NAG Library Routine Document

G11SAF

Note: before using this routine, please read the Users' Note for your implementation to check the interpretation of ***bold italicised*** terms and other implementation-dependent details.

1 Purpose

G11SAF fits a latent variable model (with a single factor) to data consisting of a set of measurements on individuals in the form of binary-valued sequences (generally referred to as score patterns). Various measures of goodness-of-fit are calculated along with the factor (theta) scores.

2 Specification

```
SUBROUTINE G11SAF (IP, N, GPROB, NS, X, LDX, IRL, A, C, IPRINT, CGETOL,        &
                   MAXIT, CHISQR, ISHOW, NITER, ALPHA, PIGAM, CM, LDCM, G,      &
                   EXPP, LDEXPP, OBS, EXF, Y, XL, IOB, RLOGL, CHI, IDF,         &
                   SIGLEV, W, LW, IFAIL)

INTEGER            IP, N, NS, LDX, IRL(NS), IPRINT, MAXIT, ISHOW, NITER,        &
                   LDCM, LDEXPP, IOB(NS), IDF, LW, IFAIL
REAL (KIND=nag_wp) A(IP), C(IP), CGETOL, ALPHA(IP), PIGAM(IP),                  &
                   CM(LDCM,2*IP), G(2*IP), EXPP(LDEXPP,IP),                     &
                   OBS(LDEXPP,IP), EXF(NS), Y(NS), XL(NS), RLOGL, CHI,          &
                   SIGLEV, W(LW)
LOGICAL            GPROB, X(LDX,IP), CHISQR
```

3 Description

Given a set of p dichotomous variables $\tilde{x} = \left(x_1, x_2, \ldots, x_p\right)'$, where $'$ denotes vector or matrix transpose, the objective is to investigate whether the association between them can be adequately explained by a latent variable model of the form (see Bartholomew (1980) and Bartholomew (1987))

$$G\{\pi_i(\theta)\} = \alpha_{i0} + \alpha_{i1}\theta. \qquad (1)$$

The x_i are called item responses and take the value 0 or 1. θ denotes the latent variable assumed to have a standard Normal distribution over a population of individuals to be tested on p items. Call $\pi_i(\theta) = P(x_i = 1 \mid \theta)$ the item response function: it represents the probability that an individual with latent ability θ will produce a positive response (1) to item i. α_{i0} and α_{i1} are item parameters which can assume any real values. The set of parameters, α_{i1}, for $i = 1, 2, \ldots, p$, being coefficients of the unobserved variable θ, can be interpreted as 'factor loadings'.

G is a function selected by you as either Φ^{-1} or logit, mapping the interval $(0, 1)$ onto the whole real line. Data from a random sample of n individuals takes the form of the matrices X and R defined below:

$$X_{s\times p} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{s1} & x_{s2} & \ldots & x_{sp} \end{bmatrix} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_s \end{bmatrix}, \qquad R_{s\times 1} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_s \end{bmatrix}$$

where $\tilde{x}_l = \left(x_{l1}, x_{l2}, \ldots, x_{lp}\right)$ denotes the lth score pattern in the sample, r_l the frequency with which $\tilde{x}_l$ occurs and s the number of different score patterns observed. (Thus $\sum_{l=1}^{s} r_l = n$). It can be shown that the log-likelihood function is proportional to

$$\sum_{l=1}^{s} r_l \log P_l,$$

where

$$P_l = P(\tilde{x} = \tilde{x}_l) = \int_{-\infty}^{\infty} P(\tilde{x} = \tilde{x}_l \mid \theta)\phi(\theta)\, d\theta \tag{2}$$

($\phi(\theta)$ being the probability density function of a standard Normal random variable).

