arXiv:2305.05099v1 [stat.ME] 9 May 2023

用于分析重复尝试设计的狄利克雷过程混合模型

Daniels, M.J.[1], Lee, M.[2], 和 Feng, W.[3]

[1]mdaniels@stat.ufl.edu，佛罗里达大学统计学系，盖恩斯维尔，佛罗里达州

[2]minjilee101@gmail.com，爱德华兹生命科学，尔湾，加利福尼亚州[3]
weifeng.vivi@gmail.com，Keros Therapeutics，列克星敦，马萨诸塞州

2023年5月10日

摘要:

在纵向研究中，基线测量后进行多次尝试以收集测量值并不罕见。记录这些尝试是否成功，有助于评估缺失数据假设。这是因为，在多次尝试失败后才提供数据的受试者，其测量值可能不同于经过较少尝试后便提供测量值的受试者。以往针对这类设计的模型是参数模型，和/或不允许进行敏感性分析。对于前者，人们始终会担心模型设定错误；对于后者，在存在缺失数据的情况下进行推断时，敏感性分析至关重要。在此，我们提出一种新方法，通过使用贝叶斯非参数方法对观测数据分布进行建模，尽量减少模型设定错误带来的问题。我们还提出了一种用于识别的新方法

和敏感性分析。我们重新分析了一项涉及严重精神疾病患者的临床试验中的重复尝试数据，并开展模拟，以更好地了解我们方法的性质。

关键词：贝叶斯非参数方法；信息先验；缺失数据

1 引言

在随访研究中，通常会在基线测量后进行多次尝试，以收集测量数据（例如，Wood 等人，2006年）。我们将这类设计称为重复尝试设计。联系尝试的次数反映了获取结局数据的难度，并可能为未观测响应的分布提供额外信息（Alho，1990；Lin 和 Schaeffer，1995；Wood 等人，2006年）。未响应或延迟响应往往与结局数据有关。参与者通常更不愿意或不太配合回答包含敏感问题的问卷，例如涉及收入、饮酒或吸烟行为的问题。在这些研究中，随机缺失（MAR）假设很可能过于严格，也不符合实际情况。针对这类设计的模型称为重复尝试模型（RAM）。

此前关于重复尝试设计的大多数论文，都是在选择模型（SM）的框架下利用多次尝试的信息（Alho, 1990; Wood et al., 2006; Jackson et al., 2010, 2012; Chen et al., 2018）。选择模型的一个优点是，其因子分解使完整数据响应模型中的参数易于解释，而这些参数通常是人们最关心的。然而，这类模型也有局限：它们对模型设定较为敏感（Kenward, 1998），且不允许像 Daniels 和 Hogan（2008）那样进行敏感性分析，而这类分析对于评估不可验证的缺失数据假设十分重要。

模式混合模型（PMMs）将受试者划分为不同模式，并将响应分布建模为各模式上的混合分布（Little，1993，1995）。PMM 与 PMM 所采用的外推因子分解之间的对应关系，使其自然适用于敏感性分析（Daniels 和 Hogan，2008），并且已有研究将其用于重复尝试设计（Daniels 等人，2015）。但是，当模式数量相对较多时，某些模式中的受试者会非常少，从而导致模式特异性参数无法识别或识别较弱（Little，1995；Daniels 和 Hogan，2000）。潜变量方法（Roy，2003；Roy 和 Daniels，2008；Lin 等人，2000，2004；Muth´en 等人，2003）

引入离散潜变量，以数据驱动的方式发现并归类模式，也有人采用临时性方法（Daniels 等人，2015）；遗憾的是，前者依赖对“类别”内分布和离散潜变量的参数设定来确保可识别性，且对模型设定错误并不稳健。贝叶斯非参数方法也可用于处理稀疏模式并避免模型设定错误带来的问题（Linero 和 Daniels，2015）。本文提出一种贝叶斯非参数方法。该方法避免了 Roy（2003）及先前 RAM 中用于合并脱落模式的方法所采用的参数假设。此外，我们将 Linero 和 Daniels（2015）中的贝叶斯非参数方法调整至不同的情境（该研究考察的是一项存在脱落的纵向研究），并在模型设定中加入协变量（这也能处理协变量中的可忽略缺失）。我们还提出了用于识别的新型灵活先验分布，以扩展 Daniels 等人（2015）此前的研究。

本研究的动机源自 QUATRO试验（Gray 等人，2006年）。QUATRO试验是一项单盲、多中心随机对照试验，旨在评估针对精神分裂症的依从性治疗的效果。该试验在四个中心开展，基线时纳入了409名受试者。随后，研究人员尝试在52周随访结束时收集生活质量结局。研究人员最多进行了九次尝试，但仍有一些未应答者。在干预组和对照组中，分别有204人中的29人（14%）和205人中的13人（6%）未能提供结局数据。由于在尝试3至9次期间收集到结局数据的受试者数量较少，先前的建模将这些受试者合并为一种模式（Daniels 等人，2015年）。

在此，我们为观测数据提出一种狄利克雷过程混合（DPM）模型。第2节介绍观测数据模型、观测数据模型参数的先验分布，以及针对给定观测数据时缺失数据的未识别条件分布参数的新型先验分布；该分布称为外推分布（Daniels 和 Hogan，2008年）。第3节中，我们使用该模型重新分析 QUATRO 数据。第4节介绍模拟研究。最后，我们进行讨论。

见第 5 节。

2 模型

我们提出了一种狄利克雷过程混合（DPM）模型，用于刻画观测数据分布，并基于观测数据中“对应的”条件分布，为外推分布中的未识别参数设定先验分布。

2.1 观测数据模型

将结果记为 Y，将基线协变量向量记为 $\boldsymbol{X}$。 R 表示收集到结果之前的尝试次数；研究中最多进行 K 次尝试。令 $R = K + 1$ 表示经过最大尝试次数后仍未获得结果。令 Z 表示干预（治疗取 1，对照取 0）。我们提出以下 DPM,

$$
\begin{aligned}
Y_i|R_i, \boldsymbol{X}_i, Z_i; \boldsymbol{\beta}_i, \sigma_i^2 &\sim p(y|r, \boldsymbol{x}, z; \boldsymbol{\beta}_i, \sigma_i^2) : r = 1, , \ldots, K \\
R_i|\boldsymbol{X}_i; \boldsymbol{\xi}_i &\sim p(r|\boldsymbol{x}; \boldsymbol{\xi}_i) \\
\boldsymbol{X}_{i\ell}; \boldsymbol{\eta}_{ix} &\sim p(\boldsymbol{x}_\ell; \boldsymbol{\eta}_{ix}), \ell = 1, \ldots, p \quad \text{independent} \\
Z_i; \boldsymbol{\eta}_{iz} &\sim p(z; \eta_{iz}) \\
(\boldsymbol{\beta}_i, \sigma_i^2, \boldsymbol{\xi}_i, \boldsymbol{\eta}_i)|F &\sim F, \\
F &\sim DP(\alpha, H_{0\boldsymbol{\beta}} \times H_{0\sigma^2} \times H_{0\boldsymbol{\xi}} \times H_{0\boldsymbol{\eta}}),
\end{aligned}
$$

其中， DP 是狄利克雷过程先验， $\boldsymbol{\eta}_i = (\boldsymbol{\eta}_{ix}, \eta_{iz})$。质量参数 α 服从伽马先验，基测度 $H.$ 则指定为模型参数的共轭先验，参数位于 $p(\cdot)$ 中。第 3 节将结合 QUATRO 数据，详细说明分布和先验分布的选择。

联合分布 $[Y, R, \boldsymbol{X}, Z]$ 可以表示为无限混合分布，

$$f(y,r,\boldsymbol{x},z;\boldsymbol{\beta},\sigma^2,\boldsymbol{\xi},\boldsymbol{\eta})=\sum_{j=1}^{\infty}\pi_j p(y|r,\boldsymbol{x},z;\boldsymbol{\beta}_j,\sigma_j^2)p(r|\boldsymbol{x};\boldsymbol{\xi}_j)p(\boldsymbol{x};\boldsymbol{\eta}_{jx})p(z;\eta_{jz}), \quad (1)$$

其中 $\pi_j=\pi_j'\prod_{s<j}(1-\pi_s')$，且 $\pi_j'\sim Beta(1,\alpha)$。

条件分布，$[Y|R,\boldsymbol{X},Z]$ 可以表示如下:

$$p(y|r,\boldsymbol{x},z)=\sum_{j=1}^{\infty}w_j(r,\boldsymbol{x},z)p(y|\boldsymbol{x},r,z;\boldsymbol{\beta}_j,\sigma_j^2),$$

其中

$$w_j(r,\boldsymbol{x},z)=\frac{\pi_j p(\boldsymbol{x},z;\boldsymbol{\eta}_j)p(r|\boldsymbol{x},z;\boldsymbol{\xi}_j)}{\sum_{h=1}^{\infty}\pi_h p(\boldsymbol{x},z;\boldsymbol{\eta}_h)p(r|\boldsymbol{x},z;\boldsymbol{\xi}_h)}.$$

显然，DPM 默认提供非常灵活的条件分布，也能为观测数据分布提供一致估计量（ Ghosal etal., 1999年）。

如果对 $[Y,R,\boldsymbol{X}]$ 的联合分布进行建模，DPM 也能处理协变量中的可忽略缺失；也就是说，它是一个生成模型。先前的 PMM 和敏感性模型方法不允许协变量缺失，因为它们不对 $\boldsymbol{X}$ 进行建模。

2.2 外推分布

此处的外推分布为 $[Y|R=K+1,\boldsymbol{X},Z]$。由于在 $R=K+1$ 时未观测到 Y，因此无法从观测数据中识别此分布。我们将在第2.4节讨论识别问题。

2.3 感兴趣的参数

此处的目标参数是与 R 和 $\boldsymbol{X}$ 无关的边际治疗效应,

$$\theta=E(Y|Z=1)-E(Y|Z=0).$$

右侧的每一项都可以计算得出。

$$E(Y|Z=z)=\int\sum_{r=1}^{K+1}E(Y|Z=z,R=r,\boldsymbol{X}=\boldsymbol{x})p(r|z,\boldsymbol{x})dF(\boldsymbol{x}). \tag{2}$$

请注意，在上述积分中，我们隐含地假设分布 $F(\boldsymbol{x})$ 不依赖于干预 Z（这是由随机化决定的）。还请注意，鉴于我们关注的是 θ，只需确定外推分布的条件期望 $E(Y|R=K+1,\boldsymbol{X},Z)$。下一节将讨论用于确定这一期望（外推分布的期望）的先验分布。

2.4 外推分布的先验分布

Daniels 等人（2015）假设 R 呈单调（线性）趋势，从而确定了上述期望；该趋势涉及在给定 Z 和 $\boldsymbol{X}$ 的条件下，Y 的期望，即未应答者的结果比提供结果的个体更差。在有许多次重复尝试的情形下，例如 QUATRO 研究（$K=9$），需要采用其他设定，因为在尝试次数很多时，线性设定不太可能合理。此外，随机缺失假设意味着 $E(Y|R=K+1,\boldsymbol{X},Z)=E(Y|R<K+1,\boldsymbol{X},Z)$；这在此处可能并不合理（因为它没有利用重复尝试设计）。

