Acta Medica et Biologica
Vol. 34, No. 4, 129-137, 1987

GENERALIZED DESCRIPTION OF CLASSICAL PROBABILITY FUNCTIONS FOR COMPUTATION

ARATA SATO

Department of Psychiatry, Niigata University School of Medicine, Niigata

(Received October 13, 1986)

ABSTRACT

This is a brief report about the technical problems with the test of significance. Analyzing the computation method of the t, x^2, and F probability functions results in an improved algorithm by which P values are efficiently and accurately obtained. I rearranged these classical probability functions in a uniformed description to make some application programs for the computer. In addition, the comparison between my calculation and that of other authors suggests there are some "rounding" errors in some of the tables of the P values published in the past.

INTRODUCTION

The logic of the test of significance is an issue in scientific philosophy. Rozeboom (1960) criticized the test of significance and stated five constructive suggestions from a practical point of view as an alternative method: (a) to analyze the meaning of "probability", (b) to develop Bayes' theorem, (c) to report a confidence interval, (d) to describe the precise P value, and (e) to be free from the traditional null-hypothesis decision procedure. There have been various kinds of discussions about the application of conventional dualistic decision procedure to psychological research, but there has not been a clearly defined approach in applying the level of significance .05 to that. The concept "level of significance" is too arbitrary to be applied to the psychological research without sufficient consideration (Bakan, 1966). The Manual of the APA (1984, p. 13, pp. 80-81) says in an orthodox manner, "report descriptive statistics when reporting inferential statistics." I think we should report any P value, whether it is over .05 or not, rather than write only the comment "significant" or "not significant". In practice, however, it was considerably difficult to describe the P value itself because intricate calculation was required to obtain a satisfactory result.

In reply to Rozeboom's fourth suggestion, this report deals with the t, x^2, and F probability functions whose degrees of freedom are positive integers, and suggests that we can unify the computation algorithms throughout these probability functions:

$$Q = W + H\,(1 + \Sigma \Pi\, Ym).$$

where Q is the upper tail area of the probability distribution, W is the initial term of the integrated probability density function, H is the term whose degree of freedom is greater than or equal to two, and Ym means the general term in which the degree of freedom is greater than or equal to four (refer to the next section for further details).

It has been well noted that Fisher (1935) presented the t, x^2, and z distributions as lecture notes. In addition to Fisher's article, many expansion and approximation formulas of the classical probability functions deduced from the incomplete beta function ratio have been reported, and have been summarized in a series of papers by Toda, Shimizu, and Takeuchi (1968-1969). Zelen and Severo (1972) have already given the equations equivalent to the method stated here. We can use, moreover, the FORTRAN code by Yamauti et al (1972) to obtain reliable P values. To find the P value with a hand calculator, Lackritz (1984) described one method of calculation. His method is also adaptable for personal computer programming within the limit of small degrees of freedom.

The expansion formulas using the integration of parts are most applicable to computer programming (Takeuchi, 1970), but it is rather complicated to understand their algorithms and programs. We can write the computer programs of the classical probability functions more efficiently by using the algorithm in the following paragraph. Although the expansion formulas of the t, x^2, and F probability density functions are already known as stated above, it has not been reported, as far as I know, that the upper probability of these three functions can be calculated by a single uniformed algorithm.

DESCRIPTION OF EQUATIONS

When the expression f is a probability density function, its integral Q, termed the upper probability function or the P value function, is described in the following way:

$$Q\ \{\omega \geqq \lambda\} = \int_{\lambda}^{\infty} f(\omega)\, d\omega.$$

In this report, each figure of the set {i, j, k, l, m, n} means natural number, and the expression,

$$Q_i = W_{i \geqq 1} + H_{i \geqq 2}\,\Big(1 + \sum_{l=1}^{\substack{i \geqq 4 \\ [i/2-1]}} \prod_{m=1}^{1} Ym\Big)$$

represents the following equations:

$$Q_1 = W_1 ,$$

$$Q_2 = W_2 + H_2,$$

$$Q_3 = W_3 + H_3,$$

and

$$Q_{n \geqq 4} = W_n + H_n \left(1 + \sum_{l=1}^{[n/2-1]} \prod_{m=1}^{l} Ym\right).$$

Degree of freedom is placed at the position i.

