Mô phỏng hierarchy MCMC cho mô hình Satlns ~ end + ses + res + insured, BayesVL v0.6.5 trên 1042 quan sát thực

Vương Quân Hoàng Lã Việt Phương

AISDL (Vuong & Associates)
SDAG, Centre for Interdisciplinary Social Research (Phenikaa University)



Note 2b

Hà Nội, 28-4-2019 (6:30 PM version 3)

Bài viết này (**Note 2b**) trình bày một mô hình và ứng dụng mô phỏng BayesVL v0.6.5 [1] trên môi trường R và Stan MCMC, cho bài toán thực tế [2] với dữ liệu thực tế, gồm 1042 điểm dữ liệu [3]. Nội dung của bài là một phần của tài liệu do AISDL biên soạn phục vụ hướng dẫn sử dụng "The BayesVL R Package", nhằm thúc đẩy việc sử dụng phương pháp thống kê Bayesian trong KHXH&NV.

Một phần mã chương trình [1] được viết dựa trên *Statistical Rethinking* của Richard McElreath [5] và *BNLearn* của Scuttari [6]. Công việc mô phỏng cơ bản sử dụng kỹ thuật Hamiltonian Markov chain Monte Carlo, dựa trên Stan và R.

1. Bài toán cần xử lý

Bài toán sử dụng cơ sở dữ liệu "Health Care, Medical Insurance, and Economic Destitution: A Dataset of 1042 Stories" [3].

*Công việc: Tiến hành lại việc đánh giá áp lực kinh kế lên bệnh nhân như thế nào khi có hoặc không có nhà ở và bảo hiểm y tế, dựa trên mô phỏng MCMC thay cho ước

lượng frequentist như trong [2], tương tự như với mô hình [4] nhưng mở rộng và cũng phức tạp hơn do tính chất "nested multi-level" (hierarchy model).

2. Dữ liệu và đánh giá mô hình (dataset and estimations)

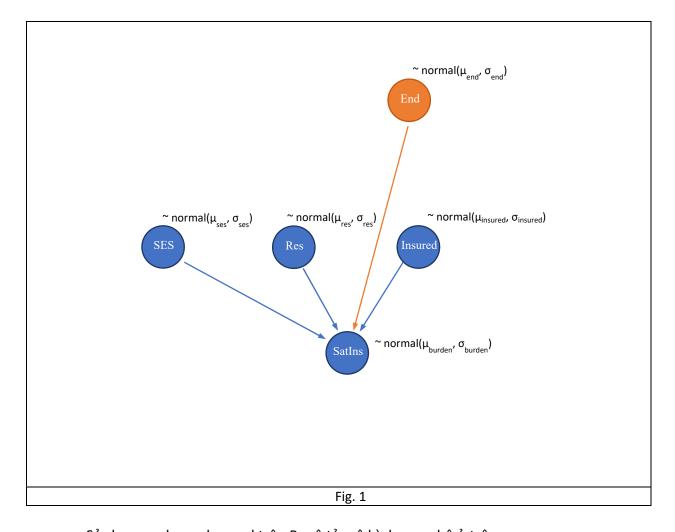
a. Dữ liệu và xây dựng mô hình

data1<-read.csv("/Statistics/1042/1042data/1042data.csv", header = TRUE) head(data1)

Các biến sử dụng:

- Res: bệnh nhân có phải là cư dân ở khu vực khám chữa bệnh không
- Insured: bệnh nhân có chế độ bảo hiểm không
- SES: tình trạng kinh tế của bệnh nhân
- End: kết thúc quá trình điều trị
- Satlns: mức độ hài lòng của bệnh nhân

Tương tự bài toán 2a, ta dựng mô hình quan hệ giữa các biến dạng lưới như sau:



Sử dụng packages bayesvl trên R mô tả mô hình quan hệ ở trên:

```
# Add nodes to model
model <- bayesvl()
model <- bvl_addNode(model, "SatIns", "norm")
model <- bvl_addNode(model, "Res", "cat")
model <- bvl_addNode(model, "Insured", "cat")
model <- bvl_addNode(model, "SES", "cat")
model <- bvl_addNode(model, "End", "cat")

# Add arcs to model
model <- bvl_addArc(model, "Res", "SatIns", "slope")
model <- bvl_addArc(model, "Insured", "SatIns", "slope")
model <- bvl_addArc(model, "SES", "SatIns", "slope")
model <- bvl_addArc(model, "End", "SatIns", "varint")
```

So với mô hình trong Note 2a [4], ta đã thêm 2 nodes cho 2 biến SES và End vào phương trình mô phỏng. Cần đặc biệt chú rằng, biến End có quan hệ với Burden thông qua arc loại varint chứ không phải slope.

Đây là loại quan hệ phân lớp multi-level dạng varying intercept.

Mô hình multi-level:

Nếu như trên mô hình hồi quy đơn giản Note 2a, công thức toán học có dạng:

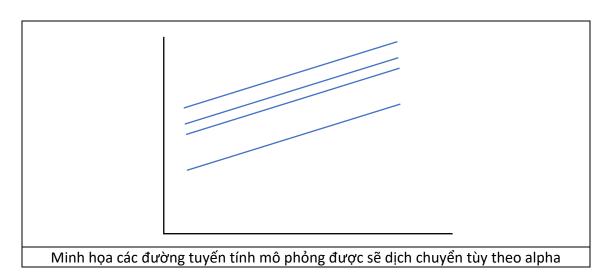
Thì với mô hình đa lớp (multilevel) varying intercept công thức toán học của mô hình sẽ có dạng:

Ví dụ mô hình quan hệ ở trên, nếu các nodes đều là kết nối slope với node Satlns, phương trình toán học mô hình sẽ có dạng:

$$y_i = \alpha_{j[i]} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon_i$$

Trong đó:

- y là biến outcome (Satlns), tức là response variable.
- x1, x2, x3 là các biến độc lập quan hệ slope tương ứng Res, Insured và
- α là hệ số varying intercept của End với j = 1,2,3,4, tương ứng các mức giá trị A, B, C, D của biến End.



So sánh 2 mô hình trên ta có thể thấy một số dạng thay đổi hệ số intercept alpha theo End như sau:

Complete pooling:

Tất cả các mức giá trị của biến End (A,B,C,D) được đối xử như nhau khi hồi quy thường.

Có thể hiểu là các giá trị biến End được cho chung hết vào 1 túi khi thực hiện hồi quy với các biến khác (complete pooling).

```
real alpha;
real beta;
real<lower=0> sigma;
y ~ normal(alpha + beta * x, sigma);
```

No pooling:

Ở mô hình multilevel, các giá trị của biến End được đánh giá độc lập khi hồi quy. Hay có thể hiểu dữ liệu được phân hoạch vào các "túi" khác nhau.

