清华大学优秀博士学位论文丛书

基于分布式鲁棒优化的应急救援系统选址模型和算法研究

刘康琳（Liu Kanglin）著

A Distributionally Robust Optimization Approach to Emergency Facility Location Problems: Models and Algorithms

清華大學出版社
北 京

内 容 简 介

高效的应急救援系统对于降低生命和财产损失具有重要意义，选址决策作为战略规划，具有长期的影响力。本书考虑了应急救援过程中可能存在的中断风险、需求波动，以及其他潜在的多重不确定性，采用分布式鲁棒优化方法决策应急设施选址和物资储备方式，改善了救援过程中的资金不足、救援质量下降等问题，利用理论性质对模型进行有效近似，提出了外逼近、分支剪界等算法加速求解效率。研究结果显示，本书的模型较为全面地刻画了实际应急系统，显著提升了传统算法的运算速度，同时有效兼顾了实际救援过程中的效率和公平。

本书可供管理科学与工程学、工业工程、交通运输工程、物流工程及物流管理方向的高年级本科生、研究生及相关领域科研人员参考。

图书在版编目（CIP）数据

基于分布式鲁棒优化的应急救援系统选址模型和算法研究 / 刘康琳著.—北京：清华大学出版社，2022.5

（清华大学优秀博士学位论文丛书）

ISBN 978-7-302-60237-8

Ⅰ. ①基…　Ⅱ. ①刘…　Ⅲ. ①突发事件-救援-公共场所-选址-布局-最优化算法-研究　Ⅳ. ①TU984.199

中国版本图书馆 CIP 数据核字（2022）第 036196 号

责任编辑：戚　亚
封面设计：傅瑞学
责任校对：王淑云
责任印制：宋　林

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　　**邮　　编：**100084
社 总 机：010-83470000　　**邮　　购：**010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：三河市东方印刷有限公司
经　　销：全国新华书店
开　　本：155mm×235mm　**印　张：**10.5　**插　　页：**1　**字　　数：**160 千字
版　　次：2022 年 7 月第 1 版　　**印　　次：**2022 年 7 月第 1 次印刷
定　　价：79.00 元

产品编号：092371-01

一流博士生教育
体现一流大学人才培养的高度（代丛书序）[①]

人才培养是大学的根本任务。只有培养出一流人才的高校，才能够成为世界一流大学。本科教育是培养一流人才最重要的基础，是一流大学的底色，体现了学校的传统和特色。博士生教育是学历教育的最高层次，体现出一所大学人才培养的高度，代表着一个国家的人才培养水平。清华大学正在全面推进综合改革，深化教育教学改革，探索建立完善的博士生选拔培养机制，不断提升博士生培养质量。

学术精神的培养是博士生教育的根本

学术精神是大学精神的重要组成部分，是学者与学术群体在学术活动中坚守的价值准则。大学对学术精神的追求，反映了一所大学对学术的重视、对真理的热爱和对功利性目标的摒弃。博士生教育要培养有志于追求学术的人，其根本在于学术精神的培养。

无论古今中外，博士这一称号都和学问、学术紧密联系在一起，和知识探索密切相关。我国的博士一词起源于 2000 多年前的战国时期，是一种学官名。博士任职者负责保管文献档案、编撰著述，须知识渊博并负有传授学问的职责。东汉学者应劭在《汉官仪》中写道：“博者，通博古今；士者，辩于然否。”后来，人们逐渐把精通某种职业的专门人才称为博士。博士作为一种学位，最早产生于 12 世纪，最初它是加入教师行会的一种资格证书。19 世纪初，德国柏林大学成立，其哲学院取代了以往神学院在大学中的地位，在大学发展的历史上首次产生了由哲学院授予的哲学博士学位，并赋予了哲学博士深层次的教育内涵，即推崇学术自由、创造新知识。哲学博士的设立标志着现代博士生教育的开端，博士则被定义为

① 本文首发于《光明日报》，2017 年 12 月 5 日。

独立从事学术研究、具备创造新知识能力的人，是学术精神的传承者和光大者。

博士生学习期间是培养学术精神最重要的阶段。博士生需要接受严谨的学术训练，开展深入的学术研究，并通过发表学术论文、参与学术活动及博士论文答辩等环节，证明自身的学术能力。更重要的是，博士生要培养学术志趣，把对学术的热爱融入生命之中，把捍卫真理作为毕生的追求。博士生更要学会如何面对干扰和诱惑，远离功利，保持安静、从容的心态。学术精神，特别是其中所蕴含的科学理性精神、学术奉献精神，不仅对博士生未来的学术事业至关重要，对博士生一生的发展都大有裨益。

独创性和批判性思维是博士生最重要的素质

博士生需要具备很多素质，包括逻辑推理、言语表达、沟通协作等，但是最重要的素质是独创性和批判性思维。

学术重视传承，但更看重突破和创新。博士生作为学术事业的后备力量，要立志于追求独创性。独创意味着独立和创造，没有独立精神，往往很难产生创造性的成果。1929 年 6 月 3 日，在清华大学国学院导师王国维逝世二周年之际，国学院师生为纪念这位杰出的学者，募款修造“海宁王静安先生纪念碑”，同为国学院导师的陈寅恪先生撰写了碑铭，其中写道：“先生之著述，或有时而不章；先生之学说，或有时而可商；惟此独立之精神，自由之思想，历千万祀，与天壤而同久，共三光而永光。”这是对于一位学者的极高评价。中国著名的史学家、文学家司马迁所讲的“究天人之际，通古今之变，成一家之言”也是强调要在古今贯通中形成自己独立的见解，并努力达到新的高度。博士生应该以“独立之精神、自由之思想”来要求自己，不断创造新的学术成果。

诺贝尔物理学奖获得者杨振宁先生曾在 20 世纪 80 年代初对到访纽约州立大学石溪分校的 90 多名中国学生、学者提出：“独创性是科学工作者最重要的素质。”杨先生主张做研究的人一定要有独创的精神、独到的见解和独立研究的能力。在科技如此发达的今天，学术上的独创性变得越来越难，也愈加珍贵和重要。博士生要树立敢为天下先的志向，在独创性上下功夫，勇于挑战最前沿的科学问题。

批判性思维是一种遵循逻辑规则、不断质疑和反省的思维方式，具有批判性思维的人勇于挑战自己，敢于挑战权威。批判性思维的缺乏往往被认为是中国学生特有的弱项，也是我们在博士生培养方面存在的一

个普遍问题。2001 年，美国卡内基基金会开展了一项“卡内基博士生教育创新计划”，针对博士生教育进行调研，并发布了研究报告。该报告指出：在美国和欧洲，培养学生保持批判而质疑的眼光看待自己、同行和导师的观点同样非常不容易，批判性思维的培养必须成为博士生培养项目的组成部分。

对于博士生而言，批判性思维的养成要从如何面对权威开始。为了鼓励学生质疑学术权威、挑战现有学术范式，培养学生的挑战精神和创新能力，清华大学在 2013 年发起“巅峰对话”，由学生自主邀请各学科领域具有国际影响力的学术大师与清华学生同台对话。该活动迄今已经举办了 21 期，先后邀请 17 位诺贝尔奖、3 位图灵奖、1 位菲尔兹奖获得者参与对话。诺贝尔化学奖得主巴里·夏普莱斯（Barry Sharpless）在 2013 年 11 月来清华参加“巅峰对话”时，对于清华学生的质疑精神印象深刻。他在接受媒体采访时谈道：“清华的学生无所畏惧，请原谅我的措辞，但他们真的很有胆量。”这是我听到的对清华学生的最高评价，博士生就应该具备这样的勇气和能力。培养批判性思维更难的一层是要有勇气不断否定自己，有一种不断超越自己的精神。爱因斯坦说：“在真理的认识方面，任何以权威自居的人，必将在上帝的嬉笑中垮台。”这句名言应该成为每一位从事学术研究的博士生的箴言。

提高博士生培养质量有赖于构建全方位的博士生教育体系

一流的博士生教育要有一流的教育理念，需要构建全方位的教育体系，把教育理念落实到博士生培养的各个环节中。

在博士生选拔方面，不能简单按考分录取，而是要侧重评价学术志趣和创新潜力。知识结构固然重要，但学术志趣和创新潜力更关键，考分不能完全反映学生的学术潜质。清华大学在经过多年试点探索的基础上，于 2016 年开始全面实行博士生招生“申请–审核”制，从原来的按照考试分数招收博士生，转变为按科研创新能力、专业学术潜质招收，并给予院系、学科、导师更大的自主权。《清华大学“申请–审核”制实施办法》明晰了导师和院系在考核、遴选和推荐上的权力和职责，同时确定了规范的流程及监管要求。

在博士生指导教师资格确认方面，不能论资排辈，要更看重教师的学术活力及研究工作的前沿性。博士生教育质量的提升关键在于教师，要让更多、更优秀的教师参与到博士生教育中来。清华大学从 2009 年开始探

索将博士生导师评定权下放到各学位评定分委员会，允许评聘一部分优秀副教授担任博士生导师。近年来，学校在推进教师人事制度改革过程中，明确教研系列助理教授可以独立指导博士生，让富有创造活力的青年教师指导优秀的青年学生，师生相互促进、共同成长。

在促进博士生交流方面，要努力突破学科领域的界限，注重搭建跨学科的平台。跨学科交流是激发博士生学术创造力的重要途径，博士生要努力提升在交叉学科领域开展科研工作的能力。清华大学于 2014 年创办了"微沙龙"平台，同学们可以通过微信平台随时发布学术话题，寻觅学术伙伴。3 年来，博士生参与和发起"微沙龙"12 000 多场，参与博士生达 38 000 多人次。"微沙龙"促进了不同学科学生之间的思想碰撞，激发了同学们的学术志趣。清华于 2002 年创办了博士生论坛，论坛由同学自己组织，师生共同参与。博士生论坛持续举办了 500 期，开展了 18 000 多场学术报告，切实起到了师生互动、教学相长、学科交融、促进交流的作用。学校积极资助博士生到世界一流大学开展交流与合作研究，超过 60% 的博士生有海外访学经历。清华于 2011 年设立了发展中国家博士生项目，鼓励学生到发展中国家亲身体验和调研，在全球化背景下研究发展中国家的各类问题。

在博士学位评定方面，权力要进一步下放，学术判断应该由各领域的学者来负责。院系二级学术单位应该在评定博士论文水平上拥有更多的权力，也应担负更多的责任。清华大学从 2015 年开始把学位论文的评审职责授权给各学位评定分委员会，学位论文质量和学位评审过程主要由各学位分委员会进行把关，校学位委员会负责学位管理整体工作，负责制度建设和争议事项处理。

全面提高人才培养能力是建设世界一流大学的核心。博士生培养质量的提升是大学办学质量提升的重要标志。我们要高度重视、充分发挥博士生教育的战略性、引领性作用，面向世界、勇于进取，树立自信、保持特色，不断推动一流大学的人才培养迈向新的高度。

邱勇

清华大学校长

2017 年 12 月 5 日

丛书序二

以学术型人才培养为主的博士生教育，肩负着培养具有国际竞争力的高层次学术创新人才的重任，是国家发展战略的重要组成部分，是清华大学人才培养的重中之重。

作为首批设立研究生院的高校，清华大学自 20 世纪 80 年代初开始，立足国家和社会需要，结合校内实际情况，不断推动博士生教育改革。为了提供适宜博士生成长的学术环境，我校一方面不断地营造浓厚的学术氛围，一方面大力推动培养模式创新探索。我校从多年前就已开始运行一系列博士生培养专项基金和特色项目，激励博士生潜心学术、锐意创新，拓宽博士生的国际视野，倡导跨学科研究与交流，不断提升博士生培养质量。

博士生是最具创造力的学术研究新生力量，思维活跃，求真求实。他们在导师的指导下进入本领域研究前沿，吸取本领域最新的研究成果，拓宽人类的认知边界，不断取得创新性成果。这套优秀博士学位论文丛书，不仅是我校博士生研究工作前沿成果的体现，也是我校博士生学术精神传承和光大的体现。

这套丛书的每一篇论文均来自学校新近每年评选的校级优秀博士学位论文。为了鼓励创新，激励优秀的博士生脱颖而出，同时激励导师悉心指导，我校评选校级优秀博士学位论文已有 20 多年。评选出的优秀博士学位论文代表了我校各学科最优秀的博士学位论文的水平。为了传播优秀的博士学位论文成果，更好地推动学术交流与学科建设，促进博士生未来发展和成长，清华大学研究生院与清华大学出版社合作出版这些优秀的博士学位论文。

感谢清华大学出版社，悉心地为每位作者提供专业、细致的写作和出

版指导，使这些博士论文以专著方式呈现在读者面前，促进了这些最新的优秀研究成果的快速广泛传播。相信本套丛书的出版可以为国内外各相关领域或交叉领域的在读研究生和科研人员提供有益的参考，为相关学科领域的发展和优秀科研成果的转化起到积极的推动作用。

感谢丛书作者的导师们。这些优秀的博士学位论文，从选题、研究到成文，离不开导师的精心指导。我校优秀的师生导学传统，成就了一项项优秀的研究成果，成就了一大批青年学者，也成就了清华的学术研究。感谢导师们为每篇论文精心撰写序言，帮助读者更好地理解论文。

感谢丛书的作者们。他们优秀的学术成果，连同鲜活的思想、创新的精神、严谨的学风，都为致力于学术研究的后来者树立了榜样。他们本着精益求精的精神，对论文进行了细致的修改完善，使之在具备科学性、前沿性的同时，更具系统性和可读性。

这套丛书涵盖清华众多学科，从论文的选题能够感受到作者们积极参与国家重大战略、社会发展问题、新兴产业创新等的研究热情，能够感受到作者们的国际视野和人文情怀。相信这些年轻作者们勇于承担学术创新重任的社会责任感能够感染和带动越来越多的博士生，将论文书写在祖国的大地上。

祝愿丛书的作者们、读者们和所有从事学术研究的同行们在未来的道路上坚持梦想，百折不挠！在服务国家、奉献社会和造福人类的事业中不断创新，做新时代的引领者。

相信每一位读者在阅读这一本本学术著作的时候，在吸取学术创新成果、享受学术之美的同时，能够将其中所蕴含的科学理性精神和学术奉献精神传播和发扬出去。

清华大学研究生院院长

2018 年 1 月 5 日

导师序言

本书的编辑出版恰逢新冠疫情世界范围依然流行、极端天气导致的郑州大雨刚刚结束之时，这些事件中的应急救援活动给人们留下了深刻印象。本书针对应急救援设施的选址决策开展理论研究和案例分析，正好迎合了当前热点，反映了研究工作的重要价值。非常感谢清华大学和清华大学出版社的支持，使作者的博士学位论文研究工作能够入选“清华大学优秀博士学位论文丛书”项目。本人也非常荣幸为本书作序。

应急救援在现代社会发展中扮演着越来越重要的角色。它用来减少由于自然灾害（例如极端天气、地震、疫情等）或者人为因素（恐怖袭击、核泄漏等）给人类社会经济系统造成的巨大生命和财产损失。应急救援服务包括灾害发生前、灾害发生时和灾害发生后三个阶段。灾前决策能显著提高救援过程中的救援效率、缩短响应时间、降低救援成本。因此，针对灾前决策的研究得到了学术界的广泛关注。

本书作者针对灾害发生前的应急救援设施选址这一战略性决策开展研究。特别是考虑了救援过程中来自需求端和供给端的不确定因素，采用分布鲁棒优化方法对它们进行数学建模，进而分别针对三类重要的救援设施选址问题开展了理论研究和实证分析，包括考虑需求不确定性的救助站选址问题、在 Wasserstein 模糊集内考虑供给中断风险的选址问题、综合考虑需求端和供给端的不确定性的救助站选址问题。研究工作结合问题特征、数据可得性，以及模型易计算性的要求，对问题进行了数学建模研究；通过分析模型结构特征，开发了高效的最优算法。模型和算法在以实际案例数据为基础的数值分析中表现出优秀的性能；通过算例分析，探索了相应的应急救援设施选址决策中的若干管理规律。

本书的研究应用分布鲁棒优化方法，在应急救援设施选址领域的建

模和算法研究中具有较强的学术创新性，体现了研究工作的学术理论价值。同时，通过案例分析，展示了模型和算法在实际场景中的应用和效果，为不确定环境中应急救援系统的设施选址决策提供了一套可行的优化决策工具，反映出研究工作的应用价值。

当前，在新冠疫情世界范围的大流行、极端天气导致的郑州大雨等事件中，应急救援系统在降低灾害的影响方面起到了重要的作用，也使人们愈发认识到规划建设有效的应急救援系统对于经济社会正常运行的重要意义。本书的工作综合应用了数据驱动的决策理论、优化技术和数据科学方法，对应急救援系统的选址决策进行了开创性的学术研究工作，表明这些方法在应急救援决策中的有效性，为进一步开展本领域深入的理论研究，探索了一条可行的研究技术路线；此外，研究成果的案例研究也预示了定量的科学优化决策方法可以为决策者高效地提供实际可用的解决方案，为提升应急救援系统的决策质量和效率提供有效的工具，从而进一步提升我国应急救援系统在面临突发应急事件时的应对效率和效果。

张智海

2021 年 8 月 6 日于北京

摘　要

应急救援系统不仅可以对日常紧急情况（如火灾、地震、急救电话等）做出实时响应，更可以在面临大规模灾难（如自然灾害、大规模疫情等）时缩短救援时间、提高救援效果，对于降低生命和财产损失具有重要意义。由于突发紧急情况具有高度不确定性，灾时需求量、灾难发生频次、响应时间、道路情况、库存剩余等诸多方面的随机因素使得与应急物流相关的优化问题变得更加复杂。同时，联合国人道主义事务协调办公室的调研报告显示，救援过程中的资金不足问题日趋明显。选址决策作为战略层规划，具有长期的影响力。如何使得选址决策更好地应对可能出现的中断风险、需求波动，以及其他潜在的多重不确定性，如何避免资金不足造成的救援质量下降现象，如何权衡运营成本、效率和公平之间的关系，成为应急救援领域亟待解决的问题。

本书按照应急系统的不确定性来源将随机变量分为两类，分别考虑了需求不确定性、供给不确定性，以及联合不确定性条件下的三类数学模型，并在分布式鲁棒优化的框架下求解。分布式鲁棒优化是近年来新兴的一类研究方法，根据应急系统突发事件频次低、历史数据少的特点，分布式鲁棒优化能够充分利用有限的历史数据对随机参数可能发生的最坏情况进行预估，使系统在最坏情况下的表现不致太差，在保证应急救援领域服务质量的同时最大限度地降低运营成本。

首先，本书将需求不确定性以联合机会约束的形式嵌入救助站选址和规模设定问题，并将原始模型近似为一个带有参数的二次锥规划问题；同时利用模型理论特点，讨论了联合机会约束与独立机会约束近似结果之间的关系。其次，本书探讨了考虑中断风险的应急救援系统选址问题，利用 0-1 变量刻画设施中断，将随机变量的分布模糊集限定在以实证分

布为中心的 Wasserstein 鲁棒集内，采用数据驱动的两阶段优化方法，分析对比了分布式鲁棒优化和传统随机规划的表现；本书还通过有效不等式得到了第二阶段子问题的凸包，将混合整数规划模型等效为一个线性规划，并基于此得到了理论上的最坏情况分布，提出了两种高效的分支剪界算法。最后，本书同时考虑了应急救援系统的需求和供给随机性，利用随机变量的一二阶矩近似原始独立机会约束；在求解时，提出外逼近算法加速，并在实证数据集中验证了模型的效果。在本书的最后，我们对研究内容和创新点进行了总结，并提出了可能的拓展方向。

关键词：应急救援系统；选址问题；灾前准备；分布式鲁棒优化；不确定因素

Abstract

Emergency relief system can not only provide prompt response to common emergencies (such as fire, accidents and emergency calls), it can also shorten relief time, increase relief efficiency, decrease morbidity and mortality dramatically while under the attack of large-scale disasters (such as natural disaster, pandemic etc.). Common emergencies or large-scale disasters may face high uncertainty, such as demand, frequency, response time, traffic, inventory on-hand, etc. Meanwhile, the United Nations' Office for the Coordination of Humanitarian Assistance (OCHA) announced that funding deficit is becoming a prominent issue during the relief process. Location planning, which is a strategic planning decision, will lead to long-lasting effects. How to cope with the uncertainties from demand and supply sides, how to overcome the decrease of service quality due to limited budget, how to balance the trade-off between cost, efficiency and equity, have become urgent issues in the area of relief process.

