

Uncertainty in Non-Linear Regression

Jeremy Ash

June 13, 2018

Review

Once a non-linear regression (NLR) model is fit to a data set, standard errors for the estimates of model parameters or functions of these parameters, such as the regression curve, provide a measure of uncertainty. These standard errors can then be used to form confidence intervals. A few attempts have been made to determine exact confidence intervals for non linear regression models, but these are never used in practice because they are difficult to implement for all but the simplest models and their confidence regions are not elliptical and difficult to comprehend (Duncan 1978, @draper1974effects).

All of the “usual” methods for estimating these standard errors rely on some asymptotic argument for their validity. That is, as the sample size gets larger, the estimated standard error approaches the true standard error. Unfortunately, non-linear regression models are often fit with few observations and standard errors may either be over or under-estimated. This translates to 95% confidence/credible intervals (CIs) with coverage probabilities that are either above or below the nominal level.

This has been demonstrated in numerous simulation studies. Duncan (Duncan 1978) showed that when sample sizes are small, two of the most standard methods for approximating standard errors (linear approximation and jackknife) often under-estimate standard errors and produce confidence intervals with coverage probability far below the nominal level. In another simulation study, Simonoff and Tsai (Simonoff and Tsai 1986) examined jack-knife based estimators that have been adjusted to account for non-linearity, but found these methods have similar shortcomings. Peddada and Haseman (Peddada and Haseman 2005) showed that the standard methods for NLR CI estimation have poor coverage probabilities in the context of a non-linear dose response model similar to the one considered in this study.

In this study I consider four methods for the estimation of the standard errors of the model parameters and regression curves. The first two estimate standard errors in a frequentist parametric non-linear regression framework. The next method uses Bayesian non-linear regression. The final method uses smoothing splines to form a nonparametric estimate of the regression curve.

Parametric Non-Linear Regression

The most standard way of performing parametric non-linear regression is an adaptation of the Gauss-Markov model to a non-linear setting. Let $y \in \mathbb{R}^m$ be the response vector, $e \in \mathbb{R}^m$ a random error vector, $\beta \in \mathbb{R}^p$ the fixed but unobserved model parameters, and $X \in \mathbb{R}^{m \times p}$ the fixed design matrix. The Gauss Markov model is:

$$y = X\beta + e$$

Where $e \sim N(0, \sigma^2 I)$ and, consequently $E[y] = X\beta$. Non-linear regression maintains the assumptions of the Gauss-Markov model, but replaces $X\beta$ with a non-linear function of the predictors:

$$y_i = f(x_i) + e_i, \quad i = 1, \dots, n$$

Where $e_i \stackrel{iid}{\sim} N(0, \sigma^2)$ and, consequently $E[y_i] = f(x_i)$. In this study we assume the following parametric form of $f(x_i)$:

$$f(x) = a + \frac{b}{1 + e^{(-c-dx)}}$$

Where $\beta = (a, b, c, d)$ are the model parameters. The model parameters are estimated by finding the values of β that minimizes the error sum of squares (SSE):

$$LOSS(\beta) = \sum_{j=1}^m (y_{ij} - f(x_j))^2$$

Unlike linear regression, there is no closed form solution to this problem. Instead, an iterative convex optimization algorithm such as Gauss-Newton (GN) is used to approximate the solution. However, the loss function is not convex for the DR model (as is often the case in non-linear regression), so these algorithms are liable to be trapped in local optima if good initial estimates of β are not used. In this study, non-linear regression was performed for many data sets and any strategy for finding initial estimates that worked in all scenarios was challenging. GN found a non-singular gradient for some data sets. To remedy this, I used the Levenberg-Marquardt algorithm (LMA) (Moré 1978) implemented in the `minpack.lm` package (Elzhov et al. 2016). LMA interpolates between two convex optimization algorithms: GN and gradient descent. It is often more robust to bad initializations than GN, meaning that it will converge to global optima more often given bad initializations. The convergence criteria implemented in `minpack.lm` was almost always satisfied, but still there are no guarantees that a local optima was not obtained. This is important, as a poor estimate of the model parameter will result in poor estimates of standard errors (Bates and Watts 1988).

The model parameters and regression fit at any design point can be expressed as function of the parameters. Call the function of interest $g(\beta)$. Once the least squares fit was obtained, there is also no closed form expressions for the standard errors of $g(\hat{\beta})$. Two of the most common methods for approximating these standard errors are a linear approximation method and bootstrapping. Once these standard errors are estimated, approximate 95% confidence intervals can be obtained.

