Math 254A Lecture 7 Notes

Daniel Raban

April 12, 2021

1 Large Deviations and Affine Approximation of Semicontinuous Functions

1.1 Recap

Here is our main result so far: We have a σ-finite measure space (M, λ) and a locally convex topological vector space, and $\mathcal{U}$ as the collection of open convex sets on X. We assume that every $U \in \mathcal{U}$ is an increasing union of compact, convex sets (e.g. $\mathbb{R}^d, Y^*$). We also have a measurable map $\varphi : M \to X$ which takes values in a metrizable subset. Then

$$\lambda^{\times n}\left(\left\{p \in M^n : \frac{1}{n}\sum_{i=1}^n \varphi(p_i) \in U\right\}\right) = e^{n \cdot s(U) + o(n)}$$

for $U \in \mathcal{U}$. And if $s : \mathcal{U} \to [-\infty, \infty]$ ($\neq +\infty$ if s is locally finite), then there exists a point function $s : X \to [-\infty, \infty)$ which is upper semicontinuous and concave with $s(U) = \sup\{s(x) : x \in U\}$.

Example 1.1. In our original counting of type classes, we had $M = A$ is a finite alphabet, λ is counting measure, $p(a) = \delta_a$, and $\frac{1}{n}\sum_{i=1}^n \varphi(a_i) = p_a$ is the empirical distribution.

Example 1.2. In Large Deviations Theory, (M, λ) is a probability space, and $X = \mathbb{R}$. Then $\xi_1 = \varphi(p_1), \xi_2 = \varphi(p_2), \dots$ are iid random variables. Then the theorem says

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \xi_i \in U\right) = \exp\left(n \cdot \sup_U s(x) + o(n)\right).$$

Here, $s \leq 0$ always.

1.2 The large deviations principle

How does this fit into probability theory? Suppose $\mathbb{E}[|\xi_i|] < \infty$ (iff $\varphi \in L^1(\lambda)$). Then the Weak Law of Large Numbers says

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \xi_i \in U\right) \to \begin{cases} 1 & \mathbb{E}[\xi_i] \in U \\ 0 & \mathbb{E}[\xi_i] \notin \overline{U}. \end{cases}$$

In the case where $\sup_U s < 0$, this gives an exponential decay, upgrading the result of the Weak Law of Large Numbers. We can see Large Deviations Theory as a refinement of the convergence to zero in the WLLN.

The most "standard" formulation of the large deviations principle says

- $$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i \in U\right) \geq \exp\left(n \cdot \sup_{x\in U} s(x) + o(n)\right)$$

  for all open $U \subseteq \mathbb{R}$. [This follows from the observation that LHS $\geq \mathbb{P}(\frac{1}{n}\sum_{i=1}^{n}\xi_i \in I)$ for all open intervals $I \subseteq U$.]

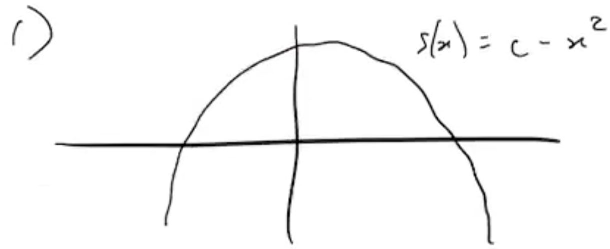
- $$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i \in C\right) \leq \exp\left(n \cdot \sup_{C} s(x) + o(n)\right)$$

  for all closed $C \subseteq \mathbb{R}$. [This follows from above if C is compact: if $\sup_C s(x) = \alpha$, then we can cover C with finitely many open intervals $I_1, \ldots, I_k$ such that $\mathbb{P}(\frac{1}{n}\sum_{i=1}^{n}\xi_i \in I_\ell) \leq e^{n\cdot\sup_{I_\ell} s + o(n)}$ for all $\ell \leq k$. We can extend this to closed sets if $s(x) \to -\infty$ as $x \to \pm\infty$, in which case s is called *good*.[1] If s is good, we can cover a general closed set with far away half infinite intervals on each side and have a compact set in the middle. The apply the previous argument.]

1.3 Approximation of concave, upper semicontinuous functions by affine functions

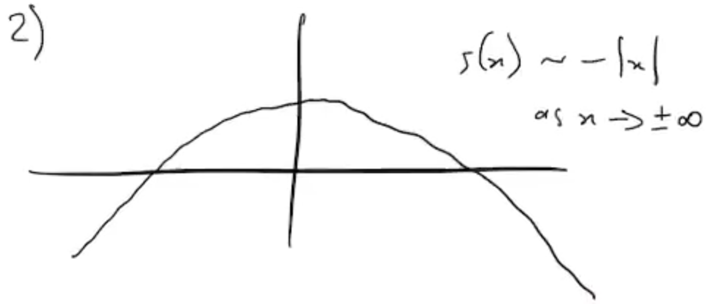
Returning to the general story, assuming local finiteness, $s : X \to [-\infty, \infty)$ is upper semicontinuous and concave. How can we describe these in general? Here are some examples where $X = \mathbb{R}$:

Example 1.3. $s(x) = c - x^2$



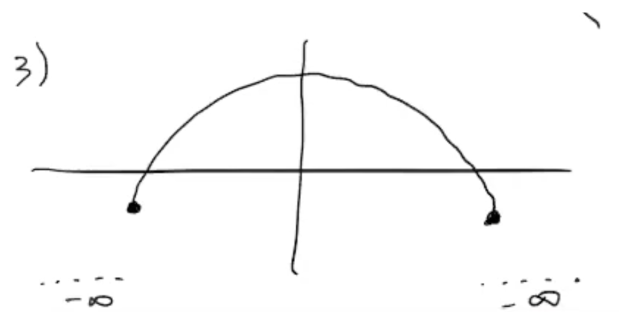
[1] This is also called *proper* in analysis.

