

生存分析法在学术期刊评价上的应用初探

李幸¹ 曾文¹ 范畅² 师兵范³ 望俊成¹

(1. 中国科学技术信息研究所, 北京 100038; 2. 武汉大学, 湖北武汉 430072;
3. 中国人民大学, 北京 100872)

摘要: 期刊评价中文献老化理论视角的研究相对薄弱。本文探究生存分析法在学术期刊评价上的应用, 在理论上克服了因领域差异导致的壁垒, 并选取CSTPCD中30家中物理学科期刊作为样本数据集进行实证研究。通过对论文被引时序数据进行分析, 研究物理学科文献老化规律, 并选取样本期刊进行对比分析, 以探究期刊评价的新思路。

关键词: 生存分析; 期刊评价; 文献老化; 被引时序分布; 生存时间函数

中图分类号: G350

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2018.01.012

Application of Survival Analysis in Evaluation of Academic Periodicals

LI Xing¹, ZENG Wen¹, FAN Chang², SHI Bingfan³, WANG Juncheng¹

(1. Institute of Scientific and Technical Information of China, Beijing 100038; 2. Wuhan University, Wuhan 430072; 3. Renmin University of China, Beijing 100872)

Abstract: The research on the aging of literature in journal evaluation is relatively weak and some scholars have applied survival analysis to the study of the law of information aging. Based on this, the author explores the application of survival analysis to journal literature. In theory, the barriers caused by differences in the field have been overcome. In practice, 30 Chinese physics journals in CSTPCD are selected as sample data sets for empirical research. By analyzing the cited time series data of literature, the aging law of the subject of physics is studied, and the sample periodicals are selected for comparative analysis to explore new ideas of periodical evaluation.

Keywords: survival analysis, journal evaluation, literature obsolescence, the distribution of citations in time, survival time function

0 引言

学术期刊是由学术论文组成, 论文质量决定

论文价值, 论文的价值大小直接影响到期刊的质量。近年来, 对论文质量的衡量研究也逐渐受到研究人员的重视^[1-3]。论文的价值可通过被引用

作者简介: 李幸 (1992—), 女, 中国科学技术信息研究所硕士研究生, 情报学理论与方法; 曾文 (1973—), 女, 博士, 中国科学技术信息研究所副研究员, 主要研究方向: 知识组织; 范畅 (1994—), 女, 武汉大学本科生, 主要研究方向: 信息分析; 师兵范 (1997—), 男, 中国人民大学本科生, 主要研究方向: 信息资源管理; 望俊成 (1984—), 男, 博士, 中国科学技术信息研究所副研究员, 中国科学技术信息研究所国家科技信息资源综合利用与公共服务中心副主任, 主要研究方向: 数据可视化、专利分析、情报学基础理论 (通讯作者)。

基金项目: 中国科学技术信息研究所创新研究基金面上项目“面向国家科技安全的情报分析方法体系研究”(MS2017-09); 国家社会科学基金项目“专利信息的生命特征揭示及老化规律研究”(12CTQ025)。

收稿时间: 2017年6月30日。

为体现,而被引行为随时间推移有兴衰变更,即论文价值的大小变化,体现出类似生物体的生命周期现象,论文价值大小的变化表现为论文的生命周期现象,是论文老化过程的体现^[4]。本文将在医学领域已得到广泛应用的生存分析法引入到期刊评价中,从论文层面入手映射到期刊层面,以期在期刊评价中信息老化理论的研究注入新的研究视角和思路。期刊文献老化速度或新颖性的测度指标是多样的,但对时间维度的纳入程度及方式却不尽相同。从共时与历时的分析角度来看,大部分指标属于共时分析,如普莱斯指数、引用半衰期,极少有指标采用历时分析的,如被引半衰期;从统计被引信息与时间的关系来看,绝大多数指标的方式是统计某段时间内期刊的被引量,较为宏观。这些衡量期刊文献老化、指标都只是对实际情况的一种理想化、概略性的测度指标^[5]。这就对文献老化理论的研究与发展提出了迫切的要求。

相对于已有的这些老化指标,本研究从历时分析入手,微观统计在监测时间段内期刊单位时间的被引频次分布数据,尝试性引入生存分析法,初步探究生存函数在期刊评价方面的应用,探索期刊文献的生命规律,以期在期刊评价工作注入新的视角。

1 生存分析方法

生存分析法源于医学领域,由于其在分析时序数据等方面的优势,逐渐被应用到多个行业领域,如工程学、金融业、保险业等。近年来,有学者将生存分析法应用于网络信息、专利信息等不同类型信息的老化规律研究^[6-9]。这为生存分析在期刊评价上的应用提供了理论基础与参照。

1.1 生存分析的含义

生存分析是指将终点事件和出现此事件所经历的时间结合起来进行分析的一种统计分析方法,是研究生存现象和现象的相应时间数据及其规律的方法^[10]。通俗来说,生存分析是研究一个或多个生物,在经过某一特定的时间后,将发生某种特定事件(称为终点事件)的概率,是一

种动态研究某个现象或因素随时间变化的规律的统计方法^[11]。生存分析法主要通过构建生存模型(生存函数、概率密度函数等),计算出各个时间点上的危险率,即所研究的对象的生存时间及其发生状态转变的可能性。

1.2 生命长度及生存数据

在生存分析法中,生存时间是指从某点开始到被观测对象发生终点事件所经历的时间。对于期刊论文来说,生存时间可以称为生命长度,是指论文自见刊之日起至失效(即失去利用价值)所经历的时间跨度。判断信息失效是生存分析法运用到期刊评价过程中至关重要的阶段。

在大多数研究中,实验的时间检测窗口是固定的,但研究对象进入检测窗口的时间是不同的,并且在检测时期内生存时间情况也不尽相同。因此,生存数据可以分为完全数据和删失数据。

