Analytical calculation of Sobol sensitivity indices for Gaussian Processes with a squared exponential covariance function

Milias Argeitis, Andreas; Kurdyaeva, Tamara

Publication date:
2018

Analytical calculation of Sobol sensitivity indices for Gaussian Processes with a squared exponential covariance function

Notation and modeling assumptions

- Column vector of random inputs: $\mathbf{X} = [X_1 ... X_d]^T \in \mathbb{R}^d$, $\mathbf{X}' = [X'_1 ... X'_d]^T \in \mathbb{R}^d$,
- Column subvector of $\mathbf{X}$ without X_s element: $\mathbf{X}_{-s} = [X_1 \dots X_{s-1}\ X_{s+1} \dots X_d]^T \in \mathbb{R}^{d-1}$
- Column vector consists of elements $\mathbf{X}'_{-s}$ and X_s: $\hat{\mathbf{X}} = [X'_1 \dots X_s \dots X'_d]^T \in \mathbb{R}^d$
- Random inputs X_s, $s = 1, ..., d$ are mutually independent and each input follows a standard uniform distribution: $X_s \sim \mathcal{U}(0, 1)$, $s = 1, ..., d$
- Function F represents a black-box model of interest: $Y = F(X)$, where Y is the model output
- Observed n training data points is denoted by $\widetilde{X} = (\widetilde{\mathbf{x}}^{(1)}, \dots, \widetilde{\mathbf{x}}^{(n)})$
- Test n_* data points is denoted by $X_* = \left(\mathbf{x}_*^{(1)}, \dots, \mathbf{x}_*^{(n_*)}\right)$
- Vector of function F values at n training points $\widetilde{X}$: $\widetilde{\mathbf{y}} = \left[\widetilde{y}^{(1)}, \dots, \widetilde{y}^{(n)}\right]^T$, $\widetilde{y}^{(i)} = F\left(\widetilde{\mathbf{x}}^{(i)}\right)$, $i = 1, ..., n$
- Gaussian process (GP) parametrization $f(x) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$

$$
\begin{aligned}
k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \prod_{j=1}^{d} \exp\left(-\frac{(x_j - x'_j)^2}{2\, l_j^2}\right) &= * \left| h_j = \frac{1}{2} l_j^{-2} \right| \\
= \sigma_f^2 \prod_{j=1}^{d} \exp\left(-h_j (x_j - x'_j)^2\right) & \qquad (1)
\end{aligned}
$$

- Posterior GP distribution $f_* | \widetilde{X}, \widetilde{\mathbf{y}}, X_* \sim \mathcal{N}(m(X_*), v(X_*))$, where

$$
m(X_*) = K(X_*, \widetilde{X}) \left[K(\widetilde{X}, \widetilde{X}) + \sigma^2 I\right]^{-1} \widetilde{\mathbf{y}}, \qquad (2)
$$

$$
v(X_*) = K(X_*, X_*) - K(X_*, \widetilde{X}) \left[K(\widetilde{X}, \widetilde{X}) + \sigma^2 I\right]^{-1} K(\widetilde{X}, X_*), \qquad (3)
$$

  where the additive term $\sigma^2 I$ reflects the fact that the function values are corrupted by additive noise with variance σ^2. Even if $\sigma^2 = 0$ (as is the case when the function is exactly evaluated), a small noise term is typically inserted to regularize the inversion of the covariance matrix $K(\widetilde{X}, \widetilde{X})$.

- $K = K(\widetilde{X}, \widetilde{X})$
- k_{pq} for $p = 1, ..., n$ and $q = 1, ..., n$ denotes the entries of the matrix $\left[K + \sigma^2 I\right]^{-1}$
- w_p, $p = 1, ..., n$ are the elements of the vector $\left[K + \sigma^2 I\right]^{-1} \widetilde{\mathbf{y}}$
- Expectations and variances with respect to the GP posterior distribution are denoted as $\mathbb{E}_*$ and $\mathbb{V}\mathrm{ar}_*$, respectively
- $\mathrm{erf}(\cdot)$ denotes the error function.

