新兴出行服务下通勤过程中的病毒传播风险建模：新冠肺炎案例研究

Baichuan Mo[a]、Peyman Noursalehi[b]、Haris N. Koutsopoulos[c]、Jinhua Zhao[b]

[a]麻省理工学院土木与环境工程系，马萨诸塞州剑桥，邮编 *02139*[b]麻省理工学院城市研究与规划系，马萨诸塞州剑桥，邮编 *20139*[c]东北大学土木与环境工程系，马萨诸塞州波士顿，邮编 *02115*

arXiv:2301.02594v1 [stat.AP] 6 Jan 2023

摘要

通勤是日常生活的重要组成部分。随着新冠肺炎疫情逐渐缓解，越来越多的人重返办公室工作，通勤过程中的新冠肺炎传播问题日益受到关注。近期兴起的出行服务（如网约车和共享单车）由于出行过程中涉及共享车辆或空间，进一步加剧了感染风险。因此，量化通勤过程中的感染风险十分重要。本文提出了一个概率框架，用于估算个体通勤过程中的感染风险，并考虑公共交通、拼车、骑自行车和步行等不同出行方式。研究旨在根据起讫点（OD）、出发时间和出行方式，评估感染概率及其估计误差（即不确定性量化）。我们首先定义了一个通用行程规划函数，根据起讫点、出发时间和出行方式生成出行轨迹，并计算选择不同路径的概率。随后，我们考虑两种感染途径：1）通过密切接触感染；2）通过接触表面感染。感染风险按行程路段进行计算。我们使用不同来源的数据（如智能卡数据、出行调查和人口数据），估算个体与传染环境之间可能发生的接触。我们采用一阶近似来降低计算复杂度。此外，我们还推导出估计误差的闭式表达式，从而能够量化估计结果的不确定性。我们将模型应用于麻省理工学院社区，并以此作为案例研究，评估员工和学生的通勤感染风险。结果显示，大多数个体的感染概率接近于零，最高感染概率约为 0.8%，这表明通勤过程中感染的概率较低。出行距离较长、乘坐公共交通以及在高峰时段出行的个体更有可能感染。本文还讨论了该模型的实际应用。

关键词： 新冠肺炎；感染风险；新兴出行方式。

1. 引言

新冠肺炎极大地影响了全世界人们的生活。近期，随着疫苗接种工作的开展以及人们防疫意识的增强，我们正迈入与病毒共存的新时代。随着大流行病逐渐得到控制，越来越多的人开始回到办公室工作。

通勤是办公室工作人员日常生活的重要组成部分。鉴于新冠肺炎具有传染性，通勤过程中的感染风险令人担忧，尤其是

*通讯作者

对于乘坐公共交通的人而言，许多既有研究均指出了这一点（Mo 等人，2021；Zhou 和 Koutsopoulos，2021）。另一方面，近期兴起的出行服务（如网约车和共享单车）会因出行过程中共用车辆或空间而进一步加剧感染风险。因此，有必要更全面地量化通勤中的感染风险；相关结果有助于人们评估自身的健康风险，并更好地做出通勤路线和出行方式选择，也有助于政策制定者做出知情决策。

许多既有研究对公共交通系统中的新冠肺炎感染风险进行了建模。公共交通是重要的通勤出行方式。这些研究可分为城市尺度的宏观层面研究（Mo 等人，2021）和车辆尺度的微观尺度研究（Shinohara 等人，2021；Zhou 和 Koutsopoulos，2021）。研究人员还考察了通勤对新冠肺炎更广泛空间传播的影响及相关防控策略（Mitze 和 Kosfeld，2022；Ando 等人，2021；Fajgelbaum 等人，2021；Kondo，2021）。此外，也有研究利用实证数据（如调查）评估新冠肺炎对通勤过程的影响，例如对客流变化、出行方式选择（Tan 和 Ma，2021；Medlock 等人，2021）以及出发时间变化的影响（Ecke 等人，2022）。

然而，仍存在两个研究空白。首先，先前研究均未考虑将多种出行方式和多模式出行行程纳入整体通勤过程的感染建模。其次，大多数关于感染风险建模的先前研究仅输出感染概率（或表示传播强度的 R_0 值），并未提供估计不确定性（即估计结果的准确程度）。

本研究提出一个概率框架，用于估计个体采用不同出行方式通勤期间的感染风险，包括公共交通、拼车、骑自行车和步行。该模型既能评估感染概率，也能评估估计误差（即不确定性量化）。我们首先定义一个通用的行程规划函数，根据出发地、目的地、出发时间和出行方式，生成出行轨迹以及选择不同路径的概率。我们考虑两种感染途径：1）通过密切接触感染；2）通过接触表面感染。感染风险按行程路段进行计算。我们使用不同来源的数据（如智能卡数据、出行调查和人口数据）来估计个体与传染环境之间可能发生的互动。我们采用一阶近似技术来降低计算复杂度。我们推导出估计误差的闭式表达式，从而能够量化估计结果的不确定性。我们以麻省理工学院社区为案例研究，实施该模型，并评估员工和学生的通勤感染风险。结果表明，大多数个体的感染概率接近于零。感染概率最高约为 0.8%，这意味着通勤过程中感染的概率较低。出行距离较长、乘坐公共交通以及在高峰时段出行的个体更容易感染。

本文的主要贡献有两点：

- 本研究首次专门探讨通勤过程中的病毒传播建模，并考虑了各种出行方式和多模式出行行程。研究同时涵盖了因密切接触和接触表面导致的感染。

- 除了估算感染概率外，本文还计算了用于不确定性量化的估计误差（即估算概率的标准差），而这在现有文献中尚未涉及。

其余部　　本文的组织结构如下。文献综述见第2节。在

第3节介绍问题并讨论解决方法。第4节将所提出的框架应用于麻省理工学院社区，作为一项案例研究。最后，我们在第5节总结本研究并概述主要发现。

2. 文献综述

2.1. 公共交通中的感染建模

公共交通是重要的通勤出行方式。既有研究探讨了公共交通网络中的疫情传播和感染风险建模。 Mo 等人(2021) 提出了一种时变加权接触网络，用于模拟传染病通过公共交通系统传播。该模型在大都市层面实施，用于人群感染计算。 Zhou 和 Koutsopoulos (2021) 提出了一种改进的 Wells-Riley 模型，用于在车辆层面计算公共交通系统中的感染概率。该模型通过上车乘客和下车乘客的数量以及感染者的数量，捕捉客流的时空特征。类似地， Ku 等人 (2021) 通过模拟乘客在旅途中相遇并相互传染的过程，分析了公共交通中的感染暴露程度。 Shinohara 等人(2021) 通过收集各种条件下的换气率数据，采用双区域指数模型计算通勤列车中的感染风险。 Zhao 等人 (2022) 开发了一种基于 Wells-Riley 模型的方法，用于定量评估乘坐公共交通的感染风险。他们比较了不同应对措施在控制疫情传播方面的效果。

2.2. 新冠肺炎对通勤的影响

新冠肺炎可能会在许多方面影响通勤过程，例如乘客量和服务频率下降、乘客出行方式选择、路线选择以及出发时间选择的变化。以往研究利用不同来源的实证数据评估了这些影响。例如，许多研究使用智能卡数据分析新冠肺炎对公共交通客流变化的影响（Ahangari 等人，2020；Chang 等人，2021；Wilbur 等人，2020；Jenelius 和 Cebecauer，2020）。此外，也有研究考察网约车系统（Meredith-Karam 等人，2021）和共享单车系统（Wang 和 Noland，2021）的客流变化。 Tan 和 Ma （2021）开展了一项调查，以了解新冠肺炎疫情期间通勤者的出行方式选择变化。他们基于一项包含559份调查答卷的样本，采用逻辑回归模型分析个人属性、出行属性以及对新冠肺炎的认知。 Ecke 等人（2022）考察了人们在新冠肺炎疫情前后通勤行为的变化。结果显示，就通勤时间和通勤方式而言，人们的通勤行为并未发生显著变化。

2.3. 通勤对新冠肺炎传播的影响

通勤可能通过将具有传染性的乘客运送到不同地区，促进新冠肺炎传播。例如， Mitze 和 Kosfeld (2022) 提出了一个疫情传播空间计量经济模型，以识别通勤模式在疾病空间传播中发挥的作用，并探讨实施的遏制政策是否影响了这一传播渠道的强度。 Ando 等人(2021) 利用来自 27,036 名受访者的互联网调查数据，研究通勤、新冠肺炎感染风险与新冠肺炎引发的焦虑之间的关系。 Fajgelbaum 等人 (2021) 在通勤网络中设计了一种应对新冠肺炎的最优动态封锁策略。 Kondo(2021) 开发了空间易感者–暴露者–感染者–康复者模型，以分析限制跨区域通勤流动对日本新冠肺炎感染空间传播的影响。

2.4. 研究空白

据作者所知，现有研究尚未针对包含多种出行方式和多模式出行行程的整个通勤过程，开展专门的感染建模。大多数研究聚焦于单一出行方式（例如公共交通）的感染建模，或是在城市层面一般性地对疫情传播进行建模；在这类研究中，通勤过程仅是宏大框架的一部分，并未采用针对出行方式或行程方案的建模方法。另一方面，以往大多数有关感染风险建模的研究仅输出感染概率或表征传播强度的 R0 值，并未计算估计不确定性（即结果的准确程度）。

