

航线网络中需求无约束估计的非参数选择模型^①

郭 鹏

(贵阳学院经济管理学院, 贵阳 550005)

摘要: 传统的航空客运需求无约束估计方法仅针对平行直达航班中的顾客需求“溢出”和“再现”问题, 未能考虑航空网络中直达和中转联程航班之间的网络替代效应。基于顾客偏好排序列表定义了航空网络顾客类型集合, 建立了考虑顾客策略行为的网络型非参数离散选择模型。考虑到网络环境下历史预售数据的不完备性, 站在线上和线下交易平台的角度, 分别建立了非截尾和截尾需求情况下的完备数据对数似然函数。采用 EM 算法对顾客到达率和概率质量函数进行联合估计, 并提出了网络环境下的顾客“初始需求”、“再现需求”和“溢出需求”无约束估计计算方法。通过数值模拟验证了所提方法的可行性和有效性, 相较于现有方法能准确反映产品间网络替代效应对顾客选择行为影响, 从而更加有效地避免对历史顾客“初始需求”的高估问题。

关键词: 航空客运网络; 需求无约束估计; 网络替代效应; 非参数离散选择模型; EM 算法

中图分类号: F272.1; O211.6 **文献标识码:** A **文章编号:** 1007-9807(2020)12-0030-22

0 引言

近年来, 在由若干出发地和目的地 (origin and destination, O&D) 组成的航空客运网络市场中, 中转联程航线产品的顾客需求量不断增加。随着航空公司成功应用收益管理 (revenue management, RM) 所依赖的计算机性能和产品分销能力的日益提升, 这些因素都为在 O&D 网络层面上成功实施具有竞争力的需求预测提供了条件。在航线网络中, 若一条直达 (nonstop) 航线中的某个票价舱位预售开放, 而相同票价舱位在与之具有相同出发地和目的地的一条中转联程 (connecting) 航线上预售关闭, 则该票价舱位的顾客需求在后一条航线上将受到“随机独立截尾 (random independent censoring)”。基于航段的预测无法对直达

和中转联程航线中大量嵌套票价舱位的需求截尾机理 (demand censoring mechanism)、以及顾客选择行为进行有效区分, 而 O&D 预测可更好地对需求结构变化做出反应, 在出发地—目的地—行程—票价舱位—销售点 (origin-destination-itinerary-fare class-point of sale, ODIF-POS) 水平上, 基于历史受截尾可观察预订数据 (historical observed censored booking data) 和产品可用性信息 (availability information) 对未来顾客的需求无约束需求 (unconstrained demand) 进行预测^[1]。O&D 预测能为收益管理应用企业带来 2%~4% 的收入增长^[2]。

文献 [3-6] 分别对收益管理系统 (revenue management systems, RMSs) 实施过程中的需求截尾及其无约束估计问题进行了研究综述, 通过系统总结和梳理近 40 余年所提收益管理需求无约束估计方法, 均得出如下结论: 在 ODIF-POS 水平

① 收稿日期: 2018-08-20; 修订日期: 2020-03-03。

基金项目: 国家社会科学基金资助项目 (15BGL198); 贵州省教育厅高校人文社会科学研究青年项目 (2019qn009); 贵阳市财政支持贵阳学院学科建设与研究生教育项目 (JG-2020)。

作者简介: 郭鹏 (1984—), 男, 四川宜宾人, 博士, 副教授。Email: pengguo1984@126.com

上精确地捕捉短视型 (myopic) 和策略型 (strategic) 顾客对机票的当期或跨期选择行为, 并对收益管理实施过程中历史顾客的真实“初始需求 (primary demand)”、“再现需求 (recaptured demand)”和“溢出需求 (spilled demand)”进行无约束估计 (unconstraining estimation), 一直以来, 这始终是航空客运 O&D 无约束需求预测 (unconstrained demand forecasting) 中最具挑战的工作之一。

收益管理中存量控制和动态定价策略的实施造成了顾客需求的“溢出 (spill)”和“再现 (recapture)”效应, 为了对历史顾客选择行为进行描述, 需求无约束估计的首要关键是指定选择模型 (choice model), 无论是参数型还是非参数型。虽然文献 [7-10] 分别提出了适用于航空客运网络需求无约束估计的参数模型和方法, 考虑到 O&D 网络中不同行程之间的顾客承运量比例差异巨大^[11] 并且顾客在预售提前期中一般服从非齐次到达过程^[12,13] 因此, 在不同需求密度水平下, 需事先针对影响顾客偏好结构的相关协变量和需求分布形式提出假设, 并要在变量设置的粒度水平与参数估计的误差之间做出权衡; 上述因素影响了参数模型和方法在高密度和高强度网络收益管理实践中的应用效率, 其鲁棒性受到了极大挑战。

van Ryzin 和 Vulcano^[12,13] 基于顾客偏好排序列表 (rank-based preference list) 建立了针对需求无约束估计问题的非参数离散选择模型 (non-parametric discrete choice model), 并提出了相应需求模型估计方法, 但所提模型未将网络环境下的“顾客策略行为 (strategic customer behavior)”^[14-16] 考虑在内。郭鹏等^[17] 基于非参数模型提出了针对平行直达航班 (parallel flights) 的需求无约束估计方法, 并通过数值算例说明: 忽略策略型顾客选择行为会造成公司对顾客“初始需求”的高估。若在 O&D 网络应用环境下, 该高估问题会持续影响后续 O&D 预测和 O&D 控制 (O&D control) 等的准确度, 使整套收益管理优化策略失效, 导致公司收入损失^[18]。因此, 非参数模型在航空网络需求无约束估计中的适用性和准确性还有

待进一步挖掘。基于此, 本文建立了考虑顾客策略行为的网络型非参数离散选择模型。

另外, 由于航空客运中央预订系统 (central reservation systems, CRSs) 既与互联网环境下的在线旅行社 (online travel agencies, OTAs) 和公司官网保持对接, 又同线下传统销售渠道共享信息, 考虑到上述两方面来源的需求数据均具有不同程度的不完备性^[12,13,17] 这增加了基于现有历史销售数据对模型参数进行估计的难度。针对该问题, 本文分别站在“非截尾需求 (uncensored demand)”和“截尾需求 (censored demand)”的角度, 提出了模型参数联合估计的 EM 算法、以及网络需求的无约束估计计算方法。最后, 通过数字模拟说明了所提模型和方法在航空网络需求无约束估计应用中的适用性和有效性。

表 1 本文的研究对象

Table 1 Research objects of this paper

编号	要素	类别
1	资源特征	离散
2	容量特征	固定
3	舱位需求特征	随机 - 相关
4	价格设置特征	动态
5	价格等级数量	$M^1 + M^2 + \dots + M^k$
6	价格支付意愿	由低到高
7	再现需求	考虑
8	跨期替代行为	考虑
9	航空网络产品替代效应	考虑
10	取消预订和 no-shows	不考虑
11	批量需求	不考虑
12	超售	不考虑

1 研究对象

本文研究主要针对航空客运网络需求无约束估计问题, 所提方法还适用于其他网络收益管理应用行业, 比如铁路客运和汽车租赁。根据“易逝资产收益管理 (perishable asset revenue management, PARM)”概念^[19] 其研究对象如表 1 所示。

2 考虑顾客策略行为的网络型非参数离散选择模型

基于顾客偏好排序列表的非参数离散选择模型最初被广泛应用于经济学和心理学等研究领域, Mahajan 和 van Ryzin^[20] 首次将该类模型应用到了运营管理中的零售产品组合规划问题, 并指出: 文献研究中用于描述顾客选择行为的几种常用选择模型均可被表示为它的特殊情况, 其中包括: 多项式 Logit 模型 (multinomial logit model, MNL)、马尔科夫第二选择模型 (Markovian second choice model)、通用备份模型 (universal backup model)、兰卡斯特需求模型 (Lancaster demand model)、以及独立需求模型 (independent demand model). 虽然其后的文献^[12, 13, 21-25] 分别使用非参数离散选择模型对收益管理领域相关问题进行研究, 但所考虑的顾客偏好排序列表也仅涉及短视型顾客类型. 本节将构建考虑了顾客策略行为的网络型非参数离散选择模型, 为提出 O&D 预测中的网络需求无约束估计方法提供理论依据.

2.1 符号说明

结合航空客运网络收益管理实践, 令机票在时间点 $t = 0$ 开始预售, 停止预售时刻为 $t = T$, $t = (1, 2, \dots, T)$. 航空公司根据 O&D 控制的需要, 可将整个机票预售期拆分为若干个连续相邻时刻的“预售提前期间隔 (lead-time interval)” Δ_t^k .

1) O&D 行程和票价舱位集合

D : 航线网络出发地集合, $d = |D|$, $k \in D$;

H^k : 出发地 k 中的航线网络行程集合, $h^k = |H^k|$, $l \in H^k$;

$M^k = \{M^{1,k}, M^{2,k}, \dots, M^{h^k}\}$: H^k 中的舱位集合, $m^k = |M^k \cup 0|$, $i \in M^k \cup 0$, $i = 0$ 表示顾客选择“不购买”、或转移购买竞争者机票;

S_t^k : M^k 中预售开放的舱位集合, 即 Δ_t^k 中的报价集合.

2) 已知参数

ρ^k : 以出发地 k 为枢纽的客运网络市场份额, $0 < \rho^k < 1$;

I_{ii}^k : M^k 所列舱位 i 在 Δ_t^k 中的历史预售开放状态, 取值 0 或 1, 分别表示舱位预售关闭或开放;

B_{ii}^k : M^k 中的舱位 i 在 Δ_t^k 中的历史可观察订购量;

R_t^k : 各 Δ_t^k 中的机票销售子时间点, $r = (1, 2, \dots, R_t^k)$;

$\Delta_t^k(r)$: 时间点 r 与 $r-1$ 之间“销售子时段”;

$I_{ii}^k(r)$: 各 $\Delta_t^k(r)$ 中的舱位预售开放状态, 取值 0 或 1, 表示舱位的预售关闭或预售开放状态;

$B_{ii}^k(r)$: M^k 中的舱位 i 在 $\Delta_t^k(r)$ 中的历史可观察订购量.

3) O&D 顾客类型集合

σ_t^{jk} : Δ_t^k 中到来的顾客类型 j 的偏好排序列表;

α_t^k : Δ_t^k 中到来的短视型顾客类型集合;

β_t^k : Δ_t^k 中到来的策略型顾客类型集合;

N_t^k : Δ_t^k 中到来的顾客类型全集, $j \in N_t^k$;

AN_t^k : 能够在任意 Δ_t^k 中初次到来的顾客类型集合 $AN_t^k \in (\alpha_t^k \cup \beta_t^k)$;

BN_t^k : 由 $(\Delta_1^k, \dots, \Delta_{t-1}^k)$ “跨期再现”到 Δ_t^k 中进行延期购买的策略型顾客类型集合, $BN_t^k \in (\beta_1^k \cup \beta_2^k \cup \dots \cup \beta_{t-1}^k)$, $N_t^k = AN_t^k \cup BN_t^k$;

CN_t^k : 在 Δ_t^k 中到来后不会转移到 $(\Delta_{t+1}^k, \dots, \Delta_T^k)$ 中进行延期购买的顾客类型集合, 由于不会产生“跨期替代行为 (inter-temporal substitution behavior)”, 它包括由 $(\Delta_1^k, \dots, \Delta_{t-1}^k)$ 转移到 Δ_t^k 中的策略型顾客、以及在 Δ_t^k 中初次到来的短视型顾客, $CN_t^k \in (\beta_1^k \cup \beta_2^k \cup \dots \cup \beta_{t-1}^k \cup \alpha_t^k)$;

DN_t^k : 由于无法成功预订到所偏好舱位, 在 Δ_t^k 中到来后会选择“跨期再现”到 $(\Delta_{t+1}^k, \dots, \Delta_T^k)$ 中进行购买的顾客类型集合, 包括由 $(\Delta_1^k, \dots, \Delta_{t-1}^k)$ 转移到 Δ_t^k 中、以及 Δ_t^k 中初次到来的策略型顾客, $DN_t^k \in (\beta_1^k \cup \beta_2^k \cup \dots \cup \beta_t^k)$, $N_t^k = CN_t^k \cup DN_t^k$;

$C_{i(r)}(i, S_t^k, N_t^k)$: N_t^k 中以 S_t^k 所列舱位 i 为偏好列表首选、并在 $\Delta_t^k(r)$ 中到来的顾客类型集合

$$C_{i(r)}(i, S_t^k, N_t^k) = \left\{ j: \sigma_t^{jk}(i) < \sigma_t^{jk}(w), \forall \sigma_t^{jk} \in N_t^k, \forall w \in S_t^k, w \neq i \right\};$$

$C_{i(r)}(0, S_t^k, N_t^k)$: 在提供报价集合 S_t^k 的情况下, N_t^k 中选择“不购买” ($i=0$) 的顾客类型集合,

$$C_{i(r)}(0, S_i^k, N_i^k) = \left\{ \begin{array}{l} j: I_{\sigma_i^{jk}(v)^{-1}}(r) = 0, \forall \sigma_i^{jk} \in N_i^k, \\ \forall \sigma_i^{jk}(v)^{-1} \in \bar{S}_i^k, p=1, 2, \dots, |\sigma_i^{jk}| - 1 \end{array} \right\};$$

其中 $\sigma_i^{jk}(v)^{-1}$ 是顾客偏好排序列表 σ_i^{jk} 的反函数, $v = 1, 2, \dots, |\sigma_i^{jk}|$, 用于表示偏好列表中相应排序所对应的舱位.

4) 未知参数

λ_i^k : 客运网络各 Δ_i^k 中的顾客到达率;

$\lambda_{i(r)}^k$: 客运网络各 $\Delta_i^k(r)$ 中的顾客到达率;

$x_i^k(j)$: 顾客类型 σ_i^{jk} 在 Δ_i^k 中到达的概率 $P(\sigma_i^{jk})$;

$P_i(i | S_i^k, N_i^k)$: N_i^k 中的顾客类型在 $\Delta_i^k(r)$ 中购买 S_i^k 所列舱位 i 的概率, 可将其简化表示为 $y_{ii}^k(r)$;

$P_i(0 | S_i^k, N_i^k)$: 在提供报价集合 S_i^k 的情况下, N_i^k 中的顾客类型选择“不购买”($i = 0$) 的概率, 可简化表示为 $y_{0i}^k(r)$.

2.2 顾客非齐次到达过程的定义

基于一般情况的考虑, 同一顾客在 24 h 内由同一出发地不会乘坐多于一架次的航班前往同一或不同目的地, 即同一顾客在 24 h 内仅会从同一出发地搭乘航班出发一次, 因此, 令集合 H^k 仅由同一出发地 24 h 内起飞的不同行程航班组成; 由于可获得的历史预售数据 B_{ii}^k 均来自航空公司自身经营的航线, 可令集合 M^k 中所考虑舱位具有相差不超过 24 h 的预售起始时间. 上述假设保证了顾客需求可在各 Δ_i^k 的舱位集合 M^k 中发生“溢出”、“当期再现”和“跨期再现”效应.

