北京林业大学

本科毕业论文

（普通高等教育）

论文题目	基于贝叶斯网络的幸福感数据分析		
学　　院	经济管理学院		
专业名称	统计学		
班　　级	统计 13 班		
学　　号	130634112		
姓　　名	张晗雨		
指导教师	王兰会	**职　　称**	副教授
指导老师	胡大源	**职　　称**	教授

基于贝叶斯网络的幸福感数据分析

统计 13 张晗雨
指导教师 王兰会 胡大源

摘要

本文基于贝叶斯网络，在大样本下，分析、估计和推理不同因素相互关联作用下的幸福感。与现有研究幸福感方法相比，贝叶斯能够从大量复杂的数据中发现知识和结构，对不确定性进行探索，且对线性等统计假设要求较少。结构学习、参数学习和推理，是基于贝叶斯网络对幸福感研究的三个步骤。本文首先对幸福感数据进行结构学习，得到幸福感的父节点为预期收入、是否有收入困难问题和婚姻状态，也即影响幸福感的关键变量与收入和婚姻状态相关；其次对已确定的幸福感贝叶斯网络进行参数估计，估算出幸福感各离散变量的分布概率，发现预期收入对幸福感的条件概率分布影响较大；最后在幸福感贝叶斯学习的基础上，加入先验信息，对其幸福感状态进行概率推理，探寻每个节点对幸福感的影响，发现预期收入和婚姻状态影响最为明显。

关键词：贝叶斯网络，条件独立性，幸福感

Analysis of Happiness Data Based on Bayesian Network

Statistics 13 Hanyu Zhang

Supervisor Lanhui Wang, Dayuan Hu

Abstract

Based on the Bayesian network, the sense of happiness under the influence of different factors is analyzed, estimated and inferred in this paper. Compared with the existing methods, Bayesian network can discover knowledge and structure from a large number of complex data, explore the uncertainty, and less demand for linear statistical assumptions. The happiness research based on Bayesian network is divided into three steps: structural learning, parameter learning and inference. Firstly, the structure of happiness data is studied in this paper. In this network, the father nodes of happiness are the expected income, whether there are income problems and marital status, that is, the key variables that affect happiness are related to income and marital status. Secondly, The Bayesian network is used to estimate the distribution probability of the discrete variables, and it is found that the expected income has a great influence on the conditional probability distribution of happiness. Finally, based on the study of happiness Bayesian study, adding a prior information, probabilistic inference of its happiness, and exploring the influence of each node on happiness, that the impact of the expected income and marital status are the most obvious is pointed out in this paper.

Key Words: Bayesian Network，Conditional Independence，Happiness

目录

一 绪论

（一）研究背景

长期以来，经济发展近乎等同于经济增长，许多国家都把经济增长率等经济指标作为提高国民福祉的主要手段，然而过度追求经济利润最大化，使得社会付出高昂的代价，如生态成本、资源成本、社会成本等，甚至连整个人文发展情况也呈现逆态势。当前世界各国在关注经济发展的同时，也开始将注意力转向国民幸福问题，提升国民的幸福感作为一项涉及民生的重要工作。中央电视台“你幸福吗？”的调查采访节目，再一次引发国民对幸福感的关注。近三十多年来，我国经济发展水平已经显著提高，居民财富不断积累。经济总量方面，GDP 由 1978 年的 3645 亿增长到 2016 年的 744127 亿元，增长近 204 倍；人均收入方面，人均消费水平由 1978 年的 184 元增长到 2016 年的 17111 元，增加了近 93 倍。发展经济只作为一种国家强大的手段，其最终目的在于使其国民获得生存保障和幸福增进。我们彻底告别了短缺经济，居民生活水平大幅提高，可经济收入的增长与幸福感的提升是否一致？大量经验事实表明，中国人均产出的增长，并没有使幸福感出现相应提升，反而出现了令人担忧的下降趋势。“幸福悖论”现象是否已经在中国出现？幸福感及其众多影响因素的作用机制是怎样?这些都是值得关注和深思的问题。

（二）文献综述

有关幸福感的文献主要集中在以下三个方面：幸福悖论的提出及验证、幸福感及其影响因素的研究和幸福感相关理论与方法，下面将依次进行梳理。

1 幸福悖论的提出及验证

幸福及其决定因素长期以来是心理学的研究重点。主观幸福感被定义为人们对其生活质量上情感性和认知性的整体评价。幸福感是一种主观感受，没有客观标准，决定幸福与否不是实际发生了什么，而是对于实际情况所做出的情绪反应。自 1974 年 Easterlin 正式提出“幸福悖论”，即当国家变得更富有时，人们平均幸福水平并未随之相应提高，人均实际收入和幸福感不存在显著的正向关系[1]，对幸福感的研究开始纳入经济学范畴。

随着时间和数据推移，一些学者基于微观数据的实证研究也支持“幸福悖论”。Charles Kenny（1999）基于 1952 年到 1988 年的数据，发现在美国这一关系甚至是负相关的[2]。Easterlin（2010）利用 37 个国家的数据再次验证了悖论的存在性[3]。“幸福悖论”已在美国、日本、英国等很多发达国家出现，对中国的实证研究也有类似结论。改革开放以来，经济增长并不一定与居民幸福感同步提升（陈统奎等，2005）[4]。我国 1990 年国民幸福指数为 6.64，1995 年继续上升为 7.08，

但 2001 年反而下降为 6.60，即国民幸福的持续增加并不能为经济持续快速增长所保证（Veenhoven，2005）[5]。中国幸福感发展轨迹，与中东欧转型国家基本一致，为 U 型走势（丁云等，2013）[6]。

2 幸福感的影响因素研究

对于“收入增加而幸福没有增加”的悖论，国外学者们主要从两个视角进行研究。一是扩大影响幸福感影响因素的范围。Diener（1984）探讨事件、情境和人口统计变量（如年龄、性别）等外部因素如何影响人的幸福感[7]。Graham and Pettinato (2001) 认为收入或经济增长以外的因素，比如心理满足感、生活质量、健康水平等心理情感因素和失业与通货膨胀等社会环境因素，会显著地影响个人幸福水平。Deaton（2007）在幸福感的度量、幸福感与经济发展、年龄、身高、收入、健康、宗教、居住模型的关系等方面都进行了深入的研究[8][9]。Deaton（2010）提出主观幸福感包括情感福祉、生活质量评价和人生观三个方面，每一类都受到不同因素的影响[10]。问卷调查主要是考察生活质量评价，即人们对生活整体的满意程度，主要受到收入和教育的影响，受社会经济形态影响较大，经济情况好的地区人们的生活质量评价普遍偏高；情感福祉更多地体现人一天内情绪的情况，波动性更大，整体生活水平上升时不会受到较大影响，但会因为低收入而显著降低。另一个视角是对影响幸福感的最主要因素——收入进行更深入的挖掘，如绝对收入、相对收入以及收入的相对剥夺。Diener（2000）等研究发现绝对收入越高的群体，幸福感也越高，但当收入增加到一定水平，绝对收入的提高对幸福感的作用会变弱，与幸福感的相关程度降低，解释力有限。Easterlin（1995）强调相对收入对幸福感起决定作用。Graham and Pettinato（2001）通过受访者当前与过去的比较、未来经济状况的预期、所处经济地位的自我评价三个方面度量相对收入，发现相对收入对幸福感作用逐渐增强。Jones and Wildman（2008）探索得到非经济因素影响相对剥夺感，进而影响幸福感，且相对剥夺与幸福感之间呈反方向变动关系，即内心感觉良好、地位优越的个体其相对剥夺较低，从而主观幸福感较高，反之亦然[11]。

幸福感研究同样也引起国内学者的关注。任国强（2012）等研究发现与绝对收入相比，相对收入、相对剥夺对主观幸福感起决定性作用[12]。樊丹花（2013）通过构建结构方程模型发现，收入通过个体状况与社会状况作为中间变量对幸福感的间接影响大于直接影响[13]。黄嘉文（2013）在不同空间和时间条件下测度了教育回报对居民幸福感的影响[14]。鲁元平（2012）[15] 和徐映梅（2014）[16] 等学者也得出了如下的一些结论，幸福感与年龄呈现 U 型关系，已婚与幸福感呈正相关，男性比女性更快乐、阶层固化与机会不均等负向影响幸福感等等。

3 幸福感的研究理论与方法

国外对幸福感的理论分析和经验性分析主要以人格理论、相对理论以及适应理论展开，其中，相对理论又由期望值理论、社会比较理论以及目标实现理论三个分支构成（Easterlin，1995）。虽然经验和描述性分析能在一定程度上解释，但实证方法对于进一步研究十分重要。Alesina(2004)选用有序 Probit 模型处理有序变量相邻选项之间的距离问题[17]。Rojas（2007）建立多元线性回归方程验证收入对幸福感的影响。Jones and Wildman（2008）构建了参数、半参数面板数据模型。John Knight（2009，2010）使用中国 2002 年的全国住户调查数据库，采用简单线性回归模型，分别估计了农村幸福方程、城市幸福方程及作为农村人口子集的农民工幸福方程。

在幸福感涵义方面，我国学者受美国学者 Diener 影响很大，其阐释幸福感为个人根据自定标准对其生活质量的整体性评估，具有主观性、稳定性和整体性的特点。由于其定义涵盖了幸福感研究的主要内容，又便于进行调查研究，受到我国学者的认可（李志，2006）[18]。在理论研究方面，田国强和杨立岩（2006）通过构建一个规范的经济学理论模型—门槛模型，把攀比理论和“忽视变量”理论结合，来解释幸福悖论[19]。陆士桢（2011）以青年保安人员为研究对象，以质性研究方法对保安人员主观幸福感及其影响因素进行研究，解决了特殊群体样本量不足和调查问卷先入为主的缺陷[20]。吴丽民和陈惠雄（2010）和樊丹花（2013）构建“收入—中间变量—幸福”模型，并指出收入直接影响幸福感，或通过个体状况、社会状况等中间变量间接影响幸福感[21]。

大多数学者在前人研究的基础上，采用统计、计量分析方法进一步做实证研究。王忠郴（2007）利用层次分析法建立幸福感程度等价评价指标体系，然后对该体系从下层到上层分别进行模糊综合评判[22]。何立新（2011）采用有序 Probit 模型在全样本、收入分层和城乡分层中研究收入差距和机会不均对幸福感的影响，并提出一种处理内生性问题的方案[23]。张良桥（2014）以经济先发居民为研究对象，通过建立半参数及分位数回归模型，系统研究绝对收入、相对收入、期望收入对幸福的影响[24]。莫世亮（2014）等人将数据挖掘中的 CHAID 方法引入社会学，探讨了高校青年教师的职业幸福感状况及其影响因素[25]。