P_l denotes the unconditional probability of observing score pattern $\tilde{x}_l$. The integral in (2) is approximated using Gauss–Hermite quadrature. If we take $G(z) = \operatorname{logit} z = \log\left(\frac{z}{1-z}\right)$ in (1) and reparameterise as follows,

$$\begin{aligned} \alpha_i &= \alpha_{i1}, \\ \pi_i &= \operatorname{logit}^{-1} \alpha_{i0}, \end{aligned}$$

then (1) reduces to the logit model (see Bartholomew (1980))

$$\pi_i(\theta) = \frac{\pi_i}{\pi_i + (1 - \pi_i) \exp(-\alpha_i \theta)}.$$

If we take $G(z) = \Phi^{-1}(z)$ (where Φ is the cumulative distribution function of a standard Normal random variable) and reparameterise as follows,

$$\begin{aligned} \alpha_i &= \frac{\alpha_{i1}}{\sqrt{(1 + \alpha_{i1}^2)}} \\ &, \\ \gamma_i &= \frac{-\alpha_{i0}}{\sqrt{(1 + \alpha_{i1}^2)}} \end{aligned}$$

then (1) reduces to the probit model (see Bock and Aitkin (1981))

$$\pi_i(\theta) = \phi\left(\frac{\alpha_i \theta - \gamma_i}{\sqrt{(1 - \alpha_i^2)}}\right).$$

An E-M algorithm (see Bock and Aitkin (1981)) is used to maximize the log-likelihood function. The number of quadrature points used is set initially to 10 and once convergence is attained increased to 20.

The theta score of an individual responding in score pattern $\tilde{x}_l$ is computed as the posterior mean, i.e., $E(\theta \mid \tilde{x}_l)$. For the logit model the component score $X_l = \sum_{j=1}^{p} \alpha_j x_{lj}$ is also calculated. (Note that in calculating the theta scores and measures of goodness-of-fit G11SAF automatically reverses the coding on item j if $\alpha_j < 0$; it is assumed in the model that a response at the one level is showing a higher measure of latent ability than a response at the zero level.)

The frequency distribution of score patterns is required as input data. If your data is in the form of individual score patterns (uncounted), then G11SBF may be used to calculate the frequency distribution.

4 References

Bartholomew D J (1980) Factor analysis for categorical data (with Discussion) *J. Roy. Statist. Soc. Ser. B* **42** 293–321

Bartholomew D J (1987) *Latent Variable Models and Factor Analysis* Griffin

Bock R D and Aitkin M (1981) Marginal maximum likelihood estimation of item parameters: Application of an E-M algorithm *Psychometrika* **46** 443–459

5 Parameters

1: IP – INTEGER *Input*

On entry: p, the number of dichotomous variables.

Constraint: $\text{IP} \geq 3$.

2: N – INTEGER *Input*

On entry: n, the number of individuals in the sample.

Constraint: $\text{N} \geq 7$.

3: GPROB – LOGICAL *Input*

On entry: must be set equal to .TRUE. if $G(z) = \Phi^{-1}(z)$ and .FALSE. if $G(z) = \operatorname{logit} z$.

4: NS – INTEGER *Input*

On entry: NS must be set equal to the number of different score patterns in the sample, s.

Constraint: $2 \times \text{IP} < \text{NS} \leq \min(2^{\text{IP}}, \text{N})$.

5: X(LDX,IP) – LOGICAL array *Input/Output*

On entry: the first s rows of X must contain the s different score patterns. The lth row of X must contain the lth score pattern with $\text{X}(l,j)$ set equal to .TRUE. if $x_{lj} = 1$ and .FALSE. if $x_{lj} = 0$. All rows of X must be distinct.

On exit: given a valid parameter set then the first s rows of X still contain the s different score patterns. However, the following points should be noted:

(i) If the estimated factor loading for the jth item is negative then that item is re-coded, i.e., 0s and 1s (or .TRUE. and .FALSE.) in the jth column of X are interchanged.

(ii) The rows of X will be reordered so that the theta scores corresponding to rows of X are in increasing order of magnitude.

6: LDX – INTEGER *Input*

On entry: the first dimension of the array X as declared in the (sub)program from which G11SAF is called.

