Optimization and Regularization Under Arbitrary Objectives

Jared Lakhani
Supervisor: Dr Etienne Pienaar

Introduction

Update-Step:

- Given current state $\boldsymbol{\theta}^{(j)} \in \mathbb{R}^S$.
- Propose candidate $\boldsymbol{\theta}^*$ obtained from $\boldsymbol{\theta}^* = \boldsymbol{\theta}^{(j)} + \mathbf{Q}$.

Acceptance step:

$$\boldsymbol{\theta}^{(j+1)} = \begin{cases} \boldsymbol{\theta}^*, & \text{if } U < \alpha \\ \boldsymbol{\theta}^{(j)}, & \text{otherwise.} \end{cases}$$

- $U \sim \mathit{Unif}(0, 1)$ induces randomness, so not all "better" proposals are accepted = better exploration of the parameter space.

Acceptance probability α:

$$\alpha = \min\left(\frac{p(\boldsymbol{\theta}^* \mid \mathcal{D})}{p(\boldsymbol{\theta}^{(j)} \mid \mathcal{D})} \cdot \frac{Q(\boldsymbol{\theta}^{(j)} \mid \boldsymbol{\theta}^*)}{Q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(j)})}, 1\right).$$

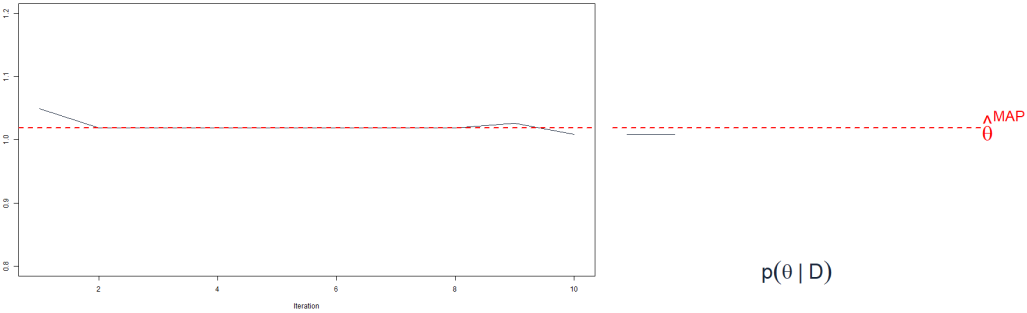
For symmetric proposals, e.g. $\mathbf{Q} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_s)$:

$$\begin{aligned} Q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(j)}) &= \frac{1}{(2\pi)^{S/2}\,|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\tfrac{1}{2}(\boldsymbol{\theta}^* - \boldsymbol{\theta}^{(j)})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}^* - \boldsymbol{\theta}^{(j)})\right) \\ &= Q(\boldsymbol{\theta}^{(j)} \mid \boldsymbol{\theta}^*). \end{aligned}$$

$$\alpha = \min \left(\underbrace{\frac{p(\mathcal{D} \mid \boldsymbol{\theta}^*)}{p(\mathcal{D} \mid \boldsymbol{\theta}^{(j)})}}_{\text{Likelihood ratio}} \cdot \underbrace{\frac{p(\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{(j)})}}_{\text{Prior ratio}}, \ 1 \right).$$

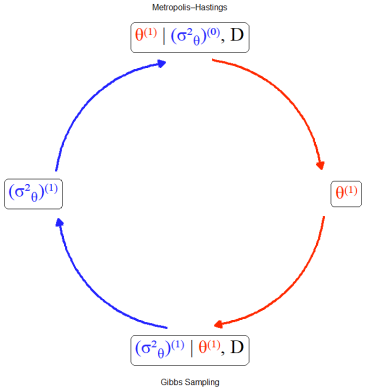
- Higher values $\Rightarrow \boldsymbol{\theta}^*$ fits the data better (via the likelihood).
- Higher values $\Rightarrow \boldsymbol{\theta}^*$ is more consistent with prior beliefs.
- $\boldsymbol{\theta}^*$ may not score highly on both the likelihood and prior.
- Even if both ratios are high, $\boldsymbol{\theta}^*$ may still be rejected if $U > \alpha$.
- If $p(\boldsymbol{\theta}^* \mid \mathcal{D}) \geq p(\boldsymbol{\theta}^{(j)} \mid \mathcal{D})$, then $\boldsymbol{\theta}^*$ is accepted automatically.