Sobol indices

The estimate of the Sobol indices for the output Y with respect to the input X_s

$$\boxed{\hat{S}_s = \frac{\mathbb{E}_*\{\mathrm{Var}\{E\,(Y|X_s)\}\}}{\mathbb{E}_*\{\mathrm{Var}(Y)\}},\ s = 1, ..., d.} \tag{4}$$

Inference for variance in numerator $\mathbb{E}_*\{\mathrm{Var}\{E\,(Y|X_s)\}\}$

Let us first calculate the numerator in the formula for Sobol indices (4).

$$\begin{aligned} \mathbb{E}_*\{\mathrm{Var}\{E(Y|X_s)\}\} &= \mathbb{E}_*\{E\,(E(Y|X_s))^2\} - \mathbb{E}_*\{E(Y)\}^2 = \\ &= \mathbb{E}_*\{E\,(E(Y|X_s))^2\} - \mathbb{V}\mathrm{ar}_*\{E(Y)\} - (\mathbb{E}_*\{E(Y)\})^2 = \\ &= I - II - III \end{aligned} \tag{5}$$

First component I from (5): $\mathbb{E}_*\{E\,(E(Y|X_s))^2\}$

We derive the formulas for every component of (5) separately. The first term I requires the computation of the following integrals:

$$\begin{aligned} \mathbb{E}_*\{E\,(E(Y|X_s))^2\} &= E\left[\mathbb{V}\mathrm{ar}_*\{E(Y|X_s)\} + (\mathbb{E}_*\{E(Y|X_s)\})^2\right] = \\ &= \int_0^1 \int_{[0,1]^{d-1}} \int_{[0,1]^{d-1}} v(\mathbf{x}, \hat{\mathbf{x}})\ d\mathbf{x}_{-s} d\mathbf{x}'_{-s} dx_s + \int_0^1 \int_{[0,1]^{d-1}} \int_{[0,1]^{d-1}} m_*(\mathbf{x}) m_*(\hat{\mathbf{x}}) d\mathbf{x}_{-s} d\mathbf{x}'_{-s} dx_s = \\ &= L + M. \end{aligned} \tag{6}$$

For the simplification we also provide the derivation of the formulas for the individual terms L and M in (6) separately.

$$\begin{aligned} L &= \int_0^1 \int_{[0,1]^{d-1}} \int_{[0,1]^{d-1}} v(\mathbf{x}, \hat{\mathbf{x}})\ d\mathbf{x}_{-s} d\mathbf{x}'_{-s} dx_s \overset{3}{=} \\ &= \int_0^1 \int_{[0,1]^{d-1}} \int_{[0,1]^{d-1}} K(\mathbf{x}, \hat{\mathbf{x}})\ d\mathbf{x}_{-s} d\mathbf{x}'_{-s} dx_s - \int_0^1 \int_{[0,1]^{d-1}} \int_{[0,1]^{d-1}} K(\mathbf{x}, \tilde{X}) \left(K + \sigma^2\right)^{-1} K(\tilde{X}, \hat{\mathbf{x}})\ d\mathbf{x}_{-s} d\mathbf{x}'_{-s} dx_s = \\ &= \int_0^1 Z_{-s}(\hat{\mathbf{x}}) dx_s - \int_0^1 \left[z_{-s}^T(\mathbf{x}) \left(K + \sigma^2\right)^{-1} z_{-s}(\hat{\mathbf{x}})\right] dx_s \end{aligned} \tag{7}$$

Now we divide the calculation of the integrals into several parts to make it easier to follow

$$\begin{aligned} z_{-s}(\mathbf{x}) &= \int_{[0,1]^{d-1}} K(\mathbf{x}, \tilde{X})\ d\mathbf{x}_{-s} \overset{(1}{)} = \sigma_f^2 \exp\left(-h_s(x_s - \tilde{x}_s)^2\right) \prod_{j=-s} \int_0^1 \exp\left(-h_j(x_j - \tilde{x}_j)^2\right)\ dx_j = \\ &= \sigma_f^2 \exp\left(-h_s(x_s - \tilde{x}_s)^2\right) \prod_{j=-s} I_j \end{aligned} \tag{8}$$

