统计预测引论

周源泉　李宝盛　丁为航　张正平　◎　著

科学出版社

统计预测引论

周源泉　李宝盛　丁为航　张正平　著

科学出版社
北京

内 容 简 介

本书研究了常见分布情况下、在未来失效时间、未来时间区间内的失效数单样、双样预测，多个未来样本顺序统计量的联合预测，以及泊松过程、幂律过程的故障与时间终止条件下，在未来失效时间和未来时间区间内失效数的单样、双样预测；给出了构造单样预测区间的条件方法及构造一般预测区间的经典、贝叶斯与信赖方法，用于预测区间与异常值检验的关系；采用数理统计的三大流派的方法进行统计推断，所有的统计预测区间都是精确预测区间。

本书可供学习和研究统计预测的高等院校教师、高年级本科生、研究生，以及科研单位研究人员和管理人员阅读。

图书在版编目(CIP)数据

统计预测引论/周源泉等著．—北京：科学出版社，2017．6

ISBN 978-7-03-053211-4

Ⅰ．①统…　Ⅱ．①周…　Ⅲ．①统计预测-研究　Ⅳ．①C8

中国版本图书馆 CIP 数据核字(2017)第 127103 号

责任编辑：刘宝莉　陈　婕　纪四稳 / 责任校对：桂伟利

责任印制：张　倩 / 封面设计：陈　敬

科学出版社 出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

中国科学院印刷厂 印刷

科学出版社发行　各地新华书店经销

*

2017 年 6 月第　一　版　开本：720×1000 1/16

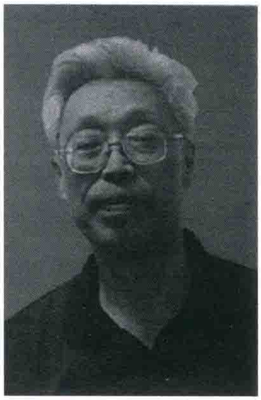
2017 年 6 月第一次印刷　印张：25 1/4

字数：490 000

定价：160.00 元

(如有印装质量问题，我社负责调换)

作 者 简 介



周源泉，男，生于1937年12月，江苏宜兴人。北京强度环境研究所研究员，多所大学的客座教授。1960年毕业于北京大学。主要研究方向为可靠性理论与工程、数理统计理论与应用。完成了百余项工程与研究项目，为国家节省了亿元以上经费，多次获中国航天工业总公司与国防科工委科技进步成果奖，获2000年国家质量管理突出贡献奖。发表学术论文220余篇，出版专著:《可靠性评定》(科学出版社)、《可靠性增长》(科学出版社)、《质量可靠性增长与评定方法》(北京航空航天大学出版社)、《可靠性基础入门》(中国统计出版社)。

李宝盛，男，生于1962年5月，陕西宝鸡人。北京航天动力研究所研究员。1983年毕业于中山大学，获理学学士学位；1988年毕业于西安交通大学，获理学硕士学位；2000年毕业于西北工业大学，获工学博士学位。主要研究方向为可靠性理论与工程、概率统计、液体火箭发动机系统。发表学术论文40余篇，完成可靠性理论与工程研究课题多项，编制可靠性分析计算软件系统多项，编写国家军用标准(GJB)、中华人民共和国航天标准(QJA)多项，出版专著2部。

丁为航，男，生于1973年12月，山东巨野人。1996年毕业于北方工业大学，获工学学士学位；2015年毕业于首都经济贸易大学，获硕士学位。就职于北京日立北工大信息系统有限公司。主要研究方向为计算机科学与应用。发表学术论文数篇，完成信息系统软、硬件的开发与安装工作多项，编制可靠性评定、可靠性增长及统计预测软件系统多项。

张正平，男，生于1968年4月，江苏高邮人。北京强度环境研究所所长，研究员，973项目首席科学家。1992年毕业于北京工业大学，获硕士学位；2009年毕业于哈尔滨工业大学，获工学博士学位。主要研究方向为可靠性试验与评估等。主持科研项目多项，获得国防技术发明奖二等奖1项、国防科技进步奖三等奖2项。发表论文20余篇，出版专著2部。

序

周源泉研究员从事航天可靠性研究和实践五十年，取得了许多显著成果。近十多年来，和其他作者合作，用 Frequentist、Bayes、Fiducial 方法解决了国际上多年来未能解决的位置尺度族分布的单样、双样和多样预测区间等一百多项统计预测学科的难题。这些成果对确定各类武器系统的贮存期、各类军民品的安全寿命与有效期，以及对各类军民品进行寿命预测与多批产品的同步验收，都具有重要意义。目前，这些成果已得到有关专家的认可。现将这些成果结集成书——《统计预测引论》，推进了可靠性工程和数理统计的发展，有助于“中国制造”产品质量进一步提高。