3. 方法论

3.1. 问题定义

考虑一组个体 $\mathcal{I}$。对于每个个体 $i \in \mathcal{I}$，假设我们知道其日常通勤的出发地 o_i、目的地 d_i、出发时间 t_i 和出行方式 m_i。本研究旨在估计个体 i 受到影响的概率： $\mathbb{P}(i \text{ 感染} \mid o_i,\ d_i,\ t_i,\ m_i)$。此外，我们还旨在量化估计的不确定性。如果将 $\mathbb{P}(i \text{ 感染} \mid o_i,\ d_i,\ t_i,\ m_i)$视为随机变量，本研究的另一目标是获得该估计的标准误：$\sqrt{\text{Var}[\mathbb{P}(i \text{ 感染} \mid o_i,\ d_i,\ t_i,\ m_i)]}$。

在以下各节中，我们首先说明如何估计 $\mathbb{P}(i \text{ 感染} \mid o_i,\ d_i,\ t_i,\ m_i)$。标准误的估计将在第3.5节中说明。

3.2. 行程方案生成

给定个人 i 的行程信息（$o_i,\ d_i,\ t_i,\ m_i$），存在一个行程规划函数 $TP(\cdot)$，以（$o_i,\ d_i,\ t_i,\ m_i$）为输入，并输出该个人 $\mathcal{R}_i$ 的一组可行路径，以及所有路径 $r \in \mathcal{R}_i$ 的相应路径选择概率 $\mathbb{P}_{\text{Path}}(r)$，即：

$$\mathcal{R}_i,\ \mathbb{P}_{\text{Path}}(r)_{r\in\mathcal{R}_i} = TP(o_i, d_i, t_i, m_i) \tag{1}$$

例如，我们可以将 $TP(\cdot)$ 定义为谷歌地图API与 C-logit 模型的复合函数：

$$TP(o_i, d_i, t_i, m_i) = \texttt{C-Logit} \circ \texttt{Google Map API}(o_i, d_i, t_i, m_i) \tag{2}$$

具体而言，谷歌地图API会返回所有 $r \in \mathcal{R}_i$ 的路径集 $\mathcal{R}_i$ 及路径属性 $\boldsymbol{X}_r$（例如，出行时间、出行成本等），即：

$$\mathcal{R}_i,\ (\boldsymbol{X}_r)_{r\in\mathcal{R}_i} = \texttt{Google Map API}(o_i, d_i, t_i, m_i) \tag{3}$$

C-Logit 模型输出每条路径的路径选择概率和标准误。c-Logit 模型是多项Logit（MNL）模型的一种扩展，用于修正路径重叠造成的路径间相关性（Cascetta 等人，1996年）。其核心思想是定义一个称为路径 r 的“共性因子”（即 CF_r），用于衡量路径 r 与同一起讫点的其他路径之间的相似程度。基于 C-logit 模型，选择路径 r 的概率可计算如下

$$\mathbb{P}_{\text{Path}}(r) = \texttt{C-Logit}(\boldsymbol{X}_r) = \frac{\exp[\boldsymbol{\beta}^T \cdot (\boldsymbol{X}_r, CF_r)]}{\sum_{r'\in\mathcal{R}_i} \exp[\boldsymbol{\beta}^T \cdot (\boldsymbol{X}_{r'}, CF_{r'})]} \tag{4}$$

其中 β 是段落　　待估计的参数向量。关于 CF_r 的公式可参见Cascetta 等人（1996年）　　.

请注意，　TP（·）可以比谷歌地图API加 C-logit 模型定义得更广泛。其他例子包括在多模式网络中使用 k最短路径，并结合任何路径选择行为模型（Moetal.，2022）。

3.3. 行程路段的感染建模

3.3.1. 行程路段的定义

路径 $r \in \mathcal{R}_i$ 通常包含多个行程路段，例如从家步行到公交车站、乘坐公交车，以及从公交车站步行到办公室。每个行程路段都可能发生感染。在本研究中，我们将路径上的一个路段 s 定义为沿路径以相同出行方式连续出行的部分。设路径 r 的所有路段集合为 $\mathcal{S}_r$。值得注意的是，如果一趟公共交通行程包含一次或多次换乘，我们会根据换乘将其拆分为多个路段，因为乘客需要先下车，再乘坐另一辆车辆，这在感染建模中相当于切换到一种“新的出行方式”。为方便建模，我们还忽略较短的步行路段（少于3分钟或少于1公里，例如换乘步行）。

我们提供三个示例来说明路段的定义（图 1）。第一个示例包括三个路段：从家步行到地铁站、乘坐地铁，以及从地铁站步行到办公室。第二个示例是仅包含一个路段的网约车行程。第三个示例展示了包含换乘的公共交通行程，根据定义，该行程被划分为两个路段。

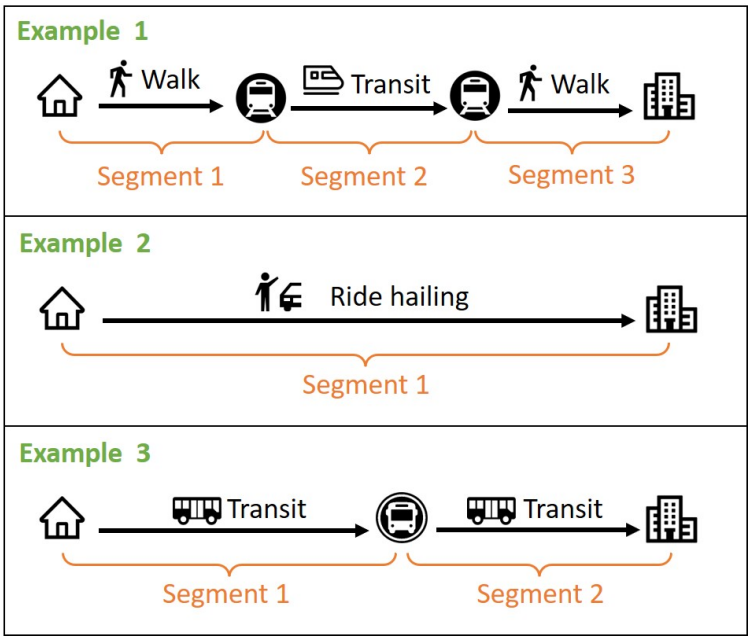


图 1：路段定义示意图

3.3.2. 每个路段的感染风险

我们考虑两种不同的感染途径：1）与具有传染性的人员密切接触而感染；2）接触受污染表面而感染。

密切接触感染：考虑一段行程路段 $s \in \mathcal{S}_r$。我们将 $\mathcal{P}_s$ 定义为曾处于个体 i 六英尺传染范围内的人员集合。对于某人 $p \in \mathcal{P}_s$，设其互动持续时间为

i 与 p 之间为 $d_{i,\ p}$。如果 p 具有传染性，i 就有可能被感染。根据互动发生在室内（例如公交车上）还是室外（例如步行经过时），我们采用两种不同的物理模型。

在室内环境中，i 被 p 感染的概率可通过著名的 Wells-Riley 模型（Riley 等人，1978）计算：

$$\mathbb{P}_{\text{Indoor}}(p \to i \mid p \text{ infectious}) = 1 - \exp\left(\frac{-b \cdot q \cdot d_{i,p}}{Q^{\text{Indoor}}}\right) \tag{5}$$

其中，b 为每人的呼吸率（m^3 /小时）；q 为量子生成率（/小时）；Q 为房间内洁净空气的通风率（m^3 /小时）。

在室外环境中，Rowe 等人（2021）提出了一种空气域模型，并推导出类似的感染概率公式：

$$\mathbb{P}_{\text{Outdoor}}(p \to i \mid p \text{ infectious}) = 1 - \exp\left(\frac{-b \cdot q \cdot d_{i,p}}{Q^{\text{Outdoor}}}\right) \tag{6}$$

其中，$Q^{\text{Outdoor}} = L \cdot H \cdot V_{\infty}$ 为洁净空气的室外通风率。H 和 L 分别为假设室外建模空间的高度和长度（与风向垂直）。V_{∞} 为风速（m/h）。建议 H 和 L 的取值分别约为 $5m$ 和 $50m$。

因此，给定接触者集合后，i 在路段 s 期间因密切接触而产生的总传染概率为

$$\mathbb{P}^{(s)}_{\text{Cont}}(i \text{ infected}) = 1 - \prod_{p \in \mathcal{P}_s} \left[\left(1 - \mathbb{P}_{\text{In/Outdoor}}(p \to i \mid p \text{ infectious})\right) \cdot \mathbb{P}(p \text{ infectious}) + \mathbb{P}(p \text{ not infectious})\right] \tag{7}$$

其中，等式7 成立，是因为被 $p \in \mathcal{P}_s$ 至少一人感染的概率等于 1 减去未被任何人感染的概率。

通过接触表面感染：Wilson 等人（2021）根据病毒生物负荷估算了单次手部与污染物接触的感染概率。其估算已考虑转移效率、用于接触表面和面部的手部比例，以及手部和可供接触的污染物表面积等方面的不确定性。将该概率定义为 $\mathbb{P}_{\text{Touch}}$（$i$ 感染者 $\mid V$），其中 V表示此次接触的病毒生物负荷。为简化起见，假设在一个行程路段内病毒生物负荷保持不变，其值为 V_s。我们还假设，路段 s （定义为 N_s）内的接触次数与 s （定义为 T_s）中的出行持续时间成正比，比例因子为 γ （即 $N_s = \gamma_s \cdot T_s$）。因此，个体 i因接触表面而产生的总感染概率为：

$$\mathbb{P}^{(s)}_{\text{Surf}}(i \text{ infected}) = 1 - \left(1 - \mathbb{P}_{\text{Touch}}(i \text{ infected} \mid V_s)\right)^{N_s} \tag{8}$$

