Chapter 30

General Theory of Large Deviations

A family of random variables follows the *large deviations principle* if the probability of the variables falling into "bad" sets, representing large deviations from expectations, declines exponentially in some appropriate limit. Section 30.1 makes this precise, using some associated technical machinery, and explores a few consequences. The central one is Varadhan's Lemma, for the asymptotic evaluation of exponential integrals in infinite-dimensional spaces.

Having found one family of random variables which satisfy the large deviations principle, many other, related families do too. Section 30.2 lays out some ways in which this can happen.

As the great forensic statistician C. Chan once remarked, "Improbable events permit themselves the luxury of occurring" (reported in Biggers, 1928). Large deviations theory, as I have said, studies these little luxuries.

30.1 Large Deviation Principles: Main Definitions and Generalities

Some technicalities:

Definition 397 (Level Sets) *For any real-valued function* $f : \Xi \mapsto \mathbb{R}$*, the* level sets *are the inverse images of intervals from* $-\infty$ *to* c *inclusive, i.e., all sets of the form* $\{x \in \Xi : f(x) \leq c\}$.

Definition 398 (Lower Semi-Continuity) *A real-valued function* $f : \Xi \mapsto \mathbb{R}$ *is* lower semi-continuous *if* $x_n \to x$ *implies* $\liminf f(x_n) \geq f(x)$.

Lemma 399 *A function is lower semi-continuous iff either of the following equivalent properties hold.*

i For all $x \in \Xi$, the infimum of f over increasingly small open balls centered at x approaches $f(x)$:

$$\lim_{\delta \to 0} \inf_{y:\ d(y,x)<\delta} f(y) = f(x) \tag{30.1}$$

ii f has closed level sets.

Proof: A character-building exercise in real analysis, left to the reader. □

Lemma 400 *A lower semi-continuous function attains its minimum on every non-empty compact set, i.e., if C is compact and $\neq \emptyset$, there is an $x \in C$ such that $f(x) = \inf_{y \in C} f(y)$.*

Proof: Another character-building exercise in real analysis. □

Definition 401 (Logarithmic Equivalence) *Two sequences of positive real numbers a_n and b_n are* logarithmically equivalent, *$a_n \simeq b_n$, when*

$$\lim_{n\to\infty} \frac{1}{n} (\log a_n - \log b_n) = 0 \tag{30.2}$$

Similarly, for continuous parameterizations by $\epsilon > 0$, $a_\epsilon \simeq b_\epsilon$ when

$$\lim_{\epsilon \to 0} \epsilon (\log a_\epsilon - \log b_\epsilon) = 0 \tag{30.3}$$

Lemma 402 ("Fastest rate wins") *For any two sequences of positive numbers, $(a_n + b_n) \simeq a_n \vee b_n$.*

Proof: A character-building exercise in elementary analysis. □

Definition 403 (Large Deviation Principle) *A parameterized family of random variables, X_ϵ, $\epsilon > 0$, taking values in a metric space Ξ with Borel σ-field $\mathcal{X}$,* obeys a large deviation principle with rate $1/\epsilon$, *or just* obeys an LDP, *when, for any set $B \in \mathcal{X}$,*

$$-\inf_{x \in \mathrm{int} B} J(x) \leq \liminf_{\epsilon \to 0} \epsilon \log \mathbb{P}(X_\epsilon \in B) \leq \limsup_{\epsilon \to 0} \epsilon \log \mathbb{P}(X_\epsilon \in B) \leq -\inf_{x \in \mathrm{cl} B} J(x) \tag{30.4}$$

for some non-negative function $J : \Xi \mapsto [0, \infty]$, its raw rate function. *If J is lower semi-continuous, it is just a* rate function. *If J is lower semi-continuous and has compact level sets, it is a* good rate function.[1] *By a slight abuse of notation, we will write $J(B) = \inf_{x \in B} J(x)$.*

[1] Sometimes what Kallenberg and I are calling a "good rate function" is just "a rate function", and our "rate function" gets demoted to "weak rate function".

Remark: The most common choices of ϵ are $1/n$, in sample-size or discrete sequence problems, or ε^2, in small-noise problems (as in Chapter 22).

Lemma 404 (Uniqueness of Rate Functions) *If X_ϵ obeys the LDP with raw rate function J, then it obeys the LDP with a unique rate function J'.*

PROOF: First, show that a raw rate function can always be replaced by a lower semi-continuous function, i.e. a non-raw (cooked?) rate function. Then, show that non-raw rate functions are unique.