为便于符号表示，我们设定 $\alpha^{(z,K+1)}=E(Y|R=K+1,\boldsymbol{X},Z)$，暂时忽略其对 $\boldsymbol{X}$ 的依赖。参数 $\alpha^{(z,K+1)}$ 无法识别，因此我们在已识别的条件均值基础上，为 $\alpha^{(z,K+1)}$ 指定一个信息性先验。我们假设，缺失模式中的未应答者比提供结果的个体有更差的结果；不过，如果认为这一假设不合理，也可以调整先验分布。与早期研究不同，我们不假设 Y 的期望值与 R 之间具有某种参数形式。

三角先验

我们首先考虑两个三角先验。为方便符号表示，我们再次令 $\alpha^{(zk)}=$

$E(Y|R=k, \boldsymbol{X}, Z)$。先验分布为

$$\alpha^{(z,K+1)}|\alpha^{(z1)},\cdots,\alpha^{(zK)} \sim T(\alpha_{min}^{(z)} - C^{(z)},\ \alpha_{min}^{(z)},\ \alpha_{min}^{(z)} - C^{(z)}),$$
$$\alpha^{(z,K+1)}|\alpha^{(z1)},\cdots,\alpha^{(zK)} \sim T(\alpha_{min}^{(z)} - C^{(z)},\ \alpha_{min}^{(z)},\ \alpha_{min}^{(z)}),$$

其中， $T(a,b,c)$是一个下限为 a、上限为 b、众数为 c的三角分布； $\alpha_{min}^{(z)} = \min(\alpha^{(z1)},\cdots,\alpha^{(zK)})$和 $C^{(z)} = (\alpha_{max}^{(z)} - \alpha_{min}^{(z)}) \times P\%$。 $C^{(z)}$规定了我们希望将“均值”参数 $\alpha^{(z,K+1)}$ 从观测结果的最小均值参数向外移动多远。 P 是敏感性参数，用于将 $C^{(z)}$ 校准为已识别 $\alpha^{(zd)}$范围的百分比。第一种设定将最大权重赋予更极端的均值；第二种设定则将最大权重赋予与观测到 Y 相对应的 R 值中的最小均值。

均匀先验

我们考虑的另一种信息先验是针对 $\alpha^{(z,K+1)}$ 的

$$\alpha^{(z,K+1)}|\alpha^{(z1)},\cdots,\alpha^{(zK)} \sim Unif(\alpha_{min}^{(z)} - C^{(z)}, \alpha_{min}^{(z)}),$$

其中， $C^{(z)}$ 的定义与三角先验中的相同，但在此处表示该均匀先验的区间长度，并说明我们会将 $\alpha^{(z,K+1)}$的下界从 $\alpha_{min}^{(z)}$向外移动多远。 P仍是敏感性参数，用于将 $C^{(z)}$ 校准为已识别 $\alpha^{(zk)}$范围的百分比。

2.5 计算

所有模型中的后验抽样和治疗效应计算 θ 均可在 R 3.3.2 软件中使用 R 程序包 `rjags` 实现，并对式（1）中的混合模型表示采用截断近似（Ishwaran and James，2001）。生成三角分布时使用 R 程序包 `triangle` 。有关如何进行蒙特卡罗积分的详细信息

第 2.3 节见补充材料。

3 QUATRO 分析

3.1 模型

在 QUATRO 试验中，关注的结局 Y 是自我报告的生活质量评分，基线协变量 $\boldsymbol{X}$ 包括基线分数和 4 个中心的指示变量；在可忽略缺失假设下，DPM 可以处理基线分数的缺失。研究人员最多尝试 9 次，以收集参与者的 52 周结局（$K = 9$）。我们将 $Z = 1$ 设为治疗组，将 $Z = 0$ 设为对照组。各治疗组中对应 3 至 9 次尝试的模式都非常稀疏，即每种模式中的受试者都很少（表1）。Daniels 等人（2015年）将这些受试者合并为一种模式。合并后，结局均值总体上随着尝试次数增加而下降，因此非随机缺失假设被认为是合理的，因为不理想（较低）的最终生活质量评分可能与迟回应者有关。

在我们的分析中，尝试次数 R 取值为 $\{1, 2, \cdots, 10\}$，而 $R = 10$对应于所有尝试结束后仍然 Y 缺失的情况。对于 Y 缺失但尝试次数少于最大次数（K）的个体，我们将其归入尝试次数

模式。我们对 $(Y, R, \boldsymbol{X}, Z)$ 的联合分布指定如下 DPM

$$
\begin{aligned}
Y_i|R_i, \boldsymbol{X}_i, Z_i; \boldsymbol{\beta}_i, \alpha_i^{(zr)}, \sigma_i^2 &\sim N(\alpha_i^{(zr)} + \boldsymbol{X}_i\boldsymbol{\beta}_i, \sigma_i^2), \\
R_i|\boldsymbol{X}_i, Z_i; \boldsymbol{\xi}_i &\sim \text{Multinomial}(1, \boldsymbol{\xi}_i), \\
\boldsymbol{X}_{1i}|Z_i; \boldsymbol{\eta}_i &\sim \text{Multinomial}(1, \boldsymbol{\eta}_i), \\
X_{2i}|m_i, \tau &\sim N(m_i, \tau_i^2), \\
Z_i &\sim \text{Ber}(p_i) \\
(\alpha_i, \boldsymbol{\beta}_i, \sigma_i^2, \boldsymbol{\xi}_i, \boldsymbol{\eta}_i, m_i, p_i)|F &\sim F, \\
F &\sim DP(\alpha, H_{0\alpha^{(zr)}} \times H_{0\boldsymbol{\beta}} \times H_{0\sigma^2} \times H_{0\boldsymbol{\xi}} \times H_{0\boldsymbol{\eta}} \times H_{0m} \times H_{0\tau} \times H_{0p}),
\end{aligned}
$$

其中，$\boldsymbol{X}_{1i}$ 是中心指示变量向量，X_{2i} 分别是第 i个观测值的基线分数。参数的基测度为 $H_{0\alpha^{(zr)}} = N(\mu_z, \pi_z^2)$、$H_{0\boldsymbol{\beta}} = N(\boldsymbol{\beta}_0, \sigma_0^2)$、$H_{0\sigma^2}$ =Inv-Gamma(a, b)、$H_{0\theta}$ =Dirichlet(κ_1)、$H_{0\boldsymbol{\eta}}$ =Dirichlet(κ_2)、$H_{0m} = N(m, \tau^2)$、$H_{0\tau}$ =Inv-Gamma(c, d)和 H_{0p} =贝塔分布(e, f)。我们参照 Linero 和 Daniels（2015）以及 Roy 等（2018）来设定分布 $H_{0\cdot}$ 的超参数（详情见补充信息）。

我们考虑了结果的条件分布在完整尝试次数 R $(K = 9)$ 下以及在合并后的尝试次数下依赖于 DPM 的模型，其中按照 Daniels 等人（2015）的做法，将 $K \in \{3, \ldots, 9\}$ 合并为一组。对于条件分布依赖于合并后尝试次数的模型，令 $R^\star$ 表示（合并后的）尝试次数，并将结果缺失的个体归入 $R^\star = 4$。对于合并尝试次数的设定，我们改变了 $[Y_i|R_i, \boldsymbol{X}_i, Z_i$的形式；$\boldsymbol{\beta}_i, \alpha^{(zr^\star)_i}, \sigma_i^2]$ 改为 $N(\alpha^{(zr^\star)_i} + \boldsymbol{X}_i\boldsymbol{\beta}_i, \sigma_i^2)$；因此，现在截距取决于 $R^\star$ 而不是 R。但请注意，我们仍然对 $[R|\boldsymbol{X}, Z]$进行建模。

未作出回应的个体未必会经历最大尝试次数，这种情况并不少见。这里我们将他们归入 $K + 1$ 组。不过，也可以将他们视为删失对象。我们将在第 5 节讨论删失 R 这一问题。

3.2 结果与敏感性分析

我们通过 R 软件包 `loo` 计算了模型的期望逐点预测对数密度；该模型的截距分别取决于 R 和 $R^{\star}$。条件分布取决于全部尝试次数（R）的模型，以及条件分布取决于合并后的尝试次数（$R^{\star}$）的模型，其点估计值（标准误）分别为 -2258.4（65.8）和 -2117.6（33.9）。因此，有一定证据支持在 DPM 中采用“较简单”的条件分布。

第2.3节中定义的不同模型和先验设定下的边际处理效应见表2。RAM-PMM 的敏感性参数记为 C，按照 Daniels 等人（2015）的做法设为 3，这相当于假设从未提供结果的那些受试者与最后一个（即第三个）合并模式中的受试者相同。外推分布的先验设定如第2.4节所述。我们用 unif_P 表示敏感性参数为 P 的均匀先验。众数为 $\alpha_{min}^{(z)} - C^{(z)}$、敏感性参数为 P 的三角先验记为 $\text{tri}1_P$，众数为 $\alpha_{min}^{(z)}$、P 的三角先验记为 $\text{tri}2_P$。这里我们考察了 $P = 10$ 和 20。点质量（pm）表示位于 $\alpha_{min}^{(z)}$ 的点质量先验。

在所有先验分布和模型下，治疗效应都相似，均支持负向治疗效应，但95%可信区间均包含零；先验分布之间缺乏差异与缺失比例较低有关（我们将在模拟中探讨这一点）。非点质量先验的置信区间均略宽于点质量先验。此外，MAR下的治疗效应有所减弱。使用PMM或选择模型时，结论并无不同，不过选择模型的点估计离零最远。

3.3 拟合优度

为了评估结果分布的拟合优度，我们按如下方式计算DPM的观测数据均值

$$\int \sum_{r=1}^{K} E(Y|Z=z,R=r,\boldsymbol{X}=\boldsymbol{x})p(r|z,\boldsymbol{x})dF(\boldsymbol{x}|z). \tag{3}$$

与（2）的区别在于，求和仅至 K ，且现在我们对 $[\boldsymbol{X}|Z]$ 的分布进行积分。针对每种合并后的尝试，给定各合并尝试的响应估计期望值记录于补充信息中的表 ??，分别对应依赖于 R （完整）和依赖于 $R^{\star}$ （合并）的条件分布。观测数据的均值与 DPM 得出的均值非常相似，且在所有情况下，95%置信区间均覆盖了观测数据的均值。因此，DPM 的拟合效果似乎很好。

