What’s for today

- EOF method
- Bayesian spatial statistics (brief introduction)
We will revisit nonstationary covariance models here and we will discuss the method for spatial and spatio-temporal models.

The lecture material is based on the notes written by Richard L. Smith (at UNC).
The EOF method and extensions

- EOF stands for *empirical orthogonal functions*
- This is also known as *Karhunen-Loève expansion*
- If we restrict the method to a finite set of observations, it is same as principal component analysis
THE EOF METHOD AND EXTENSIONS

- We consider doing regression about a variable \( y \) as a linear function of a random field \( X(s) \) defined on \( D \).
- Suppose we have replicates \((X_i, y_i), i = 1, \ldots, n\), where \( X_i(s) \) are independent.
- We consider the following decomposition

\[
y_i = \int_D X_i(s)B(s)\,ds + \epsilon_i, \quad 1 \leq i \leq n
\]

with \( \epsilon_i \) uncorrelated with \( X_i(s) \).
- Here \( B(s) \) is the least squares estimator and we get it from

\[
\sum_i y_i X_i(t) = \int_D \sum_i X_i(s)X_i(t)B(s)\,ds
\]

- We assume \( y_i \) and \( X_i(s) \) have mean zero.
The EOF method and extensions

Note as $n \rightarrow \infty$,

$$n^{-1} \sum y_i X_i(t) \rightarrow C_y(t) \quad \text{and} \quad n^{-1} \sum_i X_i(s) X_i(t) \rightarrow C(s, t),$$

where $C_y(t)$ is the covariance of $y_i$ and $X_i(t)$ and $C(s, t) = \text{Cov}[X_i(s), X_i(t)]$

Thus we get $C_y(t) = \int_D C(s, t) B(s) ds$

When we only have a finite set of observations, we split $D$ into $m$ segments $w_1, \ldots, w_m$, centered at $m$ observation points $s_1, \ldots, s_m$

We set $y_i = \sum_j w_j X_i(s_j) B(s_j) + \epsilon_i, \ i = 1, \ldots, n$

We may write this as $y_i = \sum_j x_{ij} \beta_j + \epsilon_i, \ i = 1, \ldots, n$
We also need *Karhunen-Loève* expansion of $C(s, t)$:

$$\int_D C(s, t)\psi(t)dt = \lambda \psi(s)$$

Under some conditions, $\{\psi_\nu, \nu \geq 1\}$ is complete and orthogonal with $\int_D \psi_\mu(s)\psi_\nu(s)ds = 1$ if $\nu = \mu$, and 0 otherwise.

Using the fact we write $C(s, t) = \sum_\nu \lambda_\nu \psi_\nu(s)\psi_\nu(t)$

We also write $X(s) = \sum_\nu z_\nu \lambda_\nu^{1/2} \psi_\nu(s)$ with $z_\nu, \nu = 1, 2, \ldots$, are uncorrelated random variables with mean zero and common variance 1.

We now write $B(s) = \sum_\nu \beta_\nu \psi_\nu(s)$

We get $y_i = \sum_\nu z_{i\nu} \lambda_\nu^{1/2} \beta_\nu + \epsilon_i$
For a general class of $C(s, t)$, without stationarity assumption, we find a complete orthonormal basis of eigenfunctions $\psi_\nu$

We then expand the process $X(s)$ and the covariance function using the basis

If we want to predict a random variable $y_i$ using $X_i(s)$, the optimal linear predictor is $\int_D X_i(s)B(s)ds$ and we get the explicit form for $B(s)$

If we observe the process $X_i(s)$ at only a finite number of locations, $s_1, \ldots, s_m$ and if we define a matrix $\Gamma$ with $\Gamma_{jk} = C(s_j, s_k)\sqrt{w_jw_k}$, then the Karhunen-Loève expansion is the same as a principal component decomposition for $\Gamma$ and the optimal prediction for $y_i$ is equivalent to fitting a principal components regression
Connection with Predictive process model

