

Notes on spike inference

Giovanni Diana

April 28, 2021

1 Probability distributions

- **Multivariate normal distribution:**

$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \equiv \frac{1}{(2\pi \det \boldsymbol{\Sigma})^{d/2}} \exp \left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right] \quad (1)$$

- **Beta distribution**

$$\text{beta}(x; \alpha, \beta) \equiv \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\text{B}(\alpha, \beta)} \quad (2)$$

where

$$\text{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \quad (3)$$

and Γ is the Gamma function.

- **Inverse Gamma distribution**

$$\text{InvGamma}(x; \alpha, \beta) \equiv \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp \left(-\frac{\beta}{x} \right) \quad (4)$$

2 The model

$$\begin{cases} c_t = \gamma c_{t-1} + A s_t, & \text{for } t = T, T-1, \dots, 2 \\ c_1 = c_0 + A s_1 \end{cases} \quad (5)$$

$$y_t \sim \mathcal{N}(c_t + b, \sigma^2) \quad (6)$$

$$s_t \sim \text{Bernoulli}(q) \quad (7)$$

For the prior distributions we have

$$\boldsymbol{\theta} \equiv \{A, b, c_0\} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (8)$$

$$\sigma^2 \sim \text{InvGamma}(\alpha_\sigma, \beta_\sigma) \quad (9)$$

$$q \sim \text{beta}(\alpha_q, \beta_q) \quad (10)$$

note that in Pnevmatikakis 2013 $\alpha_\sigma = 1$ and $\beta_\sigma = 0.1$.

Remark 1. Note that the empirical Bayes approach described in the arXiv version of the paper is wrong. The marginal likelihood $p(\mathbf{s}|\alpha, \beta)$ reads

$$p(\mathbf{s}|\alpha, \beta) = \frac{\alpha^{[n]} \beta^{[T-n]}}{(\alpha + \beta)^{[T]}} \quad (11)$$

where n is the number of spikes in \mathbf{s} and we used the ascending power notation

$$x^{[n]} = x \cdot (x+1) \cdots (x+n-1) \quad (12)$$

The mistake in the arXiv version of Pnevmatikakis 2013 is to replace ascending powers with regular powers, leading to the wrong conclusion that the marginal likelihood $p(\mathbf{s}|\alpha, \beta)$ only depends on the ratio α/β . This was fixed in the conference version.

Given the spike train \mathbf{s} and the parameters γ and A , we can obtain a close-form expression for the calcium at any time c_t . Note that at $t = 2$ we have

$$c_2 = \gamma c_1 + A s_1 = \gamma c_0 + A(\gamma s_1 + s_2) \quad (13)$$

Multiplying by γ and adding $A s_3$ leads to

$$c_3 = \gamma^2 c_0 + A(\gamma^2 s_1 + \gamma s_2 + \gamma s_3) \quad (14)$$

which generalizes to the equation

$$c_t = \gamma^{t-1} c_0 + A \sum_{k=1}^t \gamma^{t-k} s_k \quad (15)$$

By defining the matrix

$$\tilde{G}_{tk} = \begin{cases} \gamma^{t-k} & k \leq t \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

and the vector $\mathbf{v} \equiv \{1, \gamma, \gamma^2, \dots, \gamma^{T-1}\}$ we can rewrite the calcium in matrix form

$$\mathbf{c} = c_0 \mathbf{v} + A \tilde{\mathbf{G}} \cdot \mathbf{s} \quad (17)$$

3 Likelihood, joint probability and full conditionals

3.1 Likelihood

The data likelihood can be expressed as

$$P(\mathbf{y}|\mathbf{s}, \boldsymbol{\theta}, \sigma) = (2\pi\sigma^2)^{T/2} \cdot \exp \left[-\frac{1}{2\sigma^2} (\mathbf{y} - b\mathbf{1} - c_0 \mathbf{v} - A \tilde{\mathbf{G}} \cdot \mathbf{s})^2 \right] \quad (18)$$

where $\mathbf{1} \equiv \{1, 1, \dots, 1\}$.

