

-stable, then $[X]$ is strictly
e when X has positive or

$]_{\infty} \leq a$ or $\langle M \rangle \leq a$ a.s. for
a suitably scaled process.)

ν , show that $X \in \mathcal{M}_{loc}^2$ iff
case $\langle X \rangle_t = tEX_1^2$. (Hint:

al martingale with positive
Hint: For any such processes
4 to $M - N$.)

continuous or has accessible
Theorem 26.6 (iv).)

$\leq c < \infty$ a.s. is a special
 $M + A$ satisfying $|\Delta A| \leq c$
gale iff it has a decomposi-
Hint: Use Lemma 26.5, and

continuous or has accessible
e M and A have the same
gales, we may choose $M + A$
Use Proposition 25.17 and
rcise.)

able iff it is a special semi-
such that M is continuous

Chapter 27

Large Deviations

*Legendre–Fenchel transform; Cramér’s and Schilder’s theorems;
large-deviation principle and rate function; functional form of
the LDP; continuous mapping and extension; perturbation of
dynamical systems; empirical processes and entropy; Strassen’s
law of the iterated logarithm*

In its simplest setting, large deviation theory provides the exact rate of convergence in the weak law of large numbers. To be precise, consider any i.i.d. random variables $\xi_1, \xi_2, \dots$ with mean m and cumulant-generating function $\Lambda(u) = \log Ee^{u\xi_1} < \infty$, and write $\bar{\xi}_n = n^{-1} \sum_{k \leq n} \xi_k$. Then for any $x > m$, the tail probabilities $P\{\bar{\xi}_n > x\}$ tend to 0 at an exponential rate $I(x)$, given by the Legendre–Fenchel transform Λ^* of Λ . In higher dimensions, it is often convenient to state the result more generally in the form $n^{-1} \log P\{\bar{\xi}_n \in B\} \rightarrow -I(B)$, where $I(B) = \inf_{x \in B} I(x)$ and B is restricted to a suitable class of continuity sets. In this standard format of a *large-deviation principle with rate function I* , the result extends to an amazing variety of contexts throughout probability theory.

A striking example, of fundamental importance in statistical mechanics, is Sanov’s theorem, which provides a similar large deviation result for the empirical distributions of a sequence of i.i.d. random variables with a common distribution μ . Here the rate function I is defined on the space of probability measures ν on $\mathbb{R}$ and agrees with the relative entropy function $H(\nu|\mu)$. Another important example is Schilder’s theorem for the family of rescaled Brownian motions in $\mathbb{R}^d$, where the rate function becomes $I(x) = \frac{1}{2} \|\dot{x}\|_2^2$, the squared norm in the Cameron–Martin space considered in Chapter 18. The latter result can be used to derive the Fredlin–Wentzell estimates for randomly perturbed dynamical systems. It also provides a short proof of Strassen’s law of the iterated logarithm, a stunning extension of the classical Khinchin law from Chapter 13.

Modern proofs of those and other large deviation results rely on some general extension principles, which also serve to explain the wide applicability of the present ideas. In addition to some rather straightforward and elementary techniques of continuity and approximation, we consider the more sophisticated and extremely powerful methods of inverse continuous mapping and projective limits, both of which play a crucial role in subsequent applications. We may also call attention to the significance of

exponential tightness, and to the essential equivalence between the setwise and functional formulations of the large-deviation principle.

Large deviation theory is arguably one of the most technical branches of modern probability theory. For the nonexpert it then seems essential to avoid getting distracted by topological subtleties or elaborate computations. Many results are therefore stated here under simplifying assumptions. Likewise, we postpone our discussion of general principles until the reader has become acquainted with the basic ideas in a concrete setting. For this reason, important applications appear both at the beginning and at the end of the chapter, separated by a more abstract discussion of some general notions and principles.

Let us now return to the elementary context of i.i.d. random variables $\xi, \xi_1, \xi_2, \dots$ and write $S_n = \sum_{k \leq n} \xi_k$ and $\bar{\xi}_n = S_n/n$. If $m = E\xi$ exists and is finite, then $P\{\bar{\xi}_n > x\} \rightarrow 0$ for all $x > m$ by the weak law of large numbers. Under stronger moment conditions, the rate of convergence turns out to be exponential and can be estimated with great accuracy. This rather elementary but quite technical result lies, along with its multidimensional counterpart, at the core of large-deviation theory and provides both a pattern and a point of departure for more advanced developments. For motivation, we begin with some simple observations.

