

2015 年第十二届五一数学建模联赛

承诺书

我们仔细阅读了五一数学建模联赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与本队以外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其它公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们愿意承担由此引起的一切后果。

我们授权五一数学建模联赛赛组委会,可将我们的论文以任何形式进行公开展示(包括进行网上公示,在书籍、期刊和其他媒体进行正式或非正式发表等)。

我们参赛选择的题号为（从 A/B/C 中选择一项填写）： A

我们的参赛报名号为:

参赛组别（研究生或本科或专科）： 本科

所属学校（请填写完整的全名）

参赛队员 (打印并签名) : 1.

2.

3.

日期: 2015 年 5 月 3 日

获奖证书邮寄地址: 邮政编码:

收件人姓名: 联系电话:

2015 年第十二届五一数学建模联赛

编 号 专 用 页

竞赛评阅编号（由竞赛评委会评阅前进行编号）：

评阅记录

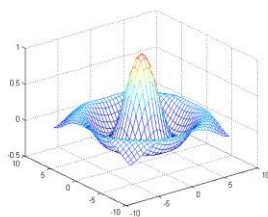
评 阅 人						
评 分						
备 注						

_____裁剪线_____裁剪线_____裁剪线

竞赛评阅编号（由竞赛评委会评阅前进行编号）：

参赛队伍的参赛号码：（请各参赛队提前填写好）：
2152#

2015 年第十二届五一数学建模联赛



题目 不确定条件下的最优路径的搜索模型

摘 要

本文以一般交通网络图为研究对象，在每条路径中每个路段的行驶时间的均值和方差已知的基础上，构造最优路径可靠性模型，保证相同高的概率到达时行驶总时间最短，然后把时间相关性和空间相关性融入建立的模型中，层层求解。最后根据所建立的模型分析算法，得出最优路径。

针对问题一，考虑到在现实中从起点到中间的路径错综复杂，且每条路径的路段数量不确定，我们绘制了一般交通网络图。在已知各路段行驶时间的均值与标准差的前提下，给出最优路径的定义。基于各路段的行驶时间服从正态分布的假设下利用卷积公式，针对每条路径给出以行驶总时间为变量的正态密度函数。最后对所得函数变形处理，得到判断最优路径的数学模型，并根据模型求出示例一的最优路径。

针对问题二，首先利用卷积公式和0-1规划，给出最优路径定义的条件和给出最优路径的搜索算法，运用 Matlab 和 excel 从所有路径中，搜索最优路径；并将其最优路径搜索算法运用到实际问题中，简化图如图 2，搜索最优路径为 $P \rightarrow v_2 \rightarrow v_1 \rightarrow v_4 \rightarrow Q$ ，该路径的均值为 17.77，标准差为 3.66。

针对问题三，在时间相关性方面构造 Copula 函数和最大似然估计函数，利用 excel 软件，得到从 $v_1 \rightarrow v_5$ ，在 7:00 ~ 8:00 堵塞时 8:00 ~ 9:00 对其他路段的均值和标准差的影响。在空间相关性方面，构造 Moren 模型和空间自相关分析，得到在一定时间下，某个路段出现交通事故，对其他相关路段的影响。结论为第 8 路段的均值为 12，标准差为 3.6。

针对问题四，赋予路段均值和方差的等权重，作为该路段的行驶参考时间，最终转化为 Dijkstra 算法，求得最优路径为 $P \rightarrow v_1 \rightarrow v_4 \rightarrow Q$ ，最优行驶时间为 13.65。

本文逻辑严谨，切入点独到，综合运用多种模型及软件，结果可靠且多样化。

关键词：最优路径；正态分布；Moren 模型；灵敏度分析；Matlab

§ 1 问题的重述

一、背景知识

1. 中国城市交通的基本情况

近年来,经济迅速的增长和城市化进程的加快使城市人口不断增加,城市人口的大量聚集和私家车人均拥有量的激增使城市交通面临着越来越艰巨的任务,交通拥堵、交通安全、高效交通等问题一涌而出,交通问题日益严峻,从一年一度的“春运”现象就可见一斑,不仅给居民的日常生活造成了严重的影响,还制约了城市及周边城市的发展。在复杂的交通环境下,如何综合考虑多方面因素,寻找一条可靠、快速、安全的最优路径,已成为每个人所迫切解决的问题。

2. 交通网

(1) **交通点**: 交通点即通常的汽车站、火车站、道路的交叉点、机场、港口等交通结点。

(2) **交通线**: 交通线即连接点与点的铁路、公路、水路以及航线的交通路线。

在特定的地域范围内,根据地区经济的发展和人们活动的需求,各种现代交通运输方式联合,各种交通线和交通点交织,形成了不同形式和层次的交通运输网,简称交通网。其布局受到经济、社会、技术和自然的影响和制约。

按交通运输方式分类,形成了铁路运输网、公路运输网、水路运输网、航空运输网和管道运输网。不同运输方式结合形成综合交通运输网。

3. 传统的最优路径

传统的最优路径是基于理想交通状况下分析得到的平均总行驶时间最短的路径。在这种情况下每条路段的行驶时间是确定的,可以用经典的最短路算法(Dijkstra算法)来寻找最短路径。这也是大多数车辆路径导航系统寻找最优路径所用的算法。但传统路径存在相当大的弊端,在现实生活中,交通事故,天气情况,流量等一系列不确定因素都会让最优路径失去最优性。

4. 最优路径的改良

不同于传统路径,改良后的最优路径把在现实生活中可能遇到的一系列不确定因素如天气、车流量等纳入考虑范围内,将每条路段的行驶时间不确定化,使其近似服从一个随机分布,建立模型,从而给出一个相对的最优路径。

二、相关数据

1. 原问题第一问的示例交通网络图及其数据;

三、要解决的问题

以中国矿业大学到徐州火车站为例,假设行驶时间是随机变量,假设已知每条路段行驶时间的均值和标准差。若走绕城快速路平均 33 分钟到达,标准差只有 1 分钟,若走市区道路,平均 30 分钟到达,标准差 15 分钟。若采用传统的最优路径算法,应选择平均时间较短的市区道路,但现实往往相反。最优路径不仅要考虑平均时间最短,还要考虑不确定性条件下车辆准时到达的可靠性,在此要求下,解决下列问题:

1. **问题一**: 建立适用于一般交通网络的数学模型,定量分析车辆行驶时间的不确定性,给出在不确定条件下最优路径才定义和数学表达式,并应用于示例一。

2. **问题二**: 根据第一问建立的模型设计算法搜索最优路径,并应用于具体的交通网络中,尽可能地从理论上分析算法的收敛性和复杂性。

3. **问题三**: 建立数学模型描述交通行驶时间之间的相关性,将其应于前二问的最

优路径搜索中，并设计算法解决这一问题，给出算例证明其有效性，尽可能的分析算法的收敛性、复杂性等性质。

4. 问题四：从不确定性条件下交通网络的实际情况出发，在合理假设下，进一步完善前三问的数学模型和相关算法。或提出一种或多种与前三问不同的最优路径的定义方法，建立相关的数学模型并设计算法，应用数值算例验证算法的有效性并尽可能从理论上分析算法的收敛性、复杂性等性质。

§ 2 模型的假设

1. 假设每条路段行驶时间服从正态分布；
2. 假设每条路段行驶时间的均值和标准差；
3. 已知假设所有不确定因素对行驶时间的影响都可以用均值和标准差体现；

§ 3 名词解释与符号说明

一、名词解释

1. **均值：**表示一组数据集中趋势的量数，反映数据集中趋势的一项指标。
2. **标准差：**标准差是方差的算术平方根，能反映一个数据集的离散程度。
3. **相关性：**数据间的关联程度，一个数据的取值会受到其他数据的影响。
4. **算法的复杂性：**算法复杂性是输入规模的函数，一般为避免不同输入对算法行为造成的巨大差别，考察所有输入规模为 n 的复杂性。

