A Proofs of clustering speedup

Proof of Lemma 1. Given portions x and y of red and blue balls, resp., in the left urn, consider the $2 \times 2 \times 2$ possible Gibbs moves: remove red/blue from left/right urn and replace in left/right urn. For large data size N, data size changes very little after a single removal, so the add and remove steps decouple into differentials dx_{rem}, dy_{rem}, dx_{add}, and dy_{add}. Compute the probabilities of each move; then compute the mean and variance in x and, by red-blue symmetry, y:

$$\mathbb{E}[dx_{\text{rem}}] = \frac{1}{N}\frac{1-2x}{2}$$

$$\mathbb{V}[dx_{\text{rem}}] = \frac{1}{N^2}\frac{x(1-x)}{r}$$

$$\mathbb{E}[dx_{\text{add}}] = \frac{1}{2N}\frac{n_2 r_1 - n_1 r_2}{n_2 r_1 + n_1 r_2} + \frac{\alpha}{N^2}\frac{2n_1}{xr + (1-x)l}$$

$$\mathbb{V}[dx_{\text{add}}] = \frac{1}{N^2}\frac{r\ n_2 r_1\ n_1 r_2}{(n_2 r_1 + n_1 r_2)^2} + \frac{\alpha}{N^3}\left[\frac{n_1 n_2}{(xn_1 + (1-x)n_2)^2} + \frac{2}{xn_1 + (1-x)n_2}\right]$$

with lower-case intrinsic quantities defined as

$$r_1 = (\#\text{red on left})/N \qquad r_2 = (\#\text{red on right})/N$$
$$b_1 = (\#\text{blue on left})/N \qquad b_2 = (\#\text{blue on right})/N$$
$$n_1 = r_1 + b_1 \qquad n_2 = r_2 + b_2$$

These moments comprise the N-scaled Fokker-Planck coefficients

$$f = N\begin{bmatrix}\mathbb{E}[dx_{\text{rem}} + dx_{\text{add}}] \\ \mathbb{E}[dy_{\text{rem}} + dy_{\text{add}}]\end{bmatrix}$$

$$D = N^2\begin{bmatrix}\mathbb{V}[dx_{\text{rem}} + dx_{\text{add}}] & 0 \\ 0 & \mathbb{V}[dy_{\text{rem}} + dy_{\text{add}}]\end{bmatrix}$$

By inspection these depend only on intrinsic quantities and the scaled hyperparameter $\frac{\alpha}{N}$. □

Proof of Theorem 2. At fixed error bound ϵ, the continuous dynamics is within ϵ of true dynamics by data size, say, N_ϵ. Thus at large data sizes, the MCMC dynamics is linear and mixing time is $T_{\text{cold}} = O\left(N^2 \log(\epsilon)\right)$. In a subsample annealing schedule $\beta(t) = t/T$, the subsample annealing dynamics at subsamples larger than N_ϵ is approximately time-scaled versions of the dynamics at full size N, so the effective schedule length is

$$T_{\text{eff}} = \int_{\frac{N_\epsilon}{N}T}^{T} \frac{dt}{\beta(t)^2} = \frac{1}{N}\left[\frac{1}{N_\epsilon} - \frac{1}{N}\right] \geq \frac{2}{NN_\epsilon}$$

Since effective time is inverse in data size, annealing mixes in time $T_{\text{anneal}} = O\left(N \log(\epsilon)\right)$. □

B Proofs of bimodal speedup

Proof of Lemma 3. Consider a two-state system $\mathbf{x} = [x, 1-x]^T$ at energy levels $[N\gamma, 0]$. The steady-state solution at temperature β should be $\pi_beta := [\sigma(\beta\gamma N), \sigma(\beta\gamma N)]$, where $\sigma(t) = \frac{1}{1+\exp(-t)}$ is the logistic sigmoid function. In continuous time mixing, we think of the state briefly jumping on to an energy barrier of height $\beta\delta N$ then jumping back down according to π_β. If the rate of jumping up to energy $\beta\delta N$ is $\exp(-\beta\delta N)$, then the dynamics is:

$$\frac{d\mathbf{x}}{dt} = \exp(-\beta\delta N)\left[\begin{bmatrix}\sigma(\beta\gamma N) & \sigma(\beta\gamma N) \\ \sigma(-\beta\gamma N) & \sigma(-\beta\gamma N)\end{bmatrix} - \mathbf{I}\right]\mathbf{x}$$

The first coordinate x determines the state; expanding yields Equation 1. □

Proof of Theorem 4. In this binary system the TVD of state x from truth is $|x - x_{\text{true}}| = |x - \sigma(\gamma N)|$. Now we seek asymptotic lower bounds on T guaranteeing TVD$< \epsilon$. To prove (a) observe that in cold inference ($\beta = 1$), the system is linear homogeneous with eigenvalue $\exp(-N\delta)$. To prove (b) we transform from time coordinates t to "natural" coordinates

$$\tau = \exp\left(\frac{T}{N\delta}\left[\exp\left(-\frac{N\delta t}{T}\right) - \exp\left(-N\delta\right)\right]\right),$$

where, assuming worst-case initial condition $x(0) = 0$, the final state x is a uniform integral

$$x = \int_0^1 \sigma(\beta(\tau)N\gamma)\,d\tau$$

involving the transformed annealing schedule

$$\beta(\tau) = \frac{-1}{N\delta}\log\left(\exp(-N\delta) - \frac{N\delta}{T}\log(\tau)\right).$$

Using the inequality $\sigma(\gamma) - \sigma(\beta\gamma) \leq \exp(-\beta\gamma)$, we can bound error by

$$\text{TVD} < \int_0^1 \exp(-\beta(\tau)N\gamma)\,d\tau \qquad (2)$$

Since the integrand $\exp(-\beta(\tau)N\gamma)$ is bounded in $(0,1)$, and $\beta(\tau)$ is increasing, Equation 2 holds if T is chosen large enough that $\exp(-\beta(\epsilon/2) > 1$, for example if

$$T > \frac{N\delta\log\left(\frac{2}{\epsilon}\right)}{\left(\frac{\epsilon}{2}\right)^{\frac{\delta}{\gamma}} - \exp(-N\delta)}$$

or more conservatively, for any $K > 1$, and sufficiently large N,

$$T > KN\delta\log\left(\frac{2}{\epsilon}\right)\left(\frac{2}{\epsilon}\right)^{\frac{\delta}{\gamma}},$$

whence the asymptotic bound. □