Engineering Tripos Part IIB FOURTH YEAR

Module 4M24: Computational Statistics and Machine Learning

4M24 Tripos 2023/24 - Cribs

1. **Monte Carlo**

(a) With $\mathbb{E}[f(X)] = I, \mathbb{E}\left[(f(X)-I)^2\right] = \sigma_f^2$, we can define a transform to zero mean and unit variance

$$\tilde{f}_n = \frac{f_n - I}{\sigma_f} \tag{1}$$

so that $\mathbb{E}\left[\tilde{f}(X)\right] = 0, \mathbb{E}\left[\tilde{f}(X)^2\right] = 1$.

Now define $S_N = \sum_{n=1}^N \tilde{f}_n(x_n)$, and $Z_N = S_N/\sqrt{N}$. Based on properties on MGF, we can write

$$\begin{aligned} M_{S_N}(t) &= \left(M_{\tilde{f}_n}(t)\right)^N \\ M_{Z_N}(t) &= \left(M_{\tilde{f}_n}(t/\sqrt{N})\right)^N = \left(\mathbb{E}\left[\exp(t\tilde{f}(x)/\sqrt{N})\right]\right)^N \end{aligned}$$

now expanding $\exp(t\tilde{f}(x)/\sqrt{N})$ using the Taylor expansion

$$\exp(t\tilde{f}(x)/\sqrt{N}) = 1 + \frac{t}{\sqrt{N}}\tilde{f}(x) + \frac{t^2}{2N}\tilde{f}(x)^2 + \mathcal{O}(t^3)$$

Taking the expectation

$$\mathbb{E}\left[\exp(t\tilde{f}(x)/\sqrt{N}) = 1 + \frac{t}{\sqrt{N}}\tilde{f}(x) + \frac{t^2}{2N}\tilde{f}(x)^2 + \mathcal{O}(N^{-2})\right] = 1 + \frac{t^2}{2N} + \mathcal{O}(N^{-2})$$

now we can write

$$M_{Z_N}(t) = \left(1 + \frac{t^2}{2N} + \mathcal{O}(N^{-2})\right)^N$$

and in the limit

$$\lim_{N\to\infty} M_{Z_N}(t) = \lim_{n\to\infty}\left(1 + \frac{t^2}{2N} + \mathcal{O}(N^{-2})\right)^N = \exp(t^2/2)$$

which is the MGF of a standard normal, hence in the limit $Z_N \sim \mathcal{N}(0,1)$. Now we just need to find the distribution of $\hat{I}$. We have

$$\begin{aligned} S_N &= \sqrt{N} Z_N \sim \mathcal{N}(0, N) \\ &= \sum_{n=1}^N \tilde{f}_n(x_n) = \sum_{n=1}^N \frac{f_n(x_n) - I}{\sigma_f} \end{aligned}$$

rearranging

$$\frac{1}{N}\sum_{n=1}^N f_n(x_n) = \frac{\sigma_f S_N}{N} + I$$

and since $\frac{\sigma_f S_N}{N} \sim \mathcal{N}(0, \sigma_f^2/N)$, we conclude

$$\hat{I} = \frac{1}{N}\sum_{n=1}^N f_n(x_n) \sim \mathcal{N}\left(I, \frac{\sigma_f^2}{N}\right).$$

60%

(b) Variance of $\hat{I}$

$$\mathrm{Var}\left(\sum_{n=1}^{N} w_n f(x_n)\right) = \sum_{n=1}^{N} \mathrm{Var}\left(w_n f(x_n)\right) = \sum_{n=1}^{N} w_n^2 \mathrm{Var}\left(f(x_n)\right) = \sigma_f^2 \sum_{n=1}^{N} w_n^2$$

minimise $\sigma_f^2 \sum_{n=1}^N w_n^2$, subject to constraint $\sum_{n=1}^N w_n = 1$ (unbiased estimator). Lagrange multiplier method

$$L(\mathbf{w}, \lambda) = \sigma_f^2 \sum_n w_n^2 + \lambda(\sum_n w_n - 1)$$
$$\frac{\partial L}{\partial w_i} = 2\sigma_f^2 w_i + \lambda = 0$$

giving $w_i = \frac{-\lambda}{2\sigma_f^2}$. Constraint

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

hence $\lambda = \frac{-2\sigma_f^2}{N}$ and finally

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

40%

2. **Thermodynamic Integration**

(a)

$$p(\theta|y,t) = \frac{p(y|\theta,t)p(\theta)}{p(y|t)}$$

where $p(y|t) = \int p(y|\theta', t)p(\theta')d\theta' = \int p(y|\theta')^t p(\theta')d\theta'$ 10%