.....的经验值 V_s 可从Harvey 等人(2020收集的日常活动环境中的病毒生物负荷数据中获得。由于单次接触的感染风险较低，式8 可用一阶泰勒级数近似：

$$\mathbb{P}^{(s)}_{\text{Surf}}(i \text{ infected}) \approx \gamma_s \cdot T_s \cdot \mathbb{P}_{\text{Touch}}(i \text{ infected} \mid V_s) \tag{9}$$

其中，式9 在计算上比式8更高效。

3.3.3. 不同出行方式的感染风险

每个路段 $s \in \mathcal{S}_r$ 都对应一种特定的出行方式。尽管我们在第 3.3.2 节提供了通用的感染风险计算模型，但仍需针对每种出行方式明确模型参数和变量。本节假设，在某一路段的行程中，传染环境保持不变 s，且取值为 b_s、q_s、Q_s^{Indoor} 和 Q_s^{Outdoor}。

公共交通路段的感染风险：公共交通路段（公交车或轨道交通）通常包含多个站点。设公共交通路段的站点集合（不包括最后一站）为 s $\mathcal{A}_s$。我们排除最后一站，是因为个体 i 到达最后一站时会下车。设车辆从车站 $a \in \mathcal{A}_s$ 出发时，车内乘客集合（不包括个体 i）为 $\mathcal{P}_{s,\ a}$。$\mathcal{P}_{s,\ a}$ 可根据智能卡数据获得，且 $\mathcal{P}_s = \cup_{a\in\mathcal{A}_s}\mathcal{P}_{s,\ a}$。设车辆从车站 a 到下一站的行驶时间为 TT_a。

那么，当车辆从车站 a 驶向下一站时，个体 i 因在车内密切接触而感染的概率为：

$$\mathbb{P}_{\text{Cont}}^{(s,a)}(i \text{ infected}) = 1 - \prod_{p\in\mathcal{P}_{s,a}} \left[\mathbb{P}(p \text{ infectious}) \cdot \exp\left(\frac{-b_s \cdot q_s \cdot TT_a}{Q_s^{\text{Indoor}}}\right) + (1 - \mathbb{P}(p \text{ infectious}))\right] \tag{10}$$

对于任意$p \in \mathcal{P}_{s,\ a}$，我们可以利用智能卡数据获取他们的出发站。因此，基于其出发站的区域感染统计数据可以近似表示 $\mathbb{P}$（p具有传染性的）和$\mathbb{P}$（p 不具有传染性的）。公共交通路段的总感染概率为：

$$\mathbb{P}_{\text{Cont}}^{(s)}(i \text{ infected}) = 1 - \prod_{a\in\mathcal{A}_s} \left(1 - \mathbb{P}_{\text{Cont}}^{(s,a)}\right) \quad \text{when } s \text{ is a transit segment} \tag{11}$$

公式 10 和 11 的计算效率可能较低。当 $\mathbb{P}$（p具有传染性的）较小时，我们可以忽略所有包含 $\left(\mathbb{P}\text{（}p\text{ 具有传染性的）}\right)^2$的二阶乘积项，采用以下近似：

$$\begin{aligned}
&\prod_{p\in\mathcal{P}_{s,a}} \left[\mathbb{P}(p \text{ infectious}) \cdot \exp\left(\frac{-b_s \cdot q_s \cdot TT_a}{Q_s^{\text{Indoor}}}\right) + (1 - \mathbb{P}(p \text{ infectious}))\right] \\
&= \prod_{p\in\mathcal{P}_{s,a}} \left[1 - \mathbb{P}(p \text{ infectious}) \cdot \left(1 - \exp\left(\frac{-b_s \cdot q_s \cdot TT_a}{Q_s^{\text{Indoor}}}\right)\right)\right] \\
&\approx 1 - \sum_{p\in P_{s,a}} \mathbb{P}(p \text{ infectious}) \cdot \left(1 - \exp\left(\frac{-b_s \cdot q_s \cdot TT_a}{Q_s^{\text{Indoor}}}\right)\right)
\end{aligned} \tag{12}$$

那么我们有

$$\mathbb{P}_{\text{Cont}}^{(s,a)}(i \text{ infected}) \approx \sum_{p\in\mathcal{P}_{s,a}} \mathbb{P}(p \text{ infectious}) \cdot \left(1 - \exp\left(\frac{-b_s \cdot q_s \cdot TT_a}{Q_s^{\text{Indoor}}}\right)\right) \tag{13}$$

这就是 被$\mathcal{P}_{s,a}$中任意一人传染的概率之和。同理，如果 $\mathbb{P}^{(s,a)}$ Cont i s

较小时，我们可以近似表示 $\mathbb{P}^{(s)}$ Cont：

$$\mathbb{P}^{(s)}_{\text{Cont}}(i \text{ infected}) \approx \sum_{a \in \mathcal{A}_s} \sum_{p \in \mathcal{P}_{s,a}} \mathbb{P}(p \text{ infectious}) \cdot \left(1 - \exp\left(\frac{-b_s \cdot q_s \cdot TT_a}{Q_s^{\text{Indoor}}}\right)\right)$$
$$\text{when } s \text{ is a transit segment} \tag{14}$$

其中 E q秒。13和1 4由于将乘积替换为求和，计算效率更高 .

对于表面接触感染，我们只需指定 V_s （病毒生物负荷）， γ_s （接触率），以及 T_s （出行时间），即可使用公式9。值得注意的是， V_s 会因需求水平不同而随一天中的时段和公共交通线路而变化。一般来说，需求较高的时段和线路应具有较高的 V_s。

步行/骑行路段的感染风险：对于步行或骑行路段，我们假设不存在表面接触感染风险，因为通勤者在通勤过程中无需接触公共接触表面：

$$\mathbb{P}^{(s)}_{\text{Surf}}(i \text{ infected}) = 0 \quad \text{when } s \text{ is a bike/walk segment} \tag{15}$$

对于近距离接触感染，我们可以根据行人密度数据集近似估算 $\mathcal{P}_s$ 。将一次接触的平均接触时间记为 $d_{\text{W/B}}$。采用相同的近似方法，我们可以计算感染概率：

$$\mathbb{P}^{(s)}_{\text{Cont}}(i \text{ infected}) \approx |\mathcal{P}_s| \cdot \mathbb{P}(p \text{ infectious}) \cdot \left(1 - \exp\left(\frac{-b_s \cdot q_s \cdot d_{\text{W/B}}}{Q_s^{\text{Outdoor}}}\right)\right)$$
$$\text{when } s \text{ is a bike/walk segment} \tag{16}$$

其中， $\mathbb{P}$（p具有传染性的）可根据步行/骑行路线，利用区域感染统计数据获得；$|\mathcal{P}_s|$ 为该路段的平均接触次数。

网约车路段的感染风险：通勤时乘坐网约车， $\mathcal{P}_s$ 包括司机和可能的拼车乘客。 $\mathcal{P}_s$ 可从网约车开源数据中获得，以估算车辆内的平均人数。司机与每位 i 的接触时间$d_{i,\ p}$ 为 T_s （路段总出行时间）。而$d_{i,}$ 每位 i 与拼车乘客的$_p$ 将通过拼车数据进行估算。采用相同的一阶近似，我们可以计算密切接触导致的感染概率：

$$\mathbb{P}^{(s)}_{\text{Cont}}(i \text{ infected}) \approx \sum_{s \in \mathcal{P}_s} \mathbb{P}(p \text{ infectious}) \cdot \left(1 - \exp\left(\frac{-b_s \cdot q_s \cdot d_{i,p}}{Q_s^{\text{Indoor}}}\right)\right)$$
$$\text{when } s \text{ is a ride-hailing segment} \tag{17}$$

就接触表面的风险而言，之前的乘客可能具有传染性，因此座椅可能受到污染。我们还可以指定 V_s 和 γ_s ，然后使用公式9 计算感染风险。不过，值得注意的是 V_s 对于网约车而言，应远低于公共交通。

驾车路段的感染风险：驾车是一种私人通勤方式，出行过程中通常不会发生密切接触或接触具有传染性的表面。因此，我们简单假设，当 s 为驾车路段时， $\mathbb{P}^{(s)}$ Surf($i$infected) $=\mathbb{P}^{(s)}$ Cont (i infected) $= 0$ 。

3.4. 感染风险整合

第3.3 节给出了特定路段感染风险的计算公式。由于一条通勤路径 r 由多个路段组成，使用路径 $r \in \mathcal{R}_i$ 时的总感染概率为：

$$\begin{aligned}\mathbb{P}(i \text{ infected} \mid r) &= 1 - \prod_{s \in \mathcal{S}_r} \mathbb{P}(i \text{ not infected in segment } s) \\ &= 1 - \prod_{s \in \mathcal{S}_r} \left(1 - \mathbb{P}^{(s)}_{\text{Surf}}(i \text{ infected})\right) \cdot \left(1 - \mathbb{P}^{(s)}_{\text{Cont}}(i \text{ infected})\right)\end{aligned} \tag{18}$$