(1) 完全数据:实验对象的起点和终点均落在时间窗口内。在本研究中,如论文的发表时间及失效的时间点都在2008年1月至2015年12月,该论文就属于完全数据。

(2) 删失数据:或称为结尾数据,包括左删失、右删失和区间删失。左删失数据指检测窗口起点前样本中已经发生了研究事件的数据。在本研究中,是指发表时间早于2008年1月的论文,在本研究的数据集中不涉及该类数据。右删失数据指在监测时间终点仍未发生终点事件的样本。在本研究中,是指在达到监测时间2015年12月时,仍未失效的期刊论文。区间删失数据指只能确定事件发生在某一区间内,但无法准确得知发生的时间点。在本研究中,并不涉及该类型删失数据。

1.3 生存函数

在生存分析法中,生存时间的分布常用下列4个函数来描述生存过程,即累积生存函数、分布函数、密度函数、危险率函数^[11]。这4个函数在数学上是等价的,给出其中一个函数,可以派生出其余3个函数。以下从论文角度对生存函数进行介绍。

1.3.1 累积生存函数

累积生存函数, 记作 $S(t)$, 指论文的生命长度 T 超过时间 t 的概率, 其函数表达式为:

$$S(t) = P(T > t) \quad (1)$$

其中, $S(t)$ 是时间 t 的非增函数, 并且具有 $S(0) = 1$, $S(\infty) = 0$ 的属性。

Berkson^[2]于 1942 年提出了 $S(t)$ 的图形表示, 称为生存曲线。生存曲线的走势表征生存规律变化情况。一条陡峭的生存曲线, 表示较低的生存率或较短的生命长度。一条平缓的生存曲线, 表示较高的生存率或较长的生命长度^[1]。生存函数或生存曲线可以用来确定生命长度的四分位数, 并且可以用来比较两个或多个数据集的生存分布。

1.3.2 累积分布函数

累积分布函数又称累积死亡率函数或累积风险率函数, 记作 $F(t)$, 表示论文的生命长度不超过时间 t 的概率, 其含义与生存函数相反, 其表达式为:

$$F(t) = P(T < t) = 1 - S(t) \quad (2)$$

1.3.3 概率密度函数

概率密度函数又称作密度函数, 记作 $f(t)$, 指某一特定事件发生于区间 $(t, t + \Delta_t)$ 内的概率, 它表示事件发生速度的快慢, 其函数表达式为:

$$f(t) = \lim_{\Delta_t \rightarrow 0} \frac{P(t < T < t + \Delta_t)}{\Delta_t} \quad (3)$$

$f(t)$ 的图像被称为密度曲线。密度函数有以下两个属性: 一是 $f(t)$ 为非负函数: $f(t) \geq 0$, $t \geq 0$; 二是密度函数与 t 时间轴之间的面积为 1^[1]。在任一时间段内的失效个体比率及失效频率峰值能够通过密度函数得到。

1.3.4 风险函数

风险函数又称瞬时失效率、条件死亡率或风险率, 记作 $h(t)$, 指当个体在 t 时刻存活条件下, 在以后的单位时间 $(t, t + \Delta_t)$ 内发生某一特定事件的条件概率, 其函数形式为:

$$h(t) = \lim_{\Delta_t \rightarrow 0} \frac{P(t < T < t + \Delta_t | T > t)}{\Delta_t} \quad (4)$$

风险函数也能通过累积分布函数 $F(t)$ 与概率密度函数 $f(t)$ 进行定义^[1]:

$$h(t) = \frac{f(t)}{1 - F(t)} \quad (5)$$

风险函数的曲线有多种形式, 如递增、递减、为常数等, 对风险函数唯一的限制是非负性, 即 $h(t) \geq 0$ ^[1]。

在实践中, 这 4 个函数可以用于说明数据的不同方面。生存数据分析的一个基本研究内容就是, 使用这 4 个函数中的一个或多个对样本数据集进行估计, 得出数据主题生存规律的相关结论。

1.4 双比例失效法

在评价中应用生存分析法时, 终点事件即期刊论文“失效”的判定颇为复杂。本研究认为, 期刊论文自发表之日起便拥有了一定的价值, 其价值体现在被其他文献引用的行为中, 随着论文发表时间的延长, 论文将经历单位时间内被引频次越来越高、在某单位时间点被引频次达到峰值、随后逐渐降低至很少被人引用甚至不再被引用的生命周期。现假定某篇论文在连续 T 个单位时间内没有被引用, 可以认为该论文在这 T 个时间后被大量关注或使用的概率不高, 以此来判定论文失效, 其中连续零被引时长 T 被称为判定论文失效的有效时间间隔。因此, 定义生命长度为自论文见刊之日起至判定该论文失效的有效时间间隔 T 的起始时刻为止所经历的时间长度。可见, T 值是判定论文失效与计算论文生存时间的关键。为确定 T 值, 引入了双比例失效法, 该方法有以下几个主要步骤。

(1) 预设双比例阈值 P_1 、 P_2 。经验性通常设为 80%、90%。其中, P_1 规定失效论文在生存时间内的累积被引频次占检测时间段内总被引频次的比例, 即有效累计被引率应达到的值; P_2 规定达到 P_1 值的论文量应达到的比例。

(2) 递归判断。当 $T = i$ ($i = 1, 2, \dots$) 时, 判别失效论文, 分别计算每篇失效论文在生存时间内

的有效累计被引率。

(3) 判断每篇失效论文的有效被引率是否达到P1。

(4) 统计满足条件(3)的论文数量, 计算该论文量占样本总量的百分比, 判断符合P1的论文量是否达到阈值P2: 如果该比例小于P2, 则T往后取值, 从第二步重新计算; 如果该比例大于等于P2, 则将对应T值作为判断该领域论文失效的时间间隔。