Linear approximation

The most typical way of estimating standard errors in parameteric linear regression is by linear approximation (LA) of the non-linear model function, $f(x_i)$. A Taylor expansion of $f(x_i)$ is used to derive an approximate linear function of the parameters: $f(x) \approx V\beta$. Then the usual formulas for the standard error in linear regression are used for $g(\hat{\beta})$, which can be approximated by a linear function of the parameters, $c^T\beta$.

$$SE(g(\hat{\beta})) \approx SE(c^T\hat{\beta}) \approx \hat{\sigma} \sqrt{c^T(V^TV)^{-1}c}$$

Where $\hat{\sigma}$ is estimated by the mean squared error of the residual with $m-4$ degrees of freedom: $LOSS(\beta)/(m-4)$. $(1 - \alpha)100\%$ confidence intervals can then be obtained by:

$$g(\hat{\beta}) \pm z_{\alpha/2} SE(g(\beta))$$

Where $z_{\alpha/2}$ is the upper $\alpha/2$ quartile of the standard normal distribution. Note that these confidence intervals assume that σ is known. It was actually not known in the simulation study, it was estimated. I should have used the upper $\alpha/2$ quartile of student's t-distribution with $n-4$ degrees of freedom, but I realized this too late to change the simulation.

The Taylor approximation works well when the sample size is large, as the higher order terms in the Taylor expansion go to zero, but this is not the case when the sample size is small. The accuracy of this asymptotic approximation depends in a complex way on the functional form of $f(x)$, the design points, the error distribution and the true values of β (Duncan 1978). The accuracy of the linear approximation is quantified by so called “curvature effects” which can be decomposed into two components: the *intrinsic effect* due to the shape of the response function and *parametric effect* due to the functional form of the parameters

(Bates and Watts 1988, @Peddada2005). Some jackknife estimators of standard errors attempt to correct for this curvature to improve their estimates (Simonoff and Tsai 1986).

Bootstrapping

Similar to the jackknife procedures mentioned previously, the bootstrap is a data resampling method that presents several advantages over the linear approximation method. First, the bootstrap excels at estimating the sampling distribution of estimators when deriving their analytical sampling distributions is tedious (Fox 2015). As in NLR, a closed form expression for standard errors is challenging, but it is easy to generate bootstrap samples numerically and use these samples to estimate the standard error. Importantly, bootstrapping does not require a linear approximation to obtain the standard error estimate. Also, the parametric assumption in NLR that the errors are iid normal is likely unreasonable as the sample size is small and the central limit theorem cannot be invoked. Bootstrapping is intrinsically a nonparametric method and makes no assumptions on the error distribution. Often, relying on the asymptotic results in small samples results in underestimation of standard errors and confidence intervals that are too tight. The bootstrap often provides more robust inference procedures for the standard error and interval estimation.

Denote the the empirical DR curve as the set of dose response pairs $(y_1, x_1), \dots, (y_m, x_m)$. If the true population, P , that the empirical DR curve was sampled from was known, then many DR curves (each of size m) could be sampled from this population. The standard error of the estimator $g(\hat{\beta})$ could be estimated by Monte Carlo sampling. That is, for samples $i = 1, \dots, B$, the NLR model would be fit to the data and $g(\hat{\beta})_i^*$ estimated. The standard deviation of the $g(\hat{\beta})^*$ Monte Carlo samples is a consistent estimator of $SE(g(\hat{\beta}))$. As is often the case in life, we do not have P , but we have something similar. We have an empirical population P_m , which is the empirical DR curve. Non-parametric bootstrapping (NB) relies on the fact that P_m is a consistent estimator ($P_m \rightarrow P$, as $m \rightarrow \infty$). Many dose response pairs of size m can then be sampled from P_m instead of P and estimation of $SE(g(\hat{\beta}))$ can proceed as in Monte Carlo sampling. Non-parametric confidence interval estimates can be obtained by the upper and lower $\alpha/2$ quartiles of the empirical distribution of the $g(\hat{\beta})^*$ bootstrap samples (Fox 2015). These are called percentile bootstrap CI estimators. The problem with this procedure is that it still relies on asymptotic argument for the validity its standard error and interval estimates. When m is small the accuracy of the approximation of P with P_m is suspect. Percentile bootstrap CI estimators are often too tight and have poor coverage when the sample size is small (Fox 2015).

An alternative to the non-parametric bootstrap is the parametric bootstrap (PB). The only difference between these two procedures is how the data is sampled for each bootstrap iteration. The parametric bootstrap assumes that the parametric form of the distribution the data was sampled from is known. In this simulation study, I am able to assume that the functional form of the DR curve is known, $f(x)$, and we have that $y_i \sim N(f(x_i), \sigma^2)$. The parameters of this distribution are estimated by the least squares method mentioned previously. This provides estimates for $f(x_i)$ and σ^2 : $\hat{f}(x_i)$ and $\hat{\sigma}^2$. For each design point: x_i , $i = 1, \dots, m$ a new y_i is sampled from the distribution $N(\hat{f}(x_i), \hat{\sigma}^2)$. Standard error and interval estimation then proceeds as in the non-parametric bootstrap. Algorithm 1 provides pseudocode for the procedure.

