1 Variational Bayesian inference

In the true generative model of olfactory scenes we assumed that the spike counts were Poisson,

$$P(\mathbf{r}|\mathbf{c}) = \prod_i \frac{\left(r_0 \Delta t + \Delta t \sum_j w_{ij} c_j\right)^{r_i}}{r_i!} e^{-\left(r_0 \Delta t + \Delta t \sum_j w_{ij} c_j\right)} \tag{1.1a}$$

and we used a "spike and slab" prior on the concentrations, and a Bernoulli prior on s_j, the variable that indicates presence or absence of odor j,

$$P(\mathbf{c}|\mathbf{s}) = \prod_j (1 - s_j)\delta(c_j) + s_j \Gamma(c_j|\alpha_1, \beta_1) \tag{1.1b}$$

$$P(\mathbf{s}) = \prod_j \pi^{s_j}(1 - \pi)^{1-s_j} \tag{1.1c}$$

where $\delta(c)$ is the Dirac delta-function and $\Gamma(c|\alpha, \beta)$ is the Gamma distribution,

$$\Gamma(c|\alpha, \beta) = \frac{\beta^\alpha c^{\alpha-1} e^{-\beta c}}{\Gamma(\alpha)}. \tag{1.2}$$

Here $\Gamma(\alpha)$ is the ordinary Gamma function: $\Gamma(\alpha) = \int_0^\infty dx\, x^{\alpha-1} e^{-x}$.

Because of the delta-function in the prior, performing efficient variational inference in our model is, as far as we know, difficult. Therefore, we approximate it with a Gamma distribution, $\Gamma(c_j|\alpha, \beta)$.

$$P_{var}(\mathbf{r}|\mathbf{c}) = \prod_i \frac{\left(\Delta t \sum_j w_{ij} c_j\right)^{r_i}}{r_i!} e^{-\Delta t \left(\sum_j w_{ij} c_j\right)} \tag{1.3a}$$

$$P_{var}(\mathbf{c}) = \prod_j \Gamma(c_j|\alpha, \beta). \tag{1.3b}$$

Collecting the terms in Eq. (1.3a), we see that the posterior distribution over $\mathbf{c}$ and $\mathbf{s}$ is given by

$$P(\mathbf{c}|\mathbf{r}) \propto \prod_i \frac{\left(\Delta t \sum_j w_{ij} c_j\right)^{r_i}}{r_i!} e^{-\left(\Delta t \sum_j w_{ij} c_j\right)} \times \prod_j \Gamma(c_j|\alpha, \beta). \tag{1.4}$$

Variational inference with this posterior is hard, primarily because the likelihood consists of products over sums. We can, however, turn those products over sums into sums over products by using the multinomial theorem,

$$\left(\Delta t \sum_j w_{ij} c_j\right)^{r_i} = \sum_{N_{ij}} \Delta\left(r_i - \sum_j N_{ij}\right) r_i! \prod_{j=0} \frac{(\Delta t w_{ij} c_j)^{N_{ij}}}{N_{ij}!} \tag{1.5}$$

where the sum over N_{ij} is shorthand for a set of sums in which N_{i1}, N_{i2}, ... all run from 0 to r_i, and Δ is the Kronecker delta: $\Delta(n) = 1$ if $n = 0$ and 0 otherwise. The posterior distribution can now be written as

$$P(\mathbf{c}|\mathbf{r}) = \sum_{\mathbf{N}} P(\mathbf{N}, \mathbf{c}|\mathbf{r}) \tag{1.6}$$

where, inserting Eq. (1.3b) into (1.4), $P(\mathbf{N}, \mathbf{c}|\mathbf{r})$ is given by

$$\begin{aligned} P(\mathbf{N}, \mathbf{c}|\mathbf{r}) \propto & \prod_i \Delta\Big(r_i - \sum_j N_{ij}\Big) \prod_j \frac{(\Delta t w_{ij} c_j)^{N_{ij}} e^{-\Delta t w_{ij} c_j}}{N_{ij}!} \\ & \times \prod_j \Gamma(c_j|\alpha, \beta) \,. \end{aligned} \tag{1.7}$$

The variational approach we use approximates the augmented posterior distribution, $P(\mathbf{N}, \mathbf{c}|\mathbf{r})$, rather than the original one, $P(\mathbf{c}|\mathbf{r})$. We use a factorized variational distribution of the form

$$Q(\mathbf{N}, \mathbf{c}|\mathbf{r}) = Q(\mathbf{N}|\mathbf{r}) Q(\mathbf{c}|\mathbf{r}) \tag{1.8}$$

where we are using the notation that a probability distribution is labeled by its argument. This can in principle produce confusion, but it won't for this problem.

