dr4pl: A Stable Convergence Algorithm for the 4 Parameter Logistic Model

by Hyowon An, Justin T. Landis, Aubrey G. Bailey, James S. Marron and Dirk P. Dittmer

Abstract The *4 Parameter Logistic* (*4PL*) *model* has been recognized as a major tool to analyze the relationship between doses and responses in pharmacological experiments. A main strength of this model is that each parameter contributes an intuitive meaning enhancing interpretability of a fitted model. However, implementing the 4PL model using conventional statistical software often encounters numerical errors. This paper highlights the issue of convergence failure and presents several causes with solutions. These causes include outliers and a non-logistic data shape, so useful remedies such as robust estimation, outlier diagnostics and constrained optimization are proposed. These features are implemented in a new R package dr4pl (Dose-Response analysis using the 4 Parameter Logistic model) whose code examples are presented as a separate section. Our R package dr4pl is shown to work well for data sets where the traditional dose-response modelling packages drc and nplr fail.

Introduction

Dose-response models are a primary tool of pharmacological and toxicological studies. After fitting a suitably chosen model to experimental data, the fitted parameters are used to evaluate the potency of a drug or compare the potencies of different drugs. Based on outcomes of the analysis, drug candidates may be advanced to further development. One of the most popular models among dose-response models is the *4 Parameter Logistic* (*4PL*) *model* whose detailed explanation is given in Subsection 4 Parameter Logistic model. In this paper, we focus on some important challenges in the 4PL model and present methods to handle those challenges.

Fitting dose-response models to data is usually a nonlinear regression problem. One of the problems with nonlinear regression is that numerical errors often occur during a model fitting optimization process. The 4PL model is not an exception. Examples of problem data sets and diagnostics of their problems are presented in Subsection Motivation: convergence failure.

In this paper, we analyze the convergence failure problem within the context of the 4PL model. Descriptions of the 4PL model and existing methods are presented in Section Existing methods. Problems are clearly illustrated using the contour of the likelihood function, shown in Section Diagnosis of convergence failure. This suggests a useful set of constraints to put on the parameter space, and their effectiveness is analyzed in that section as well. Non-convergence is indicated by our R package **dr4pl** (Landis et al., 2018) when crossing one of these boundaries, in which case robust parameter estimates are returned as described in Section Remedy for outliers. Section Remedy for the support problem presents a new method of selecting initial parameter estimates which is not currently implemented in other dose-response modelling R packages. This method will be shown to be effective in handling a non-logistic data shape in that section. Walk-through of R functions in **dr4pl** with real-world data sets will be presented in Section Walk-through in R.

Existing methods

This section describes the 4PL model and existing methods that have been implemented for the 4PL model. Example data sets that motivate our research are presented as well.

4 Parameter Logistic model

The response in a pharmacological process usually slowly trends downward at low dose levels, then rapidly drops at a certain dose level and finally stabilizes. Tracking this sort of change in response results in a reverse S shaped curve motivating sigmoidal curves as primary models of interest. Robertson (1908) suggested that the biochemical nature of such processes is well modelled by the differential equation

$$\frac{\mathrm{d}g}{\mathrm{d}z} = \frac{\theta_3}{\theta_1} g(z) \left(\theta_1 - g(z) \right),$$ (1)

where z is a dose level, $g(z)$ is the response at the level z and $\theta_1, \theta_3 \in \mathbb{R}$ are parameters. Solving this differential equation yields

$$g(z|\theta_1, \theta_2, \theta_3) = \theta_1 - \frac{\theta_1}{1 + (z/\theta_2)^{\theta_3}},$$

where $\theta_2 > 0$ is a parameter coming from a boundary condition given to Equation (1). Since this model allows the response value $g(z)$ to range only between 0 and θ_1, a common extension uses a shift parameter $\theta_4 \in \mathbb{R}$ yielding the well-known *4 parameter logistic (4PL) model* or *Hill model* given in Hill (1910) as

$$
\begin{aligned}
f(x|\theta) &= \theta_1 + \frac{\theta_4 - \theta_1}{1 + 10^{\theta_3(x - log_{10}\theta_2)}} \\
&= \theta_1 + \frac{\theta_4 - \theta_1}{1 + (z(x)/\theta_2)^{\theta_3}},
\end{aligned}
\tag{2}
$$

where $x \in \mathbb{R}$ is a log 10 dose, $z(x) = 10^x > 0$ is the corresponding dose and the parameter vector $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$ determines the relationship between the log 10 dose x and the response $f(x)$. Working on this log 10 scale has been shown to accurately model many physical reactions.

A major strength of the 4PL model is that each of the four parameters has an intuitive meaning. The parameters θ_1 and θ_4 are called the *upper asymptote* and *lower asymptote*, respectively, since we have from Equation (2) that

$$\text{if } \theta_3 > 0, \text{ then } \lim_{x \to -\infty} f(x|\theta) = \theta_4, \ \lim_{x \to \infty} f(x|\theta) = \theta_1,$$
$$\text{if } \theta_3 < 0, \text{ then } \lim_{x \to -\infty} f(x|\theta) = \theta_1, \ \lim_{x \to \infty} f(x|\theta) = \theta_4.$$

In this paper, we assume that $\theta_1 > \theta_4$ holds. This implies that $\theta_3 > 0$ corresponds to an increasing curve while $\theta_3 < 0$ corresponds to a decreasing curve. A response usually decreases as a dose increases in a pharmacological experiment, so we focus on decreasing curves throughout this paper. However, all the functionalities such as model fitting and confidence interval calculation are implemented in the R package **dr4pl** for both decreasing and increasing curves. The parameter θ_2 is called the *half maximum inhibitory concentration* (*IC50*) or *effective concentration* (*EC50*), depending on whether the curve is decreasing or increasing, which represents the dose level whose response lies halfway between the upper and lower asymptotes. That is, we have $f(\theta_2) = (\theta_1 + \theta_4)/2$. The parameter θ_3 is related to the slope of the function f at θ_2 as given in Appendix A of Giraldo et al. (2002) by

$$\left. \frac{\mathrm{d}}{\mathrm{d}x} f(x|\theta) \right|_{x = log_{10}\theta_2} = (\ln 10)(\theta_4 - \theta_1)\theta_3,$$

where ln indicates the natural logarithm.

The effects of the IC50 parameter θ_2 and the slope parameter θ_3 on the 4PL model is illustrated in Figure 1. For each plot in the figure, the x-axis represents dose levels on a log 10 scale and the y-axis represents response values. For both the left and right plots, the upper asymptote θ_1 is fixed at 100 and the lower asymptote θ_4, which is represented by a horizontal gray dashed line, is fixed at 0. The left plot shows the change in the dose response curve as the IC50 parameter θ_2 changes. Note that the blue dotted curve with the largest θ_2 value of 10 is located on the right while the red solid curve with the smallest θ_2 value of 0.1 is located on the left. However, the rate of decrease θ_3 of each curve is the same. The right plot shows that the blue dotted curve with the most negative θ_3 value of -2 most rapidly decreases around the IC50 value while the red solid curve with the least negative θ_3 value of -0.5 most slowly decreases. Note that all three curves in the right plot achieve the half response θ_2 at the same dose level, i.e. they have the same IC50 values.

Motivation: convergence failure

As mentioned in Section Introduction, dose-response models are prone to numerical errors in their optimization processes like other nonlinear regression models; see Bates and Watts (1988) and Seber and Wild (1989) for theoretical and practical issues related to nonlinear regression models. The specific type of errors in dose-response models depends on the data and statistical software at hand, and here we focus on R packages developed for bioassay data. The R package **drc** (Ritz et al., 2005) provided various non-linear regression models for bioassay data analysis. The R package **nplr** (Commo and Bot, 2016) focused on the family of n-parameter logistic regression models with an option to choose an optimal value n based on a goodness-of-fit measure. Unfortunately, these two R packages have only scant solutions to the convergence failure problem. For example, **drc** provides a function argument `errorm` to control whether an error or a warning message is drawn when convergence failure happens, and **nplr** provides a weighting scheme that reduces the effect of outliers. Both methods did not suffice to prevent convergence failure as shown in the following.

Figure 2 shows four real-world data sets. The R package **drc** fails to converge on all these four data sets and the R package **nplr** fails on the last data set. In each plot of the figure, the x-axis represents doses on a log 10 scale and the y-axis represents responses. Data points flagged as problematic by our
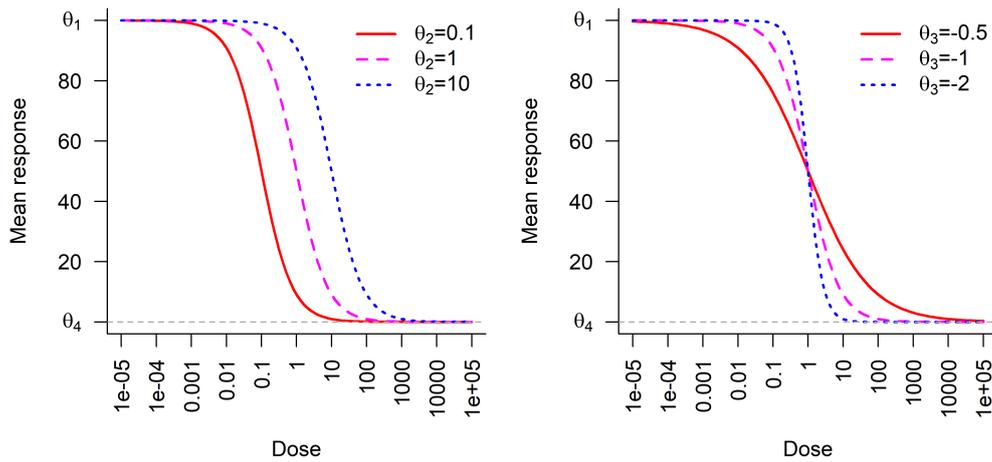
Figure 1: 4PL models with different values of the IC50 parameter θ_2 (left plot) and the slope parameter θ_3 (right plot). The other parameters are fixed at $\theta_1 = 100, \theta_4 = 0$ with $\theta_3 = -1$ in the left plot and $\theta_2 = 1$ in the right plot. The left plot implies that θ_2 controls horizontal shifts while the right plot implies that θ_3 drives rates of decrease.

