
SAMPLING FROM JOINT DENSITY OF AUTOREGRESSIVE CONTINUOUS PROCESS AND SPIKE TRAIN: A PÓLYA–GAMMA BASED DATA-AUGMENTATION APPROACH

A PREPRINT

Anonymous

June 30, 2025

ABSTRACT

Keywords

1 Generative Model

We formulate a joint generative model for the continuous local-field-potential (LFP) signal $\{X_t\}_{t=1}^T$ and the binary spike train $\{N_t\}_{t=1}^T$. The LFP X_t is treated as a random process with an autoregressive (AR) Gaussian prior, while the spike N_t is a Bernoulli random variable whose log-odds at time t depends linearly on the current and past values of X_t .

LFP dynamics: AR(k) Gaussian process. For $t > k$ we assume

$$X_t \mid X_{t-1}, \dots, X_{t-k}, \phi, \sigma^2 \sim \mathcal{N}(\phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_k X_{t-k}, \sigma^2), \quad (1)$$

where $\phi = (\phi_1, \dots, \phi_k)$ are the AR coefficients and σ^2 is the process-noise variance. The initial k values X_1, \dots, X_k are taken as known or drawn from the stationary distribution of the AR process (e.g. a multivariate Gaussian). Equation (1) encodes the assumption that the LFP is a smooth signal with temporal correlation.

Spike emission: logistic model with s lags of X . Given the current LFP X_t and its past s values,

$$N_t \mid X_{t:t-s}, \beta \sim \text{Bernoulli}(p_t), \quad p_t = \sigma(\psi_t), \quad \psi_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \dots + \beta_{s+1} X_{t-s}, \quad (2)$$

where $\sigma(u) = 1/(1 + e^{-u})$. The coefficient β_0 is a baseline log-odds; the vector $(\beta_1, \dots, \beta_L)$ quantifies how the current and the $L - 1$ most recent latent states modulate spike probability. In the simplest special case $L = 1$, $\sigma^{-1}(p_t) = \beta_0 + \beta_1 X_t$.

Fix the AR order k and the maximum spike-state lag s . Let $L = \max\{k, s\}$ so that all required lags exist. Write $\phi = (\phi_1, \dots, \phi_k)$ and $\beta = (\beta_0, \beta_1, \dots, \beta_{s+1})$. Give the first L latent samples a Gaussian prior $p(X_{1:L}) = \mathcal{N}(\mathbf{m}_0, \mathbf{P}_0)$. Then the complete-data likelihood factorises as

$$p(X_{1:T}, N_{L+1:T} \mid \beta, \phi) = p(X_{1:L}) \prod_{t=L+1}^T \underbrace{p(X_t \mid X_{t-1}, \dots, X_{t-k}, \phi)}_{\text{AR}(k) \text{ transition}} \prod_{t=L+1}^T \underbrace{p(N_t \mid X_t, X_{t-1}, \dots, X_{t-s}, \beta)}_{\text{Bernoulli observation}} \quad (3)$$

where

$$p(X_t \mid X_{t-1}, \dots, X_{t-k}, \phi) = \mathcal{N}\left(X_t \mid \sum_{j=1}^k \phi_j X_{t-j}, \sigma^2\right) \quad (4)$$

and

$$p(N_t \mid X_t, X_{t-1}, \dots, X_{t-s}, \beta) = \left[\sigma\left(\beta_0 + \sum_{j=1}^s \beta_{j+1} X_{t-j}\right) \right]^{N_t} \left[1 - \sigma\left(\beta_0 + \sum_{j=1}^s \beta_{j+1} X_{t-j}\right) \right]^{1-N_t} \quad (5)$$

This fully specifies the generative process linking the continuous-time latent (LFP-like) dynamics and the discrete spike train.