同样，如果每个路段的传染概率都很小，则有：

$$\mathbb{P}(i \text{ infected} \mid r) \approx \sum_{s \in \mathcal{S}_r} \left(\mathbb{P}^{(s)}_{\text{Cont}}(i \text{ infected}) + \mathbb{P}^{(s)}_{\text{Surf}}(i \text{ infected})\right) \tag{19}$$

结合第 3.2 节中的路径选择概率，可得个体 i 通勤期间的总传染概率：

$$\mathbb{P}(i \text{ infected} \mid o_i, d_i, t_i, m_i) = \sum_{r \in \mathcal{R}_i} \mathbb{P}_{\text{Path}}(r) \cdot \mathbb{P}(i \text{ infected} \mid r) \tag{20}$$

3.5. 不确定性量化

尽管公式 20 给出了给定行程的最终感染概率，但参数不确定性可能导致估算出现误差。本节将 $\mathbb{P}(i \text{ 感染者} \mid o_i,\ d_i,\ t_i,\ m_i)$ 视为随机变量，并旨在计算其标准误：$\sqrt{\text{Var}[\mathbb{P}(i \text{ 感染者} \mid o_i,\ d_i,\ t_i,\ m_i)]}$。

在本研究中，我们关注感染计算所导致的不确定性。因此，假设 $\mathbb{P}_{\text{Path}}(r)$为固定值。于是：

$$\text{Var}[\mathbb{P}(i \text{ infected} \mid o_i, d_i, t_i, m_i)] = \sum_{r \in \mathcal{R}_i} \left(\mathbb{P}_{\text{Path}}(r)\right)^2 \cdot \text{Var}[\mathbb{P}(i \text{ infected} \mid r)] \tag{21}$$

根据公式 19，由于各路段之间具有独立性，我们可以进一步将方差分解到不同路段：

$$\text{Var}[\mathbb{P}(i \text{ infected} \mid r)] = \sum_{s \in \mathcal{S}_r} \left(\text{Var}\left[\mathbb{P}^{(s)}_{\text{Cont}}(i \text{ infected})\right] + \text{Var}\left[\mathbb{P}^{(s)}_{\text{Surf}}(i \text{ infected})\right]\right) \tag{22}$$

因此，我们只需针对不同类型（即出行方式）的路段，分别指定 $\text{Var}\left[\mathbb{P}^{(s)}\ \text{Cont}(i\text{infected})\right]$ 和 $\text{Var}\left[\mathbb{P}^{(s)}\ \text{Surf}(i\text{infected})\right]$。

公共交通路段：根据公式 14，公共交通中因密切接触导致的感染概率，就是对不同车站和接触情况的概率求和。在现实中，$\mathcal{P}_{s,\ a}$ 会因公共交通系统中的需求变化而具有不确定性。然而，通常很难对一组数据的不确定性进行建模。因此，我们将乘客在 $\mathcal{P}_{s,\ a}$ 具有传染性的平均概率定义为 $\bar{\mu}_{s,\ a}$。那么，公式14 可以改写为：

$$\mathbb{P}^{(s)}_{\text{Cont}}(i \text{ infected}) \approx \sum_{a \in \mathcal{A}_s} |\mathcal{P}_{s,a}| \cdot \bar{\mu}_{s,a} \cdot \left(1 - \exp\left(\frac{-b_s \cdot q_s \cdot TT_a}{Q^{\text{Indoor}}_s}\right)\right) \quad \text{when } s \text{ is a transit segment} \tag{23}$$

其中，$|\mathcal{P}_{s,\ a}|$ 是车辆从车站a出发时车内乘客人数（不包括 i）。方差可表示为：

$$\text{Var}[\mathbb{P}^{(s)}_{\text{Cont}}(i \text{ infected})] = \sum_{a \in \mathcal{A}_s} \text{Var}\left[|\mathcal{P}_{s,a}| \cdot \bar{\mu}_{s,a} \cdot \left(1 - \exp\left(\frac{-b_s \cdot q_s \cdot TT_a}{Q_s^{\text{Indoor}}}\right)\right)\right] \quad \text{when } s \text{ is a transit segment} \tag{24}$$

推导若干随机变量乘积（或若干变量的指数）的方差闭式表达式十分困难。在某些简单情况下，可以使用詹森不等式得到方差的下界（或上界）。然而，对于一般函数 $f(\boldsymbol{Z})$ （该函数可能既非凸也非凹），其中 $\boldsymbol{Z}$ 是随机变量向量。通常很难得到 $\text{Var}[f(\boldsymbol{Z})]$ 的解析形式。因此，我们提出一种基于自助法的算法来估计经验方差（算法1）。该算法的输入为 $f(\boldsymbol{Z})$、 $\boldsymbol{Z}$的分布（即 $\mathbb{P}(\boldsymbol{Z})$）以及最大抽样次数 M。其思路是对 $\boldsymbol{Z} \sim \mathbb{P}(\boldsymbol{Z})$进行抽样，并利用样本估计方差。

算法1基于自助抽样的经验方差估计算法

1: **函数** 自助法方差($f(\boldsymbol{Z})$, $\mathbb{P}(\boldsymbol{Z})$, M)
2: **对于** $m = 1, 2, ..., M$ **do**
3: 抽取样本 $\boldsymbol{Z}^{(m)} \sim \mathbb{P}(\boldsymbol{Z})$
4: $\bar{f}(\boldsymbol{Z}) = \frac{1}{M}\sum_{m=1}^{M} f(\boldsymbol{Z}^{(m)})$ ▷ 估计经验均值
5: 方差$[f(\boldsymbol{Z})] = \frac{1}{M}\sum_{m=1}^{M}\left(f(\boldsymbol{Z}^{(m)}) - \bar{f}(\boldsymbol{Z})\right)^2$ ▷ 估计经验方差
6: **返回** 方差$[f(\boldsymbol{Z})]$

值得注意的是，算法1 （第 3 行）中的抽样过程可以独立、联合或顺序进行，具体取决于 $\boldsymbol{Z}$ 中的元素是独立、依赖还是条件独立。

根据算法 1，我们可以通过$|\mathcal{P}_{s、\ a}|$、 $\bar{\mu}_{s、\ a}$、 b_s、 q_s、 TT_a 和 Q_s 的分布来估计 $\text{Var}\left[|\mathcal{P}_{s,a}| \cdot \bar{\mu}_{s,a} \cdot \left(1 - \exp\left(\frac{-b_s \cdot q_s \cdot TT_a}{Q_s^{\text{Indoor}}}\right)\right)\right]$。在本研究中，我们假设根据实证数据中的观测值，$|\mathcal{P}_{s、\ a}|$（车辆载客量）和 TT_a （出行时间）均服从正态分布。其分布参数可根据智能卡数据和自动车辆定位数据进行估计。 b_s、 q_s 和 Q_s 服从均匀分布，其分布依据文献列于表1 。对于 $\bar{\mu}_{s、\ a}$，由于其由抽样得到的不同$\mathcal{P}_{s、\ a}$生成，因此难以对其分布进行参数化。因此，我们将 $\bar{\mu}_{s、\ a}$ 的所有样本保留为经验分布。于是，$\bar{\mu}_{s、\ a} \sim \mathbb{P}$（$\bar{\mu}_{s、\ a}$）的每个样本本质上都是从其样本池中进行的一次自助抽样。

关于表面接触感染，根据公式 9，我们通过设定 $f(\boldsymbol{Z}) = \gamma_s \cdot T_s \cdot \mathbb{P}_{\text{Touch}}(i \text{ 感染者} \mid V_s)$ 和 $\boldsymbol{Z} = (\gamma_s, T_s, V_s, \mathbb{P}_{\text{Touch}}(i \text{ 感染者} \mid V_s)$，使用算法1估算方差。 $\mathbb{P}_{\text{Touch}}(i$感染者 $\mid V_s)$的分布可根据 Wilson 等人（2021）的数据获得。 V_s 的分布可根据 Harvey 等人（2020）的结果获得。 T_s 的分布可根据 AVL数据获得。假设 γ_s 服从均匀分布，其分布见表1。

步行/骑行路段：步行和骑行路段中接触传播感染风险的方差，

根据公式 16，可表示为：

$$\text{Var}[\mathbb{P}^{(s)}_{\text{Cont}}(i \text{ infected})] = \text{Var}\left[|\mathcal{P}_s| \cdot \bar{\mu}_s \cdot \left(1 - \exp\left(\frac{-b_s \cdot q_s \cdot d_{\text{W/B}}}{Q_s^{\text{Outdoor}}}\right)\right)\right] \quad \text{when } s \text{ is a bike/walk segment} \tag{25}$$

其中，$\bar{\mu}_s$ 是 $\mathcal{P}_s$中接触的平均传染概率。其形式与公式24类似，因此可使用算法1计算。$|\mathcal{P}_s|$ 的分布可根据行人流量数据获得。$\bar{\mu}_s$ 的分布可根据区域统计数据获得。我们假设所有感染相关参数（包括$d_{\text{W/B}}$）均服从均匀分布，其分布如表1所示。由于我们假设步行和骑行路段不存在与接触表面相关的感染风险，因此无需考虑其方差。

