Macrologit: Logit-Space Laplace Approximation for MacroTaylor

We consider an ensemble of $N_{\rm ch}$ independent Markov channels with K microscopic states. The macroscopic occupancy vector is

$$\boldsymbol{p} = (p_1, \dots, p_K), \qquad p_i \geq 0, \quad \sum_{i=1}^{K} p_i = 1.$$

Over a single interval we observe a scalar current y with MacroTaylor likelihood

$$F_{\rm data}(\boldsymbol{p}) = \frac{1}{2} \log V(\boldsymbol{p}) + \frac{1}{2} \frac{\delta(\boldsymbol{p})^2}{V(\boldsymbol{p})},$$

where

$$\mu_y(\boldsymbol{p}) = N_{\rm ch}\,(\boldsymbol{p} \cdot \boldsymbol{\gamma}), \qquad \delta(\boldsymbol{p}) = y - \mu_y(\boldsymbol{p}), \qquad V(\boldsymbol{p}) = \epsilon^2 + N_{\rm ch}\,(\boldsymbol{p} \cdot \boldsymbol{\sigma}^2).$$

Here $\boldsymbol{\gamma}$ is the per-state single-channel current and $\boldsymbol{\sigma}^2$ the per-state intrinsic variance (MacroTaylor objects).

A "Gaussian on the simplex" prior is specified by a per-channel covariance $\boldsymbol{\Sigma}$ (rank $K-1$ on the simplex) and a prior mean $\boldsymbol{\mu}$. The corresponding fraction-level prior for $\boldsymbol{p}$ is

$$F_{\rm prior}(\boldsymbol{p}) = \frac{N_{\rm ch}}{2}\,(\boldsymbol{p} - \boldsymbol{\mu})\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{p} - \boldsymbol{\mu})^{\top}.$$

The exact negative log-posterior in $\boldsymbol{p}$ is

$$F_{\rm exact}(\boldsymbol{p}) = F_{\rm data}(\boldsymbol{p}) + F_{\rm prior}(\boldsymbol{p}).$$

MacroTaylor provides explicit formulas for the gradient $\boldsymbol{g}_p(\boldsymbol{p}) = \nabla_{\boldsymbol{p}} F_{\rm exact}(\boldsymbol{p})$ and Hessian $\boldsymbol{H}_p(\boldsymbol{p}) = \nabla^2_{\boldsymbol{p}} F_{\rm exact}(\boldsymbol{p})$, with a rank-2 data curvature in the directions spanned by $\boldsymbol{\gamma}$ and $\boldsymbol{\sigma}^2$.

1. Softmax parametrisation of the simplex

We introduce unconstrained logits $\boldsymbol{\theta} \in \mathbb{R}^{K-1}$ by choosing state K as reference (gauge). Define the extended logit vector

$$\tilde{\boldsymbol{\theta}} = (\theta_1, \dots, \theta_{K-1}, 0) \in \mathbb{R}^K,$$

and the normalising constant

$$Z(\boldsymbol{\theta}) = \sum_{j=1}^{K} \exp(\tilde{\theta}_j) = 1 + \sum_{k=1}^{K-1} \exp(\theta_k).$$

Softmax definition. The softmax (reference-state) map $\boldsymbol{\theta} \mapsto \boldsymbol{p}(\boldsymbol{\theta})$ is

$$p_i(\boldsymbol{\theta}) = \frac{\exp(\tilde{\theta}_i)}{Z(\boldsymbol{\theta})}, \quad i = 1, \ldots, K.$$

Explicitly:

$$p_i(\boldsymbol{\theta}) = \frac{\exp(\theta_i)}{Z(\boldsymbol{\theta})}, \quad i = 1, \ldots, K-1, \qquad p_K(\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})}.$$

By construction

$$p_i(\boldsymbol{\theta}) > 0, \qquad \sum_{i=1}^{K} p_i(\boldsymbol{\theta}) = 1 \quad \text{for all } \boldsymbol{\theta}.$$

Softmax Jacobian. The Jacobian $J(\boldsymbol{\theta}) = \partial \boldsymbol{p}/\partial \boldsymbol{\theta}$ is a $K \times (K-1)$ matrix with entries

$$J_{ij}(\boldsymbol{\theta}) = \frac{\partial p_i}{\partial \theta_j}, \quad i = 1, \ldots, K, \ j = 1, \ldots, K-1.$$