对于各 Δ_i^k 中所包含时间长度的问题, 可根据 O&D 存量控制和动态定价策略实施结果、以及各舱位的历史可观察订购量 B_{ii}^k 来确定. 为保证同一 Δ_i^k 中各票价舱位的历史预售开放状态 I_{ii}^k 和价格保持不变, 可将 Δ_i^k 设置为部分舱位预售开放状态或价格发生变化前后的时间间隔; 同时, 需保证同一 Δ_i^k 中所有舱位历史可观察订购量之和 $\sum_{i=1}^{m^k} B_{ii}^k > 0$ 即同一 Δ_i^k 中至少有一个舱位开放预售并存在订票, 否则, 所确定的 Δ_i^k 无意义.

对于任意的报价集合 $S_i^k = \{i: I_{ii}^k = 1, i \in M^k \cup \{0\}\}$, 一方面, 需保证各 S_i^k 中所列同一舱位 i

至少在一个 Δ_i^k 中的历史可观察订购量 $B_{ii}^k > 0$, 即 $\sum_{i=1}^T B_{ii}^k > 0$, 否则, 在 M^k 中对该舱位不做考虑; 另一方面, 令 $0 \in S_i^k$, 即 $|S_i^k| \geq 2$, 但 $S_i^k \neq \{0\}$.

如前所述, 本文所提非参数模型使用伯努利过程来描述顾客在各 Δ_i^k 中的非齐次到达过程, 其主要思想是: 将每个 Δ_i^k 等分成 R_i^k 个 $\Delta_i^k(r)$, 顾客在各 $\Delta_i^k(r)$ 中的到达过程被定义为满足到达率为 $\lambda_{i(r)}^k$ 的齐次伯努利过程, $0 < \lambda_{i(r)}^k < 1$, 则 $\lambda_i^k = \lambda_{i(r)}^k \times R_i^k$; 令 $R_i^k = \sum_{i=1}^{m^k} B_{ii}^k / \rho^k$, 由于各 R_i^k 在数值上存在差异, 这为使用 λ_i^k 来描述顾客在各 Δ_i^k 中的非齐次到达过程提供了条件. 针对上述处理, 需做如下假设:

1) 令 $I_{ii}^k(r) = I_{ii}^k$;

2) 每个 $\Delta_i^k(r)$ 中至多允许有一位顾客到来,

即 $\sum_{i=1}^{m^k} B_{ii}^k(r) \leq 1$. 由于 $0 < \rho^k < 1$, 则 $R_i^k > \sum_{i=1}^{m^k} B_{ii}^k$, 故被等分后的 Δ_i^k 中必然存在有顾客未成功预订到机票的“销售子时段” $\Delta_i^k(r)$, 将这些 $\Delta_i^k(r)$ 放入集合 \bar{O}_i 中, 而令有机票售出的 $\Delta_i^k(r) \in O_i$, $|O_i| = \sum_{i=1}^{m^k} B_{ii}^k$, 则 $|\bar{O}_i| = R_i^k - |O_i|$. 同时, 令每个舱位 i 所对应的有机票售出的 $\Delta_i^k(r)$ 总数 $|O_i| = B_{ii}^k$, 可按一定的舱位顺序, 将 O_i 分为 m^k 个子集合: $O_i = O_{i1} \cup O_{i2} \cup \dots \cup O_{im^k}$. 在具体构造各舱位在 O_i 中所对应的 $B_{ii}^k(r)$ 时, 依次在各 $\Delta_i^k(r) \in O_{i\hat{i}}$ 中对舱位 i 的 $B_{ii}^k(r)$ 进行赋值

$$B_{ii}^k(r) = \begin{cases} 1, & i = \hat{i} \\ 0, & i \neq \hat{i} \end{cases}$$

换句话说, 当所考虑的 $\Delta_i^k(r)$ 属于舱位 i 所对应的有机票售出的“销售子时段”子集合 $O_{i\hat{i}}$ 时, 则令舱位 i 的 $B_{ii}^k(r) = 1$, 而其余舱位 i' 的 $B_{i'i}^k(r) = 0, i' \neq i, i' = (1, 2, \dots, m^k)$, 以此类推, 直到将 O_i 遍历完毕. 当考虑的 $\Delta_i^k(r) \in \bar{O}_i$ 时, 令所有舱位的 $B_{ii}^k(r) = 0$.

因此, 上述假设保证了每个舱位所对应的 $B_{ii}^k(r)$ 和 B_{ii}^k 满足关系式 $\sum_{\Delta_i^k(r) \in O_i} B_{ii}^k(r) = B_{ii}^k$ 和

$\sum_{\Delta_i^k(r) \in O_i} B_{ii}^k(r) = 0$, 由此可得 $\sum_{i=1}^{m^k} \sum_{\Delta_i^k(r) \in (O_i \cup \bar{O}_i)} B_{ii}^k(r) = \sum_{i=1}^{m^k} B_{ii}^k$. 从本质上讲, 确定 $B_{ii}^k(r)$ 取值的依据是各舱位的 B_{ii}^k , 而非 I_{ii}^k . 也就是说, 如果某舱位的 $B_{ii}^k > 0$ (其 I_{ii}^k 也必然等于 1), 上述假设会在 O_i 中专门为该舱位分配长度为 $|O_i|$ 的“销售子时段”子集合. 在该舱位专属的所有 $\Delta_i^k(r) \in O_i$ 中, 令该舱位的 $B_{ii}^k(r) = 1$, 即便存在其余舱位 i' 的 $I_{i'i}^k = 1$, 也还是令 $B_{i'i}^k(r) = 0, i' \neq i$; 以此类推. 因此, 上述假设保证了在有机票售出的同一个 $\Delta_i^k(r)$ 中不会出现多于一个舱位的 $B_{ii}^k(r) = 1$ 的情况, 并且每个 $\Delta_i^k(r) \in O_i \cup \bar{O}_i$ 中所有舱位到来的顾客数量 $\sum_{i=1}^{m^k} B_{ii}^k(r) \leq 1$.

考虑到在没有机票售出的 $\Delta_i^k(r) \in \bar{O}_i$ 中存在如下两种情况:

- 1) 有顾客到来但选择不购买;
- 2) 没有顾客到来.

因此, 将属于前者的 $\Delta_i^k(r) \in \bar{O}_i$ 归入集合 \bar{O}_i^λ 中, 而用集合 \bar{O}_i^λ 来概括后者, 并有如下关系 $|\bar{O}_i| = |\bar{O}_i^\lambda| + |\bar{O}_i^\lambda|, R_i^k = |O_i| + |\bar{O}_i^\lambda| + |\bar{O}_i^\lambda|$. 由于当 $\Delta_i^k(r) \in O_i$ 时, 必然会有顾客到来, 故令 \hat{R}_i^k 表示各 Δ_i^k 中有顾客到来的“销售子时段”数量, 则有 $\hat{R}_i^k = |O_i| + |\bar{O}_i^\lambda|$ 和 $R_i^k = \hat{R}_i^k + |\bar{O}_i^\lambda|$.

2.3 O&D 顾客类型的定义

本文所提非参数离散选择模型中所考虑的 O&D 顾客类型可由 σ_i^{jk} 来定义:

1) 短视型顾客: 购买偏好仅限于在同一 Δ_i^k 中不同票价舱位间进行“当期替代”的顾客.

例如, 某位偏好排序列表为 $\sigma_i^{jk} = (a, b, c, 0)$ 的顾客, 其首要偏好为 Δ_i^k 中的舱位 a . 若 a 预售关闭 ($I_{aa}^k = 0$), 顾客会降舱 (buy-down) 或升舱 (buy-up) 购买同一航班相同 Δ_i^k 中预售开放且价格较低的舱位 b 或价格较高的舱位 c , 该情况属于短视型顾客需求当期“垂直再现 (vertical recapture)”; 另外, 如果舱位 b 与舱位 a 属于同一直达航班, 而舱位 c 是 O&D 网络中相同目的地和出发地的另一中转联程航班中的舱位, 则上述偏好排序列表

描述了顾客需求在 O&D 网络不同行程航班舱位间的当期“水平再现 (horizontal recapture)”.

2) 策略型顾客: 各 Δ_i^k 中的购买偏好除具有上述“当期替代”行为特征外, 还具备“跨期替代行为”, 即能够在不同 Δ_i^k 中的不同票价舱位间进行延期购买的顾客^[14-17].

例如, 本文基于偏好排序列表将某策略型顾客定义为 $\sigma_i^{jk} = (a, a_{t+1}, b_{t+1}, a_{t+2}, b_{t+2}, c_{t+2}, 0)$; 若 $I_{aa}^k = 0$, 其会选择延期到 Δ_{t+1}^k 中再对舱位 a 和舱位 b 进行购买; 同样地, 如果 $I_{a_{t+1}a_{t+1}}^k = 0$ 和 $I_{b_{t+1}b_{t+1}}^k = 0$, 则该顾客还会继续跨期到 Δ_{t+2}^k 中选择购买 O&D 网络不同行程航班舱位.

有关短视型和策略型顾客选择行为、及其需求“溢出”和“再现”效应的详细描述, 可参见文献 [17] 中的 1.2 节. 为更好地说明偏好排序列表的结构, 还需做如下的限定:

1) 本文所考虑的顾客类型仅限于理性消费者, 顾客被假定会首选偏好排序列表中排名最靠前的票价舱位 $i_t = \arg \min_{i \in S_t^k \cup 0} \sigma_i^{jk}(i)$;

2) “不购买” ($i = 0$) 选项不能成为首要偏好, 需保证顾客类型的偏好列表长度 $|\sigma_i^{jk}| \geq 2$, 换句话说, 除去排序最靠后的“不购买” ($i = 0$) 选项外, 有效的顾客类型需表现出对至少 1 个航班票价舱位有购买意愿的特征;

3) 仅考虑策略型顾客需求在不同 Δ_i^k 之间的“跨期再现”, 由同一 Δ_i^k 等分成的不同 $\Delta_i^k(r)$ 间的顾客到达过程相互独立, 不考虑不同 $\Delta_i^k(r)$ 间的“跨期替代行为”.

由上述限定可知, 由于各 Δ_i^k 中的票价舱位数 $|M^k \cup 0| = m^k$, “预售提前期间隔”时期数为 T , 它们都是有限的, 若考虑了策略型顾客的“跨期替代行为”, 则以 k 为出发地的 O&D 网络市场在 Δ_i^k 中所能够到来的顾客类型数最多为

$$n = \sum_{j=1}^{T \times m^k} \binom{T \times m^k}{j} j!$$

因此, 计算顾客偏好排序列表数量的时间复杂度可用“组合”阶 $O((T \times m^k)!)$ 来表示.

2.4 网络型非参数离散选择模型的建立

1) 当 $i \neq 0$ 时:

$$\begin{aligned}
 P_r(i | S_i^k, N_i^k) &= P_r(i | S_i^k, AN_i^k) + P_r(i | S_i^k, BN_i^k) \\
 &= \sum_{j \in C_{i(r)}(i, S_i^k, AN_i^k)} x_i^k(j) + \sum_{j \in C_{i(r)}(i, S_i^k, BN_i^k)} x_i^k(j) \\
 &= \sum_{j \in C_{i(r)}(i, S_i^k, N_i^k)} x_i^k(j) \quad (1)
 \end{aligned}$$

2) 当 $i = 0$ 时:

$$\begin{aligned}
 P_r(0 | S_i^k, N_i^k) &= P_r(0 | S_i^k, CN_i^k) + P_r(0 | S_i^k, DN_i^k) \\
 &= \sum_{j \in C_{i(r)}(0, S_i^k, CN_i^k)} x_i^k(j) + \sum_{j \in C_{i(r)}(0, S_i^k, DN_i^k)} x_i^k(j) \\
 &= \sum_{j \in C_{i(r)}(0, S_i^k, N_i^k)} x_i^k(j) \quad (2)
 \end{aligned}$$

其中 $\sum_{i \in S_i^k} P_r(i | S_i^k, N_i^k) + P_r(0 | S_i^k, N_i^k) = 1$.

虽然本文和郭鹏等^[17]均在 van Ryzin 和 Vulcano^[13]所提同类模型基础上考虑了顾客策略行为,但本文的主要贡献是将该类模型扩展到了网络环境,其网络型特征主要体现在:

1) 由于航空客运网络中不同城市的机场相距较远,顾客进行航空出行决策时一般不考虑出发城市之间的替代效应,因此,在定义顾客偏好排序列表 σ_i^k 和舱位报价集合 S_i^k 时,所考虑的舱位类型涵盖了航线网络中基于同一出发地前往相同或不同目的地的直达和中转联程航线航班舱位, $i \in M^k \cup 0$,这使得集合 $C_{i(r)}(i, S_i^k, N_i^k)$ 和 $C_{i(r)}(0, S_i^k, N_i^k)$ 能够充分体现顾客需求在 O&D 票价舱位间的网络替代效应。

2) 郭鹏等^[17]和 van Ryzin 和 Vulcano^[13]所提模型仅将同一目的地前往相同目的地的平行直达航班舱位作为研究对象,未考虑相同目的地情况下的中转联程、以及不同目的地情况下直达和中转联程航班舱位间的替代效应,因此,其参数设置无法对顾客需求在网络环境下的选择行为和到达过程进行描述。而本文在考虑舱位间网络替代效应的情况下,使用 $x_i^k = P(\sigma_i^k)$ 对顾客到达的离散概率质量函数(probability mass function, PMF) 进行描述,其到达率 λ_i^k 也能够反映出 O&D 环境下顾客的非齐次到达过程。

3 联合估计问题的完备数据对数似然函数

在建立网络型非参数离散选择模型后,接下来所关注的问题是怎样对 $(x_i^k, \lambda_{i(r)}^k)$ 进行联合估计,本文采用极大似然估计方法,以最好地实现估计结果对可观察预订量 B_{it}^k 的解释。除 B_{it}^k 之外,可从 OTAs、公司官网和传统线下旅行社等渠道获得的已知数据还包括 ρ^k 、 I_{it}^k 和潜在的 N_{it}^k 。但是,若要对参数进行准确估计,还需其他两方面的重要信息:一方面,在特定时期所到来顾客的具体类型;另一方面,当没有舱位机票成功售出时,是否有顾客到来并开展购买问询。由于航空公司无法完全获取上述两方面的信息,这便构成了历史预售数据的不完备性问题^[12, 13, 17]。其中,所到达顾客的类型数据在现有渠道下均无法直接获取,这是因为顾客偏好属于主观意愿,顾客类型信息无法在海量的预订和问询过程中被实时地精确判断和区分;而第二方面的不完备性问题却可通过计算机技术手段克服^[6]。

与 van Ryzin 和 Vulcano^[13]的研究思路类似,本文拟提出解决极大似然估计问题的简化方法,其主要构想是利用完备数据对数似然函数(complete data log-likelihood function),以克服上述两方面数据不完备性对参数估计造成的影响,它既能反映顾客需求“溢出”效应的可能性,又可将每位到来顾客所属精确类型的概率考虑在内。按所获得上述第二方面数据的完备程度不同,可将本文的参数估计过程分为“非截尾需求”情况和“截尾需求”情况,它们的区别在于:

1) 在“非截尾需求”情况下,由于在线交易平台能够对 O&D 市场中的顾客访问记录、购买行为等信息进行跟踪和记录,可对 \bar{O}_i^k 和 \bar{O}_i^k 进行准确区分,能够对顾客需求是否受到截尾进行精确地记录,用于参数估计的预订数据属于非截尾数据。

2) 相反地,“截尾需求”情况则是指预订数据中无法反映顾客到达过程中的需求“溢出”效应,当没有机票售出时,无法通过对是否有顾客到来的跟踪和记录来判断当期顾客需求是否受到截尾,该情况常见于传统线下旅行社和部分无法收集到在线访问信息的机票预售系统,参数估计中可用到的预订数据属于受截尾数据^[12,13].

