Computational Statistics & Machine Learning
Lecture 10
Gaussian Measures in Hilbert Space

Mark Girolami
mag92@cam.ac.uk

Department of Engineering
University of Cambridge

November 3, 2021

Overview

- Lebesgue Measure in infinite dimensions
- Gaussian Measure in Hilbert space
- Equivalence and Singularity of Gaussian Measures in infinite dimensions
- Bayes formula in Hilbert space

Lebesgue Measure

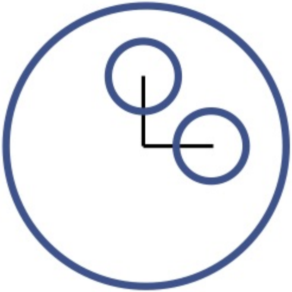
- We know that Lebesgue measure is a generalisation of notions such as length, area and volume
  - For every interval $I = [a, b]$ on $\mathbb{R}$ then $\mu(I) = l(I) = b - a$
  - If $A \subset B \subset \mathbb{R}$ then $0 \leq \mu(A) \leq \mu(B) \leq \infty$
  - For each subset A of $\mathbb{R}$ and each point $x_0 \in \mathbb{R}$ then $A + x_0 = \{x + x_0 : x \in A\}$ and $\mu(A + x_0) = \mu(A) + \mu(x_0) = \mu(A)$
- Finiteness
- Monotonicity
- Translation invariance

Lebesgue Measure in infinite dimensions

- Lebesgue measure in D dimensions $\prod_{d=1}^{D}(b_d - a_d)$
- What happens when $D = \infty$?
- Lets take a Hilbert space, H, and see what happens
- We know there is a countable orthonormal set acting as a basis in H, i.e. $\{e_i | i = 1, \cdots, \infty\}$ with $\langle e_i, e_j \rangle = \delta_{ij}$
- How far apart are they?
  $\|e_i - e_j\|^2 = \|e_i\|^2 + \|e_j\|^2 = 1 + 1 = 2$, if $i \neq j$
- Lets take a ball centered at 0 with radius 2 denoted $B(0, 2)$
- Lets place balls of radius $\frac{1}{2}$ at each e_i i.e. $\{B(e_i, \frac{1}{2}), i = 1, \cdots, \infty\}$
- It should be clear that $B(e_i, \frac{1}{2}) \cap B(e_j, \frac{1}{2}) = \emptyset$
- ... and $\bigcup_{i \in \mathbb{N}} B(e_i, \frac{1}{2}) \subset B(0, 2)$

Lebesgue Measure in infinite dimensions

- Lets take a ball centered at 0 with radius 2 denoted $B(0,2)$
- Lets place balls of radius $\frac{1}{2}$ at each e_i i.e. $\{B(e_i,\frac{1}{2}), i = 1, \cdots, \infty\}$
- Then $B(e_i,\frac{1}{2}) \cap B(e_j,\frac{1}{2}) = \emptyset$ and $\bigcup_{i \in \mathbb{N}} B(e_i,\frac{1}{2}) \subset B(0,2)$



Lebesgue Measure in infinite dimensions

- Now if Lebesgue measure in infinite dimensions is translation invariant then $\mu(B(e_i, \frac{1}{2})) = \mu(B(e_j, \frac{1}{2}))$, for all $i, j \in \mathbb{N}$
- The monotonicity and countable additivity of Lebesgue measure tells that

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

- But wait..... .... Lebesgue measure is (sigma) finite so $\mu(B(0,2)) < \infty$
- But $\sum_{i \in \mathbb{N}} \mu\left(B\left(e_i, \frac{1}{2}\right)\right) = constant \times \sum_{i \in \mathbb{N}} = \infty$
- Meaning that $\mu(B(0,2)) \geq \infty$
- There is no Lebesgue measure in infinite dimensional Hilbert (or Banach) spaces
- This is a problem discuss why this is a problem (think about how probability densities are defined)

Measures in infinite dimensions

- Most of Machine Learning takes place in infinite dimensional spaces
  - Function approximation - Kernel methods
  - Stochastic processes in Signal Processing
  - Dirichlet Processes in probabilistic modelling
  - Gaussian Processes in Bayesian Machine Learning
  - Inverse Problems in all aspects of engineering science
- Require a reference measure for the function spaces we are reliant upon in Machine Learning