通常, 经过双比例失效法得到的T值有多个, 而在本研究中, 取最小值作为判定论文失效的有效时间间隔值T。实质上, T值与双比例阈值P1、P2之间是相互影响的, 由预设的P1、P2值, 通过双比例失效法遍历寻找到的有效T值, 并在此阈值对下确定有效T值的大小及各类型论文的比例情况, 可以反向判断预设的双比例阈值是否合适, 从而对阈值对进行调整。

2 实证研究

2.1 数据来源与处理

评价对象为30种物理学科中文期刊, 数据来源于中国科技论文与引文数据库(CSTPCD), 以2016年为原点, 向前追溯收集这些期刊集合在2008—2015年发表的论文及在被引关系中涉及的论文信息(共86375条记录)、30种期刊所发表论文的被引用关系记录(共115248条记录)及施引和被引期刊的年发文频次信息(共1698条信息)。数据清洗(去除冗余记录、重要信息缺失记录等, 补充或修正期刊年发文频次信息等)及数据表结构转换过程, 在此不作赘述。在此需要说明的是, 在确定论文出版时间时, 由于CSTPCD不直接提供期刊每期出版具体时间, 并且对于双

月刊、季刊等类型期刊, 不同的期刊选择发表的月份是不同的, 考虑到研究的可行性问题, 统一把出版周期内最后一个月作为出版月。由此得到可用于生存分析的论文被引时序分布表, 如表1所示。其中, -1对应论文见刊前的时间, 0说明论文在该时间节点未被引用, 正整数为论文在该时间节点处对应的被引用频次。

2.2 数据统计

CSTPCD收录的母刊集在2008年1月至2015年12月的发文量、被引频次、施引论文数、施引期刊数量等信息如表2所示。样本数据集中, 《物理学报》发文量最高, 其次为《光学学报》, 而《光谱学与光谱分析》的施引期刊数量最高。

2.3 数据分类及T值判定

根据生存数据的分类特点及论文在监测时间段内的被引和失效情况, 现将论文分为三类: 第一类为无生命论文; 第二类为失效论文; 第三类为未失效论文。第一类论文是在发表后T+De(判定信息失效的时间间隔T及首次被引时滞均值De, De是本研究数据集物理学各个科期刊论文首次被引用的时间与其见刊时间的时差的算数平均数)时间间隔内均未被引, 这类论文自发表后就直接失效, 无生命长度; 第二类论文是根据前面介绍的方法, 判定在监测时间段内失效了的论文, 生命长度为失效点时间减去发表点时间, 属于完全数据; 第三类论文是自发表后一直被引, 监测时间结束生命仍在延续, 规定这类论文的生存时间为监测时间终点减去发表点时间, 属于右删失数据。

在利用“双比例”失效法判定有效时间间隔T值前, 需要首先预设双比例P1、P2阈值对的值。通常学者根据“二八法则”将双比例阈值定

表1 论文被引时序表

ID	被引论文编码	出版时间	监测时间段被引情况列						
			200801	……	201003	201004	201005	……	201512
32780	A103200802007	200802	-1	……	0	1	1	……	0
32821	A103200806004	200806	-1	……	0	2	0	……	0
34827	C006200903057	200902	-1	……	0	0	1	……	0
36827	C006201012013	201006	-1	……	-1	-1	-1	……	1
……	……	……	……	……	……	……	……	……	……

表 2 母刊发文及被引情况统计表 (中文版)

母刊	发文量/篇	被引频次/次	施引论文数/篇	施引期刊数量/种
波谱学杂志	381	693	465	174
低温物理学报	269	278	227	77
发光学报	1092	1921	1452	302
高压物理学报	450	781	670	197
工程热物理学报	2453	4775	4054	552
光谱学与光谱分析	4331	13503	10136	1171
光散射学报	386	583	464	178
光学技术	905	1746	1643	370
光学学报	4605	18042	10682	645
光学与光电技术	498	485	428	136
光子学报	2921	6713	4984	521
核聚变与等离子体物理	306	438	341	105
红外与毫米波学报	635	2149	1840	386
计算物理	525	827	724	266
量子电子学报	687	929	689	150
量子光学学报	264	256	217	58
热科学与技术	281	401	353	136
声学技术	617	1141	1031	334
声学学报	615	1539	1198	265
数学物理学报	903	1308	1094	162
物理	369	658	583	234
物理学报	10413	29046	17992	992
物理学进展	68	173	166	87
物理与工程	89	25	21	8
应用光学	1356	2574	2175	344
应用声学	336	549	522	217
原子核物理评论	319	450	364	121
真空与低温	235	278	243	90
中国光学	303	654	489	67
中国科学物理学力学天文学	898	2204	1750	475
共计	37510	—	—	—

为 80%。考虑物理领域的上述学科论文被引分布特点,再根据阈值对与T值之间的制约关系,经过多次实践调整最终将双比例阈值P1和P2均设定为71%。在此阈值对P1、P2下采用穷举法,对各个T值的有效性进行判定。如图1所示,在双比例阈值对71%—71%的条件下,随着横轴时间间隔T值从小到大变动,三类论文的数量也发生了明显的变动。在此过程中,无生命论文数量单调减少后变为0,失效论文量先上升后下降最后变为0,未失效论文量单调增加,最终所有样本论文都成为未失效论文。可以看出,物理学科作为基础学科,其论文右删失情况严重,这符合我们对物理学科发展及其老化速度相对迟缓的认知。

表3列出了部分T值下论文类别分布及失效比例情况,其中“失效—P1数量”是指失效论文在有效生命长度内被引频次占观测时间内总被引频次的比例大于P1(71%)的论文数量;在考虑到物理学科论文右删失现象较为显著的情况下,将“失效—P2占比”定义为失效—P1论文数量和未失效论文数量占失效论文数量和未失效论文总量的比例。实质上,双比例阈值是不影响遍历T值时三类论文的变化规律的,而是用来确定“失效—P1数量”与“失效—P2占比”达到阈值所对应的T值。