Our goal is to choose $Q(\mathbf{N}, \mathbf{c}|\mathbf{r})$ so that it minimizes the KL distance between $Q(\mathbf{N}, \mathbf{c}|\mathbf{r})$ and $P(\mathbf{N}, \mathbf{c}|\mathbf{r})$. To see what this implies, we explicitly minimize the KL distance with respect to $Q(\mathbf{N}|\mathbf{r})$. To do that we differentiate with respect to $Q(\mathbf{N}|\mathbf{r})$,

$$\begin{aligned} & \frac{d}{dQ(\mathbf{N}|\mathbf{r})} \sum_{\mathbf{N},\mathbf{s}} \int d\mathbf{c}\, Q(\mathbf{N}, \mathbf{c}|\mathbf{r}) \log \frac{Q(\mathbf{N}, \mathbf{c}|\mathbf{r})}{P(\mathbf{N}, \mathbf{c}|\mathbf{r})} \\ & \qquad = 1 + \log Q(\mathbf{N}|\mathbf{r}) - \sum_{\mathbf{s}} \int d\mathbf{c}\, Q(\mathbf{c}|\mathbf{r}) \log P(\mathbf{N}, \mathbf{c}|\mathbf{r}) \,, \end{aligned} \tag{1.9}$$

and set the right hand side to zero. This yields

$$\log Q(\mathbf{N}|\mathbf{r}) \sim \langle \log P(\mathbf{N}, \mathbf{c}|\mathbf{r}) \rangle_{Q(\mathbf{c}|\mathbf{r})} \tag{1.10a}$$

where "$\sim$" indicates equality up to constants. An essentially identical calculation yields

$$\log Q(\mathbf{c}|\mathbf{r}) \sim \langle \log P(\mathbf{N}, \mathbf{c}|\mathbf{r}) \rangle_{Q(\mathbf{N}|\mathbf{r})} \,. \tag{1.10b}$$

To proceed, then, we simply need to average $\log P(\mathbf{N}, \mathbf{c}|\mathbf{r})$ with respect to the variational distributions. We start by writing down an explicit expression for this quantity,

$$\log P(\mathbf{N}, \mathbf{c}|\mathbf{r}) \sim \sum_i \log \Delta\Big(r_i - \sum_j N_{ij}\Big) - \log N_{ij}! - \log[\beta^{-\alpha}\Gamma(\alpha)] + \sum_{ij} N_{ij} \log(\Delta t w_{ij} c_j) - \Delta t w_{ij} c_j + (\alpha - 1)\log c_j - \beta c_j \,.$$

Using Eq. (1.10), and performing averages over either $Q(\mathbf{N}|\mathbf{r})$ or $Q(\mathbf{c}|\mathbf{r})$ in Eq. (1.11a), whichever is appropriate, tells us that

$$\log Q(\mathbf{N}|\mathbf{r}) \sim \sum_i \log \Delta\Big(r_i - \sum_j N_{ij}\Big) + \sum_{ij} N_{ij} \langle \log(\Delta t w_{ij} c_j)\rangle_{Q(\mathbf{c}|\mathbf{r})} - \log N_{ij}! \tag{1.11a}$$

$$\log Q(\mathbf{c}|\mathbf{r}) \sim \sum_j \left[\big(\alpha + \langle N_{ij}\rangle_{Q(\mathbf{N}|\mathbf{r})} - 1\big)\log c_j - \left(\beta + \Delta t \sum_i w_{ij}\right) c_j\right] - \sum_j \log\left[\frac{\Gamma(\alpha)}{\beta^{\alpha}}\right] . \tag{1.11b}$$

Examining these expressions, we see that $Q(\mathbf{N}|\mathbf{r})$ is multinomial and $Q(\mathbf{c}|\mathbf{r})$ is a Gamma distribution,

$$Q(\mathbf{N}|\mathbf{r}) = \prod_i \Delta\Big(r_i - \sum_j N_{ij}\Big)\, r_i! \prod_j \frac{p_{ij}^{N_{ij}}}{N_{ij}!} \tag{1.12a}$$

$$Q(\mathbf{c}|\mathbf{r}) = \prod_j \Gamma(c_j|\alpha_j, \beta_j) \tag{1.12b}$$

where

$$p_{ij} = \frac{e^{\langle \log(w_{ij} c_j)\rangle_{Q(\mathbf{c}|\mathbf{r})}}}{\sum_j e^{\langle \log(w_{ij} c_j)\rangle_{Q(\mathbf{c}|\mathbf{r})}}} \tag{1.13a}$$

$$\alpha_j = \alpha + \sum_i r_i p_{ij} \tag{1.13b}$$

$$\beta_j = \beta + \Delta t \sum_i w_{ij} \,. \tag{1.13c}$$

Our one remainig task is to compute $\langle \log c_j\rangle_{Q(\mathbf{c}|\mathbf{r})}$. Using the fact that in general,

$$\langle \log c\rangle_{\Gamma(c|\alpha,\beta)} = \Psi(\alpha) - \log\beta \,, \tag{1.14}$$

