What’s for today

- Asymptotic standard errors of MLE
- Problems in estimating Matérn covariance parameters
Suppose, in the previous example of Carbon and Nitrogen in soil data, that we get the parameter estimates.

For maximum likelihood estimation, we can use Hessian matrix of the loglikelihood function to get the asymptotic standard errors of the maximum likelihood estimates.

Hessian matrix is the matrix of the second-order partial derivatives of a function.

The observed information matrix is the negative of the Hessian matrix of the loglikelihood function evaluated at the maximum likelihood estimators.
It is known in statistics theory that maximum likelihood estimators are asymptotically normal with the mean being the true parameter values and the covariance matrix being the inverse of the observed information matrix.

In particular, the square root of the diagonal entries of the inverse of the observed information matrix are asymptotic standard errors of the parameter estimates.
Note though that these asymptotic results are based on the asymptotic framework that the spatial domain is expanding as number of observations increase (\textit{increasing domain asymptotics}).

There are other asymptotic frameworks:

- Infill asymptotics (or fixed domain asymptotics): in this framework, the spatial domain is fixed (bounded) and location of observation gets denser as the number of observation increases.
- Mixed domain asymptotics: combination of increasing domain asymptotics and infill asymptotics.

There are theories available for increasing domain asymptotics but theories under infill asymptotics have not been developed much, although it is common in spatial statistics that infill asymptotics is more natural asymptotic framework.
In practice, how do we get the asymptotic standard errors of the parameter estimates?

Suppose the loglikelihood function is \( I(\alpha, \beta) \) and we perform numerical minimization of \(-I(\alpha_0, \beta_0)\) where \( \alpha = e^{\alpha_0} \) and \( \beta = 3\frac{e^{\beta_0}}{1+e^{\beta_0}} \)

Suppose the numerical minimization algorithm returns \( J \) as its Hessian matrix

If we denote the “true” Hessian evaluated at the minimum values of the parameters (MLE) \( H \), we have the following relationship:

\[
H_{1,1} = J_{1,1} \left( \frac{\partial \alpha_0}{\partial \alpha} \right)^2, \quad H_{1,2} = J_{1,2} \frac{\partial \alpha_0}{\partial \alpha} \frac{\partial \beta_0}{\partial \beta}, \quad H_{2,2} = J_{2,2} \left( \frac{\partial \beta_0}{\partial \beta} \right)^2
\]

Therefore, we obtain the asymptotic SE of MLE from the inverse of \( H \) in the end
For Carbon-Nitrogen data, we get

\[ \hat{\alpha} = 0.040(0.021), \hat{\beta} = 50.73(33.10), \hat{\nu} = 0.52(0.25), \hat{\delta} = 0.029(0.0038), \]

\[ \hat{\beta}_0 = -3.12(0.14), \hat{\beta}_1 = 0.80(0.019) \]
On parameterizations of Matérn covariance function

- In statistics, we care about the property of estimators of parameters.
- Roughly speaking, as we have infinite number of data points, we want our estimators converge to the true values of the parameters.
- However, for Matérn covariance model, if our spatial domain is fixed and if we assume we get more dense observations, we cannot get such nice property for all of the parameters.
In particular, Zhang (2004, *JASA*) showed that we can only have such nice property for a combination of Matérn parameters. Use the parameterization that

\[ K(x) = \frac{\sigma^2 (x/\alpha)^\nu}{2^{\nu-1} \Gamma(\nu)} K_{\nu}(x/\alpha) \]

Then only \( \sigma^2 \alpha^{-2\nu} \) can be estimated *consistently*. Now I will show two figures that illustrate some of the above points (pictures provided by Jonathan Stroud).
Figure 1
About Figure 1

- It shows the likelihood function for the range vs sill for the exponential ($\nu=1/2$) and Whittle models ($\nu=1$)
- The bottom row shows the corresponding posterior distributions using Berger, DeOlivera & Sanso’s objective prior
- This is for a small dataset with n=100 obs
About Figure 2

- It shows the likelihood for the range vs sill in the exponential model.
- The bottom row shows the likelihood function for range vs the transformed parameter $\psi = (\text{sill} / \text{range})$.
- This is for a large dataset with $n=262144$ observations.
A Few Remarks

- Note the long ridge in the likelihood function for range vs sill, especially as \( n \) gets large.
- The curvature of this ridge is determined by the slope of the covariogram at the origin.
- This suggests the transformation \( \psi = \frac{sill}{range^{2\nu}} \), as proposed by Zhang (2004).
- The transformation fixes the problem for large \( n \), as range and \( \psi \) are nearly orthogonal.