Also, the inverse function of these equations can be used to obtain the percentage point values directly.

"Student's" t Distribution

The probability density function of the t distribution is defined as follows:

$$f_n(t) = \frac{1}{n^{1/2} \cdot B(1/2, n/2)} \cdot (1 + (t^2/n))^{-(n+1)/2}$$

where $B(\beta, \gamma)$ means the beta function and n shows the degree of freedom. The upper probability function is expressed as follows:

$$2Q_{odd\ n}\{t \geqq \lambda\} = 1 - (2/\pi) \cdot \tan^{-1}((1/y) - 1)^{1/2} - ((2/\pi) \cdot (y \cdot (1-y))^{1/2})_{n \geqq 2} \cdot \left(1 + \sum_{l=1}^{\substack{n \geqq 4 \\ [n/2-1]}} \prod_{m=1}^{l} ((2m/(2m+1)) \cdot y)\right)$$

and

$$2Q_{even\ n}\{t \geqq \lambda\} = 1 - ((1-y)^{1/2})_{n \geqq 2} \cdot \left(1 + \sum_{l=1}^{\substack{n \geqq 4 \\ [n/2-1]}} \prod_{m=1}^{l} (((2m-1)/(2m)) \cdot y)\right)$$

where $y = n/(\lambda^2 + n)$.

The reciprocal expressions of this upper probability function set with fixed n values are deduced as:

$$\lambda\{2Q_1\{t \geqq \lambda\} = \alpha\} = \tan((\pi/2) \cdot (1-\alpha))$$

and

$$\lambda\{2Q_2\{t \geqq \lambda\} = \alpha\} = 2^{1/2} \cdot \tan(\sin^{-1}(1-\alpha))$$
$$= ((2/(\alpha \cdot (2-\alpha))) - 2)^{1/2}.$$

Pearson's x^2 Distribution

The probability density function of the x^2 distribution is:

$$f_n(x^2) = \frac{1}{2\Gamma(n/2)} \cdot (x^2/2)^{n/2-1} \cdot \exp(-x^2/2)$$

where $\Gamma(\gamma)$ means the gamma function. When the symbol $Q\{z \geqq \lambda\}$ represents the upper probability function of the standard normal distribution, the upper tail area of the x^2 distribution is described as follows:

$$Q_{odd\ n}\{x^2 \geqq \lambda\} = 2Q\{z \geqq \lambda^{1/2}\} + ((2 \cdot \lambda/\pi)^{1/2} \cdot \exp(-\lambda/2))_{n \geqq 2} \cdot (1 + \sum_{l=1}^{\overset{n \geqq 4}{[n/2-1]}} \prod_{m=1}^{l} (\lambda/(2m+1)))$$

and

$$Q_{even\ n}\{x^2 \geqq \lambda\} = (\exp(-\lambda/2))_{n \geqq 2} \cdot (1 + \sum_{l=1}^{\overset{n \geqq 4}{[n/2-1]}} \prod_{m=1}^{l} (\lambda/(2m))).$$

If we want the P value of the standard normal distribution, we are able to use, for example, the following continued fractions:

$$Q\{z \geqq \lambda\} = 1/2 - (\lambda \cdot g(\lambda)/a_1)$$
$$a_n = 2n - 1 + ((-1)^n n\lambda^2/a_{n+1})$$

(Shenton ,1954)

or

$$Q\ \{z \geqq \lambda\} = g(\lambda)/a_1$$
$$a_n = \lambda + (n/a_{n+1})$$

(Laplace, 1805
[cited in Shenton, 1954])

where

$$g(z) = \frac{1}{(2\pi)^{1/2}} \cdot \exp(-z^2/2).$$

The next equation is well-known one:

$$\lambda\{Q_2\{x^2 \geqq \lambda\} = \alpha\} = -2\ln\alpha.$$

Snedecor's F Distribution

The probability density function of the F distribution is:

$$f_{n1,n2}(F)=\frac{n_1^{n_1/2}\cdot n_2^{n_2/2}}{B(n_1/2,n_2/2)}\cdot\frac{F^{(n_1-1)/2}}{(n_1\cdot F+n_2)^{(n_1+n_2)/2}}\quad .$$