Constraint: $\text{LDX} \geq \text{NS}$.

7: IRL(NS) – INTEGER array *Input/Output*

On entry: the ith component of IRL must be set equal to the frequency with which the ith row of X occurs.

Constraints:

$$\text{IRL}(i) \geq 0, \text{ for } i = 1, 2, \ldots, s;$$
$$\sum_{i=1}^{s} \text{IRL}(i) = n.$$

On exit: given a valid parameter set then the first s components of IRL are reordered as are the rows of X.

8: A(IP) – REAL (KIND=nag_wp) array *Input/Output*

On entry: $\text{A}(j)$ must be set equal to an initial estimate of α_{j1}. **In order to avoid divergence problems with the E-M algorithm you are strongly advised to set all the A(j) to 0.5.**

On exit: $\text{A}(j)$ contains the latest estimate of α_{j1}, for $j = 1, 2, \ldots, p$. (Because of possible recoding all elements of A will be positive.)

9: C(IP) – REAL (KIND=nag_wp) array *Input/Output*

On entry: $\mathrm{C}(j)$ must be set equal to an initial estimate of α_{j0}. **In order to avoid divergence problems with the E-M algorithm you are strongly advised to set all the** $\mathbf{C}(j)$ **to** $\mathbf{0.0}$**.**

On exit: $\mathrm{C}(j)$ contains the latest estimate of α_{j0}, for $j = 1, 2, \ldots, p$.

10: IPRINT – INTEGER *Input*

On entry: the frequency with which the maximum likelihood search routine is to be monitored.

$\mathrm{IPRINT} > 0$
: The search is monitored once every IPRINT iterations, and when the number of quadrature points is increased, and again at the final solution point.

$\mathrm{IPRINT} = 0$
: The search is monitored once at the final point.

$\mathrm{IPRINT} < 0$
: The search is not monitored at all.

IPRINT should normally be set to a small positive number.

Suggested value: $\mathrm{IPRINT} = 1$.

11: CGETOL – REAL (KIND=nag_wp) *Input*

On entry: the accuracy to which the solution is required.

If CGETOL is set to 10^{-l} and on exit $\mathrm{IFAIL} = 0$ or 7, then all elements of the gradient vector will be smaller than 10^{-l} in absolute value. For most practical purposes the value 10^{-4} should suffice. You should be wary of setting CGETOL too small since the convergence criterion may then have become too strict for the machine to handle.

If CGETOL has been set to a value which is less than the square root of the ***machine precision***, ϵ, then G11SAF will use the value $\sqrt{\epsilon}$ instead.

12: MAXIT – INTEGER *Input*

On entry: the maximum number of iterations to be made in the maximum likelihood search. There will be an error exit (see Section 6) if the search routine has not converged in MAXIT iterations.

Suggested value: $\mathrm{MAXIT} = 1000$.

Constraint: $\mathrm{MAXIT} \geq 1$.

13: CHISQR – LOGICAL *Input*

On entry: if CHISQR is set equal to .TRUE., then a likelihood ratio statistic will be calculated (see CHI).

If CHISQR is set equal to .FALSE., no such statistic will be calculated.

14: ISHOW – INTEGER *Input*

On entry: indicates which of the following three quantities are to be printed before exit from the routine (given a valid parameter set):

(a) Table of maximum likelihood estimates and standard errors (as returned in the output arrays A, C, ALPHA, PIGAM and CM).

(b) Table of observed and expected first and second order margins (as returned in the output arrays EXPP and OBS).

(c) Table of observed and expected frequencies of score patterns along with theta scores (as returned in the output arrays IRL, EXF, Y, XL and IOB) and the likelihood ratio statistic (if required).

ISHOW $= 0$
None of the above are printed.

ISHOW $= 1$
(a) only is printed.

ISHOW $= 2$
(b) only is printed.

ISHOW $= 3$
(c) only is printed.