Lemma 27.1 (convergence) *Let $\xi, \xi_1, \xi_2, \dots$ be i.i.d. random variables. Then*

- (i) $n^{-1} \log P\{\bar{\xi}_n \geq x\} \rightarrow \sup_n n^{-1} \log P\{\bar{\xi}_n \geq x\} \equiv -h(x)$ for all x ;
- (ii) h is $[0, \infty]$ -valued, nondecreasing, and convex;
- (iii) $h(x) < \infty$ iff $P\{\xi \geq x\} > 0$.

Proof: (i) Writing $p_n = P\{\bar{\xi}_n \geq x\}$, we get for any $m, n \in \mathbb{N}$

$$\begin{aligned} p_{m+n} &= P\{S_{m+n} \geq (m+n)x\} \\ &\geq P\{S_m \geq mx, S_{m+n} - S_m \geq nx\} = p_m p_n. \end{aligned}$$

Taking logarithms, we conclude that the sequence $-\log p_n$ is subadditive, and the assertion follows by Lemma 10.21.

(ii) The first two assertions are obvious. To prove the convexity, let $x, y \in \mathbb{R}$ be arbitrary, and proceed as before to get

$$P\{S_{2n} \geq n(x+y)\} \geq P\{S_n \geq nx\} P\{S_n \geq ny\}.$$

Taking logarithms, dividing by $2n$, and letting $n \rightarrow \infty$, we obtain

$$h(\tfrac{1}{2}(x+y)) \leq \tfrac{1}{2}(h(x) + h(y)), \quad x, y > 0.$$

(iii) If $P\{\xi \geq x\} = 0$, then $P\{\bar{\xi}_n \geq x\} = 0$ for all n , and so $h(x) = \infty$. Conversely, (i) yields $\log P\{\xi \geq x\} \leq -h(x)$, and so $h(x) = \infty$ implies $P\{\xi \geq x\} = 0$. $\square$

To determine the limit in Lemma 27.1 we need some further notation, which is given here for convenience directly in d dimensions. For any random

vector ξ in $\mathbb{R}^d$, we introduce

$$\Lambda(u) = \Lambda_\xi$$

known in statistics as the *cumulant generating function*, since by Hölder's inequality with $p + q = 1$

$$\begin{aligned} \Lambda(pu + qv) &= \log E e^{(pu+qv) \cdot \xi} \\ &\leq \log(E e^{pu \cdot \xi}) + \log(E e^{qv \cdot \xi}) \end{aligned}$$

The surface $z = \Lambda(u)$ in $\mathbb{R}^d$ is called the *rate function*. Note that the plane with slope u is tangent to the surface at u .

$$z + \Lambda(u)$$

where Λ^* denotes the *Legendre transform* of Λ :

$$\Lambda^*(x) = \sup_{u \in \mathbb{R}^d} \{x \cdot u - \Lambda(u)\}$$

We can often compute Λ^* explicitly. We shall need the results

Lemma 27.2 (Gaussian and Bernoulli) *If $\xi = (\xi_1, \dots, \xi_d)$ is standard Gaussian or Bernoulli, then*

- (i) *If $\xi = (\xi_1, \dots, \xi_d)$ is standard Gaussian, then $\Lambda^*(x) = \frac{1}{2} \|x\|^2$.*
- (ii) *If $\xi \in \{0, 1\}^d$ with $P\{\xi = x\} = p_1 \cdots p_d$ for $x = (x_1, \dots, x_d) \in \{0, 1\}^d$, then $\Lambda^*(x) = \sum_{i=1}^d x_i \log \frac{x_i}{p_i}$.*

$$\Lambda_\xi^*(x) = x \log \frac{x}{p}$$

The function Λ^* is again convex and lower semicontinuous, as before.

$$\begin{aligned} \Lambda^*(px + qy) &= \sup_{u \in \mathbb{R}^d} \{ (px + qy) \cdot u - \Lambda(u) \} \\ &\leq p \sup_{u \in \mathbb{R}^d} \{ xu \cdot u - \Lambda(u) \} + q \sup_{u \in \mathbb{R}^d} \{ yu \cdot u - \Lambda(u) \} \\ &= p \Lambda^*(x) + q \Lambda^*(y). \end{aligned}$$

If $\Lambda < \infty$ near the origin, then Λ^* is finite near 0. Thus, the surface $z = \Lambda(u)$ is convex, and we conclude that Λ^* is also truly d -dimensional, the *rate function*, and continuous near m . For one-sided statements, which are often needed, we assume that m is a vector in $\mathbb{R}^d$.

