a spatial lme example

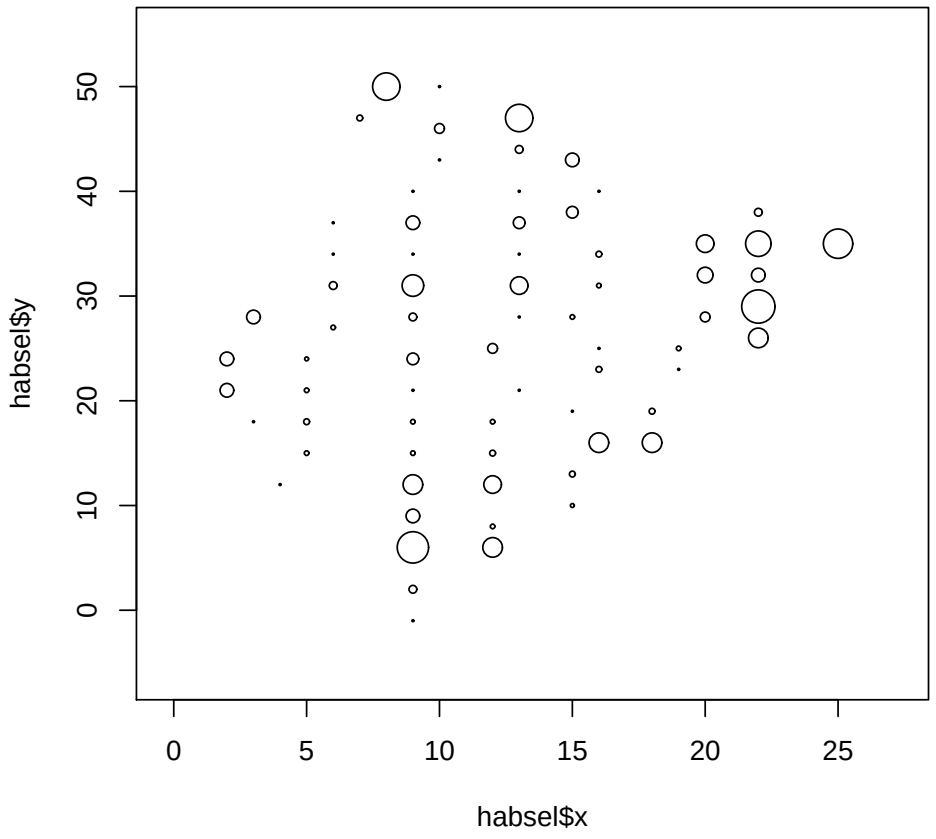
Ottar Bjørnstad

June 2, 2008

spatial dependence

The data contained in **habsel.txt** represents the number of animals captured ($captures) in 71 permanent traps located at co-ordinates $x and $y on a small Norwegian island. Also reported is various habitat variables in the viscinity of each trap. There is a bunch of things measured but we will focus on $lichen (the cover of lichen), $veg (the total vegetation cover, $heather (the cover of heather), $moss (the cover of moss) and $stdp (a measure of structural complexity).

