Large deviation theory

Hendrik Weber
hendrik.weber
@warwick.ac.uk

Literature:

- Frank den Hollander — Large deviations
- Deuschel & Stroock — "
- Dembo & Zeitouni — LD, Techniques and application

Lecture notes Wolfgang König (online)
Feng & Kurtz "Large deviations for stochastic processes"

Introduction

LLN: (X_i) iid real valued random var

$$S_n = \sum_{i=1}^{n} X_i \qquad \frac{1}{n} S_n \xrightarrow{a.s.} \mathbb{E}[X_1]$$

if $\mathbb{E}[|X_1|] < \infty$

CLT: typical fluctuations are of order $\sqrt{n}$.

Question: How does $\mathbb{P}\left[\frac{1}{n} S_n \geq x\right]$ behave for $x \geq \mathbb{E}[X_1]$?

Example: (Markov chain Monte Carlo)

μ proba-measure on E

Look for $K(x, dy)$ on E such that μ is unique invariant measure for Markov-chain X_i

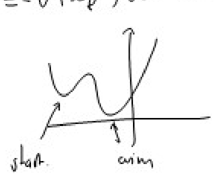
$\Rightarrow$ Ergodic Thm:

$$\mu_N(A) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{A\}}(X_i) \xrightarrow{a.s} \mu(A)$$

$\mathbb{P}[\mu_N \simeq \nu]$: how does this decay?

Example: Suppose we have

$$dx_t^\varepsilon = -V(x_t^\varepsilon)\,dt + \varepsilon\, dw_t \qquad x \in \mathbb{R}^n$$

V 

t fix $\varepsilon \to$

$x_t^\varepsilon \xrightarrow[a.s.]{} x_t$

Outline: First: Classical "Cramér
Schilder
Sanov
Gärtner-Ellis

Second: Markov processes
Freidlin-Wentzell
Donsker-Varadhan

Chapter 1 Large deviations for iid sequences
'Cramér Thm' '38

$\boxed{A}$ Coin tossing

$$X_i = \begin{cases} 1 & \text{with proba } \frac{1}{2} \\ 0 & \end{cases} \quad \text{independent}$$

$$\mathbb{E}[X_i] = \frac{1}{2} \qquad \text{fix } x > \frac{1}{2}$$

$$S_n = \sum_{i=1}^{n} X_i$$

Question: How does $\mathbb{P}\left[\frac{1}{n} S_n \geq x\right]$ behave for n large?

Answer:

$$\mathbb{P}\left[S_n \geq nx\right]$$

$$= \sum_{k \geq nx} \binom{n}{k} \cdot 2^{-n}$$

$$2^{-n} \cdot \max_{k \geq nx} \binom{n}{k} \leq \mathbb{P}\left[S_n \geq nx\right] \leq (n+1)\, 2^{-n} \max_{k \geq nx} \binom{n}{k}$$

max is attained at $k = \lceil nx \rceil$

$$\max_{k \geq nx} \binom{n}{k} = \frac{n!}{(n - \lceil nx \rceil)!\ (\lceil nx \rceil !)} = *$$

$$n! \sim n^n 2^{-n} \sqrt{2\pi n}$$

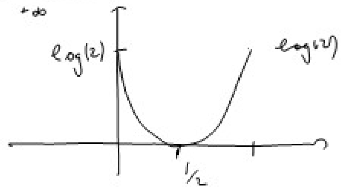
$$\Rightarrow * \sim \frac{\sqrt{2\pi n}}{\sqrt{2\pi(n-nx)}\sqrt{2\pi n x}} \cdot \underbrace{\frac{2^{-n}}{2^{-(n-nx)} \cdot 2^{-nx}}}_{=1} \cdot \underbrace{\frac{n^n}{(n(1-x))^{n(1-x)}\, nx^{nx}}}_{= \frac{1}{(1-x)^{n(1-x)}\, x^{nx}}}$$

$$\Rightarrow \log\left(\mathbb{P}\left[S_n \geq nx\right]\right) = -n\log(2) - n(1-x)\log(1-x) - n(x)\log x + \text{lower order}$$

$$\Rightarrow \frac{1}{n}\log \mathbb{P}\left[S_n \geq nx\right] \xrightarrow{n \to \infty} -I(x)$$

$$I(x) = \log(2) + (1-x)\log(1-x) + x\log x$$

I



convex

Conclusion: (1) $\mathbb{P}[X_n \geq x]$ decay exponentially fast. Question is right exponential rate.

Procedure of taking

$$\frac{1}{n}\log \mathbb{P}(A) \longrightarrow ?$$

corresponds to throwing away lower order.

unique zero (at $\frac{1}{2}$)

$\Rightarrow$ LLN:

For every $\delta > 0$ $\mathbb{P}\left[\frac{1}{n} S_n \geq \frac{1}{2} + \delta\right]$

summable. By Borel-Cantelli ok.

