5 LDP in spaces of functions

5a The simplest case of Mogulskii's theorem

Tossing a fair coin n times we get a random element of $\{0,1\}^n$. We embed all these spaces $\{0,1\}^n$ into a single metrizable compact space

(5a1) $$K = \{\varphi \in L_\infty(0,1) : \mathbf{0} \le \varphi \le \mathbf{1}\}$$

as follows: given $\beta = (\beta_1, \dots, \beta_n) \in \{0,1\}^n$, we define $\varphi_\beta \in K$ by

(5a2) $$\varphi_\beta(t) = \beta_k \quad \text{for } t \in \Big(\frac{k-1}{n}, \frac{k}{n} \Big) .$$

The relevant metrizable topology on K, well-known as the weak* topology, may be described as follows: for $\varphi, \varphi_1, \varphi_2, \dots \in K$,

(5a3) $$\varphi_k \to \varphi \quad \text{if and only if} \quad \forall \eta \in L_1(0,1) \; \int \varphi_k \eta \to \int \varphi \eta .$$

Here is an example of a metric that generates this topology:

(5a4) $$\operatorname{dist}(\varphi, \psi) = \max_k \frac{1}{k} \bigg| \int \varphi \eta_k - \int \psi \eta_k \bigg| ,$$

where $\eta_1, \eta_2, \dots$ are a sequence dense in the unit ball of $L_1(0,1)$. The choice of $\eta_1, \eta_2, \dots$ influences the metric but not the topology. Another metric (for the same topology):

(5a5) $$\operatorname{dist}(\varphi, \psi) = \max_{t \in [0,1]} \bigg| \int_0^t \varphi - \int_0^t \psi \bigg| .$$

We consider the distribution μ_n of the random function φ_β,

(5a6) $$\mu_n \in P(K) , \quad \int f \, \mathrm{d}\mu_n = \frac{1}{2^n} \sum_{\beta \in \{0,1\}^n} f(\varphi_\beta) .$$

5a7 Exercise. Assume that $(\mu_n)_n$ satisfies LDP with a rate function I. Then

$$\min\{I(\varphi) : \varphi \in K,\ \int \varphi = u\} = I_{0.5}(u)\,,$$

where $I_{0.5}(u) = u \ln \frac{u}{0.5} + (1-u) \ln \frac{1-u}{0.5} = u \ln u + (1-u)\ln(1-u) + \ln 2$ (recall (3a5) and (3a9)).

Prove it.

Hint: the contraction principle (Th. 2b1), and 3a4.

5a8 Exercise. Assume that $(\mu_n)_n$ satisfies LDP with a rate function I. Then

$$I(\varphi) = \frac{I(\varphi_{\text{left}}) + I(\varphi_{\text{right}})}{2}$$

for all $\varphi \in K$; here $\varphi_{\text{left}}, \varphi_{\text{right}} \in K$ are defined by

$$\varphi_{\text{left}}(t) = \varphi(0.5t)\,,\ \varphi_{\text{right}}(t) = \varphi(0.5 + 0.5t) \quad \text{for } t \in (0,1)\,.$$

Prove it.

Hint: $K = K_1 \times K_2$, $K_1 \subset L_\infty(0, 0.5)$, $K_2 \subset L_\infty(0.5, 1)$; $\mu_{2n} = \mu_n^{(1)} \times \mu_n^{(2)}$; $2I(\varphi) = I_1(\varphi_1) + I_2(\varphi_2)$ by 4d1, 4d2 and 2a17. On the other hand, the natural one-to-one correspondence between K and K_1 transforms μ_n to $\mu_n^{(1)}$, thus, I to I_1.

Applying the same formula to $I(\varphi_{\text{left}})$ and $I(\varphi_{\text{right}})$ we split $I(\varphi)$ into four terms. And so on.

Now you could guess the rate function!

5a9 Theorem. $(\mu_n)_n$ satisfies LDP with the rate function

$$I(\varphi) = \int_0^1 I_{0.5}(\varphi(t))\,\mathrm{d}t\,.$$

See [1, Th. 5.1.2].