Next we evaluate the integral defined as I_j in (8)

$$I_j = \int_0^1 \exp\left(-h_j(x_j - \tilde{x}_j)^2\right)\, dx_j = * \left| a_j = \sqrt{h_j}(x_j - \tilde{x}_j) \right| = \frac{1}{\sqrt{h_j}} \int_{-\sqrt{h_j}\tilde{x}_j}^{\sqrt{h_j}(1-\tilde{x}_j)} \exp\left(-a_j^2\right)\, da_j =$$

$$= \frac{\sqrt{\pi}}{2}\frac{1}{\sqrt{h_j}}\left(\mathrm{erf}\left(\sqrt{h_j}(1-\tilde{x}_j)\right) + \mathrm{erf}\left(\sqrt{h_j}\tilde{x}_j\right)\right) = \sqrt{\frac{\pi}{2}} l_j \left(\mathrm{erf}\left(\frac{1}{\sqrt{2}l_j}(1-\tilde{x}_j)\right) + \mathrm{erf}\left(\frac{1}{\sqrt{2}l_j}\tilde{x}_j\right)\right) \tag{9}$$

We can substitute the exact formula of I_j (9) into the final formula of z_{-s} in (8)

$$z_{-s}(\mathbf{x}) = \sigma_f^2 \exp\left(-\frac{1}{2l_s^2}(x_s - \tilde{x}_s)^2\right) \prod_{j=-s} \sqrt{\frac{\pi}{2}} l_j \left(\mathrm{erf}\left(\frac{1}{\sqrt{2}l_j}(1-\tilde{x}_j)\right) + \mathrm{erf}\left(\frac{1}{\sqrt{2}l_j}\tilde{x}_j\right)\right) \tag{10}$$

Finally we calculate the first double integral in (7) which was marked as Z_{-s}

$$Z_{-s} = \int_{[0,1]^{d-1}} \int_{[0,1]^{d-1}} K(\mathbf{x}, \hat{\mathbf{x}}) d\mathbf{x}_{-s} d\mathbf{x}'_{-s} = \sigma_f^2 \prod_{j=-s} \int_0^1 \left[\int_0^1 \exp\left(-h_j(x_j - x'_j)^2\right)\, dx_j \right] dx'_j = \sigma_f^2 \prod_{j=-s} \int_0^1 I_j dx'_j =$$

$$= \sigma_f^2 \prod_{j=-s} \sqrt{\frac{\pi}{2}} l_j \int_0^1 \left(\mathrm{erf}\left(\frac{1}{\sqrt{2}l_j}(1-x'_j)\right) + \mathrm{erf}\left(\frac{1}{\sqrt{2}l_j}x'_j\right)\right) dx'_j = \sigma_f^2 \prod_{j=-s} J_j \tag{11}$$

We present the exact analytic calculation of the J_j integral in (11)

$$J_j = \sqrt{\frac{\pi}{2}} l_j \left[\int_0^1 \mathrm{erf}\left(\frac{1}{\sqrt{2}l_j}(1-x'_j)\right) dx'_j + \int_0^1 \mathrm{erf}\left(\frac{1}{\sqrt{2}l_j}x'_j\right) dx'_j \right] =$$

$$= * \left| b_j = \frac{1}{\sqrt{2}l_j}(1-x'_j),\ c_j = \frac{1}{\sqrt{2}l_j}x'_j \right| =$$

$$= \sqrt{\frac{\pi}{2}} l_j \left[-\sqrt{2}l_j \int_{1/\sqrt{2}l_j}^{0} \mathrm{erf}(b_j) db_j + \sqrt{2}l_j \int_0^{1/\sqrt{2}l_j} \mathrm{erf}(c_j) dc_j \right] =$$