Unfortunately, the parametric bootstrap also relies on an asymptotic argument for the validity of its estimators. The estimators of parameters in the NLR model are only accurate at large sample sizes. This means that the distribution of y_i will not be specified correctly for small m . However, Lee (Man and Lee 2018) has shown that when the parametric model in PB is at least approximately accurate, the PB estimate may be more accurate than the NB estimate. They derive the asymptotic MSE for the PB and NB estimators and show that it is asymptotically slightly larger for NB. Therefore, PB is the asymptotically optimal choice. On these grounds I have selected the parametric bootstrap procedure for estimating the standard errors and CIs. Actually, it is still unclear whether PB or NB will provide better estimates and both should be examined. Also, bootstrapping the NLR residuals is another strategy frequently used for linear regression. For the sake of time, I have only examined the parametric bootstrap.

Algorithm 1 Non-Parametric Bootstrap Standard Estimation

```
1: for  $j = 1$  to  $B$  do
2:   for  $i = 1$  to  $m$  do
3:     sample:  $y_i^* \sim N(\hat{f}(x_i), \hat{\sigma}^2)$ 
4:   end for
5:   estimate:  $g(\hat{\beta})_j^*$  using NLR
6: end for
7:  $s^2 = \frac{1}{B} \sum_{j=1}^B (g(\hat{\beta})_j^*)^2 - (\frac{1}{B} \sum_{j=1}^B g(\hat{\beta})_j^*)^2$ 
8:  $s = \sqrt{s^2}$ 
```

Bayesian Non-Linear Regression

A Bayesian NLR model (BNLR) was also considered. The likelihood in the model was the same as the frequentist NLR model. That is, $y_i \sim N(f(x_i), \sigma^2)$. Sensible priors were generated for parameters (a, b, c, d) by considering a reparametization of the curve commonly referred to as the Hill slope:

$$f(x) = \theta_1 + \frac{\theta_2}{1 + e^{\frac{-d+\theta_3}{\theta_4}}}$$

where θ_1 is the lower bound, θ_2 is the upper bound, θ_3 is the 50% dose, and θ_4 is a shape parameter. Certain constraints needed to be accounted for: $\theta_2 > \theta_1 > 0, \theta_3 \in (0, 1)$, which is the range of the doses, and $\theta_4 > 0$. This suggested the following prior distributions:

$$\begin{aligned}\theta_1 &= \exp(\alpha_1) \\ \theta_2 &= \theta_1 + \exp(\alpha_2) \\ \theta_3 &= \exp(\alpha_3) / [1 + \exp(\alpha_3)] \\ \theta_4 &= \exp(\alpha_4)\end{aligned}$$

Where $\alpha_i \stackrel{iid}{\sim} N(0, 100)$. A non-informative prior was assumed for the error variance: $1/\sigma^2 \sim \text{Gamma}(.01, .01)$. This is identical to the model proposed by Dr. Reich **here**. Posterior distributions of the (a, b, c, d) parameters were obtained by sampling from the posterior of the θ parameters and performing the following transformations: $a = \theta_1$, $b = \theta_2$, $c = -\theta_3/\theta_4$, $d = 1/\theta_4$.

While the Bayesian method does not estimate the standard error of the parameters and regression fits, the model does provide an estimate of the standard deviation of each function of the model parameters given the data, $g(\beta)|x$, according to the posterior distribution. For large sample sizes, the Bayesian central limit theorem can be invoked and the standard deviations of $g(\beta)|x$ should agree with the standard errors of $g(\hat{\beta})$ estimated by the previously discussed frequentist methods. The 95% credible intervals of $g(\beta)|x$ should also agree with the 95% confidence intervals for $g(\beta)$. However, for small sample sizes these estimates may differ.

At small sample sizes, Bayesian methods can improve the MSE of an estimator over frequentist methods when informative priors are used for the model parameters and they are correctly specified (i.e., the prior probability is concentrated near the true parameter values). One method for obtaining these informative priors was proposed by Dr. Reich **here**. DR data sets often contain a large number of empirical DR curves, but there are only a few observations for each curve. If the parameters in the NLR models are treated as random effects, each drawn from the same population with a prior distribution specified in the Bayesian model, information can be borrowed across curves to provide more stable parameter estimates. This Bayesian hierarchical model approach was employed by Wilson et al (Wilson et al. 2013). Informative prior distributions centered on these parameter estimates may provide better standard error and interval estimators. In the current study, only a model with non-informative priors was considered, but the model could easily be extended in this way to further improve these estimators.

Smoothing Splines

A standard way of performing non-linear regression is to make no assumptions about the functional form of $f(x)$ and perform nonparametric regression. There are numerous methods for nonparametric regression (e.g., local linear regression, nearest neighbor analysis, gaussian process regression). In this study, the smoothing spline regression method (SS) implemented in the `mgcv` package was used (Wood 2017). Briefly, smoothing splines approximate $f(x)$ with a natural cubic spline with knots at each of the design points: $h(x)$ (James et al. 2013). To estimate $h(x)$ the error sum of squares is minimized as in parametric NLR, but a penalty term based on the second derivative of $h(x)$ prevents overfitting:

$$LOSS = \sum_{i=1}^m (y_i - h(x_i))^2 + \delta \int h''(t)^2 dt$$

Larger values of δ parameter constrain the estimate of the response function to be more “smooth”, reducing overfitting. `mgcv` uses cross validation to select on optimal value of δ .

