2 Basic notions

The formalism of the probability theory grows on a probability space $(\Omega, \mathcal{F}, P)$ and the corresponding spaces of random variables, $L_p(\Omega, \mathcal{F}, P)$. In spite of their names, these notions belong to analysis (measure theory, functional analysis) rather than probability theory.

Likewise, the formalism of the large deviations theory grows on notions (LD-convergence, rate function) of analytical nature. They are explained in this section.[1]

2a Large deviation principle (LDP)

Let K be a compact metrizable space.

All continuous functions $K \to \mathbb{R}$ are a separable Banach space $C(K)$.

All (Borel) probability measures on K are a set $P(K)$. Every $\mu \in P(K)$ gives us a linear functional $C(K) \to \mathbb{R}$,

$$f \mapsto \int f \, \mathrm{d}\mu \, ,$$

satisfying two conditions,

$$f \ge 0 \implies \int f \, \mathrm{d}\mu \ge 0 \qquad \text{and} \qquad \int \mathbf{1} \, \mathrm{d}\mu = 1 \, .$$

The linear functional determines μ uniquely.[2] The weak convergence of measures[3] is defined by

$$\mu_n \to \mu \quad \Longleftrightarrow \quad \forall f \in C(K) \ \int f \, \mathrm{d}\mu_n \to \int f \, \mathrm{d}\mu$$

[1] See [1], Sect. 4.3 ('Varadhan's integral lemma') and 4.4 ('Bryc's inverse Varadhan lemma'). "The next theorem could actually be used as a starting point for developing the large deviation paradigm" [1, before Th. 4.3.1]. See also [4, Def. 6.8 and Th. 6.9] ('Laplace principle'), [5, Sect. III.3], [3, Sect. 1.3], [7, Th. 2.2], [2, Th. 2.1.10], [6, Th. 2.6].

[2] In fact, every such functional corresponds to some measure (Riesz-Markov theorem).

[3] Sometimes called 'weak* convergence' by functional analysts.

for $\mu, \mu_n \in P(K)$.

Given $\mu \in P(K)$ and $p \in [1, \infty)$, we have a seminorm $\|\cdot\|_{L_p(\mu)}$ on $C(K)$,

$$\|f\|_{L_p(\mu)} = \Big(\int |f|^p \, \mathrm{d}\mu \Big)^{1/p} \quad \text{for } f \in C(K) \, ,$$

satisfying

$$|f| \le |g| \quad \Longrightarrow \quad \|f\| \le \|g\| \, , \tag{2a1}$$

$$\|\mathbf{1}\| \le 1 \, , \tag{2a2}$$

$$f, g \ge 0 \quad \Longrightarrow \quad \|f \vee g\| \le 2^{1/p} (\|f\| \vee \|g\|) \tag{2a3}$$

for $f, g \in C(K)$; here $a \vee b = \max(a, b)$. Indeed, $\int (f \vee g)^p \, \mathrm{d}\mu \le \int (f^p + g^p) \, \mathrm{d}\mu \le 2 \big((\int f^p \, \mathrm{d}\mu) \vee (\int g^p \, \mathrm{d}\mu) \big)$.

2a4 Exercise. The following two conditions on $\mu, \mu_n \in P(K)$ are equivalent:

(a) $\|f\|_{L_p(\mu_n)} \to \|f\|_{L_p(\mu)}$ for all $f \in C(K)$;

(b) $\mu_n \to \mu$ (weakly).

(As before, p is a *given* number of $[1, \infty)$.)

Prove it.

Hint: $f = |g|^p - |h|^p \dots$

Let $\mu_n \in P(K)$, $p_n \in [1, \infty)$, $p_n \to \infty$. It happens often[1] that the limit

$$\lim_{n \to \infty} \|f\|_{L_{p_n}(\mu_n)} \tag{2a5}$$

exists for all $f \in C(K)$. Then the limit is another seminorm $\|\cdot\|_{\lim}$ on $C(K)$, satisfying (2a1), (2a2) and

$$f, g \ge 0 \quad \Longrightarrow \quad \|f \vee g\|_{\lim} \le \|f\|_{\lim} \vee \|g\|_{\lim} \tag{2a6}$$

for all $f, g \in C(K)$. In order to describe this new seminorm we introduce a function $\Pi : K \to [0, 1]$ by

$$\frac{1}{\Pi(x)} = \sup\{ f(x) : \|f\|_{\lim} \le 1 \} \, . \tag{2a7}$$

It need not be continuous. Rather, $1/\Pi$ is lower semicontinuous (see below), thus, Π is upper semicontinuous. (But why $\Pi(\cdot) \le 1$? Just try $f = \mathbf{1}$.)

[1] And no wonder: in fact, the seminorms on $C(K)$ satisfying (2a1), (2a2) are a *compact* metrizable space…

2a8 Definition. A function $\varphi : K \to \mathbb{R}$ is *lower semicontinuous,* if it satisfies the following equivalent conditions:

(a) $\liminf_{y\to x, y\ne x} \varphi(y) \ge \varphi(x)$ for every $x \in K$;

(b) the set $\{x \in K : \varphi(x) \le c\}$ is closed for every $c \in \mathbb{R}$;

(c) φ is the (pointwise) supremum of some set of continuous functions;

(d) there exist $f_n \in C(K)$ such that $f_n(x) \uparrow \varphi(x)$ for every x.