The following key result is the *Sanov theorem*. For simplicity, we assume that m is a vector in $\mathbb{R}^d$.

ivalence between the setwise
tion principle.

the most technical branches
pert it then seems essential
leties or elaborate computa-
der simplifying assumptions.
d principles until the reader
a concrete setting. For this
t the beginning and at the
t discussion of some general

t of i.i.d. random variables
= S_n/n . If $m = E\xi$ exists
> m by the weak law of
ons, the rate of convergence
d with great accuracy. This
along with its multidimen-
n theory and provides both
dvanced developments. For
tions.

be i.i.d. random variables.

$x\} \equiv -h(x)$ for all x ;
vex;

any $m, n \in \mathbb{N}$

$\geq nx\} = p_m p_n$.

ce $-\log p_n$ is subadditive,

ve the convexity, let $x, y \in$

$P\{S_n \geq ny\}$.

$\rightarrow \infty$, we obtain

$x, y > 0$.

r all n , and so $h(x) = \infty$.
nd so $h(x) = \infty$ implies
□

ed some further notation,
ensions. For any random

vector ξ in $\mathbb{R}^d$, we introduce the function

$$\Lambda(u) = \Lambda_\xi(u) = \log Ee^{u\xi}, \quad u \in \mathbb{R}^d, \quad (1)$$

known in statistics as the *cumulant-generating function* of ξ . Note that Λ is convex, since by Hölder's inequality we have for any $u, v \in \mathbb{R}^d$ and $p, q > 0$ with $p + q = 1$

$$\begin{aligned} \Lambda(pu + qv) &= \log E \exp((pu + qv)\xi) \\ &\leq \log((Ee^{u\xi})^p (Ee^{v\xi})^q) = p\Lambda(u) + q\Lambda(v). \end{aligned}$$

The surface $z = \Lambda(u)$ in $\mathbb{R}^{d+1}$ is determined by the family of supporting hyperplanes (d -dimensional affine subspaces) with different slopes, and we note that the plane with slope $x \in \mathbb{R}^d$ (or normal vector $(1, -x)$) has equation

$$z + \Lambda^*(x) = xu, \quad u \in \mathbb{R}^d,$$

where Λ^* denotes the *Legendre-Fenchel transform* of Λ , given by

$$\Lambda^*(x) = \sup_{u \in \mathbb{R}^d} (ux - \Lambda(u)), \quad x \in \mathbb{R}^d. \quad (2)$$

We can often compute Λ^* explicitly. Here we list two simple cases that will be needed below. The results are proved by elementary calculus.

Lemma 27.2 (*Gaussian and Bernoulli distributions*)

- (i) If $\xi = (\xi_1, \dots, \xi_d)$ is standard Gaussian in $\mathbb{R}^d$, then $\Lambda_\xi^*(x) \equiv |x|^2/2$.
- (ii) If $\xi \in \{0, 1\}$ with $P\{\xi = 1\} = p \in (0, 1)$, then $\Lambda_\xi^*(x) = \infty$ for $x \notin [0, 1]$ and

$$\Lambda_\xi^*(x) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}, \quad x \in [0, 1].$$

The function Λ^* is again convex, since for any $x, y \in \mathbb{R}^d$ and for p and q as before

$$\begin{aligned} \Lambda^*(px + qy) &= \sup_u [p(ux - \Lambda(u)) + q(uy - \Lambda(u))] \\ &\leq p \sup_u (ux - \Lambda(u)) + q \sup_u (uy - \Lambda(u)) \\ &= p\Lambda^*(x) + q\Lambda^*(y). \end{aligned}$$

If $\Lambda < \infty$ near the origin, then $m = E\xi$ exists and agrees with the gradient $\nabla \Lambda(0)$. Thus, the surface $z = \Lambda(u)$ has tangent hyperplane $z = mu$ at 0, and we conclude that $\Lambda^*(m) = 0$ and $\Lambda^*(x) > 0$ for $x \neq m$. If ξ is also truly d -dimensional, then Λ is strictly convex at 0, and Λ^* is finite and continuous near m . For $d = 1$, we sometimes need the corresponding one-sided statements, which are easily derived by dominated convergence.

