Minimum Reduced Order Modelling

Robert A. Milton, Solomon F. Brown

Department of Chemical and Biological Engineering, University of Sheffield, Sheffield, S1 3JD, United Kingdom

Abstract

Keywords: Gaussian Process, Global Sensitivity Analysis, Sobol' Index, Surrogate Model

1. Notation

Tensors axes are multi-indexed as boldface subscripts

$$() =: \mathbf{0} \subseteq \mathbf{n} := (1, \ldots, n) \subseteq \mathbf{N} := (1, \ldots, N) \subset \mathbb{Z}^{+} \qquad 0 \leq n \leq N \in \mathbb{N}$$

which precede any superscript operation (e.g. inversion, transposition, exterior power). Subtraction of multi-indices is set-theoretic difference, for example

$$\mathbf{lN} - (\mathbf{l} - \mathbf{1})\mathbf{N} := ((l-1)N + 1, \ldots, lN)$$

Prime diacritics are used for bookeeping only, and will appear and disappear quite freely. We always demand that constant $N^{\cdots\prime} := N$, but do not constrain $n^{\cdots\prime} = n$ except explicitly. Multi-indexed quantities are square bracketed, and broadcast to fill every explicit axis. The matrix $[1]_{\mathbf{N}\times\mathbf{N}}$ filled with 1s should not be confused with the diagonal (identity) matrix $\langle 1\rangle_{\mathbf{N}\times\mathbf{N}} =: \langle [1]_{\mathbf{N}}\rangle$.

The response $[Y]_{\mathbf{L}\times\mathbf{N}} \in \mathbb{R}^{LN}$ to the design matrix $[X]_{\mathbf{M}\times\mathbf{N}} \in \mathbb{R}^{MN}$ of observed inputs is assumed standardized to multivariate normal sampling

$$[0]_{\mathbf{M}} = \mathbb{E}_{\mathbf{N}}[X] := \sum_{n\in\mathbf{N}} \frac{[X]_{\mathbf{M}\times n}}{N} = \frac{[X]_{\mathbf{M}\times\mathbf{N}}\,[1]_{\mathbf{N}}}{N} \quad ; \quad [1]_{\mathbf{M}} = \mathrm{tr}(\mathbb{V}[X])$$

$$[0]_{\mathbf{L}} = \mathbb{E}_{\mathbf{N}}[Y] := \sum_{n\in\mathbf{N}} \frac{[Y]_{\mathbf{L}\times n}}{N} = \frac{[Y]_{\mathbf{L}\times\mathbf{N}}\,[1]_{\mathbf{N}}}{N} \quad ; \quad [1]_{\mathbf{L}} = \mathrm{tr}(\mathbb{V}[Y])$$

Email addresses: `r.a.milton@sheffield.ac.uk` (Robert A. Milton), `s.f.brown@sheffield.ac.uk` (Solomon F. Brown)

Adjacent tensors are multiplied following the Einstein summation convention by boldface multi-index only. Because tensors arise in this work almost exclusively as covariances or exterior products, they are usually symmetric under permutations of ranks of the same dimension (e.g $\mathbf{N}, \mathbf{N}''$ or $\mathbf{L}, \mathbf{L}'$).

Syntax for exterior powers of tensors and their expectations is

$$\begin{aligned}[\cdot]^k_{\dots} &:= [\cdot]^{(k-1)}_{\dots} \otimes [\cdot]_{\dots'} \\ \mathbb{E}^k[\cdot] &:= \mathbb{E}[\cdot]^k \\ \mathbb{V}[\cdot] := \mathbb{V}[\cdot,\cdot] &:= \mathbb{E}\left[[\cdot]^2\right] - \mathbb{E}^2[\cdot]\end{aligned}$$

Expectations always carry a multi-index indicating (the dimensions of) the probability space over which they are taken.

Tensor quotients denote the inverse of the Hadamard (element-wise) product $\circ$

$$[q] = \frac{[a]}{[b]} \quad \Longleftrightarrow \quad [q] \circ [b] = [a]$$

where every tensor is broadcast to the same dimensions.

Unbounded multi-indices will use $\mathbf{o}$ in place of $\mathbf{n}$. A tensor Gaussian like

$$\begin{aligned}&\left[\mathsf{p}\big([\mathsf{z}]_{\mathbf{m}\times\mathbf{o}}\big|\, [Z]_{\mathbf{m}\times\mathbf{L}\times\mathbf{L}''\times\mathbf{o}'}\,, [\Sigma]_{\mathbf{L}\times\mathbf{L}'\times\mathbf{L}''\times\mathbf{L}'''\times\mathbf{m}\times\mathbf{m}}\big)\right]_{l\times l'\times l''\times l'''\times o\times o'} \\ &\qquad := (2\pi)^{-M/2} \left|[\Sigma]_{l\times l'\times l''\times l'''}\right|^{-1/2} \\ &\exp\left(-\frac{[\mathsf{z}-Z]^{\intercal}_{\mathbf{m}\times l\times l'\times l''\times l'''\times o\times o'}\, [\Sigma]^{-1}_{l\times l'\times l''\times l'''}\, [\mathsf{z}-Z]_{\mathbf{m}'\times l\times l'\times l''\times l'''\times o\times o'}}{2}\right)\end{aligned} \tag{1}$$

is defined in terms of the matrix

$$[\Sigma]_{l\times l'\times l''\times l'''} := [\Sigma]_{l\times l'\times l''\times l'''\times\mathbf{m}\times\mathbf{m}'}$$

and the transpose $\intercal$ (moving first multi-index to last) of the broadcast difference between two tensors

$$[\mathsf{z}-Z]_{\mathbf{m}\times l\times l'\times l''\times l'''\times\mathbf{o}\times\mathbf{o}'} := [\mathsf{z}]_{\mathbf{m}\times 1\times 1\times 1\times 1\times\mathbf{o}\times 1} - [Z]_{\mathbf{m}\times l\times 1\times l''\times 1\times 1\times\mathbf{o}'}$$

To be clear, this tensor definition applies only to explicit p, it *never* underpins a normal distibution N. The algebraic development which follows relies exclusively on trivial normal marginalization and scaling

$$\begin{aligned}&[\mathsf{z}]_{\mathbf{M}} \sim \mathsf{N}\big([Z]_{\mathbf{M}}\,, [\Sigma]_{\mathbf{M}\times\mathbf{M}}\big) \Rightarrow [\mathsf{z}]_{\mathbf{m}} \sim \mathsf{N}\big([Z]_{\mathbf{m}}\,, [\Sigma]_{\mathbf{m}\times\mathbf{m}}\big) \\ &[\mathsf{z}]_{\mathbf{M}} \sim \mathsf{N}\big([Z]_{\mathbf{M}}\,, [\Sigma]_{\mathbf{M}\times\mathbf{M}}\big) \Rightarrow [\Theta]^{\intercal}_{\mathbf{M}\times\mathbf{M}}\, [\mathsf{z}]_{\mathbf{M}} \sim \mathsf{N}(\Theta^{\intercal} Z, \Theta^{\intercal}\Sigma\Theta)\end{aligned} \tag{2}$$

$$(3)$$

together with an invaluable product formula reported in [1]

$$\begin{aligned}\mathsf{p}(\,\mathsf{z}|\,a,A)\,\mathsf{p}(\,\Theta^{\intercal}\mathsf{z}|\,b,B) &= \mathsf{p}(\,0|\,(b-\Theta^{\intercal}a),(B+\Theta^{\intercal}A\Theta))\\ &\times \mathsf{p}\big(\mathsf{z}|\,(A^{-1}+\Theta B^{-1}\Theta^{\intercal})^{-1}(A^{-1}a+\Theta B^{-1}b),(A^{-1}+\Theta B^{-1}\Theta^{\intercal})^{-1}\big) \quad (4)\end{aligned}$$

In referring back to these formulae, remember that $\mathsf{z}, Z, \Sigma, \Theta$ are arbitrary vectors and matrices here – within the dictates of the minimal dimension, sign, symmetry and invertibility requirements for these formulae to make sense – and not restricted to any particular values these quantities may later take.

