Engineering Tripos Part IIB FOURTH YEAR

Module 4M24: Computational Statistics and Machine Learning

4M24 Tripos 2021/22 - Cribs

1. **Control Variates**

(a)

$$\begin{aligned}\mathrm{Var}(\hat{\mu}) &= \mathrm{Var}\left(\frac{1}{N}\sum_{n=1}^{N} X_n\right) \\ &= \frac{1}{N^2}\sum_{n=1}^{N}\mathrm{Var}(X_n) = \frac{1}{N^2}\sum_{n=1}^{N}\sigma^2 \\ &= \frac{\sigma^2}{N}\end{aligned}$$

20 marks available

(b) (i)

$$\hat{\mu}_{CV} = \hat{\mu} + c\,[m - \hat{m}]$$

Unbiased because $\mathbb{E}[\hat{\mu}] = \mu$, and $\mathbb{E}[\hat{m}] = m\,(= \mathbb{E}[g(x)])$.

10 marks available

(ii)

$$\mathrm{Var}(\hat{\mu}_{CV}) = \mathrm{Var}(\hat{\mu} + c\,[m - \hat{m}])$$

Using $\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X, Y)$ we have the following:

$$\begin{aligned}\mathrm{Var}(\hat{\mu}_{CV}) &= \mathrm{Var}(\hat{\mu}) + c^2\mathrm{Var}(m - \hat{m}) + 2c\mathrm{Cov}(\hat{\mu}, m - \hat{m}) \\ &= \mathrm{Var}(\hat{\mu}) + c^2\mathrm{Var}(\hat{m}) + 2c\mathrm{Cov}(\hat{\mu}, \hat{m})\end{aligned}$$

Now we can find the optimal variance by finding the stationary point of the variance wrt the coefficient c.

$$\frac{\partial}{\partial c}\mathrm{Var}(\hat{\mu}_{CV}) = 2c\mathrm{Var}(\hat{m}) + 2\mathrm{Cov}(\hat{\mu}, \hat{m}) = 0$$

Solving for c:

$$c = -\frac{\mathrm{Cov}(\hat{\mu}, \hat{m})}{\mathrm{Var}(\hat{m})}$$

Substituting back into the expression for the control variate estimator variance $\mathrm{Var}(\hat{\mu}) + c^2\mathrm{Var}(\hat{m}) + 2c\mathrm{Cov}(\hat{\mu}, \hat{m})$ gives the answer:

$$\mathrm{Var}(\hat{\mu}_{CV}) = \mathrm{Var}(\hat{\mu}) - \frac{\mathrm{Cov}(\hat{\mu}, \hat{m})^2}{\mathrm{Var}(\hat{m})}$$

30 marks available

(c) $X \sim \mathcal{U}[0, 1] \in \mathbb{R}$. First we can find the mean $\mu = \mathbb{E}[f(x)]$.

$$\mu = \mathbb{E}\left[\frac{1}{1+x}\right] = \int_0^1 \frac{1}{1+x}dx = [\log(1+x)]_0^1 = \log 2$$

Now to find the variance $\sigma^2 = \mathbb{E}\left[\frac{1}{(1+x)^2}\right] - \mu^2$.

$$\begin{aligned}
\sigma^2 &= \mathbb{E}\left[\frac{1}{(1+x)^2}\right] - (\log 2)^2 \\
&= \int_0^1 \frac{1}{(1+x)^2} dx - (\log 2)^2 \\
&= \left[-\frac{1}{1+x}\right]_0^1 - (\log 2)^2 \\
&= \frac{1}{2} - (\log 2)^2 \approx 0.0195
\end{aligned}$$

which gives

$$\mathrm{Var}(\hat{\mu}) = \frac{\sigma^2}{N} \approx \frac{0.0195}{N}$$

Defining $\mathrm{Var}(\hat{m}) = \frac{\sigma_g^2}{N}$. We find the mean and variance of $g(x)$:

$$m = \mathbb{E}\left[1+x\right] = \int_0^1 (1+x) dx = \left[x + \frac{x^2}{2}\right]_0^1 = \frac{3}{2}$$

$$\begin{aligned}
\sigma_g^2 = \mathbb{E}\left[(1+x)^2\right] - \frac{3}{2} = \int_0^1 (1+x)^2 dx - \frac{9}{4} &= \int_0^1 (1+2x+x^2) dx - \frac{9}{4} \\
&= \left[x + x^2 + \frac{x^3}{3}\right]_0^1 - \frac{9}{4} = \frac{1}{12}
\end{aligned}$$

which gives

$$\mathrm{Var}(\hat{m}) = \frac{1}{12N}$$

Finally we must find the covariance $\mathrm{Cov}(\hat{\mu}, \hat{m})$:

$$\begin{aligned}
\mathrm{Cov}(\hat{\mu}, \hat{m}) &= \frac{1}{N}\mathrm{Cov}(f(x), g(x)) \\
&= \frac{1}{N}\mathbb{E}\left[f(x)g(x)\right] - \frac{3}{2N}\log 2 \\
&= \frac{1}{N}\int_0^1 \frac{1+x}{1+x} dx - \frac{3}{2N}\log 2 \\
&= \frac{1}{N}\int_0^1 dx - \frac{3}{2N}\log 2 \\
&= \frac{1}{N} - \frac{3}{2N}\log 2
\end{aligned}$$