For $i \leq K-1$ and $j \leq K-1$:

$$\frac{\partial p_i}{\partial \theta_j} = p_i(\delta_{ij} - p_j),$$

and for the reference state K:

$$\frac{\partial p_K}{\partial \theta_j} = -\,p_K\,p_j, \qquad j = 1, \ldots, K-1.$$

Equivalently, in block form, write

$$\boldsymbol{p}_{1:K-1} = \begin{bmatrix} p_1 \\ \vdots \\ p_{K-1} \end{bmatrix}, \qquad \boldsymbol{p}_{1:K-1}\boldsymbol{p}_{1:K-1}^{\top} \in \mathbb{R}^{(K-1)\times(K-1)}.$$

Then

$$J(\boldsymbol{\theta}) = \begin{bmatrix} \operatorname{diag}(\boldsymbol{p}_{1:K-1}) - \boldsymbol{p}_{1:K-1}\boldsymbol{p}_{1:K-1}^{\top} \\ -\,p_K\,\boldsymbol{p}_{1:K-1}^{\top} \end{bmatrix}.$$

Because $\sum_i p_i(\boldsymbol{\theta}) \equiv 1$, we have

$$\mathbf{1}^{\top} J(\boldsymbol{\theta}) = \mathbf{0}^{\top}, \qquad J(\boldsymbol{\theta})^{\top}\mathbf{1} = \mathbf{0},$$

so the image of J lies in the tangent space of the simplex: all allowed perturbations $\delta\boldsymbol{p} = J\,\delta\boldsymbol{\theta}$ satisfy $\mathbf{1}^{\top}\delta\boldsymbol{p} = 0$.

2. Prior and posterior in logit space

We place a Gaussian prior on $\boldsymbol{\theta}$:

$$\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{S}),$$

with mean $\boldsymbol{m} \in \mathbb{R}^{K-1}$ and covariance $\boldsymbol{S} \in \mathbb{R}^{(K-1)\times(K-1)}$. The prior contribution to the energy is

$$F_{\text{prior}}(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{m})^\top \boldsymbol{S}^{-1}(\boldsymbol{\theta} - \boldsymbol{m}) \quad (+\text{ constant}).$$

The MacroTaylor data term becomes a function of $\boldsymbol{\theta}$ via $\boldsymbol{p}(\boldsymbol{\theta})$:

$$F_{\text{data}}(\boldsymbol{\theta}) = F_{\text{data}}\big(\boldsymbol{p}(\boldsymbol{\theta})\big).$$

The exact negative log-posterior in logit space is therefore

$$F_{\text{exact}}(\boldsymbol{\theta}) = F_{\text{data}}\big(\boldsymbol{p}(\boldsymbol{\theta})\big) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{m})^\top \boldsymbol{S}^{-1}(\boldsymbol{\theta} - \boldsymbol{m}).$$

3. Gradient in logit space

Let $\boldsymbol{g}_p(\boldsymbol{p})$ denote the MacroTaylor gradient of F_{exact} with respect to $\boldsymbol{p}$, treated as a row vector:

$$\boldsymbol{g}_p(\boldsymbol{p}) = \nabla_{\boldsymbol{p}} F_{\text{exact}}(\boldsymbol{p}).$$

By the chain rule,

$$\boldsymbol{g}_\theta(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} F_{\text{exact}}(\boldsymbol{\theta}) = (\boldsymbol{\theta} - \boldsymbol{m})^\top \boldsymbol{S}^{-1} + J(\boldsymbol{\theta})^\top \boldsymbol{g}_p\big(\boldsymbol{p}(\boldsymbol{\theta})\big).$$

This is the gradient used in Newton or quasi-Newton updates in logit space.

4. Hessian and Laplace covariance in logit space

For the data term,

$$F_{\text{data}}(\boldsymbol{\theta}) = F_{\text{data}}\big(\boldsymbol{p}(\boldsymbol{\theta})\big),$$

we have

$$\nabla_{\boldsymbol{\theta}} F_{\text{data}} = J(\boldsymbol{\theta})^\top \boldsymbol{g}_p(\boldsymbol{p}),$$

$$\nabla^2_{\boldsymbol{\theta}} F_{\text{data}} = J(\boldsymbol{\theta})^\top \boldsymbol{H}_p(\boldsymbol{p}) J(\boldsymbol{\theta}) + \sum_{i=1}^{K} g_{p,i}(\boldsymbol{p})\, \nabla^2_{\boldsymbol{\theta}} p_i(\boldsymbol{\theta}),$$

where $\boldsymbol{H}_p(\boldsymbol{p})$ is the MacroTaylor Hessian in $\boldsymbol{p}$ and $g_{p,i}$ is the i-th component of $\boldsymbol{g}_p$.

