Random effects in RTMB

Anders Nielsen & Ben Bolker

an@aqua.dtu.dk

Example: Paired observations

- Two methods A and B to measure blood cell count (to check for the use of doping).
- Paired study.

Person ID	Method A	Method B
1	5.5	5.4
2	4.4	4.9
3	4.6	4.5
4	5.4	4.9
5	7.6	7.2
6	5.9	5.5
7	6.1	6.1
8	7.8	7.5
9	6.7	6.3
10	4.7	4.2

- It must be expected that two measurements from the same person are correlated, so a paired t-test is the correct analysis
- The t-test gives a p-value of 5.1%, which is a borderline result...
- But more data is available

- In addition to the planned study 10 persons were measured with only one method
- Want to use all data, which is possible with *random effects*
- Assume these 20 are randomly selected from a population where the blood cell count is normally distributed
- Consider the following model:
  $C_i = \alpha(M_i) + B(P_i) + \varepsilon_i, \quad i = 1 \dots 30$
  $\alpha(M_i)$ the two fixed method effects
  $B(P_i) \sim \mathcal{N}(0, \sigma_P^2)$ the 20 random effects
  $\varepsilon_i \sim \mathcal{N}(0, \sigma_R^2)$ measurement noise
  All $B(P_i)$ and ε_i are assumed independent
- This model uses all data and gives a 95% CI for the method bias $\alpha(A) - \alpha(B)$ which is: $(0.05; 0.41)$.
- Notice that now there is a (slightly) significant method bias.

Person ID	Method A	Method B
1	5.5	5.4
2	4.4	4.9
3	4.6	4.5
4	5.4	4.9
5	7.6	7.2
6	5.9	5.5
7	6.1	6.1
8	7.8	7.5
9	6.7	6.3
10	4.7	4.2
11		5.1
12		4.4
13		4.5
14		5.3
15		7.5
16	5.7	
17	6.0	
18	7.5	
19	6.5	
20	4.2	

Random effects

Random effect model

- In purely fixed effects models we have
  - Random variables we observe
  - Model parameters we want to estimate
- In random effects models we have
  - Random variables we observe
  - Random variables we do NOT observe
  - Model parameters we want to estimate
- This model class is very useful and goes by many names: random effects models, mixed models, latent variable models, state-space models, frailty models, hierarchical models, ...

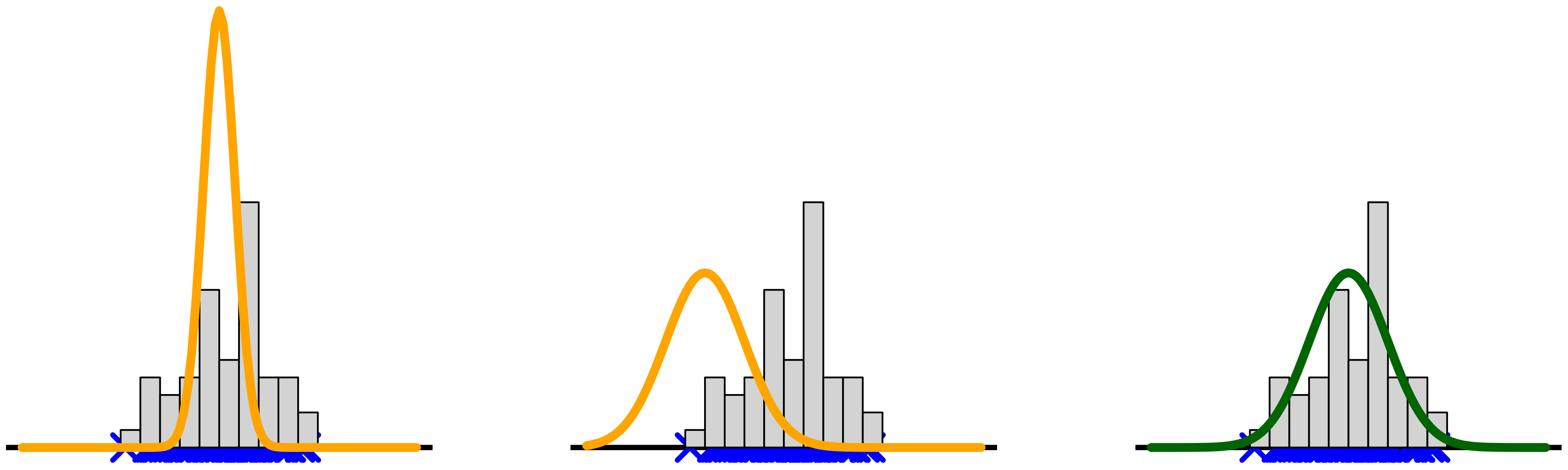
So the difference is that we have some quantity with a distribution, which is unobserved.

