SL Math Stochastic Quantization Summer School
TA Session on tightness of probability measures

Evan Sorensen

Monday, July 1, 2024

1 Definitions

For a metric space S, let $\mathcal{B}(S)$ be its Borel σ-algebra. Let $\mathcal{P}(S)$ be the set of probability measures on $(S, \mathcal{B}(S))$. Let $C_b(S)$ be the set of all bounded continuous functions $S \to \mathbb{R}$, and let $B_b(S)$ be the set of all bounded Borel measurable functions $S \to \mathbb{R}$.

We say that a set collection of probability measures $\Lambda \subseteq \mathcal{P}(S)$ is **tight** if, for every $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subseteq S$ so that, for all $\mu \in \mathcal{P}(S)$, $\mu(K_\varepsilon) > 1 - \varepsilon$.

We may also refer to tightness of a set of random variables. For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a collection $\mathcal{X}$ of random variables $X : \Omega \to S$, is tight, if for every $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subseteq S$ so that, for every $X \in \mathcal{X}$, $\mathbb{P}(X \in K_\varepsilon) > 1 - \varepsilon$. Note that, if for a random variable $X : \Omega' \to \Omega$, we let μ_X be push-forward of $\mathbb{P}$ under X, then this definition is equivalent to saying that the set $\{\mu_X\}_{x \in \mathcal{X}}$ is a tight family of measures in $\mathcal{P}(S)$.

We say a sequence of probability measures $(\mu_n)_{n \in \mathbb{N}} \subseteq \mathcal{P}(S)$ converges weakly to $\mu \in \mathcal{P}(S)$, if for every $f \in C_b(S)$,

$$\int_S f(x)\, \mu_n(dx) \overset{n \to \infty}{\longrightarrow} \int_S f(x)\, \mu(dx).$$

We also say a sequence of random variables $X_n : \Omega \to S$ converges in distribution to a random variable $X : \Omega \to S$ if, for every $f \in C_b(S)$,

$$\mathbb{E}[f(X_n)] \overset{n \to \infty}{\longrightarrow} \mathbb{E}[f(X)],$$

where $\mathbb{E}$ denotes expectation under $\mathbb{P}$.

We say that a set of measures $\Lambda \subseteq \mathcal{P}(S)$ is **relatively compact** if, for every sequence in Λ, there exists a weakly convergent subsequence (the limit lies in $\mathcal{P}(S)$, but not necessarily Λ).

The following is proved, for example, in [Bil99, Section 5]. See also [DP06, Section 6.3] for a proof of the forward implication in the case where Ω is a Hilbert space.

Theorem 1.1 (Prohorov's Theorem)**.** *Let S be a metric space. If $\Lambda \subseteq \mathcal{P}(S)$ is tight, then Λ is relatively compact. If S is Polish, then the converse also holds.*

Recall that a topological space S is **Polish** if it is separable and completely metrizable. The "metrizable" part is important because a Polish space may be complete with respect to one metric but not another. A classic example is the space $(0,1)$, which is homeomorphic to $\mathbb{R}$ and hence Polish; however, it is not complete under the Euclidean metric. In this summer school, we will typically take S to be a separable Hilbert space, which is Polish.

Recall that a G_δ set is a countable intersection of open sets. We will not make use of the following fact in these notes, but it may be helpful to know. (For a proof, see, for example, [Sri98, Theorems 2.2.1 and 2.2.7])

Theorem 1.2 (Alexandrov's Theorem)**.** *Let S be a Polish space and $A \subseteq S$. Then, A is a Polish space under the subspace topology if and only if A is a G_δ subset of S.*

2 Some examples

Example 2.1. *On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables taking values in $\mathbb{R}^d$, and let $|\cdot|$ be the Euclidean norm on $\mathbb{R}^d$. If there exist $C, p > 0$ so that $\mathbb{E}|X_n|^p \le C$ for all $n \in \mathbb{N}$, then $(X_n)_{n\in\mathbb{N}}$ is tight.*

Proof. For $\varepsilon > 0$, let K_ε be the compact set $\{x \in \mathbb{R}^d : |x| \le \left(\frac{C}{\varepsilon}\right)^{1/p}\}$ By Markov's inequality,

$$\mathbb{P}(X_n \notin K_\varepsilon) = \mathbb{P}\left(|X_n|^p > \frac{C}{\varepsilon}\right) \le \varepsilon. \qquad \square$$

The above example is quite simple in finite dimensions. Suppose we instead have random variables $(X_n)_{n\in\mathbb{N}}$ taking values in an infinite-dimensional normed vector space S, with norm $|\cdot|_S$ and we have uniform bounds $\mathbb{E}|X_n|_B^p \le C$ for some $C, p > 0$. It turns out that this is not enough to conclude tightness in this more general setting, because, for $T > 0$, the set $\{x \in S : |x|_S \le T\}$ is not necessarily contained in a compact set. However, these uniform moment bounds are enough to conclude tightness in a larger space S', if S is compactly embedded in S'.

