Zdeněk Režný; Josef Jirkovský

Algorithms. 47. ALPHA CHISQ. An algorithm for the chi-square test without application of critial values

Aplikace matematiky, Vol. 28 (1983), No. 2, 153–155

Persistent URL: http://dml.cz/dmlcz/104018

47. ALPHA CHISQ

AN ALGORITHM FOR THE CHI-SQUARE TEST WITHOUT APPLICATION OF CRITICAL VALUES

Ing. ZDENĚK REŽNÝ, CSc., Ing. JOSEF JIRKOVSKÝ

Ústav biofyziky a nukleární medicíny Fakulty všeobecného lékařství Karlovy univerzity, Salmovská 3, 120 00 Praha 2

The algorithm suggested in this paper computes the probability that a statistic having the χ^2-distribution will exceed the value x actually observed, i.e.

$$(1) \qquad \alpha_n(x) = \frac{1}{2^{n/2}\,\Gamma(n/2)} \int_x^\infty y^{n/2-1}\,e^{-y/2}\,dy \quad (n \geqq 1, x \geqq 0),$$

where n is the number of degrees of freedom. This probability may be hence called the significance degree, similarly as in the cases of Student's t- and Snedecor's F-statistic treated in our previous papers [1] and [2], respectively.

The algorithm is based, for the values of n not exceeding 30, on the following recurrent relations and initial conditions

$$(2) \qquad \alpha_n(x) = \alpha_{n-2}(x) + \frac{(x/2)^{n/2-1}\,e^{-x/2}}{\Gamma(n/2)},$$

$$(3) \qquad \alpha_0(x) \overset{\text{def}}{=} 0\,, \quad \alpha_1(x) = \sqrt{(2/\pi)} \sum_{k=0}^{\infty} \frac{(-1)^k\,x^{k+1/2}}{2^k k!\,(2k+1)}$$

$$(n \geqq 2, x \geqq 0);$$

these statements can be proved by differentiation, as they obviously hold for $x = 0$. For greater values of n, however, the same normal approximation is used as in the subroutine CDTR in the IBM packet SSP3 [6], with the difference that in the evaluation of the normal distribution function, the function $\alpha_1(x)$ (see (1), (3)) is used instead of Hasting's approximation.

The above mentioned subroutine CDTR is widely branched with respect to various domaines of the (x, n)-plane. On the contrary, the algorithm presented is very concise and is suitable for cases of x values lying in a currently limited diapazon. Thus, it is more comparable with another competitive algorithm *khi* of Barra [7].

However, it is shorter, especially because the algorithm *khi* uses a certain asymptotic expression for the calculation of $\alpha_1(x)$ at higher value of x, while the algorithm presented simply omits this calculation at all. Besides, let us note that the series with alternating signs in (3) arises from the term by term integration of the series gained by the Maclaurin expansion of the exponential in the integrand, while in *khi*, another series (for smaller values of x) is used, but with positive terms such that it is difficult to estimate its rest.

```
real procedure ALPHA CHISQ(x, n); real x; integer n;
begin
  real a;
  real procedure P(x, n); value x; real x; integer n;
  begin real a, b, c, d, e, f; integer j;
       j := n − (n ÷ 2) × 2;
       if j = 0 then begin f := 1; L1 : a := 0; go to L2 end;
       b := f := 0·79788456080 × sqrt(x);
       if x ≧ 16·5 then go to L1;
       a := 1 − b: d := e := 6;
       for c := 1, c + 2 while abs(b) = 0.00005 do
           begin b := −b × c × x/d; a := a − b;
                 e := e + 8; d := d + e
           end;
    L2: b := f × exp (−0.5 × x);
       for d := j + 2 step 2 until n do
           begin a := a + b; b := b × x/d end;
       P := a
  end P;
if n ≦ 30 then ALPHA CHISQ := P(x, n)
         else
         begin a := 4.5 × n;
               a := ((x/n) ↑ (1/3) − 1 + 1/a) × sqrt(a);
               ALPHA CHISQ := 0.5 × (1 + sign(a) × (P(a ↑ 2, 1) − 1))
         end
end
```

As is seen from (2), (3), if n is not greater than 30, the result in case of an even n is given by a sum of a finite number, viz. $n/2$, of terms, which are calculated without any limitation of precision, whereas for an odd n, the constants 16·5 and 0·00005 used in the algorithm lead to the accuracy of 4 decimal places. For n greater than 30, however, the normal approximation can sometimes have a worse precision.

We give some check values:

$$ALPHA\ CHISQ\ (12{\cdot}116,\ 1) = 0{\cdot}0005$$
$$ALPHA\ CHISQ\ (4{\cdot}0,\ 1) = 0{\cdot}0455$$

$$
\begin{aligned}
&ALPHA\ CHISQ\ (0{\cdot}001, 2) &&= 0{\cdot}9995\\
&ALPHA\ CHISQ\ (3{\cdot}0, 3) &&= 0{\cdot}3916\\
&ALPHA\ CHISQ\ (1{\cdot}0, 4) &&= 0{\cdot}9098\\
&ALPHA\ CHISQ\ (15{\cdot}987, 10) &&= 0{\cdot}1000\\
&ALPHA\ CHISQ\ (26{\cdot}0, 19) &&= 0{\cdot}1302\\
&ALPHA\ CHISQ\ (31{\cdot}1382, 40) &&= 0{\cdot}8413\\
&ALPHA\ CHISQ\ (48{\cdot}8617, 40) &&= 0{\cdot}1587
\end{aligned}
$$

For further checking of the performance of the procedure it is possible to use not only the tables of critical values or distribution function of χ^2, e.g. [3], [4], but also those of the incomplete Γ-function or, for the cases $n = 1$ and $n > 30$, of the normal distribution, after performing the corresponding necessary transforms.

The program has been tested in the symbolic language FORTRAN IV-PLUS [5] and implemented in the Institute of Biophysics and Nuclear Medicine, Faculty of General Medicine, Charles University, for the computer PDP 11/34.

References

[1] *Z. Režný, J. Jirkovský:* STUDENT. An algorithm for Student's t-test without application of critical values, Aplikace matematiky 19 (1974), 133—135.

[2] *Z. Režný, J. Jirkovský:* SNEDECOR. An algorithm for Fisher-Snedecor's F-test without application of critical values, Aplikace matematiky 19 (1974), 286—289.

[3] *J. Janko:* Statistical Tables (in Czech), Publ. House of the Czechoslovak Acad. of Sci., Prague 1958.

[4] *J. Likeš, J. Laga:* Basic Statistical Tables (in Czech), Nat. Publ. House of Technical Literature (SNTL), Prague 1978.

[5] PDP-11 FORTRAN. Language Reference Manual, Digital Equipment Corporation, Maynard, Mass., 1979.

[6] *R. E. Bargmann, S. P. Ghosh:* Statistical distribution programs for a computer language, IBM Research Report RC-1094, 1963.

[7] *J.-R. Barra*: Notions fondamentales de statistique mathématique, Dunod, Paris, 1971.