网约车路段：对于网约车路段，由于 $\mathcal{P}_s$ 相对较小（每辆车最多三名乘客），我们假设 $|\mathcal{P}_s|$ 在 $\{0, 1, 2, 3\}$ 中服从多项分布（即最多有另外两名乘客和一名司机）。具体分布可根据拼车数据获得。为了计算网约车路段的 $\text{Var}[\mathbb{P}^{(s)}_{\text{Cont}}(i\text{感染者})]$，抽样过程如下（与算法1略有不同）：

- 步骤 1：从多项分布中抽取样本 $|\mathcal{P}_s|$ 。
- 步骤 2：根据 $|\mathcal{P}_s|$ 的值，抽取数量相同的乘客和司机，以生成$|\mathcal{P}_s|$；对于每位乘客 $p \in \mathcal{P}_s$，我们还会抽取$d_{i,\ p}$（拼车行程时间）。
- 步骤 3：从表 1 中定义的分布抽取其他感染相关参数。根据公式 17 计算密切接触的感染概率。

经过 M 次抽样后，我们可以估算样本方差，以近似表示 $\text{Var}[\mathbb{P}^{(s)}_{\text{Cont}}(i \text{ 感染})]$，与算法 1 类似。基于接触表面的感染概率的方差计算方式与公共交通路段相同，只需替换 T_s、V_s 和 γ_s 的分布。

对于驾车路段，由于我们假设不存在感染风险，因此不计算方差。

4. 案例研究

本案例研究基于麻省理工学院现有的数据。该模型针对大波士顿地区所有麻省理工学院学生和教职工的通勤情况进行实施。

4.1. 数据来源和参数设置

本研究使用来自多个来源的数据来估算感染风险并设置参数。第一组数据是麻省理工学院员工通勤调查。该调查收集每位个体的（o_i、d_i、t_i、m_i）$i \in \mathcal{I}$。根据这些信息，我们为每位个体 i生成路径集合 $\mathcal{R}_i$ 及相应的路径属性（即步行时间、等待时间、乘车时间、出行成本），这些数据基于谷歌地图 API 获取。为简化起见，我们仅考虑谷歌地图 API 针对出行方式 m_i 推荐的首条路径（即忽略路径选择估计）。这一假设是合理的，因为在大多数情况下，如果是 $m_i =$ 公共交通，只有一条公共交通路线可用。对于驾车、骑自行车或网约车，个体通常会遵循导航系统，因此选择谷歌地图 API 提供的第一个选项。

我们用于计算公共交通路段感染风险的另一个数据来源，是来自马萨诸塞湾交通局（MBTA）的智能卡数据。我们使用过去两周内每天同一时段（一个小时的间隔）的历史智能卡数据，计算乘客人数的均值和方差，适用于$\mathcal{P}_{s,\ a}$。具体而言，智能卡数据提供刷卡进站的时间和地点。我们采用了Sánchez-Martínez (2017)提出的目的地估计模型，以获取每段行程的目的地。然后，使用网络加载模型（Mo 等人，2020）获取各时间间隔内的车辆载客量。根据区域统计数据和乘客的出发地，计算车辆内乘客的平均传染概率 $\bar{\mu}_{s,\ a}$。站点之间的出行时间（即 TT_a）取自自动车辆定位（AVL）数据。

关于骑行和步行路段，本研究没有可用的开源行人和骑行者密度数据。因此，我们结合波士顿的人口数据（Cubit 马萨诸塞州人口统计数据，2020年）、马萨诸塞州出行调查（MTS）（波士顿地区大都会规划组织，2011年）以及全国住户出行调查（NHTS）数据（美国联邦公路管理局，2017），生成合成行人和骑行者密度数据。合成行人密度数据包含每个时间间隔（本研究中为每分钟）特定街道上骑行者和行人数目的均值与方差。详细的数据生成过程见附录 A。基本思路是利用现有数据集生成大量骑行和步行轨迹样本，并将这些样本汇总到街道-分钟级。因此，给定个人的骑行/步行路段 s 的 i，我们可以根据其轨迹，利用上述生成的骑行者和行人密度数据，对骑行/步行接触次数进行抽样。重复这一过程多次，即可得到 $|\mathcal{P}_s|$的分布。 $\bar{\mu}_s$ 的分布根据社区感染率计算（Fisher，2020）。接触时间（$d_{\text{W/B}}$）假定服从均匀分布，步行行程的接触时间为 $\mathcal{U}(4,\ 6)$秒，骑行行程的接触时间为 $\mathcal{U}(2,\ 4)$秒。

对于网约车路段，我们只需要 $|\mathcal{P}_s|$（乘客数量）和$d_{i,\ p}$（拼车行程时长）的分布。然而，波士顿没有公开的网约车数据。本研究使用芝加哥的开源交通网络公司（TNC）数据（芝加哥数据门户，2018）计算 $|\mathcal{P}_s|$ 和$d_{i,\ p}$ 的分布，作为波士顿相应分布的近似值。

通过谷歌地图API，我们可以获取T_s的值。然而， T_s 的分布尚不明确。理想情况下，可以通过GPS数据获得 T_s 的分布。鉴于数据有限，我们假设 T_s 服从正态分布，并将通过谷歌地图API获取的值作为均值。根据实证研究（Li等人，2013），标准差假设为 $T_s \times 30\%$。值得注意的是，对于公共交通路段，我们不使用谷歌地图数据，而是利用AVL数据获取站点间出行时间。因此，我们将路段出行时间计算为$T_s = \sum_{a \in \mathcal{A}_s} TT_a$。 T_s 的均值和方差，分别是 TT_a的均值和方差之和，假设各路段之间相互独立。

感染相关参数汇总于表1：

4.2. 数据描述与统计

麻省理工学院员工通勤调查包含974份个人问卷答复，涵盖以下信息：$(o_i,\ d_i,\ t_i,\ m_i)$。请注意，此处仅考虑前往麻省理工学院的通勤出行。其出发时间、出行方式和职业类别分布如图2所示。

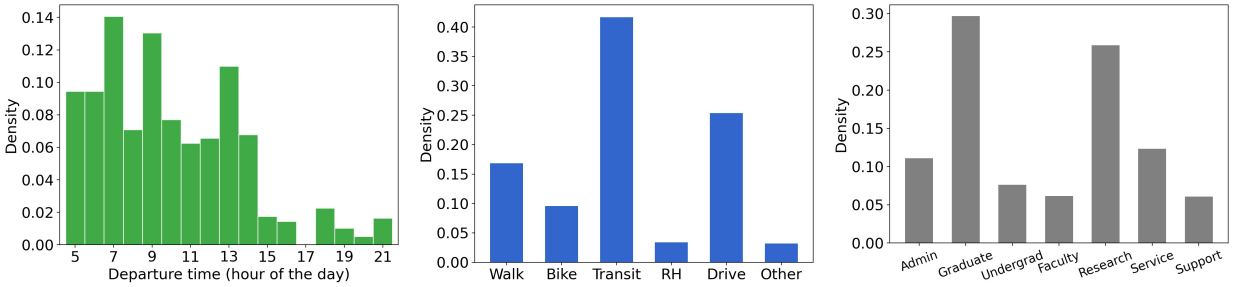
由于仅考虑前往麻省理工学院的出行，大多数人的出发时间都在上午（图2a）。傍晚也有一些人前往校园，他们可能是住在附近的学生，也可能是上夜班的教职工。就出行方式而言（图2b），超过40%的麻省理工学院教职工和学生选择公共交通作为通勤方式。驾车是第二受欢迎的模式。在所有填写

表1：感染相关参数

Mode	参数与分布							
	b_s (m^3/h)	q_s $(/h)$	V_∞ (m/s)	L (m)	H (m)	Q_s^{Indoor} (m^3/h)	γ_s $(/h)$	V_s (gc/cm^2)
公共交通（火车）	$\mathcal{U}$(0。65，0。79)	$\mathcal{U}$(1，31)	N.A.	N.A.	N.A.	$\mathcal{U}$(800，1100)	$\mathcal{U}$(3，5)	$\mathcal{U}$(30，100)
公共交通（公交车）	$\mathcal{U}(0.65, 0.79)$	$\mathcal{U}(1, 31)$	N.A.	N.A.	N.A.	$\mathcal{U}$(300，500)	$\mathcal{U}$(3，5)	$\mathcal{U}(30, 100)$
Bike	$\mathcal{U}(1.4, 1.8)$	$\mathcal{U}$(2，100)	$\mathcal{U}$(2，4)	$\mathcal{U}(30, 60)$	$\mathcal{U}(2.5, 5)$	N.A.	N.A.	N.A.
Walk	$\mathcal{U}(1.2, 1.6)$	$\mathcal{U}(2, 100)$	$\mathcal{U}$(1，2)	$\mathcal{U}(30, 60)$	$\mathcal{U}(2.5, 5)$	N.A.	N.A.	N.A.
网约车	$\mathcal{U}$(0。65，0。79)	$\mathcal{U}$(1，31)	N.A.	N.A.	N.A.	$\mathcal{U}$(90，120)	$\mathcal{U}$(1，3)	$\mathcal{U}$(2，50)

N.A.：不适用 公共交通感染的参数取自Zhou 和 Koutsopoulos (2021）。量子生成率的参数取自Buonanno 等人 (2020b）和Buonanno等人 (2020a）。户外感染模型的参数取自Rowe 等人 (2021）。网约车通风率基于Ott 等人(2008）。病毒生物负荷数据改编自Harvey 等人(2020）（假设 10% 的病毒具有传染性）。