从研究内容来看，从经济、人口学、家庭、工作、人际关系和情感等内外部因素解释幸福感问题一直引发关注，尽管已经建立了比较全面的理论框架，数据来源也得到补充和完善，但不同国家的幸福感研究结果也是有所区别的，到目前为止幸福感的研究在不同国家，甚至在同一国家内都很难达到统一的结论。从研究方法来看，从经验事实、统计描述、质性研究，到复杂计量分析方法均有涉及，甚至有少量文章引入数据挖掘的方法。目前对幸福感的研究方法各有优劣。传统的计量模型，如广泛使用的最小二乘法（OLS）、有序 Probit 及分位数回归模型等，尽管对待估参数的解释更为方便直观，但是对变量间相互作用识别不够，忽略了幸福感各影响因素之间的交互影响，无法同时处理多个因变量，也无法同时得到各变量之间的结构关系。结构方程模型（SEM）包含测量模

型和结构模型，可以识别幸福感和其影响因素之间的直接或间接关系，允许相对弹性的模型设定，并同时检验一系列回归方程，但是无法利用验证的结构与传导机理对其幸福感进行推理。目前部分幸福感研究也引入数据挖掘的方法，如 CHAID，其针对幸福感这一结果变量，对其影响因素分成一系列二维分类表，找到最佳分类变量及结果，从而分析出哪些群体更可能幸福或不幸福。虽然此类方法有助于推理出一个人的幸福状态，但以数据为导向却无法解释各分类节点是否有因果关系，且其在推理过程中无法加入先验信息。概率图模型结合概率和图论的知识，利用图来表示与模型有关变量的联合概率分布，已成为不确定性研究的热点。作为一种重要的概率图模型，贝叶斯网以随机变量作为节点，通过有向无环图，用网络中节点来表示概率依赖关系，借助条件独立性来简化概率推理，为各种随机现象建模分析提供了有效工具[26]。以变量间相互作用为前提、通过结构学习、参数估计以及概率推理来获取结果分布信息的贝叶斯网络已广泛用于风险分析、可靠性分析、评价问题等，顾客满意度研究[27]等社会经济问题也有所涉及，研究显示贝叶斯网络在不确定性概率分析具有很大优势，然而，利用贝叶斯网络对幸福感的分析和研究却为数不多。

对于幸福感在中国的研究，特别是对影响因素的分析还存在以下难点和不足：首先，幸福感的人口社会特征之间的关系是极其复杂的系统问题，影响因素繁杂与作用机制模糊给研究带来了困难。目前多依靠抽样调查完成实证分析，且问卷数据以分类数据居多，传统的线性模型、计量模型对变量间相互作用识别不够，忽略了各因素之间的交互影响，结果可能会产生一定偏差。其次，缺乏大样本实证证据及实验支持。数据来源大多是国际数据，而非中国样本，即使为中国样本，结论也受到研究对象群体数量、地域及层级的局限性。而这些问题是本文需要解决的，也正是本文的创新和贡献所在。

（三）研究目的及视角

直接测度主观福利的幸福感往往是多个因素相互作用的结果，表现出明显的层次性和不确定性表示幸福感因果关系的作用机制值得探索，尤其是在大样本容量下，适用于研究、分析、估计和推理不同因素相互关联作用下的幸福感的方法是十分有必要的。哪些才是影响其幸福感的关键因素？这些因素和幸福感又是怎样的因果关系、作用机制？如何根据观察到的因素对一个人的幸福状态进行推理？这些问题都是值得深入研究的。与现有研究幸福感方法相比，贝叶斯能够从大量复杂的数据中发现知识和结构，且对线性等统计假设要求较少。因此，本文认为基于贝叶斯网络，对居民主观幸福感研究构建有效的不确定性知识结构、并进行概率的推理计算是可行且有必要的。

（四）研究内容

本文拟基于贝叶斯网络，在大样本下，分析、估计和推理不同因素相互关联作用下的幸福感。

首先对幸福感数据进行结构学习，构建出幸福感及其影响因素间的网络结构图；其次对已确定的幸福感贝叶斯网络进行参数估计，估算出幸福感各离散变量的分布概率；最后在幸福感贝叶斯学习的基础上，加入先验信息，对其幸福感状态进行概率推理。

本文其他部分结构安排如下：第二部分是数据和样本简介。详细比较了各数据集的优劣与可得性，阐述了本文数据来源，样本限定、变量描述和处理方法，包含数据清理和插补的过程。第三部分是理论与方法。首先阐释了贝叶斯网络理论，包括图、数据和模型的关系，用一个学生实例简单阐述其实质，并对贝叶斯网络建立的流程、逻辑及基本问题进行了总结；然后围绕幸福感对贝叶斯网络理论阐释。第四部分是幸福感实证分析。首先使用 R 对幸福感数据进行结构学习，介绍了重要概念、完整数据集的结构学习理论、算法及优缺点，构建出幸福感及其影响因素间的网络结构图，并用损失函数选择最优结构，并在此基础上论述了其结构及相互作用的合理性，分析各变量之间的依赖关系和条件独立性；然后在幸福感结构图基础上进行参数估计，得到各个节点的概率，并与实际计算结果对比，验证了其贝叶斯网络估计参数的精度；最后利用联合树算法对幸福感进行贝叶斯网络的概率推理分析，具体分析了幸福感节点处的推理概率，并讨论不同节点信息的加入对幸福感的推理概率的影响。第五部分是结论与讨论。

二 幸福感数据及样本选择

为了表示幸福感多个因素的相互关系，探索幸福感因果关系的作用机制，尤其是在大样本容量下，研究、分析、估计和推理不同因素相互关联作用下的幸福感，本文拟基于贝叶斯网络，对幸福感数据依次进行结构学习、参数估计和推理分析。目前中国学者研究幸福感使用数据库有世界价值普查（World Value Survey）、中国综合社会调查项目（CGSS）或是学者针对某一地区、某一特定人群采集的一手数据。研究幸福感数据来源大多是国际数据，而非中国样本，即使为中国样本，结论也受到研究对象群体数量、地域及层级的局限性。因此，本文希望通过更权威、更匀质、最新的大样本数据对幸福感进行深度研究。

（一）数据库介绍

本论文使用“CCTV 经济生活大调查”项目数据。该项目由是中央电视台联合国家统计局、中国邮政集团公司推出的年度经济调查活动，调查内容主要包括全年经济印象、百姓收入支出、投资理财、市场热点、民生焦点等。旨在了解及剖析百姓生活各个方面，为政府决策提供指导性意见，是国内覆盖面最广的民间调查。

项目自 2006 年起，每年年底组织一次，每次约发放 10 万张问卷，已经连续组织 11 年（即 2006 年至 2017 年）。由于每年我国的经济社会状况以及影响国家百姓经济生活的重大热点事件不同，项目每年问卷结构、问题设计都有较大变动，故本文仅使用 2016 年的截面数据。

（二）数据调查方式

调查问卷以“特殊的明信片形式”和“中国邮政网络”两个渠道进行抽样调查，前者利用明信片小巧方便、容易寄递的特点有助于调查的快速回收；后者可利用邮政便捷的服务和投递网络的优势保证调查的顺利进行和回收率，并且节省了调查费用。通过以上两种渠道，调查覆盖了全国 31 个省、直辖市和自治区、104 个城市、300 个县，共计 10 万个中国家庭。以期最大限度使样本更广泛、更匀质，尽量避免由于交通便利差异、贫富差异所导致的样本有偏性，以便有效地反映和估计整个总体特征。调查问卷通过中国邮政回收后由国家统计局负责数据录入分析。

