S2F2 - Hauptseminar Stochastische Prozesse und Stochastische Analysis
(WS2019/20)

Große Abweichungen

Large deviations

M.Gubinelli/N.Barashkov

Thema

Die Theorie der großen Abweichungen behandelt in systematischen Weise die Berechnung von Wahrscheinlichkeiten “exponentiell unwahrscheinlicher” Ereignisse.

Diese Theorie ist zu einem der wichtigsten Instrumente der Wahrscheinlichkeitstheorie geworden und erlaubt die Behandlung zahlreicher Anwendungsprobleme. In dem Seminar wollen wir die wichtigsten Grundlagen dieser Theorie erarbeiten und auch einige interessante Anwendungen kennenlernen.

Grundlage bildet das Buch "A Weak Convergence Approach to the Theory of Large Deviations", von Dupuis, Paul, and Richard S. Ellis.

Vorkenntnisse.

Mindestens Einführung in die W-Theorie, ein bisschen auch noch Stochastische Prozesse

Literatur.

Dupuis, Paul, and Richard S. Ellis. 1997. *A Weak Convergence Approach to the Theory of Large Deviations.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York. https://doi.org/10.1002/9781118165904.

Introduction

Wolf's dice data

Rudolph Wolf (1816-1893, swiss astronomer)

Table 1. Wolf's Dice Data

		White Die						Row Total
		1	2	3	4	5	5	
Red Die	1	547	587	500	462	621	690	3407
	2	609	655	497	535	651	684	3631
	3	514	540	468	438	587	629	3176
	4	462	507	414	413	509	611	2916
	5	551	562	499	506	658	672	3448
	6	563	598	519	487	609	646	3422
	Column Total	3246	3449	2897	2841	3635	3932	20000

$$\sum_i \frac{(N_i - pN)^2}{N} \approx (76.87)^2$$

"die Wurfelseiten nicht als gleichmögliche Fälle sich darstellen"

References

https://en.wikipedia.org/wiki/Edwin_Thompson_Jaynes

http://bayes.wustl.edu/etj/articles/entropy.concentration.pdf

Boltzmann's discovery: Sanov's theorem

For sequences $(X_n)_{n\geqslant 1}$ of iid variables on the finite set $\mathcal{K}=\{1,\ldots,N\}$ with common law $\rho\in\Pi(\mathcal{K})$ we can define the *empirical vector* L_n with values in the compact metrizable space $\Pi(\mathcal{K})=\{p\in[0,1]^N : p_1+\cdots+p_N=1\}$ as

$$L_n(i)=\frac{1}{n}\sum_{k=1}^{n} 1_{X_k=i}=\frac{\#\{1\leqslant k\leqslant n: X_n=i\}}{n}$$

and let μ_n to be the law on L_n (thus $\mu_n\in\Pi(\Pi(\mathcal{K}))$).

$$\text{Relative entropy of } \nu \text{ wrt } \mu: \qquad H(\nu|\mu)=\sum_{i=1}^{N}\nu(i)\log\frac{\nu(i)}{\mu(i)}.$$

Theorem. *The sequence $(\mu_n)_n$ satisfy*

$$-\inf_{\nu\in\mathring{A}} H(\nu|\rho)=\liminf_n\frac{1}{n}\log\mu_n(\mathring{A})\leqslant\limsup_n\frac{1}{n}\log\mu_n(\bar{A})=-\inf_{\nu\in\bar{A}} H(\nu|\rho).$$

That is $(\mu_n)_n$ satisfies a ***large deviation*** *principle on $\Pi(\mathcal{K})$ with rate function $I(\nu)=H(\nu|\rho)$.*

This formulation of large deviations have been introduced by Donsker and Varadhan.

Laplace's principle

Theorem. ***Laplace principle*** *$(\mu_n)_n$ has large deviations on $\mathcal{X}$ with rate n and rate function I iff*

$$\frac{1}{n}\log \lim_n \int_{\mathcal{X}} e^{-nf(x)} \mu_n(\mathrm{d}x) = -\inf(f(x) + I(x))$$

for all bounded (Lipshitz) continuous $f: \mathcal{X} \longrightarrow \mathbb{R}$.

Example. Revisiting dice throwing. Observed entropy of Wolf's data:

$$h = H(\hat{L}_n | \rho) = 0.0067696, \qquad n = 200000$$

by large deviations:

$$\mathbb{P}(H(L_n|\rho) \geq h) \approx e^{-nh} \approx 6 \times 10^{58}!!$$

Gibbsian conditioning

In the previous setting fix some integer $k \geqslant 1$ and consider the law $\mu_n \in \Pi(\mathcal{K}^k)$ of $(X_1, \ldots, X_k)$ conditional of an event involving $L_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$, the empirical measure of the vector $(X_1, \ldots, X_n)$:

$$\mu_n(f) = \int_{\mathcal{K}^k} f(x)\,\mu_n(\mathrm{d}x) = \mathbb{E}[f(X_1, \ldots, X_k) | L_n \in B]$$

where $A \in \mathcal{B}(\mathcal{K}^k)$ and $B \in \mathcal{B}(\Pi(\mathcal{K}))$. We will work with $k = 1$ generalization to higher k being easy.

Lemma. *Assume that* B *is closed and* $\inf_{B^\circ} H_\rho = \min_B H_\rho = H_\rho(\hat{\nu})$ *for a unique* $\hat{\nu}$ *then* $\mu_n(f) \longrightarrow \hat{\nu}(f)$.

