Summary of Gaussian process and Gaussian measure

Tianpei Xie

Jul. 16th., 2015

Contents

1 Definitions and summary

1.1 Measure space on infinite dimensional vector space

- **Definition** [Taylor and Lay, 1958]
  A *topological vector space (TVS)* X is a vector space equipped with a topology on X so that the vector addition and scalar product operations are both continuous.

  A complete normed linear space is called a *Banach space*; A complete inner product space is called a *Hilbert space*.

  The *(algebraic) dual* space V^* of a vector space V consists of all linear functionals on V together with a naturally induced linear structure. The *continuous dual space* X^* of a topological vector space X corresponds to all continuous linear functionals, which is a subspace of (algebraic) dual space.

- **Definition** [Schaefer and Wolff, 1999]
  A topological space X is *locally convex Hausdorff* if it is a Hausdorff space such that every neighborhood of any $x \in E$ contains a convex neighborhood of x.

  Similarly, X is *locally compact Hausdorff (LCH)*, if every neighborhood U for every $x \in U$ has a compact neighborhood $C \subset U$, $x \in C$.

  Equivalently, a *locally convex space (LCS)* is defined to be a vector space V along with a family of *seminorms* $\{p_\alpha\}_{\alpha \in A}$ on V, a semi-norm with $p_\alpha(u) = 0 \Rightarrow u = 0$ is a norm.

- **Definition** Let X be locally convex space, a n-dimensional *cylinder set* as [Lifshits, 2013]

  $$C_A[f_1, \ldots, f_n] \equiv \{\boldsymbol{x} \mid (f_1(\boldsymbol{x}), \ldots, f_n(\boldsymbol{x})) \in A\}, n = 1, 2, \ldots,$$

  for any $A \in \mathcal{B}(\mathbb{R}^n)$, $A_i \in \mathcal{B}(\mathbb{R})$, $f_i \in X^* \subset \mathbb{R}^X$, the dual space of continuous linear functional on X.

  The collection of $C_A[f_1, \ldots, f_n]$ with all possible $A \in \mathcal{B}(\mathbb{R}^n)$, and all $f_i \in X^* \subset \mathbb{R}^X$ is denoted as $\mathscr{C}_n$.

- If the underling space is sample space $X \equiv \Omega$, then $C_A[\xi_1, \ldots, \xi_n]$ is a measureable set induced by a collection of random variables $\{\xi_t, t \geq 1\}$.

- **Definition** [Lifshits, 2013]
  The collection of all cylinder sets $\mathscr{C}_n$ for all finite dimensions $n \geq 1$ is referred as the *algebra of cylinder sets*, denoted as $\mathscr{C}_0$. That is, $\mathscr{C}_0 \equiv \bigcup_{n=1}^{\infty} \mathscr{C}_n$, where $\mathscr{C}_n$ is denotes as $\mathcal{B}^n \times X^* \times X^* \times X^* \times \cdots$.

  The collection $\mathscr{C}_0$ forms an algebra (closed under complements and finite union). If $X = X^*$, the $\mathscr{C}_0$ is the basis for the product topology X^∞

- **Definition** The σ-algebra $\mathscr{C} = \sigma(\mathscr{C}_0)$ generated from the algebra of cylinders $\mathscr{C}_0$ is called *cylinderical σ-algebra.*

  If $X \equiv \Omega$, with random variables $\{\xi_t, t \geq 1\}$, $\mathscr{C} \supset \sigma(\xi_t, t \geq 1)$ is the sigma-algebra generated by random variables $\{\xi_t, t \geq 1\}$.

- **Definition** The *Borel σ-algebra* $\mathscr{B}$ on the TVS X is generated by all open/closed sets in topology of X.

- $\mathscr{C} \subset \mathscr{B}$.

- Let X be a (infinite dimensional) vector space. A cylindrical measure space is denoted as $(X, \mathscr{C}, \mu)$, where μ is a measure on $\mathscr{C}$.

  In particular, X is locally compact Hausdorff, and $\mathscr{C}$ is the cylindrical σ-algebra generated by all cylinder sets via continuous linear functionals on X,

  If μ is a *Radon measure* on LCH X, i.e. it is *inner regular* (inner approximated via compact set), *outer regular* (outer approximated by open set) and *locally finite* (every point is covered by an open set with finite measure), then it can be uniquely extended to the Borel σ-algebra. That is, $(X, \mathscr{B}, \mu)$ is defined. [Lifshits, 2013]

- [Folland, 2013] A Radon measure is an extension of Lebesgue measure in $\mathbb{R}^d$ to a LCH X. A *Radon measure* on a *locally compact Hausdorff* space can be expressed in terms of *continuous linear functionals* on the space of *continuous functions with compact support*. (A Radon measure is real then it can be decomposed into the difference of two positive measures.)

- In sum, the measure space of interest is $(X, \mathscr{B}, \mu)$, where X is *locally compact Hausdorff (LCH)* space (a topological vector space), $\mathscr{B}$ is the Borel σ-algebra, including a collection $\mathscr{C}_0$ of all cylinder sets for all continuous linear functionals on X, μ is a *(Radon) measure* on $\mathscr{B}$.