样本数量看，2016 年调查发放十万份，2016 年回收 88415 份，回收率达 88%。

（三）样本结构

调查样本中覆盖了中国居民的各个年龄段、各种收入水平、各种文化程度、各种职业及各常驻地区，2016 年的样本结构显示：

样本分布以较发达地区为主，中西部样本较少；

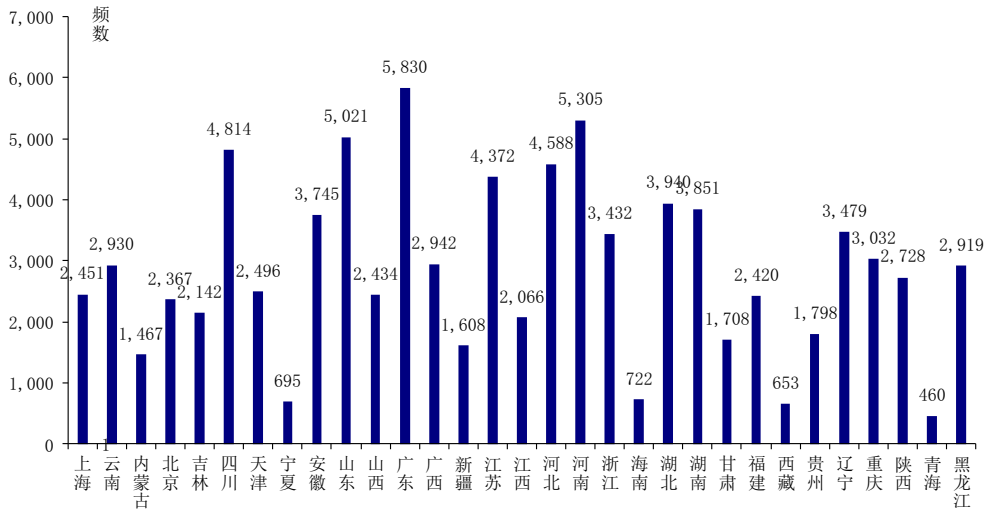

图 2.1 样本地区分布

Fig2.1 The distribution of sample area

按样本来源以城镇为主，69%样本来自城镇，31%来自农村；

按年龄结构偏青中年人，其中 50%为 18-35 岁，47%为 35-59 岁，3%为 60 岁以上；

按教育程度看，10%为小学及以下，40%为中学，50%为大专及以上；

按收入主要集中在年收入 5 万以下人群，其中 34%在 2 万及以下，32%在 2-5 万，31%在 5-10 万，3%在 10 万及以上。

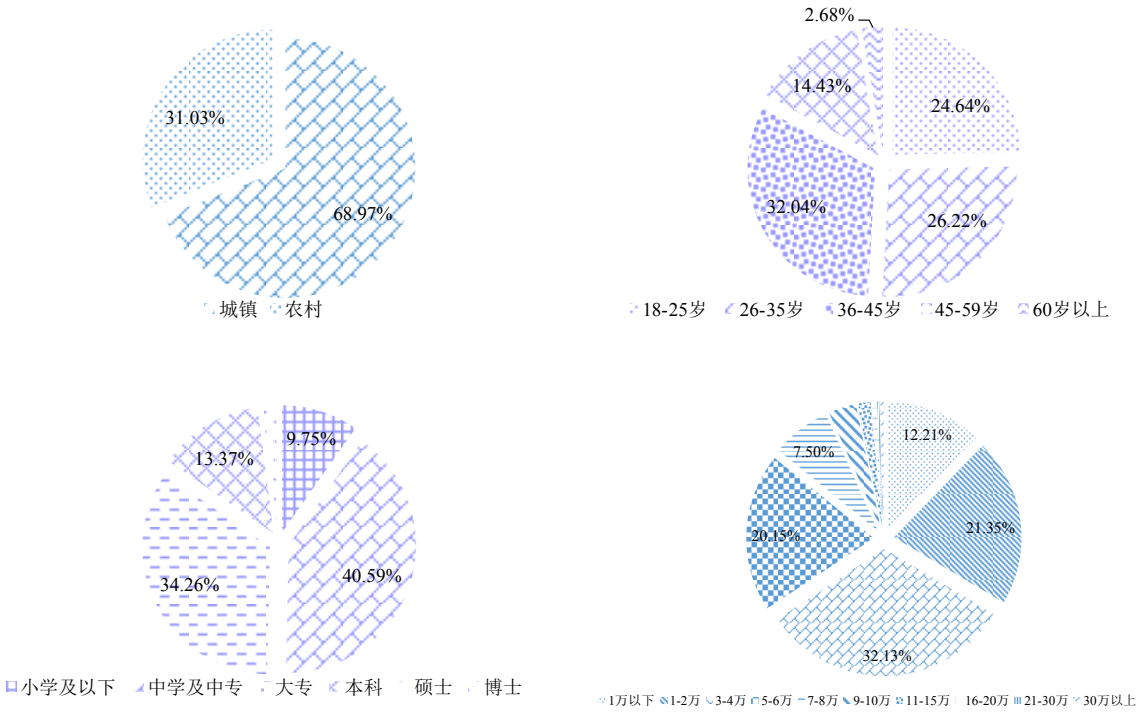

图 2.2 样本结构

Fig 2.2 Sample structure

（四）变量选择和数据预处理

本文使用“CCTV 经济生活大调查”的数据。主要用到的变量包括幸福感，收入，财富，个人基本情况等，具体的指标如下表所示：

表 2.1 变量描述

Table 2.1 Variable description

变量		取值个数	变量描述
生活幸福感		5	很幸福；比较幸福；一般；比较不幸福；很不幸福
预期收入		5	增加 20%以上；增加 10-20%；增加 10%以内；持平；减少 10%以内；减少 10 以上
是否有生活困难	住房困难	2	有；没有
	收入问题	2	有；没有
	医疗困难	2	有；没有
教育		6	小学及以下；中学及中专；大专；本科；硕士；博士
婚姻		5	未婚无恋人；未婚有恋人；已婚；离异；丧偶
职业		9	行政事业单位人员；企业管理人员；城市户籍企业职工；在校学生；务农农民；进城务工人员；离退休人员；待业/失业；自由职业者
常住地		2	城市；农村
个人收入		10	1 万以下；1-2 万；3-4 万；5-6 万；7-8 万；9-10 万；11-15 万；16-20 万；21-30 万；30 万以上
家庭年收入		10	1 万以下；1-2 万；3-4 万；5-6 万；7-8 万；9-10 万；11-15 万；16-20 万；21-30 万；30 万以上

根据问卷设计，我们把幸福指数分为 5 级，分别是很幸福、比较幸福、一般、比较不幸福、很不幸福。根据 2016 年样本，除去缺失数据的无效样本后，在 87810 份有效样本分布中，幸福指数呈现左偏分布，其中 17.92 的人很幸福，比较幸福占 33.95%，一般的占 37.38%，而比较不幸福和很不幸福的分别仅占 6.72%和 4.3%。

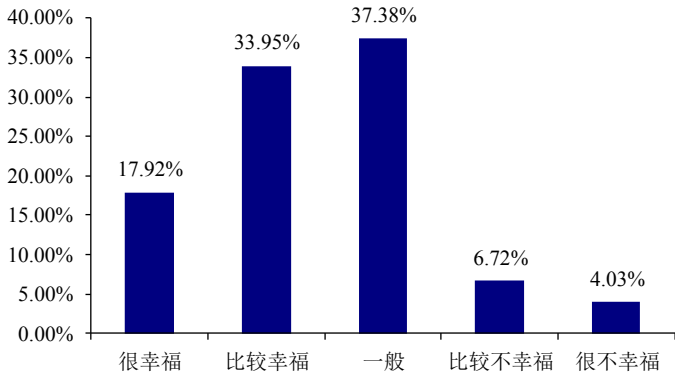

图 2.3 幸福感分布

Fig 2.3 The distribution of happiness

三 理论与方法

（一）贝叶斯网络概述

1 图、数据与模型

贝叶斯网络结合图和概率论的知识，将数据中变量间的概率分布对应到网络图中，更形象紧凑地展示各个变量间的依赖关系和条件独立关系，并形成一条判断网络。首先，我们先理解一下图形、数据以及模型之间的关系。如下图所示，贝叶斯网络的有项无环图（G）中的节点表示数据中随机变量$\{X_1，X_1，\ \ldots，X_n\}$，连接两个节点的箭头表示两个随机变量有因果关系或非条件独立，从而形成变量间概率分布的紧凑表示，即通过链式法则将变量间的联合概率分解成各自的局部条件概率分布相乘，实现数据到图的对应。简言之，把所要研究问题中涉及的随机变量，根据条件独立性绘制在一个有向图中，就形成了贝叶斯网络。

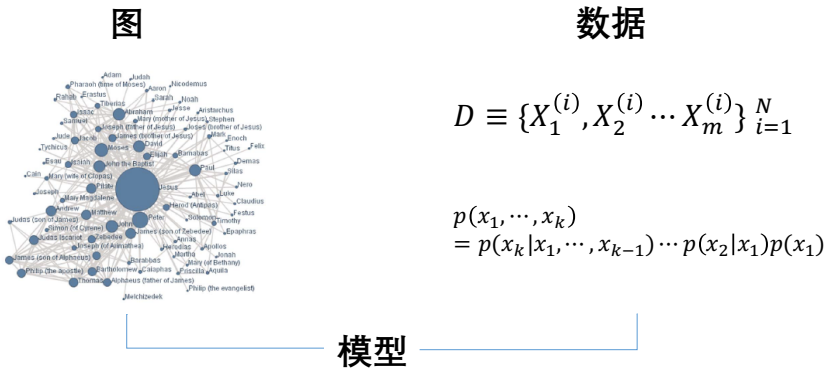

图 3.1 图、数据与模型

Fig 3.1 Graph, data, and model

2 示例

贝叶斯网主要利用图形和概率分布的形式，将各种复杂关系给出一个系统模型。考虑一个简单的学生推荐信质量的例子。课程难度（D）和学生智商（I）同时影响考试成绩（G），SAT 成绩（S）仅取决于学生智商（I），推荐信的质量（L）仅依赖于考试成绩（G），该问题中五个随机变量，除了考试成绩可以取三个值，其他均为二值变量，总的来说，联合概率分布的取值总数为 2*2*2*4*2=48。当得知联合概率分布时，就可以根据学生的情况对推荐信质量进行查询和判断。那么，这 48 个取值的概率分布利用图结构进行表达，如图 3.2 所示，即为贝叶斯网络结构。从中可以推出独立关系如下：

$$(G \perp S|I)，(L \perp I|G)，(L \perp D|G)，(L \perp S|G)$$

根据这种独立关系进行因子分解，可以判断“课程难度大，学生智商高，考试成绩中等，SAT

成绩高且得到高质量推荐信”的概率，可以简化为只利用五个数值计算得到：

$$P(D=1)，P(I=1)，P(G=2|D=1,I=1)，P(L=1|G=2)，P(S=1|I=1)$$

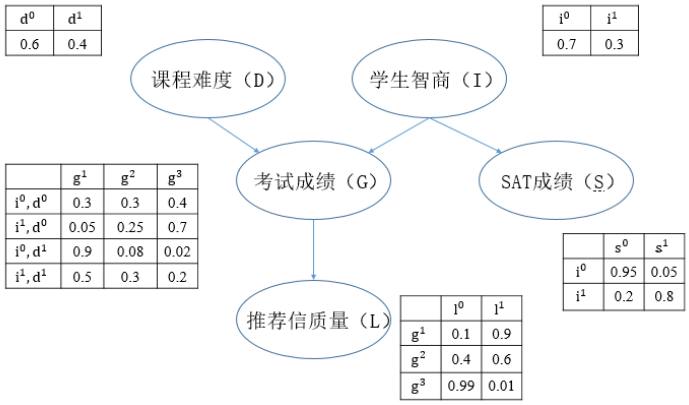

图 3.2 贝叶斯网络示例

Fig 3.2 An example for bayesian network

相对于原始联合概率分布的 47 个必要参数，加入条件独立性简化后的结构只需要 1+1+8+3+2=15 个必要参数。也就是说，条件独立性使得分布以因子形式更紧凑的表示，当数据中变量增大的时，参数简化的效果就会更加明显。由此可见，贝叶斯网络在于找到条件独立性，并以概率的关系选择主次关系，从而简化了复杂问题，进行分析和推理。以上是贝叶斯网络图结构应用的一个简单例子，其中，图结构刻画了随机变量的依赖关系和条件独立关系，节点代表随机变量，且每一个节点附有一个概率分布，节点之间的箭头则代表了随机变量间的统计关系。下面我们讨论贝叶斯网络图模型要探讨的问题和流程。

3 贝叶斯网络概述

贝叶斯网络可以将复杂的变量关系通过网络图形象表示出来，并通过条件独立性简化其联合概率分布，因此，贝叶斯网网络的应用主要是从不确定性中找到变量的可能关系，并根据可能的关系探讨要查询的问题。由此就引出三个最基本的问题，如何从复杂的不确定性的现实问题中找到主要的知识、假设和影响因素？如何根据数据建立“正确”的模型？如何根据建立好的模型和已观测到的数据来回答要推理的问题？

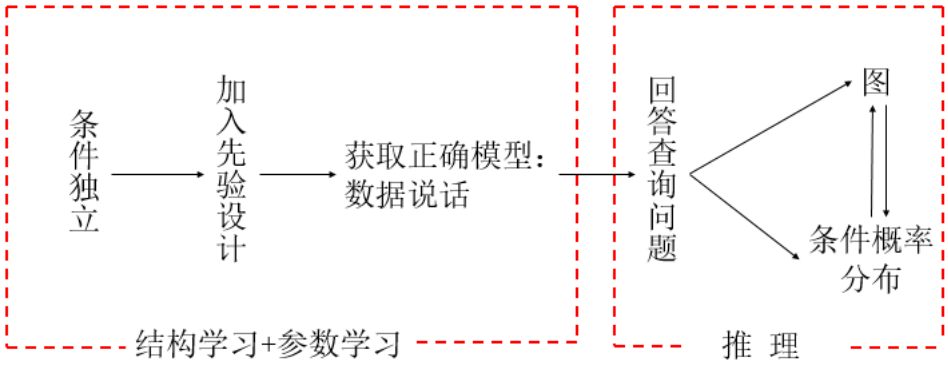

图 3.3 贝叶斯网络概述

Fig 3.3 Bayesian Network Overview

解决前两个问题，一是要求在建立模型前，研究者已经对研究的问题有一定的认识，根据前人文献和专家知识设定一定的条件独立关系，对已经确定的事实给出先验设计；二是研究问题中不确定的部分需要根据数据信息决定模型，即采取某种方法来构造网络，并在已有观测值中，根据极大似然估计或贝叶斯估计来选择最优结构图。 第三个基本问题也即贝叶斯网络的目的，即利用条件独立关系找到变量间的主次关系和联合概率分布，那么推理问题就是在学习的结构图基础上，给定变量的观测值，探讨其他变量的概率分布。由此可见，贝叶斯网络图是根据概率选择主次关系，并遵循“设定依据现实经验、方法是利用数据进行估计、目的是获得概率分布”这一逻辑来研究不确定性问题。

