The \texttt{bbmle} package, designed to simplify maximum likelihood estimation and analysis in R, extends and modifies the \texttt{mle} function and class in the \texttt{stats4} package that comes with R by default. \texttt{mle} is in turn a wrapper around the \texttt{optim} function in base R. The maximum-likelihood-estimation function and class in \texttt{bbmle} are both called \texttt{mle2}, to avoid confusion and conflict with the original functions in the \texttt{stats4} package. The major differences between \texttt{mle} and \texttt{mle2} are:

- \texttt{mle2} is more robust, with additional warnings (e.g. if the Hessian can’t be computed by finite differences, \texttt{mle2} returns a fit with a missing Hessian rather than stopping with an error)

- \texttt{mle2} uses a \texttt{data} argument to allow different data to be passed to the negative log-likelihood function

- \texttt{mle2} has a formula interface like that of (e.g.) \texttt{gls} in the \texttt{nlme} package. For relatively simple models the formula for the maximum likelihood
can be written in-line, rather than defining a negative log-likelihood function. The formula interface also simplifies fitting models with categorical variables. Models fitted using the formula interface also have applicable `predict` and `simulate` methods.

- `bbmle` defines `anova`, `AIC`, `AICc`, and `BIC` methods for `mle2` objects, as well as `AICtab`, `BICtab`, `AICctab` functions for producing summary tables of information criteria for a set of models.

Other packages with similar functionality (extending GLMs in various ways) are

- on CRAN: `aods3` (overdispersed models such as beta-binomial); `vgam` (a wide range of models): `betareg` (beta regression); `pscl` (zero-inflated, hurdle models); `maxLik` (another general-purpose maximizer, with a different selection of optimizers)

- In Jim Lindsey’s code repository ([http://popgen.unimaas.nl/~jlindsey/rcode.html](http://popgen.unimaas.nl/~jlindsey/rcode.html)): `gnlr` and `gnlr3`

1 Example: *Orobanche*/overdispersed binomial

This example will use the classic data set on *Orobanche* germination from Crowder (1978) (you can also use `glm(...,family="quasibinomial")` or the `aods3` package to analyze these data).

1.1 Test basic fit to simulated beta-binomial data

First, generate a single beta-binomially distributed set of points as a simple test.

Load the `emdbook` package to get functions for the beta-binomial distribution (random-deviate function `rbetabinom` — these functions are also available in Jim Lindsey’s `rmutil` package).

```r
library(emdbook)

set.seed(1001)
x1 <- rbetabinom(n=1000,prob=0.1,size=50,theta=10)

load the package:

library(bbmle)

Construct a simple negative log-likelihood function:
mtmp <- function(prob, size, theta) {
  -sum(dbetabinom(x1, prob, size, theta, log=TRUE))
}

Fit the model — use data to pass the size parameter (since it wasn’t hard-coded in the mtmp function):

(m0 <- mle2(mtmp, start=list(prob=0.2, theta=9), data=list(size=50)))

## Call:
## mle2(minuslogl = mtmp, start = list(prob = 0.2, theta = 9), data = list(size = 50))
## Coefficients:
## prob theta
## 0.1030974 10.0758090
## Log-likelihood: -2723.5

(here and below, I’m suppressing lots of warnings about NaNs produced)

The summary method for mle2 objects shows the parameters; approximate standard errors (based on quadratic approximation to the curvature at the maximum likelihood estimate); and a test of the parameter difference from zero based on this standard error and an assumption that the likelihood surface is quadratic (or equivalently that the sampling distribution of the estimated parameters is normal).

summary(m0)

## Maximum likelihood estimation
## Call:
## mle2(minuslogl = mtmp, start = list(prob = 0.2, theta = 9), data = list(size = 50))
## Coefficients:
## Estimate Std. Error z value Pr(z)
## prob 0.1030974 0.0031626 32.599 < 2.2e-16 ***
## theta 10.0758090 0.6213319 16.216 < 2.2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## -2 log L: 5446.995

Construct the likelihood profile (you can apply confint directly to m0, but if you’re going to work with the likelihood profile [e.g. plotting, or looking for
confidence intervals at several different $\alpha$ values] then it is more efficient to compute the profile once):

```r
p0 <- profile(m0)
```

Compare the confidence interval estimates based on inverting a spline fit to the profile (the default); based on the quadratic approximation at the maximum likelihood estimate; and based on root-finding to find the exact point where the profile crosses the critical level.