1.2 Random functions, dual space and Gaussian measure

- Consider now the probability space $(\Omega, \mathscr{F}, \mathbb{P})$, where Ω is the sample space, $\mathscr{F}$ is a σ-algebra containing $\mathscr{C}_0$ of all cylinder sets for any family of random variables $\xi \equiv \{\xi_x, x \in E\}$ on Ω

  $$C_A[\xi_{x_1}, \ldots, \xi_{x_n}] \equiv \{\omega \mid (\xi_{x_1}(\omega), \ldots, \xi_{x_n}(\omega)) \in A\}, n = 1, 2, \ldots,$$

  for any $A \in \mathcal{B}(\mathbb{R}^n)$, $\mathbb{P}$ is a probability measure on $\mathscr{F}$. (Assuming topology on Ω, then P should be a Radon measure.)

- **Definition** [Lifshits, 2013, Rasmussen and Williams, 2005]
  A family of random variables $\xi_\cdot \equiv \{\xi_x, x \in E\}$ defined on $(\Omega, \mathscr{F}, P)$ is called a *random function*, where E is the index set or input domain. In specific,

  1. a *random function* is a mapping $\xi_\cdot : E \times \Omega \to \mathbb{R}$, with each finite-dimensional vector $(\xi_{x_1}, \ldots, \xi_{x_n}) : (\Omega, \mathscr{F}) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ being measureable for every $(x_1, \ldots, x_n) \subset E$, for all $n \geq 1$.

  2. Also, a *random function* is a measurable mapping $\Xi : (\Omega, \mathscr{F}) \to (\mathbb{R}^E, \mathscr{B})$, where $\mathbb{R}^E = \{f : E \to \mathbb{R}\}$ is the set of all functions from E to $\mathbb{R}$, $\mathscr{B} = \mathcal{B}(\mathbb{R}^E)$ is a Borel

σ-algebra generated by the product topology on $\mathbb{R}^E$. Here Ξ is the infinite sequence as the evaluation of $\xi.$ at all $x \in E$, i.e.

$$\Xi(\omega) \equiv (\xi_x(\omega), x \in E)\,.$$

3. If $E \equiv T \subset \mathbb{R}$, it is called a *random process*, whereas for $E \subset \mathbb{R}^n$, it is called a *random field*.
4. Assume that E is a *separable* metric space (i.e. E has dense countable subset E'); without loss of generality, assume that E is countable.
5. Given a sample point $\omega \in \Omega$, $\xi.(\omega) \in \mathbb{R}^E$ is a real-valued function on E, which is called a *sample function* of the random function, or a *sample path* of the random process for $T \subset \mathbb{R}$. A random function is an infinite-dimensional random vector

- **Definition** [Lifshits, 2013]
  1. The induced probability on $\mathbb{R}^E$ is given by
  $$\mathcal{P}_{\boldsymbol{\xi}}\left(A\right) = \mathbb{P}\left\{\omega : (\xi_x(\omega), x \in E) \in A\right\} = \mathbb{P} \circ \Xi^{-1}(A); \quad \forall A \in \mathscr{B} = \mathcal{B}(\mathbb{R}^E)$$
  is called the *distribution of random functions*, denoted as $\mathcal{P}_{\boldsymbol{xi}}$. It is a probability *measure* on the *space of sample functions*.
  2. Given the distribution $\mathcal{P}_{\boldsymbol{\xi}}$ of random functions ξ_x, the space $(\mathbb{R}^E, \mathscr{B}, \mathcal{P}_{\boldsymbol{\xi}})$ can be seen as a *finite measure space* on the space of functions $\mathbb{R}^E$.
  3. For each finite-dimensional joint distribution $\mathcal{P}_{\xi,n}, \quad d\mathcal{P}_{\xi,n} \ll dx^n$ is dominated by Lebesgue measure on $\mathbb{R}^n$.

- The *dual* space Ω^* contains the space of all random variables. Note that each random variable ξ is a linear functional on Ω, i.e. $(\xi + \eta)(\omega) = \xi(\omega) + \eta(\omega)$.

- **Definition** [Lifshits, 2013]
  The *barycenter* $\omega_m \in \Omega$ is defined for random function $\xi. \in L^1(\Omega, \mathbb{P})$ so that
  $$\xi.(\omega_m) = \int_\Omega \xi.(\omega)\, d\,\mathbb{P},$$
  where the value $m. \equiv \xi.(\omega_m) \in \mathbb{R}^E$ is referred as the *mean function (mean path)* of random function. Note that for each $x \in E$, we can find a barycenter $\omega_{m,x}$, wherease ω_m is the *common barycenter* for all $\xi_x \in L^1(\Omega, \mathbb{P})$ for all $x \in E$.

- **Definition** [Lifshits, 2013]
  The *covariance operator* $K : \Omega^* \to \Omega$ is a linear operator from the dual space to the sample space, so that for any linear functionals (random variables) $\xi_x, \xi_z \in \mathcal{L}^2(\Omega, \mathbb{P})$, then
  $$\begin{aligned}\xi_x\left(K\xi_z\right) &\equiv \int_\Omega \xi_x(\omega - \omega_m)\xi_z(\omega - \omega_m)\, d\,\mathbb{P} \\ &= \int_\Omega (\xi_x(\omega) - m_x)(\xi_z(\omega) - m_z) d\,\mathbb{P}\end{aligned}$$

K is a *self-adjoint* operator, i.e. $\xi_x (K\xi_z) = \xi_z (K\xi_x)$.

- **Definition** [Lifshits, 2013]
A measure $\mathbb{P}$ on $(\Omega, \mathscr{F})$ is *Gaussian measure* if and only if for all linear functions $f \in \Omega^*$, the induced probability

$$\mathcal{P}_f \equiv \mathbb{P} \circ f^{-1} \in \mathcal{G}$$

is a Gaussian distribution on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$; i.e. the dual Ω^* is the space of all Gaussian-distributed random variables $f(\cdot) : \Omega \to \mathbb{R}$.