ISHOW $= 4$
(a) and (b) are printed.

ISHOW $= 5$
(a) and (c) are printed.

ISHOW $= 6$
(b) and (c) are printed.

ISHOW $= 7$
(a), (b) and (c) are printed.

Constraint: $0 \leq$ ISHOW ≤ 7.

15: NITER – INTEGER *Output*

On exit: given a valid parameter set then NITER contains the number of iterations performed by the maximum likelihood search routine.

16: ALPHA(IP) – REAL (KIND=nag_wp) array *Output*

On exit: given a valid parameter set then ALPHA(j) contains the latest estimate of α_j. (Because of possible recoding all elements of ALPHA will be positive.)

17: PIGAM(IP) – REAL (KIND=nag_wp) array *Output*

On exit: given a valid parameter set then PIGAM(j) contains the latest estimate of either π_j if GPROB = .FALSE. (logit model) or γ_j if GPROB = .TRUE. (probit model).

18: CM(LDCM,$2 \times$ IP) – REAL (KIND=nag_wp) array *Output*

On exit: given a valid parameter set then the strict lower triangle of CM contains the correlation matrix of the parameter estimates held in ALPHA and PIGAM on exit. The diagonal elements of CM contain the standard errors. Thus:

$$\begin{aligned}
\mathrm{CM}(2 \times i - 1, 2 \times i - 1) &= \text{standard error } (\mathrm{ALPHA}(i)) \\
\mathrm{CM}(2 \times i, 2 \times i) &= \text{standard error } (\mathrm{PIGAM}(i)) \\
\mathrm{CM}(2 \times i, 2 \times i - 1) &= \text{correlation } (\mathrm{PIGAM}(i), \mathrm{ALPHA}(i)),
\end{aligned}$$

for $i = 1, 2, \ldots, p$;

$$\begin{aligned}
\mathrm{CM}(2 \times i - 1, 2 \times j - 1) &= \text{correlation } (\mathrm{ALPHA}(i), \mathrm{ALPHA}(j)) \\
\mathrm{CM}(2 \times i, 2 \times j) &= \text{correlation } (\mathrm{PIGAM}(i), \mathrm{PIGAM}(j)) \\
\mathrm{CM}(2 \times i - 1, 2 \times j) &= \text{correlation } (\mathrm{ALPHA}(i), \mathrm{PIGAM}(j)) \\
\mathrm{CM}(2 \times i, 2 \times j - 1) &= \text{correlation } (\mathrm{ALPHA}(j), \mathrm{PIGAM}(i)),
\end{aligned}$$

for $j = 1, 2, \ldots, i - 1$.

If the second derivative matrix cannot be computed then all the elements of CM are returned as zero.

19: LDCM – INTEGER *Input*

On entry: the first dimension of the array CM as declared in the (sub)program from which G11SAF is called.

Constraint: LDCM $\geq 2 \times$ IP.

20: G($2 \times$ IP) – REAL (KIND=nag_wp) array *Output*

On exit: given a valid parameter set then G contains the estimated gradient vector corresponding to the final point held in the arrays ALPHA and PIGAM. G$(2 \times j - 1)$ contains the derivative of the log-likelihood with respect to ALPHA(j), for $j = 1, 2, \ldots, p$. G$(2 \times j)$ contains the derivative of the log-likelihood with respect to PIGAM(j), for $j = 1, 2, \ldots, p$.

21: EXPP(LDEXPP,IP) – REAL (KIND=nag_wp) array *Output*

On exit: given a valid parameter set then EXPP(i, j) contains the expected percentage of individuals in the sample who respond positively to items i and j ($j \leq i$), corresponding to the final point held in the arrays ALPHA and PIGAM.

22: LDEXPP – INTEGER *Input*

On entry: the first dimension of the arrays EXPP and OBS as declared in the (sub)program from which G11SAF is called.

Constraint: LDEXPP $\geq$ IP.

