STAT 547E: Scalable Sampling

Assignment 2

University of British Columbia
Department of Statistics

Due: April 8, 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. A starter notebook is provided on the course page.

Setup. Let $\pi_\beta = \gamma_\beta / Z_\beta$ be an annealing path indexed by $\beta \in [0, 1]$, where γ_β is an unnormalised density we can evaluate pointwise. We interpolate between a tractable reference $\pi_0 = \eta$ and the target $\pi_1 = \pi$ via a schedule $0 = \beta_0 < \beta_1 < \cdots < \beta_T = 1$, using N particles and a local Markov kernel K_{β_t} leaving π_{β_t} invariant. The incremental weight of particle n at step t is $g_t^n = \gamma_{\beta_t}(X_{t-1}^n)/\gamma_{\beta_{t-1}}(X_{t-1}^n)$.

Algorithm 1 Annealed Importance Sampling (AIS)

```
1: Sample X_0^n ~ η and set w_0^n ← 1 for n = 1, ..., N.
2: for t = 1, ..., T do
3:     Move: sample X_t^n ~ K_{β_t}(X_{t-1}^n, ·) for each n.
4:     Reweight: w_t^n ← w_{t-1}^n · g_t^n for each n.
5: end for
6: return Ẑ_AIS = (1/N) Σ_{n=1}^N w_T^n.
```

Algorithm 2 Sequential Importance Resampling (SIR)

```
1: Sample X_0^n ~ η for n = 1, ..., N.
2: for t = 1, ..., T do
3:     Move: sample X_t^n ~ K_{β_t}(X_{t-1}^n, ·) for each n.
4:     Reweight: compute g_t^n for each n.
5:     Resample: draw N indices by systematic resampling with weights g_t^n; accumulate Ẑ *= (1/N) Σ_n g_t^n.
6: end for
7: return Ẑ_SIR = Π_{t=1}^T (1/N) Σ_{n=1}^N g_t^n.
```

Both estimators are unbiased for Z, i.e. $\mathbb{E}[\hat{Z}_{\text{AIS}}] = \mathbb{E}[\hat{Z}_{\text{SIR}}] = Z$.

Problem 1. The *Efficient Local Exploration* (ELE) assumption states that the weights g_t^n are independent in n and t, with marginal law $g_t^n \overset{d}{=} g_t(X_{t-1})$, $X_{t-1} \sim \pi_{\beta_{t-1}}$.

(a) Show that under the ELE assumption,

$$\mathbb{V}\left[\frac{\hat{Z}_{\text{AIS}}}{Z}\right] = \frac{\exp\left(\sum_{t=1}^{T} D_t\right) - 1}{N}, \qquad \mathbb{V}\left[\frac{\hat{Z}_{\text{SIR}}}{Z}\right] = \prod_{t=1}^{T}\left(1 + \frac{\exp(D_t) - 1}{N}\right) - 1,$$

where $D_t = \log\left(1 + \chi^2(\pi_{\beta_t} \| \pi_{\beta_{t-1}})\right)$,

$$\chi^2(\pi' \| \pi) = \mathbb{E}_\pi\left[\left(\frac{\mathrm{d}\pi'}{\mathrm{d}\pi} - 1\right)^2\right].$$

Hint. Under ELE, $\chi^2(\pi_{\beta_t} \| \pi_{\beta_{t-1}})$ equals the variance of the normalised incremental weight g_t^n / z_t under $\pi_{\beta_{t-1}}$, where $z_t = \mathbb{E}_{\pi_{\beta_{t-1}}}[g_t^n]$.

(b) Let $\varphi : [0,1] \to [0,1]$ be a strictly increasing differentiable function with $\varphi(0) = 0$, $\varphi(1) = 1$, and set $\beta_t = \varphi(t/T)$. Define the *kinetic energy*

$$E[\varphi] = \int_0^1 I(\varphi(u))\, \dot{\varphi}(u)^2 \, \mathrm{d}u, \qquad I(\beta) = \mathbb{E}_{\pi_\beta}\left[\left(\tfrac{d}{d\beta} \log \pi_\beta\right)^2\right].$$

You may assume all quantities are sufficiently regular to justify interchange of limits and integrals. Show that

$$\lim_{T\to\infty} T\, \mathbb{V}\left[\frac{\hat{Z}_{\text{AIS}}}{Z}\right] = \lim_{T\to\infty} \mathbb{V}\left[\frac{\hat{Z}_{\text{SIR}}}{Z}\right] = \frac{E[\varphi]}{N}.$$

(c) Let $\Lambda = \int_0^1 \sqrt{I(\beta)}\, \mathrm{d}\beta$ be the Fisher–Rao length of the path. Prove that $E[\varphi] \geq \Lambda^2$, with equality if and only if

$$\varphi(u) = F^{-1}(\Lambda\, u), \qquad F(\beta) = \int_0^\beta \sqrt{I(\beta')}\, \mathrm{d}\beta'.$$

Taking parts (a)–(c) together, explain the algorithmic importance of these results for designing AIS/SIR in practice.

Problem 2. (a) Show that the linear path between $\mathcal{N}(\mu_0, \sigma^2)$ and $\mathcal{N}(\mu_1, \sigma^2)$ satisfies $\Lambda = z$, where $z = |\mu_1 - \mu_0|/\sigma$.

(b) Show that the linear path between $\mathcal{N}(\mu, \sigma_0^2)$ and $\mathcal{N}(\mu, \sigma_1^2)$ has length

$$\Lambda = \sqrt{2}\left|\log \frac{\sigma_0}{\sigma_1}\right|.$$

(c) Use parts (a) and (b) to construct an annealing path between $\mathcal{N}(\mu_0, \sigma^2)$ and $\mathcal{N}(\mu_1, \sigma^2)$ such that $\Lambda = O(\log |z|)$ as $z \to \infty$.

(d) Explain the algorithmic significance of this construction.

Problem 3. Fix $\mu = 10$, $N = 100$, and the linear path between $\mathcal{N}(0, 1)$ and $\mathcal{N}(\mu, 1)$ with uniform schedule $\beta_t = t/T$ (optimal by Problem 1(c)); note $Z = 1$. The local kernel K_β consists of k steps of Random Walk Metropolis targeting π_β with proposal standard deviation 0.5; $k = \infty$ denotes exact sampling (ELE baseline). Run $M = 200$ replicates on a log-spaced grid of T values and plot mean $\pm$ std of $\log \hat{Z}$.

(a) For AIS, plot the mean $\pm$ std band of $\log \hat{Z}$ for $k \in \{1, 5, 20, 100, \infty\}$ on the same axes. Mark vertical lines at $T = \Lambda$ and $T = \Lambda^2$. At what k does the band converge to the ELE baseline? By what factor in T does $k = 1$ inflate the effective schedule size?

(b) Repeat for SIR (systematic resampling at every step). Does resampling change the k at which ELE kicks in? At small k, is the SIR band wider or narrower than the AIS band?

(c) Fix budget $B = T \times k = 2000$. Compare:

- **Many steps, poor mixing:** $k = 1$, $T = 2000$.
- **Few steps, good mixing:** $k = 20$, $T = 100$.

For both AIS and SIR, plot both strategies alongside the ELE baseline ($k = \infty$, $T = 100$). Which strategy wins, and does the answer differ between AIS and SIR?