Spatio-temporal statistics (MATH4341) Michaelmas term, 2023

Problem class sheet 3

(DRAFT, TO BE REFINED AFTER THE PROBLEM CLASS)

Lecturer & author: Georgios P. Karagiannis georgios.karagiannis@durham.ac.uk

Exercise 1. ($\star\star$)

Inventory of useful formulas.

[Normal distr. conditioning] Let $x_1 \in \mathbb{R}^{d_1}$, and $x_2 \in \mathbb{R}^{d_2}$. If

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathrm{N}_{d_1+d_2}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_1 & \Sigma_{21}^\top \\ \Sigma_{21} & \Sigma_2 \end{bmatrix}\right)$$

then it is

$$x_2|x_1 \sim \mathrm{N}_{d_2}\left(\mu_{2|1}, \Sigma_{2|1}\right)$$

where

$$\mu_{2|1} = \mu_2 + \Sigma_{21}\Sigma_1^{-1}\left(x_1 - \mu_1\right) \text{ and } \Sigma_{2|1} = \Sigma_2 - \Sigma_{21}\Sigma_1^{-1}\Sigma_{21}^\top$$

Consider the Bayesian Kriging from your lecture notes:

$$Z(s) = Y(s) + \varepsilon(s), \quad s \in \mathcal{S}$$

where

$$\varepsilon(\cdot) \sim \mathrm{GP}\left(0, c_\varepsilon(\cdot,\cdot|\tau)\right)$$

with $c_\varepsilon(s, s'|\tau) = \tau^2 1_{\{0\}}(\|s - s'\|)$ and

$$Y(\cdot)|\beta,\theta \sim \mathrm{GP}\left(\mu(\cdot|\beta), c_Y\left(\cdot,\cdot|\sigma^2,\phi\right)\right)$$

with mean function $\mu(\cdot|\beta)$ (to be specified later) labeled by unknown parameter β and covariance function $c_Y(\cdot,\cdot|\sigma^2,\phi)$.

Assume there is available a dataset $\{(s_i, Z_i)\}_{i=1}^n$ where $Z_i = Z(s_i)$ is a realization of a stochastic process (Z_s).

(1) Write the hierarchical spatial model $Z(\cdot)|Y(\cdot),\beta,\varphi$ and $Y(\cdot)|\beta,\varphi$ where $\varphi = (\sigma^2,\phi,\tau)^\top$.

(2) Write the marginal process $Z(\cdot)|\beta,\varphi$ where $\varphi = (\sigma^2,\phi,\tau)^\top$, its mean function denoted as $\mu(\cdot|\cdot)$, and its covariance function denoted as $c(\cdot|\cdot)$.

Hint:: Let Y and X be independent random variables with $X \sim \mathrm{N}(\mu_X, \Sigma_X)$, $Y \sim \mathrm{N}(\mu_Y, \Sigma_Y)$. Let A and B be fixed matrices. Let c be a fixed vector. Then

$$AX + BY + c \sim \mathrm{N}\left(A\mu_X + B\mu_Y + c, A\Sigma_X A^\top + B\Sigma_Y B^\top\right)$$

(3) Compute the predictive process $Z(\cdot)|Z,\beta,\varphi$ as

$$Z(\cdot)|Z,\beta,\varphi \sim \mathrm{GP}\left(\mu_1(\cdot|\beta,\varphi), c_1(\cdot,\cdot|\varphi)\right)$$

with

$$\begin{aligned} c_1(s,s'|\varphi) =& c(s,s|\varphi) + (C(S,s|\varphi))^\top (C(S,S|\varphi))^{-1} C(S,s'|\varphi) \\ \mu_1(s|\beta,\varphi) =& \mu(s|\beta) - (C(S,s|\varphi))^\top (C(S,S|\varphi))^{-1} (\mu(S|\beta) - Z) \end{aligned}$$

Hint: See the Conditional Normal formula above.

(4) Assume $\mu(s|\beta) = \psi(s)^\top \beta$. Consider a conjugate prior on β as $\beta \sim \mathrm{N}(b, B)$ where $B > 0$.

(a) Write down the Bayesian statistical model involving $[Z|\beta,\varphi]$, and $[\beta|\varphi]$.