1.3 Prescribed probability space of sample functions

- Now consider the *induced probability space* $(F, \mathscr{B}_F, \mathcal{P}_{\boldsymbol{\xi}})$, where $F \equiv F_{\boldsymbol{\xi}}$ is the space of all sample functions of $\boldsymbol{\xi} \equiv (\xi_x, x \in E)$, $\mathscr{B}_F = \mathscr{B}|_F$ is the restriction of Borel σ-algebra to F, $\mathcal{P}_{\boldsymbol{\xi}}$ is the distribution of random functions ξ. (or, a measure of sample functions). Here, each point $f \in F$ is given as $f \equiv f_\omega(\cdot) = \xi(\cdot, \omega)$ for some $\omega \in \Omega$.

- **Definition** The *distribution of a random function* ξ, $\mathcal{P}_{\boldsymbol{\xi}}$ is a *Gaussian distribution* on the induced measure space $(F, \mathscr{B}_F, \mathcal{P}_{\boldsymbol{\xi}})$, if and only if for any linear functionals $I \in F^*$, the induced probability (i.e. the distribution of linear functional I w.r.t. $\mathcal{P}_{\boldsymbol{\xi}}$) as

$$\mathcal{P}_I \equiv \mathcal{P}_{\boldsymbol{\xi}} \circ I^{-1} \in \mathcal{G}$$

is a Gaussian distribution on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Equivalently, $\mathcal{P}_{\boldsymbol{\xi}}$ is Gaussian if and only if any finite-dimensional distribution $\mathcal{P}_{\xi,n}$ is Gaussian.

- **Definition** [Lifshits, 2013]
Given the prescribed probability space $(F, \mathscr{B}_F, \mathcal{P}_{\boldsymbol{\xi}})$, where $\mathcal{P}_{\boldsymbol{\xi}} \in \mathcal{G}(F)$ is a Gaussian measure, it is possible to construct a space of $\mathcal{P}_{\boldsymbol{\xi}}$-*measureable linear functionals* $F_P^* \subset F^*$. In particular,

$$\begin{aligned}
&F^* \subset \mathcal{L}^2(F, \mathcal{P}_{\boldsymbol{\xi}}) \\
&F_P^* = \overline{F^*} \subset \mathcal{L}^2(F, \mathcal{P}_{\boldsymbol{\xi}}) \\
\text{i.e.} \quad &\int_F (I(f))^2 \, \mathcal{P}_{\boldsymbol{\xi}}(df) < \infty, \quad \text{for any } I(\cdot) \in F_P^*.
\end{aligned}$$

The first statement is implied by the definition of Gaussian measure above, since the sample function $f \sim \mathcal{P}_{\boldsymbol{\xi}} \in \mathcal{G}(F)$ if and only if the distribution of linear functional $I(\cdot) \in F^*$, P_I is a univariate Normal distribution, i.e. $P_I \in \mathcal{G}(\mathbb{R})$. In the second statement, the closure is w.r.t. $\mathcal{L}^2$ metric topology induced from $\mathcal{L}^2(F, \mathcal{P}_{\boldsymbol{\xi}})$.

- **Definition** [Lifshits, 2013]
Given the prescribed probability space $(F, \mathscr{B}_F, \mathcal{P}_{\boldsymbol{\xi}})$, where $\mathcal{P}_{\boldsymbol{\xi}}$ is a measure of sample function, the *kernel* of the measure $\mathcal{P}_{\boldsymbol{\xi}}$ is a *linear* subspace of F, denoted as H_P, such that

$$H_P \equiv \{h \in F : \ \mathcal{P}_{ah,\boldsymbol{\xi}} \ll \mathcal{P}_{\boldsymbol{\xi}}, \forall a \text{ const.}\}$$

where $\mathcal{P}_{h,\boldsymbol{\xi}}(A) \equiv \mathcal{P}_{\boldsymbol{\xi}}(A-h)$ for any $A \in \mathscr{B}_F$. $h \in H_P$ is called an *admissible shift*.

If $\mathcal{P}_{\boldsymbol{\xi}} \in \mathcal{G}(F)$, then the kernel H_P of measure $\mathcal{P}_{\boldsymbol{\xi}}$ is a Reproducing Kernel Hilbert space (RKHS) in which the covariance function is the reproducing kernel. In fact, $H_P \simeq F_P^*$, therefore H_P is a Hilbert space. Also

$$K(F^*) \subseteq H_P \subseteq F,$$

where $K : F^* \to F$ is the covariance operator. In fact, we will show later that H_P is a RKHS.