R package are shown as red triangles. These include strong outliers in every case, and are found as described in Subsection Robust regression and outlier detection. Error Case 1 in the upper left plot has an outlier at the dose level 10^{-5} which seems to result from an error in measurement. Error Case 2 in the upper right plot has more complicated problems. There exists a strong outlier at the dose level 1.35, and the two responses measured at the dose level 0.135 seem to be measured at a wrong scale. Both of them have abnormally large values and their values are far more different from each other than the responses measured at other dose levels. Even though these two responses were not detected as outliers by our method, this plot indicates that the data set of Error Case 2 requires users' attention.

The data in the lower two plots have a different problem in addition to outliers. They seem to have insufficient data points on the right side of their IC50 parameters. This situation is well known in dose-response experiments in cases where solubility issues limit the maximal drug concentration that can be used in dose titration. Error Case 3 in the lower left plot shows two strong outliers with abnormally small response values at the dose level 10. In addition, the response starts decreasing at the dose levels 10 and 100 but no more responses are measured at higher dose levels. This is because the drug fell out of solution at higher concentrations in this experiment. We call this type of problem a *support problem* because of a lack of data points on the right side, and possibly also on the left side, of the IC50 parameter. This lack in Error Case 3 can result in instability of an optimization process due to difficulty in estimation of the IC50 parameter θ_2 and the slope parameter θ_3. Indeed, the package **drc** draws an error message for that data. Error Case 4 in the lower right plot has a similar support problem also resulting in a **drc** error.

Convergence failure of numerical optimization methods can be overcome in various ways. Experimenters often try other starting parameter values or adopt a simpler mathematical model. However, both solutions have a drawback in dose-response modelling since they require arbitrary decisions by experimenters. This limits their utility for automated, high-throughput screening and may be cost-prohibitive depending on the expense per data point. It can be foolhardy to decide which starting values should be tried next unless problems with data are diagnosed. Performing a grid search on the parameter space can be an alternative as claimed in Wang et al. (2010), but even this does not fundamentally solve the problems with the data sets in Figure 2. Shifting to a simpler model can result in loss of interpretability of the 4PL model. As pointed out in Di Veroli Giovanni et al. (2015) and Lim (2015), the 4PL model conveys useful intuitive content through their parameters.

Diagnosis of convergence failure

The 4PL model given in Equation (2) is fitted to a data set $\{(x_i, y_i) \,|\, i = 1, 2, \cdots, n\}$ where x_i is the log 10 dose level of the i-th observation and y_i is the corresponding response value. The most commonly
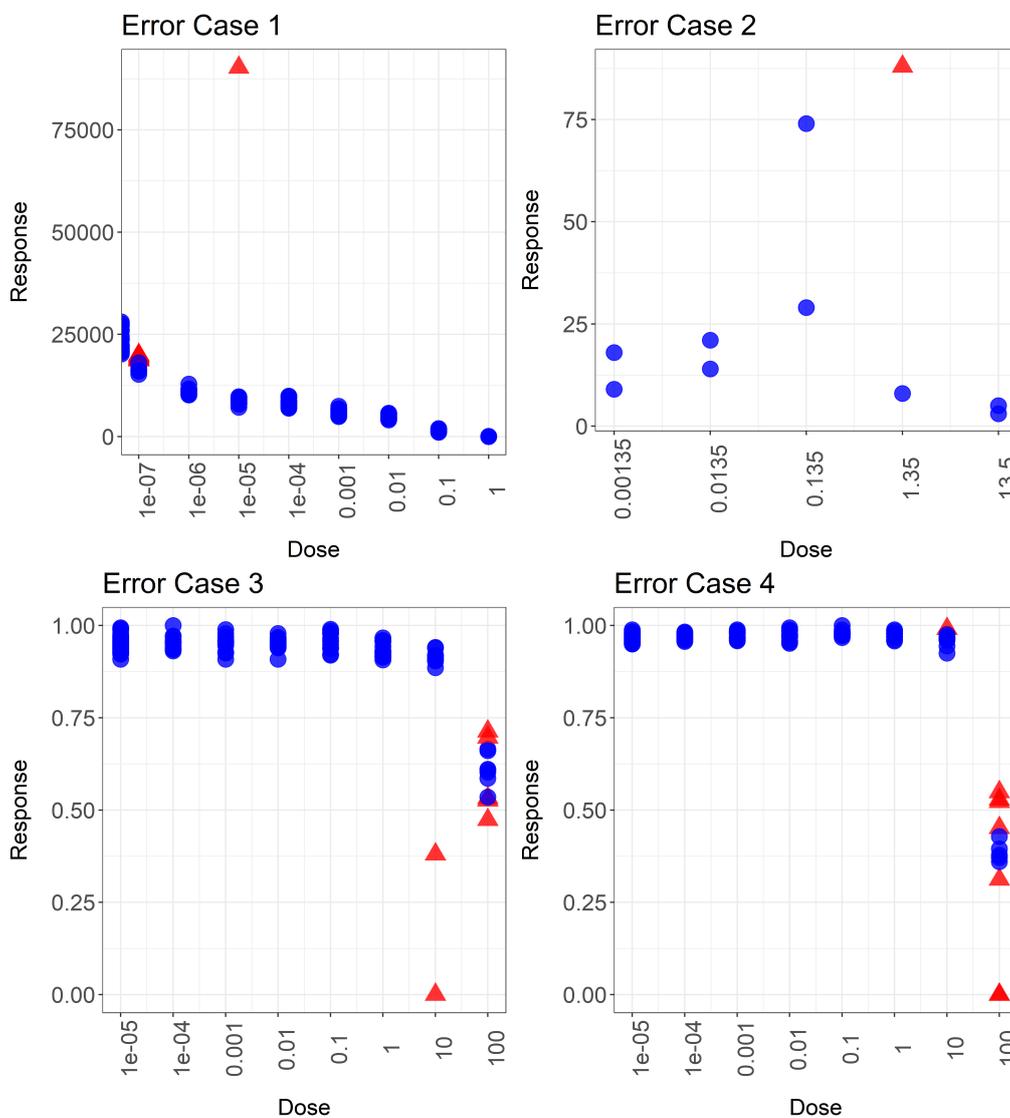
Figure 2: Data that generate numerical errors in the R packages **drc** and **nplr**. The y-axis represents responses and the x-axis represents doses in a log 10 scale. All data sets have at least one strong outlier which are denoted by red triangles. Data in the lower row also have insufficient numbers of data points on the right sides of the dose-response curves.

chosen loss function is the sum-of-squares errors essentially given as

$$\mathcal{L}(\theta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i|\theta))^2,$$ (3)

where the function f was given in Equation (2) as

$$f(x|\theta) = \theta_1 + \frac{\theta_4 - \theta_1}{1 + 10^{\theta_3(x - log_{10}\theta_2)}}.$$ (4)

Since this function is nonlinear in the parameters θ, minimization of $\mathcal{L}$ is a typical nonlinear regression problem. Gadagkar and Call (2015) mentioned the optimization problem is indeed a non-convex optimization, so different initial parameter estimates can result in different fitted models or even convergence failure. Example data sets with such a convergence failure problem were given in Figure 2. In this section, we present one possible solution to the convergence problem based on statistical theory, which can help users diagnose problems with their data.

Shape of a loss function

To find a reason behind the convergence failure, we draw contour plots of the sum-of-squares loss function given in Equation (3). To visualize the loss function $\mathcal{L}(\theta|x)$, we first fix the values of θ_1 and θ_4 at $y_{\max}$ and $y_{\min}$ defined as

$$y_{\max} = \max_{1 \le i \le n} y_i + 0.001 \cdot \left(\max_{1 \le i \le n} y_i - \min_{1 \le i \le n} y_i \right),$$

$$y_{\min} = \min_{1 \le i \le n} y_i - 0.001 \cdot \left(\max_{1 \le i \le n} y_i - \min_{1 \le i \le n} y_i \right),$$ (5)

respectively. The reason for choosing the above estimates is that the R package **drc** uses them as a default. Then we draw contour plots of the loss functions in terms of θ_2 and θ_3 yielding Figure 3 in which the contour plots of the data of Figure 2 are shown. In each plot, the x-axis represents the IC50 parameter θ_2 in a log 10 scale and the y-axis represents the slope parameter θ_3. The parameter values corresponding to the same loss function value are connected as a solid line and the error value is given as text inserted into that line. It can be seen from the upper two plots that the loss function monotonically decreases as the IC50 parameter θ_2 decreases towards zero. This results in numerical errors in computing the estimated mean response in Equation (4), since as $\theta_2 \to 0$, the $log_{10}(\theta_2) \to -\infty$. The lower two plots also indicate that the loss function tends to decrease as the values of the IC50 parameter θ_2 increase towards infinity. Similarly to the upper two plots, this causes numerical errors in finding the minimizer of the loss function.