（二）贝叶斯网络理论阐释：以幸福感为例

1 理论阐释：图

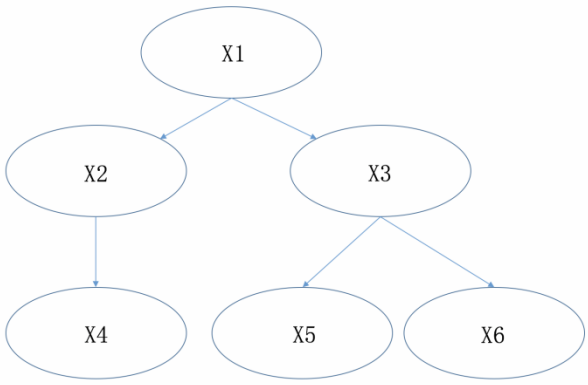

图 3.4 贝叶斯网络

Fig 3.4 Bayesian Network

准确理解一个完整的贝叶斯网络，需要把握以下三点含义：

- 节点表示随机变量。
- 节点之间的有向边表示随机变量的相互依赖关系，且每一个节点都附有一个分布，其中，根节点代表的是其边缘概率分布，而其他节点是条件概率分布。
- 数据中随机变量的联合分布，可通过链式法则，简化成条件概率连乘的形式，从而的到分布的分解，实现贝叶斯网络的紧凑表示。

因此，分解幸福感数据的联合概率分布，根据条件独立和依赖关系绘制在一个有向图中，这种贝叶斯网络降低了概率模型的复杂程度，也更加直观易懂，方便根据概率选择影响幸福感的主次因素，从而进行深层次的研究。

2 贝叶斯网络的结构学习

用贝叶斯网络对幸福感数据进行结构学习，目的在于构建一个有向无环图，用于表示网络中幸福感及其影响因素之间的概率依赖关系和条件独立关系。一般需要在复杂度和模型精度之间寻找平衡：一边需要构建详尽的网络拓扑结构，以获得对幸福感数据集足够的学习精度；另一边要求构建的模型应尽可能简单，以降低幸福感数据构建网络的结构成本和复杂度。

基于评分搜索的方法、基于约束的方法和混合学习方法，是对幸福感数据进行结构学习的三类算法。基于约束的学习算法视为约束满足问题，主要是通过检验幸福感及其影响因素的条件独立性来构建结构。其低阶数据处理简单，但是高阶的条件独立性检验很复杂而且结果不一定可靠。而基于评分搜索的学习算法视为结构优化问题，主要是利用得分函数评价每个幸福感网络结构优劣，然后用搜索算法来寻找出分数最高的最优结构以优化。其可以把专家经验知识以结构先验概率分布的形式融入到过程中，但是算法收敛速度慢，计算复杂，容易陷入局部最优问题。混合算法结合两种方法，可显著提高学习速度并可处理大型网络，一般先用幸福感各变量间条件独立性检验减少搜索空间或构造便利序列，然后再通过评分搜索算法对幸福感数据进行结构学习。

综上，对幸福感数据进行贝叶斯网络的结构学习，归纳起来都需要经过三个步骤：首先，设定某个“标准”，比如评分搜索方法中的评分函数，依赖分析方法中的独立性条件（CI）。其次，随机或者根据一些前提要求，确定一个幸福感各变量的初始图。最后，基于某种高效准确的算法规则，对幸福感及其影响因素之间进行加边、减边、换向操作，把第一步中设定的“标准”实现的最好，从而确定最终的网络结构。

3 贝叶斯网络的参数估计

幸福感数据的参数学习过程是在网络结构已知的情况下，从数据中学习幸福感及其影响因素的条件概率分布。条件概率分布的参数模型已预先指定，只需估计其中的参数，而极大似然估计和贝

叶斯方法是最常用的两种参数学习方法。极大似然估计把待估参数看作取值未知的确定性量，依据参数与数据集的似然程度，来选择使似然函数值最大的参数值作为学习的结果。贝叶斯估计是基于贝叶斯公式，根据样本信息修正先验信息，由先验知识和观察到的数据集共同决定不确定性的概率参数。

4 贝叶斯网络的推理分析

幸福感贝叶斯网络的拓扑结构给出了幸福感及其影响因素的联合概率分布，其推理是当某些变量给定其它变量的状态作为证据时，如何推断它们的状态，即已知某些变量的取值，计算另一些变量的后验概率分布。按证据变量和因果变量的不同逻辑关系，这种概率推理又可分成：从结果到原因的推理、从原因到结果的推理、同一结果的不同原因的关联推理以及包含以上三种的混合推理。

对幸福感贝叶斯网络概率推理的算法可分成两种：精确推理（exact inference）和近似推理(approximate inference)。精确推断在幸福感贝叶斯网络上重复应用局部计算的贝叶斯定理来得到条件概率或条件概率密度的精确值。最著名的精确推断算法是联合树算法，其是把贝叶斯网变换成一个联合树（junction tree）来执行再用原贝叶斯网的局部分布参数去计算联合树中复合节点的参数集合。近似推理以牺牲推导结果的精确度来换取推到效率的提高，最常用的是基于随机抽样的蒙特卡罗方法，即通过抽样得到一组满足概率分布的样本,然后用这些样本进行统计估计与计算。然而，但当网络中有新的证据变量加入时,尤其是证据变量的先验概率相对较小时,其收敛将会非常慢。因此，本文使用 R 对幸福感数据进行贝叶斯推理分析，主要采用的是精确推理算法中的联合树（Junction Tree）。

四 幸福感的实证分析

幸福感往往是多个因素相互作用的结果，涉及经济、人口学、家庭、工作、人际关系和情感等内外部因素，表现出明显的层次性和不确定性，但是哪些才是影响其的主要因素，这些因素与幸福感又是怎样的因果关系，如何根据已观测到数据对其幸福感进行推理？前面已经提及，贝叶斯网络根据概率选择主次关系，能够紧凑地展示变量间的条件独立关系，并形成一条清晰的推理网络。此外，选用贝叶斯网络还可以减少参数的估计量并加入研究者的先验设计，更符合认知世界，探讨幸福感的逻辑。因此，本文选择贝叶斯网络对幸福感进行实证分析。

（一）幸福感数据的贝叶斯网络结构学习

为了找到影响幸福感的主要因素，并将其相互关系直观地表现出来，首先要依据数据集寻找到幸福感及其各影响因素的最优结构，也是后续对不同因素相互关联作用下的幸福感进行分析、估计和推理的基础。因此，本节使用 R 对幸福感数据进行结构学习，构建出幸福感及其影响因素间的网络结构图，并用损失函数选择最优结构，并在此基础上论述了其结构及相互作用的合理性，分析幸福感各变量之间的依赖关系和条件独立性。

1 算法实现及选择

本文使用 R 对幸福感数据进行贝叶斯网络学习，并通过损失函数确定最优网络结构。测试共计 6 种算法，分别是基于约束的 Grow-Shrink 算法和 Incremental Association 算法，基于评分搜索的 Hill Climbing 算法和 Tabu Search 算法，以及混合算法中的 Max-Min Hill Climbing 算法和 Restricted Maximization 算法。评分搜索方法使用的评分函数是 BIC 评分函数，其核心思想是找到可以让使用的数据集出现概率最大的网络，依赖分析方法是通过相互信息（Mutual information）来量化独立性，其核心是独立性测试。

在进行完整数据集贝叶斯网络学习前，首先使用算法对数据进行测试，即不对结构设定任何先验条件，仅通过对样本数据的学习，由机器算法自动获取网络结构。然而这种方法对数据的依赖度太高，不同算法的差异较大，且会因为过度追求最好“标准”舍弃一些变量和条件关系，容易与所研究的问题不符，因此一般选择数据融合的方式，将专家认识作为先验知识，再用算法对样本数据学习，以期避免单一方法确定网络结构的弊端。由文献综述得到理想的贝叶斯网络图应该是经济因素和非经济因素共同决定幸福感，也是我们通过贝叶斯网络学习希望得到的结构图，因此，对于本文的研究问题，幸福感作为最终被验证的变量，即其作为其他变量的子节点出现，而非父节点。基

于此，我们给出网络结构一个简单的先验条件，即将“幸福感⟶其他变量”此路径设为不通，这一先验条件仅因文章研究目的所定，不是对变量间的依赖关系的假定，不会对后续结构和参数学习造成较大的影响。

使用上述六种算法对样本数据进行学习，考虑到算法的不稳定性，进行十次交叉验证，并对得到的对数损失函数取均值，以此评估模型好坏。损失函数一般是用来估计模型的预测值与真实值的不一致程度，损失函数越小则模型的稳定性越好，本文采用离散模型，因此选择对数似然函数。比较结果如下图所示：

表 4.1 各算法的损失函数比较

Table 4.1 Comparison of loss functions for each algorithm

类别	算法	损失函数	得分排序	稳定性排序
基于约束	Grow-Shrink	13.26908	2	5
	Incremental Association	13.09093	3	4
基于评分搜索	Hill Climbing	12.43564	6	1
	Tabu Search	12.43592	5	2
混合学习	Max-Min Hill Climbing	12.62837	4	3
	Restricted Maximization	13.33626	1	6

