Chapter 34

Large Deviations for Weakly Dependent Sequences: The Gärtner-Ellis Theorem

This chapter proves the Gärtner-Ellis theorem, establishing an LDP for not-too-dependent processes taking values in topological vector spaces. Most of our earlier LDP results can be seen as consequences of this theorem.

34.1 The Gärtner-Ellis Theorem

The Gärtner-Ellis theorem is a powerful result which establishes the existence of a large deviation principle for processes where the cumulant generating function tends towards a well-behaved limit, implying not-too-strong dependence between successive values. (Exercise 34.5 clarifies the meaning of "too strong".) It will imply our LDPs for IID and Markovian sequences. I could have started with it, but its proof, as you'll see, is pretty technical, and so it seemed better to use the more elementary arguments of the preceding chapters.

To fix notation, Ξ will be a real topological vector space, and Ξ^* will be its dual space, of continuous linear functions $\Xi \mapsto \mathbb{R}$. (If $\Xi = \mathbb{R}^d$, we can identify Ξ and Ξ^* by means of the inner product. In differential geometry, on the other hand, Ξ might be a space of tangent vectors, and Ξ^* the corresponding one-forms.) X_ϵ will be a family of Ξ-valued random variables, parameterized by $\epsilon > 0$. Refer to Definitions 423 and 424 in Section 31.1 for the definition of the cumulant generating function and its Legendre transform (respectively), which I will denote by $\Lambda_\epsilon : \Xi^* \mapsto \mathbb{R}$ and $\Lambda^*_\epsilon : \Xi \mapsto \overline{\mathbb{R}}$.

The proof of the Gärtner-Ellis theorem goes through a number of lemmas.

Basically, the upper large deviation bound holds under substantially weaker conditions than the lower bound does, and it's worthwhile having the partial results available to use in estimates even if the full large deviations principle does not apply.

Definition 444 *The* upper-limiting cumulant generating function *is*

$$\overline{\Lambda}(t) \equiv \limsup_{\epsilon \rightarrow 0} \epsilon \Lambda_{\epsilon}(t/\epsilon) \tag{34.1}$$

and its Legendre transform is written $\overline{\Lambda}^{*}(x)$.

The point of this is that the limsup always exists, whereas the limit doesn't, necessarily. But we can show that the limsup has some reasonable properties, and in fact it's enough to give us an upper bound.

Lemma 445 $\overline{\Lambda}(t)$ *is convex, and* $\overline{\Lambda}^{*}(x)$ *is a convex rate function.*

PROOF: The proof of the convexity of $\Lambda_{\epsilon}(t)$ follows the proof in Lemma 425, and the convexity of $\overline{\Lambda}(t)$ by passing to the limit. To etablish $\overline{\Lambda}^{*}(x)$ as a rate function, we need it to be non-negative and lower semi-continuous. Since $\Lambda_{\epsilon}(0) = 0$ for all ϵ, $\overline{\Lambda}(0) = 0$. This in turn implies that $\overline{\Lambda}^{*}(x) \geq 0$. Since the latter is the supremum of a class of continuous functions, namely $t(x) - \overline{\Lambda}(t)$, it must be lower semi-continuous. Finally, its convexity is implied by its being a Legendre transform. □

Lemma 446 (Upper Bound in Gärtner-Ellis Theorem: Compact Sets) *For any compact set* $K \subset \Xi$,

$$\limsup_{\epsilon \rightarrow 0} \epsilon \log \mathbb{P}(X_{\epsilon} \in K) \leq -\overline{\Lambda}^{*}(K) \tag{34.2}$$

PROOF: Entirely parallel to the proof of the upper bound in Cramér's Theorem (429), up through the point where closed sets are divided into a compact part and a remainder far from the origin, of exponentially-small probability. Because K is compact, we can proceed as though the remainder is empty. □

Lemma 447 (Upper Bound in Gärtner-Ellis Theorem: Closed Sets) *If the family of distributions* $\mathcal{L}(X_{\epsilon})$ *are exponentially tight, then for all closed* $C \subset \Xi$,

$$\limsup_{\epsilon \rightarrow 0} \epsilon \log \mathbb{P}(X_{\epsilon} \in C) \leq -\overline{\Lambda}^{*}(C) \tag{34.3}$$

PROOF: Exponential tightness, by definition (417), will let us repeat Theorem 429 trick of dividing closed sets into a compact part, and an exponentially-vanishing non-compact part. □