Hill bounds: safeguards on the parameter space

The primary resolution of the convergence failure problem comes from imposing appropriate boundaries on the parameter space to prevent parameter estimates from diverging to infinity or converging to 0. These bounds should be broad enough to contain the global minimum of the loss function when it has any local minimum. To obtain reasonable sizes of bounds on the parameters, we focus on the *Hill plot*. Recall from Equation (2) that the mean response is modelled as

$$f(x|\theta) = \theta_1 + \frac{\theta_4 - \theta_1}{1 + 10^{\theta_3(x - log_{10}\theta_2)}},$$

where x is a log 10 dose. Note that the nonlinear nature of the 4PL model comes from the fact that the linear function of x, $\theta_3(x - log_{10}\theta_2)$, appears as an exponential in the denominator of the expression. This motivates us to rearrange the terms to yield the *Hill equation* modeled as

$$log_{10}\left(\frac{f(x|\theta) - \theta_4}{\theta_1 - f(x|\theta)} \right) = \theta_3 x - \theta_3 log_{10}(\theta_2).$$ (6)

The left hand side of this equation is the logit-transformed mean response and the right hand side is the linear function of log 10 dose x. Since both the asymptotes θ_1 and θ_4 are unknown, we use $y_{\max}$ and $y_{\min}$ defined in Equation (5) as their estimates in real data approximation as

$$log_{10}\left(\frac{y_i - y_{\min}}{y_{\max} - y_i} \right) \approx -\theta_3 log_{10}(\theta_2) + \theta_3 x_i.$$ (7)

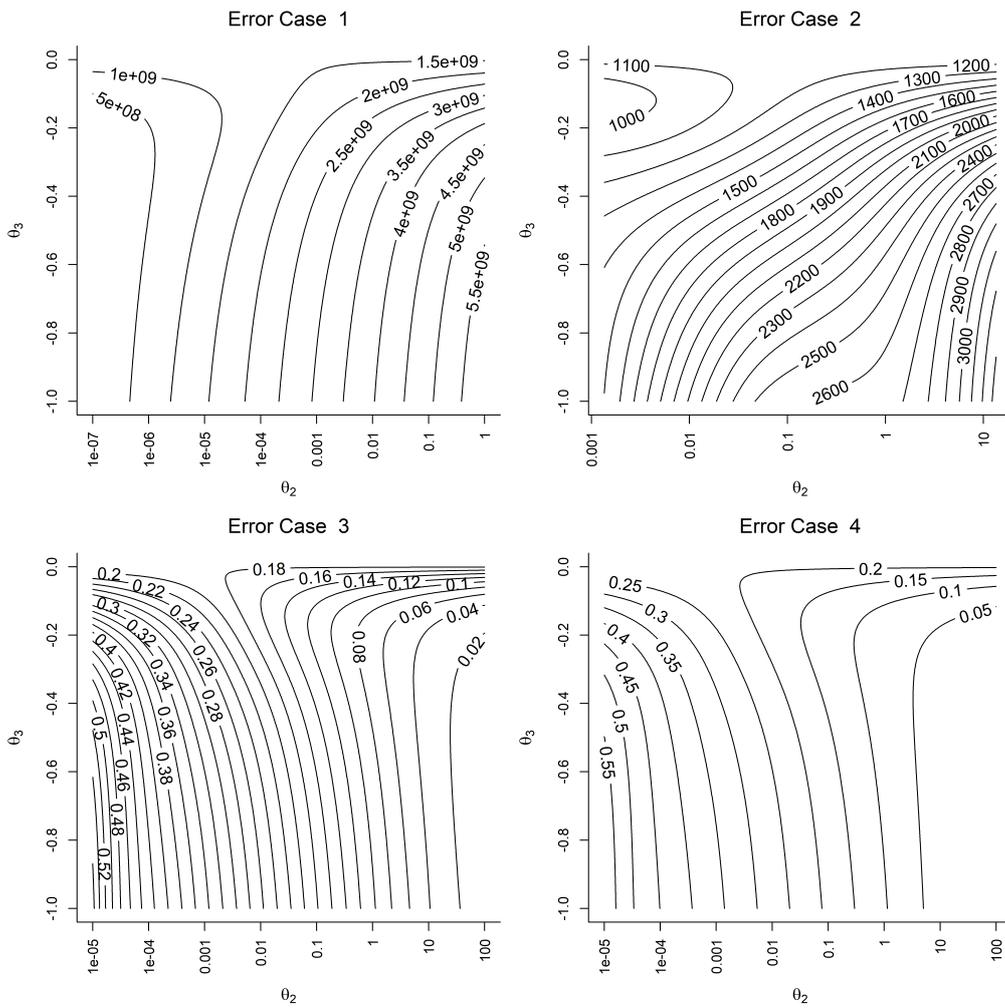See Chapter 2 of Pratt and Taylor (1990) for detailed discussion of the Hill plot and Hill equation.

Figure 3: Contour plots of data shown in Figure 2. The *x*-axis represents the IC50 parameter θ_2 and the *y*-axis represents the slope parameter θ_3. The loss functions in the upper two plots decrease as the IC50 θ_2 decreases, and the loss functions in the lower two plots decrease as θ_2 increases; a minimum is not achieved in the plots.

To obtain reasonable bounds on the parameters θ_2 and θ_3, we first view Equation (7) as a linear regression problem. If we re-parameterize Equation (7) as

$$log_{10}\left(\frac{y_i - y_{\min}}{y_{\max} - y_i}\right) = \beta_2 + \beta_3 x_i + \epsilon_i, \tag{8}$$

where $\beta_2 = -\theta_3 log_{10}(\theta_2)$, $\beta_3 = \theta_3$ and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, we can obtain interval estimators of β_2 and β_3 based on $(1 - \alpha)$ confidence intervals based on the normality assumption. That is, we compute

$$\begin{aligned}
&\left(\tilde{\beta}_2 - z_{1-\alpha/2}\text{s.e.}(\tilde{\beta}_2), \tilde{\beta}_2 + z_{1-\alpha/2}\text{s.e.}(\tilde{\beta}_2)\right), \\
&\left(\tilde{\beta}_3 - z_{1-\alpha/2}\text{s.e.}(\tilde{\beta}_3), \tilde{\beta}_3 + z_{1-\alpha/2}\text{s.e.}(\tilde{\beta}_3)\right),
\end{aligned} \tag{9}$$

where $\tilde{\beta}_2$ and $\tilde{\beta}_3$ are the least squares solutions of Equation (8), z_p is the p-th quantile of the standard normal distribution and s.e. is the abbreviation for the standard errors of the regression coefficients. If the linear model of Equation (7) well approximates the original 4PL model of Equation (2), then parameter estimates will rarely be out of the bounds in Equation (9) during the optimization process.

However, using these bounds in optimization have limitations in practice in that they may miss important nonlinear structure inherent in the 4PL model. In real data analysis, such confidence intervals imposed too narrow bounds on the parameters and hindered optimization processes from convergence even when data have good logistic shapes. Section 5.1 of Seber and Wild (1989) present approximate confidence intervals of the true parameters of nonlinear regression models based on the derivatives of the mean response function f. When data have sufficiently many data points, the least squares estimators $\hat{\theta}$ have asymptotically normal distributions

$$\hat{\theta} - \theta^* \xrightarrow{d} \mathcal{N}\left(0, C^{-1}\right), \tag{10}$$

where θ^* is the true parameter vector and $C = J^T J$ with J being the Jacobian matrix whose (i, j)-th element is given as $\partial f(x_i|\theta)/\partial\theta_j$. In actual computation of confidence intervals, the parameter estimators $\hat{\theta}$ are used to approximate J, i.e. we substitute $\hat{J}$ whose (i, j)-th element is $\partial f(x_i|\theta)/\partial\theta_j\big|_{\theta=\hat{\theta}}$ for J.

Based on this asymptotic theory, the approximate confidence intervals for the true parameters are given in Section 5.1 of Seber and Wild (1989) as

$$\left(\hat{\theta}_j + t_{n-p}^{\alpha/(2\cdot p)}s\sqrt{\hat{c}_{j,j}}, \hat{\theta}_j + t_{n-p}^{1-\alpha/(2\cdot p)}s\sqrt{\hat{c}_{j,j}}\right), \tag{11}$$

for $1 \le j \le p$ where p is the number of parameters, s^2 is the residual sum of squares, t_n^α is the α-th quantile of the t-distribution with n degrees of freedom and $\hat{c}_{j,j}$ is the j-th diagonal element of the matrix $\hat{C}^{-1} = (\hat{J}^T\hat{J})^{-1}$. In the context of the 4PL model, we evaluate the interval estimators based on parameter estimators $\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3, \tilde{\theta}_4$ obtained from the Hill equation in Equation (7) and standard errors obtained from Equation (11). That is, we have

$$\left(\tilde{\theta}_j + t_{n-4}^{\alpha/8}s\sqrt{\tilde{c}_{j,j}}, \tilde{\theta}_j + t_{n-4}^{1-\alpha/8}s\sqrt{\tilde{c}_{j,j}}\right), \tag{12}$$

for $1 \le j \le 4$ where $\tilde{c}_{jj}$ is the j-th diagonal element of the variance-covariance matrix $\tilde{C}^{-1}$ evaluated at the initial parameter estimators $\tilde{\theta}$ obtained from the Hill equation. Since the bounds in Equation (12) originate from the Hill model, we call these the *Hill bounds*. Subsection 5.2.2 of Seber and Wild (1989) suggests using $\hat{H}/2$ instead of $\hat{J}^T\hat{J}$ when computing confidence intervals where

$$\hat{H}_{j,k} = \left.\frac{\partial^2 \mathcal{L}}{\partial\theta_j\partial\theta_k}\right|_{\theta_j=\hat{\theta}_j,\theta_k=\hat{\theta}_k},$$

is an estimated Hessian matrix of the loss function $\mathcal{L}$. We implement both versions of the Hill bounds based on $\tilde{H}/2$ and $\tilde{J}^T\tilde{J}$ in our R package with the default being the latter, since the latter tends to yield broader Hill bounds. Users can try using both options in real data analysis. The confidence level α is set at 0.0001 as a default and can be given by users as a tuning parameter.