For any raw rate function J, define $J'(x) = \liminf_{y \rightarrow x} J(x)$. This is clearly lower semi-continuous, and $J'(x) \leq J(x)$. However, for any open set B, $\inf_{x \in B} J'(x) = \inf_{x \in B} J(x)$, so J and J' are equivalent for purposes of the LDP.

Now assume that J is a lower semi-continuous rate function, and suppose that $K \neq J$ was too; without loss of generality, assume that $J(x) > K(x)$ at some point x. We can use semi-continuity to find an open neighborhood B of x such that $J(\mathrm{cl}B) > K(x)$. But, substituting into Eq. 30.4, we obtain a contradiction:

$$
\begin{aligned}
-K(x) &\leq -K(B) && (30.5)\\
&\leq \liminf_{\epsilon \rightarrow 0} \epsilon \log \mathbb{P}(X_\epsilon \in B) && (30.6)\\
&\leq -J(\mathrm{cl}B) && (30.7)\\
&\leq -K(x) && (30.8)
\end{aligned}
$$

Hence there can be no such rate function K, and J is the unique rate function. □

Lemma 405 *If X_ϵ obeys an LDP with rate function J, then $J(x) = 0$ for some x.*

PROOF: Because $\mathbb{P}(X_\epsilon \in \Xi) = 1$, we must have $J(\Xi) = 0$, and rate functions attain their infima. □

Definition 406 *A Borel set B is J-continuous, for some rate function J, when $J(\mathrm{int}B) = J(\mathrm{cl}B)$.*

Lemma 407 *If X_ϵ satisfies the LDP with rate function J, then for every J-continuous set B,*

$$
\lim_{\epsilon \rightarrow 0} \epsilon \log \mathbb{P}(X_\epsilon \in B) = -J(B) \quad (30.9)
$$

PROOF: By J-continuity, the right and left hand extremes of Eq. 30.4 are equal, so the limsup and the liminf sandwiched between them are equal; consequently the limit exists. □

Remark: The obvious implication is that, for small ϵ, $\mathbb{P}(X_\epsilon \in B) \approx ce^{-J(B)/\epsilon}$, which explains why we say that the LDP has rate $1/\epsilon$. (Actually, c need not be constant, but it must be at least $o(\epsilon)$, i.e., it must go to zero faster than ϵ itself does.)

There are several equivalent ways of defining the large deviation principle. The following is especially important, because it is often simplifies proofs.

Lemma 408 *X_ϵ obeys the LDP with rate $1/\epsilon$ and rate function $J(x)$ if and only if*

$$\limsup_{\epsilon\rightarrow 0} \epsilon \log \mathbb{P}\left(X_\epsilon \in C\right) \leq -J(C) \quad (30.10)$$

$$\liminf_{\epsilon\rightarrow 0} \epsilon \log \mathbb{P}\left(X_\epsilon \in O\right) \geq -J(O) \quad (30.11)$$

for every closed Borel set C and every open Borel set $O \subset \Xi$.

PROOF: "If": The closure of any set is closed, and the interior of any set is open, so Eqs. 30.10 and 30.11 imply

$$\limsup_{\epsilon\rightarrow 0} \epsilon \log \mathbb{P}\left(X_\epsilon \in \text{cl}B\right) \leq -J(\text{cl}B) \quad (30.12)$$

$$\liminf_{\epsilon\rightarrow 0} \epsilon \log \mathbb{P}\left(X_\epsilon \in \text{int}B\right) \geq -J(\text{int}B) \quad (30.13)$$

but $\mathbb{P}\left(X_\epsilon \in B\right) \leq \mathbb{P}\left(X_\epsilon \in \text{cl}B\right)$ and $\mathbb{P}\left(X_\epsilon \in B\right) \geq \mathbb{P}\left(X_\epsilon \in \text{int}B\right)$, so the LDP holds. "Only if": every closed set is equal to its own closure, and every open set is equal to its own interior, so the upper bound in Eq. 30.4 implies Eq. 30.10, and the lower bound Eq. 30.11. □

A deeply important consequence of the LDP is the following, which can be thought of as a version of Laplace's method for infinite-dimensional spaces.

