Module 4M24: Computational Statistics and Machine Learning

4M24 Tripos 2024/25 - Cribs

1. **Monte Carlo Estimator**

(a) $\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} X_n$

$$\mathbb{E}[\hat{\mu}] = \mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N} X_n\right]$$
$$= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[X_n] = \frac{1}{N}\sum_{n=1}^{N}\mu = \frac{1}{N}N\mu = \mu$$

Expected value of the estimator is the true mean so estimator is unbiased.

Five marks for Monte Carlo estimate and 5 marks for showing mean of empirical estimate is actual desired value

(b) Now we have

$$\hat{\mu} = \sum_{n=1}^{N} w_n X_n$$

Mean of the estimator

$$\mathbb{E}[\hat{\mu}] = \sum_{n=1}^{N} w_n \mathbb{E}[X_n] = \mu \sum_{n=1}^{N} w_n$$

Variance is found using the following property of the expectation: $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$.

$$\mathrm{Var}(\hat{\mu}) = \mathrm{Var}\left(\sum_{n=1}^{N} w_n X_n\right) = \sum_{n=1}^{N} w_n^2 \mathrm{Var}(X_n)$$
$$= \sigma^2 \sum_{n=1}^{N} w_n^2$$

ten marks for the mean and 20 for the variance

(c) For this to be an unbiased estimate, $\mathbb{E}[\hat{\mu}] = \mu$

$$\mathbb{E}[\hat{\mu}] = \mu \sum_{n=1}^{N} w_n = \mu$$

Giving

$$\sum_{n=1}^{N} w_n = 1$$

So the weights must sum to 1 to give an unbiased estimator.

ten marks if identify correctly summability to one

(d) For a minimum variance unbiased estimator, we must minimise the variance w.r.t the weights, subject to the constraint of the estimator being unbiased, i.e. the sum of the weights must equal 1. So we perform constrained optimisation:

Minimise $\mathrm{Var}(\hat{\mu}) = \sigma^2 \sum_{n=1}^{N} w_n^2$ w.r.t $w_1, \ldots, w_N$ s.t. $\sum_{n=1}^{N} w_n = 1$

Using the Lagrange multiplier

$$\underset{w_1,\ldots,w_N}{\operatorname{argmin}} \left\{ \sigma^2 \sum_{n=1}^{N} w_n^2 + \lambda \left(\sum_{n=1}^{N} w_n - 1 \right) \right\}$$

let $L = \sigma^2 \sum_{n=1}^{N} w_n^2 + \lambda \left(\sum_{n=1}^{N} w_n - 1 \right)$

$$\frac{\partial L}{\partial w_i} = 2\sigma^2 w_i + \lambda = 0 \tag{1}$$

$$w_i = \frac{-\lambda}{2\sigma^2} \tag{2}$$

Now using the constraint:

$$\sum_{n=1}^{N} w_n = 1$$

$$\sum_{n=1}^{N} \frac{-\lambda}{2\sigma^2} = \frac{-N\lambda}{2\sigma^2} = 1$$

$$\lambda = \frac{-2\sigma^2}{N}$$

$$w_i = \frac{2\sigma^2/N}{2\sigma^2} = \frac{1}{N}$$

Equally weighted samples gives the minimum variance Monte Carlo estimate.

ten marks to define constrained optimisation, twenty marks for Lagrangian with constraints, ten for correct derivative, ten for expression for lambda and each w

2. **Gradient of Marginal Density**

(a) Noting that

$$\nabla_\theta \log p(y|\theta_n) = \frac{\nabla_\theta p(y|\theta_n)}{p(y|\theta_n)} = \frac{1}{p(y|\theta_n)} \nabla_\theta \int p(y, x|\theta_n) dx$$

and that

$$\frac{1}{p(y|\theta_n)} \nabla_\theta \int p(y, x|\theta_n) dx = \frac{1}{p(y|\theta_n)} \int \nabla_\theta \log p(y, x|\theta_n) \times p(y, x|\theta_n) dx = \int \nabla_\theta \log p(y, x|\theta_n) \times \frac{p(y, x|\theta_n)}{p(y|\theta_n)} dx$$

further

$$\int \nabla_\theta \log p(y, x|\theta_n) \times \frac{p(y, x|\theta_n)}{p(y|\theta_n)} dx = \int \nabla_\theta \log p(y, x|\theta_n) \times p(x|y, \theta_n) dx = \mathbb{E}_{x|y,\theta} \left\{ \nabla_\theta \log p(y, x|\theta_n) \right\}$$

45 Marks available. 15 marks for each step.

(b) Metropolis Hastings algorithm

---

for $m = 1, \ldots, M$ **do**
  sample $e_m \sim \mathcal{U}[0, 1]$.
  propose $x'|x^m$ using ULA for $q(x'|x^m)$ with $x' = x^m + \delta \nabla \log p(x^m|y, \theta_n) + \sqrt{2\delta}\mathcal{N}(0, 1)$
  evaluate $\alpha(x', x^m) = \min\left(1, \frac{p(x'|y,\theta_n)}{p(x^m|y,\theta_n)} \times \frac{q(x^m|x')}{q(x'|x^m)}\right)$
  if $\epsilon_m \leq \alpha(x', x^m)$ **then**
    $x^{m+1} = x'$
  else
    $x^{m+1} = x^m$
  end
end
Return $x_1, \cdots, x_M$

---

40 Marks available. 10 marks for each step including ULA.

(c)

$$\mathbb{E}_{x|y,\theta} \left\{ \nabla_\theta \log p(y, x|\theta_n) \right\} \approx \frac{1}{M} \sum_{m=1}^{M} \log p(y, x_m|\theta_n)$$

with $x_1, \cdots, x_M$ from Markov chain targeting $p(x|y, \theta_n)$ 15 Marks available.