- Define: Maximum a Posteriori estimate: $\hat{\boldsymbol{\theta}}^{MAP} = \text{argmax}_{\boldsymbol{\theta}}\, p(\boldsymbol{\theta} \mid \mathcal{D})$ = the mode of the posterior.



$\theta^{(j)}$ vs. MCMC iterations, and formation of $p(\theta \mid \mathcal{D})$.

- Gaussian prior for parameters $\boldsymbol{\theta} \in \mathbb{R}^S \implies \boldsymbol{\theta} \sim \mathcal{N}(\mathbf{o}, \sigma_\theta^2 \mathbf{I}_S)$.
- Express uncertainty about disperion $\sigma_\theta^2 \rightarrow$ give it its own distribution to sample from.
- σ_θ^2 now also needs a prior $\rightarrow$, assume $\sigma_\theta^2 \sim$ Inv-Gamma(a, b) with $a, b \approx 0 \implies$ nearly uninformative.
- $\left(\sigma_\theta^2\right)^{(j+1)} \mid \boldsymbol{\theta}^{(j+1)}, \mathcal{D} \sim$ Inv-Gamma $\left(a + \frac{S}{2}, b + \frac{\|\boldsymbol{\theta}^{(j+1)}\|^2}{2}\right) \rightarrow$ Gibbs sampling i.e. directly sample from known distribution.



- 2-Block MCMC samples from the joint, $p(\boldsymbol{\theta}, \sigma_\theta^2 \mid \mathcal{D})$.
- By **alternately** drawing samples from the conditionals $p(\boldsymbol{\theta} \mid \sigma_\theta^2, \mathcal{D})$ and $p(\sigma_\theta^2 \mid \boldsymbol{\theta}, \mathcal{D})$.
- I.e. $p(\boldsymbol{\theta} \mid \sigma_\theta^2, \mathcal{D}) = \frac{p(\boldsymbol{\theta}, \sigma_\theta^2 \mid \mathcal{D})}{p(\sigma_\theta^2 \mid \mathcal{D})}$, when sampling $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta} \mid \sigma_\theta^2, \mathcal{D})$, σ_θ^2 is fixed $\implies p(\boldsymbol{\theta} \mid \sigma_\theta^2, \mathcal{D}) \propto p(\boldsymbol{\theta}, \sigma_\theta^2 \mid \mathcal{D})$.
- I.e. $p(\sigma_\theta^2 \mid \boldsymbol{\theta}, \mathcal{D}) = \frac{p(\boldsymbol{\theta}, \sigma_\theta^2 \mid \mathcal{D})}{p(\boldsymbol{\theta} \mid \mathcal{D})}$, when sampling σ_θ^2 from $p(\sigma_\theta^2 \mid \boldsymbol{\theta}, \mathcal{D})$, $\boldsymbol{\theta}$ is fixed $\implies p(\sigma_\theta^2 \mid \boldsymbol{\theta}, \mathcal{D}) \propto p(\boldsymbol{\theta}, \sigma_\theta^2 \mid \mathcal{D})$.
- Later we'll show $\nu \propto \frac{1}{\sigma_\theta^2} \implies \hat{\boldsymbol{\theta}}^{MAP}$ inherently has degree of regularization 'baked in' since $p(\boldsymbol{\theta} \mid \mathcal{D}) = \int p(\boldsymbol{\theta}, \sigma_\theta^2 \mid \mathcal{D})\, d\sigma_\theta^2 \rightarrow$ training data infers regularization!

- To improve mixing (reduced autocorrelation between samples) $\rightarrow$ better match the geometry of the stationary distribution with proposal distribution.
- Proposal for $\boldsymbol{\theta} \in \mathbb{R}^S$ at current iteration $j \rightarrow \mathbf{Q}_\theta \sim \mathcal{N}(\boldsymbol{\theta}^{(j)}, \boldsymbol{\Sigma}_j)$.
- $\boldsymbol{\Sigma}_j = \mathrm{Cov}\left(f\left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(j)}\right)\right) + \epsilon \mathbf{I}$.