3.2 Joint probability

The joint probability (probability of everything) is

$$P(\mathbf{y}, \mathbf{s}, \boldsymbol{\theta}, \sigma^2) = \text{beta}(q; \alpha_q, \beta_q) \cdot \text{InvGamma}(\sigma^2; \alpha_\sigma, \beta_\sigma) \cdot \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ q^{N_s} (1-q)^{T-N_s} \cdot (2\pi\sigma^2)^{T/2} \cdot \exp \left[-\frac{1}{2\sigma^2} (\mathbf{y} - b\mathbf{1} - c_0 \mathbf{v} - A \tilde{\mathbf{G}} \cdot \mathbf{s})^2 \right] \quad (19)$$

where N_s is the number of spikes in the vector \mathbf{s} .

3.3 Full conditionals

Let's consider the probability of the spike vector \mathbf{s} conditional to everything else. The general trick to calculate full conditionals is to look at the joint probability and only keep factors that depend on the argument of the conditional. We can deal later with the normalization factor if we need to (alternatively we can use the Metropolis rejection method to sample from unnormalized distributions). In the case of \mathbf{s} , by expanding the square in the exponential and keeping only terms that depend on \mathbf{s} we obtain

$$P(\mathbf{s}|\mathbf{y}, \boldsymbol{\theta}, \sigma^2) \propto \exp \left[-\frac{A^2}{2\sigma^2} \mathbf{s}^T \tilde{\mathbf{G}}^T \tilde{\mathbf{G}} \cdot \mathbf{s} + \frac{A}{\sigma^2} \mathbf{s}^T \tilde{\mathbf{G}}^T \cdot \tilde{\mathbf{y}} + N_s \log \frac{q}{1-q} \right] \quad (20)$$

where $\tilde{\mathbf{y}} = \mathbf{y} - b\mathbf{1} - c_0 \mathbf{v}$.

Note that for this type of distribution the normalization factor cannot be obtained analytically. To sample the spikes we will need to introduce a Metropolis-Hastings acceptance rule. Let's derive first the remaining conditionals. The full conditional of the model parameters $\boldsymbol{\theta}$ is obtained by keeping factors from the joint probability (Eq. (19)) which only depend on $\boldsymbol{\theta}$:

$$P(\boldsymbol{\theta}|\mathbf{s}, \sigma^2, \mathbf{y}) \propto \exp \left[-\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})^T (\boldsymbol{\Sigma})^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{S}\boldsymbol{\theta})^2 \right] \quad (21)$$

where we introduced the $T \times 3$ matrix \mathbf{S}

$$\mathbf{S} = (A\tilde{\mathbf{G}} \cdot \mathbf{s}, \mathbf{1}, \mathbf{v}). \quad (22)$$

This notation exploits the linearity of the residual $\mathbf{y} - b\mathbf{1} - c_0\mathbf{v} - A\tilde{\mathbf{G}} \cdot \mathbf{s}$ with respect to A , b and c_0 . The quadratic form in the exponential of Eq. (21) implies that the full conditional will be a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}', \boldsymbol{\Lambda})$. We can determine the covariance matrix $\boldsymbol{\Lambda}$ from the quadratic term in $\boldsymbol{\theta}$

$$\boldsymbol{\Lambda}^{-1} = \boldsymbol{\Sigma}^{-1} + \frac{1}{\sigma^2} \mathbf{S}^T \mathbf{S} \quad (23)$$

The mean vector $\boldsymbol{\mu}'$ can be obtained by matching the linear terms in the exponent of the conditional $(\boldsymbol{\theta}^T(\boldsymbol{\Lambda})^{-1}\boldsymbol{\mu}')/2$ with the linear in Eq. (21) proportional to $\boldsymbol{\theta}^T$

$$\frac{1}{2} \boldsymbol{\theta}^T (\boldsymbol{\Lambda})^{-1} \boldsymbol{\mu}' = \frac{1}{2} \boldsymbol{\theta}^T (\boldsymbol{\Sigma})^{-1} \boldsymbol{\mu} + \frac{1}{2\sigma^2} \boldsymbol{\theta}^T \mathbf{S}^T \mathbf{y} \quad (24)$$

therefore we have

$$\boldsymbol{\mu}' = \boldsymbol{\Lambda} \left(\boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \frac{1}{\sigma^2} \mathbf{S}^T \mathbf{y} \right) \quad (25)$$

Remark 2. Note that in the conference version of Pnevmatikakis 2013 there is a typo in the mean of this conditional distribution. The typo is not present in the arXiv version.