Như vậy việc chọn mô hình multilevel với biến End, ta đã giả định các bệnh nhân sẽ có mức độ hài lòng rất khác nhau ở các kết quả kết thúc điều trị khác nhau.

Bằng cách này mô hình cho phép chia sẻ sức mạnh trong từng mức giá trị của biến End, cho phép có kết quả suy luận hợp lý hơn ở các mức giá trị End có ít dữ liệu, và tính đại diện của mẫu khảo sát không tốt lắm.

```
real alpha[4]; // 4 mức giá trị A, B, C, D
real beta;
real<lower=0> sigma;
y ~ normal(alpha[end] + beta * x, sigma);
```

Partial pooling:

Mô hình không phân hoạch (complete pooling) và mô hình phân hoạch (no pooling) ở trên đều có hạn chế. Với mô hình hồi quy thường, giá trị apha sẽ sẽ được mô phỏng giao động quanh 1 giá trị trung bình mu và không phân hoạch theo các mức biến End.

Ở mô hình phân hoạch no-pooling, các giá trị alpha khác nhau sẽ được mô phỏng cho từng mức giá trị của biến End các giá trị của biến End được đánh giá độc lập khi hồi quy.

Tuy nhiên mô hình này cũng có vấn đề, với một số mức giá trị biến End có số mẫu dữ liệu ít hơn nhiều so với các mức khác có thể cho các mức giá trị alpha (α) cực đoan, ta cũng thường không tin tưởng nhiều nếu mức này chỉ có một vài quan sát (không đại diện tốt cho tập giá trị thực tế).

Mô hình partial pooling tương tự như mô hình đa cấp no pooling nhưng gộp một phần các mô phỏng vector hệ số alpha $_{\rm j}$ ($\alpha_{\rm j}$) về dao động quanh mức trung bình mu (μ) của chúng.

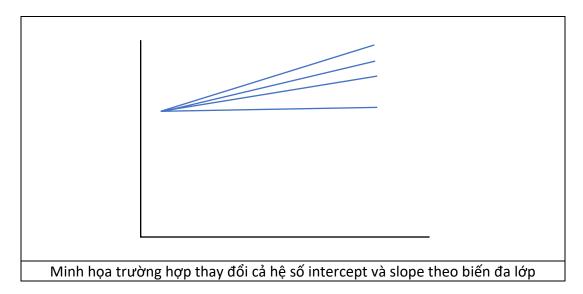
```
real alpha[4]; // 4 mức giá trị A, B, C, D
real beta;
real<lower=0> sigma;
real mu_alpha;
real<lower=0> sigma_alpha;
```

```
mu_alpha ~ normal(0, 100);
alpha ~ normal(mu_alpha, sigma_alpha);
// quy chuẩn alpha quanh mức trung bình
y ~ normal(alpha[end] + beta * x, sigma);
```

Các dạng mô hình multi-level khác:

Tương tự như vậy, với mô hình đa lớp (multilevel) varying slope công thức toán học của mô hình sẽ có dạng:

Có thể nhận thấy lúc này hệ số góc β (β) của biến độc lập x đã được thể hiện dưới dạng đa lớp.



Khi thay đổi cả hệ số intercept (α) và slope (β) theo biến đa lớp ta có mô hình varying intercept and varying slope có dạng:

b. Đánh giá mô hình

Đánh giá mô hình bằng bnlearn

Tính xác suất có điều kiện các biến trong mô hình:

```
bvl_bnBayes(model, data1[c("Res","SatIns","Insured","SES","End")])
```