Moreover, H_P is equal to the topological support, i.e. $\mathcal{P}_{\boldsymbol{\xi}}(H_P(K)) = 1$,

$$\begin{aligned} H_P &= \operatorname{supp}(\mathcal{P}_{\boldsymbol{\xi}}) \\ &= \bigcap_{\substack{\eta \text{ degenerate} \\ I^*\eta=0}} \{f \in F : \eta(f) = 0\} \end{aligned}$$

where the degenerate means that $\eta : F \to \mathbb{R}$ as a random variable has zero variance

- **Definition** Suppose given the finite measure space $(F, \mathscr{B}_F, \mathcal{P}_{\boldsymbol{\xi}})$ induced from the random function $\boldsymbol{\xi}$. If, furthermore, $F \equiv C(E) \subset \mathbb{R}^E$ is the space of all *continuous* sample functions on E, then

  1. dual space F^* is isomorphic to the space of all *Baire measures* on E so that for every $\eta \in F^*$, a unique representation is given as

     $$\eta(f) = \int_E f d\mu_\eta$$

     where μ_η is the Baire measures on $(E, \mathscr{A}, \nu)$ associated with η, and $\mathscr{A}$ is the Borel (Baire) σ-algebra on index set E.[Reed and Simon, 1980, Lifshits, 2013]

  2. Define a *barycenter* $f_m \in F$ for a L^1 linear functional $\eta \in L^1(F, \mathcal{P}_{\boldsymbol{\xi}})$ as

     $$\eta(f_m) = \int_F \eta(f) d\mathcal{P}_{\boldsymbol{\xi}},$$

     with the mean value $m_\eta = \eta(f_m)$ as a linear functional $m_\eta(f_m) = \eta(f_m)$.

  3. [Lifshits, 2013]
     Define the *covariance operator* $K : F^* \to F$ as a linear operator from the space of $\mathcal{L}^2$ integrable linear functionals to the function space, so that for any continuous linear

functionals $\zeta, \eta \in F^* \subset \mathcal{L}^2(F, \mathcal{P}_{\boldsymbol{\xi}})$, then

$$
\begin{aligned}
\zeta\left(K\eta\right) &\equiv \int_F \zeta(f - f_m)\eta(f - f_m)\mathcal{P}_{\boldsymbol{\xi}}(df) \\
&= \int_F (\zeta(f) - m_{\zeta})(\eta(f) - m_{\eta})\mathcal{P}_{\boldsymbol{\xi}}(df) \\
&\text{suppose zero mean } m_{\zeta} = m_{\eta} = 0 \\
&= \int_F \left[\int_E f(t)\mu_{\zeta}(dt) \int_E f(s)\mu_{\eta}(ds)\right] \mathcal{P}_{\boldsymbol{\xi}}(df) \\
&= \int_{E\times E} \left[\int_F f(t)f(s)\mathcal{P}_{\boldsymbol{\xi}}(df)\right] \mu_{\zeta}(dt)\mu_{\eta}(ds) \\
&\equiv \int_{E\times E} K(t,s)\; \mu_{\zeta}(dt)\mu_{\eta}(ds) \\
&\equiv \zeta\left(\int_E K(\cdot, s)\mu_{\eta}(ds)\right)
\end{aligned}
$$

where μ_{η}, μ_{ζ} are the associated Baire measure on E. K is a integral kernel with

$$
f(\cdot) \equiv (K\eta)(\cdot) = \int_E K(\cdot, s)\mu_{\eta}(ds) \in F
$$

Still, K is a *self-adjoint* operator, i.e. $\zeta\left(K\eta\right) = \eta\left(K\zeta\right)$. Here the covariance of output $f_t = \pi_t(f)$ and $f_s = \pi_s(f)$ is given by

$$
\begin{aligned}
K(t,s) &\equiv \pi_t\left(K\pi_s\right) \\
&= \int_F f(t)f(s)\mathcal{P}_{\boldsymbol{\xi}}(df) = \mathbb{E}_{\mathcal{P}_{\boldsymbol{\xi}}(f)}\left[f(t)f(s)\right]
\end{aligned}
$$

where π_t, π_s is the evaluation functional in F^* and $K : E \times E \to \mathbb{R}$ is the associated kernel function, which is continuous on $E \times E$.

4. The reproducing property: Since $F = C(E) \subset \mathcal{L}^2(E, \nu)$ forms a Hilbert space, then

$$
\begin{aligned}
F_P^* \ni \eta(g) &= \int_E g(s)\mu_{\eta}(ds) \quad \text{(by Riesz-Markov theorem)} \\
&= \int_E g(s)f_{\eta}(s)\nu(ds) \quad \text{(by Riesz Representation theorem)} \\
H_P \ni f_{\eta}(t) &= \int_E K(t,s)\mu_{\eta}\nu(ds) = \int_E K(t,s)f_{\eta}(s)\nu(ds), \qquad t \in E
\end{aligned}
$$

for any $f_{\eta} = I\eta = K\eta \in H_P$.

1.4 The properties of Gaussian measure

- The unique properties of the family of Gaussian measure $\mathcal{G}(X)$ on X,

  1. Sufficient statistic: A Gaussian distribution $\mathcal{N} \in \mathcal{G}(X)$ is determined uniquely by the mean m and covariance operator K.

2. Invariant under linear transformation: For $f \sim \mathcal{N} \in \mathcal{G}(X)$, then any linear functional $I(f) \sim \mathcal{N}_I \in \mathcal{G}(\mathbb{R})$.
3. Invariant under translation operation: $f \sim \mathcal{N}(m, K)$, then for any $h \in \text{supp}(\mathcal{N})$, $f + h \sim \mathcal{N}(m + h, K)$.

   The closure of the kernel of Gaussian measure $\overline{H_{\mathcal{N}}}$ is equal to the topological support $\text{supp}(\mathcal{N})$ of the measure $\mathcal{N}$.