- From the above, we write $X(s) = \sum_{\nu} z_{\nu} \lambda_{\nu}^{1/2} \psi_{\nu}(s)$
- Suppose we approximate the above to $X(s) = \sum_{\nu=1}^{m} z_{\nu} \lambda_{\nu}^{1/2} \psi_{\nu}(s)$ for some finite $m$
- Now the covariance matrix of the above approximation is of rank $m$
- However, the difficulty here is to find the eigen functions from the integral equation
- Consider a set of knots $S = \{s_1^*, \ldots, s_m^*\}$
- Then we write $\frac{1}{m} \sum_{k=1}^{m} C(s, s_k^*) \psi_i(s_k^*) \approx \lambda_i \psi_i(s)$
- By plugging in the points in the knot to $s$ of the above equation we can solve for $\psi_i$
- Therefore the idea of predictive process is based on two approximations: (1) finite approximation of K-L expansion (2) the approximation of eigenfunctions using knots
Picture taken from Sudipto Banerjee’s website
Predictive process model

Picture taken from Sudipto Banerjee’s website
We will first go over some basics of Bayesian statistics.

I will also give you a very simple example of Bayesian method for spatial statistics.

We will then see more interesting examples.
So far we’ve seen ways of estimating parameters using likelihood or OLS (WLS, ...)

Bayesian method gives another way of estimating parameters as well as predicting values

So far we did point estimation given the assumption that parameters are fixed unknown numbers

In Bayesian world, you do not do point estimation, rather you get distributions for the parameters (or prediction values) given the assumption that parameters are random variables
Bayes’ theorem relates the conditional and marginal probabilities of events A and B:

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

The above is a fundamental theorem for Bayesian statistics.

Then we talk about “prior”, “likelihood”, and “posterior”
- Prior is the distribution of parameters (based on your prior belief)
- You know what likelihood is, basically the probability distribution for the data with parameters as arguments
- Then posterior is the distribution of parameters conditioning on the data (parameter distribution after incorporating the information from the data)
Suppose \( y = (y_1, \cdots, y_n) \) is the data.

We have unknown parameters \( \theta = (\theta_1, \cdots, \theta_k) \).

In Bayesian world, we assume \( \theta \) is random (not a fixed unknown number).

Then we think about a probability distribution of the parameters \( \theta \), \( \pi(\theta) \), this is called prior.

We can write our likelihood as \( L(y|\theta) \).

Then applying Bayes’ Theorem, we get the posterior distribution of \( \theta \) by

\[
p(\theta|y) = \frac{p(\theta, y)}{p(y)} = L(y|\theta) \cdot \pi(\theta) \cdot \text{constant}
\]

That is, we get the posterior distribution of the parameters from the product of prior and the likelihood.
In practice

- In many situations, closed form the probability distribution of posterior may not exist or the distribution is not one of the known distributions.
- In that situation, roughly speaking, what you would do is to get the posterior draws.
- You look at the histogram of the posterior draws or mean and median of posterior draws to get estimates for the parameters.
- There are many R packages for Bayesian statistics and there is a package called spbayes for spatial statistics with Bayesian methods.
Suppose we consider a random field $Z(s)$ observed on $s_1, \cdots, s_n \in \mathbb{R}^2$

We consider a model for $Z$ as $Z(s) = \mu + e(s)$ with $e \sim N(0, \Sigma)$

Here let’s model $\Sigma$ using an exponential covariance function:

$$\text{Cov}[e(s_1), e(s_2)] = \alpha \exp(-|s_1 - s_2|)$$

with unknown $\alpha$

Suppose we give a flat prior for $\mu$ and the prior for $\alpha$ is $p(\alpha) \propto \frac{1}{\alpha}$ (and they are independent)

The likelihood is $f(Z|\mu, \alpha) \propto \frac{1}{\alpha^{n/2}} \exp\left[-\frac{1}{\alpha}(Z - \mu 1)^T \Sigma_0^{-1} (Z - \mu 1)\right]$ when $\Sigma = \alpha \Sigma_0$

What is $p(\mu|\alpha, y)$?

The above won’t works as a posterior for $\mu$ since $\alpha$ is unknown

What is $p(\mu, \alpha|y)$?
In the method that we’ve learned so far for Kriging, we ignore the uncertainty in the covariance parameter estimates.

Bayesian approach will take into account the uncertainty about the parameters on prediction (Kriging).