3.1 非截尾需求情况

为建立该情况下的完备数据对数似然函数,由于所获得的非截尾数据不存在第二方面的不完备性,仅需通过设置所到来顾客所属具体类型的可能性参数来处理第一方面的数据不完备性.由于 $C_{i(r)}(i, S_i^k, N_i^k)$ 需要与在 $\Delta_i^k(r)$ 中的 $B_{ii}^k(r)$ 和 S_i^k 保持一致,首先定义 N_i^k 中的顾客到达可能性状态矩阵: $G_{ii}^k \in \{0, 1\}^{R_i \times |N_i^k|}$, 其中每一行对应顾客到来的 $\Delta_i^k(r)$, 每一列对应一种顾客类型; 当 $G_{ii}^k(r, j) = I\{\sigma_i^{jk} = \sigma_i^k, j \in C_{i(r)}(i, S_i^k, N_i^k)\} = 1$ 时, 表示顾客类型 σ_i^{jk} 与 $\Delta_i^k(r)$ 中舱位 i 的交易数据 $B_{ii}^k(r)$ 相符, 其中 i 也可不为 0; 否则 $G_{ii}^k(r, j) = 0$ 表示顾客类型 σ_i^{jk} 在 $\Delta_i^k(r)$ 中不可能到来, $j \notin C_{i(r)}(i, S_i^k, N_i^k)$. 相应的, 令 AG_{ii}^k 、 BG_{ii}^k 、 CG_{ii}^k 和 DG_{ii}^k 分别表示与 AN_i^k 、 BN_i^k 、 CN_i^k 和 DN_i^k 相对应的顾客到来可能性状态矩阵, 其元素分别用 $AG_{ii}^k(r, j)$ 、 $BG_{ii}^k(r, j)$ 、 $CG_{ii}^k(r, j)$ 和 $DG_{ii}^k(r, j)$ 表示. 因此, 当 $\Delta_i^k(r) \in O_i$ 时, 若 $B_{ii}^k(r) = 1$, 则

$$C_{i(r)}(i, S_i^k, AN_i^k) = \left\{ \begin{array}{l} j: \sigma_i^{jk}(i) < \sigma_i^k(w), \\ \forall \sigma_i^{jk} \in AN_i^k, \forall w \in S_i^k, w \neq i \end{array} \right\}$$

$$C_{i(r)}(i, S_i^k, BN_i^k) = \left\{ \begin{array}{l} j: \sigma_i^{jk}(i) < \sigma_i^k(w), \\ \forall \sigma_i^{jk} \in BN_i^k, \forall w \in S_i^k, w \neq i \end{array} \right\}$$

$$AG_{ii}^k(r, j) = I\{\sigma_{i(r)}^k = \sigma_i^k, j \in C_{i(r)}(i, S_i^k, AN_i^k)\}$$

$$BG_{ii}^k(r, j) = I\{\sigma_{i(r)}^k = \sigma_i^k, j \in C_{i(r)}(i, S_i^k, BN_i^k)\}$$

同时, 当 $\Delta_i^k(r) \in \bar{O}_i$ 时, 所有舱位的 $B_{ii}^k(r) = 0$, 即顾客选择 $i = 0$, 需根据上述表达式确定 $C_{i(r)}(0, S_i^k, CN_i^k)$ 、 $C_{i(r)}(0, S_i^k, DN_i^k)$ 、 CG_{0i}^k 、 DG_{0i}^k 和 G_{0i}^k ; 最后, 由于 $N_i^k = AN_i^k + BN_i^k = CN_i^k + DN_i^k$, 因此

$$\begin{aligned} C_{i(r)}(i, S_i^k, N_i^k) &= C_{i(r)}(i, S_i^k, AN_i^k) \cup C_{i(r)}(i, S_i^k, BN_i^k) \\ &= C_{i(r)}(0, S_i^k, CN_i^k) \cup C_{i(r)}(0, S_i^k, DN_i^k) \end{aligned}$$

在上述定义过程中, 令各 $AN_i^k = \sigma_i^k$. 当 $t = 1$ 时, $BN_i^k = \emptyset$, 当 $t = (2, 3, \dots, T)$ 时, 需进行如下判断: 如果顾客类型 $\sigma_i^{jk} \in DN_{i-1}^k$ 在 $\Delta_{i-1}^k(r) \in \bar{O}_{i-1}$ 中确定到来, 即 $G_{i-1}^k(r, j) = 1$, 但由于其所偏好舱位在 Δ_{i-1}^k 中均关闭预售, 于是“跨期再现”到 Δ_i^k 中再次购买, 则 $\sigma_i^{jk} \in BN_i^k$; 相反, 若 $G_{ii}^k(r, j) = 0$ 或 $\sigma_i^{jk} \notin DN_{i-1}^k$, 则 $\sigma_i^{jk} \notin BN_i^k$. 在得到 N_i^k 后, 可通过判断获得 CN_i^k 和 DN_i^k 集合. 由此可得非截尾情况下的完备数据对数似然函数, 如式(3)所示.

3.2 截尾需求情况

为克服第二方面的数据不完备性, 该情况下需引入特定时期顾客到来的状态变量 $a_i^k(r)$. 令 $a_i^k(r) = 1$ 表示确定 $\Delta_i^k(r)$ 中有顾客到来但选择不购买, 相反, 若确定没有顾客到来, 令 $a_i^k(r) = 0$. 由于 $\Delta_i^k(r) \in O_i$ 中一定有顾客到来, 故 $a_i^k(r) = 1$; 而对于时期 $\Delta_i^k(r) \in \bar{O}_i$, 事先无法确定 $C_{i(r)}(0, S_i^k, N_i^k)$ 是否为 \emptyset , 因此 $a_i^k(r)$ 的取值仅表示一种可能性大小. 当面对第二方面的数据不完备性时, 对于顾客具体类型信息的处理方法与“非截尾需求”情况类似. 因此, 该情况下的完备数据对数似然函数可由式(4)来表示.

4 $(x_i^k, \lambda_{i(r)}^k)$ 联合估计问题的 EM 算法

本节在考虑舱位间网络替代效应的情况下, 提出了参数联合估计的 EM 算法, 虽然所提方法与文献[13, 17]在计算顺序上有相似之处, 但本文选择模型中的 x_i^k 和 $\lambda_{i(r)}^k$ 在计算结构、范围和数量上与文献[13, 17]有显著差异. 值得注意的是, 式(3)和式(4)在 x_i^k 和 $\lambda_{i(r)}^k$ 上是可分的, 并且是全局凹函数^[13], 这一特性为采用 EM 算法对 $\theta_i^k = (x_i^k, \lambda_{i(r)}^k)$ 进行联合估计提供了便利, 其基本思想如下所示

$$\begin{aligned}
 L(x_i^k, \lambda_{i(r)}^k) &= \sum_{\substack{\Delta_i^k(r) \in O_i, \\ C_{i(r)}(i, S_i^k, N_i^k) \neq \emptyset}} \left[\ln \lambda_{i(r)}^k + \ln \left(\sum_{j \in C_{i(r)}(i, S_i^k, AN_i^k)} I\{\sigma_i^k(r) = \sigma_i^{jk}\} x_i^k(j) + \sum_{j \in C_{i(r)}(i, S_i^k, BN_i^k)} I\{\sigma_i^k(r) = \sigma_i^{jk}\} x_i^k(j) \right) \right] + \\
 &\quad \sum_{\substack{\Delta_i^k(r) \in \bar{O}_i, \\ C_{i(r)}(0, S_i^k, N_i^k) \neq \emptyset}} \left[\ln \lambda_{i(r)}^k + \ln \left(\sum_{j \in C_{i(r)}(0, S_i^k, LN_i^k)} I\{\sigma_i^k(r) = \sigma_i^{jk}\} x_i^k(j) + \sum_{j \in C_{i(r)}(0, S_i^k, DN_i^k)} I\{\sigma_i^k(r) = \sigma_i^{jk}\} x_i^k(j) \right) \right] + \\
 &\quad \sum_{\Delta_i^k(r) \in \bar{O}_i} \ln(1 - \lambda_{i(r)}^k) \\
 &= \sum_{\substack{\Delta_i^k(r) \in O_i, \\ C_{i(r)}(i, S_i^k, N_i^k) \neq \emptyset}} \left[\ln \lambda_{i(r)}^k + \ln \left(\sum_{j \in AN_i^k} AG_{ii}^k(r, j) x_i^k(j) + \sum_{j \in BN_i^k} BG_{ii}^k(r, j) x_i^k(j) \right) \right] + \\
 &\quad \sum_{\substack{\Delta_i^k(r) \in \bar{O}_i, \\ C_{i(r)}(0, S_i^k, N_i^k) \neq \emptyset}} \left[\ln \lambda_{i(r)}^k + \ln \left(\sum_{j \in CN_i^k} CG_{0i}^k(r, j) x_i^k(j) + \sum_{j \in DN_i^k} DG_{0i}^k(r, j) x_i^k(j) \right) \right] + \sum_{\Delta_i^k(r) \in \bar{O}_i} \ln(1 - \lambda_{i(r)}^k) \\
 &= \sum_{j=1}^{|N_i^k|} \left(\sum_{\substack{\Delta_i^k(r) \in O_i, \\ C_{i(r)}(i, S_i^k, N_i^k) \neq \emptyset}} G_{ii}^k(r, j) + \sum_{\substack{\Delta_i^k(r) \in \bar{O}_i, \\ C_{i(r)}(0, S_i^k, N_i^k) \neq \emptyset}} G_{0i}^k(r, j) \right) \ln x_i^k(j) + (|O_i| + |\bar{O}_i|) \ln \lambda_{i(r)}^k + |\bar{O}_i| \ln(1 - \lambda_{i(r)}^k)
 \end{aligned} \tag{3}$$

$$\begin{aligned}
 L(x_i^k, \lambda_{i(r)}^k) &= \sum_{\substack{\Delta_i^k(r) \in O_i, \\ C_{i(r)}(i, S_i^k, N_i^k) \neq \emptyset}} \left[\ln \lambda_{i(r)}^k + \ln \left(\sum_{j \in C_{i(r)}(i, S_i^k, AN_i^k)} I\{\sigma_i^k(r) = \sigma_i^{jk}\} x_i^k(j) + \sum_{j \in C_{i(r)}(i, S_i^k, BN_i^k)} I\{\sigma_i^k(r) = \sigma_i^{jk}\} x_i^k(j) \right) \right] + \\
 &\quad \sum_{\Delta_i^k(r) \in \bar{O}_i} a_i^k(r) \left[\ln \lambda_{i(r)}^k + \ln \left(\sum_{j \in C_{i(r)}(0, S_i^k, LN_i^k)} I\{\sigma_i^k(r) = \sigma_i^{jk}\} x_i^k(j) + \sum_{j \in C_{i(r)}(0, S_i^k, DN_i^k)} I\{\sigma_i^k(r) = \sigma_i^{jk}\} x_i^k(j) \right) \right] + \\
 &\quad \sum_{\Delta_i^k(r) \in \bar{O}_i} (1 - a_i^k(r)) \ln(1 - \lambda_{i(r)}^k) \\
 &= \sum_{\substack{\Delta_i^k(r) \in O_i, \\ C_{i(r)}(i, S_i^k, N_i^k) \neq \emptyset}} \left[\ln \lambda_{i(r)}^k + \ln \left(\sum_{j \in AN_i^k} AG_{ii}^k(r, j) x_i^k(j) + \sum_{j \in BN_i^k} BG_{ii}^k(r, j) x_i^k(j) \right) \right] + \\
 &\quad \sum_{\Delta_i^k(r) \in \bar{O}_i} a_i^k(r) \left[\ln \lambda_{i(r)}^k + \ln \left(\sum_{j \in CN_i^k} CG_{0i}^k(r, j) x_i^k(j) + \sum_{j \in DN_i^k} DG_{0i}^k(r, j) x_i^k(j) \right) \right] + \\
 &\quad \sum_{\Delta_i^k(r) \in \bar{O}_i} (1 - a_i^k(r)) \ln(1 - \lambda_{i(r)}^k) \\
 &= \sum_{j=1}^{i, N_i^k} \left(\sum_{\substack{\Delta_i^k(r) \in O_i, \\ C_{i(r)}(i, S_i^k, N_i^k) \neq \emptyset}} G_{ii}^k(r, j) + \sum_{\Delta_i^k(r) \in \bar{O}_i} a_i^k(r) G_{0i}^k(r, j) \right) \ln x_i^k(j) + \\
 &\quad \left(|O_i| + \sum_{\Delta_i^k(r) \in \bar{O}_i} a_i^k(r) \right) \ln \lambda_{i(r)}^k + \left(|\bar{O}_i| - \sum_{\Delta_i^k(r) \in \bar{O}_i} a_i^k(r) \right) \ln(1 - \lambda_{i(r)}^k)
 \end{aligned} \tag{4}$$

步骤 1 使用已知历史需求和顾客信息对

步骤 2 在第 $p - 1$ 次迭代得到参数估计值

$\hat{\theta}_i^{k(0)} = (\hat{x}_i^{k(0)}, \lambda_{i(r)}^{k(0)})$ 进行初始化 并满足 $\sum_{j \in N_i^k} x_i^k(j)^{(0)} = 1, x_i^k(j)^{(0)} \geq 0, \forall j \in N_i^k, 0 \leq \lambda_{i(r)}^{k(0)} \leq 1$.

$\hat{\theta}_i^{k(p-1)} = (\hat{x}_i^{k(p-1)}, \lambda_{i(r)}^{k(p-1)})$ 后, 可确定完备数据对数似然函数的条件期望值 $E[L(\theta_i^k) | \hat{\theta}_i^{k(p-1)}]$, 即

E (expectation) 步骤.

步骤 3 最大化上述 E 步骤中所得期望对数似然函数,并得到第 p 次迭代后新的估计值 $\hat{\theta}_i^{k(p)}$,即 M (maximization) 步骤.

步骤 4 重复上述 E 步骤和 M 步骤,获得一个参数迭代序列 $\{\hat{\theta}_i^{k(p)} \quad p = 1 \ 2 \ , \dots\}$. 在第 p 次迭代后,需将 $\hat{\theta}_i^{k(p)}$ 与第 $p - 1$ 次迭代后所得 $\hat{\theta}_i^{k(p-1)}$ 进行比较,若满足收敛条件 $\|\hat{\theta}_i^{k(p)} - \hat{\theta}_i^{k(p-1)}\| < \varepsilon$,则终止迭代,否则,返回 E 步骤,直到收敛.