```r
confint(p0)
## 2.5 % 97.5 %
## prob 0.09709228 0.1095103
## theta 8.91708205 11.3559592

confint(m0, method = "quad")
## 2.5 % 97.5 %
## prob 0.09689875 0.1092961
## theta 8.85802088 11.2935972

confint(m0, method = "uniroot")
## 2.5 % 97.5 %
## prob 0.09709908 0.1095033
## theta 8.91691019 11.3559746
```

All three types of confidence limits are similar.

Plot the profiles:

```r
par(mfrow = c(1, 2))
plot(p0, plot.confstr = TRUE)
```

---

Note: The plots are not rendered in this text format, but they would show the likelihood profiles for `prob` and `theta` with confidence intervals at different levels.
By default, the plot method for likelihood profiles displays the square root of
the deviance difference (twice the difference in negative log-likelihood from
the best fit), so it will be V-shaped for cases where the quadratic approximation
works well (as in this case). (For a better visual estimate of whether the profile
is quadratic, use the absVal=FALSE option to the plot method.)

You can also request confidence intervals calculated using unirho, which
may be more exact when the profile is not smooth enough to be modeled ac-
curately by a spline. However, this method is also more sensitive to numeric
problems.

Instead of defining an explicit function for minuslogl, we can also use the
formula interface. The formula interface assumes that the density function
given (1) has x as its first argument (if the distribution is multivariate, then x
should be a matrix of observations) and (2) has a log argument that will return
the log-probability or log-probability density if log=TRUE. Some of the extended
functionality (prediction etc.) depends on the existence of an s- variant function
for the distribution that returns (at least) the mean and median as a function
of the parameters (currently defined: snorm, sbinom, sbeta, snbinom, spois).

```
m0f <- mle2(x1~dbetabinom(prob,size=50,theta),
    start=list(prob=0.2,theta=9),data=data.frame(x1))
```

Note that you must specify the data via the data argument when using the
formula interface. This may be slightly more unwieldy than just pulling the
data from your workspace when you are doing simple things, but in the long
run it makes tasks like predicting new responses much simpler.

It’s convenient to use the formula interface to try out likelihood estimation
on the transformed parameters:

```
m0cf <- mle2(x1~dbetabinom(prob=plogis(lprob),size=50,theta=exp(ltheta)),
    start=list(lprob=0,ltheta=2),data=data.frame(x1))
```

```
confint(m0cf,method="uniroot")
## 2.5 % 97.5 %
## lprob -2.229960 -2.095759
## ltheta 2.187954 2.429738
```

```
confint(m0cf,method="spline")
## 2.5 % 97.5 %
## lprob -2.229963 -2.095756
## ltheta 2.187948 2.429742
```

In this case the answers from unirho and spline (default) methods barely
differ.
1.2 Real data (*Orobanche*, Crowder (1978))

Data are copied from the *aods3* package (but a copy is saved with the package to avoid depending on the *aods3* package):

```r
load(system.file("vignetteData","orob1.rda",package="bbmle"))
summary(orob1)
```

```
## dilution  n   m
## 1/1 :6  Min. : 7.00 Min. : 0.00
## 1/25:5 1st Qu.:17.50 1st Qu.: 8.00
## 1/625:5 Median :47.50 Median :13.50
##      Mean :44.25 Mean :27.19
##      3rd Qu.:57.50 3rd Qu.:46.25
##      Max. :104.00 Max. :90.00
```

Now construct a negative log-likelihood function that differentiates among groups:

```r
ML1 <- function(prob1,prob2,prob3,theta,x)
{
    prob <- c(prob1,prob2,prob3)[as.numeric(x$dilution)]
    size <- x$n
    -sum(dbetabinom(x$m,prob,size,theta,log=TRUE))
}
```

Results from Crowder (1978):

<table>
<thead>
<tr>
<th>model</th>
<th>prop</th>
<th>diffs</th>
<th>theta</th>
<th>sd.prob1</th>
<th>sd.prob2</th>
<th>sd.prob3</th>
<th>NLL</th>
</tr>
</thead>
<tbody>
<tr>
<td>prop diffs</td>
<td>0.132</td>
<td>0.871</td>
<td>0.839</td>
<td>78.424</td>
<td>0.027</td>
<td>0.028</td>
<td>−34.991</td>
</tr>
<tr>
<td>full model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−34.829</td>
</tr>
<tr>
<td>homog model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−56.258</td>
</tr>
</tbody>
</table>

