Engineering Tripos Part IIB — FOURTH YEAR

Module 4M24: Computational Statistics and Machine Learning

4M24 Tripos 2020/21 - Cribs

1. **Monte Carlo Estimator**

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

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

   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

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

   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, $\mathbb{E}\left[\hat{\mu}\right] = \mu$

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

   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. **Gibbs Sampling**

(a) Gibbs uses exact conditionals for proposals. Moving from u_0, v_0 to u_1, v_1, which is achieved by the two step procedure:

Step 1: (u_0, v_0) to (u_1, v_0)

Step 2: (u_1, v_0) to (u_1, v_1)

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

$$\alpha((u_0, v_0), (u_1, v_0)) = \frac{\pi(u_1, v_0) q(u_0, v_0 | u_1, v_0)}{\pi(u_0, v_0) q(u_1, v_0 | u_0, v_0)} = \frac{\pi(u_1|v_0)\pi(v_0)p(u_0|v_0)}{\pi(u_0|v_0)\pi(v_0)p(u_1|v_0)} = 1$$

then

$$\alpha((u_1, v_0), (u_1, v_1)) = \frac{\pi(v_1|u_1)\pi(u_1)p(v_0|u_1)}{\pi(v_0|u_1)\pi(u_1)p(v_1|u_1)} = 1$$

So every sample is accepted using this Gibbs sampling scheme.
20 marks showing interleaving of steps and ten marks each for acceptance probabilities for each step

(b) The full posterior is given by the priors × likelihood

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

The likelihood for the model is given by:

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

Now the conditionals can be defined.

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

Giving

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

Now for the conditional for the precision

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

Giving

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

Where

$$\alpha' = \alpha + \frac{N}{2}$$

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

30 marks for each conditional - ten for unnormalised conditional, ten marks for inetrmediate steps (completing square etc), te marks for correct exact conditional

3. **Measures**

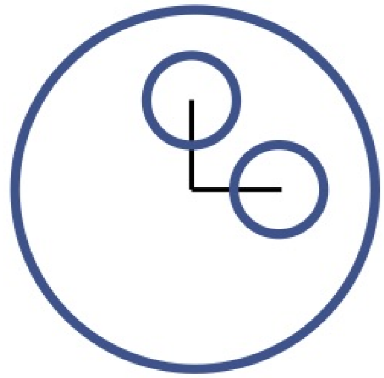
(a) From slides

1. Hilbert space has orthonormal set $e_i | i = 1, \ldots, \infty$

2. $\|e_i - e_j\|^2 = 0$ for $i = j$ and $= 2$ for $i \neq j$

3. $B(0, 2)$ and $B(e_i, \frac{1}{2})$ to show $\bigcup_{i\in\mathbb{N}} B(e_i, \frac{1}{2}) \subset B(0, 2)$.

These balls are also disjoint, i.e. $B(e_i, \frac{1}{2}) \cap B(e_j, \frac{1}{2}) = \emptyset$ for $i \neq j$



Translation invariance means that the measure of all the sub balls are equivalent, i.e.

$$\mu\left(B(e_i, \frac{1}{2})\right) = \text{constant} \qquad \forall i$$

Countable Additivity gives the inequality

$$\sum_{i\in\mathbb{N}} \mu(B(e_j, \frac{1}{2})) \leq \mu(B(0,2))$$

Giving $\mu(B(0,2)) \geq \text{constant} * \sum_{i\in\mathbb{N}} = \infty$ This is a contradiction, so the Lebesgue measure in infinite dimensional Hilbert space does not exist.

5 marks each for measure properties and 15 marks using these to show 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 marks indicting function needs to be buonded - 5 marks showing sum should be finite, 5 marks identifying space is l_2

(c)

$$\begin{aligned} H(\mu,\nu) &= \int \sqrt{d\mu d\nu} = \int \sqrt{\mathcal{N}(0,a)\mathcal{N}(0,b)dx^2} \\ &= \int \frac{1}{(2\pi a)^{1/4}} \frac{1}{(2\pi a)^{1/4}} \exp\left(-\frac{x^2}{4a} - \frac{x^2}{4b}\right) dx \\ &= \frac{1}{(ab)^{1/4}} \int \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\left(\frac{a+b}{2ab}\right)\right) dx \end{aligned}$$

Defining $\sigma^2 = \frac{2ab}{a+b}$

$$\begin{aligned} H(\mu,\nu) &= \frac{1}{(ab)^{1/4}} \underbrace{\int \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx}_{\sigma} \\ &= (ab)^{1/4}\sqrt{\frac{2}{a+b}} \end{aligned}$$

10 marks being able to write integral with remaining 25 for steps to correct expression

(d)

$$H(\mu,\nu) = \prod_{k=1}^{\infty} (\alpha a^2)^{1/4} \sqrt{\frac{2}{a(1+\alpha)}}$$
$$= \prod_{k=1}^{\infty} \sqrt{\frac{2\sqrt{\alpha}}{1+\alpha}}$$

$$\prod_{k=1}^{N} \sqrt{\frac{2\sqrt{\alpha}}{1+\alpha}} = \left(\frac{2\sqrt{\alpha}}{1+\alpha}\right)^{N/2}$$

Taking the limit

$$H(\mu,\nu) = \lim_{N\to\infty} \left(\frac{2\sqrt{\alpha}}{1+\alpha}\right)^{N/2} = \begin{cases} 0 & \alpha \neq 1 \\ 1 & \alpha = 1 \end{cases}$$

So the product measures are equivalent if $\alpha = 1$, and singular if $\alpha \neq 1$.