   In general, a Borel set A that is invariant w.r.t. the kernel H_p has Gaussian measure $\mathcal{N}(A) = 0$ or $\mathcal{N}(A) = 1$.
4. For a set of Gaussian random variables $\xi_n \xrightarrow{P} \xi$ if and only if $\xi_n \xrightarrow{L^1} \xi$.
5. For stationary Gaussian process, the covariance kernel has a spectral representation.
6. The unique probability measure that has maximum (differential) entropy under the second-order and first-order moment constraint is the Gaussian measure.
7. For a Gaussian measure, the sample path is either bounded almost surely or unbounded almost surely.
8. The oscillation of a Gaussian random function (supremum of absolute difference) is equal to a constant value almost surely. In specific, it does not depend on the sample points.

2 Useful facts

- (The *standard Gaussian measure* on $\mathbb{R}^\infty$). [Lifshits, 2013]
  Assume a Gaussian sequence $\{\xi_t, t \in \mathbb{N}\}$ is defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. To construct $\mathcal{P}$ on sequences in $\mathbb{R}^\infty$, consider a mapping $\Xi : \Omega \to \mathbb{R}^\infty$ by formula,

$$\Xi(\omega) \equiv \{\xi_t(\omega) : t \in \mathbb{N}\} \equiv \{z_t, t \geq 1\} \in \mathbb{R}^\infty.$$

  Then the distribution of Gaussian sequence $\{\xi_t, t \geq 1\}$ is given by $\mathcal{P}_{\boldsymbol{\xi}} = \mathbb{P} \circ \Xi^{-1}$. In general, for any n-dimensional joint distribution

$$P_{\boldsymbol{\xi}}(d\boldsymbol{z}) \equiv d\mathcal{P}_{\boldsymbol{\xi},n}(z_1, \ldots, z_n) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\sum_{k=1}^{n} z_k^2\right\} d\boldsymbol{z}.$$

  The prescribed probability space of sample Gaussian sequences is $(\mathbb{R}^\infty, \mathscr{B}, \mathcal{P}_{\boldsymbol{\xi}})$, where $\mathscr{B} = \Xi\mathscr{F} = \{A \subset \mathbb{R}^\infty \mid \Xi^{-1}(A \cap \ell^2) \in \mathscr{F}\}$. Both the *topological support* and the *kernel* of the Gaussian measure $\mathcal{P}_{\boldsymbol{\xi}}$ is $\ell^2 = \left\{\boldsymbol{z} : \sum_{i=1}^\infty z_i^2 < \infty\right\}$, the most important *Hilbert* space in $\mathbb{R}^\infty$. Also $(\ell^2)^* = \ell^2$ is closed, so it is equivalent to define the probability space as $(\ell^2, \mathscr{B}, \mathcal{P}_{\boldsymbol{\xi}})$.

  Note that by definition, the cylindrical σ-algebra $\mathscr{C}$ is equal to the Borel σ-algebra $\mathscr{B}$ on $\mathbb{R}^\infty$, since the cylinder set $\{\boldsymbol{z} \in \mathbb{R}^\infty : (z_1, \ldots, z_n) \in A\}$ is the subbasis for the product topology on $\mathbb{R}^\infty$. The distribution of $\mathcal{P}_{\boldsymbol{\xi}}$ is in fact a Radon Gaussian measure.

  We have

$$\begin{aligned}
m_t &= \int_\Omega \xi_t(\omega) d\mathbb{P} = \int_{\ell^2} z_t P_{\boldsymbol{\xi}}(d\boldsymbol{z}) \\
&= 0 \\
\operatorname{cov}(\xi_t, \xi_s) \equiv \xi_t(K\xi_s) &= \int_\Omega (\xi_t(\omega) - m_t)(\xi_s(\omega) - m_s)\, d\mathbb{P} \\
&= \int_{\ell^2} (z_t - m_t)(z_s - m_s)\, P_{\boldsymbol{\xi}}(d\boldsymbol{z}) \\
&= \delta_s(t) = \begin{cases} 1 & s = t \\ 0 & s \neq t \end{cases}
\end{aligned}$$

  The dual space $(\mathbb{R}^\infty)^*$ has basis as $\{\pi_k, k \geq 1\}$, where $\pi_k(\boldsymbol{z}) = z_k$ is the evaluation map. In particular, for any $\mathcal{P}_{\boldsymbol{\xi}}$-measureable linear functional $z(\cdot) \in (\mathbb{R}^\infty)^*_P \equiv \overline{(\mathbb{R}^\infty)^*} \subset \mathcal{L}^2(\mathbb{R}^\infty, \mathcal{P}_{\boldsymbol{\xi}})$, we have

$$z(\cdot) = \sum_{k=1}^\infty z_k\, \pi_k(\cdot), \quad \sum_{k=1}^\infty z_k^2 < \infty. \tag{1}$$

  Each functional z is in fact in $\ell^2 = \overline{(\mathbb{R}^\infty)^*} = (\mathbb{R}^\infty)^*_P$ by assumption that $\mathcal{P}_{\boldsymbol{\xi}} \in \mathcal{G}(\mathbb{R}^\infty)$. In specific, z yields a *univariate Normal distribution* $P_z = \mathcal{N}(0, \sum_k z_k^2)$ with respect to the measure $\mathcal{P}_{\boldsymbol{\xi}}$. P_z is a probability measure on $\mathcal{B}(\mathbb{R})$.

  The adjoint operator $I : (\mathbb{R}^\infty)^* \to \mathbb{R}^\infty$ is just the natural mapping which translates each functional $z = \sum_{j \in \mathbb{N}} z_j\, \pi_k$ into sequence of $\{z_j, j \in \mathbb{N}\}$. And, the embedding mapping $I^* : (\mathbb{R}^\infty)^* \to (\mathbb{R}^\infty)^*_P \Leftrightarrow \ell^2 \to \ell^2$ is an identity map, since $(\ell^2)^*$ is closed.