(b) Compute the posterior distribution as

$$\beta|Z,\varphi \sim \mathrm{N}(b_n(\varphi), B_n(\varphi))$$

with

$$\begin{aligned} B_n =& \left(B^{-1} + \Psi^\top (C(S,S|\varphi))^{-1} \Psi\right)^{-1} \\ b_n =& B_n \left(B^{-1} b + \Psi^\top (C(S,S|\varphi))^{-1} Z\right) \end{aligned}$$

where $R(S,S|\varphi)$ is a matrix with $[R(S,S|\varphi)]_{i,j} = r(s_i, s_j|\varphi)$.

Hint: Use the following identity

$$(y-\Phi\beta)^\top \Sigma^{-1}(y-\Phi\beta) + (\beta-\mu)^\top V^{-1}(\beta-\mu) = (\beta-\mu^*)^\top (V^*)^{-1} (\beta-\mu^*) + S^*;$$

$$V^* = \left(V^{-1} + \Phi^\top \Sigma^{-1} \Phi\right)^{-1}; \qquad \mu^* = V^* \left(V^{-1}\mu + \Phi^\top \Sigma^{-1} y\right)$$

$$S^* = \mu^\top V^{-1} \mu - (\mu^*)^\top (V^*)^{-1} (\mu^*) + y^\top \Sigma^{-1} y;$$

(c) Compute the (posterior) predictive process $Z(\cdot)|Z,\varphi$ given the data Z and given the parameters φ as

$$Z(\cdot)|Z,\varphi \sim \mathrm{GP}\left(\mu_2(\cdot|\varphi), c_2(\cdot,\cdot|\varphi)\right)$$

with

$$\begin{aligned} \mu_2(s|\varphi) =& \left(\Psi C^{-1} (C(s))^\top - \psi(s)\right)^\top \left(B^{-1} + \Psi^\top C^{-1} \Psi\right)^{-1} B^{-1} b \\ &+ \left[(C(s))^\top + \left(\Psi C^{-1} (C(s))^\top - \psi(s)\right)^\top \left(B^{-1} + \Psi^\top C^{-1} \Psi\right)^{-1} \Psi\right] C^{-1} Z \\ c_2(s,s'|\varphi) =& c(s,s'|\varphi) - (C(s))^\top C^{-1} C(s') \\ &+ \left(\Psi C^{-1} (C(s))^\top - \psi(s)\right)^\top \left(B^{-1} + \Psi^\top C^{-1} \Psi\right)^{-1} \left(\Psi C^{-1} (C(s'))^\top - \psi(s')\right) \end{aligned}$$

with column vector $C(s) := (c(s, s_1|\varphi), ..., c(s, s_n|\varphi))^\top$, and matrix $C := C(S, S|\varphi)$.

(d) Compute the marginal likelihood $\Pr(Z|\varphi)$ in the form

$$\Pr\left(Z|\sigma^2, \varphi\right) = \mathrm{N}\left(Z|\Psi b, \left(C^{-1} - C^{-1}\Psi\left(B^{-1} + \Psi^\top B^{-1}\Psi\right)^{-1}\Psi^\top C^{-1}\right)^{-1}\right)$$

where Ψ is a matrix with $[\Psi]_{i,j} = \psi_j(s_i)$, and R is a matrix with $[C]_{i,j} = c(s_i, s_j|\varphi)$.

Hint-2:: It is

$$\int \mathrm{N}(Z|\Psi\beta, C)\,\mathrm{N}(\beta|b, B)\,\mathrm{d}\beta = \mathrm{N}\left(Z|\Psi b, C + \Psi B\Psi^\top\right)$$

Hint 3:: [Woodbury matrix identity]

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U\left(C^{-1} + VA^{-1}U\right)^{-1}VA^{-1}$$

(5) Consider non-informative prior $\Pr(\beta) \propto 1$ for β by specifying $b \to 0$ and letting $B^{-1} \to 0$. Argue whether such a prior can be used. Recompute the (asymptotic) quantities $\Pr(Z|\varphi)$, $[Z(\cdot)|Z, \varphi]$ under this new prior.

Solution.

