

Gaussian Measures on Hilbert Spaces

Kevin O'Connor

Abstract

We begin by providing some necessary mathematical background and defining Gaussian measures. Reproducing kernel spaces are then introduced and a proof of their existence and uniqueness is provided. Finally, we use these concepts to prove a fundamental result about the Gaussian white noise expansion.

1 Background

In this section, we briefly provide the reader with a few definitions and a basic result which will be useful later. These concepts lie mostly within the domain of functional analysis. Readers looking for a (much) more detailed treatment might refer to [4].

The types of spaces we will be interested in are Banach and Hilbert spaces, defined as follows:

Definition 1.1. A *Banach space* is a normed vector space which is complete with respect to the metric induced by its norm.

Definition 1.2. A pair $(E, \langle \cdot, \cdot \rangle)$ is a *Hilbert space* if the norm, $\|x\| = \sqrt{\langle x, x \rangle}$ is complete, or equivalently, $(V, \|\cdot\|)$ is a Banach space.

Why these two types of spaces? We will see later that completeness and inner product spaces will be defining features of reproducing kernel spaces. Furthermore, we will use Banach spaces as the sample space of the Gaussian measures that we consider. More generally, many of the most widely studied spaces of functions are Banach spaces and viewing them as Hilbert spaces allows us to consider them in ways similar to euclidean spaces of vectors in \mathbb{R}^n that one encounters in a standard linear algebra course.

We will also make use of the dual spaces of Banach spaces, defined below.

Definition 1.3. Let $(E, \|\cdot\|)$ be a Banach space. The *topological dual space* or simply *dual space* of E , denoted E^* is defined as

$$E^* = \{\varphi : E \rightarrow \mathbb{R} : \varphi \text{ is continuous and linear}\}$$

The definition above assumes that \mathbb{R} is the field of scalars underlying the vector space, E . At other times, this field may be \mathbb{C} instead. It should also be noted that frequently E will be a space of functions and so E^* will be a space of functionals, or maps from functions to the scalar field. A special property of Hilbert spaces is that their duals are also Hilbert spaces. Thus we get a natural inner product and completeness of the dual of any given Hilbert space as a result of the Riesz-Frèchet representation theorem which we state next without proof.

Theorem 1.4. (Riesz-Fréchet Representation Theorem) Let $(E, \langle \cdot, \cdot \rangle)$ be a Hilbert space and let $\varphi : E \rightarrow \mathbb{R}$ be a map. Then the following are equivalent:

1. φ is continuous and linear
2. There is a unique $x \in E$ with $\varphi(y) = \langle y, x \rangle, \forall y \in E$

The representation theorem says that if you have a continuous linear map on a Hilbert space, you can write it as an inner product with some unique point in the space. In the opposite direction, linearity follows simply by the definition of an inner product and continuity of the inner product can be proven using the Cauchy-Schwarz inequality. We will use this theorem later when discussing reproducing kernel spaces to show the existence of a unique reproducing kernel.

Finally, we define a continuous embedding.

Definition 1.5. Let $(X, |\cdot|_X), (Y, |\cdot|_Y)$ be two Banach spaces with $X \subseteq Y$. Then X is said to be **continuously embedded** in Y if the identity map,

$$id : X \hookrightarrow Y : x \mapsto x$$

is continuous.

We will see later that this property is a necessary aspect of the relationship between the RKHS and the ambient Banach space in which it resides.

2 Gaussian Measures

The objects of study in this document will be Gaussian measures, denoted $\mathcal{N}(\mu, \sigma^2)$. The reader will likely have encountered the Gaussian (Normal) distribution in an introductory statistics class and Gaussian measures are defined similarly. Recall that the density for a one-dimensional Gaussian distribution is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} \quad (2.1)$$

In multiple dimensions, if Σ is positive definite and thus Σ^{-1} exists,

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu) \right\} \quad (2.2)$$

Note that when we wish to make it clear what μ and σ^2 are in the density, we will use the notation, $f(x; \mu, \sigma^2)$. So then what makes a measure Gaussian? It is informative to consider the one-dimensional case and use this to define the multi-dimensional case.