2. Gaussian Process (GP) Regression

Conventionally, a Gaussian process is viewed either as a random function, or an indexed collection of random variables (RVs). In either case the argument or index is the singular datum $[x]_{\mathbf{M}} \in \mathbb{R}^{M}$. We shall instead adopt the perspective of formal definition, wherein the input (index) set is a variable-sized design matrix $[x]_{\mathbf{M}\times\mathbf{o}} \in \mathbb{R}^{Mo}$ and the response (state) space is $\mathbb{R}^{Lo} \ni \big[[\mathsf{y}]_{\mathbf{L}}\,|\,[x]_{\mathbf{M}\times\mathbf{o}}\big]_{\mathbf{L}\times\mathbf{o}}(\omega)$. The argument (ω) indicates a realization of the random variable which formally defines and fully specifies the GP:

$$[\mathsf{y}]_{\mathbf{L}}\,\big|\,[x]_{\mathbf{M}\times\mathbf{o}} \sim \mathsf{N}^{\dagger}\big([y(x)]_{\mathbf{L}\times\mathbf{o}}\,, [k_{\mathsf{y}}(x,x)]_{\mathbf{L}\times\mathbf{L}\times\mathbf{o}\times\mathbf{o}}\big) \quad \forall o \in \mathbb{Z}^{+}$$

Tensor axes must concatenate into a multivariate normal distribution

$$\begin{gathered}[]_{\mathbf{L}\times\mathbf{o}} \sim \mathsf{N}^{\dagger}\big([]_{\mathbf{L}\times\mathbf{o}}\,, []_{\mathbf{L}\times\mathbf{L}'\times\mathbf{o}\times\mathbf{o}'}\big) \Longleftrightarrow []^{\dagger}_{\mathbf{L}\times\mathbf{o}} \sim \mathsf{N}\Big([]^{\dagger}_{\mathbf{L}\times\mathbf{o}}\,, []^{\dagger}_{\mathbf{L}\times\mathbf{L}'\times\mathbf{o}\times\mathbf{o}'}\Big)\\ \Big[[]^{\dagger}_{\mathbf{L}\times\mathbf{o}}\Big]_{\mathbf{lo}-(\mathbf{l}-\mathbf{1})\mathbf{o}} := []_{l\times\mathbf{o}}\\ \Big[[]^{\dagger}_{\mathbf{L}\times\mathbf{L}'\times\mathbf{o}\times\mathbf{o}'}\Big]_{(\mathbf{lo}-(\mathbf{l}-\mathbf{1})\mathbf{o})\times(\mathbf{l}'\mathbf{o}'-(\mathbf{l}'-\mathbf{1})\mathbf{o}')} := []_{l\times l'\times\mathbf{o}\times\mathbf{o}'}\end{gathered}$$

supporting the fundamental definition of the Gaussian process kernel, as a covariance over response space

$$[k_{\mathsf{y}}(x,x)]_{l\times l'\times o\times o'} := \mathbb{V}_{\mathbf{Lo}}\big[[\mathsf{y}|x]_{l\times o}\,, [\mathsf{y}|x]_{l'\times o'}\big]$$

2.1. Prior GP

Gaussian Process regression decomposes output $[\mathsf{y}]_{\mathbf{L}}$ into signal GP $[\mathsf{f}]_{\mathbf{L}}$, and independent noise GP $[\hat{\mathsf{e}}]_{\mathbf{L}}$ with constant noise covariance $[E]_{\mathbf{L}\times\mathbf{L}}$

$$\begin{aligned}[\mathsf{y}|E]_{\mathbf{L}} &= [\mathsf{f}]_{\mathbf{L}} + [\hat{\mathsf{e}}|E]_{\mathbf{L}} \\ [\hat{\mathsf{e}}|E]_{\mathbf{L}} \,\big|\, [x]_{\mathbf{M}\times\mathbf{o}} &\sim \mathsf{N}^{\dagger}\big([0]_{\mathbf{L}\times\mathbf{o}}\,, [E]_{\mathbf{L}\times\mathbf{L}} \otimes \langle 1\rangle_{\mathbf{o}\times\mathbf{o}}\big)\end{aligned}$$

The ARD kernel is hyperparametrized by signal covariance $[F]_{\mathbf{L}\times\mathbf{L}}$ and the matrix $[\Lambda]_{\mathbf{L}\times\mathbf{M}}$ of characteristic lengthscales for each output/input combination. Using the broadcast Hadamard product $\circ$ we define

$$\begin{aligned}\big\langle \Lambda^2 \pm I\big\rangle_{l\times l'\times\mathbf{M}\times\mathbf{M}} &:= \big\langle [\Lambda]_{l\times\mathbf{M}} \circ [\Lambda]_{l'\times\mathbf{M}} \pm [I]_{\mathbf{M}}\big\rangle \qquad I \in \mathbb{Z} - \mathbb{Z}^{-} \\ \big\langle \Lambda^2\big\rangle_{l\times l'\times\mathbf{M}\times\mathbf{M}} &:= \big\langle \Lambda^2 \pm 0\big\rangle_{l\times l'} \\ [\pm F]_{l\times l'} &:= (2\pi)^{M/2} \left|\big\langle \Lambda^2\big\rangle_{l\times l'}\right|^{1/2} [F]_{l\times l'}\end{aligned}$$

and implement the objective ARD prior using Eq. (1)

$$[\mathsf{f}|F,\Lambda]_{\mathbf{L}} \,\big|\, [x]_{\mathbf{M}\times\mathbf{o}} \sim \mathsf{N}^{\dagger}\Big([0]_{\mathbf{L}\times\mathbf{o}}\,, [\pm F]_{\mathbf{L}\times\mathbf{L}'} \circ \mathsf{p}\Big([x]_{\mathbf{M}\times\mathbf{o}} \big|\, [x]_{\mathbf{M}\times\mathbf{o}}\,, \big\langle \Lambda^2\big\rangle_{\mathbf{L}\times\mathbf{L}'}\Big)\Big)$$

2.2. Predictive GP

Bayesian inference for GP regression further conditions the hyper-parametrized GP $\mathsf{y}|E,F,\Lambda$ on the observed realization of the random variable $[\mathsf{y}|X]$

$$[Y]^{\dagger}_{\mathbf{L}\times\mathbf{N}} := \big[[\mathsf{y}|E,F,\Lambda]_{\mathbf{L}} \,\big|\, [X]_{\mathbf{M}\times\mathbf{N}}\big]^{\dagger}(\omega) \in \mathbb{R}^{LN}$$

To this end we define

$$\begin{aligned}[K_{\hat{\mathsf{e}}}]_{\mathbf{Lo}\times\mathbf{Lo}} &:= \mathbb{V}_{\mathbf{Lo}}\Big[\big[[\hat{\mathsf{e}}|E]_{\mathbf{L}} \,\big|\, [x]_{\mathbf{M}\times\mathbf{o}}\big]^{\dagger}\Big] \\ &= \big[[E]_{\mathbf{L}\times\mathbf{L}} \otimes \langle 1\rangle_{\mathbf{o}\times\mathbf{o}}\big]^{\dagger} \\ [k(x,x')]_{\mathbf{Lo}\times\mathbf{L}'\mathbf{o}'} &:= \mathbb{V}_{\mathbf{Lo}}\Big[\big[[\mathsf{f}|F,\Lambda]_{\mathbf{L}} \,\big|\, [x]_{\mathbf{M}\times\mathbf{o}}\big]^{\dagger}, \big[[\mathsf{f}|F,\Lambda]_{\mathbf{L}'} \,\big|\, [x']_{\mathbf{M}\times\mathbf{o}'}\big]^{\dagger}\Big] \\ &= \Big[[\pm F]_{\mathbf{L}\times\mathbf{L}'} \circ \mathsf{p}\Big([x]_{\mathbf{M}\times\mathbf{o}} \big|\, [x']_{\mathbf{M}\times\mathbf{o}'}\,, \big\langle \Lambda^2\big\rangle_{\mathbf{L}\times\mathbf{L}'}\Big)\Big]^{\dagger} \\ [K_Y]_{\mathbf{LN}\times\mathbf{LN}} &:= \mathbb{V}_{\mathbf{Lo}}\Big[\big[[\mathsf{y}|E,F,\Lambda]_{\mathbf{L}} \,\big|\, [X]_{\mathbf{M}\times\mathbf{N}}\big]^{\dagger}\Big] \\ &= k([X]_{\mathbf{M}\times\mathbf{N}}\,, [X]_{\mathbf{M}\times\mathbf{N}}) + [K_{\hat{\mathsf{e}}}]_{\mathbf{LN}\times\mathbf{LN}}\end{aligned} \tag{5}$$

Applying Bayes' rule

$$\begin{aligned}\mathrm{p}(\mathsf{f}|Y)\mathrm{p}(Y) = \mathrm{p}(Y|\mathsf{f})\mathrm{p}(\mathsf{f}) &= \mathsf{p}\big(Y^{\dagger}\big|\,\mathsf{f}^{\dagger}, K_{\hat{\mathsf{e}}}\big)\,\mathsf{p}\big(\mathsf{f}^{\dagger}\big|\,[0]_{\mathbf{LN}}\,, k(X,X)\big)\\ &= \mathsf{p}\big(\mathsf{f}^{\dagger}\big|\,Y^{\dagger}, K_{\hat{\mathsf{e}}}\big)\,\mathsf{p}\big(\mathsf{f}^{\dagger}\big|\,[0]_{\mathbf{LN}}\,, k(X,X)\big)\end{aligned}$$

Eq. (4) immediately reveals the marginal likelihood

$$\mathrm{p}\big([Y|E,F,\Lambda]\,\big|X\big) = \mathsf{p}\Big([Y]^{\dagger}_{\mathbf{L}\times\mathbf{N}}\Big|\,[0]_{\mathbf{LN}}\,, K_Y\Big) \tag{6}$$

and the posterior distribution

$$\begin{aligned}&\big[[\mathsf{f}|Y|E,F,\Lambda]\,\big|X\big]^{\dagger}_{\mathbf{L}\times\mathbf{N}} \sim\\ &\qquad\qquad \mathsf{N}\big(k(X,X)K_Y^{-1}Y^{\dagger},\ k(X,X) - k(X,X)K_Y^{-1}k(X,X)\big)\end{aligned}$$