4 模拟研究

我们开展了一项模拟研究，旨在更好地了解使用所提出的方法与其他参数方法估计治疗效应的情况。

4.1 设计

我们研究了若干数据生成情景。在情景（1）中，我们考虑了真实情况基于 QUATRO 数据分析结果估计值的数据集，但采用了两种不同的样本量：与 QUATRO 相近的样本量（409）和较大的样本量（1000）。在其他情景中，我们使用样本量 $N = 500$ 和 $N = 1000$ ，并针对 $E[Y|R,Z]$ 使用不同的模型和函数形式来生成数据集。对于所有这些情景，我们考虑了几种不同的误差分布。对于尝试次数的分布，我们计算了第 3 次尝试之后的尝试次数概率，并将其平均分配，以计算第 4 至第 9 次尝试的概率； $R =$ 的概率

$K+1=10$ 基于 QUATRO 数据。我们利用这些概率生成了 R 的分布。将经过 $R=K+1$ 次尝试后获得的结局数据设为缺失。在其中两个情景中，我们还考虑了比 QUATRO 中观察到的更多缺失情况。

对于所有情景，我们仅考虑一个协变量 X，并假设该协变量的系数在各组之间保持不变。对于前四种情景，我们为 $Y|X=x$、$R=r$、$Z=z$ 假设了三种不同的分布：正态分布、自由度为 3 的学生 t 分布，以及偏度参数为 α =3 的偏态正态分布（Azzalini，2013年）。对于前三种情景，我们将 $\alpha^{(zr)}+x\beta_1$ 设为位置参数，将 σ^2 设为尺度参数。生成偏态正态分布数据时，使用了 R 软件包 `sn`。

对于所有情景，除情景（4）外，我们均拟合了 DPM（其先验设定与数据示例中所用设定相同）以及第 3 节所述的合并尝试次数的 PMM-RAM 模型（敏感性参数设为 $C=3$）。在情景（4）中，还拟合了未合并尝试次数的 PMM-RAM 模型。对于情景（4）和（5），我们拟合了选择模型（SM）。

关于观测数据模拟情景的更多细节如下。

(1) 在此情景中，为了生成复制数据，我们使用了 QUATRO 数据，并拟合了 PMM-RAM 的结果条件分布；其中，将尝试次数为 3 至 9 次的受试者合并，并将基线评分作为协变量。

(2) 在此情景（以及情景 3）中，我们再次拟合了 PMM-RAM，但假设$\alpha^{(zr)}$是(z、 r)、$z=0$、1、 $r=1$、.....、 R的确定性函数。我们将函数形式 h^* 设为线性，如下所示，

$$h^\star(z,r^\star)=z(27.24-1.91r^\star)+(1-z)(25.58-1.65r^\star),$$

其中 $r^\star=1$，.....，3。我们假设对于 $r=4$，.....，9，有 $\alpha^{(z,r)}=h^\star$(z, 3)。对于该情景（以及情景 3），协变量 X 来自正态分布 N(2, 0.2)。

我们设定系数 $\beta_1 = 0.4$。

(3) 这与情景 (2) 相同，但此处采用非线性函数形式 h，如下所示：

$$h(z, r) = z(30\exp(-0.13r)) + (1 - z)(29\exp(-0.15r)),$$

其中 $r = 1$，.....，9。

(4) 我们使用 Stata 的 alho 模块，基于基线分数对 QUATRO 数据拟合了敏感性模型。由于该模块要求每次尝试至少有一个观测值，因此我们使用了 8 次尝试，并将第 9 次尝试视为缺失。随后，结果和尝试次数按顺序生成，依据如下：

$$Y|X = x, Z = z \sim N(\beta_0^\star + \xi z + x\beta^\star, \sigma^\star),$$

$$P(R = r|R \geq r, Z = z, X = x, Y = y) = \text{expit}(\lambda_{0r} + \gamma_r z + \lambda_r x + \delta_1 y + \delta_2 yz)$$

其中 $\beta^\star$、ξ、$\sigma^\star$、λ_{0r}、γ_r、λ_r、δ_1 和 δ_2 均根据 QUATRO 数据估计得到。

如上所述，要运行 alho 模块，每种尝试次数都必须至少有一个观测值。DPM 不受此限制。因此，我们比较了仅使用符合该限制的数据集的结果，以及使用全部数据集（但不包含敏感性模型）的结果。请注意，在样本量为 500 时，49% 的数据集至少有一次尝试没有结果。

(5) 我们考虑了一个复杂的双成分混合模型，其中一个成分的均值采用线性函数形式（$M = 1$），另一个成分的均值采用非线性函数形式（$M = 2$），如下所示：

$$Y|X = x, R = r, Z = z, M = 1 \sim N(g(z, r) + x\beta_1, \sigma^2),$$

$$Y|X = x, R = r, Z = z, M = 2 \sim N(g^\star(z, r) + x\beta_2, \sigma^2),$$

其中 $g(z,r) = (60.24 - 1.91r)z + (60.58 - 1.65r)(1-z)$, $g^{\star}(z,r) = z(30\exp(-0.13r)) + (1-z)(29\exp(-0.15r))$, $\beta_1 = 0.4$、$\beta_2 = 1$ 和 $\sigma = 2$。参数值与基于 QUATRO 的其他情景相似。

我们考虑了以下权重：

$$\pi_{M=1}(Z,R) = \frac{\exp(2Z - 0.2R - 1)}{1 + \exp(2Z - 0.2R - 1)}, \quad \pi_{M=2} = 1 - \pi_{M=1}.$$

(6) 该情景明确采用了潜在类别模型（LCM）。我们假设结果截距 $\alpha^{(zc)}$ 是治疗（$Z = z$）和潜在类别（$C = c$）的确定性函数，其中 $c = 1, \ldots, 4$。我们将函数形式 h^* 设为线性，如下所示，

$$h^{\star}(z,c) = z(27.24 - 1.91c) + (1-z)(25.58 - 1.65c).$$

类别 C 与尝试次数 R 之间的关系如下：$pr(C = 1|R = 1) = pr(C = 2|R = 2) = pr(C = 3|3 \leq R \leq 9) = 0.8$, $pr(C = 2|R = 1) = pr(C = 3|R = 1) = pr(C = 1|R = 2) = pr(C = 3|R = 2) = pr(C = 1|R = 3) = pr(C = 2|R = 3) = 0.1$ 以及 $pr(C = 4|R = 10) = 1$。与此前的一些情景一样，协变量 X 取自正态分布 $N(2, 0.2)$。我们设定系数$\beta_1 = 0.4$。

外推分布中的未识别位置参数 $\alpha^{(z,K+1)}$ 在每种情景下具体设定如下：(1) $C = 3$（与上一个已识别模式相同）；(2) $h^{\star}(z, 4)$（比上一个已识别的合并模式多一个单位）；(3) $h(z, 10)$（比上一个已识别模式多一个单位）；(4) NA（在参数化敏感性模型中隐式识别）；(5) $E(Y|X, R = 10, Z)$；(6) NA（在设定中定义）。

每种情景均模拟了 1000 个数据集。计算偏差 $\frac{1}{1000}\sum_{k=1}^{1000}(\hat{\theta}_k - \theta)$、均方误差 (MSE) $\frac{1}{1000}\sum_{k=1}^{1000}(\hat{\theta}^{(k)} - \theta)^2$, 以及 95%（等尾）可信区间的经验覆盖概率（和长度），以评估 θ 的估计，其中 θ 是

边际处理效应的真实值，而 $\hat{\theta}^{(k)}$ 是第 k 个模拟数据集中的估计值。

4.2 结果

下面总结各个情景的结果。对于每个情景，DPM 的“none”列（见表格）展示了 DPM 对任意观测数据分布的稳健性（忽略 $K+1$ 模式，即外推分布）；其他行则展示了敏感性方法的影响。

情景 (1)–(3)

对于情景 (1)，不同先验设定下 DPM 的结果显示，覆盖率接近名义覆盖率；对于合并模式数据，偏差和均方误差更小。结果对不同的误差分布设定也具有稳健性（符合预期）。参见补充信息中的表格 ??、??、??、?? 和 ??。

PMM-RAM 只能拟合合并模式数据。正态误差下的结果与 DPM 相似，但在 t 误差和偏态正态误差下，均方误差更大；不过，覆盖率仍接近名义水平。偏态正态误差下，选择模型（SM）的均方误差也更大。

情景（2）和（3）的结果与情景（1）的主要结果一致。参见表格 3 和表格 4，以及补充信息中的表格 ??、??、??、??、??、??、?? 和 ??。情景（4）

在选择模型真实情况下，DPM 的覆盖率在所有先验分布下都接近名义水平，并且就均方误差而言，结果与敏感性模型拟合非常相似。请参阅补充材料中的表格 ??、?? 和 ??。

情景 (5)

正如预期，DPM 的均方误差和偏差远小于 PMM-RAM 和敏感性模型，且覆盖率接近名义覆盖率；PMM-RAM 和敏感性模型的覆盖率随着样本量增大而下降，最低至 .87；见表格 5 和 6。情景 (6)

在此情景下，DPM 表现良好，偏差较小，且两种样本量下的覆盖率都略高于名义水平；见补充信息中的表格 ?? 和 ??。不同缺失程度的探索

我们考察了情景（2）和（6）中不同缺失比例（10%、25%、35%）的影响；参见补充信息中的表格??-??。正如预期，缺失比例越高，可信区间的长度越长。此外，缺失比例增加时，不同先验设定的影响也更大。我们还注意到，先验分布之间关于 θ 的推断差异取决于缺失比例、范围（即 $\alpha_{max} - \alpha_{min}$）以及敏感性参数 P。

5 讨论

我们提出了一种用于重复尝试设计的贝叶斯非参数方法，该方法采用狄利克雷过程混合模型。观察到稀疏模式并不少见。与 Daniels 等人（2015）采用临时方式合并稀疏模式不同，这种方法以依赖于数据的方式隐式处理稀疏模式，允许协变量存在（可忽略的）缺失，并支持敏感性分析。然而，由于缺失比例较低，本文的结论与 Daniels 等人（2015）报告的结论在实质上并无不同。

为了研究我们的 DPM 模型的性能，包括缺失程度高于 QUATRO 的设置，我们在第4节开展了模拟研究。DPM 方法在所有情景下均表现出理想的频率学派运行特征（偏差、均方误差、覆盖率）。使用不同先验分布来设定外推分布均值的敏感性分析方法也表现出良好的频率学派性质。PMM-RAM 和敏感性模型在所有情景下均运行良好，但最后一个情景中的复杂分布除外。然而，我们还指出，PMM-RAM 无法拟合包含大量稀疏模式的设置（我们在模拟中将其合并），而敏感性模型在处理稀疏模式时也存在问题，因为该宏要求每种模式中至少有一个观测值。

我们的方法有许多扩展方向。在分析 QUATRO 数据时，无论为获取结果数据进行了多少次尝试，所有未能提供结果数据的个体都被归入 $R = K + 1$ 尝试组。如果将这些个体视为删失，他们会仅根据 $[R, X]$ 被分配到“模式”中，而在这里，这导致大多数个体不属于 $K + 1$ 组；我们认为，在本应用中这样做并不合理。

