第15卷 第2期　　水生生物学报　　Vol. 15, No. 2
1991年6月　　ACTA HYDROBIOLOGICA SINICA　　June, 1991

研究简报

关于分布类型各检验方法的讨论

王继忠　　徐小清
（中国科学院水生生物研究所，武汉 430072）

DISCUSSIONS ON THE METHODS OF DISTRIBUTION TESTS

Wang Jizhong and Xu Xiaoqing
(*Institute of Hydrobiology, Academia Sinica, Whuan* 430072)

关键词　正态检验，计算机模拟，检验效率

Key words　Normality tests, Computer Simulation, Test efficiency

在现代生物数学中，常常需要对数据的分布类型作正态性检验。然而，在常用的分布类型检验方法中，由于各方法的侧重点不同，应用条件也不尽相同，因而对于某一类具体问题，各检验方法的检验效率也不同。因此，有必要对常用的几种分布类型检验方法的检验效率和应用条件进行讨论，以便能在实际应用中有选择地使用这些方法。

检 验 方 法

正态性检验的零假设为 H_0：总体服从正态分布，备择假设为 H_1：总体不服从正态分布[1]。

本文约定：$x_1, x_2, \cdots, x_n$ 为来自总体的子样，$X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ 为相应的顺序统计量，n 为子样容量。$\bar{x}$ 表示样本算术平均值，

$$m_i = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^i$$

为样本的 i 阶中心矩，α 为检验显著性水平，Z_α 为在 H_0 成立时所述统计量 α 分位数。

1. 偏度检验 ($8 \leqslant n \leqslant 5000$)[4]

在 H_0 成立时，偏度系数 $E(C_s) = 0$，故：统计量为：

$$C_s = \frac{m_3}{(m_2)^{3/2}}$$

则检验的拒绝域为 $|C_s| > Z_{1-\alpha}$

2. 峰度检验 ($7 \leqslant n \leqslant 5000$)[4]

在 H_0 成立时，峰度系数

$$E(C_k) = -\frac{6}{n+1},$$

故其统计量为：

$$C_k = \frac{m_4}{m_2^2} - 3$$

则拒绝域为：$|C_k| > Z_{1-\alpha}$

3. 偏度峰度检验 ($8 \leqslant n \leqslant 5000$)[2,3,4]

此检验实际上是偏度检验和峰度检验的联合检验，当偏度检验和峰度检验的均接受 H_0 时，此检验也接受 H_0，即

其统计量为：(C_s, C_k)

则拒绝域为：$|C_s| > Z_{1-\alpha}$ 或 $|C_k| > Z_{1-\alpha}$

4. Kolmogorov-Smirnov 检验 ($n > 3$)[3,5]

该检验方法的统计量为：

$$D_n = \max\left(\left|\frac{i-1}{n} - \Phi(u_i)\right|, \left|\frac{i}{n} - \Phi(u_i)\right|\right)$$

其中：$u_i = (x_i - \bar{x})/S$

1990年3月26日收到。

则检验的拒绝域为 $D_n > Z_\alpha$

5. W^2 和 A^2 检验 ($n \geqslant 5$)[5]

实际上 W^2 和 A^2 检验在一定程度上与 Kolmogorov Smirnov 验的思路是一致的，即都是比较理论分布函数和经验分布函数的差异。对于母体均值和方差均未知的情况：

则 W^2 和 A^2 的统计量分别为：

$$W^2 = \sum_{i=1}^{n} [\Phi(u_i) - (2i-1)/(2n)]^2 + \frac{1}{12n}$$

$$A^2 = \left(-n - \frac{1}{n}\right)\sum_{i=1}^{n} [(2i-1)Ln(\Phi(u_i)) + (2n+1-2i)Ln(1-\Phi(u_i))]$$

则拒绝域为：

W^2 检验：$W^2 > Z_\alpha$

A^2 检验：$A^2 > Z_\alpha$

W^2 和 A^2 检验：$W^2 > Z_\alpha$ 或 $A^2 > Z_\alpha$

6. W 检验 ($3 \leqslant n \leqslant 50$)[2,3,4,6]

W 检验的统计量为：

$$W = \frac{\left[\sum_{i=1}^{\left[\frac{n}{2}\right]} a_{in}(X_{(n-i+1)} - X_{(i)})\right]^2}{n \cdot m_2}$$

其中 a_{in} 为 W 检验系数，有表可查。

则拒绝域为 $W < Z_\alpha$

7. D 检验 ($10 \leqslant n \leqslant 2000$)[2,4]

D 检验的统计量为：

$$D = \frac{\sum_{i=1}^{n}\left[i - \frac{n+1}{2}\right] X_{(i)}}{n^2 \cdot \sqrt{m_2}}$$