Note that I is far from being continuous. In fact,

$$\liminf_{\psi \to \varphi} I(\psi) = I(\varphi) \quad \text{but} \quad \limsup_{\psi \to \varphi} I(\psi) = \ln 2$$

for all $\varphi \in K$. Note also that

$$\mu_n\{\varphi \in K : I(\varphi) = \ln 2\} = 1 \quad \text{for all } n\,.$$

How could we prove the theorem? The approach of 3a does not work here, since the number of atoms of μ_n is exponentially large. No binomial coefficients, just 2^n atoms of probability 2^{-n} each. However, we may apply Sanov's theorem to $\int_0^1 \varphi$, $\int_0^{0.5} \varphi$, $\int_{0.5}^1 \varphi$ and so on. Doing so in the next section, we'll prove the theorem for $n \in \{1, 2, 4, 8, \dots\}$. Here we just discuss it.

The map $K \to C[0,1]$,

$$\varphi \mapsto w\,, \quad w(t) = \int_0^t \varphi(s)\,\mathrm{d}s\,,$$

is continuous and one-to-one, therefore (by compactness) a homeomorphism. Thus, the LDP on K leads to LDP on the set of functions $w : [0,1] \to \mathbb{R}$ such that

(5a10) $0 \le w(t) - w(s) \le t - s \quad \text{whenever } 0 \le s \le t \le 1\,, \quad \text{and} \quad w(0) = 0$

with the rate function

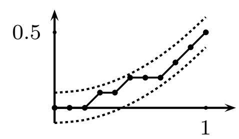
(5a11) $$J(w) = \int_0^1 I_{0.5}(w'(t))\,\mathrm{d}t\,.$$

(The derivative exists almost everywhere.) Note that the random function w_β (corresponding to φ_β) is piecewise linear, with the derivative $\beta_k \in \{0,1\}$ on $\left(\frac{k-1}{n}, \frac{k}{n}\right)$. It is a (rescaled) path of a random walk.

Do not hesitate to use Theorem 5a9 in the exercises below.

5a12 Exercise. A fair coin is tossed n times, giving $(\beta_1, \dots, \beta_n) \in \{0,1\}^n$. Consider

$$p_{n,\varepsilon} = \mathbb{P}\left(\forall k = 1, \dots, n \quad \left|\frac{\beta_1 + \dots + \beta_k}{n} - \frac{1}{2}\Bigl(\frac{k}{n}\Bigr)^2\right| \le \varepsilon\right).$$



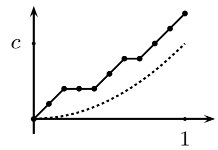
Prove that

$$\limsup_{n\to\infty} \left| \sqrt[n]{p_{n,\varepsilon}} - \frac{\sqrt{\mathrm{e}}}{2} \right| \to 0 \quad \text{as } \varepsilon \to 0+ \,.$$

Hint: use 4b12.

5a13 Exercise. A fair coin is tossed n times, giving $(\beta_1, \dots, \beta_n) \in \{0,1\}^n$. Given $c \in [0,1]$, we consider

$$p_n = \mathbb{P}\left(\forall k = 1, \dots, n \quad \frac{\beta_1 + \dots + \beta_k}{n} \ge c\Bigl(\frac{k}{n}\Bigr)^2\right).$$



Prove that

$$\sqrt[n]{p_n} \to 1 \qquad \text{for } 0 \le c \le 0.5\,,$$
$$\sqrt[n]{p_n} \to \frac{1}{2c^c(1-c)^{1-c}} \qquad \text{for } 0.5 \le c \le 1$$

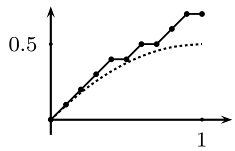
($0^0 = 1$, as before).

Hint: use 4b6; guess the extremal function; prove your guess, taking into account that $\int_0^1 I_{0.5}(\varphi(t))\,\mathrm{d}t \ge I_{0.5}\big(\int_0^1 \varphi(t)\,\mathrm{d}t\big)$.

5a14 Exercise. In the situation of 5a13, formulate and prove a statement about the conditional distribution (in the spirit of 4c5).