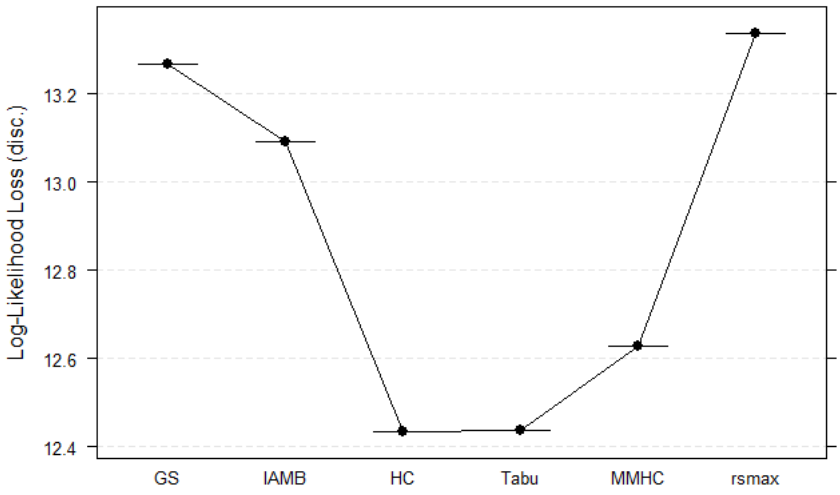

图 4.1 各算法的损失函数比较

Fig 4.1 Comparison of loss functions for each algorithm

可以发现，根据损失函数得分的排序为 Hill Climbing < Tabu Search < Max-Min Hill Climbing < Incremental Association < Grow-Shrink < Restricted Maximization。因此选择基于评分搜索的算法最有，本文选择 Hill Climbing 算法得到的贝叶斯网络图进行分析。

2 贝叶斯网络结构学习结果

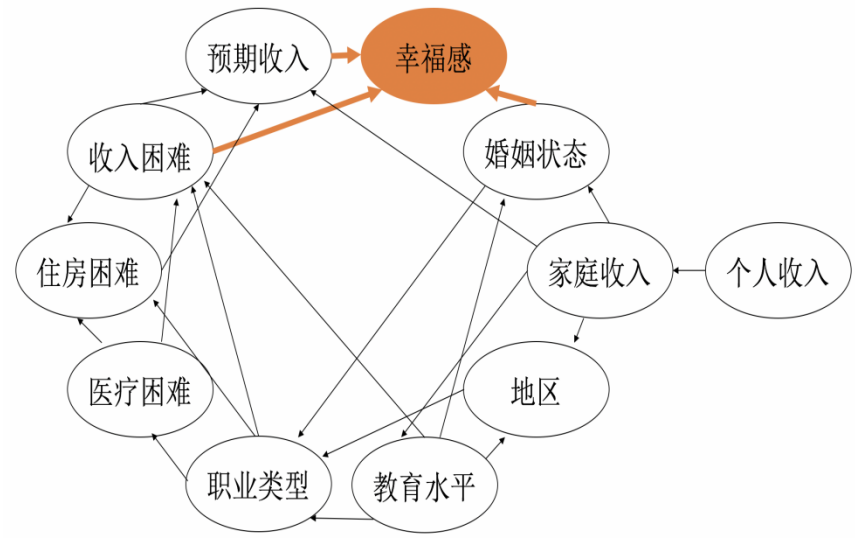

图 4.2 基于 Hill Climbing 的贝叶斯网络结构

Fig 4.2 Bayesian Network Structure Based on Hill Climbing

基于爬山评分搜索算法构建的贝叶斯网络结构主要节点如下：

[幸福感|预期收入，是否有收入困难，婚姻状况]

[预期收入|是否有收入困难，是否有住房困难，家庭收入状况]

[是否有收入困难|是否有就医困难，职业类型，教育水平]

[婚姻状况|教育水平，家庭收入状况]

从图中可以看出，变量有层次地呈现出影响和被影响关系。在“幸福感”所在节点处，父节点为“预期收入”，“是否有收入困难”，“婚姻状况”。当父节点已知时，可以认为其他变量与幸福感是条件独立的，即仅通过父节点就可以判断幸福感的状况；当父节点数据无法获得时，其他变量如家庭“收入状况”、“职业类型”、“教育水平”等也会通过观察不到的父节点变量影响到幸福感的状况。

此外，图中的其他节点也符合事实逻辑。如“预期收入”直接受到“是否有收入困难”，“是否与住房困难”和“家庭收入状况”的影响；“是否有收入困难”直接受到“是否有就医困难”，“职业类型”和“教育水平”的影响；而“婚姻状况”也直接受到“教育水平”和“家庭收入状况”影响。以此类推，直到没有其他父节点的“个人收入状况”，即可以得到多条影响“幸福感”的路径，如“个人收入状况→家庭收入状况→婚姻状况→幸福感”，“个人收入状况→家庭收入状况→预期收入→幸福感”，或者“个人收入状况→家庭收入状况→教育水平→职业类型→是否有收入困难→幸福感”等。即基于贝叶斯网络的幸福感网络结构，以及结构背后的依赖关系和条件独立性，均与事实认知相符，其结构学习的结果是合理可信的，也即影响幸福感的关键变量与收入和婚姻状态相关。

（二）幸福感数据的贝叶斯网络参数估计

从幸福感的贝叶斯网络结构图中可以发现影响幸福感的关键变量与收入和婚姻状态相关，为了进一步探究影响作用大小，本节在确定好的幸福感结构图基础上进行参数估计，得到各个节点的概率，并将参数估计得到的后验概率与实际计算结果进行了对比分析，进一步验证了其贝叶斯网络估计参数的精度。

基于爬山评分搜索算法构建出贝叶斯网络结构后，就可以估算出幸福感及其他所有离散变量的条件概率。本文表 4.2 列出了幸福感在婚姻状态、是否有收入问题和预期收入共同作用下的参数学习结果，即后验概率。由于数据较多，本文仅列出幸福感这一变量的条件概率分布，参数学习结果如下：

表 4.2 幸福感节点条件概率分布

Table 4.2 Happiness node conditional probability distribution

婚姻状态	是否有收入问题	预期收入	幸福感				
			很幸福	比较幸福	一般	比较不幸福	很不幸福
未婚无恋人	否	增长 20%以上	0.813	0.072	0.083	0.018	0.015
		增长 10%-20%	0.15	0.432	0.282	0.094	0.042
		增加 10%以内	0.156	0.316	0.341	0.127	0.06
		持平	0.143	0.223	0.439	0.114	0.081
		减少 10%以内	0.092	0.177	0.45	0.133	0.148
		减少 10%以上	0.123	0.118	0.296	0.138	0.325
未婚无恋人	是	增长 20%以上	0.267	0.285	0.331	0.066	0.052
		增长 10%-20%	0.148	0.442	0.296	0.074	0.04
		增加 10%以内	0.123	0.291	0.444	0.098	0.044
		持平	0.125	0.254	0.483	0.08	0.058
		减少 10%以内	0.07	0.218	0.497	0.092	0.123
		减少 10%以上	0.09	0.137	0.448	0.118	0.208
未婚有恋人	否	增长 20%以上	0.383	0.306	0.218	0.057	0.035
		增长 10%-20%	0.157	0.495	0.256	0.06	0.033
		增加 10%以内	0.139	0.34	0.386	0.106	0.029
		持平	0.149	0.281	0.417	0.101	0.052
		减少 10%以内	0.061	0.233	0.424	0.147	0.135
		减少 10%以上	0.137	0.225	0.291	0.165	0.181
未婚有恋人	是	增长 20%以上	0.257	0.326	0.305	0.063	0.049
		增长 10%-20%	0.107	0.499	0.288	0.074	0.032
		增加 10%以内	0.114	0.336	0.426	0.084	0.04
		持平	0.108	0.279	0.475	0.095	0.042
		减少 10%以内	0.087	0.288	0.399	0.139	0.087
		减少 10%以上	0.054	0.197	0.435	0.136	0.177

表 4.3 续表

Table 4.3 Continued table

婚姻状态	是否有收入问题	预期收入	幸福感				
			很幸福	比较幸福	一般	比较不幸福	很不幸福
已婚	否	增长 20%以上	0.372	0.316	0.25	0.039	0.023
		增长 10%-20%	0.206	0.443	0.291	0.038	0.022
		增加 10%以内	0.18	0.402	0.338	0.057	0.023
		持平	0.192	0.316	0.423	0.049	0.02
		减少 10%以内	0.104	0.231	0.478	0.115	0.072
		减少 10%以上	0.136	0.208	0.421	0.095	0.14
已婚	是	增长 20%以上	0.272	0.341	0.337	0.03	0.02
		增长 10%-20%	0.151	0.439	0.35	0.043	0.018
		增加 10%以内	0.127	0.373	0.423	0.057	0.02
		持平	0.101	0.299	0.531	0.046	0.023
		减少 10%以内	0.094	0.253	0.488	0.105	0.06
		减少 10%以上	0.105	0.201	0.457	0.105	0.133
离异	否	增长 20%以上	0.358	0.225	0.299	0.064	0.054
		增长 10%-20%	0.121	0.374	0.353	0.082	0.07
		增加 10%以内	0.104	0.286	0.433	0.13	0.047
		持平	0.13	0.227	0.443	0.143	0.057
		减少 10%以内	0.093	0.198	0.419	0.209	0.081
		减少 10%以上	0.118	0.235	0.294	0.132	0.221
离异	是	增长 20%以上	0.221	0.243	0.354	0.083	0.099
		增长 10%-20%	0.152	0.37	0.351	0.104	0.024
		增加 10%以内	0.096	0.38	0.39	0.086	0.049
		持平	0.107	0.181	0.542	0.128	0.043
		减少 10%以内	0.11	0.154	0.407	0.165	0.165
		减少 10%以上	0.105	0.079	0.395	0.145	0.276
丧偶	否	增长 20%以上	0.258	0.21	0.371	0.081	0.081
		增长 10%-20%	0.129	0.233	0.388	0.164	0.086
		增加 10%以内	0.136	0.273	0.344	0.084	0.162
		持平	0.105	0.224	0.483	0.105	0.084
		减少 10%以内	0.086	0.171	0.343	0.143	0.257
		减少 10%以上	0.03	0.212	0.212	0.182	0.364
丧偶	是	增长 20%以上	0.245	0.286	0.286	0.061	0.122
		增长 10%-20%	0.16	0.42	0.24	0.1	0.08
		增加 10%以内	0.14	0.318	0.341	0.132	0.07
		持平	0.147	0.259	0.397	0.112	0.086
		减少 10%以内	0.071	0.25	0.357	0.143	0.179
		减少 10%以上	0	0.083	0.133	0.083	0.7

总体而言，给定婚姻状况和收入困难状态，预期收入较高的，其幸福感高的概率较大；给定婚姻状况和预期收入，有收入困难的，其幸福感高的概率较小；给定预期收入和收入困难状态，未婚无恋人幸福感高的概率最大，其次是未婚有恋人和已婚状态，而离异和丧偶幸福感高的概率较小。