③ ν, μ are measures on some space E

$H(\nu \| \mu)$ relative entropy

Kullback-Leibler divergence

$$= \begin{cases} \int \log \frac{d\nu}{d\mu} \, d\nu & \text{if } \nu \text{ a.c. } \mu \\ +\infty & \text{else.} \end{cases}$$

μ_x on $\{0,1\}$ such that under this measure

$\int y \, \mu_x(dy) = x$

$I(x) = H\left(\mu_x \| \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1\right)$ Exo

$\boxed{B}$ From now on X_i iid real valued random variables

$$\varphi(t) = \mathbb{E}\left[e^{tX}\right] < \infty \quad \forall t$$

$\mathbb{E}[X_i] = m$ $\quad \mathrm{Var}(X_i) = \sigma^2$ $\quad S_n = \sum_{i=1}^n X_i$

Question: How does the probability

$\mathbb{P}\left[\frac{1}{n} S_n \geq x\right]$ decay to 0 for $x > m$?

Calculation: $t > 0$

$$\mathbb{P}\left[\frac{1}{n} S_n \geq x\right] = \mathbb{P}\left[\exp(tS_n) \geq \exp(tnx)\right]$$

$$\overset{\text{Chebyshev}}{\leq} \exp(-tnx) \, \mathbb{E}\left[\exp(tS_n)\right]$$

$$= \exp(-tnx) \; \varphi(t)^n$$

$$= \exp\left(-n\left(tx - \log \varphi(t)\right)\right).$$

$$\Rightarrow \mathbb{P}\left[\frac{1}{n} S_n \geq x\right] \leq \exp\left(-n I(x)\right)$$

$$I(x) = \sup_{t > 0} \left(tx - \log \varphi(t)\right).$$

Comments: Legendre transform of φ

$$\varphi^*(x) = \sup_{t \in \mathbb{R}} \left(tx - \varphi(x)\right)$$

Rate function is the Legendre transform of log of moment generating function.

Exo. The Rate function I from $\boxed{A}$ is also of that form.

Assume that t^* minimises

$tx - \log \varphi(t)$

$$\Rightarrow x - \frac{\varphi'(t_*)}{\varphi(t_*)} = 0$$

$$0 = x - \frac{\varphi'(t)}{\varphi(t)} \qquad \varphi(t) = \mathbb{E}[e^{tX}]$$
$$\varphi'(t) = \mathbb{E}[Xe^{tX}]$$

$$\Rightarrow x = \frac{\mathbb{E}[Xe^{tX}]}{\mathbb{E}[e^{tX}]} \qquad (*)$$

μ = distribution of X_i

$\nu = \frac{e^{tx}}{\varphi(t)}$ defines a new proba. meas μ

then $(*)$ states that $\mathbb{E}_\nu[X] = x$.

$\boxed{C}$ Cramér Theorem,

Theorem: X_i iid ; $\varphi(t) = \mathbb{E}[e^{tX}] < \infty \ \forall t$

$S_n = \sum_{i=1}^{n} \qquad m = \mathbb{E}[X_i]$, $\sigma^2 = \mathrm{Var}(X_i)$

Then for all $x \geq m$

$$\lim_{n\to\infty} n \log \mathbb{P}\left[\tfrac{1}{n} S_n \geq x\right] = -I(x)$$
$$I(x) = \sup_{t\in\mathbb{R}} \left(tx - \log \varphi(t)\right)$$

proof: We assume X is not deterministic.

Step 1 We can reduce ourselves to case

$x = 0 \qquad m < 0.$

proof of 1: Assume, we have proved that case.

Then for general X consider $Y_i = \cancel{X} - x$

$$\mathbb{P}\left[\tfrac{1}{n}\sum_{i=1}^{n} X_i \geq x\right] = \mathbb{P}\left[\tfrac{1}{n}\sum Y_i \geq 0\right]$$

taking log, multiply with n $\longrightarrow -I_Y(0)$

$$I_Y(0) = \sup_{t\in\mathbb{R}} \left(-\log \varphi_Y(t)\right) = * \qquad \varphi_Y(t) = \mathbb{E}[e^{tY}]$$
$$= \mathbb{E}[e^{tX}]\, e^{-tx}$$
$$* = \sup_{t\in\mathbb{R}} \left(tx - \log \varphi(t)\right) = I(x)$$

Hence, we want to proof

$$n \log \mathbb{P}[S_n \geq 0] \longrightarrow -I(0) = -\sup_{t\in\mathbb{R}}(-\log \varphi(t))$$
$$= \log \inf_{t\in\mathbb{R}} \varphi(t)$$
$$= \log \rho.$$

Step 2 Observations: $\varphi \in C^\infty$

$$\varphi(t) = \mathbb{E}[e^{tX}]$$
$$\varphi'(t) = \mathbb{E}[Xe^{tX}]$$
$$\varphi''(t) = \mathbb{E}[X^2 e^{tX}] > 0$$

φ strictly convex , inf can be attained at <u>at most</u> 1 point.