Another example:

$$p_n = \mathbb{P}\Big(\forall k = 1,\dots,n \quad \frac{\beta_1 + \dots + \beta_k}{n} \ge \frac{k}{n} - \frac{1}{2}\Big(\frac{k}{n}\Big)^2\Big)\,.$$



It appears that

$$\sqrt[n]{p_n} \to \frac{\mathrm{e}^{1/4}}{\sqrt{2}} \quad \text{as } n \to \infty\,.$$

The extremal function is

$$w(t) = \begin{cases} t - 0.5t^2 & \text{for } 0 \le t \le 0.5, \\ 0.5t + 0.125 & \text{for } 0.5 \le t \le 1. \end{cases}$$

In order to prove its extremality, the following lemma helps: $J(w \wedge v) \le J(w)$ for every *linear* function $v : [0,1] \to \mathbb{R}$ such that $v(0) \ge 0$ and $v'(\cdot) \ge 0.5$; here $w \wedge v$ is the pointwise minimum.

Two-dimensional random arrays are quite similar. The interval $(0,1)$ and the square $(0,1) \times (0,1)$ are isomorphic measure spaces, thus, $L_\infty(0,1)$ and $L_\infty\big((0,1)\times(0,1)\big)$ are isomorphic. But moreover, the natural partition of the interval into 2^{2n} parts corresponds to that of the square. And the natural correspondence between the compact sets K in dimensions 1 and 2 is a homeomorphism. Thus, Theorem 5a9 implies the corresponding result in two (and more) dimensions. Note also that the metric

$$\operatorname{dist}(\varphi,\psi) = \max_{s,t\in[0,1]} \bigg| \iint_{(0,s)\times(0,t)} (\varphi - \psi) \bigg|$$

generates the considered topology on the space K (over the square). Thus, we may consider two-dimensional 'paths', getting the rate function

$$J(w) = \iint_{(0,1)\times(0,1)} I_{0.5}\Big(\frac{\partial^2}{\partial s \partial t} w(s,t)\Big)\,\mathrm{d}s\mathrm{d}t\,.$$

5b Infinite dimension as the limit of finite dimensions: the Dawson-Gärtner theorem

We return for a while to the general situation: a compact metrizable space K and a sequence $(\mu_n)_n$ of probability measures on K.

Given $g \in C(K)$, we may consider the distribution ν_n of g w.r.t. μ_n, that is, the probability measure on $\mathbb{R}$ defined by $\nu_n(B) = \mu_n(\{x : g(x) \in B\}) = \mu_n(g^{-1}(B))$ for Borel sets $B \subset \mathbb{R}$; equivalently, $\int_K f_1(g(\cdot))\,\mathrm{d}\mu_n = \int_\mathbb{R} f_1\,\mathrm{d}\nu_n$ for all continuous (or bounded Borel) functions $f_1 : \mathbb{R} \to \mathbb{R}$. Clearly, ν_n are concentrated on the compact set $g(K) \subset \mathbb{R}$. If $(\mu_n)_n$ is LD-convergent (on K) then $(\nu_n)_n$ is also LD-convergent (on $g(K)$) by the contraction principle. The opposite is generally wrong.

5b1 Exercise. The sequence $(\nu_n)_n$ is LD-convergent if and only if the limit $\lim_n \|f\|_{L_n(\mu_n)}$ exists for all $f \in C(K)$ of the form $f(\cdot) = f_1(g(\cdot))$ for continuous $f_1 : \mathbb{R} \to \mathbb{R}$.

Prove it.

Hint: $\|f\|_{L_n(\mu_n)} = \|f_1\|_{L_n(\nu_n)}$.

Given $g, h \in C(K)$, we may consider the joint distribution ν_n of g, h w.r.t. μ_n, that is, the probability measure on $\mathbb{R}^2$ defined by $\nu_n(B) = \mu_n(\{x : (g(x), h(x)) \in B\})$ for Borel sets $B \subset \mathbb{R}^2$. Similarly to 5b1, LD-convergence of $(\nu_n)_n$ means convergence of $\|f\|_{L_n(\mu_n)}$ for all $f \in C(K)$ of the form $f(\cdot) = f_2(g(\cdot), h(\cdot))$ for continuous $f_2 : \mathbb{R}^2 \to \mathbb{R}$.