对预期收入而言，第一，在未婚无恋人且不存在收入困难问题条件下，预期收入增加 20%以上，其生活感受很幸福的概率达 80%以上，随着预期收入减少，其很幸福的概率先大幅度减少，后平稳减少，如从“增加 20%”降到“增加 10%-20%”，很幸福的概率减少了 60%，随后从“增加 10%-20%”“增加 10%以内”“持平”“减少 10%以内”“减少 10%以上”很幸福概率小幅波动减少 1%-2%。第二，在未婚无恋人且不存在收入问题条件下，在各个幸福状态下，预期收入“增加 20%”其很幸福的概率最大，为 81.3%；“增加 10%-20%”其比较幸福的概率最大，为 43.2%；而对于“增加 10%以内”“持平”“减少 10%以内”的人，其幸福感一般的概率最大，大约为 30%-50%；最后对于“减少 10%以上”的人，其很不幸福的概率（32.5%）明显高于其他预期人群。上述分析对于其他固定婚姻和收入困难问题状态也适宜，即给定婚姻状况和收入困难状态，预期收入较高的，其幸福感高的概率较大。

对收入困难状态而言，给定婚姻状态和预期收入，看是否有收入困难问题如何影响幸福感。第一，对于未婚无恋人和预期收入“增加 20%以上”的人，其无收入困难的很幸福概率为 81.3%，高于其有收入困难 26.7%的概率，约 55%。第二，对于未婚无恋人和预期收入“增加 20%以上”的人，其无收入困难的很幸福的概率最大，有收入困难的一般幸福的概率最大。但随着预期收入的降低，这种差别不明显，尽管概率不同，但两者幸福层级一致，即在预期收入降低情况下，幸福感对有收入困难与否不敏感。如当预期收入“增加 10%以内”、“持平”和“减少 10%以内”，有收入困难和无收入困难都最大概率为一般幸福。可能是因为是否有收入困难问题，会直接或通过预期收入间接两种途径影响幸福感。

对婚姻状态而言，给定收入困难状态和预期收入，看不同婚姻状态影响幸福感程度多少。对于不存在收入困难，预期收入“增长 20%以上”，其幸福的概率（包含“很幸福”和“比较幸福”），从未婚无恋人、未婚有恋人、已婚、离异、丧偶逐步降低，分别是 88.5%、68.9%、68.8%、58.3%、46.8%，其中未婚有恋人与已婚区别不大。但对于存在收入困难，预期收入“增长 20%以上”，其幸福的概率发生变化，已婚的概率最高，为 61.3%，其次是未婚有恋人为 58.3%，然后是丧偶和未婚无恋人，分别为 53.1%和 55.2%，最低是离异概率为 46.4%。从侧面反应，没有收入困难时，未婚无恋人的幸福感高的可能性较大，而存在收入困难问题时，两人相互扶持使得已婚幸福感高的可能性更大。上述分析对于其他固定预期收入状态也适宜。

综上所述，在判断一个人的幸福感状态时，可以参考这样的逻辑:1.如果知道一个人的婚姻状态和收入困难状况，那么预期收入高的，其感到幸福的概率更高。2.如果知道一个人的婚姻状态，有收入困难的比没有收入困难感到幸福的概率低，但随着预期收入降低，两者差别变小。3.如果知道一个人的预期收入，没有收入困难时，未婚无恋人的幸福感高的可能性较大，而存在收入困难问题时，两人相互扶持使得已婚幸福感高的可能性更大，而离异和丧偶均以较低概率感到幸福。

为验证幸福感贝叶斯网络模型参数学习的精度，本文将参数估计得到的后验概率与实际计算结果进行了对比分析。表 3.3 仅列出在预期收入、是否有收入困难问题以及婚姻状况(未婚无恋人)共同作用下，幸福感概率分布的实际计算结果。通过实际计算结果和贝叶斯网络模型预测结果的对比分析发现，由贝叶斯网络得出的后验概率的绝对误差为均为 10^{-4} 级的，最大为 0.0005。说明贝叶斯网络模型具有很高的精确度，因此，运用贝叶斯网络对幸福感各个变量进行数据分析和结果推理是可行的。

表 4.4 幸福感节点实际计算结果

Table 4.4 Calculated results of happiness nodes

婚姻状态	是否有收入问题	预期收入	幸福感				
			很幸福	比较幸福	一般	比较不幸福	很不幸福
未婚无恋人	否	增长 20%以上	0.813	0.072	0.083	0.018	0.015
		增长 10%-20%	0.150	0.432	0.282	0.094	0.042
		增加 10%以内	0.156	0.316	0.341	0.127	0.060
		持平	0.143	0.223	0.439	0.114	0.081
		减少 10%以内	0.092	0.177	0.450	0.133	0.148
		减少 10%以上	0.123	0.118	0.296	0.138	0.325
未婚无恋人	是	增长 20%以上	0.267	0.285	0.331	0.066	0.052
		增长 10%-20%	0.148	0.442	0.296	0.074	0.040
		增加 10%以内	0.123	0.291	0.444	0.098	0.044
		持平	0.125	0.254	0.483	0.080	0.058
		减少 10%以内	0.070	0.218	0.497	0.092	0.123
		减少 10%以上	0.090	0.137	0.448	0.118	0.208

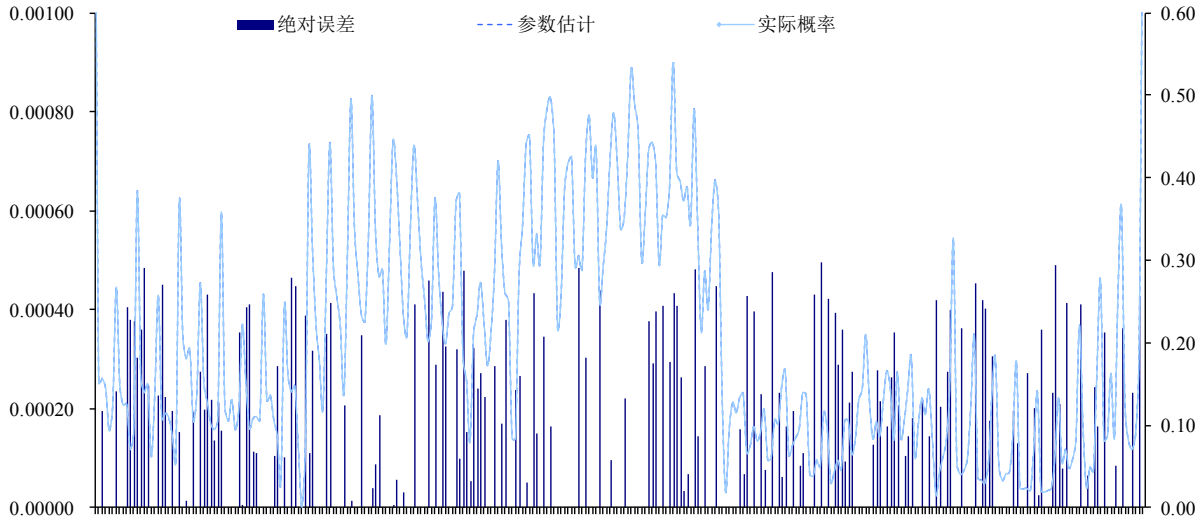

图 4.3 贝叶斯网络参数学习结果的绝对误差分布

Figure 4.3 Absolute error distribution of Bayesian network parameters learning results

（三）幸福感数据的贝叶斯网络概率推理

通过对幸福感数据的学习，本文发现影响幸福感的关键变量与收入和婚姻状态相关，且预期收入对幸福感的条件概率分布影响较大，接下来考虑如何根据已观测到数据对其幸福感进行推理？对幸福感贝叶斯网络的概率推理主要是其后验概率或密度的计算。例如，我们根据幸福感数据分布的假设已经学习了一个幸福感贝叶斯网络结构和参数，随后，我们想要利用贝叶斯网络的知识，计算在该幸福感数据集分布下一个新的“证据”（即信息）的影响，就是要去计算幸福感的后验概率。本节利用联合树算法对幸福感进行贝叶斯网络的概率推理分析，具体分析了幸福感节点处的推理概率，并讨论不同节点信息的加入对幸福感的推理概率的影响。

1 幸福感的推理分析

利用联合树算法对幸福感进行贝叶斯网络的推理分析，表 3.4 列出了预期收入、是否有收入困难问题、婚姻状态对幸福感的推理结果。此外，还可以推理出图 3.5 中任一节点或任多个节点已知条件下幸福感的条件概率。

表 4.5 预期收入、是否有收入困难问题、婚姻状态对幸福感的推理结果

Table 4.5 Expected income, whether there are income difficulties, Marriage status Reason for reason of happiness

变量	离散化取值	幸福感				
		很幸福	比较幸福	一般	比较不幸福	很不幸福
预期收入	增长 20%以上	0.380	0.286	0.266	0.039	0.028
	增长 10%-20%	0.164	0.443	0.314	0.053	0.026
	增加 10%以内	0.144	0.367	0.390	0.071	0.029
	持平	0.136	0.290	0.478	0.063	0.032
	减少 10%以内	0.093	0.237	0.469	0.117	0.084
	减少 10%以上	0.111	0.189	0.414	0.113	0.173
是否有收入困难问题	否	0.224	0.339	0.334	0.065	0.039
	是	0.139	0.346	0.417	0.062	0.035
婚姻状态	未婚无恋人	0.237	0.276	0.340	0.085	0.062
	未婚有恋人	0.151	0.356	0.359	0.087	0.047
	已婚	0.174	0.359	0.390	0.051	0.027
	离异	0.143	0.280	0.403	0.111	0.063
	丧偶	0.147	0.272	0.352	0.112	0.117

根据表 5.1 的推理结果，具体分析预期收入、是否有收入困难问题、婚姻状态对幸福感的影响。

（1）预期收入对幸福感的影响

根据建立的幸福感分析的贝叶斯网络模型和推理结果，即可得到不同预期收入下幸福感的概率分布：

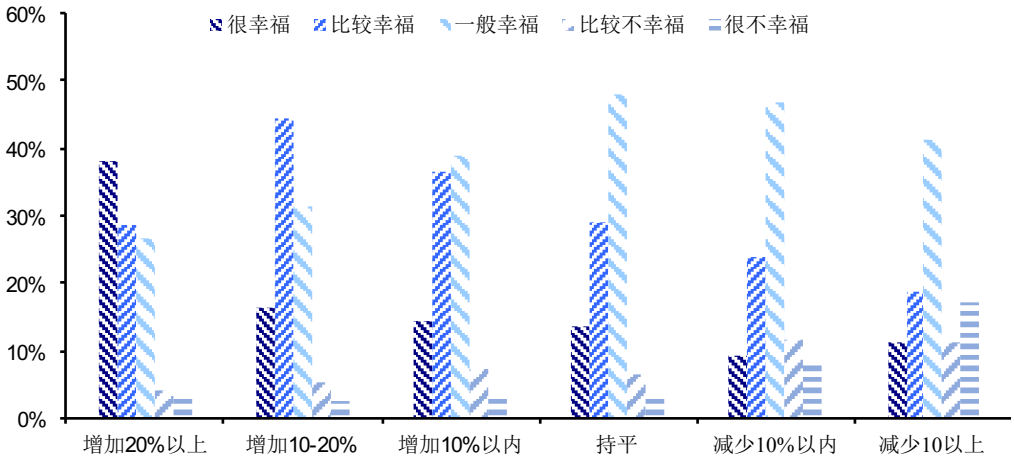

图 4.4 预期收入对幸福感的影响

Figure 4.4 Impact of expected income on happiness

具体来说，预期收入“增加 20%以上”很幸福的概率最高，其推理结果为 38%，其他幸福感状态的推理概率随幸福感减少而减少；“增加 10%-20%”比较幸福的推理概率最高，为 44.3%，其次是一般幸福概率为 31.4%；“增加 10%以内”“持平”等最大推理概率均为一般幸福，且不幸福状态