Trivial cases:

(i) $\mathbb{P}[X_i \geq 0] = 0 \implies \mathbb{P}(S_n \geq 0] = 0$

$-\log \infty$

in this case

$$\varphi(t) = \mathbb{E}(e^{tX}] > 0$$

and $\lim_{t\to\infty} \varphi(t) = \mathbb{E}(e^{tX}] = 0$

(ii) Assume that $\mathbb{P}[X_i \leq 0] = 1$ but
$\mathbb{P}[X_i = 0] > 0$

Then $\varphi(t) = \mathbb{E}[e^{tX}]$

$$= \underbrace{\mathbb{E}(e^{tX} \mathbb{1}_{\{X<0\}}]}_{\to 0 \ t\to\infty} + \underbrace{\mathbb{E}(e^{tX}\mathbb{1}_{\{X=0\}})}_{\mathbb{P}(X=0]}$$

$\implies \rho = \mathbb{P}(X=0]$

$\mathbb{P}[S_n \geq 0] = \rho^n$.

Step 3 $\mathbb{P}(X>0) > 0$, $\mathbb{P}(X<0) > 0$.

$\varphi(t) \longrightarrow \pm\infty \quad t \to \pm\infty$

φ strictly convex

$\implies$ there is a unique minimum attained at t_*.

$t_* > 0$ because $\varphi'(0) = \mathbb{E}(X] = m < 0$.

Upper bound: Already proved that $t > 0$

$$\mathbb{P}(S_n \geq 0] = \mathbb{P}(\exp(t_n S_n) \geq 1]$$

$$\leq \mathbb{E}(\exp(t_n S_n)) = \varphi^n(t)$$

infimum is attained for a positive t.

Lower bound: Recall μ was ~~the mean~~ the distribution of X and
ν as $e^{t_* x} \cdot \rho^{-1} \mu(dx)$

under ν $\mathbb{E}(X] = 0$.

$$\mathbb{P}_\mu [S_n \geq 0] = \rho^n \mathbb{E}_\mu \left[e^{-tS_n} \frac{\prod_{i=1}^n e^{tX_i}}{\rho^n} \mathbb{1}_{\{S_n \geq 0\}} \right]$$

$$= \rho^n \mathbb{E}_\nu \left[e^{-tS_n} \mathbb{1}_{\{S_n \geq 0\}} \right]$$

It only remains $\mathbb{E}_\nu (e^{-tS_n} \mathbb{1}_{\{S_n \geq 0\}}]$

under ν $\frac{S_n}{\sqrt{n}} \implies \mathcal{N}(0, \hat{\sigma}^2)$

$$\mathbb{E}_\nu [e^{-tS_n} \mathbb{1}_{\{S_n \geq 0\}}] \geq$$

$$\mathbb{E} \left[e^{-t\sqrt{n}\sigma} \mathbb{1}_{\{S_n \in [0, \sqrt{n}\sigma]\}} \right]$$

$$\geq e^{-t\sqrt{n}\sigma} C \qquad \square$$

D Discussion of rate function:

$$I(x) = \sup_t (tx - \log \varphi(t))$$

Lemma: (i) $\log \varphi$ convex; $I(x)$ convex

(ii) $I \geq 0$ $\quad I(x) = 0$ iff $x = m = \mathbb{E}(X)$

(iii) Sublevelsets of I are compact, in particular I is l.s.c.

proof of (i) $\log \varphi(\lambda x_1 + (1-\lambda)x_2)$

$$= \log \mathbb{E}\left[e^{\lambda x_1 + (1-\lambda)x_2 X} \right]$$

$$\overset{\text{Hölder}}{\leq} \log \mathbb{E}\left(e^{x_1 X}\right)^{\lambda} \mathbb{E}\left(e^{x_2 X}\right)^{(1-\lambda)}$$

$$= \lambda \log \varphi(x_1) + (1-\lambda)\varphi(x_2)$$

$$I(\lambda x_1 + (1-\lambda)x_2) = \sup_t \left((\lambda x_1 + (1-\lambda)x_2)t - \log \varphi(t) \right)$$

$$\leq \lambda \sup_t (x_1 t - \log \varphi(t)) + (1-\lambda) \sup_t ((1-\lambda)tx - \log \varphi(t))$$

rest exercise.

Remark: This implies LLN.

- $\varphi(t) < \infty \ \forall t$ is too strong. We will discuss general case later.

E Relative entropy

E some nice space, ν, μ proba-measures on E

$$H(\nu \| \mu) = \begin{cases} \int \log \frac{d\nu}{d\mu} \, d\nu & \text{if } \nu \ll \mu \\ +\infty & \text{else} \end{cases}$$

relative entropy – Kullback-Leibler divergence.

Lemma: $H(\nu \| \mu) \geq 0$, equality iff $\nu = \mu$.

proof: $f = \frac{d\nu}{d\mu}$ $\rightarrow$

$$H(\nu \| \mu) = \int (\log f) f \, d\mu$$

we know that $\int f \, d\mu = 1$

$(\log f) f \geq f - 1$ equality iff $f = 1$

Lemma: X_i iid random variables, $\varphi(t)$ = moment gen. fct.

μ = distribution of X $\quad I(x) = \sup_t (tx - \log(t))$

- ν_x = "shifted measure"

(i) $H(\nu | \mu)$ is minimized by ν_x among all proba. measures that have expectation x.

(ii) $H(\nu_x | \mu) = I(x)$

proof Exo.