

非中心分布非中心参数的数值计算^{*1)}

魏 公 毅

(中国科学院计算中心)

A NUMERICAL METHOD FOR NON-CENTRALITY PARAMETER OF NONCENTRAL DISTRIBUTION

Wei Gong-yi

(Computing Center, Academia Sinica)

Abstract

In this paper, some problems in iterative computation for non-centrality parameters of noncentral t , χ^2 and F distributions are discussed. Based on a large amount of computational simulation a selective approach for the initial formula is presented, the practical effectiveness of two iterative formulas is analyzed, and a few better iterative intervals are suggested. Thus computational efficiency increases greatly.

§ 1. 引 言

在应用 t , χ^2 和 F 分布作检验时, 为了计算检验的功效, 必然要引出相应的非中心 t , χ^2 和 F 分布, 具体公式参看[1]. 通常, 人们最关心的是, 对于某种检验方法, 当犯第一类错误的概率 α 固定以后, 非中心分布非中心参数 λ 应取什么样值, 才能使检验功效达到 $1 - \beta$, 其中 β 是犯第二类错误的概率. 也就是说, 给定 α 和 $1 - \beta$, 通过求解非线性方程来确定非中心参数 λ .

对于非中心 t 分布, 非中心参数 λ 应满足方程

$$\begin{cases} Q(t; \nu) = \alpha, \\ Q(t; \nu, \lambda) = 1 - \beta, \end{cases} \quad (1.1)$$

其中 ν 为自由度, $Q(t; \nu)$ 和 $Q(t; \nu, \lambda)$ 分别表示 t 分布和非中心 t 分布的上侧概率.

对于非中心 χ^2 分布, 非中心参数 λ 应满足方程

$$\begin{cases} Q(\chi^2; \nu) = \alpha, \\ Q(\chi^2; \nu, \lambda) = 1 - \beta, \end{cases} \quad (1.2)$$

其中 ν 为自由度, $Q(\chi^2; \nu)$ 和 $Q(\chi^2; \nu, \lambda)$ 分别表示 χ^2 分布和非中心 χ^2 分布的上侧概

* 1989年6月21日收到.

1) 国家自然科学基金资助项目.

率。

对于非中心 F 分布,非中心参数 λ 应满足方程

$$\begin{cases} Q(F; \nu_1, \nu_2) = \alpha, \\ Q(F; \nu_1, \nu_2, \lambda) = 1 - \beta, \end{cases} \quad (1.3)$$

其中 ν_1, ν_2 为自由度, $Q(F; \nu_1, \nu_2)$ 和 $Q(F; \nu_1, \nu_2, \lambda)$ 分别表示 F 分布和非中心 F 分布的上侧概率。

很明显,对于方程(1.1),(1.2),(1.3),使用迭代法计算参数 λ 要花费大量的时间。为了节省计算量,提高计算效率,对于初值公式及迭代公式的选取,对于迭代区间的确定是极其重要的。下面将分别对这三个问题进行讨论。

§2. 初值公式

在参数 λ 的迭代计算中,初值公式选得好,初值靠近真解,迭代过程收敛得快,节省计算时间;初值公式选得不好,初值偏离真解,不仅增加迭代次数,而且还有可能导致迭代过程发散,无法求出真解。参数 λ 的近似式一般都是 $\alpha, 1 - \beta$ 和自由度 ν (或 ν_1, ν_2) 的函数。当参数 λ 有许多种近似式时,总希望选一种较好的近似式作为初值。然而,用解析方法区分众多近似式的优劣有较大的困难。在对非中心参数 λ 进行了大量模拟计算的基础上,本文对初值公式的相对误差进行了分析,从而选出一种较好的近似式。作为选择初始公式的例子,下面给出自由度为 ν 的非中心 t 分布非中心参数 λ 的四个近似式:

(1) 正态近似式^[2]

$$\lambda \simeq u_\beta \sqrt{1 + (1 - d^2)t^2} + d \cdot t,$$

式中 u_β 表示正态分布上侧分位数, $t = t_{\alpha/2}(\nu)$ 表示中心 t 分布上侧分位数,

$$d \simeq 1 - \frac{1}{4\nu} + \frac{1}{32\nu^2}.$$

(2) 正态近似式^[3]

$$\lambda \simeq (u_\alpha + u_\beta) / \sqrt{1 - u_\alpha^2 / (2\nu)},$$

式中 u_α 和 u_β 为正态分布上侧分位数。

(3) Hodges-Lehmann 近似式^[1]

参数 λ 满足下面近似式:

$$\begin{aligned} x + y - \lambda \left(1 - \frac{x^2}{4\nu}\right) &\simeq \lambda \left(\frac{-6x^2 + 5x^4}{96\nu^2} + \frac{4x^2 + 6x^4 - x^6}{128\nu^3} \right) \\ &+ \lambda^2 \left(\frac{-x^3}{24\nu^2} + \frac{-2x^3 + x^5}{48\nu^3} \right), \end{aligned}$$

式中 $x = u_\alpha, y = u_\beta$ 。

(4) 山内近似式^[1]

$$\begin{aligned} \lambda &\simeq t - x + \frac{1}{114\nu}(-xt^2 - t) + \frac{1}{21(4\nu)^2} \left[xt^4 + \frac{4}{3}(x^2 - 1)t^3 + 2xt^2 + t \right] \\ &+ \frac{1}{31(4\nu)^3} [-3xt^6 - 8(x^2 - 1)t^5 - 6xt^4 + 4(x^2 - 1)t^3 + 12xt^2 + 15t], \end{aligned}$$

式中 $z = t_{\alpha/2}(\nu)$, $x = u_{\beta}$.

取 $\alpha = 0.005, 0.01, 0.05, 0.1(0.1)0.5, 0.7, 0.9$; $1 - \beta = 0.1(0.1)0.9, 0.995$; $\nu = 1, 2, 5, 10, 15, 20, 30, 50, 100, 120$ 的各种组合, 对于(1.1), 使用 Newton 法计算满足 10^{-4} 精度的 λ 值. 初值 $\hat{\lambda}_i (i = 1, 2, 3, 4)$ 分别用公式(1)–(4)产生. 初值的相对误差用 $R_i = (\hat{\lambda}_i - \lambda) / \lambda (i = 1, 2, 3, 4)$ 表示. 令 $R_{\min} = \min(|R_1|, |R_2|, |R_3|, |R_4|)$. 与 R_{\min} 相对应的公式编号记为 $N (N = 1, 2, 3, 4)$. 很明显, 编号为 N 的公式的相对误差 $|R_N|$ 小于或等于其余 3 个公式的相对误差的绝对值, 迭代次数也是如此. 用公式(N)作初值比用其余 3 个公式作初值能减少迭代次数, 因而能节省计算量. 对于一个固定的 α 值, $1 - \beta$ 和 ν 的数值共有 100 种组合, 因此可以得到 100 个 N 值的频数统计量. 频数的大小指出了某个公式的相对误差在 4 个公式中取得最小值的次数. 某个公式的频数越大, 则说明该公式在迭代中越有效. 我们给出 10 个 α 值, 4 个公式的频数表(见表 1)及频数折线图(见图 1). 对于所讨论的 1000 种组合, 使用公式(1)可以保证有 70% 以上的相对误差值 $|R_1|$ 是最小的. 另外, 公式(1)较简单, 计算量小. 显然这对于迭代计算是有利的. 所以使用公式(1)作为初值是比较合适的.