Kết quả:

```
Bayesian network parameters
Parameters of node SatIns (multinomial distribution)
Conditional probability table:
, , Insured = No, SES = Hi, End = A
   Res
SatIns
           No
                   Yes
Avg 0.6396103896 0.1694630872
NoCmt 0.3279220779 0.6526845638
Satis 0.0162337662 0.1694630872
Unsat 0.0162337662 0.0083892617
, , Insured = Yes, SES = Hi, End = A
   Res
SatIns
           No
                   Yes
Avg 0.4916107383 0.6526845638
NoCmt 0.0083892617 0.0083892617
Satis 0.4916107383 0.3305369128
Unsat 0.0083892617 0.0083892617
, , Insured = No, SES = Lo, End = A
   Res
SatIns
           No
                   Yes
Avg 0.0162337662 0.0431034483
NoCmt 0.6396103896 0.8706896552
Satis 0.0162337662 0.0431034483
Unsat 0.3279220779 0.0431034483
, , Insured = Yes, SES = Lo, End = A
   Res
SatIns
           No
                   Yes
Avg 0.3940000000 0.7780000000
NoCmt 0.0100000000 0.0100000000
```

```
Satis 0.3940000000 0.0100000000
 Unsat 0.2020000000 0.2020000000
, , Insured = No, SES = Med, End = A
   Res
SatIns
           No
                  Yes
Avg 0.0237983035 0.2500000000
NoCmt 0.9738454288 0.7467658473
Satis 0.0011781338 0.0016170763
Unsat 0.0011781338 0.0016170763
, , Insured = Yes, SES = Med, End = A
   Res
SatIns
           No
                  Yes
Avg 0.6723722276 0.9632504056
NoCmt 0.0012054002 0.0001352082
Satis 0.3020732883 0.0286911844
Unsat 0.0243490839 0.0079232017
, , Insured = No, SES = Hi, End = B
   Res
SatIns
           No
                  Yes
Avg 0.0235849057 0.0235849057
NoCmt 0.9292452830 0.9292452830
Satis 0.0235849057 0.0235849057
Unsat 0.0235849057 0.0235849057
, , Insured = Yes, SES = Hi, End = B
   Res
SatIns
           No
                   Yes
Avg 0.5860000000 0.6396103896
NoCmt 0.0100000000 0.0162337662
Satis 0.2020000000 0.3279220779
Unsat 0.2020000000 0.0162337662
, , Insured = No, SES = Lo, End = B
   Res
SatIns
           No
                  Yes
Avg 0.1068329718 0.2020000000
NoCmt 0.8356832972 0.7780000000
Satis 0.0027114967 0.0100000000
Unsat 0.0547722343 0.0100000000
```

```
, , Insured = Yes, SES = Lo, End = B
   Res
SatIns
           No
                  Yes
Avg 0.4673350582 0.6908163265
NoCmt 0.0016170763 0.0051020408
Satis 0.4673350582 0.0051020408
Unsat 0.0637128072 0.2989795918
, , Insured = No, SES = Med, End = B
   Res
SatIns
           No
                  Yes
Avg 0.1070851837 0.1521329987
NoCmt 0.8725272507 0.8146173149
Satis 0.0005046427 0.0015683814
Unsat 0.0198829229 0.0316813049
, , Insured = Yes, SES = Med, End = B
   Res
SatIns
           No
                  Yes
Avg 0.4947524752 0.7851800554
NoCmt 0.0004950495 0.0006925208
Satis 0.3997029703 0.1868421053
Unsat 0.1050495050 0.0272853186
, , Insured = No, SES = Hi, End = C
   Res
SatIns
           No
                   Yes
Avg 0.2500000000 0.4764150943
NoCmt 0.2500000000 0.4764150943
Satis 0.2500000000 0.0235849057
Unsat 0.2500000000 0.0235849057
, , Insured = Yes, SES = Hi, End = C
   Res
SatIns
           No Yes
Avg 0.0431034483 0.0431034483
NoCmt 0.0431034483 0.0431034483
Satis 0.0431034483 0.8706896552
Unsat 0.8706896552 0.0431034483
, , Insured = No, SES = Lo, End = C
```

```
Res
SatIns
           No
                  Yes
Avg 0.2500000000 0.2500000000
NoCmt 0.7252475248 0.2500000000
Satis 0.0123762376 0.2500000000
Unsat 0.0123762376 0.2500000000
, , Insured = Yes, SES = Lo, End = C
   Res
SatIns
           No
                  Yes
Avg 0.5860000000 0.2500000000
NoCmt 0.0100000000 0.2500000000
Satis 0.394000000 0.2500000000
Unsat 0.0100000000 0.2500000000
, , Insured = No, SES = Med, End = C
   Res
SatIns
           No
                 Yes
Avg 0.1015463918 0.4764150943
NoCmt 0.7943298969 0.4764150943
Satis 0.0025773196 0.0235849057
Unsat 0.1015463918 0.0235849057
, , Insured = Yes, SES = Med, End = C
   Res
SatIns
           No
                  Yes
Avg 0.4138225256 0.2500000000
NoCmt 0.0042662116 0.2500000000
Satis 0.4957337884 0.2500000000
Unsat 0.0861774744 0.2500000000
, , Insured = No, SES = Hi, End = D
   Res
SatIns
           No
                  Yes
Avg 0.2500000000 0.0431034483
NoCmt 0.2500000000 0.8706896552
Satis 0.2500000000 0.0431034483
Unsat 0.2500000000 0.0431034483
, , Insured = Yes, SES = Hi, End = D
   Res
```

```
SatIns
           No
                   Yes
Avg 0.2500000000 0.2500000000
NoCmt 0.2500000000 0.2500000000
Satis 0.2500000000 0.2500000000
Unsat 0.2500000000 0.2500000000
, , Insured = No, SES = Lo, End = D
   Res
SatIns
           No
                  Yes
Avg 0.2500000000 0.2500000000
NoCmt 0.2500000000 0.2500000000
Satis 0.2500000000 0.2500000000
Unsat 0.2500000000 0.2500000000
, , Insured = Yes, SES = Lo, End = D
   Res
SatIns
           No
                   Yes
Avg 0.7008670520 0.2500000000
NoCmt 0.0072254335 0.2500000000
Satis 0.2846820809 0.2500000000
Unsat 0.0072254335 0.2500000000
, , Insured = No, SES = Med, End = D
   Res
SatIns
           No
                   Yes
Avg 0.0015683814 0.0162337662
NoCmt 0.9952948557 0.9512987013
Satis 0.0015683814 0.0162337662
Unsat 0.0015683814 0.0162337662
, , Insured = Yes, SES = Med, End = D
   Res
SatIns
           No
                   Yes
Avg 0.5776450512 0.9748322148
NoCmt 0.0042662116 0.0083892617
Satis 0.1680887372 0.0083892617
Unsat 0.2500000000 0.0083892617
Parameters of node Res (multinomial distribution)
Conditional probability table:
          Yes
    No
```

0.4458175 0.5541825

Parameters of node Insured (multinomial distribution)

Conditional probability table:

No Yes

0.3070342 0.6929658

Parameters of node SES (multinomial distribution)

Conditional probability table:

Hi Lo Med

0.03929024 0.09442332 0.86628644

Parameters of node End (multinomial distribution)

Conditional probability table:

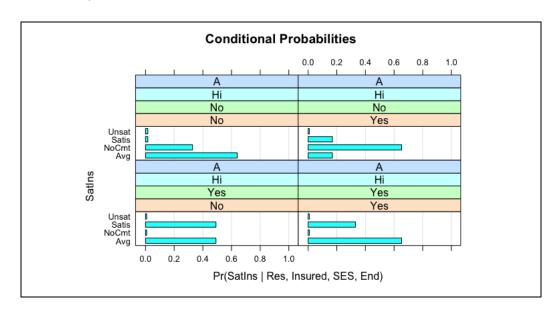
A B C I

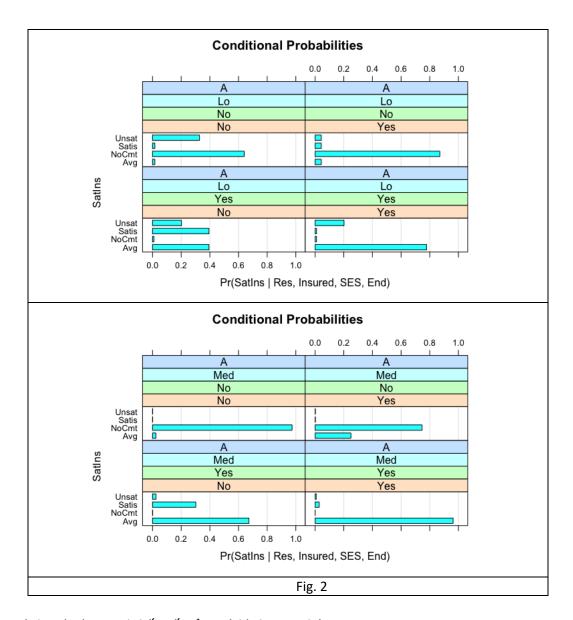
 $0.51473384\ 0.37690114\ 0.04705323\ 0.06131179$

Nếu vẽ các xác suất có điều kiện này:

bvl_bnBarchart(model, data1[c("Res","SatIns","Insured","SES","End")])

Ta sẽ thu được:





Đánh giá sức mạnh kết nối của mô hình quan hệ:

```
>bvl_bnStrength(model, data1[c("Res","SatIns","Insured","SES","End")])
from to strength
1 Res SatIns 5.314485e-03
2 Insured SatIns 4.063128e-125
3 SES SatIns 1.014490e-01
4 End SatIns 9.498561e-01
```

Chức năng này gọi package bnlearn để đo lường quan hệ giữa các biến trong lưới model thông qua xác suất tương ứng từng cung (arc). Giá trị strength được thể hiện chính là giá trị p-value.