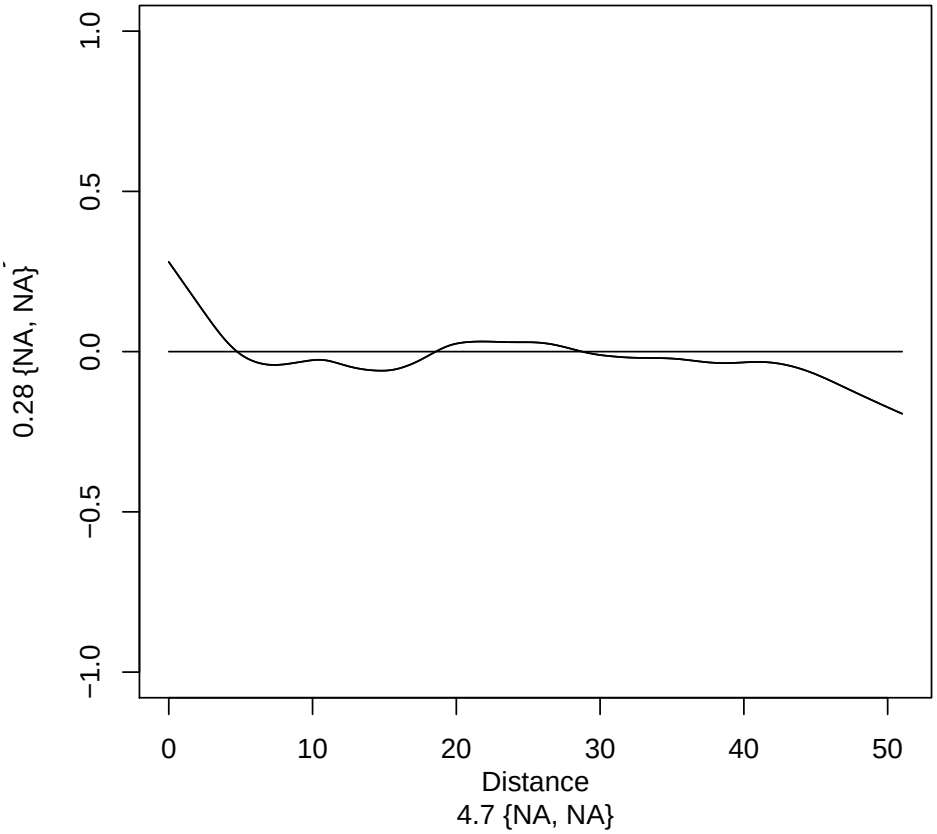
```
> habsel = read.table("habsel.txt", header = TRUE)
> symbols(x = habsel$x, y = habsel$y, circles = habsel$captures, inches = 0.1)
```



```
> fit1 = lm(log(captures + 1) ~ lichen + veg + heather + moss + stdp,
+     data = habsel)
> summary(fit1)

Call:
lm(formula = log(captures + 1) ~ lichen + veg + heather + moss +
    stdp, data = habsel)

Residuals:
     Min       1Q   Median       3Q      Max
-1.41327 -0.61763  0.07103  0.54326  1.37054

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.31972    0.73337   0.436  0.66432
lichen      -0.48353    0.23575  -2.051  0.04431 *
veg          0.05878    0.02125   2.766  0.00738 **
heather      0.28951    0.16541   1.750  0.08480 .
moss        -0.31274    0.14465  -2.162  0.03431 *
stdp         0.12136    0.09234   1.314  0.19338
---
Signif. codes:  0 Ś***Š 0.001 Ś**Š 0.01 Ś*Š 0.05 Ś.Š 0.1 Ś Š 1

Residual standard error: 0.7598 on 65 degrees of freedom
Multiple R-Squared: 0.304,  Adjusted R-squared: 0.2505
F-statistic: 5.678 on 5 and 65 DF,  p-value: 0.0002098
```

So it looks like there is some abundance-environment association. HOwever, the issue with these types of observational data is that there may be strong spatial autocorrelation that may invalidate a simple regression of abundance against environment. We can do a quick look at this correlation in residuals from an 'independence' model. (the log-transformation is to make the capture variable a bit more Gaussian-looking):

```
> require(ncf)
> cres = spline.correlog(x = habsel$x, y = habsel$y, z = resid(fit1), resamp = 0)
> plot(cres)
```



So there seems to be some spatial autocorrelation. To do the correct model taking into account the autocorrelation, we use corr-argument available in mle. First, however, we need to make sure mle understands that all the data belongs to a single dependence group. And then we fit the model assuming exponential spatial dependence:

– OTTAR TALK ABOUT PARAMETRIC CORRELATION FUNCTIONS HERE

```
> require(nlme)
> habsel$grp = rep(1, dim(habsel)[1])
> fit2 = lme(log(captures + 1) ~ lichen + veg + heather + moss + stdp, data = habsel,
+     random = ~1 | grp,
+     corr = corSpatial(form = ~x + y, type ="exponential", nugget = F), method = "ML")
> summary(fit2)

Linear mixed-effects model fit by maximum likelihood
 Data: habsel
       AIC      BIC   logLik
  169.2112 189.5753 -75.6056

Random effects:
 Formula: ~1 | grp
         (Intercept)  Residual
StdDev: 2.337264e-05 0.7340087
```

```
Correlation Structure: Exponential spatial correlation
 Formula: ~x + y | grp
 Parameter estimate(s):
   range
2.053232
Fixed effects: log(captures + 1) ~ lichen + veg + heather + moss + stdp
                Value Std.Error DF    t-value p-value
(Intercept) -0.0338972 0.7069316 65 -0.0479497  0.9619
lichen      -0.3456476 0.2333066 65 -1.4815165  0.1433
veg          0.0633296 0.0200605 65  3.1569395  0.0024
heather      0.1579573 0.1640640 65  0.9627785  0.3392
moss        -0.2814743 0.1362936 65 -2.0652058  0.0429
stdp         0.1803396 0.0864901 65  2.0850892  0.0410
 Correlation:
        (Intr) lichen veg    heathr moss
lichen  -0.447
veg     -0.761  0.444
heather -0.160  0.132  0.238
moss    -0.370 -0.059  0.043 -0.206
stdp    -0.485  0.129  0.066 -0.097 -0.142

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-1.97650808 -0.93407055 -0.02748326  0.68969420  1.83586377

Number of Observations: 71
Number of Groups: 1
```

Excersize: interpret the results! What does 'range' mean?