(b)

$$p(y|t=0) = \int p(\theta')d\theta' = 1$$
$$p(y|t=1) = \int p(y|\theta')p(\theta')d\theta' = p(y)$$

15%

(c)

$$\begin{aligned}\frac{d}{dt}\log p(y|t) &= \frac{1}{p(y|t)}\frac{d}{dt}p(y|t) = \frac{1}{p(y|t)}\int \frac{d}{dt}\left[p(y|\theta)^t\right]p(\theta)d\theta \\ &= \frac{1}{p(y|t)}\int p(y|\theta)^t p(\theta)\log p(y|t)d\theta \\ &= \int \frac{p(y|\theta)^t p(\theta)}{p(y|t)}\log p(y|t)d\theta \\ &= \int p(\theta|y,t)\log p(y|t)d\theta = \mathbb{E}_{\theta|y,t}\left\{\log p(y|\theta)\right\}\end{aligned}$$

30%

(d)

$$\int_0^1 \frac{d}{dt}\log p(y|t)dt = \log p(y|t=1) - \log p(y|t=0) = \log p(y)$$

$$\int_0^1 \int p(\theta|y,t)\log p(y|t)d\theta dt = \int_0^1 \int p(\theta|y,t)p(t)\log p(y|t)d\theta dt$$
$$= \int_0^1 \int p(\theta,t|y)\log p(y|t)d\theta dt = \mathbb{E}_{\theta,t|y}\left\{\log p(y|\theta)\right\}$$

since $p(t)=1, t\in[0,1]$.
finally

$$\log p(y) = \mathbb{E}_{\theta,t|y}\left\{\log p(y|\theta)\right\}$$

15%

(e) Sampling algorithm

---

for $n=1,\ldots,N$ **do**
  sample $t_n \sim \mathcal{U}[0,1]$.
  run any MCMC to sample $\theta|y,t_n$
  for $m=1,\ldots,M$ **do**
    sample $\theta_m \sim p(\theta|y,t_n)$.
  end
end

$$\mathbb{E}_{\theta|y,t_n}\left\{\log p(y|\theta)\right\} \approx \frac{1}{M}\sum_{m=1}^{M}\log p(y|\theta_m), \quad \log p(y) \approx \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}_{\theta|y,t_n}\left\{\log p(y|\theta)\right\}$$

30%

3. **Gaussian Measures**

(a) Bookwork - See lecture slides.
A Hilbert space H has an associated basis $e_i, i=1,\ldots,\infty$. We denote a ball or radius r, centered at x as $B(x,r)$. Consider a ball $B(0,2)$, and consider placing for each basis vector a ball of radius a half inside this, centered at the basis vector: $\{B(e_i,1/2), i=1,\ldots,\infty\}$. Clearly they do not intersect, hence $B(e_i,1/2)\cap B(e_j,1/2)=\emptyset$ if $i\neq j$. We also know the union of all the balls will be a subset of $B(0,2)$, i.e. $\cup_{i\in\mathbb{N}}B(e_i,1/2)\subset B(0,2)$. Since the Lebesgue measure in infinite dimensions is **translation invariant**, we have the measure of each of the balls is constant, $\mu\left(B(e_i,1/2)\right)=\mu\left(B(e_j,1/2)\right), \forall i,j\in\mathbb{N}$. **Monotonicity** and **countable additivity** gives

$$\sum_{i\in\mathbb{N}}\mu\left(B(e_i,1/2)\right)\leq\mu\left(B(0,2)\right)$$

however, the Lebesgue measure is (sigma) finite, so $\mu\left(B(0,2)\right)<\infty$, but $\sum_{i\in\mathbb{N}}\mu\left(B(e_i,1/2)\right)=\text{constant}\times\sum_{i\in\mathbb{N}}=\infty$ hence we have a contradiction ($\mu\left(B(0,2)\right)<\infty$ and $\mu\left(B(0,2)\right)\geq\infty$), and therefore the Lebesgue measure is not well defined in an infinite dimensional Hilbert space.
5% for each measure point in bold. 20% for showing contradiction.