One can wonder if bounds on the upper asymptote θ_1 and lower asymptote θ_4 are needed as well. The following theorem shows that such bounds are unnecessary since diverging values of θ_1 and θ_4 always increase loss function values.
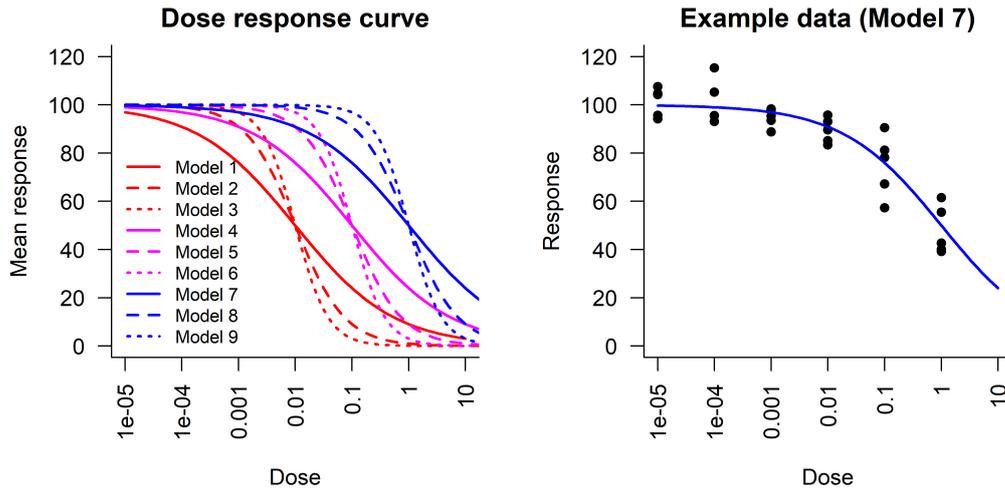
Figure 4: Dose response models presented in Table 2 and an example data set generated by Model 7. The left plot shows the dose response curves of the models in Table 2. The right plot shows data generated by Model 7 of Table 2.

Model	Parameters $(\theta_1, \theta_2, \theta_3, \theta_4)$	Model	Parameters $(\theta_1, \theta_2, \theta_3, \theta_4)$
1	(100,0.01,-0.5,0)	6	(100,0.1,-1.5,0)
2	(100,0.01,-1,0)	7	(100,1,-0.5,0)
3	(100,0.01,-1.5,0)	8	(100,1,-1,0)
4	(100,0.1,-0.5,0)	9	(100,1,-1.5,0)
5	(100,0.1,-1,0)		

Table 1: Model numbers and their parameters used in our simulation.

Theorem 2.3.1. *Let $\theta_2 > 0$ and $\theta_3 \in \mathbb{R}$ be fixed. Then there exist constants $\theta_1', \theta_4' \in \mathbb{R}$ such that*

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta_1} > 0 \; \forall \theta_1 \geq \theta_1', \quad \frac{\partial \mathcal{L}(\theta)}{\partial \theta_4} < 0 \; \forall \theta_4 \leq \theta_4'$$

for fixed $\theta_4 \in \mathbb{R}$ and $\theta_1 \in \mathbb{R}$, respectively. That is, loss function values increase as the values of parameters θ_1 and θ_4 grow to the infinity and the negative infinity, respectively, which prevents their estimates from diverging.

Proof. *See Appendix.* ■

Analysis of the Hill bounds

In this subsection, we evaluate whether the Hill bounds are appropriate as a criterion for convergence failure. We simulate data following the 4PL model and check whether the Hill bounds prevent optimization processes from finding the global minimum of the loss function $\mathcal{L}$. To evaluate this, we generate data and count the numbers of times the Hill bounds miss the true parameters. We also count the numbers of times that optimization procedures encounter the Hill bounds at the end of optimization. Lower counts in both cases indicate that the Hill bounds do not hinder optimization processes from converging to desired parameter estimates.

Figure 4 shows parameter settings of our simulation. The left plot shows nine dose-response curves that generate data whose values of parameters are shown in Table 1. The levels of slope parameter θ_3 are denoted by line types while the levels of IC50 parameter are denoted by colors. The right plot shows the dose-response curve of Model 7 with data generated by the model. This model is chosen since it exhibits the most severe support problem among all the nine models. Even though data in the right plot are generated by a normal distribution, it seems that any optimization procedure can have trouble fitting the 4PL model to that data. We simulate 1,000 data sets with five observations at each dose level resulting in 25 observations in total. The standard deviation of the underlying normal distribution of a 4PL model is fixed at 5 throughout the simulations.

Parameters	$\theta_3 = -0.5$	$\theta_3 = -1$	$\theta_3 = -1.5$
$\theta_2 = 0.01$	0.054	0.014	0
$\theta_2 = 0.1$	0	0	0.002
$\theta_2 = 1$	0.181	0	0

Table 2: Proportions of times out of 1,000 simulations that the Hill bounds miss any of the true IC50 and slope parameters. The Hill bounds contain the true parameters for most of the cases, except when data have a support problem (bottom row) and the slope is gradual (leftmost column).

Parameters	$\theta_3 = -0.5$	$\theta_3 = -1$	$\theta_3 = -1.5$
$\theta_2 = 0.01$	0.038	0.009	0.1
$\theta_2 = 0.1$	0.12	0.002	0.195
$\theta_2 = 1$	0.398	0.258	0.079

Table 3: Proportions of times out of 1,000 simulations that the Hill bounds are attained by optimization procedures. The Hill bounds do not prevent optimization processes from achieving local minima of $\mathcal{L}$ for most of cases, except when data have support problems as shown in the bottom row.

Table 2 shows the proportions of times that the Hill bounds miss true parameters of the nine models given in Figure 4. For each simulated data set, we counted the number of times that the Hill bounds miss any of the IC50 and slope parameters. Note from the table that the Hill bounds contain the true parameters for most of the simulated data. The only exception is when data have a severe support problem ($\theta_2 = 1$) and the slope of the dose-response curve is very gradual ($\theta_3 = -0.5$). Since we do not have well-shaped data for this case, we take the fact that the Hill bounds miss the true parameters for these data as an advantage. When the Hill bounds are attained at the end of the optimization, it is recommended to look at the scatter plot, to investigate potential support problems. As illustrated in the top two rows of the table, in cases where the range of the data is enough for estimation on both sides of the IC50, the Hill bounds successfully contain the true parameters.

Table 3 presents the proportion of times the Hill bounds are hit at the end of optimization processes out of 1,000 simulations. The final parameter estimates are inside the Hill bounds for most cases. This indicates that the Hill bounds are not too restrictive for optimization procedures to find a local minimum of the loss function $\mathcal{L}$. However, when there are severe support problems as shown in the bottom left and bottom middle cells, the Hill bounds restrict optimization processes to be inside the bounds. As implied in the previous paragraph, this can help users realize that their data suffer from support problems and actions should be taken.

Remedy for outliers

As seen in the previous section, it seems that the convergence failure results from either outliers or the support problem. We tackle each of the causes with a different approach. Outlier problems are often dealt with by robust estimation methods. The paper Riazoshams et al. (2009) reviewed various methods of outlier detection in the context of nonlinear regression, especially when the form of the mean response function f is known. On the other hand, Dixon's Q test in Dean and Dixon (1951) can be used to detect an outlier without any assumptions on the form of f, but this method is most effective when a data size is less than or equal to 10. In this paper, we use a novel method presented in Motulsky and Brown (2006) that combines robust estimation and outlier detection. This method is currently adopted in GraphPad Prism and called the ROUT method therein as a default outlier detection method. The indices of outliers in data and a scatter plot with a robust fit are provided to users to enhance their understanding of problems.

Robust regression and outlier detection

As noted in Subsection Motivation: convergence failure, one of the main reasons for convergence failure is strong outliers. In this case, the contour of the loss function $\mathcal{L}(\theta|x, y)$ can exhibit monotonicity as a function of the IC50 parameter θ_2 as shown in Figure 2. This implies that the Hill bounds of Subsection Hill bounds: safeguards on the parameter space can be attained at the end of an optimization process. The R package **dr4pl** reports in this case that there seems to exist a problem in data and it requires a

user's attention. Meanwhile, the R package fits a robust regression model to data and reports indices of outliers in data with a plot highlighting the outliers in red triangles, which will be shown in Figure 6. By studying the robust fit and the indicated outliers, the user can get some insight into the data and decide whether the observations denoted as outliers should be removed or not.

To determine which data observations are outliers, we use the method given in Motulsky and Brown (2006). Define the i-th residual of a fitted dose-response model

$$r_i = y_i - f(x_i|\hat{\theta}),$$

where the mean response f is given in Equation (2) and $\hat{\theta}$ is the estimator obtained by a robust regression model. The method of that paper approximates the distribution of r_i by a t-distribution and then adjusts p-values by the *false discovery rate* (FDR). See Benjamini and Yekutieli (2001) for explanation of FDR. Adapting their algorithm in the context of the 4PL model yields the following.

1. Fit a dose-response model to data using a robust regression model to obtain residuals r_i for $i = 1, 2, \cdots, n$. Detailed description of a robust regression model is given in the following paragraphs.

2. Compute the *median absolute deviation* (MAD) of absolute valued residuals $r_i^{(a)} = |r_i|$ as

$$\hat{\sigma} = \text{med}_{1 \leq i \leq n} \left\{ \left| r_i^{(a)} - \text{med}_{1 \leq i \leq n} \left\{ r_i^{(a)} \right\} \right| \right\},$$

   where med stands for the sample median.

3. Rank the absolute values of the residuals from the smallest to the largest, so that $r_{i:n}^{(a)}$ is the i-th order statistic of the absolute valued residuals.

4. Loop from $i = \lfloor 0.7 \times n \rfloor$ to n

   (a) Compute $\alpha_i = 0.01(n - (i-1))/n$ and $t_{i:n} = r_{i:n}^{(a)}/\hat{\sigma}$ for $i = 1, 2, \cdots, n$.

   (b) Compute the two-tailed p-value p_i of the statistic $t_{i:n}$ from the t distribution with $n - 4$ degrees of freedom.