二、符号说明

序号	符号	符号说明
1	P	路径起点
2	Q	路径终点
3	L_i	第 i 条路径
4	C_i	第 i 条路径所用的总时间
5	T_i	每条路径的第 i 个路段所用时间
6	μ_i	每条路径的第 i 个路段所用时间的均值
7	σ_i	每条路径的第 i 个路段所用时间的方差
8	$P_T(t)$	T 的正态分布密度函数
9	v_i	第 i 交通点
10	y_i	0-1变量
11	F	多维分布函数
12	C	<i>Copula</i> 函数
13	$\hat{\theta}_i$	估计边缘分布参数

§ 4 模型的建立与求解

一、问题一的分析与求解

1. 对问题的分析

相比于传统最优路径，改良后的最优路径需要把现实中的不确定因素纳入考虑范围，对此我们将各路段行驶时间变量化，引入正态分布，首先将现实中的交通网络简化

为一般网络，基于各路段的行驶时间服从正态分布的假设，利用正态分布可加性把每条路径上的多个路段整合为一个路段。然后利用卷积公式，给出每条路径的正态密度函数，变形得出可靠性模型，并将相同概率下及时到达的路径中用时最短的路径定义为最优路径。

定义 1 最优路径

定义在相同概率到达终点的条件下，平均行驶总时间最短的路径即为最优路径。

2. 对问题的求解

(1) 模型的准备

如下图所示，绘制如下图。不妨假设以 P 为起点， Q 为终点的道路有若干条途径，中每条路径的均值与方差已知。



图 1 一般路径图

设随机变量 $T_1 \sim N(\mu_1, \sigma_1^2)$, $T_2 \sim N(\mu_2, \sigma_2^2)$ 且 T_1 与 T_2 独立，则有

$$T = T_1 + T_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)。$$

证明： $T = T_1 + T_2$ 仍在 $(-\infty, +\infty)$ 上取值，利用卷积公式

$$P_T(t) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left[\frac{(t-t_1-\mu_1)^2}{\sigma_1^2} + \frac{(t-\mu_2)^2}{\sigma_2^2}\right]\right\} dt_2$$

对上式被积公式的指数部分按 t_2 的幂次展开，合并同类项得：

$$\frac{(t-t_1-\mu_1)^2}{\sigma_1^2} + \frac{(t-\mu_2)^2}{\sigma_2^2} = A\left(t_2 - \frac{B}{A}\right)^2 + \frac{(t-\mu_1-\mu_2)^2}{\sigma_1^2 + \sigma_2^2}$$

其中

$$A = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}, \quad B = \frac{t-\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}。$$

代回原式可得：

$$P_T(t) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{1}{2} \frac{(t-\mu_1-\mu_2)^2}{\sigma_1^2 + \sigma_2^2}\right\} \cdot \int_{-\infty}^{+\infty} \exp\left[-\frac{A}{2}\left(t_2 - \frac{B}{A}\right)^2\right] dt_2$$

利用正态密度函数的正则性，上式的积分应为 $\frac{\sqrt{2\pi}}{\sqrt{A}}$ ，于是

$$P_T(t) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp\left\{-\frac{1}{2} \frac{(t-\mu_1-\mu_2)^2}{\sigma_1^2 + \sigma_2^2}\right\}$$

这正是均值为 $\mu_1 + \mu_2$ ，方差为 $\sigma_1^2 + \sigma_2^2$ 的正态密度函数，有下式成立：

$$T = T_1 + T_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)。$$

下用数学归纳法证明：

对于任意 n ，有下式成立：

$$\sum_{i=1}^n T_i \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)。$$

当 $k=1,2$ 时，显然成立。

假设当 $n=k$ 时成立，即为

$$\sum_{i=1}^k T_i \sim N(\sum_{i=1}^k \mu_i, \sum_{i=1}^k \sigma_i^2)，$$

则对于当 $n=k+1$ 时，

$$T = \sum_{i=1}^{k+1} T_i$$

$$P_T(t) = \frac{1}{2\pi\sigma_{k+1}\sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_k^2}} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \left[\frac{(t - t_{k+1} - \sum_{i=1}^k \mu_i)^2}{\sum_{i=1}^k \sigma_i^2} + \frac{(t_{k+1} - \mu_{k+1})^2}{\sigma_{k+1}^2} \right] \right\} dt_{k+1}$$

同样对于上式函数中指数部分按照 t_{k+1} 展开可得：

$$P_T(t) = \frac{1}{2\pi\sigma_1\sigma_2 \dots \sigma_{k+1}} \exp \left\{ -\frac{1}{2} \frac{(t - \sum_{i=1}^{k+1} \mu_i)^2}{\sum_{i=1}^{k+1} \sigma_i^2} \right\} \int_{-\infty}^{\infty} \exp \left\{ -\frac{A}{2} (t_{k+1} - \frac{B}{A})^2 \right\} dt_{k+1}$$

其中

$$A = \sum_{i=1}^{k+1} \frac{1}{\sigma_i^2}$$

$$B = \frac{t - \sum_{i=1}^k \mu_i}{\sum_{i=1}^k \sigma_i^2} + \frac{\mu_{k+1}}{\sigma_{k+1}^2}。$$

于是

$$P_T(t) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2 + \dots + \sigma_{k+1}^2)}} \exp \left\{ -\frac{1}{2} \frac{(t - \sum_{i=1}^{k+1} \mu_i)^2}{\sum_{i=1}^{k+1} \sigma_i^2} \right\}$$

即当 $n=k+1$ 时成立。

综上，得证。

(2) 模型的建立与求解

①模型的建立

模型 I 最优路径的可靠性模型

假设到达终点的概率为 p ，由卷积公式知某路径 $L_i \sim N(\sum_{j=1}^k \mu_{ij}, \sum_{j=1}^k \sigma_{ij}^2)$ ， k 为路径 L_i

的路段数。对其进行标准化处理，可得下式

$$\Phi\left(\frac{C_i - \sum_{j=1}^k \mu_{ij}}{\sqrt{\sum_{j=1}^k \sigma_{ij}^2}}\right) = p$$

可解得

$$C_i = \sum_{j=1}^k \mu_{ij} + u_p \sqrt{\sum_{j=1}^k \sigma_{ij}^2}$$

比较 C_i ，得到 $\min\{C_1, C_2, \dots, C_n\}$ 所对应路径即为最优路径。

②模型的求解

不妨取到达终点的把握性 p 为 0.9015，对应附录，查表可知对应分位数 $u_p = 1.2$ ，

$$\begin{cases} \Phi\left(\frac{C_1 - 33}{1}\right) = 0.9015 \\ \Phi\left(\frac{C_2 - 30}{15}\right) = 0.9015 \end{cases}$$

化简得：

$$C_1 = 1.2 * 1 + 33 = 34.2, \quad C_2 = 1.2 * 15 + 30 = 48。$$

故选取 C_1 ，即均值为 33，标准差为 1 的绕城快速路为最优路径。

③最优路径的不确定性的分析

从起始点到终点的每段路径的所需时间为 c_i ，由于 c_i 是随机的，所以每个路段所用时间对应的概率为 $p_i = \Phi\left(\frac{C_i - \mu_i}{\sigma_i}\right)$ ，由于 $0 < p_i < 1$ ，因为每个路段的行驶时间都是不确定性的，因而从起始点到终点也所用时间也是不确定性的。

二、问题二的分析与求解

1. 对问题的分析

由于从起始点到终点道路的复杂性。由上我们已经给出不同路径给出最优路径的最优方法。故本题所要考虑的主要问题即为给出已知各路段的情况下，给出求出所有路径的一般算法，然后根据第一问结论，得出最优路径。

2. 对问题的求解

(1)最优路径的算法设计

①算法的文字表述

设从起始点到终点总共有 $n+2$ 个交通点，除去起始点和终点还剩下 n 个，以下为最优路径的搜索步骤：

第一步：

从起始点，寻找与起始点直接相连的交通点，设交通点为 $\{v_1, v_2, \dots, v_k\}$ ，起点到其各个交通点的 0-1 变量为 $y_1, y_2, \dots, y_k$ ，则 $y_1 + y_2 + \dots + y_k = 1$ 。