Theorem 409 (Varadhan's Lemma) *If X_ϵ are random variables in a metric space Ξ, obeying an LDP with rate $1/\epsilon$ and rate function J, and $f : \Xi \mapsto \mathbb{R}$ is continuous and bounded from above, then*

$$\Lambda_f \equiv \lim_{\epsilon\rightarrow 0} \epsilon \log \mathbf{E}\left[e^{f(X_\epsilon)/\epsilon}\right] = \sup_{x\in\Xi} f(x) - J(x) \quad (30.14)$$

PROOF: We'll find the limsup and the liminf, and show that they are both $\sup f(x) - J(x)$.

First the limsup. Pick an arbitrary positive integer n. Because f is continuous and bounded above, there exist finitely closed sets, call them $B_1, \ldots B_m$, such that $f \leq -n$ on the complement of $\bigcup_i B_i$, and within each B_i, f varies by at most $1/n$. Now

$$\limsup \epsilon \log \mathbf{E}\left[e^{f(X_\epsilon)/\epsilon}\right] \quad (30.15)$$

$$\leq (-n) \vee \max_{i\leq m} \limsup \epsilon \log \mathbf{E}\left[e^{f(X_\epsilon)/\epsilon}\mathbf{1}_{B_i}(X_\epsilon)\right]$$

$$\leq (-n) \vee \max_{i\leq m} \sup_{x\in B_i} f(x) - \inf_{x\in B_i} J(x) \quad (30.16)$$

$$\leq (-n) \vee \max_{i\leq m} \sup_{x\in B_i} f(x) - J(x) + 1/n \quad (30.17)$$

$$\leq (-n) \vee \sup_{x\in\Xi} f(x) - J(x) + 1/n \quad (30.18)$$

Letting $n \to \infty$, we get $\limsup \epsilon \log \mathbf{E}\left[e^{f(X_\epsilon)/\epsilon}\right] = \sup f(x) - J(x)$.

To get the liminf, pick any $x \in Xi$ and an arbitrary ball of radius δ around it, $B_{\delta,x}$. We have

$$\liminf \epsilon \log \mathbf{E}\left[e^{f(X_\epsilon)/\epsilon}\right] \geq \liminf \epsilon \log \mathbf{E}\left[e^{f(X_\epsilon)/\epsilon}\mathbf{1}_{B_{\delta,x}}(X_\epsilon)\right] \quad (30.19)$$

$$\geq \inf_{y \in B_{\delta,x}} f(y) - \inf_{y \in B_{\delta,x}} J(y) \quad (30.20)$$

$$\geq \inf_{y \in B_{\delta,x}} f(y) - J(x) \quad (30.21)$$

Since δ was arbitrary, we can let it go to zero, so (by continuity of f) $\inf_{y \in B_{\delta,x}} f(y) \to f(x)$, or

$$\liminf \epsilon \log \mathbf{E}\left[e^{f(X_\epsilon)/\epsilon}\right] \geq f(x) - J(x) \quad (30.22)$$

Since this holds for arbitrary x, we can replace the right-hand side by a supremum over all x. Hence $\sup f(x) - J(x)$ is both the liminf and the limsup. □

Remark: The implication of Varadhan's lemma is that, for small ϵ, $\mathbf{E}\left[e^{f(X_\epsilon)/\epsilon}\right] \approx c(\epsilon)e^{\epsilon^{-1}(\sup_{x\in\Xi} f(x)-J(x))}$, where $c(\epsilon) = o(\epsilon)$. So, we can replace the exponential integral with its value at the extremal points, at least to within a multiplicative factor and to first order in the exponent.

An important, if heuristic, consequence of the LDP is that "Highly improbable events tend to happen in the least improbable way". Let us consider two events $B \subset A$, and suppose that $\mathbb{P}(X_\epsilon \in A) > 0$ for all ϵ. Then $\mathbb{P}(X_\epsilon \in B | X_\epsilon \in A) = \mathbb{P}(X_\epsilon \in B)/\mathbb{P}(X_\epsilon \in A)$. Roughly speaking, then, this conditional probability will vanish exponentially, with rate $J(A) - J(B)$. That is, even if we are looking at an exponentially-unlikely large deviation, the vast majority of the probability is concentrated around the *least unlikely* part of the event. More formal statements of this idea are sometimes known as "conditional limit theorems" or "the Gibbs conditioning principle".

30.2 Breeding Large Deviations

Often, the easiest way to prove that one family of random variables obeys a large deviations principle is to prove that another, related family does.

