MACS 201 : Hilbert spaces and probability

Hilbert spaces

Def. Let $\mathcal{H}$ be a complex linear space. An **inner-product** on $\mathcal{H}$ is a function $\langle \cdot \mid \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbf{C}$ which satisfies the following properties :

(i) $\forall (x, y) \in \mathcal{H} \times \mathcal{H}, \langle x \mid y \rangle = \overline{\langle y \mid x \rangle}$,

(ii) $\forall x, y, z \in \mathcal{H} \forall (\alpha, \beta) \in \mathbf{C} \times \mathbf{C}, \langle \alpha x + \beta y \mid z \rangle = \alpha \langle x \mid z \rangle + \beta \langle y \mid z \rangle$,

(iii) $\forall x \in \mathcal{H}, (\langle x \mid x \rangle = 0) \iff (x = 0)$

Then $\|\cdot\| : x \mapsto \sqrt{\langle x \mid x \rangle} \geqslant 0$ defines a norm on $\mathcal{H}$. Both are continuous.

Th. *For all $x, y \in \mathcal{H}$, we have :*

a) Cauchy-Schwarz inequality : $|\langle x \mid y \rangle| \leqslant \|x\| \cdot \|y\|$,

b) triangular inequality : $|\|x\| - \|y\|| \leqslant \|x - y\| \leqslant \|x\| + \|y\|$,

c) Parallelogram inequality : $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$.

Def. An inner-product space $\mathcal{H}$ is called an Hilbert space if it is complete.

Prop. *For all measured space $(\Omega, \mathcal{F}, \mu)$, the space $L^2(\Omega, \mathcal{F}, \mu)$ endowed with $\langle f \mid g \rangle = \int f \bar{g} \, \mathrm{d}\mu$ is a Hilbert space.*

Def. Two vectors $x, y \in \mathcal{H}$ are orthogonal if $\langle x \mid y \rangle = 0$ which we denoted by $x \perp y$. If $\mathcal{S}$ is a subspace of $\mathcal{H}$, we write $x \perp \mathcal{S}$ if $\forall s \in \mathcal{S}, x \perp s$. Also we write $\mathcal{S} \perp \mathcal{T}$ if all vectors in $\mathcal{S}$ are orthogonal to $\mathcal{T}$.

Not. If $\mathcal{H} = \mathcal{A} + \mathcal{B}$ and $\mathcal{A} \perp \mathcal{B}$ we will denote $\mathcal{H} = \mathcal{A} \overset{\perp}{\oplus} \mathcal{B}$.

Def. Let $\mathcal{E}$ be a subset of an Hilbert space $\mathcal{H}$. The orthogonal set of $\mathcal{E}$ is $\mathcal{E}^\perp = \{x \in \mathcal{H} \mid \forall y \in \mathcal{E}, \langle x \mid y \rangle = 0\}$.

Th. *If $\mathcal{E}$ is a subset of an Hilbert space $\mathcal{H}$, then $\mathcal{E}^\perp$ is closed.*

Orthogonal and orthonormal bases

Def. Let E be a subset of $\mathcal{H}$. It is an orthogonal set if for all $(x, y) \in E \times E, x \neq y, x \perp y$. If moreover $\forall x \in E, \|x\| = 1$, we say that E is orthonormal.

Th. *Let $(e_i)_{i \geqslant 1}$ be an orthonormal sequence of an Hilbert space $\mathcal{H}$ and let $(\alpha_i)_{i \geqslant 1} \in \mathbf{C}^\mathbf{N}$. The series $\sum_{i=1}^\infty \alpha_i e_i$ converges in $\mathcal{H}$ if and only if $\sum_i |\alpha_i|^2 < \infty$, in which case $\|\sum_{i=1}^\infty \alpha_i e_i\|^2 = \sum_{i=1}^\infty |\alpha_i|^2$.*