- (The Gaussian measure in Hilbert space X).

  Let X be a separable Hilbert space, and $\{\boldsymbol{e}_k\}_{k=1}^{\infty}$ be a basis in X. Let $\{w_k\}$ a sequence of *independent* $\mathcal{N}(0,1)$-distributed random variables defined in probability space $(\Omega, \mathscr{F}, \mathbb{P})$. In other word, $\{w_k\}$ is a standard Gaussian sequence in $(\ell^2, \mathscr{B}_w, \mathcal{W})$. Finally let $\{\sigma_k\}$ be a sequence of nonegative numbers with

  $$\sum_{k=1}^{\infty} \sigma_k^2 \leq \infty.$$

  Define a mapping $\Xi : \Omega \to X$ by formula

  $$\Xi(\omega) = \boldsymbol{a} + \sum_{k=1}^{\infty} \sigma_k w_k(\omega) \boldsymbol{e}_k, \tag{2}$$

  where $\boldsymbol{a} \in X$ is a constant (function). Here Ξ is a random function in X with $\boldsymbol{\eta} \equiv \Xi(\omega)$ a sample path.

  Then $\mathcal{P}_{\boldsymbol{\xi}} = \mathbb{P} \circ \Xi^{-1} \in \mathcal{G}(X)$ is the Gaussian measure on X. In specific, let $\boldsymbol{a} = 0$ for $P \in \mathcal{G}_0(X)$ to be a centered Gaussian measure on X. The prescribed probability space is given as $(X, \mathscr{B}, \mathcal{P}_{\boldsymbol{\xi}})$. The Gaussian measure $\mathcal{P}_{\boldsymbol{\xi}}$ is a Radon measure on the Borel σ-algebra $\mathscr{B}$ of X and it can be shown that *any* Gaussian measure in X may be built using this construction.

  To construct the space of all $\mathcal{P}_{\boldsymbol{\xi}}$-measureable linear functionals $X_P^* \subset X^*$, we see that, due to the Riesz representation theorem, the space X^* dual to a Hilbert space consists of linear functionals of the form

  $$I(\cdot) = \sum_{k=1}^{\infty} I_k \left\langle \cdot \,, \, \boldsymbol{e}_k \right\rangle, \quad \sum_{k=1}^{\infty} I_k^2 < \infty,$$

  where $\boldsymbol{\eta} \in X$ is a sample path. Take the closure of X^* in $\mathcal{L}^2(X, \mathcal{P}_{\boldsymbol{\xi}})$ to obtain X_P^*, which consists of the functional of the form

  $$\zeta(\cdot) = \sum_{k=1}^{\infty} \zeta_k \left\langle \cdot \,, \, \boldsymbol{e}_k \right\rangle_{\mathcal{L}^2(\mathcal{P}_{\boldsymbol{\xi}})}, \quad \sum_{k=1}^{\infty} \zeta_k^2 \sigma_k^2 < \infty, \tag{3}$$

  where the inner product in X_P^* is given as

  $$\begin{aligned} \left\langle \zeta_1(\cdot) \,, \, \zeta_2(\cdot) \right\rangle_{X_P^*} &= \int \zeta_1(\boldsymbol{\eta}) \zeta_2(\boldsymbol{\eta}) P_{\boldsymbol{\xi}}(d\boldsymbol{\eta}) \\ &= \sum_{k=1}^{\infty} \sigma_k^2 \, \zeta_{k,1} \zeta_{k,2}. \end{aligned}$$

  The distribution of functional $\zeta(\cdot) = \sum_{k=1}^{\infty} \zeta_k \left\langle \cdot \,, \, \boldsymbol{e}_k \right\rangle$ w.r.t. $\mathcal{P}_{\boldsymbol{\xi}}$ is $\mathcal{N}(0, \sum_{k=1}^{\infty} \zeta_k^2 \sigma_k^2)$ on $\mathcal{B}(\mathbb{R})$.

  The topological support $\mathrm{supp}(P) = H_P = \{h \in X : h \text{ is an admissible shift of } \mathcal{P}_{\boldsymbol{\xi}}\}$ is the whole space X where $\mathcal{P}_{\boldsymbol{\xi}}$ is defined.

- (The Gaussian measure in Reproducing Kernel Hilbert space X).
  Consider the space $\mathcal{H}$ of the sample functions is a Reproducing Kernel Hilbert space on E, associated with kernel function $K : E \times E \to \mathbb{R}$. Let $\{\phi_i(\cdot), i \geq 1\}$ be the eigenfunctions of K w.r.t. measure μ, where ϕ_j is associated with eigenvalue λ_j and for $\{\lambda_i(\cdot), i \geq 1\}$

$$\lambda_i \phi_i(x) = \langle K(\cdot, x)\,,\, \phi_i \rangle = \int_{\mathcal{X}} \phi_i(z) K(z,x) d\mu(z), \forall x \in E$$

$$\sum_{i=1}^{\infty} \lambda_i < \infty; \quad \lambda_j \geq 0, j \geq 1.$$

  The collection of eigenfunctions $\{\phi_i(\cdot), i \geq 1\}$ forms an orthogonal basis for $\mathcal{H}$, such that for any sample function $f : E \to \mathbb{R}$, $f \in \mathcal{H}$,

$$f(\cdot) = \sum_{i=1}^{\infty} f_i \phi_i(\cdot)$$

  Moreover, the following property holds

$$f(x) = \langle f\,,\, K(\cdot, x) \rangle_{\mathcal{H}}\,. \quad \text{(reproducing property)}$$

$$\text{where } \langle f\,,\, g \rangle_{\mathcal{H}} = \sum_{i=1}^{\infty} \frac{f_i g_i}{\lambda_i} = \left\langle K^{-1} f\,,\, g \right\rangle,$$

$$K(x, x') = \sum_{i=1}^{\infty} \lambda_i \phi_i(x) \phi_i(x') \quad \text{(Mercer's theorem)}$$

$$f(\cdot) = \sum_m \widehat{\beta}_m K(\cdot, x_m) \quad \text{(Representor's theorem)}$$