The ultimate goal is the posterior predictive GP which extends the posterior distribution to arbitrary – usually unobserved – $[x]_{\mathbf{M}\times\mathbf{o}}$. This is traditionally derived from the definition of conditional probability, but this seems unnecessary, for the extension must recover the posterior distribution when $x = X$. There is only one way of selectively replacing X with x in the posterior formula which preserves the coherence of tensor ranks:

$$\begin{aligned}&\big[[\mathsf{f}|Y|E,F,\Lambda]\,\big|x\big]^{\dagger}_{\mathbf{L}\times\mathbf{o}} \sim\\ &\qquad\qquad \mathsf{N}\big(k(x,X)K_Y^{-1}Y^{\dagger},\ k(x,x) - k(x,X)K_Y^{-1}k(X,x)\big)\end{aligned} \tag{7}$$

2.3. GP Optimization

Henceforth we implicitly condition on optimal hyperparameters, which maximise the marginal likelihood Eq. (6).

$$[E]_{\mathbf{L}\times\mathbf{L}}\,, [F]_{\mathbf{L}\times\mathbf{L}}\,, [\Lambda]_{\mathbf{L}\times\mathbf{M}} := \operatorname{argmax} \mathsf{p}\Big([Y]^{\dagger}_{\mathbf{L}\times\mathbf{N}}\Big|\,[0]_{\mathbf{LN}}\,, K_Y\Big) \tag{8}$$

The lengthscale tensor could feasibly have been of maximal rank $[\Lambda]_{\mathbf{L}\times\mathbf{L}\times\mathbf{M}}$. We have restricted this to $[\Lambda]_{\mathbf{L}\times\mathbf{M}}$, as one set of ARD lengthscales per output is heuristically satisfying and enables effective optimization as follows. For each output $l \in \mathbf{L}$ construct a separate GP to optimize the diagonal hyperparameters

$$\begin{aligned}&[E]_{l\times l}\,, [F]_{l\times l}\,, [\Lambda]_{l\times\mathbf{M}} =\\ &\qquad\qquad \operatorname{argmax} \mathsf{p}\Big([Y]^{\dagger}_{l\times\mathbf{N}}\Big|\,[0]_{\mathbf{N}}\,, [K_Y]_{(\mathbf{lN}-(\mathbf{l}-\mathbf{1})\mathbf{N})\times(\mathbf{lN}-(\mathbf{l}-\mathbf{1})\mathbf{N})}\Big)\end{aligned}$$

From this starting point, E, F may be optimized (off-diagonal elements in particular) in the full multi-output GP Eq. (8). One may then attempt to re-optimize lengthscales according to Eq. (8), and iterate, although this may be gilding the lily.

3. Reduction of Order by Marginalization (ROM)

As sample data we take three standardized normal random variables

$$[\mathsf{z}]_{\mathbf{M}\times 3} \sim \mathsf{N}^{\dagger}\left([0]_{\mathbf{M}\times\mathbf{3}}\,, \langle 1\rangle_{\mathbf{M}\times\mathbf{M}\times\mathbf{3}\times\mathbf{3}}\right) \tag{9}$$

The sample basis is rotated to the input data

$$[\mathsf{x}]_{\mathbf{M}'\times i} := [\Theta]^{\intercal}_{\mathbf{M}\times\mathbf{M}'\times i}\,[\mathsf{z}]_{\mathbf{M}\times i} \tag{10}$$

The three datapoints represent an arbitrary datum, marginalized differently in 3 ROMs. This is needed to ascertain covariances between these ROMs. For bookeeping we define

$$[\mathbf{m}] := [\mathbf{m}]_{\mathbf{3}} := [\dot{\mathbf{m}}, \ddot{\mathbf{m}}, \mathbf{0}] \quad ; \quad \mathbf{m} \in \{\dot{\mathbf{m}} \times 1, \ddot{\mathbf{m}} \times 2, \mathbf{0} \times 3\}$$

and the ragged tensors

$$[\mathsf{z}]_{[\mathbf{m}]} := [[\dot{\mathsf{z}}]_{\dot{\mathbf{m}}}\,, [\ddot{\mathsf{z}}]_{\ddot{\mathbf{m}}}\,, [\,\dddot{\mathsf{z}}\,]_{\mathbf{0}}] \quad ; \quad [\Theta]_{[\mathbf{m}]\times[\mathbf{M}]} := \left[[\Theta]_{\dot{\mathbf{m}}\times\mathbf{M}}\,, [\Theta]_{\ddot{\mathbf{m}}\times\mathbf{M}}\,, [\Theta]_{\mathbf{0}\times\mathbf{M}}\right]$$

The non-ragged versions straightforwardly replace $[\mathbf{m}]$ with $[\mathbf{M}] := [\mathbf{M}, \mathbf{M}, \mathbf{M}]$, eliciting the surrogate response RV

$$\begin{aligned}\left[[\mathsf{y}|Y]_{\mathbf{L}}\,\middle|\,[\Theta]^{\intercal}_{[\mathbf{M}]\times[\mathbf{M}]}[\mathsf{z}]_{[\mathbf{M}]}\right]^{\dagger}_{\mathbf{L}\times\mathbf{3}} = \left[[\mathsf{y}|Y]_{\mathbf{L}}\,\middle|\,[\mathsf{x}]_{\mathbf{M}\times\mathbf{3}}\right]^{\dagger}_{\mathbf{L}\times\mathbf{3}} \sim \\ \mathsf{N}\left(k(\mathsf{x}, X)K_Y^{-1}Y^{\dagger},\ k(\mathsf{x},\mathsf{x}) - k(\mathsf{x}, X)K_Y^{-1}k(X,\mathsf{x}) + K_{\hat{\mathsf{e}}}\right)\end{aligned} \tag{11}$$

The marginal response

$$\begin{aligned}[\mathsf{e}]_{\mathbf{L}\times\mathbf{3}} := \mathbb{E}_{[\mathbf{M}]-[\mathbf{m}]}\left[[\mathsf{y}|Y]_{\mathbf{L}}\,\middle|\,[\Theta]^{\intercal}_{[\mathbf{m}]\times[\mathbf{M}]}[\mathsf{z}]_{[\mathbf{m}]}\right]_{\mathbf{L}\times\mathbf{3}} \\ \sim \mathsf{N}^{\dagger}\left([f(\mathsf{z};\Theta)]_{\mathbf{L}\times\mathbf{3}}\,, [\sigma(\mathsf{z};\Theta)]_{\mathbf{L}\times\mathbf{L}\times\mathbf{3}\times\mathbf{3}}\right)\end{aligned} \tag{12}$$

serves to define marginal expectation f and covariance σ. For lucidity we may directly subscript output sub-tensors by their $[\mathbf{m}]$-element, as in

$$\mathsf{e}_{[\mathbf{m}]_i} := [\mathsf{e}]_{\mathbf{L}\times i} \quad ; \quad f_{[\mathbf{m}]_i} := [f]_{\mathbf{L}\times i} \quad ; \quad \sigma_{[\mathbf{m}]_i,[\mathbf{m}]_j} := [\sigma]_{\mathbf{L}\times\mathbf{L}\times i\times j}$$

Knowledge ranges from the totally marginal RV $\sim \mathsf{N}(f_{\mathbf{0}}, \sigma_{\mathbf{0},\mathbf{0}})$ to the totally conditioned RV $\sim \mathsf{N}(f_{\mathbf{M}}, \sigma_{\mathbf{M},\mathbf{M}})$. The question is how to calculate these quantities, and this deserves some clarity. The marginal response lies at the root of ROM, stratifying the overall response e_{M} into the response e_{m} to the first $m \in (0, \ldots, M)$ sample dimensions. We do not regard a marginal response as a GP because its input (index) is necessarily a single datum, moreover a random variable $[\mathsf{z}]_{\mathbf{m}}$. Rather than a GP, we envisage each marginal response in Eq. (12) as a normally distributed RV on response $\mathbf{L}$-space. The parameters f, σ specifying each normal RV are functions of $[\mathsf{z}]_{\mathbf{m}}$, an RV on input $\mathbf{M}$-space. The input and response probability spaces are entirely separate. According to this discussion – and, for purely technical reasons, Fubini's Theorem – we may re-order expectations taken over the two probability spaces

$$\mathbb{E}_{\mathbf{L}}\,\mathbb{E}_{\mathbf{m}} = \mathbb{E}_{\mathbf{m}}\,\mathbb{E}_{\mathbf{L}}$$