4.1 E 步骤

4.1.1 非截尾需求情况

在将有关 $\lambda_{i(r)}^k$ 的各项从式 (3) 中提取出来后,可得到对数似然函数

$$L(\lambda_{i(r)}^k) = (|O_i| + |\bar{O}_i^\lambda|) \ln \lambda_{i(r)}^k + |\bar{O}_i^\lambda| \ln (1 - \lambda_{i(r)}^k)$$

由此可获得一个极大似然估计封闭解

$$\lambda_{i(r)}^{k*} = (|O_i| + |\bar{O}_i^\lambda|) / R_i^k = \hat{R}_i^k / R_i^k \quad (5)$$

$0 < \lambda_{i(r)}^{k*} < 1$, 由于 $\lambda_{i(r)}^{k*} = \lambda_{i(r)}^{k*} \times R_i^k$, 显然,

$\lambda_{i(r)}^{k*} = (|O_i| + |\bar{O}_i^\lambda|) = \hat{R}_i^k$. 式 (3) 剩余部分组成了似然函数 $L(x_i^k)$ 定义

$$b_i^k(j) = \sum_{\substack{\Delta_i^k(r) \in O_i, \\ C_{i(r)}(i, S_i^k, N_i^k) \neq \emptyset}} G_{ii}^k(r, j) + \sum_{\substack{\Delta_i^k(r) \in \bar{O}_i^\lambda, \\ C_{i(r)}(0, S_i^k, N_i^k) \neq \emptyset}} G_{0i}^k(r, j) \quad (6)$$

上述表达式定义了 N_i^k 中顾客类型 σ_i^{jk} 在非截尾需求情况下选择购买和不购买所到来的总次数, $b_i^k(j) > 0, j = 1 \ 2 \ , \dots, |N_i^k|$, 则

$$L(x_i^k) = \sum_{j=1}^{|N_i^k|} b_i^k(j) \ln x_i^k(j) \quad (7)$$

该情况下 E 步骤中的未知数据是式 (6) 和式 (7) 中的顾客到来状态矩阵 G_{ii}^k . 但在得到 $\hat{x}_i^{k(p-1)}$ 后,可首先确定 $E[b_i^k(j) | \hat{x}_i^{k(p-1)}]$, 故需求解 $\hat{G}_{ii}^{k(p-1)}$. 参考 van Ryzin 和 Vulcano^[13] 所提 EM 算法在完备数据情况下 E 步骤的处理方式,令 $P_r(\sigma_{i(r)}^k = \sigma_i^{jk} | i, S_i^k, N_i^k)$ 表示 $C_{i(r)}(i, S_i^k, N_i^k)$ 中的顾客类型 σ_i^{jk} 在 $\Delta_i^k(r)$ 中选择购买 S_i^k 中舱位 i

的条件概率,利用贝叶斯定理对 $\hat{G}_{ii}^{k(p-1)}$ 进行更新

$$\begin{aligned} \hat{G}_{ii}^k(r, j)^{(p-1)} &= E[G_{ii}^k(r, j) | i, S_i^k, N_i^k; \hat{x}_i^{k(p-1)}] \\ &= P_r(\sigma_{i(r)}^k = \sigma_i^{jk} | i, S_i^k, N_i^k; \hat{x}_i^{k(p-1)}) \\ &= \frac{P_r(i | \sigma_i^{jk}, S_i^k, N_i^k; \hat{x}_i^{k(p-1)}) P_r(\sigma_i^{jk} | N_i^k, \hat{x}_i^{k(p-1)})}{P_r(i | S_i^k, N_i^k; \hat{x}_i^{k(p-1)})} \\ &= \frac{I\{j \in C_{i(r)}(i, S_i^k, N_i^k)\} \hat{x}_i^k(j)^{(p-1)}}{\sum_{j \in C_{i(r)}(i, S_i^k, N_i^k)} \hat{x}_i^k(j)^{(p-1)}} \quad (8) \end{aligned}$$

通过式 (8) 也可对 $\hat{G}_{0i}^{k(p-1)}$ 进行更新. 因此

$$\begin{aligned} \hat{b}_i^k(j)^{(p-1)} &= E[b_i^k(j) | \hat{x}_i^{k(p-1)}] \\ &= \sum_{\substack{\Delta_i^k(r) \in O_i, \\ C_{i(r)}(i, S_i^k, N_i^k) \neq \emptyset}} P_r(\sigma_i^{jk} | i, S_i^k, N_i^k; \hat{x}_i^{k(p-1)}) + \\ &\quad \sum_{\substack{\Delta_i^k(r) \in \bar{O}_i^\lambda, \\ C_{i(r)}(0, S_i^k, N_i^k) \neq \emptyset}} P_r(\sigma_i^{jk} | 0, S_i^k, N_i^k; \hat{x}_i^{k(p-1)}) \quad (9) \\ &= \sum_{\substack{\Delta_i^k(r) \in O_i, \\ C_{i(r)}(i, S_i^k, N_i^k) \neq \emptyset}} \hat{G}_{ii}^k(r, j)^{(p-1)} + \\ &\quad \sum_{\substack{\Delta_i^k(r) \in \bar{O}_i^\lambda, \\ C_{i(r)}(0, S_i^k, N_i^k) \neq \emptyset}} \hat{G}_{0i}^k(r, j)^{(p-1)} \end{aligned}$$

换言之, $\hat{b}_i^k(j)^{(p-1)}$ 代表了第 $p - 1$ 次迭代计算得到的 N_i^k 中顾客类型 σ_i^{jk} 在 Δ_i^k 中到达次数的条件期望值,将其代入式 (7) 便得到了非截尾需求情况下完备数据对数似然函数 $L(x_i^k)$ 的条件期望

$$E[L(x_i^k) | \hat{x}_i^{k(p-1)}] = \sum_{j=1}^{|N_i^k|} \hat{b}_i^k(j)^{(p-1)} \ln \hat{x}_i^k(j)^{(p-1)} \quad (10)$$

4.1.2 截尾需求情况

在将有关 $\lambda_{i(r)}^k$ 的各项从式 (4) 中提取出来后,可得到对数似然函数

$$\begin{aligned} L(\lambda_{i(r)}^k) &= \left(|O_i| + \sum_{\Delta_i^k(r) \in \bar{O}_i} a_i^k(r) \right) \ln \lambda_{i(r)}^k + \\ &\quad \left(|\bar{O}_i| - \sum_{\Delta_i^k(r) \in \bar{O}_i} a_i^k(r) \right) \ln (1 - \lambda_{i(r)}^k) \end{aligned}$$

可获得 $L(\lambda_{i(r)}^k)$ 的唯一极值点解

$$\lambda_{i(r)}^{k*} = \frac{|O_i| + \sum_{\Delta_i^k(r) \in \bar{O}_i} a_i^k(r)}{R_i^k} \quad (11)$$

显然 $\lambda_i^{k*} = \sum_{\Delta_i^k(r) \in (O_i \cup \bar{O}_i)} a_i^k(r)$. 式(4) 剩余部分组成了似然函数 $L(x_i^k)$. 定义

$$b_i^k(j) = \sum_{\substack{\Delta_i^k(r) \in O_i, \\ C_{i(r)}(i, S_i^k, N_i^k) \neq \emptyset}} G_{ii}^k(r, j) + \sum_{\Delta_i^k(r) \in \bar{O}_i} a_i^k(r) G_{0i}^k(r, j) \quad (12)$$

也就是说 $b_i^k(j)$ 计算 N_i^k 中顾客类型 σ_i^{jk} 在截尾需求情况下的各 Δ_i^k 中到达的总次数. 对于 $b_i^k(j) > 0, j = 1, 2, \dots, |N_i^k|$ 则 $L(x_i^k)$ 可表示为

$$L(x_i^k) = \sum_{j=1}^{|N_i^k|} b_i^k(j) \ln x_i^k(j) \quad (13)$$

该情况下 E 步骤中的未知数据是式(12) 和式(13) 中的顾客到来状态矩阵 G_{ii}^k 和状态变量 $a_i^k(r)$. 在第 $p-1$ 次迭代得到参数估计值 $\hat{x}_i^{k(p-1)}$ 后, 可首先确定 $b_i^k(j)$ 的期望值 $E[b_i^k(j) | \hat{x}_i^{k(p-1)}]$. 因此, 需要求解 G_{ii}^k 和 $a_i^k(r)$ 的期望值 $\hat{G}_{ii}^{k(p-1)}$ 和 $\hat{a}_i^k(r)^{(p-1)}$. 根据 van Ryzin 和 Vulcano^[13] 基于贝叶斯理论的处理方式, 可获得 $\hat{a}_i^k(r)^{(p-1)}$ 概率质量函数的更新值. 如式(4) 中的描述, 当 $\Delta_i^k(r) \in O_i$ 时, $\hat{a}_i^k(r)^{(p-1)} = 1$; 而对于 $\Delta_i^k(r) \in \bar{O}_i$ 时, 有

$$\begin{aligned} \hat{a}_i^k(r)^{(p-1)} &= E[a_i^k(r) | \Delta_i^k(r) \in \bar{O}_i; \hat{x}_i^{k(p-1)}] \\ &= P_r(a_i^k(r) = 1 | \Delta_i^k(r) \in \bar{O}_i; \hat{x}_i^{k(p-1)}) \\ &= \frac{P_r(\Delta_i^k(r) \in \bar{O}_i | a_i^k(r) = 1; \hat{x}_i^{k(p-1)}) P_r(a_i^k(r) = 1 | \hat{x}_i^{k(p-1)})}{P_r(\Delta_i^k(r) \in \bar{O}_i | \hat{x}_i^{k(p-1)})} \\ &= \frac{P_r(0 | S_i^k, N_i^k; \hat{x}_i^{k(p-1)}) \hat{\lambda}_{i(r)}^{k(p-1)}}{\hat{\lambda}_{i(r)}^{k(p-1)} P_r(0 | S_i^k, N_i^k; \hat{x}_i^{k(p-1)}) + (1 - \hat{\lambda}_{i(r)}^{k(p-1)})} \end{aligned} \quad (14)$$

其中

$$P_r(0 | S_i^k, N_i^k; \hat{\theta}_i^{k(p-1)}) = \sum_{j \in C_{i(r)}(0, S_i^k, N_i^k)} \hat{x}_i^k(j)^{(p-1)}$$

代表有顾客到来但选择不购买的条件概率.

如遇到 $C_{i(r)}(0, S_i^k, N_i^k) = \emptyset$ 的情况, 则 $P_r(0 | S_i^k, N_i^k; \hat{\theta}_i^{k(p-1)}) = 0$, 从而 $\hat{a}_i^{k(p-1)}(r) = 0$; 它描述的是所有产品都提供预售, 但却没有成功预订交易发生, 反应了没有顾客真实到来的情况.

因此, 由式(12) 可获得如下估计值

$$\begin{aligned} \hat{b}_i^k(j)^{(p-1)} &= E[b_i^k(j) | \hat{\theta}_i^{k(p-1)}] \\ &= \sum_{\substack{\Delta_i^k(r) \in O_i, \\ C_{i(r)}(i, S_i^k, N_i^k) \neq \emptyset}} P_r(\sigma_i^{jk} | i, S_i^k, N_i^k; \hat{\theta}_i^{k(p-1)}) + \sum_{\Delta_i^k(r) \in \bar{O}_i} \hat{a}_i^{k(p-1)}(r) P_r(\sigma_i^{jk} | 0, S_i^k, N_i^k; \hat{\theta}_i^{k(p-1)}) \\ &= \sum_{\Delta_i^k(r) \in O_i, C_{i(r)}(i, S_i^k, N_i^k) \neq \emptyset} \hat{G}_{ii}^k(r, j)^{(p-1)} + \sum_{\Delta_i^k(r) \in \bar{O}_i} \hat{a}_i^k(r)^{(p-1)} \hat{G}_{0i}^k(r, j)^{(p-1)} \end{aligned} \quad (15)$$

其中 $\hat{G}_{ii}^k(r, j)^{(p-1)}$ 或 $P_r(\sigma_i^{jk} | i, S_i^k, N_i^k; \hat{\theta}_i^k)$ 已在式(8) 中得到了定义, 包括 $i=0$ 的情况. 当 $\Delta_i^k(r) \in O_i$ 时

$$\sum_{j \in N_i^k} \sum_{\Delta_i^k(r) \in O_i} \hat{G}_{ii}^k(r, j)^{(p-1)} = \sum_{i=1}^{mk} B_{ii}^k$$

同时, 由于

$$\sum_{j \in N_i^k} G_{ii}^k(r, j)^{(p-1)} = \sum_{j \in N_i^k} \hat{G}_{0i}^k(r, j)^{(p-1)} = 1$$

所以

$$\sum_{j \in N_i^k} \hat{b}_i^k(j)^{(p-1)} = \sum_{\Delta_i^k(r) \in (O_i \cup \bar{O}_i)} \hat{a}_i^k(r)^{(p-1)}$$

将 $\hat{b}_i^k(j)^{(p-1)}$ 代入式(13) 便得到了截尾需求情况下完备数据对数似然函数 $L_{N_i^k}$ 的条件期望

$$E[L(x_i^k) | \hat{x}_i^{k(p-1)}] = \sum_{j=1}^{|N_i^k|} \hat{b}_i^k(j)^{(p-1)} \ln \hat{x}_i^k(j)^{(p-1)} \quad (16)$$

4.2 M 步骤

为了取得使式(3) 和式(4) 最大化的 $\hat{x}_i^{k(p)}$, 考虑到函数 $E[L(x_i^k) | \hat{x}_i^{k(p-1)}]$ 是全局凹的^[13],

则 $\hat{x}_i^{k(p)}$ 的极大似然估计问题可表示如下

$$\begin{aligned} \max \quad & \sum_{j=1}^{|N_i^k|} \hat{b}_i^k(j)^{(p-1)} \ln \hat{x}_i^k(j)^{(p-1)} \\ \text{s. t.} \quad & \sum_{j=1}^{|N_i^k|} \hat{x}_i^k(j)^{(p-1)} = 1, \\ & \hat{x}_i^k(j)^{(p-1)} \geq 0, \forall j \in N_i^k \end{aligned} \quad (17)$$

这是一个凹规划问题,根据卡罗需-库恩-塔克条件(Karush-Kuhn-Tucker conditions, KKT)^[26,27]可得到 $x_i^k(j)^*$ 唯一的极大值封闭解

$$\hat{x}_i^k(j)^{(p)*} = \frac{\hat{b}_i^k(j)^{(p-1)}}{\sum_{h=1}^{|N_i^k|} \hat{b}_i^k(h)^{(p-1)}}, j = 1, 2, \dots, |N_i^k| \quad (18)$$

M 步骤的简便性是本文所提 EM 算法最吸引人的特性之一。

4.3 收敛性检验

与 van Ryzin 和 Vulcano^[13] 所提 EM 算法类似,本文式(10)和式(16)的拟凹性保证了上述迭代过程是一种 EM 算法,是所谓广义 EM 算法(generalized EM algorithm, GEM)^[28]的一个特例,该过程满足一个适当的正则条件,由 Wu^[29] 所提定理 2、Boyles^[30] 所提定理 5、以及 McLachlan 和 Krishnan^[28] 中的定理 3.2 可证明上述结论。

5 网络需求的无约束估计计算

与文献[17]的研究不同,本节不仅给出了历史顾客“初始需求”总量的计算方法,还提出了历史顾客“再现需求”和“溢出需求”总量的估计方法,并在此基础上,总结了历史顾客网络无约束估计需求量之间的关系、以及需求无约束估计效果的评价指标。

5.1 历史顾客“初始需求”总量的估计

由于在非截尾需求和截尾需求情况下均无法获得各 Δ_i^k 中到来顾客的具体类型信息,因此,为了对 M^k 中所有舱位在各 Δ_i^k 中的历史顾客“初始需求”总量 UP_{ii}^k 进行无约束估计计算, $i = (1, 2, \dots, m^k)$, 首先需假设各 Δ_i^k 的 M^k 中所有舱位均开放预售,由于集合 AN_i^k 可通过市场调研确定,于是便可推断出在 AN_i^k 中以舱位 i 为初次购买“首选”的顾客类型集合

$$C_i(i, M^k, AN_i^k) = \left\{ j: \sigma_i^k(j) < \sigma_i^k(w), \forall \sigma_i^k \in AN_i^k, \forall w \in M^k, w \neq i \right\}$$