Parameter priors

We now adopt a Bayesian framework and place prior over the parameters and Let $\theta = \{\phi, \sigma^2, \beta\}$. We assume independent priors for the three parameter blocks:

- **AR coefficients:** $\phi \sim \mathcal{N}(\mu_\phi, \Sigma_\phi)$, e.g. $\mu_\phi = \mathbf{0}$ and $\Sigma_\phi = \tau_\phi^2 I_k$ with large τ_ϕ^2 .
- **Noise variance:** $\sigma^2 \sim \text{InvGamma}(\alpha_0, \beta_0)$, giving a diffuse scale prior when (α_0, β_0) are small.
- **Logistic coefficients:** $\beta \sim \mathcal{N}(\mu_\beta, \Sigma_\beta)$, typically with $\mu_\beta = \mathbf{0}$ and $\Sigma_\beta = \tau_\beta^2 I_{s+2}$.

These priors are convenient because they are conjugate or conditionally conjugate once a suitable data–augmentation is applied (as we will discuss in the next section), enabling efficient Gibbs sampling for posterior inference.

2 Bayesian Inference via Pólya–Gamma Augmentation

2.1 Inference challenge.

Direct Bayesian updating in this model is complicated by the non-conjugate logistic likelihood in Eq. (3): the posterior does not admit closed-form expressions because the Bernoulli likelihood $\sigma(z_t)^{N_t} (1 - \sigma(z_t))^{1-N_t}$ is not conjugate to a Gaussian prior on X_t or β . We therefore pursue a *data-augmentation* strategy, introducing latent variables that render the conditional posteriors tractable. In particular, we leverage the Pólya–Gamma (PG) augmentation scheme of Polson et al. [2013], developed for logistic regression models, which creates an auxiliary variable that makes the log-odds appear in Gaussian form. Our Gibbs sampler iteratively samples

- the regression coefficients β given the current state sequence $X_{1:T}$,
- the latent states $X_{1:T}$ given the spike data, and
- a set of PG variables given the current (β, X) ,

thereby cycling through the full conditional distributions. We first state the Pólya–Gamma identity and its application to our model, then detail the resulting Gibbs updates.

2.2 Pólya–Gamma based data-augmentation.

Pólya–Gamma identity. A random variable ω is said to follow a Pólya–Gamma distribution with parameters (b, c) , written $\omega \sim \text{PG}(b, c)$, if it can be represented as an infinite mixture of Gammas (see Polson et al., 2013 for definitions). The crucial identity (Polson et al., 2013, Thm. 1) is that for any $b > 0$ and $\psi \in \mathbb{R}$

$$\frac{e^{a\psi}}{(1 + e^\psi)^b} = 2^{-b} e^{\kappa\psi} \int_0^\infty \exp\left(-\frac{\omega\psi^2}{2}\right) p(\omega) d\omega, \quad \kappa = a - \frac{b}{2}, \quad (6)$$

where $p(\omega)$ is the density of $\text{PG}(b, 0)$. Thus the troublesome factor $(1 + e^\psi)^{-b}$ becomes a Gaussian kernel in ψ at the cost of introducing ω . We now specialize to the Bernoulli case in our model to make obvious how this identity is crucial in designing an efficient Gibbs sampler.

For each bin t , recall that we impose the logistic link, $\sigma(\psi_t) = \frac{1}{1 + e^{-\psi_t}}$, so that the Bernoulli likelihood is

$$p(N_t | \psi_t) = \sigma(\psi_t)^{N_t} [1 - \sigma(\psi_t)]^{1-N_t} = \frac{e^{N_t \psi_t}}{1 + e^{\psi_t}}. \quad (7)$$

Introduce the centred quantity $\kappa_t = N_t - \frac{1}{2} \in \{-\frac{1}{2}, +\frac{1}{2}\}$ and set $b = 1$ and $a = N_t$ in (6); this yields

$$\frac{e^{N_t \psi_t}}{1 + e^{\psi_t}} = \frac{1}{2} e^{\kappa_t \psi_t} \int_0^\infty \exp\left(-\frac{\omega_t \psi_t^2}{2}\right) p_{\text{PG}}(\omega_t | 1, 0) d\omega_t. \quad (8)$$