A clear disadvantage of this method is that because it is nonparametric, standard errors for (a, b, c, d) cannot be estimated. However standard errors and CIs can be estimated for the response function at any design point. These estimators may be much more accurate when the form of the response function is misspecified. This will often be the case in real DR data, but in this simulation study the form of $f(x)$ is known, putting the other methods at an advantage. However, if this nonparametric approach has standard error and interval estimates that are comparable to the parametric NLR models, it is likely to outperform these methods when the functional form of $f(x)$ is misspecified.

To estimate standard errors, `mgcv`, uses a Bayesian view of the smoothing process. Let $W_{m \times l}$ be the design matrix for the natural cubic spline. $\beta_s \in \mathbb{R}^l$ is given a zero mean improper Gaussian prior distribution with precision matrix proportional to S_δ , where $S_\delta = (\delta \int h''(t)^2 dt) I_{l \times l}$ (Wood 2017). This simplification of the matrix S_δ occurs because we only have one predictor variable. The posterior distribution of the β_s can be shown to be:

$$\beta_s | y, \delta = N(\hat{\beta}_s, V_\beta)$$

where $V_\beta = (W^T W + S_\delta)^{-1} \sigma^2$. Then the estimate of the posterior standard deviation of $f(x_i)$ is the square root of the i th diagonal element of $X V_\beta X^T$. Call this $\sqrt{v_i}$. A $(1 - \alpha)100\%$ credible interval for $f(x_i)$ is:

$$\hat{f}(x_i) \pm z_{\alpha/2} \sqrt{v_i}$$

Marra and Wood (Marra and Wood 2011) have shown in an simulation study that these credible intervals have surprisingly good coverage probabilities even at sample sizes as small as 200. The complication when estimating confidence intervals for smoothing splines is the smoothing bias term. These Bayesian confidence intervals happen to correct for this bias term, resulting in excellent coverage properties.

#Simulation

Simulation of the data

DR data was simulated according to a model previously developed by Dr. Reich and described in further detail [here](#).

$$y_i = f(x_i) + e_i$$

Where y_i is the response measured at dose x_i for $i = 1, \dots, m$ observations. $f(x)$ is the function:

$$f(x) = a + \frac{b}{1 + e^{(-c-dx)}}$$

And $e_i \stackrel{iid}{\sim} N(0, \sigma^2)$. Where $\sigma^2 = .5$. In this logistic curve, a and b define the upper and lower bound. c is the intercept and d the slope in the logistic regression function.

Sensible values were generated for parameters (a, b, c, d) by considering a reparamtization of the curve commonly referred to as the Hill slope:

$$f(x) = \theta_1 + \frac{\theta_2}{1 + e^{\frac{-d+\theta_3}{\theta_4}}}$$

Constraints that needed to be accounted for were $\theta_2 > \theta_1 > 0, \theta_3 \in (0, 1)$, which is the range of the doses, and $\theta_4 > 0$. This suggested the distributions for the simulation of θ .

$$\begin{aligned}\theta_1 &= 0.1 + \alpha_1, & \alpha_1 &\sim Exp(25) \\ \theta_2 &= \theta_1 + \alpha_2, & \alpha_2 &\sim Gamma(10, 10) \\ \theta_3 &\sim Beta(5, 3) \\ \theta_4 &\sim Gamma(5, 100)\end{aligned}$$

Theta values were simulated from these distributions, and the following transformations provided (a, b, c, d): $a = \theta_1, b = \theta_2, c = -\theta_3/\theta_4, d = 1/\theta_4$.

Throughout this report, one set of parameters were simulated and selected to represent the “true” DR curve. These were: $a = 0.128, b = 0.848, c = -5.515, d = 9.996$. Data was simulated according to this DR model using code modified from code by Dr. Reich available [here](#).

Three different values for m, the number of dose response pairs, were considered: $m = 7, 10, 15$. This was to simulated the typical DR modeling scenario, where there are a large number of DR curves, but the data for each curve is sparse [cite reich]. For each sample size, doses: $x_1 < x_2 < \dots < x_m$ were equally spaced on the interval $[0, 1]$. 100 DR datasets were simulated for each sample size according to the known DR model.

Results

Figure 1 shows an example of one of the DR curves simulated when $m = 10$. The estimate of the non-linear response function $\hat{f}(x)$ is shown according to each of the four non-linear regression methods. 95% confidence/credible intervals are shown for $f(x_i)$ at each dosage: x_1, \dots, x_m . For comparison, the true $f(x)$ curve is also plotted. For each method, $\hat{f}(x)$ matches $f(x)$ closely. The intervals estimated by all four methods contain the true curve. The intervals are tightest for the PB and SS methods. The LA and BNLR methods estimate confidence intervals that are substantially wider, suggesting that they are overestimating the standard error of the response function for this data set.