第二步：

在第一步的 $\{v_1, v_2, \dots, v_k\}$ 的 k 个交通点中，逐个寻找相邻交通线的条数，如果 $\{v_1, v_2, \dots, v_k\}$ 中的点 v_j 相邻边数为 k_j 条，则 $\sum_{k=1}^{k_j} y_k = 2$ 。

第三步：

重复做第二步，直到到达与终点相邻的边的条数时，与第一步类似。

第四步：

导出关于非齐次线性方程的通解，但由于 y_k 是 0-1 变量只取值于 0 或者 1，所以存在有限个解使得起点到终点单线联通。

第五步：

将上步找到各个若干条单联通线，带入到第一问所给的数学表达式中，通过比较时间大小即得到了最优路径。

②算法的方程表示

将上述步骤用非齐次线性方程可以表示为：

$$\left\{ \begin{array}{l} \text{起点的约束: } y_1 + y_2 + \cdots y_k = 1 \\ v_1 \text{的约束: } \sum_{k=1}^{1_j} y_k = 2 \\ v_2 \text{的约束: } \sum_{k=1}^{2_j} y_k = 2 \\ \vdots \\ \text{终点的约束: } y_{s-1} + y_s + \cdots y_n = 1 \end{array} \right.$$

其中， $y_1, y_2, \cdots, y_k$ 为与起点相连接的 0-1 变量，

$$y_i = \begin{cases} 1, \text{该路径选择此路;} & (i=1,2,3\cdots n) \\ 0, \text{该路径不选择此路;} \end{cases}$$

起点的约束条件 $y_1 + y_2 + \cdots y_k = 1$ 说明的是从起点开始到相邻的点只选择一条路， v_1 的约束 $\sum_{k=1}^1 y_k = 2$ 说明的是每条路径经过每一交通点时，有且只选择两条路段。终点的约束： $y_{s-1} + y_s + \cdots + y_n = 1$ 说明的从相邻的点只选择一条路到终点。其中 $y_{s-1}, y_s, \cdots, y_n$ 是与终点相连接边的个数， $1_j, 2_j, \cdots, n_j$ 为各个节点相邻的边的个数。

(2) 该算法实际交通中的运用

交通网络图简化如下：

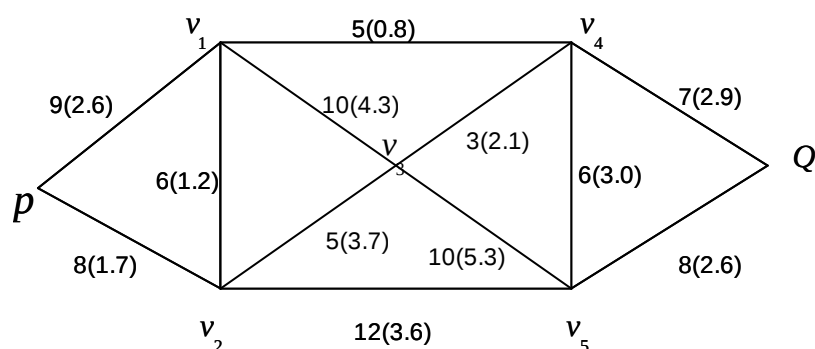


图 2 交通网络简化图

利用(1)中算法结论，可得如下非齐次线性方程组

$$\left\{ \begin{array}{l} P \text{点约束: } y_1 + y_2 = 1; \\ v_1 \text{点约束: } y_1 + y_3 + y_4 + y_5 = 2; \\ v_2 \text{点约束: } y_2 + y_3 + y_6 + y_7 = 2; \\ v_3 \text{点约束: } y_5 + y_6 + y_8 + y_9 = 2; \\ v_4 \text{点约束: } y_4 + y_8 + y_{10} + y_{11} = 2; \\ v_5 \text{点约束: } y_7 + y_9 + y_{11} + y_{12} = 2; \\ Q \text{点约束: } y_{11} + y_{12} = 2; \\ y_i = 0 \text{或} 1 (i = 1, 2, 3, \dots, 12); \end{array} \right.$$

求解具体过程步骤见附录程序一。

最后得出在到达终点的把握性 p 为 0.9015 时, L_{21} 为最优路径。即最优路径如下:

$$p \rightarrow v_2 \rightarrow v_1 \rightarrow v_4 \rightarrow Q。$$

(3) 分析算法有效性, 收敛性

算法的有效性分为两个方面, 第一是算法的正确性; 其二还必须考虑执行算法所消耗的时间和执行算法所耗费的空间 (主要是指辅助空间), 以及算法是否易读, 易编码和易于调试。

三、问题三的分析与求解

1. 对问题的分析

对于路段的不确定性, 由上第一问与第二问我们探讨知, 我们给出了在只考虑一般道路状况下 (此处道路交通状况不包括对行驶时间造成较大影响的情况) 给出了一般最优路径的求解一般算法。对于问题三, 我们主要将交通事故、恶劣天气、突发事件等考虑在内, 并从时间相关性和空间相关性两个方面。并利用控制变量法, 即在考虑该路段时间相关性时, 不考虑空间相关性对其造成的影响。分析空间相关性时, 不对时间相关性加以考虑。

2. 对问题的求解

(1) 模型的准备

时间相关性和空间相关性的定义如下:

①**时间相关性**: 对于路段 a , 不同时间段的相关性, 例如 7:00–8:00 和 8:00–9:00 之间的相关性。

②**空间相关性**: 同一个时间段 (例如 7:00–8:00 之间), 路段 a 和路段 b 的相关性。

(2) 模型的建立与求解

① 时间相关性

下面引入 Copula 函数对时间相关性进行描述。Copula 可以解释为“相依函数”或“连接函数”, 是把多维随机变量的联合分布用其一维边际分布连接起来的函数。下面首先引入 Sklar 定理。

Sklar 定理:

假设一个多维分布函数 F 的边际分布函数为:

$$F_1(\bullet), F_2(\bullet), F_3(\bullet) \cdots F_n(\bullet),$$

则存在一个 Copula 函数 C 满足:

$$F(x_1, x_2, x_3 \cdots x_n) = C(F_1(\bullet), F_2(\bullet), F_3(\bullet) \cdots F_n(\bullet))$$

如果 $F_1(\bullet), F_2(\bullet), F_3(\bullet) \cdots F_n(\bullet)$ 是连续的, 则 Copula 函数是唯一确定的, 反之亦然。由这个定理可以推出当确定了多个路段行程时间边际分布和选定一个合适的 Copula 函

数后，就可以方便地计算出这些路段行程时间的联合分布，这正是 Copula 函数在实际中解决交通网络行程时间可靠性的优势所在。Sklar 定理的数学表达如下：

$$f(x_1, x_2, x_3 \cdots x_n; \theta) =$$

$$c(F_1(x_1, \theta_1), F_2(x_2, \theta_2), F_3(x_3, \theta_3) \cdots F_n(x_n, \theta_n); \theta_c) \prod_{i=1}^n f_i(x_i, \theta_i) =$$

$$c(u_1, u_2, u_3 \cdots u_n; \theta_c); \theta_c) \prod_{i=1}^n f_i(x_i, \theta_i)$$

其中 $c(F_1(x_1), F_2(x_2), F_3(x_3) \cdots F_n(x_n)) =$

$$\frac{\partial C(F_1(x_1), F_2(x_2), F_3(x_3) \cdots F_n(x_n))}{\partial F_1(x_1) \partial F_2(x_2) \partial F_3(x_3), \cdots, \partial F_n(x_n)}$$

$f(x_1, x_2, x_3 \cdots x_n; \theta)$ 为具有参数 θ 的联合概率密度函数； $f_i(x_i; \theta_i)$ 为具有参数

$\theta_i (i=1, 2, \cdots, n)$ 的边缘分布函数。根据 Copula 函数的相关理论，可以运用两阶段法构建 Copula 模型。第一阶段，确定边缘分布；第二阶段，选取一个适当的 Copula 函数，以便能很好地描述出随机变量之间的相关结构。