This will be used repeatedly, starting with the normal RV parameters

$$\begin{aligned}
[f(\mathsf{z};\Theta)]^{\dagger}_{\mathbf{L}\times\mathbf{3}} &:= \mathbb{E}_{\mathbf{L}}\,\mathbb{E}_{[\mathbf{M}]-[\mathbf{m}]}\left[[\mathsf{y}|Y]_{\mathbf{L}}\,\middle|[\Theta]^{\intercal}_{[\mathbf{m}]\times[\mathbf{M}]}[\mathsf{z}]_{[\mathbf{m}]}\right]^{\dagger}_{\mathbf{L}\times\mathbf{3}} \\
&= \mathbb{E}_{[\mathbf{M}]-[\mathbf{m}]}\,\mathbb{E}_{\mathbf{L}}[\mathsf{y}|Y]^{\dagger} \\
&= \mathbb{E}_{[\mathbf{M}]-[\mathbf{m}]}\left[k(\mathsf{x}, X)K_Y^{-1}Y^{\dagger}\right] \\
[\sigma(\mathsf{z};\Theta)]^{\dagger}_{\mathbf{L}\times\mathbf{L}\times\mathbf{3}\times\mathbf{3}} &:= \mathbb{V}_{\mathbf{L}}\,\mathbb{E}_{[\mathbf{M}]-[\mathbf{m}]}\left[[\mathsf{y}|Y]_{\mathbf{L}}\,\middle|[\Theta]^{\intercal}_{[\mathbf{m}]\times[\mathbf{M}]}[\mathsf{z}]_{[\mathbf{m}]}\right]^{\dagger}_{\mathbf{L}\times\mathbf{3}} \\
&= \mathbb{E}_{[\mathbf{M}]-[\mathbf{m}]}\,\mathbb{V}_{\mathbf{L}}\,[\mathsf{y}|Y]^{\dagger} \\
&= \mathbb{E}_{[\mathbf{M}]-[\mathbf{m}]}\left[k(\mathsf{x},\mathsf{x}) - k(\mathsf{x}, X)K_Y^{-1}k(X,\mathsf{x}) + K_{\hat{\mathsf{e}}}\right]
\end{aligned}$$

which uses the shorthand

$$\mathbb{E}_{[\mathbf{M}]-[\mathbf{m}]}[\cdot] := \mathbb{E}_{[\mathbf{M}]-[\mathbf{m}]}\left[\cdot\,\middle|[\Theta]^{\intercal}_{[\mathbf{m}]\times[\mathbf{M}]}[\mathsf{z}]_{[\mathbf{m}]}\right]$$

Equations (9) to (12) support analytic expectations of f and σ using Eq. (5) and Eqs. (2) to (4), reported in the following Subsections.

3.1. Marginal Expectation

The marginal expectation in Eq. (12) is given by

$$[f(\mathsf{z};\Theta)]^{\dagger}_{\mathbf{L}\times\mathbf{3}} = \mathbb{E}_{[\mathbf{M}]-[\mathbf{m}]}\left[k(\mathsf{x}, X)K_Y^{-1}Y^{\dagger}\right] = [g(\mathsf{z};\Theta)]^{\dagger}_{\mathbf{L}\times\mathbf{L}'\times\mathbf{3}\times\mathbf{N}'}\left[K_Y^{-1}Y^{\dagger}\right]_{\mathbf{L}'\mathbf{N}'}$$

where $\mathbf{m} = [\mathbf{m}]_i \times i$ and

$$\begin{aligned}
[g(\mathsf{z};\Theta)]_{l\times l'\times i\times\mathbf{N}'} &:= [\pm F]_{l\times l'} \circ \mathsf{p}\Big([0]_{\mathbf{M}} \big| \, [X]_{\mathbf{M}\times\mathbf{N}'}, \big\langle \Lambda^2 + 1 \big\rangle_{l\times l'} \Big) \\
&\qquad \circ \frac{\mathsf{p}\big([\mathsf{z}]_{\mathbf{m}} \big| \, [G]_{\mathbf{m}\times l\times l'\times\mathbf{N}'}, [\Gamma]_{l\times l'} \big)}{\mathsf{p}\big([\mathsf{z}]_{\mathbf{m}} \big| \, [0]_{\mathbf{m}}, \langle 1 \rangle \big)} \\
&= [\pm F]_{l\times l'} \circ \frac{\mathsf{p}\big([0]_{\mathbf{M}} \big| \, [X]_{\mathbf{M}\times\mathbf{N}'}, \langle \Lambda^2 + 1 \rangle_{l\times l'} \big)}{\mathsf{p}\big([0]_{\mathbf{m}} \big| \, [\Theta]_{\mathbf{m}\times\mathbf{M}} [X]_{\mathbf{M}\times\mathbf{N}'}, [\Phi]^{-1}_{l\times l'} \big)} \\
&\qquad \circ \mathsf{p}\big([\mathsf{z}]_{\mathbf{m}} \big| \, [\Theta]_{\mathbf{m}\times\mathbf{M}} [X]_{\mathbf{M}\times\mathbf{N}'}, [\Phi]^{-1}_{l\times l'\times\mathbf{m}\times\mathbf{m}'} [\Gamma]_{l\times l'} \big)
\end{aligned}$$

and

$$\begin{aligned}
[G]_{\mathbf{m}\times l\times l'\times\mathbf{N}'} &:= [\Theta]_{\mathbf{m}\times\mathbf{M}} \big\langle \Lambda^2 + 1 \big\rangle^{-1}_{l\times l'\times\mathbf{M}\times\mathbf{M}'} [X]_{\mathbf{M}'\times\mathbf{N}'} \\
[\Gamma]_{l\times l'\times\mathbf{m}\times\mathbf{m}'} &:= [\Theta]_{\mathbf{m}\times\mathbf{M}} \Big\langle \big\langle \Lambda^2 \big\rangle_{l\times l'} \big\langle \Lambda^2 + 1 \big\rangle^{-1}_{l\times l'} \Big\rangle_{\mathbf{M}\times\mathbf{M}'} [\Theta]^{\intercal}_{\mathbf{m}'\times\mathbf{M}'} \\
[\Phi]_{l\times l'\times\mathbf{m}\times\mathbf{m}'} &:= [\Theta]_{\mathbf{m}\times\mathbf{M}} \big\langle \Lambda^2 + 1 \big\rangle^{-1}_{l\times l'\times\mathbf{M}\times\mathbf{M}'} [\Theta]^{\intercal}_{\mathbf{m}'\times\mathbf{M}'}
\end{aligned}$$

Of particular importance

$$\begin{aligned}
f_{\mathbf{0}}(\mathsf{z};\Theta) \; &= [g_{\mathbf{0}}(\mathsf{z};\Theta)]^{\dagger}_{\mathbf{L}\times\mathbf{L}'\times\mathbf{N}'} \left[K_Y^{-1} Y^{\dagger} \right]_{\mathbf{L}'\mathbf{N}'} \\
g_{\mathbf{0}}(\mathsf{z};\Theta) &:= [g(\mathsf{z};\Theta)]_{\mathbf{L}\times\mathbf{L}'\times 3\times\mathbf{N}'} \\
&:= [\pm F]_{\mathbf{L}\times\mathbf{L}'} \circ \mathsf{p}\Big([0]_{\mathbf{M}} \big| \, [X]_{\mathbf{M}\times\mathbf{N}'}, \big\langle \Lambda^2 + 1 \big\rangle_{\mathbf{L}\times\mathbf{L}'\times\mathbf{M}\times\mathbf{M}} \Big)
\end{aligned}$$

Standardization of X and Y instills a totally marginal expectation of $f_{\mathbf{0}}(\mathsf{z};\Theta) \approx [0]_{\mathbf{L}}$, but this is usually inexact.

3.2. Marginal Covariance

The marginal covariance in Eq. (12) is given by

$$\begin{aligned}
&[\sigma(\mathsf{z};\Theta)]_{\mathbf{L}\times\mathbf{L}'\times\mathbf{3}\times\mathbf{3}'} = \\
&\qquad [F]_{\mathbf{L}\times\mathbf{L}'} \circ [\phi(\mathsf{z};\Theta)]_{\mathbf{L}\times\mathbf{L}'\times\mathbf{3}\times\mathbf{3}'} - [\psi(\mathsf{z};\Theta)]_{\mathbf{L}\times\mathbf{L}'\times\mathbf{3}\times\mathbf{3}'} + [E]_{\mathbf{L}\times\mathbf{L}'\times\mathbf{3}\times\mathbf{3}'}
\end{aligned}$$

For $\mathbf{m} := [\mathbf{m}]_i \times i, \; \mathbf{m'} := [\mathbf{m}]_{i'} \times i'$

$$\begin{aligned}
&[\phi(\mathsf{z};\Theta)]^{\dagger}_{l\times l'\times i\times i'} := \frac{\mathbb{E}_{[\mathbf{M}]-[\mathbf{m}]} [k([\mathsf{x}]_i, [\mathsf{x}]_{i'})]_{l\times l'}}{[F]_{l\times l'}} = \\
&\frac{\left| \langle \Lambda^2 \rangle_{l\times l'\times\mathbf{M}\times\mathbf{M}} \right|^{1/2} \mathsf{p}\big([\mathsf{z}]_{\mathbf{m}} \big| \, [0]_{\mathbf{m}}, [\Upsilon]_{l\times l'} \big) \, \mathsf{p}\big([\mathsf{z}]_{\mathbf{m}'} \big| \, [Z]_{l\times l'\times i\times\mathbf{m}'}, [\Pi]_{l\times l'\times i} \big)}{\left| \langle \Lambda^2 + 2 \rangle_{l\times l'\times\mathbf{M}\times\mathbf{M}} \right|^{1/2} \mathsf{p}\big([\mathsf{z}]_{\mathbf{m}} \big| \, [0]_{\mathbf{m}}, \langle 1 \rangle \big) \, \mathsf{p}\big([\mathsf{z}]_{\mathbf{m}'} \big| \, [0]_{\mathbf{m}'}, \langle 1 \rangle \big)}
\end{aligned}$$