Reminder: Estimation with purely fixed effects

- We have:

  Observations: $y = (y_1, y_2, \ldots, y_n)$

  Parameters (μ, σ) in model: $y_i \sim N(\mu, \sigma^2)$



- Choose parameters which make our model best match the data (optimize likelihood).

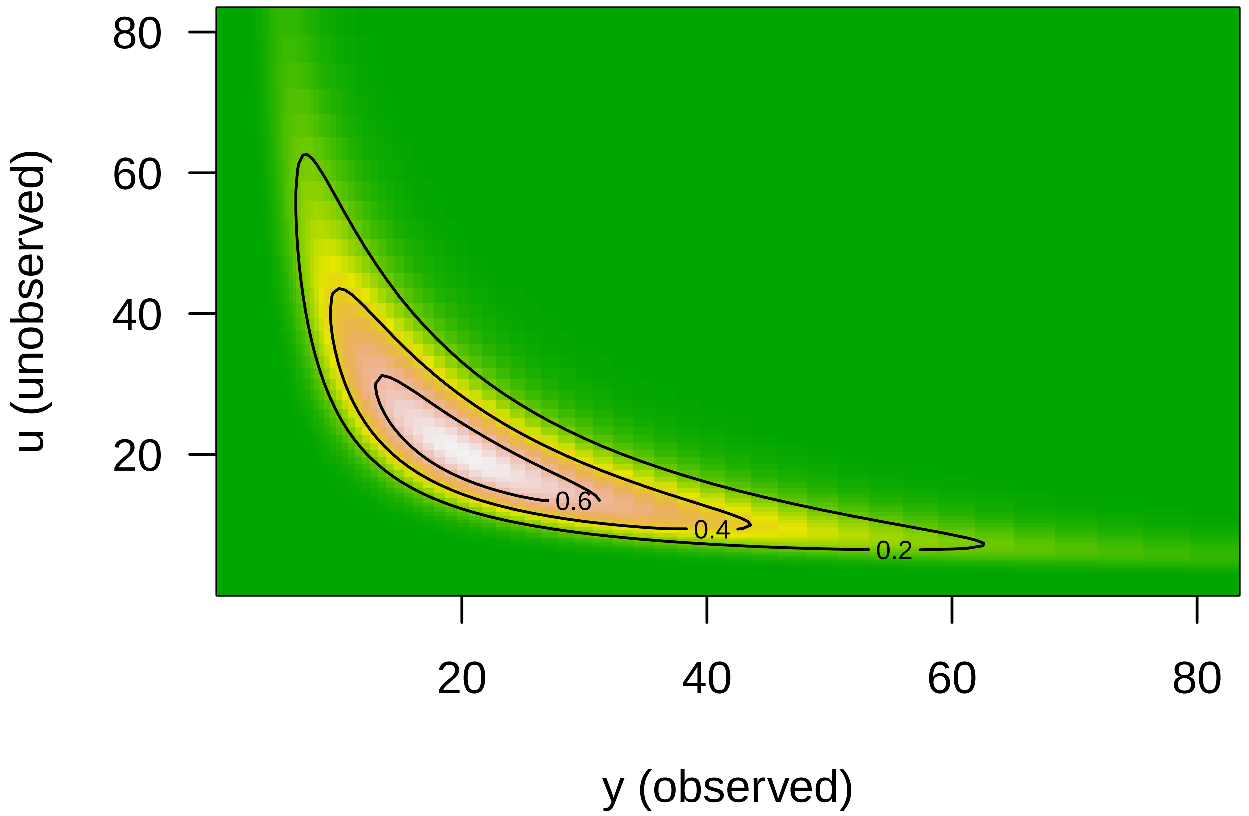
Estimation with random effects

- We have:

  Observations: y

  Unobserved random effects: u

  Parameters (θ) in model: $(y, u) \sim D(\theta)$



- How do we estimate our parameters when some of our observations are not observed?

