Calculating the likelihood

Criteria:

- Correct model formulation.
- Speed: fitting in real time.
- Ease of user specification.
- Default information displayed should not be erroneous or misleading in any context.
- Ability to handle large data sets.

Current major problems

- Interpretation of nonlinear model formulæ.
- Vectorized multidimensional integration for nonlinear parameters.

Specifying the model

Two basic parts:

1. Probability distribution:
    - list of choices for user.

2. Ways in which (location, dispersion, shape) parameters of this distribution depend on covariates:
    - regression model formulæ to be specified by user.

Probability distribution

- Need a wide choice.

- Stochastic dependence relationships:

  - Longitudinal dependence: except in special cases, requires recursive updating.

  - Clustering: random effects (multidimensional integration) or recursive updating.

For speed, recursive likelihoods (e.g. Kalman filtering) must be calculated in C.

Covariate dependence

Example: open first-order one-compartment model.

Location parameter over time as varies as

$$\mu_t \;=\; \frac{V k_a}{k_a - k_e}\left(\mathrm{e}^{-k_e t} - \mathrm{e}^{-k_a t}\right)$$

```
mu <- function(p){
  p[1]*p[2]/(p[2]-p[3])*
    (exp(-p[3]*t)-exp(-p[2]*t))}
```

Dispersion parameter depends on time through the location parameter,

$$\sigma^2 = \mu_t^\delta$$

```
disp <- function(p,mu) p[1]*log(mu)
```

But k_a, k_e, V may depend, in various ways, on other covariates.

Linear (in parameters) part

- Specified by W&R model formula.

- However, link functions only useful for a (transformed) parameter depending on covariates through a strictly linear model.

```
mu <- function(p,linear){
  tmp <- exp(linear)
  p[4]*tmp/(tmp-p[5])*(exp(-p[5]*t)-exp(-tmp*t))}
modelfn(..., mu=mu, linear=~height+gfr, ...)
```

Nonlinear specification

Ability to distinguish variables and parameters (existing vectors and unknowns) in formulæ (like Maple, Mathematica).

```
mu=~ volume*absorption/(absorption-elimination)*

  (exp(-elimination*time)-exp(-absorption*time))
```

with initial estimates

```
init=list(volume=3, absorption=0.1,

  elimination=0.05)
```

Some criteria:

- Detect existing ‘vectors’.

- Ignore conflicts of unknowns with existing (non-variable) object names.

- Can be substituted into probability distribution function to construct likelihood *function* that can be rapidly evaluated by a nonlinear optimizer (nlm).

- Function of *one* vector of unknown parameters for nlm (pointers?).

- Evaluated in parent environment (or, if necessary, higher environments according to R scoping rules) to that of the function in which specification is used.