(b) Clearly $\prod_{k=1}^{\infty} g$ requires that

$$\exp\left(-\frac{1}{2}\sum_{k=1}^{\infty}x_k^2\right)$$

is finite and non-zero, which means that $\{x_k\}$ must satisfy

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

This is equivalent to the definition of l_2 - the space of bounded sequences.

$$\{x\in\mathbb{R};\sum_{k=1}^{\infty}x_k^2<\infty\}=l_2$$

5% indicting function needs to be bounded - 5% showing sum should be finite, 5% identifying space is l_2

(c) trace$(C) \leq \infty$ so C must be **trace class**. 15% for trace$(C) \leq \infty$- 5% for trace class.

(d) Define $\langle e_n, Ce_n \rangle$ as

$$\begin{aligned}
\langle e_n, Ce_n \rangle &= \mathbb{E}(e_n, u)(u, e_n) \text{ with } u \in H \\
&= \mathbb{E} \int_\Omega \int_\Omega e_n(x)(u(x)u(y))e_n(y)dydx \\
&= \mathbb{E} \int_\Omega e_n(x) \left(\int_\Omega u(x)u(y)e_n(y)dy \right) dx \\
&= \int_\Omega e_n(x) \left(\int_\Omega \mathbb{E}\{u(x)u(y)\}e_n(y)dy \right) dx \\
&= \int_\Omega e_n(x) \left(\int_\Omega c(x,y)e_n(y)dy \right) dx
\end{aligned}$$

$Ce_n(x) = \int_D c(x,y)e_n(y)dy$ 30%

4. **Product of Gaussians**

(a)

$$f(x)g(x) = \frac{1}{2\sigma_f\sigma_g} \exp\left[-\frac{x^2 - 2x\mu_f + \mu_f^2}{2\sigma_f^2} - \frac{x^2 - 2x\mu_g + \mu_g^2}{2\sigma_g^2} \right]$$

consider the term inside the exponential

$$\begin{aligned}
-\frac{x^2 - 2x\mu_f + \mu_f^2}{2\sigma_f^2} - \frac{x^2 - 2x\mu_g + \mu_g^2}{2\sigma_g^2} &= \frac{-2\sigma_g^2x^2 + 4x\sigma_g^2\mu_f - 2\sigma_g^2\mu_f^2 - 2\sigma_f^2x^2 + 4x\sigma_f^2\mu_g - 2\sigma_f\mu_g^2}{4\sigma_f^2\sigma_g^2} \\
&= \frac{-(\sigma_f^2 + \sigma_g^2)x^2 + 2x(\sigma_g^2\mu_f + \sigma_f^2\mu_g) - (\sigma_g^2\mu_f^2 + \sigma_f^2\mu_g^2)}{2\sigma_f^2\sigma_g^2} \\
&= \frac{-x^2 + 2x\left(\frac{\sigma_g^2\mu_f + \sigma_f^2\mu_g}{\sigma_f^2 + \sigma_g^2}\right) - \left(\frac{\sigma_g^2\mu_f^2 + \sigma_f^2\mu_g^2}{\sigma_f^2 + \sigma_g^2}\right)}{\frac{2\sigma_f^2\sigma_g^2}{\sigma_f^2 + \sigma_g^2}}
\end{aligned}$$

so we have

$$\mu_{fg} = \frac{\sigma_g^2\mu_f + \sigma_f^2\mu_g}{\sigma_f^2 + \sigma_g^2}, \quad \sigma_{fg}^2 = \frac{\sigma_f^2\sigma_g^2}{\sigma_f^2 + \sigma_g^2}$$

40%

(b) write down

$$p_{fg}(x) = \frac{1}{\sqrt{2\pi\sigma_{fg}^2}} \exp\left(-\frac{(x - \mu_{fg})^2}{2\sigma_{fg}^2} \right)$$

20%

(c)

$$\mu_{fg} = 1/2, \sigma_{fg} = 1/2, \quad p_{fg}(x) = \frac{1}{\sqrt{\pi}} \exp\left(-(x - \frac{1}{2})^2 \right)$$

(10%)

$$\nabla_x \log p_{fg}(x) = -2(x - \frac{1}{2}) = -2x + 1$$

(10%)

so ULA

$$x_{n+1} = x_n + \delta(1 - 2x) + \sqrt{2\delta}\zeta_n, \quad \zeta_n \sim \mathcal{N}(0, 1)$$

(20%)
TOTAL 40%