The exact posterior Hessian in logit space is

$$\boldsymbol{H}_\theta(\boldsymbol{\theta}) = \nabla^2_{\boldsymbol{\theta}} F_{\text{exact}}(\boldsymbol{\theta}) = \boldsymbol{S}^{-1} + J(\boldsymbol{\theta})^\top \boldsymbol{H}_p(\boldsymbol{p}) J(\boldsymbol{\theta}) + \sum_{i=1}^{K} g_{p,i}(\boldsymbol{p})\, \nabla^2_{\boldsymbol{\theta}} p_i(\boldsymbol{\theta}).$$

Near the MAP $\boldsymbol{\theta}^\star$ the last term (curvature of the softmax mapping itself) is typically small when the posterior is concentrated. A Gauss–Newton approximation drops this term:

$$\boldsymbol{H}_\theta(\boldsymbol{\theta}^\star) \approx \boldsymbol{S}^{-1} + J(\boldsymbol{\theta}^\star)^\top \boldsymbol{H}_p\big(\boldsymbol{p}(\boldsymbol{\theta}^\star)\big) J(\boldsymbol{\theta}^\star).$$

The logit-space Laplace covariance is then

$$\boldsymbol{\Sigma}_\theta^{\text{post}} \approx \boldsymbol{H}_\theta(\boldsymbol{\theta}^\star)^{-1}.$$

5. Induced mean and covariance in occupancy space

The posterior over $\boldsymbol{p}$ induced by the Gaussian posterior in $\boldsymbol{\theta}$ is logistic-normal and skewed. We approximate its first two moments by a Taylor expansion of $\boldsymbol{p}(\boldsymbol{\theta})$ around $\boldsymbol{\theta}^\star$.

Write the i-th component as $p_i(\boldsymbol{\theta})$ and denote its gradient and Hessian with respect to $\boldsymbol{\theta}$ at $\boldsymbol{\theta}^\star$ by

$$\boldsymbol{j}_i := \nabla_{\boldsymbol{\theta}} p_i(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}^\star} \in \mathbb{R}^{K-1}, \qquad \boldsymbol{H}_i := \nabla^2_{\boldsymbol{\theta}} p_i(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}^\star} \in \mathbb{R}^{(K-1)\times(K-1)}.$$

By definition,

$$\boldsymbol{j}_i^\top \text{ is the } i\text{-th row of } J(\boldsymbol{\theta}^\star).$$

Second-order mean (default). Using a second-order Taylor expansion and taking the Gaussian expectation, the mean of p_i is approximated by

$$\mathbb{E}[p_i] \approx p_i(\boldsymbol{\theta}^\star) + \frac{1}{2}\operatorname{tr}\big(\boldsymbol{H}_i\, \boldsymbol{\Sigma}_\theta^{\text{post}}\big), \qquad i = 1, \dots, K.$$

Thus our default Macrologit posterior mean in occupancy space is

$$(\boldsymbol{\mu}_p)_i \approx p_i(\boldsymbol{\theta}^\star) + \frac{1}{2}\operatorname{tr}\big(\boldsymbol{H}_i\, \boldsymbol{\Sigma}_\theta^{\text{post}}\big).$$

The softmax Hessians $\boldsymbol{H}_i$ can be computed analytically from the definition $p_i(\boldsymbol{\theta}) = \exp(\tilde{\theta}_i)/Z(\boldsymbol{\theta})$. (For implementation: derive $\nabla^2_{\boldsymbol{\theta}} p_i$ once symbolically and reuse it.)

