Basic structure and methods of proof for random probability measures

Kalle Koskinen

University of Helsinki

2022

Probability theory

- Topological spaces offer a suitable framework for many natural constructions in probability theory
- If X is a topological space then the Borel σ-algebra $\mathcal{B}(X)$ is the smallest σ-algebra containing the open sets of X
- If X and Y are topological spaces, then any continuous function $f : X \to Y$ is measurable

A topology can thus generate a measurable structure and provide a large class of measurable functions

Probability theory cont.

- If μ is a measure on a measurable space X, then the mapping $T : L_1(\mu) \to \mathbb{R}$ defined by

$$T[f] := \int_X \mu(dx)\ f(x) := \mu[f]$$

  is a positive linear functional known as the **evaluation map**
- One can "recover" measures of sets via the indicator function

$$\mu(A) = \int_X \mu(dx)\ \mathbb{1}(x \in A) = \mu[\mathbb{1}(\cdot \in A)]$$

Probability theory cont.

- Supopose that the topology on X is metrizable by a metric d
- For any two probability measures on X, it follows that $\mu = \nu$ if and only if $\mu[f] = \nu[f]$ for all continuous bounded functions $f \in C_b(X)$

The metric structure allows one to shift their view from probabilities of events to expectations of continuous bounded functions

A Bourbakian approach to probability theory

Theorem (Riesz-Markov-Kakutani)

Let X be a locally compact Hausdorff space. For any positive linear functional ψ on $C_c(X)$, there is a unique Radon measure μ on X such that

$$\psi[f] = \int_X \mu(dx)\, f(x) := \mu[f]$$

for all $f \in C_c(X)$.

- If $\mu[1] = 1$, then μ is a probability measure
- A radon probability measure is both inner- and outer regular
- Outer regular: $\mu(E) = \inf\{\mu(U) : E \subset U,\ U \text{ is open}\}$
- Inner regular: $\mu(E) = \sup\{\mu(K) : K \subset E,\ K \text{ is compact}\}$

Examples

- Let $\rho \in L^1(\mu)$ such that $\rho \geq 0$ almost everywhere and $||\rho||_1 = 1$ then

$$\psi[f] := \int_X \mu(dx)\ \rho(x) f(x)$$

- Let $X = [0,1]$, define the integral as the Riemann integral then

$$\psi[f] := \int_0^1 dx\ f(x)$$

  defines a measure on $[0,1]$ with $\mathcal{B}([0,1])$ this is the pre-completion Lebesgue measure

Uniqueness of Radon probability measures

- Suppose that X is a separable locally compact metric space
- For Radon probability measures μ and ν it follows that $\mu = \nu$ if and only if $\mu[f] = \nu[f]$ for all $f \in C_c(X)$

Thus one can begin from a theory of positive linear functionals on a separable locally compact metric space and generate measurable structure, highly regular probability measures, and criteria for the uniqueness of said measures

Abstract approximation results

- A subset of functions $M \subset C_b(X)$ is called separating if $\mu[f] = \nu[f]$ for any $f \in M$ implies that $\mu = \nu$
- By the previous remarks, if X is a metric space then $C_b(X)$ is separating, and if X is a separable locally compact metric space then $C_c(X)$ is separating

Abstract approximation results cont.

- A set of functions $M \subset C_b(X)$ is said to separate points if for every pair of distinct points $x, y \in X$ there exists $f \in M$ such that $f(x) \neq f(y)$
- The same set of functions is called nowhere vanishing if for every $x \in X$ there exists a function $f \in M$ such that $f(x) \neq 0$

Theorem (Stone-Weierstrass)

Let X be a compact Hausdorff space and let $\mathcal{A} \subset C(X)$ be a nowhere vanishing subalgebra. The subalgebra $\mathcal{A}$ is dense in $C(X)$ if and only if it separates points.

Abstract approximation results cont.

- An application of the Stone-Weierstrass theorem yields the following theorem

Theorem

Let X be a complete separable metric space. Any nowhere vanishing subalgebra $\mathcal{A} \subset C_b(X)$ which separates points is separating.

