Statistics 306 **Homework 1** **Due Oct 13**

1. This question is related to the data analysis in the paper *Phonemic diversity supports a serial founder effect model of language expansion from Africa* published by Q.D. Atkinson in *Science* (15 April 2011). It is recommended that you read the paper and the supplementary material, which are available at

`http://www.sciencemag.org/content/332/6027/346.full.pdf`

`.../content/suppl/2011/04/13/332.6027.346.DC1/1199295-Atkinson-SOM.pdf`

These articles and other relevant files can be found at

`http://www.stat.uchicago.edu/~pmcc/prelims/2011`

Download the files `S1.dat` and `S1.R`, modify the file address in first line of the latter as needed, and execute it. The response vector `S1$Tnpd` contains the total phoneme diversities for 504 languages, and `S1$Fam` is the family to which each language belongs. There are 109 distinct families. Each language has a geographical centroid, and `S1$Dbo` is the distance (linguistic distance) from a certain point in central west Africa. In addition, `gcd` is the matrix of great-circle distances between centroids. All distances are given in km, but 1000 km is a better scale, and it is assumed in what follows that all distances are converted to this scale. Set $\nu = 0.5, \lambda = 1$ and $K(d) = (d/\lambda)^\nu K_\nu(d/\lambda)$, i.e.

```
K <-(gcd/lambda)^nu * besselK(gcd/lambda, nu);  diag(K) <- 2^(nu-1)*gamma(nu).
```

For $\nu = 1/2$, this is the exponential function; for $\nu < 1$, the limit as $\lambda \to \infty$, suitably scaled, is $-d^{2\nu}$. For $\nu = 0, \lambda = \infty$, the scaled limit is $-\log(d)$ with some constant (zero) on the diagonal; for $\nu = 1, \lambda = \infty$, the limit is $d^2 \log(d)$.

(i) Use `fit0 <- regress(S1$Tnpd ~ S1$Dbo, ~Fam)` to fit the linear gaussian model in which

$$E(Y_i) = \beta_0 + \beta_1 \mathtt{Dbo}_i$$
$$\mathrm{cov}(Y_i, Y_j) = \sigma_0^2 \delta_{ij} + \sigma_1^2 \delta_{F(i),F(j)}$$

where $F(i)$ is the family to which language i belongs. `summary(fit0)` gives parameter estimates. Explain what this means.

(ii) Use `fit1 <- regress(S1$Tnpd ~S1$Dbo, ~Fam+K)` to check whether there is spatial correlation in addition to within-family correlation. Try a range of λ-values, and compare models via the log likelihood `fit$llik`. Does spatial correlation suffice on its own? Clifford (2004) asserts that, for naturally occurring processes, $nu = 0, \lambda = \infty$ is invariably the 'right' value. Does the likelihood-based confidence region include the Clifford point?

(iii) It is possible that spatial correlations might be stronger between languages in the same family than between languages in different families. Check this by including a further term `Fam*K` where `Fam` is the block-factor (matrix) version of `S1$Fam`. For this purpose, use the exponential model with range 1000 km. What do you conclude?

(iv) Use the model in part (b) with $\nu = 1/2, \lambda = 1$ to predict the `Tnpd`-value for an Indo-European language whose centroid is Lisbon, $(-9.15, 38.75)$.

2. The singular-value decomposition of a real matrix A of order $n \times m$ is $A = H_1 D H_2$ where H_1 is orthogonal of order n, H_2 is orthogonal of order m, D is of order $n \times m$ with $D_{ii} \geq 0$, and all off-diagonal components zero. The singular values of A are the diagonal components $\{D_{ii} \geq 0\}$, conventionally arranged in decreasing order. Sometimes the term refers to the non-zero values. Let $J(A)$ be the product of the non-zero singular values, if there are any, and zero otherwise.

(i) Show that if A is symmetric and positive semi-definite, the singular values coincide with the non-zero eigenvalues, and thus that $J(A) = \mathrm{Det}(A)$.

(ii) True or False? $J(AB) \leq J(A)\, J(B)$. Give a proof or counter-example.

(iii) Under what conditions on the image of B and kernel of A is it the case that $J(AB) = J(A)\, J(B)$? Construct an example using 2×2 matrices where the product rule fails. Deduce that $J(AA') = J(A'A) = J^2(A)$.

(iv) Let A be of order $m \times n$ with rank $m \leq n$. Let $B = (A\Sigma A')^{-1}$, where Σ is symmetric and strictly positive definite. Deduce that $J(ABA') = J(B)\, J^2(A)$. Show that $Q = \Sigma ABA$ is a projection matrix. Express Q in terms of the matrix K whose columns span $\ker(A)$. Deduce the relation between $\det(B)$ and $\mathrm{Det}(WQ)$.

3. This question refers to the experiment described in

`http://www.stat.uchicago.edu/~pmcc/courses/stat306/rats.pdf`

and the related files `.../rats.R` and `.../rats.dat`.

(i) Making due allowance for correlations between observations on the same rat, and systematic variation from site to site, estimate the treatment effect and its standard error. Explain the details of the model fitted.

(ii) Compute a likelihood ratio statistic suitable for testing the hypothesis of no treatment effect making due allowance for correlations and site-to-site variations. Explain carefully how you computed this statistic. Report the value, the null distribution and the p-value.