Given $g_1, g_2, \dots \in C(K)$, we may consider the joint distribution $\nu_n^{(j)}$ of $g_1, \dots, g_j$ w.r.t. μ_n. LD-convergence of $(\nu_n^{(j)})_n$ for all j means convergence of $\|f\|_{L_n(\mu_n)}$ for all $f \in C(K)$ of the form $f(\cdot) = f_j(g_1(\cdot), \dots, g_j(\cdot))$, for all j. Are all such f dense in $C(K)$? They are a subalgebra of $C(K)$, thus, the answer is given by the Stone-Weierstrass theorem:

> A subalgebra of $C(K)$ is dense if and only if
> it separates points of K.

5b2 Theorem. Let $g_1, g_2, \dots \in C(K)$ separate points of K, and $\nu_n^{(j)}$ be the joint distribution of $g_1, \dots, g_j$ w.r.t. μ_n. Then

(a) If for each j the sequence $\big(\nu_n^{(j)}\big)_n$ is LD-convergent (on the image $K_j \subset \mathbb{R}^j$ of K under the map $x \mapsto (g_1(x), \dots, g_j(x))$), then the sequence $(\mu_n)_n$ is LD-convergent.

(b) If for each j the sequence $\big(\nu_n^{(j)}\big)_n$ satisfies LDP with a rate function I_j on K_j then the sequence $(\mu_n)_n$ satisfies LDP with the rate function

$$I(x) = \sup_j I_j\big(g_1(x), \dots, g_j(x)\big)\,.$$

(See also [1, Th. 4.6.1].)

Proof. By the Stone-Weierstrass theorem, functions $f \in C(K)$ of the form $f(\cdot) = f_j(g_1(\cdot), \dots, g_j(\cdot))$ are a dense set $D \subset C(K)$.

(a) Convergence of $\| \cdot \|_{L_n(\mu_n)}$ on D implies convergence on the whole $C(K)$, since

$$\limsup_n \|f\|_{L_n(\mu_n)} - \liminf_n \|f\|_{L_n(\mu_n)} \le \\ \le 2\|f - \tilde f\|_{C(K)} + \limsup_n \|\tilde f\|_{L_n(\mu_n)} - \liminf_n \|\tilde f\|_{L_n(\mu_n)} = 2\|f - \tilde f\|_{C(K)}$$

for $f \in C(K)$, $\tilde f \in D$.

(b) We will prove that

$$\mathrm{e}^{I(x)} = \sup\{ f(x) : \|f\| \le 1 \} \, ,$$

where $\|f\| = \lim_n \|f\|_{L_n(\mu_n)}$. For each j it is given that

$$\mathrm{e}^{I_j(y_j)} = \sup\{ f_j(y_j) : \|f_j\|_j \le 1 \} \, ,$$

where $y_j = \big(g_1(x), \dots, g_j(x) \big)$ and $\|f_j\|_j = \lim_n \|f_j\|_{L_n(\nu_n^{(j)})}$. If $f(\cdot) = f_j(g_1(\cdot), \dots, g_j(\cdot))$ then $\|f\| = \|f_j\|_j$ (since $\|f\|_{L_n(\mu_n)} = \|f_j\|_{L_n(\nu_n^{(j)})}$) and $f(x) = f_j(y_j)$. Thus, $\sup\{ f(x) : \|f\| \le 1 \} \ge \sup\{ f_j(y_j) : \|f_j\| \le 1 \} = \mathrm{e}^{I_j(y_j)}$ for all j, therefore

$$\sup\{ f(x) : \|f\| \le 1 \} \ge \sup_j \mathrm{e}^{I_j(y_j)} = \mathrm{e}^{I(x)} \, .$$