$$= 2\sqrt{\pi} l_j^2 \int_0^{1/\sqrt{2}l_j} \mathrm{erf}(t_j) dt_j =$$

$$= * \left| u = \mathrm{erf}(t_j), dv = dt_j, du = \frac{2}{\sqrt{\pi}} \exp\left(-t_j^2\right) dt_j, v = t_j \right| =$$

$$= 2\sqrt{\pi} l_j^2 \left[\mathrm{erf}(t_j) t_j \Big|_0^{1/\sqrt{2}l_j} - \frac{2}{\sqrt{\pi}} \int_0^{1/\sqrt{2}l_j} \exp\left(-t_j^2\right) t_j dt_j \right] =$$

$$= 2\sqrt{\pi} l_j^2 \left[\frac{1}{\sqrt{2}l_j} \mathrm{erf}\left(\frac{1}{\sqrt{2}l_j}\right) + \frac{1}{\sqrt{\pi}} \exp\left(-t_j^2\right) \Big|_0^{1/\sqrt{2}l_j} \right] =$$

$$= 2\sqrt{\frac{\pi}{2}} l_j \mathrm{erf}\left(\frac{1}{\sqrt{2}l_j}\right) + 2l_j^2 \left(\exp\left(-\frac{1}{2l_j^2}\right) - 1\right) \tag{12}$$

Then we can write down a closed form to compute Z_{-s} in (11)

$$Z_{-s} = \sigma_f^2 \prod_{j=-s} \left[2\sqrt{\frac{\pi}{2}} l_j \mathrm{erf}\left(\frac{1}{\sqrt{2} l_j} \right) + 2 l_j^2 \left(\exp\left(-\frac{1}{2 l_j^2} \right) - 1 \right) \right] \tag{13}$$

Finally, the analytic formula for L in (7) is simplified to the following form

$$\begin{aligned} L &= \sigma_f^2 \prod_{j=-s} J_j - \sigma_f^4 \sum_{p=1}^{n} \sum_{q=1}^{n} \left[k_{pq} \prod_{j=-s} I_j(\mathbf{x}_p) \prod_{j=-s} I_j(\mathbf{x}_q) \int_0^1 \exp\left(-h_s \left[(x_s - \tilde{x}_s^p)^2 + (x_s - \tilde{x}_s^q)^2 \right] \right) dx_s \right] = \\ &= \sigma_f^2 \prod_{j=-s} J_j - \sigma_f^4 \sum_{p=1}^{n} \sum_{q=1}^{n} \left[k_{pq} \prod_{j=-s} I_j(\mathbf{x}_p) \prod_{j=-s} I_j(\mathbf{x}_q)\, R_{pqs} \right] \end{aligned} \tag{14}$$

The derivation of the integral R_{pqs} is presented separately:

$$\begin{aligned} R_{pqs} &= \int_0^1 \exp\left(-h_s \left[(x_s - \tilde{x}_s^p)^2 + (x_s - \tilde{x}_s^q)^2 \right] \right) dx_s = \\ &= * \left| a^2 + b^2 = \frac{1}{2} \left[(a+b)^2 + (a-b)^2 \right] \right| = \\ &= \exp\left[-\frac{h_s}{2} \left(\tilde{x}_s^p - \tilde{x}_s^q \right)^2 \right] \int_0^1 \exp\left[-\frac{h_s}{2} \left(2x_s - \tilde{x}_s^p - \tilde{x}_s^q \right)^2 \right] dx_s = \\ &= * \left| g_s = \sqrt{\frac{h_s}{2}} \left(2x_s - \tilde{x}_s^p - \tilde{x}_s^q \right),\ dg_s = \sqrt{2h_s} dx_s,\ g_s(0) = -\sqrt{\frac{h_s}{2}} \left(\tilde{x}_s^p + \tilde{x}_s^q \right),\ g_s(1) = \sqrt{\frac{h_s}{2}} \left(2 - \tilde{x}_s^p - \tilde{x}_s^q \right) \right| = \\ &= \exp\left[-\frac{h_s}{2} \left(\tilde{x}_s^p - \tilde{x}_s^q \right)^2 \right] \frac{1}{\sqrt{2h_s}} \int_{g_s(0)}^{g_s(1)} \exp\left(-g_s^2 \right) dg_s = \\ &= \exp\left[-\frac{h_s}{2} \left(\tilde{x}_s^p - \tilde{x}_s^q \right)^2 \right] \frac{1}{\sqrt{2h_s}} \frac{\sqrt{\pi}}{2} \left(\mathrm{erf}\left[g_s(1) \right] - \mathrm{erf}\left[g_s(0) \right] \right) = \\ &= \frac{\sqrt{\pi}}{2} l_s \exp\left[-\frac{1}{4 l_s^2} \left(\tilde{x}_s^p - \tilde{x}_s^q \right)^2 \right] \left(\mathrm{erf}\left[\frac{1}{2 l_s} \left(2 - \tilde{x}_s^p - \tilde{x}_s^q \right) \right] + \mathrm{erf}\left[\frac{1}{2 l_s} \left(\tilde{x}_s^p + \tilde{x}_s^q \right) \right] \right) \end{aligned} \tag{15}$$