Suppose we want to predict $Z(s_0)$ based on $Z(s_1), \cdots, Z(s_n)$.

Bayesian solution of Kriging is the posterior predictive distribution of $Z(s_0)$ given the observations $Z$,

$$
\pi(Z(s_0)|Z) \approx \int \pi(Z(s_0)|Z, \phi)\pi(\phi|Z)d\phi
$$
In the previous equation, how do we get $\pi(Z(s_0)|Z, \phi)$?

We approximate the PPD by

$$\pi(Z(s_0)|Z) = \frac{1}{m} \sum_{i=1}^{m} \pi(Z(s_0)|Z, \phi^{(i)})$$

where $\phi^{(i)}$ is the $i$th draw from the posterior distribution.
The lecture notes below are written by Huiyan Sang.
Spatial model specifications: $P(y|X, \theta)$.
Prior specifications: $P(\theta)$.
Posterior inference of model parameters: $P(\theta|y)$.
Predictions at new locations: $P(y_0|y)$.
Model comparisons.
A spatial Gaussian regression model for point referenced data:

\[ Y(s) = x^T(s)\beta + w(s) + \epsilon(s) \]

- \( w(s) \sim GP(0, \sigma^2 \rho(\cdot; \phi)) \).
- \( w(s_1, ..., s_n) \sim N(0, \sigma^2 H(\phi)) \).
- \( \epsilon(s) \sim N(0, \tau^2) \).
- \( \theta = (\beta, \sigma^2, \phi, \tau^2) \).

\[ f(y|X, \theta) = N(X\beta, \sigma^2 H(\phi) + \tau^2 I) \]
Data facts:

- 74 houses in Baton Rouge, LA.
- Response: (log) house price.
- Explanatory variables: LivingArea, OtherArea, Age, Bedrooms, Baths, HalfBaths.
- Locations: Easting, Northing.
- Exploratory data analysis.
- Spatial Gaussian regression model
  \[ \mathbf{X} = (\text{LivingArea}, \text{OtherArea}, \text{Age}) \]
Bayesian spatial modelling

- Spatial model specifications:
  \[ P(y|X, \theta) = N(X\beta, \sigma^2 \text{Matérn}(\phi, \nu) + \tau^2 I). \]
- Prior specifications: \( P(\beta, \sigma^2, \phi, \nu, \tau^2). \)
  \[
  \begin{align*}
  \pi(\beta) & \sim 1{or}N(\mu_\beta^0, \sigma_\beta^0) \\
  \pi(\sigma^2) & \sim IG(2, 1) \\
  \pi(\tau^2) & \sim IG(2, 1) \\
  \pi(\phi) & \sim U(a_\phi, b_\phi) \\
  \pi(\nu) & \sim U(a_\nu, b_\nu)
  \end{align*}
  \]
- Posterior inference: \( P(\theta|y) \)
We seek the posterior $P(\theta|y)$: use MCMC algorithm to draw samples from $P(\theta|y)$.

- Step 0: give initial values for $\theta$.
- Step $t + 1$:
  - Sample $P(\beta|\theta^t, y, x)$ from a multivariate normal.
  - Sample $P(\sigma^2|\theta^t, y, x)$ using Metropolis-Hasting (MH).
  - Sample $P(\tau^2|\theta^t, y, x)$ using MH.
  - Sample $P(\phi|\theta^t, y, x)$ using MH.
  - Sample $P(\nu|\theta^t, y, x)$ using MH.

- Repeat until MCMC converges.
**R package: spBayes**

Input: model, prior.
Output: posterior samples for each parameter.

library(spBayes)

m.1 <- spLM(LogSellingPrice~LivingArea+Age,
coords=cbind(Easting,Northing),
starting=list("phi"=3/5,"sigma.sq"=1,"tau.sq"=1,"nu"=1),
sp.tuning=list("phi"=0.01,"sigma.sq"=0.05,
"tau.sq"=0.05,"nu"=0.001),
priors=list("phi.Unif"=3/c(15,.5),"nu.Unif"=c(0.3,4),
"sigma.sq.IG"=c(2,1),
"tau.sq.IG"=c(2,1)),
cov.model="matern",
n.samples=5000, sub.samples=c(1000,6000,3),verbose=TRUE,
n.report=100, sp.effects=TRUE)