Definition 448 *The* limiting cumulant generating function *is*

$$\Lambda(t) \equiv \lim_{\epsilon \rightarrow 0} \epsilon \Lambda_{\epsilon}(t/\epsilon) \tag{34.4}$$

when the limit exists. Its domain of finiteness $D \equiv \{t \in \Xi^* : \ \Lambda(t) < infty\}$*. Its* limiting Legendre transform *is* Λ^**, with domain of finiteness* D^**.*

Lemma 449 *If* $\Lambda(t)$ *exists, then it is convex,* $\Lambda^*(x)$ *is a convex rate function, and Eq. 34.2 applies to the latter. If in addition the process is exponentially tight, then Eq. 34.3 holds for* $\Lambda^*(x)$*.*

Proof: Because, if a limit exists, it is equal to the limsup. □

Lemma 450 *If* $\Xi = \mathbb{R}^d$*, then* $0 \in \text{int} D$ *is sufficient for exponential tightness.*

Proof: Exercise. □

Unfortunately, this is *not* good enough to get exponential tightness in arbitrary vector spaces.

Definition 451 (Exposed Point) *A point* $x \in \Xi$ *is* exposed for $\overline{\Lambda}^*(\cdot)$ *when there is a* $t \in \Xi^*$ *such that* $\overline{\Lambda}^*(y) - \overline{\Lambda}^*(x) > t(y-x)$ *for all* $y \neq x$*.* t *is the* exposing hyper-plane *for* x*.*

In $\mathbb{R}^1$, a point x is exposed if the curve $\overline{\Lambda}^*(y)$ lies strictly above the line of slope t through the point $(x, \overline{\Lambda}^*(x))$. Similarly, in $\mathbb{R}^d$, the $\overline{\Lambda}^*(y)$ surface must lie strictly above the hyper-plane passing through $(x, \overline{\Lambda}^*(x))$ with surface normal t. Since $\overline{\Lambda}^*(y)$ is convex, we could generally arrange this by making this the tangent hyper-plane, but we do not, yet, have any reason to think that the tangent is well-defined. — Obviously, if $\Lambda(t)$ exists, we can replace $\overline{\Lambda}^*(\cdot)$ by $\Lambda^*(\cdot)$ in the definition and the rest of this paragraph.

Definition 452 *An exposed point* $x \in \Xi$ *with exposing hyper-plane* t *is* nice*,* $x \in N$*, if*

$$\lim_{\epsilon \rightarrow 0} \epsilon \Lambda_{\epsilon}(t/\epsilon) \tag{34.5}$$

exists, and, for some $r > 1$*,*

$$\overline{\Lambda}(rt) < \infty \tag{34.6}$$

Note: Most books on large deviations do not give this property any particular name.

Lemma 453 (Lower Bound in Gärtner-Ellis Theorem) *If the* X_{ϵ} *are exponentially tight, then, for any open set* $O \subseteq \Xi$*,*

$$\liminf_{\epsilon \rightarrow 0} \epsilon \log \mathbb{P}\left(X_{\epsilon} \in O\right) \geq -\overline{\Lambda}^*(O \cap N) \tag{34.7}$$

PROOF: If we pick any nice, exposed point $x \in O$, we can repeat the proof of the lower bound from Cramér's Theorem (429). In fact, the point of Definition 452 is to ensure this. Taking the supremum over all such x gives the lemma. □

Theorem 454 (Abstract Gärtner-Ellis Theorem) *If the X_ϵ are exponentially tight, and $\overline{\Lambda}^*(O \cap E) = \overline{\Lambda}^*(O)$ for all open sets $O \subseteq \Xi$, then X_ϵ obey an LDP with rate $1/\epsilon$ and good rate function $\overline{\Lambda}^*(x)$.*

PROOF: The large deviations upper bound is Lemma 447. The large deviations lower bound is implied by Lemma 453 and the additional hypothesis of the theorem. □

Matters can be made a bit more concrete in Euclidean space (the original home of the theorem), using, however, one or two more bits of convex analysis.

Definition 455 (Relative Interior) *The* relative interior *of a non-empty and convex set A is*

$$\text{rint}A \equiv \{x \in A: \ \forall y \in A, \exists \delta > 0, \ x - \delta(y - x) \in A\} \tag{34.8}$$

Notice that $\text{int}A \subseteq \text{rint}A$, since the latter, in some sense, doesn't care about points outside of A.