$$f\left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(j)}\right) = \begin{cases} \left(\boldsymbol{\theta}^{(j-\delta\Delta)}, \boldsymbol{\theta}^{(j-\delta\Delta+\delta)}, \ldots, \boldsymbol{\theta}^{(j-\delta\Delta+(\Delta-1)\delta)}, \boldsymbol{\theta}^{(j)}\right), & \text{if } j > \delta\Delta \\ \left(\boldsymbol{\theta}^{(j-\lfloor \frac{j}{\delta} \rfloor \delta)}, \boldsymbol{\theta}^{(j-\lfloor \frac{j}{\delta} \rfloor \delta+\delta)}, \ldots, \boldsymbol{\theta}^{(j-\lfloor \frac{j}{\delta} \rfloor \delta+(\lfloor \frac{j}{\delta} \rfloor-1)\delta)}, \boldsymbol{\theta}^{(j)}\right), & \text{if } \delta < j \leq \delta\Delta \\ \left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(j)}\right), & \text{if } j \leq \delta. \end{cases}$$

- δ serves as a stride parameter $\rightarrow$ non-consecutive $\boldsymbol{\theta}$ used (thinning).
- $\Delta \rightarrow$ window of accepted proposals used (eases computation).
- $\boldsymbol{\Sigma}_j$ now asymmetric $\rightarrow$ use Adaptive MH only in burn-in.

Adaptive Scaling

- Want acceptance probability α toward theoretical optimum of ≈ 0.234 per Roberts et al. 1997.
- Scale $\boldsymbol{\Sigma_j} \implies (s^2)^{(j)}\, \boldsymbol{\Sigma_j}$.

$$(s^2)^{(j+1)} = (s^2)^{(j)} \times \exp\left(\gamma_j \cdot (\alpha_j - 0.234)\right).$$

- γ_j is a sequence of diminishing adaptation rates $\rightarrow \gamma_j = \frac{1}{j^k}$ per Roberts et al. 1997.
- $(s^2)^{(j)}\, \boldsymbol{\Sigma_j}$ now asymmetric $\rightarrow$ use adaptive scaling only in burn-in.

The Objective & Regularization

- L2 penalized objective: $\text{argmax}_{\boldsymbol{\theta}} \left(\text{Obj}(\boldsymbol{\theta}) - \nu \sum_{i=1}^{S} \theta_i^2 \right)$.
- For Gaussian prior: $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{o}, \sigma_\theta^2 \mathbf{I}_S) \rightarrow \log[p(\boldsymbol{\theta} \mid \mathcal{D})] \propto \log\left[p(\mathcal{D} \mid \boldsymbol{\theta})\right] - \frac{1}{2\sigma_\theta^2} \sum_{i=1}^{S} \theta_i^2$.
- If we assume $p(\mathcal{D} \mid \boldsymbol{\theta}) \propto \text{Obj}\,(\boldsymbol{\theta}) \rightarrow \log[p(\boldsymbol{\theta} \mid \mathcal{D})] \propto \text{Obj}\,(\boldsymbol{\theta}) - \nu \sum_{i=1}^{S} \theta_i^2, \quad \text{with } \nu \propto \frac{1}{\sigma_\theta^2}$.
- $\implies \hat{\boldsymbol{\theta}}^{MAP} = \text{argmax}_{\boldsymbol{\theta}}\, p(\boldsymbol{\theta} \mid \mathcal{D}) = \text{argmax}_{\boldsymbol{\theta}} \left(\text{Obj}(\boldsymbol{\theta}) - \nu \sum_{i=1}^{S} \theta_i^2 \right)$.

- **2-Block MCMC:** since we sample σ_θ^2 at each iteration $\implies$ L2 shrinkage $\nu^{(j)} \propto \frac{1}{(\sigma_\theta^2)^{(j)}}$ on each proposal $\boldsymbol{\theta}^*$.
- Marginal $p(\sigma_\theta^2 \mid \mathcal{D})$ reflects posterior uncertainty about shrinkage strength.
- Previously: fixed prior $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{o}, \sigma_\theta^2 \mathbf{I}_S)$.
- Now: hierarchical prior $\boldsymbol{\theta} \mid \sigma_\theta^2 \sim \mathcal{N}(\mathbf{o}, \sigma_\theta^2 \mathbf{I}_S)$ with $\sigma_\theta^2 \sim \text{Inv-Gamma}(a, b)$ $\implies$ marginal prior $\boldsymbol{\theta} \sim t_{2a}\left(\mathbf{o}, \frac{b}{a}\mathbf{I}_S\right)$.
- MAP estimator:

$$\hat{\boldsymbol{\theta}}^{MAP} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left(\text{Obj}(\boldsymbol{\theta}) - \left(a + \tfrac{S}{2}\right) \log\left(b + \tfrac{1}{2}\sum_{i=1}^{S} \theta_i^2\right) \right).$$

- Student-t prior $\approx$ Gaussian prior near o **but** much weaker shrinkage for large $|\theta_i|$.

