补充材料

潘晚坷　　温秀娟　　金海洋

2023-05-15

目录

1 数据模拟

我们借助一个假想的心理学实验展示如何模拟虚拟数据。在模拟实验中，40 名抑郁症患者和 40 名健康对照组被试观看 30 张积极和 30 张中性图片，期间我们采集了他们的脑电数据。因变量是晚期正电位（late positive potentials, LPP）的波幅。简单来说，这是一个 2 (组别 `group`：抑郁症组 `depression`、对照组 `control`) × 2 (图片类型 `type`：积极 `positive`、中性 `neutral`) 的混合实验设计，其中组别为被试间因素，图片类型为被试内因素。该假想实验的数据是使用 faux 工具包生成 (DeBruine 2023)，下面是模拟这个实验所预设的参数。

```
subj_n <- 80   # 总被试量：抑郁患者 30 人，健康对照组被试 30 人
trial_n <- 30  # 每张图片呈现的次数

# 固定效应
```

```
b0 <- 0.5       # 截距 (所有条件的均值)
b1 <- 6.5       # 图片类型的固定效应 (主效应)
b2 <- 0.1       # 组别的固定效应 (主效应)
b3 <- 0.1       # 图片类型与组别的交互作用

# 随机效应
u0s <- 2        # 被试的随机截距
u1s <- 2        # 被试的随机斜率 (图片类型)

# 误差项
sigma <- 2

# 生成假定实验的条件的数据矩阵
df_simu <- add_random(subj = subj_n) %>%
  # 添加被试的组别信息（被试间）
  add_between("subj", group = c("depression", "control")) %>%
  # 添加图片类型的信息（被试内）
  add_within("subj", type = c("netural","positive")) %>%
  # 每种图片呈现 30 次
  add_random(trial = trial_n) %>%
  # 图片类型的编码：中性 =-0.5；正性 =0.5
  add_contrast("type", "anova", colnames = "type_code") %>%
  # 被试组别的编码：抑郁症组 =-0.5；控制组 =0.5
  add_contrast("group", "anova",colnames = "group_code") %>%
  # 添加基于被试的随机截距和斜率 (图片类型)
  add_ranef("subj", u0s = u0s, u1s = u1s, .cors=0.5) %>%
  # 添加观察值的误差项
  add_ranef(sigma = sigma) %>%
  # 最后根据设置的固定效应和随机效应参数值，生成因变量。
  mutate(LPP = (b0+u0s) +          # 截距
           (b1+u1s) * type_code + # 图片材料的斜率
           b2 * group_code +       # 组别的斜率
           b3 * type_code * group_code +    # 交互作用
           sigma)             # 误差项
```

```
df_simu <- df_simu %>%
  select(subj, group, type, LPP) # 去除冗余的信息

# 保存生成的数据
save(df_simu, file = "simulated_data.rdata")

# 查看生成的数据
head(df_simu,10)
```

```
## # A tibble: 10 x 4
##    subj   group      type      LPP
##    <chr>  <fct>      <fct>   <dbl>
##  1 subj01 depression netural -1.64
##  2 subj01 depression netural -4.00
##  3 subj01 depression netural -1.53
##  4 subj01 depression netural -3.22
##  5 subj01 depression netural -4.34
##  6 subj01 depression netural  1.49
##  7 subj01 depression netural -3.46
##  8 subj01 depression netural  2.93
##  9 subj01 depression netural -3.27
## 10 subj01 depression netural -1.44
```

2 方差分析

我们使用 bruceR 包进行混合实验设计方差分析 (Bao 2023)。以下为示例代码与结果。其中，图片类型的主效应显著 ($F(1,78) = 983.78, p < 0.01, \eta_p^2 = .93$)，而组别的主效应、组别与图片类型的交互作用不显著 ($p > 0.1$)。

```
df_simu = df_simu |> mutate(
  组别 = factor(group, labels = c(" 抑郁组"," 控制组")),
  图片类型 = factor(type, labels = c(" 中性图片"," 积极图片"))
)
```

```r
# Two-way mixed ANOVA test
df_simu |> bruceR::MANOVA(
  subID = "subj",
  between = " 组别",
  within = " 图片类型",
  dv = "LPP",
  digits = 2
  # file = " 重复测量方差分析结果.doc"
)
```

```
##
##     * Data are aggregated to mean (across items/trials)
##     if there are >=2 observations per subject and cell.
##     You may use Linear Mixed Model to analyze the data,
##     e.g., with subjects and items as level-2 clusters.

##
## ====== ANOVA (Mixed Design) ======
##
## Descriptives:
##
##  "组别" "图片类型"  Mean   S.D.  n
##
##  抑郁组    中性图片 -2.71 (2.12) 40
##  抑郁组    积极图片  3.78 (2.56) 40
##  控制组    中性图片 -2.43 (1.97) 40
##  控制组    积极图片  4.07 (2.62) 40
##
## Total sample size: N = 80
##
## ANOVA Table:
## Dependent variable(s):      LPP
## Between-subjects factor(s): 组别
## Within-subjects factor(s):  图片类型
## Covariate(s):               -
```

```
##
##                        MS  MSE df1 df2      F     p      ²p [90% CI of  ²p]  ²G
##
## 组别                 3.19 9.18   1  78   0.35  .557          .00 [.00, .06] .00
## 图片类型          1689.20 1.72   1  78 983.78 <.001 ***      .93 [.90, .94] .67
## 组别 * 图片类型      0.00 1.72   1  78   0.00  .969          .00 [.00, .00] .00
##
## MSE = mean square error (the residual variance of the linear model)
##  ²p = partial eta-squared = SS / (SS + SSE) = F * df1 / (F * df1 + df2)
##  ²p = partial omega-squared = (F - 1) * df1 / (F * df1 + df2 + 1)
##  ²G = generalized eta-squared (see Olejnik & Algina, 2003)
## Cohen’s f² =  ²p / (1 -  ²p)
##
## Levene’s Test for Homogeneity of Variance:
##
##                Levene’s F df1 df2     p
##
## DV: 中性图片       0.772   1  78  .382
## DV: 积极图片       0.027   1  78  .869
##
##
## Mauchly’s Test of Sphericity:
## The repeated measures have only two levels. The assumption of sphericity is always met.
```