In this book, we classify the uncertainties into two categories according to their sources, i.e., demand uncertainty and supply uncertainty. After that, we formulate three different models to describe the characteristics mentioned above, including uncertain demand, facility failure, and uncertainties from both sides. All of the problems are solved within the framework of distributionally robust optimization (DRO). DRO is an emerging approach to solve stochastic problems. Because the disasters are less frequent and lack of historical data, DRO can utilize limited empirical data to analysis the worst-case scenario from the ambiguity set, which could guarantee a satisfying system performance even under the worst-case. In the field of emergency relief process, it could ensure system efficiency with a relatively low budget.

Firstly, we consider the emergency medical facility location and sizing problem with joint chance constraints, where demand is assumed to be random. The original problem is reformulated as a parametric second-order cone program, we continue to develop some theoretical comparisons between the approximations of individual and joint chance constraints. Secondly, the book investigates a reliable facility location problem within Wasserstein ambiguity set, where facility failure is characterized as binary random variables. The superiority of the two-stage robust model is highlighted through a comprehensive comparison with classical stochastic programming methods. Notably, the second-stage subproblem can be simplified from integer programming to linear programming by adding some valid inequalities, we can figure out the convex hull of the subproblem. Based on the favorable properties of the reformulated model, two branch-and-cut algorithms are also implemented to increase computational efficiency. Last but not least, we combine the uncertainties from both of the demand and supply sides, and use the first two moments to approximate individual chance constraints, the original problem is finally reformulated as a second-order cone program. By proposing an efficient outer approximation algorithm, the problem is solved and verified through practical datasets. In the end, we conclude this book by summarizing the contributions, and propose several directions for future research.

Key Words: emergency relief system; location; disaster preparedness; distributionally robust optimization; uncertain factors

目　录

Contents

第 1 章　引　　言

1.1　研究背景及意义

近年来，应急救援过程在业界和学术界均受到越来越多的重视[1]。自然灾害（如地震、海啸、飓风、洪水等）[2-4] 和人为灾害（恐怖袭击、大规模游行罢工、核泄漏等）均可能对现有供应链产生毁灭性打击，从而造成巨大的生命和财产损失。例如，1999 年的台湾地震使戴尔公司被迫采取降价策略来保持产品竞争力[3]；2000 年 3 月，由闪电造成的火灾使菲律宾本土的半导体生产工厂瘫痪，电子产品经销商因此遭受了长达六个月的原材料短缺，致使爱立信在北美地区造成了高达 40 亿美元的经济损失[3]；2015 年，尼泊尔发生 8.1 级地震，共造成 8786 人死亡，经济损失 52.3 亿美元，超过 560 万人受灾[5]。Guha-Sapir 等人[5] 的研究显示，2015 年，自然灾害造成的经济财产和人身安全损失巨大，在已报道的 376 次自然灾害中，共造成 22765 人死亡，1.1 亿人受到灾害影响，所造成的财产损失超过 700 亿美元。

在此背景下，运营管理领域旗舰期刊 *Production and Operations Management* 设立了人道主义救援与灾难管理分部（Humanitarian Operations and Crisis Management，HO & CM)，并在 2014 年发布了与之相关的特刊[6]。该特刊指出，针对应急救援过程的优化问题越来越受到重视的主要原因包括：①灾难造成的损失不可忽视，且针对救援过程的优化研究明显不足；②红十字会、联邦应急管理局、牛津饥荒救济委员会等国际组织开始针对不同灾难启动应急管理优化策略；③新兴的研究方法为解决救援过程中的优化问题开辟了新途径。

与人道主义救援相关的优化决策按照突发紧急事件的顺序具体可以

分为三个阶段，即灾害发生前、灾害发生时和灾害发生后[7]，如图 1.1所示。在灾难发生前，优化策略主要体现在设施建设、设施加固、库存备货、服务匹配等战略性策略上。在灾难发生时，主要的优化任务是救灾物资的运输和人员疏散，需要结合灾情作出实时响应，包括资源再分配[8]、现场人员调度[9] 等。在灾难发生后，决策者需对灾后设施重建[10]、受损物资的修复、回收丢弃问题[11] 进行优化。

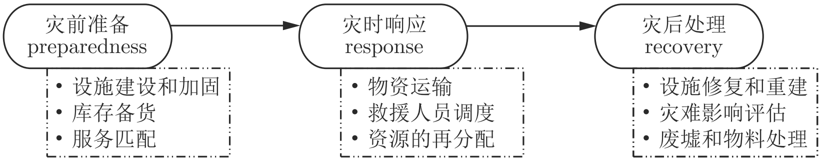


图 1.1　应急救援服务的阶段示意图

本书所研究的选址问题属于**灾前战略性决策**。Wassenhove 等人[12] 认为，灾前决策能显著提高救援过程中的救援效率、缩短响应时间、降低救援成本。考虑到自然灾害的不确定性和不可预测性，在灾前规划阶段，决策者需要通过有效的优化手段，预估灾害中可能发生的各种情况，提供足够的救灾物资，最大限度地减少伤亡人数[13]。此外，联合国人道主义事务协调办公室（OCHA）的调研报告显示，截至 2018 年 8 月，由于资金限制，在世界范围内共有 3800 万人无法接受任何人道主义援助，且资金不足的问题越来越明显[14]。近年来，大量研究开始关注救援系统的优化问题，如何在救援过程中兼顾效率和公平，确保在有限资的源下最大限度地提高救援质量；如何在有限成本下，提供高效的灾前选址、库存和分配决策，越来越受到研究人员的重视，详情可参见近年来的一些综述类文章[15-17]。

救援过程中的随机性可以按照来源大致分为两个部分：需求端和供给端[18]。需求端的不确定性与不同地区的受灾程度、人员密集程度、人口数等因素有关，在数学建模时主要表现为随机需求量。供给端的不确定性主要包括设施的运营情况（如设施在受灾情况下的剩余可用库存、能否提供正常服务等）和运输配送情况（如道路损毁情况、救援响应时间等），在数学建模时主要以设施是否中断、可用库存比例、道路是否中断等随机变量刻画。

除上述现实意义外，研究应急救援系统的选址问题还具有重要的理

论意义。大部分选址问题对应的数学模型均为混合整数线性规划，且已有研究表明，即使最为基础的选址决策都已经被证明是 NP 难（NP-hard）的[19]。本书所考虑的救援系统选址问题在传统研究的基础上考虑了多种来源的不确定因素，将模型重构或近似为混合整数二次锥规划或两阶段随机规划，求解难度进一步加大。为加速求解效率，我们设计了多种精确算法，如分支剪界法、列和约束生成法、外逼近法等。此外，本书还利用了模型的理论性质（如最坏情况分析、有效不等式、最优条件等）对原始模型进行转换，进一步降低了求解难度。

1.2　研究内容及方法

已有文献大多采用随机规划（stochastic programming）的方式刻画系统不确定性，为了降低多重积分的计算负担，大部分研究均采用基于场景的优化策略。与基于确定性模型和人工决策的选址策略相比，基于场景的优化模型可以进一步提高系统的运营效率。例如，2011 年土耳其东部的凡省遭受了 7.2 级大地震，Kilci 等人[156] 采用基于场景的随机规划方法改进了土耳其红新月协会提出的避难所选址策略，实证结果显示，新提出的随机规划方法在降低加权运输距离、提高避难场所使用率方面具有明显优势。

然而，Snyder 在 2007 年的综述中[20] 指出，基于场景的求解方法存在两点不足之处：一是难于精确选择所选用的场景（scenario）；二是场景数量和计算效率之间存在权衡关系，当数量过大时，将难以求解；当数量较少时，则不能很好地表征系统特点。鲁棒优化（robust optimization）完美地解决了上述两个问题，与之相关的概念、定义和文献将在 2.5节详细介绍。Ni 等人[21] 以 2010 年玉树地震为研究背景，对比了确定性模型、两阶段随机规划模型和鲁棒优化模型在仓库选址和应急资源储备等方面的表现。研究结果显示，相对其他两类模型，鲁棒优化模型虽然在建造和库存成本上具有更大的投入，但显著降低了灾难发生时的缺货惩罚，使系统总成本与确定性模型和随机规划模型相比，分别降低了 19% 和 15% 左右。

为更好地处理应急救援过程中的极端不确定性，本书的三个研究内

容全部采用分布式鲁棒优化（distributionally robust optimization）方法。分布式鲁棒优化是近年来新兴的描述事物不确定性的理论研究方法，其原理主要为将不确定因素限定在一个具有某些特定分布特征的集合中，在充分利用当前信息的基础上，最大限度地考虑可能发生的所有情况，从而提出鲁棒、可靠的优化策略。在应急救援领域，与随机规划和鲁棒优化相比，分布式鲁棒优化具备如下两个显著优势：①能够考虑应急救援过程中的最坏情况，使策略选择更为鲁棒；②分布式鲁棒优化能够最大限度地利用已有数据信息和分布信息，使策略选择更具针对性、减少传统鲁棒优化过于保守的特点。

除此之外，分布式鲁棒优化在运算效率和求解算法上均具备一定优势，现有研究成果可以保证算法在可接受时间内求得可行解，并已经成功应用到很多领域，如投资组合优化[22]、生产批量问题[23] 和预约调度[24]。然而，在应急救援领域，分布式鲁棒优化的应用尚处于起步阶段。本书的研究内容和对应具体解决方法主要包括如下方面。

研究内容一：考虑需求不确定性的紧急救助站选址问题

该研究内容主要考虑需求端不确定性对紧急救助站选址问题的影响。主要研究对象是急救中心。急救系统是现代社会医疗服务体系的重要组成部分，也在应急救援过程中发挥着重要作用。它不仅能对日常发生的紧急情况（如火灾、车祸和急救电话）作出实时响应，而且能在大规模灾害发生时，为实施救援、运送物资和紧急治疗提供基础保障，高效的救援系统能有效降低伤亡率，减轻灾害造成的损失。同时，急救中心可以作为库存中心储存一定的药品、急救设备和医疗物资，在发生紧急情况时为受影响地区提供物资支援。

在技术层面，无论是日常紧急情况还是大规模灾难，管理者都将面临相似的优化决策。以本书研究的灾前战略性决策为例，二者均需考虑设施选址、物资库存以及分配规则等优化目标，数学模型具有一定的内在联系。Gralla 等人[25] 认为，衡量救援效率的三个指标为效率、效果和公平，并分别以系统运行成本、服务质量和需求点接受服务的公平性刻画，这三个指标同样在针对急救中心选址的模型（第一个研究内容）中被一一展现。

在考虑需求不确定性时，我们采用了两类随机变量：日常需求和可能

同时发生的最大需求，这两类变量分别被用来描述日常急救情况和灾难发生时的最坏情况。同时，由于日常需求发生频次相对较高，研究人员可以利用历史数据进行有效预测，因而本书采用了数据驱动的分布式鲁棒优化方法，根据估计的样本的均值和方差限定随机变量的不确定集。而对于灾难发生时可能同时发生的最大需求来说，历史数据少、发生频次低，较难采用数据驱动方法精准预测参数范围，本书采用了传统鲁棒优化方法预估参数不确定集合，并引入联合机会约束定量刻画系统满足需求的比例。

由于联合机会约束涉及多重积分且难于求解，我们将原始模型近似为带有参数的二次锥规划（second-order cone problem, SOCP）问题，并提出迭代的 SOCP 方法，求得近似后模型的精确解。不仅如此，我们还将联合机会约束与独立机会约束的理论近似值进行了对比，给出了独立机会约束近似更加保守的理论条件，并在该特殊条件下提出了相应的外逼近算法，加速求解。最后，在数值实验阶段，本书将基于分布式鲁棒优化的救助站选址问题与基于随机优化的选址问题，在成本和稳定性等方面进行了对比。最后，我们将该模型应用于实证数据，对比相关研究，提出管理建议。

研究内容二：在 Wasserstein 模糊集内考虑中断风险的选址问题

该研究内容主要针对供给端的不确定性，并将其运用到应急救援系统的选址问题。在对设施中断进行建模时，本书将其描述为 0-1 随机变量。由于紧急事件大多发生频次低且不确定性极高，历史数据少，随机规划的研究方法不能完整刻画可能出现的所有情况。鲁棒优化考虑了系统最差的可能情况，保证系统在灾难发生时的最坏情况不至具有太差表现，既避免了过高的日常运营成本，又能保证救援的效率和效果。Wasserstein 不确定集是一类新兴的数据驱动的分布式鲁棒优化方法，自 2015 年以来广受关注，具有良好的理论性质：如有限样本保证（finite sample guarantee）、渐进一致性（asymptotic consistency）和易求解性（tractability）[26]。在选取随机变量模糊集时，我们采用 Wasserstein 函数定义实证分布和真实分布之间的距离，并将具体决策分为如下两个阶段：在第一阶段决策系统选址，在第二阶段根据随机参数的实现值决策分配。

问题的求解瓶颈为第二阶段的子问题。第二阶段问题是最大化-最小

化的优化问题。首先，我们利用第二阶段问题的特殊性质给出了模型的最优条件，并证明第二阶段子问题具有超模性（supermodular）。然后，通过有效不等式将原始最大-最小化问题转化成一个完整的最大化问题，并证明了重构模型的参数矩阵为全幺模矩阵（totally unimodular matrix），将混合整数规划问题简化为线性规划。最后，将第一、第二阶段优化决策合并，重构原始模型为一个混合整数线性规划。除此之外，我们还从理论上证明了给定选址决策之后的最坏情况分布。

在求解时，我们采用了两种分支剪界算法，分别基于最坏情况分布和最坏可能场景。在利用最坏情况分布加入有效不等式时，利用现代求解器的 lazycallback 模块加入 Benders' 切和次模切，并在根节点借鉴 Fischetti 等人[27] 的 in-out 方法加速求解。在利用最坏可能场景加入有效不等式时，与传统列和约束生成原理相似，在求解器的 lazycallback 模块加入子问题最坏场景对应的 Benders' 切。

在进行数值验证时，我们对比了商业求解器和两种分支剪界算法的求解效率，利用交叉验证方法确定了 Wasserstein 球的最优半径，利用样本外分析方法对比了随机规划模型、Lu 等人的模型[28] 和 Wasserstein 鲁棒模型，同时计算了鲁棒模型的成本和收益，证明了鲁棒模型可以利用较少的投资成本获得较大的系统稳定性收益。

研究内容三：考虑需求不确定性和中断风险的救助站选址问题

该研究内容综合考虑了需求端和供给端的不确定性。需求端不确定性用随机需求量表示，供给端的不确定性包括设施剩余库存可用比例和道路是否中断两类变量。需求和设施剩余库存用连续随机变量表示、道路是否中断用 0-1 伯努利变量表示。对应于应急救援系统的三个评价指标——效率、效果和公平，此模型分别用最小化系统成本、带有服务水平的机会约束和带有覆盖比例的机会约束表示。

在对模型进行近似时，根据随机变量的分布特点（任意分布、对称分布和单峰对称分布），提出了三种近似服务水平约束的方法。同时，对考虑覆盖比例的机会约束，提出了一系列线性逼近近似方法，将原始多重积分的机会约束线性化。近似后的模型可以用混合整数 SOCP 表示。

在求解时，采用外逼近算法。首先，证明原 SOCP 问题的线性松弛问题具有凸性，然后利用一阶泰勒展开求得有效不等式，最后利用循环加

切和分支剪接两种方法实现算法。

在进行数值验证时，首先验证了外逼近算法能够显著提高商业求解器的运算效率，并对比了三种近似方法下的最优解；其次，对相关参数进行了灵敏度分析；然后，将现有模型框架与随机规划方法在效率、效果和公平三个角度进行对比；最后，将实证数据应用于本模型，探索有价值的管理建议。

1.3 研究框架及本书结构

本书共分为 6 章，其中第 1 章为引言部分，第 2 章为文献综述，第 3~5 章为主要研究内容，第 6 章为总结和展望。本书的研究框架如图 1.2 所示。

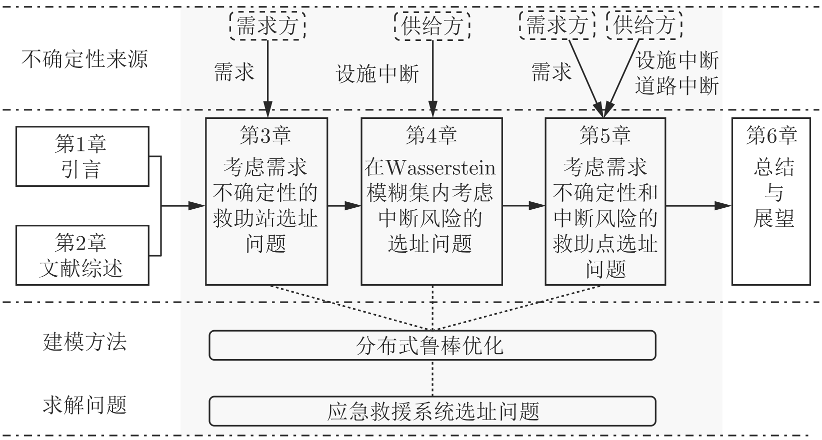


图 1.2 章节结构图

第 1 章为引言，介绍了应急救援过程在现实和理论研究中的重要意义、灾前战略性选址决策在应急救援领域的分类，以及系统的两种不确定性来源。

第 2章为文献综述，首先介绍了经典选址模型的理论基础及其在应急救援领域的应用；随后按照系统随机性的来源，对相关文献进行综述；最后，对本书采用的主要研究方法——分布式鲁棒优化的概念、模型以及应用进行了综述。

第 3章主要考虑了需求不确定性，将数据驱动的分布式鲁棒优化方法应用于救助站选址和规模设定问题，用迭代的二次锥算法求得联合机会约束精确解，既降低了传统 Bonferroni 近似的保守性，又提高了已有的基于场景方法的系统稳定性。本书还从理论上证明了联合机会约束和独立机会约束之间的优势条件 (dominant condition)，指出在特殊情况下独立机会约束可能比联合机会约束的近似结果更为保守，并据此提出外逼近算法，大大提高了问题的求解效率。

第 4章主要考虑供给不确定性对于应急救援系统选址问题的影响。该研究将随机中断变量限定在 Wasserstein 分布模糊集中，建立了两阶段优化模型，利用模型的特殊性质证明了第二阶段子问题的次模性，通过有效不等式将混合整数线性规划问题转化为线性规划。最后提出两类高效的分支剪界算法求解大规模问题。