23: OBS(LDEXPP,IP) – REAL (KIND=nag_wp) array *Output*

On exit: given a valid parameter set then OBS(i, j) contains the observed percentage of individuals in the sample who responded positively to items i and j ($j \leq i$).

24: EXF(NS) – REAL (KIND=nag_wp) array *Output*

On exit: given a valid parameter set then EXF(l) contains the expected frequency of the lth score pattern (lth row of X), corresponding to the final point held in the arrays ALPHA and PIGAM.

25: Y(NS) – REAL (KIND=nag_wp) array *Output*

On exit: given a valid parameter set then Y(l) contains the estimated theta score corresponding to the lth row of X, for the final point held in the arrays ALPHA and PIGAM.

26: XL(NS) – REAL (KIND=nag_wp) array *Workspace*

If GPROB has been set equal to .FALSE. (logit model) then on exit, given a valid parameter set, XL(l) contains the estimated component score corresponding to the lth row of X for the final point held in the arrays ALPHA and PIGAM.

If GPROB is set equal to .TRUE. (probit model), this array is not used.

27: IOB(NS) – INTEGER array *Output*

On exit: given a valid parameter set then IOB(l) contains the number of items in the lth row of X for which the response was positive (.TRUE.).

28: RLOGL – REAL (KIND=nag_wp) *Output*

On exit: given a valid parameter set then RLOGL contains the value of the log-likelihood kernel corresponding to the final point held in the arrays ALPHA and PIGAM, namely

$$\sum_{l=1}^{s} \mathrm{IRL}(l) \times \log(\mathrm{EXF}(l)/n).$$

29: CHI – REAL (KIND=nag_wp) *Output*

On exit: if CHISQR was set equal to .TRUE. on entry, then given a valid parameter set, CHI will contain the value of the likelihood ratio statistic corresponding to the final parameter estimates held in the arrays ALPHA and PIGAM, namely

$$2 \times \sum_{l=1}^{s} \mathrm{IRL}(l) \times \log(\mathrm{EXF}(l)/\mathrm{IRL}(l)).$$

The summation is over those elements of IRL which are positive. If $\mathrm{EXF}(l)$ is less than 5.0, then adjacent score patterns are pooled (the score patterns in X being first put in order of increasing theta score).

If CHISQR has been set equal to .FALSE., then CHI is not used.

30: IDF – INTEGER *Output*

On exit: if CHISQR was set equal to .TRUE. on entry, then given a valid parameter set, IDF will contain the degrees of freedom associated with the likelihood ratio statistic, CHI.

$$\begin{array}{ll} \mathrm{IDF} = s_0 - 2 \times p & \text{if } s_0 < 2^p; \\ \mathrm{IDF} = s_0 - 2 \times p - 1 & \text{if } s_0 = 2^p, \end{array}$$

where s_0 denotes the number of terms summed to calculate CHI ($s_0 = s$ only if there is no pooling).

If CHISQR has been set equal to .FALSE., then IDF is not used.

31: SIGLEV – REAL (KIND=nag_wp) *Output*

On exit: if CHISQR was set equal to .TRUE. on entry, then given a valid parameter set, SIGLEV will contain the significance level of CHI based on IDF degrees of freedom. If IDF is zero or negative then SIGLEV is set to zero.

If CHISQR was set equal to .FALSE., then SIGLEV is not used.

32: W(LW) – REAL (KIND=nag_wp) array *Workspace*
33: LW – INTEGER *Input*

On entry: the dimension of the array W as declared in the (sub)program from which G11SAF is called.

Constraint: $\mathrm{LW} \geq 4 \times \mathrm{IP} \times (\mathrm{IP} + 16)$.

34: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. If you are unfamiliar with this parameter you should refer to Section 3.3 in the Essential Introduction for details.

