The prediction problem

• Suppose that \( \{Z(s) : s \in D\} \), for \( D \subset \mathbb{R}^p \) is a stationary process with mean \( \mu_Z \) and covariance \( C_Z(h) \).

• The problem: Based on data \( z = (z(s_1), \ldots, z(s_n))^T \), what is the best prediction of \( Z(s_0) \) at a new location \( s_0 \)?

• This is a hard problem – let us simplify a little:
  - Based on data \( z = (z(s_1), \ldots, z(s_n))^T \), what is the best linear unbiased prediction (BLUP) of \( Z(s_0) \) at a new location \( s_0 \)?
  - We would also like a measure of uncertainty for this prediction.

• This is the spatial prediction or kriging [Krige, 1951] problem [See Cressie, 1990, for a discussion of the origins of kriging.]
A linear predictor

- The solution to the kriging problem should look like linear regression, but with random variables instead of data.
- Let us write our linear predictor as
  \[ PZ(s_0) = a_0 + \sum_{k=1}^{n} a_k Z(s_k). \]
- Here \( P \) denotes the linear prediction operator.
- We want to select the coefficients \( \{a_0, \ldots, a_n\} \) that minimize the mean squared prediction error which is defined by
  \[ \text{MSPE} = E[U(s_0)]^2, \]
  where
  \[ U(s_0) = Z(s_0) - PZ(s_0), \]
  is the prediction error in predicting \( Z(s_0) \) using \( PZ(s_0) \).
- As with linear regression, to minimize, we take partial derivatives of this expression with respect to each parameter, set equal to zero, and solve for the parameters.

The intercept term \( a_0 \)

- We have
  \[
  \frac{\partial}{\partial a_0} \text{MSPE} = \frac{\partial}{\partial a_0} E \left[ Z(s_0) - a_0 - \sum_{k=1}^{n} a_k Z(s_k) \right]^2 \\
  = (2) E \left[ (Z(s_0) - a_0 - \sum_{k=1}^{n} a_k Z(s_k))(-1) \right].
  \]
- Setting this expression equal to zero and simplifying we get
  \[ E(Z(s_0)) - a_0 - \sum_{k=1}^{n} a_k E(Z(s_k)) = 0. \]
- Since \( \{Z(s)\} \) is stationary this simplifies to
  \[ \mu_Z - a_0 - \sum_{k=1}^{n} a_k \mu_Z = 0. \]
- Thus \( a_0 \) satisfies
  \[ a_0 = \mu_Z \left( 1 - \sum_{k=1}^{n} a_k \right), \]
  a function of the mean of the process and the other prediction coefficients \( a_1, \ldots, a_n \).
The other terms

- For \( j = 1, \ldots, n \) we obtain that
  \[
  \frac{\partial}{\partial a_j} \text{MSPE} = \frac{\partial}{\partial a_j} E \left[ (Z(s_0) - a_0 - \sum_{k=1}^{n} a_k Z(s_k))^2 \right] 
  = (2) E \left[ (Z(s_0) - a_0 - \sum_{k=1}^{n} a_k Z(s_k))(-Z(s_j)) \right] 
  \]
- Again setting equal to zero we get
  \[
  E \left[ (Z(s_0) - a_0 - \sum_{k=1}^{n} a_k Z(s_k)) Z(s_j) \right] = 0.
  \]
- Or,
  \[
  E[Z(s_0)Z(s_j)] - a_0 E(Z(s_j)) - \sum_{k=1}^{n} a_k E[Z(s_k)Z(s_j)] = 0.
  \]
- With our definition of \( a_0 \)
  \[
  E[Z(s_0)Z(s_j)] - \mu_Z^2 (1 - \sum_{k=1}^{n} a_k) - \sum_{k=1}^{n} a_k E[Z(s_k)Z(s_j)] = 0.
  \]
- Thus
  \[
  \left\{ E[Z(s_0)Z(s_j)] - \mu_Z^2 \right\} = \sum_{k=1}^{n} a_k \left\{ E[Z(s_k)Z(s_j)] - \mu_Z^2 \right\},
  \]
  or more simply,
  \[
  C_Z(s_0 - s_j) = \sum_{k=1}^{n} a_k C_Z(s_k - s_j).
  \]