第 5章同时考虑了供给端和需求端的不确定性，对灾前设施选址问题提供了更加符合现实的建模方法和求解思路。在建模时，同时考虑了应急救援过程的效率、效果和公平原则，本书提出的两类机会约束，分别限定了设施满足需求的概率和需求被覆盖的概率。在求解时，外逼近算法能够显著降低求解器的运算时间。基于实证数据的测试结果进一步证实了鲁棒模型在多重随机条件下的优越表现。

第 6章对全书的要点和贡献进行了总结，并对可能的研究方向进行展望。

第 2 章　文献综述

近年来，国内外学者开始关注应急救援系统中的设施选址问题，涌现出大量相关论文，为本研究的开展奠定了坚实的基础。本章将结合具体研究内容对相关领域的文献进行综述，其中，2.1节介绍了经典选址问题的建模方法及其在应急救援系统中的应用，2.2节概括了考虑需求不确定性的应急系统选址问题，2.3节对考虑中断风险的可靠选址问题进行了梳理和回顾，2.4节介绍了同时考虑需求和供给端不确定性的选址问题及其相关文献，2.5节主要围绕本研究所应用的方法——分布式鲁棒优化进行阐述。

2.1　选址问题及其在应急救援系统中的应用

选址问题主要决策系统中公共设施的地点，以及设施与顾客之间的服务关系。这些基础设施可以包括消防站、急救中心、物流配送中心、社区服务站、学校、电信服务站等，涵盖了日常生活的各个方面。20 世纪 60 年代以来，针对选址的理论优化问题在学术界引起高度关注，至今长盛不衰，具体原因可以概括为如下五点[19]。

第一，选址策略在各个社会层级的团体组织中均普遍出现，从家庭、企业、政府机构到跨国集团，与选址相关的决策屡见不鲜。第二，选址属于战略性决策，相较于其他优化问题，需要花费大量启动成本与社会资源，对组织运营具有长期意义。第三，选址通常与包括污染、拥堵、经济发展在内的经济外部性（economic externalities）因素紧密相连，且与库存、路径规划、用户匹配等其他优化问题息息相关，具体问题往往十分复杂。第四，选址问题大多难以求解，问题的求解难度使其在大规模实证问

题中的算法设计更具挑战。第五，不同背景下的选址问题往往差距很大，目标函数、约束和变量的变化会造成模型和算法的显著差异，因而并不存在一个普适的方法可以应用于一切实际背景。

按照模型的目标函数和约束，经典选址问题可以分为如下几类，包括：覆盖集模型（set-covering model），最大覆盖集模型（maximum covering model）、p 中心模型（p-center model）、p 均值模型（p-median model）、固定费用模型（fixed-charge model）以及 p 枢纽选址模型（p-hub location model），各模型的具体特点及代表文献如表 2.1所示。

表 2.1 经典选址问题综述

模型名称	首次提出文献	概述
覆盖集模型（set-covering）	Toregas 等人[37]	建设最小数量的设施来覆盖所有的需求点
最大覆盖集模型（max covering）	Church 等人[38]	最大化被覆盖的需求，且最多只能建设 p 个设施
p 中心模型（p-center）	Hakimi[39-40]	最小化各个需求点到其最近设施的最大运输成本，且最多只能建立 p 个设施
p 中值模型（p-median）	Hakimi[39-40]	最小化平均运输成本（与需求成正比），且最多只能建立 p 个设施
固定费用模型（fixed-charge）	Balinski[41]	最小化固定建设成本与平均运输成本（与需求成正比）
p 枢纽选址问题（p-hub location）	Okelly[42]	假设存在 p 个枢纽，最小化枢纽之间、枢纽与目的地、枢纽与非枢纽之间的平均运输成本

应急救援领域的选址问题大都基于上述几种经典选址模型、结合具体研究背景对经典问题进行拓展。例如，Badri 等人[29] 考虑了覆盖集问题在消防站选址过程中的特殊性，把运行时间、救援效率等多个冲突的评价指标引入目标函数。Adenso-Díaz 和 Rodriguez[30] 基于传统最大覆盖集模型对乡村地区救护车分配和选址问题建模，通过引入响应时间分析，实现了成本和服务水平的双优化，并利用启发式算法求解。Jia 等人[31] 针对洛杉矶的实际情况调整了选址模型、确定了最优的救助站位置。Ndiaye 和 Alfares[32] 为游牧民族的基本医疗救护点选址问题提出了有针对性的 0-1 规划算法。Scherrer[33] 对新社区健康服务中心的选址问题进行了探

讨。Ares 等人[34] 结合非洲的具体情况，在兼顾公平和效率的基础上，提出了相应的救助站选址模型和列生成算法。孙庆珍等人[35] 将 p 中心模型应用到基于多目标决策的城市应急设施选址问题上，解决了某市新区的消防站选址问题。詹斌和吕腊梅[36] 将选址问题应用到高速公路应急资源配置这一具体背景上。上述论文均假设系统的所有参数为固定值，我们将在 2.2节和 2.3节对考虑随机性的应急救援系统选址问题进行综述。

2.2　考虑需求不确定性的应急系统选址问题

考虑需求不确定的救助站选址问题大致可以分为三类：概率模型、随机优化模型和鲁棒优化模型[16]。概率模型大多以排队论为基础。Larson 等人[43] 在 1973 年首次提出了超立方体排队模型（hypercube queuing model，HQM）；1975 年，Larson[44] 改进了之前模型难以求解的问题，利用近似算法提高可解性。自此，HQM 被广泛应用在类似研究中[45-47]。

随机优化模型大多可分为两个阶段。在问题的第一阶段，求解与随机因素无关的决策（包括选址、设施规模等）；在第二阶段，求解与随机因素相关的决策（如配送等）。Beraldi 等人[48] 在 2004 年的研究中率先将应急救援系统中的需求不确定性纳入研究范畴，建立混合整数规划模型，同时优化了急救中心的选址和规模设定。Beraldi 和 Bruni[49] 拓展了他们在 2004 年的研究，利用机会约束限制了最低的服务水平，并通过可能发生的需求场景对急救系统选址问题进行建模，松弛了排队模型中各服务台独立的假设。文献 [50] 考虑了两级供应链系统的网络规划问题，同时将选址、库存、缺货、分配纳入优化范畴，建立两阶段随机规划模型，利用拉格朗日松弛和启发式算法求解。Erbeyoğlu 和 Bilge[51] 在灾前选址决策中同时考虑了永久设备和临时设备，保证所有可能场景下的需求均可以被满足，并利用基于 logic 的 Benders 分解算法求得了模型的精确解。

在上述研究中，大部分学者假设决策者为风险中立型，仅考虑系统的平均表现。然而，在救援过程中，罕见的极端灾难会造成不可估量的损失，因而将最坏的可能情况纳入优化模型具有重要意义。近年来，部分学者开始考虑风险厌恶的决策问题，例如，Özgün 等人[52] 考虑了两级供应链网络规划问题，在模型中引入关于系统稳定性和风险容忍度的两类机

会约束，并且提出了与风险价值（value-at-risk，VaR）相关的成本函数。Mostajabdaveh 等人[53] 同时将平均运输成本和基尼系数的绝对差值作为优化目标，考虑了灾时通信中断无法交流的困境。

然而，上述两阶段随机优化模型中大部分论文均利用基于场景（scenario-based）的方法建模。为了克服随机规划需要大量历史数据的弊端，Zhang 和 Jiang[54] 用鲁棒优化方法描述了救助站选址问题中的需求不确定性，提出双目标模型权衡了效率与公平之间的关系。Zhang 和 Li[55] 利用联合机会约束刻画了需求被满足的概率，同时引入次模切加速求解。Liu 等人[56] 首次利用分布式鲁棒优化方法将联合机会约束引入模型，并通过迭代的二次锥规划算法求解。姜冬青[57] 利用鲁棒优化方法考虑了库存、选址和路径的联合优化问题。

2.3 考虑中断风险的应急系统选址问题

考虑设施中断风险的选址问题可追溯到 Drezener 在 1987 年发表的论文[58]。该文章利用 p 中值和 p 中心模型刻画设施的中断风险，并采用基于近邻搜索的启发式方法对该问题进行了求解。在此基础上，Lee[59] 采用基于空间曲线的启发式算法，将考虑设施中断的 p 中值模型拆分成 p 个 1 中值模型，使计算复杂度显著降低。与考虑设施中断风险的选址问题相关的研究成果汇总在表 2.2中。

表 2.2 考虑设施中断风险的选址问题相关文献综述

文献	分类	模型		失效概率		是否相关		算法
		UFLP	RPMP	相同	不同	是	否	
随机规划								
Lee[59]	NP		√		√		√	He
Snyder 和 Daskin[60]	NP	√	√	√			√	LR
Berman 等人[61]	NP		√		√		√	He
Cui 等人[62]	NP,CA	√			√		√	CAA
Li 和 Ouyang[63]	CA	√			√	√		CAA
Shen 等人[64]	NP, SB	√		√	√		√	He
Mak 和 Shen[65]	NP,SB	√		√	√		√	He

续表

文献	分类	模型		失效概率		是否相关		算法
		UFLP	RPMP	相同	不同	是	否	
Aboolian 等人[66]	NP	√			√		√	DA,He
Alcaraz 等人[67]	NP	√		√			√	Cplex
Lim 和 Daskin[68]	RB		√		√		√	LR
Lim 等人[69]	CA,RB	√			√	√		CA
Li 等人[70]	RB	√	√		√		√	LR
Li 和 Zhang[71]	SB	√			√	√		SBA
Zarrinpoor 等人[72]	SB	√			√	√		BD
Yu 等人[73]	NP,SB	√			√	√	√	LR
Yu 和 Zhang[74]	NP,SB	√			√	√	√	DA
鲁棒优化								
An 等人[75]	RO		√		√		√	CCG
Lu 等人[28]	RO	√	√		√	√		Cplex
第二个研究内容	RO	√	√		√	√		BD,B&C

分类	NP: 非线性概率模型；CA：连续近似模型；SB：基于场景的模型
	RB: 可靠备份模型；RO: 鲁棒优化模型
模型	UFLP：无能力限制的固定成本选址问题
	RPMP：考虑中断风险的 p 中值模型
算法	He：启发式算法；LR：拉格朗日松弛；CAA：连续近似算法
	DA：分解算法；SBA：基于场景的算法：BD：Benders 分解
	CCG：列与约束生成算法；Cplex：Cplex 求解器直接求解
	B&C：分支剪界

Synder 等人在 2016 年的文献综述中将考虑设施中断风险的选址问题分为四类[18]，包括非线性概率模型（nonlinear probability, NP）、可靠备份模型（reliable backup, RB）、基于场景的模型（scenario-based, SB）和连续近似模型（continuum approximation, CA）。

（1）非线性概率模型大都利用非线性项计算各个需求点被第 r 级设施服务的概率，由 Snyder 和 Daskin 在 2005 年首次提出[60]，并在后续研究中被广泛应用 [4,61-62,64-67,73-74]。例如，Snyder 和 Daskin 在 2005 年的论文[60] 中假设全部设施的中断概率相同且固定，目标函数由不考虑中断发生的期望成本和考虑中断发生的非线性成本加权构成，利用拉格朗

日松弛算法求得了模型的最优解。Shen 等人 [64] 松弛了各点失效概率均必须相同的假设，提出了多种启发式求解算法。Mak 和 Shen[65] 在此基础上考虑了需求不确定性，使用库存策略应对设施的中断风险，提高了系统稳定性。Alcaraz 等人[67] 将覆盖集约束引入模型，通过加入有效不等式提高计算效率。

（2）可靠备份模型将所有设施分为可能发生中断和不会发生中断两类，每个需求点均能够被分配至一个可中断和一个不可中断的设施。Lim 等人[68] 在 2010 年首次提出了设施加固概念和可靠备份模型。Lim 等人[69] 在 2013 年拓展了他们在 2010 年的研究，提出连续空间内的可靠备份的选址问题，重点研究了设施中断估计偏差对于系统成本的影响。研究结果表明，低估风险所造成的损失远大于高估风险时的保守性投入，而各中断事件的相关性对系统成本的影响则相对较小。Li 等人[69] 将可靠备份模型应用于 p 中值和固定成本模型，提出利用建设防御工事降低失效概率的方法，并用拉格朗日松弛算法求解模型。

（3）基于场景的模型通过枚举全部或部分灾害发生的可能情况建立随机规划模型，此类方法直观且能很好地刻画不同场景之间的相关性。然而随着枚举数量的增多，求解难度指数增加，且决策者较难选择最有代表性的场景。例如 Yu 等人在 2017 和 2018 年的研究中 [73-74] 探讨了设施中断风险独立和相关的两种情况，分别利用基于场景法和非线性概率法建模，引入风险评定因子，建立了风险厌恶的可靠设施选址问题。另外，在处理一些较为复杂的问题时，基于场景的建模方法更加常见。例如，Zarrinpoor 等人[72] 考虑了供给端、需求端和中间渠道的不确定性，在选址的基础上引入了排队模型，利用 Benders 分解方法求解。除此之外，利用此方法的文献还有 [64-65,71,76]。

（4）连续近似模型假设顾客的需求在一个连续平面上均匀分布，各个参数均可表示为关于地理位置的连续函数。Cui 等人[62] 利用混合整数规划模型和连续近似方法描述了固定费用选址问题。文章指出，尽管拉格朗日松弛算法能得到理论最优解，但求解速度往往较慢，在计算大规模实证案例时，连续近似算法能在较短时间内得到近优解。与 Cui 等人[62] 的研究成果不同，Li 和 Ouyang[68] 假设顾客完全了解设施失效概率信息，并可以据此自主选择服务。

上述四类模型大都基于传统随机规划，在进行优化时，假设设施中断概率为一个固定的值。然而，在实际生活中，造成设施中断的原因往往是自然灾害或人为事件，发生的概率相对较小，且无足够历史数据支撑，对于失效概率的估计并非准确。近年来，鲁棒优化的出现为此类研究开辟了新的途径，相关的研究现状综述详见 2.5节。目前仅有两篇文章是基于鲁棒优化方法研究这一问题的。An 等人[75] 在 2014 年的研究中考虑了 p 中值模型中的最差情况和平均情况，建立两阶段鲁棒优化模型，并提出列和约束生成的算法。Lu 等人在 2015 年的论文中[28] 通过假设设施中断概率的均值，求得其最坏理论分布，重点考虑了设施中断相关性对于选址策略的影响。

2.4　考虑需求和供给不确定性的应急系统选址问题

在同时考虑供给端和需求端的不确定性时，大部分研究均通过随机优化的方式枚举可能出现的所有场景。在灾前规划阶段，优化内容包括设施选址、库存和缺货等战略性决策，其中，供给方的不确定性可以分为点和边两类。

在考虑点的随机性时，大部分学者采用 2.3节提到的 0-1 随机变量刻画点的中断状态。除此之外，还可以引入 0~1 的连续随机变量描述设施可用库存的剩余比例 [13,77-78]。在考虑边的随机性时，Ball 和 Lin[79] 定义不能在规定时间内进行服务的边为中断边，建立了基于覆盖集的可靠选址模型；Elç 等人[52] 利用易接近指数（score for accessibility）刻画需求点和设施服务中心的道路情况，当且仅当易接近指数大于等于一个预先设定的阈值时，该需求才能被相应的设施服务；Paul 等人在 2019 年的两篇文章中 [80-81] 考虑了随机运输时间、随机灾民痛苦程度和设施受灾程度，分别利用鲁棒优化和两阶段随机规划方法求解。

近年来，同时考虑供给和需求端不确定性的研究日益增多。例如，Rawls 和 Turnquist 在 2010 年的研究[13] 中同时将设施在灾难过程中的可用库存、边的可用运力以及各点的需求定义为随机变量，利用基于场景的随机规划方法建模，并在墨西哥湾易受飓风影响的实证数据集上验证了模型的效果。Raws 和 Turnquis[77] 拓展了他们在 2010 年的研究，通过

引入机会约束保障了需求被满足的概率。Noyan[82] 将目标函数定义为平均运营成本和风险评价因子——条件风险价值（conditional value-at-risk，CVaR）的凸组合，讨论了考虑两种不确定性来源的选址决策。Lu[83] 在利用鲁棒的 p 中心模型的同时考虑了点的随机权重和边的不确定到达时间。Hong 等人[78] 提出了风险厌恶的灾前设施选址问题，利用联合机会约束刻画了系统整体需求被满足的概率。Sanci 和 Daskin[84] 考虑了灾后网络重建过程中的随机性：包括需求不确定性、点的受损情况及边的修复时间。表 2.3汇总了考虑随机因素的应急救援系统选址问题的文献。

表 2.3　考虑多种随机性的应急救援系统选址问题相关文献综述

文献	机会约束	公平性	不确定性			求解方法
			需求	点	边	
随机规划						
Beraldi 和 Bruni[49]	√	√	√			SB,He
Beraldi 等人[48]	√	√	√			Cplex
Rawls 和 Turnquist[77]	√		√	√		SB
Hong 等人[78]	√	√	√		√	SB
Elçi 等人[52]	√		√	√		SB,BD
Özgün 等人[85]	√	√	√		√	SB
Mostajabdaveh 等人[53]	√	√	√		√	SB,GA
Döyen[50]			√	√	√	LR,He
Rawls 和 Turnquist[13]			√	√	√	SB,LR,AA
Noyan[82]			√	√	√	SB,Dep
Paul 和 Wang[80]			√	√	√	SB
Tofighi 等人[86]		√	√	√	√	SB,TDEA
Erbeyoğlu 和 Bilge[51]		√	√			SB,BD
Paul 和 Zhang[81]		√	√	√	√	SB
Sanci 和 Daskin[84]			√	√	√	SB,AA
鲁棒优化						
Zhang 和 Li[55]	√		√			MIP,SC
Liu 等人[56]	√		√			MIP,OA
Ni 等人[21]			√	√	√	MIP

续表

文献	机会约束	公平性	不确定性			求解方法
			需求	点	边	
Wang 和 Paul [87]			√	√		BD
第三个研究内容	√	√	√	√	√	OA

SB: 基于场景的随机规划方法；BD: Benders 分解；
LR: 拉格朗日松弛；He: 启发式算法；
Dep: 分解算法；AA: 近似算法
SC: 次模切；MIP: 混合整数线性规划；
TDEA: 定制的进化算法；GA: 遗传算法

2.5 分布式鲁棒优化及机会约束

本节将具体介绍本书所应用的研究方法——分布式鲁棒优化和机会约束。由于多重积分的存在，大量已有研究成果利用分布式鲁棒优化方法对机会约束进行了近似，因而两种方法之间存在密不可分的联系。

2.5.1 分布式鲁棒优化

不确定性大量存在于现实生活。然而，研究者对于随机参数的估计常常存在系统误差，这些误差的来源包括对系统缺乏足够的了解、缺乏历史数据、建模水平的限制等。在处理参数不确定优化问题时，有两种建模方式被广泛使用，即随机规划和鲁棒优化。

大部分随机规划问题均假设参数的分布形式已知或可以被精确预测，在满足此假设时，随机规划问题的相关求解算法相对成熟，详情可参考一些高质量的综述性文章或教科书[88-90]。

与之对比，鲁棒优化并不假设随机变量的具体分布形式，而是将随机参数限定在一个可能的不确定集合内，对可能出现的最坏情况进行分析和评估。在应急救援领域，灾难所造成的损失往往超出预期，考虑系统最坏情况下的表现与事实更为贴近。不仅如此，大量研究表明，鲁棒优化框架下的数学模型更加易于求解[91-94]。鲁棒优化最初于 1958 年被应用于报童模型，在该问题中，研究者推导了随机需求的最坏分布为两点分布，