It follows that in order to characterize probability measures on complete separable metric spaces, it is enough to study their expectations of functions contained in nowhere vanishing subalgebras which separate points

Summary

- Let us now remark that a separable topological space which is metrizable by a complete metric is called a Polish space
- A topological space is second countable if it has a countable base
- Every locally compact Hausdorff space which is second countable is Polish

Normalizable positive linear functionals on locally compact second-countalbe Hausdorff spaces define probability measures on a Polish space... and every such probability measure is uniquely defined by its expectations on nowhere vanishing subalgebras of continuous bounded functions which separate points

How I learned to love useless algebraic structures like magmas

- Suppose that $\mathcal{B} \subset C_b(X)$ is collection of functions which is nowhere vanishing, separates points, and is closed under multiplication(a magma)
- Let $\mathcal{A}$ be the subalgebra generated by $\mathcal{B}$
- It can be shown that $\mathcal{A}$ is a nowhere vanishing subalgebra that seperates points consisting of all finite linear combinations of elements in $\mathcal{B}$
- By linearity of the expectation, it can be shown that $\mu[f] = \nu[f]$ for $f \in \mathcal{A}$ if and only if $\mu[g] = \nu[f]$ for $g \in \mathcal{B}$

By linearity, expectations of nowhere vanishing magmas which separates points are separating

How I learned to love useless algebraic structures like magmas cont.

- A large number of elementary results in probability theory are examples of this theorem, but they are almost always proven in a different way... here are some examples

Examples

- The factorizable continuous bounded functions on $\mathbb{R}^n$ defined by $\{\prod_{i=1}^n f_i(x_i) : f_i \in C_b(\mathbb{R})\}$
- Equivalent to stating that the distribution of the random vector (X, Y) is determined by expectations $\mathbb{E}f(X)g(Y)$ for $f, g \in C_b(\mathbb{R})$
- The monomials on the interval $[0, 1]$ defined by $\{x^k : \ k \in \mathbb{N} \cup \{0\}\}$
- Equivalent to stating that for random variables on the unit interval $X = Y$ if $\mathbb{E}X^k = \mathbb{E}Y^k$ for all $k \in \mathbb{N} \cup \{0\}$
- The trigonometric polynomials on $\mathbb{R}$ defined by $\{\sum_{i=1}^k (a_i sin(s_i x) + b_i \cos(c_i x) : a_i, b_i, s_i, c_i \in \mathbb{R}\}$
- Equivalent to stating that for random variables X, Y we have $X = Y$ if $\mathbb{E}\sin(tX) = \mathbb{E}\sin(tY)$ and $\mathbb{E}\cos(tX) = \mathbb{E}\cos(tY)$ for all $t \in \mathbb{R}$
- This is equivalent to stating that $X = Y$ if $\mathbb{E}e^{itX} = \mathbb{E}e^{itY}$ for all $t \in \mathbb{R}$

Weak convergence and the space of probability measures

- We say that a sequence μ_n of probability measures converges weakly to another probability measure μ if $\mu_n[f] \to \mu[f]$ for all $f \in C_b(X)$
- If X is a Polish space, then there exists a (Levy) metric d on the space of probability measures $\mathcal{M}_1(X)$ which makes $\mathcal{M}_1(X)$ into a Polish space itself and we have $\mu_n \to \mu$ weakly if and only if $d(\mu_n, \mu) \to 0$

Prokhorov's theorem

- One of the most remarkable abstract results in probability theory is Prokhorov's theorem and one of its corollaries
- A probability measure μ is said to be tight if for any $\varepsilon > 0$ there exists a compact set K such that $\mu(K) \geq 1 - \varepsilon$
- A family of probability measures μ_i is said to be uniformly tight if for any $\varepsilon > 0$ and all $i \in I$ there exists a compact set K such that $\mu_i(K) \geq 1 - \varepsilon$

Theorem (Prokhorov)

Let X be a Polish space and let $\mathcal{F} := \{\mu_i\}_{i \in I}$ be a family of probability measures on X. It follows that $\overline{\mathcal{F}} \subset \mathcal{M}_1(X)$ is compact if and only if $\mathcal{F}$ is uniformly tight.

Interlude on metric compactness

- In a metric space, compactness is equivalent to sequential compactness
- A set K is compact if and only if every subsequence has a convergent subsubsequence with a limit in K
- If every subsequence of a sequence in a compact set converges to the same limit, then it follows that the sequence converges to the same limit

Example

- Suppose that $f_n : K \to \mathbb{R}$ converges uniformly on f and that f has a unique maximizing point x^* such that $f(x^*) = \max_{x \in K} f(x)$
- Let x_n^* be a maximizing point for f_n i.e. $f_n(x_n^*) = \max_{x \in K} f_n(x)$
- Since x_n^* belong to a compact set there exists a convergent subsequence $x_{n_k}^*$ with a limit y^*
- By definition of a maximizing point, we have $f_n(x_{n_k}^*) \geq f_{n_k}(x^*)$
- Taking limits we have $f(y^*) \geq f(x^*)$, but this implies that $y^* = x^*$ by uniqueness of x^*, thus $y^* = x^*$ and every subsequence of x_n^* converges to the same limit x^* which implies that the sequence of maximizing points converges to the unique maximizing point of f

Prokhorov cont.