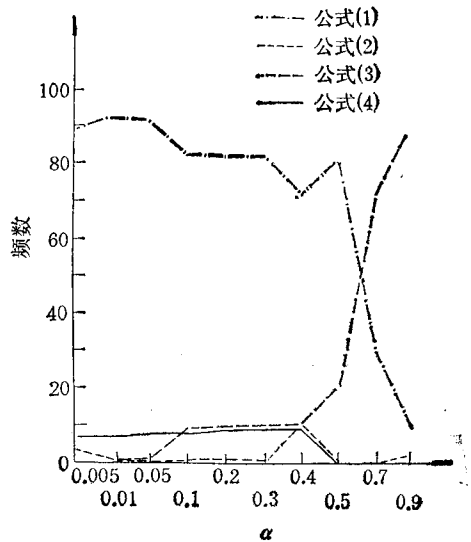


图 1 频数折线图

如果能把计算结果进行较细致的分析, 则可以为使用者提供更多的参考信息, 即在什么样的 α 值条件下, 用哪个公式最有利. 例如, 当 $0.005 \leq \alpha \leq 0.5$ 时, 使用公式(1); 当 $0.7 \leq \alpha \leq 0.9$ 时, 使用公式(3), 这种作法在

表 1 频数表

α	公 式			
	(1)	(2)	(3)	(4)
0.005	89	4	0	7
0.01	92	1	0	7
0.05	91	0	1	8
0.1	82	1	9	8
0.2	81	1	9	9
0.3	81	0	10	9
0.4	71	10	10	9
0.5	80	0	20	0
0.7	30	0	70	0
0.9	9	2	89	0
合 计	706	19	218	57

1000组计算中,将会有92.4%的相对误差值是最小的,显然对迭代计算将更有利。

对于非中心 χ^2 和 F 分布,其参数 λ 的近似式可分别由[3]中的公式(6.24)及[2]中的公式(5.5)给出。

§3. 迭代公式

对于非中心 t 和 χ^2 分布,通常使用 Newton 迭代法求解参数 λ 值。关于 λ 的一阶偏导数已经存在^[1]。

对于非中心 χ^2 分布,在关于 λ 的一阶偏导数的基础上还可以求出二阶偏导数

$$\frac{\partial^2 P(\chi^2; \nu, \lambda)}{\partial \lambda^2} = \frac{1}{4} [P(\chi^2; \nu + 4, \lambda) - 2P(\chi^2; \nu + 2, \lambda) + P(\chi^2; \nu, \lambda)],$$

其中 $P(\chi^2; \nu, \lambda)$ 是非中心 χ^2 分布的下侧概率。于是可以使用基于二阶台劳展开的 Hitotumatu 迭代法求解 λ ,具体公式可参看[4]。

为了比较两种迭代公式的实际效果,取 $\alpha = 0.01, 0.05; 1 - \beta = 0.1(0.1)0.9, 0.95; \nu = 1, 2, 5, 10, 15, 20, 30, 50, 100, 120$ 的各种组合,对于(1.2),使用§2提到的初值 $\hat{\lambda}$,用 Newton 迭代法和 Hitotumatu 迭代法分别计算满足 10^{-4} 精确度的 λ 值以及相对误差值 $R = (\hat{\lambda} - \lambda)/\lambda$ 。在200组计算中,使用 Newton 法的结果是,166组收敛,34组发散;使用 Hitotumatu 法的结果是,197组收敛,3组发散。对结果的进一步分析表明:

(1) 若 Newton 法收敛(令 n_1 为迭代次数),则 Hitotumatu 法肯定收敛(令 n_2 为迭代次数),且 $n_2 \leq n_1$ 。这种情况有166组(占83%),数值例子见表2。

(2) Newton 法发散,而 Hitotumatu 法收敛的情况有31组(占15.5%)。

(3) Newton 法和 Hitotumatu 法均发散的情况有3组(占1.5%)。

综上所述,Hitotumatu 法比 Newton 法的优越之处仅在于,对于 Newton 法发散的情况,使用 Hitotumatu 法有可能收敛。例如,在上述计算中,对于 Newton 法发散的34组情况,使用 Hitotumatu 法以后有31组收敛,约占发散组数的91%,这个结果是很吸引人的。

