# Bayesian-Inspired Upgrades for Proteus: Mathematical Formulation

Dated: 2025-??-??

## 1 Overview

This note provides formal definitions, assumptions, and problem statements for a set of Bayesian-inspired upgrades to the Proteus architecture (Stage 1/2). For each item we: (i) state the mathematical problem with the current system, (ii) introduce the proposed technique and its assumptions, and (iii) show how it addresses the problem.

Throughout, let $G = (\mathcal{V}, \mathcal{E})$ be the GNG/Hebbian graph with nodes $\{i\}$, let $n_{i\to j}$ denote directed transition counts ($\mathrm{BMU}_1 \to \mathrm{BMU}_2$ tallies), and let $\mathcal{S}$ denote the dual simplicial complex with simplex masses $m = \{m_S\}_{S\in\mathcal{S}}$.

## 2 Evidence-Based Split/Stop (Bayes Factor / WAIC / LOO)

### 2.1 Problem: Geometry-Only Splits Ignore Sample Size

Current split triggers use geometric proxies (variance $\sigma_i^2$ or torsion ratio $R_S$) against fixed thresholds. Formally, the decision rule is

$$\text{Split at node } i \iff \sigma_i^2 > \tau_{\text{local},i} \quad \text{or} \quad R_S > r_0.$$

This rule does not condition on local evidence size $n_{\text{eff}}(i)$, making Type-I errors (false splits) likely when $n_{\text{eff}}(i)$ is small.

### 2.2 Model: Transition Likelihood With Dirichlet Smoothing

Let $q(j \mid i; m)$ be the geometry-induced router (cf. SI S13):

$$q(j \mid i; m) = \frac{\sum_{S\ni i,j} \kappa_{ijS}\, m_S}{\sum_{S\ni i} \kappa_{iS}\, m_S}, \quad \kappa_{ijS} \geq 0.$$

For node $i$, transitions are modeled as

$$\{n_{i\to j}\}_j \mid n_i, m \sim \text{Multinomial}\big(n_i,\, q(\cdot \mid i; m)\big),$$

with a symmetric Dirichlet prior $q(\cdot \mid i) \sim \text{Dir}(\alpha_0 \mathbf{1})$ integrated out, yielding the Dirichlet–multinomial (DM) marginal likelihood

$$\mathcal{L}_{\text{DM}}(\{n_{i\to j}\}; \alpha_0, \theta_i) \;=\; \frac{\Gamma(\alpha_0 J)}{\Gamma(\alpha_0 J + n_i)} \prod_{j=1}^{J} \frac{\Gamma(\alpha_0 + n_{i\to j})}{\Gamma(\alpha_0)},$$

where $J = |\mathcal{N}(i)|$ and $\theta_i$ denotes the implied $q(\cdot \mid i; m)$ ordering (used to align categories).

### 2.3 Split Hypotheses and Bayes Factor

Consider a candidate split of node $i$ along direction $\mathbf{u}_i$ into children $i_1, i_2$. Using a rolling buffer of routed events at $i$, partition events into two buckets by sign/proximity relative to $\mathbf{u}_i$ (cheap proxy for the split). Let $\{n^{(1)}_{i\to j}\}, \{n^{(2)}_{i\to j}\}$ be child tallies and $\{n_{i\to j}\} = \{n^{(1)}_{i\to j} + n^{(2)}_{i\to j}\}$ the parent tallies.
Hypotheses:

$$H_0 : \text{ no split (one router } q \text{ for all events)} \qquad H_1 : \text{ split (two routers } q^{(1)}, q^{(2)}).$$

Using DM marginals with the same $\alpha_0$ for parent and children, the Bayes factor is

$$\mathrm{BF}_{\text{split}} = \frac{\mathcal{L}_{\text{DM}}(\{n^{(1)}_{i\to j}\}; \alpha_0)\, \mathcal{L}_{\text{DM}}(\{n^{(2)}_{i\to j}\}; \alpha_0)}{\mathcal{L}_{\text{DM}}(\{n_{i\to j}\}; \alpha_0)}.$$

Decision: split iff $\mathrm{BF}_{\text{split}} > \tau_{\text{BF}}$ and $n_{\text{eff}}(i) \geq n_{\min}$.