Ordinary kriging

- Thus the values of \( a_1, \ldots, a_n \) that minimizes the MSPE solves the following system of simultaneous equations:
  \[
  C_Z(s_0 - s_j) = \sum_{k=1}^{n} a_k C_Z(s_k - s_j), \quad j = 1, \ldots, n.
  \]
- Given their solution,
  \[
  a_0 = \mu_Z \left( 1 - \sum_{k=1}^{n} a_k \right).
  \]
- The above equations are the ordinary kriging (K) equations.
- Then with the solution to these equations
  \[
  P_Z(s_0) = a_0 + \sum_{k=1}^{n} a_k Z(s_k)
  \]
  is called the ordinary kriging predictor.
Ordinary kriging written in matrix form

- Let $\mathbf{Z} = (Z(s_1), \ldots, Z(s_n))^T$ be the vector of random variables associated with our data $\mathbf{z}$.

- Let $\Sigma_Z = \text{cov}(\mathbf{Z})$ denote the $n \times n$ nonnegative definite covariance matrix for $\mathbf{Z}$ with $(i, j)$ element $C_Z(s_i - s_j)$.

- Define $c_Z(s_0) = (C_Z(s_0 - s_1), \ldots, C_Z(s_0 - s_n))^T$, to be the vector of covariances between the process $Z(\cdot)$ at $s_0$ and all the other locations.

- Then $\mathbf{a} = (a_1, \ldots, a_n)^T$ satisfies $c_Z(s_0) = \Sigma_Z \mathbf{a}$.

- Since $\Sigma_Z$ is invertible we get that $\mathbf{a} = \Sigma_Z^{-1} c_Z(s_0)$, with $a_0 = \mu_Z(1 - \mathbf{a}^T \mathbf{1}_n)$.

Prediction in terms of conditional expectations

- From the first problem set we know that
  
  $$E([Y - f(X)]^2)$$

  is minimized by
  
  $$f(X) = E(Y|X).$$

- Similarly it can be shown that
  
  $$E([Z(s_0) - PZ(s_0)]^2)$$

  is minimized by
  
  $$PZ(s_0) = E(Z(s_0)|Z(s_1), \ldots, Z(s_n)).$$
Properties of the predictor

• The ordinary kriging predictor is indeed 
  unbiased – to see this note that we can write,

  \[ PZ(s_0) = a_0 + \sum_{k=1}^{n} a_k Z(s_k) \]

  \[ = \mu_Z(1 - \sum_{k=1}^{n} a_k) + \sum_{k=1}^{n} a_k Z(s_k) \]

  \[ = \mu_Z + \sum_{k=1}^{n} a_k(Z(s_k) - \mu_Z), \]

  which has mean \( \mu_Z = E(Z(s_0)) \).

• This implies that the prediction error

  \[ U(s_0) = Z(s_0) - PZ(s_0), \]

  has mean zero.

• Also the prediction error is uncorrelated with the observed random variables: for each \( k = 1, \ldots, n \)

  \[ \text{cov}(U(s_0), Z(s_k)) = 0. \]

The minimum value of the MSPE

• Since the predictor is unbiased the minimum value of the MSPE attained by the predictor is

  \[
  E(Z(s_0) - a_0 - a^T(Z - 1\mu_Z))^2
  = \text{var}(Z(s_0) - a_0 - a^T(Z - 1\mu_Z))
  = \text{var}(Z(s_0) - a^T Z)
  = \text{var}(Z(s_0)) - 2\text{cov}(Z(s_0), a^T Z) + \text{var}(a^T Z)
  = C_Z(0) - 2a^T c_Z(s_0) + a^T \text{cov}(Z)a.
  \]

• We call this quantity the 
  kriging variance.
Gaussian processes, & producing prediction intervals

- In the Gaussian case, all these results can be re-derived using conditional distributions involving multivariate normal distributions.