表2 $\alpha = 0.01, 1 - \beta = 0.2$

ν	$\hat{\lambda}$	λ	$ R $	n_1	n_2
15	25.4902	9.0724	1.8096	8	4
20	28.2201	10.3123	1.7365	7	4
30	32.7796	12.3952	1.6445	6	4
50	39.9806	15.7062	1.5455	6	4
100	53.1687	21.8100	1.4378	6	4
120	57.4597	23.8028	1.4140	6	4

对于非中心 F 分布,无法给出关于 λ 的偏导数。但是,当自由度 ν_1, ν_2 和 F 固定时,非中心 F 分布函数是参数 λ 的单调函数参看[5]。因此,使用二分法求解 λ 是最合适的。另外,对于非中心 t 和 χ^2 分布,当 Newton 法或 Hitotumatu 法失效

时,也能用二分法求解.

§ 4. 迭代区间

使用二分法时, 迭代区间的确定是非常重要的. 包含真解的区间选得大, 迭代次数多; 区间选得小, 迭代次数少, 效率高.

非中心 t, χ^2 和 F 分布参数 λ 的变化范围是 $0 \leq \lambda < \infty$. 显然, 在实际中这个区间是无法使用的. 使用已有的统计数值表(例如, [1, 2]), 能得到一定 $\alpha, 1 - \beta, \nu$ (或 ν_1, ν_2) 范围内的参数 λ 的最大值 λ_{\max} . 因此, 取区间 $0 \leq \lambda \leq \lambda_{\max}$ 也是可行的.

假设非中心参数的真解为 λ , 近似解为 $\hat{\lambda}$, 近似解的绝对误差取为 $A = |\hat{\lambda} - \lambda|$, 相对误差取为 $R = (\hat{\lambda} - \lambda)/\lambda$. 如果通过对非中心参数 λ 的大量模拟计算能求出近似解的最大绝对误差 A_{\max} , 则可以取区间 $\lambda_1 \leq \lambda \leq \hat{\lambda} + A_{\max}$, 其中 $\lambda_1 = \max(0, \hat{\lambda} - A_{\max})$. 当 A_{\max} 较小时, 这个区间也是合适的.

如果能得到最大相对误差 R_{\max} , 则可以取区间 $\hat{\lambda}/(1 + R_{\max}) \leq \lambda \leq \hat{\lambda}/(1 - R_{\max})$, 其中 $0 < R_{\max} < 1$. 如果能同时得到最大相对误差 R_{\max} 和最小相对误差 R_{\min} , 则可以取区间 $\hat{\lambda}/(1 + R_{\max}) \leq \lambda \leq \hat{\lambda}/(1 + R_{\min})$, 其中 $R_{\min} > -1$.

例如, 对于非中心 F 分布, 取 $\alpha = 0.01, 0.05, 0.1, 0.5$; $1 - \beta = 0.5, 0.7, 0.8, 0.9$; $\nu_1 = 1, 2, 4, 6, 8, 10, 20, 30, 60, 120$; $\nu_2 = 2, 5, 10, 20, 30, 40, 50, 60, 120, 240$ 的 1600 种组合情况, 对于方程(1.3), 使用二分法计算满足 10^{-4} 精确度的参数 λ , 取[2]中公式(5.5)作为近似值 $\hat{\lambda}$. 于是得到 $R_{\max} = 0.34$ 及 $R_{\min} = -0.04$. 进而可以获到两个较好的估计区间: $0.746\hat{\lambda} \leq \lambda \leq 1.515\hat{\lambda}$ 及 $0.746\hat{\lambda} \leq \lambda \leq 1.042\hat{\lambda}$. 在实际计算中, 使用这两个区间是令人满意的.

参 考 文 献

- [1] 山内二郎等, 统计数值表, 日本规格协会, 1972.
- [2] 山内二郎等, 统计数值表 ユンサイス版, 日本规格协会, 1977.
- [3] 竹内 啓, 確率分布と统计解析, 日本规格协会, 1975.
- [4] S. Hitotumatu, A method of Successive approximation based on the expansion of second order. Math. Japonicae, 7(1962), 31-50.
- [5] 方开泰, 许建伦, 统计分布, 科学出版社, 1987.