Mô hình hiện tại có 4 arc, các giá trị p-values đều nhỏ hơn 0,05 và support rất tốt từ dữ liệu.

c. Mô hình hồi quy bayesian

Với mô hình quan hệ đã tạo ở trên, nếu viết ở dạng hồi quy tuyến tính, ta có phương trình:

```
y_{\text{satins}}\left[i\right] = \alpha_{\text{end}}[x_{\text{end}}[i]] + \beta_{\text{res}}x_{\text{res}}[i] + \beta_{\text{insured}}x_{\text{insured}}[i] + \beta_{\text{ses}}x_{\text{ses}}[i] + \varepsilon_{\text{satins}}[i]
```

Trên phương trình ta thấy có 3 hệ số β_{res} , $\beta_{insured}$ và β_{ses} là 3 hệ số hồi quy "slope" là hệ số ảnh hưởng đến hệ số góc của đường hồi quy, 1 hệ số varying intercept là α_{end} . Khi tạo mô hình quan hệ các biến ở trên ta đã định nghĩa arc với loại quan hệ "slope":

```
model <- bvl_addArc(model, "Res", "Burden", "slope")
model <- bvl_addArc(model, "Insured", "Burden", "slope")
model <- bvl_addArc(model, "SES", "Burden", "slope")
```

Và quan hệ varying intercept

```
model <- bvl_addArc(model, "SES", "Burden", "varint")
```

Vì vậy, để tạo mô hình hồi quy trên bayesvl chỉ cần thực hiện lệnh:

```
# Generate the stan code for model model_string <- bvl_model2Stan(model) cat(model_string)
```

Mô hình được tạo sẽ ở dạng no-pooling varying intercept with mixed effects model_string chứa code stan cho mô hình hồi quy tuyến tính.

Code stan được tạo:

```
data{
    // Define variables in data
    int<lower=1> Nobs; // Number of observations (an integer)
    real SatIns[Nobs]; // outcome variable
    int Res[Nobs];
    int Insured[Nobs];
    int SES[Nobs];
```

```
int End[Nobs];
   int<lower=1> NEnd; // Level of End (an integer)
}
parameters{
  // Define parameters to estimate
   real<lower=0> sigma SatIns;
   real b Res SatIns;
   real b Insured SatIns;
   real b SES SatIns;
  real a End 0;
   real<lower=0> sigma End;
  vector[NEnd] u_End;
transformed parameters{
  // Transform parameters
  vector[NEnd] a End;
   real mu_SatIns[Nobs];
  // Varying intercepts definition
  for(k in 1:NEnd) {
    a_{end}[k] = a_{end} + u_{end}[k];
  }
  for (i in 1:Nobs) {
    mu SatIns[i] = b Res SatIns * Res[i] + b Insured SatIns * Insured[i] +
b_SES_SatIns * SES[i] + a_End[End[i]];
  }
model{
  // Priors
  b Res SatIns ~ normal(0,100);
   b_Insured_SatIns ~ normal(0,100);
  b SES SatIns ~ normal(0,100);
   sigma End \sim normal(0,100);
   u End ~ normal(0, sigma End);
   // Likelihoods
   SatIns ~ normal(mu_SatIns, sigma_SatIns);
}
```

Chạy mô phỏng mô hình:

```
dat1042 <- with(data1,
list(Nobs = length(Res),
```

```
NEnd = length(unique(End)),
SES = as.numeric(SES),
End = as.numeric(End),
Res = as.numeric(Res),
Insured = as.numeric(Insured),
SatIns = as.numeric(SatIns)))

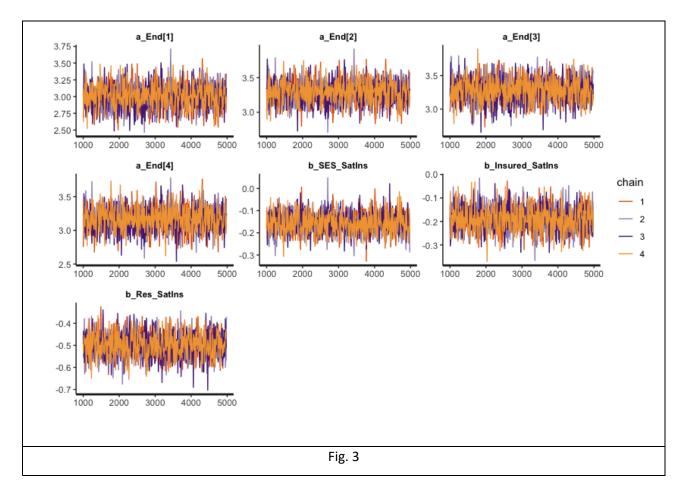
# Fit the model
model <- bvl_modelFit(model, dat1042, warmup = 2000, iter = 5000, chains
= 4, cores = 4)
```

d. Đánh giá mô hình

Đánh giá sự hội tụ của mô hình có thể thực hiện thông qua trace plot:

```
traceplot(fit)
```

Đồ họa thu được trong Figure 3:



Ở mô hình này ta thấy trace plots tốt, không có các chuỗi dị thường divergent chains

Bước tiếp theo ta có thể xem summary output.

```
summary(fit)
```

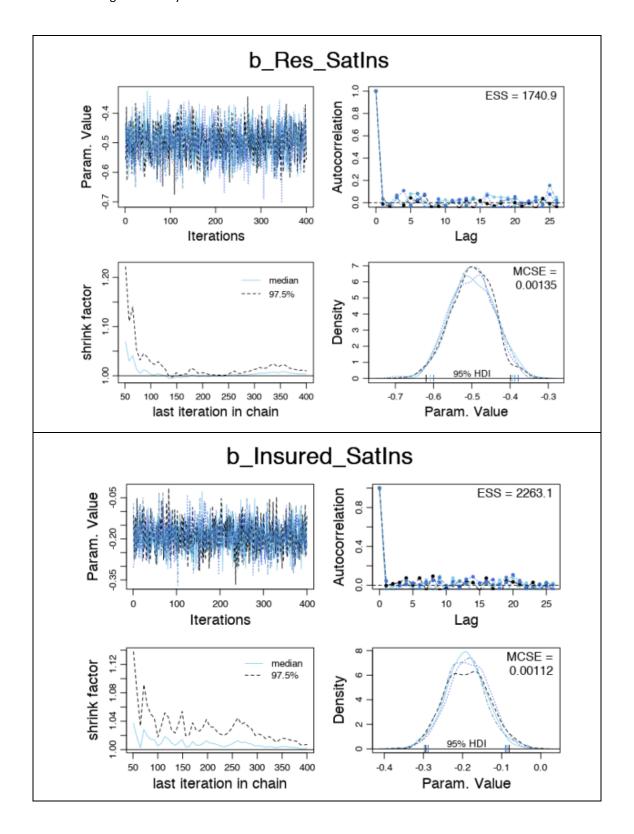
Cơ bản kết quả thu được khá tốt, đảm bảo quy cách kỹ thuật của mô phỏng:

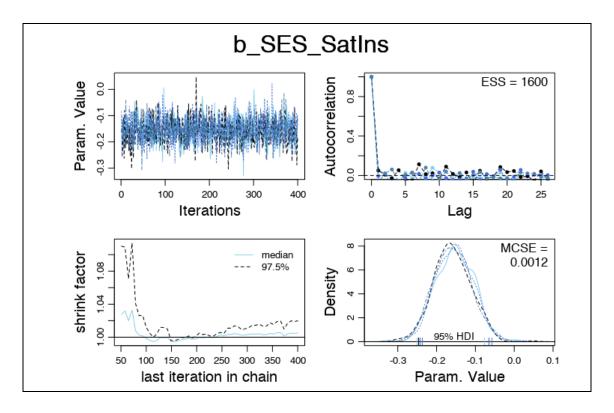
```
> summary(fit)
Model Info:
Nodes:
             3
            2
Arcs:
Scores:
Formulas:
             Satins ~ a_End[End] + b_Res * Res + b_Insured * Insured + b_SES * SES
Estimates:
4 chains, each with iter=5000; warmup=1000; thin=10;
         mean se mean sd 2.5% 25% 50% 75% 97.5% n eff Rhat
a End[1]
             3.00 0.00 0.18 2.65 2.88 3.00 3.13 3.36 1450 1.00
a End[2]
             3.29 0.00 0.17 2.96 3.18 3.30 3.41 3.61 1433 1.00
a_End[3] 3.29 0.00 0.18 2.93 3.16 3.29 3.41 3.64 1660 1.00 a_End[4] 3.17 0.00 0.18 2.81 3.05 3.17 3.20 3.51 1500 1.00
b SES Satins -0.16 0.00 0.05 -0.24 -0.19 -0.16 -0.12 -0.06 1567 1.00
b Res Satins -0.50 0.00 0.06 -0.61 -0.54 -0.50 -0.46 -0.39 1603 1.00
              0.72  0.00  0.02  0.69  0.71  0.72  0.73  0.75  1899  1.00
sigma SatIns
              0.34  0.03  0.71  0.08  0.15  0.21  0.34  1.19  418  1.01
sigma End
```

Chúng ta có thể thấy rằng effective sample size (n_eff) là ổn (trên 1000 thường là một dấu hiệu tốt), một số liệu chẩn đoán khác là giá trị Rhat cũng cho thấy sự hội tụ (Rhat bằng 1 là một dấu hiệu tốt, hơn 1.1 có thể chỉ ra là có vấn đề).

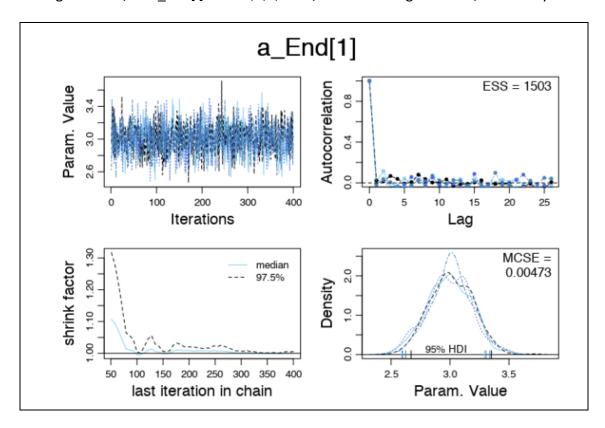
```
bvl_mcmcDiag(model, "a_End")
bvl_mcmcDiag(model, "b_Res_SatIns")
bvl_mcmcDiag(model, "b_Insured_SatIns")
bvl_mcmcDiag(model, "b_SES_SatIns")
```

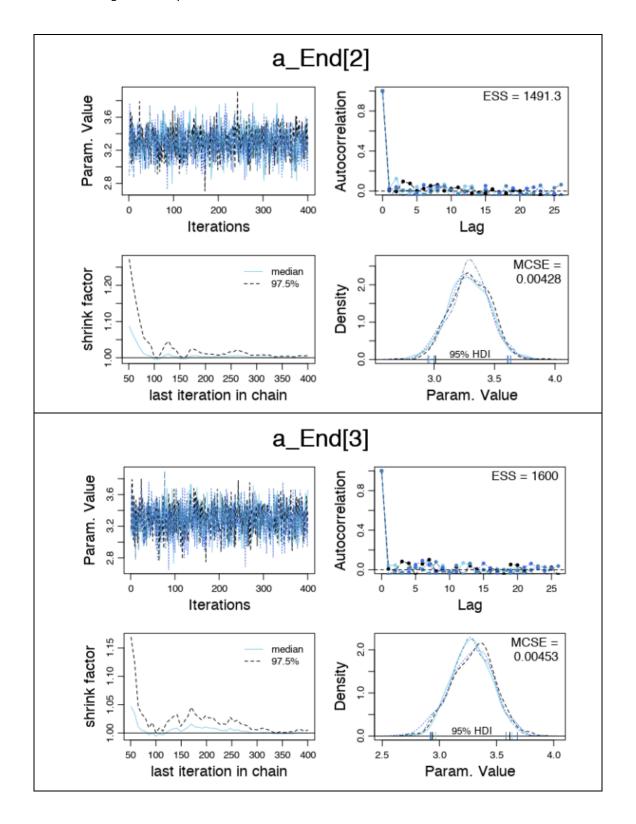
Chúng ta có được các đồ thị kiểm tra quy cách mô phỏng cho bộ hệ số b_SES_SatIns và b_Insured_SatIns và b_Insured_SatIns như sau:

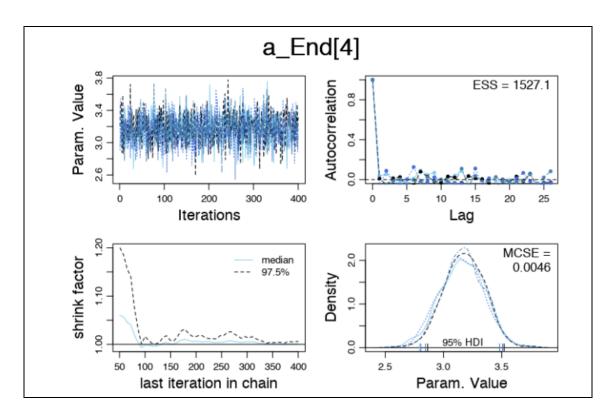




Riêng vector hệ số a_End[i] với i=1,2,3,4 được tách ra trong các đồ họa dưới đây:

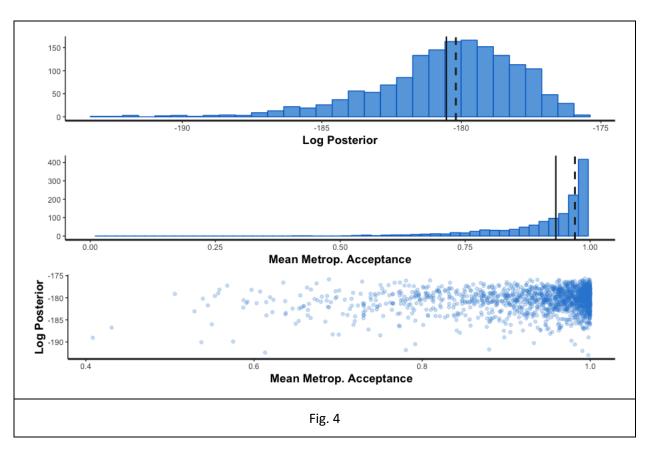






Việc kiểm tra tính hợp lệ, tốc độ hội tụ của xích, Rhat và eff có thể tiến hành hoàn toàn bình thường.

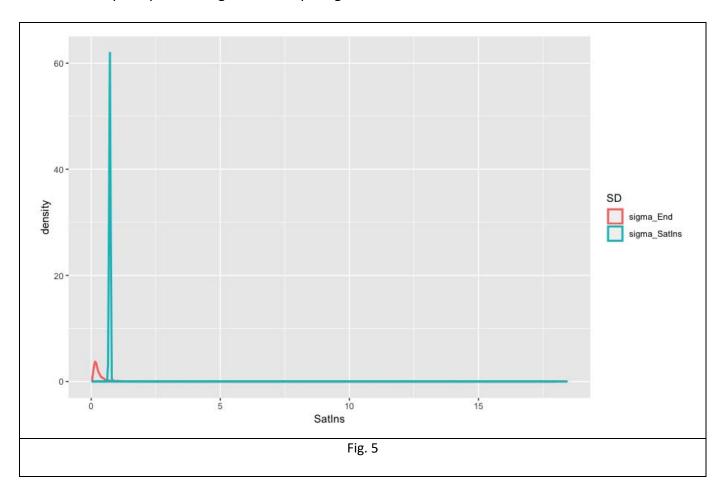
Nếu vẽ log posterior:



Đến đây, các kiểm tra posterior samples cơ bản hoàn thành. Trên phương diện kỹ thuật thống kê, mô hình có thể coi là đáp ứng được các yêu cầu (satisfactory).

e. Kết quả hồi quy

Các phân phối của sigma sau mô phỏng:

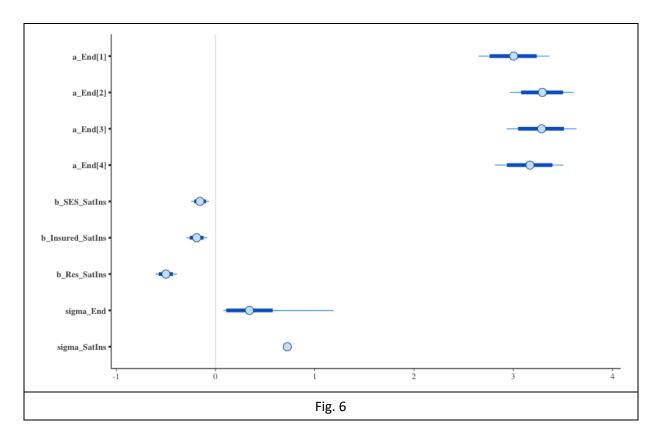


Phân phối các hệ số (Fig. 5):

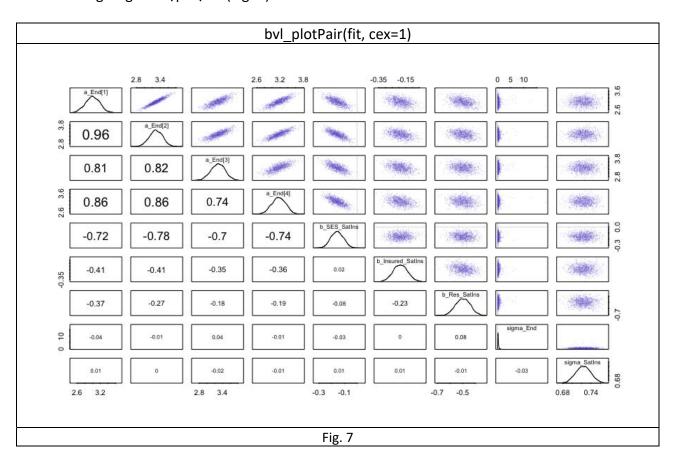
Nếu vẽ interval phân phối với mức đánh giá 80% (Fig. 6)

```
bvl_plotIntervals(fit, prob = 0.8, prob_outer = 0.95, color_scheme = "brightblue")
```

Figure 6:

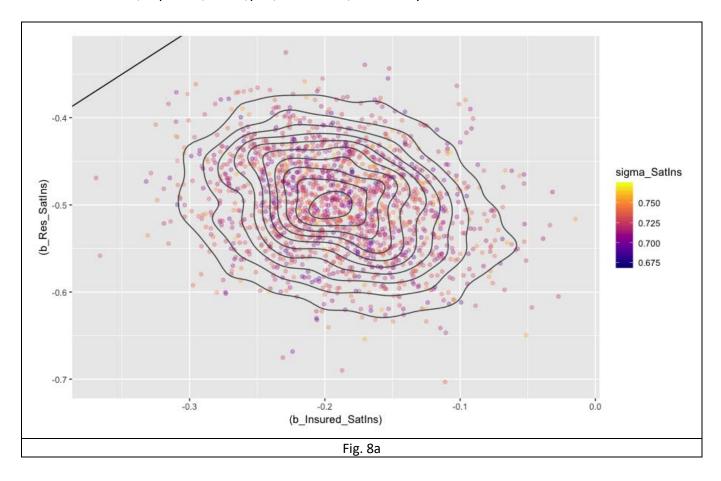


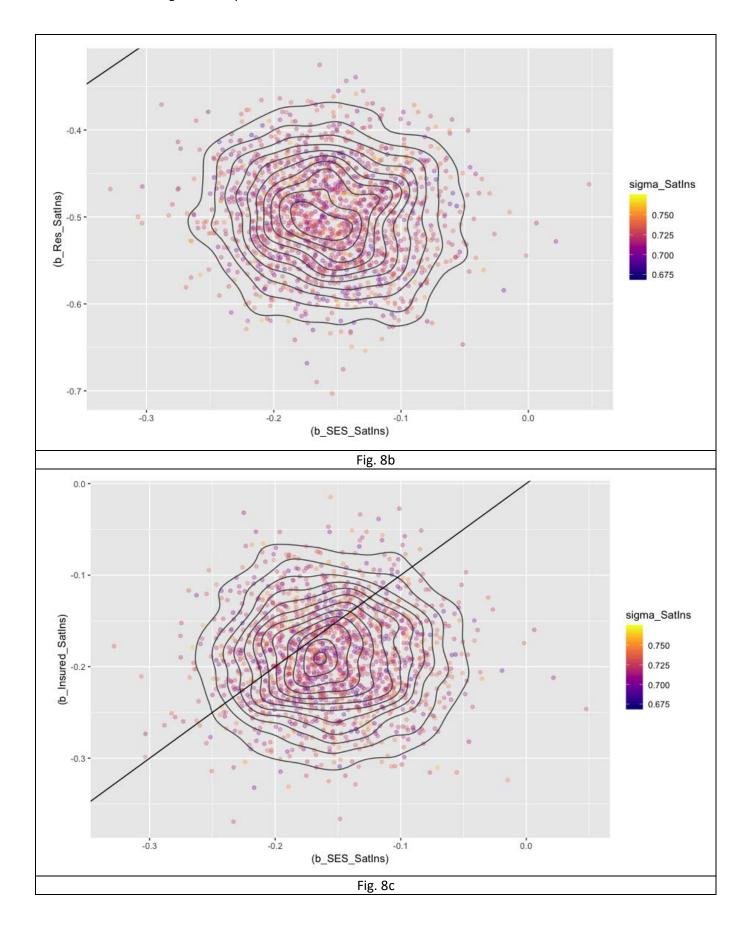
Tương ứng các cặp hệ số (Fig. 7):



Nếu so sánh riêng 2 tham số b_Res, b_Insured:

Ba đồ họa quan hệ các cặp hệ số thu được dưới đây:





Các hình thuộc Figure 8 rất đáng chú ý và cần giải thích kỹ lưỡng. Chúng cũng có giá trị biện luận (logic) rất tốt, đặc trưng cho phương pháp Bayesian. Một vài ý giải thích cơ bản như sau.

Trên hình (Fig. 8a) ta thấy hệ số nghiêng nhiều về Res. Ảnh hưởng Res lên SatIns mạnh hơn so với insured. Tuy nhiên 2 hệ số này cùng dấu (-) ảnh hưởng cùng xu hướng.

Trên hình (Fig. 8b). Ảnh hưởng Res lên SatIns mạnh hơn so với SES. Tuy nhiên 2 hệ số này cùng dấu (-) ảnh hưởng cùng xu hướng.

Trên hình (Fig. 8c). Ảnh hưởng SES và Insured lên SatIns gần tương đương nhau, 2 hệ số này cùng dấu (-) ảnh hưởng cùng xu hướng.

Khôi phục lại phương trình hồi quy ta sẽ có 4 phương trình:

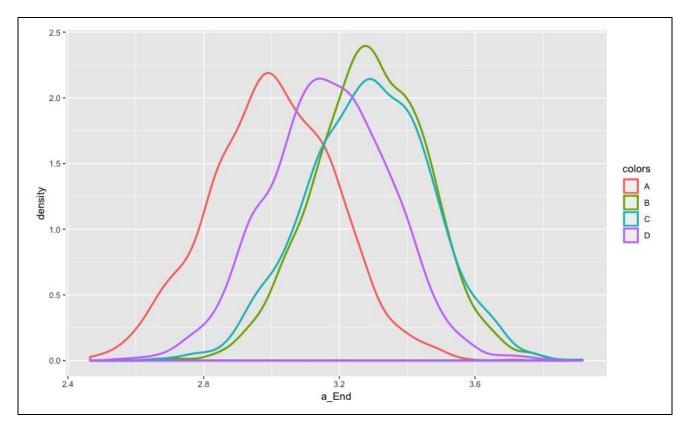
Đối với các bệnh nhân khỏi bệnh End = A ta có:

- Đối với các bệnh nhân cần tiếp tục điều trị End = B ta có:

- Đối với các bệnh nhân dừng giữa chừng End = C ta có:

- Đối với các bệnh nhân kết thúc sớm End = D ta có:

So sánh hệ số alpha của 4 giá trị End được biểu diễn qua đồ thị dưới đây:



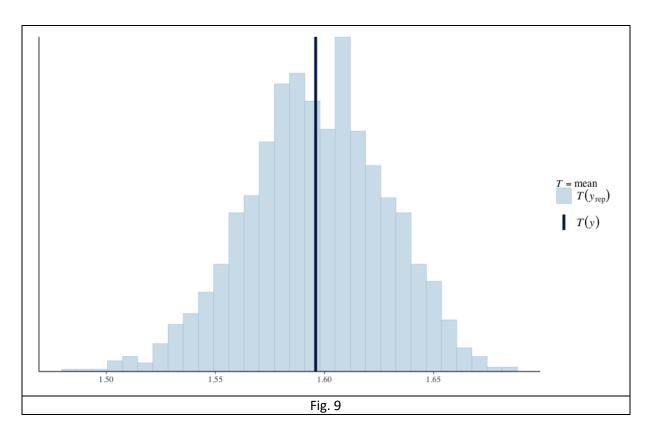
*Câu hỏi: Liệu có thể suy nghĩ gì khi dường như simulated data chỉ ra rằng:

"Những bệnh nhân khỏi bệnh ít hài lòng với dịch vụ nhất?!"