The upper tail area of the F distribution is expressed as follows:

$$\begin{aligned}Q_{odd\ n_1,any\ n_2}\{F\geqq\lambda\}&=2Q_{n_2}\{t\geqq(n_1\cdot\lambda)^{1/2}\}\\&+(2\cdot((1-y)\cdot y^{n2})^{1/2}/B(1/2,n_2/2))_{n_1\geqq2}\\&\cdot\Big(1+\sum_{l=1}^{\substack{n_1\geqq4\\ [n_1/2-1]}}\prod_{m=1}^{l}(((n_2+2m-1)/(2m+1))\cdot(1-y))\Big)\end{aligned}$$

and

$$\begin{aligned}Q_{even\ n_1,any\ n_2}\{F\geqq\lambda\}&=(y^{n_2/2})_{n_1\geqq2}\\&\cdot\Big(1+\sum_{l=1}^{\substack{n_1\geqq4\\ [n_1/2-1]}}\prod_{m=1}^{l}(((n_2+2m-2)/(2m))\cdot(1-y))\Big)\end{aligned}$$

where $y=n_2/(n_1\cdot\lambda+n_2)$.

The above equation set is equivalent to the next one:

$$\begin{aligned}Q_{any\ n_1,odd\ n_2}\{F\geqq\lambda\}&=1-2Q_{n_1}\{t\geqq(n_2/\lambda)^{1/2}\}\\&-(2\cdot(y\cdot(1-y)^{n_1})^{1/2}/B(1/2,n_1/2))_{n_2\geqq2}\\&\cdot\Big(1+\sum_{l=1}^{\substack{n_2\geqq4\\ [n_2/2-1]}}\prod_{m=1}^{l}(((n_1+2m-1)/(2m+1))\cdot y)\Big)\end{aligned}$$

and

$$\begin{aligned}Q_{any\ n_1,even\ n_2}\{F\geqq\lambda\}&=1-((1-y)^{n_1/2})_{n_2\geqq2}\\&\cdot\Big(1+\sum_{l=1}^{\substack{n_2\geqq4\\ [n_2/2-1]}}\prod_{m=1}^{l}(((n_1+2m-2)/(2m))\cdot y)\Big)\end{aligned}$$

where $y=n_2/(n_1\cdot\lambda+n_2)$,

which is deduced from the relation:

$$Q_{n_1,n_2}\{F\geqq\lambda\}=1-Q_{n_2,n_1}\{F\geqq1/\lambda\}.$$

The beta function is defined by the following equation:

$$B(\beta, \gamma) = \int_0^1 x^{\beta-1}(1-x)^{\gamma-1}dx,$$

and here is denoted B (1/2, k/2):

$$B(1/2, 4m/2) = 4/(4m-1) \cdot \prod_{j=1}^{m-1} (8j/(2j+2m-1)),$$

$$B(1/2, (4m-1)/2) = \pi/2 \cdot \prod_{j=1}^{m-1} ((2j+2m-1)/(8j)),$$

$$B(1/2, (4m-2)/2) = 2/(2m-1) \cdot \prod_{j=1}^{m-1} (8j/(2j+2m-1)),$$

and

$$B(1/2, (4m-3)/2) = (2m-1) \cdot \pi/(4m-3) \cdot \prod_{j=1}^{m-1} ((2j+2m-1)/(8j))$$

where $m = [(k+3)/4]$.