Theorem 410 (Contraction Principle) *If X_ϵ, taking values in a metric space Ξ, obeys an LDP, with rate ϵ and rate function J, and $f : \Xi \mapsto \Upsilon$ is a continuous function from that metric space to another, then $Y_\epsilon = f(X_\epsilon)$ also obeys an LDP, with rate ϵ and raw rate function $K(y) = J(f^{-1}(y))$. If J is a good rate function, then so is K.*

PROOF: Since f is continuous, f^{-1} takes open sets to open sets, and closed sets to closed sets. Pick any closed $C \subset \Upsilon$. Then

$$\begin{aligned}
& \limsup_{\epsilon \rightarrow 0} \epsilon \log \mathbb{P}\left(f(X_\epsilon) \in C\right) && (30.23)\\
&= \limsup_{\epsilon \rightarrow 0} \epsilon \log \mathbb{P}\left(X_\epsilon \in f^{-1}(C)\right) \\
&\leq -J(f^{-1}(C)) && (30.24)\\
&= -\inf_{x \in f^{-1}(C)} J(x) && (30.25)\\
&= -\inf_{y \in C} \inf_{x \in f^{-1}(y)} J(x) && (30.26)\\
&= -\inf_{y \in C} K(y) && (30.27)
\end{aligned}$$

as required. The argument for open sets in Υ is entirely parallel, establishing that K, as defined, is a raw rate function. By Lemma 404, K can be modified to be lower semi-continuous without affecting the LDP, i.e., we can make a rate function from it. If J is a good rate function, then it has compact level sets. But continuous functions take compact sets to compact sets, so $K = J \circ f^{-1}$ will also have compact level sets, i.e., it will also be a good rate function. □

There are a bunch of really common applications of the contraction principle, relating the large deviations at one level of description to those at coarser levels. To make the most frequent set of implications precise, let's recall a couple of definitions.

Definition 411 (Empirical Mean) *If $X_1, \ldots X_n$ are random variables in a common vector space Ξ, their* empirical mean *is $\overline{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$.*

We have already encountered this as the sample average or, in ergodic theory, the finite time average. (Notice that nothing is said about the X_i being IID, or even having a common expectation.)

Definition 412 (Empirical Distribution) *Let $X_1, \ldots X_n$ be random variables in a common measurable space Ξ (not necessarily a vector or metric space). The* empirical distribution *is $\hat{P}_n \equiv \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$, where δ_x is the probability measure that puts all its probability on the point x, i.e., $\delta_x(B) = \mathbf{1}_B(x)$. $\hat{P}_n$ is a random variable taking values in $\mathcal{P}(\Xi)$, the space of all probability measures on Ξ. (Cf. Example 10 in chapter 1 and Example 43 in chapter 4.) $\mathcal{P}(\Xi)$ is a metric space under any of several distances, and a complete separable metric space (i.e., Polish) under, for instance, the total variation metric.*

Definition 413 (Finite-Dimensional Empirical Distributions) *For each k, the* k-dimensional empirical distribution *is*

$$\hat{P}_n^k \equiv \frac{1}{n}\sum_{i=1}^{n} \delta_{(X_i, X_{i+1}, \ldots X_{i+k})} \tag{30.28}$$

where the addition of indices for the delta function is to be done modulo n, i.e., $\hat{P}_3^2 = \frac{1}{3}\left(\delta_{(X_1,X_2)} + \delta_{(X_2,X_3)} + \delta_{(X_3,X_1)}\right)$. $\hat{P}_n^k$ takes values in $\mathcal{P}\left(\Xi^k\right)$.

Definition 414 (Empirical Process Distribution) *With a finite sequence of random variables* $X_1, \ldots X_n$*, the* empirical process *is the periodic, infinite random sequence* $\tilde{X}_n$ *as the repetition of the sample without limit, i.e.,* $\tilde{X}_n(i) = X_{i \bmod n}$*. If* T *is the shift operator on the sequence space, then the* empirical process distribution *is*

$$\hat{P}_n^{\infty} \equiv \frac{1}{n} \sum_{i-0}^{n-1} \delta_{T^i \tilde{X}_n} \tag{30.29}$$

$\hat{P}_n^{\infty}$ *takes values in the space of infinite-dimensional distributions for one-sided sequences,* $\mathcal{P}\left(\Xi^{\mathbb{N}}\right)$*. In fact, it is always a stationary distribution, because by construction it is invariant under the shift* T*.*

Be careful not to confuse this empirical process with the quite distinct empirical process of Examples 10 and 43.