Copula 模型的参数估计可以采用极大似然估计法和矩估计法。其中极大似然估计是最常用的 Copula 模型的估计方法。采用极大似然估计法，Copula 模型的参数估计可以分为两步：

Step1: 估计边缘分布参数

$$\hat{\theta}_i = \arg \max \sum_{t=1}^T \ln f_i(x_{it}, \theta_i)$$

Step2: 估计 Copula 函数中的参数

$$\hat{\theta}_c = \arg \max \sum_{t=1}^T \ln c(F(x_{1t}, \hat{\theta}_1), F(x_{2t}, \hat{\theta}_2), F(x_{3t}, \hat{\theta}_3), \cdots, F(x_{nt}, \hat{\theta}_n); \theta_c)$$

即，首先估计出边缘分布函数的参数 θ_i ，然后利用估计值 $\hat{\theta}_i$ 作为已知数代入 Copula 模型中，进而得到 Copula 函数中参数 θ_c 的估计值 $\hat{\theta}_c$ 。

本题中，当考虑路段间行程时间相关性时，计算路径行程时间可靠性便不能够通过简单的串联关系得到，即路径的行程时间概率分布函数不再是各路段行程时间的概率分布函数的乘积。此时需要利用 Copula 函数推导出路径行程时间的随机分布函数，然后再利用路径的行程时间可靠度定义计算出路径行程时间可靠度。

已知假设为路段行程时间分布符合参数为 (μ, σ) 的对数正态分布。其概率密度及累计分布函数分别为

$$f_{T_a}(t) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma(t-t_a^0)}} e^{\frac{-(\ln(t-t_a^0)-\mu)^2}{2\sigma^2}} & t > t_a^0 > 0 \\ 0 & \text{其他} \end{cases}$$

$$F_T(t) = \int_{t_a^0}^t \frac{1}{\sqrt{2\pi\sigma(t-t_a^0)}} e^{\frac{-(\ln(t-t_a^0)-\mu)^2}{2\sigma^2}} dt$$

其中： t_a^0 为 a 的自由行驶时间。

如果起点 P 到 Q 终点中路段 $v_1 \rightarrow v_4$ 路段出 7:00 ~ 8:00 现重大交通事故,就是说此路段的均值和方差瞬间变得充分大,可以看作此路断路,下面我们做九条路段 $P \rightarrow v_1$, $P \rightarrow v_2$, $v_1 \rightarrow v_2$, $v_1 \rightarrow v_3$, $v_2 \rightarrow v_3$, $v_2 \rightarrow v_5$, $v_3 \rightarrow v_4$, $v_3 \rightarrow v_5$, $v_4 \rightarrow v_5$ 在时间 8:00 ~ 9:00 内除此路段外其他路段的通行时间和标准差。

表 1 各路段参数值

路段	参数估计				假设检验		
	$\hat{\mu}$	$\hat{\sigma}$	均值	方差	自由度 $(k-r-1)$	$\hat{\chi}^2$	$\chi_{0.05}^2(k-r-1)$
Link1	4.171	0.147	9	2.6	7	5.981	14.067
Link2	4.295	0.097	8	7.1	7	2.563	14.067
Link3	2.788	0.077	5	0.8	5	8.565	11.071
Link4	2.865	0.085	10	4.3	6	10.364	12.592
Link5	2.862	0.090	5	3.7	7	8.973	14.067
Link6	2.469	0.050	12	3.6	4	1.456	9.488
Link7	2.983	0.058	3	2.1	6	2.716	12.592
Link8	2.956	0.064	10	5.3	5	3.283	11.071
Link9	2.658	0.070	6	3.0	6	6.162	12.592

从检验结果可以看出,各个路段行程时间分布的 $\chi^2 < \chi_{0.05}^2(k-r-1)$,表明不能拒绝原假设,即认为各路段时间服从正态分布。程序见附录程序二。

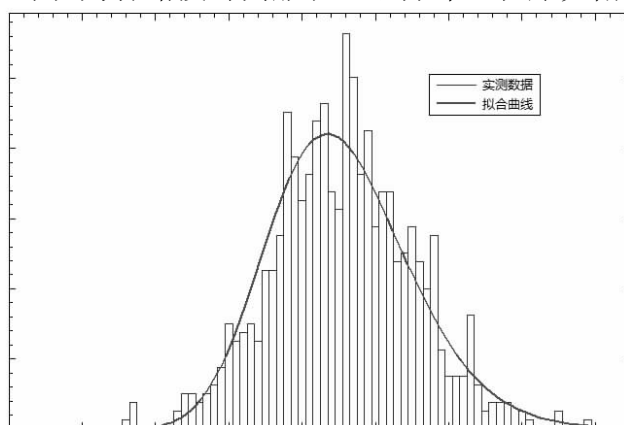


图 3 路段一行驶时间分布拟合曲线

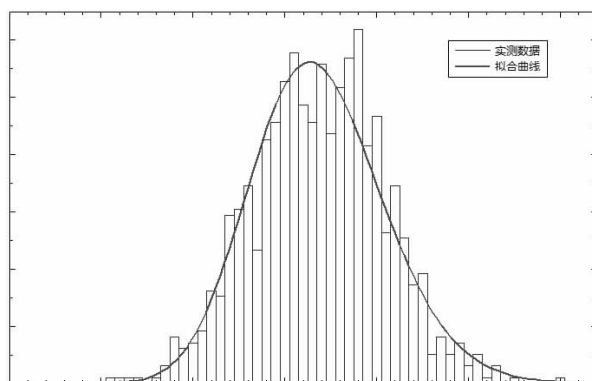


图 4 路段二行驶时间分布拟合曲线

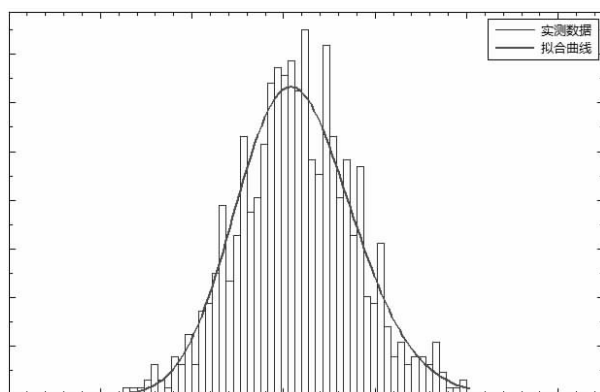


图 6 路段四行驶时间分布拟合曲线

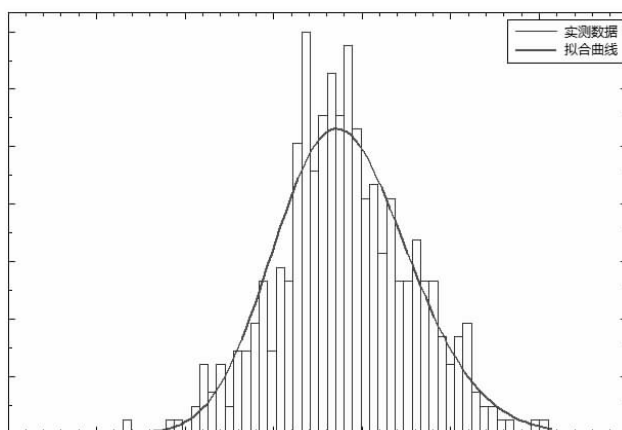


图 6 路段五行行驶时间分布拟合曲线

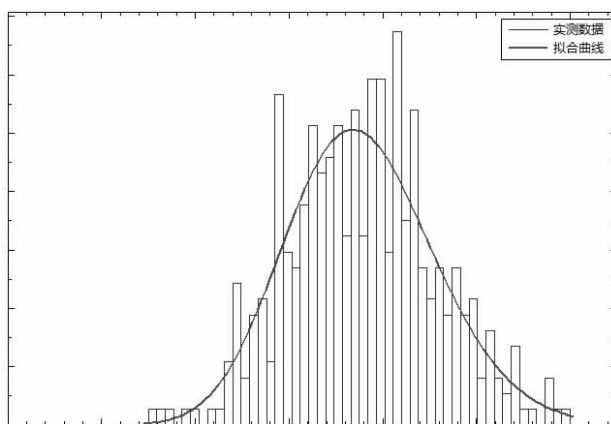


图 7 路短七行驶时间分布拟合曲线

采用最大似然估计方法得到最大似然估计，得出下表：

表 2 路段的参数值

路段	$\hat{\mu}$	$\hat{\sigma}$	实际均值	实际标准差	均值残差平方	标准差平方和
Link1	11.8	3.3	9	2.6	7.84	0.49
Link2	8.8	2.8	8	2.6	0.64	0.04
Link3	5.2	1.1	5	0.8	0.04	0.09
Link4	11.4	4.1	10	4.3	1.96	0.04
Link5	7	3.9	5	3.7	4	0.04
Link6	11.6	3.8	12	3.6	0.16	0.04
Link7	3.3	2.9	3	2.1	0.09	0.64