我们假设，随着联系尝试次数增加，未作出响应的个体得分会降低。对于其他分析，未观察到响应的个体可能与经过少数几次联系尝试后观察到的个体更为相似，而不是与经过多次联系尝试后观察到的个体相似；可以通过为外推分布期望指定替代先验来处理这一问题。为了进一步改善必要条件期望的估计，$E(Y|R, \boldsymbol{X}, Z)$ 可以用扩展 DPM 替代 DPM（Wade 等人，2011）。另一个值得研究的方向是，将当前方法扩展到具有多种模式的重复尝试情境，如 M¨ussner 等人（2015）所述。

致谢

本研究部分由美国国立卫生研究院 R01 CA183854、HL 158963 和 HL 166324 资助。

补充信息

第 3 节和第 4.2 节中提及的网络附录和表格可在 Wiley 在线图书馆的 Biometrics 网站上查阅。

参考文献

Alho, J. M.（1990）。使用逻辑回归调整无应答偏倚。Biometrika，77(3)：617–624。

Azzalini, A.（2013年）。《偏态正态及相关分布族》。Institute of Mathematical StatisticsMonographs。剑桥大学出版社。

Chen, B.、Li, P. 和 Qin, J.（2018年）。利用回访信息将Heckman选择模型推广至不可忽略的无响应。《中华统计学志》，28(4)：1761–1785。

Daniels, M. J. 和 Hogan, J. W.（2000年）。针对信息性脱落下敏感性分析的模式混合模型重新参数化。Biometrics，56(4)：1241–1248。

Daniels, M. J. 和 Hogan, J. W.（2008年）。纵向研究中的缺失数据：贝叶斯建模与敏感性分析策略，统计学与应用概率专著第109卷。Chapman & Hall/CRC，佛罗里达州博卡拉顿。

Daniels, M. J.、Jackson, D.、Feng, W. 和 White, I. R.（2015年）。重复尝试设计分析的模式混合模型。Biometrics，71(4)：1160–1167。

Ghosal, S.、Ghosh, J. K. 和 Ramamoorthi, R. V.（1999年）。密度估计中狄利克雷混合的后验一致性。《统计年鉴》，27(1):143–158。

Gray, R.、Leese, M.、Bindman, J.、Becker, T.、Burti, L.、David, A.、Gournay, K.、Kikkert, M.、Koeter, M.、Puschner, B.、Schene, A.、Thornicroft, G. 和Tansella, M.（2006年）。针对精神分裂症患者的依从性治疗：欧洲多中心随机对照试验。《英国精神病学杂志》，189(6):508–514。

Ishwaran, H. 和 James, L. F.（2001年）。断棒先验的吉布斯抽样方法。《美国统计学会会刊》，96(453):161–173。

Jackson, D.、Mason, D.、White, I. R. 和 Sutton, S.（2012年）。对一项互联网戒烟试验中缺失数据机制的探讨。《BMC医学研究方法学》，12(1):157。

Jackson, D.、White, I. R. 和 Leese, M.（2010年）。关于缺失，我们能了解多少。

数据？：对精神病学临床试验的一项探索。《英国皇家统计学会会刊：A辑（社会统计学）》，173(3):593–612。

Kenward, M. G.（1998）。非随机脱落的重复测量选择模型：敏感性分析示例。《医学统计》，17(23):2723–2732。

Lin, H., McCulloch, C. E., 和 Rosenheck, R. A.（2004）。纵向研究中信息性间歇缺失数据的潜在模式混合模型。《Biometrics》，60(2):295–305。

Lin, H., McCulloch, C. E., Turnbull, B. W., Slate, E. H., 和 Clark, L. C.（2000年）。用于分析非规则安排观测值的生物标志物轨迹的潜在类别混合模型。《医学统计》，19(10):1303–1318。

Lin, I.-F. 和 Schaeffer, N. C.（1995）。利用调查参与者估计未参与的影响。《公共舆论季刊》，59:236–258。

Linero, A. R. 和 Daniels, M. J.（2015年）。一种灵活的贝叶斯方法，用于处理非可忽略缺失的纵向研究中的单调缺失数据，并应用于一项急性精神分裂症临床试验。《美国统计学会会刊》，110(509)：45–55。

Little, R. J. A.（1993）。多变量不完整数据的模式混合模型。《美国统计学会会刊》，88(421)：125–134。

Little, R. J. A.（1995）。重复测量研究中脱落机制的建模。《美国统计学会会刊》，90(431)：1112–1121。

M¨ussener, U.、Bendtsen, M.、Karlsson, N.、White, I. R.、McCambridge, J. 和 Bendtsen, P.（2015年）。面向大学生的基于短信的戒烟干预：一项随机对照试验（NEXit 试验）的研究方案。《试验》，16(1)：1–9。

Muth´en, B., Jo, B. 和 Brown, C. H.（2003）。针对失效的分层随机试验的主要分层方法

：纽约市择校券案例研究［（含评论）］。《美国统计学会会刊》，98(462)：311–314。

Roy, J. (2003). 使用潜在脱落类别模型对存在不可忽略脱落的纵向数据进行建模。Biometrics, 59(4):829–836。

Roy, J. 和 Daniels, M. J. (2008年). 一类适用于存在不可忽略脱落且可能脱落时间众多情形的一般模式混合模型。Biometrics, 64(2):538–545。

Roy, J., Lum, K. J., Zeldow, B., Dworkin, J. D., Re III, V. L. 和 Daniels, M. J. (2018年). 用于基于随机缺失协变量进行因果推断的贝叶斯非参数生成模型。Biometrics,74(4):1193–1202。

Wade, S., Mongelluzzo, S., Petrone, S., 等（2011年）。一种用于贝叶斯非参数推断的扩展共轭先验。Bayesian Analysis, 6(3):359–385。

Wood, A. M., White, I. R. 和 Hotopf, M.（2006年）。利用联系失败的尝试次数对不可忽略的未响应进行调整。《英国皇家统计学会杂志：A辑（社会统计学）》，169(3)：525–542。

表1：QUATRO 数据：按尝试次数（k）和随机分组（Z）划分的计数（结局均值）。

	Y 在 k 次尝试后观察到									Y 缺失
尝试次数（k）	1	2	3	4	5	6	7	8	9	
对照组（$n = 205$）	77(42.4)	94(41.3)	7(38.7)	7(34.7)	3(34.2)	2(32.9)	1(40.7)	1(62.98)	0(NA)	13
治疗 ($n = 204$)	73(40.7)	90(40.2)	7(38.6)	1(45.7)	3(35.0)	0(NA)	0(NA)	1(30.3)	0(NA)	29

表2：不同先验和模型设定下治疗效应 θ 的后验汇总。Full 对应于对完整尝试次数 ($K = 9$) 进行建模（即，$Y|R$），而 merged 对应于在第3.1节给出的 DPM 中对合并后的尝试次数 $K = 3$ 进行建模（即，将 R 替换为 $R^\star$）；full with merged 对应于条件分布仅取决于合并后的尝试次数的 DPM（即，$Y|R^\star$），但仍对 R 进行建模。前八行对应于对外推分布条件均值采用不同先验的 DPM 模型设定。PMM 和 SM 分别对应于早期工作中的模式混合模型和选择模型。MAR 对应于可忽略的MAR。

	Full		完整合并条件		合并	
模型	mean	95%置信区间 / 置信区间长度	mean	95%置信区间 / 置信区间长度	mean	95%置信区间 / 置信区间长度
完成者	-0.44	(-2.39, 1.61) / 4.00	-0.44	(-2.47, 1.58) / 4.05	-0.48	(-2.59, 1.57) / 4.16
点质量	-0.43	(-2.30, 1.48) / 3.78	-0.39	(-2.37, 1.69) / 4.06	-0.48	(-2.69, 1.75) / 4.44
均匀分布	-0.43	(-2.31, 1.48) / 3.79	-0.39	(-2.38, 1.70) / 4.08	-0.48	(-2.69, 1.75) / 4.44
均匀分布	-0.43	(-2.31, 1.48) / 3.79	-0.39	(-2.38, 1.70) / 4.08	-0.48	(-2.69, 1.76) / 4.44
tri1_{10}	-0.43	(-2.31, 1.48) / 3.79	-0.39	(-2.38, 1.70) / 4.08	-0.48	(-2.69, 1.75) / 4.44
tri1_{20}	-0.43	(-2.31, 1.48) / 3.79	-0.39	(-2.38, 1.70) / 4.08	-0.48	(-2.69, 1.75) / 4.44
tri2_{10}	-0.43	(-2.31, 1.48) / 3.79	-0.39	(-2.38, 1.69) / 4.07	-0.48	(-2.69, 1.75) / 4.44
tri2_{20}	-0.43	(-2.31, 1.47) / 3.79	-0.39	(-2.38, 1.70) / 4.08	-0.48	(-2.69, 1.76) / 4.44
PMM					-0.7	(-3.5, 2.0) / 5.5
SM	-1.6	(-3.9, 0.7) / 4.6			-1.5	(-3.9, 0.8) / 4.7
MAR					-0.50	(-3.19, 2.04) / 5.23

表 3：情景 2，偏态正态分布：估计的治疗效应 θ 的偏差、均方误差、区间覆盖率概率和长度（括号内为蒙特卡洛标准误），基于 1000 次重复模拟。样本量分别为 500 和 1000。下标 4 表示将尝试次数为 3 至 9 次的受试者合并为 3 次（$K = 3$）。先验等于 'none' 对应于不含缺失组的治疗效应估计$K + 1$ （因此仅包括完成者）。不同先验的符号表示见第 3.2 节。

情景	500_4			500			1000_4			1000		
先验	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率
None	-0.010	0.358	0.956	-0.009	0.361	0.959	-0.021	0.193	0.949	-0.019	0.191	0.948
p.m	-0.008	0.359	0.960	-0.011	0.368	0.954	-0.020	0.195	0.949	-0.023	0.191	0.947
unif_{10}	-0.009	0.359	0.959	-0.012	0.369	0.955	-0.022	0.196	0.948	-0.024	0.192	0.948
unif_{20}	-0.010	0.360	0.959	-0.014	0.369	0.954	-0.023	0.196	0.949	-0.026	0.192	0.949
tri1_{10}	-0.009	0.360	0.959	-0.013	0.369	0.955	-0.022	0.196	0.948	-0.025	0.192	0.948
tri1_{20}	-0.011	0.360	0.959	-0.014	0.370	0.954	-0.024	0.196	0.948	-0.027	0.193	0.948
tri2_{10}	-0.008	0.359	0.959	-0.012	0.368	0.954	-0.021	0.195	0.948	-0.024	0.192	0.947
tri2_{20}	-0.009	0.360	0.959	-0.013	0.369	0.955	-0.022	0.196	0.948	-0.025	0.192	0.949