For environments where it might be inappropriate to halt program execution when an error is detected, the value -1 or 1 is recommended. If the output of error messages is undesirable, then the value 1 is recommended. Otherwise, because for this routine the values of the output parameters may be useful even if $\mathrm{IFAIL} \neq 0$ on exit, the recommended value is -1. **When the value -1 or 1 is used it is essential to test the value of IFAIL on exit.**

On exit: $\mathrm{IFAIL} = 0$ unless the routine detects an error or a warning has been flagged (see Section 6).

6 Error Indicators and Warnings

If on entry IFAIL $= 0$ or -1, explanatory error messages are output on the current error message unit (as defined by X04AAF).

Note: G11SAF may return useful information for one or more of the following detected errors or warnings.

Errors or warnings detected by the routine:

IFAIL $= 1$

On entry, IP < 3,
or N < 7,
or NS $\leq 2 \times$ IP,
or NS $>$ N,
or NS $> 2^{\mathrm{IP}}$,
or two or more rows of X are identical,
or LDX $<$ NS,
or $\sum_{l=1}^{\mathrm{NS}} \mathrm{IRL}(l) \neq \mathrm{N}$,
or at least one of IRL$(l) < 0$, for $l = 1, 2, \ldots, \mathrm{NS}$,
or MAXIT < 1,
or ISHOW < 0,
or ISHOW > 7,
or LDCM $<$ IP $+$ IP,
or LDEXPP $<$ IP,
or LW $< 4 \times \mathrm{IP} \times (\mathrm{IP} + 16)$.

IFAIL $= 2$

For at least one of the IP items the responses are all at the same level. To proceed, you must delete this item from the dataset.

IFAIL $= 3$

There have been MAXIT iterations of the maximum likelihood search routine. If steady increases in the log-likelihood kernel were monitored up to the point where this exit occurred, then the exit probably occurred simply because MAXIT was set too small, so the calculations should be restarted from the final point held in A and C. This type of exit may also indicate that there is no maximum to the likelihood surface.

IFAIL $= 4$

One of the elements of A has exceeded 10.0 in absolute value (see Section 8.3). If steady increases in the log-likelihood kernel were monitored up to the point where this exit occurred then this exit may indicate that there is no maximum to the likelihood surface. You are advised to restart the calculations from a different point to see whether the E-M algorithm moves off in the same direction.

IFAIL $= 5$

This indicates a failure in F01ABF which is used to invert the second derivative matrix for calculating the variance-covariance matrix of parameter estimates. It was also found that MAXIT iterations had been performed (see IFAIL $= 3$). The elements of CM will then have been set to zero on exit. You are advised to restart the calculations with a larger value for MAXIT.

IFAIL $= 6$

This indicates a failure in F01ABF which is used to invert the second derivative matrix for calculating the variance-covariance matrix of parameter estimates. It was also found that one of the elements of A had exceeded 10.0 in absolute value (see IFAIL $= 4$). The elements of CM will have then been set to zero on exit. You are advised to restart the calculations from a different point to see whether the E-M algorithm moves off in the same direction.

IFAIL = 7

If CHISQR was set equal to .TRUE. on entry, so that a likelihood ratio statistic was calculated, then IFAIL = 7 merely indicates that the value of IDF on exit is ≤ 0, i.e., the χ^2 statistic is meaningless. In this case SIGLEV is returned as zero. **All other output information should be correct, i.e., can be treated as if IFAIL = 0 on exit.**

7 Accuracy

On exit from G11SAF if IFAIL = 0 or 7 then the following condition will be satisfied:

$$\max_{1 \leq i \leq 2 \times p} \{|\mathrm{G}(i)|\} < \mathrm{CGETOL}.$$

If IFAIL = 3 or 5 on exit (i.e., MAXIT iterations have been performed but the above condition does not hold), then the elements in A, C, ALPHA and PIGAM may still be good approximations to the maximum likelihood estimates. You are advised to inspect the elements of G to see whether this is confirmed.

8 Further Comments

8.1 Timing

The number of iterations required in the maximum likelihood search depends upon the number of observed variables, p, and the distance of the starting point you supplied from the solution. The number of multiplications and divisions performed in an iteration is proportional to p.