Link8	12.8	4.1	10	5.3	7.84	0.81
Link9	5.5	3.8	6	3.0	0.25	0.64

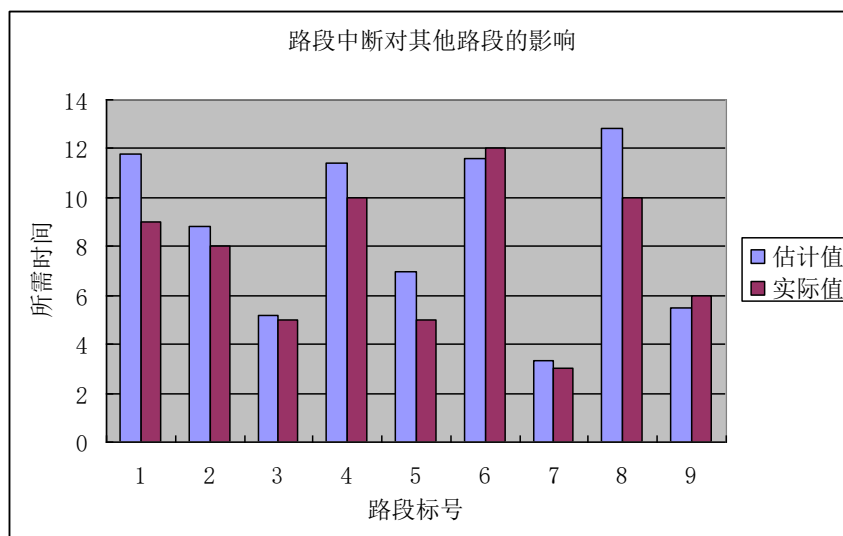


图 8 路段时间均值的影响

由上图可以看出,当如果起点 P 到 Q 终点中路段 $v_1 \rightarrow v_4$ 路段出7:00~8:00 现重大交通事故,就是说此路段的均值和方差瞬间变得充分大,可以看作此路断路,下面我们做九条路段 $P \rightarrow v_1$, $P \rightarrow v_2$, $v_1 \rightarrow v_2$, $v_1 \rightarrow v_3$, $v_2 \rightarrow v_3$, $v_2 \rightarrow v_5$, $v_3 \rightarrow v_4$, $v_3 \rightarrow v_5$, $v_4 \rightarrow v_5$ 在时间8:00~9:00,通过各个道路所需的时间都有明显的上升趋势,但是可以看出唯独最后一条,却有减少的现象,观察该道路的特点可知,该路段的均值为12,标准差为3.6,途径改路耗时长,易于拥堵,所以人们绕路,导致该路段比平时要畅通,与直观感受相符合。

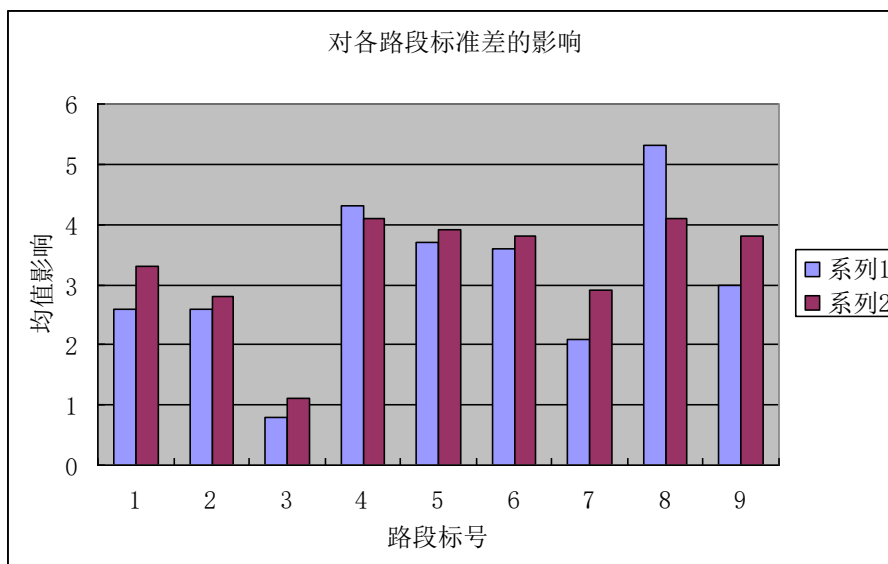


图 9 各个路段标准差的变化

如果起点 P 到 Q 终点中路段 $v_1 \rightarrow v_4$ 路段出7:00~8:00 现重大交通事故,就是说此路段的均值和方差瞬间变得充分大,可以看作此路断路,下面我们做九条路段 $P \rightarrow v_1$, $P \rightarrow v_2$, $v_1 \rightarrow v_2$, $v_1 \rightarrow v_3$, $v_2 \rightarrow v_3$, $v_2 \rightarrow v_5$, $v_3 \rightarrow v_4$, $v_3 \rightarrow v_5$, $v_4 \rightarrow v_5$ 在时间8:00~9:00可以看出该路段均值为3,标准差为2.1,也就是说该路段耗时较少,但是

易于拥挤，所以人们更愿意选择路径耗时较少的路通行，所以人们不会选择该路通行，因而该路会忽然之间变得畅通无阻，这与图中所给的标准差骤减相符合。

②空间相关性：

空间相关是根据位置相似性和属性相似性的匹配情况来测度的。位置的相似可通过空间接近性矩阵或权重矩阵 w 来描述，而属性值的相似一般通过交叉乘积 $x_i x_j$ 或平方差异 $(x_i - x_j)^2$ ，或绝对差异 $|x_i - x_j|$ 来描述。若存在正空间自相关，则在近邻的空间位置上属性值的差异小；若存在负的空间自相关，则近邻的位置上属性值的差异大。此外空间自相关程度各不相同，其强度是可测度的。强的空间自相关意味着近邻对象的属性值高度接近，而无需考虑是正值还是负值。空间自相关的测量是建立在空间单元相邻与否的二元逻辑基础上的。按照这一定义，邻边的结构由 0-1 来表达：空间相邻指 2 个空间单元共有非零长度的边界，从而赋予 1 的空间贴近度。通常定义一个二元对称空间权重矩阵 w 来表达， n 个位置的空间区域的邻近关系，其形式如下：

$$W = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & & \vdots \\ w_{n1} & \cdots & w_{nn} \end{bmatrix}$$

式中： w_{ij} 为区域 i 与 j 的邻近关系，它可以根据邻接标准或距离标准来度量。

2 种最常用的确定空间权重矩阵的规则：

1) 简单的二进制邻接矩

$$W_{ij} = \begin{cases} 1 & \text{当区域 } i \text{ 和 } j \text{ 相连接} \\ 0 & \text{其他} \end{cases}$$

2) 基于距离的二进制空间权重矩阵

$$W_{ij} = \begin{cases} 1 & \text{当区域 } i \text{ 和 } j \text{ 的距离小于 } d \text{ 时} \\ 0 & \text{其他} \end{cases}$$

