Derivation of various types of intervals from a `selm` object

Adelchi Azzalini
adelchi.azzalini@unipd.it

29th June 2016

This document is connected to the R *package* sn *and it represents a technical complement to its manual. The main purpose of the document is to describe how certain interval-type objects are derived from an object generated by a call to the function* selm*. The terminology adopted here is in agreement with the one of the package documentation; hence refer to the package manual for any term or concept not explained in this document. If you are interested in using the package but not so much in its internal working, this document may be not of key concern to you.*

Preliminaries

It is assumed that the function `selm` has been called to fit a model with a univariate response variable and its execution has been successfully completed with a solution *not* on the boundary of the parameter space. Such a call generates an object of `selm-class`, denoted `object` in the following. Since version 1.4-0 of the package, one can apply methods `confint` and `predict` to such objects. The main role of this document is to describe the computing procedures underlying these methods.

In case the `selm` fitting process has landed on the boundary of the parameter space, consider refitting the model with estimation `method="MPLE"` or switching from `family="ST"` to `family="SN"`, depending on which side of the parameter space has been hit. This should lead to a new fit in the interior of the parameter space.

Objects of `mselm-class`, generated by a call to `selm` with a multivariate response, are not handled by `confint` and `predict`.

References to supporting theoretical results are made via the monograph by Azzalini & Capitanio (2014), later denoted 'the SN book'. This choice is made merely for simplicity, to avoid a long list of individually focused sources, some of which are quite technical. The book points to the original publications of the pertaining results where full details are available.

Confidence intervals via `"confint"`

This section describes the computing schemes involved to obtain the confidence intervals generated by a call of type `confint(object, ...)`.

In the majority of cases, the distribution of the estimates is approximated by a normal distribution with covariance matrix given by the inverse matrix of the observed information matrix. This is possibly adjusted by a penalty term if the option `method="MPLE"` had been adopted at the fitting stage; in case the penalty function was user-defined, that function must still be accessible. This normal-based approximation applies for all families (`"SN"`, `"ST"`, "SC") and all forms of parameterization (CP, DP, pseudo-CP, if admissible for the pertaining family), with a few exceptions described below.

The special treatment of some cases reflects the qualitatively different nature of the log-likelihood asymptotics near $\alpha = 0$ for the SN family and the other families, if the DP parameterization is adopted; see § 3.1.1, 4.3.3 and 6.3.1 of the SN book.

In the SN case, a confidence interval for α is obtained from the profile (penalized) log-likelihood function $\ell_p(\alpha)$. More specifically, if q denotes the upper quantile of the χ^2_1 distribution at the chosen confidence level, two solutions of the equation for α

$$2\{\ell_p(\hat\alpha) - \ell_p(\alpha)\} = q \tag{1}$$

are searched for, one on each side of the MLE, $\hat\alpha$. Computation of ℓ_p is accomplished by using the function `profile.selm`.

In principle, results about normal asymptotic distribution of the MLE, leading to confidence intervals derived from (1), hold for the CP parameterization; see § 3.1.4–3.1.6 of the SN book. However, equivariance of the MLE and of likelihood-based intervals (1) allow to work equivalently with the DP component, taking into account the one-to-one correspondence between α and γ_1. This has the numerical advantage of avoiding boundary values in the search of the solution and it avoids repeated transformation between CP and DP values, given the way of working of `profile.selm`. When the required parameter type is CP, the confidence interval for α is then mapped on the γ_1 scale at the end of the process.

For the scale parameter (ω in the DP set, σ in the CP set, $\tilde\sigma$ in the CP set), a normal approximation to the asymptotic distribution formally holds. However, to avoid intervals which include negative values, the normal approximation is applied to the logarithmic transform of the parameter. This produces a confidence interval of the real axis, which is then back-transformed into the original parameter space, $\mathbb{R}^+$. A similar device is applied to the tail-weight parameter of the ST family, that is, ν, γ_2 or $\tilde\gamma_2$, depending on the adopted parameter set.

For the DP parameterization of a SN family with a non-significant slant parameter, no confidence interval is constructed for the location parameter ξ in case of a simple sample or, analogously, for the intercept term of the β parameters in a regression model. The decision that the slant is not significant is taken by examining whether the confidence interval for α (or equivalently for γ_1), as derived above, includes the point 0 or not.

Prediction intervals via `"predict.selm"`

Interval type `"confidence"`

The key fact here is normality of the asymptotic distribution of $x_i^T\hat\beta$, where x_i denotes a vector of covariates supplied to `predict.selm` and $\hat\beta$ are estimates of the regression parameters, β, obtained either from MLE or MPLE estimation. This asymptotic normality holds in all cases but one, analogously to what has already been discussed in the previous section.

The one exception occurs with the SN family using the DP parameterization when α is near zero. Although non-normality of $\hat\beta$ pertains only to the intercept term, it affects the overall distribution of $x_i^T\hat\beta$. Therefore a confidence interval is produced only if $\hat\alpha$ is significantly non-zero.

Interval type "prediction"

In all pertaining cases, we are working with a regression model which, for a generic value of the response variable, y, can be written equivalently in either of the two forms

$$y = x^\top \beta^{\text{DP}} + \varepsilon \tag{2}$$
$$= x^\top \beta^{\text{CP}} + (\varepsilon - \mu_\varepsilon) \tag{3}$$

where $\beta^{\text{DP}}, \beta^{\text{CP}}$ denote the vectors of regression parameters in the DP and CP variant, the first element of the covariate vector x is 1 and ε has either SN or ST distribution with null DP-location. The SC case is included within the ST family. For brevity, we not explicitly write down an analogous expression for pseudo-CP, but this case is incorporated in the development below, as it will be clear in a moment. For mathematical details, see Sections 3.1.4 and 4.3.4 of the SN book.