根据表3的统计数据,当T值取14时(字体已被加粗的行),被引频次大于P1的论文数量所占比例73.2%首次超过P2(71%)。并且,当

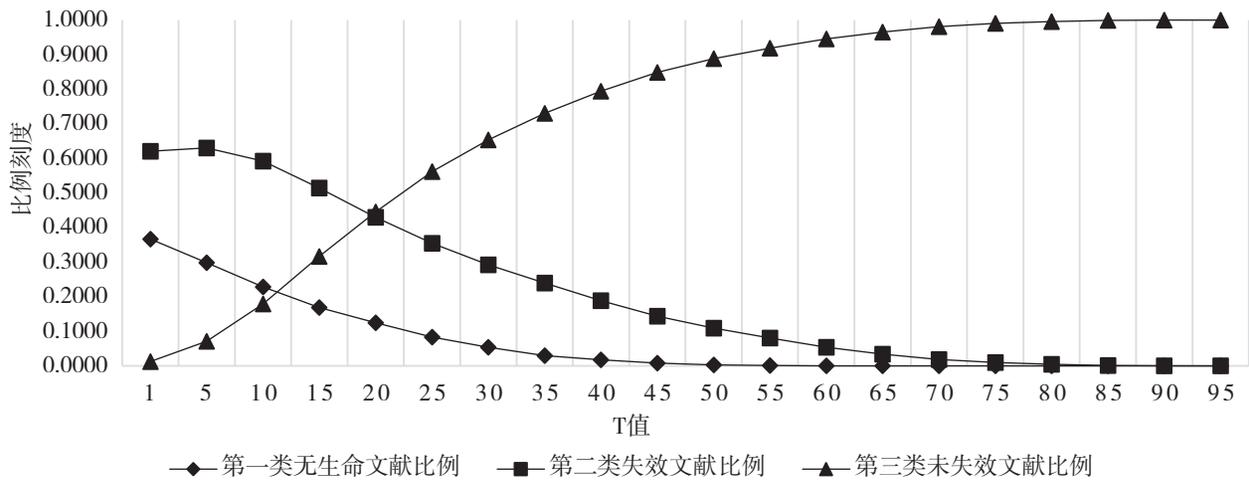


图 1 三类论文随 T 值的变化情况

表 3 判定 T 值数据节选

T 值	第一类无生命论文		第二类失效论文		第三类未失效论文		失效-P1 数量	失效-P2 占比
	数量	比例	数量	比例	数量	比例		
1	11629	0.367	19678	0.621	398	0.013	6458	0.342
2	11124	0.351	19973	0.630	608	0.019	6902	0.365
3	10531	0.332	20031	0.632	1143	0.036	7130	0.391
4	10075	0.318	19998	0.631	1632	0.051	7533	0.424
5	9452	0.298	19987	0.630	2266	0.071	7822	0.453
6	9030	0.285	19917	0.628	2758	0.087	8301	0.488
7	8484	0.268	19674	0.621	3547	0.112	8507	0.519
8	8108	0.256	19471	0.614	4126	0.130	8937	0.554
9	7637	0.241	19105	0.603	4963	0.157	9101	0.584
10	7253	0.229	18779	0.592	5673	0.179	9453	0.619
11	6777	0.214	18411	0.581	6517	0.206	9628	0.648
12	6450	0.203	18019	0.568	7236	0.228	9946	0.680
13	5983	0.189	17501	0.552	8221	0.259	9934	0.706
14	5722	0.180	16956	0.535	9027	0.285	9983	0.732
15	5363	0.169	16311	0.514	10031	0.316	9780	0.752
16	5122	0.162	15664	0.494	10919	0.344	9646	0.774
17	4772	0.151	15106	0.476	11827	0.373	9477	0.791
18	4576	0.144	14604	0.461	12525	0.395	9410	0.809
19	4181	0.132	14076	0.444	13448	0.424	9228	0.824
20	3960	0.125	13624	0.430	14121	0.445	9171	0.840

T=14 时，失效论文中满足 P1 条件的论文量（即失效-P1 数量）达到最大，并在此后逐渐下降。由此可见，在双比例 71%—71% 的判定标准下，T 取 14 满足判定条件，并且满足 P1 条件的论文量达到最大。因此，本研究判定实验样本论文失效的有效时间间隔的 T 值为 14。

2.4 样本数据集的生存函数曲线分析

在论文生命长度分布基础上，本研究依据论文生命表及生存时间函数，针对样本论文的生存

情况进行进一步的分析。以 5 个月为单位区间跨度，将论文生命长度划分为 19 个时间段，利用生存曲线、概率密度曲线及风险曲线，对物理领域样本论文整体情况进行分析。

2.4.1 累计生存曲线

在本研究中，物理学科样本论文集的累积生存曲线如图 2 所示，反映了对应时间仍处于存活状态的论文数量占总体样本数量的比例。由于研究领域的单一性，所以选取时间前后纵向对比分

析，物理学科论文随着时间的增加，累计生存率整体呈现下降趋势，前 11 个时间区间下降显著，大约有 80% 的论文生命长度在 11 个时间区间内。突破 11 个时间区间存活下来的论文累计生存曲线变缓，失效速度减缓，生存时间较长。在检测时间为结束时，仍然有将近 10% 存活率，说明物理学科论文数据的右删失情况显著。

2.4.2 概率密度曲线

样本论文集的概率密度曲线，如图 3 所示（经过平滑后的折线图），反映了样本论文在各个寿命区间的分布情况，即具有相同寿命的论文在总体中所占的比例。由图 3 可见，在前 11 个寿命区间的概率密度数值较大，随后曲线呈现下降趋于平稳的状态。此外，样本论文见刊后第 7 个