并在最坏情况下分析了报童模型[95]。经过近四十年的沉寂，Ben-Tal 和 Nemirovski[96-98]、El Ghaoui 等人 [99-100] 在 1997 年所开创的基础研究，重新推动了鲁棒优化的发展。

已有文献在定义随机参数的不确定集时，大体可以概括为如下几个：

- 椭球集（ellipsoidal uncertainty set）：假设随机参数的分布范围被限制在一个椭球内，通过控制椭球的中心和半径，调节近似的保守程度。研究表明，在不确定集被表示为椭球时，相应的鲁棒优化问题可进一步近似为二次锥问题，便于求解 [99-101]。

- 多面体集（polyhedral uncertainty set）：多面体集被认为是椭球集的一个特例[101]，可以用一系列线性约束表示。然而，该不确定集会随着随机参数维度的增加呈现显著指数增长的趋势，需权衡随机参数维度和模型准确性之间的关系。

- 基数集（cardinality constrained uncertainty set）：Bertsimas 和 Sim (2004)[102] 通过定义一系列多面体集来表征不确定性的预算（budget of uncertainty）。不确定参数可以在估计值的附近上下波动，且波动总量不能超过预先设定的最大成本。尽管基于基数集的最坏情况分析是非凸优化，但通过对偶理论和凸近似，可以用一系列线性约束对该集合进行重构。

- 范数集（norm uncertainty set）：基于范数集的鲁棒优化问题可被描述成包含对偶范数（dual norm）约束的凸优化问题[103]。特别地，L_1 范数和 L_∞ 范数可简化为线性问题，L_2 范数可被简化为二次锥问题。

鲁棒优化能够保证系统在最坏可能的情况下仍然具有可靠的运行状态，因而，相较于传统随机规划来说更为保守。实际上，尽管随机变量的真实分布难以被确切估计，研究人员仍然可以通过历史数据等相关信息估计随机变量的均值、方差或其他分布信息[104]。分布式鲁棒优化在传统鲁棒优化的基础上，充分考虑了变量的分布信息，从而避免过度保守。考虑随机参数分布特征的鲁棒优化模型被称为“分布式鲁棒优化”。分布式鲁棒优化大致可分为两类：基于矩信息的分布式鲁棒优化和基于距离信息的分布式鲁棒优化。

在探讨基于矩信息的分布式鲁棒优化时，大部分研究考虑了随机变量的均值和方差。Scarf 等人在 1958 年[95] 首次将一、二阶距应用于报

童模型。Becker[105] 采用分解算法求解分布式鲁棒优化问题，利用投影高维变量的迭代算法进行求解。Zymler 等人[106] 将均值方差信息应用于联合机会约束，提出半定规划算法。基于均值和方差的鲁棒优化文献还有 [23-24, 102, 107-108]。与前述论文相比，Delage 和 Ye[22] 并未假设固定的均值方差，而是增加了矩信息估计的误差范围，并利用数据驱动的方式从理论上证明了估计误差与数据量之间的关系，此方法在近期的研究中被广泛应用，如装箱问题[109] 和批量生产问题[23]。近年来，大量文献开始考虑基于距离的不确定分布集，即假设随机变量的分布与实际分布之间的距离为一个定值，分布之间的距离可以被定义为 Prohorov 距离[110], Kullback-Leibler 距离 [111], Kantorovich 距离 [112], L_1 范数[113] 和 Wasserstein 距离[114] 等。

2.5.2 机会约束

机会约束由 Charnes 等人在 1958 年[115] 首次提出，定义如下：

$$\mathbb{P}\{\boldsymbol{A}(\tilde{\boldsymbol{z}}) \geqslant \boldsymbol{b}(\tilde{\boldsymbol{z}})\} \geqslant 1-\epsilon \tag{2-1}$$

其中，$\boldsymbol{A}(\tilde{\boldsymbol{z}}) \geqslant \boldsymbol{b}$ 代表与 n 维随机变量 $\tilde{\boldsymbol{z}}$ 相关的 m 个线性约束。机会约束 (2-1) 保证 m 个线性约束同时成立的概率不低于 $1-\epsilon$。当 $m=1$ 时，被称为“独立机会约束”，当 $m>1$ 时，被称为“联合机会约束”。

与机会约束相关的问题颇具挑战，具体原因有如下三个[116]。

（1） 机会约束是非凸问题。

（2） 多重积分使得不等式左边部分的精确概率难以计算。

（3） 随机变量的分布形式 $\mathbb{P}$ 难以估计。

为解决第一个问题，科研工作者通过寻找特殊分布将非凸优化近似为凸优化[111, 115, 117], 或提出相应的保守近似[118-120]。为解决第二个问题，已有研究提出了基于场景的近似策略[111, 119, 121]、整数规划建模 [122-123] 和鲁棒优化算法[102, 124]。为解决第三个问题，现有文献利用实证数据估计随机参数的分布情况，建立了分布 $\mathbb{P}$ 的模糊集（ambiguity set），提出了相应的近似方法 [106,110-111,116,119]。上述大量研究均关注独立机会约束的近似方法，只有少量文献 [106, 116, 125] 考虑了联合机会约束。

2.6 本 章 小 结

本章结合选题背景、研究内容和研究方法对应急救援系统设施选址问题的相关文献进行了综述。首先，我们回顾了经典选址问题的一般建模方法，并结合救援背景简单介绍了几类选址问题的应用场景。随后，我们按照不确定性来源对考虑随机条件的应急救援领域选址问题进行了梳理，分别从需求不确定性、供给不确定性和二者结合的不确定性三个角度汇总和对比了已有研究成果。最后，结合本书所采用的分布式鲁棒优化方法，简要介绍了该领域的概念、定义、建模方式和应用背景，并重点对研究内容一和研究内容三中提到的机会约束进行了系统阐述。

第 3 章　考虑需求不确定性的救助站选址问题

高效的应急救援系统不仅能应对日常生活中的突发紧急情况（如火灾和车祸、突发疾病等），还能在城市发生大规模紧急事件（如地震、台风、海啸、大规模传染性疾病）时，提供及时有效的救助支持，最大限度地挽回人民的生命和财产损失。救助站的选址问题与传统选址问题有如下几点差异。

第一，必须保证服务质量，也就是在建模时需考虑救援的响应时间、救援的覆盖面等指标。第二，在保证服务质量的同时，需要降低运营成本，兼顾救援的效率和公平。第三，由于紧急事件具有高度不确定性，如何将救援过程中的不确定性因素用抽象的数学模型来刻画颇具挑战。第四，新的优化策略和研究手段的出现，为解决救助站选址问题提供了新的研究思路。综上所述，考虑急救系统救助站选址问题具有重要意义。

3.1　问题描述与建模

急救网络由一系列需求点和可选设施点集组成，救援物资和设备（以救护车为例）被储藏在急救中心以满足附近需求点的可能需求。本节引入两个随机变量刻画需求的不确定性：一是各点的日常需求，用 Θ 表示；二是在大规模灾难发生时可能发生的最大需求（maximum number of concurrent demands, MNCD），用 D 表示。前者通过日常 24 小时内收到的急救电话数刻画；而后者则代表了在大规模灾难发生时，在处理单个紧急任务的平均时长内可能接到的急救电话数。本研究的主要任务是在同时考虑日常和极端情况下需求不确定性的基础上，找到最优的急救中心选址、物资储藏量和匹配方案，使需求被满足的概率大于或等于预先设

定的服务水平，且最小化总运营成本。本章建立了两阶段分布式鲁棒优化模型（distributionally robust model, DRM），所用的符号体系如下。

随机变量：

I 需求点集合，用 i 代表序号；

J 可能的救助站点集合，用 j 代指序号；

I_j 能够被设施 j 覆盖的需求点集，即 $I_j = \{i \in I | c_{ij} \leqslant T\}$；

J_i 能够覆盖的需求点 i 的设施集，即 $J_i = \{j \in J | c_{ij} \leqslant T\}$；

T 能在规定响应时间内到达的最大距离；

f_j 设施 j 的日常建造成本；

a_j 设施 j 存储救援物资和设备的日常运营成本；

c_{ij} 设施 j 和需求点 i 之间的距离；

β 单位运输成本；

Θ_i 随机变量，代表需求点 i 的日常需求；

D_i 随机变量，代表需求点 i 可能同时发生的最大需求。

决策变量：

X_{ij} 连续变量，需求点 i 的需求被设施 j 满足的比例；

Y_j 0-1 变量，当在设施 j 建立救助站时为 1，否则为 0；

N_j 整数变量，在设施点 j 储藏的救援物资（如救护车）数量。

另外，本书中使用***黑斜体***的字母代表向量或者矩阵。

本研究引入联合机会约束刻画系统需求在整个地理区域内被满足的概率。联合机会约束是独立机会约束的拓展，定量刻画了系统的稳定性，与考虑独立机会约束的救助站选址论文[55] 相比，本研究的不同主要体现在四个方面：① 拓展独立机会约束为联合机会约束，提高了系统整体需求被满足的概率；② 从理论上证明了独立机会约束和联合机会约束之间的关系，具有一定的理论贡献；③ 考虑了风险厌恶的目标函数，弥补了风险中立研究方法的不足；④ 提出外逼近方法加速模型求解，而 Zhang 和 Li[55] 仅采用商业软件求解。

与 MNCD 相关的联合机会约束可以被描述为

$$\mathbb{P}\left\{\sum_{i \in I_j} D_i X_{ij} \leqslant N_j, \forall j \in J\right\} \geqslant \alpha \tag{3-1}$$

综上，风险厌恶的两阶段救助站选址和规模设定模型汇总如下：

$$
\begin{aligned}
\mathrm{P}:\min \quad & \left(\sum_{j\in J} f_j Y_j + \sum_{j\in J} a_j N_j + \sup_{\substack{F\in\mathcal{F}\\ G\in\mathcal{G}}} \mathbb{E}_{F,G}\left[g(\boldsymbol{Y},\boldsymbol{N},\boldsymbol{\varTheta},\boldsymbol{D})\right]\right) & \text{(3-2)}\\
\text{s.t.} \quad & N_j \leqslant MY_j, \forall j\in J & \text{(3-3)}\\
& Y_j \in \{0,1\}, \forall j \in J & \text{(3-4)}\\
& N_j \in \mathbb{Z}^+, \forall j\in J & \text{(3-5)}
\end{aligned}
$$

目标函数 (3-2) 用以最小化在不确定集合内的最大成本，其中集合 $\mathcal{F}$ 和 $\mathcal{G}$ 分别代表随机变量 $\boldsymbol{\varTheta}$ 和 $\boldsymbol{D}$ 的不确定分布集，分布 F 和 G 分别属于集合 $\mathcal{F}$ 和 $\mathcal{G}$。系统总成本由急救中心建造成本、设备购买和运维成本，以及运输成本组成。$g(\boldsymbol{Y},\boldsymbol{N},\boldsymbol{\theta},\boldsymbol{d})$ 是当固定 $\boldsymbol{Y},\boldsymbol{N},\boldsymbol{\varTheta}=\boldsymbol{\theta},\boldsymbol{D}=\boldsymbol{d}$ 时的第二阶段成本。约束 (3-3) 限制了救护车只能被安置在已经建立的设施点。约束 (3-4) 和约束 (3-5) 是 0-1 变量约束和非负整数约束。

考虑第二阶段问题时的模型如下：

$$
\begin{aligned}
g(\boldsymbol{Y},\boldsymbol{N},\boldsymbol{\theta},\boldsymbol{d}) = \min \ & \beta\sum_{i\in I}\left(\theta_i \sum_{j\in J} c_{ij}X_{ij}\right) & \text{(3-6a)}\\
\text{s.t.} \ & \sum_{j\in J_i} X_{ij} = 1, \forall i\in I & \text{(3-6b)}\\
& X_{ij} \leqslant Y_j, \forall i\in I, \forall j \in J & \text{(3-6c)}\\
& \mathbb{P}\left\{\sum_{i\in I_j} d_i X_{ij} \leqslant N_j, \forall j\in J\right\} \geqslant \alpha & \text{(3-6d)}\\
& 0\leqslant X_{ij}\leqslant 1. & \text{(3-6e)}
\end{aligned}
$$

目标函数 (3-6a) 旨在最小化运输成本。约束 (3-6b) 保证每个需求点 i 的需求被各个设施分担。约束 (3-6c) 表示需求只能被分配给已经建立的设施。约束 (3-6d) 是联合机会约束。约束 (3-6e) 限制了决策变量 X_{ij} 的范围。

3.2 模型近似

由于联合机会约束是非凸的且难于求解，我们利用数据驱动的鲁棒优化方法[22] 将关于第一个随机变量——日常需求（D）的目标函数近似

成一个带有参数的二次锥规划（second-order cone program, SOCP），并把随机变量 MNCD 限定在一个给定均值和方差的椭球内[55]。在考虑矩信息的分布式鲁棒模型优化中，考虑一阶矩和二阶矩是一种常见的建模形式。一阶距和二阶矩分别可以用均值和方差表示，研究人员能够直观地从历史数据中得到相应的估计值，对真实的分布情况进行判断，大量文献采用类似方法建模[22,24,95]。在研究中引入两个随机变量主要是出于变量物理意义的考虑：日常需求较为平稳，且历史数据较大，均值和方差均较为稳定，可以充分利用已有的历史数据估计矩信息；而 MNCD 用来描述极端大规模灾害时的需求，历史数据较少、难于估计、波动巨大，因而采用较为保守的传统鲁棒优化方法建模。

3.2.1 目标函数

由目标函数 (3-6a) 不难看出，当建造变量 $\boldsymbol{Y}$ 和救护车数量 $\boldsymbol{N}$ 固定时，第二阶段问题的最优解仅仅依赖于单位运输成本 β，距离参数 c_{ij} 和日常需求 $\boldsymbol{\theta}$。由于目标函数 (3-6a) 是日常需求 $\boldsymbol{\theta}$ 的线性函数，第二阶段的平均成本 $g(\boldsymbol{Y},\boldsymbol{N},\boldsymbol{\theta},\boldsymbol{d})$ 仅仅依赖于随机变量 $\boldsymbol{\theta}$。考虑日常需求 $\boldsymbol{\Theta}$ 的一阶矩足以表征问题的不确定性。因此，可以基于上述性质得到目标函数的等价转化，即将原始最小化-最大化目标函数 (3-2) 重构为一个整体的最小化问题。在本研究中，我们将一阶矩 $\mathbb{E}_F[\boldsymbol{\Theta}]$ 限定在一个椭球内，且假设其估计的均值为 $\boldsymbol{\mu}\in\mathbb{R}^I$，估计的协方差矩阵为 $\boldsymbol{\Sigma}\geqslant 0$。在此基础上，不确定分布集可以概括为

$$\mathcal{F}=\left\{F:(\mathbb{E}_F[\boldsymbol{\Theta}]-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbb{E}_F[\boldsymbol{\Theta}]-\boldsymbol{\mu})\leqslant\epsilon^2\right\} \tag{3-7}$$

其中，ϵ 为不确定集合的半径，控制不确定集范围，ϵ 的取值直接表示了对一阶矩近似的准确程度。

由 Delage 和 Ye 在 2010 年的研究可知[22]，参数 ϵ 的大小可以根据历史数据的个数来控制，若日常需求的历史数据有 M 个，即 $\{\theta^i\}_{i=1}^M$，且通过历史数据计算的随机变量均值和协方差矩阵分别为 $\boldsymbol{\mu}_0$ 和 $\boldsymbol{\Sigma}_0$。假设存在实数 $R\geqslant 0$ 和 $\delta>0$，使 $\mathbb{P}\{(\boldsymbol{\Theta}-\boldsymbol{\mu_0})^{\mathrm{T}}\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\Theta}-\boldsymbol{\mu_0})\leqslant R^2\}=1$，则我们能够以不低于 $1-\delta$ 的概率得到

$$(\boldsymbol{\mu_0}-\hat{\boldsymbol{\mu}})^{\mathrm{T}}\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\mu_0}-\hat{\boldsymbol{\mu}})\leqslant\eta(\delta) \tag{3-8}$$

其中，$\hat{\boldsymbol{\mu}} = (1/M)\sum\limits_{i=1}^{M}\theta^i$，$\eta(\delta) = (R^2/M)[2+\sqrt{2\ln(1/\delta)}]^2$。

命题 3.1　模型 P 可以被等价转换为如下形式，

$$\min_{r,\boldsymbol{q},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{N}} \quad \beta(\epsilon r + \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{q}) + \sum_{j\in J}\{f_jY_j + a_jN_j\} \tag{3-9}$$

$$\text{s.t.} \quad q_i = \sum_{j\in J_i} c_{ij}X_{ij}, \forall i \tag{3-10}$$

$$\|\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{q}\| \leqslant r \tag{3-11}$$

$$r \geqslant 0, \boldsymbol{q} \geqslant 0 \tag{3-12}$$

$$\text{式 (3-3)} \sim \text{式 (3-5)}, \text{式 (3-6b)} \sim \text{式 (3-6e)}$$

其中，r 和 $\boldsymbol{q}$ 是辅助决策变量。

证明　由式 (3-6a) 可知,

$$\mathbb{E}_F[g(\boldsymbol{Y},\boldsymbol{N},\boldsymbol{\theta})] = \beta(\boldsymbol{q}^*)^{\mathrm{T}}\mathbb{E}_F[\boldsymbol{\Theta}] = \min_{(\boldsymbol{X},q)\in\Omega(\boldsymbol{Y},\boldsymbol{N})} \boldsymbol{q}^{\mathrm{T}}\mathbb{E}_F[\boldsymbol{\Theta}],$$

其中，$\Omega(\boldsymbol{Y},\boldsymbol{N})$ 由约束 (3-6b) $\sim$ 约束 (3-6e) 定义。因此，

$$\begin{aligned}\sup_{F\in\mathcal{F}} \mathbb{E}_F[g(\boldsymbol{Y},\boldsymbol{N},\boldsymbol{\theta})] &= \beta\sup_{F\in\mathcal{F}}\min_{(\boldsymbol{X},q)\in\Omega(\boldsymbol{Y},\boldsymbol{N})} \boldsymbol{q}^{\mathrm{T}}\mathbb{E}_F[\boldsymbol{\Theta}] \\ &= \beta\max_{F\in\mathcal{F}}\min_{(\boldsymbol{X},q)\in\Omega(\boldsymbol{Y},\boldsymbol{N})} \boldsymbol{q}^{\mathrm{T}}\mathbb{E}_F[\boldsymbol{\Theta}]\end{aligned} \tag{3-13}$$

式 (3-13) 中的最大化-最小化运算等价于其对应的最小化-最大化表示方法。对于最优解 $\boldsymbol{q}^*$ 而言，有

$$\begin{aligned}&\max_{F\in\mathcal{F}}\min_{(\boldsymbol{X},q)\in\Omega(\boldsymbol{Y},\boldsymbol{N})} \boldsymbol{q}^{\mathrm{T}}\mathbb{E}_F[\boldsymbol{\Theta}] \\ =&\max_{F\in\mathcal{F}}(\boldsymbol{q}^*)^{\mathrm{T}}\mathbb{E}_F[\boldsymbol{\Theta}] \geqslant \min_{(\boldsymbol{X},q)\in\Omega(\boldsymbol{Y},\boldsymbol{N})}\max_{\mathbb{E}[\boldsymbol{\Theta}]\in\Lambda}(\boldsymbol{q})^{\mathrm{T}}\mathbb{E}_F[\boldsymbol{\Theta}]\end{aligned}$$

其中，$\Lambda = \{\boldsymbol{x}\in\mathbb{R}^{\mathrm{T}} : (\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \leqslant \epsilon^2\} \subseteq \mathbb{R}_+^{\mathrm{T}}$。

