1 SUMMARY

To evaluate the **cumulative chi-squared probability function,** i.e. given a statistic $\hat{x}$ distributed as χ^2 with n degrees a freedom evaluate the probability of $\hat{x} \geq x$

$$Q(x,n) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \int_x^\infty t^{\frac{n}{2}-1} e^{-\frac{t}{2}}\, dt$$

to an absolute accuracy, where $x \geq 0$ and $n \geq 1$ an integer.

The approximations used are; a convergent series based on the expansion for the integral $Q(x,n)$; an asymptotic series for the complement probability $P(x,n) = 1 - Q(x,n)$; and for very large x and n an approximation based on a Gaussian distribution. Empirical boundaries $Q(x,n) \leq 10^{-\beta}$ and $Q(x,n) \geq 1 - 10^{-\beta}$ are used to define regions in which the probability, within the guaranteed accuracy, can be set to 0 or 1 and allow truncated forms of the series to be used giving savings in computer time.

ATTRIBUTES — **Version:** 1.0.0. **Types:** `SA01A`; `SA01AD`. **Calls:** `FC03, FC07`. **Original date:** August 1971. **Origin:** M.J.Hopper and J. Hedger, Harwell.

2 HOW TO USE THE PACKAGE

2.1 Argument list

The single precision version

```
CALL SA01A(XE,N,Q)
```

The double precision version

```
CALL SA01AD(XE,N,Q)
```

`XE` is a `REAL` (`DOUBLE PRECISION` in the D version) variable and must be set by the user to x the estimate of the chi-squared value. **Restriction:** $x \geq 0$.

`N` is an `INTEGER` and must be set by the user to n the number of degrees of freedom. **Restriction:** $n \geq 1$.

`Q` is a `REAL` (`DOUBLE PRECISION` in the D version) variable which is set by the subroutine to the value of the probability $Q(x,n)$. If the restrictions on x and n are violated Q is returned set to 2 to indicate an error; a diagnostic message is also printed.

3 GENERAL INFORMATION

Use of common: none.

Workspace: none.

Other routines called directly: none.

Input/output: a diagnostic message is printed when the restrictions on x and n are violated.

Restrictions: $x \geq 0$, $n \geq 1$.

Accuracies:

5×10^{-5} 4-byte arithmetic.

5×10^{-10} 8-byte arithmetic.

4 METHOD

The region defined by the restrictions $x \geq 0$, $n \geq 1$ is divided into subregions. Two empirical boundaries define the subregion

$$10^{-\beta} \leq Q(x,n) \leq 1-10^{-\beta} \quad (1)$$

where $\beta = 6$ for single precision and $\beta = 11$ for double precision. Outside region (1) the value of Q can be safely set to 0 (or 1) within the accuracy guaranteed.

The inside of region (1) is further subdivided into the regions

$$x + n > \gamma \quad (2)$$

and

$$x + n \leq \gamma \quad (3)$$

where $\gamma = 170$ single precision and $\gamma = 40\,000$ double precision.

For region (2) an extension to the approximation based upon a Gaussian distribution, (N. C. Severo and M. Zelen, Biometrika, Vol. 47, 1960, p.411-413, is used and is very cheap to compute.

Region (3) is further divided into

$$x \geq \tfrac{1}{2}(n - 2 - x\,(\mathrm{mod}2))$$

where a series expansion for the integral Q is used, and

$$x < \tfrac{1}{2}(n - 2 - x\,(\mathrm{mod}2))$$

where an asymptotic expansion for the complement of Q, $P(x,n) = 1 - Q(x,n)$ is used. Where possible, truncated forms of the series are used truncated at points calculated from the empirical boundaries. In the double precision version the truncation can result in considerable savings in computer time.