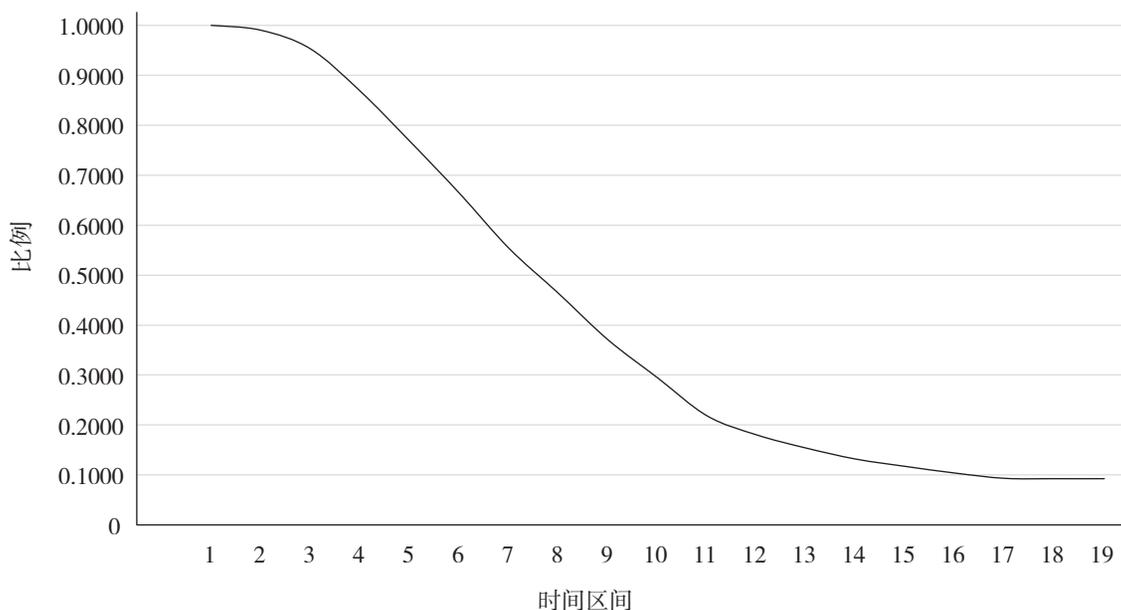


图 2 样本集的累积生存曲线

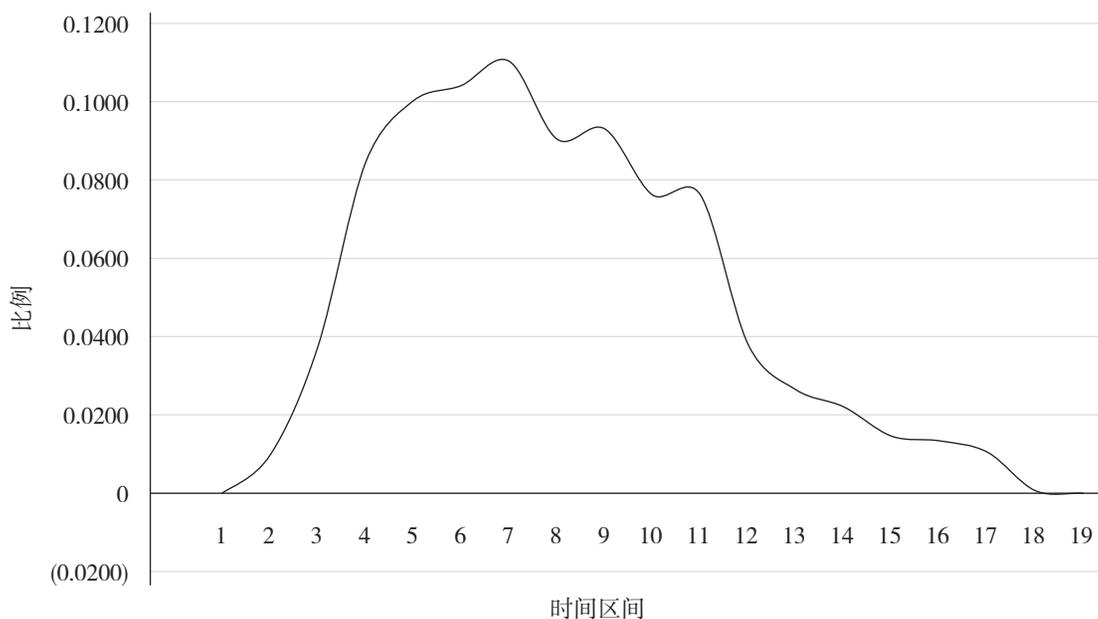


图 3 样本集的概率密度曲线

时间区间出现了寿命分布最大值,说明论文见刊后第7个时间区间失效的概率相比于其他时间要高。见刊第11个时间区间后的论文生命长度分布较为分散。

2.4.3 风险曲线

样本集的风险曲线如图4所示(经过平滑后的折线图),反映了每个寿命区间内未失效论文的失效概率,曲线出现多次升落现象,但从整体来看,在前10个时间区间内,论文在单位时间内的失效风险呈现逐渐增大的趋势;在第11个时间区间失效风险达到最大;随后迅速回落,逐渐趋于平缓。在风险迅速下降的过程中,在第13、15时间区间内再次迎来失效小高峰。度过失

效高峰期的样本论文失效的概率不断变小,成为该领域内长寿论文。

2.5 样本期刊生存曲线对比分析

在分析了数据集整体生存规律的基础上,将生存分析尝试性地应用在期刊间的对比上。在此选取《低温物理学报》《红外与毫米波学报》及《应用光学》3种期刊,对其生存函数曲线进行对比分析。笔者统计了这3种期刊的发文量及被引分布表中的记录数(表4),其中发文被引率是CSTPCD中收录的被引论文量占发文量的比值。从2016年版《中国科技期刊引证报告(核心板)》(自然科学卷)的2015年中国科技核心期刊综合评价总分排名表中,摘录出样本期刊的