Definition 2.1. (Gaussian Measure)

1. A measure, \mathbb{P}_1 , on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a Gaussian measure if it is either concentrated at a point, i.e. $\mathbb{P}_1 = \delta_\mu$, or has a density of the form, (2.1).
2. A probability measure, \mathbb{P} on a space $(E, \mathcal{B}(E))$, is called a Gaussian measure if

for any $x \in E$, $\exists \mu \in \mathbb{R}$, $\sigma \geq 0$, such that

$$\mathbb{P}(\{y \in E : \langle x, y \rangle \in A\}) = \mathbb{P}_1(A; \mu, \sigma^2)$$

What intuition can we take away from this definition? For $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, we expect that a Gaussian measure would measure sets in a Gaussian way, i.e. using a Gaussian density. And in arbitrary measure spaces with an inner product, a Gaussian measure should measure one-dimensional projections of points, as a Gaussian measure in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ would.

It can be shown that for a Gaussian measure \mathbb{P} with density $f(x; \mu, \sigma^2)$,

$$\int_{\mathbb{R}} x f(x; \mu, \sigma^2) dx = \mu$$

or one with density $f(\mathbf{x}; \mu, \sigma^2)$,

$$\int_{\mathbb{R}^n} (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T f(\mathbf{x}; \mu, \Sigma) d\mathbf{x} = \Sigma$$

Thus we refer to μ as the mean, σ^2 as the variance, and Σ as the covariance of \mathbb{P} . A Gaussian measure with $\mu = 0$ will be referred to as *symmetric*. Next we give the form of the characteristic function, $\phi(x)$, of a Gaussian measure on \mathbb{R} .

Theorem 2.2. A Gaussian measure on \mathbb{R} has a characteristic function of the form,

$$\phi(t) = \exp \left\{ i\mu t - \frac{1}{2} \sigma^2 t^2 \right\}$$

A Gaussian measure on an arbitrary Hilbert space, $(E, \langle \cdot, \cdot \rangle)$, has a characteristic function of the form,

$$\phi(\mathbf{t}) = \exp \left\{ i \langle \mathbf{t}, \mu \rangle - \frac{1}{2} \langle \Sigma \mathbf{t}, \mathbf{t} \rangle \right\}$$

Proof. We will prove the case when $n = 1$.

$$\begin{aligned} \phi(t) &= \int \exp \{ itx \} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} dx \\ &= \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} \left(x^2 - 2(\sigma^2 it + \mu)x + \mu^2 \right) \right\} dx \\ &= \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} \left(x^2 - 2(\sigma^2 it + \mu)x + (\sigma^2 it + \mu)^2 \right. \right. \\ &\quad \left. \left. - (\sigma^2 it + \mu)^2 + \mu^2 \right) \right\} dx \\ &= \exp \left\{ -\frac{1}{2\sigma^2} (\sigma^4 t^2 - 2\mu\sigma^2 it) \right\} \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} \left(x - (\sigma^2 it + \mu) \right)^2 \right\} dx \\ &= \exp \left\{ \mu it - \frac{\sigma^2 t^2}{2} \right\} \end{aligned}$$

■

One benefit of using the characteristic function instead of the density is the fact that the characteristic function exists even when Σ is singular, while the density does not. Later we will also use the characteristic function of a Gaussian measure to show weak convergence.

3 Reproducing Kernel Spaces for Gaussian Measures

In this section, we formally introduce the Reproducing Kernel Hilbert Space for Gaussian measures. This is an important subspace such that any element of the dual space has a symmetric Gaussian law with variance determined by the norm defined on the subspace. The details are given below.