Corollary 415 *The following chain of implications hold:*

i If the empirical process distribution obeys an LDP, so do all the finite-dimensional distributions.

ii If the n*-dimensional distribution obeys an LDP, all* $m < n$ *dimensional distributions do.*

iii If any finite-dimensional distribution obeys an LDP, the empirical distribution does.

iv If the empirical distribution obeys an LDP, the empirical mean does.

Proof: In each case, we obtain the lower-level statistic from the higher-level one by applying a continuous function, hence the contraction principle applies. For the distributions, the continuous functions are the projection operators of Chapter 2. □

Corollary 416 ("Tilted" LDP) *In set-up of Theorem 409, let* $\mu_\epsilon = \mathcal{L}(X_\epsilon)$*. Define the probability measures* $\mu_{f,\epsilon}$ *via*

$$\mu_{f,\epsilon}(B) \equiv \frac{\mathbf{E}\left[e^{f(X_\epsilon)/\epsilon}\mathbf{1}_B(X_\epsilon)\right]}{\mathbf{E}\left[e^{f(X_\epsilon)/\epsilon}\right]} \tag{30.30}$$

Then $Y_\epsilon \sim \mu_{f,\epsilon}$ *obeys an LDP with rate* $1/\epsilon$ *and rate function*

$$J_F(x) = -(f(x) - J(x)) + \sup_{y \in \Xi} f(y) - J(y) \tag{30.31}$$

PROOF: Define a set function $F_\epsilon(B) = \mathbf{E}\left[e^{f(X_\epsilon)/\epsilon}\mathbf{1}_B(X_\epsilon)\right]$; then $\mu_{f,\epsilon}(B) = F_\epsilon(B)/F_\epsilon(\Xi)$. From Varadhan's Lemma, we know that $F_\epsilon(\Xi)$ has asymptotic logarithm $\sup_{y\in\Xi} f(y) - J(y)$, so it is just necessary to show that

$$\limsup_{\epsilon} \epsilon \log F_\epsilon(B) \leq \sup_{x\in \mathrm{cl}B} f(x) - J(x) \quad (30.32)$$

$$\liminf_{\epsilon} \epsilon \log F_\epsilon(B) \geq \sup_{x\in \mathrm{int}B} f(x) - J(x) \quad (30.33)$$

which can be done by imitating the proof of Varadhan's Lemma itself. □

Remark: "Tilting" here refers to some geometrical analogy which, in all honesty, has never made any sense to me.

Because the LDP is about exponential decay of probabilities, it is not surprising that several ways of obtaining it require a sort of exponential bound on the dispersion of the probability measure.

Definition 417 (Exponentially Tight) *The parameterized family of random variables X_ϵ, $\epsilon > 0$, is* exponentially tight *if, for every finite real M, there exists a compact set $C \subset \Xi$ such that*

$$\limsup_{\epsilon\rightarrow 0} \epsilon \log \mathbb{P}\left(X_\epsilon \not\in C\right) \leq -M \quad (30.34)$$

The first use of exponential tightness is a converse to the contraction principle: a high-level LDP is implied by the combination of a low-level LDP and high-level exponential tightness.

Theorem 418 (Inverse Contraction Principle) *If X_ϵ are exponentially tight, f is continuous and injective, and $Y_\epsilon = f(X_\epsilon)$ obeys an LDP with rate function K, then X_ϵ obeys an LDP with a good rate function $J(x) = K(f(x))$.*

PROOF: See Kallenberg, Theorem 27.11 (ii). Notice, by the way, that the proof of the upper bound on probabilities (i.e. that $\limsup \epsilon \log \mathbb{P}\left(X_\epsilon \in B\right) \leq -J(B)$ for closed $B \subseteq \Xi$) does not depend on exponential tightness, just the continuity of f. Exponential tightness is only needed for the lower bound. □

Theorem 419 (Bryc's Theorem) *If X_ϵ are exponentially tight, and, for all bounded continuous f, the limit*

$$\Lambda_f \equiv \lim_{\epsilon\rightarrow 0} \epsilon \log \mathbf{E}\left[e^{f(X_\epsilon/\epsilon)}\right] \quad (30.35)$$

exists, then X_ϵ obeys the LDP with good rate function

$$J(x) \equiv \sup_f f(x) - \Lambda_f \quad (30.36)$$

where the supremum extends over all bounded, continuous functions.