的概率随之增加。即，预期收入越高，其高程度幸福感的概率越大；随着预期收入的降低，高程度幸福感概率降低，不幸福的概率上升。预期收入是从相对比较的角度出发，拿未来的状况与当下做比较，当未来预期优于当下状况时，则会体验到幸福，反之则体验到不幸。

（2）是否有收入困难问题对幸福感的影响。

不同收入困难状态下幸福感的推理概率分布如下:

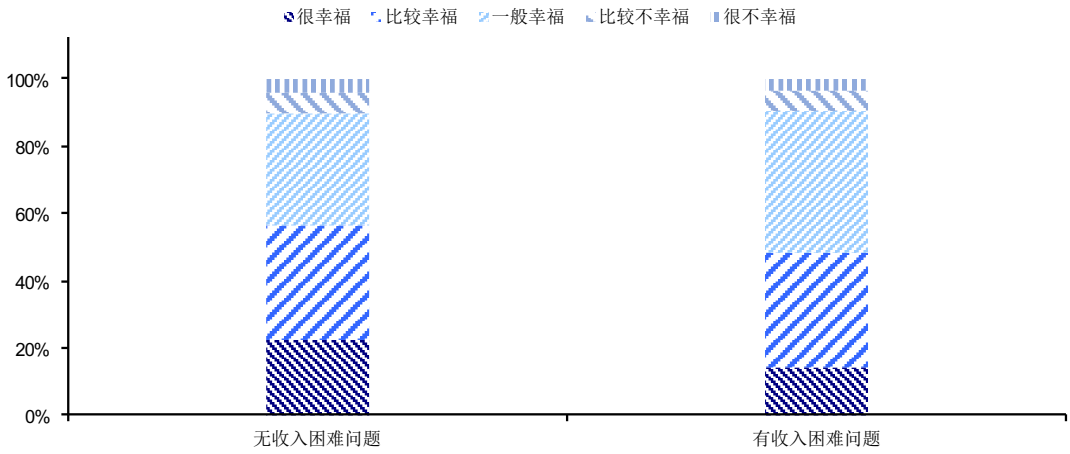

图 4.5 是否有收入困难对幸福感的影响

Figure 4.5 whether there are income difficulties on the impact of happiness

无收入困难问题其很幸福的推理概率为 22.4%，略高于有收入困难问题 13.9%的推理概率；而有收入困难问题其一般幸福的推理概率，高于无收入困难 8.3%；而比较幸福、比较不幸福和很不幸福的推理概率差异不超过 0.7%。即相比于有收入困难的，无收入困难的在高程度幸福感的推理概率较大，不幸福的推理概率两者差别不大。说明，解决收入困难问题，可以提高幸福程度，但对于不幸福的人来说，收入困难不是唯一问题，还有其他影响因素如婚姻等需要考虑。

（3）婚姻状态对幸福感的推理结果。

由建立的幸福感分析的贝叶斯网络模型和推理结果，即婚姻状态影响幸福感的概率分布：

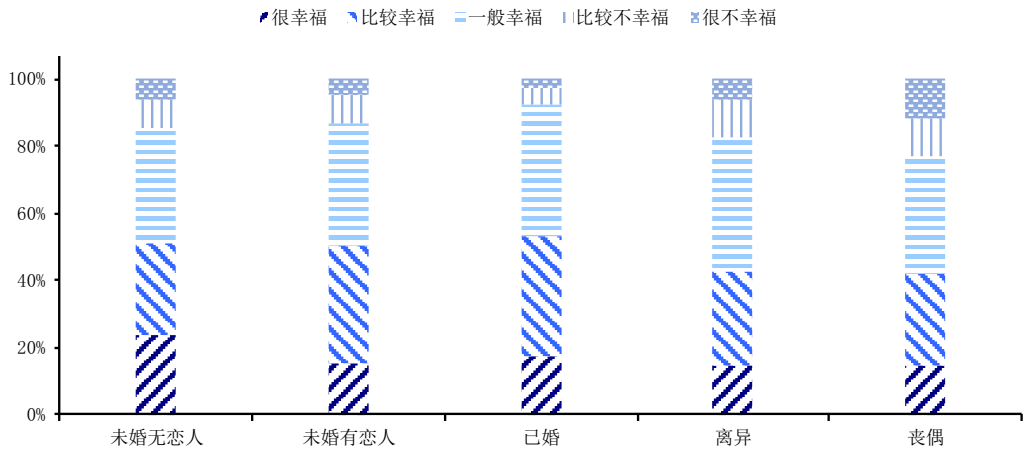

图 4.6 婚姻状态对幸福感的影响

Figure 4.6 The impact of marital status on happiness

“未婚无恋人”在很幸福的推理概率上，高于“已婚”，高于“未婚有恋人”，依次为 23.7%、17.4%、15.1%； “已婚”幸福的推理概率（很幸福、比较幸福和一般幸福）最高，为 92.2%，其次是“未婚有恋人”为 86.5%，最后是“未婚无恋人”为 85.3%。对于“离异”和“丧偶”，其一般幸福的推理概率均最大，很幸福和比较幸福的差别不大，很不幸福的概率，“丧偶”为 11.7%，“离异”为 6.3%。即，婚姻状态影响幸福感，“未婚无恋人”“未婚有恋人”和“已婚”其幸福的概率较大，其中，“未婚无恋人”很幸福的推理概率最大。此外，“离异”和“丧偶”会使得不幸福的推理概率较大。

2 对幸福感推理的进一步讨论

根据图 4.2，以上只分析了父节点“预期收入”、“是否有收入问题”及“婚姻状况”对幸福感的影响，下面我们通过一个例子来讨论不同节点信息的加入对幸福感的推理概率的影响。

首先看这样一个例子，不妨假设已知一个人的预期收入为“增长 20%以上”、“无收入困难问题”且“未婚无恋人”，根据已建构的幸福感贝叶斯网络进行分析，在已知父节点条件下，幸福感与网络中其他变量均为条件独立，因此得到其幸福（很幸福与比较幸福之和）的推理概率为 88.5%。倘若此时不知道其预期收入的信息，那么其幸福的推理概率降低为 56.8%，原因是前文分析过预期收入是幸福感重要的父节点，缺少“增长 20%以上”便以网络中预期收入的平均信息作为替代推理条件。此时，在未观察到预期收入的条件下，网络中其他变量与幸福感不再是条件独立，可以通过未观察到的预期收入影响幸福感，如“个人收入⟶家庭收入⟶预期收入⟶幸福感”，“职业类型⟶是否有住房困难⟶预期收入⟶幸福感”。但若给出预期收入的父节点（是否有住房困难、家庭收入状况）的信息，幸福感又条件独立于其他变量，因此，加入“无住房困难”和家庭年收入

为“30 万以上”，其幸福感推理概率上升为 59.5%。

同理，缺少“未婚无恋人”使幸福的推理概率下降到 58.5%，在未观察到婚姻状态的条件下，对于其父节点，我们可以观察到 “家庭收入⟶婚姻状态⟶幸福感”和“教育水平⟶婚姻状态⟶幸福感”。根据前文介绍到 d-分离概念，同样其子节点“职业类型⟵婚姻状态⟶幸福感”也可以影响到幸福感的推理概率。我们给出“在校学生”、“博士”信息，幸福感又条件独立于其他变量，其幸福的推理概率为 56.9%

若缺少“是否有收入困难”信息，其幸福的推理概率为 51.1%，幸福感在网络中的条件独立关系又会发生变化，其中“是否有医疗困难”有两种途径影响幸福感，一是“是否有医疗困难⟶是否有收入困难⟶幸福感”，二是根据 d-分离条件，在已知“无住房困难”条件下，“是否有医疗困难⟵是否有住房困难⟶是否有收入困难⟶幸福感”的路径也是可行的。若“是否有医疗困难”和“是否有住房困难信息”缺失，由于职业类型、教育水平、家庭收入已知，其所在地区及个人收入条件独立于幸福感。但若进一步家庭收入也缺失，那么网络中的条件独立会更加复杂。

表 4.6　不同节点信息的加入对幸福感的推理概率的影响

Table 4.6 Impact of the addition of different nodes on the probability of reasoning of happiness

序号	节点信息	幸福的推理概率
1	[幸福感|预期收入=增长 20%以上，是否有收入困难=否，婚姻状况=未婚无恋人]	0.885
2	[幸福感|是否有收入困难=否，婚姻状况=未婚无恋人]	0.568
3	[幸福感|是否有住房困难=否，是否有收入困难=否，婚姻状况=未婚无恋人]	0.601
4	[幸福感|是否有住房困难=否，家庭收入状况=30 万以上，是否有收入困难=否，婚姻状况=未婚无恋人]	0.595
5	[幸福感|是否有住房困难=否，家庭收入状况=30 万以上，是否有收入困难=否]	0.585
6	[幸福感|是否有住房困难=否，家庭收入状况=30 万以上，是否有收入困难=否，教育水平=博士，职业=在校学生]	0.569
7	[幸福感|家庭收入状况=30 万以上，教育水平=博士，职业=在校学生]	0.511
8	[幸福感|个人收入状况=9-10 万，教育水平=博士，职业=在校学生]	0.508

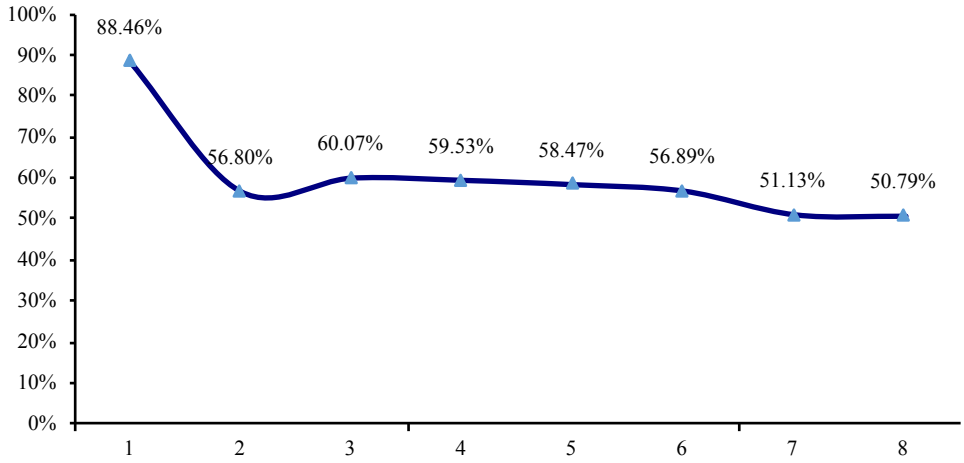

图 4.7 不同节点信息的加入对幸福感的推理概率的影响

Figure 4.7 The effect of the addition of different nodes on the probability of reasoning of happiness