Example 1.4. $s(x) \sim -|x|$ as $x \to \pm\infty$.

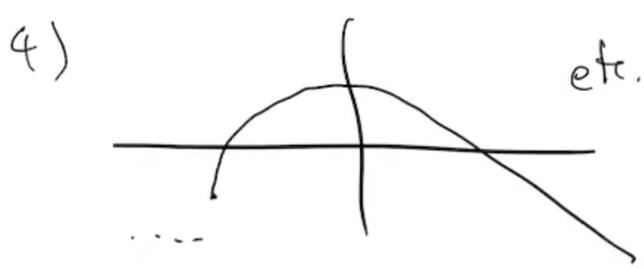


Example 1.5. Upper semicontinuous example with

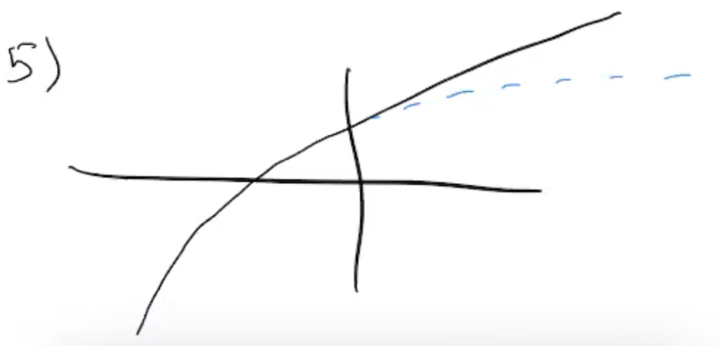
$$s(x) = \begin{cases} c - x^2 & |x| \le M \\ -\infty & |x| > M. \end{cases}$$



Example 1.6. An example that tends to $-\infty$ on the right:



Example 1.7. An example which tends to $+\infty$ on the right:



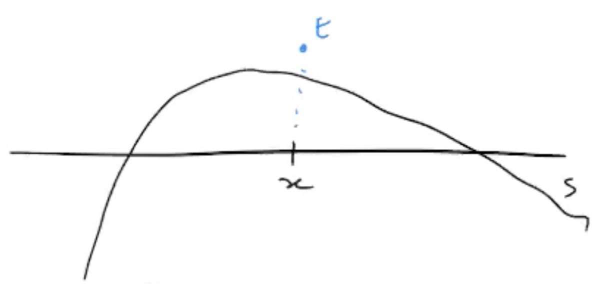
The key to all these cases is whether we can draw a straight tangent line that lies entirely above the graph. In example 3, we run into a bit of trouble at the endpoints, since we cannot draw vertical line (with infinite slope), so we may need an ε bit of wiggle room. What this idea leads to is the fact that any upper semicontinuous function can be written as an infimum of affine functions. Here is a lemma that we need.

Lemma 1.1. *Let X be a locally convex topological vector space, and let $s : X \to [-\infty, \infty)$ be an upper semicontinuous concave function with $x \in X$. If $t > s(x)$, then there exists a $c \in \mathbb{R}$ and a continuous functional y on X such that*

- $c + \langle y, z \rangle \geq s(z)$ *for all* $z \in X$,
- $c + \langle y, x \rangle < t$.

This is the infinite dimensional analogue of whether we can place a line above the graph of $s(x)$ which stays below any point above the graph.

Proof.



We want to think of this as a picture in a larger topological vector space that includes the vertical coordinate. Let $\widetilde{X} = X \times \mathbb{R}$, which is a locally convex topological vector space with the product topology. The point (x, t) lies above the subset $C := \{(x, \theta) : \theta \leq s(z)\}$.

This subset is closed because s is upper semicontinuous and is convex because s is concave. By the Hahn-Banach separation theorem, there exists a $\widetilde{y} \in \widetilde{X}^*$ such that $\widetilde{y}(x,t) > \sup_C \widetilde{y}$. Also, $\widetilde{y}$ can be written as $\widetilde{y}(z,\theta) + \langle y, z \rangle + \alpha\theta$ for some $y \in X^*$ and $\alpha \in \mathbb{R}$. If we let c be the y-intercept of the hyperplane given by Hahn-Banach and rewrite the inequality $\widetilde{y}(x,t) > \sup_C \widetilde{y}$ in terms of c, we get the result. □

Proposition 1.1. *A function $s : X \to [-\infty, \infty)$ is upper semicontinuous and concave if and only if*

$$s(x) = \inf\{c + \langle y, x \rangle : c \in \mathbb{R}, y \in X^*, c + \langle y, z \rangle \geq s(z)\ \forall z \in X\}.$$

1.4 The Fenchel-Legendre transform

How can we give a canonical family in here? For fixed $y \in X^*$, what is the best c to use? We want $c + \langle y, z \rangle \geq s(z)$ for all z. That is, we want $c \geq s(z) - \langle y, z \rangle$, so we want to take

$$c = \sup_{z \in X} s(z) - \langle y, z \rangle =: s^*(y).$$

This is known as the **Fenchel-Legendre** transform of s. Here are some properties of s^*:

Proposition 1.2.

1. *s^* is lower semicontinuous and convex.*

   Proof. s^* is the supremum of lower semicontinuous, convex (affine) functions. □

2. *Provided $s \not\equiv -\infty$, we get $s^* : X^* \to (-\infty, \infty]$ and s^* is not always $+\infty$.*

3. *$s(z) = \inf_y s^*(y) - \langle y, z \rangle$. That is, $s = (s^*)_*$.*