This concludes the evaluation of the integral L in (14) with the exact analytic form for the components $I_j(\mathbf{x})$ in (9), J_j in (12) and R_{pqs} in (15).

On the next step to finalize the calculation of (6), we evaluate the second part M in the equation.

$$\begin{aligned} M &= \int_0^1 \int_{[0,1]^{d-1}} \int_{[0,1]^{d-1}} m_*(\mathbf{x}) m_*(\hat{\mathbf{x}}) d\mathbf{x}_{-s} d\mathbf{x}'_{-s} dx_s = \\ &= \int_0^1 \left[z_{-s}^T(\mathbf{x}) \left(K + \sigma^2 \right)^{-1} \widetilde{\mathbf{y}}\, z_{-s}^T(\hat{\mathbf{x}}) \left(K + \sigma^2 \right)^{-1} \widetilde{\mathbf{y}} \right] dx_s = \\ &= \sigma_f^4 \sum_{p=1}^{n} \sum_{q=1}^{n} \left[w_p\, w_q \prod_{j=-s} I_j(\mathbf{x}_p) \prod_{j=-s} I_j(\mathbf{x}_q) \int_0^1 \exp\left(-h_s \left[(x_s - \tilde{x}_s^p)^2 + (x_s - \tilde{x}_s^q)^2 \right] \right) dx_s \right] = \\ &= \sigma_f^4 \sum_{p=1}^{n} \sum_{q=1}^{n} \left[w_p\, w_q \prod_{j=-s} I_j(\mathbf{x}_p) \prod_{j=-s} I_j(\mathbf{x}_q) R_{pqs} \right] \end{aligned} \tag{16}$$

Second component II from (5): $\mathbb{V}\mathbf{ar}_*\{E(Y)\}$

Let us return to the equation in (5) and focus on the second term II

$$\begin{aligned}\mathbb{V}\mathrm{ar}_*\{E(Y)\} &= \int\limits_{[0,1]^d}\int\limits_{[0,1]^d} k_*(\mathbf{x},\mathbf{x}')\, d\mathbf{x}d\mathbf{x}' = \\ &= \int\limits_{[0,1]^d}\int\limits_{[0,1]^d} K(\mathbf{x},\mathbf{x}')d\mathbf{x}d\mathbf{x}' - \int\limits_{[0,1]^d}\int\limits_{[0,1]^d} K(\mathbf{x},\tilde{X})\left(K+\sigma^2\right)^{-1} K(\tilde{X},\mathbf{x}')\, d\mathbf{x}d\mathbf{x}' = \\ &= Z - z^T\left(K+\sigma^2\right)^{-1} z \end{aligned} \tag{17}$$

Again we recast the calculation in more simple steps and use the formulas that was presented previously. For instance, the computation of the integral z in (17) is reduced to the evaluation of I_j from the equation (9).