Definition 3.1. For a Gaussian measure \mathbb{P} defined on a Banach space E , a linear subspace, $H \subset E$ with Hilbert norm $|\cdot|_H$ is a **reproducing (Hilbert) kernel space (RKHS)** for \mathbb{P} if H is complete, continuously embedded in E and for arbitrary $\varphi \in E^*$,

$$\mathcal{L}(\varphi) = \mathcal{N}(0, |\varphi|_H^2) \quad \text{where} \quad |\varphi|_H = \sup_{|x|_H \leq 1} |\varphi(x)|$$

Let's dig deeper into this definition. Let $H \subset E$ be a RKHS and fix an arbitrary $\varphi \in E^*$. It is a continuous linear map from $E \rightarrow \mathbb{R}$ by definition. Then we can apply the representation theorem to conclude that there is a unique point $h \in H$ such that $\varphi(f) = \langle h, f \rangle_H, \forall f \in H$. The function h is said to be the **reproducing kernel** for φ . We will refer to this characteristic in general as the **reproducing property** of the RKHS. Importantly, we can extend this property to the whole space! Here completeness and continuity allow us to take sequences of points in the RKHS converging to \mathbb{P} -almost any given point in the ambient space and show that the reproducing property holds more generally for $f \in E$ and $\varphi \in L^2(E, \mathcal{B}(E), \mathbb{P})$.

Remark 3.2. This leads us to the **reproducing kernel formula**,

$$\int_E \langle h, x \rangle_H \langle g, x \rangle_H \mathbb{P}(dx) = \langle h, g \rangle_H, \quad h, g \in H$$

You might be wondering, “Does this space even exist?” For a Gaussian measure on a Banach space, it does. In fact, for such measure spaces, it is unique! Below we walk through Da Prato and Zabczyk’s proof of these facts.

Theorem 3.3. For a Gaussian measure \mathbb{P} defined on a separable Banach space, there exists a unique reproducing kernel space $(H, |\cdot|_H)$.

Proof. (Existence): Fix $\varphi \in E^*$. Then since any Gaussian measure has finite moments, $\varphi \in L^2(E, \mathcal{B}(E), \mathbb{P})$. Furthermore, for any $\varphi \in \overline{E^*}$, where $\overline{E^*}$ denotes the closure of E^* in L^2 , it is an almost sure limit of functions $\{\varphi_n\}_{n \geq 1} \subset E^*$ with Gaussian distributions, so

$$\mathcal{L}(\varphi) = \mathcal{N}(0, |\varphi|_{L^2}^2)$$

Since φ is continuous and linear with symmetric law, $\varphi(0) = 0$. Now define the functional $J : \overline{E^*} \rightarrow E$ as

$$J(\varphi) = \int_E x \varphi(x) \mathbb{P}(dx)$$

Then J is one-to-one and continuous. To see that it is one-to-one, consider the case when $J(\varphi) = 0$. Then since φ is linear, we can distribute it into $J(\varphi)$. I.e.,

$$0 = \varphi(J(\varphi)) = \varphi\left(\int_E x\varphi(x)\mathbb{P}(dx)\right) = \int_E |\varphi(x)|^2\mathbb{P}(dx)$$

This implies $\varphi = 0$ a.s. Similarly, if $J(\varphi) \neq 0$, then

$$\varphi(J(\varphi)) = \int_E |\varphi(x)|^2\mathbb{P}(dx)$$

and since φ is uniquely identified by its L^2 norm, we may conclude that J is one-to-one. To see continuity, notice that linearity of φ and the integral gives us

$$\|J(\varphi)\|^2 = \left|\int_E x\varphi(x)\mathbb{P}(dx)\right|^2 = \left|\int_E |\varphi(x)|^2\mathbb{P}(dx)\right| \left|\int_E |x|^2\mathbb{P}(dx)\right| = \|\varphi\|_{L^2}^2 \int_E |x|^2\mathbb{P}(dx)$$

Notice then for $\psi \in \overline{E}^*$,

$$\begin{aligned} \|J(\varphi) - J(\psi)\|^2 &= \left|\int_E x\varphi(x)\mathbb{P}(dx) - \int_E x\psi(x)\mathbb{P}(dx)\right|^2 \\ &= \left|\int_E x(\varphi(x) - \psi(x))\mathbb{P}(dx)\right|^2 \\ &= \left|\int_E x(\varphi - \psi)(x)\mathbb{P}(dx)\right|^2 \\ &= \|J(\varphi - \psi)\|^2 \end{aligned}$$