- For Gaussian processes linear prediction is optimal; these equations give the best unbiased predictor amongst all such predictors (linear or otherwise).

- If \( \{Z(s) : s \in D\} \) is a Gaussian process then a \( 100(1 - \alpha)\% \) prediction interval for \( Z(s_0) \) is given by
  \[
PZ(s_0) = \mu_Z(s_0) + \sum_{k=1}^{n} a_k (Z(s_k) - \mu_Z(s_k)),
  \]
  where MSPE is the kriging variance, and \( z_{1-\alpha/2} \) is the \( 1 - \alpha/2 \)th quantile of the standard normal distribution.

- For non-Gaussian processes this interval is approximate.

Assuming nonstationarity: universal kriging

- The prediction equations generalize to the nonstationary case (assuming variances exist).

- Let \( \mu_Z(s) = E(Z(s)) \) be the spatially-varying mean and let \( C_Z(s, t) \) be the covariance between \( Z(s) \) and \( Z(t) \).

- Then the universal kriging (UK) predictor of \( Z(s_0) \) that minimizes the MSPE is
  \[
PZ(s_0) = \mu_Z(s_0) + \sum_{k=1}^{n} a_k (Z(s_k) - \mu_Z(s_k)),
  \]
  where
  \[
a = (a_1, \ldots, a_n)^T = \Sigma_Z^{-1} c_Z(s_0),
  \]
  with
  \[
  \Sigma_Z = \text{cov}(Z) = [C_Z(s_i, s_j)]_{i,j}
  \]
  and
  \[
c_Z(s_0) = (C_Z(s_0, s_k) : k = 1, \ldots, n)^T.
  \]

- The minimum MSPE that this predictor attains is
  \[
  C_Z(s_0, s_0) - a^T c_Z(s_0).
  \]
Universal kriging with a regression mean structure

- In the case that \( \mu_Z(s) = x^T(s)\beta \) the universal kriging predictor of \( Z(s_0) \) becomes

\[
P_Z(s_0) = x^T(s_0)\beta + a^T(Z - X\beta),
\]

where

\[
X = \begin{bmatrix}
x^T(s_1) \\
\vdots \\
x^T(s_n)
\end{bmatrix}.
\]

Kriging for intrinsic processes

- It is possible to re-derive the kriging equation assuming only that the geostatistical process \( \{Z(s) : s \in D\} \) is \textbf{intrinsic}, with variogram \( 2\gamma_Z(\cdot) \).

- Assuming we have a linear unbiased predictor, the mean squared prediction error is

\[
MSPE = -\sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \gamma_Z(s_j - s_k).
\]

- We minimize this with respect to \( a_1, \ldots, a_n \), but now write the equations in terms of the variogram instead of the covariance function.
The practicalities of prediction

- In practice we do not know the true mean or covariance function for \( \{Z(s)\} \).

- The typical strategy is to estimate the mean and covariance parameters using the data \( z \), and then to plug them into the kriging prediction and variance equations.

  - We call such predictors **estimated best linear unbiased predictors** (EBLUPs).

- There is a suggestion in the literature [e.g., Zimmerman and Zimmerman, 1991] that we lose coverage when we estimate the parameters;
  e.g., upon building a 95% prediction interval, we actually get a 86% interval in practice.

Assessing predictive performance via crossvalidation

- It is common to **crossvalidate** to assess the performance of the predictions.

- For each \( i = 1, \ldots, n \):
  1. Remove \( z(s_i) \) from the dataset.
  2. Using the remaining data points, we predict \( Z(s_i) \).

- The empirical mean squared prediction over all locations is given by

  \[
  \frac{1}{n} \sum_{i=1}^{n} (z(s_i) - PZ(s_i))^2.
  \]
Optimal prediction? – an example

• Example: Suppose that in an environmental monitoring situation that we want to produce a map of those points in our domain of interest $D$ that exceed some threshold $\tau$.
  
  – How could we do it?
  
  – Is this an optimal prediction?
References