五 结论与讨论

本文以研究幸福感数据中的不确定性为核心，基于贝叶斯网络给出了影响幸福感各因素之间相互关系建模及推断的方法。首先，介绍了数据和样本，详细阐述了数据来源，样本限定、变量描述和处理方法。然后，进一步明确理论与方法，阐释了贝叶斯网络理论，包括图、数据和模型的关系，用一个学生实例简单阐述其实质，并对贝叶斯网络建立的流程、逻辑及基本问题进行了总结，围绕幸福感对贝叶斯网络进行理论阐释。其次，根据初探模型确定先验条件，对贝叶斯网络进行结构学习，并根据损失函数选择稳定结构的算法（Hill Climbing），并在此基础上论述了其结构及相互作用的合理性，分析幸福感各变量之间的依赖关系和条件独立性。再者，在确定网络结构基础上，对参数进行估计，并将参数估计得到的后验概率与实际计算结果进行了对比分析以验证精度。最后，利用联合树算法对幸福感进行贝叶斯网络的精确推理。得出预期收入、是否有收入困难问题、婚姻状态对幸福感的推理结果，并进一步讨论了不同节点信息的加入对幸福感的推理概率的影响。

（一）主要结论

幸福感是经济与非经济因素相互作用的结果，各变量间有明显的影响与被影响的层次性关系。幸福感的贝叶斯网络结构充分体现了变量间的层次关系，借助条件独立性，根据概率选择主次关系，且有向无环图的表达形式形象直观，更加接近人认知的思维特征和推理方式。“幸福感”父节点为“预期收入”，“是否有收入困难”，“婚姻状况”。当父节点已知时，可以认为其他变量与幸福感是条件独立的，即仅通过父节点就可以判断幸福感的状况；当父节点数据无法获得时，其他变量如家庭“收入状况”、“职业类型”、“教育水平”等也会通过观察不到的父节点变量影响到幸福感的状况。

如果知道一个人的婚姻状态和收入困难状况，那么预期收入高的，其感到幸福的概率更高。如果知道一个人的婚姻状态，有收入困难的比没有收入困难感到幸福的概率低，但随着预期收入降低，两者差别变小。如果知道一个人的预期收入，没有收入困难时，未婚无恋人的幸福感高的可能性较大，而存在收入困难问题时，两人相互扶持使得已婚幸福感高的可能性更大，而离异和丧偶均以较低概率感到幸福。与实际计算结果对比分析，贝叶斯网络模型的后验概率具有很高的精确度，运用贝叶斯网络对幸福感各个变量进行数据分析和结果预测是可行的。

利用贝叶斯网络的知识，可以计算出该数据集分布下任意信息的影响。其结果表明，预期收入是从相对比较的角度出发，当未来预期优于当下状况时，则会体验到幸福，反之则体验到不幸。同时，解决收入困难问题，可以提高幸福程度，但对于不幸福的人来说，收入困难不是唯一问题，其

他影响因素如婚姻等需要纳入考虑。婚姻状态影响幸福感，“未婚无恋人”“未婚有恋人”和“已婚”其幸福的概率较大，其中，“未婚无恋人”很幸福的推理概率最大，而“离异”和“丧偶”会使得不幸福的推理概率较大。此外，不同节点信息的加入，会影响贝叶斯网络中的条件独立关系，进而对幸福感的推理概率产生影响。

（二）研究不足与展望

对于不确定性研究，贝叶斯网络根据概率选择主次关系，能够紧凑地展示变量间的条件独立关系，可以用来探讨影响幸福感的主要因素及其因果关系，并形成一条清晰的推理网络，根据已观测到数据对其幸福感状态进行推理。这一套完整的贝叶斯网络能够在幸福感数据分析中得到应用，为幸福感各因素相互作用建模及各因素相互影响作用推理，提供了一套可行性方案。

然而，由于文章是尝试性的使用此类方法并应用于幸福感数据分析中，存在以下局限问题：一是，由于幸福感是主观认知评价，没有办法剔除无效说谎样本，但是随着检测手段的提高和数据集的完善，会得到更为精确的分析结果，但在现有数据下，基于贝叶斯网络研究幸福感的方法仍然可行。二是，以目前已有的认知，学习结果中有部分依赖关系难以被直观解释，随着专家知识的进步，相应的数据处理过程及变量选择应在更加完善的算法下得到加强。若要更为精确的研究此类问题，需要结合专家知识，根据问题选择合适的方法，对模型加入更为严格的先验设计进行建模和分析，并探索更具一般性和高效性、支持任何概率推理的算法。

致谢

四年一瞬，感慨良多，念及别离，多有不舍，提笔小抒，自叹七幸。

一幸正逢韶华。学问如浩瀚宇宙，选择亦不胜枚举，未来尚未定论，变化无所不在。然恰同学少年，风华正茂，希望尚存，试错尚可。时进不可骄，时滞不可丧，亦歌于途，可休于树，何心有所惧？

二幸家人相伴。爱之亲之，养之育之，责之教之。吾家倡和而不同，虽有争辩也求同存异。自高中起，母之书信六十有余，言词真切乃日常小悟，不盼几多成功，但念女真实自由，寻心之所幸。

三幸得与良师。四年遇数位良师大家，恩言传身教面命耳提。恩师之提携，授东坡八面受敌之法，勿生余念，低调谦逊，厚积方能薄发。众师之关爱，得见识之开阔，学问之趣味，心境之豁达。

四幸总角之好存长久。相识相知十余年，暂别重逢仍牵挂，两眼相看无需答。谈天下论时事，聊人情说世故，虽相隔万里，情仍在意长存。

五幸新知变旧雨。四年得数友，悲不离，喜不妒，共历新生迷茫、繁重学业、彷徨惆怅、岔口选择，期间相扶相伴，终等得花开。

六幸难时终有助。学业之难，未来之思，生活之惑，感情之困。皆得师兄师姐之提点，同窗之善待，友人之鼓励，方能回首向来萧瑟处，也无风雨也无晴，心之感念，不胜言表。

七幸千帆过尽，归来仍是少年心。对知识的追求，对爱的渴望，对未知的敬畏，愿始终保持着对生命单纯而强烈的激情。

离别在即，不胜悲伤，然点滴在心，终不悔四年路。

参考文献

[1] Easterlin R.A. Does economic growth improve the human lot? some empirical evidence [J]. Stanford university Press,Paolo Alto,1974,89-125.

[2] Kenny,Charles. Does growth cause happiness or does happiness cause growth?[J].KYKLOS, 1999,52(1): 3-26.

[3] Easterlin R.A., Laura Angelescu.Modern Economic Growth: Cross Sectional and Time Series Evidence.Handbook of Social Indicators and Quality of Life Research, forthcoming

[4] 陈统奎,刘劭.从 GDP 到 GNH:中国经济增长但人民并不幸福[J].新民周刊,2005(41) .

[5] Veenhoven R. Apparent quality-of-life in nations: How long and happy people live[M].Quality of Life Research in Chinese, Western and Global Contexts. Springer Netherlands, 2005: 61-86.

[6] 理查德·伊斯特林,丁云,么莹莹.中国的主观幸福感研究(1990—2010) [J].国外理论动态, 2013(07):24-31

[7] Diener E, Suh E.M., Lucas R.E. Subjective well-being: three decades of progress[J].Psychological Bulletin,1999,125(2):276-302.

[8] Angus Deaton.Income, aging, health and wellbeing around the world:Evidence from the Gallup World Poll.2007. Web.

[9] Angus Deaton.The Financial Crisis and the Well-Being of Americans.NBER Working Papers 2011: n. pag. Print.

[10] Kahneman, D., A. Deaton.High income improves evaluation of life but not emotional well-being[J]. Proceedings of the National Academy of Sciences, 2010. 107(38): 16489-16493.

[11] Jones A.M.,Wildman J.Health,income and relative deprivation: Evidence from the BHPS[J].Journal of Health Economics,2008,27:308-324.

[12] 任国强,桂玉帅,刘刚.收入对主观幸福感的影响-国际的经验与国内的证据[J].经济问题探索, 2012(07):23-32.

[13] 樊丹花.收入与主观幸福感的关系研究[D]. 2013, 上海师范大学.

[14] 黄嘉文.教育程度、收入水平与中国城市居民幸福感——一项基于 CGSS2005 的实证分析[J].社会, 2013(05):181-203.

[15] 徐映梅,夏伦.中国居民主观幸福感影响因素分析——一个综合分析框架[J].中南财经政法大学学报, 2014(02):12-19.

[16] 鲁元平.中国“幸福—收入之谜”的作用机制研究[D]. 2012, 华中科技大学.

[17] Alesina A.,R. Di Tell,R. MacCulloch.Inequality and Happiness : Are Europeans and Americans different? [J].Journal of Public Economics,2004,88,2009-2042.

[18] 李志,谢朝晖.国内主观幸福感研究文献述评[J].重庆大学学报(社会科学版), 2006(04): 83-88.

[19] 田国强,杨立岩.对“幸福－收入之谜”的一个解答[J]. 经济研究,2006(11):4-15.

[20] 陆士桢,徐选国.青年保安人员主观幸福感及影响因素研究——从质性研究方法角度进行分析[J].青年探索, 2011(05): 49-55.

[21] 吴丽民,陈惠雄.收入与幸福指数结构方程模型构建——以浙江省小城镇为例[J].中国农村经济, 2010(11): 63-74.

[22] 王忠彬,徐辉.幸福感程度等级的二层模糊综合评判方法[J].企业经济, 2007(08):119-121.

[23] 何立新,潘春阳.破解中国的 Easterlin 悖论:收入差距、机会不均与居民幸福感[J].管理世界, 2011(08):11-22

[24] 张良桥.经济先发地区居民收入对幸福感影响的实证研究——基于非参数、半参数及分位数回归方法[J].西安航空学院学报, 2014(4):24-29.

[25] 莫世亮等.浙江省高校青年教师职业幸福感人群细分及其影响因素的研究——基于数据挖掘 CHAID 方法的应用[J].重庆电子工程职业学院学报, 2014(01):78-81.

[26] Koller D, Firedman N. Probabilistic Graphical Models:Principles and Techniques[M]. Cambridge: The MIT Press,2009.

[27] Kenett, S.S.R.S. Bayesian networks of customer satisfaction survey data. Working Paper n. 2007-33.