在第二阶段，存在一个 $g(\boldsymbol{Y},\boldsymbol{N},\boldsymbol{\theta})$ 的最优解 $(\boldsymbol{X}^*,\boldsymbol{q}^*)$ 使

$$\max_{\mathbb{E}[\boldsymbol{\Theta}]\in\Lambda}\min_{\boldsymbol{X}\in\Omega(\boldsymbol{Y},\boldsymbol{N})} \boldsymbol{q}^{\mathrm{T}}\mathbb{E}_F[\boldsymbol{\Theta}] = \max_{\mathbb{E}[\boldsymbol{\Theta}]\in\Lambda}(\boldsymbol{q}^*)^{\mathrm{T}}\mathbb{E}[\boldsymbol{\Theta}] \geqslant \min_{\boldsymbol{X}\in\Omega(\boldsymbol{Y},\boldsymbol{N})}\max_{\mathbb{E}[\boldsymbol{\Theta}]\in\Lambda} \boldsymbol{q}^{\mathrm{T}}\mathbb{E}_F[\boldsymbol{\Theta}]$$

第一个等式成立的原因是 $\boldsymbol{q}^*$ 为第二阶段问题的最优解。另外，由于 $\boldsymbol{X}^*\in\Omega(\boldsymbol{Y},\boldsymbol{N})$，第二个不等式显然成立；并且，根据 minimax 不等式[126]，该

不等式的反方向也成立。由 $\boldsymbol{\Theta}$ 的均值和协方差信息可知，$\max\limits_{\boldsymbol{x}\in\Lambda}\boldsymbol{q}^{\mathrm{T}}\boldsymbol{x}=\epsilon\sqrt{\boldsymbol{q}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{q}}+\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{q}$，模型 P 被重构成

$$
\begin{aligned}
\min_{r,\boldsymbol{q},\boldsymbol{X},\boldsymbol{Y},\boldsymbol{N}} \quad & \epsilon\sqrt{\boldsymbol{q}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{q}}+\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{q}+\sum_{j\in J}\left\{f_jY_j+a_jN_j\right\}, \\
\text{s.t.} \quad & q_i=\sum_{j\in J}c_{ij}X_{ij},\forall i, \\
& q\geqslant 0, \text{式 (3-3)} \sim \text{式 (3-5)}, \text{式 (3-6b)} \sim \text{式 (3-6e)}.
\end{aligned}
$$

在引入辅助变量 r 后，命题 3.1 得证。 □

3.2.2 机会约束

假设随机变量 $\boldsymbol{D}$ 的分布 G 被限制在如下模糊集中，

$$
\mathcal{G}=\left\{G:\ \begin{array}{c}(\boldsymbol{D}-\boldsymbol{u})^{\mathrm{T}}\boldsymbol{\Gamma}^{-1}(\boldsymbol{D}-\boldsymbol{u})\leqslant Q^2\\ \mathbb{E}_G(\boldsymbol{D})=\boldsymbol{u}\\ \mathbb{E}_G(\boldsymbol{D}^2)=\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}+\boldsymbol{\Gamma}\end{array}\right\} \tag{3-14}
$$

其中，参数 $Q>0$ 控制模糊集 $\mathcal{G}$ 的大小。由于盒子不确定集具有较强的保守性，本书仅考虑了椭球不确定集[127]。

鉴于联合机会约束的近似方法十分复杂，在处理时，我们首先通过命题 3.2介绍独立机会约束的近似方法，然后在此基础上对联合机会约束进行近似。

定义 $v(N_j,\boldsymbol{X}_j)=\sum\limits_{i\in I_j}d_iX_{ij}-N_j=\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{d}-N_j$，由于独立机会约束的二次锥近似函数 $\mathbb{E}(v(N_j,\boldsymbol{X}_j)^+)$ 的上界，本书在引理 3.1中给出了该上界的表达方式并在引理 3.2中证明了其具有可加性。

引理 3.1 假设 $\boldsymbol{u}$ 和 $\boldsymbol{\Gamma}$ 是随机变量 $\boldsymbol{d}$ 的均值和协方差矩阵，$\pi(N_j,\boldsymbol{X}_j)$ 是函数 $\mathbb{E}(v(N_j,\boldsymbol{X}_j)^+)$ 的一个上界，则

$$
\pi(N_j,\boldsymbol{X}_j)=\frac{1}{2}(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j)+\frac{1}{2}\sqrt{(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j)^2+\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j} \tag{3-15}
$$

证明 由 $w^+=(w+|w|)/2$, 可得 $\mathbb{E}\left[v(N_j,\boldsymbol{X}_j)^+\right]=\frac{1}{2}\mathbb{E}(v(N_j,\boldsymbol{X}_j)+|v(N_j,\boldsymbol{X}_j)|)$。对凸函数 $\psi(\cdot)$ 而言，Jensen's 不等式 $\psi[\mathbb{E}(X)]\leqslant\mathbb{E}[\psi(X)]$

恒成立；因此，

$$\begin{aligned}\left[\mathbb{E}\left|v(N_j,\boldsymbol{X}_j)\right|\right]^2 \leqslant & \mathbb{E}\left[\left|v(N_j,\boldsymbol{X}_j)\right|^2\right]=\mathbb{E}\left[\left|\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{d}-N_j\right|^2\right]\\ =&\mathbb{E}\left[\left(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{d}\right)^2+N_j^2-2N_j\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{d}\right]\\ =&N_j^2+\mathbb{E}\left[\left(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{d}\right)^2\right]-2N_j\boldsymbol{X}_j^{\mathrm{T}}\mathbb{E}(\boldsymbol{d})\\ =&\left[\boldsymbol{X}_j^{\mathrm{T}}\mathbb{E}(\boldsymbol{d})\right]^2+\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j+N_j^2-2N_j\boldsymbol{X}_j^{\mathrm{T}}\mathbb{E}(\boldsymbol{d})\\ =&\left(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}\right)^2+\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j+N_j^2-2N_j\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}\\ =&(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j)^2+\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j\end{aligned}$$

其中，第一个不等式成立的原因是函数 $\psi(x)=x^2$ 为凸函数。由此可知，$\mathbb{E}\left[v(N_j,\boldsymbol{X}_j)^+\right]\leqslant\frac{1}{2}(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j)+\frac{1}{2}\sqrt{(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j)^2+\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j}=\pi(N_j,\boldsymbol{X}_j)$，引理得证。□

引理 3.2　函数 $\pi(N,\boldsymbol{X})$ 是可加函数（subadditive function），即 $\pi(N_1,\boldsymbol{X}_1)+\pi(N_2,\boldsymbol{X}_2)\geqslant\pi(N_1+N_2,\boldsymbol{X}_1+\boldsymbol{X}_2)$。

证明　假设$S=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{u}-N$，$\boldsymbol{U}=(S,\boldsymbol{X})$，$\boldsymbol{\Gamma}_1=\left(\begin{array}{c|ccc}1&0&\cdots&0\\ \vdots&&\boldsymbol{\Gamma}&\\ 0&&&\end{array}\right)$，

其中，$\boldsymbol{\Gamma}_1$ 是半正定矩阵。式（3-15）中的非线性项可用二范数（euclidean norm）表示，即 $\sqrt{(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{u}-N)^2+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}}=\sqrt{S^2+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}}=\sqrt{\boldsymbol{U}^{\mathrm{T}}\boldsymbol{\Gamma_1}\boldsymbol{U}}=\left\|\boldsymbol{\Gamma}_1^{\frac{1}{2}}\boldsymbol{U}\right\|$。由范数的可加性 $||\cdot||$ 可知，$||A||+||B||\geqslant||A+B||$，因而，$\left\|\boldsymbol{\Gamma}_1^{\frac{1}{2}}\boldsymbol{U}_1\right\|+\left\|\boldsymbol{\Gamma}_1^{\frac{1}{2}}\boldsymbol{U}_2\right\|\geqslant\left\|\boldsymbol{\Gamma}_1^{\frac{1}{2}}(\boldsymbol{U}_1+\boldsymbol{U}_2)\right\|$。引理得证。□

在得到 $\mathbb{E}(N_j,\boldsymbol{X}_j)$ 的上界之后，独立机会约束的二次锥近似可以由命题 3.2获得。

命题 3.2　独立机会约束

$$\mathbb{P}\left\{\sum_{i\in I_j}d_iX_{ij}-N_j\leqslant 0\right\}\geqslant 1-\varepsilon,\forall j\in J \tag{3-16}$$

的二次锥近似为

$$\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j+\sqrt{\frac{1-\varepsilon}{\varepsilon}}\sqrt{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j}\leqslant 0,\forall j\in J \tag{3-17}$$

其中，$\varepsilon=1-\alpha$。

证明 由于 CVaR（conditional-value-at-risk）函数的易解性，以往文献大都利用 CVaR 对独立机会约束进行凸近似[128]。CVaR 的经典定义 $\varrho_{1-\varpi}(\cdot)$ 如下：

$$\varrho_{1-\varpi}(\tilde{v})\triangleq\min_{\iota}\left\{\iota+\frac{1}{\varpi}\mathbb{E}[(\tilde{v}-\iota)^{+}]\right\} \tag{3-18}$$

其中，$\tilde{v}$ 是随机变量，$\varpi\in\{0,1\}$ 是一个接近于 0 的安全因子。CVaR 代表概率分布中不小于 $1-\varpi$ 置信水平的自变量的均值。以往研究证明，CVaR 约束是独立机会约束 $\mathbb{P}\{y(\tilde{\boldsymbol{z}})\leqslant 0\}\geqslant 1-\varpi$ 的最紧近似，其中 $y(\tilde{\boldsymbol{z}})$ 与随机向量 $\tilde{\boldsymbol{z}}$ 仿射相关 [118-119]。由此可知，若约束

$$\min_{\iota}\left\{\iota+\frac{1}{\varepsilon}\mathbb{E}\left[\left(\sum_{i\in I_j}d_iX_{ij}-N_j-\iota\right)^{+}\right]\right\}\leqslant 0$$

成立，则独立机会约束 (3-16) 一定成立。

由于均值函数 $\mathbb{E}\left[\left(\sum\limits_{i\in I_j}d_iX_{ij}-N_j-\iota\right)^{+}\right]$ 仍然难以求解，我们采用引理 3.1获得的上界 $\mathbb{E}[(\cdot)^{+}]$ 近似该均值，即

$$\begin{aligned}
&\varrho_{1-\varepsilon}[v(N_j,\boldsymbol{X}_j)]\\
\leqslant&\min_{\iota}\left(\iota+\frac{\pi(N_j+\iota,\boldsymbol{X}_j)}{\varepsilon}\right)\\
=&\min_{\iota}\left(\iota+\frac{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j-\iota}{2\varepsilon}+\frac{\sqrt{(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j-\iota)^2+\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j}}{2\varepsilon}\right)\\
=&\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j+\sqrt{\frac{1-\varepsilon}{\varepsilon}}\sqrt{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j}
\end{aligned} \tag{3-19}$$

其中，最后一个约束成立的条件是 $\iota^{*}=\dfrac{(1-2\varepsilon)\sqrt{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j}}{2\sqrt{\varepsilon(1-\varepsilon)}}+\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j$。因此，不等式 (3-17) 是独立机会约束 (3-16) 的一个有效近似，命题得证。 □

命题 3.2得到的独立机会约束近似可以被拓展到联合机会约束。根据集合性质，联合机会约束 (3-6d) 等价于$\mathbb{P}\Big(\bigcup_{j\in J}\Big\{\sum_{i\in I_j} d_i X_{ij} > N_j\Big\}\Big) \leqslant \varepsilon$。Bonferroni 不等式可以将联合机会约束简化为独立机会约束，具体操作如下：首先将联合机会约束放缩为一系列独立机会约束的和，

$$\mathbb{P}\left(\bigcup_{j\in J}\left\{\sum_{i\in I_j} d_i X_{ij} - N_j > 0\right\}\right) \leqslant \sum_{j\in J}\left[\mathbb{P}\left(\sum_{i\in I_j} d_i X_{ij} - N_j > 0\right)\right] \leqslant \varepsilon \tag{3-20}$$

然后对求和公式中的每个元素利用独立机会约束进行近似，即，对于任意 $j \in J$,

$$\begin{aligned}\mathbb{P}\left(\sum_{i\in I_j} d_i X_{ij} - N_j > 0\right) \leqslant \varepsilon_j, \forall j \Rightarrow \mathbb{P}\left(\sum_{i\in I_j} d_i X_{ij} - N_j \leqslant 0\right)\\ \geqslant 1-\varepsilon_j, \forall j\end{aligned} \tag{3-21}$$

其中，$\sum_{j\in J}\varepsilon_j = \varepsilon$。由于式 (3-16) 与式 (3-21) 相比唯一的不同点是右端项 ε 和 ε_j 的值，联合机会约束 (3-6d) 可被近似为式 (3-17)。

然而，在实际运算中如何找到最优的 ε_j 是极为困难的。尽管近似程度较为保守，根据 Nemirovski 和 Shapiro[119]，以及 Chen 等人[129] 的论文，大部分研究均直接令 $\varepsilon_j = \varepsilon/|J|$。

定义集合 $\mathcal{W}$ 为 MNCD 的不确定集合，根据 Chen 等人[118] 在 2010 年提出的方法，我们在此基础上提出了含参二次锥规划（parametric SOCP）近似方法，并在命题 3.3中证明。在含参 SOCP 近似方法中，引入两个参数：集合 $\mathcal{J}$（全集 $\mathcal{W}$ 的一个子集）和常数 $\lambda_j > 0, \forall j \in \mathcal{J}$。由于 λ_j 的值不受联合机会约束 (3-1) 可行域的影响，能够提高近似的准确性。集合 $\mathcal{J}$ 选出了能够根据不确定集合的大小自动满足需求约束的设施点，减少了运算负担。在利用 Bonferroni 不等式进行近似时，近似后的独立机会约束共有 $|J|$ 个，分别对应第 j 个可用设施，即 $\mathbb{P}\left\{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{d} - N_j\right\} \geqslant 0, \forall j \in J$。而含参 SOCP 近似方法仅考虑了左侧概率最大情况下的唯一一个二次锥近似，即将在命题 3.3中提到的 $\mathbb{P}\{\max_{j\in\mathcal{J}}\left(\lambda_j\left[\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{d} - N_j\right]\right) \geqslant 0\}$。

命题 3.3 定义

$$\Upsilon(\boldsymbol{N},\boldsymbol{X},\boldsymbol{\lambda},\mathcal{J})\stackrel{\text{def}}{=}\min_{w_0,\boldsymbol{w}}\left\{\min_{\iota}\left[\iota+\frac{\pi(w_0+\iota,\boldsymbol{w})}{\varepsilon}\right]+\frac{1}{\varepsilon}\left[\sum_{j\in\mathcal{J}}\pi\left(\lambda_j N_j-w_0,\lambda_j\boldsymbol{X}_j-\boldsymbol{w}\right)\right]\right\}$$

则

$$\Upsilon(\boldsymbol{N},\boldsymbol{X},\boldsymbol{\lambda},\mathcal{J})\leqslant 0 \tag{3-22}$$

和

$$\max_{j\in\mathcal{W}\setminus\mathcal{J}}\left[\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{d}-N_j\right]\leqslant 0 \tag{3-23}$$

成立是联合机会约束 (3-6d) 成立的充分条件。

证明 与 Chen 等人[118] 在 2010 年提出的方法相似。当 $j\notin\mathcal{J}$ 时，由式 (3-23) 可得

$$\mathbb{P}\left(\mathbf{X}_j^{\mathrm{T}}\boldsymbol{d}-N_j>0\right)=0,\forall j\in\mathcal{W}\setminus\mathcal{J}$$

而在其他情况下，对于任意的 $\boldsymbol{\lambda}>0$，$\mathbb{P}\left(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{d}-N_j\leqslant 0,\forall j\in\mathcal{W}\right)$ 等价于仅考虑左侧函数中的最大值，即 $\mathbb{P}\{\max_{j\in\mathcal{J}}\left(\lambda_j\left[\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{d}-N_j\right]\right)\leqslant 0\}$。由命题 3.2可得, CVaR 约束 (3-6d) 是对独立机会约束最紧的近似方法，因而对于可行解 ($\boldsymbol{N}$, $\boldsymbol{X}$) 而言，我们仅需考虑 $\max_{j\in\mathcal{J}}\left[\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{d}-N_j\right]$ 对应的 CVaR 约束，即 $\varrho_{1-\varepsilon}[\max_{j\in\mathcal{J}}[\lambda_j\left(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{d}-N_j\right)]]\leqslant 0$。由期望的最大化公式[130] 可知,

$$\begin{aligned}&\mathbb{E}\left(\max_{i=1,2,\cdots,n}A_i-\iota\right)^{+}\\ \leqslant\ &\mathbb{E}(B-\iota)^{+}+\sum_{i=1}^{n}\mathbb{E}(A_i-B)^{+},\text{对于任意参数 } B \text{ 都成立}\end{aligned} \tag{3-24}$$

令 $B=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{d}-w_0$，可得

$$\begin{aligned}&\varrho_{1-\varepsilon}\left[\max_{j\in\mathcal{J}}\left[\lambda_j\left(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{d}-N_j\right)\right]\right]\\ =&\min_{\iota}\left\{\iota+\frac{1}{\varepsilon}\mathbb{E}\left[\left(\max_{j\in\mathcal{J}}\left[\lambda_j\left(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{d}-N_j\right)\right]-\iota\right)^{+}\right]\right\}\\ \leqslant&\min_{\iota,\boldsymbol{w},w_0}\left\{\iota+\frac{1}{\varepsilon}\Big[\mathbb{E}[(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{d}-w_0-\iota)^{+}]+\right.\end{aligned}$$

$$
\begin{aligned}
&\sum_{j\in\mathcal{J}}\mathbb{E}\left[\left(\left([\lambda_j\boldsymbol{X}_j-\boldsymbol{w}]^{\mathrm{T}}\boldsymbol{d}-\lambda_jN_j+w_0\right)^+\right]\right]\Bigg\}\\
\leqslant&\min_{\iota,\boldsymbol{w},w_0}\left\{\iota+\frac{1}{\varepsilon}\left[\pi(w_0+\iota,\boldsymbol{w})+\sum_{j\in\mathcal{J}}\pi(\lambda_jN_j-w_0,\lambda_j\boldsymbol{X}_j-\boldsymbol{w})\right]\right\}\\
=&\min_{w_0,\boldsymbol{w}}\left\{\min_{\iota}\left[\iota+\frac{\pi(w_0+\iota,\boldsymbol{w})}{\varepsilon}\right]+\frac{1}{\varepsilon}\left[\sum_{j\in\mathcal{J}}\pi\left(\lambda_jN_j-w_0,\lambda_j\boldsymbol{X}_j-\boldsymbol{w}\right)\right]\right\}\\
=&\Upsilon(\boldsymbol{N},\boldsymbol{X},\boldsymbol{\lambda},\mathcal{J})
\end{aligned}
$$

其中，第一个不等式根据式 (3-24) 可知，第二个不等式由引理 3.1可知。综上所述，若 $\Upsilon(\boldsymbol{N},\boldsymbol{X},\boldsymbol{\lambda},\mathcal{J})\leqslant 0$ 成立，则 $\varrho_{1-\varepsilon}\left[\max_{j\in\mathcal{J}}\left[\lambda_j\left(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{d}-N_j\right)\right]\right]\leqslant 0$ 一定成立。命题得证。 □

定义 $\varrho_{1-\varepsilon}[v(N_j,\boldsymbol{X}_j)]$ 的上界为 $\phi_{1-\varepsilon}(N_j,\boldsymbol{X}_j)$，根据式 (3-19)，$\phi_{1-\varepsilon}(N_j,\boldsymbol{X}_j)=\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j+\sqrt{\frac{1-\varepsilon}{\varepsilon}}\sqrt{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j}$。命题 3.3中的不等式 (3-22) 和不等式 (3-23) 构成了联合机会约束 (3-6d) 的二次锥近似。通过引入两个辅助决策变量 s_0 和 s_j，不等式 (3-22) 等价于如下三个约束：

$$s_0+\frac{1}{\varepsilon}\sum_{j\in\mathcal{J}}s_j\leqslant 0 \tag{3-25}$$

$$\phi_{1-\varepsilon}(w_0,\boldsymbol{w})\leqslant s_0 \tag{3-26}$$

$$\pi\left(\lambda_jN_j-w_0,\lambda_j\boldsymbol{X}_j-\boldsymbol{w}\right)\leqslant s_j,\forall j\in\mathcal{J} \tag{3-27}$$

另外，由 $\boldsymbol{D}$ 所在的不确定分布集的定义 (3-14) 和 Chen 等人在 2009 年发表的文章[131] 中的定理 3 可知，

$$\max\left[\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{d}-N_j\right]=\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j+Q\sqrt{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j} \tag{3-28}$$

因而，不等式 (3-23) 可以被表示为 $\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j+Q\sqrt{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j}\leqslant 0$。

综上所述，模型 P 可以被近似成带有两个参数 λ_j 和 $\mathcal{J}$ 二次锥规划问题，在后文中被简称为“RP-SOCP”。

$$\text{RP-SOCP}:\min\quad \beta(\epsilon r+\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{q})+\sum_{j\in J}\{f_jY_j+a_jN_j\} \tag{3-29}$$

$$\text{s.t.}\quad \boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j+Q\sqrt{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j}\leqslant 0,\forall j\in\mathcal{W}\setminus\mathcal{J} \tag{3-30}$$

式 (3-4), 式 (3-5), 式 (3-6b), 式 (3-6c), 式 (3-6e),

式 (3-10), 式 (3-11), 式 (3-12), 式 (3-25),

式 (3-26), 式 (3-27).