从上面的分析可以明显地看出，对于一个线性路线上任何集合的行个部分，将会有 $2(n-1)$ 个连接。

模型 II Moran 模型

Moran 模型其表达式如下

$$I = \frac{n}{\sum \sum w_{ij}} \cdot \frac{\sum \sum w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum (X_i - \bar{X})^2}$$

式中： n 为研究对象的数目， X_i 为观测值； $\bar{X}$ 为 X_i 的平均值， $Moran's I$ 的期望值为：

$$E(I) = -\frac{1}{n-1}$$

方差为：

$$Var(I) = \frac{n^2 S_1 - n S_2 + 3(\sum \sum w_{ij})^2}{(\sum \sum w_{ij})^2 (n^2 - 1)}$$

式中：

$$S_1 = \frac{1}{2} \sum \sum (w_{ij} + w_{ji})^2 ;$$

$$S_2 = \sum_i (\sum_i W_{ij} + \sum_j W_{ji})^2$$

本文使用的空间自相关统计量由 *Moran's I* 衍生，即以近似正态假设为前提对所计算出的 *Moran's I* 值进行标准化，得到检验统计量 *Z*。根据 *Z* 值对原假设“变量在所讨论的空间区域上呈随机分布”得到检验统计量 *Z*。然后根据 *Z* 值对原假设进行检验，以判断空间自相关是否存在。检验时取显著性水平 $\alpha = 0.10$ （双侧检验）。其中：

$$Z = [I - E(I)] / [VAR(I)]^{1/2}$$

当 *Z* 值为正且显著时，表明存在正的空间自相关，也就是说相似的观测值（高值或低值）趋于空间集聚；当 *Z* 值为负且显著时，表明存在负的空间自相关，相似的观测值趋于分散分布；当 *Z* 值为零时，观测值呈独立随机分布。

下给出例子加以说明：

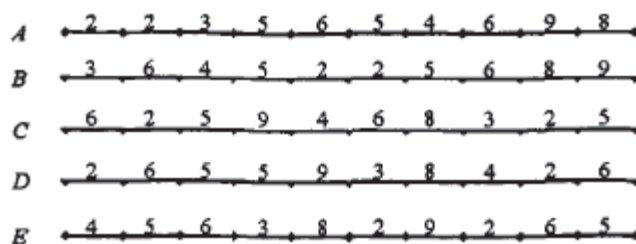


图 10 各路段交通事故图

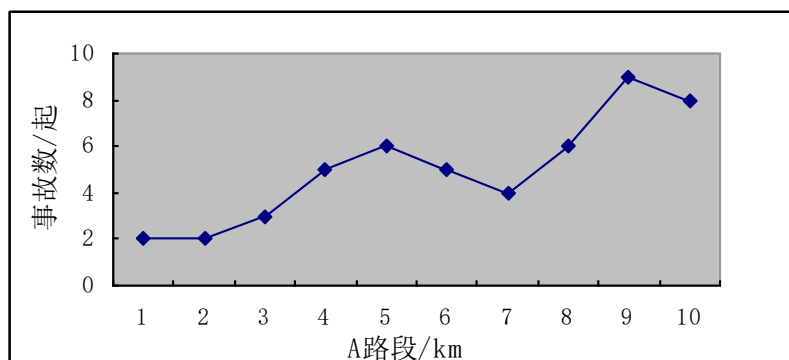


图 11 A 路段交通事故折线图

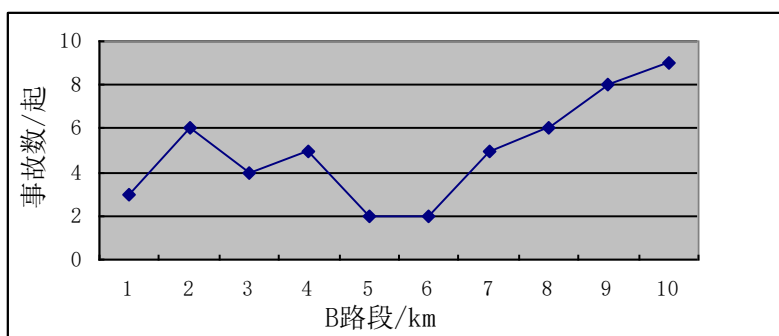


图 12 B 路段交通事故折线图

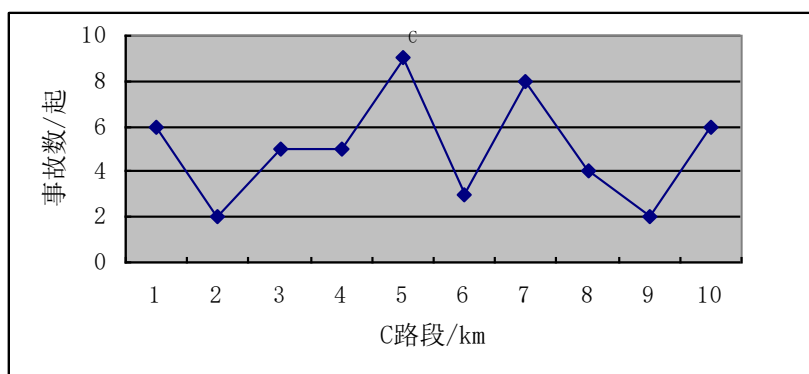


图 13 C 路段交通事故折线图

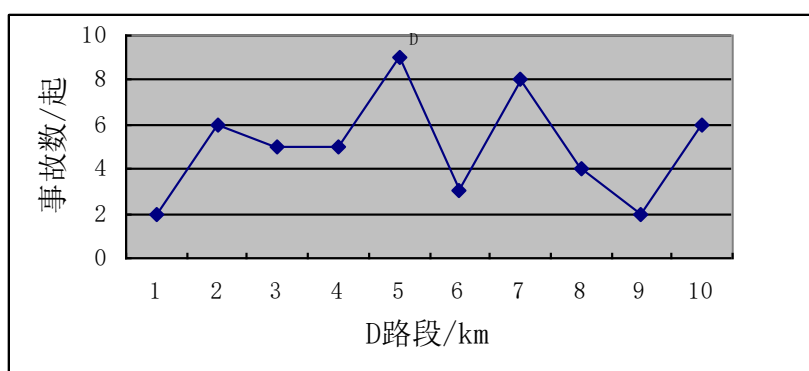


图 14 D 路段交通事故折线图

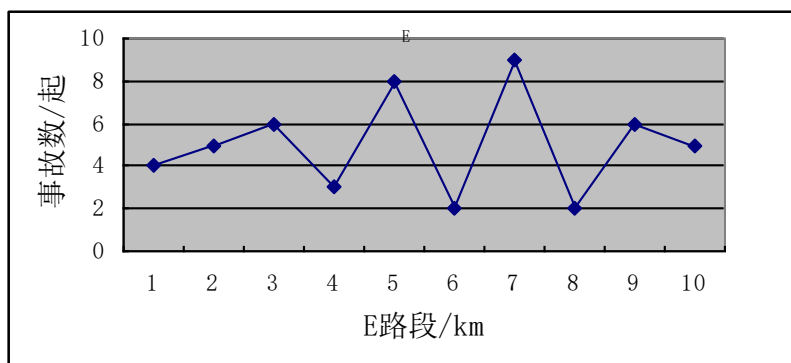


图 15 E 路段交通事故折线图

通过计算得到这 5 个路段上的事故空间自相关指数，结果如下表

表 3 路段自相关指数

路段	<i>Moran's I</i>	<i>Z</i>
A	0.667	2.63
B	0.467	1.95
C	-0.111	0.00
D	-0.444	-1.13
E	-0.978	-2.93

从上表可以看出，在 0.01 显著性水平（ $Z = 2.58$ ）下，A 为正自相关，E 为负自相

关；在 0.05 显著性水平（ $Z=1.96$ ）下 A,B 为正自相关，D,E 为负自相关；C 为非自相关。

正自相关在公路上相邻的路段具有相似的事故几率。可以表明这一路段上的某些因素(如天气或设计因素等)与相邻路段事故的增加有着因果关系。负网络自相关可以表明公路上相邻路段有着不同事故率的趋势。这是由分析得出的结果，但在观察试验中却很少发生，除非所在其他的路段上都有一个出口或通道，能在那些路段上造成额外的事故。故在对已知路段事故数考察以后即可的出交通道路的空间相关性。