(1) The hierarchical model was

$$\begin{aligned} Z(\cdot)|Y(\cdot), \tau &\sim \mathrm{GP}\left(Y(\cdot), c_\varepsilon\left(\cdot, \cdot|\sigma^2, \xi\right)\right) \\ Y(\cdot)|\beta, &\sim \mathrm{GP}\left(\mu(\cdot|\beta), c_Y\left(\cdot, \cdot|\sigma^2, \phi\right)\right) \end{aligned}$$

(2) We use the additive property of the Gaussian distribution (Hint-1) it is

$$Z(\cdot)|\beta, \varphi \sim \mathrm{GP}(\mu(\cdot|\beta), c(\cdot, \cdot|\varphi))$$

where

$$c(s, s'|\varphi) = c_Y\left(s, s'|\sigma^2, \phi\right) + c_\varepsilon\left(s, s'|\sigma^2, \xi\right)$$

(3) Assume a vector of "unseen" sites $S_* = (s_{*,1}, ..., s_{*,q})^\top$ for any $q \in \mathbb{N}_0$. Let convenient notation $Z := Z(S)$, and $Z_* := Z(S_*)$. The join marginal distribution of $(Z_*, Z)^\top$ given β, $\varphi = (\sigma^2, \phi, \tau)^\top$ is

$$\begin{pmatrix} Z_* \\ Z \end{pmatrix}|\beta, \varphi \sim \mathrm{N}\left(\begin{pmatrix} \mu(S_*; \beta) \\ \mu(S; \beta) \end{pmatrix}, \begin{pmatrix} C(S_*, S_*|\varphi) & (C(S_*, S|\varphi))^\top \\ C(S_*, S|\varphi) & C(S, S|\varphi) \end{pmatrix}\right)$$

by using convenient notation $[C(S_*, S|\varphi)]_{i,j} = s(s_{*,i}, s_j|\varphi)$ and $[\mu(S; \beta)]_i = \mu(s_i; \beta)$.

Using the Normal distribution conditioning, I get

$$Z_*|Z, \beta, \varphi \sim \mathrm{N}(\mu_*(S_*|\beta, \varphi), C_*(S_*, S_*|\varphi))$$

where

$$\begin{aligned} C_*\left(S_*,S_*|\varphi\right) =& C\left(S_*,S_*|\varphi\right)+\left(C\left(S,S_*|\varphi\right)\right)^\top\left(C\left(S,S|\varphi\right)\right)^{-1}C\left(S,S_*|\varphi\right) \\ \mu_*\left(S_*|\beta,\varphi\right) =& \mu\left(S_*|\beta\right)-\left(C\left(S,S_*|\varphi\right)\right)^\top\left(C\left(S,S|\varphi\right)\right)^{-1}\left(\mu\left(S|\beta\right)-Z\right) \end{aligned}$$

As it is for any length of of any vector S_*, then it is a Gaussian process

$$Z\left(\cdot\right)|Z,\beta,\varphi\sim\mathrm{GP}\left(\mu_1\left(\cdot|\beta,\varphi\right),c_1\left(\cdot,\cdot|\varphi\right)\right)$$

with

$$\begin{aligned} c_1\left(s,s'|\varphi\right) =& c\left(s,s|\varphi\right)+\left(C\left(S,s|\varphi\right)\right)^\top\left(C\left(S,S|\varphi\right)\right)^{-1}C\left(S,s'|\varphi\right) \\ \mu_1\left(s|\beta,\varphi\right) =& \mu\left(s|\beta\right)-\left(C\left(S,s|\varphi\right)\right)^\top\left(C\left(S,S|\varphi\right)\right)^{-1}\left(\mu\left(S|\beta\right)-Z\right) \end{aligned}$$

(4)

(a) the Bayesian model is

$$\begin{cases} Z|\beta,\varphi\sim\mathrm{N}\left(\Psi\beta,C\left(S,S|\varphi\right)\right) \\ \beta|\varphi\sim\mathrm{N}\left(b,B\right) \end{cases} \tag{0.1}$$

(b) Let $C:=C\left(S,S|\varphi\right)$. The posterior distribution (by using Bayes theorem) is

$$\begin{aligned} \Pr\left(\beta|Z,\varphi\right) \propto& \Pr\left(Z|\beta,\varphi\right)\Pr\left(\beta|\varphi\right) \\ =& \mathrm{N}\left(Z|\Psi\beta,C\right)\mathrm{N}\left(\beta|b,B\right) \\ \propto& \exp\left(-\frac{1}{2}\left(Z-\Psi\beta\right)^\top C^{-1}\left(Z-\Psi\beta\right)\right)\exp\left(-\frac{1}{2}\left(\beta-b\right)^\top B^{-1}\left(\beta-b\right)\right) \\ =& \exp\left(-\frac{1}{2}\left[\left(Z-\Psi\beta\right)^\top C^{-1}\left(Z-\Psi\beta\right)+\left(\beta-b\right)^\top B^{-1}\left(\beta-b\right)\right]\right) \end{aligned}$$

