STAT 547E: Scalable Sampling

Assignment 1

University of British Columbia
Department of Statistics

Due: March 11, 2026

Instructions. Submit your solutions as a single PDF containing both derivations and code/figures. For computational questions, include well-commented code and clearly labelled plots.

Setup. Let $\pi(x) = \gamma(x)/Z$ be a target distribution on $\mathbb{X}$, where $\gamma(x) > 0$ is an unnormalised density we can evaluate pointwise and $Z = \int \gamma \, \mathrm{d}x$ is the unknown normalising constant. Let η be a *reference distribution* on $\mathbb{X}$ from which we can sample and whose density we can evaluate. Define the weight function

$$w(x) = \frac{\gamma(x)}{\eta(x)},$$

and for a pair $(x, y) \in \mathbb{X}^2$ the swap acceptance probability

$$\alpha(x, y) = \min\left\{1, \ \frac{w(y)}{w(x)}\right\}.$$

Let $K(x, \mathrm{d}x')$ be a π-invariant Markov kernel on $\mathbb{X}$ that we can efficiently simulate. Consider the Stabilised MCMC algorithm below.

Algorithm 1 Stabilised MCMC

Initialise: draw $X_0 \sim \eta, \ Y_0 \sim \eta$
for $t = 1, \ldots, T$ **do**
  Draw $X_t' \sim K(X_{t-1}, \mathrm{d}x)$
  Draw $Y_t' \sim \eta$
  Compute $\alpha_t \leftarrow \alpha(X_t'.Y_t')$
  Draw $U_t \sim \mathrm{Uniform}(0, 1)$
  if $U_t \leq \alpha_t$ **then**
    $(X_t, Y_t) \leftarrow (Y_t', X_t')$
  else
    $(X_t, Y_t) \leftarrow (X_t', Y_t')$
  end if
end for

Define the empirical swap rejection rate and the running IS estimate of Z:

$$\hat{r}_T = 1 - \frac{1}{T} \sum_{t=1}^{T} \alpha_t, \qquad \hat{Z}_T = \frac{1}{T} \sum_{t=1}^{T} w(Y_t).$$

Problem 1. (a) Write down the joint Markov kernel $\bar{K}\big((x,y), \mathrm{d}(x',y')\big)$ and show that $\bar{K}$ is $\bar{\pi}$-invariant, where $\bar{\pi} = \pi \otimes \eta$.

(b) Show that the stationary swap acceptance rate satisfies

$$\bar{\pi}[\alpha] := \mathbb{E}_{\bar{\pi}}[\alpha(X,Y)] = 1 - \|\pi \otimes \eta - \eta \otimes \pi\|_{\mathrm{TV}}.$$

Problem 2. For each statement below, either prove it or provide a counterexample.

(a) If K is π-irreducible then $\bar{K}$ is $\bar{\pi}$-irreducible.

(b) If K is not π-irreducible then $\bar{K}$ is not $\bar{\pi}$-irreducible.

(c) If K is π-reversible then $\bar{K}$ is $\bar{\pi}$-reversible.

Problem 3. Suppose $\pi = \mathcal{N}(\mu, \sigma^2)$ and $\eta = \mathcal{N}(0, \sigma^2)$, and let $\rho = \mu/\sigma$.

(a) Show that $\hat{r}_T$ converges almost surely to

$$r(\rho) = 2\Phi\left(\frac{\rho}{\sqrt{2}}\right) - 1,$$

where Φ is the standard normal CDF.

Hint: the total variation between two Gaussians with the same covariance Σ *and means* a, b *is*

$$\big\|\mathcal{N}(a,\Sigma) - \mathcal{N}(b,\Sigma)\big\|_{\mathrm{TV}} = 2\Phi\big(\tfrac{1}{2}\|a-b\|_{\Sigma^{-1}}\big) - 1, \qquad \|a-b\|_{\Sigma^{-1}} := \sqrt{(a-b)^\top \Sigma^{-1}(a-b)}.$$

(b) Show that $\hat{Z}_T$ is unbiased with variance

$$\mathbb{E}[\hat{Z}_T] = 1, \qquad \mathbb{V}[\hat{Z}_T] = \frac{\exp(\rho^2) - 1}{T}.$$

(c) Compare how the expected waiting time $r(\rho)/(1 - r(\rho))$ and the variance $T\mathbb{V}[\hat{Z}_T]$ scale as $\rho \to 0$ and $\rho \to \infty$. Do they deteriorate at the same rate?

Problem 4. Let π be the **GMM-40** distribution: an equally-weighted mixture of 40 bivariate Gaussians,

$$\pi(x) = \frac{1}{40}\sum_{k=1}^{40} \mathcal{N}(x \mid \mu_k,\, 0.1\, I_2),$$

with means μ_k arranged on a $5{\times}8$ grid of spacing $\Delta = 3$, i.e. $\mu_k \in \{0,3,6,9,12\} \times \{0,3,6,9,12,15,18,21\}$.

Implement and run the following samplers for $T = 10{,}000$ iterations:

(i) Random Walk Metropolis (RWM) with isotropic proposal standard deviation $\sigma_{\mathrm{prop}} \in \{0.3,\ 1,\ 3\}$.

(ii) Metropolis-Adjusted Langevin Algorithm (MALA) with step size $\epsilon \in \{0.05,\ 0.1,\ 0.5\}$.

For each sampler and tuning choice, report: (a) a scatter plot of samples overlaid on contours of π; (b) a traceplot of the x_1-component; (c) the ESS of the x_1-component. Comment on how the proposal scale affects performance.

Problem 5. Apply the stabilised MCMC algorithm to the GMM-40 target from Problem 4, using both the RWM ($\sigma_{\text{prop}} = 1$) and MALA ($\epsilon = 0.1$) local kernels.

(a) Run with reference $\eta = \mathcal{N}(0, \sigma_\eta^2 I_2)$, $\sigma_\eta = 10$, and $T = 10{,}000$ iterations. Report: (i) scatter plots of (X_t) and (Y_t) overlaid on contours of π; (ii) traceplots of the x_1-component; (iii) the final ESS of x_1 and the empirical rejection rate $\hat{r}_T$.

(b) Repeat for $\sigma_\eta \in \{2^{-5}, \dots, 2^5\}$ and plot the final ESS and $\hat{r}_T$ as functions of σ_η. Discuss how the reference distribution influences performance.

Compare throughout with your results from Problem 4.