PROOF: See Kallenberg, Theorem 27.10, part (ii). □

Remark: This is a converse to Varadhan's Lemma.

Theorem 420 (Projective Limit) *Let $\Xi_1, \Xi_2, \ldots$ be a countable sequence of metric spaces, and let X_ϵ be a random sequence from this space. If, for every n, $X^n_\epsilon = \pi_n X_\epsilon$ obeys the LDP with good rate function J_n, then X_ϵ obeys the LDP with good rate function*

$$J(x) \equiv \sup_n J_n(\pi_n x) \tag{30.37}$$

PROOF: See Kallenberg, Theorem 27.12. □

Definition 421 (Exponentially Equivalent Random Variables) *Two families of random variables, X_ϵ and Y_ϵ, taking values in a common metric space, are* exponentially equivalent *when, for all positive δ,*

$$\lim_{\epsilon \rightarrow 0} \epsilon \log \mathbb{P}\left(d(X_\epsilon, Y_\epsilon) > \delta\right) = -\infty \tag{30.38}$$

Lemma 422 *If X_ϵ and Y_ϵ are exponentially equivalent, one of them obeys the LDP with a good rate function J iff the other does as well.*

PROOF: It is enough to prove that the LDP for X_ϵ implies the LDP for Y_ϵ, with the same rate function. (Draw a truth-table if you don't believe me!) As usual, first we'll get the upper bound, and then the lower.

Pick any closed set C, and let C_δ be its closed δ neighborhood, i.e., $C_\delta = \{x : \exists y \in C,\ d(x,y) \leq \delta\}$. Now

$$\mathbb{P}\left(Y_\epsilon \in C_\delta\right) \leq \mathbb{P}\left(X_\epsilon \in C_\delta\right) + \mathbb{P}\left(d(X_\epsilon, Y_\epsilon) > \delta\right) \tag{30.39}$$

Using Eq. 30.38 from Definition 421, the LDP for X_ϵ, and Lemma 402

$$\begin{aligned}
& \limsup \epsilon \log \mathbb{P}\left(Y_\epsilon \in C\right) && (30.40)\\
\leq\ & \limsup \epsilon \log \mathbb{P}\left(X_\epsilon \in C_\delta\right) + \epsilon \log \mathbb{P}\left(d(X_\epsilon, Y_\epsilon) > \delta\right) &&\\
\leq\ & \limsup \epsilon \log \mathbb{P}\left(X_\epsilon \in C_\delta\right) \vee \limsup \epsilon \log \mathbb{P}\left(d(X_\epsilon, Y_\epsilon) > \delta\right) && (30.41)\\
\leq\ & -J(C_\delta) \vee -\infty && (30.42)\\
=\ & -J(C_\delta) && (30.43)
\end{aligned}$$

Since J is a good rate function, we have $J(C_\delta) \uparrow J(C)$ as $\delta \downarrow 0$; since δ was arbitrary to start with,

$$\limsup \epsilon \log \mathbb{P}\left(Y_\epsilon \in C\right) \leq -J(C) \tag{30.44}$$

As usual, to obtain the lower bound on open sets, pick any open set O and any point $x \in O$. Because O is open, there is a $\delta > 0$ such that, for some open neighborhood U of x, not only is $U \subset O$, but $U_\delta \subset O$. In which case, we can say that

$$\mathbb{P}\left(X_\epsilon \in U\right) \leq \mathbb{P}\left(Y_\epsilon \in O\right) + \mathbb{P}\left(d(X_\epsilon, Y_\epsilon) > h\right) \tag{30.45}$$

Proceeding as for the upper bound,

$$
\begin{aligned}
-J(x) &\leq -J(U) && (30.46)\\
&\leq \liminf \epsilon \log \mathbb{P}(X_\epsilon \in U) && (30.47)\\
&\leq \liminf \epsilon \log \mathbb{P}(Y_\epsilon \in O) \vee \limsup \epsilon \log \mathbb{P}(d(X_\epsilon, Y_\epsilon) > \delta) && (30.48)\\
&= \liminf \epsilon \log \mathbb{P}(Y_\epsilon \in O) && (30.49)
\end{aligned}
$$

(Notice that the initial arbitrary choice of δ has dropped out.) Taking the supremum over all x gives $-J(O) \leq \liminf \epsilon \log \mathbb{P}(Y_\epsilon \in O)$, as required. □