```r
(m1 <- mle2(ML1,start=list(prob1=0.5,prob2=0.5,prob3=0.5,theta=1),
data=list(x=orob1)))
```

```r
##
## Call:
## mle2(minuslogl = ML1, start = list(prob1 = 0.5, prob2 = 0.5, prob3 = 0.5, theta = 1),
##     data = list(x = orob1))
##
## Coefficients:
## prob1   prob2   prob3   theta
## 0.1318261 0.8706219 0.8382650 73.7195747
##
## Log-likelihood: -34.99
##
## Warning: optimization did not converge (code 1: )
```
Or:

```r
## would prefer ~dilution-1, but problems with starting values ...
(m1B <- mle2(m~dbetabinom(prob,size=n,theta),
  param=list(prob~dilution),
  start=list(prob=0.5,theta=1),
  data=orob1))
```

The result warns us that the optimization has not converged; we also don’t match Crowder’s results for \( \theta \) exactly. We can fix both of these problems by setting `parscale` appropriately.

Since we don’t bound \( \theta \) (or below, \( \sigma \)) we get a fair number of warnings with this and the next few fitting and profiling attempts. We will ignore these for now, since the final results reached are reasonable (and match or nearly match Crowder’s values); the appropriate, careful thing to do would be either to fit on a transformed scale where all real-valued parameter values were legal, or to use `method="L-BFGS-B"` (or `method="bobyqa"` with the `optimx` package) to bound the parameters appropriately. You can also use `suppressWarnings()` if you’re sure you don’t need to know about any warnings (beware: this will suppress all warnings, those you weren’t expecting as well as those you were . . . )

```r
(m2 <- mle2(ML1,start=as.list(coef(m1)),
  control=list(parscale=coef(m1)),
  data=list(x=orob1)))
```

```r
## Call:
## mle2(minuslogl = ML1, start = as.list(coef(m1)), data = list(x = orob1),
## control = list(parscale = coef(m1)))
##
## Coefficients:
##     prob1  prob2  prob3     theta
## 0.1322123 0.8708914 0.8393195 78.4227877
##
## Log-likelihood: -34.99
```

Calculate likelihood profile (restrict the upper limit of \( \theta \), simply because it will make the picture below a little bit nicer):

```r
(p2 <- profile(m2,prof.upper=c(Inf,Inf,Inf,theta=2000)))
```

Get the curvature-based parameter standard deviations (which Crowder used rather than computing likelihood profiles):
We are slightly off Crowder’s numbers — rounding error? Crowder also defines a variance (overdispersion) parameter \( \sigma^2 = 1/(1 + \theta) \).

Using the delta method (via the `deltavar` function in the `emdbook` package) to approximate the standard deviation of \( \sigma \):

```r
sqrt(deltavar(sqrt(1/(1+theta)),meanval=coef(m2)["theta"],
    vars="theta",Sigma=vcov(m2)[4,4]))
```

Another way to fit in terms of \( \sigma \) rather than \( \theta \) is to compute \( \theta = 1/\sigma^2 - 1 \) on the fly in a formula:

```r
m2b <- mle2(m~dbetabinom(prob,size=n,theta=1/sigma^2-1),
    data=orob1, parameters=list(prob~dilution,sigma~1),
    start=list(prob=0.5,sigma=0.1))
```

As might be expected since the standard deviation of \( \sigma \) is large, the quadratic approximation is poor:

```r
ri <- rbind(confint(p2)["theta",],
    confint(m2,method="quad")["theta",])
rownames(ri) <- c("spline","quad")
ri
```

### Data

```
round(stdEr(m2),3)
```

```
## prob1 prob2 prob3 theta
## 0.028 0.029 0.032 74.238
```

### sqrt(1/(1+coef(m2)["theta"]))

```
## theta
## 0.1122089
```

```r
sqrt(1/(1+theta))
```

```
## theta
## 0.1122089
```

```r
sqrt(1/(1+theta)),meanval=coef(m2)["theta"],
    vars="theta",Sigma=vcov(m2)[4,4])
```

```
## [1] 0.05244185
```

```r
sqrt(deltavar(sqrt(1/(1+theta)),meanval=coef(m2)["theta"],
    vars="theta",Sigma=vcov(m2)[4,4]))
```

```
## sigma
## 0.052
```

p2b <- profile(m2b,prof.lower=c(-Inf,-Inf,-Inf,0))
Plot the profile:

```
plot(p2, which="theta", plot.confstr=TRUE)
```

What does the profile for $\sigma$ look like?