问卷的个体中，约30%是研究生。受访者还包括教职员工和教职工（例如行政、服务、支持和研究人员）。



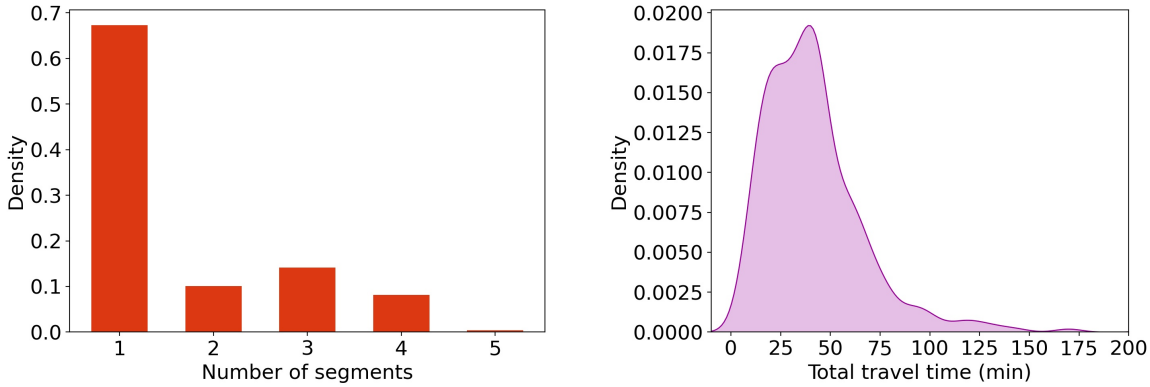
(a) 出发时间分布 (b) 出行方式分布 (c) 人员类型分布

图2：麻省理工学院员工通勤调查的描述性统计（RH 表示网约车）

图3 展示了通过 Google地图API 收集的出行信息分布。大多数通勤出行只有一个路段（图3a）。路段数量最多为五个。所有出行的出行时间近似呈长尾对数正态分布（图 3b）。大多数出行的出行时间在 1 小时以内。

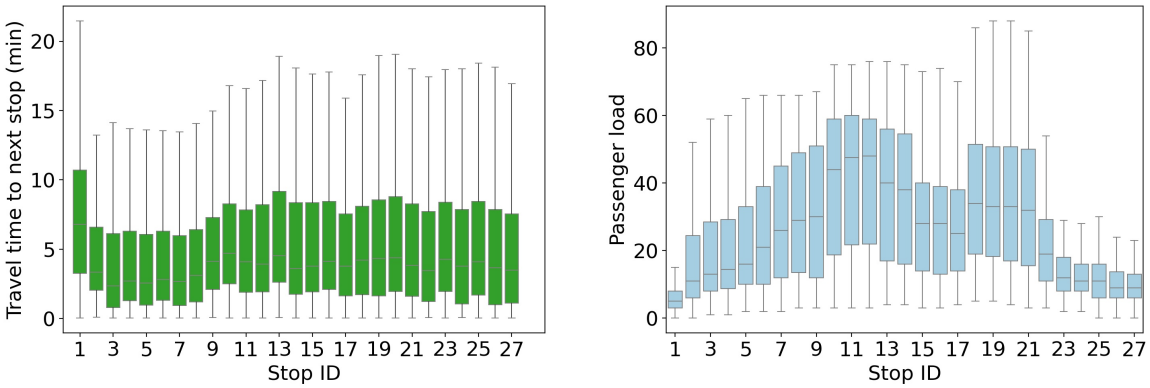
对于公共交通出行，图 4展示了波士顿 1号公交线路相邻车站之间出行时间的箱线图，以及相应的乘客载客量。1号线路是沿马萨诸塞大道运行的一条热门公交线路。前往第一个站点的出行时间相对较长，因为车辆通常从车库发车（图4a）。从乘客载客量来看（图 4b），线路中段站点车上的乘客人数相对较多。图表显示，出行时间和乘客载客量都存在不确定性，因此病毒传播建模应考虑这些不确定性。

图 5 展示了上午8:00推断出的步行和骑自行车出行次数的空间分布（详见 附录 A）。两者的空间分布相似。人口密度较高的地点出行次数更多。步行出行次数多于骑自行车出行次数。



（a）路段数量分布　　（b）出行方式分布

图3：路段数量和出行时间的分布



（a）到下一站的行程时间　　（b）乘客载客量

图4：1号公交线路的出行时间和乘客载客量箱线图（站点编号表示从北到南的站点顺序）

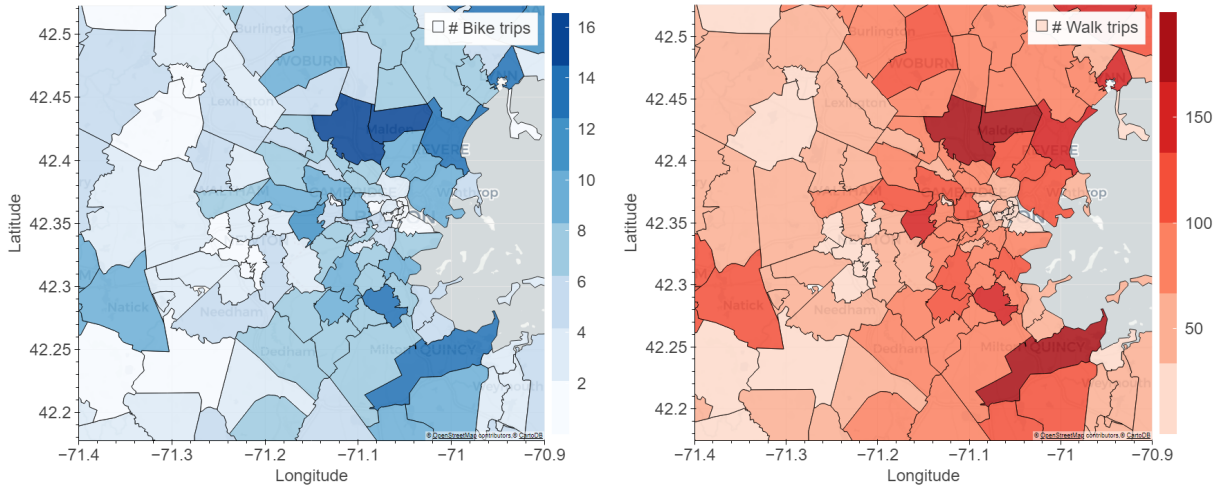
4.3. 结果

4.3.1. 个体的感染风险

排除出行方式为“其他”的个体后，还剩下943名个体。我们使用所提出的方法计算他们的感染概率和标准差。结果见图6。大多数个体的感染概率接近于零。最高感染概率约为0.8%。这表明通勤前往麻省理工学院期间感染的概率相当低。结果与此前关于公共交通感染风险的研究一致（Zhou 和 Koutsopoulos，2021）。就估计误差而言，标准差约为估计概率的50%。

由于出行方式、出发时间和出行距离不同的个体可能具有不同的感染概率，我们采用线性回归模型分析不同因素对感染风险的影响。因变量为推断的感染概率，自变量如下：

- 是否为教职工：该个人是否为麻省理工学院的教职员工（是 = 1）。



(a) 自行车出行次数　　(b) 步行出行次数

图5：上午8:00步行和自行车出行次数的空间分布

- 距离（公里）：该个人住所到麻省理工学院的欧几里得距离。
- 是否乘坐公共交通：个体的通勤方式是否为公共交通（是 = 1）。
- 是否处于早高峰：个体的出发时间是否在上午6:00至上午9:00之间（是 =1）。

线性回归的结果见表 2。“距离×是否乘坐公共交通”用于区分距离对公共交通用户和非公共交通用户的影响。我们发现，教职员工感染的可能性较低。这可能是因为他们通常驾车上学，而就感染防护而言，驾车是最安全的出行方式。居住地离学校较远且乘坐公共交通通勤的个体感染风险较高。这可能是因为他们的出行时间更长，接触更多，因此更容易暴露于病毒之中。我们还观察到，早高峰时段出发的人感染概率较高，这可能是因为早高峰时段接触人数更多。值得注意的是，除非将“距离”和“是否乘坐公共交通”结合起来（即“距离×是否乘坐公共交通”），否则二者均不显著。这表明，距离对感染风险的影响主要适用于公共交通用户。

表2：感染概率影响因素（%）

变量	系数（p 值）	变量	系数（p 值）
截距	-0.015 (0.997)	距离	0.048 (0.834)
是否乘坐公共交通	4.409 (0.545)	是否处于早高峰	16.658** (0.005)
是否为教职工	-16.881* (0.095)	距离×是否乘坐公共交通	3.404** (0.000)

