4M24 Computational Statistics and Machine Learning, 2026

Exam Question 1

We wish to estimate

$$\mathbb{E}_p[f(X)] = \int_{\mathbb{R}^d} f(x)p(x)\,dx,$$

where $p(x) = \mathcal{N}(0, I_d)$, using importance sampling with proposal $q(x) = \mathcal{N}(1_d, I_d)$ and importance weights $w(x) = \frac{p(x)}{q(x)}$.

Solution

(a) Form of the Importance Weights

The densities are

$$p(x) = (2\pi)^{-d/2} \exp\left(-\tfrac{1}{2}\|x\|^2\right), \quad q(x) = (2\pi)^{-d/2} \exp\left(-\tfrac{1}{2}\|x - 1_d\|^2\right).$$

Therefore,

$$w(x) = \frac{p(x)}{q(x)} = \exp\left(-\tfrac{1}{2}\|x\|^2 + \tfrac{1}{2}\|x - 1_d\|^2\right).$$

Expanding,

$$\|x - 1_d\|^2 = x^\top x - 2\,1_d^\top x + d,$$

so that

$$\log w(x) = -\tfrac{1}{2}x^\top x + \tfrac{1}{2}(x^\top x - 2\,1_d^\top x + d) = -\,1_d^\top x + \tfrac{d}{2}.$$

Thus

$$\boxed{w(x) = \exp\left(\tfrac{d}{2} - 1_d^\top x\right)}.$$

At the proposal mean $x = 1_d$,

$$w(1_d) = \exp\left(\tfrac{d}{2} - d\right) = \exp\left(-\tfrac{d}{2}\right).$$

Hence the weights decay like

$$w(x) \approx \exp(-\alpha d), \quad \alpha = \tfrac{1}{2}.$$

Marks available 30%

(b) Effective Sample Size (ESS)

The ESS for N samples is

$$\text{ESS} = N \frac{\left(\mathbb{E}_q[w(X)]\right)^2}{\mathbb{E}_q[w(X)^2]}.$$

First moment.

$$\mathbb{E}_q[w(X)] = \int \frac{p(x)}{q(x)} q(x)\, dx = \int p(x)\, dx = 1.$$

Second moment.

$$\mathbb{E}_q[w(X)^2] = \int \frac{p(x)^2}{q(x)}\, dx.$$

Now

$$p(x)^2 = (2\pi)^{-d} \exp(-\|x\|^2), \quad q(x) = (2\pi)^{-d/2} \exp\left(-\tfrac{1}{2}\|x - 1_d\|^2\right).$$

Thus

$$\frac{p(x)^2}{q(x)} = (2\pi)^{-d/2} \exp\Big(- \|x\|^2 + \tfrac{1}{2}\|x - 1_d\|^2 \Big).$$

Simplify:

$$-\|x\|^2 + \tfrac{1}{2}\|x - 1_d\|^2 = -\tfrac{1}{2}\|x\|^2 - 1_d^\top x + \tfrac{d}{2}.$$

So

$$\mathbb{E}_q[w^2] = (2\pi)^{-d/2} e^{d/2} \int_{\mathbb{R}^d} \exp\Big(- \tfrac{1}{2}\|x\|^2 - 1_d^\top x \Big)\, dx.$$

Completing the square:

$$-\tfrac{1}{2}\|x\|^2 - 1_d^\top x = -\tfrac{1}{2}\|x + 1_d\|^2 + \tfrac{d}{2}.$$

Thus

$$\mathbb{E}_q[w^2] = (2\pi)^{-d/2} e^{d} \int_{\mathbb{R}^d} \exp\Big(- \tfrac{1}{2}\|x + 1_d\|^2 \Big)\, dx.$$

Changing variable $y = x + 1_d$:

$$\int_{\mathbb{R}^d} \exp\Big(- \tfrac{1}{2}\|y\|^2 \Big)\, dy = (2\pi)^{d/2}.$$

Therefore

$$\boxed{\mathbb{E}_q[w^2] = e^d.}$$

ESS.

$$\mathrm{ESS} = N \cdot \frac{1^2}{e^d} = \boxed{Ne^{-d}}.$$

Marks available 40%

(c) Implications as $d \to \infty$

- From (a), typical importance weights decay exponentially in d, e.g. $w(1_d) = e^{-d/2}$.
- From (b), $\mathrm{ESS} = Ne^{-d}$, so to maintain a constant ESS one requires $N \sim e^d$ samples.
- Therefore, in high dimensions, the estimator becomes exponentially inefficient: almost all weight concentrates on extremely rare samples, making naive importance sampling computationally infeasible.