Estimation with random effects — 2

- The banana is only an intermediate calculation

  1: Joint model (banana) is determined from model parameters θ

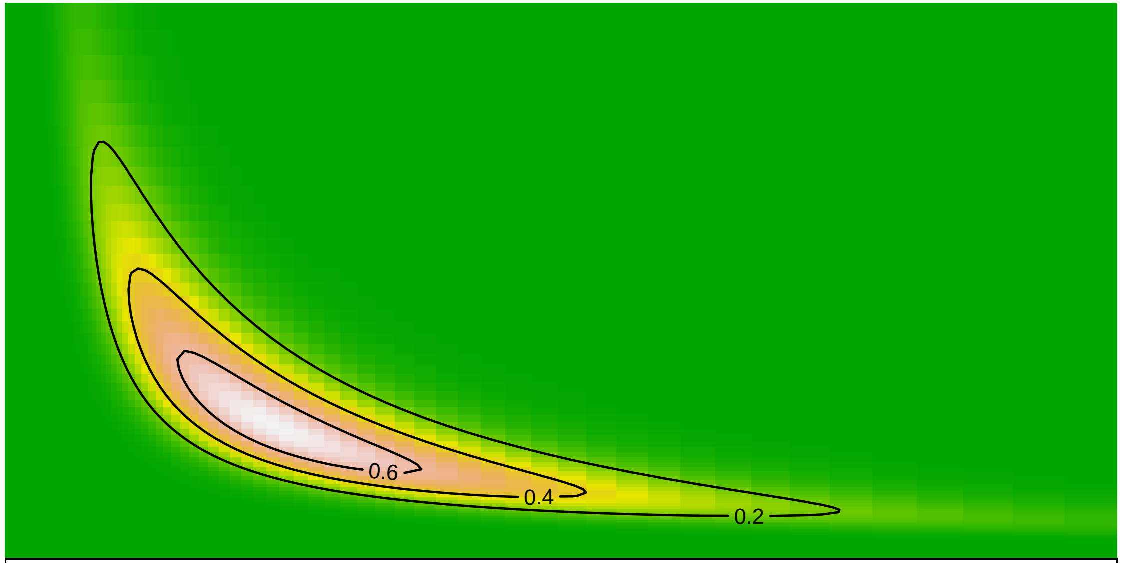
  2: Marginal model is calculated from joint by integration

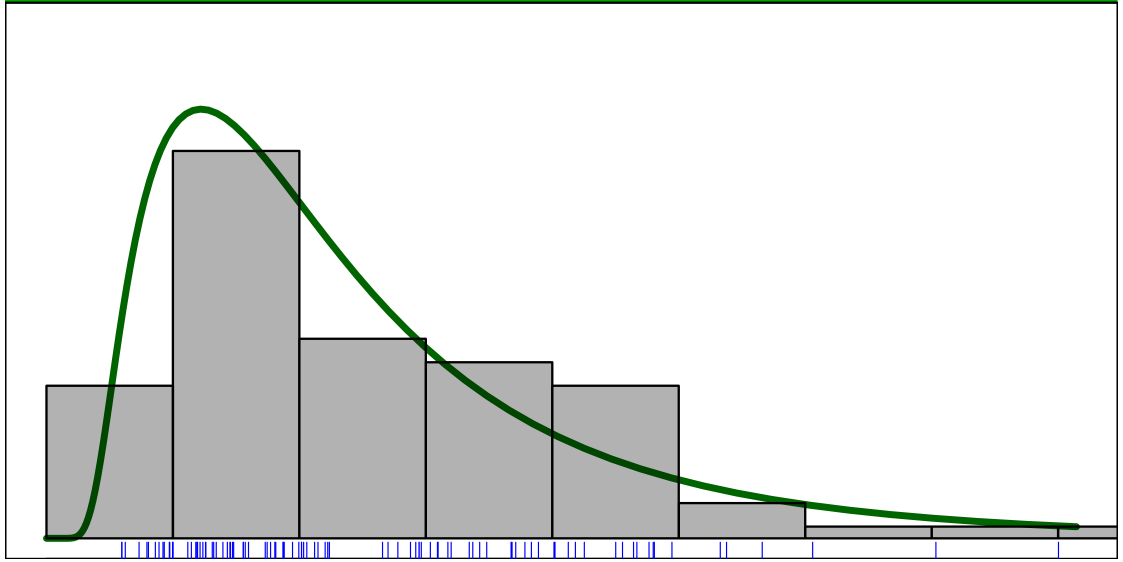
  3: Marginal is matched to data as always

- Imagine the distribution $D(\theta)$ is described by a likelihood function $L(y, u, \theta)$, then:

  $$L_M(y, \theta) = \int L(y, u, \theta)\, du$$

  is the *marginal* likelihood.