   (c) If any of p_j is less than or equal to α_j, then end the loop and report all the observations for $1 \leq i \leq j$ as outliers.

Contrary to the suggestion of Motulsky and Brown (2006) to use their quantile based measure of scale,

$$\tilde{\sigma} = \frac{n}{n-4} q_{0.6827},$$

where $q_{0.6827}$ is the 0.6827-th quantile of absolute valued residuals, we use the median absolute deviation. The reason is that for small sample sizes around 10, the constant factor $n/(n-4)$ in this equation significantly increases a scale estimate, often resulting in non-rejection of any observations in a data set. For example, the method of Motulsky and Brown (2006) with their estimator $\tilde{\sigma}$ did not detect any observation in Error Case 2 in Figure 2 as an outlier. We adopt the constant 0.01 in the definition of α_i as recommended by Motulsky and Brown (2006).

For robust regression, we adopt M-estimators. The letter M implies that the M-estimators are obtained from equations that are in similar forms with *m*aximum likelihood estimators. The M-estimators are obtained by minimizing

$$\mathcal{L}_\rho(\theta|x, y) = \frac{1}{n} \sum_{i=1}^{n} \rho(r_i),$$

where $\rho : \mathbb{R} \to \mathbb{R}_+$ is a differentiable function. See Huber and Ronchetti (2009) for detailed discussion of the M-estimators. Note that the loss function in Equation (3) is a special case of the loss function $\mathcal{L}_\rho$ when $\rho(r) = r^2$. Various forms of the functions ρ for the M-estimation suggested in the literature are as follows,

$$\text{Absolute} : \rho_1(r) = |r|,$$

$$\text{Huber} : \rho_H(r) = \begin{cases} r^2, & |r| \leq A_H, \\ 2A_H|r| - A_H^2, & |r| > A_H, \end{cases}$$

$$\text{Tukey's biweight} : \rho_T(r) = \begin{cases} r^6/A_T^4 - 3r^4/A_T^2 + 3r^2, & |r| \leq A_T, \\ A_T^2, & |r| > A_T, \end{cases} \tag{13}$$

where $A_H = 1.345$ was suggested by Huber (1964) and $A_T = 4.685$ was suggested by Mosteller and Tukey (1977). The constant A_T was shown in Mosteller and Tukey (1977) to possess 95% efficiency

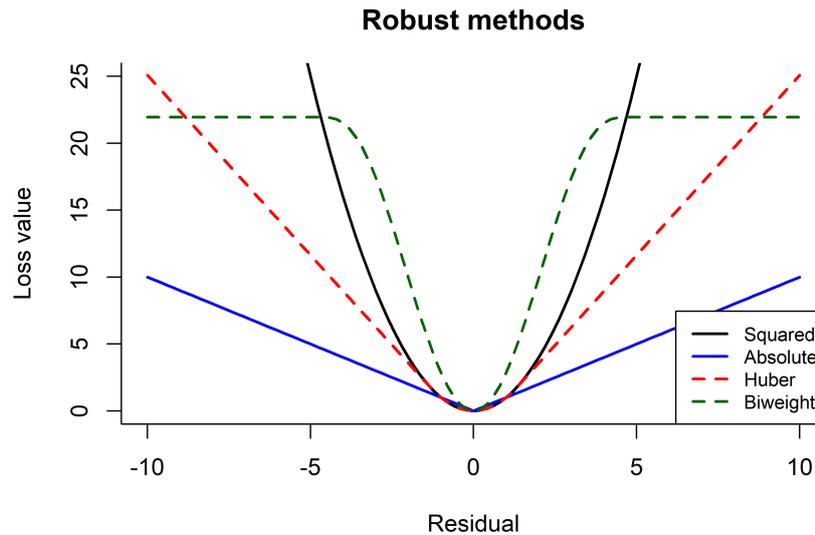**Figure 5:** The functions ρ for different M-estimators. The x-axis represents residual values while the y-axis represents the values of the function ρ.

when errors follow a normal distribution, while guaranteeing resistance to contamination of up to 10% of outliers in terms of the breakdown point. These definitions are valid when the standard deviation of the residuals r_i is 1, otherwise a robust estimator of the scale of residuals should divide the residuals.

Figure 5 shows the shape of the ρ functions corresponding to different M-estimators. The x-axis represents residual values while the y-axis represents the values of the function ρ. Note from the figure that the Huber loss and squared loss coincide with each other for residual values less than A_H, but they deviate from each other as the value of the residual r grows larger. The linearity of Huber's loss above A_H prevents a regression model from being heavily affected by outliers that have large absolute values of residuals. Tukey's biweight loss can be thought of as being more robust since it gives constant values to all residuals whose absolute values exceed the cutoff A_T. This implies that a strong outlier with a large absolute value will have the same effect on a regression fit with an observation with an absolute valued residual equal to A_T.

Using M-estimators in nonlinear regression has a few problems. First of all, the conditions for M-estimators to have asymptotic normality are complicated and difficult to be verified as mentioned in Section 12.3 of Seber and Wild (1989). This prevents us from obtaining approximate confidence intervals of true parameters based on M-estimators and their standard errors. On top of that, there are no known general results about consistency of M-estimators in nonlinear regression, which hinders us from trusting fitted parameter estimates by M-estimators. However, the absolute loss function was shown in Oberhofer (1982) to exhibit weak consistency in a nonlinear regression setting. Hence, we set the absolute loss estimator to be a default robust estimation method in our R package when the Hill bounds are attained.

Figure 6 shows the results of applying the R package **dr4pl** to the error cases of Figure 2. Unlike the R packages **drc** and **nplr**, the package **dr4pl** succeeded in fitting a robust regression model to all four data sets. The x-axis represents doses in a $log\,10$ scale and the y-axis represents responses. The black solid line shows a fit obtained by the absolute loss function and red triangles show the outliers detected by the method described in this section. Note from all the plots that the outliers that we pointed out in Section Motivation: convergence failure are marked as outliers. This implies that the method of Motulsky and Brown (2006) is successful in solving the convergence failure problem of the 4PL model.

Remedy for the support problem

The next cause of the convergence failure is the support problem whose possible solution is presented in this section. Final parameter estimates of the 4PL model obtained by an optimization procedure depend on initial estimates supplied to the procedure. The traditional R packages **drc** and **nplr** use
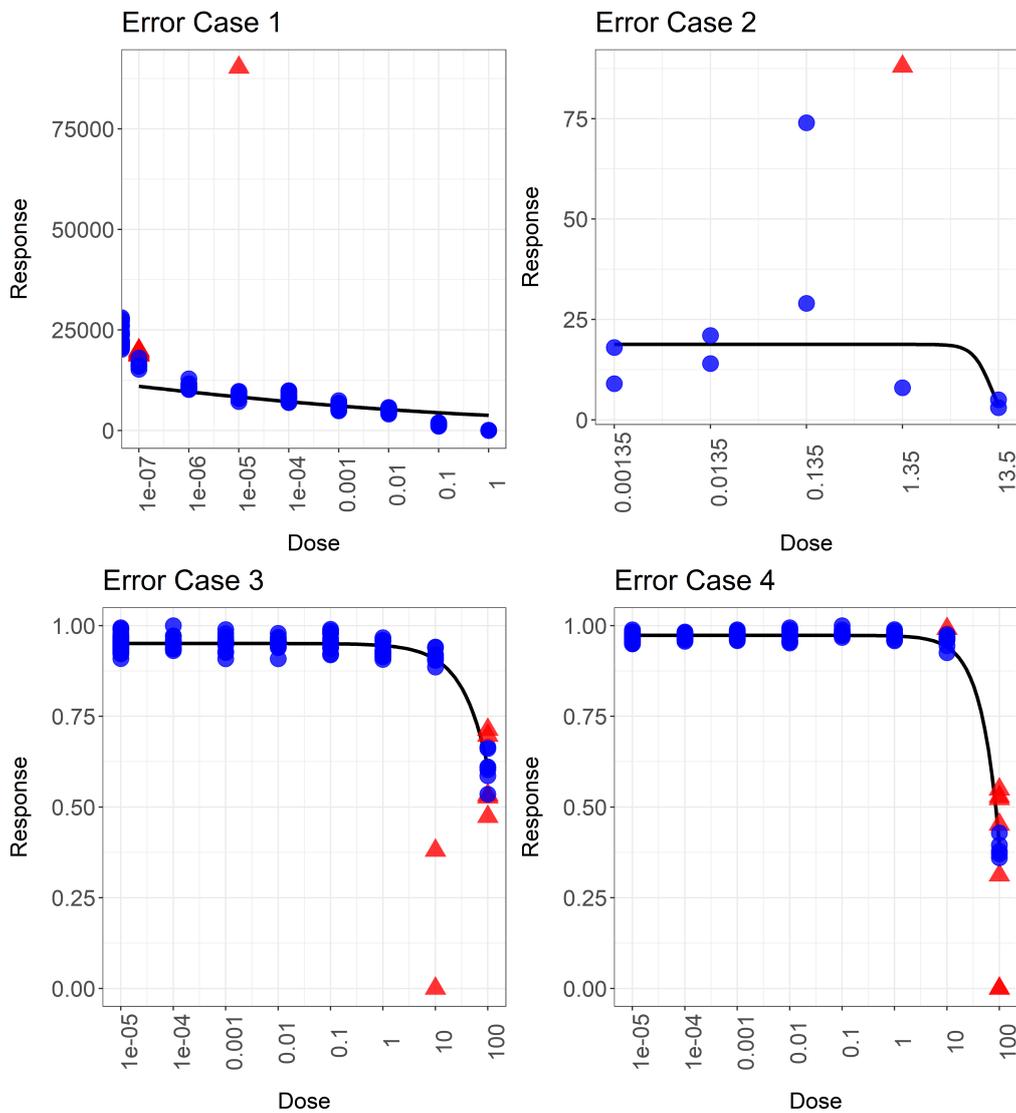
Figure 6: The dose-response curves fitted by the absolute loss function in Equation (13) denoted by black curves and outliers detected by the method of Motulsky and Brown (2006) denoted by red triangles. Note that the curves are reasonably estimated.

the *logistic method* which computes the initial parameter estimates as

$$\tilde{\theta}_1^{(L)} = y_{\max}, \tilde{\theta}_2^{(L)} = 10^{-\tilde{\beta}_2/\tilde{\beta}_3}, \tilde{\theta}_3^{(L)} = \tilde{\beta}_3, \tilde{\theta}_4^{(L)} = y_{\min}$$

where $y_{\min}, y_{\max}, \tilde{\beta}_2$ and $\tilde{\beta}_3$ are defined in Subsection Hill bounds: safeguards on the parameter space. As demonstrated in Figure 2, those initial parameter estimates do not always lead the optimization process to convergence. We present possible improvements to the logistic method and show a new initialization method that is effective in handling the support problem with data.