```
plot(p2b, which="sigma", plot.confstr=TRUE, show.points=TRUE)
```
Now fit a homogeneous model:

```r
ml0 <- function(prob, theta, x) {
  size <- x$n
  -sum(dbetabinom(x$m, prob, size, theta, log=TRUE))
}
m0 <- mle2(ml0, start=list(prob=0.5, theta=100),
          data=list(x=orob1))

The log-likelihood matches Crowder’s result:

```r
logLik(m0)
```

```
## 'log Lik.' -56.25774 (df=2)
```

It’s easier to use the formula interface to specify all three of the models fitted by Crowder (homogeneous, probabilities differing by group, probabilities and overdispersion differing by group):
\[
\begin{align*}
m0f & \leftarrow \text{mle2}(m \sim \text{dbetabinom}(\text{prob}, \text{size}=n, \text{theta}), \\
& \quad \text{parameters=list}(\text{prob} \sim 1, \text{theta} \sim 1), \\
& \quad \text{data=orob1}, \\
& \quad \text{start=list}(\text{prob}=0.5, \text{theta}=100)) \\
m2f & \leftarrow \text{update}(m0f, \\
& \quad \text{parameters=list}(\text{prob} \sim \text{dilution}, \text{theta} \sim 1), \\
& \quad \text{start=list}(\text{prob}=0.5, \text{theta}=78.424)) \\
m3f & \leftarrow \text{update}(m0f, \\
& \quad \text{parameters=list}(\text{prob} \sim \text{dilution}, \text{theta} \sim \text{dilution}), \\
& \quad \text{start=list}(\text{prob}=0.5, \text{theta}=78.424))
\end{align*}
\]

\text{anova} runs a likelihood ratio test on nested models:

\texttt{anova(m0f,m2f,m3f)}

```
## Likelihood Ratio Tests
## Model 1: m0f, m \sim \text{dbetabinom}(\text{prob}, \text{size}=n, \text{theta}): \text{prob} \sim 1, \text{theta} \sim 1
## Model 2: m2f, m \sim \text{dbetabinom}(\text{prob}, \text{size}=n, \text{theta}): \text{prob} \sim \text{dilution}, \text{theta} \sim 1
## Model 3: m3f, m \sim \text{dbetabinom}(\text{prob}, \text{size}=n, \text{theta}): \text{prob} \sim \text{dilution}, \text{theta} \sim \text{dilution}
## Tot Df Deviance Chisq Df Pr(>Chisq)
## 1 2 112.515 
## 2 4 69.981 42.534 1 5.805e-10 ***
## 3 6 69.981 0.0008 2 0.9996 
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The various \texttt{ICtab} commands produce tables of information criteria; by default the results are sorted and presented as $\Delta IC$; there are various options, including printing model weights.

\texttt{AICtab(m0f,m2f,m3f,weights=TRUE)}

```
## dAIC df weight
## m2f 0.0 4 0.88
## m3f 4.0 6 0.12
## m0f 38.5 2 <0.001
```

\texttt{BICtab(m0f,m2f,m3f,nobs=nrow(orob1),weights=TRUE)}

```
## dBIC df weight
## m2f 0.0 4 0.941
## m3f 5.5 6 0.059
## m0f 37.0 2 <0.001
```

\texttt{AICctab(m0f,m2f,m3f,nobs=nrow(orob1),weights=TRUE)}
## dAICc df weight
## m2f 0.0 4 0.9922
## m3f 9.7 6 0.0078
## m0f 35.8 2 <0.001

2 Example: reed frog size predation

Data from an experiment by Vonesh (Vonesh and Bolker, 2005)