3.3　模型 RP-SOCP 的理论性质

定理 3.1 详细对比了含参 SOCP 近似方法与独立机会约束近似方法在系统稳定性方面的表现。

定理 3.1　对于所有的 $j\in\mathcal{J}$，不等式 (3-25)~ 不等式 (3-27) 比不等式 (3-17) 更紧。

证明　由于约束 (3-25)~ 约束 (3-27) 等价于 $\Upsilon(\boldsymbol{N},\boldsymbol{X},\boldsymbol{\lambda},\mathcal{J})\leqslant 0$，则

$$\begin{aligned}
0\geqslant&\min_{\iota,\boldsymbol{w},w_0}\left\{\iota+\frac{1}{\varepsilon}\left[\pi(w_0+\iota,\boldsymbol{w})+\sum_{j\in\mathcal{W}}\pi(\lambda_jN_j-w_0,\lambda_j\boldsymbol{X}_j-\boldsymbol{w})\right]\right\}\\
\geqslant&\min_{\iota,\boldsymbol{w},w_0}\left\{\iota+\frac{1}{\varepsilon}\left[\pi(w_0+\iota,\boldsymbol{w})+\pi(\lambda_jN_j-w_0,\lambda_j\boldsymbol{X}_j-\boldsymbol{w})\right]\right\}\\
\geqslant&\min_{\iota}\left\{\iota+\frac{1}{\varepsilon}\left[\pi(\lambda_jN_j+\iota,\lambda_j\boldsymbol{X}_j)\right]\right\}\\
\overset{\lambda_j=1}{=}&\min_{\iota}\left\{\iota+\frac{1}{\varepsilon}\left[\pi(N_j+\iota,\boldsymbol{X}_j)\right]\right\}\\
=&\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j+\sqrt{\frac{1-\varepsilon}{\varepsilon}}\sqrt{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j}
\end{aligned}$$

其中，由 3.3节中对于 $\Upsilon(\boldsymbol{N},\boldsymbol{X},\boldsymbol{\lambda},\mathcal{J})$ 的定义可知，第一个不等式成立；由于第二个不等式的右端值中去掉了最后一项的求和符号，该不等式显然成立；由引理 3.2可知，函数 $\pi(\cdot,\cdot)$ 具有可加性，第三个不等式成立；当 $\lambda_j=1$ 时，第一个等式成立；由式 (3-19)，最后一个不等式成立。　□

此外，当 MNCD 的不确定集合被限定在椭球集内时，在 RP-SOCP 中，与椭球集半径 Q 相关的命题 3.4、引理 3.4和定理 3.2成立。

命题 3.4　当 $Q \leqslant \sqrt{\dfrac{1-\varepsilon}{\varepsilon}}$ 时，RP-SOCP 在 $\mathcal{J}=\varnothing$ 时得到最优解。

证明　利用反证法证明此命题。不失一般性，假设在最优解中存在一个序号 $j^\circ \in \mathcal{J}$ 使 $\mathcal{J} \neq \varnothing$。对于所有 $j \in \mathcal{J}$ 而言，需保证 $\Upsilon(\boldsymbol{N},\boldsymbol{X},\boldsymbol{\lambda},\mathcal{J}) \leqslant 0$，同时

$$
\begin{aligned}
0 \geqslant \Upsilon(\boldsymbol{N},\boldsymbol{X},\boldsymbol{\lambda},j^\circ) \geqslant & \boldsymbol{X}_{j^\circ}^{\mathrm{T}}\boldsymbol{u} - N_{j^\circ} + \sqrt{\frac{1-\varepsilon}{\varepsilon}}\sqrt{\boldsymbol{X}_{j^\circ}^{\mathrm{T}}\boldsymbol{\varGamma}\boldsymbol{X}_{j^\circ}} \\
\geqslant & \boldsymbol{X}_{j^\circ}^{\mathrm{T}}\boldsymbol{u} - N_{j^\circ} + Q\sqrt{\boldsymbol{X}_{j^\circ}^{\mathrm{T}}\boldsymbol{\varGamma}\boldsymbol{X}_{j^\circ}}
\end{aligned}
$$

由定理 3.1可知，第一个不等式成立；由于 $Q \leqslant \sqrt{\dfrac{1-\varepsilon}{\varepsilon}}$，第二个不等式成立。因此，不等式 $\Upsilon(\boldsymbol{N},\boldsymbol{X},\boldsymbol{\lambda},j^\circ) \leqslant 0$ 比约束 $\boldsymbol{X}_{j^\circ}^{\mathrm{T}}\boldsymbol{u} - N_{j^\circ} + Q\sqrt{\boldsymbol{X}_{j^\circ}^{\mathrm{T}}\boldsymbol{\varGamma}\boldsymbol{X}_{j^\circ}} \leqslant 0$ 更为保守，也就是说，如果 $j^\circ \in \mathcal{W} \setminus \mathcal{J}$，即 $j^\circ \notin \mathcal{J}$，我们能找到一个更好的解。此结论与该解为最优解的推论矛盾，命题得证。 $\square$

命题 3.4表示，当 $Q \leqslant \sqrt{\dfrac{1-\varepsilon}{\varepsilon}}$ 时，RP-SOCP 等价于一个与参数 λ_j 和 $\mathcal{J}$ 不相关的标准 SOCP [132]。另外，在此条件下，这个标准的 SOCP 能够有效地被外逼近算法求解，详见 3.4节。

定义 Z^D，Z^B 和 Z^I 分别为模型 RP-SOCP、基于 Bonferroni 近似的联合机会约束模型，以及基于 CVaRD 的独立机会约束近似模型的最优（最小）成本，则如下定理成立。

定理 3.2　Z^B，Z^I 和 Z^D 的大小关系为

（1） 如果 $Q < \sqrt{\dfrac{1-\varepsilon}{\varepsilon}}$，则 $Z^D < Z^I$；

（2） 如果 $Q = \sqrt{\dfrac{1-\varepsilon}{\varepsilon}}$，则 $Z^D = Z^I$；

（3） 如果 $Q > \sqrt{\dfrac{1-\varepsilon}{\varepsilon}}$，则 $Z^I < Z^D < Z^B$。

证明　定理按照如下三种情况展开证明，

（1）如果 $Q < \sqrt{\dfrac{1-\varepsilon}{\varepsilon}}$，由命题 3.4可知 $\mathcal{J}=\varnothing$，则 $\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u} - N_j +$

$Q\sqrt{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j} < \boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u} - N_j + \sqrt{\dfrac{1-\varepsilon}{\varepsilon}}\sqrt{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j} \leqslant 0$，$\forall j \in \mathcal{W}$，即第一个约束成立，则第二个约束一定成立，因而模型 RP-SOCP 是独立机会约束模型的松弛形式，对于一个最小化问题来说，$Z^D < Z^I$。

（2）当 $Q = \sqrt{\dfrac{1-\varepsilon}{\varepsilon}}$ 时，证明过程与第一种情况类似。

（3）当 $Q > \sqrt{\dfrac{1-\varepsilon}{\varepsilon}}$ 时，如果 $j \in \mathcal{J}$，根据定理 3.1，不等式 (3-25)～ 不等式 (3-27) 比约束 (3-17) 更紧。当 $j \in \mathcal{W} \setminus \mathcal{J}$ 时，$\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u} - N_j + Q\sqrt{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j} \leqslant 0$ 是约束 (3-17) 成立的充分条件，RP-SOCP 的可行域比独立机会约束限制下更紧，则 $Z^D > Z^I$。不等式的另一边由 Chen 等人[118] 的论文中的定理 3.2 中给出，即 $Z^D < Z^B$。 □

3.4 求解方法

对于含参 SOCP 问题，本书提出改进的参数迭代优化算法求解，详见 3.4.1节。该算法在求解大规模问题时仍然十分困难。根据 3.3节的性质，本书采用外逼近算法对特殊情况下的算例加速求解。

3.4.1 改进的参数迭代算法

由于 $\Upsilon(\boldsymbol{N}, \boldsymbol{X}, \boldsymbol{\lambda}, \mathcal{J})$ 并非关于 $\boldsymbol{Y}$，$\boldsymbol{N}$ 和 $\boldsymbol{\lambda}$ 的凸函数，求解命题 3.3所述问题的最大困难在于如何找到合适的 $\boldsymbol{\lambda}$ 和 $\mathcal{J}$。幸运的是，当 $\boldsymbol{Y}$ 和 $\boldsymbol{N}$ 的值固定时，我们可以通过迭代的方法改进当前得到的 SOCP，并找到使模型 RP-SOCP 可行的参数 $\boldsymbol{\lambda} > 0$ 和集合 $\mathcal{J} \subseteq \mathcal{W}$[118]。为了改进当前目标函数，我们需要找到模型 RP-SOCP 中更大的松弛变量。定义 $\mathcal{H}(\boldsymbol{X}, \boldsymbol{N}) = \left\{j : \max\limits_{j \in \mathcal{W}} \left[\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{d} - N_j\right] > 0\right\}$，目标函数 (3-29) 能够通过调整 λ_j 和 $j \in \mathcal{H}(\boldsymbol{X}, \boldsymbol{N})$ 的方式不断改进。对于 RP-SOCP 的一个可行解 $(\boldsymbol{X}, \boldsymbol{N})$ 来说，我们可以通过求解模型 (3-31) 调整集合 $\mathcal{H}(\boldsymbol{X}, \boldsymbol{N})$。

$$\min_{t} \quad \sum_{j=1}^{J} t_j,$$

$$
\begin{aligned}
\text{s.t.} \quad & \boldsymbol{X}_j^{\mathrm{T}} \boldsymbol{u} - N_j + Q\sqrt{\boldsymbol{X}_j^{\mathrm{T}} \boldsymbol{\Gamma} \boldsymbol{X}_j} \leqslant t_j, \forall j \in \mathcal{W}, \\
& t \in \mathbb{R}
\end{aligned} \tag{3-31}
$$

在得到模型 (3-31) 的最优解后，更新集合 $\mathcal{H}(\boldsymbol{X},\boldsymbol{N}) = \{j : t_j^* > 0\}$。如果集合 $\mathcal{H}(\boldsymbol{X},\boldsymbol{N})$ 是非空的，最优的 $\boldsymbol{\lambda}^*$ 可通过求解模型 (3-32) 来获得。

$$
\begin{aligned}
\min_{\lambda} \quad & \Upsilon(\boldsymbol{N},\boldsymbol{X},\boldsymbol{\lambda},\mathcal{H}), \\
\text{s.t.} \quad & \sum_{j \in \mathcal{H}(\boldsymbol{X},\boldsymbol{N})} \lambda_j = 1, \\
& \lambda_j \geqslant 0,\ \forall j \in \mathcal{H}(\boldsymbol{X},\boldsymbol{N})
\end{aligned} \tag{3-32}
$$

迭代改进参数 $\boldsymbol{\lambda}$ 和 $\mathcal{H}(\boldsymbol{X},\boldsymbol{N})$ 的方法如算法 1 所示。

算法 1 求解 RP-SOCP 的迭代算法

输入： TC：模型 RP-SOCP 的目标函数，初始解为 $\mathrm{TC}^1 = 0$

$\mathcal{H}(\boldsymbol{X},\boldsymbol{N})$：全集 $\mathcal{W}$ 的一个子集，且 $\mathcal{H}^1(\boldsymbol{X},\boldsymbol{N}) = \mathcal{W}$

$\boldsymbol{\lambda}$：参数向量，初始解为 $\lambda_j^1 = 1/|J|, \forall j \in \mathcal{H}^1(\boldsymbol{X},\boldsymbol{N})$

τ：一个比较小的数

K：最大迭代次数

算法流程：

```
for k = 1 : K do
    输入 λ^k 和 H^k(X,N) 求解模型 RP-SOCP。得到最优解 (Y*, X*, N*) 和
    最小的目标函数 TC*。令 (Y^k, X^k, N^k)=(Y*, X*, N*)，TC^{k+1}=TC*
    if TC^{k+1} − TC^k ⩽ τ 或者 H^k(X,N) = ∅ then
        跳出循环
    end if
    固定 Y^k, X^k, N^k，找到模型 (3-31) 的最优解 t*，H^k(X,N)，λ^k
    令 H^{k+1}(X,N) := {j|t_j* > 0, j ∈ W}
    固定 Y^k, X^k, N^k, H^{k+1}(X,N)，求解模型 (3-32)。得到最优的 λ*，令 λ^{k+1} =
    λ*
end for
```

尽管在一般情况下，算法 1需要反复迭代才能得到最终的最优解，但当 $Q \leqslant \sqrt{\dfrac{1-\varepsilon}{\varepsilon}}$ 时，仅需两次迭代即可跳出循环、达到最优，如引理 3.3 所示。

引理 3.3 当 $Q \leqslant \sqrt{\dfrac{1-\epsilon}{\epsilon}}$ 时，算法 1在第二次迭代时停止，终止的条件为 $\mathcal{H}^2(\boldsymbol{X},\boldsymbol{N})=\varnothing$。

证明 在算法 1中，由于在第一次迭代时，集合 $\mathcal{H}^1(\boldsymbol{X},\boldsymbol{N})=\mathcal{W}$，根据定理 3.1，对于所有的 $j\in\mathcal{W}$，约束 (3-25)~ 约束 (3-27) 比约束 (3-17) 更紧。又因为 $Q \leqslant \sqrt{\dfrac{1-\varepsilon}{\varepsilon}}$，可知

$$\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j+Q\sqrt{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\varGamma}\boldsymbol{X}_j}\leqslant \boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j+\sqrt{\frac{1-\varepsilon}{\varepsilon}}\sqrt{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\varGamma}\boldsymbol{X}_j}\leqslant 0$$

因此，约束 (3-30) 相较于 (3-17) 冗余，即模型 (3-30) 在仅含有式 (3-25) ~ 式 (3-27) 时仍然成立。由于模型 (3-31) 计算了模型 (3-30) 的松弛情况，对于冗余约束来说，相应的松弛变量为 0（$t_j\leqslant 0,\forall j\in\mathcal{W}$），也就是说 $\mathcal{H}^2(\boldsymbol{X},\boldsymbol{N})=\varnothing$，达到了跳出条件，算法结束。引理得证。 □

引理 3.4 当 $Q \leqslant \sqrt{\dfrac{1-\varepsilon}{\varepsilon}}$ 时，RP-SOCP 等价于如下模型 RP-1。

$$\begin{aligned}
\text{RP-1}: \quad \min \quad & \beta(\epsilon r+\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{q})+\sum_{j\in J}\{f_jY_j+a_jN_j\}, \\
\text{s.t.} \quad & \text{式 (3-4)}, \text{式 (3-5)}, \text{式 (3-6b)}, \text{式 (3-6c)}, \\
& \text{式 (3-6e)}, \text{式 (3-10)}, \text{式 (3-11)}, \text{式 (3-12)}, \\
& \text{式 (3-30)},
\end{aligned} \tag{3-33}$$

证明 根据引理 3.3，当 $Q \leqslant \sqrt{\dfrac{1-\varepsilon}{\varepsilon}}$ 时，算法 1在第二次迭代时触发跳出条件 $\mathcal{H}^2(\boldsymbol{X},\boldsymbol{N})=\varnothing$ 结束。此条件显示约束 (3-25) ~ 约束 (3-27) 是无效的。因而，当模型 RP-SOCP 把无效约束都去掉之后等价于模型 RP-1。 □

3.4.2 求解模型 RP-1 的外逼近算法

外逼近（OA）算法是求解切平面问题的一类经典算法，最初由 Duran 和 Grossmann 在 1987 年提出[133]，主要用来求解混合整数非线性规划问题（mixed-integer nonlinear programs，MINLP），该算法的详细介绍在附录 A给出。在求解 RP-1 时，我们首先在引理 3.5中证明该问题的线性松弛的凸性，再进一步设计算法求解。

引理 3.5　当 $\boldsymbol{Y}$ 和 $\boldsymbol{N}$ 为连续变量时，RP-1 是凸的。

证明　除了非线性约束 (3-11) 和约束 (3-30) 之外，RP-1 是一个完整的线性规划，因此，我们只需证明非线性项 (3-11) 和约束 (3-30) 的凸性。定义

$$\Phi(\boldsymbol{q}, r) = \sqrt{\boldsymbol{q}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{q}} - r \tag{3-34a}$$

$$\Psi(\boldsymbol{X}_j, \boldsymbol{N}_j) = Q\sqrt{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{X}_j} + \boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u} - N_j \tag{3-34b}$$

与 Shahabi 等人在 2014 年[134] 的论文相似。$\Phi(\boldsymbol{q}, r)$ 和 $\Psi(\boldsymbol{X}, \boldsymbol{N})$ 是凸函数，引理得证。 □

3.4.2.1　初始化

初始化的目的在于找到原问题的一个可行解。首先，我们假设具有最低建造成本的设施被建立，即 $Y_{j^*}^0 = 1$, $j^* = \{j|f_j = \min_{s\in J} f_s\}$。然后，对于所有不能在规定时长内被所建立的设施覆盖的需求点来说，建立其周围的一个设施来满足该点需求，即 $Y_{j^\circ}^0 = 1$, $j^\circ = \{j|I \setminus I_{j^*}\}$。最后，根据约束 (3-6b) 和约束 (3-6c)，$X_{ij}^0 = 1$ if $j = j^*$且$j = j^\circ, \forall i \in I$；根据式 (3-30), 可得相应的救护车数量为 $N_j = \lceil {\boldsymbol{X}_j^0}^{\mathrm{T}}\boldsymbol{u} + Q\sqrt{{\boldsymbol{X}_j^0}^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j^0}\rceil$。