Prop. *Let $x \in \mathcal{H}$ (Hilbert space) and $E = \{e_1, \ldots, e_n\}$ a finite orthonormal set of vectors. Then $\|x - \sum_{k=1}^n \langle x \mid e_k \rangle e_k\|^2 = \|x\|^2 - \sum_{k=1}^n |\langle x \mid e_k \rangle|^2 = \inf\{\|x - y\|^2, y \in \mathrm{Span}(e_1, \ldots, e_n)\}$.*

Cor (Bessel inequality). *Let $(e_i)_{i \geqslant 1}$ be an orthonormal sequence of a Hilbert space $\mathcal{H}$. Then$\forall, x \in \mathcal{H}, \sum_{i=1}^\infty |\langle x \mid e_i \rangle|^2 \leqslant \|x\|^2$.*

Def. A subset E of a Hilbert space $\mathcal{H}$ is said dense id $\overline{\mathrm{Span}}(E) = \mathcal{H}$. An orthonormal dense sequence is called a Hilbert basis.

Prop. *Consider the measured space $(\Omega, \mathcal{F}, \mu)$ and the Hilbert space $\mathcal{H} = L^2(\Omega, \mathcal{F}, \mu)$, $\overline{\mathrm{Span}}(\mathbf{1}_A, A \in \mathcal{F}) = \mathcal{H}$.*

Th. *Let $(e_i)_{i \geqslant 1}$ be a Hilbert basis of the Hilbert space $\mathcal{H}$. Then $\forall x \in \mathcal{H}, x = \sum_{i=1}^\infty \langle x \mid e_i \rangle e_i$.*

Th. *Let $(e_i)_{i \geqslant 1}$ be an orthonormal sequence of the Hilbert space $\mathcal{H}$. The following assertions are equivalent :*

(i) $(e_i)_{i \geqslant 1}$ is a Hilbert basis,

(ii) if some $x \in \mathcal{H}$ satisfies $\forall i \geqslant 1, \langle x \mid e_i \rangle = 0$ then $x = 0$,

(iii) $\forall x \in \mathcal{H}, \|x\|^2 = \sum_{i=1}^\infty |\langle x \mid e_i \rangle|^2$.

Th. *A Hilbert space $\mathcal{H}$ if separable (i.e. contains a countable dense subset) if and only if it admits a Hilbert basis.*

Fourier series

Let $\psi_n : x \mapsto \frac{1}{\sqrt{2\pi}} e^{inx}, n \in \mathbf{Z}$. Let $L^1(\mathbf{T})$ denote the set of 2π-periodic locally integrable functions. For $f \in L^1(\mathbf{T})$, set $\forall n \in \mathbf{N}, f_n = \sum_{k=-n}^n \left(\int_\mathbf{T} f \bar{\phi}_k\right) \phi_k$.

Th. *Let f be a continuous 2π-periodic function. Then the Cesaro sequence $\frac{1}{n} \sum_{k=0}^{n-1} f_k$ converges uniformly to f.*

Cor. *Let μ be a finite measure on the Borel sets of $\mathbf{T} = \mathbf{R}/(2\pi\mathbf{Z})$. The sequence $(\phi_n)_{n \in \mathbf{Z}}$ is dense in the Hilbert space $L^2(\mathbf{T}, \mathcal{B}(\mathbf{T}), \mu)$.*

Cor. *The sequence $(\phi_n)_{n \in \mathbf{Z}}$ is a Hilbert basis in $L^2(\mathbf{T})$. In particular, $\forall f \in L^2(\mathbf{T}), f = \sum_{k=-\infty}^\infty \alpha_k \phi_k$ with $\alpha_k = \frac{1}{\sqrt{2\pi}} \int_\mathbf{T} f(x) e^{-ikx} \, \mathrm{d}x$ when the infinite sum converges in $L^2(\mathbf{T})$. The Parseval identity then reads $\int_\mathbf{T} |f(x)|^2 \, \mathrm{d}x = \sum_{k=-\infty}^\infty |\alpha_k|^2$.*

Projection and orthogonality principle

Th (Projection theorem). *Let $\mathcal{E}$ be a closed convex subset of a Hilbert space $\mathcal{H}$ and $x \in \mathcal{H}$. Then the following holds :*

(i) *There exists a unique vector* $\operatorname{proj}(x \mid \mathcal{E}) \in \mathcal{E}$ *such that* $\|x - \operatorname{proj}(x \mid \mathcal{E})\| = \inf_{w\in\mathcal{E}} \|x - w\|$.