则拒绝域为：$D < Z_\alpha$ 或 $D > Z_{1-\alpha}$

模拟计算结果与讨论

为了考察各种正态性检验方法的应用条件和检验功效，以便有选择地加于应用，我们利用计算机提供的伪随机数发生函数，根据随机变量的抽样原理，构造具有正态分布，标准均匀分布，Logistic 分布，Rayleigh 分布和 Weibull 分布的子样，以考察各检验方法对样本容量、分布的对称性的识别功效。在 Micro VAX II 计算机上随机模拟一万多次，(对所考察的分布类型和样本容量模拟重复次数不低于 100 次)，超过[4]，[7]和[8]中所记载的次数。主要结果如下面各表：

表 1 正态性检验的模拟计算结果

Tab. 1 Results of the Normality Tests

理论分布类型：标准正态分布 N(0,1)

Theoretical Distribution Type: Standard Normal Distribution

样品量	偏度法	校正偏度法	峰度法	校正峰度法	偏度峰度法	校正偏度峰度法	K-S 法	W^2 法	A^2 法	W^2&A^2 法	W法	D法	X^2 法
10	84	84	57	89	47	84	94	94	94	94	94	88	≡
20	96	96	88	93	86	93	93	96	93	93	89	90	≡
30	93	93	92	96	85	90	95	95	94	94	94	90	≡
40	94	94	95	95	88	89	97	98	98	96	96	97	≡
50	92	92	95	96	88	89	97	97	97	97	97	94	89
60	91	91	96	93	87	84	99	99	98	98	≡	97	92
70	95	95	96	97	92	93	98	97	97	96	≡	98	87
80	93	93	97	97	90	91	92	93	94	92	≡	95	90
90	91	91	95	97	87	89	98	96	95	95	≡	95	90
100	92	92	97	99	89	91	96	97	97	97	≡	98	94
300	92	92	87	92	80	85	92	95	95	94	≡	91	90
500	90	90	83	88	75	80	93	95	95	94	≡	87	95
700	94	94	86	88	81	83	89	93	94	93	≡	91	91
1000	91	91	84	87	76	79	95	96	96	96	≡	93	97

理论分布类型：标准均匀分布 U(0,1)

Theoretical Distribution Type: Standard Uniform Distribution

样品量	偏度法	校正偏度法	峰度法	校正峰度法	偏度峰度法	校正偏度峰度法	K-S 法	W^2 法	A^2 法	W^2&A^2 法	W法	D法	X^2 法
10	0	0	61	0	61	0	0	0	0	0	0	0	≡
20	12	12	33	0	45	4	14	10	14	14	13	5	≡
30	0	0	53	10	53	14	7	14	21	21	27	15	≡
40	0	0	79	53	79	53	25	36	54	54	86	58	≡
50	2	2	82	60	84	62	24	42	57	57	83	59	85
60	1	0	95	90	95	90	39	57	74	74	≡	75	94
70	0	0	95	94	95	94	39	61	80	80	≡	84	93
80	1	1	97	96	97	96	50	74	86	86	≡	89	98
90	1	1	100	99	100	99	58	86	94	94	≡	93	99
100	0	0	100	99	100	99	59	86	96	96	≡	97	100
300	0	0	100	100	100	100	100	100	100	100	≡	100	100
500	0	0	100	100	100	100	100	100	100	100	≡	100	100
700	0	0	100	100	100	100	100	100	100	100	≡	100	100
1000	0	0	100	100	100	100	100	100	100	100	≡	100	100

理论分布类型：标准 Logistic 分布 L(0,1)

Theoretical Distribution Type: Standard Logistic Distribution

样品量	偏度法	校正偏度法	峰度法	校正峰度法	偏度峰度法	校正偏度峰度法	K-S 法	W^2 法	A^2 法	W^2&A^2 法	W法	D法	X^2 法
10	17	17	26	6	43	17	3	0	0	0	0	0	≡
20	35	31	40	40	46	42	5	13	11	21	11	12	≡
30	33	33	24	33	44	44	6	9	9	9	3	15	≡
40	31	31	33	36	42	42	17	17	17	17	11	20	≡
50	26	26	30	35	39	44	14	14	16	16	10	21	24
60	30	29	33	39	43	47	14	22	24	24	≡	28	16
70	25	25	31	31	40	40	11	14	17	17	≡	24	14
80	26	26	42	49	48	54	14	16	19	19	≡	31	15
90	36	36	45	47	52	54	18	18	22	23	≡	41	14
100	51	50	75	78	84	85	21	28	32	32	≡	59	22
300	33	33	87	88	87	88	41	56	63	64	≡	85	32
500	42	42	92	93	94	95	53	67	77	77	≡	89	60
700	39	39	98	99	99	99	69	83	91	91	≡	98	60
1000	42	42	100	100	100	100	85	91	94	94	≡	100	83