So

$$\|J(\varphi) - J(\psi)\|^2 = \|\varphi - \psi\|_{L^2}^2 \int_E |x|^2\mathbb{P}(dx)$$

and J is continuous. Then $H = J(E)$ is a Hilbert space with inner product,

$$\langle J(\varphi), J(\psi) \rangle_H = \int_E \varphi(x)\psi(x)\mathbb{P}(dx)$$

Then letting $|\cdot|_H$ denote the norm in H , we will show that $|\cdot|_H = |\cdot|_{L^2}$. For any $\psi \in E^*$,

$$|\psi|_H = \sup_{|x|_H=1} |\psi(x)| = \sup_{|\varphi|_{L^2}=1} |\psi(J(\varphi))| = \sup_{|\varphi|_{L^2}=1} \left| \int_E \varphi(x)\psi(x)\mathbb{P}(dx) \right| = \sup_{|\varphi|_{L^2}=1} |\langle J(\varphi), J(\psi) \rangle_H|$$

Then using the Cauchy-Schwarz inequality,

$$|\psi|_H \leq |\psi|_{L^2} \sup_{|\varphi|_{L^2}=1} |\varphi|_{L^2} = |\psi|_{L^2}$$

The upper bound is achieved by taking $\varphi = \psi/|\psi|_{L^2}$, which gives us $|\psi|_H = |\psi|_{L^2}$. Then we conclude that H is a reproducing kernel space for \mathbb{P} .

(Uniqueness): Suppose there is another Hilbert space, $\tilde{H} \subset E$, with inner product, $\langle \cdot, \cdot \rangle_{\tilde{H}}$,

which is a reproducing kernel space for \mathbb{P} . Let $\varphi \in E^*$ be arbitrary. By the reproducing property, $\exists \tilde{J}\varphi \in \tilde{H}$ unique such that $\forall h \in \tilde{H}$,

$$\varphi(h) = \langle \tilde{J}\varphi, h \rangle_{\tilde{H}}, \quad \forall h \in \tilde{H}$$

and

$$|\tilde{J}\varphi|_{\tilde{H}}^2 = \langle \tilde{J}\varphi, \tilde{J}\varphi \rangle_{\tilde{H}} = \int_E |\varphi(x)|^2 \mathbb{P}(dx) = |\varphi|_{L^2}^2$$

In other words, we are denoting the reproducing kernel for a given $\varphi \in E^*$ in RKHS \tilde{H} as $\tilde{J}\varphi$. Then for $\varphi \in E^*$ and $\psi \in \bar{E}^*$,

$$\varphi(\tilde{J}\psi) = \langle \tilde{J}\varphi, \tilde{J}\psi \rangle_{\tilde{H}} = \int_E \varphi(x)\psi(x)\mathbb{P}(dx) = \varphi(J(\psi))$$

Thus $\tilde{J}(\psi) = J(\psi)$ so H is isometrically embedded in \tilde{H} . Now if $H \neq \tilde{H}$, $\exists \tilde{h} \in \tilde{H}$, with $\tilde{h} \notin H$ and $\tilde{h} \neq 0$. But for all $\varphi \in E^*$,

$$\varphi(\tilde{h}) = \langle \tilde{h}, J\varphi \rangle_{\tilde{H}} = 0$$

This implies that $\tilde{h} = 0$ and $\tilde{H} = H$, which proves uniqueness. ■

We will now use the concept of a RKHS to give a result white noise expansions. In particular, this result shows that a Gaussian measure can be expressed as in terms of the orthonormal basis elements of its RKHS. Again, we follow Da Prato and Zybczyk's proof, filling in details when necessary.