The following key result identifies the function h in Lemma 27.1. For simplicity, we assume that $m = E\xi$ exists in $[-\infty, \infty)$.

Theorem 27.3 (rate function, Cramér, Chernoff) Let $\xi, \xi_1, \xi_2, \dots$ be i.i.d. random variables with $m = E\xi < \infty$. Then for any $x \geq m$, we have

$$n^{-1} \log P\{\bar{\xi}_n \geq x\} \rightarrow -\Lambda^*(x). \quad (3)$$

Proof: Using Chebyshev's inequality and (1), we get for any $u > 0$

$$P\{\bar{\xi}_n \geq x\} = P\{e^{uS_n} \geq e^{nux}\} \leq e^{-nux} Ee^{uS_n} = e^{n\Lambda(u) - nux},$$

and so

$$n^{-1} \log P\{\bar{\xi}_n \geq x\} \leq \Lambda(u) - ux.$$

This remains true for $u \leq 0$, since in that case $\Lambda(u) - ux \geq 0$ for $x \geq m$. Hence, by (2) we have the upper bound

$$n^{-1} \log P\{\bar{\xi}_n \geq x\} \leq -\Lambda^*(x), \quad x \geq m, \quad n \in \mathbb{N}. \quad (4)$$

To derive a matching lower bound, we first assume that $\Lambda < \infty$ on $\mathbb{R}_+$. Then Λ is smooth on $(0, \infty)$ with $\Lambda'(0+) = m$ and $\Lambda'(\infty) = \text{ess sup } \xi \equiv b$, and so for any $a \in (m, b)$ we can choose a $u > 0$ such that $\Lambda'(u) = a$. Let $\eta, \eta_1, \eta_2, \dots$ be i.i.d. with distribution

$$P\{\eta \in B\} = e^{-\Lambda(u)} E[e^{u\xi}; \xi \in B], \quad B \in \mathcal{B}. \quad (5)$$

Then $\Lambda_\eta(r) = \Lambda_\xi(r + u) - \Lambda_\xi(u)$, and therefore $E\eta = \Lambda'_\eta(0) = \Lambda'_\xi(u) = a$. For any $\varepsilon > 0$, we get by (5)

$$\begin{aligned} P\{|\bar{\xi}_n - a| < \varepsilon\} &= e^{n\Lambda(u)} E[e^{-nu\bar{\eta}_n}; |\bar{\eta}_n - a| < \varepsilon] \\ &\geq e^{n\Lambda(u) - nu(a+\varepsilon)} P\{|\bar{\eta}_n - a| < \varepsilon\}. \end{aligned} \quad (6)$$

Here the last probability tends to 1 by the law of large numbers, and so by (2)

$$\liminf_{n \rightarrow \infty} n^{-1} \log P\{|\bar{\xi}_n - a| < \varepsilon\} \geq \Lambda(u) - u(a + \varepsilon) \geq -\Lambda^*(a + \varepsilon).$$

Fixing any $x \in (m, b)$ and putting $a = x + \varepsilon$, we get for small enough $\varepsilon > 0$

$$\liminf_{n \rightarrow \infty} n^{-1} \log P\{\bar{\xi}_n \geq x\} \geq -\Lambda^*(x + 2\varepsilon).$$

Since Λ^* is continuous on (m, b) by convexity, we may let $\varepsilon \rightarrow 0$ and combine with (4) to obtain (3).

The result for $x > b$ is trivial, since in that case both sides of (3) equal $-\infty$. If instead $x = b < \infty$, then both sides equal $\log P\{\xi = b\}$, the left side by a simple computation and the right side by an elementary estimate. Finally, assume that $x = m > -\infty$. Since the statement is trivial when $\xi = m$ a.s., we may assume that $b > m$. For any $y \in (m, b)$, we have

$$0 \geq n^{-1} \log P\{\bar{\xi}_n \geq m\} \geq n^{-1} P\{\bar{\xi}_n \geq y\} \rightarrow -\Lambda^*(y) > -\infty.$$

Here $\Lambda^*(y) \rightarrow \Lambda^*(m) = 0$ by continuity, and (3) follows for $x = m$. This completes the proof when $\Lambda < \infty$ on $\mathbb{R}_+$.