Data augmentation. Interpreting each ω_t as an *auxiliary* draw

$$\omega_t \sim \text{PG}(1, 0),$$

identity (8) becomes an equality in distribution. With the ψ_t given by the linear prediction in (2), (8) yields the *augmented* likelihood

$$p(N_t \mid \beta, X_{t:t-s}, \omega_t) = p(N_t \mid \psi_t, \omega_t) \quad (9)$$

$$\begin{aligned} &\propto \exp\left\{-\frac{\omega_t}{2} \psi_t^2 + \kappa_t \psi_t\right\} \\ &\propto \exp\left(-\frac{\omega_t}{2} [\psi_t - \kappa_t/\omega_t]^2\right), \end{aligned} \quad (10)$$

i.e. a Gaussian kernel in ψ_t .

Note that with this augmentation scheme, since ψ_t is *linear* in the coefficient vector β , and $X_{t:t-s}$ expression (10) implies that, under any Gaussian prior on β , the augmented posterior $p(\beta \mid X_{1:T}, N_{1:T}, \omega_{1:T})$ is multivariate normal. Hence the originally non-conjugate logistic likelihood becomes conditionally conjugate after augmentation, so that β admits a closed-form Gaussian Gibbs update.

In cases where X_t is latent, the same Gaussian form in (10), combined with the AR prior, provides Gaussian conjugacy for the whole latent trajectory. The full latent trajectory $X_{1:T}$ can then be drawn by forward-filtering backward-sampling.

Conditionally on (ψ_t, N_t) the density in (8) factors as

$$p(\omega_t \mid \psi_t, N_t) \propto \exp\left(-\frac{\omega_t \psi_t^2}{2}\right) p_{\text{PG}}(\omega_t \mid 1, 0),$$

which is precisely the kernel of a $\text{PG}(1, \psi_t)$ distribution (Polson et al., 2013). Hence the update in the Gibbs sampler is simply

$$\omega_t \mid \beta, X_t, N_t \sim \text{PG}(1, \psi_t),$$

and the ω_t draws are *independent* across time bins. Efficient numerical samplers for $\text{PG}(1, c)$ are available in the Python package `polyagamma` that we use in our experiments..

2.3 Gibbs sampler for observed continuous process

Throughout this section the continuous process X_t is *assumed fully observed*; that is, we infer the model parameters $\theta = (\phi, \sigma^2, \beta)$ conditioned on the observed $X_{1:T}, N_{1:T}$. Thanks to the Pólya–Gamma augmentation, every augmented-conditional enjoys close-form updates and it suffices to have efficient sampler for Gaussian and Pólya–Gamma distribution.

Notation for lags. Define the *design vector* for the logistic observation and the usual regression matrices for the AR block:

$$\begin{aligned} \varphi_t &= [1, X_t, X_{t-1}, \dots, X_{t-s}]^\top \in \mathbb{R}^{s+2}, & F &= \begin{bmatrix} \varphi_1^\top \\ \vdots \\ \varphi_T^\top \end{bmatrix} \in \mathbb{R}^{T \times (s+2)}, \\ Z &= \begin{bmatrix} X_{k-1} & \dots & X_0 \\ \vdots & & \vdots \\ X_{T-1} & \dots & X_{T-k} \end{bmatrix} \in \mathbb{R}^{(T-k) \times k}, & y &= X_{k:T-1} \in \mathbb{R}^{T-k}. \end{aligned}$$

The t -th linear predictor is $\psi_t = \beta^\top \varphi_t$.