Coverage probabilities (CPs) for parameter and response function 95% CIs were estimated by Monte Carlo simulation. At each sample size, 100 DR curves were simulated. For single parameter CPs, the fraction of the intervals that contained the true parameter values estimated the CP. The average of the CP estimates at all of the design points was used to assess the CP of the response function. If $CI(x_j)_i$ is the 95% CI for $f(x_j)$ for the ith Monte Carlo sample, then I define the estimate the CP of the response function as:

$$\widehat{CP} = \frac{1}{n * m} \sum_{i=1}^n \sum_{j=1}^m f(x_j) \in CI(x_j)_i$$

The expected width of each CI was also estimated with the average CI width for the Monte Carlo samples. The average of the CI widths over all design points was reported as the estimate of expected regression fit CI width.

Ideally a CI will cover the true parameter with high probability and will have a small expected width, meaning they are informative about the true parameter value. CIs with coverage probabilities much lower than the advertised value can be misleading. Results such as significant differences between curves can appear stronger than they actually are.

Figures 2 and 3 show the parameter and regression fit CPs and average CI widths for all four methods. In **Figure 2**, methods with CP closest to the .95 line are desirable. CPs higher than the line indicate over-estimation of standard errors, while CPs lower than the line indicate under-estimation. At the smallest sample size (7), the LA and PB CIs have CPs that are far below the nominal level. This is true for both the parameters and the regression fits. As the sample size increases to 15, however, the CPs are close .95. LA is often closer to the .95 line, suggesting that it has better CP performance than PB. Also, **Figure 3** shows the expected width of the LA parameter CI intervals is substantially smaller than PB, so LA is clearly the optimal choice for the parameters. For the regression fit, PB produces intervals with smaller expected width, but the CP is far below 95%. However, at larger sample sizes, LA is much too conservative, so neither method is ideal. For LA, the CIs are too wide and the standard errors overestimated, but this method is still preferable to PB as the coverage probability is not far below the advertised value. The PB CIs would suggest results that are stronger than actually are and should be treated with caution.

Figures 2 and 3 show the Bayesian NLR results as well, but these results are suspect as none of the models reached convergence. 100 MCMC runs per sample size were performed. Each run consisted of 10,000 burnin samples and 50,000 post burnin MCMC samples. The effective sample sizes for many of the parameters were below 1,000 and the Gelman Rubins diagnostic was above 1.1 for some parameters.

The BNLR CIs have a CPs above 95% for the parameters and regression fit at small sample sizes, but curiously this dips below 95% when the sample size increases to 15. While the CP appears good for these intervals, the expected widths of these intervals are extremely large. For the c and d parameters the expected widths were so large that they could not be plotted. I think this is likely due to the fact that I sampled θ parameters from their posterior in the Hill function and then transformed to the c and d parameters using ratios. The regression fit CIs have CPs above 95% but the expected widths are substantially larger than LA. Interestingly, the other Bayesian interval estimation method, SS resulted in intervals with CP and expected width that were nearly identical to BNLR.

If regression fit CIs are of interest, the choice of the best method is really up to the user. LA provides CIs that have small expected widths, but at small sample sizes the CP may be slightly below .95. LA will probably provide the most accurate estimate of the standard errors, but at times it underestimates. Since this will only result in slightly overconfident result, how serious of a problem this is is up to the user. BNLR and SS could be an alternative to LA with CP that is above the nominal level, but the expected widths will be much larger, meaning that true differences between DR curves may be detected with low probability. If a conservative method is desired, SS should be chosen over BNLR as BNLR is much more computationally intensive.

For NLR parameters, LA CIs are clearly optimal at large small sample sizes, with small expected widths and CP close to .95. With small sample sizes, all methods are suspect. LA and PB have CP much below .95 and BNLR results in intervals that are unreasonably large.