3.4.2.2　OA 子问题

子问题 SP 的主要任务是在整数变量值被固定时，找到其他连续变量的最佳取值。在第 h 次迭代时，输入整数变量值 $\tilde{Y}_j^h$ 和 $\tilde{N}_j^h$，对应的子问题为

$$\begin{aligned}
\mathrm{SP}: \min \quad & \beta(\epsilon r + \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{q}) + \sum_{j\in J}\left\{f_j\tilde{Y}_j^h + a_j\tilde{N}_j^h\right\}, \\
\text{s.t.} \quad & \boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u} - \tilde{N}_j^h + Q\sqrt{\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{X}_j} \leqslant 0, \forall j \in \mathcal{W}, \\
& X_{ij} \leqslant \tilde{Y}_j^h, \forall i \in I, \forall j \in J \\
& \text{式 (3-6b), 式 (3-6e), 式 (3-10), 式 (3-11), 式 (3-12)}
\end{aligned}$$

3.4.2.3　OA 主问题

OA 算法主问题 MP 为混合整数线性规划问题（mixed-integer linear program，MILP），MP 通过 SP 的最优连续变量 $\tilde{r}^h$, $\tilde{\boldsymbol{q}}^h$, and $\tilde{\boldsymbol{X}}^h$，对

非线性约束 (3-11) 和约束 (3-30) 加切，切的具体表达形式在命题 3.5中表示。

命题 3.5 在第 h 次迭代中，非线性约束 (3-11) 和非线性约束 (3-30) 的 OA 切为

$$\boldsymbol{q}^{\mathrm{T}}\boldsymbol{\Sigma}\tilde{\boldsymbol{q}}^{h}-r\tilde{r}^{h}\leqslant 0 \tag{3-35a}$$

$$\left(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{u}-N_j\right)\sqrt{\tilde{\boldsymbol{X}}_j^{h}{}^{\mathrm{T}}\boldsymbol{\Gamma}\tilde{\boldsymbol{X}}^{h}{}_j}+Q\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\Gamma}\tilde{\boldsymbol{X}}_j^{h}\leqslant 0 \tag{3-35b}$$

证明 由引理 3.5可知，$\Phi(\boldsymbol{q},\boldsymbol{r})$ 和 $\Psi(\boldsymbol{X}_j,\boldsymbol{N}_j)$ 具有凸性，对 $\Phi(\boldsymbol{q},\boldsymbol{r})$ 和 $\Psi(\boldsymbol{X}_j,\boldsymbol{N}_j)$ 进行一阶泰勒展开，可得 $\Phi(\tilde{\boldsymbol{q}}^h,\tilde{\boldsymbol{r}}^h)+\nabla\Phi(\tilde{\boldsymbol{q}},\tilde{\boldsymbol{r}})\left[\boldsymbol{q}-\tilde{\boldsymbol{q}}^h,r-\tilde{r}^h\right]^{\mathrm{T}}\leqslant\Phi(\boldsymbol{q},\boldsymbol{r})\leqslant 0$ 和 $\Psi(\tilde{\boldsymbol{X}}_j^h,\tilde{N}_j^h)+\nabla\Psi(\tilde{\boldsymbol{X}}_j^h,\tilde{\boldsymbol{N}}_j^h)[\boldsymbol{X}_j-\tilde{\boldsymbol{X}}_j^h,N_j-\tilde{N}_j^h]^{\mathrm{T}}\leqslant\Psi(\boldsymbol{X}_j,N_j)\leqslant 0$，其中 $\nabla\Phi(\tilde{\boldsymbol{X}}_j^h,\tilde{\boldsymbol{N}}_j^h)=\left[\dfrac{(\tilde{\boldsymbol{q}}^h)^{\mathrm{T}}\boldsymbol{\Gamma}}{\sqrt{(\tilde{\boldsymbol{q}}^h)^{\mathrm{T}}\boldsymbol{\Gamma}\tilde{\boldsymbol{q}}^h}},-1\right]$，$\nabla\Psi(\tilde{\boldsymbol{X}}_j^h,\tilde{\boldsymbol{N}}_j^h)=\left[\boldsymbol{\mu}^{\mathrm{T}}+\dfrac{Q(\tilde{\boldsymbol{X}}_j^h)^{\mathrm{T}}\boldsymbol{\Gamma}}{\sqrt{\tilde{\boldsymbol{X}}_j^h{}^{\mathrm{T}}\boldsymbol{\Gamma}\tilde{\boldsymbol{X}}_j^h}},-1\right]$。由于 $\tilde{r}^h=(\tilde{\boldsymbol{q}})^{\mathrm{T}}\boldsymbol{\Gamma}\tilde{\boldsymbol{q}}^h$，通过简单的代数处理，OA 切的数学表达分别为式 (3-35a) 和式 (3-35b)，命题得证。 □

OA 主问题汇总如下：

$$\begin{aligned}
\mathrm{MP}:\quad \min\quad & \eta, \\
\text{s.t.}\quad & \eta\geqslant\beta(\epsilon r+\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{q})+\sum_{j\in J}\{f_jY_j+a_jN_j\} && (3\text{-}36)\\
& \eta\leqslant \mathrm{UB}^h-\varepsilon,\forall h && (3\text{-}37)\\
& \text{式 (3-3), 式 (3-6b), 式 (3-6c), 式 (3-6e),}\\
& \text{式 (3-10), 式 (3-11), 式 (3-12), 式 (3-35a), 式 (3-35b)}
\end{aligned}$$

其中，式 (3-36) 定义了目标函数，约束 (3-37) 保证 MP 的最优值不会超过前 h 次迭代中获得的上界 (UB^h)。在这里，我们采用由 Fletcher 等人[135] 提出的 ε-最优的方法进行求解。当 MP 不可行时，停止迭代。具体算法流程见算法 2。

算法 2 OA 算法

输入： $\tilde{Y}_j^0$ 和 $\tilde{N}_j^0$：整数决策变量 Y_j 和 N_j 的初始解

LB^0：原问题下界，等于 $-\infty$

UB^0: 原问题上界，等于 ∞

$\mathbb{K}$：最大迭代次数

算法流程：

1: **for** $h = 1 : \mathbb{K}$ **do**

2: 求解 SP。得到最优解 $\tilde{X}_{ij}^h$，$\tilde{q}_i^h$ 和 $\tilde{r}^h$，令 SP 的最优目标函数值为问题上界 UB^h

3: 将 SP 求得的 $\tilde{X}_{ij}^h$，$\tilde{q}_i^h$ 和 $\tilde{r}^h$ 代入式 (3-35a) 和式 (3-35b)，建立 OA 切，求解 MP，求得 $\tilde{Y}_j^h$ 和 $\tilde{N}_j^h$，令 MP 的目标函数为当前问题下界 LB^h

4: **if** MP 不可行 **then**

5: 停止迭代，返回当前值

6: **end if**

7: **end for**

3.5 数值实验

本节将通过一系列数值实验验证模型的有效性和可靠性，并对联合机会约束的 Bonferroni 近似、独立机会约束的经典近似方法、基于场景的随机规划方法（详见 B.1）和本章提出的分布式鲁棒优化模型四种方法进行系统比较。另外，本节还验证了 OA 算法的有效性。值得注意的是，除了独立机会约束的经典近似方法外，其他三种模型均是针对联合机会约束的近似，联合机会约束能够保证整个系统的稳定性，而独立机会约束仅仅是满足每个需求点的需求。不同方法对应的最优网络结构将会在3.5.3节展示。DRM 的系统稳定性将会在 3.5.4节展示。基于实证数据的数值结果将会在 3.5.5节展示。

所有数值实验均是在具有 64 位 3.4-GHz Intel Core i9 处理器、内存为 32GB 且运行 Windows 10 操作系统的计算机上进行的，计算时间以秒为单位报告。求解器 Mosek 8.0.0.79[136] 被直接应用于求解模型 RP-SOCP，RP-1，以及 OA 算法的 MP 和 SP，跳出条件为最优间隙（optimality gap）小于 0.01。所有代码均是用 MATLAB 编程实现，且应用 YALMIP 工具包[137] 作为 MATLAB 和 Mosek 的媒介。

3.5.1 性能分析

根据 I，J 和 Q 的不同取值，本节共考虑 25 组不同规模的算例，其中 $I = J$ 的可能取值为 10，15，20，25，30，Q^2 的可能取值为 10，19，

30，40，50，对于不同问题规模，共随机生成五组算例。在基于场景的算法中，随机变量 $\boldsymbol{\Theta}$ 和 $\boldsymbol{D}$ 各生成 20 个可能取值。

在进行数值实验时，我们将整个系统限定在一个 10×10 的方框内，随机生成需求点和可能的设施位置。参数选择与 Zhang 和 Li 在 2015 年的研究[55] 类似。

f_j 在 [25,75] 区间内随机生成；

a_j 在 [1,3] 区间内随机生成；

β 等于 5；

α 等于 0.95；

Θ_i 在公式 (3-7) 定义的分布集内随机生成，$\boldsymbol{\Sigma}$ 是半正定矩阵，$\boldsymbol{\mu}$ 在 [0.1,5] 区间内均匀生成，σ_i 在 [0.5,1.5] 均匀生成；日常需求的协方差系数为 $\rho_{ij}^{\Theta} = 0.1$，$\forall i \neq j$ $\&\rho_{ij}^{\Theta} = 1$，$\forall i = j$；

D_i 在公式 (3-14) 定义的分布集中均匀生成，$\boldsymbol{\Gamma}$ 是半正定矩阵，$\boldsymbol{u}$ 在 [0.1,10] 区间内随机生成，γ_i 在 [0,2] 随机生成；MNCD 的协方差系数为 $\rho_{ij}^{D} = 0.1$，$\forall i \neq j$ $\&\rho_{ij}^{D} = 1$，$\forall i = j$。

根据公式 (3-7)，我们采用数据驱动的方式构造分布集 $\mathcal{F}$[22-23]，并按照多元正态分布，从整体中随机生成 1000 个日常需求模拟真实历史数据，再基于这些历史数据依照数据驱动方法准备输入参数和分布集合、完成数值实验。首先，计算随机变量 $\boldsymbol{\Theta}$ 的样本均值和样本方差，分别记做 $\hat{\boldsymbol{\mu}}$ 和 $\boldsymbol{\Sigma}_0$。然后，根据式 (3-7)，令 $\delta = 0.05$，$R^2 = \max\limits_{m=1,2,\cdots,M} (\boldsymbol{\Theta}_m - \hat{\boldsymbol{\mu}})^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\Theta}_m - \hat{\boldsymbol{\mu}})$，$\epsilon = (R/\sqrt{M})[2 + \sqrt{2\ln(1/\delta)}]$，估计随机变量均值 $\mathbb{E}(\boldsymbol{\Theta})$ 的分布集合。分布集 $\mathcal{G}$ 是半径为 Q（$Q = \sqrt{10}, \sqrt{19}, \sqrt{30}, \sqrt{40}, \sqrt{50}$）的椭球集。最后，令 $\tau = 0.01$，$K = 30$，实现算法 1。

我们比较了不同问题规模下五组算例的平均目标函数（最优成本）。四种算法——基于场景的建模方法、本章提出的分布式鲁棒优化方法、Bonferroni 近似方法和独立机会约束近似方法，分别被简称为“Scb”“DRM”“Bon”和“Ind”。基于场景的建模方法将在附录 B.1中详细介绍。独立机会约束的近似方法是公式 (3-17) 中提到的基于 CVaR 的近似方法。为了使比较结果更为明确，表 3.1 以比例的形式进行展示，即表中结果为其他算法的目标函数值除以独立机会约束近似下的目标函数值。

从表 3.1中不难看出，当随机变量 $\boldsymbol{D}$ 所在的不确定椭球集 (3-14) 的半

径减小时，我们对随机参数的估计值更为准确，因而可以获得更小的总成本。根据定理 3.2，当 $\varepsilon = 0.05$，$Q^2 = (1-\varepsilon)/\varepsilon = 19$ 时，$Z^I = Z^D$，表 3.1中该列比值总是为 1，实验结果与理论证明结果一致。当 $Q^2 < (1-\varepsilon)/\varepsilon$ 或 $Q^2 > (1-\varepsilon)/\varepsilon$ 时，数值实验结果仍与理论一致。另外，DRM 对应的目标函数值远小于 Bon，有效克服了 Bonferroni 近似带来的过保守性。由于 ScB 仅仅考虑了分布集中的部分样本，并不一定将最坏情况纳入优化范畴，因而其对应的目标函数在理论上应该小于或等于鲁棒优化模型。与独立机会约束相比，联合机会约束更加强调系统的稳定性，因而会带来成本的提高，与稳定性相关的比较将会在 3.5.4节详细阐述。

表 3.1　ScB，DRM，Bon 和 Ind 四种算法的目标函数值比较结果

$\|I\| = \|J\|$	ScB					DRM					Bon	Ind
	10	19	30	40	50	10	19	30	40	50		
10	0.97	0.97	0.99	0.99	1.00	0.98	1.00	1.01	1.02	1.03	1.13	1.00
15	0.92	0.93	0.94	0.94	0.96	0.96	1.00	1.12	1.15	1.18	1.38	1.00
20	0.98	0.99	0.99	1.00	1.00	0.98	1.00	1.02	1.04	1.05	1.28	1.00
25	0.83	0.84	0.84	0.85	0.85	0.96	1.00	1.03	1.06	1.08	1.51	1.00
30	0.82	0.82	0.83	0.83	0.84	0.97	1.00	1.03	1.05	1.08	1.48	1.00

表 3.2记录了四种算法运算时间的均值和标准差（括号内的值），其中 Bon 能在最短的时间内得到问题的最优解，DRM 的运行时间最长。然而，本节研究的选址问题为战略优化、而非实时优化决策，运算效率并非最重要的因素。在问题规模较小（I 和 J 较小）时，随着分布集半径 Q 的增加，DRM 的时间无显著差异；当问题规模较大（I 和 J 较大）且 $Q^2 \geqslant 19$ 时，运行时间会随着分布集合半径 Q 的增大而减小。ScB 算法的运行时间无明显趋势。

表 3.2　运行时间汇总（以秒为单位）

$\|I\| = \|J\|$	Bon	ScB				
		$Q^2 = 10$	$Q^2 = 19$	$Q^2 = 30$	$Q^2 = 40$	$Q^2 = 50$
10	1.10	3.22	3.51	3.44	3.57	5.04
	(0.36)	(0.39)	(0.41)	(0.45)	(0.72)	(0.88)
15	6.38	17.80	20.94	19.80	20.49	22.06
	(4.41)	(4.95)	(8.15)	(7.19)	(4.37)	(8.14)

续表

$\lvert I\rvert=\lvert J\rvert$	Bon	ScB				
		$Q^2=10$	$Q^2=19$	$Q^2=30$	$Q^2=40$	$Q^2=50$
20	59.29 (19.30)	78.57 (23.98)	73.80 (22.69)	85.30 (31.74)	72.65 (18.53)	79.70 (29.76)
25	141.01 (108.64)	280.97 (120.23)	285.95 (75.98)	305.04 (134.40)	265.17 (98.60)	314.62 (114.05)
30	1070.99 (521.81)	705.25 (160.50)	656.45 (81.58)	588.97 (122.75)	669.97 (113.47)	727.36 (138.55)
$\lvert I\rvert=\lvert J\rvert$	Ind	DRM				
		$Q^2=10$	$Q^2=19$	$Q^2=30$	$Q^2=40$	$Q^2=50$
10	1.23 (0.37)	15.64 (7.84)	15.56 (13.57)	17.56 (17.72)	14.90 (12.59)	14.89 (12.70)
15	11.19 (6.56)	79.09 (36.19)	75.01 (31.51)	87.98 (25.36)	87.73 (25.37)	84.68 (25.90)
20	121.72 (39.43)	285.20 (24.48)	278.42 (35.80)	262.60 (27.63)	253.34 (28.01)	248.22 (26.95)
25	502.89 (258.29)	1580.35 (341.25)	1430.90 (226.73)	1369.63 (199.38)	1352.11 (204.21)	1466.92 (370.54)
30	7553.39 (4829.04)	12403.49 (4819.81)	11594.63 (5143.79)	10967.14 (3562.55)	9826.28 (3046.06)	9166.01 (3009.20)

表 3.3记录了在 $Q^2 \leqslant \sqrt{\dfrac{1-\epsilon}{\epsilon}}$ 时，OA 算法的运算效率。在 $Q^2=10$ 和 $Q^2=19$ 时，针对不同的 $|I|$ 和 $|J|$ 规模，随机生成 5 组算例，记录三种算法运行的平均时间。列“DRM”表示直接利用算法 1求解模型 RP-SOCP 所需的时间，列“Mosek”表示直接用 Mosek 求解模型 RP-1 所需的时间，列“OA”记录了利用 OA 算法求解模型 RP-1 的时间。假设最高运行时间为 14400s，表 3.3中有 * 标记的数值代表在 14400s 内该算法并未得到最优解。

表 3.3显示，OA 算法的运算效率显著优于其他两种方法。尽管列“Mosek”和列“DRM”所求解的模型（分别为 RP-SOCP 和 RP-1）并不一致，若能在规定时间内完成计算，二者总能得到相同的最优目标值，进一步验证了模型在特殊情况下的理论性质：引理 3.3和引理 3.4。

表 3.3　OA 算法运行效率对比（以秒为单位）

$\lvert I\rvert=\lvert J\rvert$	$Q^2=10$			$Q^2=19$		
	DRM	Mosek	OA	DRM	Mosek	OA
10	15.64	0.31	0.14	15.56	0.29	0.14
15	79.09	1.07	0.30	75.01	1.14	0.32
20	285.2	6.33	1.39	278.42	2.28	0.76
25	1580.35	40.25	6.12	1430.90	37.11	5.69
30	12403.49	1216.15	17.92	11594.63	197.52	6.62
35	14400.00*	7296.62	74.80	14400.00*	7358.82	69.84
40	14400.00*	14400.00*	396.03	14400.00*	14400.00*	401.43
45	14400.00*	14400.00*	742.57	14400.00*	14400.00*	1075.07
50	14400.00*	14400.00*	4874.18	14400.00*	14400.00*	4835.95

3.5.2　灵敏度分析

本节主要对单位运输成本 (β) 和 MNCD 的均值 ($\boldsymbol{u}$) 进行灵敏度分析。在分析时，需控制其他参数保持一致。图 3.1和图 3.2显示了总成本（虚线）、建立的设施数（条形图数值）和救护车总数（括号内的值）。