Priors. Assume independent priors

$$\beta \sim \mathcal{N}(\mu_\beta^{(0)}, \Sigma_\beta^{(0)}), \quad \phi \sim \mathcal{N}(\mu_\phi^{(0)}, \Sigma_\phi^{(0)}), \quad \sigma^2 \sim \text{InvGamma}(\alpha_0, \beta_0).$$

Gibb’s sampling and update rules. Starting from initial $\{\beta^{(0)}, \phi^{(0)}, \sigma^{2(0)}\}$, we iteratively sample the following variables

- *Pólya–Gamma variables.* For every bin t draw

$$\omega_t \sim \text{PG}(1, \psi_t^{(m-1)}), \quad \psi_t^{(m-1)} = (\beta^{(m-1)})^\top \varphi_t.$$

- *Logistic coefficients* β . Let $\kappa_t = N_t - \frac{1}{2}$ and $\Omega = \text{diag}(\omega_1, \dots, \omega_T)$. Then

$$\Sigma_\beta^* = \left(F^\top \Omega F + (\Sigma_\beta^{(0)})^{-1} \right)^{-1}, \quad \mu_\beta^* = \Sigma_\beta^* \left(F^\top \kappa + (\Sigma_\beta^{(0)})^{-1} \mu_\beta^{(0)} \right),$$

$$\beta^{(m)} \sim \mathcal{N}(\mu_\beta^*, \Sigma_\beta^*).$$

For a flat (improper) prior set $(\Sigma_\beta^{(0)})^{-1} = 0$.

- *AR coefficients* ϕ . Using the Gaussian–linear regression $y = Z\phi + \varepsilon$ ($\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$),

$$\Sigma_\phi^* = \left(\sigma^{-2} Z^\top Z + (\Sigma_\phi^{(0)})^{-1} \right)^{-1}, \quad \mu_\phi^* = \Sigma_\phi^* \left(\sigma^{-2} Z^\top y + (\Sigma_\phi^{(0)})^{-1} \mu_\phi^{(0)} \right),$$

$$\phi^{(m)} \sim \mathcal{N}(\mu_\phi^*, \Sigma_\phi^*).$$

- *Noise variance* σ^2 . With current residuals $r = y - Z\phi^{(m)}$,

$$\sigma^{2(m)} \sim \text{InvGamma}\left(\alpha_0 + \frac{T-k}{2}, \beta_0 + \frac{1}{2} r^\top r\right).$$

After discarding a burn-in and (optionally) thinning, the retained draws approximate $p(\beta, \phi, \sigma^2 \mid X_{1:T}, N_{1:T})$, from which point estimates, credible intervals and predictive quantities can be computed.

2.4 Model order selection

For analyzing real data, the model order is not known and we would want to select the model order within a reasonable range. We first observe empirically that our sampler is robust to model order mis-specification so long as we choose s to be large enough. The chosen prior naturally regularize β_k to be close to 0 when there the X_{t-k+1} has no effect on the spike train. Therefore, in practice, it might suffice to choose a s sufficiently large to recover the model order.

To uncover the model order in a more principled manner, it only requires minimal changes to our sampler to include a γ term that gates whether the coefficients β_j 's enter the logistic link. Concretely, to allow the data to *select* which of the s_{\max} past LFP values truly modulate spiking, we place a spike–and–slab prior on each logistic coefficient β_j , $j = 1, \dots, s_{\max}$. Write

$$\gamma = (\gamma_1, \dots, \gamma_{s_{\max}}), \quad \gamma_j \in \{0, 1\},$$

for binary inclusion indicators on the lag terms. We then specify

$$\begin{aligned} \gamma_j &\sim \text{Bernoulli}(\pi), & j &= 1, \dots, s_{\max}, \\ \beta_j \mid \gamma_j &\sim (1 - \gamma_j) \delta_0 + \gamma_j \mathcal{N}(0, \tau^2), & j &= 1, \dots, s_{\max}, \\ \beta_0 &\sim \mathcal{N}(0, \tau^2), & & \text{(intercept term)} \end{aligned}$$

where δ_0 is a point mass at zero, τ^2 is the slab variance, and $\pi \in (0, 1)$ controls sparsity. Under this prior, the logistic link in Eq. (2) becomes

$$\psi_t = \beta_0 + \sum_{j=1}^{s_{\max}} \beta_j X_{t-j+1},$$

but only those j with $\gamma_j = 1$ contribute nonzero coefficients.