- **Classical case:** Likelihood = joint density of data given parameters:

$$p(\mathcal{D} \mid \boldsymbol{\theta}) = f(x_1, \ldots, x_n \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i \mid \boldsymbol{\theta}).$$

- **Arbitrary objective:** No assumptions on data-generating model $\rightarrow$ only an objective $\mathrm{Obj}(\boldsymbol{\theta})$ to maximise (e.g., accuracy, score, ROI).
- **Idea:** Define a **pseudo-likelihood:**

$$p(\mathcal{D} \mid \boldsymbol{\theta}) \propto \mathrm{Obj}(\boldsymbol{\theta}).$$

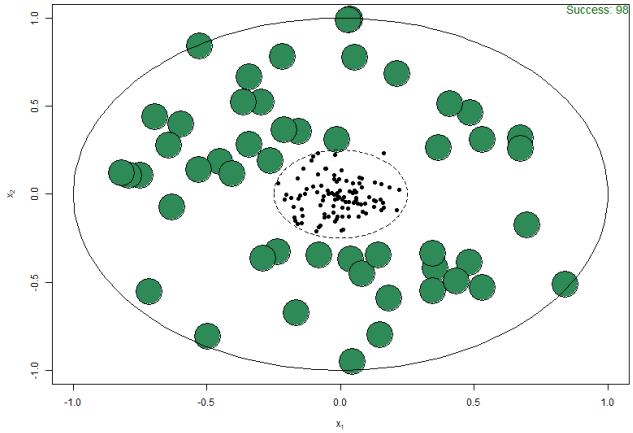
- Target the conditional **pseudo-posterior:**

$$p(\boldsymbol{\theta} \mid \sigma_\theta^2, \mathcal{D}) \propto p(\mathcal{D} \mid \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta} \mid \sigma_\theta^2).$$

  repurposing MCMC as a mode-seeking algorithm $\rightarrow$ sample around dominant mode of conditional.
- Acceptance guided by $\mathrm{Obj}(\boldsymbol{\theta})$ $\rightarrow$ higher objective $\Rightarrow$ higher acceptance.
- **Tempering:** $\left[p(\boldsymbol{\theta} \mid \mathcal{D})\right]^{\beta}$ sharpens sensitivity to $\mathrm{Obj}(\boldsymbol{\theta})$.

The Navigation Problem

- **Arena setup:** T drones navigate a 2D annular region bounded by inner radius R_{inner} and outer radius R_{outer}.
- **Objective:** Each drone attempts to exit the arena within K steps, with fixed step length δ.
- **Crash rule:** A drone crashes (and ceases navigation) if it comes within distance R_{crash} of any of the J orbiting obstacles.



- Denote: drone and obstacle coordinates $\left(\mathbf{x}_t^{(k)}, \mathbf{o}_j^{(k)}\right)$ and detection radius $R_{detection} = \boldsymbol{\phi}(\boldsymbol{\theta}_1)$.
- Network inputs for the t^{th} drone: $\boldsymbol{\Omega}\left(\mathbf{x}_t^{(k)}, \mathbf{o}_j^{(k)}, \boldsymbol{\phi}(\boldsymbol{\theta}_1)\right)$, defined as the sum of reciprocal Euclidean distances to obstacles within $R_{detection}$, capturing the caution level needed.
- Network outputs: position updates $\left(\tilde{x}_{t1}^{(k)}, \tilde{x}_{t2}^{(k)}\right)$ for the t^{th} drone at step k.
- Output activation: $\sigma_L(\cdot) = \tanh$, enabling movement in all planar directions.
- Full mapping:

$$(\mathbf{x}_t^{(k)}, \mathbf{o}_j^{(k)}, \boldsymbol{\theta}) \longrightarrow \mathbf{model}\Big(\boldsymbol{\Omega}\left(\mathbf{x}_t^{(k)}, \mathbf{o}_j^{(k)}, \boldsymbol{\phi}(\boldsymbol{\theta}_1)\right), \boldsymbol{\theta}_2\Big) \xrightarrow{\sigma_L(\cdot)} \left(\tilde{x}_{t1}^{(k)}, \tilde{x}_{t2}^{(k)}\right) \in [-1, 1]^2$$