四、问题四的分析与求解

1. 对问题的分析

由于从起点到终点每个路段都是随机的，从实际来看从起点到终点所需时间越短越好，但是有时不得不考虑图中的随机因素，既要考虑途中所耽搁的时间，如果我们将均值和方差去权重即反映了个人的行程偏好不同。

2. 对问题的求解

①最优路径新定义

在均值方差一定的权重下，均值方差在各自权重下的和即为路段的时间，从起点到终点总时间最小的路径即为最优路径。

②最优路径的搜索

由①知：

$$\begin{cases} \lambda_1 \mu_i + \lambda_2 \sigma_i, i = 1, 2, \dots, n \\ \lambda_1 + \lambda_2 = 1 \end{cases}$$

下面可以将其转化为标号算法

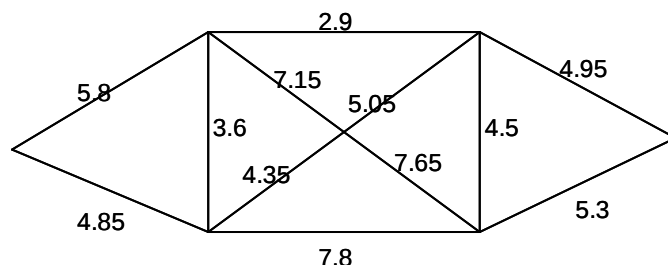


图 16 标号算法初步

利用标号算法求解过程见附录二。由上标号算法知最优路径为：

$$P \rightarrow v_1 \rightarrow v_4 \rightarrow Q$$

所用行驶时间为 13.65。由此可见，不同评判标准下所得到的最优路径不同，且这种“最优路径”是由个人喜好所决定的。

§ 5 灵敏度分析

一、灵敏度分析

①对方差一定不同概率下的分析:若将第一问的小问题中，从环城快速路由概率保证下所需时间为 $t_1 = \mu_1 + u_p \sigma_1$ ；

从城区路所需时间为 $t_2 = \mu_2 + u_p \sigma_2$ 。

表 4 正态分布的分位数

数据	1	2	3	4	5	6	7	8	9	10
概率	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
位数	0	0.13	0.26	0.39	0.53	0.68	0.85	1.04	1.29	1.65

可以看出概率随着分位数的增大而增大，运用 matlab 作出有关于 t_1 和 t_2 的图像为程序见附录程序三：

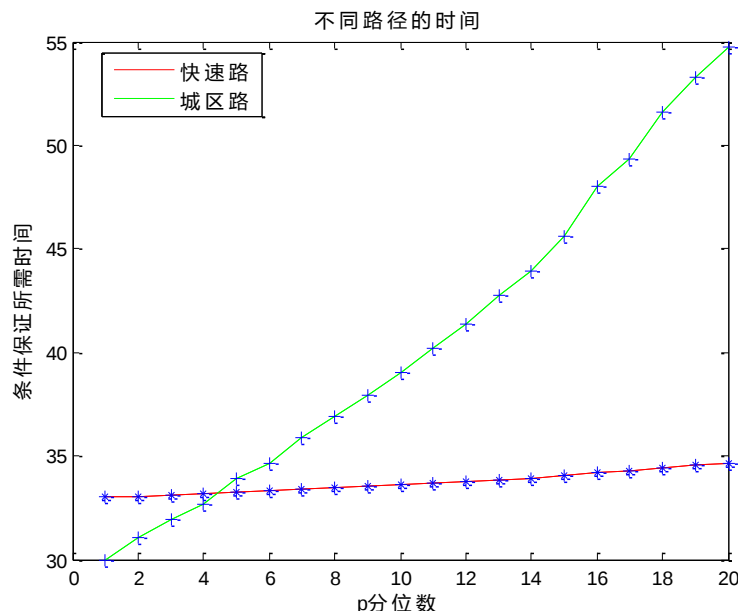


图 17 条件保证的时间图

由以上途中可以看出，两条道路无论是当到达终点的概率提高时，所需时间都有增加的趋势，但是选择城区路有着随着概率增加有着明显的增加趋势，随机性太多，途中耽搁时间也多。很有趣的发现在概率为 0.57 时，从两种途径到达的时间是一样的，但是当到达终点的概率低于 0.57 时，应该选择城区路，博取一下道路无拥堵的机会，但是，当概率要求很高时，应该选择快速路，减少途中随机因素的干扰。也由此，可以到到多个路段随机因素服从正态分布的路段，通过卷积公式，合成一条路径，也满足上述灵敏度分析。

§ 6 型的评价与推广

一、模型的优点

1. 相比于传统的解决最优路径的 Dijkstra 算法，最优路径的可靠性模型把不确定性因素纳入模型中，使模型更现实化，所得结果更实用；
2. Moren 模型不仅考虑到每个路段的行驶时间的随机性，还考虑到相邻路段之间的相关性，比模型一进一步提高了最优路径的可靠性。

二、模型的缺点

1. 在构建模型时，对问题进行多次假设、简化，导致最终结果存在误差，使得模型的实用性减弱；
2. 模型的检验和算法的分析缺乏大规模数据的参与，使得所得结果具有片面性。

三、模型的推广

本文所建立的模型一和模型二不仅适用于最优路径的选择，也适用于物流方案的调

整，航空运输业最小费用的优化。

参考文献

- [1]胡运权等，运筹学基础及应用（第六版），北京：高等教育出版社，2014.2.
- [2]茆诗松，程依明，濮晓龙，概率论与数理统计教程，北京：高等教育出版社，2011.2.
- [3]张志涌，杨祖樱等，MATLAB 教程：R2012a，北京：北京航空航天大学出版社，2010.8
- [4]万永福等，数学实验教程：Matlab 版，北京：科学杂志社，2006
- [5]杨桂元，朱家明，数学建模竞赛优秀论文评析，合肥：中国科学技术大学出版社，2013.9