Substituting this into the optimal variance expression gives the required result.

$$\mathrm{Var}\left(\hat{\mu}_{CV}\right) = \frac{0.0195}{N} - 12N\left(\frac{1}{N} - \frac{3}{2N}\log 2\right)^2$$

40 marks available

2. **Function Approximation**

(a) The set $C[-\pi,\pi]$ is the continuous functions of $[-\pi,\pi]$ and $L^2[-\pi,\pi]$ is the set of square integrable functions which will include discontinuous functions such as the step function, which would not be in $C[-\pi,\pi]$. Therefore, $C_2[-\pi,\pi]$ would be the set of L^2 integrable functions which are also continuous. 10 marks available

(b)

$$f(x) = \sum_{k=1}^{\infty} c_k \phi_k(x)$$

$$\begin{aligned}\|f(x)\|_{L^2}^2 &= \int_{-\pi}^{\pi} \left(\sum_i c_i \phi_i(x)\right)\left(\sum_j c_j^* \phi_j^*(x)\right) dx \\ &= \sum_i \sum_j c_i c_j^* \int_{-\pi}^{\pi} \phi_i(x)\phi_j^*(x) dx \\ &= \sum_i \sum_j c_i c_j^* \delta_{i,j} = \sum_{i=1}^{\infty} c_i^2\end{aligned}$$

where the last line comes from the orthonormality of the functions $\{\phi_k\}$, i.e.

$$\int_{-\pi}^{\pi} \phi_i(x)\phi_j^*(x) dx = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad = \delta_{i,j}$$

giving

$$\|f(x)\|_{L^2} = \sqrt{\sum_{i=1}^{\infty} c_i^2}$$

20 marks available

(c)

$$\|f\|_{W_2^s}^2 = \|D^s f\|_{L^2}^2$$

$$\begin{aligned}f(x) &= \sum_{k=1}^{\infty} c_k \exp(ikx) \\ Df(x) &= \sum_{k=1}^{\infty} (ik) c_k \exp(ikx) \\ D^s f(x) &= \sum_{k=1}^{\infty} (ik)^s c_k \exp(ikx)\end{aligned}$$

$$\|D^s f\|_{L^2}^2 = \sum_k \left|(ik)^{2s}\right| c_k^2 = \sum_k k^{2s} c_k^2$$

30 marks available

(d)

$$\begin{aligned}
\epsilon_N(f) = \|f - f_N\|_{L^2}^2 &= \|\sum_{k=N+1}^{\infty} c_k \phi_k(x)\|_{L^2}^2 \\
&= \sum_{k=N+1}^{\infty} c_k^2 = \sum_{k=N+1}^{\infty} c_k^2 k^{2s} \frac{1}{k^{2s}} \\
&< \frac{1}{N^{2s}} \sum_{k=N+1}^{\infty} c_k^2 k^{2s} \\
&< \frac{1}{N^{2s}} \sum_{k=1}^{\infty} c_k^2 k^{2s} = \frac{1}{N^{2s}} \|f\|_{W_2^s}^2
\end{aligned}$$

The smoother the function, the value of s will increase, and as s increases, the impact of increasing N is greater. Rate of convergence increases as the smoothness of the function increases. 40 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

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

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 Diffusion**

(a) Starting with the Langevin SDE:

$$dX_t = -\nabla U(X_t)dt + \sqrt{2}dB_t$$

which has invariant density $p(x) \propto \exp(-U(X_t))$.

From the given SDE:

$$\nabla U(X_t) = X_t - \tanh(X_t)$$

Now integrate to find $U(X_t)$:

$$\begin{aligned} U(x) &= \int (x - \tanh(x))\, dx \\ &= \frac{x^2}{2} - \ln(\cosh(x)) + C \end{aligned}$$

Which gives

$$\begin{aligned} p(x) &\propto \exp(-U(x)) \\ &= \exp\left(-\frac{x^2}{2} + \ln(\cosh(x)) - C\right) \\ &\propto \cosh(x)\exp\left(-\frac{x^2}{2}\right) \\ &= \frac{e^x + e^{-x}}{2}\exp\left(-\frac{x^2}{2}\right) \\ &\propto \exp\left(-\frac{x^2}{2} + x\right) + \exp\left(-\frac{x^2}{2} - x\right) \\ &\propto \exp\left(-\frac{1}{2}(x-1)^2\right) + \exp\left(-\frac{1}{2}(x+1)^2\right) \end{aligned}$$

and so the normalised density $p(x)$ is given by

$$p(x) = \frac{1}{2}\mathcal{N}(1, 1) + \frac{1}{2}\mathcal{N}(-1, 1)$$

60 marks available

(b)

$$q(x) = \exp\left(-x - \exp(-x)\right)$$

Need to determine $-\nabla U(x) = \nabla \ln q(x)$

$$\ln q(x) = -x - e^{-x}$$

$$\nabla \ln q(x) = -1 + e^{-x}$$

Hence

$$-\nabla U(x) = -1 + e^{-x}$$

and the corresponding SDE is given by

$$dX_t = \left(e^{-X_t} - 1\right) dt + \sqrt{2} dB_t$$

40 marks available