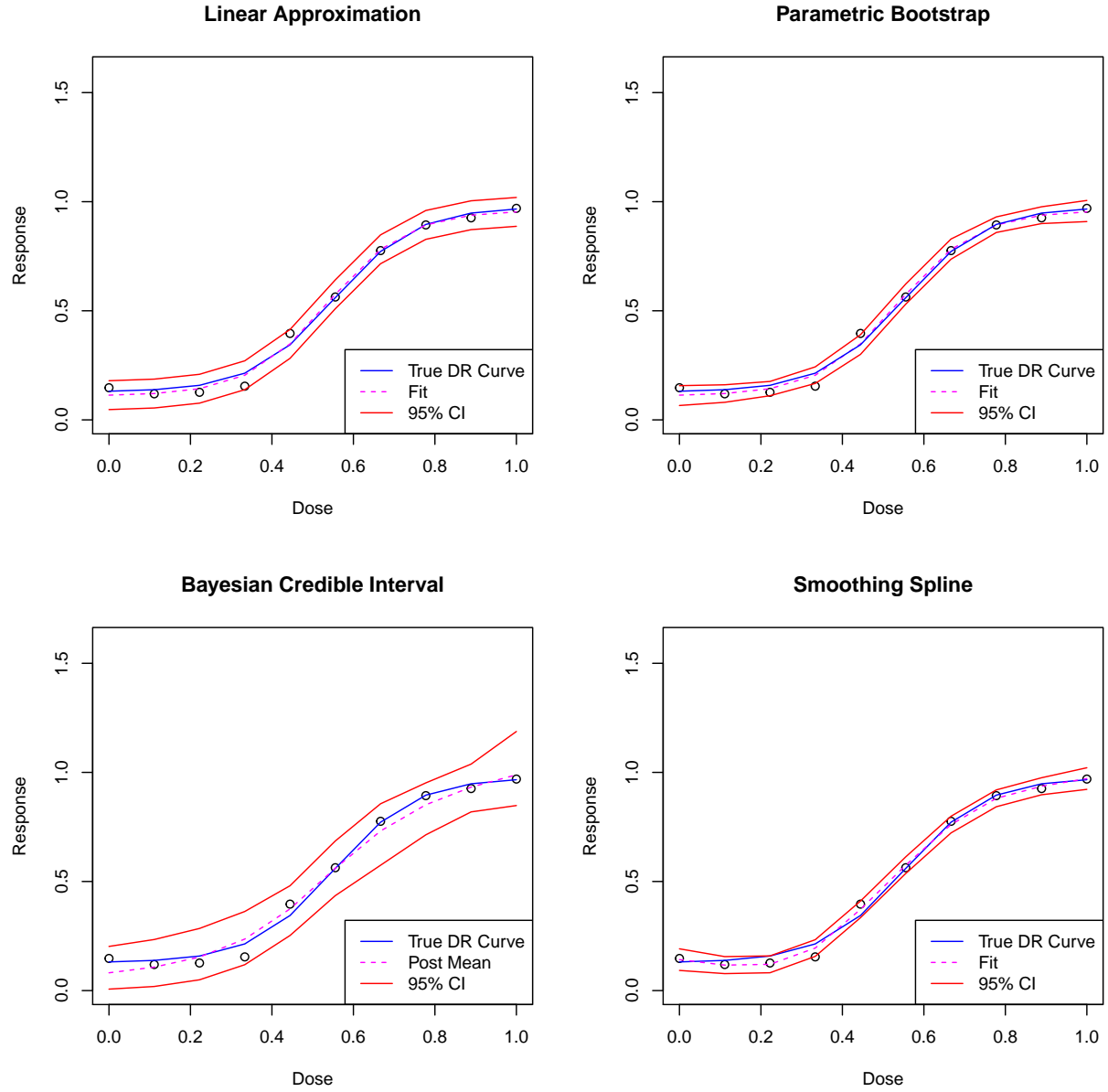


Figure 1: Example of the non-linear regression fits and interval estimates by four methods. (top) NLR fit with 95% confidence intervals (red) estimated by linear approximation (top left) or parametric bootstrap (top right). The true function used to generate the data is shown in blue. (bottom left) Bayesian NLR. Fit is the posterior mean of the regression line (purple dashed) along with a 95% credible interval (red). (bottom right) Fit estimated by a smoothing spline (purple dashed) with Bayesian 95% credible interval (red).