The case when $\Lambda(u) = \infty$ for $u > 0$ is trivial. Thus, for any $r > m$ we consider Λ_r and Λ_r^* for the associated

$$n^{-1} \log P\{\bar{\xi}_n \geq x\}$$

Now $\Lambda_r(u) \uparrow \Lambda(u)$ by monotonicity. By the previous theorem the convergence is uniform on $[r, \infty)$. Since also Λ' is unbounded, $\Lambda_r^*(x) \rightarrow \Lambda^*(x)$ for all $x \geq m$. Hence the result follows from (7).

We may now supplement the proof of the decline of the tail probability.

Corollary 27.4 (exponential bound) Let $E\xi < \infty$ and $b = \text{ess sup } \xi < \infty$. Then $P\{\bar{\xi}_n \geq x\}$ decreases exponentially for $x \geq m$. The exponential decline extends to b .

Proof: If $\Lambda(\varepsilon) < \infty$ for some $\varepsilon > 0$, then $\Lambda^*(x) > 0$ for all $x \geq m$ and the result follows from (4).

The large deviation estimate from intervals $[x, \infty)$ to arbitrary intervals $[a, b]$ is a direct consequence of the large-deviation principle for i.i.d. random vectors in $\mathbb{R}^d$. Motivated by these applications and extensions, we assume that Λ is finite on $\mathbb{R}_+$. The interior and closure of a set $A \subset \mathbb{R}^d$ are denoted by $\text{int } A$ and $\bar{A}$ respectively.

Theorem 27.5 (large deviation principle) Let $\xi_1, \xi_2, \dots$ be i.i.d. random vectors in $\mathbb{R}^d$ with $E\xi_1 = m$. Then for any $A \subset \mathbb{R}^d$ we have

$$\begin{aligned} -\inf_{x \in B^\circ} \Lambda^*(x) &\leq \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\bar{\xi}_n \in A\} \\ &\leq \lim_{n \rightarrow \infty} \frac{1}{n} \log P\{\bar{\xi}_n \in \bar{A}\} \end{aligned}$$

Proof: To derive the upper bound, for every $x \in \mathbb{R}^d$ some $u_x \in \mathbb{R}^d$

$$u_x x - \Lambda(u_x)$$

and by continuity we may choose u_x such that

$$u_x y > \Lambda(u_x)$$

By Chebyshev's inequality and (3)

$$\begin{aligned} P\{\bar{\xi}_n \in B_x\} &\leq e^{-n(u_x y - \Lambda(u_x))} \\ &\leq e^{-n(\Lambda^*(y) - \Lambda^*(x))} \end{aligned}$$

Cernoff) Let $\xi, \xi_1, \xi_2, \dots$ be i.i.d. Then for any $x \geq m$, we

$$-\Lambda^*(x). \quad (3)$$

1), we get for any $u > 0$

$$E e^{u S_n} = e^{n \Lambda(u) - n u x},$$

$$u) - u x.$$

use $\Lambda(u) - u x \geq 0$ for $x \geq m$.

$$x \geq m, n \in \mathbb{N}. \quad (4)$$

Assume that $\Lambda < \infty$ on $\mathbb{R}_+$. Let n and $\Lambda'(\infty) = \text{ess sup } \xi \equiv b$, > 0 such that $\Lambda'(u) = a$. Let

$$B], \quad B \in \mathcal{B}. \quad (5)$$

ore $E \eta = \Lambda'_\eta(0) = \Lambda'_\xi(u) = a$.

$$P\{|\bar{\eta}_n - a| < \varepsilon\} \quad (6)$$

of large numbers, and so by

$$u(a + \varepsilon) \geq -\Lambda^*(a + \varepsilon).$$

we get for small enough $\varepsilon > 0$

$$-\Lambda^*(x + 2\varepsilon).$$

ity, we may let $\varepsilon \rightarrow 0$ and

case both sides of (3) equal $\log P\{\xi = b\}$, the left side by an elementary estimate. The statement is trivial when $y \in (m, b)$, we have

$$y\} \rightarrow -\Lambda^*(y) > -\infty.$$

(3) follows for $x = m$. This

The case when $\Lambda(u) = \infty$ for some $u > 0$ may be handled by truncation. Thus, for any $r > m$ we consider the random variables $\xi_k^r = \xi_k \wedge r$. Writing Λ_r and Λ_r^* for the associated functions Λ and Λ^* , we get for $x \geq m \geq E \xi^r$

$$n^{-1} \log P\{\bar{\xi}_n \geq x\} \geq n^{-1} \log P\{\bar{\xi}_n^r \geq x\} \rightarrow -\Lambda_r^*(x). \quad (7)$$

