University of California, Los Angeles
Department of Statistics

Statistics C173/C273 **Instructor: Nicolas Christou**

Kriging revisited

We observe $\mathbf{Z} = (Z(s_1), Z(s_2), \ldots Z(s_n))'$ and we want to predict $T = Z(s_0)$ and let $\hat{T} = \hat{Z}(s_0) = h(Z)$ (a function of the data $\mathbf{Z}$. We assume $\hat{T} = \mathbf{w}'\mathbf{Z}$).

Theorem 1

As discussed, kriging minimizes the mean square prediction error, $MSE(\hat{Z}(s_0)) = E(Z(s_0) - \hat{Z}(s_0))^2$. An important result is the following: $MSE(\hat{Z}(s_0))$ takes its minimum value when $\hat{Z}(s_0) = E(Z(s_0)|\mathbf{Z})$.

Proof

The proof is based on the expectation by conditioning. Here is an introductory example first! Let X, Y be random variables. We can find the expectation of Y using $EY = \int_{-\infty}^{\infty} f(y)dy$, however we can also find EY by conditioning on X:

$$EY = E_X\left[E_Y[Y|X]\right]$$

Note: The subscripts above mean that we take expectation with respect to that random variable. The proof of this very useful result is given next. Suppose Y, X have joint pdf $f(x, y)$. If $g(x, y)$ is a function of X, Y then $Eg(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)f(x, y)dxdy$.

$$\begin{aligned}
EY &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(y,x)dydx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(y|x)f(x)dydx \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} yf(y|x)dy\right] f(x)dx \\
&= \int_{-\infty}^{\infty} E(Y|X)f(x)dx \\
&= E_X\left[E_Y[Y|X]\right]
\end{aligned}$$

How do we use this result in the spatial prediction problem? We write the $MSE(\hat{Z}(s_0))$ as follows:

$$MSE(\hat{T}) = E(T - \hat{T})^2 = E_Z[E_T[T - \hat{T})^2|Z]]$$

Now using expectation operations (e.g. $EQ^2 = var(Q) + (EQ)^2$) for the inner expectation we get:

$$E_T[(T - \hat{T})^2|Z] = var_T[(T - \hat{T})|Z] + \left[E_T[(T - \hat{T})|Z]\right]^2 \tag{1}$$

However, conditioning on $\mathbf{Z}$, any function of $\mathbf{Z}$ is a constant and therefore we note the next two results:

$$var_T[(T - \hat{T})|Z] = var_T(T|Z) \text{ and } E_T[(T - \hat{T})|Z] = E_T(T|Z) - \hat{T}. \tag{2}$$

Using (2) we write (1) as follows:

$$E_T[(T-\hat{T})^2|Z] = var_T(T|Z) + \left[E_T(T|Z) - \hat{T}\right]^2 \tag{3}$$

Finally take expectation of (3) w.r.t. $\mathbf{Z}$ to get

$$E(T-\hat{T})^2 = E_Z\left[var_T(T|Z)\right] + E_Z\left[\left[E_T(T|Z) - \hat{T}\right]^2\right] \tag{4}$$

Equation (4) decomposed the $MSE(\hat{T})$ into two parts: The first term on the right-hand side does not depend on the choice of $\hat{T}$ and therefore we cannot do anything about it! However the second term on the right-hand side can take its minimum value when $\hat{T} = E(T|Z)$ which will be our predictor!

Before we actually find the predictor in our spatial prediction problem we present another important result from multivariate normal distribution.

Theorem 2

Suppose that $\mathbf{Y}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ are partitioned as follows $\mathbf{Y} = \begin{pmatrix} \mathbf{Y_1} \\ \mathbf{Y_2} \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$, and $\mathbf{Y} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. It can be shown that the conditional distribution of $\mathbf{Y_1}$ given $\mathbf{Y_2}$ is also multivariate normal, $\mathbf{Y_1}|\mathbf{Y_2} \sim MVN(\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2})$, where
$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{Y_2} - \boldsymbol{\mu}_2)$, and $\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$.

Apply now Theorem 1 and Theorem 2 in the spatial prediction problem. Assume the distribution of $\begin{pmatrix} Z_(s_0) \\ \mathbf{Z} \end{pmatrix}$ is multivariate normal with mean vector $\mu\mathbf{1}$ and variance covariance matrix $\begin{pmatrix} \sigma^2 & \mathbf{c}' \\ \mathbf{c} & \mathbf{C} \end{pmatrix}$, i.e. $\begin{pmatrix} Z_(s_0) \\ \mathbf{Z} \end{pmatrix} \sim N_{n+1}\left(\begin{pmatrix} \mu \\ \mu\mathbf{1} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \mathbf{c}' \\ \mathbf{c} & \mathbf{C} \end{pmatrix}\right)$,

Result

Using the previous theorems, the predictor that minimizes the mean square prediction error (see Theorem 1) will be (see Theorem 2) $\hat{Z}(s_0) = \mu + \mathbf{c}'\mathbf{C}^{-1}(\mathbf{Z} - \mu\mathbf{1})$, which is the simple kriging predictor. The prediction variance (also see Theorem 2) will be $\sigma^2 - \mathbf{c}'\boldsymbol{\Sigma}_{22}^{-1}\mathbf{c} = C(0) - \mathbf{c}'\mathbf{C^{-1}}\mathbf{c}$, which is the simple kriging variance.