中国科学院院士　谢光选

2015 年 5 月

前　言

20 世纪 90 年代，经过调研，我决定利用退休后的时间学习和研究“统计预测”这个方向，一是因为它有广泛的工程应用价值，二是国内尚没有单位与个人开展此项研究。这个想法得到了时任哈尔滨理工大学副校长郭建英教授的赞同，我们联合申请与开展了一项国家自然科学基金项目，内容是了解国际上关于指数分布，双参数指数分布单、双样预测区间等的研究情况。2001 年完成此项目后，我继续开展研究，但遇到了韦布尔分布单、双样 Frequentist 预测区间这个拦路虎，它是国际上自 1973 年以来一直未能解决的难题。当刘永才院士等向我提出了攻克此难题的任务。我硬着头皮着手开展研究，经过 5 年的努力，于 2006 年解决了韦布尔分布、正态分布、对数正态分布的单、双样预测区间问题；之后，相继给出了（对数）位置尺度族分布的单、双样预测区间与以上分布的 l 个未来样本的顺序统计量的联合预测区间，以及这些分布的未来失效数的单、双样预测区间；对常用离散分布预测区间，幂律过程单、双样预测区间等进行补充。上述所有相关研究成果构成了一个比较完整的体系，合在一起共 50 多篇论文，构成了本书主要内容。

在整个研究中，李宝盛研究员、丁为航先生与我进行了高效的合作，将本书的理论部分给予了计算机编程实现，为本书的各个新模型编制了高精度软件，并用数值验证了本书的各命题，为本书成果的推广应用打开了大门。谢光选、余梦伦、黄瑞松、刘永才、朱建士五位院士在我的研究过程中也给予了极大的支持。本书的出版得到了北京强度环境研究所领导的大力支持。北京强度环境研究所原所长龚知明研究员邀请陈家鼎、茆诗松、李国英、费鹤良教授审定了本书有关统计预测理论基础的五篇论文。在此，对上述关心、帮助、支持我的朋友、领导表示衷心的感谢。

由于作者水平所限，书中难免存在不妥之处，诚请专家、读者不吝指正。

在十五年的研究中，我的父亲、母亲先后仙逝，谨以此书，以资纪念。

周源泉

2016年12月

目　录

第 1 章　指数分布的预测区间

1.1　指数分布预测区间系数表

虽然 Bain 和 Patel(1991)给出了指数分布单样预测区间(prediction intervals，PI)的系数表，但是其精度较低，有时不能满足工程实践与科学研究的需要。为此，本章给出了指数分布预测区间的高精度系数表，指出应用该表可找到指数分布的双样预测区间与双参数指数分布的单样预测区间，并通过数值例题介绍了该表的应用。

1.1.1　指数分布的单样预测区间的系数表

Y 服从具有未知失效率 λ 的指数分布，对 Y 取大小为 n、截尾数为 r 的Ⅱ型截尾样本，其前 r 个最小观测值依序为

$$y_1 \leqslant y_2 \leqslant \cdots \leqslant y_r, \quad r<n \tag{1.1.1}$$

单样预测的任务之一是，根据 $\underline{y}=(y_1, y_2, \cdots, y_r)$给出样本的未来的第 $j-r$ 个最小观测值 $y_j(r<j\leqslant n)$的预测区间。

Lawless(1971)和周源泉(1997a)都曾指出，当置信水平为 γ 时，y_j 的预测区间为$[a, b]$，其中

$$a=y_r+d_1\tau, \quad b=y_r+d_2\tau$$

式中，τ 为总试验时间，

$$\tau=\sum_{i=1}^{r} y_i+(n-r)y_r \tag{1.1.2}$$

预测系数 d_1 和 d_2 分别由式(1.1.3)和式(1.1.4)确定：

$$\sum_{k=0}^{j-r-1} C_k\left[1+d_1(n-j+k+1)\right]^{-r}=\frac{1+\gamma}{2} \tag{1.1.3}$$

$$\sum_{k=0}^{j-r-1} C_k\left[1+d_2(n-j+k+1)\right]^{-r}=\frac{1-\gamma}{2} \tag{1.1.4}$$