By using the hint I have

$$\left(Z-\Psi\beta\right)^\top C^{-1}\left(Z-\Psi\beta\right)+\left(\beta-b\right)^\top B^{-1}\left(\beta-b\right)=\left(\beta-b_n\right)^\top\left(B_n\right)^{-1}\left(\beta-b_n\right)+R_n$$

where by denoting $B_n:=B_n\left(\varphi\right)$, and $b_n:=b_n\left(\varphi\right)$ I get

$$\begin{aligned} B_n =& \left(B^{-1}+\Psi^\top C^{-1}\Psi\right)^{-1} \\ b_n =& B_n\left(B^{-1}b+\Psi^\top C^{-1}Z\right) \end{aligned}$$

and R_n is a "constant" quantity that does not contain any β. Hence

$$\begin{aligned} \Pr\left(\beta|Z,\varphi\right) \propto& \exp\left(-\frac{1}{2}(\beta-b_n)^\top\left(B_n\right)^{-1}\left(\beta-b_n\right)-\frac{1}{2}R_n\right) \\ \propto& \exp\left(-\frac{1}{2}(\beta-b_n)^\top\left(B_n\right)^{-1}\left(\beta-b_n\right)\right) \end{aligned}$$

Well, from the above, I recognize the kernel of the Multivariate Normal distribution, as

$$\beta|Z,\varphi \sim \mathrm{N}\left(b_n\left(\varphi\right), B_n\left(\varphi\right)\right)$$

(c) Assume a vector of "unseen" sites $S_* = \left(s_{*,1}, ..., s_{*,q}\right)^\top$ for any $q \in \mathbb{N}_0$. Let convenient notation $Z := Z\left(S\right)$, and $Z_* := Z\left(S_*\right)$. It is

$$\begin{aligned}\Pr\left(Z_*|Z,\varphi\right) &= \int \Pr\left(Z_*|Z,\varphi\right)\Pr\left(\beta|Z,\varphi\right)\mathrm{d}\beta \\ &= \int \mathrm{N}\left(Z_*|\mu_1\left(S_*\right), C_1\left(S_*, S_*\right)\right)\mathrm{N}\left(\beta|b_n, B_n\right)\mathrm{d}\beta\end{aligned}$$

Let $\Psi_* = \Psi\left(S_*\right)$, $C_* = C\left(S_*, S|\varphi\right)$, and $C_{**} = C\left(S_*, S_*|\varphi\right)$. Notice that

$$\begin{aligned}\mu_1\left(S_*\right) =& \Psi_*\beta - C_*C^{-1}\left(\Psi\beta - Z\right) \\ =& \left[\Psi_* - C_*C^{-1}\right]\beta + C_*C^{-1}Z\end{aligned}$$

Hence, for given/fixed Z, φ, it is

$$Z_* = C_*C^{-1}Z + \left[\Psi_* - C_*C^{-1}\right]\beta + \zeta, \quad \zeta \sim \mathrm{N}\left(0, C_1\left(S_*, S_*\right)\right)$$

Hence, because $\beta \sim \mathrm{N}\left(b_n, B_n\right)$, and because $Z_*|Z,\varphi$ is a linear combination of the Normally distributed random vector $\beta \sim \mathrm{N}\left(b_n, B_n\right)$, $Z_*|Z,\varphi$ follows a Normal distribution, with mean

$$\begin{aligned}\mu_2\left(S_*\right) =& \mathrm{E}_{\beta\sim\mathrm{N}(b_n,B_n)}\left(Z_*|\mu_1\left(S_*\right), C_1\left(S_*, S_*\right)\right) \\ =& \left(\Psi_* - C_*C^{-1}\right)\mathrm{E}_{\beta\sim\mathrm{N}(b_n,B_n)}\left(\beta\right) + C_*C^{-1}Z \\ =& \left(\Psi_* - C_*C^{-1}\right)b_n + C_*C^{-1}Z \\ =& \left(\Psi_* - C_*C^{-1}\right)\left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1}\left(B^{-1}b + \Psi^\top C^{-1}Z\right) + C_*C^{-1}Z \\ =& \left(\Psi_* - C_*C^{-1}\right)\left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1}B^{-1}b \\ &+ \left[\left(\Psi_* - C_*C^{-1}\right)\left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1}\Psi^\top + C_*\right]C^{-1}Z\end{aligned}$$