Recall from Eq. (10) that, conditional on ω , the likelihood in β is Gaussian with design columns $\{F_j\}_{j=1}^{s_{\max}}$ and centered responses $\kappa = N - \frac{1}{2}$. The full conditional of γ_j is

$$p(\gamma_j = 1 \mid \beta_{-j}, \omega, N) = \frac{\pi \text{BF}_j}{(1 - \pi) + \pi \text{BF}_j}, \quad \gamma_j \sim \text{Bernoulli}(p(\gamma_j = 1 \mid \dots)),$$

where the Bayes–factor BF_j comparing slab vs spike is

$$\text{BF}_j = \frac{\int p(\beta_j \mid \omega, N) \mathcal{N}(\beta_j \mid 0, \tau^2) d\beta_j}{p(\beta_j = 0 \mid \omega, N)} = \sqrt{\frac{\tau^2}{S_j}} \exp\left(\frac{1}{2} m_j^2 S_j\right).$$

Here, writing $z_j = F_j \odot \sqrt{\omega}$ and $\kappa = N - \frac{1}{2}$, the Gaussian conditional in β_j has precision and mean

$$S_j = z_j^\top z_j + \frac{1}{\tau^2}, \quad m_j = \frac{z_j^\top \kappa}{S_j}.$$

Equivalently,

$$\log \text{BF}_j = \frac{1}{2} \left[\ln(\tau^2 / S_j) + m_j^2 S_j \right].$$

Sampling each γ_j in this way lets the data select a sparse subset of lags via the spike–and–slab prior.

Gibbs sampler for model order selection Our full Gibbs sampler therefore cycles through

1. Draw $\omega_1, \dots, \omega_T \sim \text{PG}(1, \psi_t)$ (Polya–Gamma layer).
2. Draw each γ_j via its Bernoulli update above.
3. Draw β jointly from its Gaussian conditional $N(\mu_\beta^*, \Sigma_\beta^*)$, zeroing out those with $\gamma_j = 0$.
4. Draw AR-coefficients ϕ and noise σ^2 as in Sec. 2.3.

After discarding burn-in and (optionally) thinning, the retained draws approximate the joint posterior

$$p(\beta, \phi, \sigma^2, \gamma \mid X_{1:T}, N_{1:T}).$$

2.5 Gibbs sampler for *latent* continuous process

We now remove the assumption that the LFP-like trajectory $X_{1:T}$ is observed. All three parameter blocks $\{\beta, \phi, \sigma^2\}$, the auxiliary Pólya–Gamma variables $\omega_{1:T}$, and the latent state sequence $X_{1:T}$ must be sampled within each Gibbs iteration. The new ingredient is a forward–filtering, backward–sampling (FFBS) draw of $X_{1:T}$ from a *linear–Gaussian* state–space model whose “pseudo–observations” depend on the current (β, ω) .

Linear–Gaussian form of the augmented model. Fix iteration index $m-1$ and write $\psi_t^{(m-1)} = \beta^{(m-1)\top} \varphi_t$, with φ_t defined in (7). After drawing $\omega_t^{(m)} \sim \text{PG}(1, \psi_t^{(m-1)})$ we have, by (10),

$$\kappa_t = N_t - \frac{1}{2} = \omega_t^{(m)} \left[\psi_t - \frac{\kappa_t}{\omega_t^{(m)}} \right] + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \omega_t^{(m)-1}).$$