To show this, we first recall some important definitions: A subset A of a metric space S is **bounded** if there exists $C > 0$ so that $|x|_S \le C$ for all $x \in A$. The set A is **totally bounded** if every sequence in A contains a Cauchy subsequence.

For two normed vector spaces S_1 S_2 with norms $|\cdot|_{S_1}$ and $|\cdot|_{S_2}$, we say that S_1 is **compactly embedded** in S_2 if $S_1 \subseteq S_2$ and the following hold

1. The inclusion map $\mathfrak{i} : S_1 \to S_2$ is continuous; i.e., there exists a constant $C > 0$ such that $|\mathfrak{i}x|_{S_2} \le C|x|_{S_1}$ for all $x \in S_1$.

2. For every bounded set $A \subseteq S_1$, the set A is totally bounded in S_2.

Example 2.2. *On* $(\Omega, \mathcal{F}, \mathbb{P})$*, let* $(X_n)_{n\in\mathbb{N}}$ *be a sequence of random variables taking values in a normed vector space* $(S_1, |\cdot|_{S_1})$*, and suppose that* S_1 *is compactly embedded in the complete vector space* $(S_2, |\cdot|_{S_2})$*. Suppose also that there exist* $C, p > 0$ *so that* $\mathbb{E}|X_n|_{S_1}^p \leq C$ *for all* $n \in \mathbb{N}$*. Then, the sequence* $(X_n)_{n\in\mathbb{N}}$ *is tight in the larger space* S_2*.*

Proof. Let $\varepsilon > 0$, and set $K_\varepsilon = \{x \in S_1 : |x|_{S_1} \leq \left(\frac{C}{\varepsilon}\right)^{1/p}\}$. This set is bounded in S_1, so K_ε is totally bounded in S_2. It is a standard fact that, if S is a complete metric space, then a subset $A \subseteq S$ is has compact closure if and only if it totally bounded. Hence, the closure of K_ε in S_2, denoted $Cl_{S_2}(K_\varepsilon)$ is compact.

By the same computation as in the finite-dimensional case,

$$\mathbb{P}(X_n \notin Cl_{S_2}(K_\varepsilon)) \leq \mathbb{P}(X_n \notin K_\varepsilon) \leq \varepsilon. \qquad \square$$

We now give a concrete example of a compact embedding which will appear in the lectures.

Example 2.3. *Consider the weighted* ℓ^p *spaces n the lattice* $\mathbb{Z}^d$*:*

$$\ell^p_\sigma = \{f : \mathbb{Z}^d \to \mathbb{R} : |f|_{\ell^p_\sigma} := |\rho f|_{\ell^p} < \infty, \quad \rho(x) = (1 + b|x|)^{-\sigma}\},$$

where $b > 0$ *is a constant. If* $\sigma' > \sigma$*, then* $\ell^p_\sigma \hookrightarrow \ell^p_{\sigma'}$ *is a compact embedding.*

Proof. It follows immediately that $|f|_{\ell^p_{\sigma'}} \leq |f|_{\ell^p_\sigma}$, so the embedding is continuous. Now, let A be a bounded set in ℓ^p_σ. We show A is totally bounded in $\ell^p_{\sigma'}$. Let $(f_k)_{k\in\mathbb{N}}$ be a sequence in A. By boundedness of A, there exists a constant $C_A > 0$ so that, for all $k \in \mathbb{N}$,

$$|f_k|^p_{\ell^p_\sigma} = \sum_{x\in\mathbb{Z}^d} (1 + b|x|)^{-\sigma p} |f_k(x)|^p \leq C_A. \tag{2.1}$$

In particular, $(1 + b|x|)^{-\sigma p}|f_k(x)|^p \leq C_A$ for all $k \in \mathbb{N}$ and $x \in \mathbb{Z}^d$. Enumerate the points of $\mathbb{Z}^d$ as $x_1, x_2, \ldots$ By compactness of bounded sets in $\mathbb{R}$, there exists a subsequence $(f_{1,k})_{k\in\mathbb{N}}$ of (f_k) so that $f_{1,k}(x_1)$ converges, and we call the limit $f(x_1)$. There exists a further subsequence $(f_{2,k})_{k\in\mathbb{N}}$ so that $f_{2,k}(x_2)$ converges, and we call the limit $f(x_2)$. Continue this way, constructing sequences $(f_{j,k})_{k\in\mathbb{N}}$ for each $j \in \mathbb{N}$ so that each subsequence is a subsequence of the previous, and $f_{j,k}(x_j)$ converges to a limit, which we call $f(x_j)$. This defines a function $f : \mathbb{Z}^d \to \mathbb{R}$, and it is standard to show that the diagonal subsequence $\left(f_{k,k}\right)_{k\in\mathbb{N}}$ converges pointwise to f. In particular, for each $x \in \mathbb{Z}^d$, $\left(f_{k,k}(x)\right)_{k\in\mathbb{N}}$ is Cauchy.