We can test for significant spatial correlation by comparing this model with the mle-model without spatial dependence:

```
> fit3 = lme(log(captures + 1) ~ lichen + veg + heather + moss + stdp,
+     data = habsel,
+     random = ~1 | grp, method = "ML")
> anova(fit2, fit3)
```

```
     Model df      AIC      BIC    logLik   Test  L.Ratio p-value
fit2     1  9 169.2112 189.5753 -75.60560
fit3     2  8 172.2187 190.3202 -78.10936 1 vs 2 5.007518  0.0252
```

Excersize: Fit a new model with 'Gaussian' (not exponential) spatial autorrelation. Which model better fits the data?

Some other types of dependence

random blocks: maternal effects

The data contained in **mouse.txt** has the following variables (this is a slightly expanded version of the data we looked at in the Bayes class).

```
> mouse <- read.table("mouse.txt", sep = "\t", header = T)
> names(mouse)
```

```
 [1] "cage"    "mweight" "lsize"   "code"    "ind"     "sex"     "wt0"
 [8] "wt1"     "wt2"     "wt3"     "wt4"     "wt5"     "wt7"     "wt9"
[15] "wt11"    "wt13"    "wt15"    "wt18"
```

data$wtX is the weight of offspring at the age of X days; $sex is 1 = male, 2 = female; $code is the indiviudal tag, $ind is the individual number of the 95 individuals; $cage is the the identifier of the litter (as well as the identity of the mother); $mweight is the weight of the mother; and $lsize is the litter size. The biological questions are: (1) Do males and females differ with respect to their *average* growth. (2) Is there a maternal effect? Do individuals from the same litter tend to be similar? (3) Do sex effect differ by litter.

let's remove any individual with missing data:

```
> mouse = na.omit(mouse)
```

We might first take an 'empirical' tack on the question (this is not necessary for the lme modelling, but it is interesting). Lets fit a model for the weight on day 11, ignoring any interdependence between littermates. Then look whether there is any signature of dependence in the residuals. First fit the model:

```
> fit = lm(wt11 ~ as.factor(sex) + mweight + lsize, data = mouse)
```

To look at autocorrelation recall the definition of correlation: $cor(x, y) = (x - \mu_x)(y - \mu_y)/\sigma_x\sigma_y$. The following bit of R-code will will calculate the autocorrelation-matrix among all residuals:

```
> scres = (resid(fit) - mean(resid(fit)))/sd(resid(fit))
> rcor = outer(scres, scres)
```

And here is one way to flag all littermates (with a 1, and all non-littermates or 'self-comparison' with 0):

```
> mat = outer(mouse$cage, mouse$cage, "-")
> mat[mat != 0] = -1
> mat = mat + 1
> diag(mat) = 0
```

With this it is easy to calculate the within-litter autocorrelation:

```
> mean(rcor[mat == 1])

[1] 0.5170621
```

There is clearly a substantial interdependence.

Now lets model the interdependence specifically (and using weight at day 5 as a covariate. To do this we use the **lme**-function in the **nlme**-package:

```
> library(nlme)
> fit = lme(wt11 ~ as.factor(sex) + wt5 + mweight + lsize,
+     random = ~1 | cage, data = mouse)
```

The above is the random-intercept model.
Excersize: use summary() to look at the fitted model. What is your interpretation? Advanced Q: What's up with the degrees-of-freedom for the fixed effects (hint: think of split-plot vs randomized block design)?

Usually we may want to look at whether there is random variability in other parameters. For instance, we may be interestedin asking: Do mothers who have unusually large female offspring also tend to have unusually large offspring for the litter size? To do this (and make sure that R manages to converge on good values), we have to change some control-parameters in the model:

```
> fit = lme(wt11 ~ as.factor(sex) + wt5 + lsize,
+     random = ~as.factor(sex) + lsize | cage, data = mouse,
+     control = lmeControl(maxIter = 500,
+     msMaxIter = 500, opt = "optim"))
```

Excersize: use summary() and random.effects() to answer: *Do mothers who have unusually large female offspring also tend to have unusually large offspring for the litter size?*

To examine the model fit, we can use:

```
> plot(fit, wt11 ~ fitted(.) | cage, abline = c(0, 1))
```

repeated measures

Of course, the full data-set is really a repeated measures data set (each individual was weighed on many occations). mle() can help us the full model if we want to (we first reshape the data):

```
> mouse2 = reshape(mouse, idvar = "ind", varying = list(names(mouse)[7:18]),
+     v.names = "wt", direction = "long", times = c(0,
+     1, 2, 3, 4, 5, 7, 9, 11, 13, 15, 18))
> fit = lme(wt ~ as.factor(sex) + time + lsize,
+     random = ~time | cage/ind, data = mouse2,
+     control = lmeControl(maxIter = 500,
+     msMaxIter = 500, opt = "optim"))
```

Iterpret the result (NB! |cage/ind means that individual is nested within cage). Is most of the random variation in growth at the maternal level or the individual level?