Interesting case: $B = \{\nu \in \Pi(\mathcal{K}) : \nu(\varphi) \in [e, e + \delta]\}$. Take $\delta > 0$ small and $e \in \mathbb{R}$ is such that $\mathbb{E}[\varphi(X_1)] < e < \sup_{\mathcal{K}} \varphi$ so that $\nu(\varphi) \approx e$ is *atypical* for ρ, by the LLN we have $L_n(\varphi) \longrightarrow \mathbb{E}[\varphi(X_1)]$ a.s.

Let $\lambda \in \mathbb{R}$ and introduce the "tilted" measures $\rho_\lambda = e^{\lambda f}\rho / Z(\lambda)$ with $Z(\lambda) = \rho(e^{\lambda f})$ and observe that $H_\rho(\nu) = H_{\rho_\lambda}(\nu) + \lambda\,\nu(f) - \log Z(\lambda)$.

$$H_\rho(\rho_\lambda) = \lambda e + \log Z(\lambda) = \min_{\nu : \nu(f) \in [e, e+\delta]} [H_{\rho_\lambda}(\nu) + \lambda\,\nu(f)] + \log Z(\lambda) = \min_B H_\rho$$

so $\hat{\nu} = \rho_\lambda$.

Physical interpretation

Consider an assembly of n independent particles each of them characterized by some quantity X_i, $i = 1, \ldots, n$ taking values in $\mathcal{K}$ (e.g. energy, momentum, position, etc...) and assume that the allowed configurations of the whole system are those compatible with a given mean value of some function $f : \mathcal{K} \longrightarrow \mathbb{R}$: $\sum_i f(X_i)/n \simeq e$ (e.g. energy per particle, density, etc..).

This constraint is macroscopic in the sense that involves only an average over all the particles. Then in the limit of a infinite system ($n \longrightarrow \infty$, in reality $n \simeq 10^{23}$) the configurations of a very small subsystem of size k (in our model k is fixed as $n \longrightarrow \infty$) are described by iid configurations, each particle distributes as ρ_λ, the Gibbs distribution compatible with the macroscopic constraint.

This is the mathematical basis of statistical mechanics.

Jupiter's red spot

Can be mathematically understood via large deviations

(see the bachelor thesis of Adrian Rieckert)

Mogulskii theorem

Let $(X_n)_{n\geqslant 1}$ be an iid sequence of Bernoulli(p) r.v. and $X_{\leqslant n} = (X_1, \ldots, X_n)$. Let

$$F_n(x_1, \ldots, x_n)(\theta) = \sum_{i=1}^{n} x_i \, 1_{\theta \in [(i-1)/, i/n)}$$

so that $F_n(X_{\leqslant n})$ is a random element in $\mathcal{K} = \{f \in L^\infty([0,1]) : \|f\|_{L^1} \leqslant 1\}$ and we denote by μ_n its law.

On $\mathcal{K}$ define a distance by taking a countable dense subset $\{\varphi_k\}_{k\geqslant 1}$ of the unit ball of L^1 and letting $d(f,g) = \sum_{k\geqslant 1} 2^{-k} |\varphi_k(f) - \varphi_k(g)|$. Another possible distance is given by $d(f,g) = \sup_{0\leqslant t\leqslant 1} |\int_0^t (f(\theta) - g(\theta)) \mathrm{d}\theta|$.

Let

$$J_p(x) = H(\mathrm{Ber}(x) | \mathrm{Ber}(p)) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}$$

Theorem. (Mogulskii) *The sequence $(\mu_n)_n$ obey the LDP on $\mathcal{K}$ with rate function*

$$I(f) = \int_0^1 J_p(f(\theta)) \mathrm{d}\theta.$$

Large deviations for random walks

Let $(X_n)_{n\geqslant 1}$ be a sequence of iid $\mathrm{Bernoulli}(p)$ random variables. Consider the process $S_n = X_1 + \cdots + X_n$ with $S_0 = 0$. Define a continuous random function φ_n on $[0,1]$ by

$$\varphi_n(t) = \frac{S_k}{n} + \frac{(S_{k+1} - S_k)}{n}(t-k) \qquad \text{for } k \leqslant t < k+1.$$

Let $\mathcal{K}$ be the subset of $C([0,1])$ such that $f \in \mathcal{K}$ if and only if $f(0) = 0$ and $|f(t) - f(s)| \leqslant |t-s|$ for all $0 \leqslant s \leqslant t \leqslant 1$. Observe that $\varphi_n(t)$ is a piecewise linear function for which $\varphi_n(k/n) = S_k/n$.

Theorem. *The sequence $(\mu_n)_n$ obey the LDP on $\mathcal{K}$ with rate function*

$$I(f) = \int_0^1 J_p(f'(s))\mathrm{d}s$$

where $f'(s)$ is the derivative of $f \in \mathcal{K}$ (which exists almost everywhere since f is Lipshitz).

Seminarplan

Wch	*Thema*	*Name*
1	Large deviations in terms of Laplace principle (1.1-1.2)	
2	Basic results in the theory (1.3)	
3	Properties of relative entropy (1.4)	
4	Γ-convergence and Gibbsian-conditioning (notes)	
5	Sanov's theorem. Statement and representation formula (2.1-2.3)	
6	Lower and upper bounds (2.4-2.5)	
7	Mogulskii's theorem. Representation formula (3.1-3.2)	
8	Upper bound and rate function (3.3)	
9	Statement of the theorem and proof + Cramérs theorem + comments (3.4-3.5-3.6)	
10	Random walk model, rep formula + compactness (5.2-5.3)	
11	Upper bound and rate function (6.2)	
12	Lower bound and statement of the theorem (6.5)	
13	Markov chains, rep formula + compactness (8.2)	
14	Upper bound and rate function (8.3-8.4)	
15	Properties of rate function and Lower bound (8.5-8.6)	
16	???	