Also, the following equations can be obtained for the percentage points:

$$\lambda\{Q_{1,1}\{F \geqq \lambda\} = \alpha\} = \tan^2((\pi/2) \cdot (1-\alpha)),$$

$$\lambda\{Q_{2,n2}\{F \geqq \lambda\} = \alpha\} = (n_2/2) \cdot (\alpha^{-2/n_2} - 1)$$

$$\because Q_{2,n_2}\{F \geqq \lambda\} = (n_2/(2\lambda + n_2))^{n_2/2},$$

$$\lambda\{Q_{n1,2}\{F \geqq \lambda\} = \alpha\} = (2/n_1) \cdot ((1-(1-\alpha)^{2/n1})^{-1} - 1)$$

$$\because Q_{n1,2}\{F \geqq \lambda\} = 1 - (n_1\lambda/(n_1\lambda + 2))^{n_1/2},$$

$$\lambda\{Q_{2,\infty}\{F \geqq \lambda\} = \alpha\} = -\ln \ \alpha,$$

and

$$\lambda\{Q_{\infty,2}\{F \geqq \lambda\} = \alpha\} = -(\ln(1-\alpha))^{-1}.$$

AN APPLICATION

We can obtain the percentage point values by applying not only Newton's method of successive approximations but also by using the following simple arithmetic relation as well.

When we read, for example, a percentage point vblue as

$$\lambda\{Q\{\omega \geqq \lambda\} = \alpha\} \fallingdotseq 1.23,$$

we are to define the set of numbers,

Table 1. Discrepancies in t values from Fisher and Yates

2Q	n	author	Fisher and Yates
.3	60	1.045	1.046
.001	2	31.599	31.598
	23	3.768	3.767

Table 2. Discrepancies in x^2 values from Pearson and Hartley

Q	n	author	Pearson and Hartley
.995	23	9.26042	9.26043
.990	7	1.239042	1.239043
	23	10.19572	10.19567
.975	21	10.28290	10.28293
.750	1	0.1015310	0.1015308
	3	1.212533	1.212534
.005	14	31.3193	31.3194
.001	8	26.124	26.125
	25	52.620	52.618

Table 3. Major Discrepancies in F values from Snedecor and Cochran

Q	n_1	n_2	author	Snedecor and Cochran
.25	30	60	1.22	1.24
	60	60	1.19	1.21
.05	6	18	2.66	3.66
.025	15	7	4.57	5.47
	15	8	4.10	4.20
	30	7	4.36	5.36
	40	7	4.31	5.31
	60	120	1.53	1.63
	∞	11	2.88	2.83
	∞	26	1.88	1.83
.01	75	16	2.90	2.98

Note. Most of these discrepancies are simply typographical errors in quotation.

$$\{\lambda \mid 1.225 \leqq \lambda < 1.235\}.$$

If the inequality,

$$Q\{\omega \geqq 1.225\} \geqq \alpha > Q\{\omega \geqq 1.235\}$$

is correct, we will say that the percentage point value is 1.23 for α. Special attention should be paid to the fact that the validity of this method depends on the accuracy of only the Q calculation.

Reexamination of percentage point values in some well-known statistical tables shows (see Table 1-3) that there are three discrepancies between the present analysis and the table of "Distribution of t" by Fisher and Yates (1982, p. 46), nine discrepancies between this author's version and the table of "Percentage points of the x^2 distribution" by Pearson and Hartley (1976, pp. 136-137), and 578 discrepancies out of 5080 between the present version and the table of "Variance ratio of F" by Snedecor and Cochran (1980, pp. 480-488). Discrepancies are not found in Kuebler's table of t (1976, p. 303) and Yamauti's tables of x^2, t, and F (1979, pp. 10-11, p. 14, pp. 16-25).

Many authors have calculated and published the percentage point values of the classical probability functions, which were collected and introduced by Greenwood and Hartley (1962). Norton (1952) reexamined three tables of the famous volume by Fisher and Yates (1948), and reported some errors contained in their tables.

It is actually impractical to reexamine all of the tables shown by Greenwood and Hartley in the same way as Norton's approach. I mentioned one method of reexamination and applied it to some statistical tables esteemed for their creative achievements. Although this kind of verification does not always guarantee us of the mathematical conclusiveness of the algorithm in this report, perhaps it indicates that a handicraft calculation is more troublesome than a computerized one.

The program sources used to obtain the values presented in this report are the NEC/PC-9801 BASIC(double precision) code and the NEC/MS-70 FORTRAN (DOUBLE PRECISION) code in Information Processing Center, Niigata University School of Medicine.