2a9 Exercise. Prove that (a)–(d) are equivalent.

Hint: (d)$\Longrightarrow$(c) is trivial, (c)$\Longrightarrow$(b) is easy; (b)$\Longrightarrow$(a): consider $\{y : f(y) \le f(x) - \varepsilon\}$; (a)$\Longrightarrow$(d) is harder, consider $f_n(x) = \inf_{y\in K}\bigl(\varphi(y) + n \operatorname{dist}(x,y)\bigr)$.

Upper semicontinuity is defined similarly. Generalization to $\varphi : K \to [-\infty, +\infty]$ is straightforward.

2a10 Exercise. Every upper semicontinuous function on K reaches its supremum (that is, $\exists x \quad \varphi(x) = \sup_y \varphi(y)$); every lower semicontinuous function on K reaches its infimum.

Prove it.

Hint: use compactness.

2a11 Proposition. For every $f \in C(K)$,

$$\|f\|_{\mathrm{lim}} = \max_{x\in K} \bigl(|f(x)|\Pi(x)\bigr) .$$

The proof is postponed to Sect. 4. The supremum is reached due to upper semicontinuity. The claim holds for every seminorm $\|\cdot\|_{\mathrm{lim}}$ satisfying (2a1), (2a2) and (2a6), irrespective of (2a5).

2a12 Exercise. If $\max_K(|f|\Pi_1) = \max_K(|f|\Pi_2)$ for all $f \in C(K)$, then $\Pi_1 = \Pi_2$ (assuming that $\Pi_1, \Pi_2 : K \to [0,1]$ are upper semicontinuous).

Prove it.

Hint: try $f(x) = \bigl(1 - M \operatorname{dist}(x, x_0)\bigr)^+$ for a large M, assuming that $\Pi_1(x_0) < \Pi_2(x_0)$.

It is custom to use the lower semicontinuous function $I : K \to [0,\infty]$ defined by

$$\Pi(x) = \mathrm{e}^{-I(x)} \quad \text{for } x \in K .$$

The function I is well-known as 'the rate function'; the function Π is sometimes called 'deviability'. Defining a seminorm $\|\cdot\|_I$ on $C(K)$ by

$$\|f\|_I = \max_{x\in K}\bigl(|f(x)|\mathrm{e}^{-I(x)}\bigr)$$

we get

$$\lim_{n\to\infty} \|f\|_{L_{p_n}(\mu_n)} = \|f\|_I \quad \text{for } f \in C(K)\,.$$

For now we are mostly interested in the case $p_n = n$. (The case $p_n = n^c$ for a given $c \in (0,1)$, relevant to so-called moderate deviations, will be used later.)

2a13 Definition. (a) A sequence $(\mu_n)_n$ of probability measures on a compact metrizable space K is *LD-convergent,* if the limit

$$\lim_{n\to\infty} \Big(\int |f|^n \, \mathrm{d}\mu_n \Big)^{1/n}$$

exists for all $f \in C(K)$.

(b) The sequence $(\mu_n)_n$ *satisfies LDP with a rate function* I (a lower semicontinuous function $K \to [0,\infty]$), if

$$\lim_{n\to\infty} \Big(\int |f|^n \, \mathrm{d}\mu_n \Big)^{1/n} = \max_{x\in K} \big(|f(x)| \mathrm{e}^{-I(x)} \big)$$

for all $f \in C(K)$.

Proposition 2a11 and Exercise 2a12 ensure the following.

2a14 Corollary. If $(\mu_n)_n$ is LD-convergent then $(\mu_n)_n$ satisfies LDP with one and only one rate function I (a lower semicontinuous function $K \to [0,\infty]$), namely,

$$\mathrm{e}^{I(x)} = \sup\{ f(x) : \lim_{n\to\infty} \|f\|_{L_n(\mu_n)} \le 1 \}\,.$$

2a15 Exercise. Let $K = [0,1]$ and $\mu_n \in P(K)$ be just the Lebesgue measure on $[0,1]$ (for all n). Prove that $(\mu_n)_n$ satisfies LDP with the rate function $I(\cdot) = 0$.

2a16 Exercise. Let $K = [0,1]$, and $\mu_\alpha \in P(K)$ be defined by

$$\int f \, \mathrm{d}\mu_\alpha = (\alpha + 1) \int_0^1 f(x) x^\alpha \, \mathrm{d}x\,.$$

(a) Prove that the sequence $(\mu_n)_n$ is LD-convergent, and find its rate function.
(b) The same for the sequence $(\mu_{2n})_n$.
(c) The same for the sequence $(\mu_{n^2})_n$.
(d) The same for the sequence $(\mu_{\sqrt{n}})_n$.