We show that $(f_{k,k})_{k\in\mathbb{N}}$ is Cauchy in $\ell^p_{\sigma'}$ for all $\sigma' > \sigma$. Let $\varepsilon > 0$, and let $q' > 1$ be sufficiently large so that

$$\sum_{x\in\mathbb{Z}^d\setminus[-M,M]^d} (1 + b|x|)^{(\sigma-\sigma')pq'} < \infty.$$

Then, we may choose $M = M(\varepsilon)$ sufficiently large so that

$$\left(\sum_{x\in\mathbb{Z}^d\setminus[-M,M]^d} (1+b|x|)^{(\sigma-\sigma')pq'}\right)^{\frac{1}{q'}} \le \frac{\varepsilon}{4C_A}.$$

Next, let $N = N(M,\varepsilon)$ be sufficiently large so that, for all $x \in [-M,M]^d$ and $k, j \ge N$, $(1+b|x|)^{-\sigma' p}|f_{k,k}(x) - f_{j,j}(x)|^p \le \frac{\varepsilon}{2(2M+1)^d}$. Let $p' > 1$ be defined by $\frac{1}{q'} + \frac{1}{p'} = 1$. Then, for $k, j \ge N$,

$$|f_{k,k} - f_{j,j}|^p_{\ell^p_{\sigma'}} \le \sum_{x\in[-M,M]^d} (1+b|x|)^{-\sigma' p}|f_{k,k}(x) - f_{j,j}(x)|$$
$$+ \sum_{\mathbb{Z}^d\setminus[-M,M]^d} (1+b|x|)^{(\sigma-\sigma')p}(1+b|x|)^{-\sigma p}|f_{k,k}(x) - f_{j,j}(x)|^p$$
$$\le \frac{\varepsilon}{2} + \left(\sum_{x\in\mathbb{Z}^d\setminus[-M,M]^d} (1+b|x|)^{(\sigma-\sigma')pq'}\right)^{\frac{1}{q'}} \left(\sum_{\mathbb{Z}^d\setminus[-M,M]^d} (1+b|x|)^{-\sigma pp'}|f_{k,k}(x) - f_{j,j}(x)|^{pp'}\right)^{\frac{1}{p'}}$$
$$\le \frac{\varepsilon}{2} + \frac{\varepsilon}{4C_A}(|f_{k,k}|^p_{\ell^{pp'}_\sigma} + |f_{j,j}|^p_{\ell^{pp'}_\sigma}) \le \varepsilon,$$

where in the second equality, we used Hölder's inequality, and in the last inequality, we used the monotonicity of ℓ_p norms $|g|_{\ell^{pp'}} \le |g|_{\ell^p}$ (see, for example, [Fol99, Proposition 6.11]), together with (2.1). □

3 The Krylov-Bogoliubov theorem

In this section, we loosley follow the notation and conventions of [DP06, Sections 5 and 7], although we shall work more generally with a metric space S, instead of a Hilbert space H.

Let $L(C_b(S))$ be the space of all linear bounded operators $C_b(S) \to C_b(S)$. A **Markov semigroup** $(P_t)_{t\ge0}$ on S, is a function $[0,\infty) \to L(C_b(S))$, denoted $t \mapsto P_t$ satisfying the following:

1. $P_0 = 1$ (the identity operator), and $P_{t+u} = P_tP_u$ for all $t, s \ge 0$.
2. For any $t \ge 0$ and $x \in S$, there exists a probability measure $\pi_t(x,\cdot) \in \mathcal{P}(S)$ so that
$$P_tf(x) = \int_S f(y)\pi_t(x,dy) \quad \text{for all } f \in B_b(S).$$
3. For all $f \in C_b(S)$ (resp. $B_b(S)$) and $x \in S$, the mapping $t \mapsto P_tf(x)$ is continuous (resp. Borel).

For a Markov chain $(X_t)_{t\ge0}$ with state space S, the probability measure $\pi_t(x, dy)$ is the probability density of X_t when started from $X_0 = x$. Then, $P_tf(x) = \mathbb{E}^x[f(X_t)]$, where $\mathbb{E}^x$ denotes the expectation with respect to the probability measure $\mathbb{P}^x$ of the chain $(X_t)_{t\ge0}$ started from $X_0 = x$.

A Markov semigroup $(P_t)_{t\geq 0}$ has the **Feller property** if $P_t f \in C_b(S)$ for all $t \geq 0$ whenever $f \in C_b(S)$.

