A concise representation of Goodness Of Fit (GOF)

by L. Lindegren (20 December 1994)

The goodness of fit of a given set of data \((y_1, y_2, \ldots, y_n)\) to a fitted model (described by the parameters \(p\)) is typically given by a chi-square type of statistic,

\[
\chi^2 = \sum_{i=1}^{n} \left( \frac{y_i - f_i}{\sigma_i} \right)^2. \tag{1}
\]

Here \(\sigma_i\) is the standard error of the \(i\)th data point and \(f_i(p)\) the value obtained from the fitted model. Usually the fitting procedure is more or less equivalent to varying \(p\) until \(\chi^2\) is minimised.

Provided the model is adequate and the standard errors correctly estimated, one expects \(\chi^2\) to follow the chi-square distribution with \(\nu = n - \text{dim}(p)\) degrees of freedom. If the model is inadequate, \(\chi^2\) will tend to be larger than can reasonably be expected of a chi-square random variable with \(\nu\) degrees of freedom.

Since the chi-square distribution depends strongly on \(\nu\), one must in general specify both \(\chi^2\) and \(\nu\). That two numbers are needed to represent the GOF is often an inconvenience, especially when \(\nu\) varies between different data sets. In the Hipparcos Catalogue, for instance, the number of degrees of freedom per star may vary from a few up to over 100. A visual inspection of the \(\chi^2\) values for different stars will thus require a rather difficult mental process of taking into account the varying \(\nu\) as well. Similarly, when plotting a histogram of the GOF statistic, one is hampered by the fact that each \(\nu\) corresponds to a different expected distribution.

This inconvenience can be eliminated by mapping \((\chi^2, \nu)\) into a single quantity. A natural choice is the probability that a chi-square variable with \(\nu\) degrees of freedom exceeds the observed value \(\chi^2\), or:

\[
Q(\chi^2, \nu) = \frac{1}{2} \left[ 2^{\nu/2} \Gamma \left( \frac{\nu}{2} \right) \right]^{-1} \int_{\chi^2}^{\infty} t^{\nu/2-1} e^{-t/2} \text{dt}. \tag{2}
\]

Adequate fits should give values of \(Q\) uniformly distributed between 0 and 1, with a mean and median value of \(\frac{1}{2}\).

Questionable fits will tend to give small \(Q\) values, clustering closer and closer to 0 for increasingly bad fits \((Q = 0.01, 0.001, 0.0001, \text{etc})\). Accommodating really bad fits within this representation will require that \(Q\) is given with many decimals, which is again inconvenient. A solution may be to apply a non-linear monotonic transformation of \(Q\), stretching the probability scale for small values. A reasonable ad hoc transformation might be \(-\ln Q\), giving the value 0 for a perfect fit, \(\ln 2 \approx 0.693\) for the 'median' good fit, and large positive values for bad fits. Although this transformation definitely has some appeal, a different one is in fact proposed below.
Scientists are used to think of uncertainties in terms of the normal or gaussian probability law. This suggests that transforming \( Q \) into a normal deviate should produce a GOF statistic that is quite readily comprehended. Thus, a value \( z \) may be determined such that \( Q \) is the probability that a random variable in \( N(0,1) \) exceeds \( z \):

\[
Q = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} \, dt.
\]

(3)

Acceptable, ‘good’ fits will produce \( z \) values following the \( N(0,1) \) distribution, centred on 0, with 68% of the values between \(-1\) and \(+1\), and with very few points beyond \( \pm 3 \). Large values of \( z \) (e.g., \( z > +3 \)) thus indicate model misfits, while \( z < -3 \) (say) on the contrary indicates an abnormally good fit to the data.

The degree of non-linearity implied by this transformation of \( Q \) can be appreciated by considering that \( z = 3, 6, \) and 9 correspond to, respectively, \( Q \approx 10^{-3}, 10^{-9}, \) and \( 10^{-18} \).