表 4：PMM-RAM 和选择模型（SM）采用偏态正态分布的情景 2：估计的治疗效应的偏差、均方误差、覆盖概率（和区间长度） θ （括号内为蒙特卡洛标准误），基于 1000 次重复模拟。样本量分别为 500 和 1000。RAM-PMM 的敏感性参数设为 $C = 3$ 。

情景	500			1000		
	bias	MSE	覆盖率	bias	MSE	覆盖率
PMM	-0.015	0.409	0.959	-0.025	0.206	0.961
SM	-0.020	0.453	0.948	-0.035	0.235	0.941

表 5：情景 5：基于 1000 次重复模拟的估计治疗效应 θ 的偏差、均方误差、区间覆盖概率和长度（括号内为蒙特卡洛标准误）。样本量分别为 500 和 1000。先验为“无”对应于不考虑缺失组 $K+1$ 时对治疗效应的估计（即仅包括完成者）。不同先验分布的符号表示见第 3.2 节。

情景		500			1000	
先验	bias	MSE	覆盖率	bias	MSE	覆盖率
None	-0.001	0.097	0.954	0.012	0.045	0.969
p.m	-0.043	0.103	0.955	-0.015	0.048	0.972
unif_{10}	-0.032	0.103	0.952	-0.003	0.048	0.973
unif_{20}	-0.020	0.103	0.954	0.009	0.049	0.976
tri1_{10}	-0.028	0.103	0.952	0.001	0.048	0.973
tri1_{20}	-0.013	0.103	0.956	0.016	0.049	0.975
tri2_{10}	-0.035	0.103	0.951	-0.007	0.048	0.972
tri2_{20}	-0.028	0.103	0.953	0.001	0.048	0.972

表6：情景 5 中采用 PMM-RAM 和选择模型（敏感性模型，SM）：估计的治疗效应 θ 的偏差、均方误差和覆盖概率（以及区间长度）（括号内为蒙特卡洛标准误），基于 1000 次重复模拟。样本量分别为 500 和 1000。RAM-PMM 的敏感性参数设为 $C=3$ 。

情景		500			1000	
	bias	MSE	覆盖率	bias	MSE	覆盖率
PMM	-0.155	0.134	0.928	-0.136	0.070	0.915
SM	-0.233	0.182	0.897	-0.211	0.104	0.868

arXiv:2305.05099v1 [stat.ME] 9 May 2023

“稀疏模式重复尝试设计分析的狄利克雷过程混合模型”的补充信息

1 计算

我们使用 R 软件包 `rjags`，采用分块 Gibbs 采样器方法（Ishwaran and James, 2001年）获得了后验样本。我们进行了 50000 次迭代，并设置了 5000 次迭代的预烧入期。为计算治疗效应 θ（见第 2.3 节），我们每隔五个后验样本抽取一个，并根据以下内容计算了 $E(Y|Z=z)$

$$
\begin{aligned}
E(Y|Z=z) &= \int \sum_{r=1}^{K+1} E[Y|Z=z, R=r, \boldsymbol{X}=\boldsymbol{x}] p(r|z,\boldsymbol{x}) dF(\boldsymbol{x}) \\
&= \int \sum_{r=1}^{K+1} \int \sum_{j=1}^{H} w_j(r,\boldsymbol{x},z)\, y\, p(y|\boldsymbol{x},r,z;\boldsymbol{\beta}_j) dy\; p(r|z,\boldsymbol{x})\; dF(\boldsymbol{x}) \\
&= \int \sum_{r=1}^{K} \int \sum_{j=1}^{H} w_j(r,\boldsymbol{x},z)\, y\, p(y|\boldsymbol{x},r,z;\boldsymbol{\beta}_j) dy\; p(r|z,\boldsymbol{x})\; dF(\boldsymbol{x}) \\
&\quad + \int E[Y|Z=z, R=K+1, \boldsymbol{X}=\boldsymbol{x}]\; p(R=K+1|z,\boldsymbol{x})\; dF(\boldsymbol{x}) \\
&= \int \sum_{r=1}^{K} \sum_{j=1}^{H} w_j(r,\boldsymbol{x},z)(\alpha^{(zr,j)} + \boldsymbol{x}\boldsymbol{\beta}^{(j)})\; p(r|z,\boldsymbol{x})\; dF(\boldsymbol{x}) \\
&\quad + \int E[Y|Z=z, R=K+1, \boldsymbol{X}=\boldsymbol{x}] p(R=K+1|z,\boldsymbol{x})\; p(\boldsymbol{x})\; d\boldsymbol{x}
\end{aligned}
$$

$$
\begin{aligned}
&= \int \sum_{r=1}^{K} \sum_{j=1}^{H} w_j(r, \boldsymbol{x}, z)(\alpha^{(zr,j)} + \boldsymbol{x}\boldsymbol{\beta}^{(j)}) \sum_{m=1}^{H} v_m(\boldsymbol{x}, z) p(r; \boldsymbol{\xi}_m)\ p(\boldsymbol{x})\ d\boldsymbol{x} \\
&\quad + \int \alpha^{(z,K+1)}(\boldsymbol{x})\ \sum_{m=1}^{H} v_m(\boldsymbol{x}, z)\ p(R = K+1; \boldsymbol{\xi}_m)\ p(\boldsymbol{x})\ d\boldsymbol{x} \\
&= \int \sum_{r=1}^{K} \sum_{j=1}^{H} w_j(r, \boldsymbol{x}, z)(\alpha^{(zr,j)} + \boldsymbol{x}\boldsymbol{\beta}^{(j)})\ \sum_{m=1}^{H} v_m(\boldsymbol{x}, z) p(r; \boldsymbol{\xi}_m) \sum_{l=1}^{H} \pi_l\ p(\boldsymbol{x}; \boldsymbol{\eta}_l)\ d\boldsymbol{x} \\
&\quad + \int \alpha^{(z,K+1)}(\boldsymbol{x})\ \sum_{m=1}^{H}\ v_m(\boldsymbol{x}, z)\ p(R = K+1; \boldsymbol{\xi}_m) \sum_{l=1}^{H} \pi_l\ p(\boldsymbol{x}; \boldsymbol{\eta}_l) d\boldsymbol{x},
\end{aligned}
$$

其中

$$
\begin{aligned}
w_j(r, \boldsymbol{x}, z) &= \frac{\pi_j p(\boldsymbol{x}, z; \boldsymbol{\eta}_j) p(r|\boldsymbol{x}, z; \boldsymbol{\xi}_j)}{\sum_{h=1}^{H} \pi_h p(\boldsymbol{x}, z; \boldsymbol{\eta}_h) p(r|\boldsymbol{x}, z; \boldsymbol{\xi}_h)}, \\
v_m(\boldsymbol{x}, z) &= \frac{\pi_m p(\boldsymbol{x}, z; \boldsymbol{\eta}_m)}{\sum_{h=1}^{H} \pi_h p(\boldsymbol{x}, z; \boldsymbol{\eta}_h)}.
\end{aligned}
$$

用于计算上述积分的蒙特卡洛步骤如下:

1. 从概率为（π_1，.....，π_H）的多项分布 $\{1, \ldots, H\}$ 中抽取 l。

2. 从 p（$\boldsymbol{x}$；$\boldsymbol{\eta}_l$）中抽取 $\boldsymbol{x}^l$。

3. 抽取 m 从概率为 $\{v_1$（$\boldsymbol{x}^l$，z）.....，v_H（$\boldsymbol{x}^l$，z）$\}$的多项分布 $\{1, \ldots, H\}$ 中抽取 .

4. 从 p（$r|\boldsymbol{x}^l$，z；$\boldsymbol{\xi}_m$）中抽取 r^m。

5. 如果 $r^m < K+1$，计算 $\sum_{h=1}^{H} w_h(r^m、\ \boldsymbol{x}^l、\ z)\{\alpha^{(zm,h)} + \boldsymbol{x}^l\boldsymbol{\beta}^{(h)}\}$

6. 如果 $r^m = K+1$，计算 $E[Y|Z = z, R = k, \boldsymbol{X} = \boldsymbol{x}^l] = \sum_{h=1}^{H} w_h(k, \boldsymbol{x}^l, z)\{\alpha^{(zk,h)} + \boldsymbol{x}^l\boldsymbol{\beta}^{(h)}\}$ 对于 $k = 1$、.....、K，计算 α^z_{min}，然后从第2.4节中指定的先验分布进行抽样。

为评估第 3.3 节中的拟合优度，我们需要按如下方式计算 $E[Y|Z=z,R=r]$：

$$\begin{aligned}
E[Y|Z=z,R=r] &= \int E[Y|Z=z,R=r,\boldsymbol{X}=\boldsymbol{x}]dF(\boldsymbol{x}|z,r) \\
&= \int \sum_{j=1}^{H} w_j(r,\boldsymbol{x},z)(\alpha^{(zr,j)}+\boldsymbol{x}\boldsymbol{\beta}^{(j)})p(\boldsymbol{x}|z,r)d\boldsymbol{x} \\
&= \int \sum_{j=1}^{H} w_j(r,\boldsymbol{x},z)(\alpha^{(zr,j)}+\boldsymbol{x}\boldsymbol{\beta}^{(j)}) \sum_{h=1}^{H} u_h(z,r)p(\boldsymbol{x};\boldsymbol{\eta}_h)d\boldsymbol{x},
\end{aligned}$$

其中

$$\begin{aligned}
u_h(z,r) &= \frac{\pi_h p(z,r;\boldsymbol{\eta}_{zh},\boldsymbol{\xi}_h)}{\sum_{l=1}^{H}\pi_l p(z,r;\boldsymbol{\eta}_{zl},\boldsymbol{\xi}_l)} \\
&= \frac{\pi_h p(z;\boldsymbol{\eta}_{zh},\boldsymbol{\xi}_h)p(r;\boldsymbol{\xi}_{rh})}{\sum_{l=1}^{H}\pi_l p(z;\boldsymbol{\eta}_{zl})p(r;\boldsymbol{\xi}_l)}
\end{aligned}$$

计算上述积分的蒙特卡洛步骤如下:

1. 从多项分布 $\{1,\ldots,H\}$ 中抽取 l，其概率为 $\{u_1(z,r)\ldots,u_H(z,r)\}$。

2. 从 $p(\boldsymbol{x}|z,r;\ \boldsymbol{\eta}_l,\boldsymbol{\xi}_l)$ 中抽取 $\boldsymbol{x}^l$。