print(summary(m.1$p.samples))
plot(m.1$p.samples)
### Posterior Estimates of Model Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>50%</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>11.199903</td>
<td>10.5109084</td>
<td>11.7631528</td>
</tr>
<tr>
<td>LivingArea</td>
<td>0.000581</td>
<td>0.0004689</td>
<td>0.0006958</td>
</tr>
<tr>
<td>Age</td>
<td>-0.010828</td>
<td>-0.0192138</td>
<td>-0.0024206</td>
</tr>
<tr>
<td>sigma.sq</td>
<td>0.216377</td>
<td>0.0973630</td>
<td>0.5804358</td>
</tr>
<tr>
<td>tau.sq</td>
<td>0.054274</td>
<td>0.0386557</td>
<td>0.0813797</td>
</tr>
<tr>
<td>phi</td>
<td>0.438799</td>
<td>0.2626667</td>
<td>0.7968029</td>
</tr>
<tr>
<td>nu</td>
<td>2.291299</td>
<td>2.1222359</td>
<td>2.4075995</td>
</tr>
</tbody>
</table>
Spatial random effects
Spatial model specifications: $P(y|X, \theta)$.
Prior specifications: $P(\theta)$.
Posterior inference of model parameters: $P(\theta|y)$.
Predictions at new locations: $P(y_0|y)$. 
Predictive distribution:

\[ P(y_0 | y, X, X_0) = \int P(y_0 | y, \theta, X, X_0) P(\theta | y, X) d\theta \]

Recall that we already obtained draws from \( \theta \). For each draw \( \theta \), sample \( y_0 \) from \( P(y_0 | y, \theta, X, X_0) \).
$[Y(s_0)|\theta, Y]$ is a Gaussian distribution with the mean and variance given by

$$E[Y(s_0)|\theta, Y] = x^T(s_0)\beta + h^T(s_0)(\sigma^2H + \tau^2I_n)^{-1}(Y - X\beta)$$

and

$$Var[Y(s_0)|\Omega, Y] = \sigma^2 - h^T(s_0)(\sigma^2H + \tau^2I_n)^{-1}h(s_0) + \tau^2.$$
Prediction at 2 locations \((-5068, 1691)\) and \((-5064, 1692)\)

\[
\text{new.coords} = \text{as.matrix(rbind(c(-5066,1691),c(-5068,1698)))}
\]
\[
\text{new.x} = \text{as.matrix(rbind(c(1,2000,5),c(1,2000,5)))}
\]

\[
\text{pred} \leftarrow \text{sp.predict(m.1, pred.coords=new.coords,}
\]
\[
\text{pred.covars=new.x,thin=5)}
\]

\[
\text{y.hat} \leftarrow \text{apply(exp(pred$y.pred),1,mean)}
\]

361072 193493
Bayesian Spatial Modelling

- Spatial model specifications: $f(y|X, \theta)$.
- Prior specifications: $f(\theta)$.
- Posterior inference of model parameters: $f(\theta|y)$.
- Predictions at new locations: $f(y_0|y)$.
- Model comparisons.
Model fitting: $\text{DIC} = p_D + \bar{D}$

$\bar{D} = E[-2 \log(p(y|\theta))]$, and $p_D = \bar{D} - D(\bar{\theta})$.

```
print(sp.DIC(m.1,DIC.marg=TRUE, DIC.unmarg=FALSE))
print(sp.DIC(m.2,DIC.marg=TRUE, DIC.unmarg=FALSE))
```

% Add Bedrooms

<table>
<thead>
<tr>
<th></th>
<th>value</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>bar.D</td>
<td>-147.84</td>
<td>-146.61</td>
</tr>
<tr>
<td>pD</td>
<td>2.34</td>
<td>6.18</td>
</tr>
<tr>
<td>DIC</td>
<td>-145.50</td>
<td>-140.44</td>
</tr>
</tbody>
</table>

Conclusion: the model without Bedrooms as covariate fits the data better.

Prediction performance for a hold out dataset:

$MSPE = E[\sum_{j=1}^{m} (y_{true}(s_j) - y_{predict}(s_j))^2]$