并用 $P_i(i | M^k, AN_i^k)$ 表示 $C_i(i, M^k, AN_i^k)$ 中的所有顾客在 Δ_i^k 中购买 M^k 所列预售开放舱位 i 的概率。

在利用第 4 节所提 EM 算法估计得到各顾客类型在 Δ_i^k 中到来次数的条件期望值 $\hat{b}_i^k(j)$ 后,有

$$UP_{ii}^k = \lambda_i^{k*} \times P_i(i | M^k, AN_i^k) = \sum_{j \in C_i(i, M^k, AN_i^k)} \hat{b}_i^k(j) \quad (19)$$

即得到 AN_i^k 中各顾客类型在各 Δ_i^k 中对 M^k 所有舱位 i 的真实历史“初始需求”总量的无约束估计值, $UP_{ii}^k \geq 0$ 。

5.2 历史顾客“再现需求”总量的估计

考虑到公司从 CRSs 中获取的基于 O&D 控制的历史可观察订购量 B_{ii}^k 中不仅包括上述计算得到的 UP_{ii}^k , 还包括由于所偏好舱位预售关闭而转移购买 Δ_i^k 中 M^k 所列舱位 i 的历史顾客“再现需求”总量 UR_{ii}^k 。 UR_{ii}^k 既是 AN_i^k 中各短视型和策略型顾客的历史“当期再现”行为结果,其取值又受到 BN_i^k 中各策略型顾客的历史“跨期再现”行为的影响。

令 $C_i(i, M^k, AN_i^k)$ 表示 AN_i^k 中所有可能初次到来后由于 M^k 中所列“首选”舱位关闭预售而“当期再现”至购买舱位 i 的顾客类型集合;同时,令 $C_i(i, M^k, BN_i^k)$ 表示 BN_i^k 中所有可能跨期转移到 Δ_i^k 中并对 M^k 中所列舱位 i 进行购买的策略型顾客类型集合。同样地,分别可用 $P_i(i | M^k, AN_i^k)$ 和 $P_i(i | M^k, BN_i^k)$ 表示 $C_i(i, M^k, AN_i^k)$ 和 $C_i(i, M^k, BN_i^k)$ 中的所有顾客在 Δ_i^k 中购买 M^k 所列预售开放舱位 i 的概率。于是有如下关系式

$$UR_{ii}^k = URA_{ii}^k + URB_{ii}^k = \lambda_i^{k*} \times P_i(i | M^k, AN_i^k) + \lambda_i^{k*} \times P_i(i | M^k, BN_i^k) = \sum_{j \in C_i(i, M^k, AN_i^k)} \hat{b}_i^k(j) + \sum_{j \in C_i(i, M^k, BN_i^k)} \hat{b}_i^k(j) \quad (20)$$

即得到 AN_i^k 和 BN_i^k 中各顾客类型在各 Δ_i^k 中对 M^k 所列舱位 i 的历史顾客“再现需求”总量的无约束估计值, $UR_{ii}^k \geq 0$ 。

5.3 历史顾客“溢出需求”总量的估计

5.3.1 非截尾需求情况

在此情况下,由于公司能够跟踪和记录历史顾客到来后未进行购买的行为,即顾客需求“溢出”效应,因此,很容易从 CRSs 中获得在 Δ_i^k 中对

预售关闭舱位 i 进行问询并最终选择“不购买”的历史顾客“溢出需求”总量 US_{ii}^k .

5.3.2 截尾需求情况

当有 S_i^k 所列舱位提供预售时,若没有购买行为发生,或顾客选择 $i = 0$,公司在截尾需求情况下无法区分各 Δ_i^k 中是否有顾客到来. 令 \bar{S}_i^k 表示 M^k 所列预售关闭 ($I_{ii}^k = 0$) 的舱位集合,将短视型和策略型顾客在 Δ_i^k 中对 \bar{S}_i^k 所列舱位的“溢出”需求量计入 US_{ii}^k ,分别通过已知的 CN_i^k 和 DN_i^k 集合,确定在 Δ_i^k 中选择预售关闭舱位并最终“不购买”、或跨期转移到 Δ_{i+1}^k 中的顾客类型集合

$$C_i(i, \bar{S}_i^k, CN_i^k) = \left\{ \begin{array}{l} j: \sigma_i^{jk}(i) < \sigma_i^{jk}(w), \\ \forall \sigma_i^{jk} \in CN_i^k, \forall w \in \bar{S}_i^k, w \neq i \end{array} \right\}$$

$$C_i(i, \bar{S}_i^k, DN_i^k) = \left\{ \begin{array}{l} j: \sigma_i^{jk}(i) < \sigma_i^{jk}(w), \\ \forall \sigma_i^{jk} \in DN_i^k, \forall w \in \bar{S}_i^k, w \neq i \end{array} \right\}$$

同时, $C_i(i, \bar{S}_i^k, CN_i^k)$ 和 $C_i(i, \bar{S}_i^k, DN_i^k)$ 中的顾客在 Δ_i^k 中对 \bar{S}_i^k 所列舱位 i 产生“溢出”效应的概率可分别用 $P_i(i | \bar{S}_i^k, CN_i^k)$ 和 $P_i(i | \bar{S}_i^k, DN_i^k)$ 表示. 在通过本文所提 EM 算法计算得到 $C_i(i, \bar{S}_i^k, CN_i^k)$ 和 $C_i(i, \bar{S}_i^k, DN_i^k)$ 中各顾客类型在 Δ_i^k 中到来次数的条件期望值 $\hat{b}_i^k(j)$ 后,有

$$US_{ii}^k = USC_{ii}^k + USD_{ii}^k$$

$$= -\lambda_i^{k*} \times P_i(i | \bar{S}_i^k, CN_i^k) - \lambda_i^{k*} \times P_i(i | \bar{S}_i^k, DN_i^k) \quad (21)$$

$$= - \sum_{j \in C_i(i, \bar{S}_i^k, CN_i^k)} \hat{b}_i^k(j) - \sum_{j \in C_i(i, \bar{S}_i^k, DN_i^k)} \hat{b}_i^k(j)$$

即截尾需求情况下的历史顾客“溢出需求”分别由 CN_i^k 和 DN_i^k 中的顾客需求在各 Δ_i^k 中对 \bar{S}_i^k 所列舱位的“溢出”效应共同构成, $US_{ii}^k \leq 0$.

5.4 历史顾客网络无约束估计需求量之间的关系

由上述估计计算过程便可对各 Δ_i^k 中不同舱位所到来的历史顾客需求总量 TD_{ii}^k 进行定义,并且截尾需求情况下历史顾客网络无约束估计需求量 UP_{ii}^k 、 UR_{ii}^k 和 US_{ii}^k 之间满足如下关系

$$TD_{ii}^k = B_{ii}^k + |US_{ii}^k| = UP_{ii}^k + UR_{ii}^k \quad (22)$$

当舱位预售开放时($I_{ii}^k = 1$)不存在历史顾客的“溢出需求”, $US_{ii}^k = 0$ 系统所记录到的可观察预售量由历史顾客网络无约束估计“初始需求”量和“再现需求”量组成;当舱位预售关闭时($I_{ii}^k = 0$)系统所记录到的可观察预售量 $B_{ii}^k = 0$ 这并不代表没有顾客需求到来的可能,而是由那些没有成功预订的历史顾客网络无约束估计“初始需求”量、“再现需求”量和“溢出需求”量共同造成了可观察预售量等于 0 这一现象,即无论是初次到来、“当期再现”、还是“跨期再现”到 Δ_i^k 中的顾客,所有在该情况下到来的顾客均发生了“溢出”效应. 令

$$BR_i^k = \sum_{i=1}^{m^k} B_{ii}^k / \sum_{i=1}^{m^k} TD_{ii}^k \quad (23)$$

$$PR_i^k = \sum_{i=1}^{m^k} UP_{ii}^k / \sum_{i=1}^{m^k} TD_{ii}^k \quad (24)$$

$$RR_i^k = RRA_i^k + RRB_i^k = \sum_{i=1}^{m^k} URA_{ii}^k / \sum_{i=1}^{m^k} TD_{ii}^k + \sum_{i=1}^{m^k} URB_{ii}^k / \sum_{i=1}^{m^k} TD_{ii}^k = \sum_{i=1}^{m^k} UR_{ii}^k / \sum_{i=1}^{m^k} TD_{ii}^k \quad (25)$$

$$SR_i^k = SRC_i^k + SRD_i^k = \sum_{i=1}^{m^k} |USC_{ii}^k| / \sum_{i=1}^{m^k} TD_{ii}^k + \sum_{i=1}^{m^k} |USD_{ii}^k| / \sum_{i=1}^{m^k} TD_{ii}^k = \sum_{i=1}^{m^k} |US_{ii}^k| / \sum_{i=1}^{m^k} TD_{ii}^k \quad (26)$$

其中 BR_i^k 、 PR_i^k 、 RR_i^k 和 SR_i^k 分别表示各 Δ_i^k 中历史顾客的成功预订率、“初始需求”到达率、“再现需求”到达/发生率和“溢出需求”发生率,它们是分别用于衡量各 Δ_i^k 中历史顾客需求组成情况的结构相对指标.

6 数值模拟

6.1 无约束估计计算过程

6.1.1 O&D 航线网络基本情况介绍

如图 1 所示,国内某航空公司的航线网络中包含了 6 条航段,出发地和目的地机场有:贵阳(龙洞堡机场)、武汉(天河机场)、北京(首都机场)和广州(白云机场);该航线网络中以出发地 k 为起点的 O&D 行程分别如表 2 所示;其中, $D =$

{ 贵阳、武汉} $d = 2$,以贵阳和武汉为出发地的 O&D 行程数量分别为 $h^1 = 6$ 和 $h^2 = 2$.

以 2019 年 5 月 21 日晚上 0 点至 5 月 22 日晚上 0 点从贵阳起飞的所有航班为观察对象 ,现考虑贵阳到北京航线($k = 1$) ,它由两条直达行程($l = 1$ 和 $l = 2$) 和一条中转联程行程($l = 5$) 构成 ,假定每条行程各自有三种类型的舱位机票可供顾客预订 ,分别是经济舱、商务舱和头等舱 , $|M^{1,1}| = |M^{2,1}| = |M^{5,1}| = 3$, $m = 9$ 表 3 列出了上述三条行程在五个 Δ_i^1 中的舱位等级情况 , $M^{1,1} = \{1, 2, 3\}$, $M^{2,1} = \{4, 5, 6\}$, $M^{5,1} = \{7, 8, 9\}$. 票价由低到高按字母顺序排列 ,各 Δ_i^k 中的舱位价格保持不变;同时 ,表 3 中还包括各舱位的历史可观察订购量 B_{it}^k 和历史预售开放状态 I_{it}^k .

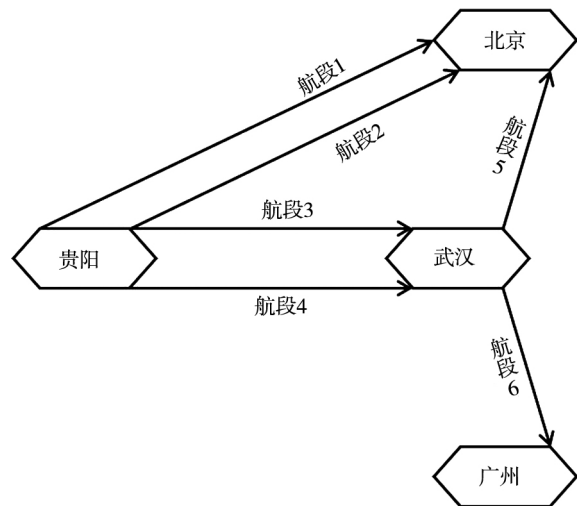


图 1 航线网络示例图
Fig. 1 Example of airline network

表 2 航线网络中的 O&D 行程示例

Table 2 Examples of O&D itinerary in airline network

出发地名称	出发地编号 k	集合 H^k		航段编号	航班时刻
		O&D 行程	编号 l		
贵阳	1	贵阳→北京	1	1	上午 08: 10
		贵阳→北京	2	2	晚上 19: 50
		贵阳→武汉	3	3	上午 09: 50
		贵阳→武汉	4	4	下午 13: 30
		贵阳→武汉→北京	5	{3, 5}	上午 09: 50、中午 12: 40
		贵阳→武汉→广州	6	{4, 6}	下午 13: 30、下午 16: 20
武汉	2	武汉→北京	1	5	中午 12: 40
		武汉→广州	2	6	下午 16: 20

表 3 贵阳→北京航线上($k = 1$) 各行程航班舱位等级在各 Δ_i^1 中价格保持不变情况下的 B_{it}^1 和 I_{it}^1

Table 3 B_{it}^1 and I_{it}^1 at the Guiyang→Beijing itinerary , where the price on each Δ_i^1 remains unchanged

行程编号 l	舱位类型	舱位编号	舱位等级(B_{it}^1, I_{it}^1)				
			Δ_1^1	Δ_2^1	Δ_3^1	Δ_4^1	Δ_5^1
1	经济	1	C(15, 1)	C(14, 1)	C(0, 0)	C(0, 0)	C(0, 0)
	商务	2	E(9, 1)	E(12, 1)	E(17, 1)	E(0, 0)	E(0, 0)
	头等	3	I(0, 1)	I(1, 1)	I(2, 1)	I(3, 1)	I(2, 1)
2	经济	4	B(19, 1)	B(0, 0)	B(0, 0)	B(0, 0)	B(0, 0)
	商务	5	F(8, 1)	F(15, 1)	F(16, 1)	F(13, 1)	F(0, 0)
	头等	6	H(1, 1)	H(1, 1)	H(1, 1)	H(2, 1)	H(3, 1)
5	经济	7	A(27, 1)	A(0, 0)	A(0, 0)	A(0, 0)	A(0, 0)
	商务	8	D(10, 1)	D(14, 1)	D(17, 1)	D(0, 0)	D(0, 0)
	头等	9	G(0, 1)	G(1, 1)	G(1, 1)	G(2, 1)	G(2, 1)

6.1.2 顾客类型集合介绍

由表 3 可知,各 Δ_i^1 中的报价集合分别为 $S_1^1 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 0\}$ 、 $S_2^1 = \{1, 2, 3, 5, 6, 8, 9, 0\}$ 、 $S_3^1 = \{2, 3, 5, 6, 8, 9, 0\}$ 、 $S_4^1 = \{3, 5, 6, 9, 0\}$ 、

$S_5^1 = \{3, 6, 9, 0\}$ 。公司通过市场调研可获得贵阳到北京航线上的公司历史市场份额,令 $\rho^1 = 0.25$, 以及针对该航线上短视型顾客和策略型顾客类型集合 $\sigma_i^1 = \alpha_i^1 \cup \beta_i^1$, 如表 4 和表 5 所示。

表 4 考虑网络替代效应的短视型顾客集 $\alpha_i^1 = \{\sigma_i^{1,1}, \sigma_i^{2,1}, \dots, \sigma_i^{68,1}\}$