3. **Gaussian Measure**

(a) (i) Since the probabilities are defined wrt the Lebesgue measure, then the definition of Bayes rule will not be valid because the Lesbesgue measure in $\mathcal{H}$ is not well defined.

10 marks available

(ii) If the reference measure μ^0 is a probability measure defined in $\mathcal{H}$, then the Radon-Nikodym derivative

$$\frac{d\mu^y}{d\mu^0}(x) \propto P(y|x)$$

where μ^y is the posterior measure, and $P(y|x)$ the likelihood of data $y \in \mathbb{R}^D$ given $x \in \mathcal{H}$.
10 marks available

(b) (i) This requires that

$$\mu^0 = \prod_{k=1}^{\infty} g \propto \exp\left(-\frac{1}{2}\sum_{k=1}^{\infty} x_k^2\right)$$

be finite, and so

$$\sum_{k=1}^{\infty} x_k^2 < \infty$$

which indicates

$$x_{k=1,\dots,\infty} \in l_2 \subset \mathbb{R}^\infty$$

30 marks available

(ii)

$$\mu^y(x) \propto P(y|x)\mu^0(x) = P(y|x)\mathcal{N}(x; 0, C)$$

For a proposal $v = u + \beta\zeta$, with $\zeta \sim \mathcal{N}(0, C)$, i.e. $v \sim \mathcal{N}(u, \beta^2 C)$, the acceptance probability is given by

$$\alpha(v, u) = \min\{J(v) - J(u), 1\}$$

where $J(v) = \log P(y|v) - \frac{1}{2}|C^{-\frac{1}{2}}v|^2$.

The problem here is that $|C^{-\frac{1}{2}}v|^2$ is unbounded and so the acceptance ratio is not well defined. 30 marks available

(iii) Alternative proposal is $v = \sqrt{1-\beta^2}u + \beta\zeta$, with $\zeta \sim \mathcal{N}(0, C)$ - the pCN proposal. The acceptance ratio becomes

$$\alpha(v, u) = \min\left\{\frac{P(y|v)}{P(y|u)}, 1\right\}$$

since $P(y, \cdot)$ is defined on $\mathbb{R}^D$, the acceptance ratio is well defined for $u, v \in \mathcal{H}$.
20 marks available

4. **Langevin Diffusions**

(a) Starting with the hyperbolic secant need to show that the density is log-concave and hence all second derivatives take negative values for all values of x.

$$
\begin{aligned}
f(x) &= \log(p(x)) \\
f'(x) &= -\frac{1}{2}\pi \tanh\left(\frac{1}{2}\pi x\right) \\
f''(x) &= -\frac{1}{4}\pi^2 \mathrm{sech}^2\left(\frac{1}{2}\pi x\right) = -\pi^2 p(x)^2 < 0
\end{aligned}
$$

The Langevin diffusion equation

$$
dX_t = \nabla \log(p(X_t))dt + \sqrt{2}dB_t = -\frac{1}{2}\pi \tanh\left(\frac{1}{2}\pi x\right) + \sqrt{2}dB_t
$$

40 marks available, 25 to show existance and 15 for describing the equation

(b) The discrete form of the diffusion is required for ULA

$$
X_{n+1} = X_n - \frac{\delta}{2}\pi \tanh\left(\frac{1}{2}\pi X_n\right) + \sqrt{2\delta}W_{n+1}
$$

where δ is the step size and $W_{n+1} \sim \mathcal{N}(0,1)$

20 marks available

(c) MALA proposal from ULA $q(x'|x) \propto \exp\left(-\frac{1}{4\delta}|x' - x + \frac{\delta}{2}\pi \tanh\left(\frac{1}{2}\pi x\right)|^2\right)$ and reverse is $q(x|x') \propto \exp\left(-\frac{1}{4\delta}|x - x' + \frac{\delta}{2}\pi \tanh\left(\frac{1}{2}\pi x'\right)|^2\right)$

Ratio of $\log \frac{p(x')}{p(x)} = \frac{\mathrm{sech}(\frac{\pi x'}{2})}{\mathrm{sech}(\frac{\pi x}{2})}$

So $\alpha(x', x) = \min\left(1, \frac{\mathrm{sech}(\frac{\pi x'}{2})}{\mathrm{sech}(\frac{\pi x}{2})} \times \frac{q(x|x')}{q(x'|x)}\right)$

Marks 40 available - 10 for each correct forward and backward proposals (20 in total) and 20 for correctly and compactly defining the acceptance probability