and with covariance matrix

$$\begin{aligned}C_2\left(S_*, S_*\right) =& \mathrm{Var}_{\beta\sim\mathrm{N}(b_n,B_n)}\left(Z_*|\mu_1\left(S_*\right), C_1\left(S_*, S_*\right)\right) \\ =& \mathrm{Var}_{\beta\sim\mathrm{N}(b_n,B_n)}\left(\left[\Psi_* - C_*C^{-1}\right]\beta\right) + \mathrm{Var}_{\zeta\sim\mathrm{N}(0,C_1(S_*,S_*))}\left(\zeta\right) \\ =& \left[\Psi_* - C_*C^{-1}\right]B_n\left[\Psi_* - C_*C^{-1}\right]^\top + C_1\left(S_*, S_*\right) \\ =& \left[\Psi_* - C_*C^{-1}\right]\left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1}\left[\Psi_* - C_*C^{-1}\right]^\top \\ &+ C_{**} + \left(C_*\right)^\top C^{-1}C_*\end{aligned}$$

Since this is for any vector S_* of any length, then

$$Z(\cdot)|Z,\varphi \sim \mathrm{GP}\left(\mu_2(\cdot|\varphi), c_2(\cdot,\cdot|\varphi)\right)$$

with mean function at s

$$\begin{aligned}\mu_2(s|\varphi) = & \left(\psi(s) - C(s)C^{-1}\right)\left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1} B^{-1} b \\ & + \left[\left(\psi(s) - C(s)C^{-1}\right)\left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1}\Psi^\top + C(s)\right] C^{-1} Z\end{aligned}$$

and covariance function as s, s'

$$\begin{aligned}c_2(s,s'|\varphi) = & \left[\psi(s) - C(s)C^{-1}\right]\left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1}\left[\psi(s) - C(s)C^{-1}\right]^\top \\ & + c(s,s'|\varphi) + (C(s))^\top C^{-1} C(s)\end{aligned}$$

(d) It is, from Hint-2,

$$\begin{aligned}\Pr(Z|\varphi) = & \int \Pr(Z|\beta,\varphi)\Pr(\beta)\,\mathrm{d}\beta \\ = & \int \mathrm{N}(Z|\Psi\beta, C(S,S|\varphi))\,\mathrm{N}(\beta|b,B)\,\mathrm{d}\beta \\ = & \int \mathrm{N}(Z|\Psi\beta, C(S,S|\varphi))\,\mathrm{N}\left(\Psi\beta|\Psi b, \Psi B\Psi^\top\right)\mathrm{d}\beta \\ = & \mathrm{N}\left(Z|\Psi b, C(S,S|\varphi) + \Psi B\Psi^\top\right)\end{aligned}$$

where by letting $C = C(S,S|\varphi)$ and using the Hint I get

$$\left(C + \Psi B\Psi^\top\right)^{-1} = C^{-1} - C^{-1}\Psi\left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1}\Psi^\top C^{-1}$$

(5) Denote $C = C(S,S|\varphi)$. It is

$$\begin{aligned}\lim_{\substack{B^{-1}\to 0\\ b\to 0}} \Pr(Z|\varphi) = & \lim_{\substack{B^{-1}\to 0\\ b\to 0}} \mathrm{N}\left(Z|\Psi b, C + \Psi B\Psi^\top\right) \\ \propto & \lim_{\substack{B^{-1}\to 0\\ b\to 0}} \mathrm{N}\left(Z|0, \left(C^{-1} - C^{-1}\Psi\left(\Psi^\top C^{-1}\Psi\right)^{-1}\Psi^\top C^{-1}\right)^{-1}\right) \\ < & \infty\end{aligned}$$

namely the bottom part of the fraction of the posterior of $\beta|Z,\varphi$ is bounded (finite), which implies a proper posterior. It is

$$\Pr(\beta|Z,\varphi) \propto \exp\left(-\frac{1}{2}(\beta - b_n)^\top B_n^{-1}(\beta - b_n)\right)$$

$$\lim_{\substack{B^{-1}\to 0\\ b\to 0}} \exp\left(-\frac{1}{2}\left(\beta-b_n\right)^{\top} B_n^{-1}\left(\beta-b_n\right)\right)=$$