Table 4 Myopic customer set $\alpha_i^1 = \{\sigma_i^{1,1}, \sigma_i^{2,1}, \dots, \sigma_i^{68,1}\}$ considering the substitution effects in network

顾客敏感类型	顾客偏好排序列表
价格	$\sigma_i^{1,1} = (A, \rho)$ 、 $\sigma_i^{2,1} = (A, B, \rho)$ 、 $\sigma_i^{3,1} = (A, B, C, D, E, \rho)$
舱位 & 价格	$\sigma_i^{4,1} = (A, B, C, \rho)$ 、 $\sigma_i^{5,1} = (A, B, C, D, E, F, \rho)$ 、 $\sigma_i^{6,1} = (D, E, F, \rho)$ 、 $\sigma_i^{7,1} = (D, E, F, G, H, I, \rho)$ 、 $\sigma_i^{8,1} = (G, H, I, \rho)$
时间 & 舱位	$\sigma_i^{9,1} = (1, \rho)$ 、 $\sigma_i^{10,1} = (2, \rho)$ 、 $\sigma_i^{11,1} = (3, \rho)$ 、 $\sigma_i^{12,1} = (4, \rho)$ 、 $\sigma_i^{13,1} = (5, \rho)$ 、 $\sigma_i^{14,1} = (6, \rho)$ 、 $\sigma_i^{15,1} = (7, \rho)$ 、 $\sigma_i^{16,1} = (8, \rho)$ 、 $\sigma_i^{17,1} = (9, \rho)$
时间 & 价格	$\sigma_i^{18,1} = (1, 2, 3, \rho)$ 、 $\sigma_i^{19,1} = (4, 5, 6, \rho)$ 、 $\sigma_i^{20,1} = (7, 8, 9, \rho)$
时间 & 舱位 & 价格	$\sigma_i^{21,1} = (1, 2, \rho)$ 、 $\sigma_i^{22,1} = (2, 3, \rho)$ 、 $\sigma_i^{23,1} = (4, 5, \rho)$ 、 $\sigma_i^{24,1} = (5, 6, \rho)$ 、 $\sigma_i^{25,1} = (7, 8, \rho)$ 、 $\sigma_i^{26,1} = (8, 9, \rho)$
行程 & 舱位 & 时间	$\sigma_i^{27,1} = (1, A, \rho)$ 、 $\sigma_i^{28,1} = (4, 1, \rho)$ 、 $\sigma_i^{29,1} = (1, A, 7, \rho)$ 、 $\sigma_i^{30,1} = (1, 7, A, \rho)$ 、 $\sigma_i^{31,1} = (4, 1, 7, \rho)$ 、 $\sigma_i^{32,1} = (7, 1, A, \rho)$ 、 $\sigma_i^{33,1} = (2, 5, \rho)$ 、 $\sigma_i^{34,1} = (5, 2, \rho)$ 、 $\sigma_i^{35,1} = (2, 5, 8, \rho)$ 、 $\sigma_i^{36,1} = (2, 8, 5, \rho)$ 、 $\sigma_i^{37,1} = (5, 2, 8, \rho)$ 、 $\sigma_i^{38,1} = (8, 2, 5, \rho)$ 、 $\sigma_i^{39,1} = (3, 6, \rho)$ 、 $\sigma_i^{40,1} = (6, 3, \rho)$ 、 $\sigma_i^{41,1} = (3, 6, 9, \rho)$ 、 $\sigma_i^{42,1} = (3, 9, 6, \rho)$ 、 $\sigma_i^{43,1} = (6, 3, 9, \rho)$ 、 $\sigma_i^{44,1} = (9, 3, 6, \rho)$
行程 & 时间 & 舱位 & 价格	$\sigma_i^{45,1} = (1, 2, A, 5, \rho)$ 、 $\sigma_i^{46,1} = (4, 5, 1, 2, \rho)$ 、 $\sigma_i^{47,1} = (1, 2, A, 5, 7, 8, \rho)$ 、 $\sigma_i^{48,1} = (1, 2, 7, 8, A, 5, \rho)$ 、 $\sigma_i^{49,1} = (4, 5, 1, 2, 7, 8, \rho)$ 、 $\sigma_i^{50,1} = (7, 8, 1, 2, A, 5, \rho)$ 、 $\sigma_i^{51,1} = (2, 3, 5, 6, \rho)$ 、 $\sigma_i^{52,1} = (5, 6, 2, 3, \rho)$ 、 $\sigma_i^{53,1} = (2, 3, 5, 6, 8, 9, \rho)$ 、 $\sigma_i^{54,1} = (2, 3, 8, 9, 5, 6, \rho)$ 、 $\sigma_i^{55,1} = (5, 6, 2, 3, 8, 9, \rho)$ 、 $\sigma_i^{56,1} = (8, 9, 2, 3, 5, 6, \rho)$
行程 & 舱位 & 时间 & 价格	$\sigma_i^{57,1} = (1, A, 2, 5, \rho)$ 、 $\sigma_i^{58,1} = (4, 1, 5, 2, \rho)$ 、 $\sigma_i^{59,1} = (1, A, 7, 2, 5, 8, \rho)$ 、 $\sigma_i^{60,1} = (1, 7, 4, 2, 8, 5, \rho)$ 、 $\sigma_i^{61,1} = (4, 1, 7, 5, 2, 8, \rho)$ 、 $\sigma_i^{62,1} = (7, 1, A, 8, 2, 5, \rho)$ 、 $\sigma_i^{63,1} = (2, 5, 3, 6, \rho)$ 、 $\sigma_i^{64,1} = (5, 2, 6, 3, \rho)$ 、 $\sigma_i^{65,1} = (2, 5, 8, 3, 6, 9, \rho)$ 、 $\sigma_i^{66,1} = (2, 8, 5, 3, 9, 6, \rho)$ 、 $\sigma_i^{67,1} = (5, 2, 8, 6, 3, 9, \rho)$ 、 $\sigma_i^{68,1} = (8, 2, 5, 9, 3, 6, \rho)$

表 5 考虑网络替代效应的策略型顾客集 $\beta_i^1 = \{\sigma_i^{69,1}, \sigma_i^{70,1}, \dots, \sigma_i^{113,1}\}$

Table 5 Strategic customer set $\beta_i^1 = \{\sigma_i^{69,1}, \sigma_i^{70,1}, \dots, \sigma_i^{113,1}\}$ considering the substitution effects in network

顾客敏感类型	顾客偏好排序列表
价格	$\sigma_i^{69,1} = (A_i, A_{i+1}, A_{i+2}, \rho)$ 、 $\sigma_i^{70,1} = (A_i, B_i, A_{i+1}, B_{i+1}, \rho)$ 、 $\sigma_i^{71,1} = (A_i, A_{i+1}, B_{i+1}, A_{i+2}, B_{i+2}, C_{i+2}, \rho)$
舱位 & 价格	$\sigma_i^{72,1} = (A_i, B_i, C_i, A_{i+1}, B_{i+1}, C_{i+1}, \rho)$ 、 $\sigma_i^{73,1} = (A_i, B_i, C_i, A_{i+1}, B_{i+1}, C_{i+1}, A_{i+2}, B_{i+2}, C_{i+2}, \rho)$ 、 $\sigma_i^{74,1} = (D_i, E_i, F_i, D_{i+1}, E_{i+1}, F_{i+1}, \rho)$
时间 & 舱位	$\sigma_i^{75,1} = (1_i, 1_{i+1}, 1_{i+2}, \rho)$ 、 $\sigma_i^{76,1} = (2_i, 2_{i+1}, \rho)$ 、 $\sigma_i^{77,1} = (4_i, 4_{i+1}, 4_{i+2}, \rho)$ 、 $\sigma_i^{78,1} = (5_i, 5_{i+1}, \rho)$ 、 $\sigma_i^{79,1} = (7_i, 7_{i+1}, 7_{i+2}, \rho)$ 、 $\sigma_i^{80,1} = (8_i, 8_{i+1}, \rho)$
时间 & 价格	$\sigma_i^{81,1} = (1_i, 1_{i+1}, 2_{i+1}, 1_{i+2}, 2_{i+2}, 3_{i+2}, \rho)$ 、 $\sigma_i^{82,1} = (4_i, 4_{i+1}, 5_{i+1}, 4_{i+2}, 5_{i+2}, 6_{i+2}, \rho)$ 、 $\sigma_i^{83,1} = (7_i, 7_{i+1}, 8_{i+1}, 7_{i+2}, 8_{i+2}, 9_{i+2}, \rho)$
时间 & 舱位 & 价格	$\sigma_i^{84,1} = (1_i, 1_{i+1}, 2_{i+1}, 1_{i+2}, 2_{i+2}, \rho)$ 、 $\sigma_i^{85,1} = (2_i, 2_{i+1}, 3_{i+1}, \rho)$ 、 $\sigma_i^{86,1} = (4_i, 4_{i+1}, 5_{i+1}, 4_{i+2}, 5_{i+2}, \rho)$ 、 $\sigma_i^{87,1} = (5_i, 5_{i+1}, 6_{i+1}, \rho)$ 、 $\sigma_i^{88,1} = (7_i, 7_{i+1}, 8_{i+1}, 7_{i+2}, 8_{i+2}, \rho)$ 、 $\sigma_i^{89,1} = (8_i, 8_{i+1}, 9_{i+1}, \rho)$
行程 & 时间 & 舱位	$\sigma_i^{90,1} = (1_i, 1_{i+1}, 1_{i+2}, 1_{i+3}, 1_{i+4}, \rho)$ 、 $\sigma_i^{91,1} = (4_i, 4_{i+1}, 4_{i+2}, 4_{i+3}, 4_{i+4}, \rho)$ 、 $\sigma_i^{92,1} = (1_i, 1_{i+1}, 1_{i+2}, 1_{i+3}, 1_{i+4}, 1_{i+5}, \rho)$ 、 $\sigma_i^{93,1} = (1_i, 1_{i+1}, 1_{i+2}, 1_{i+3}, 1_{i+4}, 1_{i+5}, 1_{i+6}, \rho)$ 、 $\sigma_i^{94,1} = (4_i, 4_{i+1}, 4_{i+2}, 4_{i+3}, 4_{i+4}, 4_{i+5}, \rho)$ 、 $\sigma_i^{95,1} = (7_i, 7_{i+1}, 7_{i+2}, 7_{i+3}, 7_{i+4}, 7_{i+5}, \rho)$ 、 $\sigma_i^{96,1} = (2_i, 2_{i+1}, 2_{i+2}, 2_{i+3}, 2_{i+4}, \rho)$ 、 $\sigma_i^{97,1} = (5_i, 5_{i+1}, 5_{i+2}, 5_{i+3}, 5_{i+4}, \rho)$ 、 $\sigma_i^{98,1} = (2_i, 2_{i+1}, 2_{i+2}, 2_{i+3}, 2_{i+4}, 2_{i+5}, \rho)$ 、 $\sigma_i^{99,1} = (2_i, 2_{i+1}, 2_{i+2}, 2_{i+3}, 2_{i+4}, 2_{i+5}, 2_{i+6}, \rho)$ 、 $\sigma_i^{100,1} = (5_i, 5_{i+1}, 5_{i+2}, 5_{i+3}, 5_{i+4}, 5_{i+5}, \rho)$ 、 $\sigma_i^{101,1} = (8_i, 8_{i+1}, 8_{i+2}, 8_{i+3}, 8_{i+4}, 8_{i+5}, \rho)$
行程 & 时间 & 舱位 & 价格	$\sigma_i^{102,1} = (1_i, 2_i, 1_{i+1}, 2_{i+1}, 1_{i+2}, 2_{i+2}, \rho)$ 、 $\sigma_i^{103,1} = (4_i, 5_i, 4_{i+1}, 5_{i+1}, 4_{i+2}, 5_{i+2}, \rho)$ 、 $\sigma_i^{104,1} = (1_i, 2_i, 1_{i+1}, 2_{i+1}, 1_{i+2}, 2_{i+2}, 1_{i+3}, 2_{i+3}, \rho)$ 、 $\sigma_i^{105,1} = (1_i, 2_i, 1_{i+1}, 2_{i+1}, 1_{i+2}, 2_{i+2}, 1_{i+3}, 2_{i+3}, 1_{i+4}, 2_{i+4}, \rho)$ 、 $\sigma_i^{106,1} = (4_i, 5_i, 4_{i+1}, 5_{i+1}, 4_{i+2}, 5_{i+2}, 4_{i+3}, 5_{i+3}, \rho)$ 、 $\sigma_i^{107,1} = (7_i, 8_i, 7_{i+1}, 8_{i+1}, 7_{i+2}, 8_{i+2}, 7_{i+3}, 8_{i+3}, \rho)$
行程 & 舱位 & 时间 & 价格	$\sigma_i^{108,1} = (1_i, 1_{i+1}, 2_i, 1_{i+2}, 1_{i+3}, 2_{i+4}, \rho)$ 、 $\sigma_i^{109,1} = (4_i, 4_{i+1}, 5_i, 4_{i+2}, 4_{i+3}, 5_{i+4}, 2_{i+5}, \rho)$ 、 $\sigma_i^{110,1} = (1_i, 1_{i+1}, 1_{i+2}, 1_{i+3}, 1_{i+4}, 1_{i+5}, 1_{i+6}, 1_{i+7}, \rho)$ 、 $\sigma_i^{111,1} = (1_i, 1_{i+1}, 1_{i+2}, 1_{i+3}, 1_{i+4}, 1_{i+5}, 1_{i+6}, 1_{i+7}, 1_{i+8}, \rho)$ 、 $\sigma_i^{112,1} = (4_i, 4_{i+1}, 4_{i+2}, 4_{i+3}, 4_{i+4}, 4_{i+5}, 4_{i+6}, 4_{i+7}, \rho)$ 、 $\sigma_i^{113,1} = (7_i, 7_{i+1}, 7_{i+2}, 7_{i+3}, 7_{i+4}, 7_{i+5}, 7_{i+6}, 7_{i+7}, \rho)$

6.1.3 网络型离散选择模型的建立

可根据上述信息得到各 Δ_i^1 中的 $C_{i(r)}^1 (i, S_i^1, N_i^1)$ 和 $C_{i(r)}^1 (0, S_i^1, N_i^1)$ 集合,利用本文第 3 节中的式(1)和式(2)建立考虑了顾客策略行为的网络型非参数离散选择模型.

6.1.4 基于 EM 算法的联合估计

本文数值模拟仅考虑“截尾需求”情况,在此情

况下,需利用第 4 节所建立的 EM 算法对各 Δ_i^1 中的 $\theta_i^1 = (x_i^1, \lambda_{i(r)}^1)$ 进行联合估计.这里将 EM 算法计算过程中的迭代收敛条件 $\|\hat{\theta}_i^{(p)} - \theta_i^{(p-1)}\| < \varepsilon$ 设置为 $\varepsilon = 0.0001$,并令各 Δ_i^1 中的参数初始化为 $\hat{x}_i^1(j)^{(0)} = 1/|N_i^1|$ 、 $\hat{\lambda}_{i(r)}^{1(0)} = \rho^1$.