Logistic method combined with a grid search

The paper Wang et al. (2010) presented a grid algorithm for fitting the 4PL model to many data sets generated by a high throughput screening experiment. A main idea of their research is that fitting the Hill equation in Equation (6) with multiple initial values of the upper and lower asymptotes, θ_1 and θ_4, and starting from the best parameter estimates can yield a better result. That is, we search the grids G_1 and G_4 on θ_1 and θ_4 for the initial estimates $(\tilde{\theta}_1, \tilde{\theta}_2, \tilde{\theta}_3, \tilde{\theta}_4)$ with the least loss value where $\tilde{\theta}_1 \in G_1, \tilde{\theta}_4 \in G_4$ and the other two estimates $\tilde{\theta}_2$ and $\tilde{\theta}_3$ are obtained by the Hill equation in Equation (6). The paper suggested naive grids on θ_1 and θ_4 based on constant multiples of $y_{\max}$ and $y_{\min}$, but we suggest new grids that statistically make sense.

The basic idea of choosing grids resembles the idea of the Hill bounds in that it uses asymptotic normality of parameter estimators presented in Equation (10). Since we use $\tilde{\theta}_1^{(L)} = y_{\max}$ and $\tilde{\theta}_4^{(L)} = y_{\min}$ as initial estimators of the upper and lower asymptote parameters in the logistic method, we can obtain approximate standard errors of those estimators from Equation (12). Based on these standard errors, we set the grids on θ_1 and θ_4 as

$$\tilde{\theta}_1 \in \left\{ \tilde{\theta}_1^{(L)} - 2\text{s.e.}\left(\tilde{\theta}_1^{(L)}\right), \tilde{\theta}_1^{(L)} - \text{s.e.}\left(\tilde{\theta}_1^{(L)}\right), \tilde{\theta}_1^{(L)}, \tilde{\theta}_1^{(L)} + \text{s.e.}\left(\tilde{\theta}_1^{(L)}\right), \tilde{\theta}_1^{(L)} + 2\text{s.e.}\left(\tilde{\theta}_1^{(L)}\right) \right\},$$

$$\tilde{\theta}_4 \in \left\{ \tilde{\theta}_4^{(L)} - 2\text{s.e.}\left(\tilde{\theta}_4^{(L)}\right), \tilde{\theta}_4^{(L)} - \text{s.e.}\left(\tilde{\theta}_4^{(L)}\right), \tilde{\theta}_4^{(L)}, \tilde{\theta}_4^{(L)} + \text{s.e.}\left(\tilde{\theta}_4^{(L)}\right), \tilde{\theta}_4^{(L)} + 2\text{s.e.}\left(\tilde{\theta}_4^{(L)}\right) \right\},$$

where s.e.$\left(\tilde{\theta}_j^{(L)}\right) = s\sqrt{\tilde{c}_{j,j}}$ for $j = 1, 4$, and the two estimators s and $\tilde{c}_{j,j}$ are defined in Equations (11) and (12), respectively. If any of data observations is greater than $\tilde{\theta}_1$ or smaller than $\tilde{\theta}_4$, then those observations are removed from data and the logistic method proceeds with remaining observations.

Mead's method

An advanced method of obtaining initial parameter estimates is illustrated in Subsection 8.3.2 of Ratkowsky (1983). The method is basically a variant of the method given in Mead (1970) which used a weighted least squares criterion to obtain an approximate solution of the ordinary least squares problem. We call this *Mead's method* in this paper. Unlike the logistic method, Mead's method estimates the upper asymptote and the IC50 parameter from a linear model by first fixing the slope and lower asymptote. Since Mead's method performs a grid search on the slope parameter, it can better obtain initial parameter estimates when there are support problems like those in the lower row of Figure 2. In such cases, obtaining a proper initial estimate of the slope parameter is difficult and performing a grid search for it can be useful.

Mead's method starts from a model

$$Y = f(x|\mu)\epsilon, \tag{14}$$

where

$$f(x|\mu) = \frac{1}{\mu_1 + \mu_2\mu_3^x}, \tag{15}$$

the parameters satisfy $\mu_1, \mu_2, \mu_3 > 0$, x is a $\log 10$ dose and ϵ follows a log-normal distribution, i.e. $\log(\epsilon) \sim \mathcal{N}(0, \sigma^2)$. We call this model the *multiplicative error model*. This model is easy to work with because its parameters are related to characteristics of dose-response curves like the 4PL model. For example,

$$\text{if } 0 < \mu_3 < 1, \text{ then } \lim_{x \to -\infty} f(x|\mu) = 0, \lim_{x \to \infty} f(x|\mu) = \frac{1}{\mu_1},$$

$$\text{if } \mu_3 > 1, \text{ then } \lim_{x \to -\infty} f(x|\mu) = \frac{1}{\mu_1}, \lim_{x \to \infty} f(x|\mu) = 0.$$

This implies that $1/\mu_1$ is the upper asymptote of the model and zero is its lower asymptote. The IC50

of the multiplicative error model is more complicated than that of the 4PL model as can be seen in the following,

$$f\left(x = log_{\mu_3}\left(\frac{\mu_1}{\mu_2}\right)\bigg| \mu\right) = \frac{1}{2\mu_1}.$$

The parameters of a multiplicative error model are related to those of a 4PL model by

$$\theta_1 = \frac{1}{\mu_1}, \ \theta_2 = \left(\frac{\mu_1}{\mu_2}\right)^{1/log_{10}\mu_3}, \ \theta_3 = log_{10}\mu_3, \ \theta_4 = 0.$$

It seems that the 4PL model has been preferred due to these more intuitive meanings of the parameters. This equation shows that the parameter μ_3 is related to the slope of a dose-response curve.

To find estimators of μ_1, μ_2, μ_3 in Equations (14) and (15), we consider a weighted least squares loss function

$$\mathcal{L}(\mu|x,y) = \sum_{i=1}^{n} \frac{\sigma^2}{Var(Y_i)} \{y_i - E(Y_i)\}^2. \tag{16}$$

For a differentiable function $g : \mathbb{R} \to \mathbb{R}$ we have

$$g(Y_i) - g(E(Y_i)) \approx \frac{dg(y)}{dy}\bigg|_{y=E(Y_i)} \{Y_i - E(Y_i)\}.$$

Hence, we can approximate $Var(g(Y_i))$ by

$$Var(g(Y_i)) \approx \frac{dg(y)}{dy}\bigg|_{y=E(Y_i)}^2 Var(Y_i).$$

If we let $g(y) = log(y)$, then we have

$$Var(Y_i) \approx Var(logY_i) \{E(Y_i)\}^2 = \sigma^2 \{E(Y_i)\}^2.$$

Substituting this into Equation (16) yields

$$\mathcal{L}(\mu|x,y) \approx \frac{1}{n} \sum_{i=1}^{n} \left\{\frac{y_i}{E(Y_i)} - 1\right\}^2.$$

Substituting the expression $E(Y_i) = f(x_i|\mu)$ into this equation yields an approximate linear model

$$\mathcal{L}(\mu|x,y) \approx \frac{1}{n} \sum_{i=1}^{n} \left\{y_i\left(\mu_1 + \mu_2\mu_3^{x_i}\right) - 1\right\}^2. \tag{17}$$

To obtain estimates of μ_1 and μ_2, we take $\mu_3 \in G$ with the grid

$$G = \{q_{0.04}, q_{0.08}, \cdots, q_{0.92}, q_{0.96}\}$$

where q_p is the p-th quantile of the Cauchy distribution. The Cauchy distribution was chosen because it can provide a broad range of quantile values due to its heavy tailedness. For each $\mu_3 \in G$, the loss function $\mathcal{L}(\mu|x,y)$ can be minimized over μ_1 and μ_2 by the standard linear regression. That is, a linear regression model with the response 1, the first predictor y_i and the second predictor $y_i\mu_3^{x_i}$ without an intercept can be fitted to obtain parameter estimates. Note that this regression model is fitted for all $\mu_3 \in G$ and parameter estimates $\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3$ with the smallest value of $\mathcal{L}(\cdot|x,y)$ are chosen as initial parameter estimates. A nonlinear optimization method starts from these estimates to find final parameter estimates.

Comparison of the logistic and Mead methods

As pointed out in the previous subsection, Mead's method is expected to find better initial parameter estimates than the logistic method when data have the support problem. By simulating data and fitting 4PL models with different initialization techniques, we can see whether the Mead method actually results in final fitted models with lower loss values than the logistic method. We also compare the two methods when data do not exhibit a support problem. By doing that, we can get insight into their relative strengths under various situations.