Equivalence of expressions (2) and (3) shows the equivalence of the prediction intervals using either of DP or CP. Thanks to this property, we can base our construction on the form which is more convenient for the case under consideration.

Once estimates of the parameters have been obtained from a set of data, a 'prediction' interval for new observations y^* involves computation of the distribution of

$$y^* = x^\top \hat\beta^{\text{DP}} + \varepsilon^* \tag{4}$$
$$= x^\top \hat\beta^{\text{CP}} + (\varepsilon^* - \hat\mu_\varepsilon) \tag{5}$$

where ε^* has the same distribution of ε above but it is independent from the data used for estimation, hence from $\hat\beta^{\text{DP}}$ and $\hat\beta^{\text{DP}}$.

Consider first the case when `object` refers to a fit with `family="SN"`. Two approximations are introduced: (a) the distribution of $\hat\beta^{\text{CP}}$, and consequently the one of $x^\top \hat\beta^{\text{CP}}$ is regarded as asymptotically normal, leading to the approximation

$$x^\top \hat\beta^{\text{CP}} \mathrel{\dot\sim} \mathrm{N}\big(x^\top \beta^{\text{CP}}, x^\top (\mathscr{I}^{\text{CP}})^{-1} x\big), \tag{6}$$

where $\mathscr{I}^{\text{CP}}$ denotes the CP observed (penalized) information matrix; (b) μ_ε and the other parameters of ε^* are treated as known, ignoring estimation error; this simplification is commonly in use. Therefore (5) represents the sum of $x^\top \hat\beta^{\text{CP}}$, with approximate distribution (6), and a SN variate $\varepsilon - \mu_\varepsilon$. The convolution distribution is still of SN type by Proposition 2.3 of the SN book, leading readily to the distribution of (5).

When `object` refers to a fit with `family="ST"` (or its special case `"SC"`), there is the complication that the convolution distribution of a Normal and an independent ST variable is not known. An additional consideration is that μ_ε does not exist when $\nu \le 1$; clearly this includes the case with `family="SC"`. However, there is the advantage that ST case is free from the problem of non-standard asymptotics of the MLE in a neighbourhood of $\alpha = 0$. This fact allows us to base our construction on (2), sidestepping the issue of μ_ε.

The distribution of $x^\top \hat\beta^{\text{DP}}$ distribution can be approximated similarly to the expression in (6), with the obvious substitution of β^{CP} and $\mathscr{I}^{\text{CP}}$ by the corresponding DP quantities, β^{DP} and $\mathscr{I}^{\text{DP}}$. Since no result analogous to Proposition 2.3 is currently available for ST variates, the distribution of (4) requires an approximation. Apart from the constant term $x^\top \beta^{\text{DP}}$, which is estimated by $x^\top \hat\beta^{\text{DP}}$, the issue now is to approximate the distribution of

$$V = U + \varepsilon^*$$

where

$$U \sim \mathrm{N}(0, \sigma_U^2), \quad \text{with } \sigma_U^2 = x^\top (\mathscr{I}^{\text{DP}})^{-1} x,$$

is independent of ε^*.

Consider first the case with $\nu > 4$, so that the first four cumulants of ε^* exist; denote them $\kappa_1, \ldots, \kappa_4$. Hence the matching cumulants of V are

$$\kappa_1, \quad \kappa_2 + \sigma_U^2, \quad \kappa_3, \quad \kappa_4 .$$

If we re-express this fact in terms of CP components of ε^*, denoted $\mu, \sigma, \gamma_1, \gamma_2$, the analogous quantities for V are

$$\mu, \quad \sigma / r, \quad \gamma_1 r^3, \quad \gamma_2 r^4 \tag{7}$$

where

$$r^2 = \sigma^2 / (\sigma^2 + \sigma_U^2) .$$

The required approximation to the distribution of V is obtained by finding the member of the ST family having CP components equal to those in (7). The task of parameter matching is accomplished using the available facilities for converting back and forth between the DP and CP parameterization. Note however that this step involves the solutions of non-linear equations; this process can then beccome lengthy when the number of prediction points x is large.

The idea of approximating an unknown distribution by one with matching four moments (or cumulants, equivalently) has been employed repeatedly in the literature. Numerical supporting evidence for this sort of procedure has been provided for instance by Solomon & Stephens (1978), although using a different parametric class.

If the condition $\nu > 4$ fails, we replace the CP components $\mu, \sigma, \gamma_1, \gamma_2$ with their pseudo-CP counterparts, which exist for all values of ν, and proceed similarly. A brief description of pseudo-CP is given in § 4.3.4 of the SN book; a full account is given in the original paper of Arellano-Valle & Azzalini (2013). One could raise the objection that the pseudo-CP set does not formally represent a proper parameterization, so that in some cases there could be two members of the ST family with the same pseudo-CP's. However, this is not an issue in the present context where the target is merely to provide a numerical approximation and the potential existence of another approximation does not cause problems.

References

R. B. Arellano-Valle & A. Azzalini (2013). The centred parameterization and related quantities of the skew-t distribution. *J. Multiv. Anal.*, **113**, 73–90. Available online 12 June 2011.

A. Azzalini with the collaboration of A. Capitanio (2014). *The Skew-Normal and Related Families*. IMS monographs, Cambridge University Press. ISBN 9781107029279.

H. Solomon & M. A. Stephens (1978). Approximation to density functions using Pearson curves. *J. Am. Stat. Assoc.*, **73**, 153–160.