8.2 Initial Estimates

You are strongly advised to use the recommended starting values for the elements of A and C. Divergence may result from values you supplied even if they are very close to the solution. Divergence may also occur when an item has nearly all its responses at one level.

8.3 Heywood Cases

As in normal factor analysis, Heywood cases can often occur, particularly when p is small and n not very big. To overcome this difficulty the maximum likelihood search routine is terminated when the absolute value of one of the α_{j1} exceeds 10.0. You have the option of deciding whether to exit from G11SAF (by setting IFAIL = 0 on entry) or to permit G11SAF to proceed onwards as if it had exited normally from the maximum likelihood search routine (setting IFAIL = -1 on entry). The elements in A, C, ALPHA and PIGAM may still be good approximations to the maximum likelihood estimates. You are advised to inspect the elements G to see whether this is confirmed.

8.4 Goodness of Fit Statistic

When n is not very large compared to s a goodness-of-fit statistic should not be calculated as many of the expected frequencies will then be less than 5.

8.5 First and Second Order Margins

The observed and expected **percentages** of sample members responding to individual and pairs of items held in the arrays OBS and EXPP on exit can be converted to observed and expected **numbers** by multiplying all elements of these two arrays by $n/100.0$.

9 Example

A program to fit the logit latent variable model to the following data:

Index	Score Pattern	Observed Frequency
1	0000	154
2	1000	11
3	0001	42
4	0100	49
5	1001	2
6	1100	10
7	0101	27
8	0010	84
9	1101	10
10	1010	25
11	0011	75
12	0110	129
13	1011	30
14	1110	50
15	0111	181
16	1111	121
Total		1000

9.1 Program Text

```
    Program g11safe

!     G11SAF Example Program Text

!     Mark 24 Release. NAG Copyright 2012.

!     .. Use Statements ..
      Use nag_library, Only: g11saf, nag_wp, x04abf
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter               :: iset = 1, nin = 5, nout = 6
!     .. Local Scalars ..
      Real (Kind=nag_wp)               :: cgetol, chi, rlogl, siglev
      Integer                          :: i, idf, ifail, ip, ip2, iprint,      &
                                          ishow, ldcm, ldexpp, ldx, lw, maxit, &
                                          n, nadv, niter, ns
      Logical                          :: chisqr, gprob
!     .. Local Arrays ..
      Real (Kind=nag_wp), Allocatable  :: a(:), alpha(:), c(:), cm(:,:),       &
                                          exf(:), expp(:,:), g(:), obs(:,:),   &
                                          pigam(:), w(:), xl(:), y(:)
      Integer, Allocatable             :: iob(:), irl(:)
      Logical, Allocatable             :: x(:,:)
!     .. Executable Statements ..
      Write (nout,*) 'G11SAF Example Program Results'
      Write (nout,*)
      Flush (nout)

!     Skip heading in data file
      Read (nin,*)

!     Read in problem size
      Read (nin,*) ip, n, ns

!     Read in control parameters
      Read (nin,*) gprob, iprint, cgetol, maxit, chisqr, ishow

!     Set the advisory channel to NOUT for monitoring information
      If (iprint>0 .Or. ishow/=0) Then
        nadv = nout
        Call x04abf(iset,nadv)
      End If
```

```
      ip2 = 2*ip
      ldx = ns
      ldcm = ip2
      ldexpp = ip
      lw = 4*ip*(ip+16)
      Allocate (x(ldx,ip),irl(ns),a(ip),c(ip),alpha(ip),pigam(ip), &
        cm(ldcm,ip2),g(ip2),expp(ldexpp,ip),obs(ldexpp,ip),exf(ns),y(ns), &
        xl(ns),iob(ns),w(lw))

!     Read in data
      Read (nin,*)(irl(i),x(i,1:ip),i=1,ns)

!     Read in intial values
      Read (nin,*) a(1:ip)
      Read (nin,*) c(1:ip)

!     Fit a latent variable model
      ifail = 0
      Call g11saf(ip,n,gprob,ns,x,ldx,irl,a,c,iprint,cgetol,maxit,chisqr, &
        ishow,niter,alpha,pigam,cm,ldcm,g,expp,ldexpp,obs,exf,y,xl,iob,rlogl, &
        chi,idf,siglev,w,lw,ifail)

    End Program g11safe
```