  where $\boldsymbol{\theta}_1 \subset \boldsymbol{\theta}, \boldsymbol{\theta}_2 \subseteq \boldsymbol{\theta}$
- For a parameter configuration $\boldsymbol{\theta} \in \mathbb{R}^S$, where $\boldsymbol{\theta}_2$ denotes network weights and biases and $R_{detection} = \boldsymbol{\phi}(\boldsymbol{\theta}_1)$, record relative success frequency of the T drones, $k(\boldsymbol{\theta})$:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \ \frac{1}{T} k(\boldsymbol{\theta}).$$

- **Environment initialization:** At iteration k, obstacles are placed at coordinates $\mathbf{o}_j^{(k)}, j = 1, \ldots, J$, determined by angular frequencies ω_j, phase shifts ϕ_j, and orbital radii r_j. Also specifies T initial drone positions $\mathbf{x}_t^{(0)}, t = 1, \ldots, T$.
- **Training & testing setup:** Initialization is controlled by a training seed ω_0^{Train}; evaluation uses 1000 test initializations with seeds $\{\omega_j^{\text{Test}}\}_{j=1}^{1000}$.
- **Preliminary regularization assessment:** Estimate $\hat{\boldsymbol{\theta}}_\nu^{\text{GA}}$ using a GA with L2 regularization of strength $\nu \rightarrow$ apply it to a newly initialized environment.
- **Result:** No consistent pattern in success rates observed $\rightarrow$ performance driven mainly by whether $\hat{R}_{detection} \approx \mathbf{0.07}$.

	In-Sample	Out-of-Sample		
$\nu \times 10^6$	$k\left(\hat{\boldsymbol{\theta}}_\nu^{\text{GA}}\right)$	$\tilde{k}\left(\hat{\boldsymbol{\theta}}_\nu^{\text{GA}}\right)$	$\bar{k}\left(\hat{\boldsymbol{\theta}}_\nu^{\text{GA}}\right)$	$\hat{R}_{detection}$
0	100	**87.00**	**78.67**	**0.0726**
1	98	54.00	50.61	0.1056
2	100	15.50	29.77	0.1206
3	98	60.00	54.6	0.0983
4	100	84.00	73.90	0.1075
5	99	66.00	58.56	0.0962
6	98	70.50	61.09	0.1016
7	100	**86.00**	**78.47**	**0.0721**
8	100	**87.00**	**79.34**	**0.0724**
9	100	80.00	69.20	0.1161

Table: Number of successes $k\left(\hat{\boldsymbol{\theta}}_\nu^{\text{GA}}\right)$ for the in-sample initialization, median ($\tilde{k}$) and mean ($\bar{k}$) for the distributions of successes on the 1000 test initializations and estimated $\hat{R}_{detection}$, against varying ν for $T = 100$ and $R_{crash} = 0.05$.

- Recall: pseudo-likelihood $\rightarrow p(\mathcal{D} \mid \boldsymbol{\theta}) \propto \text{Obj}(\boldsymbol{\theta})$.
- $\rightarrow$ repurposed MCMC as a mode-seeking algorithm $\implies$ samples concentrated around the dominant mode of conditional $p(\boldsymbol{\theta} \mid \sigma_\theta^2, \mathcal{D})$.
- Employ three variants of pseudo-likelihood, differing in their degree of sharpness.
- Recall: MH operates on ratios of likelihoods $\rightarrow$ rates of change more important than absolute scale.

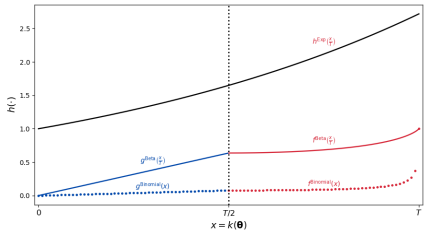


Figure: $h^{\text{Binomial}}(x)$, $h^{\text{Beta}}\left(\frac{x}{T}\right)$ and $h^{\text{Exp}}\left(\frac{x}{T}\right)$ for $x = k(\boldsymbol{\theta}) \in \{0, 1, \ldots, T\}$.