$$\exp\left(-\frac{1}{2}\left(\beta-\left(\Psi^{\top}C^{-1}\Psi\right)^{-1}\Psi^{\top}C^{-1}Z\right)^{\top}\left(\Psi^{\top}C^{-1}\Psi\right)\left(\beta-\left(\Psi^{\top}C^{-1}\Psi\right)^{-1}\Psi^{\top}C^{-1}Z\right)\right)$$

hence the limiting case is

$$\beta|Z,\varphi \overset{\text{approx}}{\sim} \mathrm{N}\left(\left(\Psi^{\top}C^{-1}\Psi\right)^{-1}\Psi^{\top}C^{-1}Z,\ \left(\Psi^{\top}C^{-1}\Psi\right)^{-1}\right)$$

The predictive process becomes

$$Z\left(\cdot\right)|Z,\varphi \overset{\text{approx}}{\sim} \mathrm{GP}\left(\mu_3\left(\cdot|\varphi\right),c_3\left(\cdot,\cdot|\varphi\right)\right)$$

$$\begin{aligned}
\mu_3\left(s|\varphi\right) &= \left[\left(\psi\left(s\right)-C\left(s\right)C^{-1}\right)\left(\Psi^{\top}C^{-1}\Psi\right)^{-1}\Psi^{\top}+C\left(s\right)\right]C^{-1}Z\\
c_3\left(s,s'|\varphi\right) &= \left(\psi\left(s\right)-C\left(s\right)C^{-1}\right)\left(\Psi^{\top}C^{-1}\Psi\right)^{-1}\left(\psi\left(s\right)-C\left(s\right)C^{-1}\right)^{\top}\\
&\quad + c\left(s,s'|\varphi\right)+\left(C\left(s\right)\right)^{\top}C^{-1}C\left(s\right)
\end{aligned}$$

Exercise 2. $(\star)$ Show that the extension variance $\sigma_E^2\left(v,V\right)$ of a small volume v to a larger volume V is obtained by

$$\sigma_E^2\left(v,V\right)=2\bar{\gamma}\left(v,V\right)-\bar{\gamma}\left(v,v\right)-\bar{\gamma}\left(V,V\right)$$

where

$$\bar{\gamma}\left(v,V\right)=\frac{1}{\left|v\right|\left|V\right|}\int_{s\in v}\int_{s\in V}\gamma\left(s-s'\right)\mathrm{d}s\mathrm{d}s'$$

Solution. Essentially I need to show that that

$$\begin{aligned}
\mathrm{Var}\left(Z\left(A\right)-Z\left(B\right)\right)=&\frac{1}{\left|A\right|\left|B\right|}\int_{x\in A}\int_{y\in B}\gamma\left(x-y\right)\mathrm{d}x\mathrm{d}y\\
&-\frac{1}{\left|A\right|\left|A\right|}\int_{x\in A}\int_{y\in A}\gamma\left(x-y\right)\mathrm{d}x\mathrm{d}y\\
&-\frac{1}{\left|B\right|\left|B\right|}\int_{x\in B}\int_{y\in B}\gamma\left(x-y\right)\mathrm{d}x\mathrm{d}y
\end{aligned}$$

where I use A,B instead of v,V and x,y instead of s,s' for clarity on notation.

It is

$$\begin{aligned}
\mathrm{Var}\left(Z\left(A\right)-Z\left(B\right)\right) =&\mathrm{Cov}\left(Z\left(A\right)-Z\left(B\right),Z\left(A\right)-Z\left(B\right)\right)\\
=&\mathrm{Cov}\left(Z\left(A\right),Z\left(A\right)\right)+\mathrm{Cov}\left(Z\left(B\right),Z\left(B\right)\right)-2\mathrm{Cov}\left(Z\left(A\right),Z\left(B\right)\right)\\
=&\frac{1}{\left|A\right|\left|A\right|}\int_{x\in A}\int_{y\in A}\mathrm{Cov}\left(Z\left(x\right),Z\left(y\right)\right)\mathrm{d}x\mathrm{d}y\\
&+\frac{1}{\left|B\right|\left|B\right|}\int_{x\in B}\int_{y\in B}\mathrm{Cov}\left(Z\left(x\right),Z\left(y\right)\right)\mathrm{d}x\mathrm{d}y\\
&-2\frac{1}{\left|A\right|\left|B\right|}\int_{x\in A}\int_{y\in B}\mathrm{Cov}\left(Z\left(x\right),Z\left(y\right)\right)\mathrm{d}x\mathrm{d}y
\end{aligned}$$