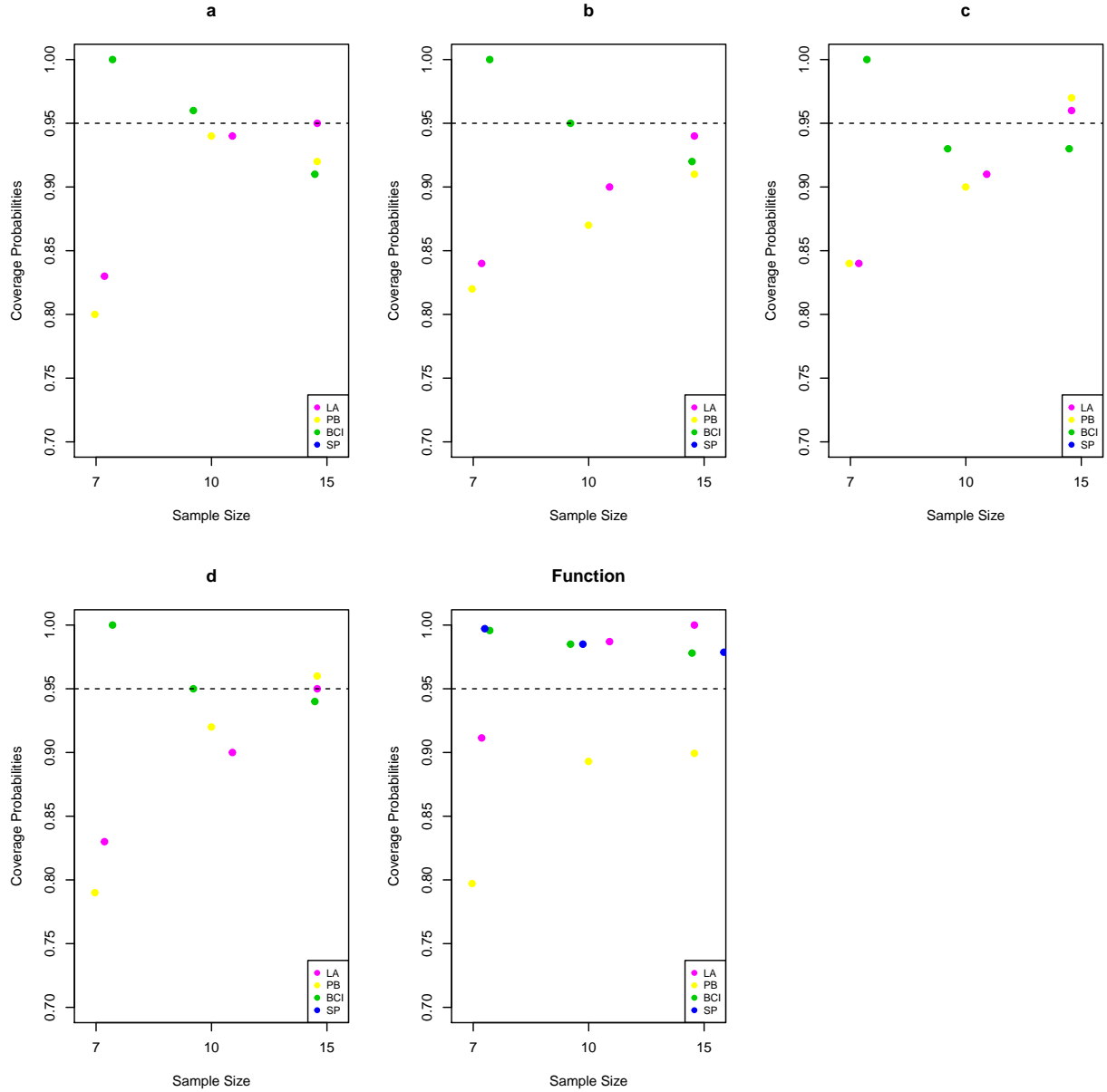


Figure 2: Coverage probabilities for the interval estimates by four non linear regression methods. 95% confidence/credible intervals were estimated for each of the parameters and the regression line over all of the design points. Coverage probabilities were estimated by Monte Carlo simulation (100 replicates). A dashed line shows the 95% coverage probability which is the nominal level for the frequentist methods. (red) NLR linear approximation 95% confidence interval. (blue) NLR parametric bootstrap confidence interval. (purple) Bayesian NLR 95% credible interval. (yellow) Smoothing spline 95% credible interval.

