

## 0.1. Notas en la obtención de las ecuaciones de Kriging

### 0.1.1. definiciones básicas <sup>1</sup>

Consideramos un campo continuo e invariante en el tiempo,

$$\phi : x \in D \rightarrow \phi(x) \in \mathbb{R} : D \subset \mathbb{R}^m$$

Modelamos  $\phi$  como un proceso gaussiano,

$$\Omega(x) = r(x)^T \beta + Z(x), \quad (1)$$

Donde  $r(x)$  es un vector de funciones de regresión conocidas,  $r(x) = [r_1(x), r_2(x), \dots, r_k(x)]^T$  y  $\beta = [\beta_1, \beta_2, \dots, \beta_k]^T$  es un vector de parámetros, también conocidos.

$Z(x)$  es un proceso gaussiano de **media zero**, cuya función de covarianza<sup>2</sup> modelamos como,

$$\begin{aligned} \text{cov}(Z(x), Z(x')) &= \sigma_z^2 \xi(x, x') \\ \xi(x, x') &= \exp \left[ - \sum_i^m \frac{1}{\theta_i} |x_i - x'_i|^{q_i} \right] \end{aligned} \quad (2)$$

Donde  $x_i$  es la componente  $i$  del vector  $x$ .

En principio, si asumimos un modelo como el de la ecuación 1 El modelo corresponde en geoestadística a lo que se conoce con el nombre de *Kriging Universal*. Hay más posibilidades si definimos, en general el modelo como,

$$Y(x) = \mu(x) + Z(x)$$

Con  $Z(\cdot)$  un proceso estocástico general de media cero y  $\mu(\cdot)$  un modelo paramétrico que expresa el valor esperado (la media) del proceso en cada punto. Podemos definir los siguientes tipos de Kriging:

1. **Kriging Simple:** Se supone conocida la estructura tanto de la media  $\mu(x)$  como de la covarianza de la parte estocástica del modelo  $\text{cov}[Z(x), Z(x')]$ .
2. **Kriging Ordinario:** La media  $\mu(x) = \mu$  es desconocida pero se asume constante. Además, la función aleatoria  $Z(x)$  se supone estacionaria.
3. **Universal Kriging:** Corresponde al modelo que define la ecuación 1, Donde se emplean funciones de regresión conocidas. En general, se supone  $Z(x)$  es estacionaria.

Que el proceso modelado mediante kriging sea estacionario, tiene como consecuencia que la función de covarianza del proceso, depende de la diferencia de posiciones  $\text{cov}(Z(x), Z(x')) = G(x - x')$  entre puntos. Esto guarda relación con la idea del semivariograma. En general un semivariograma ofrece

<sup>1</sup>Sigo siempre la notación del artículo de Kahn

<sup>2</sup>El valor esperado de una función gaussiana  $E(Z(x)) = m$  es la media de la función, la covarianza de dos variables aleatorias se define como  $\text{cov}(X, Y) = E(X - E(X)) \cdot E(Y - E(Y))$ , para el proceso gaussiano que estamos describiendo, la covarianza entre los valores de dos puntos sería  $\text{cov}(Z(x), Z(x')) = E[(Z(x) - E(Z(x))) \cdot (Z(x') - E(Z(x')))] = E(Z(x)Z(x'))$ , porque la media de  $Z$  es cero.

un modelo para estimar la varianza del proceso estocástico entre dos puntos,  $\gamma(x) = \frac{1}{2}\text{var}(Z(x) - Z(0))$ . En algunas aplicaciones se emplean modelos, como el empleado en la ecuación 2,  $\gamma(h) = \frac{1}{2}\xi(x, x+h)$ . Los resultados van a depender entonces de cómo se estimen los parámetros del semivariograma y de que modelo se use.

En aplicaciones con un número adecuado y distribuido de medidas del proceso  $(x)$ , se puede estimar con la siguiente ecuación:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x^{(i)}) - Z(x^{(i)} + h)]^2$$

donde:

- $\gamma(h)$  es la semivarianza para puntos separados una distancia  $h$ ,
- $Z(x^{(i)})$  es el valor del campo en  $x^{(i)}$ ,
- $N(h)$  es el número de pares de datos (medidas del campo) que se tiene a una distancia  $h$ .