Marks available 30%

Exam Question 2

We are given the Langevin diffusion

$$dX_t = b(X_t, \alpha)\, dt + \sqrt{2}\, dW_t, \qquad b(x, \alpha) = \alpha \tanh(\alpha x) - \alpha x,$$

with parameter $\alpha \in \mathbb{R}$. The Fokker-Planck equation for the density $\pi_t(x)$ is

$$\frac{\partial}{\partial t}\pi_t(x) \;=\; -\frac{\partial}{\partial x}\big(b(x,\alpha)\pi_t(x)\big) + \frac{\partial^2}{\partial x^2}\pi_t(x).$$

We denote the stationary (invariant) density by $\pi(x) = \pi_{t=\infty}(x)$.

(a) Invariant distribution and mixture representation

At stationarity, $\pi(x)$ satisfies

$$0 \;=\; -\frac{d}{dx}\big(b(x,\alpha)\pi(x)\big) + \frac{d^2}{dx^2}\pi(x).$$

Assuming decay at infinity, the probability current is zero, i.e.

$$\pi'(x) = b(x, \alpha)\, \pi(x).$$

Thus

$$\frac{\pi'(x)}{\pi(x)} = b(x, \alpha), \qquad \log \pi(x) = \int b(x, \alpha)\, dx + C.$$

Now compute

$$\int b(x, \alpha)\, dx = \int \Big(\alpha \tanh(\alpha x) - \alpha x\Big)\, dx = \log \cosh(\alpha x) - \tfrac{\alpha}{2}x^2.$$

Hence

$$\pi(x) \;\propto\; \exp\Big(-\tfrac{\alpha}{2}x^2\Big)\, \cosh(\alpha x).$$

Using $\cosh(\alpha x) = \frac{1}{2}(e^{\alpha x} + e^{-\alpha x})$,

$$\pi(x) \;\propto\; \tfrac{1}{2}\Big(e^{-\frac{\alpha}{2}x^2 + \alpha x} + e^{-\frac{\alpha}{2}x^2 - \alpha x}\Big).$$

Completing the square gives

$$-\tfrac{\alpha}{2}x^2 \pm \alpha x = -\tfrac{\alpha}{2}(x \mp 1)^2 + \tfrac{\alpha}{2},$$

so

$$\pi(x) \;\propto\; e^{\alpha/2}\left\{e^{-\frac{\alpha}{2}(x-1)^2} + e^{-\frac{\alpha}{2}(x+1)^2}\right\}.$$

Normalising, we obtain

$$\pi(x) = \tfrac{1}{2}\,\mathcal{N}(x; 1, 1/\alpha) + \tfrac{1}{2}\,\mathcal{N}(x; -1, 1/\alpha),$$

where $\mathcal{N}(x; \mu, \sigma^2)$ denotes the Gaussian density with mean μ and variance σ^2. This is valid only for $\alpha > 0$ (otherwise the distribution is not normalisable).

Marks available 40%

(b) Conditions for exponential convergence

Define the potential U by $\pi(x) \propto e^{-U(x)}$. From part (a),

$$U(x) = \frac{\alpha}{2}x^2 - \log\cosh(\alpha x).$$

Then

$$U'(x) = \alpha x - \alpha\tanh(\alpha x), \qquad U''(x) = \alpha - \alpha^2 sech^2(\alpha x).$$

Exponential convergence to equilibrium is guaranteed if U is uniformly strongly convex, i.e. $\exists m > 0$ with $U''(x) \geq m$ for all x.

Since $sech^2(\alpha x) \leq 1$, we have

$$U''(x) \;\geq\; \alpha - \alpha^2 = \alpha(1-\alpha).$$

Thus strong convexity holds iff

$$0 < \alpha < 1, \qquad m = \alpha(1-\alpha).$$

Therefore:

$$\boxed{0 < \alpha < 1 \quad \Longrightarrow \quad \text{exponential convergence with rate at least } m = \alpha(1-\alpha).}$$

For $\alpha \geq 1$, U is not convex (double-well potential). The spectral gap may be exponentially small, and exponential convergence fails. For $\alpha \leq 0$, no stationary distribution exists.