2a17 Exercise. (a) If $(\mu_n)_n$ is LD-convergent then $(\mu_{2n})_n$ is LD-convergent.

(b) If $(\mu_n)_n$ satisfies LDP with a rate function I, then $(\mu_{2n})_n$ satisfies LDP with the rate function $2I$.

Prove it.

Hint: $\|f\|_{L_n(\mu_{2n})} = \||f|^{1/2}\|^2_{L_{2n}(\mu_{2n})}$.

2a18 Exercise. Let $K = [0,1]$, and $(\mu_n)_n$ satisfy LDP with the rate function $I(x) = \ln(1/x)$. Prove that $\mu_n\big([0,0.5]\big) < 0.6^n$ for all n large enough.

Hint: take $f(\cdot) = 1$ on $[0,0.5]$ but $f(\cdot) = 0$ on $[0.55,1]$.

2a19 Exercise. Prove that

$$\min_{x\in K} I(x) = 0\,.$$

Hint: try $f = \mathbf{1}$.

2a20 Exercise. For every $\varepsilon > 0$,

$$\mu_n\big(\{x : I(x) \le \varepsilon\}\big) \to 1 \quad \text{as } n \to \infty\,.$$

(a) Prove it, assuming that $I(\cdot)$ is continuous.
(b) Prove it in general.

Hint: (a) take $f = \mathrm{e}^I$ and use the Markov inequality, $\mu_n\big(\{x : f^n(x) \ge \mathrm{e}^{n\varepsilon}\}\big) \le \big(\int f^n\,\mathrm{d}\mu_n\big)/(\mathrm{e}^{n\varepsilon})$; (b): use 2a8(d).

2b Contraction principle

Let K_1, K_2 be compact metrizable spaces, $F : K_1 \to K_2$ a continuous map, $(\mu_n)_n$ a sequence of probability measures on K_1, and $(\nu_n)_n$ its image on K_2 (that is, $\nu_n(B) = \mu_n(F^{-1}(B))$ for Borel sets $B \subset K_2$).

2b1 Theorem. (a) If $(\mu_n)_n$ is LD-convergent, then $(\nu_n)_n$ is LD-convergent.

(b) If $(\mu_n)_n$ satisfies LDP with a rate function I_1, then $(\nu_n)_n$ satisfies LDP with a rate function I_2 such that

$$I_2(y) = \min\{I_1(x) : x \in K_1, F(x) = y\}\,.$$

If $F^{-1}(\{y\}) = \emptyset$ then the minimum is $+\infty$ by definition. Otherwise, the minimum is reached since $F^{-1}(\{y\})$ is compact and I_1 is lower semicontinuous.

2b2 Exercise. Prove Theorem 2b1.

Hint: given $g \in C(K_2)$, introduce $f \in C(K_1)$ by $f(x) = g(F(x))$ and note that $\int |f|^n\,\mathrm{d}\mu_n = \int |g|^n\,\mathrm{d}\nu_n$.

2c Change of measure

2c1 Theorem. Let $(\mu_n)_n$, $(\nu_n)_n$ be two sequences of probability measures on a compact metrizable space K, satisfying

$$\frac{\mathrm{d}\nu_n}{\mathrm{d}\mu_n} = c_n \mathrm{e}^{-nh} \quad \text{for all } n$$

for some $h \in C(K)$ and $c_1, c_2, \dots \in (0, \infty)$.

(a) If $(\mu_n)_n$ is LD-convergent then $(\nu_n)_n$ is LD-convergent.

(b) If $(\mu_n)_n$ satisfies LDP with a rate function I, then $(\nu_n)_n$ satisfies LDP with the rate function

$$J = (I + h) - \min_K (I + h) = I + h - \lim_{n\to\infty} \frac{1}{n} \ln c_n \, .$$

2c2 Exercise. Prove Theorem 2c1.

Hint: $\big(\int |f|^n \, \mathrm{d}\nu_n\big)^{1/n} = \big(\int (|f|\mathrm{e}^{-h})^n \, \mathrm{d}\mu_n\big)^{1/n} / \big(\int (\mathrm{e}^{-h})^n \, \mathrm{d}\mu_n\big)^{1/n} \to \max(\dots)/\max(\dots)$.

See also [5, Th. III.17] ('tilted LDP').

References

[1] A. Dembo, O. Zeitouni, *Large deviations techniques and applications,* Jones and Bartlett publ., 1993.

[2] J.-D. Deuschel, D.W. Stroock, *Large deviations,* Academic Press, 1989.

[3] P. Dupuis, R.S. Ellis, *A weak convergence approach to the theory of large deviations,* Wiley, 1997.

[4] R.S. Ellis, *The theory of large deviations and applications to statistical mechanics,* 2006, http://www.math.umass.edu/∼rsellis/pdf-files/Dresden-lectures.pdf

[5] F. den Hollander, *Large deviations,* AMS, 2000.

[6] D.W. Stroock, *An introduction to the theory of large deviations,* Springer, 1984.

[7] S.R.S. Varadhan, *Large deviations and applications,* Soc. for Industrial and Appl. Math., 1984.

Index