Estrictamente hablando, el semivariograma debería ser la representación gráfica de la semivarianza y, además, para el caso estimado, al depender solo de la distancia, y no de las posiciones relativas nuestro campo ha de ser además de estacionario isotrópico.

Supongamos que tenemos un conjunto de  $n$  medidas ruidosas, de nuestro campo  $y = [y(x^{(1)}), y(x^{(2)}), \dots, y(x^{(n)})]$ . Nuestras medidas son una realización del proceso gaussiano,

$$y = R(x)\beta + \mathcal{Z} \quad (3)$$

Donde,

$$R = \begin{pmatrix} r_1(x^{(1)}) & \dots & r_k(x^{(1)}) \\ \vdots & \ddots & \vdots \\ r_1(x^{(n)}) & \dots & r_k(x^{(n)}) \end{pmatrix} \quad (4)$$

con  $\mathcal{Z} = [Z(x^{(1)}), \dots, Z(x^{(n)})]^T$ . En otras palabras; apilamos los resultados de nuestras medidas.

Supongamos que queremos construir un predictor lineal  $\hat{\Omega}(x) = a_x^T y$  a partir de las medidas. Es decir yo quiero inferir cual sería el valor de mi campo  $E[\Omega]$  en un punto  $x$ , supuesto que conozco los valores en los puntos medidos. De alguna manera, lo que estoy definiendo es una media ponderada de los valores que he medido.

Podemos definir entonces el valor esperado del error cuadrático medio a partir de la diferencia entre las medidas y el modelo gaussiano,

$$\begin{aligned} E[(a_x^T y - \Omega(x))^2] &= E[a_x^T y y^T a_x + \Omega(x)^2 - 2a_x^T y \Omega(x)] = \\ &= E[a_x^T (R\beta + \mathcal{Z})((R\beta + \mathcal{Z})^T a_x) + \\ &= E[(r^T(x)\beta + Z(x))^2] + \\ &= E[-2a_x^T (R\beta + \mathcal{Z})(r^T(x)\beta + Z(x))] \end{aligned} \quad (5)$$

Donde hemos hecho uso de que  $y = R\beta + \mathcal{Z}$  y  $\Omega(x) = r^T(x)\beta + Z(x)$ .

Desarrollamos ahora por separado cada uno de los tres sumandos de la ecuación (5),

1.

$$\begin{aligned} E[a_x^T(R\beta + \mathcal{Z})((R\beta + \mathcal{Z})^T a_x)] &= \\ a_x^T R\beta\beta^T R^T a_x + 2a_x^T R\beta E[\mathcal{Z}^T] a_x + a_x^T E[\mathcal{Z}\mathcal{Z}^T] a_x &= \\ = a_x^T R\beta\beta^T R^T a_x + a_x^T \sigma_z^2 K a_x \end{aligned} \quad (6)$$

Donde hemos hecho uso de que el valor esperado del vector  $\mathcal{Z}$  tiene que ser nulo. Y además  $E[\mathcal{Z}\mathcal{Z}^T]$  es una matriz de covarianza que podemos definir a partir de la ecuación 2 como  $\sigma_z K$  con  $K_{i,j} = \xi(x^{(i)}, x^{(j)})$ .

2.

$$\begin{aligned} E[(r^T(x)\beta + Z(x))^2] &= \\ r(x)^T \beta\beta^T r(x) + 2r(x)^T \beta E[Z(x)] + E[Z(x)^2] &= \\ = r^T \beta\beta^T r + \sigma_z^2 \end{aligned} \quad (7)$$

De nuevo, ecuación (2),  $E[Z(x)] = 0$  y  $\text{cov}(Z(x), Z(x)) = \sigma_z^2$

3.

$$\begin{aligned} E[-2a_x^T(R\beta + \mathcal{Z})(r^T(x)\beta + Z(x))] &= \\ -2a_x^T R\beta r(x)^T \beta - 2a_x^T R\beta E[Z(x)] - 2a_x^T E[\mathcal{Z}]r(x)^T \beta - 2a_x^T E[\mathcal{Z}Z(x)] &= \\ = -2a_x^T R\beta r(x)^T \beta - 2a_x^T \sigma_z k_x \end{aligned} \quad (8)$$

Donde,  $E[\mathcal{Z}Z(x)] := k_x = \sigma_z^2(\xi(x, x_1), \xi(x, x_2), \dots, \xi(x, x_n))^T$  y las esperanzas de  $Z$  y  $\mathcal{Z}$  son cero.