A probability measure $\mu \in \mathcal{P}(S)$ is **invariant** for P_t if, for all $t > 0$ and all $f \in C_b(S)$,

$$\int_S f(x)\,\mu(dx) = \int_S P_t f(x)\,\mu(dx).$$

For a Markov semigroup $(P_t)_{t\geq 0}$ on S, $x_0 \in S$, and $T > 0$, define the time-averaged measure

$$\mu_T^{x_0}(B) := \int_0^T \pi_t(x_0, B)\,dt, \qquad B \in \mathcal{B}(S).$$

Theorem 3.1 (The Krylov-Bogobliubov Theorem)**.** *Let $(P_t)_{t\geq 0}$ be a Markov Feller semigroup with state space S, where S is a matric space. Assume that for some $x_0 \in S$, the set $(\mu_T^{x_0})_{T>0}$ is tight. Then, there exists an invariant measure for $(P_t)_{t\geq 0}$. Specifically, any subsequential limit of $(\mu_T^{x_0})_{T>0}$ is invariant.*

We remark here that the assumption that $(\mu_T^{x_0})_{T>0}$ is tight is weaker than the assumption that $\big(\pi_t(x_0,\cdot)\big)_{t\geq 0}$ is tight.

Proof. We follow the proof in [DP06, Theorem 7.1]. To get a rough sense of the idea, think of starting the Markov process at time 0 from $\mu_T^{x_0}$ for some large T. Then, for $0 \leq u \ll T$, the operator P_u does not change the measure $\mu_T^{x_0}$ by much since $\mu_T^{x_0}$ is an average of measures over a long time.

We move to the full proof. By Prohorov's theorem, there exists a subsequential limit μ of $\big(\mu_T^{x_0}\big)_{T>0}$, which means there exists a sequence $T_k \to \infty$ such that, for all $f \in C_b(S)$,

$$\lim_{k\to\infty} \int_S f(x)\,\mu_{T_k}^{x_0}(dx) = \int_S f(x)\,\mu(dx). \tag{3.1}$$

We use Fubini's Theorem to rewrite the prelimiting object on the left-hand side as follows:

$$\int_S f(x)\,\mu_{T_k}^{x_0}(dx) = \int_S \frac{f(x)}{T_k} \int_0^{T_k} \pi_t(x_0,\, dx)\,dt = \frac{1}{T_k}\int_0^{T_k} P_t f(x_0)\,dt. \tag{3.2}$$

Combining (3.1) and (3.2), for all $f \in C_b(S)$,

$$\lim_{k\to\infty} \frac{1}{T_k}\int_0^{T_k} P_t f(x_0)\,dt = \int_S f(x)\,\mu(dx). \tag{3.3}$$

Let $g \in C_b(S)$. For $u \geq 0$, we seek to show that $\int_S g(x)\,\mu(dx) = \int_S P_u g(x)\,\mu(dx)$. We apply (3.3) to $f = P_u g$ (using the Feller property so that $f \in C_b(S)$) and use the semigroup property $P_{t+u} = P_t P_u$ to obtain

$$\lim_{k\to\infty} \frac{1}{T_k}\int_S P_{t+u} g(x_0)\,dt = \int_S P_u g(x)\,\mu(dx). \tag{3.4}$$

It thus suffices to show that the left-hand side of (3.4) is $\int_S g(x)\mu(dx)$. Observe that

$$\begin{aligned} &\frac{1}{0}\int_0^{T_k} P_{t+u}g(x_0)\,dt = \frac{1}{T_k}\int_u^{T_k+u} P_t g(x_0)\,dt \\ &= \frac{1}{T_k}\int_0^{T_k} P_t g(x_0)\,dt + \frac{1}{T_k}\int_{T_k}^{T_k+u} P_t g(x_0)\,dt - \frac{1}{T_k}\int_0^{u} P_t g(x_0)\,dt. \end{aligned} \tag{3.5}$$

By (3.3) applied to $f = g$, $\frac{1}{T_k}\int_0^{T_k} P_t g(x_0)\,dt$ converges to $\int_S g(x)\,\mu(dx)$ as $k \to \infty$. Hence, it remains to show the other two terms on the right of (3.5) converge to 0. This follows because $g \in C_b(S)$, so for each $t > 0$,

$$|P_t g(x_0)| \le \int_S |g(x)|\,\pi_t(x_0, dx) \le \sup_{x\in S} |g(x)|,$$

and the right-hand side is a finite bound not depending on t. □

References

[Bil99] Patrick Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[DP06] Giuseppe Da Prato. *An introduction to infinite-dimensional analysis.* Universitext. Springer-Verlag, Berlin, 2006. Revised and extended from the 2001 original by Da Prato.

[Fol99] Gerald B. Folland. *Real analysis.* Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[Sri98] S. M. Srivastava. *A course on Borel sets*, volume 180 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1998.