On the other hand, $f(x) = f_j(y_j) \le \mathrm{e}^{I_j(y_j)} \le \mathrm{e}^{I(x)}$ for $f \in D$, $\|f\| \le 1$. More generally, $f(x) \le \|f\| \mathrm{e}^{I(x)}$ for all $f \in D$. Given $\varepsilon > 0$ and an arbitrary $f \in C(K)$ such that $\|f\| \le 1$, we take $\tilde f \in D$ such that $\|f - \tilde f\|_{C(K)} \le \varepsilon$, then $f(x) \le \tilde f(x) + \varepsilon \le \|\tilde f\| \mathrm{e}^{I(x)} + \varepsilon \le (1+\varepsilon)\mathrm{e}^{I(x)} + \varepsilon$. Therefore $f(x) \le \mathrm{e}^{I(x)}$, that is,

$$\mathrm{e}^{I(x)} \ge \sup\{ f(x) : \|f\| \le 1 \} \, .$$

□

Note that

$$I_j(y_1, \dots, y_j) = \min_{y_{j+1} : (y_1, \dots, y_{j+1}) \in K_{j+1}} I_{j+1}(y_1, \dots, y_{j+1}) \quad \text{for } (y_1, \dots, y_j) \in K_j$$

by the contraction principle. Thus,

$$I_j(g_1(x), \dots, g_j(x)) \uparrow I(x) \quad \text{as } j \to \infty \, . \tag{5b3}$$

It is easy to generalize Theorem 5b2 to the situation where j runs on a subsequence (say, $j \in \{2, 4, 8, \dots\}$).

5b4 Exercise. Generalize 5b2 to continuous functions $g_j : K \to K_0$ (rather than $K \to \mathbb{R}$), where K_0 is another compact metrizable space.

5b5 Exercise. Let K be a compact metrizable space and $(\mu_n)_n$ a sequence of probability measures on K. Consider the compact metrizable space

$$K^\infty = K \times K \times \dots\,;$$

it may be metrized by

$$\operatorname{dist}_\infty\big((x_1, x_2, \dots), (y_1, y_2, \dots)\big) = \max_k \frac{1}{k} \operatorname{dist}(x_k, y_k)\,.$$

On K^∞ we consider product measures

$$\mu_n^\infty = \mu_n \times \mu_n \times \dots$$

(a) The sequence $(\mu_n^\infty)_n$ is LD-convergent if and only if the sequence $(\mu_n)_n$ is LD-convergent.

(b) If $(\mu_n)_n$ satisfies LDP with a rate function $I : K \to [0, \infty]$, then $(\mu_n^\infty)_n$ satisfies LDP with the rate function $I_\infty : K^\infty \to [0, \infty]$,

$$I_\infty\big((x_1, x_2, \dots)\big) = I(x_1) + I(x_2) + \dots$$

Prove it.

Hint: 4d1, 4d2 and 5b4.

If K is defined by (5a1), (5a3), then (up to a natural isomorphism)

$$\text{(5b6)} \qquad \begin{gathered} K^\infty = \{\varphi \in L_\infty(0,\infty) : \mathbf{0} \le \varphi \le \mathbf{1}\}\,, \\ \varphi_k \to \varphi \quad \text{if and only if} \quad \forall \eta \in L_1(0,\infty) \; \int \varphi_k \eta \to \int \varphi \eta \end{gathered}$$

for $\varphi, \varphi_1, \varphi_2, \dots \in K^\infty$. It is straightforward to adapt (5a4) to K^∞. However, (5a5) needs a modification, say,

$$\operatorname{dist}(\varphi, \psi) = \max_{t \in [0,\infty)} \frac{1}{t^2 + 1} \bigg| \int_0^t \varphi - \int_0^t \psi \bigg|\,.$$

Now we toss a coin endlessly, getting $\beta = (\beta_1, \beta_2, \dots) \in \{0,1\}^\infty$, define $\varphi_\beta \in K^\infty$ by (5a2) (waiving the restriction $k \le n$) and observe that this φ_β is distributed μ_n^∞ (μ_n being defined by (5a6)). By 5b5 and Theorem 5a9 (not proved yet), $(\mu_n^\infty)_n$ satisfies LDP with the rate function $I_\infty : K^\infty \to [0, \infty]$,

$$\text{(5b7)} \qquad I_\infty(\varphi) = \int_0^\infty I_{0.5}(\varphi(t))\, \mathrm{d}t\,.$$

This time,

$$\liminf_{\psi\to\varphi} I(\psi) = I(\varphi) \quad \text{but} \quad \limsup_{\psi\to\varphi} I(\psi) = \infty$$

for all $\varphi \in K^\infty$. Also

$$\mu_n\{\varphi \in K^\infty : I(\varphi) = +\infty\} = 1 \quad \text{for all } n\,.$$