  Now define the random function $\Xi : \Omega \to \mathcal{H}$ is given by formula

$$\Xi(\omega) = \sum_i w_i(\omega) \sqrt{\lambda_i} \phi_i(\cdot), \quad \sum_{i=1}^{\infty} w_i^2 < \infty, \tag{4}$$

  where $\{w_i, i \geq 1\}$ is a standard Gaussian sequence on $(\Omega, \mathscr{F}, \mathbb{P})$, i.e. it follows a White noise measure $\mathcal{W}$ on $\ell^2 \subset \mathbb{R}^\infty$. The probability measure of sample functions in $\mathcal{H}$ is defined as $\mathcal{P}_{\boldsymbol{\xi}} = \mathbb{P} \circ \Xi^{-1}$. The prescribed probability space is $(\mathcal{H}, \mathscr{B}_{\mathcal{H}}, \mathcal{P}_{\boldsymbol{\xi}})$.

  We can construct the space of all $\mathcal{P}_{\boldsymbol{\xi}}$-measureable linear functionals $\mathcal{H}_P^* \subset \mathcal{H}^* \simeq \mathcal{H}$, due to the Riesz representation theorem, any $I \equiv I_\eta \in \mathcal{H}_P^*$ then

$$\begin{aligned} I_\eta(\cdot) &= \langle \cdot\,,\, \eta \rangle_{\mathcal{H}} = \sum_i \eta_i \langle \cdot\,,\, \phi_i \rangle_{\mathcal{H}} = \sum_n \widehat{\alpha}_n \langle \cdot\,,\, K(\cdot, x_n) \rangle \\ &\in \mathcal{L}^2(\mathcal{H}, \mathcal{P}_{\boldsymbol{\xi}}) \\ &\Rightarrow \sum_{i=1}^{\infty} \eta_i^2 / \lambda_i < \infty \end{aligned}$$

  In other word, the distribution of I_η, $P_I = \mathcal{N}(0, \sum_{i=1}^{\infty} \eta_i^2/\lambda_i)$ on $\mathcal{B}(\mathbb{R})$.

  For the covariance operator $K : \mathcal{H}_P^* \simeq \mathcal{H} \to \mathcal{H}$, and for any $\xi, \eta \in \mathcal{H}^* \simeq \mathcal{H}$, the following equality holds,

$$\xi\left(K(\eta)\right) = \int_{\mathcal{H}} \xi(f - m) \eta(f - m) \mathcal{P}_{\boldsymbol{\xi}}(df)$$

Note that $\xi(f) \equiv f(x_\xi) = \langle f\,,\, K(\cdot, x_\xi)\rangle_{\mathcal{H}} \in \mathcal{H}^*$ and $\eta(f) \equiv f(x_\eta) = \langle f\,,\, K(\cdot, x_\eta)\rangle_{\mathcal{H}} \in \mathcal{H}^*$ are two functionals on $\mathcal{H}$. Therefore,

$$
\begin{aligned}
& cov(f(x_\xi), f(x_\eta)) \equiv \xi\left(K(\eta)\right) \\
&= \int_{\mathcal{H}} \xi(f)\eta(f)\mathcal{P}_{\boldsymbol{\xi}}(df) \\
&= \int_{\mathcal{H}} \langle f\,,\, K(\cdot, x_\xi)\rangle_{\mathcal{H}} \langle f\,,\, K(\cdot, x_\eta)\rangle_{\mathcal{H}} \mathcal{P}_{\boldsymbol{\xi}}(df) \\
&= \int_{\ell^2} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \left[\sqrt{\lambda_i} w_i \langle \phi_i\,,\, K(\cdot, x_\xi)\rangle_{\mathcal{H}}\right] \left[\sqrt{\lambda_j} w_j \langle \phi_j\,,\, K(\cdot, x_\eta)\rangle_{\mathcal{H}}\right] \mathcal{W}(d\boldsymbol{w}) \\
&= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \left(\sqrt{\lambda_i}\sqrt{\lambda_j} \int_{\ell^2} w_i w_j \mathcal{W}(d\boldsymbol{w})\right) \phi_i(x_\xi)\phi_j(x_\eta) \\
&= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \lambda_i \delta_i(j)\, \phi_i(x_\xi)\phi_j(x_\eta) \\
&= \sum_{i=1}^{\infty} \lambda_i \phi_i(x_\xi)\phi_i(x_\eta) \\
&= K(x_\xi, x_\eta)
\end{aligned}
$$

where $d\mathcal{W}(\boldsymbol{w}) = \mathcal{N}(0, I)d\boldsymbol{w}$ so that $\sum_{i=1}^{\infty} w_i^2 < \infty$. In other word, the covariance between two outputs is determined by the kernel function on their corresponding inputs.