Llegamos por tanto a que,

$$\begin{aligned} E[(a_x^T y - \Omega(x))^2] &= \\ a_x^T R\beta\beta^T R^T a_x + a_x^T \sigma_z^2 K a_x + r(x)^T \beta\beta^T r(x) + \sigma_z^2 - 2a_x^T R\beta r(x)^T \beta - 2a_x^T \sigma_z k_x &= \\ = (a_x^T R\beta - r(x)^T \beta)^2 + a_x^T \sigma_z^2 K a_x + \sigma_z^2 - 2a_x^T \sigma_z^2 k_x \end{aligned} \quad (9)$$

A nosotros nos gustaría minimizar esta expresión, ya que eso es tanto como minimizar el error esperado. Si nos fijamos en el primer término cuadrático, si nuestro ajuste es bueno, deberíamos hacer este término cero:  $R^T a_x = r(x)$ , Si lo hacemos así, estamos construyendo un estimador sin sesgo.  $Y = a_x^T y = a_x^T R\beta$  donde  $Y = E[\Omega]$  es el valor de la variable estocástica que queremos predecir  $Y = r(x)^T \beta$ .

Por tanto deberíamos minimizar el resto de expresión,

$$\begin{aligned} \min_{a_x} [a_x^T \sigma_z^2 K a_x + \sigma_z^2 - 2a_x^T \sigma_z^2 k_x] \\ \text{sujeto a:} \\ R^T a_x = 0 \end{aligned} \quad (10)$$

Si derivamos con respecto a  $a^T$ , igualamos a cero, y añadimos un multiplicadores de Lagranje  $-R\lambda$  a la expresión derivada, obtenemos:

$$\begin{aligned} K a_x - k_x - R\lambda &= 0 \\ R^T a_x &= r(x) \end{aligned} \quad (11)$$

Podemos describir las ecuaciones en forma matricial,

$$\begin{pmatrix} K & R \\ R^T & 0 \end{pmatrix} \begin{pmatrix} a_x \\ -\lambda \end{pmatrix} = \begin{pmatrix} k_x \\ r(x) \end{pmatrix} \quad (12)$$

$$\begin{pmatrix} a_x \\ -\lambda \end{pmatrix} = \begin{pmatrix} K & R \\ R^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} k_x \\ r(x) \end{pmatrix} \quad (13)$$

Podemos calcular la inversa de esta matriz, empleando los bonitos complementos de Shür y la eliminación de Gauss-Jordan por bloques<sup>3</sup>,

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \quad M^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}$$

Si lo aplicamos a nuestro sistema,

$$\begin{pmatrix} a_x \\ -\lambda \end{pmatrix} = \begin{pmatrix} K^{-1} - K^{-1}R(R^TK^{-1}R)^{-1}R^TK^{-1} & K^{-1}RR^T(K^{-1}R)^{-1} \\ (R^TK^{-1}R)^{-1}R^TK^{-1} & -(R^TK^{-1}R)^{-1} \end{pmatrix} \begin{pmatrix} k_x \\ r(x) \end{pmatrix} \quad (14)$$

Con lo que obtenemos para  $a_x$ ,

$$a_x = (K^{-1} - K^{-1}R(R^TK^{-1}R)^{-1}R^TK^{-1})k_x + K^{-1}R(R^TK^{-1}R)^{-1}r(x) \quad (15)$$

Luego nuestra estimación del valor de  $\Omega(x)$  sería,

$$\begin{aligned} \Omega(x) &= a_x^T y = \\ &= [k_x^T (K^{-1} - K^{-1}R(R^TK^{-1}R)^{-1}R^TK^{-1}) + r(x)^T (R^TK^{-1}R)^{-1}R^TK^{-1}] y = \\ &= k_x^T K^{-1}(y - R\beta) + r(x)^T \beta \end{aligned} \quad (16)$$

Que es precisamente la ecuación (18) del artículo con,

$$\beta = (R^TK^{-1}R)^{-1}R^TK^{-1}y \quad (17)$$

En realidad, si somos rigurosos, el valor de  $\beta$  que estamos empleando es una estimación, y deberíamos llamarle  $\hat{\beta}$ , siguiendo una notación que es más o menos estándar, para dejar claro que es un valor estimado.

