How to update a process so that it obeys the marginal defined by another external processes

Reed Essick
essick@cita.utoronto.ca

April 24, 2023

Derivation

Assume we have an existing Gaussian process that defines a measure for two vectors (f_a and f_b with N_a and N_b elements, repectively) and can be written as

$$p_{\mathrm{O}}(f_a, f_b) \sim \mathcal{N}\left([\mu_a, \mu_b], \begin{bmatrix} C_{aa} & C_{ab} \\ C_{ba} & C_{bb} \end{bmatrix}\right) \tag{1}$$

with mean vectors μ_a, μ_b and covariance matrix C decomposed into C_{aa} ($N_a \times N_a$ elements), C_{ab} ($N_a \times N_b$), C_{ba} ($N_b \times N_a$), and C_{bb} ($N_b \times N_b$). We note that

$$C_{aa} = C_{aa}^{\mathrm{T}} \tag{2}$$
$$C_{ab} = C_{ba}^{\mathrm{T}} \tag{3}$$
$$C_{bb} = C_{bb}^{\mathrm{T}} \tag{4}$$

We wish to update this process so that the marginal distribution for f_b follows another process, namely

$$p_{\mathrm{E}}(f_b) = \mathcal{N}\left(y_b, \Sigma_{bb}\right) \tag{5}$$

while maintaining the rest of the covariance structure encoded in C. We do this by constructing a new process

$$p_{\mathrm{N}}(f_a, f_b) = p_{\mathrm{O}}(f_a|f_b)p_{\mathrm{E}}(f_b) \tag{6}$$

where $p_{\mathrm{O}}(f_a|f_b)$ can be derived from $p_{\mathrm{O}}(f_a, f_b)$ as

$$p_{\mathrm{O}}(f_a|f_b) = \mathcal{N}\left(\mu_a + C_{ab}C_{bb}^{-1}(f_b - \mu_b),\ C_{aa} - C_{ab}C_{bb}^{-1}C_{ba}\right) \tag{7}$$

Expanding the contractions, grouping like terms, and dropping those that do not depend on either f_a or f_b, we obtain

$$\begin{aligned} -2\ln p_{\mathrm{N}}(f_a, f_b) = {} & f_a^{\mathrm{T}}\left(C_{aa} - C_{ab}C_{bb}^{-1}C_{ba}\right)^{-1} f_a \\ & - 2f_a^{\mathrm{T}}\left(C_{aa} - C_{ab}C_{bb}^{-1}C_{ba}\right)^{-1}\left(\mu_a - C_{ab}C_{bb}^{-1}\mu_b\right) \\ & - 2f_a^{\mathrm{T}}\left[\left(C_{aa} - C_{ab}C_{bb}^{-1}C_{ba}\right)^{-1} C_{ab}C_{bb}^{-1}\right] f_b \\ & + 2f_b^{\mathrm{T}}\left[C_{bb}^{-1}C_{ba}^{\mathrm{T}}\left(C_{aa} - C_{ab}C_{bb}^{-1}C_{ba}\right)^{-1}\left(\mu_a - C_{ab}C_{bb}^{-1}\mu_b\right) - \Sigma_{bb}^{-1}y_b\right] \\ & + f_b^{\mathrm{T}}\left[C_{bb}^{-1}C_{ba}\left(C_{aa} - C_{ab}C_{bb}^{-1}C_{ba}\right)^{-1} C_{ab}C_{bb}^{-1} + \Sigma_{bb}^{-1}\right] f_b \end{aligned} \tag{8}$$

This is still Gaussian in both f_a and f_b, and we obtain direct relations for the new mean vectors and (inverse) covariance defined by

$$p_{\mathrm{N}}(f_a, f_b) = \mathcal{N}\left([m_a, m_b], \begin{bmatrix}\Gamma_{aa} & \Gamma_{ab} \\ \Gamma_{ba} & \Gamma_{bb}\end{bmatrix}^{-1}\right) \tag{9}$$

as follows:

$$\Gamma_{aa} = \left(C_{aa} - C_{ab}C_{bb}^{-1}C_{ba}\right)^{-1} \tag{10}$$

$$\Gamma_{ab} = -\left(C_{aa} - C_{ab}C_{bb}^{-1}C_{ba}\right)^{-1} C_{ab}C_{bb}^{-1} \tag{11}$$