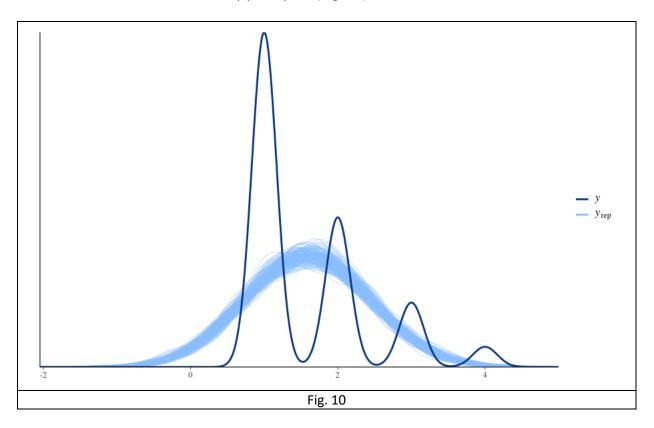
f. Posterior predictive checks

Kết quả y_rep sẽ cho ta thấy phân phối density distribution of the model predictions compare with that of the raw data

Trong đồ thị hình Fig. 9, ta thấy mean posterior (đường kẻ xanh) được align khá sát với raw data.



Nếu vẽ mật độ, ta có thể thấy phân phối (Fig. 10) như sau:



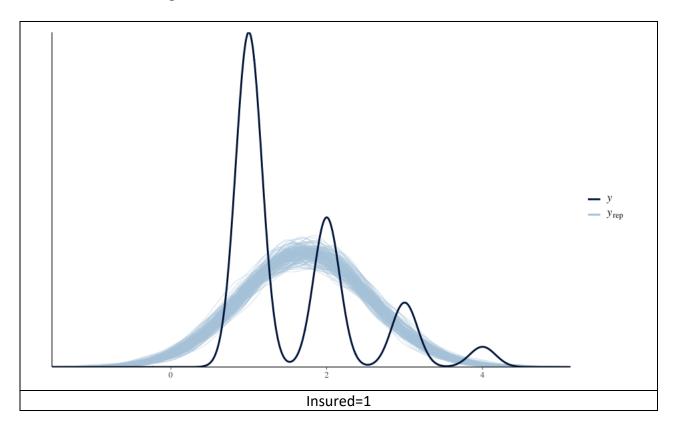
Cài thêm code để so sánh nhóm có bảo hiểm và không có bảo hiểm về mức độ hài lòng (đối với dịch vụ bảo hiểm SatIns):

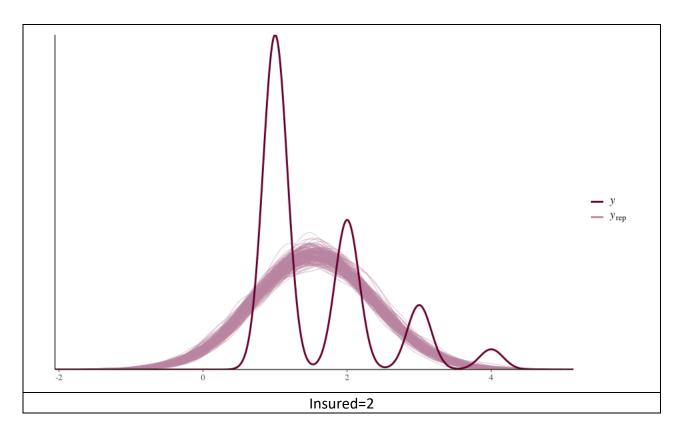
```
ppc_string <- "
  vector[Nobs] y_rep_insured1;
  vector[Nobs] y_rep_insured2;

for (i in 1:num_elements(y_rep_insured1)) {
    y_rep_insured1[i] = normal_rng(b_Res_SatIns * Res[i] + b_Insured_SatIns * 1 +
    b_SES_SatIns * SES[i] + a_End[End[i]], sigma_SatIns);
  }
  for (i in 1:num_elements(y_rep_insured1)) {
    y_rep_insured2[i] = normal_rng(b_Res_SatIns * Res[i] + b_Insured_SatIns * 2 +
    b_SES_SatIns * SES[i] + a_End[End[i]], sigma_SatIns);
  }
  "

# Fit the model
model <- bvl_modelFit(model, dat1042, warmup = 2000, iter = 5000, chains = 4,
    cores = 4, ppc = ppc_string)</pre>
```

Trong phần code đưa thêm ppc_string ở trên, ta có chèn thêm 2 biến y_rep_insured1 và y_rep_insured2 để đánh giá kết quả trong 2 trường hợp bệnh nhân có và không bảo hiểm.





*<u>Câu hỏi</u>: Liệu có thể đưa ra đánh giá gì từ simulated data khi thực tế kết quả cho thấy: Dường như có bảo hiểm và không có bảo không làm người bệnh hài lòng hơn bao nhiêu?!

Phần <u>Note 2b</u> cần được ghép với kết quả ở Note 2a [4], triển khai giải thích từ đơn giản đến phức tạp, để hoàn thiện một bức tranh chung về nghiên cứu. Các phát triển tiếp theo của BayesVL sẽ tìm hiểu tính thân thiện của rstanarm [7-8] để bổ sung tính năng, đặc biệt đồ họa và thư viện hàm dễ sử dụng.

References:

- [1] Vuong QH, & La VP. (2019). BayesVL package for Bayesian statistical analyses in R. Github: BayesVL version 0.6.5; DOI: 10.31219/osf.io/ya9u6. Available from: https://github.com/sshpa/bayesvl.
- [2] Vuong QH. (2015). Be rich or don't be sick: estimating Vietnamese patients' risk of falling into destitution. *SpringerPlus*, **4**(1), 529. DOI: 10.1186/s40064-015-1279-x.
- [3] Ho MT, La VP, Nguyen MH, Vuong TT, Nghiem KCP, Tran T, Nguyen HKT, & Vuong QH. (2019). Health care, medical insurance, and economic destitution: A dataset of 1042 stories. *Data*, 4, 57. DOI: 10.3390/data4020057.
- [4] Vuong QH, & La VP. (2019). Úng dụng BayesVL v0.6.5 mô phỏng MCMC với bài toán burden ~ res + insured sử dụng dữ liệu thực 1042 quan sát. *OSF*Preprint; DOI: 10.31219/osf.io/9rhyk. Available from: https://osf.io/9rhyk/>.

- [5] McElreath R. (2018). Statistical Rethinking: A Bayesian Course with Examples in R and Stan. London: Chapman and Hall/CRC.
- [6] Scutari M. (2010). Learning Bayesian networks with the bnlearn R package. Journal of Statistical Software, 35(3), 1-22.
- [7] Muth C, Oravecz Z, & Gabry J. (2018). User-friendly Bayesian regression modeling: A tutorial with rstanarm and shinystan. *Quantitative Methods for Psychology*, 14(2), 99-119.
- [8] Gabry J, & Goodrich B. (2016). rstanarm: Bayesian applied regression modeling via Stan. R package version 2.10.0. Available from: https://cran.r-project.org/web/packages/rstanarm/index.html