### 2.4 Predictive Alternatives (WAIC / PSIS-LOO)

Let $\ell(x; \theta)$ be per-event log predictive density under the router implied by $m$. WAIC and PSIS-LOO approximate expected out-of-sample log predictive density, yielding a scalar score ELPD. A split is favored when $\Delta\mathrm{ELPD} = \mathrm{ELPD}_{H_1} - \mathrm{ELPD}_{H_0} > 0$. A BF proxy follows via $\mathrm{BF} \approx \exp(\Delta\mathrm{ELPD})$.

### 2.5 Assumptions

- Locally stationary sampling over the buffer window; stable neighbor set for $i$.
- Router $q$ is identifiable (see Sec. 7).
- Symmetric Dirichlet smoothing with small $\alpha_0$ to regularize low counts.

### 2.6 Resolution of the Problem

The DM marginal and BF introduce an explicit dependence on $n_{\text{eff}}(i)$, shrinking spurious improvements when data are scarce. As $n_{\text{eff}}(i) \to \infty$, the criterion reduces to a likelihood-ratio test between parent and child routers, ensuring splits concentrate where transition structure is reproducible.

## 3 Split Gating Mechanics and False-Split Control

### 3.1 Gate Design

Use a composite gate:

$$\text{Split if } (\mathrm{BF}_{\text{split}} > \tau_{\text{BF}}) \;\wedge\; (n_{\text{eff}}(i) \geq n_{\min}) \;\wedge\; (t - t^{\text{last}}_{\text{decision}} \geq T_{\text{cool}}).$$

Add hysteresis: to reverse a pending split, require $\mathrm{BF}_{\text{split}} < \tau_{\text{BF}}/c$ with $c > 1$.

### 3.2 Asymptotic Control

Under $H_0$ and regularity, $2 \log \mathrm{BF}_{\text{split}}$ concentrates near 0 with variance decaying as $\mathcal{O}(1/n_{\text{eff}})$; thus the probability of false split decays exponentially in $n_{\text{eff}}$ for fixed $\tau_{\text{BF}}$.

## 4 Alpha Schedule via Dirichlet Refinement

Let a parent cell $P$ at scale $s$ have mass $m_P$. Children $\mathrm{ch}(P)$ at scale $s+1$ receive a Dirichlet refinement prior

$$m^{(s+1)}_{\mathrm{ch}(P)} \mid m^{(s)}_P \;\sim\; \mathrm{Dir}\big(\alpha_s\, \pi^{(s)}_{\mathrm{ch}(P)}\, m^{(s)}_P\big), \qquad \sum_{C\in \mathrm{ch}(P)} m^{(s+1)}_C = m^{(s)}_P.$$

**Schedule:** set $\alpha_s = \alpha_{\min} + f(\text{torsion}, n_{\text{eff}})\,(\alpha_{\max} - \alpha_{\min})$ with monotone $f \in [0,1]$ decreasing in torsion and increasing in evidence.

### 4.1 Effect

Posterior mean contracts toward the parent when $\alpha_s$ is large, reducing variance of child masses: $\mathrm{Var}[m_C] \propto 1/(\alpha_s+n_C)$. In flat, high-evidence regions, large $\alpha_s$ prevents unnecessary divergence; in curved/uncertain regions, small $\alpha_s$ permits exploration.

## 5 Smoothness Prior Scheduling (Logistic-GMRF)

Let $\theta_S \in \mathbb{R}$ be log-masses with a Gaussian Markov random field prior on the dual graph $L$:

$$\theta \sim \mathcal{N}(0,\, \tau^{-1}L^{+}), \qquad m_S = \frac{e^{\theta_S}}{\sum_{S'} e^{\theta_{S'}}}.$$

**Schedule:** set $\tau = 0$ (off) except in low-evidence patches, where $\tau \in [10^{-3}, 10^{-1}]$.

### 5.1 Effect

Adds a quadratic penalty $\frac{\tau}{2}\,\theta^\top L\,\theta$ to the log posterior, promoting neighboring simplexes to have similar log-mass. This stabilizes estimates where the transition likelihood is weak, while leaving high-evidence areas nearly unchanged (since the likelihood dominates).