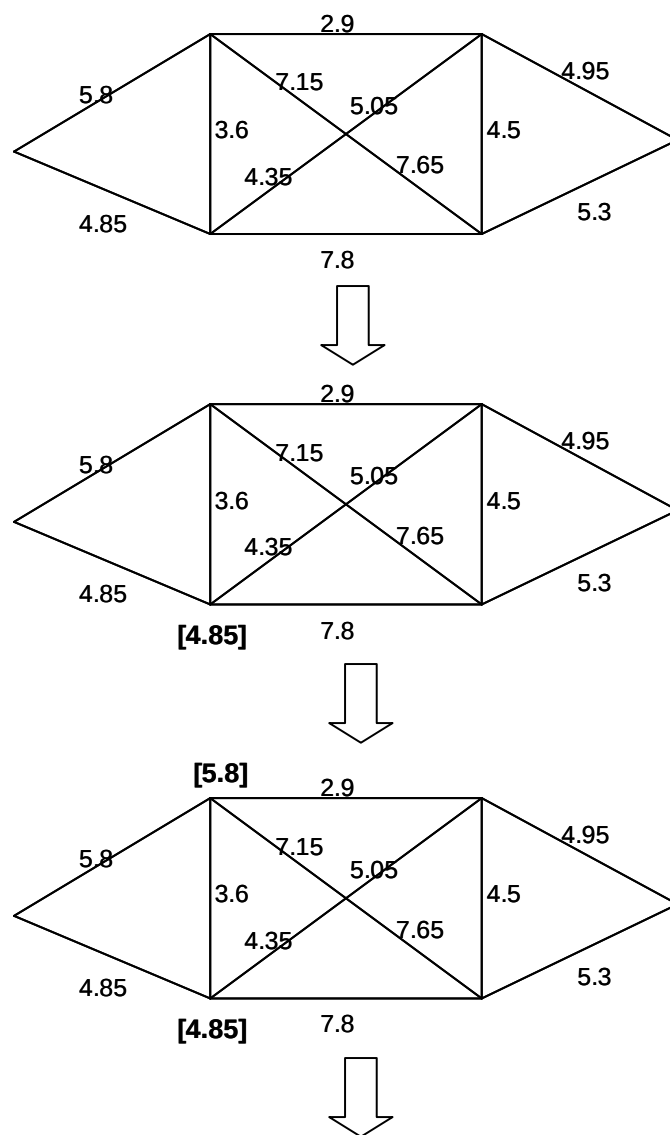
附录

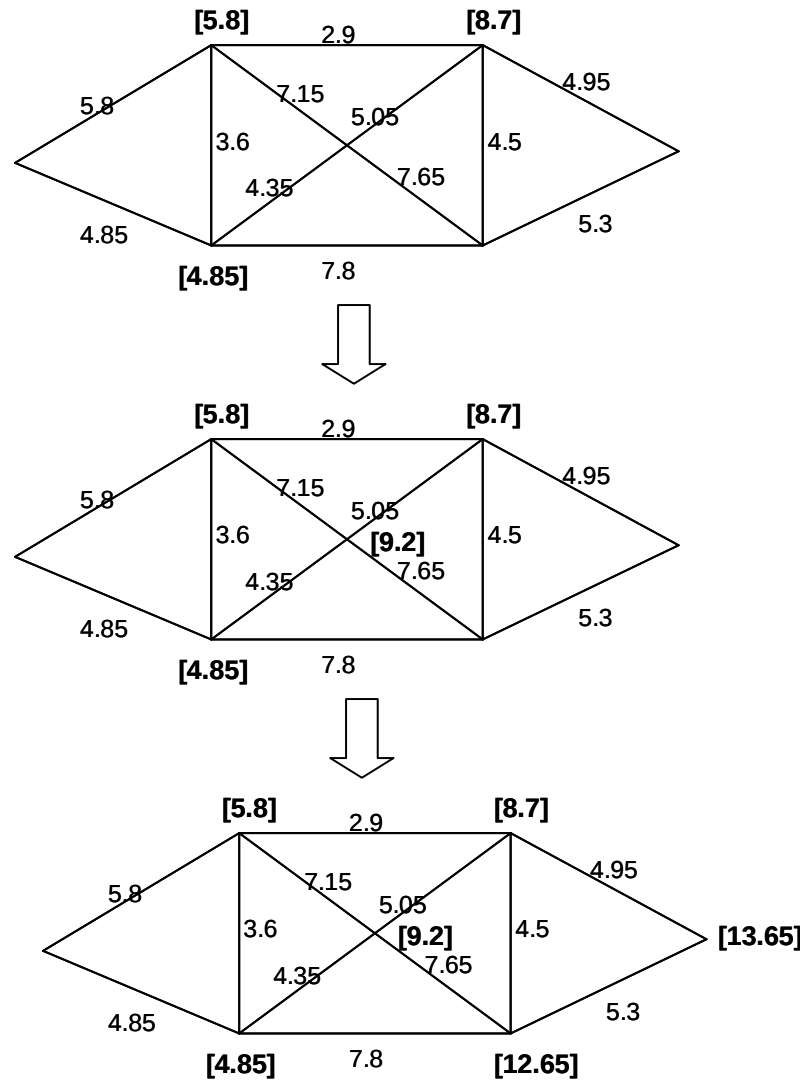
附录一：表 4 路径数据汇总

所有路径 L_i	总均值 u_i'	总方差 $\sigma_i'^2$	标准差 σ_i'	对应概率 $p = 0.9015$ 下的分位数
$L_1: p, v_1, v_2, v_3, v_4, Q$	30	34.71	5.89	41.78
$L_2: p, v_1, v_2, v_3, v_4, v_5, Q$	37	42.06	6.48	49.84
$L_3: p, v_1, v_2, v_3, v_5, Q$	38	56.74	7.53	65.78
$L_4: p, v_1, v_2, v_3, v_5, v_4, Q$	43	67.39	8.21	77.24
$L_5: p, v_1, v_2, v_5, Q$	35	27.92	5.28	34.26
$L_6: p, v_1, v_2, v_5, v_4, Q$	40	38.57	6.21	46.02
$L_7: p, v_1, v_3, v_2, v_5, Q$	39	44.97	6.71	53.02
$L_8: p, v_1, v_3, v_2, v_5, v_4, Q$	44	55.62	7.46	64.57
$L_9: p, v_1, v_3, v_4, Q$	29	38.07	6.17	45.50
$L_{10}: p, v_1, v_3, v_4, v_5, Q$	36	45.42	6.74	53.51
$L_{11}: p, v_1, v_3, v_5, Q$	37	60.1	7.75	69.40
$L_{12}: p, v_1, v_3, v_5, v_4, Q$	42	70.75	8.41	80.84
$L_{13}: p, v_1, v_4, Q$	21	15.81	3.98	20.58
$L_{14}: p, v_1, v_4, v_3, v_2, v_5, Q$	42	45.22	6.72	53.29
$L_{15}: p, v_1, v_4, v_3, v_5, Q$	35	46.66	6.83	54.86
$L_{16}: p, v_1, v_4, v_5, Q$	28	23.16	4.81	28.93
$L_{17}: p, v_2, v_1, v_3, v_5, Q$	42	57.67	7.59	66.78
$L_{18}: p, v_2, v_1, v_3, v_5, v_4, Q$	47	68.32	8.27	78.24
$L_{19}: p, v_2, v_1, v_3, v_4, Q$	24	17.15	4.14	22.12
$L_{20}: p, v_2, v_1, v_3, v_4, v_5, Q$	31	24.5	4.95	30.44
$L_{21}: p, v_2, v_1, v_4, Q$	26	13.38	3.66	17.77
$L_{22}: p, v_2, v_1, v_4, v_5, Q$	33	20.73	4.55	26.19
$L_{23}: p, v_2, v_3, v_1, v_4, Q$	35	44.12	6.64	52.09
$L_{24}: p, v_2, v_3, v_1, v_4, v_5, Q$	42	51.47	7.18	60.08

$p, v_2, v_3, v_1, v_4, v_5, Q$				
$L_{25}: p, v_2, v_3, v_5, Q$	31	51.43	7.17	60.04
$L_{26}: p, v_2, v_3, v_5, v_4, Q$	36	62.08	7.88	71.53
$L_{27}: p, v_2, v_3, v_4, Q$	23	29.4	5.42	35.91
$L_{28}: p, v_2, v_3, v_4, v_5, Q$	30	36.75	6.06	44.02
$L_{29}: p, v_2, v_5, Q$	28	22.61	4.75	28.32
$L_{30}:$	52	61.88	7.87	71.32
$p, v_2, v_5, v_3, v_1, v_4, Q$				
$L_{31}: p, v_2, v_5, v_3, v_4, Q$	40	47.16	6.88	55.40
$L_{32}: p, v_2, v_5, v_4, Q$	33	33.26	5.77	40.18

附录二：标号算法过程





程序一：求方程组程序

```

Clcf
Clear all;
Syms n1,n2,n3,n4,n5
A=load('data.txt');
B=rref(A);
C=null(A)
b=B*[n1,n2,n3,n4,n5]+null(A);
put('b')
+

```

程序二：作图程序

```

X=load('data1.txt');%各个导入不同的数据
X=x(:);
[mu,sigma]=normfit (x) ;
hist(x);
figure;
histfit(X);

```

程序三：作图程序

```
f
p=[0 0.07 0.13 0.18 0.26 0.31 0.39 0.46 0.53 0.60 0.68 0.76 0.85 0.93 1.04 1.2 1.29 1.44 1.55 1.65]
n=length(up)
x=1:n
for i=1:n
t1(i)=33+up(i)*1
t2(i)=30+up(i)*15
end
plot(x,t1,'*')
hold on
plot(x,t2,'b+')
xlabel('p 分位数')
ylabel('条件保证所需时间')
title('不同路径的时间')
legend('快速路','城区路','Location','best')
```