The accurate evaluation of \( x(\chi^2, \nu) \) according to Eqs. (2)–(3) is rather difficult because of the analytically intractable integrals. Fortunately, analytical approximations exist that are both simple and sufficiently accurate for the present purpose. The proposed approximation is the well-known ‘cube-root transformation’,

\[
x(\chi^2, \nu) \approx x^*(\chi^2, \nu) \equiv \left( \frac{9\nu}{2} \right)^{1/3} \left[ \left( \frac{\chi^2}{\nu} \right)^{1/3} + \frac{2}{9\nu} - 1 \right].
\]

(4)

This formula is usually quoted as valid for \( \nu > 30 \) (e.g., Handbook of Mathematical Functions, M. Abramowitz and I.A. Stegun, Dover, 1970), but it is in fact useful even for much smaller \( \nu \), as will be shown below. The formula is readily inverted to give \( \chi^2 \) as a function of \( x^* \) and \( \nu \):

\[
\chi^2 = \nu \left[ \left( \frac{2}{9\nu} \right)^{1/2} x^* + 1 - \frac{2}{9\nu} \right]^3.
\]

(5)

The accuracy of (4) has been numerically evaluated by means of routines from the NAG Fortran Library. For given \((x, \nu)\), \( Q \) was computed from Eq. (3) by means of the NAG routine S15ABF, then \( \chi^2 \) from Eq. (2) by means of G01CCF. Finally \( x^* \) was computed by a direct implementation of Eq. (4), and hence the error \( x^* - x \). Table 1 summarises the results of a fairly systematic mapping of this error as a function of \( x \) and \( \nu \). For any \( \nu > 0 \), the error goes through zero at three different \( x \) values that are always confined to the interval \([-1.7, +2.6] \). The error is positive for \( x < -1.7 \) and negative for \( x > +2.6 \). The approximate intervals in which the error is absolutely less than 0.1 and 0.01 are given in the table. Also given are the average absolute errors in the two intervals \(-1 < x < +1, \) which includes the majority (68%) of the well-behaved points, and \(+2 < x < +4, \) which is typically where one may want to set a limit to screen out bad fits.
The approximation in Eq. (4) is remarkably accurate even for small \( \nu \). In the most interesting region of \(-1 < x < +3.5\) the error is in fact always less than 0.1 units in \( x \). Usually, accepted fits have \( \nu \gtrsim 5 \) and \( x \lesssim +2.5 \), for which the approximation is even accurate to 0.01 units. We conclude that the formula is fully sufficient to provide a useful measure of the goodness of fit. The possible disadvantages of its slight inaccuracy are probably far outweighed by the practical advantage of having a simple, invertible analytical expression that can be coded on a single line.

The GOF represented by the 'gaussianised' \( \chi^2 \) variable \( x^* \) should not be given to more than two decimal places, and probably one is sufficient in practice. Moreover, given that \( x = 10 \) corresponds to \( Q = 10^{-23} \), it should be acceptable to enforce \( |x^*| < 10 \). Thus, not more than four characters are needed for \( x^* \), including the sign and decimal point.

Table 1. Summary of the errors of the approximation (4), expressed in the absolute deviation \( |x^* - x| \) as a function of \( x \) and \( \nu \). The approximate ranges in which \( |x^* - x| \) is smaller that 0.1 and 0.01 are given, as well as the average \( |x^* - x| \) in two intervals of \( x \).

| DOF | Error < 0.1 Error < 0.01 Average error Average error |
|-----|------------------------|------------------------|
| \( \nu \) | \( x_{\text{min}} \) \( x_{\text{max}} \) \( x_{\text{min}} \) \( x_{\text{max}} \) \(-1 < x < +1\) \(+2 < x < +4\) |
| 1   | -1.1       +3.6 | -        | 0.016  | 0.057  |
| 2   | -1.7       +3.7 | +0.1    | +1.0   | 0.009  | 0.047  |
| 3   | -2.0       +4.0 | -0.2    | +2.5   | 0.007  | 0.040  |
| 5   | -2.5       +4.3 | -1.5    | +2.5   | 0.005  | 0.030  |
| 7   | -2.9       +4.6 | -1.7    | +2.6   | 0.003  | 0.025  |
| 10  | -3.3       +5.0 | -1.9    | +2.7   | 0.002  | 0.020  |
| 20  | -4.2       +5.9 | -2.3    | +3.0   | 0.001  | 0.012  |
| 50  | -5.9       +7.6 | -3.1    | +3.6   | 0.000  | 0.006  |