Eq. (1.13a) becomes

$$p_{ij} = \frac{w_{ij} e^{\Psi(\alpha_j) - \log\beta_j}}{\sum_k w_{ik} e^{\Psi(\alpha_k) - \log\beta_k}} \tag{1.15}$$

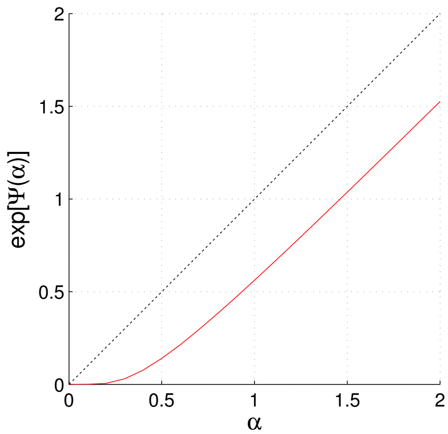

Figure 1: Function $e^{\Psi(\alpha)}$.

Introducing a different parametrisation, one that we wish to map on mitral and granule cells, we get:

$$m_i = \frac{r_i}{g_i} \tag{1.16a}$$

$$g_i = \sum_k w_{ik} e^{\Psi(\alpha_k) - \log \beta_k} \tag{1.16b}$$

$$\alpha_j = \alpha + e^{\Psi(\alpha_j) - \log \beta_j} \sum_i m_i w_{ij} \tag{1.16c}$$

$$\beta_j = \beta + \Delta t \sum_i w_{ij} \,. \tag{1.16d}$$

To solve Eqs. (1.16) in a way that mimics the kinds of operations that could be performed by neuronal circuitry, we write down a set of differential equations that have fixed points satisfying Eq. (1.16),

$$\tau_m \frac{dm_i}{dt} = r_i - m_i \sum_k w_{ik} e^{\Psi(\alpha_k) - \log \beta_k} \tag{1.17a}$$

$$\tau_g \frac{dg_i}{dt} = -g_i + \sum_k w_{ik} e^{\Psi(\alpha_k) - \log \beta_k} \tag{1.17b}$$

$$\tau_\alpha \frac{d\alpha_j}{dt} = \alpha + e^{\Psi(\alpha_j) - \log \beta_j} \sum_i m_i w_{ij} - \alpha_j \tag{1.17c}$$

Variables m_i would be encoded by M-cells, and $g_i = \sum_k w_{ik} e^{\Psi(\alpha_k) - \log \beta_k}$ by granule cells, receiving feedback projection from α_k, which should reside in the cortex. $e^{\Psi(\alpha_j)}$ is a smoothly thresholded function of α_j, see Fig. 1.

2 Algorithms to compare against

MAP with L1 prior, with scale matched to the spike-and-slab:

$$\lambda = 1/\langle c \rangle = (\pi\alpha/\beta + (1-\pi)\alpha_1/\beta_1)^{-1}$$

$$k_z(t) = k_{r0} - \Delta t \mathbf{w_i} \cdot \mathbf{c}(t) \tag{2.1}$$

$$\tau_L \frac{dc}{dt} = w'(k_r(t)/k_z(t) - 1) - \lambda \tag{2.2}$$

Koulakov

$$\tau_L \frac{du}{dt} = -u + w'(k_r(t) - k_{r0} - \mathbf{w}_i \cdot v) \tag{2.3}$$

$$v = [u - \theta]_+ \tag{2.4}$$

$$\theta \equiv \lambda \tag{2.5}$$

3 Notes

One more observation - in gamma distribution: SNR $\sim \sqrt{\alpha}$.

$$\alpha_j \sim r_i \sim (w * \alpha_1/\beta_1)dt$$

$$\beta_j \sim wdt$$

$$var \sim \alpha_1/\beta_1/(wdt)$$

so increasing concentration, increases the variance, whereas with larger w, uncertainty decreases.