5c Proof for nice n

We return to Theorem 5a9. It states LDP, namely, that $\|f\|_{L_n(\mu_n)} \to \max\bigl(|f|\mathrm{e}^{-I}\bigr)$ as $n \to \infty$ for all $f \in C(K)$. Here we prove a weaker statement (LDP along a subsequence):

$$\|f\|_{L_{2^m}(\mu_{2^m})} \to \max\bigl(|f|\mathrm{e}^{-I}\bigr) \quad \text{as } m \to \infty\,.$$

In order to use 5b, we define $g_2, g_3, \dots \in C(K)$ by[1]

$$g_j(\varphi) = \frac{1}{\operatorname{mes} I_j} \int_{I_j} \varphi\,,$$

where

$$(I_2, I_3, I_4, I_5 \dots) = \bigl((0,1), (0,0.5), (0.5,1), (0,0.25), \dots\bigr)$$

is the sequence of all dyadic intervals. Clearly, g_j separate points of K. We introduce $\nu_n^{(j)}$ on K_j as in 5b2, but we restrict ourselves to

$$j \in \{2,4,8,\dots\}\,, \quad n \in \{2,4,8,\dots\}\,, \quad n \ge j\,.$$

The set

$$K_{2j} = \{(g_2(\varphi), \dots, g_{2j}(\varphi)) : \varphi \in K\}$$

lies in $\mathbb{R}^{2j-1}$, but only the last j coordinates $g_{j+1}, \dots, g_{2j}$ are really needed; they determine $g_2, \dots, g_j$ uniquely. (For example, $g_2(\cdot) = \frac12(g_3(\cdot) + g_4(\cdot))$.)

If φ is distributed μ_n then $g_{j+1}(\varphi), \dots, g_{2j}(\varphi)$ are independent, identically distributed; namely, each of them is distributed $\mu^{3a}_{n/j}$, where μ^{3a} means 'μ of Sect. 3a'. By 3a4, $(\mu^{3a}_k)_k$ satisfies LDP with the rate function $I_{0.5}$. Thus (similarly to 2a17), for $k = j+1, \dots, 2j$,

$$\|f(g_k(\cdot))\|_{L_n(\mu_n)} = \|f\|_{L_n(\mu^{3a}_{n/j})} = \|f^j\|^{1/j}_{L_{n/j}(\mu^{3a}_{n/j})} \xrightarrow[n\to\infty]{} \\ \xrightarrow[n\to\infty]{} \bigl(\max\bigl(|f^j|\mathrm{e}^{-I_{0.5}}\bigr)\bigr)^{1/j} = \max\bigl(|f|\mathrm{e}^{-I_{0.5}/j}\bigr)\,,$$

[1]The numbers start from 2 for convenience; the natural blocks finish at $j = 2^k$.

that is, $I_{0.5}/j$ is the rate function for $g_k(\cdot)$ (along the subsequence, $n \in \{j, 2j, 3j, \dots\} \supset \{j, 2j, 4j, \dots\}$).

Prop. 4d1 (or rather, its evident generalization to the product of j measures, and n restricted to a subsequence) gives us the rate function $(y_{j+1}, \dots, y_{2j}) \mapsto \frac{1}{j}\big(I_{0.5}(y_{j+1}) + \dots + I_{0.5}(y_{2j})\big)$ for $(g_{j+1}, \dots, g_{2j})$, therefore, the rate function I_{2j} on K_{2j},

$$I_{2j}(y_2, \dots, y_{2j}) = \frac{1}{j}\big(I_{0.5}(y_{j+1}) + \dots + I_{0.5}(y_{2j})\big) \tag{5c1}$$

for distributions $\nu_n^{(2j)}$ of $g_2, \dots, g_{2j}$ (along the subsequence, still).