Theorem 3.4. *Let \mathbb{P} be a symmetric Gaussian measure on a separable Banach space E and let $H_{\mathbb{P}}$ be its reproducing kernel space. Let $\{e_n\}$ be an orthonormal and complete basis in $H_{\mathbb{P}}$ and $\{\xi_n\}$ a sequence of i.i.d. standard normal random variables. Then $\sum_{k=1}^{\infty} \xi_k e_k$ converges \mathbb{P} -a.s. in E and*

$$\mathcal{L} \left(\sum_{k=1}^{\infty} \xi_k e_k \right) = \mathbb{P}$$

Proof. As before, let \bar{E}^* be the closure of E^* in $L^2(E, \mathcal{B}(E), \mathbb{P})$. Also define the maps, $\varphi_n \in \bar{E}^*$, such that $J\varphi_n = e_n, \forall n \geq 1$. Denote

$$\mathbb{P}_n = \mathcal{L} \left(\sum_{k=1}^n \xi_k e_k \right), \quad n \geq 1$$

We will prove that $\{\mathbb{P}_n\}$ is tight. Let $S(x) = x$ and

$$S_N(x) = \sum_{n=1}^N \varphi_n(x) e_n = \sum_{n=1}^N \varphi_n(x) J\varphi_n, \quad x \in E$$

For $\varphi \in E^*$, consider the sequence,

$$\varphi(S_N(x)) = \varphi \left(\sum_{n=1}^N \varphi_n(x) e_n \right) = \sum_{n=1}^N \varphi(\varphi_n(x) e_n) = \sum_{n=1}^N \varphi_n(x) \varphi(e_n)$$

We will show that this sequence is a.s. convergent. Since $\forall n \geq 1$, φ_n has mean 0 and variance 1, $\varphi_n(x)\varphi(e_n)$ has mean 0 and variance $|\varphi(e_n)|_{H_{\mathbb{P}}}^2$, $\forall n \geq 1$. Then since

$$\sum_{n=1}^{\infty} |\varphi(e_n)|_{H_{\mathbb{P}}}^2 = \sum_{n=1}^{\infty} \langle J\varphi, e_n \rangle_{H_{\mathbb{P}}}^2 \leq \sum_{n=1}^{\infty} |J\varphi|_{H_{\mathbb{P}}}^2 |e_n|_{H_{\mathbb{P}}}^2 = |J\varphi|_{H_{\mathbb{P}}}^2 < \infty$$

the Kolmogorov two-series theorem implies that $\varphi(S_N(x))$ converges a.s. as $N \rightarrow \infty$. Additionally,

$$\mathcal{L} \left(\lim_{N \rightarrow \infty} S_N(x) \right) = \mathcal{L} \left(\sum_{n=1}^{\infty} \varphi_n(x) e_n \right) = \mathcal{N}(0, |J\varphi|_{H_{\mathbb{P}}}^2)$$

Since $\{\varphi_n\}$ is a complete orthonormal basis in $\overline{E}^* \subset L^2$, for $\varphi \in E^*$ we have in particular,

$$\varphi(x) = \sum_{n=1}^{\infty} \varphi_n(x) \langle \varphi, \varphi_n \rangle_{L^2} = \sum_{n=1}^{\infty} \varphi_n(x) \int_E \varphi_n(y) \varphi(y) \mathbb{P}(dy) \quad a.s.$$

in L^2 . Then using the reproducing kernel formula and representation of φ ,

$$\varphi(Sx) = \varphi(x) = \sum_{n=1}^{\infty} \varphi_n(x) \langle J\varphi, J\varphi_n \rangle_{H_{\mathbb{P}}} = \sum_{n=1}^N \varphi_n(x) \varphi(e_n) = \lim_{N \rightarrow \infty} \varphi(S_N(x)), \quad a.s.$$

So for $N \geq 1$, linearity of φ gives us

$$\lim_{M \rightarrow \infty} \varphi(S_M(x) - S_N(x)) = \varphi(S(x) - S_N(x))$$

and independence of $S_M - S_N$ and S_N implies that $S - S_N$ and S_N are independent. Now let \mathbb{P}_M^{\perp} denote the distribution of $S - S_M$, $M \geq 1$. Then

$$\mathbb{P}_M * \mathbb{P}_M^{\perp} = \mathbb{P}, \quad M \geq 1$$

where $*$ denotes the convolution of the two measures. For fixed ϵ , let K be a compact set with $\mathbb{P}(K) > 1 - \epsilon$. So

$$\mathbb{P}(K) = \mathbb{P}_M * \mathbb{P}_M^{\perp}(K) = \int \mathbb{P}_M(K - x) \mathbb{P}_M^{\perp}(dx) > 1 - \epsilon$$

