

# Data cloning in R with the dclone package

Peter Solymos

Quantitative Ecology Meeting - Dec 13, 2013 - Univ. Alberta

## Data cloning (DC)

DC is a computational tric to get **MLE and asymptotic variance** using general purpose Bayesian MCMC sotware

Introduced by Lele et al. 2007, see also Lele et al. 2010 ([PDFs here](#))

Motivated by the need to fit **hierarchical models** without subscribing to the Bayesian paradigm

## Hierarchical models (Normal-Normal LMM)

You often have more unknowns than observations:

$$Y_{ij} \mid \mu_{ij} \sim \text{Normal}(\mu_{ij}, \sigma^2)$$

$$i = 1, \dots, n; j = 1, \dots, m_n$$

$$\mu_{ij} = X_{ij}^T \theta + \epsilon_i$$

Often called random effects:

$$\epsilon_i \sim \text{Normal}(0, \tau^2)$$

## More generally

$$(y \mid X = x) \sim h(y; X = x, \theta_1)$$

$$X \sim g(x; \theta_2)$$

$$\theta = (\theta_1, \theta_2)$$

$$L(\theta; y) = \int h(y \mid x; \theta_1) g(x; \theta_2) dx$$

## Frequentist toolkit

It is based on a **data model** that gives a likelihood:

- which parameter values are most likely to lead to this set of observations given the model?
- the fact that we can write down the likelihood does not mean we can estimate the parameters
- for random effects it leads to a multidimensional integral, derivatives, noisy likelihood surface

LMMs and GLMMs can be fit quite efficiently using e.g. `{lme4}`.

But what do we do with growth models, differential equations, more complex mixed models, etc.?

## Bayesian toolkit

No real restrictions other than battling funny error (trap) messages or initial values.

It might take forever, but you'll get a bunch of numbers at the end, called the posterior:

$$\pi(\theta \mid y) = \frac{L(\theta; y)\pi(\theta)}{\int L(\theta; y)\pi(\theta)d\theta}$$

How do I change my belief about  $\pi(\theta)$  after observing the data?

## Binomial model

Model:  $Y_i \sim \text{Binomial}(1, p)$

Data:  $y_1, y_2, \dots, y_n$

PMF:  $P(Y = y) = p^y(1 - p)^{1-y}$

Likelihood:  $L(p; y_1, y_2, \dots, y_n) = \prod_{i=1}^n p^{y_i}(1 - p)^{1-y_i}$

Go to the code & app!



## DC theory

$$y^{(K)} = (y, \dots, y)$$

$$L(\theta; y^K) = L(\theta; y)^K$$

$$\pi_K(\theta \mid y) = \frac{[L(\theta; y)]^K \pi(\theta)}{\int [L(\theta; y)]^K \pi(\theta) d\theta}$$

$$\pi_K(\theta \mid y) \sim MVN(\hat{\theta}, \frac{1}{K} I^{-1}(\hat{\theta}))$$

## Make MLE great again!

- we can use Bayesian MCMC toolkit for frequentist inference
- mean of the posterior is the MLE ( $\hat{\theta}$ )
- $K$  times the posterior variance is the variance of the MLE
- no need to calculate high dimensional integrals and second derivatives
- noisy likelihood surface — no numerical optimization involved
- **bonus:** independent of the specification of the prior distribution.

Go to the app!

## {dclone} R package

Described in Solymos 2010 (R Journal, PDF in this repo)

On CRAN since 2009 (10 years!)

## Binomial-Binomial model

- $W_i$ : this denotes the *observed* status at the location  $i$ , can be 0 or 1,
- $Y_i$ : this denotes the true status at the location  $i$ , can be 0 or 1; this status is *unknown*.
- True status:  $Y_i \sim \text{Bernoulli}(\varphi)$
- Observed status:  $(W_i \mid Y_i = y_i) \sim \text{Bernoulli}(p^{y_i}(1 - p)^{1-y_i})$

$$L(p, \varphi; w_1, w_2, \dots, w_n) = \prod_{i=1}^n P(W_i = w_i; p, \varphi) =$$
$$\prod_{i=1}^n (p\varphi)^{w_i} (1 - p\varphi)^{1-w_i}$$

Go to the code!

## DC in action

I used DC to prove identifiability of single visit N-mixture model in SOlymos et al. 2012. Stuff implemented in `{detect}` package

Did some SAR modeling (that is being cited beside Gould and Rosenzweig), see `{sharx}` package.

`{PVAClone}` package to fit PVA models with observation error.

## Can we somehow identify the params?

- True status:  $Y_i \sim \text{Bernoulli}(\varphi)$ .
- Observed status:  $(W_{i,t} \mid Y_i = 1) \sim \text{Bernoulli}(p)$  and  $W_{i,t} \mid Y_i = 0$  equals 0 with probability 1.

The likelihood function is:

$$L(p, \varphi; w_{1,1}, \dots, w_{n,T}) = \prod_{i=1}^n \left[ \varphi \left( \binom{Y}{w_{i\cdot}} p^{w_{i\cdot}} (1-p)^{T-w_{i\cdot}} \right) + (1-\varphi) I(w_{i\cdot} = 0) \right]$$



Go back to code!

## So you say I should wait more?

Increasing  $K$  most often leads to a linear increase in DAG size

- good: easy to predict how much longer it takes
- bad: what already takes long enough will take as much longer

HPC to the rescue!

Go back to code!

# The End

For more stuff: go to [datacloning.org](https://datacloning.org)

Dclonified BUGS examples:

<https://github.com/datacloning/dcexamples/>

S4 magic:

<https://cran.r-project.org/web/packages/dcmle/index.html>

Dclone mailing list:

<https://groups.google.com/forum/#!forum/dclone-users>