$$[\psi(\mathsf{z};\Theta)]^{\dagger}_{l\times l'\times i\times i'} := \mathbb{E}_{[\mathbf{M}]-[\mathbf{m}]}\left[k([\mathsf{x}]_i\,, X)K_Y^{-1}k(X,[\mathsf{x}]_{i'})\right]_{l\times l'} =$$
$$[g(\mathsf{z};\Theta)]^{\dagger}_{l\times\mathbf{L}''\times i\times\mathbf{N}''}\left[K_Y^{-1}\right]_{\mathbf{L}''\mathbf{N}''\times\mathbf{L}'''\mathbf{N}'''}[g(\mathsf{z};\Theta)]^{\dagger}_{l'\times\mathbf{L}'''\times i'\times\mathbf{N}'''}$$

where

$$\begin{aligned}
[\Upsilon]_{l\times l'\times\mathbf{m}\times\mathbf{m}''} &:= [\Theta]_{\mathbf{m}\times\mathbf{M}}\left\langle\left\langle\Lambda^2+1\right\rangle_{l\times l'}\left\langle\Lambda^2+2\right\rangle^{-1}_{l\times l'}\right\rangle_{\mathbf{M}\times\mathbf{M}'}[\Theta]^{\intercal}_{\mathbf{m}''\times\mathbf{M}'}\\
[\Pi]^{-1}_{l\times l'\times i\times\mathbf{m}'\times\mathbf{m}'''} &:= [\Phi]^{\intercal}_{l\times l'\times\mathbf{m}\times\mathbf{m}'}[\Gamma]^{-1}_{\mathbf{m}\times\mathbf{m}''}[\Phi]_{l\times l'\times\mathbf{m}''\times\mathbf{m}'''}\\
&\quad+\left[[\Upsilon]^{-1}_{l\times l'\times\mathbf{M}\times\mathbf{M}}\right]_{\mathbf{m}'\times\mathbf{m}'''}\\
[Z]_{l\times l'\times i\times\mathbf{m}'} &:= [\Pi]_{l\times l'\times i\times\mathbf{m}'\times\mathbf{M}}[\Phi]^{\intercal}_{l\times l'\times\mathbf{m}''\times\mathbf{M}}[\Gamma]^{-1}_{\mathbf{m}''\times\mathbf{m}}[\mathsf{z}]_{\mathbf{m}}
\end{aligned}$$

3.3. Centralized Marginals

Calculations are easier with the centralized marginal responses

$$[\mathsf{c}]_{\mathbf{L}\times\mathbf{3}} := [\mathsf{e}]_{\mathbf{L}\times\mathbf{3}} - [f(\mathsf{z},\Theta)]_{\mathbf{L}\times\mathbf{3}}$$

These are normally ($\mathsf{N}^{\dagger}$) distributed with moments [2, 3]

$$\begin{aligned}
\mathbb{E}_{\mathbf{L}}[\mathsf{c}]_{\mathbf{L}\times\mathbf{3}} &= [0]_{\mathbf{L}\times\mathbf{3}}\\
\mathbb{E}_{\mathbf{L}}\left[[\mathsf{c}]^2_{\mathbf{L}\times\mathbf{3}}\right] &= [\sigma(\mathsf{z},\Theta)]_{\mathbf{L}\times\mathbf{L}'\times\mathbf{3}\times\mathbf{3}'}\\
\mathbb{E}_{\mathbf{L}}\left[[\mathsf{c}]^3_{\mathbf{L}\times\mathbf{3}}\right] &= [0]_{\mathbf{L}\times\mathbf{L}'\times\mathbf{L}''\times\mathbf{3}\times\mathbf{3}'\times\mathbf{3}''}\\
\mathbb{E}_{\mathbf{L}}\left[[\mathsf{c}]^4_{\mathbf{L}\times\mathbf{3}}\right] &= [\sigma(\mathsf{z},\Theta)]_{\mathbf{L}\times\mathbf{L}'\times\mathbf{3}\times\mathbf{3}'}\otimes[\sigma(\mathsf{z},\Theta)]_{\mathbf{L}''\times\mathbf{L}'''\times\mathbf{3}''\times\mathbf{3}'''}\\
&\quad+[\sigma(\mathsf{z},\Theta)]_{\mathbf{L}\times\mathbf{L}''\times\mathbf{3}\times\mathbf{3}''}\otimes[\sigma(\mathsf{z},\Theta)]_{\mathbf{L}'\times\mathbf{L}'''\times\mathbf{3}'\times\mathbf{3}'''}\\
&\quad+[\sigma(\mathsf{z},\Theta)]_{\mathbf{L}\times\mathbf{L}'''\times\mathbf{3}\times\mathbf{3}'''}\otimes[\sigma(\mathsf{z},\Theta)]_{\mathbf{L}'\times\mathbf{L}''\times\mathbf{3}'\times\mathbf{3}''}
\end{aligned}\tag{13}$$

4. Closed Sobol' Indices

The relevance of the first m inputs is measured by the Closed Sobol' Index

$$\left[\mathsf{S}\middle|\,[\Theta]_{\mathbf{m}\times\mathbf{M}}\right]_{\mathbf{L}\times\mathbf{L}} := \frac{\mathbb{V}_{\mathbf{m}}[\mathsf{e}_{\mathsf{m}}]}{\mathbb{V}_{\mathbf{M}}[\mathsf{e}_{\mathsf{M}}]}\tag{14}$$

In our formulation, this is an RV on response $\mathbf{L}$-space, whose distribution is effectively inexpressible. It is the quotient of two (presumably dependent) RVs from the stratified hierarchy

$$\mathbb{V}_{\mathbf{m}}[\mathsf{e}_{\mathsf{m}}] = \mathbb{E}_{\mathbf{m}}\left[\mathsf{e}^2_{\mathsf{m}}\right] - \mathbb{E}^2_{\mathbf{m}}[\mathsf{e}_{\mathsf{m}}] = \mathbb{E}_{\mathbf{m}}\left[\mathsf{e}^2_{\mathsf{m}}\right] - \mathsf{e}^2_0 = \mathbb{E}_{\mathbf{m}}\left[[\mathsf{c}_{\mathsf{m}} + f_{\mathbf{m}}]^2_{\mathbf{L}}\right] - [\mathsf{c}_0 + f_{\mathbf{0}}]^2_{\mathbf{L}}$$

Each stratum $\mathbb{V}_{\mathbf{m}}[\mathsf{e}_{\mathsf{m}}]$ is an RV on response $\mathbf{L}$-space which is the difference of two (presumably dependent) RVs, each of which has a generalized chi-squared distribution on response $\mathbf{L}$-space (because e_{m} is always normally distributed).

4.1. Expectations

The expected value of the Closed Sobol' Index is simply

$$[S_{\mathbf{m}}(\Theta)]_{\mathbf{L}\times\mathbf{L}} := \mathbb{E}_{\mathbf{L}}\left[\mathsf{S}\middle|\,[\Theta]_{\mathbf{m}\times\mathbf{M}}\right]_{\mathbf{L}\times\mathbf{L}} = \frac{V_{\mathbf{m}}(\Theta)}{V_{\mathbf{M}}(\Theta)}$$

where for any $\mathbf{m} \subseteq \mathbf{M}$

$$\begin{aligned}
[V_{\mathbf{m}}(\Theta)]_{\mathbf{L}\times\mathbf{L}} &:= \mathbb{E}_{\mathbf{L}}\mathbb{V}_{\mathbf{m}}[\mathsf{e}_{\mathsf{m}}] \\
&= \mathbb{E}_{\mathbf{m}}\,\mathbb{E}_{\mathbf{L}}\left[[\mathsf{c}_{\mathsf{m}} + f_{\mathbf{m}}]^2_{\mathbf{L}}\right] - \mathbb{E}_{\mathbf{L}}\left[[\mathsf{c}_0 + f_{\mathbf{0}}]^2_{\mathbf{L}}\right] \\
&= \mathbb{E}_{\mathbf{m}}\left[[\sigma_{\mathbf{m},\mathbf{m}}]_{\mathbf{L}\times\mathbf{L}} + [f_{\mathbf{m}}]^2_{\mathbf{L}}\right] - [\sigma_{\mathbf{0},\mathbf{0}}]_{\mathbf{L}\times\mathbf{L}} - [f_{\mathbf{0}}]^2_{\mathbf{L}} \\
&= \mathbb{E}_{\mathbf{m}}\left[[f_{\mathbf{m}}]^2_{\mathbf{L}}\right] - [f_{\mathbf{0}}]^2_{\mathbf{L}}
\end{aligned}$$

When using GPs to calculate Sobol' indices, there is no difference between the $\mathbf{L}$-expectation of the $\mathbf{m}$-covariance and the $\mathbf{m}$-covariance of the $\mathbf{L}$-expectation. This is not entirely obvious, and Oakley and O'Hagan [4] caution one to respect the (ultimately non-existent) difference. On the other hand, valid interchange is a natural consequence of the separation of probability spaces we have described in Section 3.

