MACS 201 : Hilbert spaces and probability

1 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 defined as $\mathcal{E} = \{x \in \mathcal{H} \mid \forall y \in \mathcal{E}, \langle x \mid y \rangle = 0\}$.

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. *Suppose that f is 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 $\mathrm{proj}(x \mid \mathcal{E}) \in \mathcal{E}$ such that $\|x - \mathrm{proj}(x \mid \mathcal{E})\| = \inf_{w\in\mathcal{E}} \|x - w\|$.

(ii) If moreover $\mathcal{E}$ is a linear subspace, $\mathrm{proj}(x \mid \mathcal{E})$ 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 $\mathcal{E}$.

2 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. *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 $A \in \mathcal{C}$.*

Def. Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$ and $\mathcal{G}$ a sub-σ-field of $\mathcal{F}$.

3 Mathematical statistics

3.1 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 $\frac{\mathrm{d}P^X}{\mathrm{d}\nu} = h \cdot f_P \circ g$.*