3. 计算 $\sum_{h=1}^{H} w_h(r,\boldsymbol{x}^l,z)$({v5)}。

2 默认先验分布

我们采用 Linero 和 Daniels（2015）及 Roy 等（2018）提出的分布超参数设定。数据 Y 和 X_2 经过标准化，均值为 0，方差为 0.5。对于每个混合成分 h，我们有

$$\begin{aligned}
Y_i|R_i,\boldsymbol{X}_i,Z_i;\boldsymbol{\beta}_i^{(h)},\alpha_i^{(zr,h)},\sigma_i^{(2,h)} &\sim N(\alpha_i^{(zr,h)}+\boldsymbol{X}_i\boldsymbol{\beta}_i^{(h)},\sigma_i^{(2,h)}), \\
R_i|\boldsymbol{X}_i,Z_i;\boldsymbol{\xi}_i^{(h)} &\sim \text{Multinomial}(1,\boldsymbol{\xi}_i^{(h)}), \\
X_{2i}|m_i,\tau_i^{(h)} &\sim N(m_i^{(h)},\tau_i^{(2,h)}),
\end{aligned}$$

$$\boldsymbol{X}_{1i}|Z_i;\boldsymbol{\eta}_i^{(h)} \sim \text{Multinomial}(1,\boldsymbol{\eta}_i^{(h)}),$$
$$Z_i \sim \text{Ber}(\eta_{zi}^{(h)}).$$

我们设定 $\alpha_i^{(zr,h)} \sim N(\mu_{\alpha_{z_i r_i}}, \sigma^2_{\alpha_{z_i}})$、 $\boldsymbol{\beta}_i^{(h)} \sim N(\mu\boldsymbol{\beta}_i, c\,\sigma^2_{\boldsymbol{\beta}})$ 和 $m_i \sim N(\mu_m, \sigma_m^2)$。我们假设 $\mu_{\alpha_{z_i r_i}}$和 μ_m 服从均值为 0、方差为 0.5 的正态分布。超参数 $\mu\boldsymbol{\beta}_i$ 和 $\sigma^2_{\boldsymbol{\beta}}$ 由估计得到的线性回归系数及其方差确定。这里 $c = \lceil 344/5 \rceil$，其中 344 为自由度。对于多项分布的超参数，我们设定 $\boldsymbol{\xi}_i^{(h)} \sim \text{DIR}(\mathbb{I}_{K+1}/(K+1))$ 和 $\boldsymbol{\eta}_i^{(h)} \sim \text{DIR}(\mathbb{I}_4/4)$，其中 $\mathbb{I}^n = (1,\ldots,1) \in \mathbb{R}^n$。此外， $\eta_{zi}^{(h)} \sim$服从贝塔分布 Beta(1, 1)。

方差分量的先验设定如下。

$$\sigma_i^{(2,h)} \sim \Gamma^{-1}(S_1, S_1W_1), \qquad S_1 - 2 \sim \Gamma^{-1}(1,1),$$
$$W_1 \sim \Gamma(1, \frac{2}{G_Y}), \qquad \tau_i^{(2,h)} \sim \Gamma^{-1}(S_2, S_2W_2),$$
$$S_2 - 2 \sim \Gamma^{-1}(1,1), \qquad W_2 \sim \Gamma(1, \frac{2}{G_{X_2}}),$$
$$\sigma^2_{\alpha_{z_i}} \sim \Gamma^{-1}(s_z, s_z\lambda_z g_z), \qquad \lambda_z \sim \Gamma(1,1),$$
$$s_z - 2 \sim \Gamma^{-1}(1,1), \qquad \sigma_m^2 \sim \Gamma^{-1}(s_2, s_2\lambda_2 G_{X_2}),$$
$$\lambda_2 \sim \Gamma(1,1), \qquad s_2 - 2 \sim \Gamma^{-1}(1,1),$$

其中， G_Y和 G_{X_2} 分别为 Y和基线结果的方差，而 g_z 为 $Y|Z=z$的条件方差。

3 拟合优度

表 S1：拟合优度：给定每个合并尝试 $R^\star$ 时响应的期望值。'Full' 和 'Merged' 对应 DPM 中 Y 的两种条件分布设定。

			Full		合并	
$R^\star$	Z	$\bar{Y}$	mean	95%置信区间	mean	95%置信区间
1	0	42.4	41.3	(39.0, 43.4)	41.5	(39.3, 43.6)
	1	40.7	41.1	(38.8, 43.3)	41.2	(38.9, 43.5)
2	0	41.3	40.9	(39.0, 42.9)	40.9	(39.0, 42.9)
	1	40.2	40.3	(38.3, 42.3)	40.3	（38.3，42.2）
3	0	38.7	40.4	(37.4, 43.0)	40.1	(37.1, 42.6)
	1	38.6	40.2	(37.3, 43.0)	39.7	(36.7, 42.3)

4 补充模拟

表 S2：情景 1，正态分布：基于 1000 次重复模拟，估计的治疗效应 θ 的偏差、均方误差、区间覆盖概率和长度。样本量分别为 409 和 1000。下标 4 表示将尝试次数为 3 至 9 次的受试者合并为 3 次（$K = 3$）。先验等于 'none' 表示不包含缺失组时对治疗效应的估计 $K + 1$ （因此仅包括完成者）。不同先验分布的符号表示见第 3.2 节。

情景	409_4			409			1000_4			1000		
先验	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率
None	0.034	1.145	0.951	0.033	1.163	0.954	-0.012	0.451	0.952	-0.012	0.452	0.958
p.m	-0.053	1.159	0.953	-0.031		1.167 0.954	-0.091	0.458	0.959	-0.056	0.458	0.957
unif_{10}	-0.052	1.159	0.953	-0.030	1.167	0.955	-0.090	0.458	0.960	-0.055	0.458	0.957
unif_{20}	-0.051	1.159	0.953	-0.029	1.167	0.955	-0.089	0.458	0.960	-0.053	0.458	0.959
tri1_{10}	-0.051	1.159	0.953	-0.029	1.167	0.955	-0.089	0.458	0.960	-0.054	0.458	0.959
tri1_{20}	-0.050	1.159	0.953	-0.028	1.167	0.955	-0.088	0.458	0.960	-0.052	0.458	0.959
tri2_{10}	-0.052	1.159	0.953	-0.030	1.167	0.955	-0.090	0.458	0.960	-0.055	0.458	0.959
tri2_{20}	-0.051	1.159	0.954	-0.029	1.167	0.955	-0.089	0.458	0.960	-0.054	0.458	0.958

表 S3：情景 1，采用自由度为 3 的学生 t 分布：基于 1000 次重复模拟的估计治疗效应 θ 的偏差、均方误差、区间覆盖概率和长度。样本量分别为 409 和 1000。下标 4 表示将尝试次数为 3 至 9 次的受试者合并为尝试 3 次的受试者 ($K = 3$)。先验设为 'none' 表示不包含缺失组时的治疗效应估计 $K + 1$ （因此仅包括完成者）。不同先验分布的符号表示见第 3.2 节。

情景	409_4			409			1000_4			1000		
先验	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率
None	-0.017	0.699	0.944	0.015	0.734	0.951	0.036	0.262	0.960	0.087	0.273	0.967
p.m	-0.100	0.713	0.952	-0.046	0.744	0.953	-0.044	0.264	0.967	0.046	0.271	0.969
unif_{10}	-0.099	0.713	0.951	-0.045	0.745	0.953	-0.043	0.264	0.967	0.048	0.271	0.968
unif_{20}	-0.099	0.713	0.951	-0.043	0.745	0.953	-0.041	0.264	0.968	0.046	0.272	0.968
tri1_{10}	-0.099	0.713	0.951	-0.044	0.745	0.953	-0.042	0.264	0.967	0.048	0.272	0.968
tri1_{20}	-0.098	0.713	0.953	-0.043	0.746	0.954	-0.040	0.264	0.968	0.051	0.272	0.969
tri2_{10}	-0.100	0.713	0.951	-0.045	0.744	0.953	-0.043	0.264	0.967	0.047	0.271	0.968
tri2_{20}	-0.099	0.713	0.951	-0.044	0.745	0.954	-0.042	0.264	0.968	0.048	0.272	0.968

表 S4：情景 1 采用偏态正态分布时，基于 1000 次重复模拟的估计治疗效应 θ 的偏差、均方误差、区间覆盖概率和长度。样本量分别为 409 和 1000。下标 4 表示将具有 3 至 9 次尝试的受试者合并为 3 次尝试 ($K = 3$)。先验等于 'none' 表示在不包含缺失组的情况下估计治疗效应 $K+1$ （因此仅包括完成者）。不同先验分布的符号表示见第 3.2 节。

情景	409_4			409			1000_4			1000		
先验	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率
None	-0.047	0.506	0.947	-0.042	0.509	0.957	-0.030	0.209	0.945	-0.021	0.208	0.966
p.m	-0.128	0.521	0.949	-0.086	0.515	0.958	-0.106	0.218	0.957	-0.048	0.211	0.964
unif_{10}	-0.126	0.521	0.950	-0.085	0.515	0.959	-0.104	0.218	0.957	-0.046	0.211	0.964
unif_{20}	-0.125	0.520	0.949	-0.083	0.515	0.960	-0.102	0.218	0.957	-0.044	0.211	0.965
tri1_{10}	-0.126	0.520	0.950	-0.084	0.515	0.960	-0.103	0.218	0.957	-0.046	0.211	0.965
tri1_{20}	-0.124	0.520	0.948	-0.082	0.515	0.960	-0.101	0.218	0.958	-0.043	0.211	0.965
tri2_{10}	-0.127	0.521	0.950	-0.085	0.515	0.959	-0.104	0.218	0.957	-0.047	0.211	0.964
tri2_{20}	-0.126	0.520	0.950	-0.084	0.515	0.960	-0.103	0.218	0.957	-0.046	0.211	0.965

表 S5：采用 PMM-RAM 的情景 1：基于 1000 次重复模拟的估计治疗效应 θ 的偏差、均方误差和覆盖概率（及区间长度）。样本量分别为 409 和 1000。RAM-PMM 的敏感性参数设为 $C = 3$ 。

情景	409			1000		
	bias	MSE	覆盖率	bias	MSE	覆盖率
正态分布	0.028	1.313	0.954	-0.032	0.507	0.956
学生 t 分布	-0.007	1.270	0.956	0.038	0.558	0.961
偏斜	-0.032	0.571	0.949	-0.036	0.233	0.963

表 S6：PMM-RAM 和敏感性模型（SM）采用偏态正态分布的情景 1；基于 1000 个样本的估计治疗效应 θ 的偏差、均方误差和覆盖概率。样本量分别为 409 和 1000。敏感性参数设为 $C = 3$。

情景	409			1000		
	bias	MSE	覆盖率	bias	MSE	覆盖率
PMM	-0.160	0.547	0.954	-0.153	0.227	0.961
SM	-0.163	0.572	0.927	-0.160	0.244	0.934

表 S7：情景 2：基于 1000 次重复模拟的估计治疗效应 θ 的偏差、均方误差（MSE）、区间覆盖概率及长度。样本量分别为 500 和 1000。下标 4 表示将尝试次数为 3 至 9 次的受试者合并为 3 次尝试 ($K = 3$)。先验等于 'none' 表示不包含缺失组 $K + 1$ 时治疗效应的估计值（即仅包含完成者）。不同先验分布的符号表示见第 3.2 节。