```r
code
```

```r
2 Example: reed frog size predation

Data from an experiment by Vonesh (Vonesh and Bolker, 2005)

```r
frogdat <- data.frame(
  size=rep(c(9,12,21,25,37),each=3),
  killed=c(0,2,1,3,4,5,rep(0,4),1,rep(0,4)))
frogdat$initial <- rep(10,nrow(frogdat))

library(ggplot2)

gg1 <- ggplot(frogdat,aes(x=size,y=killed))+geom_point()+
  stat_sum(aes(size=..n..))+
  labs(size="#")+scale_x_continuous(limits=c(0,40))+
  scale_size(breaks=1:3)

m3 <- mle2(killed~dbinom(prob=c*(size/d)^g*exp(1-size/d),
  size=initial),data=frogdat,start=list(c=0.5,d=5,g=1))
pdat <- data.frame(size=1:40,initial=rep(10,40))
pdat1 <- data.frame(pdat,killed=predict(m3,newdata=pdat))

m4 <- mle2(killed~dbinom(prob=c*((size/d)*exp(1-size/d))^g,
  size=initial),data=frogdat,start=list(c=0.5,d=5,g=1))
pdat2 <- data.frame(pdat,killed=predict(m4,newdata=pdat))

gg1 + geom_line(data=pdat1,colour="red")+
  geom_line(data=pdat2,colour="blue")
```
coef(m4)

## c    d    g
## 0.4138847 13.3517574 18.2511264

prof4 <- profile(m4)

Three different ways to draw the profile:
(1) Built-in method (base graphics):

plot(prof4)
(2) Using `xyplot` from the `lattice` package:

```r
prof4_df <- as.data.frame(prof4)
library(lattice)
xyplot(abs(z)~focal|param,data=prof4_df,
        subset=abs(z)<3,
        type="b",
        xlab="",
        ylab=expression(paste(abs(z),
                      " (square root of ",Delta," deviance)")),
        scale=list(x=list(relation="free")),
        layout=c(3,1))
```
Additions/enhancements/differences from \texttt{stats4::mle}

- \texttt{anova} method
- warnings on convergence failure
- more robust to non-positive-definite Hessian; can also specify \texttt{skip.hessian} to skip Hessian computation when it is problematic
• when profiling fails because better value is found, report new values
• can take named vectors as well as lists as starting parameter vectors
• added AICc, BIC definitions, ICtab functions
• added "uniroot" and "quad" options to confint
• more options for colors and line types etc etc. The old arguments are:

```r
function (x, levels, conf = c(99, 95, 90, 80, 50)/100, nseg = 50, absVal = TRUE, ...) {}
```

The new one is:

```r
function (x, levels, which=1:p, conf = c(99, 95, 90, 80, 50)/100, nseg = 50, plot.confstr = FALSE, confstr = NULL, absVal = TRUE, add = FALSE, col.minval="green", lty.minval=2, col.conf="magenta", lty.conf=2, col.prof="blue", lty.prof=1, xlabs=nm, ylab="score", show.points=FALSE, main, xlim, ylim, ...) {}
```

which selects (by character vector or numbers) which parameters to plot; nseg does nothing (even in the old version); plot.confstr turns on the labels for the confidence levels; confstr gives the labels; add specifies whether to add the profile to an existing plot; col and lty options specify the colors and line types for horizontal and vertical lines marking the minimum and confidence vals and the profile curve; xlabs gives a vector of x labels; ylab gives the y label; show.points specifies whether to show the raw points computed.

• mle.options()

• data argument

• handling of names in argument lists

• can use alternative optimizers (nlminb, nlm, constrOptim, optimx, optimize)

• uses code from numDeriv package to compute Hessians rather than built-in optimizer code

• by default, uses MASS::ginv (generalized inverse) rather than solve to invert Hessian (more robust to positive-semidefinite Hessians . . . )

• can use vecpar=TRUE (and parnames()) to use objective functions with parameters specified as vectors (for compatibility with optim etc.)
3 Newer stuff

To do:

- use `predict`, `simulate` etc. to demonstrate different parametric bootstrap approaches to confidence and prediction intervals
  - use `predict` to get means and standard deviations, use delta method?
  - use `vcov`, assuming quadratic profiles, with `predict(..., newparams=...)`
  - prediction intervals assuming no parameter uncertainty with `simulate`
  - both together ...

4 Technical details

4.1 Profiling and confidence intervals

This section describes the algorithm for constructing profiles and confidence intervals, which is not otherwise documented anywhere except in the code. * indicates changes from the version in `stats4:::mle`

4.1.1 Estimating standard error

In order to construct the profile for a particular parameter, one needs an initial estimate of the scale over which to vary that parameter. The estimated standard error of the parameter based on the estimated curvature of the likelihood surface at the MLE is a good guess.