Using the shorthand

$$[KY3]^{\dagger}_{l\times\mathbf{L}'\times\mathbf{N}'} := \left[K_Y^{-1}Y^{\dagger}\right]_{l\times\mathbf{L}'\times\mathbf{N}'} \circ [g(\mathsf{z};\Theta)]^{\dagger}_{l\times\mathbf{L}'\times 3\times\mathbf{N}'}$$

to write

$$\mathbb{E}_{\mathbf{m}}\left[[f_{\mathbf{m}}]^2_{\mathbf{L}}\right]_{l\times l'} = [KY3]^{\dagger}_{l\times\mathbf{L}''\times\mathbf{N}''}\,[H_{\mathbf{m}}(\Theta)]^{\dagger}_{l\times l'\times\mathbf{L}''\times\mathbf{L}'''\times\mathbf{N}''\times\mathbf{N}'''}\,[KY3]^{\dagger}_{l'\times\mathbf{L}'''\times\mathbf{N}'''}$$

results in

$$[H_{\mathbf{m}}(\Theta)]_{l\times l'\times l''\times l'''\times n''\times n'''} :=$$
$$\mathbb{E}_{\mathbf{m}}\left[\frac{\mathsf{p}\big([\mathsf{z}]_{\mathbf{m}}\big|\,[G]_{\mathbf{m}\times l\times l''\times n''}\,,[\Gamma]_{l\times l''\times\mathbf{m}\times\mathbf{m}}\big)\,\mathsf{p}\big([\mathsf{z}]_{\mathbf{m}}\big|\,[G]_{\mathbf{m}\times l'\times l'''\times n'''}\,,[\Gamma]_{l'\times l'''\times\mathbf{m}\times\mathbf{m}}\big)}{\mathsf{p}\big([\mathsf{z}]_{\mathbf{m}}\big|\,[0]_{\mathbf{m}}\,,\langle 1\rangle_{\mathbf{m}\times\mathbf{m}}\big)\,\mathsf{p}\big([\mathsf{z}]_{\mathbf{m}}\big|\,[0]_{\mathbf{m}}\,,\langle 1\rangle_{\mathbf{m}\times\mathbf{m}}\big)}\right]$$

Using Eq. (4) twice, with Hadamard (element-wise) division and product

$$[H_{\mathbf{m}}(\Theta)]_{l\times l'\times\mathbf{L}''\times\mathbf{L}'''\times\mathbf{N}''\times\mathbf{N}'''} := \left[|\Psi|^{-1}\right]_{l\times l'\times\mathbf{L}''\times\mathbf{L}'''} \circ$$
$$\frac{\mathsf{p}\big([G]_{\mathbf{m}\times l'\times\mathbf{L}'''\times\mathbf{N}'''}\big|\,[G]_{\mathbf{m}\times l\times\mathbf{L}''\times\mathbf{N}''}\,,[\Sigma]_{l\times l'\times\mathbf{L}''\times\mathbf{L}'''\times\mathbf{m}\times\mathbf{m}}\big)}{\mathsf{p}\big([0]_{\mathbf{m}}\big|\,[\Sigma G]_{\mathbf{m}\times l\times l'\times\mathbf{L}''\times\mathbf{L}'''\times\mathbf{N}''\times\mathbf{N}'''}\,,[\Sigma\Psi]_{l\times l'\times\mathbf{L}''\times\mathbf{L}'''\times\mathbf{m}\times\mathbf{m}}\big)}$$

where

$$
\begin{aligned}
[\Sigma]_{l\times l'\times l''\times l'''\times\mathbf{m}\times\mathbf{m}} &:= [\Gamma]_{l\times l''\times\mathbf{m}\times\mathbf{m}} + [\Gamma]_{l'\times l'''\times\mathbf{m}\times\mathbf{m}}\\
[\Psi]_{l\times l'\times l''\times l'''\times\mathbf{m}\times\mathbf{m}'} &:= [\Sigma]_{l\times l'\times l''\times l'''\times\mathbf{m}\times\mathbf{m}'}\\
&\quad - [\Gamma]_{l\times l''\times\mathbf{m}\times\mathbf{m}''}\,[\Gamma]_{l'\times l'''\times\mathbf{m}''\times\mathbf{m}'}\\
\left[|\Psi|^{-1}\right]_{l\times l'\times l''\times l'''} &:= \left|[\Psi]_{l\times l'\times l''\times l'''\times\mathbf{m}''\times\mathbf{m}'}\right|^{-1}\\
[\Sigma G]_{\mathbf{m}\times l\times l'\times l''\times l'''\times\mathbf{N}''\times\mathbf{N}'''} &:= [\Gamma]_{l'\times l'''\times\mathbf{m}\times\mathbf{m}'}\,[G]_{\mathbf{m}'\times l\times l''\times\mathbf{N}''}\\
&\quad + [\Gamma]_{l\times l''\times\mathbf{m}\times\mathbf{m}'}\,[G]_{\mathbf{m}'\times l'\times l'''\times\mathbf{N}'''}\\
[\Sigma\Psi]_{l\times l'\times l''\times l'''\times\mathbf{m}\times\mathbf{m}'} &:= [\Sigma]_{l\times l'\times l''\times l'''\times\mathbf{m}\times\mathbf{m}''}\,[\Psi]_{l\times l'\times l''\times l'''\times\mathbf{m}''\times\mathbf{m}'}
\end{aligned}
$$

4.2. Variances

Although it is not normally distributed, we shall use the **L**-variance of a stratum to measure its uncertainty

$$
\begin{aligned}
[T_{\mathbf{m}}(\Theta)]_{\mathbf{L}\times\mathbf{L}\times\mathbf{L}\times\mathbf{L}} &:= \mathbb{V}_{\mathbf{L}}\big[\mathsf{S}\big|\,[\Theta]_{\mathbf{m}\times\mathbf{M}}\big]_{\mathbf{L}\times\mathbf{L}} = \frac{[V_{\mathbf{m}}(\Theta)]^2_{\mathbf{L}\times\mathbf{L}}}{[V_{\mathbf{M}}(\Theta)]^2_{\mathbf{L}\times\mathbf{L}}}\\
&\circ\left(\frac{W_{\mathbf{m},\mathbf{m}}(\Theta)}{[V_{\mathbf{m}}(\Theta)]^2_{\mathbf{L}\times\mathbf{L}}} - 2\frac{W_{\mathbf{m},\mathbf{M}}(\Theta)}{[V_{\mathbf{m}}(\Theta)]_{\mathbf{L}\times\mathbf{L}}\otimes[V_{\mathbf{M}}(\Theta)]_{\mathbf{L}\times\mathbf{L}}} + \frac{W_{\mathbf{M},\mathbf{M}}(\Theta)}{[V_{\mathbf{M}}(\Theta)]^2_{\mathbf{L}\times\mathbf{L}}}\right)
\end{aligned}
$$

where for any $\mathbf{m},\mathbf{m}'\subseteq\mathbf{M}$, using the shorthand $\mathbf{L^4} := \mathbf{L}\times\mathbf{L}'\times\mathbf{L}''\times\mathbf{L}'''$

$$
\begin{aligned}
[W_{\mathbf{m},\mathbf{m}'}(\Theta)]_{\mathbf{L^4}} &:= \mathbb{V}_{\mathbf{L}}[\mathbb{V}_{\mathbf{m}}[\mathsf{e}_{\mathsf{m}}]\,,\mathbb{V}_{\mathbf{m}'}[\mathsf{e}_{\mathsf{m}'}]]\\
&= \mathbb{V}_{\mathbf{L}}\left[\mathbb{E}_{\mathbf{m}}\left[[\mathsf{e}_{\mathsf{m}}]^2_{\mathbf{L}} - [\mathsf{e}_0]^2_{\mathbf{L}}\right],\mathbb{E}_{\mathbf{m}'}\left[[\mathsf{e}_{\mathsf{m}'}]^2_{\mathbf{L}''} - [\mathsf{e}_0]^2_{\mathbf{L}''}\right]\right]\\
&= \mathbb{E}_{\mathbf{L}}\left[\mathbb{E}_{\mathbf{m}}\left[[\mathsf{e}_{\mathsf{m}}]^2_{\mathbf{L}} - [\mathsf{e}_0]^2_{\mathbf{L}}\right]\otimes\mathbb{E}_{\mathbf{m}'}\left[[\mathsf{e}_{\mathsf{m}'}]^2_{\mathbf{L}''} - [\mathsf{e}_0]^2_{\mathbf{L}''}\right]\right]\\
&\quad - [V_{\mathbf{m}}(\Theta)]_{\mathbf{L}\times\mathbf{L}'}\otimes[V_{\mathbf{m}'}(\Theta)]_{\mathbf{L}''\times\mathbf{L}'''}\\
&= [A_{\mathbf{m},\mathbf{m}'}(\Theta) - A_{\mathbf{0},\mathbf{m}'}(\Theta) - A_{\mathbf{m},\mathbf{0}}(\Theta) + A_{\mathbf{0},\mathbf{0}}(\Theta)]_{\mathbf{L^4}}\\
&\quad - [V_{\mathbf{m}}(\Theta)]_{\mathbf{L}\times\mathbf{L}'}\otimes[V_{\mathbf{m}'}(\Theta)]_{\mathbf{L}''\times\mathbf{L}'''}
\end{aligned}
\tag{15}
$$