Examples where this is the underlying technique

- Random effects are often used:

  Time series models via state-space models Correlation via hidden processes

  Split-plot models Correlated observations within plot

  Repeated measurements Correlated observations within subject

  Overdispersion in Poisson via Negative binomial Patchiness accounted for

  Spatial models Correlation introduced via hidden field

  Penalized splines Smoothing penalties equivalent to MVN latent variables

  ...

- Often our only option!
- When something unobserved gives us <u>extra variation</u> or <u>correlated observations</u>.

The Laplace approximation

Reminder: Multivariate normal distribution

- The density for a k-dimensional multivariate normal distribution with mean vector μ and covariance matrix Σ is:

$$L(z) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left[-\frac{1}{2}(z-\mu)^\top \Sigma^{-1}(z-\mu)\right]$$

- We write $Z \sim N_k(\mu, \Sigma)$.

- So, since it needs to integrate to one:

$$\int \exp\left[-\frac{1}{2}(z-\mu)^\top \Sigma^{-1}(z-\mu)\right] dz = \sqrt{(2\pi)^k |\Sigma|}$$

The Laplace approximation

- We need to calculate the difficult integral

$$L_M(\theta, y) = \int_{\mathbb{R}^k} L(\theta, u, y) du$$

- So we set up an approximation of $\ell(\theta, u, y) = -\log L(\theta, u, y)$

$$\ell(\theta, u, y) \approx \ell(\theta, \hat{u}_\theta, y) + \frac{1}{2}(u - \hat{u}_\theta)^\top \left(\ell''_{uu}(\theta, u, y)|_{u=\hat{u}_\theta}\right)(u - \hat{u}_\theta)$$

- Which (for given θ) is the second order Taylor approximation around:

$$\hat{u}_\theta = \underset{u}{\text{argmax}}\ L(\theta, u, y)$$

- With this approximation we can calculate:

$$
\begin{aligned}
L_M(\theta, y) &= \int_{\mathbb{R}^k} L(\theta, u, y) du \\
&\approx \int_{\mathbb{R}^k} e^{-\ell(\theta, \hat{u}_\theta, y) - \frac{1}{2}(u-\hat{u}_\theta)^\top \left(\ell''_{uu}(\theta,u,y)|_{u=\hat{u}_\theta}\right)(u-\hat{u}_\theta)} \, du \\
&= L(\theta, \hat{u}_\theta, y) \int_{\mathbb{R}^k} e^{-\frac{1}{2}(u-\hat{u}_\theta)^\top \left(\ell''_{uu}(\theta,u,y)|_{u=\hat{u}_\theta}\right)(u-\hat{u}_\theta)} \, du \\
&= L(\theta, \hat{u}_\theta, y) \sqrt{\frac{(2\pi)^k}{|\left(\ell''_{uu}(\theta, u, y)|_{u=\hat{u}_\theta}\right)|}}
\end{aligned}
$$

- In the last step we remember the normalizing constant for a multivariate normal, and that $|A^{-1}| = 1/|A|$.

Laplace approximation work flow

0. Initialize θ to some arbitrary value θ_0

1. With current value for θ optimize joint likelihood w.r.t. u to get $\hat{u}_\theta$ and corresponding Hessian $H(\hat{u}_\theta)$.

2. Use $\hat{u}_\theta$ and $H(\hat{u}_\theta)$ to approximate $\ell_M(\theta)$

3. Compute value and gradient of $\ell_M(\theta)$

4. If the gradient is greater than ϵ set θ to a different value and go to **1.**

Notice the huge number of — possibly high dimensional — optimizations that are required.