9.2 Program Data

```
G11SAF Example Program Data
 4  1000  16              :: IP,N,NS
 F  -1  1.0E-4 1000  T  7 :: GPROB,IPRINT,CGETOL,MAXIT,CHISQR,ISHOW
154 F F F F
 11 T F F F
 42 F F F T
 49 F T F F
  2 T F F T
 10 T T F F
 27 F T F T
 84 F F T F
 10 T T F T
 25 T F T F
 75 F F T T
129 F T T F
 30 T F T T
 50 T T T F
181 F T T T
121 T T T T               :: End of IRL,X
0.5  0.5  0.5  0.5        :: A (initial values)
0.0  0.0  0.0  0.0        :: C (initial values)
```

9.3 Program Results

```
 G11SAF Example Program Results


 LOG LIKELIHOOD KERNEL ON EXIT =   -0.24039E+04

 MAXIMUM LIKELIHOOD ESTIMATES OF ITEM PARAMETERS ARE AS FOLLOWS
 --------------------------------------------------------------

 ITEM J     ALPHA(J)          S.E.       ALPHA(J,0)          PI(J)            S.E.
 ------     --------          ----       ----------          -----            ----

    1         1.045          0.148          -1.276           0.218           0.017

    2         1.409          0.179           0.424           0.604           0.022

    3         2.659          0.525           1.615           0.834           0.036

    4         1.122          0.140          -0.062           0.485           0.020
```

```
EXPECTED (AND OBSERVED) PERCENTAGE OF CASES PRODUCING
POSITIVE RESPONSES FOR INDIVIDUAL AND PAIRS OF ITEMS
-----------------------------------------------------

                                        ITEM
                                        ----

ITEM      1     2     3     4
----      -     -     -     -
  1     25.9
       (25.9)
  2     19.1  57.7
       (19.1)(57.7)
  3     22.5  48.0  69.4
       (22.6)(48.1)(69.5)
  4     16.4  33.9  40.6  48.8
       (16.3)(33.9)(40.7)(48.8)


OBSERVED   EXPECTED   THETA   COMPONENT   RAW    SCORE
FREQUENCY  FREQUENCY  SCORE   SCORE       SCORE  PATTERN
---------  ---------  -----   ---------   -----  -------

   154      147.061  -1.273     0.000       0    FFFF
    11       13.444  -0.873     1.045       1    TFFF
    42       42.420  -0.846     1.122       1    FFFT
    49       54.818  -0.747     1.409       1    FTFF
     2        5.886  -0.494     2.167       2    TFFT
    10        8.410  -0.399     2.455       2    TTFF
    27       27.511  -0.374     2.531       2    FTFT
    84       92.062  -0.332     2.659       1    FFTF
    10        6.237  -0.019     3.577       3    TTFT
    25       21.847   0.027     3.705       2    TFTF
    75       73.835   0.055     3.781       2    FFTT
   129      123.766   0.162     4.069       2    FTTF
    30       26.899   0.466     4.826       3    TFTT
    50       50.881   0.591     5.114       3    TTTF
   181      179.564   0.626     5.190       3    FTTT
   121      125.360   1.144     6.236       4    TTTT
---------  ---------
  1000     1000.000


LIKELIHOOD RATIO GOODNESS OF FIT STATISTIC =      9.027
                        SIGNIFICANCE LEVEL =      0.251
(BASED ON   7 DEGREES OF FREEDOM)

Value of IFAIL parameter on exit from G11SAF =     0
```