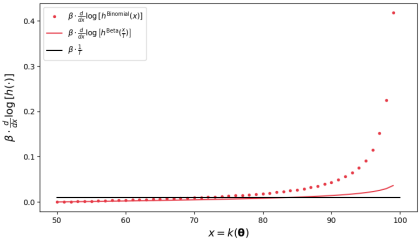


Figure: $\beta \cdot \frac{d}{dx} \log\left[h\left(\cdot\right)\right]$ for $x = k(\boldsymbol{\theta}) \in \{\frac{T}{2}, \frac{T}{2} + 1, \ldots, T\}$ for $\beta = 1$.

- Sharpest likelihood = best in-sample performance.
- Out-of-sample performance remains highly sensitive to whether $\hat{R}_{detection} \approx 0.07$.
- ESS > 100 $\implies$ satisfactory mixing (expected, since adaptive MH used).

	In-Sample	Out-of-Sample			
Likelihood Type	$k\left(\hat{\boldsymbol{\theta}}_{\nu}^{\mathrm{GA}}\right)$	$\tilde{k}\left(\hat{\boldsymbol{\theta}}_{\nu}^{\mathrm{GA}}\right)$	$\bar{k}\left(\hat{\boldsymbol{\theta}}_{\nu}^{\mathrm{GA}}\right)$	$\hat{R}_{detection}$	ESS
Binomial-based	**100**	13.00	24.45	0.3912	179.8550
Beta-based	50	8.00	12.72	0.5175	170.0205
Exponential-based	52	12.00	17.25	0.5601	317.2764

Table: Number of successes $k\left(\hat{\boldsymbol{\theta}}_{\nu}^{\mathrm{GA}}\right)$ for the in-sample initialization, summary statistics for the distributions of successes on the 1000 test initializations and estimated $\hat{R}_{detection}$, for the three likelihood types.

- Indefinite contraction of $\|\boldsymbol{\theta}^{(j)}\| \to$ for flat pseudo-likelihoods, it can be shown:

$$\log(\alpha_\theta) = \min\left(\frac{1}{2\left(\sigma_\theta^2\right)^{(j)}}\left(\|\boldsymbol{\theta}^{(j)}\|^2 - \|\boldsymbol{\theta}^*\|^2\right), 0\right).$$

- Different pseudo-likelihoods $\implies$ different $p(\sigma_\theta^2 \mid \mathcal{D}) \implies$ different ν inferred to MAP estimates.

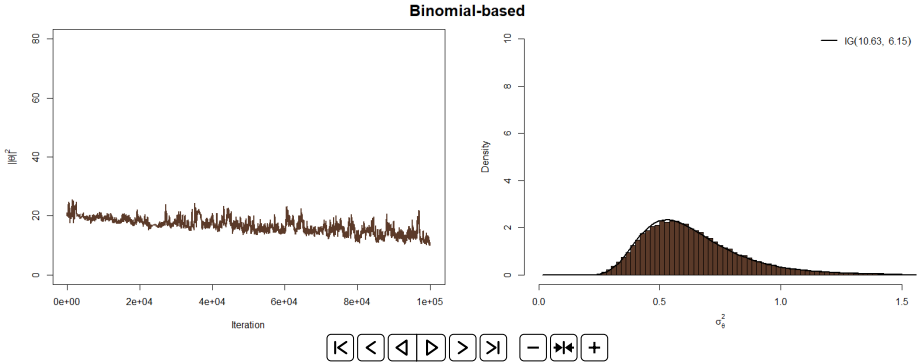
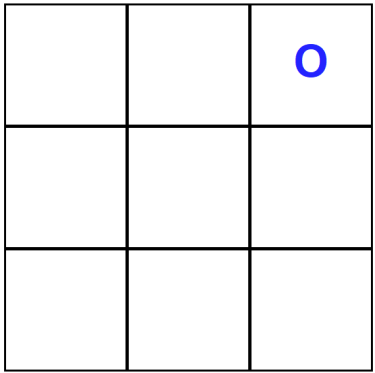


Figure: $\|\boldsymbol{\theta}^{(j)}\|^2$ for $j = 1, \ldots 100,000$ (no burn-in) in the left panel, with distribution of marginal $\sigma_\theta^2 \mid \mathcal{D}$ (using 20,000 burn-in for binomial-based and 60,000 burn-in for beta- and exponential-based) for the three likelihood types in the right panel (all plots use the same scale).