Now $\Lambda_r(u) \uparrow \Lambda(u)$ by monotone convergence as $r \rightarrow \infty$, and by Dini's theorem the convergence is uniform on every compact interval where $\Lambda < \infty$. Since also Λ' is unbounded on the set where $\Lambda < \infty$, it follows easily that $\Lambda_r^*(x) \rightarrow \Lambda^*(x)$ for all $x \geq m$. The required lower bound is now immediate from (7). $\square$

We may now supplement Lemma 27.1 with a criterion for exponential decline of the tail probabilities $P\{\bar{\xi}_n \geq x\}$.

Corollary 27.4 (exponential rate) Let $\xi, \xi_1, \xi_2, \dots$ be i.i.d. with $m = E \xi < \infty$ and $b = \text{ess sup } \xi$. Then for any $x \in (m, b)$, the probabilities $P\{\bar{\xi}_n \geq x\}$ decrease exponentially iff $\Lambda(\varepsilon) < \infty$ for some $\varepsilon > 0$. The exponential decline extends to $x = b$ iff $0 < P\{\xi = b\} < 1$.

Proof: If $\Lambda(\varepsilon) < \infty$ for some $\varepsilon > 0$, then $\Lambda'(0+) = m$ by dominated convergence, and so $\Lambda^*(x) > 0$ for all $x > m$. If instead $\Lambda = \infty$ on $(0, \infty)$, then $\Lambda^*(x) = 0$ for all $x \geq m$. The statement for $x = b$ is trivial. $\square$

The large deviation estimates in Theorem 27.3 are easily extended from intervals $[x, \infty)$ to arbitrary open or closed sets, which leads to the *large-deviation principle* for i.i.d. sequences in $\mathbb{R}$. To fulfill the needs of subsequent applications and extensions, we shall derive a version of the same result in $\mathbb{R}^d$. Motivated by the last result, and also to avoid some technical complications, we assume that $\Lambda(u) < \infty$ for all u . Write B° and B^- for the interior and closure of a set B .

Theorem 27.5 (large deviations in $\mathbb{R}^d$, Varadhan) Let $\xi, \xi_1, \xi_2, \dots$ be i.i.d. random vectors in $\mathbb{R}^d$ with $\Lambda = \Lambda_\xi < \infty$. Then for any $B \in \mathcal{B}^d$, we have

$$\begin{aligned} -\inf_{x \in B^\circ} \Lambda^*(x) &\leq \liminf_{n \rightarrow \infty} n^{-1} \log P\{\bar{\xi}_n \in B\} \\ &\leq \limsup_{n \rightarrow \infty} n^{-1} \log P\{\bar{\xi}_n \in B\} \leq -\inf_{x \in B^-} \Lambda^*(x). \end{aligned}$$

Proof: To derive the upper bound, we fix any $\varepsilon > 0$. By (2) there exists for every $x \in \mathbb{R}^d$ some $u_x \in \mathbb{R}^d$ such that

$$u_x x - \Lambda(u_x) > (\Lambda^*(x) - \varepsilon) \wedge \varepsilon^{-1},$$

and by continuity we may choose an open ball B_x around x such that

$$u_x y > \Lambda(u_x) + (\Lambda^*(x) - \varepsilon) \wedge \varepsilon^{-1}, \quad y \in B_x.$$

By Chebyshev's inequality and (1) we get for any $n \in \mathbb{N}$

$$\begin{aligned} P\{\bar{\xi}_n \in B_x\} &\leq E \exp(u_x S_n - n \inf\{u_x y; y \in B_x\}) \\ &\leq \exp(-n((\Lambda^*(x) - \varepsilon) \wedge \varepsilon^{-1})). \end{aligned} \quad (8)$$