Definition 456 (Essentially Smooth) *Λ is* essentially smooth *if (i) $\text{int}D \neq \emptyset$, (ii) Λ is differentiable on $\text{int}D$, and (iii) Λ is "steep", meaning that if D has a boundary, ∂D, then $\lim_{t \to \partial D} \|\nabla\Lambda(t)\| = \infty$.*

This definition was introduced to exploit the following theorem of convex analysis.

Proposition 457 *If Λ is essentially smooth, then Λ^* is continuous on $\text{rint}D^*$, and $\text{rint}D^* \subseteq N$, the set of exposed, nice points.*

PROOF: D^* is non-empty, because there is at least one point x_0 where $\Lambda^*(x_0) = 0$. Moreover, it is a convex set, because Λ^* is a convex function (Lemma 449), so it has a relative interior. Now appeal to Rockafellar (1970, Corollary 26.4.1). □

Remark: You might want to try to prove that Λ^* is continuous on $\text{rint}D^*$.

Theorem 458 (Euclidean Gärtner-Ellis Theorem) *Suppose X_ϵ, taking values in $\mathbb{R}^d$, are exponentially tight, and that $\Lambda(t)$ exists and is essentially smooth. Then X_ϵ obey an LDP with rate $1/\epsilon$ and good rate function $\Lambda^*(x)$.*

PROOF: From the abstract Gärtner-Ellis Theorem 454, it is enough to show that $\Lambda^*(O \cap N) = \Lambda^*(O)$, for any open set O. That is, we want

$$\inf_{x \in O \cap N} \Lambda^*(x) = \inf_{x \in O} \Lambda^*(x) \tag{34.9}$$

Since it's automatic that

$$\inf_{x \in O \cap N} \Lambda^*(x) \geq \inf_{x \in O} \Lambda^*(x) \tag{34.10}$$

what we need to show is that

$$\inf_{x \in O \cap N} \Lambda^*(x) \leq \inf_{x \in O} \Lambda^*(x) \tag{34.11}$$

In view of Proposition 457, it's really just enough to show that $\Lambda^*(O \cap \mathrm{rint} D^*) \leq \Lambda^*(O)$. This is trivial when the intersection $O \cap D^*$ is empty, so assume it isn't, and pick any x in that intersection. Because O is open, and because of the definition of the relative interior, we can pick any point $y \in \mathrm{rint} D^*$, and, for sufficiently small δ, $\delta y + (1-\delta)x \in O \cap \mathrm{rint} D^*$. Since Λ^* is convex, $\Lambda^*(\delta y + (1-\delta)x) \leq \delta\Lambda^*(y) + (1-\delta)\Lambda^*(x)$. Taking the limit as $\delta \to 0$,

$$\Lambda^*(O \cap \mathrm{rint} D^*) \leq \Lambda^*(x) \tag{34.12}$$

and the claim follows by taking the infimum over x.

34.2 Exercises

Exercise 34.1 *Show that Cramér's Theorem (429), is a special case of the Euclidean Gärtner-Ellis Theorem (458).*

Exercise 34.2 *Let $Z_1, Z_2, \ldots$ be IID random variables in a discrete space, and let $X_1, X_2, \ldots$ be the empirical distributions they generate. Use the Gärtner-Ellis Theorem to re-prove Sanov's Theorem (434). Can you extend this to the case where the Z_i take values in an arbitrary Polish space?*

Exercise 34.3 *Let $Z_1, Z_2, \ldots$ be values from a stationary ergodic Markov chain (i.e. the state space is discrete). Repeat the previous exercise for the pair measure. Again, can the result be extended to arbitrary Polish state spaces?*

Exercise 34.4 *Let Z_i be real-valued mean-zero stationary Gaussian variables, with $\sum_{i=-\infty}^{\infty} |\mathrm{cov}(Z_0, Z_i)| < \infty$. Let $X_t = t^{-1}\sum_{i=1}^{t} Z_i$. Show that these time averages obey an LDP with rate t and rate function $x^2/2\Gamma$, where $\Gamma = \sum_{i=-\infty}^{\infty} \mathrm{cov}(Z_0, Z_i)$. (Cf. the mean-square ergodic theorem 246 of chapter 21.) If Z_i are not Gaussian but are weakly stationary, find an additional hypothesis such that the time averages still obey an LDP.*

Exercise 34.5 (Too-Strong Dependence) *Let $X_i = X$, a random variable in $\mathbb{R}$, for all i. Show that the Gärtner-Ellis Theorem fails, unless X is degenerate.*