First-order covariance (delta method). For the covariance we keep the first-order (delta method) approximation: to first order in $(\boldsymbol{\theta} - \boldsymbol{\theta}^\star)$,

$$\boldsymbol{p}(\boldsymbol{\theta}) \approx \boldsymbol{p}(\boldsymbol{\theta}^\star) + J(\boldsymbol{\theta}^\star)(\boldsymbol{\theta} - \boldsymbol{\theta}^\star),$$

so

$$\boldsymbol{\Sigma}_p \approx J(\boldsymbol{\theta}^\star)\,\boldsymbol{\Sigma}_\theta^{\mathrm{post}}\,J(\boldsymbol{\theta}^\star)^\top.$$

Because $J(\boldsymbol{\theta}^\star)$ satisfies $\mathbf{1}^\top J(\boldsymbol{\theta}^\star) = \mathbf{0}^\top$, the induced covariance obeys

$$\boldsymbol{\Sigma}_p\,\mathbf{1} = \mathbf{0}, \qquad \mathbf{1}^\top \boldsymbol{\Sigma}_p = \mathbf{0}^\top,$$

so all fluctuations lie in the tangent space of the simplex and the sum of probabilities is preserved.

If desired, a purely first-order mean approximation is obtained by dropping the Hessian terms:

$$(\boldsymbol{\mu}_p)_i \approx p_i(\boldsymbol{\theta}^\star).$$

In Macrologit we use the second-order expression above as the default.

6. Macrologit interval update algorithm (summary)

For one interval and observation y:

1. **Prior in logit space.** Start from a prior $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{S})$, or map a prior in $\boldsymbol{p}$ to logits via $\boldsymbol{\theta}_0 = \mathrm{logit}(\boldsymbol{p}_0)$ (reference-state gauge).
2. **Newton iterations in $\boldsymbol{\theta}$.** At a current iterate $\boldsymbol{\theta}_k$:
   (a) Compute $\boldsymbol{p}_k = \boldsymbol{p}(\boldsymbol{\theta}_k)$ via the softmax definition.
   (b) Evaluate the MacroTaylor gradient $\boldsymbol{g}_p(\boldsymbol{p}_k)$ and Hessian $\boldsymbol{H}_p(\boldsymbol{p}_k)$ in $\boldsymbol{p}$.
   (c) Compute the softmax Jacobian $J_k = J(\boldsymbol{\theta}_k)$.
   (d) Form
   $$\boldsymbol{g}_\theta(\boldsymbol{\theta}_k) = (\boldsymbol{\theta}_k - \boldsymbol{m})^\top \boldsymbol{S}^{-1} + J_k^\top \boldsymbol{g}_p(\boldsymbol{p}_k),$$
   $$\boldsymbol{H}_\theta(\boldsymbol{\theta}_k) \approx \boldsymbol{S}^{-1} + J_k^\top \boldsymbol{H}_p(\boldsymbol{p}_k) J_k$$
   (Gauss–Newton approximation).
   (e) Take a Newton step
   $$\Delta\boldsymbol{\theta}_k = -\,\boldsymbol{g}_\theta(\boldsymbol{\theta}_k)\,\boldsymbol{H}_\theta(\boldsymbol{\theta}_k)^{-1}, \qquad \boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}_k.$$

3. **MAP and logit covariance.** Iterate until convergence to the MAP $\boldsymbol{\theta}^\star$. Approximate the logit posterior as

$$\boldsymbol{\theta} \mid y \approx \mathcal{N}(\boldsymbol{\theta}^\star, \boldsymbol{\Sigma}_\theta^{\mathrm{post}}), \qquad \boldsymbol{\Sigma}_\theta^{\mathrm{post}} \approx \boldsymbol{H}_\theta(\boldsymbol{\theta}^\star)^{-1}.$$

4. **Posterior in occupancy space.** Compute the Macrologit posterior summary in $\boldsymbol{p}$:

$$(\boldsymbol{\mu}_p)_i \approx p_i(\boldsymbol{\theta}^\star) + \frac{1}{2}\,\mathrm{tr}\big(\boldsymbol{H}_i\,\boldsymbol{\Sigma}_\theta^{\mathrm{post}}\big), \qquad \boldsymbol{\Sigma}_p \approx J(\boldsymbol{\theta}^\star)\,\boldsymbol{\Sigma}_\theta^{\mathrm{post}}\,J(\boldsymbol{\theta}^\star)^\top.$$

   This respects $p_i > 0$ and $\sum_i p_i = 1$ by construction, and fluctuations remain in the tangent space of the simplex.

Away from the boundaries of the simplex, the resulting MAP and covariance in $\boldsymbol{p}$ agree, to Laplace order, with the linear MacroTaylor approximation performed directly in $\boldsymbol{p}$-space. Near the boundaries, the logit parametrisation prevents illegal probabilities while preserving the MacroTaylor interval likelihood structure.