Re-arranging yields a *Gaussian* observation model

$$y_t^{(m)} = H_t(\beta^{(m-1)}) x_t + v_t, \quad v_t \sim \mathcal{N}(0, R_t^{(m)}), \quad (11)$$

where

$$y_t^{(m)} = \frac{\kappa_t}{\omega_t^{(m)}} - \beta_0^{(m-1)}, \quad H_t(\beta) = [\beta_1, \dots, \beta_{s+1}, 0, \dots, 0], \quad R_t^{(m)} = \omega_t^{(m)-1}.$$

(The vector H_t reduces to the scalar β_1 when the latent state is one–dimensional.) The state dynamics remain the AR(k) companion form of (1). Hence, conditional on $\Theta^{(m-1)} = \{\beta^{(m-1)}, \phi^{(m-1)}, \sigma^{2(m-1)}\}$ and $\omega^{(m)}$, the latent trajectory is Gaussian and can be drawn exactly by FFBS.

Forward–filtering. Let $(m_{t|t-1}, P_{t|t-1})$ and $(m_{t|t}, P_{t|t})$ denote the standard Kalman prediction and filtering moments. With transition matrices A, Q from the AR(k) dynamics, run, for $t = L+1, \dots, T$,

$$\begin{aligned} m_{t|t-1} &= A m_{t-1|t-1}, \\ P_{t|t-1} &= A P_{t-1|t-1} A^\top + Q, \\ S_t^{(m)} &= H_t P_{t|t-1} H_t^\top + R_t^{(m)}, \\ K_t^{(m)} &= P_{t|t-1} H_t^\top (S_t^{(m)})^{-1}, \\ m_{t|t}^{(m)} &= m_{t|t-1} + K_t^{(m)} (y_t^{(m)} - H_t m_{t|t-1}), \\ P_{t|t}^{(m)} &= [I - K_t^{(m)} H_t] P_{t|t-1}. \end{aligned}$$

Backward–sampling. Draw $x_T^{(m)} \sim \mathcal{N}(m_{T|T}^{(m)}, P_{T|T}^{(m)})$ and for $t = T-1, \dots, 1$,

$$x_t^{(m)} \sim \mathcal{N}(m_{t|t}^{(m)} + J_t^{(m)} (x_{t+1}^{(m)} - A m_{t|t}^{(m)}), P_{t|t}^{(m)} - J_t^{(m)} P_{t|t-1} J_t^{(m)\top}), \quad J_t^{(m)} = P_{t|t}^{(m)} A^\top (P_{t+1|t-1})^{-1}.$$

This produces a sample $X_{1:T}^{(m)} \sim p(X_{1:T} \mid \omega^{(m)}, \Theta^{(m-1)})$.

Revised Gibbs cycle. Each iteration now proceeds (with the superscript m omitted for brevity):

1. **PG step:** draw $\omega_t \sim \text{PG}(1, \psi_t)$ for $t = 1:T$.
2. **FFBS step:** draw $X_{1:T} \sim p(X_{1:T} \mid \omega, \Theta)$ via the Kalman filter/smoothen above.
3. **Logistic step:** $\beta \sim \mathcal{N}(\mu_\beta^*, \Sigma_\beta^*)$, rebuilding the design matrix F from the *new* state draw.
4. **AR step:** $\phi \sim \mathcal{N}(\mu_\phi^*, \Sigma_\phi^*)$ using Z, y computed from the *new* $X_{1:T}$.
5. **Variance step:** $\sigma^2 \sim \text{InvGamma}(\alpha_*, \beta_*)$ from the updated AR residuals.

Despite the additional FFBS block, every move remains a pure Gibbs update (no Metropolis steps) and the full iteration scales linearly in T when the Kalman recursions are implemented with banded-precision algebra.

3 Simulations

References

Nicholas G. Polson, James G. Scott, and Jesse Windle. Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables. *Journal of the American Statistical Association*, 108(504):1339–1349, December 2013. ISSN 0162-1459, 1537-274X. doi:10.1080/01621459.2013.829001. URL <http://www.tandfonline.com/doi/abs/10.1080/01621459.2013.829001>.