The Dawson-Gärtner theorem 5b2 (or rather, its evident generalization to subsequences) gives us LDP for $(\mu_n)_n$ with the rate function

$$I(\varphi) = \lim_j 2^{-j} \sum_{k=1}^{2^j} I_{0.5}\bigg(2^j \int_{(k-1)2^{-j}}^{k\cdot 2^{-j}} \varphi\bigg)$$

(recall 5b3). That is, $I(\varphi) = \lim_j \int I_{0.5}(\varphi_j)$, where φ_j is the orthogonal projection of φ to the 2^j-dimensional space of step functions. However, $\varphi_j \to \varphi$ in measure (in fact, almost everywhere), therefore $I_{0.5}(\varphi_j) \to I_{0.5}(\varphi)$ in measure, therefore (using boundedness), $\int I_{0.5}(\varphi_j) \to \int I_{0.5}(\varphi)$.

5d Measures coming together

A general situation, again: $(\mu_n)_n$ and $(\nu_n)_n$ be two sequences of probability measures on a compact metrizable space K. We say that they *come together,* if there exist probability measures λ_n on $K \times K$ satisfying two conditions.

First, μ_n and ν_n are the marginals of λ_n (for every n). That is, $\lambda_n(B \times K) = \mu_n(B)$ and $\lambda_n(K \times B) = \nu_n(B)$ for every Borel set $B \subset K$. Or equivalently, $\int_{K\times K} f(x)\,\lambda_n(\mathrm{d}x\mathrm{d}y) = \int_K f\,\mathrm{d}\mu$ and $\int_{K\times K} f(y)\,\lambda_n(\mathrm{d}x\mathrm{d}y) = \int_K f\,\mathrm{d}\nu$ for all $f \in C(K)$. (Every such λ_n is called a joining of μ_n and ν_n.)

Second, there exist $\varepsilon_n \to 0$ such that $\lambda_n\big(\{(x,y) : \mathrm{dist}(x,y) \le \varepsilon_n\}\big) = 1$ for all n. (The choice of the metric affects the choice of ε_n, but the condition is invariant.)

An equivalent definition without joinings exists (but will not be used). Namely, $(\mu_n)_n$ and $(\nu_n)_n$ come together, if there exist $\varepsilon_n \to 0$ such that (recall (4b9), (4b10)) $\mu_n(F) \le \nu_n(F_{+\varepsilon_n})$ and $\nu_n(F) \le \mu_n(F_{+\varepsilon_n})$ for all closed sets $F \subset K$. (The same $(\varepsilon_n)_n$ for all F, of course.)

5d1 Proposition. If $(\mu_n)_n$ and $(\nu_n)_n$ come together, then

(a) $(\mu_n)_n$ is LD-convergent if and only if $(\nu_n)_n$ is LD-convergent;

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

Proof. Given $f \in C(K)$, we introduce $f_1, f_2 \in C(K \times K)$ by $f_1(x,y) = f(x)$ and $f_2(x,y) = f(y)$. Then $\|f\|_{L_n(\mu_n)} = \|f_1\|_{L_n(\lambda_n)}$ and $\|f\|_{L_n(\nu_n)} = \|f_2\|_{L_n(\lambda_n)}$. However,

$$\max_{\mathrm{dist}(x,y)\le\varepsilon_n} |f_1(x,y) - f_2(x,y)| = \max_{\mathrm{dist}(x,y)\le\varepsilon_n} |f(x) - f(y)| \to 0 \quad \text{as } n \to \infty$$

since f is uniformly continuous (due to compactness). Thus, $\|f_1 - f_2\|_{L_n(\lambda_n)} \to 0$, therefore

$$\|f\|_{L_n(\mu_n)} - \|f\|_{L_n(\nu_n)} = \|f_1\|_{L_n(\lambda_n)} - \|f_2\|_{L_n(\lambda_n)} \to 0$$

as $n \to \infty$; (a) and (b) follow immediately. □

See also [1, Th. 4.2.13].

5e Proof for all n

Here we finish the proof of Theorem 5a9 by generalizing the argument of 5c from $n \in \{2,4,8,\dots\}$ to $n \in \{1,2,3,\dots\}$.