- Using uniform tightness and Prokhorov's theorem, we have the following highly applicable theorem

Theorem

Let X be a Polish space, $\mathcal{A} \subset C_b(X)$ a nowhere vannishing subalgebra that seperates points, and μ_n a uniformly tight sequence of probability measures on X. If there exists a probability measure μ such that for any $f \in \mathcal{A}$ and every convergent subsequence μ_{n_k} we have $\mu_{n_k}[f] \to \mu[f]$ then it follows that $\mu_n \to \mu$ weakly.

When we have a uniformly tight sequence of probability measures, it is enough to check the convergence of expectations of functions in a nowhere vanishing subalgebra which separates points

Random measures

- Recall : when X is a Polish space $\mathcal{M}_1(X)$ is a Polish space
- A measurable function $\mu : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{M}_1(X), \mathcal{B}(\mathcal{M}_1(X)))$ is called a random measure
- The distribution of a random measure is uniquely defined by $\mathbb{P}(\mu \in A)$ where A is an open set in the space of probability measures
- Depending on your predilection to topology this may or may not be a suitable starting point

A prototypical example

- Let $m \in \mathbb{R}^k$ be a vector and let $\mu(m)$ be the Gaussian measure on $\mathbb{R}^k$ with mean m and variance 1
- To be exact, we have

$$\mu(m)[f] = \frac{1}{\sqrt{2\pi}^k} \int_{\mathbb{R}^k} dx \ e^{-\frac{||x-m||^2}{2}} f(x)$$

- Suppose now that $m : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ is a measurable function i.e. a random variable
- The map $m \mapsto \mu(m)$ is sequentially continuous by dominated convergence and thus it is measurable, so the map $\omega \mapsto m(\omega) \mapsto \mu(m(\omega))$ is a random measure

A nowhere vanishing subalgebra that separates points

- Fix $f \in C_b(X)$, the mapping $\mu \mapsto \mu[f]$ is trivially sequentially continuous and thus continuous
- Let $\mathcal{A} \subset C_b(X)$ be a nowhere vanishing subalgebra that seperates points
- Let $\mathcal{A}' \subset C_b(\mathcal{M}_1(X))$ be the subalgebra generated by the maps $\mu \mapsto \mu[f]$ for $f \in \mathcal{A}$
- Let $\mu \in \mathcal{M}_1(X)$, if $\mu[f] = 0$ for all $f \in \mathcal{A}$ then μ is not even a probability measure, thus there must exist $f \in \mathcal{A}$ such that $\mu[f] \neq 0$ so $\mathcal{A}'$ is nowhere vanishing
- Let $\mu, \nu \in \mathcal{M}_1(X)$ such that $\mu \neq \nu$, if $\mu[f] = \nu[f]$ for all $f \in \mathcal{A}$ then it follows that $\mu = \nu$ which is a contradiction, thus there must exist $f \in \mathcal{A}$ such that $\mu[f] \neq \nu[f]$ and thus $\mathcal{A}'$ separates points

The generated subalgebra

- The generated subalgebra $\mathcal{A}'$ consists of finite linear combinations monomials $\mu \mapsto \prod_{i=1}^{k} \mu[f_i]$ where $f_i \in \mathcal{A}$
- By linearity, we are thus interested in expectations of the form $\mathbb{E} \prod_{i=1}^{k} \mu[f_i]$
- If we interpret this as a random vector $(\mu[f_1], \mu[f_2], ..., \mu[f_k])$, then the distribution of this random vector is uniquely defined by its characteristic function $t \mapsto \mathbb{E} e^{i \sum_{i=1}^{k} t_i \mu[f_i]}$
- Notice that $\sum_{i=1}^{k} t_i \mu[f_i] = \mu \left[\sum_{i=1}^{k} t_i f_i \right]$ and $\sum_{i=1}^{k} t_i f_i \in \mathcal{A}$
- It follows that the expectations of monomials are completely determined by the maps $s \mapsto \mathbb{E} e^{is\mu[f]}$ where $s \in \mathbb{R}$ and $f \in \mathcal{A}$
- Going backwards, we see that this is precisely the characteristic function the random variable $\mu[f]$

Penultimate theorem

Theorem

Let X be a Polish space and let $\mathcal{A} \subset C_b(X)$ be a nowhere vanishing subalgebra that separates points. It follows that $\mu = \nu$ in distribution if and only if $\mu[f] = \nu[f]$ in distribution for any $f \in \mathcal{A}$.