- **Remark** For a general locally convex Hausdorff space $X = C(T)$ of sample functions, where T is (metric) separable, the probability space is $(X, \mathscr{A}, \mathcal{P})$, where $\mathcal{P}$ is induced by some linear mapping $J : X^* \to M(\Omega, \mathscr{F}, \mathbb{P})$, for the linear space of random variables on $(\Omega, \mathscr{F}, \mathbb{P})$, i.e. any linear functional in X^* is a random variable, the joint distribution of output of f is given as

$$\mathcal{P}\left\{f : (I_1(f), \ldots, I_n(f)) \in A\right\} = \mathbb{P}\left[(J(I_1), \ldots, J(I_n)) \in A\right].$$

The topological support H_P of $\mathcal{P}$ is a RKHS in X, which is associated with the continuous covariance function $K : E \times E \to \mathbb{R}$ as defined above. In fact

$$\mathcal{P}\left\{f \in H_P\right\} = 1 \text{ or } 0.$$

- Two operators

1. **Definition** The *canonical embedding operator* $I^* : F^* \to F_P^* \subset \mathcal{L}^2(F, \mathcal{P}_{\boldsymbol{\xi}}) : I^*(f) = f$, if $f \in F_P^*$. Note that $F^* \subset \mathcal{L}^2(F, \mathcal{P}_{\boldsymbol{\xi}})$, since for each realization $f \in F$ of Gaussian random function, by definition, its linear functionals $I(f)$ has finite variance for all $\eta \in F^*$; i.e.

$$\|\eta\|_2 = \left(\int_f |\eta(f)|^2\, \mathcal{P}_{\boldsymbol{\xi}}(df)\right)^{1/2} < \infty.$$

If $\mathcal{P}_{\boldsymbol{\xi}}$ is not centered, $I^*(f) = f - m\,\mathbf{1}$, here $\mathbf{1}$ is constant function and $I^*(F^*)$ is the space of all centered linear functionals. I^* is continuous under weak topology $\mathscr{T}_{\overline{F}, F^*}$

2. **Definition** The *adjoint operator* $I : F_P^* \to F$ so that for any linear functionals $\zeta \in F^*$, $\eta \in F_P^*$

$$\begin{aligned}\zeta\,(I\eta) &= \langle \eta\,,\, I^*\zeta \rangle_{\mathcal{L}^2(F)} \\ &= \int_F \eta(f)\;(I^*\zeta)(f)\mathcal{P}_{\boldsymbol{\xi}}(df)\end{aligned}$$

3. For Gaussian measure $\mathcal{P}_{\boldsymbol{\xi}} \in \mathcal{G}(F)$, the kernel of measure $\mathcal{P}_{\boldsymbol{\xi}}$ is the domain of I under $\mathcal{P}_{\boldsymbol{\xi}}$-measureable linear functionals, i.e.,

$$H_P = I(F_P^*)$$

.

The domain of adjoint operator $I : F_P^* \to I(F_P^*) = H_P \subset F$ is an isomorphism.

4. The covariance operator

$$K = II^* : F^* \to F$$

That is,

$$K(F^*) \subseteq H_P \subseteq F$$

- Some other operators: given any Hilbert space L,

1. define $J : L \to F$ as a linear operator and the kernel $H_P = J(L)$.
2. define its adjoint $J^* : F^* \to L$
3. the kernel can be decomposed as $K = J\,J^*$
4. the variance of linear functional η, is given as

$$\|\eta\|_{\mathcal{L}^2(F,\mathcal{P}_\xi)} = \|\xi\|_L$$

where $\xi \in J^{-1}(h_\eta)$, $h_\eta = I\eta$.

5. In fact, if $L = \mathcal{L}^2(\Omega, \mathscr{F}, \mathbb{P})$ is the space of random variables with finite variance on $(\Omega, \mathscr{F}, \mathbb{P})$, then we can generate the sample function via equation

$$f(\omega) = J(\xi(\omega)),\;\; \omega \in \Omega$$

for random variables $\xi \in L$.

6. Also, define $L = \mathcal{L}^2(S, \mathscr{M}, \nu)$ to be Hilbert space of functions on domain S, with a set of basis function $\{m_t, t \in E\}$. L is considered as a generalized spectral representation with

$$\begin{aligned}J^*\left(\int_E (\cdot)\mu_\eta(dt)\right)(r) &= \int_E m_t(r)\mu_\eta(dt) \equiv m_\eta(r) \\ f_\eta(s) \equiv (J\,m_\eta)(s) = \langle m_\eta\,,\, m_s\rangle_L &= \int_E \left(\int_S m_t(u)m_s(u)\nu(du)\right)\mu_\eta(dt) \\ &= \int_E K(t,s)\mu_\eta(dt) \\ K(t,s) &= \int_S m_t(u)m_s(u)\nu(du)\end{aligned}$$

References

Gerald B Folland. *Real analysis: modern techniques and their applications.* John Wiley & Sons, 2013.

Mikhail Antolevich Lifshits. *Gaussian random functions*, volume 322. Springer Science & Business Media, 2013.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning).* The MIT Press, 2005. ISBN 026218253X.

Michael Reed and Barry Simon. *Methods of modern mathematical physics: Functional analysis*, volume 1. Gulf Professional Publishing, 1980.

Helmut H Schaefer and Manfred P Wolff. *Locally Convex Topological Vector Spaces.* Springer, 1999.

Angus Ellis Taylor and David C Lay. *Introduction to functional analysis*, volume 2. Wiley New York, 1958.