Then there must $\exists x_0 \in E$ such that $\mathbb{P}_M(K - x_0) \geq 1 - \epsilon$. Now by symmetry, we have $\mathbb{P}_M(K - x_0) = \mathbb{P}_M(-K + x_0)$. Furthermore, noticing the following set inclusion,

$$(K - x_0) \cap (-K + x_0) \subseteq \frac{K - K}{2}$$

monotonicity gives

$$\mathbb{P}_M((K - K)/2) \geq \mathbb{P}_M((K - x_0) \cap (-K + x_0))$$

By definition, we have

$$\mathbb{P}_M((K - x_0)^c) = \mathbb{P}_M((-K + x_0)^c) \leq \epsilon$$

so subadditivity gives

$$\mathbb{P}_M((K - x_0)^c \cup (-K + x_0)^c) \leq 2\epsilon$$

which implies

$$\mathbb{P}_M((K - K)/2) \geq \mathbb{P}_M((K - x_0) \cap (-K + x_0)) \geq 1 - 2\epsilon$$

Since $(K - K)/2$ is also compact, we conclude that $\{\mathbb{P}_n\}$ is tight. Similarly, one can show that $\{\mathbb{P}_M^\perp\}$ is tight. Based on our earlier work in the section on Gaussian measures, it can be shown that for $\varphi \in E^*$,

$$\begin{aligned} \hat{\mathbb{P}}(\varphi) &= \int_E e^{i\varphi(x)} \mathbb{P}(dx) = \exp \left\{ -\frac{1}{2} |J\varphi|_{H_{\mathbb{P}}}^2 \right\} \\ \hat{\mathbb{P}}_M(\varphi) &= \int_E e^{-\varphi(S_M x)} \mathbb{P}(dx) = \exp \left\{ -\frac{1}{2} \sum_{n=1}^M \langle J\varphi, e_n \rangle_{H_{\mathbb{P}}} \right\} \\ \hat{\mathbb{P}}_M^\perp(\varphi) &= \int_E e^{-\varphi(x - S_M x)} \mathbb{P}(dx) = \exp \left\{ -\frac{1}{2} \sum_{n=M+1}^{\infty} \langle J\varphi, e_n \rangle_{H_{\mathbb{P}}} \right\} \end{aligned}$$

Thus

$$\lim_{M \rightarrow \infty} \hat{\mathbb{P}}_M(\varphi) = \hat{\mathbb{P}}(\varphi) \quad \text{and} \quad \lim_{M \rightarrow \infty} \hat{\mathbb{P}}_M^\perp(\varphi) = 1$$

So any weakly convergent subsequence of $\{\mathbb{P}_M\}$ has the same limit and thus \mathbb{P}_N converges to \mathbb{P} weakly. Similarly, $\{\hat{\mathbb{P}}_M^\perp\}$ converges to $\delta_{(0)}$ weakly and thus $S - S_M \xrightarrow{\mathbb{P}} 0$. This implies that $S_M \xrightarrow{a.s.} S$. ■

References

- [1] Klenke, Achim. *Probability theory: a comprehensive course*. Springer Science & Business Media, 2013.
- [2] Da Prato, Giuseppe, and Jerzy Zabczyk. *Stochastic equations in infinite dimensions*. Cambridge university press, 2014.
- [3] Paulsen, V., & Raghupathi, M. (2016). *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces* (Cambridge Studies in Advanced Mathematics). Cambridge: Cambridge University Press. doi:10.1017/CBO9781316219232
- [4] Rudin, Walter. *Functional analysis*. International series in pure and applied mathematics. (1991).