$$z(\mathbf{x}) = \int\limits_{[0,1]^d} K(\mathbf{x},\tilde{X})d\mathbf{x} = \sigma_f^2 \prod_{j=1}^{d} I_j(\mathbf{x}) \tag{18}$$

Analogously, to compute Z we address to the integral J_j that was calculated in (12)

$$Z = \int\limits_{[0,1]^d}\int\limits_{[0,1]^d} K(\mathbf{x},\mathbf{x}')d\mathbf{x}d\mathbf{x}' = \sigma_f^2 \prod_{j=1}^{d} J_j \tag{19}$$

Third component III from (5): $\mathbb{E}_*\{E(Y)\}$

We address now to the calculation of the final third component in (5).

$$\mathbb{E}_*\{E(Y)\} = \left[\int\limits_{[0,1]^d} K(\mathbf{x},\widetilde{X})d\mathbf{x}\right]\left[K+\sigma^2 I\right]^{-1}\widetilde{\mathbf{y}} = z^T\left(K(\tilde{X},\tilde{X})+\sigma^2\right)^{-1}\widetilde{\mathbf{y}} \overset{(18)}{=} \sigma_f^2\prod_{j=1}^{d} I_j\left(K(\tilde{X},\tilde{X})+\sigma^2\right)^{-1}\widetilde{\mathbf{y}} \tag{20}$$

This derivation finalize the computation of the numerator from the formula (4), all three components, I, II and III , from (5) have been derived in analytic form.

Inference for variance in denominator $\mathbb{E}_*\{\mathrm{Var}(Y)\}$

The final step in the calculation of the estimate for the Sobol indices is the derivation of the analytic formula for the denominator in (4).

$$\begin{aligned}\mathbb{E}_*\{\mathrm{var}(Y)\} &= \mathbb{E}_*\{E(Y^2)\} - \mathbb{E}_*\{E(Y)\}^2 = E\{\mathbb{E}_*(Y^2)\} - \left(\mathbb{V}\mathrm{ar}_*\{E(Y)\} + (\mathbb{E}_*\{E(Y)\})^2\right) = \\ &= E\left(\mathbb{V}\mathrm{ar}_*\{Y\} + \{\mathbb{E}_*(Y)\}^2\right) - \left(\mathbb{V}\mathrm{ar}_*\{E(Y)\} + (\mathbb{E}_*\{E(Y)\})^2\right) = IV - II - III \end{aligned} \tag{21}$$

Similar to the previous routine for the numerator in (4), we divide the calculations in several steps. There are three components II, III and IV in (21), where II is summarized by formulas (17), (18) and (19), while III can be calculated using (20). The derivation of the forth component IV is presented in the next section.

Forth component IV **from (21):** $\mathbb{E}_*\{E(Y)\}$

Here we perform the last calculation of the forth component in (21).

$$E\{\mathbb{V}\mathrm{ar}_*\{Y\}\} + E\{\mathbb{E}_*(Y)\}^2 = \int\limits_{[0,1]^d} K(\mathbf{x},\mathbf{x})d\mathbf{x} - \int\limits_{[0,1]^d} K(\mathbf{x},\tilde{X})\left[K+\sigma^2\,\mathbb{I}\right]^{-1} K(\tilde{X},\mathbf{x})\,d\mathbf{x}+$$

$$+\int\limits_{[0,1]^d}\left[K(\mathbf{x},\tilde{X})\left(K+\sigma^2\right)^{-1}\widetilde{\mathbf{y}}\,K(\mathbf{x},\tilde{X})\left[K+\sigma^2\,\mathbb{I}\right]^{-1}\widetilde{\mathbf{y}}\,d\mathbf{x}\right] =$$

$$= \sigma_f^2 - \sigma_f^4 \sum_{p=1}^{n}\sum_{q=1}^{n}\left[k_{pq}\prod_{s=1}^{d} R_{pqs}\right] + \sigma_f^4 \sum_{p=1}^{n}\sum_{q=1}^{n}\left[w_p\; w_q \prod_{s=1}^{d} R_{pqs}\right], \qquad (22)$$

where the derivation for R_{pqs} is shown in (15).