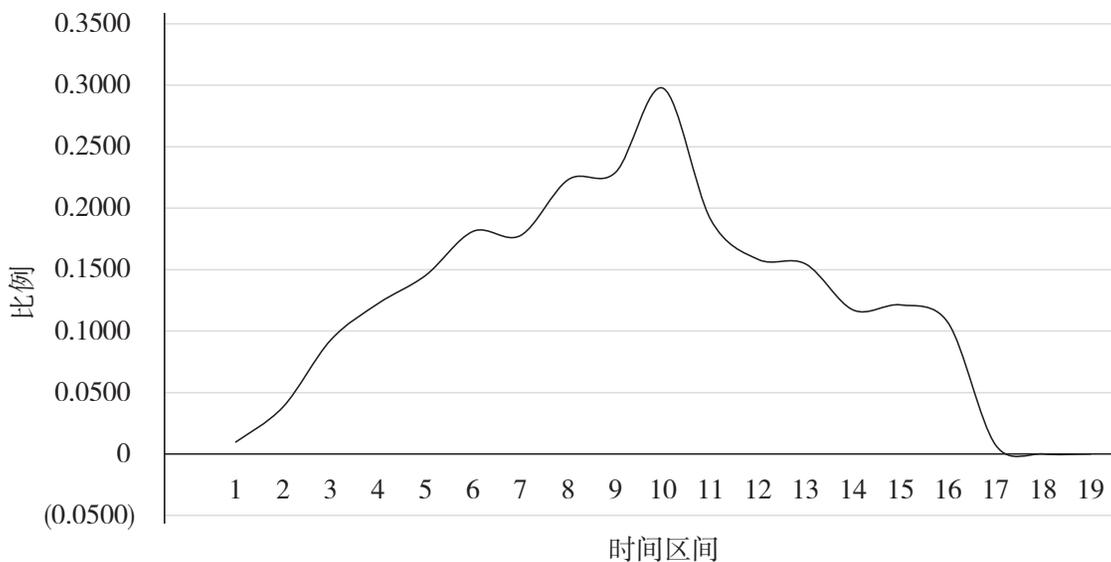


图4 样本集的风险曲线

表4 样本期刊发文及被引情况表

刊名	发文量/篇	被引频次/次	发文被引率
低温物理学报	269	179	0.6654
红外与毫米波学报	635	505	0.7953
应用光学	1356	959	0.7072

表5 样本期刊在2015年中国科技核心期刊综合评价排名表中的情况

期刊	核心总被引频次		核心影响因子		综合评价总分	
	数值/次	排名	数值	排名	数值	排名
红外与毫米波学报	822	1011	0.533	867	35.56	1147
应用光学	808	1024	0.406	1207	21.82	1771
低温物理学报	125	1960	0.136	1953	13.16	1947

评价数据情况（表 5），作为利用生存分析法对比 3 个样本期刊结果的评价标准。

3 个样本期刊的累计生存曲线对比图如图 5 所示。从总体来看，3 种刊物的累计生存曲线下下降趋势由陡到缓依次为：《低温物理学报》《应用光学》《红外与毫米波学报》。因此，整体上相对来说，在相同时刻，《红外与毫米波学报》生存率更高，《低温物理学报》的论文生存率最低。《低温物理学报》的论文生命长度集中在 11 个时间区

间内，《应用光学》的最长生存时间能达到 16 个时间区间左右，《红外与毫米波学报》在监测的末期，还能有接近 20% 的存活率。

3 个样本期刊的概率密度曲线对比图如图 6 所示（经过平滑后的折线图）。从图 6 中可见，《红外与毫米波学报》的论文寿命分布相对分散，而《低温物理学报》分别在第 6、9、11 时间区间分别出现论文寿命分布高峰。