Here, for any $\mathbf{m},\mathbf{m}'\subseteq\mathbf{M}$

$$
\begin{aligned}
[A_{\mathbf{m},\mathbf{m}'}(\Theta)]_{\mathbf{L^4}} &:= \mathbb{E}_{\mathbf{m}'}\,\mathbb{E}_{\mathbf{m}}\,\mathbb{E}_{\mathbf{L}}\left[[\mathsf{e}_{\mathsf{m}}]^2_{\mathbf{L}}\otimes[\mathsf{e}_{\mathsf{m}'}]^2_{\mathbf{L}''}\right]\\
&= \mathbb{E}_{\mathbf{m}'}\,\mathbb{E}_{\mathbf{m}}\,\mathbb{E}_{\mathbf{L}}\left[[\mathsf{c}_{\mathsf{m}} + f_{\mathbf{m}}]^2_{\mathbf{L}}\otimes[\mathsf{c}_{\mathsf{m}'} + f_{\mathbf{m}'}]^2_{\mathbf{L}''}\right]
\end{aligned}
$$

which according to Eq. (13) takes expected values over $\mathbf{L}$ of

$$\begin{aligned}
[A_{\mathbf{m},\mathbf{m}'}(\Theta)]_{\mathbf{L}^4} = \mathbb{E}_{\mathbf{m}'}\mathbb{E}_{\mathbf{m}}\big[&[\sigma_{\mathbf{m},\mathbf{m}}]_{\mathbf{L}\times\mathbf{L}'}\otimes[\sigma_{\mathbf{m}',\mathbf{m}'}]_{\mathbf{L}''\times\mathbf{L}'''}\\
&+[\sigma_{\mathbf{m},\mathbf{m}'}]_{\mathbf{L}\times\mathbf{L}''}\otimes[\sigma_{\mathbf{m},\mathbf{m}'}]_{\mathbf{L}'\times\mathbf{L}'''}+[\sigma_{\mathbf{m},\mathbf{m}'}]_{\mathbf{L}\times\mathbf{L}'''}\otimes[\sigma_{\mathbf{m},\mathbf{m}'}]_{\mathbf{L}'\times\mathbf{L}''}\\
&+[\sigma_{\mathbf{m},\mathbf{m}}]_{\mathbf{L}\times\mathbf{L}'}\otimes[f_{\mathbf{m}'}]^2_{\mathbf{L}''}+[f_{\mathbf{m}}]^2_{\mathbf{L}}\otimes[\sigma_{\mathbf{m}',\mathbf{m}'}]_{\mathbf{L}''\times\mathbf{L}'''}+[f_{\mathbf{m}}]^2_{\mathbf{L}}\otimes[f_{\mathbf{m}'}]^2_{\mathbf{L}''}\\
&+[f_{\mathbf{m}}]_{\mathbf{L}'}\oslash[\sigma_{\mathbf{m},\mathbf{m}'}]_{\mathbf{L}\times\mathbf{L}''}\oslash[f_{\mathbf{m}'}]_{\mathbf{L}'''}+[f_{\mathbf{m}}]_{\mathbf{L}}\oslash[\sigma_{\mathbf{m},\mathbf{m}'}]_{\mathbf{L}'\times\mathbf{L}''}\oslash[f_{\mathbf{m}'}]_{\mathbf{L}'''}\\
&+[f_{\mathbf{m}}]_{\mathbf{L}}\oslash[\sigma_{\mathbf{m},\mathbf{m}'}]_{\mathbf{L}'\times\mathbf{L}'''}\oslash[f_{\mathbf{m}'}]_{\mathbf{L}''}+[f_{\mathbf{m}}]_{\mathbf{L}'}\oslash[\sigma_{\mathbf{m},\mathbf{m}'}]_{\mathbf{L}\times\mathbf{L}'''}\oslash[f_{\mathbf{m}'}]_{\mathbf{L}''}\big]
\end{aligned}$$

The binary operation $\oslash$ is the exterior product $\otimes$ followed by multi-index permutation to restore the original order $\mathbf{L}\times\mathbf{L}'\times\mathbf{L}''\times\mathbf{L}'''$. These expressions shrink naturally as follows. Firstly,

$$\mathbb{E}_{\mathbf{m}'}\mathbb{E}_{\mathbf{m}}\big[[f_{\mathbf{m}}]^2_{\mathbf{L}}\otimes[f_{\mathbf{m}'}]^2_{\mathbf{L}''}\big]=\big[V_{\mathbf{m}}(\Theta)+[f_{\mathbf{0}}]^2\big]_{\mathbf{L}\times\mathbf{L}'}\otimes\big[V_{\mathbf{m}'}(\Theta)+[f_{\mathbf{0}}]^2\big]_{\mathbf{L}''\times\mathbf{L}'''}$$

which eliminates all terms free of σ from Eq. (15). Secondly, by the law of iterated expectations

$$\mathbb{E}_{\mathbf{m}}[\sigma_{\mathbf{m},\mathbf{m}}]=\mathbb{E}_{\mathbf{m}'}[\sigma_{\mathbf{m}',\mathbf{m}'}]=\sigma_{\mathbf{0},\mathbf{0}}$$

which eliminates all remaining terms free of $\sigma_{\mathbf{m},\mathbf{m}'}$ from Eq. (15). Again by the law of iterated expectations and the structure of Eq. (15), all terms containing E from $\sigma_{\mathbf{m},\mathbf{m}'}$ cancel, so we may finally write

$$[W_{\mathbf{m},\mathbf{m}'}(\Theta)]_{\mathbf{L}^4}=[B_{\mathbf{m},\mathbf{m}'}(\Theta)-B_{\mathbf{0},\mathbf{m}'}(\Theta)-B_{\mathbf{m},\mathbf{0}}(\Theta)+B_{\mathbf{0},\mathbf{0}}(\Theta)]_{\mathbf{L}^4} \tag{16}$$

component-wise

$$\begin{aligned}
[B_{\mathbf{m},\mathbf{m}'}(\Theta)]_{l\times l'\times l''\times l'''} &:= \mathbb{E}_{\mathbf{m}'}\mathbb{E}_{\mathbf{m}}\big[[\bar\sigma_{\mathbf{m},\mathbf{m}'}]_{l\times l''}[\bar\sigma_{\mathbf{m},\mathbf{m}'}]_{l'\times l'''}\\
+[\bar\sigma_{\mathbf{m},\mathbf{m}'}]_{l\times l'''}[\bar\sigma_{\mathbf{m},\mathbf{m}'}]_{l'\times l''}&+[f_{\mathbf{m}}]_{l'}[\bar\sigma_{\mathbf{m},\mathbf{m}'}]_{l\times l''}[f_{\mathbf{m}'}]_{l'''}+[f_{\mathbf{m}}]_{l}[\bar\sigma_{\mathbf{m},\mathbf{m}'}]_{l'\times l''}[f_{\mathbf{m}'}]_{l'''}\\
&+[f_{\mathbf{m}}]_{l}[\bar\sigma_{\mathbf{m},\mathbf{m}'}]_{l'\times l'''}[f_{\mathbf{m}'}]_{l''}+[f_{\mathbf{m}}]_{l'}[\bar\sigma_{\mathbf{m},\mathbf{m}'}]_{l\times l'''}[f_{\mathbf{m}'}]_{l''}\big]
\end{aligned}$$

where

$$[\bar\sigma(\mathsf{z};\Theta)]_{\mathbf{L}\times\mathbf{L}'\times\mathbf{3}\times\mathbf{3}'}:=[\pm F]_{\mathbf{L}\times\mathbf{L}'\times\mathbf{3}\times\mathbf{3}'}\circ[\phi(\mathsf{z};\Theta)]_{\mathbf{L}\times\mathbf{L}'\times\mathbf{3}\times\mathbf{3}'}-[\psi(\mathsf{z};\Theta)]_{\mathbf{L}\times\mathbf{L}'\times\mathbf{3}\times\mathbf{3}'}$$