- if `std.err` is missing, extract the standard error from the summary coefficient table (ultimately computed from `sqrt(diag(inverse Hessian))` of the fit)

- * a user-set value of `std.err` overrides this behavior unless the value is specified as `NA` (in which case the estimate from the previous step is used)

- * if the standard error value is still `NA` (i.e. the user did not specify it and the value estimated from the Hessian is missing or `NA`) use `sqrt(1/diag(hessian))`. This represents a (fairly feeble) attempt to come up with a plausible number when the Hessian is not positive definite but still has positive diagonal entries

- if all else fails, stop and * print an error message that encourages the user to specify the values with `std.err`

There may be further tricks that would help guess the appropriate scale: for example, one could guess on the basis of a comparison between the parameter values and negative log-likelihoods at the starting and ending points of the fits. On the other hand, (a) this would take some effort and still be subject to failure
for sufficiently pathological fits and (b) there is some value to forcing the user to take explicit, manual steps to remedy such problems, as they may be signs of poorly defined or buggy log-likelihood functions.

4.1.2 Profiling

Profiling is done on the basis of a constructed function that minimizes the negative log-likelihood for a fixed value of the focal parameter and returns the signed square-root of the deviance difference from the minimum (denoted by $z$). At the MLE $z = 0$ by definition; it should never be $< 0$ unless something has gone wrong with the original fit. The LRT significance cutoffs for $z$ are equal to the usual two-tailed normal distribution cutoffs (e.g. $\pm \approx 1.96$ for 95% confidence regions).

In each direction (decreasing and increasing from the MLE for the focal parameter):

- fix the focal parameter
- adjust control parameters etc. accordingly (e.g. remove the entry for the focal parameter so that the remaining control parameters match the non-fixed parameters)

- controls on the profiling (which can be set manually, but for which there is not much guidance in the documentation):
  - $z_{\text{max}}$ Maximum $z$ to aim for. (Default: $\sqrt{\text{qchisq}(1-\alpha/2, p)}$) The default maximum $\alpha$ (type I error) is 0.01. I don’t understand this criterion. It seems to expand the size of the univariate profile to match a cutoff for the multivariate confidence region of the model. The $\chi^2$ cutoff for deviance to get the $(1 - \alpha)$ multivariate confidence region (i.e., on all $p$ of the parameters) would be $\text{qchisq}(1-\alpha, p)$ — representing a one-tailed test on the deviance. Taking the square root makes sense, since we are working with the square root of the deviance, but I don’t understand (1) why we are expanding the region to allow for the multivariate confidence region (since we are computing univariate profiles) [you could at least argue that this is conservative, making the region a little bigger than it needs to be]; (2) why we are using $1 - \alpha/2$ rather than $1 - \alpha$. For comparison, MASS::profile.glm (written by Bates and Venables in 1996, ported to R by BDR in 1998) uses $z_{\text{max}} = \sqrt{\text{qchisq}(1-\alpha, 1)}$ (this makes more sense to me ...). On the other hand, the profiling code in lme4a (the profile method for merMod, in profile.R) uses $\text{qchisq}(1-\alpha_{\text{max}}, n_{\text{tot}})$...

- $\text{del}$ Step size (scaled by standard error) (Default: $z_{\text{max}}/5$.) Presumably (?) copied from MASS::profile.glm, which says (in ?profile.glm): “[d]efault value chosen to allow profiling at about 10 parameter values.”
maxsteps Maximum number of profiling steps to try in each direction. (Default: 100)

- While step < maxsteps and abs(z) < zmax, set the value of the focal parameter to its MLE + sgn*step*del*std.err where sgn represents the direction, step is the current (integer) step, and del and std.err are the step size scaling factor and standard error estimate discussed above (i.e. take steps of size (del*std.err) in the appropriate direction); evaluate z

- Stop the profiling:
  - if z doesn’t change from the previous step (stop_flat) — unless try_harder is TRUE
  - * stop if z is less than tol.newmin (default: 0.001) units better than the MLE fit, i.e. z < −tol.newmin (if −tol.newmin < z < 0, set z to zero) (newpars_found)
  - if z is NA (stop_na) — unless try_harder is TRUE
  - if z is beyond zmax (i.e., we have reached our goal: stop_cutoff)
  - if step == maxsteps
  - if the focal parameter has hit its upper/lower bound (stop_bound)

- If necessary (i.e. if applied to a fitted object and not to an existing profile), construct the profile

4.1.3 Confidence intervals

We are looking for the values where z (signed square root deviance difference) is equal to the usual two-tailed normal distribution cutoffs for a specified α level, e.g. z = ±1.96 for 95% confidence intervals (this is equivalent to a one-tailed test on the deviance difference with the cutoff value for χ²₁).

