COMPACTIFY NO MORE?

Let (X, d) be a metric space, let Γ be a family of measures in $\mathcal{P}(X)$. Prokhorov's theorem ensures that, in order to prove that Γ is relatively compact, it is enough to prove that it is *tight*. The tightness assumption reads as follows: for every $\varepsilon > 0$, there exists a compact K_ε in X such that $\mu(K_{eps}) > 1 - \varepsilon$ for every μ in Γ.

Of course, if X is compact, then every family is tight. On the other hand, if X is compact, then $\mathcal{P}(X)$ is compact (and thus every family in $\mathcal{P}(X)$ is relatively compact), but that is far less obvious. Given a sequence of probability measures μ_k on X, how would we extract a converging subsequence?

- First, let us consider the simple case where each μ_k is purely atomic, supported on 100 points in X. Each point carries a mass between 0 and 1, and we can represent μ_k as an element of $X^{100} \times [0, 1]^{100}$. By elementary compactness results, there exists a convergent subsequence of images in that space. It is then easy to check that the limit point corresponds to a purely atomic probability measure μ, and that the convergence of the atoms and the weights implies the weak convergence of the corresponding probability measures μ_k to μ.
- In the general, non-atomic case, one could attempt to use a similar argument by discretizing: split X into a finite number of small boxes that cover up the whole space, and represent μ_k in an approximate way by the data of $\mu_k(A)$ for A in the finite σ-algebra generated by the boxes. Each data point can be embedded into a finite power of $[0, 1]$, in which we can take limit points. Any equation involving a finite number of Borel sets will pass to the limit, so we obtain some sort of discretized measure in the limit, that passes sanity checks. I believe that a fun little exercise shows it can be extended into a "full" measure on X, by placing atoms at the right spots. Of course by doing that we have not constructed an actual limit point for the μ_k, but an "approximate limit".
- The approximation in the previous step comes from the discretization procedure, yielding a finite number of Borel sets. This had two advantages:
  1. We could embed measures as a finite power of the interval $[0, 1]$, and use compactness.
  2. The "sanity checks", i.e. the fact that the limit point satisfies the properties of a measure, were all finitary, and thus trivially true by passing their "finite k" version to the limit $k \to \infty$.

  In fact, for point 1. there is no actual advantage in having finitely many sets. By Tychonov's theorem (a very much non-constructive result indeed) the product of an arbitrary number of compact sets is compact for the product topology. Instead of considering a finite-scale graining of the space, we could represent each measure μ_k into an infinite product of $[0, 1]$ by listing the data of $\mu_k(A)$ for *every* Borel set A. A limit point μ for those objects is thus a way to attribute mass between 0 and 1 to every Borel set, in a fairly consistent way: the mass of $\emptyset$ is 0, the total mass is 1, and μ is finitely additive. Let us point out that (after extraction) μ is very much the weak limit of the μ_k's in the sense that e.g. $\liminf_k \mu_k(O) \geq \mu(O)$ for any open set O: in fact by construction the limit exists and there is equality! However, the requirement that μ be σ-additive is not obvious, as it requires to exchange two limits. That is precisely where compactness comes into play, as it basically allows to extract from an infinite (in particular: from a countable) family of sets a finite one: after that we can pass information about each element to the limit.

The most common way to prove that X compact implies $\mathcal{P}(X)$ compact is slightly different from the previous idea. First, one sees a probability measure as living in the dual space of continuous functions (when X is compact, there is no need to specify: bounded, compactly supported, etc.). Existence of a limit point follows from the Banach-Alaoglu-Bourbaki theorem about compactness of the unit ball for the weak-$*$ topology: this is nothing but a short application of Tychonov. The difficult step is to argue that this limit point, living in a dual space, really corresponds to some probability measure on X: this is done by applying the Riesz-(Markov-Kakutani) representation theorem (below shortened as RRT), which is valid and easy to state for X compact (there are variations when X is locally compact). The proof of RRT looks very much like the argument sketched above: one defines the corresponding measure "setwise" in a fairly natural way, the finite additivity is easy to check, and σ-additivity is reduced to finite additivity through a careful application of compactness. We sketched above a proof for Prokhorov in the compact case, whose ingredients are: a convenient representation of measures, a compactness argument, and a verification that σ-additivity works well at the level of the limit point. The first step can (but need not to) be phrased in dual terms, in which case the second step is named "Banach-Alaoglu-Bourbaki": that is, at most, a matter of convenience. The last step is the only really delicate point: it can be found among the proof of RRT and then written back in "primal" language.

What about X not compact? When X is not compact, it is no longer true that $\mathcal{P}(X)$ is compact: take $X = \mathbb{N}$ and the family of Dirac masses at each successive integer. Coincidentally, RRT fails for non-compact spaces with e.g. the exact same counter-example, but read dually: the positive linear functional on $C_b(\mathbb{N}$ given by $\limsup$ cannot be represented by a measure. In the case of $\mathbb{N}$ the lack of compactness is felt because mass can leak at infinity, but one could imagine other defects. In order to save compactness in $\mathcal{P}(X)$, or to fix the RRT, one has to add an assumption of tightness. For example, there is a "non-compact RRT" that says: any positive linear functional on $C_b(X)$ can be represented as a measure on X, *provided that it satisfies the following*: for every $\varepsilon > 0$, there exists a compact K such that $|\varphi(f)| < \varepsilon$ when f has norm at most 1 and is supported outside of K. That is good old tightness, but dual.

Where does tightness come into play, really? Take a countable bunch of open sets A_i, one wants to prove e.g.

$$\mu\left(\bigcup A_i\right) \leq \sum_i \mu(A_i).$$

Fix $\varepsilon > 0$, take a compact K big enough such that the μ_k (and thus μ) give mass at least $1 - \varepsilon$ to K. Intersecting the A_i's with K puts us back to a compact context, where we can extract a finite covering and pass to the limit in the inequalities involving μ_k's. We are making a small error on the masses, controlled by an arbitrary ε, so we are good. Most proofs use the following route:

(1) Embed X into a compactified space $\hat{X}$, and $\mathcal{P}(X)$ into the space of probability measures on $\hat{X}$, which we know is thus compact, too.
(2) Obtain a limit point that is a measure on $\hat{X}$. Tightness does not play any role here, because every family over $\hat{X}$ is tight for free!
(3) Observe that, really, it is (the image of) a measure on X, because it concentrates on $X \subset \hat{X}$. This is where tightness comes into play.

That, to me, leaves an air of mystery, if not magic. We compactify, deduce something at the level of the compactified space, and then pullback everything into the original space, as if nothing happened? The point of this note is to convince myself (and an hypothetical reader) that, indeed, nothing happened - and certainly not any magic. One could write the proof without wearing duality lenses and without compactifying.