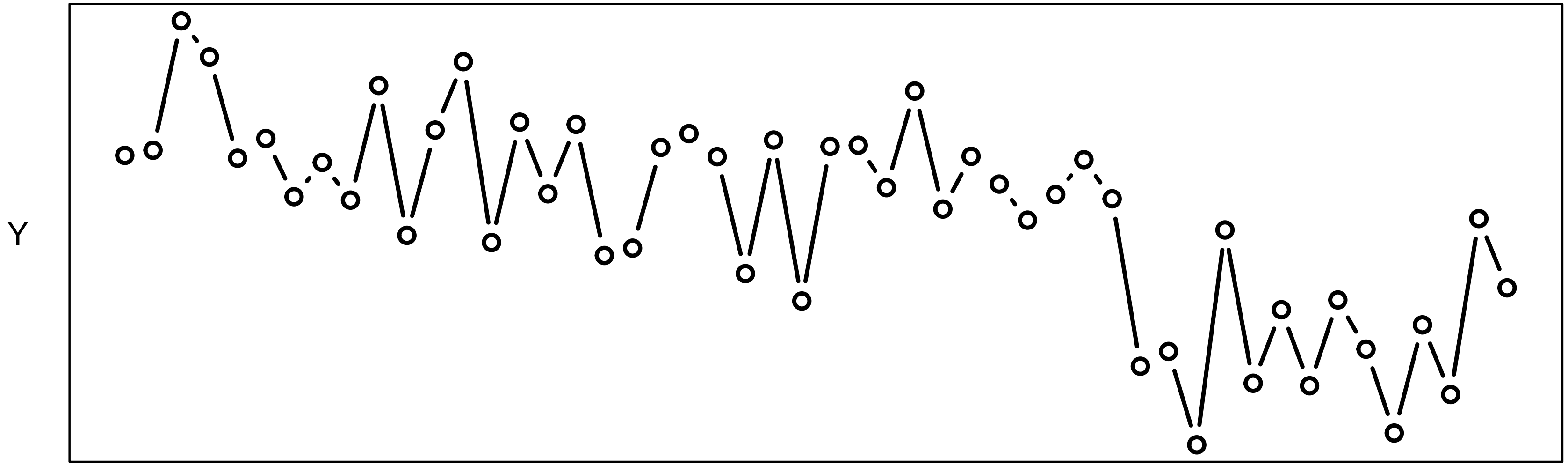
The Laplace approximation

- The Laplace likelihood only approximates the marginal likelihood for mixed models with nonlinear random effects (including GLMMs with a link function) and thus maximizing the Laplace likelihood will result in some amount of error in the resulting estimates.
- It can be shown that joint log-likelihood converges to a quadratic function of the random effect for increasing number of observations per random effect $\rightarrow$ the Laplace approximation is *asymptotically* exact.
- In practical applications the accuracy of the Laplace approximation may still be of concern, but often improved numerical approximation of the marginal likelihood (such as Gaussian quadrature) is computationally infeasible.
- We can evaluate if it is causing bias (subject for later)

Simple state-space model

State-space model (random walk+observation error)

- Observation vector Y generated from:
  - $\lambda_i = \lambda_{i-1} + \eta_i$
  - $Y_i = \lambda_i + \varepsilon_i$
  - where $i = 1 \ldots 50$, $\eta_i \sim N(0, \sigma_\lambda^2)$, and $\varepsilon_i \sim N(0, \sigma_Y^2)$ all independent.



- Notice λ vector unobserved.
- Here we wish to estimate both λ and the model parameters ($\lambda_\circ, \sigma_\lambda$, and σ_ε)

Can construct the joint likelihood

- The joint likelihood contributions from the random effects:

$$(\lambda_1 - \lambda_\circ) \sim N(0, \sigma_\lambda^2)$$

$$(\lambda_2 - \lambda_1) \sim N(0, \sigma_\lambda^2)$$

...

$$(\lambda_{50} - \lambda_{49}) \sim N(0, \sigma_\lambda^2)$$

- The joint likelihood contributions from the observations:

$$y_1 \sim N(\lambda_1, \sigma^2)$$

$$y_2 \sim N(\lambda_2, \sigma^2)$$

...

$$y_{50} \sim N(\lambda_{50}, \sigma^2)$$

- So the joint negative log likelihood becomes:

$$-\sum_{i=1}^{50} \log(\varphi_{\circ,\sigma_\lambda}(\lambda_i - \lambda_{i-1})) - \sum_{i=1}^{50} \log(\varphi_{\lambda_i,\sigma}(y_i))$$

We can solve this by Laplace approximation in RTMB

- Code up the joint negative log likelihood
  - Declare the unobserved quantities as 'random' (but part of parameter set)
  - Code as if the unobserved were observed
- Identify which quantities are to be considered as random effects (unobserved), as e.g:

```
obj <- MakeADFun(nll,par,random="lam")
```