(ii) *If moreover $\mathcal{E}$ is a linear subspace,* $\operatorname{proj}(x \mid \mathcal{E})$ *is the unique $\hat{x} \in \mathcal{E}$ such that $x - \hat{x} \in \mathcal{E}^{\perp}$. It is called the orthogonal projection of x onto $\mathcal{E}$.*

Prop. *Let $\mathcal{H}$ be a Hilbert space and $\mathcal{E}, \mathcal{E}_1, \mathcal{E}_2$ closed subspaces of $\mathcal{H}$. Then the following assertions hold.*

(i) *Suppose that* $\mathcal{E} = \overline{\operatorname{Span}}\left((e_k)_{k\in\mathbf{N}}\right)$ *with (e_k) being an orthonormal sequence. Then* $\operatorname{proj}(h \mid \mathcal{E}) = \sum_{k=0}^{\infty} \langle h \mid e_k \rangle e_k$.

(ii) *The function* $\operatorname{proj}(\cdot \mid \mathcal{H}) \colon x \mapsto \operatorname{proj}(x \mid \mathcal{E})$ *is linear and continuous on $\mathcal{H}$.*

(iii) $\|x\|^2 = \|\operatorname{proj}(x \mid \mathcal{E})\|^2 + \|x - \operatorname{proj}(x \mid \mathcal{E})\|^2$

(iv) $(x \in \mathcal{E} \iff \operatorname{proj}(x \mid \mathcal{E}) = x)$ *and* $\left(x \in \mathcal{E}^{\perp} \iff \operatorname{proj}(x \mid \mathcal{E}) = 0\right)$

(v) *If* $\mathcal{E}_1 \subset \mathcal{E}_2$ *then* $\forall x \in \mathcal{H}, \operatorname{proj}(\operatorname{proj}(x \mid \mathcal{E}_{\in} \mid \mathcal{E}_1) = \operatorname{proj}(x \mid \mathcal{E}_1)$

(vi) *If* $\mathcal{E}_1 \perp \mathcal{E}_2$ *then* $\forall x \in \mathcal{H}, \operatorname{proj}\left(x \mid \mathcal{E}_1 \overset{\perp}{\oplus} \mathcal{E}_2\right) = \operatorname{proj}(x \mid \mathcal{E}_1) + \operatorname{proj}(x \mid \mathcal{E}_2)$

Th. *Let $(M_n)_{n\in\mathbf{Z}}$ be an increasing sequence of closed subspaces of an Hilbert space $\mathcal{H}$.*

1. *Denote* $M_{-\infty} = \bigcap_n M_n$. *Then* $\forall h \in \mathcal{H}, \operatorname{proj}(h \mid M_{-\infty}) = \lim_{n\to-\infty} \operatorname{proj}(h \mid M_n)$.
2. *Denote* $M_{\infty} = \overline{\bigcup_n M_n}$. *Then* $\forall h \in \mathcal{H}, \operatorname{proj}(h \mid M_{\infty}) = \lim_{n\to\infty} \operatorname{proj}(h \mid M_n)$.

Prop. *Let $\mathcal{E}$ and $\mathcal{F}$ be two subspaces of a Hilbert space $\mathcal{H}$. If* $\mathcal{E} \overset{\perp}{\oplus} \mathcal{F} = \mathcal{H}$, *then* $\mathcal{F} = \mathcal{E}^{\perp}$.

Th. *If $\mathcal{E}$ is a closed subspace of a Hilbert space $\mathcal{H}$ then* $\mathcal{E} \overset{\perp}{\oplus} \mathcal{E}^{\perp} = \mathcal{H}$. *Moreover* $\left(E^{\perp}\right)^{\perp} = \mathcal{E}$.