Marks available 30%

(c) Effect on MALA performance

- If $\alpha \leq 0$: the SDE does not admit a normalisable stationary density, so MALA is not well-defined.
- If $0 < \alpha < 1$: the potential U is strongly convex, π is log-concave, and MALA performs well. Mixing is rapid, and tuning of the step size is straightforward.
- If $\alpha \geq 1$: the potential becomes non-convex with two sharp modes at ± 1 separated by a barrier. The chain is metastable:
  - it can spend very long times in one mode before crossing to the other;
  - the spectral gap is exponentially small in the barrier height;
  - MALA's gradient proposals further reinforce staying near a local mode.

  Hence MALA mixes extremely slowly across modes, and practical performance deteriorates badly unless additional strategies (tempering, mode-jumping, etc.) are employed.

  Marks available 30%

Exam Question 3

(a)

$$\begin{aligned}\mathbb{E}_{v\sim p(v)}\left[v^\top Jv\right] &= \mathbb{E}\left[\sum_i\sum_j v_i v_j J_{ij}\right] \\ &= \sum_i\sum_j \mathbb{E}\left[v_i v_j\right] J_{ij}\end{aligned}$$

Since v_i, v_j are independent zero-mean r.v.s with unit variance we have $\mathbb{E}[v_i v_j] = \delta_{ij}$, so

$$\begin{aligned}\sum_i\sum_j \mathbb{E}\left[v_i v_j\right] J_{ij} &= \sum_i\sum_j \delta_{ij} J_{ij} \\ &= \sum_i J_{ii} = \mathrm{tr}(J).\end{aligned}$$

Marks available 20%

(b)

i $\mathbb{E}[\hat{\mu}] = \frac{1}{N}\sum_i \mathbb{E}\left[{v^{(i)}}^\top J v^{(i)}\right] = \frac{1}{N}\sum_i \mathrm{tr}(J) = \mathrm{tr}(J) = \mu.$

Marks available 10%

ii Starting out with the variance of $\hat{\mu}$:

$$\mathrm{Var}(\hat{\mu}) = \frac{1}{N}\mathrm{Var}\left(v^\top Jv\right)$$

Individual variance of a single random projection:

$$\mathrm{Var}\left(v^\top Jv\right) = \mathbb{E}\left[\left(v^\top Jv\right)^2\right] - \mathrm{tr}(J)^2$$

Expand the square:

$$\begin{aligned}\mathbb{E}\left[\left(v^\top Jv\right)^2\right] &= \mathbb{E}\left[\left(\sum_i\sum_j v_i v_j J_{ij}\right)^2\right] \\ &= \sum_{i,j,k,l}\mathbb{E}\left[v_i v_j v_k v_l\right] J_{ij} J_{kl}\end{aligned}$$

Use the hint, Isserli's theorem:

$$\mathbb{E}\left[v_i v_j v_k v_l\right] = \mathbb{E}\left[v_i v_j\right]\mathbb{E}\left[v_k v_l\right] + \mathbb{E}\left[v_i v_k\right]\mathbb{E}\left[v_j v_l\right] + \mathbb{E}\left[v_i v_l\right]\mathbb{E}\left[v_j v_k\right] = \delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}$$

Sub in expression:

$$\begin{aligned}\sum_{i,j,k,l}\left[\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right] J_{ij}J_{kl} &= \left(\sum_{i,j}\delta_{ij}J_{ij}\right)\left(\sum_{k,l}\delta_{kl}J_{kl}\right) + \sum_{i,j,k,l}\delta_{ik}\delta_{jl}J_{ij}J_{kl} + \sum_{i,j,k,l}\delta_{il}\delta_{jk}J_{ij}J_{kl} \\ &= \left(\sum_i J_{ii}\right)\left(\sum_k J_{kk}\right) + \sum_{i,j} J_{ij}J_{ij} + \sum_{i,j} J_{ij}J_{ji}\end{aligned}$$

Using $(AB)_{ij} = \sum_k A_{ik}B_{kj}$

$$\left(\sum_i J_{ii}\right)\left(\sum_k J_{kk}\right) + \sum_{i,j} J_{ij}J_{ij} + \sum_{i,j} J_{ij}J_{ji} = \mathrm{tr}(J)^2 + \sum_i (JJ^\top)_{ii} + \sum_i (J^2)_{ii}$$
$$= \mathrm{tr}(J)^2 + \mathrm{tr}(JJ^\top) + \mathrm{tr}(J^2)$$

and since $J = \nabla_x^2 \log p_\theta(x)$ is a symmetric, we have $\mathbb{E}\left[\left(v^\top Jv\right)^2\right] = \mathrm{tr}(J)^2 + 2\mathrm{tr}(J^2)$, $\mathrm{Var}\left(v^\top Jv\right) = 2\mathrm{tr}(J^2)$ and therefore:

$$\mathrm{Var}(\hat{\mu}) = \frac{2}{N}\mathrm{tr}(J^2).$$

Marks available 40%

(c)

$$\nabla_x \log p_\theta(x) = -\frac{x}{\theta^2}, \quad \nabla_x^2 \log p_\theta(x) = -\frac{1}{\theta^2} I$$

$$\mathrm{tr}\left(-\frac{1}{\theta^2} I\right) = -\frac{n}{\theta^2}, \quad 2\mathrm{tr}\left(\frac{1}{\theta^4} I\right) = \frac{2n}{\theta^4}$$

so MC estimator is approx dist. as

$$\hat{\mu} \approx \mathcal{N}\left(-\frac{n}{\theta^2}, \frac{2n}{N\theta^4}\right).$$

Marks available 30%

Exam Question 4

(a) Gibbs uses exact conditionals for proposals. Moving from θ_0, ϕ_0 to θ_1, ϕ_1, which is achieved by the two step procedure:

$$\text{Step 1:} \qquad (\theta_0, \phi_0) \quad \text{to} \quad (\theta_1, \phi_0)$$

$$\text{Step 2:} \qquad (\theta_1, \phi_0) \quad \text{to} \quad (\theta_1, \phi_1)$$

The general acceptance probability is shown below, where the target posterior is given by π:

$$\alpha((\theta_0, \phi_0), (\theta_1, \phi_0)) = \frac{\pi(\theta_1, \phi_0) q(\theta_0, \phi_0 | \theta_1, \phi_0)}{\pi(\theta_0, \phi_0) q(\theta_1, \phi_0 | \theta_0, \phi_0)} = \frac{\pi(\theta_1 | \phi_0) \pi(\phi_0) p(\theta_0 | \phi_0)}{\pi(\theta_0 | \phi_0) \pi(\phi_0) p(\theta_1 | \phi_0)} = 1$$

then

$$\alpha((\theta_1, \phi_0), (\theta_1, \phi_1)) = \frac{\pi(\phi_1 | \theta_1) \pi(\theta_1) p(\phi_0 | \theta_1)}{\pi(\phi_0 | \theta_1) \pi(\theta_1) p(\phi_1 | \theta_1)} = 1$$

So every sample is accepted using this Gibbs sampling scheme.

Marks available 40%

(b) The full posterior is given by the priors $\times$ likelihood

$$p(\mu, \tau | x) \propto p(x | \mu, \tau) p(\mu) p(\tau)$$

The likelihood for the model is given by:

$$p(x | \mu, \tau) = \frac{\tau^{N/2}}{\sqrt{2\pi}^N} \exp\left(-\frac{\tau}{2} \sum_{n=1}^{N} (x_n - \mu)^2 \right)$$

Now the conditionals can be defined.

$$\begin{aligned} p(\mu | \tau, x) &\propto \exp\left(-\frac{\tau}{2} \sum_{n=1}^{N} (x_n - \mu)^2 \right) \exp\left(-\frac{w}{2} \mu^2 \right) \\ &\propto \exp\left(-\frac{1}{2} \left[\tau \sum_{n=1}^{N} (x_n - \mu)^2 + w\mu^2 \right] \right) \quad \text{Complete the square} \\ &\propto \exp\left(-\frac{1}{2\left(w + N\tau\right)^{-1}} \left[\mu - \frac{\tau}{w + N\tau} \sum_{n=1}^{N} y_n \right]^2 \right) \end{aligned}$$

Giving

$$p(\mu | \tau, x) = \mathcal{N}\left(\frac{\tau}{w + N\tau} \sum_{n=1}^{N} x_n, \frac{1}{w + N\tau} \right)$$

Now for the conditional for the precision

$$p(\tau|\mu, x) \propto \tau^{\alpha-1} e^{-\beta\tau} \tau^{N/2} \exp\left(-\frac{\tau}{2}\sum_{n=1}^{N}(y_n - \mu)^2\right)$$
$$= \tau^{\alpha+\frac{N}{2}-1} \exp\left(-\left[\beta + \frac{1}{2}\sum_{n=1}^{N}(x_n - \mu)^2\right]\tau\right)$$

Giving

$$p(\tau|\mu, x) = \text{Gamma}(\alpha', \beta')$$

Where

$$\alpha' = \alpha + \frac{N}{2}$$
$$\beta' = \beta + \frac{1}{2}\sum_{n=1}^{N}(x_n - \mu)^2$$

Marks available 60%