Random walk + noise in RTMB

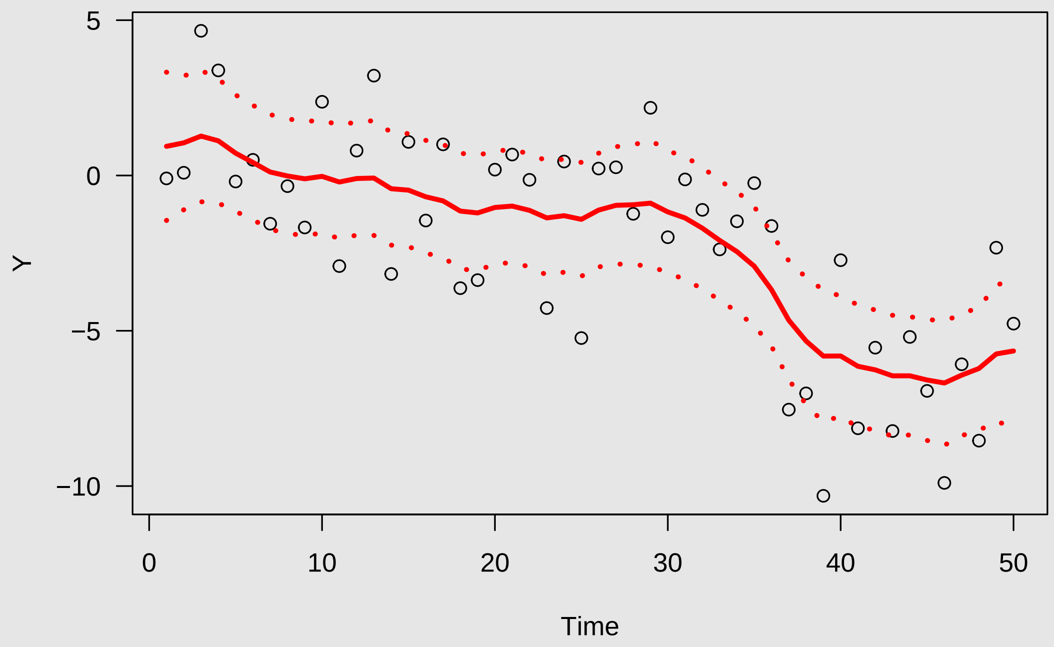
```r
library(RTMB)
dat <- list(y=scan("Y.dat"))
par <- list(logSdRw=0, logSdObs=0, lam0=0, lam=numeric(length(dat$y)))

jnll <- function(par){
  getAll(par, dat)
  sdRw <- exp(logSdRw)
  sdObs <- exp(logSdObs)
  ret <- -dnorm(lam[1], mean=lam0, sd=sdRw, log=TRUE)
  ret <- ret - sum(dnorm(diff(lam), sd=sdRw, log=TRUE))
  ret <- ret - sum(dnorm(y, mean=lam, sd=sdObs, log=TRUE))
  ret
}

obj <- MakeADFun(jnll, par, random="lam", silent=TRUE)
opt <- nlminb(obj$par, obj$fn, obj$gr)

sdr <- sdreport(obj)
pl <- as.list(sdr, "Est")
plsd <- as.list(sdr, "Std")

plot(dat$y, xlab="Time", ylab="Y", las=1)
lines(pl$lam, lwd=3, col="red")
lines(pl$lam-2*plsd$lam, lty="dotted", lwd=3, col="red")
lines(pl$lam+2*plsd$lam, lty="dotted", lwd=3, col="red")
```



ss.R

Exercise 1: RW + noise with missing observations

a) Consider again the simple random walk plus noise example. In the file: `rwmissing.dat` six of the observations are missing (coded as 'NA'). Consider and implement changes in the program to deal with that.

b) Make a plot of the estimated process and its confidence interval (similar to the one on the previous slide) to illustrate the effect of the missing observations.

Hint: The following code can be used to identify which observations are not missing:

```
idx <- which(!is.na(y))
```

That's basically it(!)

- Seems like too simple an example?
- This approach can be useful very generally, e.g:
  - Higher dimension
  - Some non-linearity
  - Some non-normality in the process and observations

Exercise 2: A theta logistic population model

A theta logistic population model is defined for the log-transformed population size as a nonlinear function of its previous size in the following way:

$$X_t = X_{t-1} + r_0 \left(1 - \left(\frac{\exp(X_{t-1})}{K}\right)^{\theta}\right) + e_t,$$

$$Y_t = X_t + u_t,$$

where $e_t \sim N(0, Q)$ and $u_t \sim N(0, R)$.

Data for this model is available in the file: `theta.dat`

a) Fit the model to data and estimate the five model parameters.

Hint: It is helpful to log transform the parameters and initialize $\log(K)$ to around 6.