Spline method (default)

- If necessary (i.e. if applied to a fitted object and not to an existing profile), construct the profile
• * If the profile of the signed square root is non-monotonic, warn the user and revert to linear approximation on the profiled points to find the cutoffs:

• Otherwise, build an interpolation spline of \( z \) (signed square root deviance difference) based on profiled points (the default is \( n = 3 \times L \) where \( L \) is the length of the original vector). Then use linear approximation on the \( y \) and \( x \) (focal parameter value) of the spline to find the cutoffs (Why construct a spline and then interpolate linearly? Why not use \texttt{backSpline} as in the profile plotting code? )

Quad method Use a quadratic approximation based on the estimated curvature (this is almost identical to using \texttt{confint.default}, and perhaps obsolete/could be replaced by a pointer to \texttt{confint.default} . . . )

Unirroot For each direction (up and down):

• start by stepping \( 5 \sigma \) away from the MLE, or to the box constraint on the parameter, whichever is closer (this standard error is based on the curvature; I should allow it, or the intervals themselves, to be overridden via a \texttt{std.err} or \texttt{interval} parameter)

• compute the difference between the deviance and the desired deviance cutoff at this point; if it is \( \texttt{NA} \), reduce the distance in steps of 0.25 \( \sigma \) until it is not, until you reduce the distance to zero

• if the product of the deviance differences at the MLE and at the point you stopped at is \( \texttt{NA} \) or positive (indicating that you didn’t find a root-crossing in the range \([0, 5\sigma]\)), quit.

• otherwise, apply \texttt{uniroot} across this interval

method="\texttt{uniroot}" should give the most accurate results, especially when the profile is wonky (it won’t care about non-smooth profiles), but it will be the slowest — and different confidence levels will have to be computed individually, whereas multiple confidence levels can be computed quickly from a single computed profile. A cruder approach would be to use profiling but decrease \texttt{std.err} a lot so that the profile points were very closely spaced.

4.1.4 Profile plotting
Plot the signed (or unsigned) square root deviance difference, and \((1 - \alpha)\) confidence regions/critical values designated by \texttt{conf} (default: \{0.99, 0.95, 0.9, 0.8, 0.5\}).

• * If the (signed) profile is non-monotonic, simply plot computed points with \texttt{type="1"} (i.e., with the default linear interpolation)

• Construct the interpolation spline (using \texttt{splines::interpSpline} rather than \texttt{spline} as in the confidence interval method (why this difference? )
• attempt to construct the inverse of the interpolation spline (using `backSpline`)

• if this fails warn the user (assume this was due to non-monotonicity) and try to use `uniroot` and `predict` to find cutoff values

• otherwise, use the inverse spline to find cutoff values

_Why is there machinery in the plotting code to find confidence intervals? Shouldn’t this call `confint`, for consistency/fewer points of failure?_

**Bugs, wishes, to do**

• **WISH**: further methods and arguments: `subset`, `predict`, `resid: sim`?

• **WISH**: extend ICtab to allow DIC as well?

• minor **WISH**: better methods for extracting `nobs` information when possible (e.g. with formula interface)

• **WISH**: better documentation, especially for S4 methods

• **WISH**: variable-length (and shaped) chunks in argument list – cleaner division between linear model specs/list of arguments/vector equivalent

• **WISH**: limited automatic differentiation (add capability for common distributions)

• **WISH**: store `objectivefunction` and `objectivefunctiongr` (vectorized objective/gradient functions) in the `mle2` object (will break backward compatibility!!); add accessors for these and for `minuslogl`

• **WISH**: document use of the objective function in `MCMCpack` to do post hoc MCMC sampling (or write my own Metropolis-Hastings sampler . . .)

• **WISH**: polish profile plotting, with lattice or ggplot2 methods

• **WISH**: add in/document/demonstrate “slice” capabilities

• **WISH**: refactor profiling to use stored objective functions rather than recalling `mle2` with `fixed` values mucked around with in the calls?? Strip out and make generic for vectorized objective function? (profileModel package only works for glm-like objects, with a linear predictor)

**References**