- This result follows by noting that the previous slide proved that $\mathcal{A}'$ is separating and its elements expectations are uniquely determined by the distributions of $\mu[f]$ for $f \in \mathcal{A}$
- One should now understand that the "distribution of a distribution" are terms of the form $\mathbb{P}(\mu \in A)$ for open sets A and "expectations of expectations" are the terms of the form $\mathbb{E}g(\mu[f])$ for $g \in C_b(\mathbb{R})$ and $f \in \mathcal{A}$

Uniform tightness of random measures

- The last object to discuss is the uniform tightness of a collection of random probability measures
- We begin with a "definitive" compact set in the space of probability measures
- Let $\varepsilon > 0$ and let $K_{k,\varepsilon}$ be a sequence of compact sets in X

Uniform tightness of random measures cont.

- Consider the set

$$\mathcal{K}_\varepsilon := \bigcap_{k=1}^{\infty} \left\{ \mu : \mu(K_{k,\varepsilon}) \geq 1 - \frac{\varepsilon}{2^k} \right\}.$$

- Observe that if we take a sequence $\mu_n \in \mathcal{K}_\varepsilon$, then by definition this sequence is uniformly tight and thus there exists a weakly convergent subsequence
- One consequence of weak convergence is that $\mu(F) \geq \limsup_n \mu_n(F)$ for any closed sets and thus the weakly convergent limit satisfies the inequalities in the intersection
- It follows that $\mathcal{K}_\varepsilon$ is compact for any $\varepsilon > 0$

Uniform tightness of random measures cont.

- For a random measure μ, we introduce the intensity measure $\mathbb{E}\mu$ by duality

$$(\mathbb{E}\mu)[f] := \mathbb{E}\mu[f]$$

  for $f \in C_b(X)$
- If $K \subset X$ is a compact set then $\mu(K)$ is a random variable and by Markov's inequality, we have

$$\mathbb{P}(\mu(K) \geq 1 - \sqrt{\varepsilon}) \geq 1 - \frac{1 - \mathbb{E}\mu(K)}{\sqrt{\varepsilon}}$$

FUniform tightness of random measures cont.

- It follows that if the intensity measures $\mathbb{E}\mu_n$ are uniformly tight, then for every $\varepsilon > 0$ there exists a compact set K_ε such that

$$\mathbb{P}(\mu_n(K_\varepsilon) \geq 1 - \sqrt{\varepsilon}) \geq 1 - \sqrt{\varepsilon}$$

  for all n

- Returning to the set $\mathcal{K}_\varepsilon$, we have

$$\begin{aligned} 1 - \mathbb{P}(\mu_n \in \mathcal{K}_\varepsilon) &\leq \sum_{k=1}^{\infty} \left(1 - \mathbb{P}\left(\mu_n(K_{k,\varepsilon}) \geq 1 - \frac{\varepsilon}{2^k}\right)\right) \\ &\leq \sqrt{\varepsilon} \sum_{k=1}^{\infty} \frac{1}{2^{\frac{k}{2}}} \end{aligned}$$

- It follows that μ_n is uniformly tight

Converse result

- The converse result is "trivial"
- If μ_n is uniformly tight and $\mathbb{E}\mu_{n_k}$ is a subsequence, by uniform tighhtness there always exists a subsubsequence $\mu_{n_{k_j}}$ which converges weakly
- Because $\mu \mapsto \mu[f]$ is a continuous function for all $f \in C_b(X)$, it follows that $\mathbb{E}\mu_{n_{k_j}}[f] \to \mathbb{E}\mu[f]$ for any $f \in C_b(X)$ which implies that the intensity measure is uniformly tight

Ultimate theorem

Theorem

Let μ_n be a sequence of random measures such that the sequence of intensity measures $\mathbb{E}\mu_n$ is uniformly tight. Let $\mathcal{A} \subset C_b(X)$ be a nowhere vanishing subalgebra that separates points. It follows that if $\mu_n[f] \to \mu[f]$ weakly for any $f \in \mathcal{A}$ then $\mu_n \to \mu$ weakly.

- It follows that for such uniformly tight random measures it is enough to study the weak convergence of the evaluation maps $\mu_n[f]$ for $f \in \mathcal{A}$

Final example

- Recall that random mean Gaussian measure $\mu(m)$ defined by

$$\mu(m)[f] := \frac{1}{\sqrt{2\pi}^k} \int_{\mathbb{R}^k} dx \ e^{-\frac{||x-m||^2}{2}} f(x)$$

- The intensity measure $\mathbb{E}\mu$ is equivalent to

$$\mathbb{E}\mu[f] = \frac{1}{\sqrt{2\pi}^k} \int_{\mathbb{R}^k} dx \ e^{-\frac{||x||^2}{2}} \mathbb{E}f(x+m)$$

- Moments

$$\mathbb{E}\mu[x_i x_j] = \delta_{ij} + \mathbb{E}m_i m_j, \ \mathbb{E}\mu[x_i]\mu[x_j] = \mathbb{E}m_i m_j$$

Thank you!