$$\Gamma_{bb} = C_{bb}^{-1}C_{ba}\left(C_{aa} - C_{ab}C_{bb}^{-1}C_{ba}\right)^{-1} C_{ab}C_{bb}^{-1} + \Sigma_{bb}^{-1} \tag{12}$$

and

$$\Gamma_{aa}m_a + \Gamma_{ab}m_b = \left(C_{aa} - C_{ab}C_{bb}^{-1}C_{ba}\right)^{-1}\left(\mu_a C_{ab}C_{bb}^{-1}\mu_b\right) \tag{13}$$

$$\Gamma_{ba}m_a + \Gamma_{bb}m_b = \Sigma_{bb}^{-1}y_b - C_{bb}^{-1}C_{ba}\left(C_{aa} - C_{ab}C_{bb}^{-1}C_{ba}\right)^{-1}\left(\mu_a - C_{ab}C_{bb}^{-1}\mu_b\right) \tag{14}$$

which simply to

$$m_a = \mu_a + C_{ab}C_{bb}^{-1}\left(y_b - \mu_b\right) \tag{15}$$

$$m_b = y_b \tag{16}$$

Finally, we can solve for

$$\gamma = \Gamma^{-1} = \begin{bmatrix}\gamma_{aa} & \gamma_{ab} \\ \gamma_{ba} & \gamma_{bb}\end{bmatrix} \tag{17}$$

by recognizing that

$$\begin{bmatrix}\gamma_{aa} & \gamma_{ab} \\ \gamma_{ba} & \gamma_{bb}\end{bmatrix}\begin{bmatrix}\Gamma_{aa} & \Gamma_{ab} \\ \Gamma_{ba} & \Gamma_{bb}\end{bmatrix} = \mathbb{1} \tag{18}$$

and therefore

$$\gamma_{aa}\Gamma_{aa} + \gamma_{ab}\Gamma_{ba} = \mathbb{1} \tag{19}$$

$$\gamma_{aa}\Gamma_{ab} + \gamma_{ab}\Gamma_{bb} = \mathbb{0} \tag{20}$$

$$\gamma_{ba}\Gamma_{aa} + \gamma_{bb}\Gamma_{ba} = \mathbb{0} \tag{21}$$

$$\gamma_{ba}\Gamma_{ab} + \gamma_{bb}\Gamma_{bb} = \mathbb{1} \tag{22}$$

Further simplification yields

$$\gamma_{aa} = \left[\Gamma_{aa} - \Gamma_{ab}\Gamma_{bb}^{-1}\Gamma_{ba}\right]^{-1} \tag{23}$$

$$\gamma_{ab} = C_{ab}C_{bb}^{-1}\Sigma_{bb} \tag{24}$$

$$\gamma_{ba} = \Sigma_{bb}C_{bb}^{-1}C_{ba} \tag{25}$$

$$\gamma_{bb} = \Sigma_{bb} \tag{26}$$

where we've left γ_{aa} in terms of Γ because of the length of the expression but have subsituted and simplified the rest of the terms. Note that the marginal distribution $p_{\mathrm{N}}(f_b) = \mathcal{N}(y_b, \Sigma_{bb}) = p_{\mathrm{E}}(f_b)$, as desired.

Modifications for numerical stability

In general, we find that m_a and γ_{aa} can suffer from issues associated with numerical stability. This is because they involve the inversion of (possibly) high-dimensional matrices that may be ill-conditioned. While the preceding is exact, we therefore implement two additional approximations to help better control the calculations.

Damping C_{ab}, C_{ba}, and C_{bb} to make them easier to invert

One issue we have found is that strong correlations in C_{bb} can make numerical inversion difficult. Given that we wish to replace C_{bb} with Σ_{bb} anyway, and really only wish there to be a relatively smooth transition between f_b and f_a, we modify C_{ab}, C_{ba}, and C_{bb} in order to damp the off-diagonal elements (and therefore make them easier to invert).