Gaussian Measures in infinite dimensions

- Lets take a function we are familiar with the standard Gaussian in $\mathbb{R}$

$$g(B) = \frac{1}{\sqrt{2\pi}} \int_B \exp\left(-\frac{1}{2}x^2\right) dx$$

- Define the infinite dimensional product measure on $\mathbb{R}^\infty$ as $\mu = \prod_{n=1}^{\infty} \mu_n$ where each $\mu_n = g$
- Consider the function

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

- Well defined only if $\sum_{k=1}^{\infty} a_k x_k^2 < \infty$
- $x \in l_{2,a} = \{x \in \mathbb{R}^\infty; \sum_{k=1}^{\infty} a_k x_k^2 < \infty\}$
- $l_{2,a}$ a Hilbert space, inner product $\langle x, y\rangle = \sum_{k=1}^{\infty} a_k x_k y_k$

Gaussian Measures in infinite dimensions

- By using the infinite dimensional standard Gaussian product measure we obtain the result

$$\int_{\mathbb{R}^\infty} f_\lambda d\mu = \frac{1}{\sqrt{\prod_{k=1}^{\infty}(1+\lambda a_k)}}$$

- There are a number of things we can see here
  - We have replaced the Lebesgue measure with the product Gaussian measure in $d\mu$
  - If the infinite product converges then $\int_{\mathbb{R}^\infty} f_\lambda d\mu$ is finite and positive
  - If $\sum_{k=1}^{\infty} a_k < \infty$ then the product converges
  - Note that $\lambda \to 0$ then $f_\lambda(x) = 1$ if $x \in l_{2,a}$
  - Also as $\lambda \to 0$ then $\int_{\mathbb{R}^\infty} f_\lambda d\mu = \int_{\mathbb{R}^\infty} I_{x \in l_{2,a}} d\mu = \int_{l_{2,a}} d\mu = \mu(l_{2,a}) = 1$
  - This is awesome - we have a probability measure which gives full measure to the whole sequence space - and as it is isomorphic to L^2 to the corresponding function space

Gaussian Measures in infinite dimensions

- Passing over many technical details it follows that Gaussian measure in l^2 (Hilbert space), with mean zero and diagonal covariance operator $Sx = (a_n x_n)_{n=1}^{\infty}$, $x \in l^2$ if $\sum_{k=1}^{\infty} a_k < \infty$
- It means that the infinite dimensional sequence $x \in l^2$ has finite Gaussian measure we can then define a probability space over functions this is getting to be practically very important in Machine learning
- We have a measure which is finite and well defined in an infinite dimensional space and can be used in taking Radon-Nikodym derivatives in changing measures in e.g. spaces of functions

Full covariance operators in infinite dimensions

- We have considered only diagonal covariance operators so far
- More general full covariance operators can be defined and used in the definition of the Gaussian measure
- Given the convergence criterion to be met for the summation of the diagonal terms a similar requirement must be met for a full covariance operator
- If we have a linear operator C on a Hilbert space, H, it has to be *trace class* (or nuclear) meaning that $trace(C) < \infty$
- The trace operator in Hilbert space with an orthonormal basis is defined by

$$trace(C) = \sum_{n=1}^{\infty} \langle e_n, Ce_n \rangle$$

what does $\langle e_n, Ce_n \rangle$ look like?

Full covariance operators in infinite dimensions

- Let f and g be two zero-mean functions defined on a domain D such that $f, g \in H$
- Define $\langle f, Cg \rangle$ as