情景	500_4			500			1000_4			1000		
先验	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率
None	0.017	0.832	0.947	0.019	0.839	0.951	-0.015	0.409	0.957	-0.013	0.411	0.956
p.m	0.026	0.828	0.951	0.024	0.841	0.957	-0.014	0.413	0.956	-0.020	0.418	0.960
unif_{10}	0.026	0.828	0.952	0.023	0.841	0.957	-0.015	0.413	0.956	-0.022	0.419	0.961
unif_{20}	0.025	0.829	0.952	0.022	0.842	0.958	-0.017	0.414	0.956	-0.024	0.420	0.961
tri1_{10}	0.025	0.828	0.952	0.023	0.841	0.957	-0.016	0.413	0.956	-0.023	0.419	0.961
tri1_{20}	0.024	0.829	0.952	0.022	0.842	0.958	-0.018	0.414	0.956	-0.025	0.420	0.961
tri2_{10}	0.026	0.828	0.952	0.024	0.841	0.957	-0.015	0.413	0.956	-0.021	0.418	0.960
tri2_{20}	0.025	0.828	0.952	0.023	0.841	0.958	-0.016	0.413	0.956	-0.023	0.419	0.961

表 S8：情景 2 中服从自由度为 3 的学生 t 分布时，基于 1000 次重复模拟得到的估计治疗效应 θ 的偏差、均方误差、区间覆盖概率和长度。样本量分别为 500 和 1000。下标 4 表示将尝试次数为 3 至 9 次的受试者合并为 3 次尝试 ($K = 3$)。先验设为 'none' 对应于不包含缺失组时治疗效应的估计值 $K + 1$ （因此仅包括完成者）。不同先验分布的符号表示可参见第 3.2 节。

情景	500_4			500			1000_4			1000		
先验	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖
None	0.012	0.471	0.956	0.007	0.605	0.957	0.036	0.243	0.956	0.025	0.247	0.9
p.m	0.020	0.473	0.958	0.010	0.499	0.959	0.039	0.246	0.959	0.020	0.252	0.9
unif_{10}	0.019	0.473	0.957	0.009 0.500 0.959	0.038	0.246	0.959	0.019	0.252	0.962		
unif_{20}	0.018	0.474	0.956	0.008 0.501 0.959	0.036	0.246	0.959	0.017	0.253	0.962		
tri1_{10}	0.019	0.474	0.957	0.009	0.500	0.959	0.037	0.246	0.959	0.018	0.253	0.9
tri1_{20}	0.017	0.474	0.956	0.007	0.502	0.959	0.035	0.246	0.960	0.016	0.253	0.9
tri2_{10}	0.019	0.473	0.957	0.010	0.500	0.959	0.038	0.246	0.959	0.019	0.252	0.9
tri2_{20}	0.019	0.474	0.957	0.009	0.500	0.959	0.037	0.246	0.959	0.018	0.253	0.9

表S9：情景 2 下使用 PMM-RAM：基于 1000 次重复模拟的估计治疗效应 θ 的偏差、均方误差和覆盖概率（以及区间长度）。样本量分别为 500 和 1000。对于 RAM-PMM，敏感性参数设为 $C = 3$ 。

情景	500			1000		
	bias	MSE	覆盖率	bias	MSE	覆盖率
正态分布	-0.044	0.282	0.947	-0.060	0.136	0.945
学生 t 分布	-0.054	1.098	0.960	-0.006	0.566	0.952
偏斜	-0.062	0.479	0.953	-0.079	0.247	0.942

表 S10：情景 3：基于 1000 次重复模拟的估计治疗效应 θ 的偏差、均方误差、区间覆盖概率和长度。样本量分别为 500 和 1000。下标 4 表示将具有 3 至 9 次尝试的受试者合并为 3 次尝试（$K = 3$）。等于“none”的先验对应于不包含缺失组 $K + 1$ 的治疗效应估计（因此仅包括完成者）。不同先验的符号表示见第 3.2 节。

情景	500_4			500			1000_4			1000		
先验	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率
None	0.017	0.938	0.962	0.013	0.911	0.963	-0.012	0.475	0.966	-0.012	0.445	0.971
p.m	0.027	0.951	0.962	0.028	0.923	0.971	-0.009	0.487	0.964	-0.003	0.464	0.973
unif_{10}	0.029	0.953	0.964	0.030	0.927	0.971	-0.007	0.489	0.964	-0.002	0.466	0.974
unif_{20}	0.031	0.955	0.964	0.032	0.930	0.971	-0.006	0.490	0.965	0.000	0.468	0.973
tri1_{10}	0.030	0.954	0.964	0.031	0.928	0.971	-0.007	0.489	0.965	-0.001	0.466	0.974
tri1_{20}	0.032	0.956	0.964	0.034	0.933	0.971	-0.005	0.491	0.966	0.001	0.469	0.976
tri2_{10}	0.029	0.952	0.963	0.029	0.925	0.971	-0.008	0.488	0.964	-0.002	0.465	0.974
tri2_{20}	0.030	0.954	0.964	0.031	0.928	0.971	-0.007	0.489	0.965	-0.001	0.466	0.974

表 S11：情景 3 采用自由度为 3 的学生 t 分布：基于 1000 次重复模拟的估计治疗效应 θ 的偏差、均方误差、区间覆盖概率和长度。样本量分别为 500 和 1000。下标 4 表示将尝试次数为 3 至 9 次的受试者合并为 3 次尝试 ($K = 3$)。先验设为 'none' 对应于不包含缺失组时治疗效应的估计 $K + 1$ （即仅包含完成者）。不同先验分布的符号表示见第 3.2 节。

情景	500_4			500			1000_4			1000		
先验	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率
None	0.011	0.628	0.971	0.017	0.610	0.975	0.047	0.352	0.969	0.037	0.335	0.979
p.m	0.027	0.647	0.974	0.029	0.635	0.977	0.049	0.360	0.971	0.038	0.327	0.982
unif_{10}	0.028	0.649	0.975	0.031	0.639	0.978	0.050	0.361	0.972	0.040	0.328	0.982
unif_{20}	0.030	0.651	0.975	0.033	0.645	0.979	0.052	0.361	0.972	0.041	0.330	0.982
tri1_{10}	0.029	0.650	0.975	0.032	0.641	0.978	0.051	0.361	0.972	0.040	0.329	0.982
tri1_{20}	0.031	0.652	0.975	0.034	0.648	0.979	0.053	0.363	0.972	0.042	0.331	0.982
tri2_{10}	0.028	0.648	0.975	0.031	0.638	0.979	0.050	0.361	0.972	0.039	0.328	0.982
tri2_{20}	0.029	0.650	0.975	0.032	0.641	0.978	0.051	0.361	0.972	0.040	0.329	0.982

表 S12：情景 3 采用偏态正态分布：基于 1000 次重复模拟的估计治疗效应 θ 的偏差、均方误差、区间覆盖概率和长度。样本量分别为 500 和 1000。下标 4 表示将尝试次数为 3 至 9 次的受试者合并为 3 次尝试 ($K = 3$)。先验设为 'none' 对应于不包含缺失组时治疗效应的估计 $K + 1$ （即仅包含完成者）。不同先验分布的符号表示见第 3.2 节。

情景		500_4			500			1000_4			1000	
先验	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率
None	-0.015	0.437	0.959	0.020	0.924	0.961	-0.029	0.248	0.974	-0.012	0.444	0.969
p.m	-0.007	0.448	0.962	0.035	0.936	0.970	-0.025	0.247	0.970	-0.003	0.463	0.974
unif_{10}	-0.005	0.449	0.963	0.037	0.939	0.970	-0.023	0.247	0.971	-0.001	0.465	0.975
unif_{20}	-0.004	0.450	0.964	0.039	0.943	0.973	-0.022	0.248	0.972	0.001	0.467	0.975
tri1_{10}	-0.005	0.449	0.963	0.038	0.940	0.970	-0.023	0.247	0.972	-0.001	0.465	0.975
tri1_{20}	-0.003	0.451	0.963	0.041	0.945	0.973	-0.021	0.248	0.972	0.002	0.468	0.975
tri2_{10}	-0.006	0.449	0.962	0.037	0.938	0.970	-0.024	0.247	0.972	-0.002	0.464	0.975
tri2_{20}	-0.005	0.450	0.963	0.038	0.940	0.970	-0.023	0.247	0.971	0.000	0.465	0.975

表S13：情景 3 中采用 PMM-RAM：基于 1000 次重复模拟的估计治疗效应 θ 的偏差、均方误差、覆盖概率（及区间长度）。样本量分别为 500 和 1000。RAM-PMM 的敏感性参数设为 $C = 3$ 。

情景	500			1000		
	bias	MSE	覆盖率	bias	MSE	覆盖率
正态分布	0.062	1.622	0.948	0.046	0.782	0.957
学生 t 分布	0.041	1.659	0.952	0.118	0.812	0.949
偏斜	0.033	1.038	0.949	0.041	0.509	0.942

表 S14：PMM-RAM 和选择模型（敏感性模型）采用偏态正态分布的情景 3：基于 1000 次重复模拟的估计治疗效应 θ 的偏差、均方误差和覆盖概率（以及区间长度）。样本量分别为 500 和 1000。对于 RAM-PMM，敏感性参数设为 $C = 3$ 。

情景	500			1000		
	bias	MSE	覆盖率	bias	MSE	覆盖率
PMM	-0.019	0.613	0.959	-0.014	0.296	0.945
SM	-0.027	0.789	0.953	-0.020	0.385	0.945

表 S15：情景 4：基于 1000 次重复模拟的估计治疗效应 θ 的偏差、均方误差（MSE）、区间覆盖概率及长度。样本量分别为 500 和 1000。下标 4 表示将具有 3 至 9 次尝试的受试者合并为 3 次尝试 ($K = 3$)。先验分布等于 'none' 对应于不包含缺失组 $K + 1$ 的治疗效应估计（即仅包含完成者）。不同先验分布的符号表示见第 3.2 节。

情景		500_4			1000_4			500			1000	
先验	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率
None	-0.167	0.890	0.941	-0.158	0.428	0.962	-0.140	0.885	0.946	-0.145	0.422	0.957
p.m	-0.165	0.885	0.940	-0.152	0.427	0.955	-0.146	0.885	0.949	-0.142	0.424	0.964
unif_{10}	-0.165	0.886	0.940	-0.152	0.427	0.955	-0.146	0.885	0.948	-0.142	0.425	0.965
unif_{20}	-0.165	0.886	0.940	-0.152	0.427	0.955	-0.147	0.885	0.948	-0.143	0.425	0.965
tri1_{10}	-0.165	0.886	0.940	-0.152	0.427	0.955	-0.147	0.885	0.948	-0.143	0.425	0.965
tri1_{20}	-0.165	0.886	0.940	-0.152	0.427	0.954	-0.147	0.885	0.949	-0.144	0.425	0.965
tri2_{10}	-0.165	0.885	0.940	-0.152	0.427	0.955	-0.146	0.885	0.949	-0.142	0.425	0.964
tri2_{20}	-0.165	0.886	0.940	-0.152	0.427	0.955	-0.147	0.885	0.948	-0.143	0.425	0.965