OK, now I need to write all these Cov as γ ; I know that

$$\begin{aligned}
\gamma\left(x-y\right) &= \frac{1}{2}\mathrm{Var}\left(Z\left(x\right)-Z\left(y\right)\right)\\
&= \frac{1}{2}\mathrm{Var}\left(Z\left(x\right)\right)+\frac{1}{2}\mathrm{Var}\left(Z\left(y\right)\right)-\mathrm{Cov}\left(Z\left(x\right),Z\left(y\right)\right)
\end{aligned}$$

that is

$$\mathrm{Cov}\left(Z\left(x\right),Z\left(y\right)\right)=\frac{1}{2}\mathrm{Var}\left(Z\left(x\right)\right)+\frac{1}{2}\mathrm{Var}\left(Z\left(y\right)\right)-\gamma\left(x-y\right)$$

Now I 'll gonna put all these in the quantity of interest, one by one

$$\begin{aligned}
\frac{1}{\left|A\right|\left|A\right|}\int_{x\in A}\int_{y\in A}\mathrm{Cov}\left(Z\left(x\right),Z\left(y\right)\right)\mathrm{d}x\mathrm{d}y =&\frac{1}{\left|A\right|\left|A\right|}\int_{x\in A}\int_{y\in A}\frac{1}{2}\mathrm{Var}\left(Z\left(x\right)\right)\mathrm{d}x\mathrm{d}y\\
&+\frac{1}{\left|A\right|\left|A\right|}\int_{x\in A}\int_{y\in A}\frac{1}{2}\mathrm{Var}\left(Z\left(y\right)\right)\mathrm{d}x\mathrm{d}y\\
&-\frac{1}{\left|A\right|\left|A\right|}\int_{x\in A}\int_{y\in A}\gamma\left(x-y\right)\mathrm{d}x\mathrm{d}y\\
=&\frac{1}{\left|A\right|}\int_{x\in A}\mathrm{Var}\left(Z\left(x\right)\right)\mathrm{d}x\\
&-\frac{1}{\left|A\right|^2}\int_{x\in A}\int_{y\in A}\gamma\left(x-y\right)\mathrm{d}x\mathrm{d}y
\end{aligned}$$

and by symmetry

$$\begin{aligned}
\frac{1}{\left|B\right|\left|B\right|}\int_{x\in B}\int_{y\in B}\mathrm{Cov}\left(Z\left(x\right),Z\left(y\right)\right)\mathrm{d}x\mathrm{d}y =&\frac{1}{\left|B\right|}\int_{x\in B}\mathrm{Var}\left(Z\left(x\right)\right)\mathrm{d}x\\
&-\frac{1}{\left|B\right|^2}\int_{x\in B}\int_{y\in B}\gamma\left(x-y\right)\mathrm{d}x\mathrm{d}y
\end{aligned}$$

and finally,

$$
\begin{aligned}
\frac{1}{|A|\,|B|}\int_{x\in A}\int_{y\in B}\operatorname{Cov}\left(Z\left(x\right),Z\left(y\right)\right)\mathrm{d}x\mathrm{d}y =&\frac{1}{|A|\,|B|}\int_{x\in A}\int_{y\in B}\frac{1}{2}\operatorname{Var}\left(Z\left(x\right)\right)\mathrm{d}x\mathrm{d}y\\
&+\frac{1}{|A|\,|B|}\int_{x\in A}\int_{y\in B}\frac{1}{2}\operatorname{Var}\left(Z\left(y\right)\right)\mathrm{d}x\mathrm{d}y\\
&-\frac{1}{|A|\,|B|}\int_{x\in A}\int_{y\in B}\gamma\left(x-y\right)\mathrm{d}x\mathrm{d}y\\
=&\frac{1}{2}\frac{1}{|A|}\int_{x\in A}\operatorname{Var}\left(Z\left(x\right)\right)\mathrm{d}x\\
&+\frac{1}{2}\frac{1}{|B|}\int_{x\in B}\operatorname{Var}\left(Z\left(x\right)\right)\mathrm{d}x\\
&-\frac{1}{|A|\,|B|}\int_{x\in A}\int_{y\in B}\gamma\left(x-y\right)\mathrm{d}x\mathrm{d}y
\end{aligned}
$$

Putting all these together, we get the result.