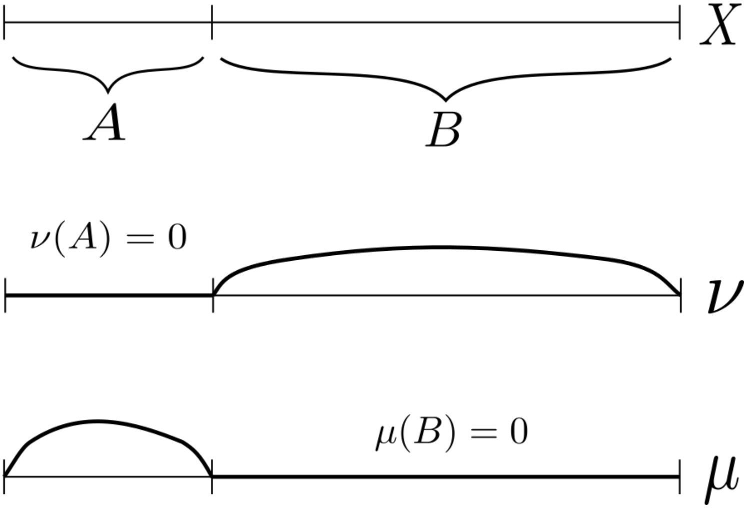
$$\begin{aligned}
\langle f, Cg \rangle &= \mathbb{E}(f,u)(u,g) \text{ with } u \in H \\
&= \mathbb{E} \int_D \int_D f(x)(u(x)u(y))g(y)dydx \\
&= \mathbb{E} \int_D f(x) \left(\int_D u(x)u(y)g(y)dy \right) dx \\
&= \int_D f(x) \left(\int_D \mathbb{E}\{u(x)u(y)\} g(y)dy \right) dx \\
&= \int_D f(x) \left(\int_D c(x,y)g(y)dy \right) dx
\end{aligned}$$

- So $Ce_n(x) = \int_D c(x,y)e_n(y)dy$
- For those of you studying Probabilistic Machine Learning recognise the *covariance function* defining a Gaussian Process

Gaussian Measures in Hilbert Space

- We have a finite measure in Hilbert spaces
- The Gaussian is a reference measure in Hilbert space
- It is clearly not translation invariant it is quasi invariant
- Consider the Radon-Nikodym derivative between two Gaussian measures in $\mathbb{R}$ i.e. $\mu = \mathcal{N}(0, \sigma)$ and $\mu_m = \mathcal{N}(m, \sigma)$
- Then
$$\frac{d\mu_m}{d\mu}(x) = \exp\left(-\frac{m^2}{2\sigma^2} + \frac{xm}{\sigma^2}\right)$$
- Both μ and μ_m have positive densities with respect to Lebesgue measure
- It follows that for $\sigma \neq 0$ then for any $A \in \mathbb{R}$, $\mu(A) = 0$ if and only if $\mu_m(A) = 0$ if and only if Lebesgue$(A) = 0$
- The measures are *absolutely continuous* (or equivalent) with respect to each other

Singular Measures



Equivalence of Measures

- Things get a little more extreme in infinite dimensions
- Two measures are absolutely continuous if they both assign zero measure to the same sets
- Two measures are singular with respect to each other if one measure assigns zero to a set and the other measure assigns zero measure to the set complement
- In Hilbert space two Gaussian measures will either be equivalent or singular see Jupyter Notebook for example
- This has serious implications in the design of various Computational Statistical methods and Machine Learning algorithms as the following lecture will demonstrate

Equivalence of Measures

- Given a Gaussian measure $\mathcal{N}(0, C)$ on l^2 the shifted Gaussian measure $\mathcal{N}(v, C)$ is either equivalent to $\mathcal{N}(0, C)$ or they are singular
- There is a subspace containing all v for which the measures are equivalent
- The subspace for all translations v for which the measures are absolutely continuous, is the image of the space l^2 under the operator $C^{\frac{1}{2}}$
- This space is known as the Cameron-Martin space (also the Reproducing kernel Hilbert Space - RKHS)

Bayes Rule in Hilbert Space

- Bayes formula defines a posterior probability measure in terms of a likelihood and prior probability measure

$$\mathbb{P}(u|y) = \frac{\mathbb{P}(y|u)\mathbb{P}(u)}{\mathbb{P}(y)}$$

- defining this on an infinite dimensional Hilbert space the correspoding densities cannot be obtained
- However the Radon-Nikodym derivative between a posterior measure μ^y and a prior measure μ^0 can define the corresponding *likelihood function*

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

- The reference measure for a Hilbert space is the Gaussian Measure and so the prior measure over functions $\mu^0 = \mathcal{N}(m, c)$
- This is formal definition of a Gaussian Process prior - one of the major tools in Machine Learning