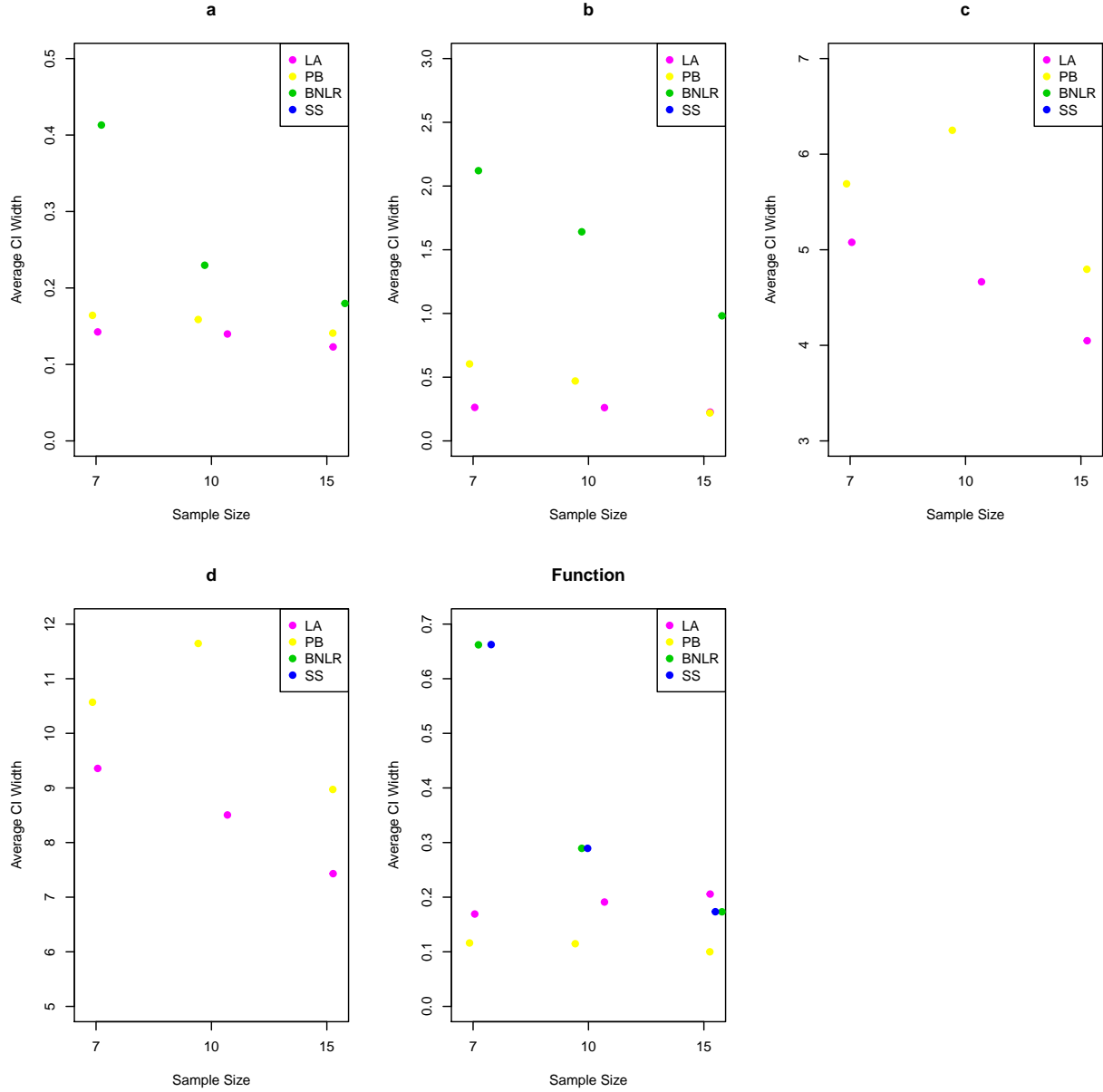


Figure 3: Expected width for the interval estimates by four non linear regression methods. 95% confidence/credible intervals were estimated for each of the parameters and the regression line over all of the design points. Expected widths were estimated by Monte Carlo simulation (100 replicates). (red) NLR linear approximation 95% confidence interval. (blue) NLR parametric bootstrap confidence interval. (purple) Bayesian NLR 95% credible interval. (yellow) Smoothing spline 95% credible interval.

References

- Bates, Douglas M., and Donald G. Watts. 1988. “Nonlinear Regression Analysis and Its Applications.” *New York: Wiley* 85 (410):594. <https://doi.org/10.1002/9780470316757>.
- Draper, Norman R, and Douglas E Shaw. 1974. “The Effects of Nonlinearity in Regression Models. Part 1. A Procedure for Examination of Adequacy of the Linear Approximation.” WISCONSIN UNIV MADISON MATHEMATICS RESEARCH CENTER.
- Duncan, George T. 1978. “An empirical study of jackknife-constructed confidence regions in nonlinear regression.” *Technometrics*. <https://doi.org/10.1080/00401706.1978.10489636>.
- Elzhov, Timur V, Katharine M Mullen, Andrej-Nikolai Spiess, and Ben Bolker. 2016. “minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds.” <https://cran.r-project.org/package=minpack.lm>.
- Fox, John. 2015. *Applied Regression Analysis and Generalized Linear Models*. Sage Publications.
- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. Vol. 112. Springer.
- Man, Stephen, and Sing Lee. 2018. “Optimal choice between parametric and non-parametric bootstrap estimates.” *Math. Proc. Camb. Phil. Soc* 115 (335):3–3. <https://doi.org/10.1017/S0305004100072121>.
- Marra, Giampiero, and Simon N Wood. 2011. “Coverage Properties of Confidence Intervals for Generalized Additive Model Components *.” <http://opus.bath.ac.uk/26662/1/MarraWoodCI.pdf>.
- Moré, Jorge J. 1978. “The Levenberg-Marquardt algorithm: Implementation and theory.” In. <https://doi.org/10.1007/BFb0067700>.
- Peddada, Shyamal, and Joseph Haseman. 2005. “Analysis of Nonlinear Regression Models: A Cautionary Note.” *Dose-Response* 3 (3):342–52. <https://doi.org/10.2203/dose-response.003.03.005>.
- Simonoff, Jeffrey S., and Chih-Ling Tsai. 1986. “Jackknife-Based Estimators and Confidence Regions in Nonlinear Regression.” *Technometrics* 28 (2):103–12. <https://doi.org/10.1080/00401706.1986.10488111>.
- Wilson, Ander, David M Reif, Brian J Reich, North Carolina, and North Carolina. 2013. “Supplemental Material for Hierarchical Dose-Response Modeling for High-Throughput Toxicity Screening of Environmental Chemicals,” no. September:0–10.
- Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with R*. CRC press.