式中

$$C_k=\frac{(-1)^k\cdot(n-r)!}{(n-j)!\ k!\ (j-r-1-k)!\ (n-j+k+1)}$$

$$\sum_{k=0}^{j-r-1} C_k=1$$

记 $d_1=d_1(n,j,r;\gamma)$，$d_2=d_2(n,j,r;\gamma)$。其中 $1\leqslant r<j\leqslant n=10$，$\gamma=0.8,0.9,0.95,0.98,0.99$；$d_1$，$d_2$ 可查表 1.1.1 得到，表值范围以外的 d_1 和 d_2 可根据我们编制的统计预测软件系统(SPSS)算出。

1.1.2 对指数分布双样预测区间的应用

Y 服从有未知失效率 λ 的指数分布，对 Y 取大小为 n、截尾数为 r 的Ⅱ型截尾的过去样本，其前 r 个最小观测值依序为

$$y_1\leqslant y_2\leqslant\cdots\leqslant y_r,\quad r\leqslant n$$

双样预测的任务之一是，根据 $\underset{\sim}{y}=(y_1,y_2,\cdots,y_r)$ 给出与 Y 独立同分布的 X 的大小为 N 的未来样本的第 i 个最小观测值 $x_i(i=1,2,\cdots,N)$ 的预测区间。

Lawless(1972)和周源泉(1997b)曾指出，当置信水平为 γ 时，x_i 的预测区间为 $[c,d]$，其中

$$c=e_1\tau,\quad d=e_2\tau$$

式中，e_1，e_2 为预测系数，分别由式(1.1.5)和式(1.1.6)确定：

$$\sum_{k=0}^{i-1}b_k\left[1+e_1(N-i+k+1)\right]^{-r}=\frac{1+\gamma}{2}\tag{1.1.5}$$

$$\sum_{k=0}^{i-1}b_k\left[1+e_2(N-i+k+1)\right]^{-r}=\frac{1-\gamma}{2}\tag{1.1.6}$$

式中

$$b_k=\frac{(-1)^k\cdot N!}{(N-j)!\ k!\ (i-1-k)!\ (N-i+k+1)}$$

记 $e_1=e_1(N,i,r;\gamma)$，$e_2=e_2(N,i,r;\gamma)$，比较 e_1 与 d_1 可得

$$e_1(N,i,r;\gamma)=d_1(N+r,i+r,r;\gamma)$$

平行地，有

$$e_2(N,i,r;\gamma)=d_2(N+r,i+r,r;\gamma)$$

1.1.3 对双参数指数分布的单样预测区间的应用

Y 服从有未知位置参数 μ 与未知尺度参数 $\sigma(\lambda=\sigma^{-1})$ 的双参数指数分布，对 Y 取大小为 n、截尾数为 r 的Ⅱ型截尾样本，其前 r 个最小观测值依序为

$$y_1\leqslant y_2\leqslant\cdots\leqslant y_r,\quad r<n$$

单样预测的任务之一是，根据 $\underset{\sim}{y}=(y_1,y_2,\cdots,y_r)$ 给出样本的未来的第 $j-r$ 个最小观测值 $y_j(r<j\leqslant n)$ 的预测区间。

Likes(1974)和周源泉(1997c)曾指出，当置信水平为 γ 时，y_j 的预测区间为 $[e,f]$，其中

$$e=y_r+f_1s,\quad f=y_r+f_2s$$

式中

$$s=\sum_{i=1}^{r}(y_i-y_1)+(n-r)(y_r-y_1)$$

预测系数 f_1 和 f_2 分别由式(1.1.7)和式(1.1.8)确定：

$$\sum_{k=0}^{j-r-1}C_k\left[1+f_1(n-j+k+1)\right]^{-r+1}=\frac{1+\gamma}{2} \tag{1.1.7}$$

$$\sum_{k=0}^{j-r-1}C_k\left[1+f_2(n-j+k+1)\right]^{-r+1}=\frac{1-\gamma}{2} \tag{1.1.8}$$

记 $f_1=f_1(n,j,r;\gamma)$，$f_2=f_2(n,j,r;\gamma)$，比较 d_1，f_1 可得

$$f_1(n,j,r;\gamma)=d_1(n-1,j-1,r-1;\gamma)$$

平行地，有

$$f_2(n,r,j;\gamma)=d_2(n-1,j-1,r-1;\gamma)$$

1.1.4 数值例子

例 1.1.1　指数分布的Ⅱ型截尾样本为 $n=4$，$r=2$，$y_1=71.5$，$y_2=84.7$，求 $\gamma=0.9$时，y_4 的预测上限 $y_{4,\mathrm{U}}$。