ACKNOWLEDGEMENTS

I wish to thank Prof Yamamoto and Dr Takayama, Department of Hygiene and Preventive Medicine, Niigata University School of Medicine, for their suggestive reading, and Prof Ihda, Department of Psychiatry, Niigata University School of Medicine, for his helpful comments.

REFERENCES

1) American Psychological Association (1984): *Publication Manual of the American Psychological*

Association, 3rd ed.. Washington, DC, American Psychological Association, 1984, p. 13, pp. 80-81.

2) Bakan, D.: The test of significance in psychological research. *Psychological Bulletin,* **66**: 423-437, 1966.

3) Fisher, R. A.: The mathematical distributions used in the common tests of significance. *Econometrica,* **3**: 353-365, 1935.

4) Fisher, R. A., and Yates, F.: *Statistical Tables for Biological, Agricultural and Medical Research,* 3rd ed.. London, Oliver and Boyd, 1949, p. 32.

5) Fisher, R. A., and Yates, F.: *Statistical Tables for Biological, Agricultural and Medical Research,* 6th ed.. Hong Kong, Longman, 1982, p. 46.

6) Greenwood, J. A., and Hartley, H. O.: *Guide to Tables in Mathematical Statistics.* Princeton, Princeton University Press, 1962.

7) Kuebler, R. R., and Smith, H. Jr.: *Statistics, a Beginning.* NY, John Wiley and Sons, 1972, p. 303.

8) Lackritz, J. R.: Exact p values for F and t tests. *The American Statistician,* **38**: 312-314, 1984.

9) Norton, H. W.: Mathematical tables-errata 199. *Mathematical Tables and other Aids to Computation,* **6**: 35-38, 1952.

10) Pearson, E. S., and Hartley, H. O.: *Biometrika Tables for Statisticians,* Vol. 1, London, Biometrika trust, 1976, pp. 136-137.

11) Rozeboom, W. W.: The fallacy of the null-hypothesis significance test. *Psychological Bulletin,* **57**: 416-428, 1960.

12) Shenton, L. R.: Inequalities for the normal integral including a new continued fraction. *Biometrika,* **41**: 177-189, 1954.

13) Snedecor, G. W., and Cochran, W. G.: *Statistical Methods,* 7th ed.. Ames, The Iowa State University Press, 1980, pp. 480-488.

14) Takeuchi, J.: Statistical distribution by utilization of computer (18). *Standardization and Quality Control,* **23(4)**: 49-54, 1970.

15) Toda, H., Shimizu, R., and Takeuchi, J.: Statistical distributions and their electronic computation (1). *Standardization and Quality Control,* **21(9)**: 59-64, 1968.

16) Toda, H., Shimizu, R., and Takeuchi, J.: Statistical distributions and their electronic computation (2). *Standardization and Quality Control,* **21(11)**: 45-52, 1968.

17) Toda, H., Shimizu, R., and Takeuchi, J.: Statistical distributions and their electronic computation (3). *Standardization and Quality Control,* **21(12)**: 35-42, 1968.

18) Toda, H., Shimizu, R., and Takeuchi, J.: Statistical distributions and their electronic computation (4). *Standardization and Quality Control,* **22(1)**: 49-57, 1969.

19) Toda, H., Shimizu, R., and Takeuchi, J.: Statistical distributions and their electronic computation (5). *Standardization and Quality Control,* **22(2)**: 51-60, 1969.

20) Toda, H., Shimizu, R., and Takeuchi, J.: Statistical distributions and their electronic computation (6). *Standardization and Quality Control,* **22(3)**: 44-58, 1969.

21) Toda, H., Shimizu, R., and Takeuchi, J.: Statistical distributions and their electronic computation (11). *Standardization and Quality Control,* **22(8)**: 67-71, 1969.

22) Yamauti, Z. (ed.): *Statistical Tables and Formulas with Computer Applications.* Tokyo, Japanese Standards Association, 1972.

23) Yamauti, Z. (ed.): *Concise Statistical Tables.* Tokyo, Japanese Standards Association, 1979.

24) Zelen, M., and Severo, N. C.: Probability functions. In *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables,* edited by Abramowitz, M., Stegun, I. A., and National Bureau of Standards. NY, John Wiley and Sons, 1972, pp. 925-995.