Th (Riesz representation theorem). *Let $\mathcal{H}$ be a Hilbert space. Then $F \colon \mathcal{H} \to \mathbf{C}$ is a non-zero continuous linear form if and only if* $\exists x \in \mathcal{H} \setminus \{0\}, \forall y \in \mathcal{H}, F(y) = \langle y \mid x \rangle$.

Unitary Operator

Def. Let $\mathcal{H}$ and $\mathcal{I}$ be two Hilbert spaces. An **isometric** operator $S \colon \mathcal{H} \to \mathcal{I}$ is a linear application such that $\forall (v, w) \in \mathcal{H}^2, \langle Sv \mid Sw \rangle_{\mathcal{I}} = \langle v \mid w \rangle_{\mathcal{H}}$. If it is moreover bijective, it is a **unitary** operator. In this case we also says that $\mathcal{H}$ and $\mathcal{I}$ are isomorphic.

Th. *Let $\mathcal{H}$ be a separable Hilbert space.*

(i) *If $\mathcal{H}$ has infinite dimension, it is isomorphic to l^2.*

(ii) *If $\mathcal{H}$ has dimension n, it is isomorphic to $\mathbf{C}^n$.*

Th. *Let $\mathcal{H}$ and $\mathcal{I}$ be two Hilbert spaces and $\mathcal{G}$ a subspace of $\mathcal{H}$.*

(i) *Let $S \colon \mathcal{G} \to \mathcal{I}$ be isometric on $\mathcal{G}$. Then S admits a unique isometric extension $\bar{S} \colon \bar{\mathcal{G}} \to \mathcal{I}$ and $\bar{S}\left(\bar{\mathcal{G}}\right)$ is the closure of $S(\mathcal{G})$ in $\mathcal{I}$.*

(ii) *Let $(v_t)_{t\in T}$ and $(w_t)_{t\in T}$ be two set of vectors in $\mathcal{H}$ and $\mathcal{I}$ indexed by an arbitrary index set T. Suppose* $\forall (s,t) \in T^2, \langle v_t \mid v_s \rangle_{\mathcal{H}} = \langle w_t \mid w_s \rangle_{\mathcal{I}}$. *Then, there exists a unique isometric operator* $S \colon \overline{\operatorname{Span}}\left((v_t)_{t\in T}\right) \to \overline{\operatorname{Span}}\left((w_t)_{t\in T}\right)$ *such that* $\forall t \in T, Sv_t = w_t$. *Moreover,* $S\left(\overline{\operatorname{Span}}\left((v_t)_{t\in T}\right)\right) = \overline{\operatorname{Span}}\left((w_t)_{t\in T}\right)$.

Probability

Th (π - λ theorem). *If $\mathcal{A} \subset \mathcal{C}$ with $\mathcal{A}$ a π-system and $\mathcal{C}$ a λ-system, then $\sigma(\mathcal{A}) = \mathcal{C}$.*

Th (Characterization of probability measures). *Let $\mathcal{C}$ be a π-system on Ω and $\mathcal{F} = \sigma(\mathcal{C})$ the smallest σ-field containing $\mathcal{C}$. Then a probability measure μ on $(\Omega, \mathcal{F})$ is uniquely characterized by $\mu(A)$ on $A \in \mathcal{C}$.*

Not. For $p > 0$, we denote by $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbf{P})$ the space of random variables X such that $\mathbf{E}\left(|X|^p\right) < \infty$ and by $L^p(\Omega, \mathcal{F}, \mathbf{P})$ the one identifying random variables that are equal $\mathbf{P}$-a.s.

Conditional calculus

Lem. *Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$ and $\mathcal{G}$ a sub-σ-field of $\mathcal{F}$. Then there exists $Y \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathbf{P})$ such that*

$$\forall A \in \mathcal{G}, \mathbf{E}(X \mathbf{1}_A) = \mathbf{E}(Y \mathbf{1}_A) \tag{1}$$

Moreover the following assertions hold.

(i) *If $Y' \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathbf{P})$ also satisfies (1) then $Y' = Y$ $\mathbf{P}$-a.s.*

(ii) *If $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$, then* $Y = \operatorname{proj}(X \mid L^2(\Omega, \mathcal{G}, \mathbf{P}))$.