## 6 Uncertainty-Aware Gates (Laplace Variance)

Approximate the posterior around the MAP by a Gaussian via Laplace: $\theta \approx \mathcal{N}(\hat\theta, H^{-1})$ with Hessian $H = -\nabla^2_\theta \log p(\theta \mid \text{data})$. Gate irreversible actions if a local uncertainty metric exceeds a threshold, e.g.,

$$\max_{S\in\mathcal{P}} \mathrm{Var}[m_S] \text{ or } \kappa(H) > \kappa_{\max}.$$

This blocks splits/prunes/warps when posterior curvature is weak (ill-posed decision).

## 7 Router Identifiability and Conditioning

Impose $\kappa_{ijS} \geq 0$, normalization $\sum_{j:S\ni i,j} \kappa_{ijS} = \kappa_{iS}$, and a minimum separation constraint ensuring the Jacobian $\partial q/\partial m$ has bounded condition number on stars $\mathcal{S}(i)$. Sufficient condition: there exists $\delta > 0$ such that for adjacent faces $S, S'$ sharing $i$,

$$\left.\frac{\kappa_{ijS}}{\sum_{S\ni i}\kappa_{iS}} - \frac{\kappa_{ijS'}}{\sum_{S'\ni i}\kappa_{iS'}}\right| \geq \delta \;\text{ for some } j \in \mathcal{N}(i).$$

This avoids degenerate mappings where distinct local mass patterns induce indistinguishable transitions.

## 8 Variational/BP Refinement Near Convergence

Define a variational family over $\theta$ (e.g., mean-field Gaussian) and optimize the ELBO

$$\mathcal{L}(\lambda) = \mathbb{E}_{q_\lambda(\theta)}[\log p(\text{data} \mid \theta)] - \mathrm{KL}\big(q_\lambda(\theta) \| p(\theta)\big).$$

Run a small, fixed number of iterations near convergence to reduce posterior bias from pure MAP. Alternatively, run loopy BP on the dual graph factors (cf. SI S13) for $K$ iterations.

## 9 Posterior Predictive Reporting

Report credible intervals for $m_S$ via variational covariance or Laplace: $m_S \approx \mathrm{softmax}(\theta_S)$ with $\theta \sim \mathcal{N}(\hat{\theta}, H^{-1})$. For transitions, propagate uncertainty through $q(j \mid i; m)$ to obtain intervals on $\hat{p}(j \mid i)$.

## 10 Dual-Flow Hyperparameter Tying

Let the dual-flow quadratic objective be

$$\min_{\{p_f\}} \lambda \sum_f (p_f - \hat{p}_f)^2 + \mu \sum_S \|A_S\, \mathbf{p}_S\|_2^2.$$

Tie $\lambda$ to Dirichlet strength (effective sample size) and $\mu$ to GMRF precision $\tau$: $\lambda \propto (\alpha_0 + n_S)$, $\mu \propto \tau$. This aligns smoothing in flow reconstruction with smoothing in mass inference.

## 11 Streaming-Friendly Conjugacy

Maintain node-local Dirichlet pseudo-counts $\tilde{n}_{i\to j}$ and update online:

$$\tilde{n}_{i\to j} \leftarrow \rho\, \tilde{n}_{i\to j} + \Delta n_{i\to j}, \qquad q(j \mid i) = \frac{\tilde{n}_{i\to j} + \alpha_0}{\sum_{j'} (\tilde{n}_{i\to j'} + \alpha_0)}.$$

Periodically (or near convergence), run a short variational/BP pass to reconcile local posteriors into a coherent global $m$.