3 个样本期刊的风险曲线对比图如图 7 所示

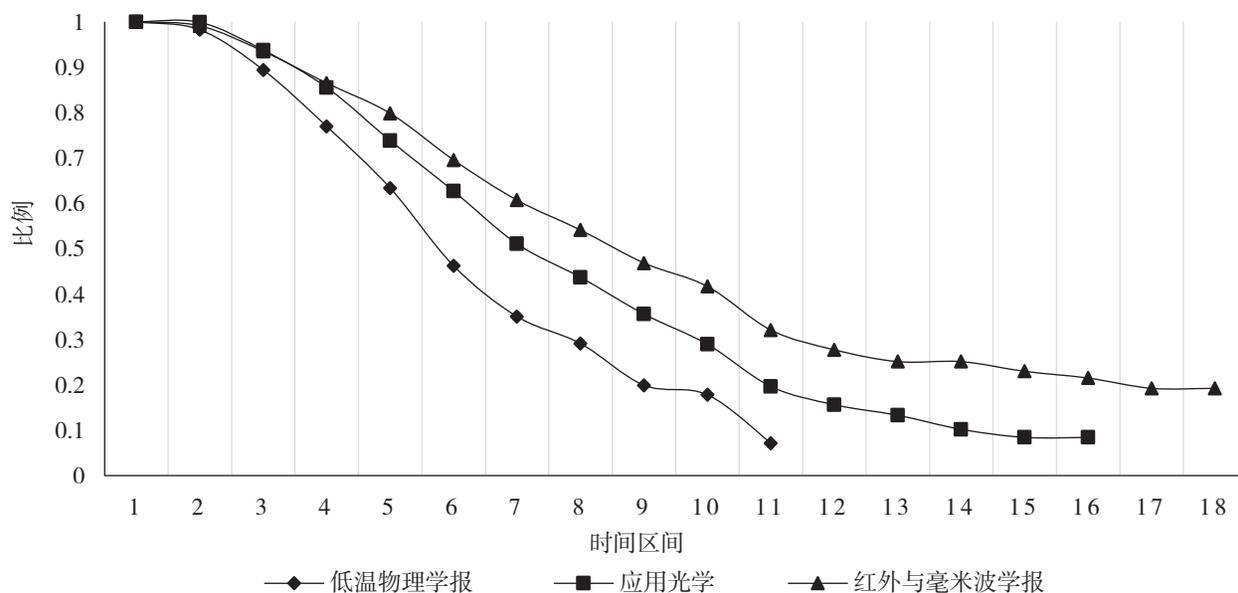


图 5 样本期刊累计生存曲线对比图

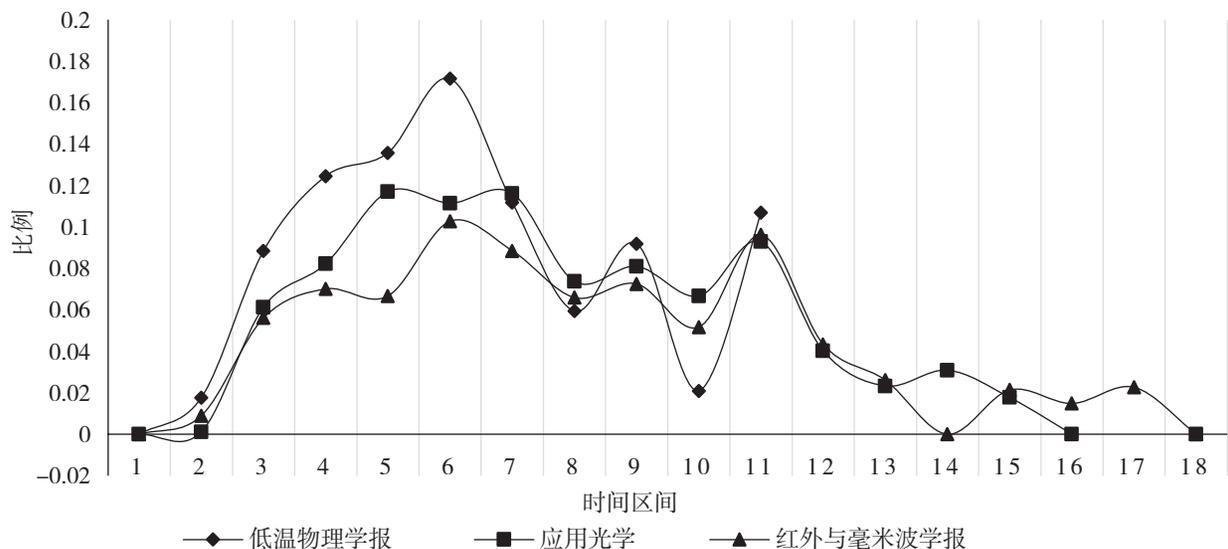


图 6 期刊概率密度曲线对比图

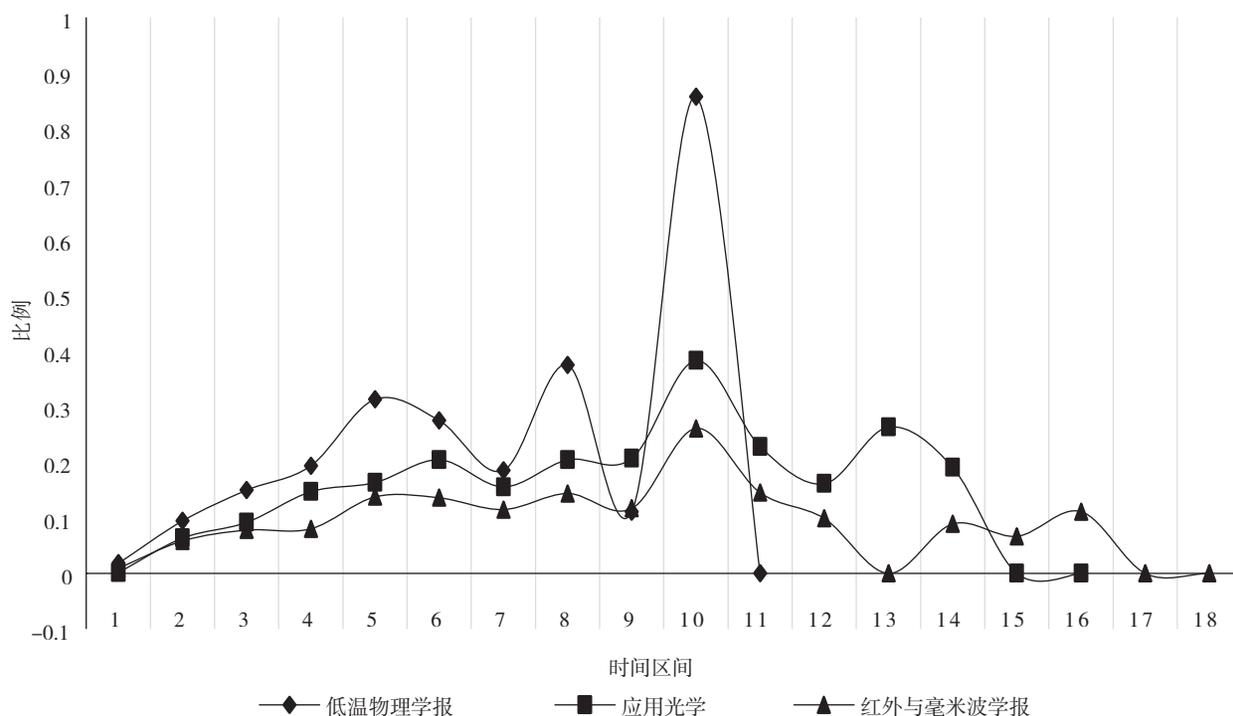


图7 期刊风险曲线对比图

(经过平滑后的折线图)。从图7中可见,3家样本期刊在第10个时间区间内论文失效率均达到最大值,这说明期刊在该时间区间内失效的论文量都很大,并且《低温物理学报》远远高于另外两种期刊。此外,《应用光学》在单位区间内的失效风险普遍高于《红外与毫米波学报》。