5. spare

The binary operation $\oslash$ is the exterior product $\otimes$ followed by multi-index permutation to restore the original order $\mathbf{L} \times \mathbf{L}' \times \mathbf{L}'' \times \mathbf{L}'''$

$$
\begin{aligned}
[A_{\mathbf{m},\mathbf{m}'}]_{\mathbf{L}\times\mathbf{L}'\times\mathbf{L}''\times\mathbf{L}'''} &= \mathbb{E}_{\mathbf{m}'}\,\mathbb{E}_{\mathbf{m}}\,\mathbb{E}_{\mathbf{L}}\big[[\mathsf{c}_{\mathsf{m}}]^2_{\mathbf{L}} \otimes [\mathsf{c}_{\mathsf{m}'}]^2_{\mathbf{L}''} \\
&+ [\mathsf{c}_{\mathsf{m}}]^2_{\mathbf{L}} \otimes [f_{\mathbf{m}'}]^2_{\mathbf{L}''} + [f_{\mathbf{m}}]^2_{\mathbf{L}} \otimes [\mathsf{c}_{\mathsf{m}'}]^2_{\mathbf{L}''} + [f_{\mathbf{m}}]^2_{\mathbf{L}} \otimes [f_{\mathbf{m}'}]^2_{\mathbf{L}''} \\
&+ [\mathsf{c}_{\mathsf{m}}]_{\mathbf{L}} \otimes [f_{\mathbf{m}}]_{\mathbf{L}'} \otimes [\mathsf{c}_{\mathsf{m}'}]_{\mathbf{L}''} \otimes [f_{\mathbf{m}'}]_{\mathbf{L}'''} + [f_{\mathbf{m}}]_{\mathbf{L}} \otimes [\mathsf{c}_{\mathsf{m}}]_{\mathbf{L}'} \otimes [\mathsf{c}_{\mathsf{m}'}]_{\mathbf{L}''} \otimes [f_{\mathbf{m}'}]_{\mathbf{L}'''} \\
&+ [f_{\mathbf{m}}]_{\mathbf{L}} \otimes [\mathsf{c}_{\mathsf{m}}]_{\mathbf{L}'} \otimes [f_{\mathbf{m}'}]_{\mathbf{L}''} \otimes [\mathsf{c}_{\mathsf{m}'}]_{\mathbf{L}'''} + [\mathsf{c}_{\mathsf{m}}]_{\mathbf{L}} \otimes [f_{\mathbf{m}}]_{\mathbf{L}'} \otimes [f_{\mathbf{m}'}]_{\mathbf{L}''} \otimes [\mathsf{c}_{\mathsf{m}'}]_{\mathbf{L}'''}\big] \\
&= \mathbb{E}_{\mathbf{m}'}\,\mathbb{E}_{\mathbf{m}}\big[[\sigma_{\mathbf{m},\mathbf{m}}]_{\mathbf{L}\times\mathbf{L}'} \otimes [\sigma_{\mathbf{m}',\mathbf{m}'}]_{\mathbf{L}''\times\mathbf{L}'''} + [\sigma_{\mathbf{m},\mathbf{m}'}]_{\mathbf{L}\times\mathbf{L}''} \otimes [\sigma_{\mathbf{m},\mathbf{m}'}]_{\mathbf{L}'\times\mathbf{L}'''} \\
&+ [\sigma_{\mathbf{m},\mathbf{m}'}]_{\mathbf{L}\times\mathbf{L}'''} \otimes [\sigma_{\mathbf{m},\mathbf{m}'}]_{\mathbf{L}'\times\mathbf{L}''} + [\sigma_{\mathbf{m},\mathbf{m}}]_{\mathbf{L}\times\mathbf{L}'} \otimes [f_{\mathbf{m}'}]^2_{\mathbf{L}''} \\
&+ [f_{\mathbf{m}}]^2_{\mathbf{L}} \otimes [\sigma_{\mathbf{m}',m'}]_{\mathbf{L}''\times\mathbf{L}'''} + [f_{\mathbf{m}}]^2_{\mathbf{L}} \otimes [f_{\mathbf{m}'}]^2_{\mathbf{L}''} \\
&+ [f_{\mathbf{m}}]_{\mathbf{L}'} \oslash [\sigma_{\mathbf{m},\mathbf{m}'}]_{\mathbf{L}\times\mathbf{L}''} \oslash [f_{\mathbf{m}'}]_{\mathbf{L}'''} + [f_{\mathbf{m}}]_{\mathbf{L}} \oslash [\sigma_{\mathbf{m},\mathbf{m}'}]_{\mathbf{L}'\times\mathbf{L}''} \oslash [f_{\mathbf{m}'}]_{\mathbf{L}'''} \\
&+ [f_{\mathbf{m}}]_{\mathbf{L}} \oslash [\sigma_{\mathbf{m},\mathbf{m}'}]_{\mathbf{L}'\times\mathbf{L}'''} \oslash [f_{\mathbf{m}'}]_{\mathbf{L}''} + [f_{\mathbf{m}}]_{\mathbf{L}'} \oslash [\sigma_{\mathbf{m},\mathbf{m}'}]_{\mathbf{L}\times\mathbf{L}'''} \oslash [f_{\mathbf{m}'}]_{\mathbf{L}''}\big]
\end{aligned}
$$

while for $i = i', \mathbf{m} := [\mathbf{m}]_i \times i$

$$
[\phi(\mathsf{z};\Theta)]_{l\times l'\times i\times i} := \frac{\mathbb{E}_{[\mathbf{M}]-[\mathbf{m}]}\,[k([\mathsf{x}]_i\,,[\mathsf{x}]_i)]_{l\times l'}}{[\pm F]_{l\times l'}} = 1
$$

$$
\begin{aligned}
[\psi(\mathsf{z};\Theta)]^{\dagger}_{l\times l'\times i\times i} := [g(\mathsf{z};\Theta)]^{\dagger}_{l\times\mathbf{L}''\times 3\times\mathbf{N}''}\Big(&\big[K_Y^{-1}\big]_{\mathbf{L}''\mathbf{N}''\times\mathbf{L}'''\mathbf{N}'''} \\
&\circ \left[\frac{\mathsf{p}\big([\mathsf{z}]_{\mathbf{m}}|\,[Q]_{\mathbf{m}\times l\times l'\times\mathbf{L}''\times\mathbf{L}'''\times\mathbf{N}''\times\mathbf{N}'''}\,,[\Psi]_{l\times l'\times\mathbf{L}''\times\mathbf{L}'''\times\mathbf{m}\times\mathbf{m}}\big)}{\mathsf{p}\big([\mathsf{z}]_{\mathbf{m}}|\,[0]_{\mathbf{m}}\,,\langle 1\rangle_{\mathbf{m}\times\mathbf{m}}\big)}\right]^{\dagger} \\
&\circ \left[\frac{[P]_{l\times l'\times\mathbf{L}''\times\mathbf{L}'''\times\mathbf{N}''\times\mathbf{N}'''}}{\mathsf{p}\big([0]_{\mathbf{M}}|\,[X]_{\mathbf{M}\times\mathbf{N}'''}\,,\langle\Lambda^2+1\rangle_{l'\times\mathbf{L}'''\times\mathbf{M}\times\mathbf{M}}\big)}\right]^{\dagger}\Big)[g(\mathsf{z};\Theta)]^{\dagger}_{l'\times\mathbf{L}'''\times 3\times\mathbf{N}'''}
\end{aligned}
$$

where

$$
\begin{aligned}
[P]_{l\times l'\times l''\times l'''\times \mathbf{N}''\times \mathbf{N}'''} &:= \mathsf{p}\left(\left\langle \Lambda^2+1\right\rangle^{-1}_{l\times l''\times \mathbf{M}\times \mathbf{M}'}[X]_{\mathbf{M}'\times \mathbf{N}''}\middle| [X]_{\mathbf{M}\times \mathbf{N}'''},\right.\\
&\left.\left\langle \Lambda^2+1\right\rangle_{l'\times l'''\times \mathbf{M}\times \mathbf{M}} - \left\langle \Lambda^2+1\right\rangle^{-1}_{l\times l''\times \mathbf{M}\times \mathbf{M}}\right)\\
[\Psi]_{l\times l'\times l''\times l'''\times \mathbf{m}\times \mathbf{m}} &:= [\Theta]_{\mathbf{m}\times \mathbf{M}''}\left\langle \left\langle \Lambda^2\right\rangle_{l\times l''}\left\langle \Lambda^2\right\rangle_{l'\times l'''}\right\rangle_{\mathbf{M}''\times \mathbf{M}'}\\
&\left\langle \left\langle \Lambda^2+1\right\rangle_{l\times l''}\left\langle \Lambda^2+1\right\rangle_{l'\times l'''}-1\right\rangle^{-1}_{\mathbf{M}'\times \mathbf{M}}[\Theta]^{\intercal}_{\mathbf{m}\times \mathbf{M}}\\
[Q]_{\mathbf{m}\times l\times l'\times l''\times l'''\times \mathbf{N}''\times \mathbf{N}'''} &:= [\Psi]_{l\times l'\times l''\times l'''\times \mathbf{m}\times \mathbf{M}''}[\Theta]_{\mathbf{M}''\times \mathbf{M}}\\
&\left(\left\langle \left\langle \Lambda^2\right\rangle^{-1}_{l\times l''}\left\langle \Lambda^2+1\right\rangle_{l\times l''}\right\rangle_{\mathbf{M}\times \mathbf{M}'}[X]_{\mathbf{M}'\times \mathbf{N}''}\right.\\
&\left.+\left\langle \Lambda^2\right\rangle^{-1}_{l'\times l'''\times \mathbf{M}\times \mathbf{M}'}[X]_{\mathbf{M}'\times \mathbf{N}'''}\right)
\end{aligned}
$$

References

[1] C. E. Rasmussen, Some useful gaussian and matrix equations (2016).

[2] L. Isserlis, On certain probable errors and correlation coefficients of multiple frequency distributions with skew regression 11 (3) 185–190. doi:10.1093/biomet/11.3.185.

[3] L. Isserlis, On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables 12 (1-2) 134–139. doi:10.1093/biomet/12.1-2.134.

[4] J. E. Oakley, A. O'Hagan, Probabilistic sensitivity analysis of complex models: a bayesian approach, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66 (3) (2004) 751–769. doi:10.1111/j.1467-9868.2004.05304.x.