(iii) *(1) continues to hold extended as $\mathbf{E}(XZ) = \mathbf{E}(YZ)$ for all $\mathcal{G}$-measurable r.v. Z such that $\mathbf{E}(|XZ|) < \infty$.*

Def. Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$ and $\mathcal{G}$ a sub-σ-field of $\mathcal{F}$. The unique $Y \in L^1(\Omega, \mathcal{G}, \mathbf{P})$ defined by (1) is called the **conditional expectation** of X given $\mathcal{G}$, and denoted by $Y = \mathbf{E}(x \mid \mathcal{G})$.

Prop. *Suppose that $X, Y, Z, (X_n)_{n\geqslant 1} \in L^1(\Omega, \mathcal{F}, \mathbf{P})$. The following hold $\mathbf{P}$-a.s.*

(i) *(linearity) $\forall a, b \in \mathbf{R}, \mathbf{E}(aX + bY \mid \mathcal{G}) = a\mathbf{E}(X \mid \mathcal{G}) + b\mathbf{E}(Y \mid \mathcal{G})$*

(ii) *If X is $\mathcal{G}$-measurable, $\mathbf{E}(X \mid \mathcal{G}) = X$*

(iii) *If $\mathcal{G} = \{\varnothing, \Omega\}$ is the trivial σ-field, then $\mathbf{E}(X \mid \mathcal{G}) = \mathbf{E}(X)$*

(iv) *If X is independent of $\mathcal{G}$ then $\mathbf{E}(X \mid \mathcal{G}) = \mathbf{E}(X)$*

(v) *(positivity) If $X \leqslant Y$ then $\mathbf{E}(X \mid \mathcal{G}) \leqslant \mathbf{E}(Y \mid \mathcal{G})$*

(vi) *$\mathbf{E}(X \mid \mathcal{G}) \vee \mathbf{E}(Y \mid \mathcal{G}) \leqslant \mathbf{E}(X \vee Y \mid \mathcal{G})$, $\mathbf{E}(X \mid \mathcal{G})_+ \leqslant \mathbf{E}(X_+ \mid \mathcal{G})$ and $|\mathbf{E}(X \mid \mathcal{G})| \leqslant \mathbf{E}(|X| \mid \mathcal{G})$*

(vii) *(tower property) If $\mathcal{H}$ is a sub-σ-field of $\mathcal{F}$ such that $\mathcal{G} \subset \mathcal{H}$ then $\mathbf{E}(\mathbf{E}(X \mid \mathcal{H}) \mid \mathcal{G}) = \mathbf{E}(X \mid \mathcal{G})$*

(viii) *The expectation is not modified by conditional expectation : $\mathbf{E}(\mathbf{E}(X \mid \mathcal{G})) = \mathbf{E}(X)$*

(ix) *If X is $\mathcal{G}$-measurable and $XY \in L^1(\Omega, \mathcal{F}, \mathbf{P})$, then $\mathbf{E}(XY \mid \mathcal{G}) = X \cdot \mathbf{E}(Y \mid \mathcal{G})$*

Def. Let Y be a r.v. and $\sigma(X)$ the sub-σ-field generated by a r.v. X. If $\mathbf{E}(Y \mid \sigma(X))$ is well-defined, it is written as $\mathbf{E}(Y \mid X)$ and is called the **conditional expectation** of Y given X.

Mathematical statistics

Statistical modeling

Def. Let $(\Omega, \mathcal{F})$ be a measurable space and $\mathcal{P}$ a collection of probabilities on this space. Let X be a measurable function from $(\Omega, \mathcal{F})$ to the observation space $(\mathsf{X}, \mathcal{X})$. We say that $\mathcal{P}$ is a **statistical model** for the observation variable X and denote $\mathcal{P}^X = \left(P^X\right)_{P\in\mathcal{P}}$ the corresponding collection of probability distributions.

It is usual in statistics to consider $\Omega = \mathsf{X}$, $\mathcal{F} = \mathcal{X}$ and $X(\omega) = \omega$, in which case $\forall P \in \mathcal{P}, P = P^X$.