从参数估计结果来看,所有 Δ_i^1 中本文所提

EM 算法的迭代次数不超过 2 次,并且随着预售时间的推移, O&D 航线网络中关闭预售舱位的比例也随之增加,估计所得的截尾需求情况下 $\Delta_i^1(r)$ 中的顾客到达率 $\lambda_{i(r)}^{1*}$ 也相应增大,如图 2 所示;这表明随着预售关闭的舱位比例增加,越来越多的策略型顾客需求会因为“溢出”效应而产生跨期替代行为,即造成了往后各 $\Delta_i^1(r)$ 中 $\lambda_{i(r)}^{1*}$ 的增大。

6.1.5 网络需求无约束估计计算结果

在使用式(15)得到截尾需求情况下 N_i^1 中顾客类型 σ_i^j 在 Δ_i^1 中到来次数的条件期望值 $\hat{b}_i^1(j)^{(1)*}$ 后,可利用式(19)~式(21)分别对 UP_i^1 、 UP_i^1 和 US_i^1 进行计算,并根据式(22)~式(26)对相应的性能指

标进行评价。结果如表 6 和图 3 所示。

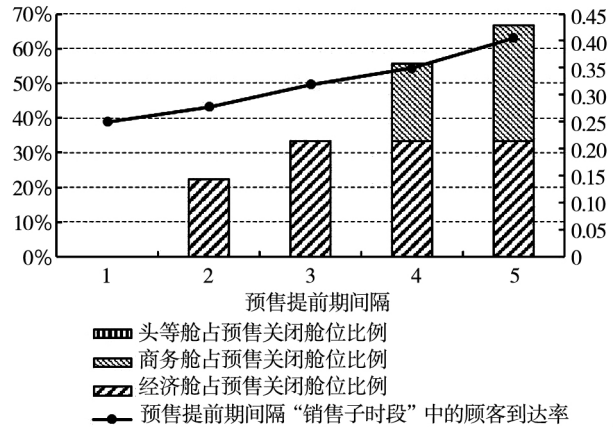


图 2 预售关闭舱位比例与顾客到达率

Fig. 2 Closed booking classes proportion and customer arrival rate

表 6 贵阳→北京航线上 ($k = 1$) 各行程航班舱位等级在 Δ_i^1 中的无约束估计计算结果

Table 6 The unconstraining estimation results of fare classes on Δ_i^1 at Guiyang→Beijing itinerary

时间	舱位类型	舱位集合	$\sum_i B_{it}^1$	$\sum_i UP_{it}^1$	再现需求总量		$\sum_i UR_{it}^1$	溢出需求总量		$\sum_i US_{it}^1$
					$\sum_i URA_{it}^1$	$\sum_i URB_{it}^1$		$\sum_i USC_{it}^1$	$\sum_i USD_{it}^1$	
Δ_1^1	经济	{1 4 7}	61	61.000 0	0	0	0	0	0	0
	商务	{2 5 8}	27	27.000 0	0	0	0	0	0	0
	头等	{3 6 9}	1	1.000 0	0	0	0	0	0	0
Δ_2^1	经济	{1 4 7}	14	29.159 5	8.500 9	0	8.500 9	-12.702 6	-10.957 8	-23.660 4
	商务	{2 5 8}	41	32.266 7	8.733 3	0	8.733 3	0	0	0
	头等	{3 6 9}	3	3	0	0	0	0	0	0
Δ_3^1	经济	{1 4 7}	0	33.939 7	27.398 4	5.766 3	33.164 7	-33.934 5	-33.169 9	-67.104 4
	商务	{2 5 8}	50	25.971 5	21.076 2	2.952 3	24.028 5	0	0	0
	头等	{3 6 9}	4	4	0	0	0	0	0	0
Δ_4^1	经济	{1 4 7}	0	11.365 0	11.690 4	5.527 5	17.217 9	-13.517 3	-15.065 6	-28.582 9
	商务	{2 5 8}	13	9.027 4	19.048 0	0.970 6	20.018 6	-10.390 4	-5.655 6	-16.046 0
	头等	{3 6 9}	7	4.333 3	2.666 7	0	2.666 7	0	0	0
Δ_5^1	经济	{1 4 7}	0	2.888 7	2.611 4	2.664 1	5.275 5	-4.609 8	-3.554 4	-8.164 2
	商务	{2 5 8}	0	4.300 9	6.672 4	1.048 3	7.720 7	-9.736 7	-2.284 9	-12.021 6
	头等	{3 6 9}	7	2.171 4	4.142 9	0.685 7	4.828 6	0	0	0

从具体数值上看,在 Δ_1^1 中,预售关闭舱位总比例为 0, $RR_1^1 = SR_1^1 = 0$,而 $BR_1^1 = PR_1^1 = 100\%$;当预售关闭舱位总比例在 Δ_3^1 中超过 30% 后, PR_3^1 和 RR_3^1 的数值接近,而 BR_3^1 和 SR_3^1 的数值接近;在 Δ_4^1 过后,预售关闭舱位的总比例超过了

50%,其中,经济舱占预售关闭舱位比例为 33.33%,商务舱占预售关闭舱位比例为 22.22%, RR_i^1 和 SR_i^1 的数值开始明显大于 PR_i^1 和 BR_i^1 的数值。也就是说,随着预售关闭舱位比例的增加,顾客网络需求“再现”效应也随之增加,直

接导致顾客“再现需求”到达/发生率 RR_i^1 的增加;与此同时,表6所列数据表明,“再现需求”中初次到来顾客的发生率 RRA_i^1 随着预售关闭舱位比例的增加而增加,并且占到整个“溢出需求”总量的绝大部分比例;而“溢出需求”发生率 SR_i^1 增加的直接原因也是因为预售关闭舱位比例的上升,顾客越往后越难买到自己首要偏好的舱位,于是成功预订率 BR_i^1 也开始随时间推移而下降。

因此,图3所列性能指标验证了网络环境下顾客需求的截尾机理,证明了本文所提网络需求无约束估计计算方法以及所提性能指标的有效性。同时,在所有的无约束估计过程中,本文所提EM算法的计算过程均能较快收敛,说明所提EM算法对O&D网络需求的无约束估计计算过程简单易行。

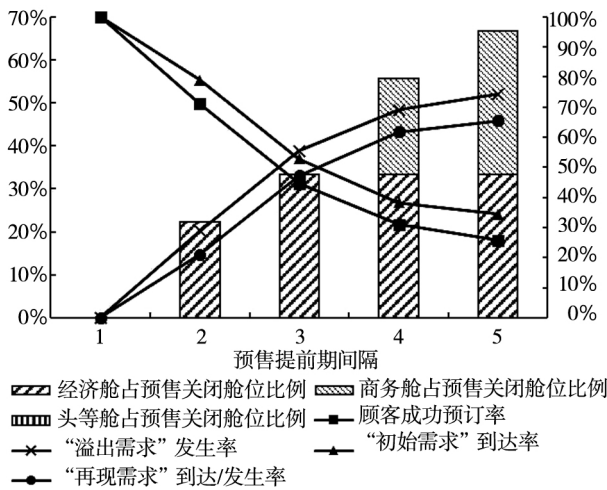


图3 预售关闭舱位比例与网络需求无约束估计结果性能指标
Fig. 3 Closed booking classes proportion and performance indexes of network demand unconstrained estimation results

6.2 无约束估计效果比较

为了说明O&D产品网络替代效应、顾客策略行为、市场份额、以及产品价格变化对本文所提方法无约束估计效果的影响,首先,令表3中所列各行程舱位价格等级在各 Δ_i^1 中会发生变化,而 B_{ii}^1 和 I_{ii}^1 相对表3均保持不变,如表7所示。

当采用本文所提方法对表7中所列数据进行无约束估计时,用“EM-RPL-O&D-M&S”表示考虑

了第6.1节表4和表5中 σ_i^1 集合所定义的全部顾客选择行为, $\sigma_i^1 = \alpha_i^1 \cup \beta_i^1$; 符号“EM-RPL-O&D-M”表示将 β_i^1 中的策略型顾客当作短视型顾客处理,不考虑其延期购买行为,由表5所列策略型顾客偏好列表可知,若不考虑策略型顾客的跨期替代偏好,则表5中所列 β_i^1 中的策略型顾客与表4中所列 α_i^1 中的短视型顾客重合,因此“EM-RPL-O&D-M”情况下所考虑的顾客集合仅由短视型顾客组成, $\sigma_i^1 = \alpha_i^1$ 。针对上述问题,若采用郭鹏等^[17]、以及 van Ryzin 和 Vulcano^[13] 所提方法,由于其未考虑O&D行程替代效应,其无约束估计计算过程无法区分直达航班和中转联程航班需求,仅能处理平行直达航班间的需求“溢出”和“再现”效应,即仅考虑表3和表7中行程1和行程2所代表的两条平行直达航班间的需求替代效应,由于行程5的第一航段由贵阳到武汉直达航班构成,其所考虑的顾客偏好列表不包括表4和表5中涉及到行程5中舱位等级偏好的顾客类型。于是,可分别用“EM-RPL-M&S”和“EM-RPL-M”表示郭鹏等^[17]、以及 van Ryzin 和 Vulcano^[13] 所提方法在考虑和未考虑策略型顾客行为情况下的计算结果。

考虑到 Δ_4^1 过后预售关闭舱位总比例超过了50%,顾客需求“再现”和“溢出”效应非常显著,因此在 $\rho = (0.15, 0.25, \dots, 0.85)$ 的情况下,分别使用上述各方法对表3和表7所列 Δ_4^1 和 Δ_5^1 中各类型舱位的历史顾客“初始需求”进行无约束估计。由于文献[13, 17]所提方法无法对行程5中的舱位需求进行考虑,为了结果的可比性,仅列出由各方法计算得出的行程1和行程2这两条直达航线中各类型舱位在 Δ_4^1 和 Δ_5^1 中的“初始需求”无约束估计平均到达率 $\overline{PR}_i^1 = \frac{\sum_{i=4}^5 UP_{ii}^1}{\sum_{i=4}^5 TD_{ii}^1}$,它是用于衡量顾客“初始需求”在各类型舱位顾客需求到达总量中所占比重的性能指标。结果如表8所示。同样,整个计算过程借助MATLAB 9.7实现。

表 7 贵阳→北京航线上($k = 1$) 各行程航班舱位等级在各 Δ_i^1 中价格变化情况下的 B_{it}^1 和 I_{it}^1

Table 7 B_{it}^1 and I_{it}^1 at the Guiyang→Beijing itinerary , where the price changes on each Δ_i^1

行程 编号 l	舱位 类型	舱位 编号	舱位等级 (B_{it}^1 I_{it}^1)				
			Δ_1^1	Δ_2^1	Δ_3^1	Δ_4^1	Δ_5^1
1	经济	1	C(15 ,1)	B(14 ,1)	B(0 0)	B(0 0)	B(0 0)
	商务	2	E(9 ,1)	E(12 ,1)	F(17 ,1)	F(0 0)	F(0 0)
	头等	3	I(0 ,1)	I(1 ,1)	I(2 ,1)	H(3 ,1)	I(2 ,1)
2	经济	4	B(19 ,1)	C(0 0)	C(0 0)	C(0 0)	C(0 0)
	商务	5	F(8 ,1)	F(15 ,1)	E(16 ,1)	E(13 ,1)	E(0 0)
	头等	6	H(1 ,1)	H(1 ,1)	H(1 ,1)	I(2 ,1)	H(3 ,1)
5	经济	7	A(27 ,1)	A(0 0)	A(0 0)	A(0 0)	A(0 0)
	商务	8	D(10 ,1)	D(14 ,1)	D(17 ,1)	D(0 0)	D(0 0)
	头等	9	G(0 ,1)	G(1 ,1)	G(1 ,1)	G(2 ,1)	G(2 ,1)

表 8 不同市场份额下 Δ_4 和 Δ_5 中各类型舱位历史顾客无约束估计“初始需求”的平均到达率(单位: %)

Table 8 Average arrival rate of unconstrained primary demand on both Δ_4 and Δ_5 with different market shares (%)

舱位 价格	舱位 类型	方法名称	市场份额(ρ)							
			0. 15	0. 25	0. 35	0. 45	0. 55	0. 65	0. 75	0. 85
保持 不 变	经济	EM-RPL-O&D-M&S	35. 93	36. 33	36. 75	37. 20	37. 96	39. 05	39. 90	41. 50
		EM-RPL-O&D-M	45. 79	46. 07	46. 38	46. 64	47. 21	48. 03	48. 50	49. 68
		EM-RPL-M&S	47. 46	47. 85	48. 08	48. 49	48. 96	50. 05	51. 15	52. 89
		EM-RPL-M	64. 94	65. 07	65. 09	65. 15	65. 25	65. 69	65. 97	66. 82
	商务	EM-RPL-O&D-M&S	32. 58	33. 00	33. 45	33. 77	34. 39	35. 06	35. 38	36. 04
		EM-RPL-O&D-M	33. 43	33. 81	34. 22	34. 49	35. 11	35. 82	36. 13	36. 87
		EM-RPL-M&S	37. 69	37. 94	38. 03	38. 18	38. 39	38. 91	39. 24	39. 83
		EM-RPL-M	39. 15	39. 46	39. 52	39. 68	39. 92	40. 71	41. 22	42. 22
	头等	EM-RPL-O&D-M&S	47. 00	47. 00	47. 00	47. 00	47. 00	47. 00	47. 00	47. 00
		EM-RPL-O&D-M	48. 67	48. 67	48. 67	48. 67	48. 67	48. 67	48. 67	48. 67
		EM-RPL-M&S	49. 05	49. 05	49. 05	49. 05	49. 05	49. 05	49. 05	49. 05
		EM-RPL-M	50. 00	50. 00	50. 00	50. 00	50. 00	50. 00	50. 00	50. 00
发 生 变 化	经济	EM-RPL-O&D-M&S	35. 73	36. 11	36. 51	36. 93	37. 65	38. 68	39. 46	40. 97
		EM-RPL-O&D-M	45. 50	45. 75	46. 03	46. 24	46. 76	47. 47	47. 83	48. 85
		EM-RPL-M&S	48. 08	48. 44	48. 63	48. 97	49. 38	50. 39	51. 36	53. 01
		EM-RPL-M	64. 69	64. 81	64. 82	64. 84	64. 91	65. 31	65. 51	66. 33
	商务	EM-RPL-O&D-M&S	33. 16	33. 57	34. 02	34. 33	34. 95	35. 60	35. 92	36. 57
		EM-RPL-O&D-M	34. 25	34. 62	35. 03	35. 29	35. 90	36. 59	36. 88	37. 60
		EM-RPL-M&S	39. 75	39. 98	40. 05	40. 18	40. 36	40. 85	41. 13	41. 69
		EM-RPL-M	41. 76	42. 06	42. 11	42. 25	42. 47	43. 23	43. 69	44. 68
	头等	EM-RPL-O&D-M&S	49. 05	49. 05	49. 05	49. 05	49. 05	49. 05	49. 05	49. 05
		EM-RPL-O&D-M	50. 00	50. 00	50. 00	50. 00	50. 00	50. 00	50. 00	50. 00
		EM-RPL-M&S	50. 00	50. 00	50. 00	50. 00	50. 00	50. 00	50. 00	50. 00
		EM-RPL-M	51. 67	51. 67	51. 67	51. 67	51. 67	51. 67	51. 67	51. 67

由表 8 可知:

1) 在舱位价格发生变化前后,被测试的所有方法在不同市场份额情况下所得结果均发生了变化,这表明基于顾客偏好排序列表的非参数离散选择模型能有效反应舱位价格变化对无约束估计结果的影响.