Specifically, we define a squared-exponential damping term

$$D(x_i, x_j) = \exp\left(-\frac{(x_i - x_j)^2}{l^2}\right) \tag{27}$$

and a white noise contribution that modify C so that

$$\begin{aligned}
(C_{ab})_{ij} &\to (C_{ab})_{ij} D(x_i, x_j) && (28)\\
(C_{bb})_{ij} &\to (C_{bb})_{ij} D(x_i, x_j) + \sigma_{\mathrm{W}}^2 \delta_{ij} && (29)\\
& && (30)
\end{aligned}$$

We then use these modified C_{ab} and C_{bb} within the expressions in the previous section.

This modifies the original process, but as long as l is relatively large and σ_{W} is relatively small, the modifications will be minor over the transition between f_b and f_a. Empirically, we find that

$$\begin{aligned}
l &= 5.0 && (31)\\
\sigma_{\mathrm{W}} &= 0.01 && (32)
\end{aligned}$$

work well when updating our our model-agnostic priors.

Approximation for γ_{aa} when Σ_{bb} is small

Finally, we note that it will often be the case that Σ_{bb} will be much smaller than C_{bb} with respect to an appropriate matrix norm. That is, we wish to update a process to restrict the marginals of certain covariates to be more tightly constrained than they otherwise would be.

By repeated use of the approximation

$$(A + X)^{-1} \approx A^{-1} - A^{-1} X A^{-1} \tag{33}$$

we can see that this limit corresponds to

$$\Gamma_{bb}^{-1} \approx \Sigma_{bb} - \Sigma_{bb} C_{bb}^{-1} C_{ba} \left(C_{aa} - C_{ab} C_{bb}^{-1} C_{ba}\right)^{-1} C_{ba} C_{bb}^{-1} \Sigma_{bb} \tag{34}$$

and (retaining terms linear in Σ_{bb})

$$\begin{aligned}
\gamma_{aa} &\approx C_{aa} - C_{ab} C_{bb}^{-1} C_{ba} + C_{ab} C_{bb}^{-1} \Gamma_{bb}^{-1} C_{bb}^{-1} C_{ba} \\
&\approx C_{aa} - C_{ab} C_{bb}^{-1} C_{ba} + C_{ab} C_{bb}^{-1} \Sigma_{bb} C_{bb}^{-1} C_{ba}
\end{aligned} \tag{35}$$

If we examine only terms up to linear order in Σ_{bb}, we obtain the relatively intuitive expression

$$\gamma_{aa} \approx C_{aa} - C_{ab} C_{bb}^{-1} (C_{bb} - \Sigma_{bb}) C_{bb}^{-1} C_{ba} \tag{36}$$

This makes sense in two limiting cases

- $\Sigma_{bb} = 0$: we know f_b exactly and obtain the standard expression for the covariance for $f_a|f_b$
- $\Sigma_{bb} = C_{bb}$: we do not update the original process, and as such we obtian $\gamma_{aa} = C_{aa}$.

Finally, if we offer one more interpretation of this expression. If we considered the standard expression for $f_a|f_b$ with some covariance for f_b, say $\mathcal{C}_{bb}$, we would obtain

$$\gamma_{aa} = C_{aa} - C_{ab}\mathcal{C}_{bb}^{-1}C_{ba} \tag{37}$$

and therefore, by matching this to our approximation, we see that

$$\begin{aligned} \mathcal{C}_{bb} &= C_{bb}\left(C_{bb} - \Sigma_{bb}\right)^{-1} C_{bb} \\ &\approx C_{bb}\left(C_{bb}^{-1} + C_{bb}^{-1}\Sigma_{bb}C_{bb}^{-1}\right) C_{bb} = C_{bb} + \Sigma_{bb} \end{aligned} \tag{38}$$

In this limit, then, we can interpret updating the marignal distribution as equivalent to the standard procedure of conditioning the process on a noisey observation of f_a with mean y_b and covariance Σ_{bb}. Historically, this is what was actually done, and we now see why it provided a decent approximation. However, it also caused issues with numerical stability that sometimes (often?) resulted in different marginal distributions for f_b than we desired (i.e., $\gamma_{bb} \neq \Sigma_{bb}$).