Let us now obtain the full conditional on the noise variance σ^2 . By keeping terms from the joint distribution only dependent on σ^2 (residual from the exponential of the likelihood and the prior)

$$P(\sigma^2 | \mathbf{s}, \mathbf{y}, \boldsymbol{\theta}) \propto (\sigma^2)^{-\alpha_\sigma - 1} e^{-\beta_\sigma / \sigma^2} \cdot (2\pi\sigma^2)^{-T/2} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{S}\boldsymbol{\theta})^2 \right] \quad (26)$$

which is an InvGamma with parameters

$$\alpha'_\sigma = \alpha_\sigma + T/2 \quad (27)$$

$$\beta'_\sigma = \beta_\sigma + \frac{1}{2} (\mathbf{y} - \mathbf{S}\boldsymbol{\theta})^2 \quad (28)$$

Finally by keeping terms from the joint that only depend on q we obtain the conditional

$$P(q | \dots) = \text{beta}(q; \alpha'_q, \beta'_q) \quad (29)$$

with

$$\alpha'_q = \alpha_q + N_{\mathbf{s}} \quad (30)$$

$$\beta'_q = \beta_q + (T - N_{\mathbf{s}}) \quad (31)$$

In practice, when T and $N_{\mathbf{s}}$ are large, we can simply set q to its maximum-likelihood value $q_{\text{MLE}} \equiv N_{\mathbf{s}}/T$.

4 Sampling

According to the above construction, it is easy to sample model parameters as the full conditional distributions are known distributions. In order to sample the binary vector \mathbf{s} representing the presence of a spike in each time bin we need to use the Metropolis-Hastings acceptance rule, which corresponds to flip one element of \mathbf{s} at a time and accept the move with probability

$$\alpha_{MH} = \min \left(1, \frac{\text{Prob}(\text{flip})}{\text{Prob}(\text{orig})} \right) \quad (32)$$

for a given element s_t we can compute the log of the ratio in Eq. (32) using the full conditional derived in Eq. (20)

$$\log \frac{P(\mathbf{s}_{\text{flip}} | \dots)}{P(\mathbf{s}_{\text{orig}} | \dots)} \quad (33)$$

the first contribution to the log ratio is of the form

$$\mathbf{s}^T \mathbf{W} \mathbf{s}|_{\text{flip}} - \mathbf{s}^T \mathbf{W} \mathbf{s}|_{\text{orig}} = \sum_{i,j} W_{ij} s_i s_j|_{\text{flip}} - \sum_{i,j} W_{ij} s_i s_j|_{\text{orig}} \quad (34)$$

Where $\mathbf{W} = A^2 \tilde{\mathbf{G}}^T \tilde{\mathbf{G}} / \sigma^2$ is a symmetric matrix. If we flip the value of σ_t to σ'_t the terms contributing to the difference between flipped and original correspond to instances of $i = t$ or $j = t$ (or both). We can simplify the difference as

$$(1 - 2s_t) \left[\underbrace{\sum_j (W_{tj} s_j + W_{jt} s_j)}_{j \neq t} - 2W_{tt} s_t + \underbrace{W_{tt}}_{j=t} \right] = (1 - 2s_t) \left[\sum_j 2W_{tj} s_j + (1 - 2s_t) W_{tt} \right] \quad (35)$$

Then we have a second contribution of the form

$$\mathbf{s}^T \cdot \mathbf{z}|_{\text{flip}} - \mathbf{s}^T \cdot \mathbf{z}|_{\text{orig}} = (1 - 2s_t) z_t \quad (36)$$

By combining the two types of contributions we obtain

$$\log \frac{P(\mathbf{s}_{\text{flip}}|\dots)}{P(\mathbf{s}_{\text{orig}}|\dots)} = (1 - 2s_t) \left(- \sum_j W_{tj} s_j - (1 - 2s_t) W_{tt} / 2 + \frac{A}{\sigma^2} \sum_j \tilde{G}_{jt} \tilde{y}_j + \log \frac{q}{1 - q} \right) \quad (37)$$