Def. Let $\nu \in \mathbf{M}_+(\mathsf{X}, \mathcal{X})$ and $\mathcal{P}$ be a statistical model for X. We say that $\mathcal{P}$ is a ν-dominated model for X, or that $\mathcal{P}^X$ is ν-dominated, if $\forall P \in \mathcal{P}, P^X \ll \nu$.

Lem. *Let $\nu \in \mathbf{M}_+(\mathsf{X}, \mathcal{X})$. Consider a ν-dominated model $\mathcal{P}$ for the variable X. Then there exists a countable collection $(P_n)_{n\geqslant 1}$ in $\mathcal{P}$ such that $\mathcal{P}^X$ is also dominated by $\mu = \sum_{n\geqslant 1} 2^{-n} P_n^X$.*

Def. Let $\mathcal{P}$ be a statistical model for the observation variable X. We say that $\mathcal{P}$ is a **parametric model** for X if there exists a finite dimensional set Θ such that $\mathcal{P} = (P_\theta)_{\theta\in\Theta}$.

Def. Let $\mathcal{P}$ be a statistical model for X. Any finite dimensional quantity $t(P^X)$ only depending on P^X as $P \in \mathcal{P}$ is called an **identifiablee parameter**.

Def. Let $\mathcal{P}$ be a statistical model for X. A **statistic** in this context is any random variable T valued in $\left(\mathbf{R}^d, \mathcal{B}\left(\mathbf{R}^d\right)\right)$ with $d \geqslant 1$, defined by $T = g(X)$ where g is a Borel function not depending on $P \in \mathcal{P}$.

If a statistic is used as a guess for a parameter $t(P) \in \mathbf{R}^d$, it is called an **estimator** of $t(P)$. In this case, the **bias** of T for estimating $t(P)$ is defined as $\mathrm{Bias}(T, P) = \int T \,\mathrm{d}P - t(P)$ whenever $\int |T| \,\mathrm{d}P < \infty$. We say that T is an *unbiased* estimator of $t(P)$ if $\forall P \in \mathcal{P}, \int T \,\mathrm{d}P = t(P)$. The **quadratic risk** or **mean squared error** (in the case $d = 1$) is defined by $\mathrm{MSE}(T, P) = \int (T - t(P))^2 \,\mathrm{d}P = \mathrm{Var}(T) + \mathrm{Bias}(T, P)^2$.

Def. Let T be a statistic valued in $(\mathbf{R}^d, \mathbf{R}^{\mathcal{D}})$ with $d \geqslant 1$. We say that T is a **sufficient statitstic** for the model $\mathcal{P}$ if, for all $P \in \mathcal{P}$, the conditional distribution of X given T does not depend on P, that is, there exists a probability kernel $Q \subset \mathbf{R}^d \times \mathcal{X}$ such that, for all $P \in \mathcal{P}$, Q is a regular version of $P^{X|T}$.

Lem. *Let S be a sufficient statistic associated to the Markov kernel Q and let $T = g(X)$ be an unbiased estimator of the parameter $t(P)$ (both real valued). Define $T^R = \int g(x) Q(S, \mathrm{d}x)$. Then T^R is an unbiased estimator of the parameter t and its variance is smaller than that of T. As a consequence we have, $\forall P \in \mathcal{P}, \mathrm{MSE}\left(T^R, P\right) \leqslant \mathrm{MSE}(T, P)$.*

Th (Fisher Factorization theorem). *Let $\nu \in \mathbf{M}_+(\mathsf{X}, \mathcal{X})$. Consider a ν-dominated model $\mathcal{P}$ for X and let $S = g(X)$ be a d-dimensional statistic. Then S is a sufficient statistic for the model $\mathcal{P}$ if and only if there exists a non-negative Borel function h on X such that $\forall P \in \mathcal{P}$, there exists a Borel function $f_P \colon \mathbf{R}^d \to \mathbf{R}_+$ such that $\frac{\mathrm{d}P^X}{\mathrm{d}\nu} = h \cdot f_P \circ g$.*