3 不收敛 MCMC 链演示

我们使用 bayesplot 和 posterior 包模拟和绘制 MCMC 链 (Gabry et al. 2019; Vehtari et al. 2021)。图 1 为 MCMC 链不收敛的示例。其中，四条链存在明显的分离，并且第四条链并没有达到稳定分布。$\hat{R} = 2.73$ 的结果大于 1.1，说明该 MCMC 链的收敛结果很差。

```r
# 模拟生成 4 条不收敛的 MCMC 链，每条链包含 4000 个样本
n_chains <- 4
chain_length <- 4000
```

```
# 生成三种链，一种收敛的链 good_chains，两种不收敛的链 bad_chains0 和 bad_chains1
good_chains <- rbeta(n = chain_length*n_chains, shape1 = 2, shape2 = 5)
good_chains <- matrix(good_chains, nrow = n_chains)
bad_chains <- matrix(
  rnorm(chain_length*n_chains, mean = sort(good_chains), sd = 0.05),
  nrow = n_chains)

chains <- array(0, dim = c(chain_length, n_chains,1))
chains[,,1] = bad_chains
dimnames(chains) <- list(
  Iteration = NULL,
  Chain = paste0("chain:",1:n_chains),
  Parameter = c("bad_chains")
)

# 绘制轨迹图
mcmc_trace(chains)
```

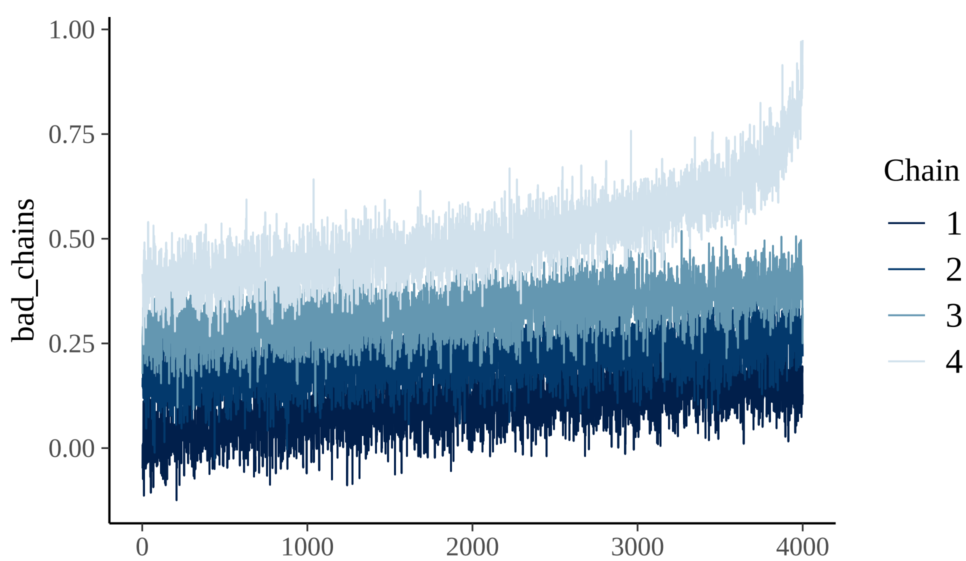

图 1: 不收敛 MCMC 链演示图

```r
# 计算 rhat
rhat( extract_variable_matrix(chains, "bad_chains") ) # 2.73128
```

```
## [1] 2.728107
```

参考文献

Bao, Han-Wu-Shuang. 2023. *bruceR: Broadly Useful Convenient and Efficient r Functions*. https://CRAN.R-project.org/package=bruceR.

DeBruine, Lisa. 2023. *Faux: Simulation for Factorial Designs*. Zenodo. https://doi.org/10.5281/zenodo.2669586.

Gabry, Jonah, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman. 2019. "Visualization in Bayesian Workflow." *J. R. Stat. Soc. A* 182: 389–402. https://doi.org/10.1111/rssa.12378.

Vehtari, Aki, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. 2021. "Rank-Normalization, Folding, and Localization: An Improved Rhat for Assessing Convergence of MCMC (with Discussion)." *Bayesian Analysis*.