样本数量：943；R^2：0.313；
**：p-值 < 0.05；*：p-值 < 0.1；
所有系数均乘以 1000。

为了更直观地展示距离的影响，图7 显示了公共交通用户和非公共交通用户的感染概率随距离变化的预测结果。总体而言，公共交通用户更容易感染，且

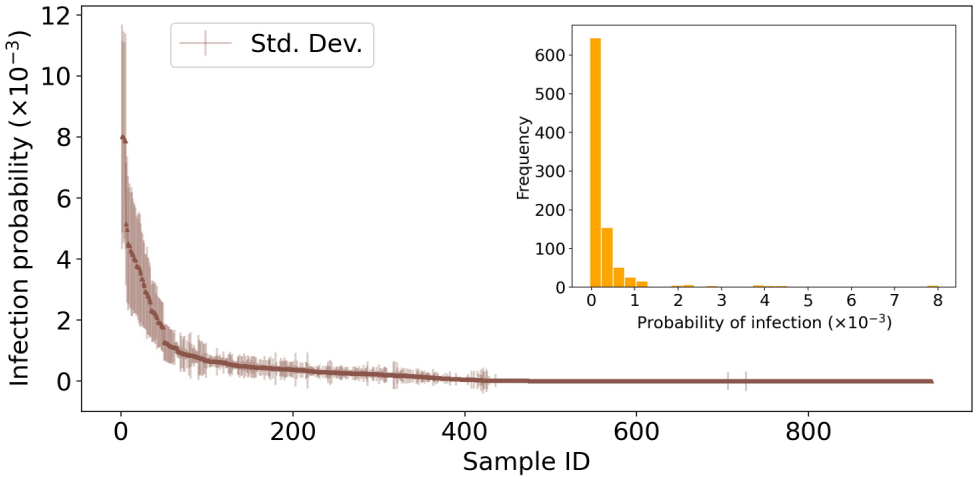


图6：推断的个体感染概率（按数值降序排列）

感染风险会随着通勤距离的增加而大幅上升。然而，距离对非公共交通用户的感染风险没有显著影响。

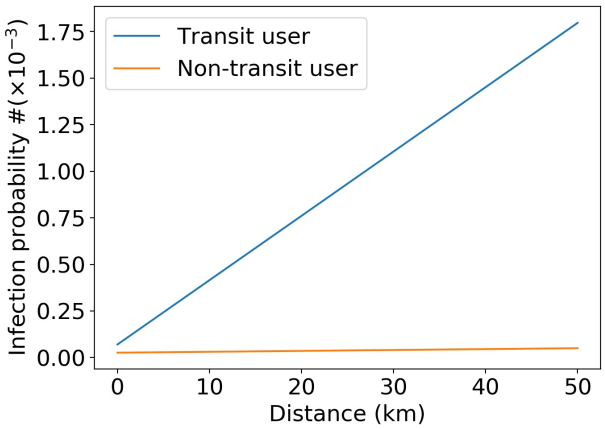


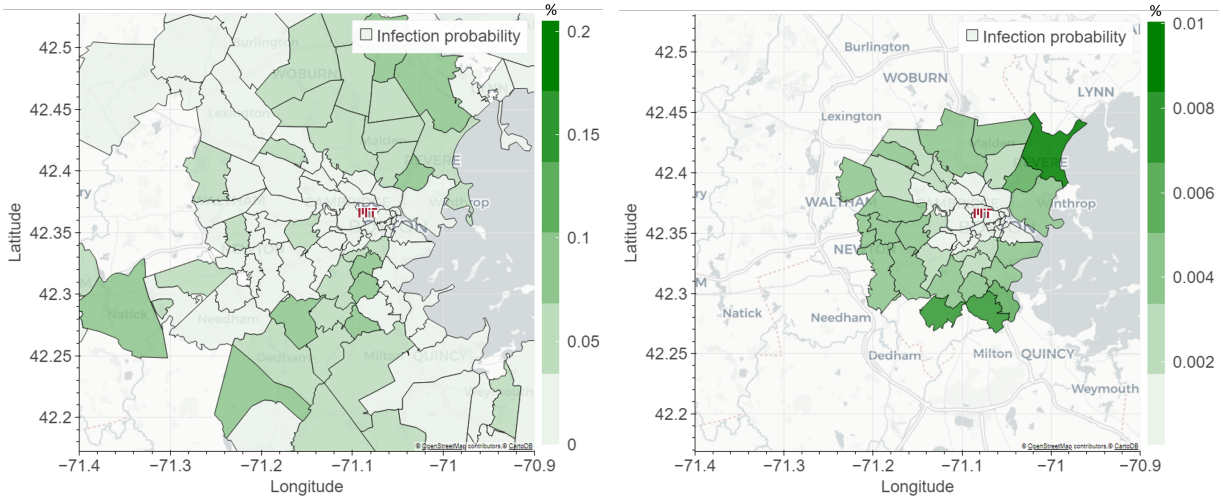
图7：预测感染概率与距离的关系

4.3.2. 感染风险的时空分布

除了计算实际调查受访者的感染风险外，还可以通过生成合成观测，利用所提出的方法评估感染风险的时空分布。

对于空间分布，我们生成家庭住址位于不同社区的合成观测。我们考虑两种出行方式：公共交通和步行。请注意，对于出行方式为步行的虚拟样本，我们还考虑距离麻省理工学院10公里以内的家庭住址。所有虚拟样本的出发时间均设为上午8:00。

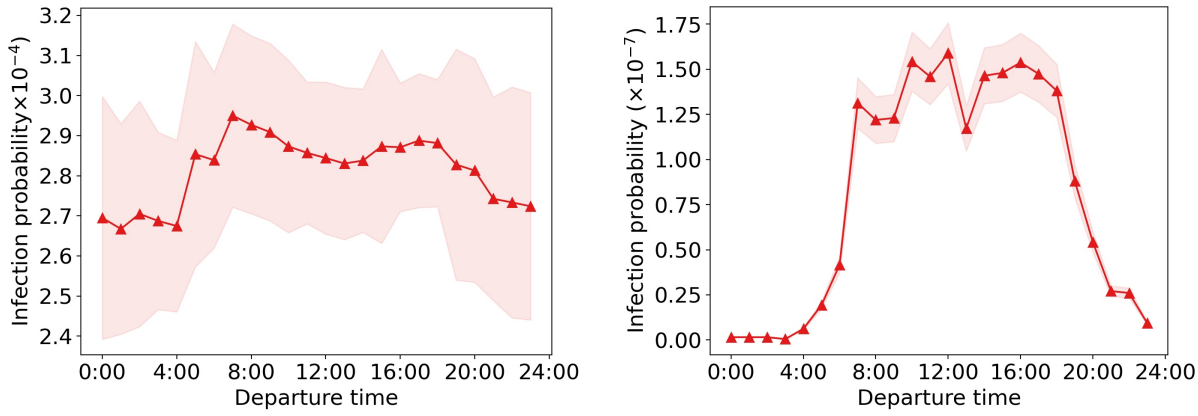
图 8a 展示了居住地址位于不同社区的公共交通出行的感染概率。总体而言，居住地点离麻省理工学院越远的人，感染风险越高。但感染风险也会受到具体公共交通线路的影响（即与距离并非完全成正比）。步行的空间分布（图8b）相似，但其感染风险远低于公共交通出行。



(a) 公共交通出行　　(b) 步行出行

图 8：感染概率的空间分布

就时间分布而言，我们生成了出发时间从 0:00 到 24:00 的合成观测，并假设这些观测的居住地点服从均匀分布，遍及所有社区。图9a 展示了公共交通出行的感染风险随出发时间的变化。阴影区域表示估计误差的 $1\ 10$。人们在早晚高峰时段出发时，感染概率较高，这合乎情理，因为交通高峰时段通常乘客载客量较大（即密切接触更多）。步行出行的感染概率呈现出相似的模式（图 9b）。感染风险最高的时段是白天（8:00 至 18:00），这很可能是因为行人数量较多。



(a) 公共交通出行　　(b) 步行出行

图 9：感染概率的时间分布

5. 结论

本文提出了一种概率框架，用于估算通勤期间在不同出行方式（包括公共交通、拼车、骑自行车和步行）下的感染风险。该模型能够评估感染概率及估计误差（即不确定性量化）。研究利用不同来源的数据（如智能卡数据、出行调查和人口数据）估算通勤个体与传染环境的交互情况。本文以麻省理工学院社区的相关数据为案例研究，评估员工和学生通勤中的感染风险。结果表明，大多数个体的感染概率都很低，最高感染概率约为0.8%。出行距离较长、乘坐公共交通以及在高峰时段出行的个体，更容易受到感染。

该模型具有多种实际应用。1）在大流行病期间或之后，企业、学校或社区可利用该模型支持有关返岗办公的决策。它们可以收集员工的通勤信息，并使用该模型评估通勤风险，以便更好地制定规划。例如，可为感染风险较高的员工安排与低风险员工不同的座位。2）该模型的另一种应用方式，是将个体层面的感染风险添加到行程规划工具（如谷歌地图）中。例如，在图 10中，行程规划工具不仅会根据出行时间显示推荐路线，还会显示感染风险。个体可以利用这些额外信息，更好地选择出行路径。



图10：将模型应用于行程规划的示例

6. 致谢

作者谨向麻省理工学院智能探索计划表示感谢，感谢其为本研究提供支持和数据。

7. 作者贡献声明

Baichuan Mo：概念化、方法论、软件、正式分析、数据整理、初稿撰写、审阅与编辑、可视化。**Peyman Noursalehi**：概念化、软-件、数据整理。**Haris N. Koutsopoulos**：概念化、审阅与编辑、指导。**Jinhua Zhao:** 概念化、指导、项目管理、经费获取。