We consider the distribution $\nu_n^{(j)}$ on K_j; still, $j \in \{2,4,8,\dots\}$, but now $n \in \{1,2,3,\dots\}$. It is sufficient to prove that $(\nu_n^{(j)})_n$ satisfies LDP with the rate function I_j (recall (5c1)), that is,

(5e1) $$\|f\|_{L_n(\nu_n^{(j)})} \to \max_{K_j} \big(|f| \mathrm{e}^{-I_j} \big) \quad \text{as } n \to \infty$$

for all $f \in K_j$ and all $j \in \{2,4,8,\dots\}$. Recall that the argument of 5c gives us a weaker statement, namely,

$$\|f\|_{L_{mj}(\nu_{mj}^{(2j)})} \to \max_{K_j} \big(|f| \mathrm{e}^{-I_{2j}} \big) \quad \text{as } m \to \infty\,.$$

(Only $m \in \{1,2,4,\dots\}$ are used there, but the argument works for all m.)

Let us start with $2j = 4$. The measure $\nu_n^{(4)}$ is basically the joint distribution of $g_3(\varphi) = 2\int_0^{0.5} \varphi$ and $g_4(\varphi) = 2\int_{0.5}^1 \varphi$, when φ is distributed μ_n. These two are independent for even n, but not for odd n; this is the problem. The solution: $\nu_{2m}^{(4)}$ and $\nu_{2m+1}^{(4)}$ are close enough.

5e2 Lemma. $(\nu_{2m}^{(4)})_m$ and $(\nu_{2m+1}^{(4)})_m$ come together.

Proof. Basically, $\nu_{2m}^{(4)}$ is the joint distribution of $(\beta_1 + \dots + \beta_m)/m$ and $(\beta_{m+1} + \dots + \beta_{2m})/m$, where $(\beta_1, \dots, \beta_{2m}) \in \{0,1\}^{2m}$ is distributed uniformly. Similarly, $\nu_{2m+1}^{(4)}$ is the joint distribution of $(\beta_1 + \dots + \beta_m + 0.5\beta_{m+1})/(m+0.5)$

and $(0.5\beta_{m+1}+\beta_{m+2}+\dots+\beta_{2m+1})/(m+0.5)$. We construct a joining λ_m of $\nu^{(4)}_{2m}$ and $\nu^{(4)}_{2m+1}$ as the joint distribution of two pairs,

$$\Big(\frac{\beta_1+\dots+\beta_m}{m},\frac{\beta_{m+2}+\dots+\beta_{2m+1}}{m}\Big) \quad \text{and} \\ \Big(\frac{\beta_1+\dots+\beta_m+0.5\beta_{m+1}}{m+0.5},\frac{0.5\beta_{m+1}+\beta_{m+2}+\dots+\beta_{2m+1}}{m+0.5}\Big)\,;$$

of course, $(\beta_1,\dots,\beta_{2m+1})$ is distributed uniformly on $\{0,1\}^{2m+1}$. We estimate the distance between the two pairs:

$$\Big|\frac{\beta_1+\dots+\beta_m}{m}-\frac{\beta_1+\dots+\beta_m+0.5\beta_{m+1}}{m+0.5}\Big| \le \\ \le m\Big(\frac{1}{m}-\frac{1}{m+0.5}\Big)+\frac{0.5}{m+0.5}=\frac{1}{m+0.5}\to 0\,;$$

the same holds for the second coordinate. □

By 5d1, $\|f\|_{L_{2m}(\nu^{(4)}_{2m+1})}$ behaves similarly to $\|f\|_{L_{2m}(\nu^{(4)}_{2m})}$, namely, converges to $\max\big(|f|\mathrm{e}^{-I_4}\big)$. The same holds for $\|f\|_{L_{2m+1}(\nu^{(4)}_{2m+1})}$, since $\frac{2m+1}{2m}\to 1$ (recall the argument of 2a17). Thus, $\|f\|_{L_n(\nu^{(4)}_n)}\to\max\big(|f|\mathrm{e}^{-I_4}\big)$.

Similarly, for every $j\in\{2,4,8,\dots\}$ and every $k\in\{0,1,\dots,j-1\}$, $\big(\nu^{(2j)}_{jm+k}\big)_m$ and $\big(\nu^{(2j)}_{j(m-1)}\big)_m$ come together, which implies (5e1) and completes the proof of Theorem 5a9.

References

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