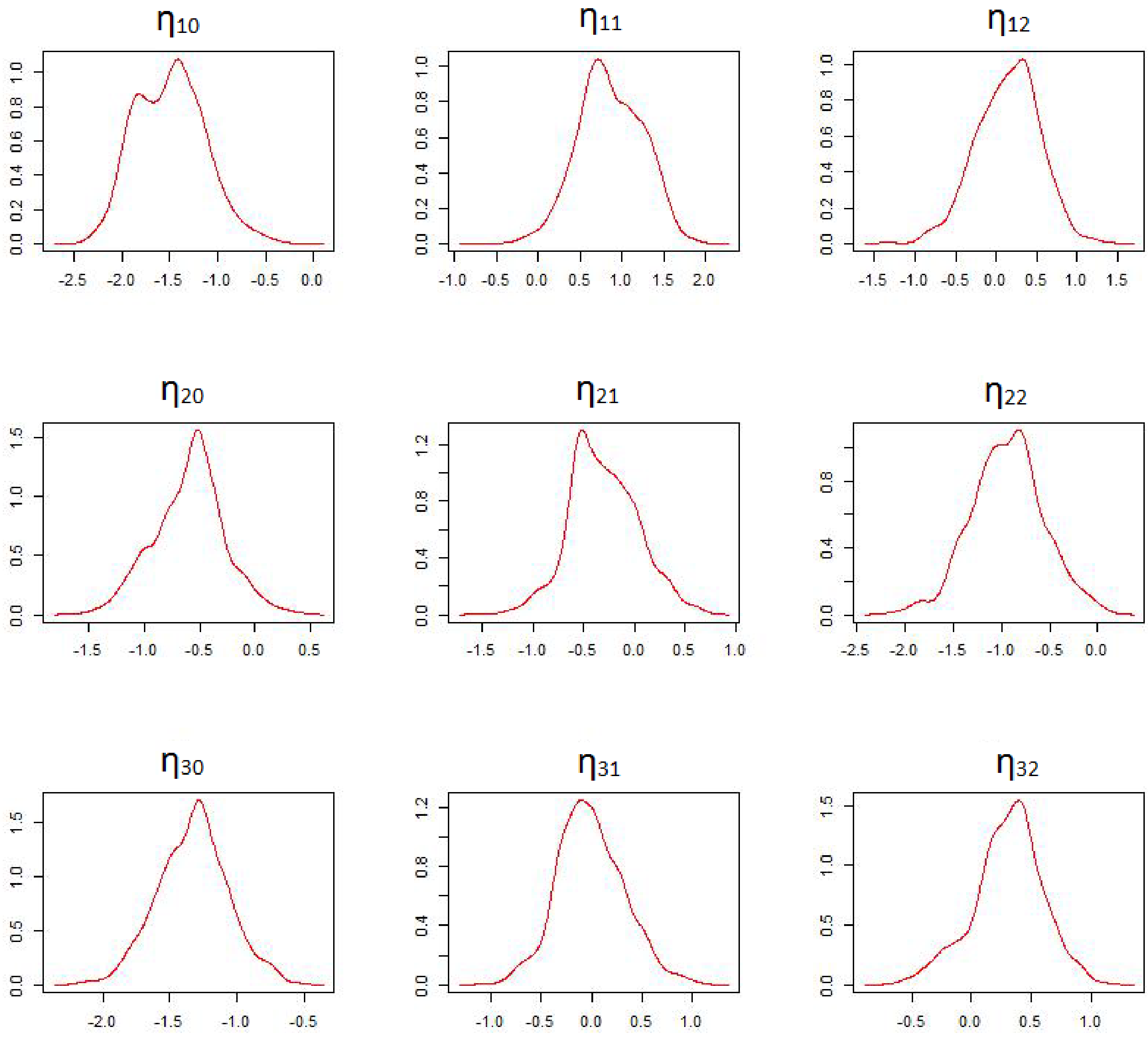
```
Model <- function() {

    for (i in 1:N){
        z[i]   <- 1
        z[i]    ~ dbern(p[i])
        p[i]   <- L[i]
        L[i]   <-  Y0[i]*pi0[i]+Y1[i]*pi1[i]+Y2[i]*pi2[i]+Y3[i]*pi3[i]

        #Probability for each of the outcomes Y=1,2,3.

        logit(pi1[i])<-eta[1]+eta[2]*X1[i]+eta[3]*X2[i]
        logit(pi2[i])<-eta[4]+eta[5]*X1[i]+eta[6]*X2[i]
```

```
        logit(pi3[i])<-eta[7]+eta[8]*X1[i]+eta[9]*X2[i]
        pi0[i]  <- 1-pi1[i]-pi2[i]-pi3[i]
    }
    #Priors for parameters in missing data model
    for (j in 1:9) {
        eta[j] ~ dnorm(0, 0.01)
    }

    # implementing the constraints
    for (k in 1:N){
        ones[k] <- 1
        ones[k] ~ dbern(C[k])
        C[k] <- step(pi0[k])
    }
}
```

Trace Plots and Posterior Densities of η in a Typical Simulation Replicate



Web Figure 1: Trace plots for posterior sampling of η in a typical simulation replicate ($n = 500$), after an initial burn-in of 5000 iterations.



Web Figure 2: Smoothed kernel estimated posterior densities of η in a typical simulation replicate ($n = 500$).

Web Appendix 2. Implementation of Constrained Bayesian Estimation in Application Data-set

The OpenBUGS code for posterior computation in the application data-set is shown below. The first part of the code describes each individual's contribution to the observed data likelihood, where the $N \times 1$ dummy vector Z of all 1's is assumed to be the results of Bernoulli trials with probabilities $p[i]$. By making each p[i] equal to the probability of observing the outcome for the i^{th} individual, the required likelihood term is provided. *surv* is the $N \times 1$ vector with 1 for survivals and 0 for deaths. Similarly, *chd* is the $N \times 1$ vector with 1 for deaths from CHD and 0 otherwise, *str* is the $N \times 1$ vector with 1 for deaths from stroke and 0 otherwise, *can* is the $N \times 1$ vector with 1 for deaths from cancer and 0 otherwise. The covariates are $N \times 1$ vectors *age* for age, *sex* for gender, *sc* for serum cholesterol, *bmi* for BMI and *hbp* for high blood pressure.

```
CBmodel <- function() {

for (i in 1:N){
z[i]   <- 1
z[i]    ~ dbern(p[i])
p[i]   <- L[i]
L[i]   <-   step(surv[i]-0.5)*(pi1[i])
    + step(chd[i]-0.5)*(pi2[i])
    + step(str[i]-0.5)*(pi3[i])
    + step(can[i]-0.5)*(pi4[i])

logit(pi2[i])<- g[1]+g[2]*age[i]+g[3]*sex[i]+g[4]*sc[i]+g[5]*bmi[i]+g[6]*hbp[i]
logit(pi3[i])<- g[7]+g[8]*age[i]+g[9]*sex[i]+g[10]*sc[i]+g[11]*bmi[i]+g[12]*hbp[i]
logit(pi4[i])<- g[13]+g[14]*age[i]+g[15]*sex[i]+g[16]*sc[i]+g[17]*bmi[i]+g[18]*hbp[i]
pi1[i]       <- 1-pi2[i]-pi3[i]-pi4[i]
}
for (k in 1:18) {
g[k] ~ dnorm(0, 0.001)
}
for (j in 1:N) {
ones[j] <- 1
ones[j] ~ dbern(CONSTR[j])
CONSTR[j] <- step(pi1[j])
}
}
```