理论分布类型：Rayleigh 分布 R(X;1)

Theoretical Distribution Type: Rayleigh Distribution

样品量	偏度法	校正偏度法	峰度法	校正峰度法	偏度峰度法	校正偏度峰度法	K-S 法	W^2 法	A^2 法	W^2&A^2 法	W法	D法	X^2 法
10	19	19	35	19	35	19	12	12	12	12	12		
20	21	21	12	12	36	21	4	8	8	8	12	12	≡
30	59	59	18	15	61	59	19	33	33	33	36	4	≡
40	39	39	11	8	42	39	16	18	26	26	33	6	≡
50	68	63	18	19	70	63	21	27	33	35	46	7	≡
60	68	68	15	16	68	68	32	45	49	50	≡	9	34
70	71	71	14	15	73	71	31	40	47	47	≡	10	27
80	75	73	14	13	77	74	39	47	56	56	≡	13	37
90	84	84	17	17	87	87	42	55	61	62	≡	11	37
100	83	83	26	20	87	84	46	52	61	61	≡	19	35
300	100	100	30	30	100	100	92	98	100	100	≡	16	79
500	100	100	30	31	100	100	98	99	100	100	≡	28	94
700	100	100	37	38	100	100	100	100	100	100	≡	40	99
1000	100	100	40	41	100	100	100	100	100	100	≡	45	100

续表 1

理论分布类型: Weibull 分布 W(X;2,1/4,0)

Theoretical Distribution Type: Weibull Distribution

样品量	偏度法	校正偏度法	峰度法	校正峰度法	偏度峰度法	校正偏度峰度法	K-S 法	W^2 法	A^2 法	W^2&A^2 法	W法	D法	X^2 法
10	6	6	34	6	34	6	0	0	0	0	6	0	≡
20	25	21	4	4	29	21	20	21	13	21	16	4	≡
30	40	40	6	9	40	40	18	21	21	24	24	7	≡
40	54	54	10	12	54	54	18	23	27	27	36	3	≡
50	53	52	9	8	57	54	28	39	43	44	53	15	22
60	57	57	14	16	62	60	34	39	43	43	≡	12	32
70	61	61	15	14	64	62	28	46	54	54	≡	14	31
80	73	71	17	15	79	75	38	49	58	58	≡	12	40
90	72	72	12	13	75	75	41	53	61	61	≡	10	30
100	78	78	31	27	90	82	49	61	69	69	≡	18	36
300	100	100	28	29	100	100	95	97	98	98	≡	21	79
500	100	100	39	41	100	100	98	100	100	100	≡	34	97
700	100	100	29	29	100	100	100	100	100	100	≡	32	99
1000	100	100	42	44	100	100	100	100	100	100	≡	38	100

注: 表中的数字为检验成功百分率。

通过随机模拟,我们可以看出:

1.偏度峰度法的功效最高, 其中偏度法对非对称分布较敏感,峰度法对对称性分布较敏感。而且偏度峰度法不需要对样品排序,计算速度快,对样品量要求不高,是一种值得推荐的好方法。

2.在大样本的情况下, 文献[5]介绍的 W^2法和 A^2 法也有较满意的功效。

3.在医学统计上广泛应用的 D 法对对称性分布敏感,但不适合非对称性分布。

4.X^2 法只在样本量足够大时才有可接受的检验功效,但功效仍不及 W^2 和 A^2 法。

5.检验功效随样品量的增加而增加。

参 考 文 献

[1] 中国科学院计算中心概率统计组, 1972。概率统计计算。472 页。科学出版社。

[2] 郭祖超, 1988。医用数理统计方法。939 页。人民卫生出版社。

[3] 夏增禄, 1987。土壤元素背景值及其研究方法。338 页。气象出版社。

[4] 梁小筠, 1988。正态性检验,数学的实践与认识, 1: 45—50。

[5] 方开泰,许建伦,1987。统计分布。366 页。科学出版社。

[6] 中山大学数学系, 1984。概率论及数理统计(上册)。362 页。高等教育出版社。

[7] 奥野忠一, 1976。分布の正规性の检定について,标准化と品质管理。

[8] Shapiro, S. S. and Wilk, M. B. 1965. An analysis of variance test for normality (Complete samples), *Biometrika*, **52**: 591—611.