表 S16：PMM-RAM 下的情景 4；基于 1000 个样本的治疗效应估计值 θ 的偏差、均方误差和覆盖概率。样本量分别为 500 和 1000。下标 4 表示将尝试次数为 3 至 8 次的受试者合并为 3 次尝试。($K = 3$)。敏感性参数设为 $C = 3$。

情景		500_4			1000_4			500			1000	
	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率
	-0.027	0.799	0.952	-0.011	0.366	0.970	-0.039	0.801	0.956	-0.017	0.369	0.976

表 S17：情景 4 采用敏感性模型；基于 1000 个样本的估计治疗效应 θ 的偏差、均方误差和覆盖概率。样本量分别为 500 和 1000。下标 4 表示将尝试次数为 3 至 8 次的受试者合并为 3 次尝试。（$K = 3$）。

情景		500_4			1000_4			500			1000	
	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率
	-0.054	0.945	0.944	-0.016	0.435	0.960	-0.041	0.921	0.946	-0.019	0.430	0.961

表 S18：情景 4：基于 1000 次重复模拟的估计治疗效应 θ 的偏差、均方误差、区间覆盖概率和长度。样本量分别为 500、1000 和 10000。先验设为“无”时，对应于不包含缺失组 $K+1$ 的治疗效应估计（因此仅包括完成者）。不同先验分布的符号表示见第 3.2 节。

情景	500			1000			10000		
先验	bias	MSE	覆盖率	bias	MSE	覆盖率	bias	MSE	覆盖率
None	-0.140	0.885	0.946	-0.145	0.422	0.957	-0.124	0.059	0.959
p.m	-0.146	0.885	0.949	-0.142	0.424	0.964	-0.105	0.057	0.972
unif_{10}	-0.146	0.885	0.948	-0.142	0.425	0.965	-0.105	0.057	0.972
unif_{20}	-0.147	0.885	0.948	-0.143	0.425	0.965	-0.105	0.057	0.976
tri1_{10}	-0.147	0.885	0.948	-0.143	0.425	0.965	-0.105	0.057	0.974
tri1_{20}	-0.147	0.885	0.949	-0.144	0.425	0.965	-0.105	0.057	0.978
tri2_{10}	-0.146	0.885	0.949	-0.142	0.425	0.964	-0.105	0.057	0.972
tri2_{20}	-0.147	0.885	0.948	-0.143	0.425	0.965	-0.105	0.057	0.973

表 S19：情景 4（不限制尝试次数）：基于 1000 次重复模拟的估计治疗效应 θ 的偏差、均方误差、区间覆盖概率和长度。样本量分别为 500 和 1000。下标 4 表示将尝试次数为 3 至 9 次的受试者合并为 3 次（$K=3$）。先验为 'none' 对应于不含缺失组的治疗效应估计（$K+1$，因此仅包含完成者）。不同先验的符号表示见第 3.2 节。

情景	500			1000		
先验	bias	MSE	覆盖率	bias	MSE	覆盖率
None	-0.137	0.883	0.950	-0.146	0.420	0.958
p.m	-0.141	0.889	0.954	-0.144	0.420	0.962
unif_{10}	-0.142	0.889	0.954	-0.145	0.420	0.962
unif_{20}	-0.143	0.889	0.954	-0.146	0.420	0.964
tri1_{10}	-0.142	0.889	0.954	-0.145	0.420	0.963
tri1_{20}	-0.143	0.890	0.954	-0.146	0.420	0.964
tri2_{10}	-0.142	0.889	0.954	-0.145	0.420	0.963
tri2_{20}	-0.142	0.889	0.954	-0.145	0.420	0.964

表 S20：情景 4 中 PMM-RAM 不限制尝试次数：基于 1000 次重复模拟，估计的治疗效应 θ 的偏差、均方误差、覆盖概率（及区间长度）。样本量分别为 500 和 1000。RAM-PMM 的敏感性参数设为 $C = 3$ 。

情景	500			1000		
	bias	MSE	覆盖率	bias	MSE	覆盖率
	-0.028	0.816	0.952	-0.012	0.364	0.978

表 S21：情景 6：基于 1000 次重复模拟，估计的治疗效应 θ 的偏差、均方误差、覆盖概率和区间长度。样本量为 500。

缺失率		none	点质量	均匀分布$_{10}$	均匀分布$_{20}$	tri1$_{10}$	tri1$_{20}$	tri2$_{10}$	tri2$_{20}$
10 %	bias	-0.018	-0.017	-0.018	-0.019	-0.018	-0.019	-0.017	-0.018
	MSE	0.896	0.900	0.901	0.902	0.901	0.902	0.901	0.901
	覆盖率	0.960	0.954	0.955	0.954	0.953	0.955	0.953	0.953
	置信区间长度	3.760	3.824	3.833	3.844	3.837	3.851	3.830	3.837
25 %	bias	-0.015	-0.019	-0.018	-0.019	-0.018	-0.019	-0.021	-0.024
	MSE	1.112	1.153	1.160	1.168	1.163	1.173	1.158	1.163
	覆盖率	0.955	0.966	0.968	0.970	0.968	0.970	0.967	0.969
	置信区间长度	4.114	4.515	4.567	4.624	4.586	4.664	4.549	4.586
35 %	bias	-0.028	-0.029	-0.034	-0.039	-0.036	-0.042	-0.032	-0.036
	MSE	1.293	1.355	1.368	1.381	1.372	1.391	1.363	1.372
	覆盖率	0.950	0.978	0.980	0.981	0.980	0.982	0.980	0.980
	置信区间长度	4.425	5.260	5.362	5.472	5.398	5.550	5.327	5.398

表 S22：场景 6：基于 1000 次重复模拟的估计的治疗效应 θ 的偏差、均方误差、覆盖概率、区间长度。样本量为 1000。

缺失率		none	点质量	均匀分布$_{10}$	均匀分布$_{20}$	三角分布$_{10}$	tri1$_{20}$	tri2$_{10}$	tri2$_{20}$
10 %	bias	0.003	-0.004	-0.006	-0.008	-0.007	-0.010	-0.005	-0.006
	MSE	0.461	0.468	0.470	0.471	0.470	0.472	0.469	0.470
	覆盖率	0.956	0.955	0.955	0.956	0.955	0.956	0.955	0.954
	置信区间长度	2.673	2.723	2.732	2.742	2.735	2.749	2.729	2.736
25 %	bias	0.023	0.020	0.015	0.011	0.014	0.008	0.017	0.014
	MSE	0.545	0.569	0.575	0.582	0.578	0.587	0.573	0.578
	覆盖率	0.956	0.976	0.977	0.977	0.977	0.977	0.976	0.977
	置信区间长度	2.933	3.279	3.327	3.379	3.344	3.415	3.310	3.344
35 %	bias	0.019	-0.004	-0.012	-0.019	-0.014	-0.024	-0.009	-0.014
	MSE	0.604	0.702	0.720	0.739	0.726	0.753	0.713	0.726
	覆盖率	0.957	0.983	0.986	0.988	0.988	0.988	0.984	0.987
	置信区间长度	3.154	3.892	3.986	4.087	4.019	4.157	3.955	4.019

表 S23：情景 2：基于 1000 次重复模拟的估计治疗效应 θ 的偏差、均方误差、覆盖概率和区间长度。样本量为 500。

缺失率		none	点质量	均匀分布$_{10}$	均匀分布$_{20}$	tri1$_{10}$	tri1$_{20}$	tri2$_{10}$	tri2$_{20}$
10 %	bias	0.050	0.055	0.053	0.052	0.053	0.052	0.054	0.053
	MSE	0.985	0.990	0.990	0.991	0.991	0.992	0.990	0.991
	覆盖率	0.938	0.935	0.935	0.935	0.935	0.935	0.935	0.935
	置信区间长度	3.720	3.777	3.786	3.795	3.789	3.803	3.783	3.789
25 %	bias	0.022	0.036	0.035	0.033	0.034	0.032	0.035	0.034
	MSE	1.032	1.055	1.056	1.058	1.057	1.059	1.056	1.057
	覆盖率	0.942	0.954	0.955	0.957	0.956	0.958	0.955	0.957
	置信区间长度	3.951	4.181	4.213	4.247	4.224	4.272	4.202	4.224
35 %	bias	0.003	0.035	0.032	0.029	0.031	0.027	0.033	0.031
	MSE	1.183	1.205	1.211	1.218	1.213	1.222	1.209	1.213
	覆盖率	0.944	0.972	0.976	0.980	0.976	0.981	0.975	0.977
	置信区间长度	4.205	4.810	4.882	4.962	4.908	5.020	4.857	4.908

表 S24：情景 2：基于 1000 次重复模拟，估计的治疗效应 θ 的偏差、均方误差、覆盖概率和区间长度。样本量为 1000。

缺失率		none	点质量	均匀分布$_{10}$	均匀分布$_{20}$	三角分布1$_{10}$	三角分布1$_{20}$	三角分布2$_{10}$	三角分布2$_{20}$
10 %	bias	-0.014	-0.012	-0.013	-0.014	-0.013	-0.015	-0.013	-0.013
	MSE	0.491	0.494	0.494	0.495	0.495	0.496	0.494	0.495
	覆盖率	0.945	0.957	0.958	0.959	0.958	0.959	0.958	0.959
	置信区间长度	2.675	2.731	2.739	2.748	2.742	2.754	2.736	2.742
25 %	bias	-0.024	-0.004	-0.005	-0.007	-0.006	-0.008	-0.005	-0.006
	MSE	0.520	0.535	0.537	0.541	0.538	0.543	0.536	0.538
	覆盖率	0.947	0.958	0.961	0.961	0.961	0.962	0.960	0.961
	置信区间长度	2.841	3.052	3.081	3.113	3.091	3.135	3.071	3.091
35 %	bias	-0.011	0.019	0.016	0.013	0.015	0.011	0.017	0.015
	MSE	0.550	0.604	0.611	0.620	0.614	0.626	0.609	0.614
	覆盖率	0.955	0.972	0.976	0.980	0.976	0.981	0.975	0.977
	置信区间长度	3.044	3.552	3.616	3.686	3.639	3.735	3.594	3.639

参考文献

Ishwaran, H. 和 James, L. F.（2001年）。断棒先验的吉布斯抽样方法。《美国统计学会会刊》，96(453)：161–173。

Linero, A. R. 和 Daniels, M. J.（2015年）。一种灵活的贝叶斯方法：用于具有非可忽略缺失的纵向研究中的单调缺失数据，并应用于急性精神分裂症临床试验。《美国统计学会会刊》，110(509)：45–55。

Roy, J.、Lum, K. J.、Zeldow, B.、Dworkin, J. D.、Re III, V. L. 和 Daniels, M. J.（2018年）。用于在随机缺失协变量条件下进行因果推断的贝叶斯非参数生成模型。《Biometrics》，74(4)：1193–1202。