10 marks for correct expression in N and 10 marks for correct limit and conclusion

4. **Hilbert Space**

(a)

$$\lim_{n\to\infty} \int_0^1 f_n(x)dx = \lim_{n\to\infty} \int_0^1 nx(1-x^2)^n dx = \lim_{n\to\infty} \left[\frac{-n}{2(n+1)}(1-x^2)^{n+1}\right]_0^1$$
$$= \lim_{n\to\infty} \frac{n}{2n+2} = \frac{1}{2}$$

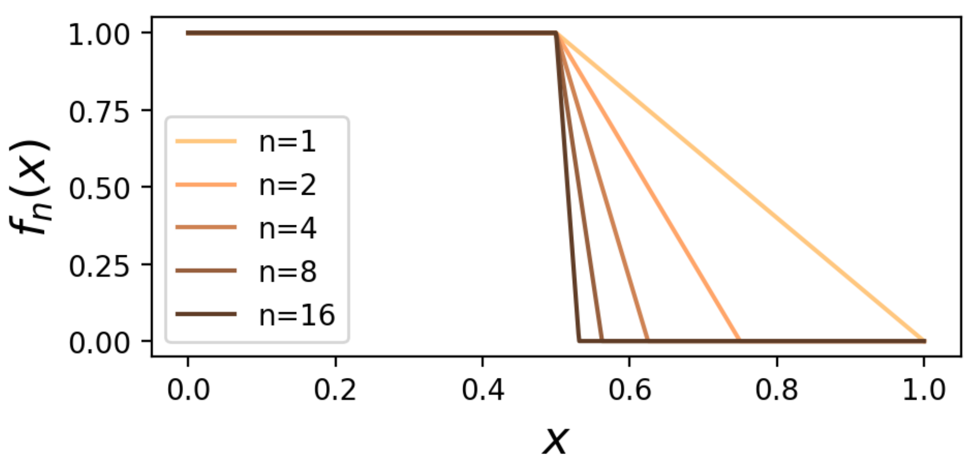
Now to check for exchangeability

$$\int_0^1 \lim_{n\to\infty} f_n(x)dx = 0$$

So this equality does not hold.

(b) Pointwise convergence does not preserve the continuity of functions so we cannot exchange the limit and the integral operations.
Uniform convergence provides convergence of the functions independent of x, so preserves continuity so the limit and integral can be exchanged.

(c) Plots of f_n are shown below for increasing n



Determine whether the sequence is Cauchy in $L^2[0,1]$, find the norm of the difference of f_n, f_m as $n, m \to \infty$. Let $n > m$

$$\int_0^1 |f_n(x) - f_m(x)|^2 \, dx = \int_{\frac{1}{2}}^{\frac{1}{2}(1+\frac{1}{n})} \left(2(m-n)\left(x - \frac{1}{2}\right)\right)^2 dx + \int_{\frac{1}{2}(1+\frac{1}{n})}^{\frac{1}{2}(1+\frac{1}{m})} \left(1 - 2m\left(x - \frac{1}{2}\right)\right)^2 dx$$

$$= \int_0^{\frac{1}{2n}} (2(m-n)x)^2 \, dx + \int_{\frac{1}{2n}}^{\frac{1}{2m}} (1 - 2mx)^2 \, dx$$

$$= 4(m-n)^2 \left[\frac{x^3}{3}\right]_0^{\frac{1}{2n}} + \left[x - 2mx^2 + \frac{4}{3}m^2x^3\right]_{\frac{1}{2n}}^{\frac{1}{2m}}$$

$$= \frac{(m-n)^2}{6n^3} + \left(\frac{1}{2m} - \frac{1}{2n}\right) - 2m\left(\frac{1}{4m^2} - \frac{1}{4n^2}\right) + \frac{4}{3}m^2\left(\frac{1}{8m^3} - \frac{1}{8n^3}\right)$$

Which tends to zero as $m, n \to \infty$ so this is indeed a Cauchy sequence in $L^2[0,1]$.

(d) The limiting function is given by

$$f(x) = \begin{cases} 1 & 0 \le x \le \frac{1}{2} \\ 0 & \frac{1}{2} < x \le 1 \end{cases}$$

And since

$$\int_0^1 f(x)^2 dx = \int_0^{1/2} dx = \frac{1}{2} < \infty$$

This limiting function $\in L^2[0,1]$ and $f_n \in L^2[0,1] \forall n$ so this forms a (complete) Hilbert space.

part a. 10 marks for correct lim-integral amd 5 marks of int-lim. part b 5 marks for each sensible statement.part c. 20 marks showing a series f_n - 30 marks showing it is cauchy. part.d 10 marks defining limit function - 10 marks showing L2 norm bounded

5. Assessors Report

Q1. Monte Carlo This was the most popular question answered and most students did well. The manipulation of the variance of the sum of variables caught some students out, and the construction of the constrained optimisation for the optimal weights of a generalised Monte Carlo estimator was tackled with not much issue. In general the question was answered well.

Q2. Gibbs Sampling This was the least popular question and the one which gained the lowest average mark. The detailed derivation of the exact conditionals was the most challenging part of the question given the mix of details and techniques required to obtain them.

Q3. Probability Measures This was attempted by around half of the candidates and overall there were no issues. The first part showing that Lebesgue Measure has no definition in an infinite dimensional Hilbert space tested knowledge and understanding, whilst the second part obtaining an expression for the Hellinger product of two measures required more technical manipulation.

Q4. Functional Analysis This question along with Q1 were the most popular with 54 of the students attempting it. The proof of the given series of functions forming a Cauchy space caused problems due mainly to the basic integration required. There were some pleasantly surprising approaches such as the use of Cauchy-Schwartz to prove convergence. Overall most students were able to distinguish between pre-Hilbert and complete spaces for the last part with a few students demonstrating a high degree of rigour in showing that all of the functions in the sequence are in L2, quite impressive.