The Tic-Tac-Toe Problem

- Classical tic-tac-toe game is a two-player, deterministic, turn-based game in which the player and opponent alternately place their respective tokens X and O on a 3×3 grid.
- The objective of the player (who plays first) is to be the first to align three of their tokens consecutively in a row, column, or diagonal, and likewise for the opponent.
- Agent trained to play against opponent whose decisions are random.



XO: Feature engineering

- Represent 3×3 board as matrix $\mathbf{M}_{3\times 3}$, vectorized row-wise into $\mathbf{m}$.
- Terminal condition: compute $\mathbf{m}'\mathbf{S}_{9\times 8}$, where each column of $\mathbf{S}$ encodes a row, column, or diagonal.
  - If any entry $= +3 \implies X$ wins.
  - If any entry $= -3 \implies O$ wins.
  - Draws are identified from $f(\mathbf{m}'\mathbf{S})$ when no moves remain.
- Network inputs at time τ_k: (i) current board $\mathbf{m}^{\tau_k - 1}$, (ii) evaluation state $\left(\mathbf{m}^{\tau_k-1}\right)'\mathbf{S}$.
- Network output: action a^* from available board positions: $\mathcal{A}_{\tau_k} = \{i \in \{1, \ldots, 9\} : m_i^{\tau_k - 1} = 0\}$.
- Output activation:

$$\sigma_L(a \mid \mathbf{m}^{\tau_k-1}, \boldsymbol{\theta}) = \frac{\exp(\ell_a)}{\sum_{a' \in \mathcal{A}_{\tau_k}} \exp(\ell_{a'})},$$

  with action a^* corresponding to action with the highest probability, $a^* = \operatorname{argmax}_{a \in \mathcal{A}_{\tau_k}} \sigma_L(a \mid \cdot)$
- Objective: maximize win frequency of player O (encoded $+1$) over K games:

$$\hat{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{\theta}} \frac{1}{K} \sum_{k=1}^{K} \mathbb{I}\left(\rho_{T_k}(\boldsymbol{\theta}) = +1\right).$$

XO: The effects of regularization

- **Training setup:** Simulate 100 games against a random-policy opponent, using seed values $\{\omega_i^{\text{Train}}\}_{i=1}^{100}$.
- Learned solution $\hat{\boldsymbol{\theta}}$ influences opponent trajectories $\implies$ each training run corresponds to a distinct environment.
- $\implies$ The optimization surface varies across training iterations.
- **Test setup:** Evaluate on 10,000 games with disjoint seeds $\{\omega_i^{\text{Test}}\}$, ensuring $\omega_i^{\text{Test}} \notin \{\omega_j^{\text{Train}}\}$.
- Question: Are all test opponent trajectories truly unseen compared to training?
- **Result:** Performance decreases (in- and out-of-sample) as regularization strength ν increases.
  - Excessive regularization $\implies$ underfitting.
  - Best out-of-sample % for $\nu \in [10^{-6}, 10^{-4}]$ (beats random vs random hit-rate $\approx 57\%$).

ν	In-Sample (%)	Out-of-Sample (%)
10^{-6}	99	83.92
10^{-5}	99	74.84
10^{-4}	98	66.96
10^{-3}	93	59.96
10^{-2}	85	59.36
10^{-1}	74	69.10
1	52	46.04

Table: Normalized win percentage of O tokens across training ($K = 100$) and test ($K = 10{,}000$) games, for varying regularization strengths ν.

- **Recall:** pseudo-likelihood tempering $\left[p(\boldsymbol{\theta} \mid \mathcal{D})\right]^{\beta} \implies$ sharpens sensitivity to Obj($\boldsymbol{\theta}$).
- β controls the trade-off between exploration and exploitation.
- As β increases, in-sample O wins improve:
  - β amplifies the likelihood ratio: $\left(\frac{p(\mathcal{D}|\boldsymbol{\theta}^*)}{p(\mathcal{D}|\boldsymbol{\theta}^{(j)})}\right)^{\beta}$.
  - Larger $\beta \implies$ chain favors higher-Obj proposals.
- Empirically: $\beta \geq 100 \implies$ convergence to same dominant mode.
- Caution: overly large $\beta \implies$ Markov chain stagnation.