解
$$\tau=\sum_{i=1}^{2}y_i+2y_2=325.6$$

查表 1.1.1 可得，当 $n=4$，$j=4$，$r=2$，$\gamma=0.8$ 时，$d_2=3.0934$，故 $y_{4,\mathrm{U}}=y_2+d_2\tau=1091.9$。

例 1.1.2　指数分布的Ⅱ型截尾的过去样本为 $n=4$，$r=2$，$y_1=71.5$，$y_2=84.7$，求与 Y 独立同分布的 X 的大小为 $N=4$ 的未来样本的第 4 个最小观测值 x_4 在 $\gamma=0.9$ 时的预测下限 $x_{4,\mathrm{L}}$。

解
$$\tau=325.6$$

$$e_1(N,i,r;\gamma)=d_1(N+r,i+r,r;\gamma)=d_1(6,6,2;0.8)=0.34452$$

故有 $x_{4,\mathrm{L}}=e_1\tau_1=112.18$。

例 1.1.3　双参数指数分布的Ⅱ型截尾样本为 $n=6$，$r=4$，$y_i=3657.5,81.5,112.5$，求 $\gamma=0.9$ 时 y_6 的预测上限 $y_{6,\mathrm{U}}$。

解
$$s=\sum_{i=1}^{4}y_i+(n-r)y_r-ny_1=296.5$$

$$\gamma=0.8\text{ 时},f_2(n,j,r;\gamma)=d_2(5,5,3;0.8)=1.6026$$

故 $y_{6,\mathrm{U}}=y_r+f_2s=587.68$。

1.1.5 指数分布预测区间系数表

指数分布预测区间系数表如表 1.1.1 所示。

表 1.1.1　指数分布预测区间系数表

γ			d_1					d_2				
n	j	r	0.8	0.9	0.95	0.98	0.99	0.8	0.9	0.95	0.98	0.99
2	2	1	0.11111	0.05263	0.02564	0.0101	0.00503	9	19	39	99	199
3	2	1	0.05556	0.02632	0.01282	0.00505	0.00251	4.5	9.5	19.5	49.5	99.5
3	3	1	0.33333	0.20643	0.13408	0.07905	0.05404	13.825	28.83	58.831	148.83	298.83
3	3	2	0.05409	0.02598	0.01274	0.00504	0.00251	2.1623	3.4721	5.3246	9	13.142
4	2	1	0.03704	0.01754	0.00855	0.00337	0.00168	3	6.3333	13	33	66.333
4	3	1	0.19004	0.11812	0.07692	0.04545	0.03111	7.6949	16.031	32.699	82.7	166.03
4	3	2	0.02705	0.01299	0.00637	0.00252	0.00125	1.0811	1.7361	2.6623	4.5	6.5711
4	4	1	0.5	0.33333	0.2343	0.15447	0.1155	17.034	35.373	72.043	182.04	365.38
4	4	2	0.17491	0.11191	0.07419	0.04447	0.03064	3.0934	4.8325	7.287	12.152	17.633
4	4	3	0.03574	0.01724	0.00847	0.00336	0.00167	1.1544	1.7144	2.42	3.6416	4.848
5	2	1	0.02778	0.01316	0.00641	0.00253	0.00126	2.25	4.75	9.75	24.75	49.75
5	3	1	0.13394	0.08333	0.0543	0.03211	0.02198	5.3892	11.224	22.892	57.893	116.23
5	3	2	0.01803	0.00866	0.00425	0.00168	0.00084	0.72076	1.1574	1.7749	3	4.3807
5	4	1	0.30599	0.20524	0.14496	0.09603	0.072	10.088	20.926	42.594	107.6	215.93
5	4	2	0.09993	0.06414	0.04262	0.02559	0.01765	1.7139	2.6699	4.0185	6.691	9.7013
5	4	3	0.01787	0.00862	0.00424	0.00168	0.00084	0.57722	0.85721	1.21	1.8208	2.424
5	5	1	0.62671	0.43201	0.31457	0.21793	0.16953	19.435	40.276	81.946	206.95	415.28