In this simulation study, we generate data from the models of Subsection Analysis of the Hill bounds whose mean response curves are presented in the left plot of Figure 4 and whose parameter values are given in Table 2. As mentioned in that subsection, Models 1, 2 and 3 represent models without any support problem, while Models 7, 8 and 9 represent severe support problems. Models 4, 5

Parameters	$\theta_3 = -0.5$	$\theta_3 = -1$	$\theta_3 = -1.5$
$\theta_2 = 0.01$	442	443	697
$\theta_2 = 0.1$	651	568	682
$\theta_2 = 1$	661	767	843

Table 4: The numbers of times that the Mead method outperforms the logistic method in the sense that it has a lower loss value for final fitted models. The bottom row ($\theta_2 = 1$) represents data with a support problem. It can be seen that the superiority of Mead's method is maximal when data have a support problem.

and 6 exhibit mild levels of support problems. For each model, 1,000 data sets are generated and the standard deviation of normally distributed errors was fixed at 5.

Table 4 shows the numbers of times that the Mead method outperforms the logistic method in the sense that it has a lower loss value for final fitted models out of 1,000 simulations. The top row of the table represents Models 1, 2 and 3 from left to right. As can be seen from the top row, the Mead method generally performs worse than the logistic method except when the slope is very steep ($\theta_3 = -1.5$). In the middle and bottom rows, the Mead method performs better than the logistic method. The degree of outperformance becomes larger when the support problem is more severe ($\theta_2 = 1$) or the slope parameter is more negative ($\theta_3 = -1.5$). This coincides with our conjecture made in the previous subsection that Mead's method can be particularly useful when data do not sufficiently exist on the right side of the dose-response curves so that estimating the slope and lower asymptote parameters is difficult. In addition, the degree of inferiority of the Mead method to the logistic method when there is no support problem (upper left cell) is not severe, which supports our use of the Mead method as the default initialization method in **dr4pl**.

Walk-through in R

In this section, we walk through example R code in which dose-response data are analyzed using our R package **dr4pl**. By following steps shown below, users can understand how to use **dr4pl** to handle data sets that have the convergence failure issue and what kind of detailed analysis can be done by setting options of R functions in **dr4pl**.

Main R function dr4pl

We start with Error Case 1 shown in Figures 2 and 6. As can be seen in those figures, this data set contains an extreme outlier at the dose level 10^{-5}. If the R package **drc** is used to fit the 4PL model to this data set, the following error message is returned.

```
> library(drc)
>
> drm(Response~Dose, data = drc_error_1, fct = LL.4())
Error in optim(startVec, opfct, hessian = TRUE, method = optMethod, control =
  list(maxit = maxIt,  :  non-finite finite-difference value [4]
Error in drmOpt(opfct, opdfct1, startVecSc, optMethod, constrained, warnVal,  :
  Convergence failed
```

Note that this error message indicates convergence failure but no explanation about its cause is given. On the contrary, fitting the 4PL model with **dr4pl** returns an object of the class "dr4pl" which contains useful information about the convergence failure and a possible solution.

```
> library(dr4pl)
>
> dr4pl.error.1 <- dr4pl(Response~Dose, data = drc_error_1, method.init = "logistic",
+                   use.Hessian = TRUE)
> class(dr4pl.error.1)
[1] "dr4pl"
> dr4pl.error.1$convergence  # TRUE indicates convergence success,
+                   # FALSE indicates convergence failure.
[1] FALSE
```

In this code, some input arguments such as `method.init` and `use.Hessian` are specified as different values from the default values of `dr4pl`. For example, the default value of `method.init` is Mead but it is specified as `logistic` in the code. This is to ensure that the function `dr4pl` works with the same parameter setting as **drc**. The R package **drc** uses the logistic method as its default initialization method and the Hessian matrix in its confidence interval estimation. Note that the function `dr4pl` does not throw an error but returns the member `dr4pl.error.1$convergence` which indicates that convergence failure happens during the optimization process.

As a possible solution to the convergence failure, the object `dr4pl.error.1` provides a member object `dr4pl.robust` that results from robust estimation and the outlier detection algorithm explained in Subsection Robust regression and outlier detection. We call the object `dr4pl.error.1$dr4pl.robust` the *robust child object* of its *parent object* `dr4pl.error.1`.

```
> dr4pl.robust.1 <- dr4pl.error.1$dr4pl.robust
> class(dr4pl.error.1$dr4pl.robust)
[1] "dr4pl"
> dr4pl.robust.1$convergence
[1] TRUE
```

It can be seen from these code lines that the robust child object does not report the convergence failure issue. To check results of robust estimation and outlier detection, we check the plot of the object `dr4pl.robust.1`.

```
> dr4pl.robust.1$idx.outlier
 [1] 132 140 127 134 133 137 141 139 130 129 128 102
> plot(dr4pl.robust.1, indices.outlier = dr4pl.robust.1$idx.outlier)
```

The member `idx.outlier` of the object `dr4pl.robust.1` shows the indices of the outliers in the input data frame `drc_error_1` detected by our algorithm. It can be seen that the last sentence in this code generates the same plot as the upper left plot of Figure 6. As mentioned in Subsection Robust regression and outlier detection, the absolute loss is given as the default robust estimation method. This implies that adopting the absolute loss successfully resolves the convergence failure problem of Error Case 1.

Our R package **dr4pl** provides other real-world data sets than Error Cases 1 - 4. We analyze the data set `sample_data_5` to compare the two different initialization methods, the logistic and Mead methods, which were presented in Section Remedy for the support problem.

```
> dr4pl.logistic.5 <- dr4pl(Response~Dose,
+                           data = sample_data_5,
+                           method.init = "logistic")
> ggplot.logistic <- plot(dr4pl.logistic.5, text.title = "Logistic method")
>
> dr4pl.Mead.5 <- dr4pl(Response~Dose,
+                       data = sample_data_5)
> ggplot.Mead <- plot(dr4pl.Mead.5, text.title = "Mead's method")
>
> grid.arrange(ggplot.logistic, ggplot.Mead, nrow = 1, ncol = 2)
```

Running this code generates the two plots in Figure 7. Note that outliers are not denoted by red triangles in these plots since they are generated by the parent objects but not by their robust child objects. It can be seen from the figure that the data set has the support problem in the sense that it is hard to determine an IC50 parameter value visually and data points on the right side of possible IC50 parameter values are not sufficiently provided. Mead's method shown in the right plot seems to yield a closer fit to the right-most data point. This coincides with our explanation in Subsection Mead's method that the method can result in better final parameter estimates particularly when data suffer from the support problem.

Auxiliary R functions

The R function summary of the R package **dr4pl** shows the 95% asymptotic confidence intervals of the final parameter estimates of a "dr4pl" object. The following code lines show example output of summary applied to the data set `sample_data_3`.

```
> dr4pl.3 <- dr4pl(Response~Dose, data = sample_data_3)
> summary(dr4pl.3)
$call
```
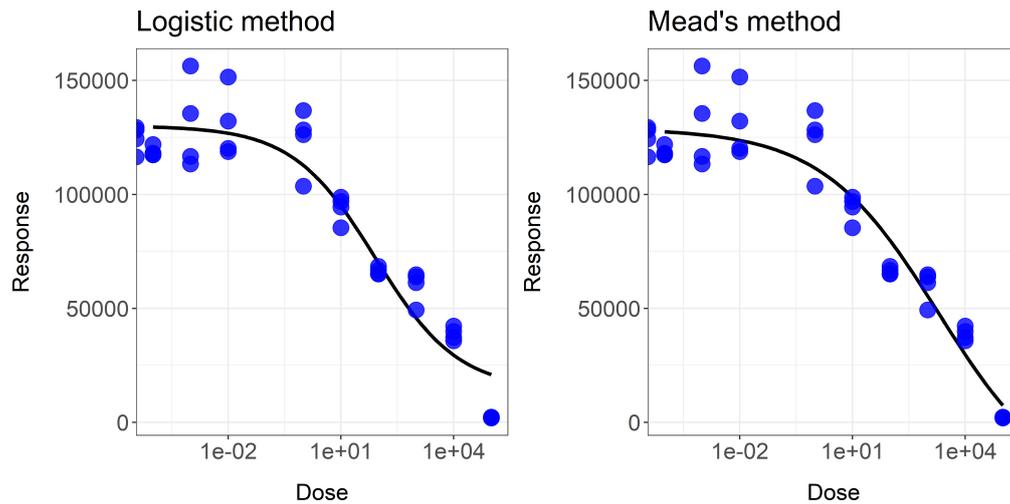
Figure 7: Dose-response curves fitted by **dr4pl** with two initialization methods: the logistic (left plot) and Mead (right plot) method. It can be seen that Mead's method yields a closer fit to the right-most data point than the logistic method.

```
dr4pl.formula(formula = Response ~ Dose, data = sample_data_3)

$coefficients
                 Estimate        StdErr          2.5 %         97.5 %
UpperLimit 59858.0700194 1.328219e-01 58678.1496415 61037.9903974
IC50           5.0794918 1.531608e-05     3.7133017      6.9483278
Slope         -0.6117371 1.440712e-05    -0.7397224     -0.4837518
LowerLimit   980.8205686 1.354448e-01  -222.4002527   2184.0413899

attr(,"class")
[1] "summary.dr4pl"
```

Recall from Subsection 2.2.1 that the IC50 is the dose level whose response value is halfway between the upper and lower asymptotes θ_1 and θ_4. Similarly, other inhibitory concentrations of interest can be defined according to different proportions between θ_1 and θ_4 as ι_p which satisfies the following equation,

$$f(\iota_p) = p\theta_1 + (1-p)\theta_4$$

where $0 \le p \le 1$. It can be easily seen that $\iota_{0.5}$ corresponds to the IC50 parameter θ_2.