从分析结果可以看出,在这3种期刊中,《低温物理学报》的论文失效率较高,生存时间比较集中且普遍较短。笔者认为这反映了该刊的论文质量要普遍低于另外2种期刊;《红外与毫米波学报》因其分散的生命分布、普遍偏低的失效风险及相对平稳降低的生存曲线成为3个样本期刊中质量最高的期刊;《应用光学》处于上述两种期刊之间。这个评价结果与2016年版的《中国科技期刊引证报告(核心板)》的2015年中国科技核心期刊综合评价排名表的结果相符合,在一定程度上说明将生存分析法应用到期刊论文数据上是有一定效用的。

3 总结与讨论

(1) 本研究将医学领域的生存分析法移用到

学术期刊评价,在此过程中克服了遇到的障碍性问题,从而使得应用生存分析法评价期刊论文成为可能。首先,对论文失效进行了定义,并利用“双比例”失效法作出判定;其次,利用期刊的年发文频次与论文见刊的年卷期计算了见刊时间(精确到月),解决了原始论文被引频次的分布。

(2) 在选用CSTPCD收录的30种物理学科中文期刊中,对整体样本集进行分析后,选取3个样本期刊,经生命曲线对比分析表明,《应用光学》的论文质量相对更高些,其次依次为《应用光学》《低温物理学报》,与2016年版的《中国科技期刊引证报告(核心板)》的评价结果相符。

(3) 本研究仅实现了生存分析法在期刊评价上应用,并利用生存函数曲线对比分析了样本期刊间的差异,对计算结果进行了初步分析,但没有深入探究在期刊评价应用方向上的相关问题,例如指标构建及评价结果效用性等问题。

(4) 生存分析法用于期刊评价尚存在许多需要完善与优化的地方:在实证数据方面,应该

(下转第91页)

有限性，本文制定的分类体系可能没有涵盖资源环境领域科学数据的所有方面，但随着资源环境领域项目的不断开展、数据资源的不断增加，科学数据分类体系也将不断被修改与完善。

参考文献

- [1] 中华人民共和国科技部. 国家“十五”科技基础性工作专项实施意见[J]. 中国基础科学, 2001(8): 31-34.
- [2] 孙九林, 孙晓华. 自然资源信息分类体系[J]. 资源科学, 1989(2): 62-69.
- [3] 廖顺宝, 蒋林. 地球系统科学数据分类体系研究[J]. 地理科学进展, 2005, 24(6): 93-98.
- [4] 王卷乐, 林海, 冉盈盈, 等. 面向数据共享的地球系

统科学数据分类探讨[J]. 地球科学进展, 2014, 29(2): 265-274.

- [5] 国家标准化管理委员会. GB/T 13745-92 学科分类与代码[S]. 2009.
- [6] 国家地球系统科学数据共享服务平台[DB/OL]. [2016-12-11]. <http://www.geodata.cn/index.html>.
- [7] 海洋科学数据中心. 海洋学科分类[DB/OL]. [2017-01-22]. <http://msdc.qdio.ac.cn/>.
- [8] 中国科技资源共享网[DB/OL]. [2017-02-06]. <http://www.escience.gov.cn/>.
- [9] 中国南北极数据中心[DB/OL]. [2017-03-25]. <http://www.chinare.org.cn/index/>.
- [10] 中国气象数据网[DB/OL]. [2017-04-24]. <http://data.cma.cn/site/index.html>.

(上接第 84 页)

延伸时间检测窗口，扩大样本数据范围，在更加宽泛的数据集上监测更长的时间；在理论方法方面，对于期刊评价上的探究过少，缺乏对评价指标及评价效用性的探究。

(5) 本文为文献老化视角下的期刊评价研究工作提供了新的思路与视角，丰富了文献信息老化的理论“基石”。通过在理论与实证两方面的研究，体现了生存分析法对被引时序数据具有普遍适用性。笔者认为，该方法可以在以论文为评价主体的期刊评价、学科信息老化速度评价、高校科研产出水平评价等方面进行推广与应用。

参考文献

- [1] 马费成, 望俊成. 信息生命周期研究述评(I): 价值视角[J]. 情报学报, 2010, 29(5): 939-947.
- [2] 王伟, 胡西厚, 黄亚明, 等. 信息计量学及其医学应用[M]. 北京: 人民卫生出版社, 2009: 63.
- [3] 望俊成. 网络信息生命周期规律研究[M]. 北京: 科学技术文献出版社, 2014.
- [4] 周晓丹. 基于专利生命周期的企业技术竞争力研究:

以无线充电领域为例[D]. 北京: 中国科学技术信息研究所, 2016.

- [5] 李翔. 基于专利生命周期的无线充电技术创新研究[D]. 武汉: 武汉大学, 2015.
- [6] 邢晓昭. 专利信息生命长度测度及规律研究: 以运算处理和计算领域为例[D]. 北京: 中国科学技术信息研究所, 2013.
- [7] 蔡梦. 生存分析理论及其应用研究综述[J]. 价值工程, 2016, 35(10): 19-21.
- [8] 张旭尧, 张卫民. 基于生存分析法的财务困境预警研究: 以我国农业上市公司为例[J]. 商业会计, 2016(3): 38-41.
- [9] 李文娟. 中国ST企业的生存分析[D]. 厦门: 厦门大学, 2008.
- [10] BERKSON J. The calculation of survival rates, in carcinoma and other malignant lesions of the stomach[J]. American Journal of Clinical Pathology, 1942, 119(8): 1534.
- [11] LEE E T, WANG J W. Statistical methods for survival data analysis[M]. New Jersey: Wiley, 1992.
- [12] BERKSON J, GAGE R P. Calculation of survival rates for cancer[J]. Proceedings of the Staff Meetings Mayo Clinic, 1950, 25(11): 270-286.