2) 随着公司市场份额的增大,被测试的所有方法所得结果仅在经济舱和商务舱中发生递增变化,而在头等舱中保持不变,这是由所测试方法的计算原理决定的.当舱位预售关闭时,参与比较的所有方法所得到的历史顾客“初始需求”总量均与 $|O_i|$ 和 $|\bar{O}_i|$ 的大小有关,而其中的 $|\bar{O}_i| = R_i^k - \sum_{i=1}^m B_u^k, R_i^k = \sum_{i=1}^m B_u^k / \rho^k$,因此,由于表 3 和表 7 所列 Δ_4^1 和 Δ_5^1 中的经济舱和商务舱存在预售关闭情况,它们的历史顾客“初始需求”总量会受到市场份额大小的影响,从而表现为“初始需求”无约束估计平均到达率随市场份额的增大而增大;同时,当舱位预售开放时,其历史顾客“初始需求”的计算仅需考虑 O_i 中的顾客选择行为,而 $|O_i| = \sum_{i=1}^m B_u^k$,由于 Δ_4^1 和 Δ_5^1 中的头等舱均预售开放,故针对它们的无约束估计结果不会随市场份额大小的变化而变化.

3) 从是否考虑策略型顾客的角度看,“EM-RPL-O&D-M&S”和“EM-RPL-M&S”方法在不同市场份额的情况下所得结果均分别明显小于“EM-RPL-O&D-M”和“EM-RPL-M”方法.“EM-RPL-O&D-M”和“EM-RPL-M”方法在无约束估计过程中误将策略型顾客对舱位的“跨期再现需求”作为了“初始需求”,而“EM-RPL-O&D-M&S”和“EM-RPL-M&S”则能避免该问题.该结论在经济舱表现得最为明显,而对于商务舱和头等舱来讲,虽然“EM-RPL-O&D-M&S”和“EM-RPL-M&S”方法分别相对于“EM-RPL-O&D-M”和“EM-RPL-M”方法所得结果仍是最小,但它们之间的结果差异开始变得不大.由表 5 中的 β_i^1 所定义的策略型顾客延期购买行为可知,当 β_i^1 中的顾客在 Δ_4^1 或 Δ_5^1

之前无法购买到当期偏好的舱位时,他们延期到 Δ_4^1 或 Δ_5^1 中选择购买经济舱机票的顾客类型数量明显多于“跨期再现”购买商务舱和头等舱的顾客类型,即由于 Δ_4^1 或 Δ_5^1 之前预售关闭舱位主要集中在经济舱,若以经济舱为“初始需求”偏好,则 β_i^1 中所列策略型顾客更加偏好延期到 Δ_4^1 或 Δ_5^1 中对经济舱机票进行购买.因此,相较于 van Ryzin 和 Vulcano^[13] 的研究,本文和郭鹏等^[17] 所提方法能准确反应顾客策略行为对无约束估计结果的影响,能够更加有效地避免顾客策略行为而导致无约束估计过程对历史顾客“初始需求”的过量估计,从而有效控制预测模型对未来顾客“初始需求”的高估.

4) 当考虑了中转联程航线产品与直达航线产品之间的 O&D 行程网络替代效应时,“EM-RPL-O&D-M&S”和“EM-RPL-O&D-M”方法在不同市场份额的情况下所得结果均分别明显小于“EM-RPL-M&S”和“EM-RPL-M”方法.本文表 4 和表 5 中所定义的顾客类型集考虑进了 O&D 行程的当期和跨期网络替代效应,即当顾客无法购买到所偏好的中转联程舱位时,他们会当期或跨期再现购买平行直达航班产品.由于文献 [13, 17] 所提方法仅针对行程 1 和行程 2 所代表的两条平行直达航班产品间替代效应,在无约束估计过程中无法考虑上述网络替代效应,会误将顾客对平行直达航班的当期或跨期“再现需求”作为当期“初始需求”,而“EM-RPL-O&D-M&S”和“EM-RPL-O&D-M”则能避免该问题.该结论在经济舱和商务舱表现得最为明显,而对于头等舱来讲,虽然“EM-RPL-O&D-M&S”和“EM-RPL-O&D-M”方法分别相对于“EM-RPL-M&S”和“EM-RPL-M”方法所得结果仍是最小,但它们之间的结果差异开始变得不大.表 4 和表 5 中选择购买平行直达航班经济舱和商务舱机票的顾客类型数量明显多于购买头等舱的顾客类型,即由于 Δ_4^1 和 Δ_5^1 预售关闭舱位主要集中在经济舱和商务舱,若以中转联程航线行程 5 中的经济舱和商务舱为“初始需求”偏好,则表 4 和表 5 中所列顾客更加偏好

“再现”到 Δ_4^1 和 Δ_5^1 中平行直达航班的经济舱和商务舱机票进行购买. 因此, 在网络需求无约束估计计算过程中, 由于中转联程航线与平行直达航线之间存在网络替代效应, 本文所提方法能够有效地避免对历史顾客“初始需求”的过量估计问题, 从而有效控制预测模型对未来顾客“初始需求”的高估.

7 结束语

来自于乘客 O&D 仿真系统(Passenger O&D Simulation, PODS) 的研究结论表明, 相较于航段控制(leg control), O&D 控制能够为收益管理应用企业带来 1% ~ 2% 的收入增长^[5]. O&D 控制层面所需数据主要来源于 O&D 预测, 后者的准确度高直接关系到收益管理系统实施的成败. 本文首先基于顾客偏好排序列表定义了 O&D 短视型和策略型顾客类型集合, 使用离散概率质量函数对顾客选择行为进行描述, 建立了考虑顾客策略行为的网络型非参数离散选择模型. 其次, 针对顾客到达率和所提选择模型的概率质量函数的联合估计问题, 分别站在线上交易平台和线下实体交易平台的角度, 在非截尾需求情况和截尾需求情况下建立了完备数据对数似然函数. 然后, 考虑到网络环境下历史预售数据的不完备性, 在仅能获取 CRSs 中基于 O&D 控制的历史可观察订购

量、历史产品预售开放状态和市场份额信息的情况下, 分别针对完备数据对数似然函数的参数联合估计问题, 提出了相应的 EM 算法、以及考虑历史顾客策略行为的 O&D 顾客“初始需求”、“再现需求”和“溢出需求”的无约束估计计算方法. 最后, 通过数值模拟验证了所提网络需求无约束估计方法的可行性和有效性, 在准确反映产品间网络替代效应对短视型和策略型顾客选择行为共同影响的同时, 能有效避免现有非参数多航班方法对历史顾客“初始需求”的高估问题.

值得注意的是, 随着 O&D 网络规模的扩大, 以顾客偏好排序列表为基础的顾客类型数量呈指数级增长, 公司通过市场调研方法获得准确 O&D 顾客类型信息的难度也随之增大, 极有可能由于对顾客信息的判断失误而对网络需求无约束估计的准确性造成影响, 使得本文所提网络型无约束估计方法的应用成本急剧上升. 因此, 仍然需要提出一种能够自动改进完备数据对数似然函数的“市场发现算法”, 即可在本文所提基于 EM 算法的联合估计过程中, 通过自动发现新顾客类型对事先预定义的初始顾客类型集合进行自动扩充, 这符合大数据时代收益管理需求无约束估计的应用要求^[6]. 同时, 在对公司网络收益管理实践的收入影响等性能指标上, 有必要对本文所提网络型方法和传统方法作进一步地仿真比较研究.

参 考 文 献:

- [1] Tirumalachetty S, Kodam V, Doreswamy G, et al. Unlocking the value from origin and destination revenue management [J]. Journal of Revenue and Pricing Management, 2017, 16(6): 607-620.
- [2] Doreswamy G R, Kothari A S, Tirumalachetty S. Simulating the flavors of revenue management for airlines [J]. Journal of Revenue and Pricing Management, 2015, 14(6): 421-432.
- [3] Guo P, Xiao B C, Li J. Unconstraining methods in revenue management systems: Research overview and prospects [J]. Advances in Operations Research, 2012, 2012(270910): 1-23.
- [4] Azadeh S S, Marcotte P, Savard G. A taxonomy of demand uncensoring methods in revenue management [J]. Journal of Revenue and Pricing Management, 2014, 13(6): 440-456.
- [5] Weatherford L. The history of unconstraining models in revenue management [J]. Journal of Revenue and Pricing Management, 2016, 15(3): 222-228.
- [6] 郭 鹏. 网络环境下收益管理系统需求无约束估计综述 [J]. 计算机工程与应用, 2017, 53(19): 17-25.

- Guo Peng. Review of demand unconstraining estimation in revenue management systems from perspective of network [J]. *Computer Engineering and Application*, 2017, 53(19): 17–25. (in Chinese)
- [7] Nikseresht A, Ziarati K. Estimating true demand in airline's revenue management systems using observed sales [J]. *International Journal of Advanced Computer Science and Applications*, 2017, 8(7): 361–369.
- [8] 郭 鹏. 收益管理中的多分布网络需求无约束估计模型 [J]. *统计与决策*, 2018, 497(5): 44–49.
- Guo Peng. A model for multi-distribution network demand unconstraining estimation in revenue management systems [J]. *Statistics & Decision*, 2018, 497(5): 44–49. (in Chinese)
- [9] Price I, Fowkes J, Hopman D. Gaussian processes for demand unconstraining [J]. *European Journal of Operational Research*, 2019, 275(2): 621–634.
- [10] Drabas T, Wu C L. A market spill-recapture unconstraining model for estimating airline true demand [J]. *Journal of Revenue and Pricing Management*, 2020, 19(1): 43–60.
- [11] Lapp M, Weatherford L. Airline network revenue management: Considerations for implementation [J]. *Journal of Revenue and Pricing Management*, 2014, 13(2): 83–112.
- [12] van Ryzin G, Vulcano G. A market discovery algorithm to estimate a general class of nonparametric choice models [J]. *Management Science*, 2015, 61(2): 281–300.
- [13] van Ryzin G, Vulcano G. Technical note: An expectation-maximization method to estimate a rank-based choice model of demand [J]. *Operations Research*, 2017, 65(2): 396–407.
- [14] 李 贺, 张玉林, 仲伟俊. 考虑战略消费者行为风险的动态定价策略 [J]. *管理科学学报*, 2012, 15(10): 11–25.
- Li He, Zhang Yulin, Zhong Weijun. Dynamic pricing strategies in the presence of strategic consumer behavior risks [J]. *Journal of Management Sciences in China*, 2012, 15(10): 11–25. (in Chinese)
- [15] 曾贺奇, 张玉林. 考虑策略消费者行为及模仿产品质量不确定的两竞争商定价 [J]. *管理工程学报*, 2019, 33(1): 214–221.
- Zeng Heqi, Zhang Yulin. Two retailers' competitive pricing considering strategic consumer behavior and imitation products with quality uncertainty [J]. *Journal of Industrial Engineering/Engineering Management*, 2019, 33(1): 214–221. (in Chinese)
- [16] 李宗活, 杨文胜, 司银元, 等. 短视型与策略型消费者并存的双渠道两阶段动态定价策略 [J]. *系统工程理论与实践*, 2019, 39(8): 2080–2090.
- Li Zonghuo, Yang Wensheng, Si Yinyuan, et al. Two-period dynamic pricing strategy of dual-channel with myopic and strategic consumers [J]. *Systems Engineering: Theory & Practice*, 2019, 39(8): 2080–2090. (in Chinese)
- [17] 郭 鹏, 萧柏春, 李 军. 收益管理中考虑顾客策略行为的多航班无约束估计方法 [J]. *系统工程理论与实践*, 2015, 35(5): 1103–1115.
- Guo Peng, Xiao Baichun, Li Jun. Multi-flight unconstraining method considering strategic customer behavior in revenue management [J]. *Systems Engineering: Theory & Practice*, 2015, 35(5): 1103–1115. (in Chinese)
- [18] Weatherford L R. Intelligent aggressiveness: Using forecast multipliers, hybrid forecasting, fare adjustment, and unconstraining methods to increase revenue [J]. *Decision Sciences*, 2017, 48(3): 391–419.
- [19] Weatherford L R, Bodily S E. A taxonomy and research overview of perishable-asset revenue management: Yield management, overbooking, and pricing [J]. *Operations Research*, 1992, 40(5): 831–844.
- [20] Mahajan S, van Ryzin G. Stocking retail assortments under dynamic consumer substitution [J]. *Operations Research*, 2001, 49(3): 334–351.
- [21] Farias V F, Jagabathula S, Shah D. A nonparametric approach to modeling choice with limited data [J]. *Management Science*, 2013, 59(2): 305–322.
- [22] Jagabathula S, Rusmevichientong P. A nonparametric joint assortment and price choice model [J]. *Management Science*, 2017, 63(9): 2773–3145.
- [23] Paul A, Feldman J, Davis J M. Assortment optimization and pricing under a nonparametric tree choice model [J]. *Manufacturing & Service Operations Management*, 2018, 20(3): 550–565.

- [24] Jagabathula S , Vulcano G. A partial-order-based model to estimate individual preferences using panel data [J]. *Management Science* , 2018 , 64(4) : 1609 – 1628.
- [25] Jagabathula S , Rusmevichientong P. The limit of rationality in choice modeling: Formulation , computation , and implications [J]. *Management Science* , 2019 , 65(5) : 2196 – 2215.
- [26] Karush W. *Minima of Functions of Several Variables with Inequalities As Side Constraints* [D]. Chicago: University of Chicago , 1939.
- [27] Kuhn H W , Tucker A W. *Nonlinear Programming* [C] // *Proc of 2nd Berkeley Symposium* , Berkeley: University of California Press , 1951: 481 – 492.
- [28] McLachlan G J , Krishnan T. *The EM Algorithm and Extensions* , 2nd ed. [M]. Hoboken: John Wiley & Sons , Inc. , 2008: 18 – 29 , 77 – 82.
- [29] Wu C F. On the convergence properties of the EM algorithm [J]. *The Annals of Statistics* , 1983 , 11: 95 – 103.
- [30] Boyles R A. On the convergence of the EM algorithm [J]. *Journal of the Royal Statistical Society (Series B)* , 1983 , 45 (1) : 47 – 50.

A nonparametric discrete choice model for unconstrained demand estimation in airline network

GUO Peng

School of Economics and Management , Guiyang University , Guiyang 550005 , China

Abstract: Traditional unconstrained demand estimation methods in airline industry only address the spillover and recapture problem of customer demand in parallel nonstop flights , and fail to consider the substitution effects in airline networks between direct and connecting flights. Based on the ranking list of customer preferences , customer type set in airline network is defined. Meanwhile , a network nonparametric discrete choice model considering strategic customer behavior is developed. In light of the incompleteness of the historical booking data in network environment , from the perspective of online and offline trading platforms , the complete data log-likelihood functions under the conditions of uncensored and censored demand are established respectively. The EM algorithm is applied to jointly estimating customer arrival rate and the probability mass function. After that , the unconstrained estimation calculation methods of primary demand , recapture demand as well as spillover demand of customers in airline network are proposed. The feasibility and effectiveness of the proposed methods are verified by numerical simulations. Compared with the existing methods , our methods can accurately reflect the impact of substitution effects in networks between products on customer choice behavior , and thereby more effectively avoid overestimating the primary demand of historical customers.

Key words: airline network; unconstrained demand estimation; substitution effects in network; nonparametric discrete choice model; EM algorithm