附录

附录 A. 行人和骑行者密度计算

计算步行和骑行出行的感染风险时，一个重要输入是行程中的近距离接触次数（即 $|\mathcal{P}_s|$的分布）。该分布可根据每个时间间隔 τ内特定街道 b 上的骑行者数量（记为$C_{b,\tau}$）和行人数量（记为$W_{b,\tau}$）的均值和方差得到。由于本研究没有可用的全球定位系统或轨迹数据，我们使用以下方法生成$C_{b,\tau}$和$W_{b,\tau}$ 的分布。

首先，我们收集波士顿大都会区各街区（邮政编码级别）的人口数据（Cubit 马萨诸塞州人口统计数据，2020年）。我们使用马萨诸塞州出行调查（MTS）（波士顿地区大都会规划组织，2011年），根据人口计算出行生成率。由于 MTS 未包含骑行和步行方式分担率，因此我们使用全国住户出行调查（NHTS）数据（美国联邦公路管理局，2017），并根据马萨诸塞州的样本获取骑行和步行出行的比例及其时间分布。结合出行生成率，我们模拟各街区在不同时间段内的骑行和步行出行数量。由于 NHTS 数据中波士顿的实际骑行和步行出行数量有限，难以获得起讫点（OD）分布，但可以获得出行距离和出发时间的分布。对于每次骑行或步行出行，我们按如下方式生成“假设”轨迹：1）首先，在街区内随机抽取一个出发地。2）然后，从预先定义的分布中抽取出行距离和出发时间，并抽取一个特定方向（均匀 $0 \sim 2\pi$）。出行距离和方向决定目的地。3）生成该行程的轨迹（即一系列街道）。根据这些轨迹，我们获得每条街道在各个时间段内的行人和骑行者数量 b （即 τ 和 $C_{b,\ \tau}$ 和 $W_{b,\ \tau}$的样本）。重复此过程多次，以获得 $\mathbb{E}[C_{b,\tau}]$，$\mathbb{E}[W_{b,\tau}]$ 和 $\mathrm{Var}[C_{b,\tau}]$、$\mathrm{Var}[W_{b,\tau}]$。

生成的步行和骑自行车行程分布如图5 所示（该图按邮政编码级别汇总）。

参考文献

Ahangari, S., Chavis, C., Jeihani, M., 2020. 新冠肺炎疫情期间的公共交通客流量分析。Medrxiv。Ando, H., Ikegami, K., Nagata, T., Tateishi, S., Eguchi, H., Tsuji, M., Matsuda, S., Fujino, Y., Ogami, A., 2021. 2020年12月日本通勤对新冠肺炎感染风险及新冠肺炎引发的焦虑的影响。《公共卫生档案》79, 1–10. 波士顿地区大都会规划组织，2011. 探索2011年马萨诸塞州出行调查：大都会规划组织出行概况。https://www.ctps.org/travel-profiles，最后访问于2022年11月4日。Buonanno, G., Morawska, L., Stabile, L., 2020a. 定量评估SARS-CoV-2感染的病毒气溶胶传播风险：前瞻性和回顾性应用。《国际环境》145, 106112. Buonanno, G., Stabile, L., Morawska, L., 2020b. 病毒气溶胶排放估算：用于感染风险评估的SARS-CoV-2量子排放率。《国际环境》141, 105794. Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A., 1996. 克服路径重叠问题的修正Logit路径选择模型：城际网络的模型设定及若干校准结果，载于：交通与交通理论。第十三届交通与交通流理论国际研讨会论文集，法国里昂，1996年7月24日至26日。Trans-

portation and Traffic Theory. 第十三届交通与交通流理论国际研讨会论文集，法国里昂，1996年7月24日至26日。Chang, H.H., Lee, B., Yang, F.A., Liou, Y.Y., 2021. 新冠肺炎是否影响台北的地铁使用？《交通地理学杂志》91, 102954. 芝加哥数据门户，2018. 交通网络服务提供商—出行。
h
t
t
p
s://
d
ata.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips/m6dm-c72p,
最后访问于2022年11月4日。Ecke, L., Magdolen, M., Chlond, B., Vortisch, P., 2022. 新冠肺炎疫情如何改变日常通勤习惯——德国出行调查的洞见。《交通政策案例研究》。Fajgelbaum, P.D., Khandelwal, A., Kim, W., Mantovani, C., Schaal, E., 2021. 通勤网络中的最优封锁。《美国经济评论：见解》3, 503–22. 美国联邦公路管理局，2017. 全国住户出行调查。https://nhts.ornl.gov/，最后访问于2022年11月4日。Fisher, A.R., 2020. 美国邮政编码层级的新冠肺炎估算。
https://github.com/adamrfisher/Estimated_COVID19_by_US_Zip_Code，最后访问于2022年11月4日。Harvey, A.P., Fuhrmeister, E.R., Cantrell, M.E., Pitol, A.K., Swarthout, J.M., Powers, J.E., Nadimpalli, M.L., Julian, T.R., Pickering, A.J., 2020. 社区环境中高频接触表面SARS-CoV-2 RNA的纵向监测。《环境科学与技术快报》8, 168–175. Jenelius, E., Cebecauer, M., 2020. 新冠肺炎对瑞典公共交通客流量的影响：车票验证记录、售票量和乘客人数分析。《交通研究：跨学科视角》8, 100242. Kondo, K., 2021. 模拟跨区域流动限制对日本新冠肺炎空间传播的影响。《科学报告》11, 1–15. Ku, D., Yeon, C., Lee, S., Lee, K., Hwang, K., Li, Y.C., Wong, S.C., 2021. 新冠肺炎疫情期间的安全公共交通出行。《科学进展》7, eabg3691. Li, R., Chai, H., Tang, J., 2013. 旅行时间估计及其可靠性的实证研究。《工程中的数学问题》2013. Cubit 马萨诸塞州人口统计数据，2020. 马萨诸塞州各邮政编码人口。
https://www.massachusetts-demographics.com/zip_codes_by_population，最后访问于2022年11月4日。Medlock, K.B., Temzelides, T., Hung, S.Y.E., 2021. 新冠肺炎与美国安全交通的价值。《科学报告》11, 1–12. Meredith-Karam, P., Kong, H., Wang, S., Zhao, J., 2021. 网约车与芝加哥公共交通之间的关系：新冠肺炎疫情前后的比较。《交通地理学杂志》97, 103219. Mitze, T., Kosfeld, R., 2022. 通勤对新冠肺炎空间传播的扩散效应。《地理系统杂志》24, 5–31. Mo, B., Feng, K., Shen, Y., Tam, C., Li, D., Yin, Y., Zhao, J., 2021. 使用时变接触网络模拟公共交通中的疫情传播。《交通研究C辑：新兴技术》122, 102893. Mo, B., Ma, Z., Koutsopoulos, H.N., Zhao, J., 2020. 城市轨道交通系统的容量约束网络性能模型。《交通研究记录》2674, 59–69. Mo, B., Von Franque, M.Y., Koutsopoulos, H.N., Attanucci, J.P., Zhao, J., 2022. 非计划长期服务中断对城市公共交通系统的影响。《IEEE智能交通系统开放期刊》3, 551–569。

Ott, W., Klepeis, N., Switzer, P., 2008. 机动车的换气率及二手烟导致的车内污染物浓度。暴露科学与环境流行病学杂志 18, 312–325. Riley, E., Murphy, G., Riley, R., 1978. 郊区小学中的麻疹空气传播。美国流行病学杂志 107, 421–432. Rowe, B.R., Canosa, A., Drouffe, J.M., Mitchell, J., 2021. 对病毒和新冠肺炎室外与室内空气传播的简易定量评估。环境研究 198, 111189. Sánchez-Martínez, G.E., 2017. 利用票务交易和车辆位置数据推断公共交通出行目的地：动态规划方法。《交通研究记录》2652, 1–7. Shinohara, N., Sakaguchi, J., Kim, H., Kagi, N., Tatsu, K., Mano, H., Iwasaki, Y., Naito, W., 2021. 通勤列车换气率调查及新冠肺炎空气传播感染风险评估。《国际环境》157, 106774. Tan, L., Ma, C., 2021. 基于逻辑回归模型的新冠肺炎疫情期间通勤者轨道交通方式选择行为。《交通运输工程学报（英文版）》8, 186–195. Wang, H., Noland, R.B., 2021. 新冠肺炎疫情期间纽约市共享单车和地铁客流量变化。《交通政策》106, 262–270. Wilbur, M., Ayman, A., Ouyang, A., Poon, V., Kabir, R., Vadali, A., Pugliese, P., Freudberg, D., Laszka, A., Dubey, A., 2020. 新冠肺炎对公共交通可达性和乘客量的影响。arXiv 预印本 arXiv:2008.02413。Wilson, A.M., Weir, M.H., Bloomfield, S.F., Scott, E.A., Reynolds, K.A., 2021. 对单一手部接触污染物情景下的新冠肺炎感染风险进行建模，并评估表面消毒可能带来的风险降低。《美国感染控制杂志》49, 846–848. Zhao, Q., Qi, Y., Wali, M.M., 2022. 一种评估乘坐公共交通新冠肺炎感染风险的方法。《国际交通科学与技术杂志》。Zhou, J., Koutsopoulos, H.N., 2021. 城市轨道交通系统中的病毒传播风险：基于微观仿真的时空特征分析。《交通研究记录》2675, 120–132.