4  Kriging the Mean

The mean value of samples from a geographical space can be estimated, either using the arithmetic mean or by constructing a weighted average integrating the knowledge of the spatial correlation of the samples. The two approaches are developed and it turns out that the solution of kriging the mean reduces to the arithmetic mean when there is no spatial correlation between data locations.

Arithmetic mean and its estimation variance

We denote by \( z \) the realization of a random variable \( Z \). Suppose we have \( n \) measurements \( z_\alpha \), where \( \alpha \) is an index numbering the samples from 1 to \( n \). This is our data. We introduce a probabilistic model by considering that each \( z_\alpha \) is a realization of a random variable \( Z_\alpha \) and we further assume that the random variables are independent and identically distributed (iid). This means that each sample is a realization of a separate random variable and that all these random variables have the same distribution. The random variables are assumed independent which implies that they are uncorrelated (the converse being true only for Gaussian random variables).

Let us assume we know the variance \( \sigma^2 \) of the random variables \( Z_\alpha \) (they have the same variance as they all have the same distribution). We do not know the mean \( m \) and wish to estimate it from the data \( z_\alpha \) using the arithmetic mean as an estimator

\[
m^*_A = \frac{1}{n} \sum_{\alpha=1}^{n} z_\alpha.
\]

(4.1)

Imagine we would take many times \( n \) samples \( z_\alpha \) under unchanged conditions. As we assume some randomness we clearly would not get identical values each time and the arithmetic mean would each time different. Thus we can consider \( m^*_A \) itself as a realization of a random variable and write

\[
M^*_A = \frac{1}{n} \sum_{\alpha=1}^{n} Z_\alpha.
\]

(4.2)

What is the average value of \( M^*_A \)? Our probabilistic model allows to compute it easily as

\[
E[ M^*_A ] = E\left[ \frac{1}{n} \sum_{\alpha=1}^{n} Z_\alpha \right] = \frac{1}{n} \sum_{\alpha=1}^{n} E[ Z_\alpha ] = m.
\]

(4.3)
So the estimated mean fluctuates around the true mean and is on average equal to it. If we call \( M_A^* - m \) the estimation error associated with the arithmetic mean, we see that the error is zero on average

\[
\mathbb{E}[M_A^* - m] = 0. \tag{4.4}
\]

The use of the arithmetic mean as an estimator of the true mean does not lead to a systematic error and the estimator is said to be unbiased.

What is the average fluctuation of the arithmetic mean around the true mean? This can be characterized by the variance \( \sigma_A^2 \) of the estimator

\[
\sigma_A^2 = \text{var}(M_A^*) = \text{var}(\frac{1}{n} \sum_{\alpha=1}^{n} Z_\alpha) = \frac{1}{n^2} \text{var}(\sum_{\alpha=1}^{n} Z_\alpha) \tag{4.5}
\]

using relation (2.20). Applying relation (2.25) to the uncorrelated identically distributed random variables we get

\[
\frac{1}{n^2} \sum_{\alpha=1}^{n} \text{var}(Z_\alpha) = \frac{\sigma^2}{n}. \tag{4.6}
\]

So the distribution of \( M_A^* \) has a variance \( n \) times smaller than that of the random variables \( Z \). We can view the variance of the estimator also as an estimation variance \( \sigma_E^2 \), i.e. the variance of the estimation error \( M_A^* - m \), because by relation (2.21) we have

\[
\sigma_E^2 = \text{var}(M_A^* - m) = \text{var}(M_A^*) = \frac{\sigma^2}{n}. \tag{4.7}
\]

**Estimating the mean with spatial correlation**

When samples have been taken at irregular spacing in a domain \( D \) like on Figure 4.1, a quantity of interest is the value of the mean \( m \). We assume again that each sample \( z(x_\alpha) \) is a realization of a random variable \( Z(x_\alpha) \) and that the random variables are identically distributed.

A first approach for estimating \( m \) is to use the arithmetic mean

\[
M_A^* = \frac{1}{n} \sum_{\alpha=1}^{n} Z(x_\alpha). \tag{4.8}
\]

However the samples from a spatial domain cannot in general be assumed independent. This implies that the random variables at two different locations are usually correlated, especially when they are near to each other in space.

A second approach to estimate \( m \) is to use a weighted average

\[
M^* = \sum_{\alpha=1}^{n} w_\alpha Z(x_\alpha) \tag{4.9}
\]

with unknown weights \( w_\alpha \).

How best choose the weights \( w_\alpha \)? We have to specify the problem further.