续表

			d_1					d_2				
γ												
n	j	r	0.8	0.9	0.95	0.98	0.99	0.8	0.9	0.95	0.98	0.99
5	5	2	0.27042	0.18796	0.13588	0.09186	0.0696	3.703	5.7255	8.578	14.231	20.598
5	5	3	0.11927	0.07725	0.05161	0.03113	0.02152	1.6026	2.2991	3.173	4.6828	6.1722
5	5	4	0.02669	0.01291	0.00635	0.00252	0.00125	0.77828	1.1147	1.5149	2.1623	2.7606
6	2	1	0.02222	0.01053	0.00513	0.00202	0.00101	1.8	3.8	7.8	19.8	39.8
6	3	1	0.10362	0.06449	0.04204	0.02486	0.01702	4.1582	8.6597	17.66	44.661	89.661
6	3	2	0.01352	0.00649	0.00318	0.00126	0.00063	0.54057	0.86803	1.3311	2.25	3.2855
6	4	1	0.22352	0.1502	0.10623	0.07047	0.05289	7.2994	15.136	30.804	77.805	156.14
6	4	2	0.07047	0.04527	0.0301	0.01808	0.01247	1.1988	1.8661	2.8073	4.6721	6.7726
6	4	3	0.01191	0.00575	0.00282	0.00112	0.00056	0.38481	0.57147	0.80665	1.2139	1.616
6	5	1	0.40127	0.27865	0.20413	0.14233	0.11115	12.001	24.84	50.509	127.51	255.84
6	5	2	0.16616	0.11615	0.08433	0.05725	0.04348	2.1783	3.3559	5.0157	8.304	12.007
6	5	3	0.06821	0.04431	0.02966	0.01792	0.0124	0.88593	1.2657	1.7416	2.5631	3.373
6	5	4	0.01335	0.00645	0.00317	0.00126	0.00063	0.38914	0.55737	0.75743	1.0811	1.3803
6	6	1	0.72783	0.51135	0.37981	0.27046	0.21503	21.354	44.196	89.867	226.87	455.2
6	6	2	0.34452	0.24882	0.18715	0.13361	0.10568	4.1588	6.3946	9.547	15.793	22.828
6	6	3	0.18714	0.13225	0.0967	0.06606	0.05033	1.8919	2.6772	3.6612	5.3597	7.0345
6	6	4	0.09062	0.05907	0.03963	0.02399	0.0166	1.0627	1.4647	1.9402	2.707	3.4145
6	6	5	0.0213	0.01031	0.00508	0.00201	0.001	0.58489	0.82056	1.0913	1.5119	1.8854

续表

γ			d_1					d_2				
n	j	r	0.8	0.9	0.95	0.98	0.99	0.8	0.9	0.95	0.98	0.99
7	2	1	0.01852	0.00877	0.00427	0.00168	0.00084	1.5	3.1667	6.5	16.5	33.167
7	3	1	0.08455	0.05263	0.03431	0.02029	0.0139	3.3885	7.0564	14.39	36.391	73.057
7	3	2	0.01082	0.0052	0.00255	0.00101	0.0005	0.43246	0.69443	1.0649	1.8	2.6284
7	4	1	0.17674	0.11886	0.08412	0.05584	0.04192	5.748	11.917	24.252	61.252	122.92
7	4	2	0.05453	0.03504	0.0233	0.014	0.00966	0.92455	1.4387	2.1638	3.6005	5.2187
7	4	3	0.00894	0.00431	0.00212	0.00084	0.00042	0.28861	0.4286	0.60499	0.9104	1.212
7	5	1	0.30059	0.20923	0.15357	0.1073	0.0839	8.8904	18.394	37.396	94.397	189.4
7	5	2	0.12152	0.08509	0.06186	0.04204	0.03195	1.573	2.4207	3.6154	5.982	8.6473
7	5	3	0.04811	0.03128	0.02095	0.01266	0.00876	0.61933	0.88382	1.2151	1.7867	2.3502
7	5	4	0.0089	0.0043	0.00212	0.00084	0.00042	0.25943	0.37158	0.50496	0.72076	0.9202
7	6	1	0.4811	0.34073	0.25479	0.18278	0.14599	13.594	28.1	57.103	144.1	289.11
7	6	2	0.22172	0.16128	0.12199	0.0876	0.06952	2.5478	3.9022	5.8108	9.5911	13.848
7	6	3	0.1152	0.08186	0.0601	0.04121	0.03147	1.1089	1.5606	2.1257	3.1001	4.0603
7	6	4	0.05185	0.0339	0.02278	0.01381	0.00957	0.58667	0.80461	1.0617	1.4757	1.8573
7	6	5	0.01065	0.00516	0.00254	0.00101	0.0005	0.29245	0.41028	0.54564	0.75594	0.9427
7	7	1	0.81162	0.57728	0.43425	0.31467	0.25364	22.952	47.462	96.466	243.47	488.47
7	7	2	0.40423	0.29845	0.22959	0.16901	0.13691	4.5235	6.9309	10.324	17.047	24.619
7	7	3	0.24046	0.17704	0.135	0.09764	0.07779	2.1074	2.9595	4.0263	5.867	7.6816