Lo que hemos obtenido es el valor esperado para el modelo lineal que suponemos siguen nuestros predictores. Nos faltaría calcular la varianza. En realidad, coincide con el valor cuadrático medio que hemos optimizado,

$$\begin{aligned} \sigma_\phi^2(x) &= E[(a_x^T y - \Omega(x))^2] = \\ &= E[(k_x^T K^{-1}(y - R\beta) + r(x)^T \beta - r(x)^T \beta - Z(x))^2] = \\ &= E[(k_x^T K^{-1}(y - R\beta) + Z(x))^2] = \\ &= E[(k_x^T K^{-1}(R\beta + \mathcal{Z} - R\beta) - Z(x))^2] = \\ &= E[(k_x^T K^{-1}\mathcal{Z} - Z(x))^2] \end{aligned} \quad (18)$$

Si desarrollamos el cuadrado de la última expresión de la ecuación (18).

---

<sup>3</sup>[https://en.wikipedia.org/wiki/Schur\\_complement](https://en.wikipedia.org/wiki/Schur_complement)

$$\begin{aligned}
\sigma_\phi^2(x) &= E[(k_x^T K^{-1} \mathcal{Z} + Z(x))^2] = \\
E [k_x^T K^{-1} \mathcal{Z} \mathcal{Z}^T K^{-1} k_x - 2k_x^T K^{-1} \mathcal{Z} Z(x) + Z(x)^2] &= \\
k_x^T K^{-1} E[\mathcal{Z} \mathcal{Z}^T] K^{-1} k_x - 2k_x^T K^{-1} \sigma_z k_x + \sigma_z^2 &= \\
k_x^T K^{-1} \sigma_z K K^{-1} k_x - 2k_x^T K^{-1} \sigma_z k_x + \sigma_z^2 &= \\
= (1 - k_x^T K^{-1} k_x) \sigma_z^2 &
\end{aligned} \tag{19}$$

## 0.2. variogramas, semivariogramas, (auto)correlación y covarianza

Las tres cosas están relacionadas. Hay que ir poco a poco con los detalles pero en primera aproximación todas dan una medida de la correlación espacial entre los valores de un campo aleatorio  $f(\mathbf{x}) \in \mathbb{R}$ .

Esta correlación se hace depender en principio solo de la distancia entre puntos  $y$ , si acaso, también en la dirección. Lo habitual es definir un modelo para la correlación  $y$ , a partir de él definir las otras dos magnitudes,

Variograma:

$$\gamma(r) = \sigma^2 \cdot (1 - \rho(r)) + n \tag{20}$$

Debería comprobarlo, pero no estoy seguro si esto es el variograma o el semivariograma. En cualquier caso uno y otro se diferencian en un factor 2.

Covarianza:

$$C(r) = \sigma^2 \cdot \rho(r) \tag{21}$$

Veamos los detalles.  $r$  es la distancia entre los puntos para los que se quiere calcular el valor del variograma o del de la covarianza  $\sigma^2$  es la varianza del proceso.  $n$  es el llamado nugget nos da un valor distinto de cero para  $r = 0$ . Más adelante lo veremos más claro con un ejemplo.

Función de correlación,

$$\rho(r) = \text{cor}\left(s \cdot \frac{r}{l}\right) \tag{22}$$

Se trata de una función de correlación normaliza.  $s$  es un factor de escala para normalizar y  $l$  representa la distancia máxima de correlación. Es decir, la distancia a partir de la cual los valores del campo ya no se consideran correlacionados.

Veamos un ejemplo típico, supongamos que elegimos para la correlación un modelo gaussiano,

$$\phi(r) = e^{-(\frac{r}{l})^2} \tag{23}$$

Donde tomamos directamente  $s = 1$ . Es facil ver que el modelo nos da una correlación 1 para  $r = 0$  y la correlación tiende a 0 cuando  $r \rightarrow \infty$