Sharpness β	In-Sample (%)	Out-of-Sample (%)
0.1	56	50.08
1	70	59.86
10	63	56.68
50	86	72.98
100	98	72.62
1000	96	68.58

Table: Normalized O win rate across training ($K = 100$) and test ($K = 10{,}000$) games for varying sharpness β.

- For low β (flat likelihoods) $\implies$ indefinite contraction of $\|\boldsymbol{\theta}^{(j)}\|$.
- Different $\beta's$ $\implies$ different $p(\sigma_\theta^2 \mid \mathcal{D})$ $\implies$ different ν inferred to MAP estimates.

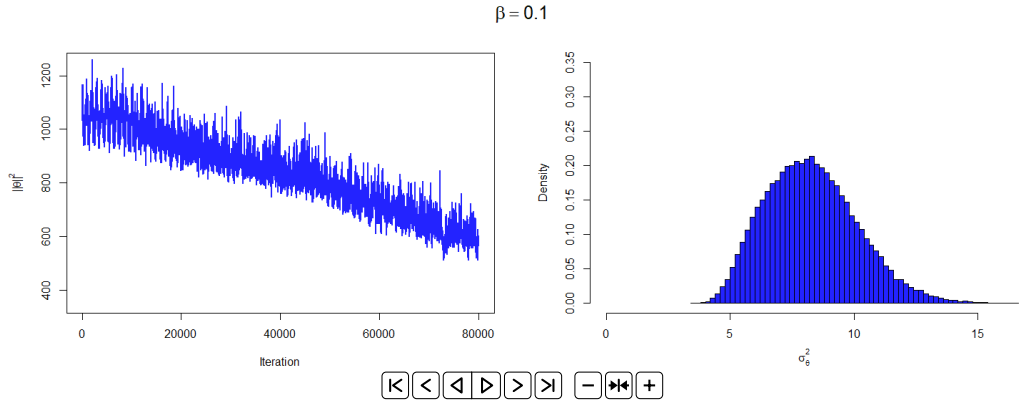


Figure: $\|\boldsymbol{\theta}^{(j)}\|^2$ for $j = 1, \ldots 80,000$ (post burn-in) with distribution of marginal $\sigma_\theta^2 \mid \mathcal{D}$ for varying likelihood sharpness β.

- Markov chain initialized with variance $\sigma^2_{\text{Init}} \to \boldsymbol{\theta}^{(1)} \sim \mathcal{N}\left(\mathbf{o}_{S\times 1},\ \sigma^2_{\text{Init}}\mathbf{I}_S\right)$.
- Marginals $p(\sigma^2_\theta \mid \mathcal{D})$ follow an inverse-gamma distribution $\to$ shape parameters remain approximately constant, rate parameters grow with $\sigma^2_{\text{Init}} \implies$ both the mean and variance shift upward.

σ^2_{Init}	In-Sample (%)	Out-of-Sample (%)	Shape	Rate
0.1	92	71.18	41.75	3.14
1	88	65.40	46.23	41.31
10	98	72.62	44.93	413.30
100	92	68.92	47.28	5704.89

Table: The normalized number of O wins, as a percentage for in-sample ($K = 100$) and out-of-sample ($K = 10,000$) sets across various initial variances σ^2_{Init} accompanied by shape and rate parameters of marginal $\sigma^2_\theta \mid \mathcal{D} \sim$ Inv-Gamma for likelihood sharpness $\beta = 100$.

- **Motivation:** Introduced uncertainty about the dispersion parameter $\sigma_\theta^2 \propto \frac{1}{\nu} \rightarrow$ training set informs ν.
- **Hyperparameters:** pseudo-likelihood form, likelihood sharpness β, and initial variance σ_{Init}^2 are user-specified $\rightarrow$ strongly influences marginal $p(\sigma_\theta^2 \mid \mathcal{D})$.
- **Limitation:** Extent of regularization is not fully data-driven.
- **Conclusion:** Using a hierarchical Bayesian model solely to infer σ_θ^2 may be unnecessary $\rightarrow$ instead fix σ_θ^2 (and thus ν) directly.
  - In this sense, the two-block MCMC procedure acts as the *user* inferring a specific regularization strength but with "extra steps".

Go further

References I

Roberts, Gareth O et al. (1997). "Weak convergence and optimal scaling of random walk Metropolis algorithms". In: *The Annals of Applied Probability* 7.1, pp. 110–120.