Kriging distribution under HSGP proof

Christine Shen

April 1, 2025

On page 18 of the scalable GP 1 slides, we claim that if $m \leq n$, the kriging distribution

$$\boldsymbol{\theta}^* \mid (\boldsymbol{\theta}, \Omega) = (\Phi^* \mathbf{S} \Phi^\top)(\Phi \mathbf{S} \Phi^\top)^\dagger \boldsymbol{\theta},$$

where $\mathbf{A}^\dagger$ denotes a generalized inverse of matrix $\mathbf{A}$ such that $\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}$.

Proof. First we note that by properties of multivariate normal, $\boldsymbol{\theta}^* \mid (\boldsymbol{\theta}, \Omega) \sim N_q(\mathbb{E}^{HS}_{\boldsymbol{\theta}^*}, \mathbb{V}^{HS}_{\boldsymbol{\theta}^*})$,

$$\begin{aligned}
\mathbb{E}^{HS}_{\boldsymbol{\theta}^*} &= (\Phi^* \mathbf{S} \Phi^\top)(\Phi \mathbf{S} \Phi^\top)^\dagger \boldsymbol{\theta} \\
\mathbb{V}^{HS}_{\boldsymbol{\theta}^*} &= (\Phi^* \mathbf{S} \Phi^{*\top}) - (\Phi^* \mathbf{S} \Phi^\top)(\Phi \mathbf{S} \Phi^\top)^{\dagger\top}(\Phi \mathbf{S} \Phi^{*\top}).
\end{aligned}$$

I won't prove this statment here. But just to highlight that if we don't insist on using a symmetric generalized inverse matrix, then there needs to be a "transpose" in the expression for the covariance matrix, i.e., it is not $(\Phi\mathbf{S}\Phi)^\dagger$, but rather $(\Phi\mathbf{S}\Phi)^{\dagger\top}$ that should be used. This is different from the expression you might find on the Wiki page for conditional distribution for multivariate normal. This is because, the source for the Wiki results is actually using a specific form of generalized inverse which is symmetric.

Next, we show that if $m \leq n$, $\mathbb{V}^{HS}_{\boldsymbol{\theta}^*} \equiv \mathbf{0}$ in order to prove the statement. This is equivalent to showing

$$(\Phi^* \mathbf{S} \Phi^{*\top}) = (\Phi^* \mathbf{S} \Phi^\top)(\Phi \mathbf{S} \Phi^\top)^{\dagger\top}(\Phi \mathbf{S} \Phi^{*\top}).$$

It's also equivalent to showing

$$\mathbf{S} = \mathbf{S}\Phi^\top(\Phi \mathbf{S} \Phi^\top)^{\dagger\top}\Phi\mathbf{S}.$$

Note that under HSGP, if $m \leq n$, Φ has full column rank. Therefore for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$,

$$\Phi\mathbf{x} = \Phi\mathbf{y} \implies \mathbf{x} = \mathbf{y}, \quad \mathbf{x}^\top\Phi^\top = \mathbf{y}^\top\Phi^\top \implies \mathbf{x} = \mathbf{y}. \tag{1}$$

By the definition of generalized inverse, we have

$$(\Phi \mathbf{S} \Phi^\top)(\Phi \mathbf{S} \Phi^\top)^\dagger(\Phi \mathbf{S} \Phi^\top) = (\Phi \mathbf{S} \Phi^\top). \tag{2}$$

Taking transpose of both sides of the equation, we have

$$(\Phi \mathbf{S} \Phi^\top)(\Phi \mathbf{S} \Phi^\top)^{\dagger\top}(\Phi \mathbf{S} \Phi^\top) = (\Phi \mathbf{S} \Phi^\top).$$

Now applying equation (1), we have

$$\mathbf{S}\Phi^\top(\Phi \mathbf{S} \Phi^\top)^{\dagger\top}\Phi\mathbf{S} = \mathbf{S}.$$

which is exactly what's needed to show $\mathbb{V}^{HS}_{\boldsymbol{\theta}^*} \equiv \mathbf{0}$ as argued above.

Applying the same arguments to equation (2) gives

$$\mathbf{S}\Phi^\top(\Phi\mathbf{S}\Phi^\top)^\dagger\Phi\mathbf{S} = \mathbf{S}.$$

This is needed to simplify the expression for $\boldsymbol{\theta}^*$ under the reparameterized model on slide 29. $\square$