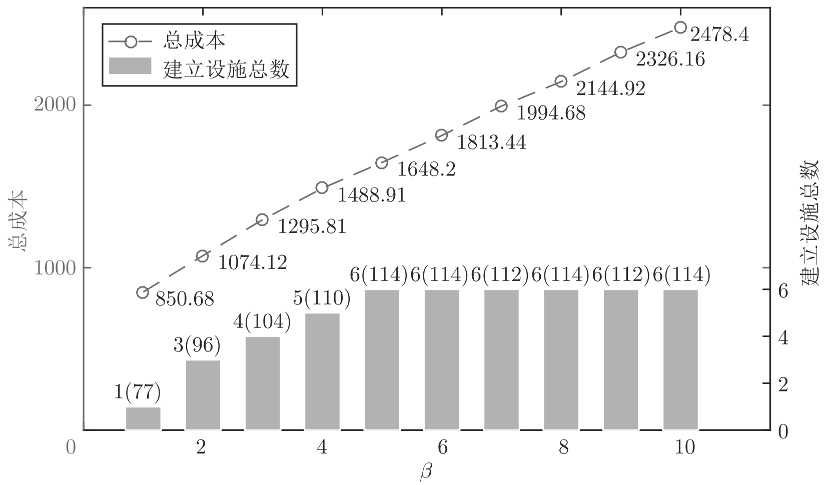


图 3.1　运输成本 (β) 的灵敏度分析

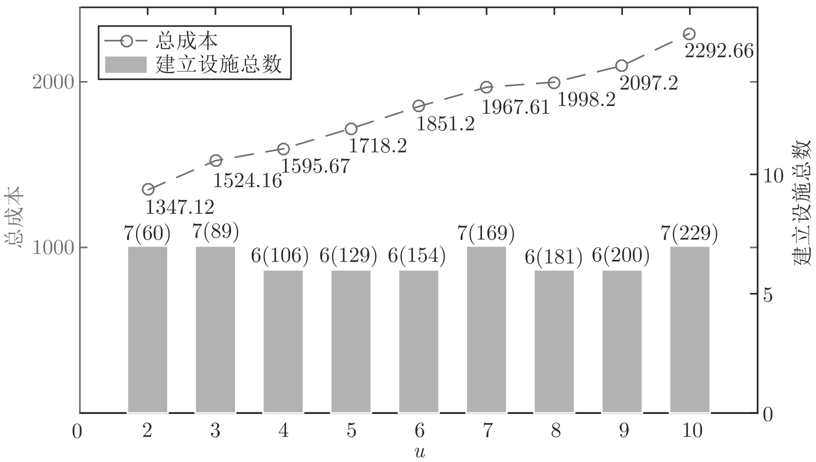


图 3.2 MNCD 均值 (u) 的灵敏度分析

3.5.2.1 对单位运输成本 (β) 的灵敏度分析

由图 3.1所示，系统总成本、救护车总数和建立的设施数均会随着 β 的提高而增加。呈现这种递增趋势的原因是，在单位运输成本增加时，决策者需要通过建立更多的急救中心或者配备更为充足的救援资源来达到降低运输成本的目的。

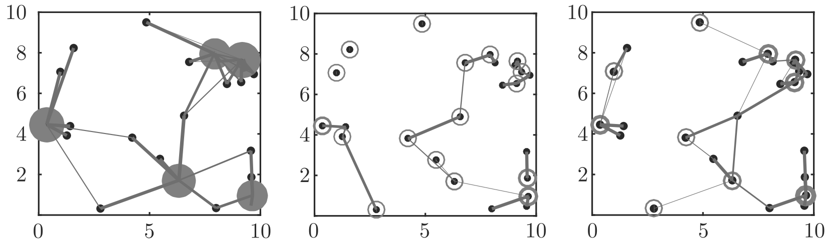
3.5.2.2 对 MNCD 均值 (u) 的灵敏度分析

本节考虑 MNCD 均值 (u) 对于最优决策的影响。图 3.2显示，最优条件下的总成本和救护车数量会随着 u 的增大而增大；而建设的设施数量大致保持不变。结果显示，当紧急情况下同时发生的需求数增大时，管理者可以通过加大备货量、购置救援设备的方法应对库存风险。

3.5.3 拓扑结构分析

在本节中，我们比较了联合机会约束的三种近似方法对于网络拓扑结构的影响，即 Bon，Scb 和 DRM。首先，在 10×10 的方格中生成 25 个随机节点，所有节点均可作为可选设施或需求点；然后，按照 3.5.1节中的定义生成其他参数。图 3.3展示了一组最优的拓扑结构，其中黑点代表随机生成的 25 个节点、红圈代表建立的设施位置、红圈的边框粗度代

表该点储藏的救护车数量、蓝线代表需求的匹配结果（线的粗度代表需求分配比例）。



（a） Bon　　（b） ScB　　（c） DRM

图 3.3　算法 Bon、ScB 和 DRM 的拓扑结构分析（前附彩图）

为了定量化比较几种模型的拓扑结构，定义单位设施对应的平均边数（M#EF）和单位设施对应的平均需求（M#DF）作为设施提供服务程度的判别指标；同时定义需求点连接的平均设施数（M#CF）和各个需求点的平均服务距离（MTD）作为需求点接受服务程度的判别指标，见表 3.4。

表 3.4　指标 M#EF，M#DF，M#CF 和 MTD 的具体定义

设施提供服务程度的判别指标	M#EF	M#DF
定义	$\sum_j \sum_i \mathcal{I}\{X_{ij}>0\} \Big/ \sum_j Y_j$	$\sum_j \sum_i [d_i X_{ij}] \Big/ \sum_j Y_j$
需求点接受服务程度的判别指标	M#CF	MTD
定义	$\sum_j \sum_i \mathcal{I}\{X_{ij}>0\} / I$	$\sum_j \sum_i [c_{ij} X_{ij}] / I$

表 3.5记录了系统总成本（TC）、建立的设施总数（TY）和所需救护车的总量（TN），M#EF，M#DF，M#CF 和 MTD，通过对比分析，我们可以观测到如下几个现象：

（1）基于随机规划的 ScB 方法倾向于建设更多的设施和储备更少的救护车，而 Bonferroni 近似方法的结论则恰恰与之相反。本节提出的分布式鲁棒优化方法 DRM 是两种方法的中和，DRM 倾向于建造数量适中的设施数量，并通过匹配适宜数量的救护车来满足各个需求点的需求，它集成了随机规划和 Bonferroni 近似的优点，使决策方案不致太过极端。

（2）在评估急救中心工作负担时，Bonferroni 近似方法要求每个设施所连接的平均需求点数和平均需求量分别为 8.0 和 10.8，显著高于 ScB 方法和 DRM 方法。而在大规模紧急情况发生时，设施中断和失效时有发生，由 Bonferroni 近似给出的过于集中的最优拓扑结构，更易受到中断风险的影响，而 ScB 和 DRM 则在一定程度上减轻了中断风险造成的损失。

（3）在评估各个需求点的服务水平时，三种算法表现相似。各个需求点所连接的设施平均值大致相等，且大多由多个设施服务同一需求点，这种现象表明，在大规模灾难发生、面临设施中断服务风险时，多重匹配原则能够在运输成本层面最大限度地降低需求未被满足的概率，ScB 和 DRM 对应的加权运输成本分别为 Bon 的 31.85% 和 47.46%，也就是说，前两种模型能在更短的时间内满足需求、提高服务质量。

表 3.5 模型 Bon，ScB 和 DRM 的拓扑结构比较

	总成本	TY	TN	M#EF	M#DF	M#CF	MTD
Bon	4065.090	5	461	8.000	10.800	1.600	1.474
ScB	2208.111	17	194	2.059	3.176	1.400	0.469
DRM	2292.330	10	219	3.800	5.400	1.520	0.699

3.5.4 稳定性验证

本节将利用蒙特卡罗仿真方法来验证 DRM 在系统稳定性上的贡献。假设 MNCD 在各点的相关性为 0，即 $\rho_{ij}^{D}=0,\forall i\neq j$，其余参数与 3.5.1节的生成方式一致。当服务水平 α 从 0.6 到 0.99 变化时，通过求解模型 Bon，Ind，ScB 和 DRM，并计算相应方法在系统整体的需求满足比例，来比较系统整体的稳定性，具体方法如下。

步骤 1： 假设 $I=J=30,Q^2=50$，计算最优的网络结构和各个设施中心救护车储备数量。

步骤 2： 在不确定集 (3-14) 中，通过蒙特卡罗仿真方法随机生成 100000 个 MNCD 的可能取值。

步骤 3： 计算 100000 个样本中需求被满足的概率。

需求在整个地理空间上被满足的概率被汇总在表 3.6中。其中，表的

第一列列举了被建立的设施序号，用符号 $\hat{j}$ 表示，符号“–”代表该节点未被建立。如表 3.6所述，由蒙特卡罗仿真得到的各点需求被满足概率均大于预先设定的服务水平 α 值。除此之外，表 3.6还展示了四种方法在整个系统上的需求满足比率。定义

$$\text{系统稳定性（}\boldsymbol{R}\text{）}=\prod_{j\in\mathcal{P}}p_j$$

其中，$\mathcal{P}=\{j|Y_j=1\}$，p_j 代表各个蒙特卡罗方针下各点需求被满足的比率。

表 3.6　四种方法蒙特卡罗方法结果比较

$\hat{j}$	$\alpha=0.60$				$\alpha=0.70$				$\alpha=0.80$			
	Bon	Ind	ScB	DRM	Bon	Ind	ScB	DRM	Bon	Ind	ScB	DRM
2	1.00	0.89	1.00	1.00	1.00	0.94	1.00	1.00	1.00	0.98	1.00	1.00
5	—	—	1.00	—	—	0.94	1.00	—	—	—	—	—
6	1.00	0.89	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.98	1.00	1.00
7	—	0.89	1.00	—	—	0.98	1.00	—	—	0.98	—	—
8	1.00	0.90	1.00	1.00	1.00	0.96	1.00	1.00	1.00	0.98	1.00	1.00
9	1.00	0.97	1.00	1.00	1.00	0.95	1.00	1.00	1.00	0.98	1.00	1.00
12	1.00	0.98	1.00	1.00	1.00	0.97	1.00	1.00	1.00	0.98	1.00	1.00
14	1.00	0.97	1.00	1.00	1.00	0.98	1.00	1.00	1.00	0.98	1.00	1.00
15	—	0.89	—	—	—	—	—	—	—	—	—	—
16	1.00	0.92	1.00	1.00	1.00	0.94	1.00	1.00	1.00	0.98	1.00	1.00
17	1.00	0.89	1.00	1.00	1.00	0.94	1.00	1.00	1.00	0.98	1.00	1.00
18	1.00	0.89	0.50	1.00	1.00	0.94	0.76	1.00	1.00	0.98	0.76	1.00
19	—	0.89	0.50	1.00	—	0.94	0.76	—	—	0.98	0.76	—
20	—	0.89	0.92	—	—	0.94	0.92	—	—	0.98	1.00	—
21	—	0.89	1.00	—	—	0.93	1.00	—	—	0.98	1.00	—
22	—	0.97	1.00	1.00	—	0.97	1.00	1.00	—	0.98	1.00	1.00
24	—	0.96	1.00	—	—	0.94	—	—	—	0.98	1.00	—
25	1.00	0.92	—	1.00	1.00	—	—	1.00	1.00	0.98	1.00	1.00
26	1.00	0.96	0.98	1.00	1.00	0.94	0.98	1.00	—	0.98	0.98	1.00
27	—	0.89	0.98	—	—	0.94	0.98	—	—	0.98	0.98	—
29	—	0.89	0.22	—	—	0.94	0.98	—	—	0.98	0.98	—
30	1.00	0.92	1.00	1.00	1.00	0.94	1.00	1.00	1.00	0.98	1.00	1.00
R	1.00	0.16	**0.05**	1.00	1.00	0.34	**0.50**	1.00	1.00	0.66	**0.54**	1.00

续表

$\hat{j}$	$\alpha=0.90$				$\alpha=0.95$				$\alpha=0.99$			
	Bon	Ind	ScB	DRM	Bon	Ind	ScB	DRM	Bon	Ind	ScB	DRM
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	—	—	1.00	—	—	—	1.00	—	—	—	1.00	—
6	—	1.00	1.00	1.00	—	1.00	1.00	1.00	—	1.00	1.00	1.00
7	—	1.00	1.00	—	—	1.00	1.00	—	—	—	1.00	—
8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	—	1.00	1.00	1.00
12	1.00	1.00	1.00	1.00	—	1.00	1.00	1.00	—	1.00	1.00	1.00
14	1.00	1.00	1.00	1.00	—	1.00	1.00	1.00	—	1.00	1.00	1.00
16	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
17	—	1.00	1.00	1.00	—	1.00	1.00	1.00	—	1.00	1.00	1.00
18	1.00	1.00	0.81	1.00	1.00	1.00	0.81	1.00	1.00	1.00	0.81	1.00
19	—	1.00	0.76	—	—	1.00	0.76	—	—	—	0.76	—
20	—	1.00	1.00	—	—	1.00	1.00	—	—	—	1.00	—
21	—	1.00	1.00	—	—	1.00	1.00	—	—	—	1.00	—
22	1.00	1.00	1.00	1.00	—	1.00	1.00	1.00	—	—	1.00	1.00
24	—	1.00	1.00	—	—	—	1.00	—	—	—	—	—
25	—	1.00	—	1.00	—	1.00	—	1.00	—	1.00	—	1.00
26	—	1.00	1.00	1.00	—	1.00	1.00	1.00	—	1.00	1.00	1.00
27	—	1.00	0.98	—	—	1.00	1.00	—	—	—	1.00	—
29	—	—	0.98	—	—	—	0.99	—	—	—	1.00	—
30	—	1.00	1.00	1.00	—	1.00	1.00	1.00	—	1.00	1.00	1.00
R	1.00	0.98	**0.59**	1.00	1.00	1.00	**0.61**	1.00	1.00	1.00	**0.62**	1.00

尽管独立机会约束和基于场景的随机规划方法能够在各个独立的需求点获得较高的服务水平，但就整体而言，整个系统所有需求被满足的概率会大大降低。Bonferroni 近似方法和 DRM 方法的结果与之形成鲜明对比，能够保证系统稳定性不低于预先设定的服务水平 α。从保守性上来说，DRM 方法大大降低了 Bon 的过保守性，在保证系统高服务水平的基础上，节约了大量经济成本（表 3.1）。另外，当 α 变大时，ScB 可能出现系统可靠性变小的情况，这是由于基于场景的建模方法仅考虑了部分可能样本，具有一定的随机性，而鲁棒优化总是考虑分布集中的最坏情况，结果更为可靠。

3.5.5　DRM 在实证数据集中的表现

我们将本章提出的 DRM 应用在具有 30 个节点的北京市急救中心选址问题的实证算例中，该数据由 Zhang 和 Li 在 2015 年的论文中提出[55]，服务水平 $\alpha = 0.95$。我们获取了 1000 个日常需求和 MNCD 的历史数据，并采用与 3.5.1节相同的方式生成 $\mathbb{E}_F(\boldsymbol{\Theta})$ 的不确定集合 (3-7)。MNCD ($\boldsymbol{D}$) 的不确定椭球集半径被设定为 $Q^2 = \max\limits_{m=1,2,\cdots,M}(\boldsymbol{D}_m - \hat{\boldsymbol{u}})^{\mathrm{T}} \cdot \boldsymbol{\Gamma}_0^{-1}(\boldsymbol{D}_m - \hat{\boldsymbol{u}})$，其中 $\hat{\boldsymbol{u}}$ 和 $\boldsymbol{\Gamma}_0$ 与由变量 $\boldsymbol{D}$ 的样本均值和样本方差代替。

我们将 DRM，Bon 以及 Zhang 和 Li 论文[55] 中的三种方法在同一数据集上进行了比较，并在表 3.7和表 3.8中进行汇总。在表的第一行，“I-A”“I-S”“I-U”分别代表随机变量为随机、对称、单峰对称时的独立机会约束近似结果。注意，“I-A”与前文所述的基于 CVaR 的经典近似方法相同，而“I-S”和“I-U”则在“I-A”的基础上增添了随机变量的更多信息。

表 3.7　五种算法在实证数据集中的比较

	Bon	I-A	I-S	I-U	DRM
总成本比例	1.234	0.966	0.950	0.935	1.000
运算时间比例	0.212	0.619	0.754	0.868	1.000

表 3.8　基于实证数据集的蒙特卡罗仿真结果

	启用设施编号	Bon	I-A	I-S	I-U	DRM
随机抽取	8	1.0000	0.9860	1.0000	1.0000	1.0000
	9	1.0000	1.0000	0.9560	0.9690	1.0000
	24	1.0000	0.9900	0.9070	0.8820	1.0000
	系统稳定性	1.0000	0.9761	0.8671	0.8547	1.0000
在多维正态分布总体中抽取	8	1.0000	1.0000	1.0000	1.0000	1.0000
	9	1.0000	1.0000	0.9993	0.9997	1.0000
	24	1.0000	1.0000	0.9988	0.9977	1.0000
	系统稳定性	1.0000	1.0000	0.9981	0.9974	1.0000

表 3.7汇总了五种算法的最优成本和运算时间，为了使对比更加清晰明确，表 3.7中的数值均是对应算法数值除以 DRM 对应数值的比值。由于独立机会约束仅保证单个需求点的服务水平，而 DRM 和 Bon 考虑系统全局的服务水平，故 DRM 和 Bon 的总成本更高。与 DRM 相比，Bon 使系统成本提高了 23.4%，DRM 的目标函数值仅比独立机会约束近似多了 7%，却带来了在系统稳定性方面的巨大提高（3.5.4节和表 3.8）。

表 3.8采用与 3.5.4节类似的蒙特卡罗仿真方法对系统稳定性进行验证。首先，利用数据驱动方法计算日常需求（$\boldsymbol{\Theta}$）的椭球集 (3-7) 半径 $\epsilon = 8.493$。然后，在两个不同的 MNCD 总体中分别抽取 100 000 个样本进行蒙特卡罗仿真：第一个总体在椭球集中均匀分布；第二个总体来自均值为 $\hat{\boldsymbol{u}}$、方差为 $\boldsymbol{\Gamma}_0$ 的多维正态分布。

如果 MNCD 为随机抽取的变量，各点需求被满足的概率在模型 Bon，I-A 和 DRM 中应当全部大于或等于 0.95，由于样本不一定具备对称或者单峰结构，I-S 和 I-U 的各点需求被满足的概率不一定大于或等于 0.95。由表 3.8可知，I-S 和 I-U 的系统稳定性分别为 0.8671 和 0.8547，小于预先设定的 0.95。而当全体样本符合多元正态分布时，各点需求被满足的概率均大于 0.95，符合预期。

综合考虑表 3.7和表 3.8的结果，DRM 能以较小的成本最大限度地满足系统可靠性。与 Bon 相比，降低了近似的保守性；与独立机会约束近似相比，提高了系统的稳定性。

3.6 本 章 小 结

本章主要介绍了应急救援系统中急救中心的选址和规模设定问题，重点考虑了需求的不确定性和系统需求被满足的比例，并在保证服务水平的基础上尽可能降低了运营成本。研究的主要贡献包括：

（1）在建模方面，本研究提出的两阶段分布式鲁棒优化模型与传统随机规划和独立机会约束建模方式相比，能够在全局系统上保证更大的稳定性、以更高概率满足需求，同时减少了传统近似方法过于保守的弊端。

（2）在对模型的理论性质进行分析时，我们对联合机会约束与独立机会约束的近似程度进行了理论分析，得到了可以推广于其他使用场景的

结论。当随机变量的分布不确定集较小、即对随机变量的估计更为准确时，联合机会约束能够保证更高的系统稳定性，甚至能够比基于独立机会约束的建模方法更加准确。

（3）在管理意见方面，由于应急救援系统的主要评价指标为系统的运行效率（在本书中以需求被满足的概率描述），本章所尝试的多种随机建模方法均显示，同一需求点被多个设施服务时能够更好地应对需求不确定性，从而提高调度的灵活程度和需求满足比例。

未来的研究方向包括：如何设计更为高效的算法求解大规模含参 SOCP 问题；当随机变量服从其他分布形式、或者其他不确定分布集时，系统表现是否有所不同；在考虑应急救援系统的其他特点（如公平性、响应时间等）时，模型的结果是否不同，能否得到一些更有价值的管理建议。