Like the R function ED of **drc** and the R function getEstimates of **nplr**, the R package **dr4pl** implements the function IC to compute inhibitory concentration parameters ι_p for $0 \le p \le 1$. The following code shows a simple use of IC to compute $\iota_{0.1}, \iota_{0.3}, \iota_{0.5}, \iota_{0.7}, \iota_{0.9}$ for the data sample_data_3.

```
> IC(dr4pl.3, c(10, 30, 50, 70, 90))
InhibPercent:10 InhibPercent:30 InhibPercent:50 InhibPercent:70 InhibPercent:90
    184.3784198      20.2930770       5.0794918       1.2714305       0.1399363
```

Summary and future extension

The 4PL model has often suffered from convergence failure problems preventing experimenters from obtaining fitted 4PL models to data sets. Without diagnosing problems with data sets, experimenters should often try many different initial parameter estimates, which still does not guarantee stable convergence of optimization processes. In this paper, we presented a novel method based on studying convergence failure problems. By imposing proper conditions on the parameter space of the 4PL models, our new R package **dr4pl** reports convergence failure during an optimization process and presents possible resolutions to users.

Another important issue can be memory usage and speed of **dr4pl**, relative to other R packages **nplr** and **drc**. Measuring time and memory consumed by the three R packages for the 13 data sets provided in **dr4pl** showed that **nplr** is much slower, and used much more memory, than either of the

other two methods. The enhanced features of **dr4pl** resulted in its being only slightly slower, with somewhat more memory usage than **drc**.

Asymptotic confidence intervals of M-estimators have been developed in Lim et al. (2013). An interesting open problem is the numerical implementation of these in the 4PL context. Another area of potential future extension is modelling interaction between different drugs. The R packages **drc** and **MixLow** support functionality for drug interaction modelling. While we are not aware of data sets reported in the literature that cause the convergence failure problem when there are more than one drug, it seems that outliers or the support problem may cause instability of optimization processes in those cases as well. Analysis of the Hill bounds can be a remedy for those, which is an important area of future research.

Appendix: proof of Theorem 2.3.1

Let $\Theta = \{ \theta = (\theta_1, \theta_2, \theta_3, \theta_4) \,|\, \theta_1, \theta_3, \theta_4 \in \mathbb{R}, \theta_2 > 0 \}$. From Equation (3), it can be seen that

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta_j} = -\frac{2}{n} \sum_{i=1}^{n} \{ y_i - f(x_i|\theta) \} \frac{\partial f(x_i|\theta)}{\partial \theta_j},\qquad(18)$$

for $j = 1, 2, 3, 4$. In addition, we have from Equation (2) that

$$\frac{\partial f(x|\theta)}{\partial \theta_1} = 1 - \frac{1}{1 + (10^x/\theta_2)^{\theta_3}} > 0,$$
$$\frac{\partial f(x|\theta)}{\partial \theta_4} = \frac{1}{1 + (10^x/\theta_2)^{\theta_3}} > 0,\qquad(19)$$

for all $x \in \mathbb{R}$ and $\theta \in \Theta$ since we have $\theta_2 > 0$. From these equations, it can be seen that the function f is increasing in θ_1 and θ_4 and

$$\lim_{\theta_1 \to \infty} f(x|\theta) = \infty, \quad \lim_{\theta_4 \to -\infty} f(x|\theta) = -\infty.$$

for all $x \in \mathbb{R}$ and $\theta \in \Theta$. Fix $\theta_2 > 0, \theta_3 \in \mathbb{R}$. Then there exist two constants $-\infty < \theta_1', \theta_4' < \infty$ such that

$$f(x_i|\theta) > y_i \,\forall \theta_1 \geq \theta_1', \; f(x_i|\theta) < y_i \,\forall \theta_4 \leq \theta_4',\qquad(20)$$

for fixed $\theta_4 \in \mathbb{R}$ and $\theta_1 \in \mathbb{R}$, respectively, for $i = 1, 2, \cdots, n$. Substituting Equations (19) and (20) into Equation (18) yields

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta_1} > 0 \,\forall \theta_1 \geq \theta_1', \; \frac{\partial \mathcal{L}(\theta)}{\partial \theta_4} < 0 \,\forall \theta_4 \leq \theta_4',$$

for those fixed $\theta_4 \in \mathbb{R}$ and $\theta_1 \in \mathbb{R}$, respectively.

Bibliography

D. M. Bates and D. G. Watts. Nonlinear regression analysis and its applications. *Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons*, 1988. ISBN 0-471-81643-4. URL https://doi.org/10.1002/9780470316757. [p172]

Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, 29(4):1165–1188, 2001. ISSN 0090-5364. [p180]

F. Commo and B. M. Bot. nplr: n-parameter logistic regression, 2016. URL https://CRAN.R-project.org/package=nplr. R package version 0.1-7. [p172]

R. B. Dean and W. J. Dixon. Simplified statistics for small numbers of observations. *Analytical Chemistry*, 23(4):636–638, 1951. URL https://doi.org/10.1021/ac60052a025. [p179]

Y. Di Veroli Giovanni, C. Fornari, I. Goldlust, G. Mills, S. B. Koh, J. L. Bramhall, F. M. Richards, and D. I. Jodrell. An automated fitting procedure and software for dose-response curves with multiphasic features. *Scientific Reports (Nature Publisher Group)*, 5:14701, 2015. URL https://doi.org/10.1038/srep14701. [p173]

S. R. Gadagkar and G. B. Call. Computational tools for fitting the Hill equation to dose-response curves. *J Pharmacol Toxicol Methods*, 71:68–76, 2015. URL https://doi.org/10.1016/j.vascn.2014.08.006. [p175]

J. Giraldo, N. M. Vivas, E. Vila, and A. Badia. Assessing the (a) symmetry of concentration-effect curves: empirical versus mechanistic models. *Pharmacology & Therapeutics*, 95(1):21–45, 2002. URL https://doi.org/10.1016/s0163-7258(02)00223-1. [p172]

A. V. Hill. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J Physiol (Lond)*, 40:4–7, 1910. [p172]

P. J. Huber. Robust estimation of a location parameter. *Ann. Math. Statist.*, 35:73–101, 1964. ISSN 0003-4851. URL http://dx.doi.org.libproxy.lib.unc.edu/10.1214/aoms/1177703732. [p180]

P. J. Huber and E. M. Ronchetti. Robust statistics. *Wiley Series in Probability and Statistics. John Wiley & Sons, 2nd edition*, 2009. ISBN 978-0-470-12990-6. URL http://dx.doi.org.libproxy.lib.unc.edu/10.1002/9780470434697. [p180]

J. T. Landis, H. An, and A. G. Bailey. dr4pl: dose response data analysis using the 4 Parameter Logistic (4PL) model, 2018. URL https://bitbucket.org/dittmerlab/dr4pl. R package version 1.1.8. [p171]

C. Lim. Robust ridge regression estimators for nonlinear models with applications to high throughput screening assay data. *Stat. Med.*, 34(7):1185–1198, 2015. ISSN 0277-6715. URL https://doi.org/10.1002/sim.6391. [p173]

C. Lim, P. K. Sen, and S. D. Peddada. Robust nonlinear regression in applications. *J. Indian Soc. Agricultural Statist.*, 67(2):215–234, 291, 2013. ISSN 0019-6363. [p188]

R. Mead. Plant density and crop yield. *Applied statistics*, pages 64–81, 1970. URL https://doi.org/10.2307/2346843. [p183]

F. Mosteller and J. W. Tukey. Data analysis and regression: a second course in statistics. *Addison-Wesley Series in Behavioral Science: Quantitative Methods*, 1977. [p180]

H. J. Motulsky and R. E. Brown. Detecting outliers when fitting data with nonlinear regression - a new method based on robust nonlinear regression and the false discovery rate. *BMC Bioinformatics*, 7: 123, 2006. URL https://doi.org/10.1186/1471-2105-7-123. [p179, 180, 181, 182]

W. Oberhofer. The consistency of nonlinear regression minimizing the L_1-norm. *Ann. Statist.*, 10(1): 316–319, 1982. ISSN 0090-5364. [p181]

W. B. Pratt and P. Taylor. Principles of drug action: the basis of pharmacology. *Churchill Livingstone, 3rd edition*, 1990. URL https://doi.org/10.1021/jm00296a900. [p175]

D. A. Ratkowsky. Nonlinear regression modeling. A unified practical approach. *Marcel Dekker, Inc., New York*, 1983. [p183]

A. H. Riazoshams, B. M. Habshah Jr, and A. Adam. On the outlier detection in nonlinear regression. *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering*, 3 (12):1105–1111, 2009. [p179]

C. Ritz, J. C. Streibig, and others. Bioassay analysis using R. *Journal of Statistical Software*, 12(5):1–22, 2005. URL https://doi.org/10.18637/jss.v012.i05. [p172]

T. B. Robertson. On the normal rate of growth of an individual, and its biochemical significance. *Archiv für Entwicklungsmechanik der Organismen*, 25(4):581–614, 1908. URL https://doi.org/10.1007/bf02163864. [p171]

G. A. F. Seber and C. J. Wild. Nonlinear regression. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons*, 1989. ISBN 0-471-61760-1. URL https://doi.org/10.1002/0471725315. [p172, 177, 181]

Y. Wang, A. Jadhav, N. Southal, R. Huang, and D. T. Nguyen. A grid algorithm for high throughput fitting of dose-response curve data. *Curr Chem Genomics*, 4:57–66, 2010. URL https://doi.org/10.2174/1875397301004010057. [p173, 183]

Hyowon An
Department of Statistics and Operations Research
The University of North Carolina at Chapel Hill
USA
ahwbest@gmail.com

Justin T. Landis
Lineberger Comprehensive Cancer Center
The University of North Carolina at Chapel Hill
USA
jtlandis314@gmail.com

Aubrey G. Bailey
Lineberger Comprehensive Cancer Center
The University of North Carolina at Chapel Hill
USA
aubreybailey@gmail.com

J. S. Marron
Department of Statistics and Operations Research
The University of North Carolina at Chapel Hill
USA
marron@unc.edu

Dirk P. Dittmer
Lineberger Comprehensive Cancer Center
The University of North Carolina at Chapel Hill
USA
dirkdittmer@me.com