2 Large deviations, Gibbs measures

Theorem 1b4 is proved via the large deviations theory; a formula for $\varphi = [g|f]$ is given in terms of Gibbs measures.

2a Who needs ridiculously small probabilities?

2a1 Example. A fair coin is tossed 200 times. The probability of 10 "heads" is

$$2^{-200}\binom{200}{10} \approx 1.4 \cdot 10^{-44}\,.$$

We are pretty sure that this event will not occur in practice. Then, does it matter, is it 10^{-44}, or 10^{-40}, or 10^{-50}?

For coin tossing it does not matter, but for statistical physics it does!

2a2 Example. Consider a system of 200 spins[1] $\omega_1, \dots, \omega_{200} = \pm 1$ with a one-particle Hamiltonian $h(\omega_1) = \omega_1$ and a given energy per particle

$$h^{(200)}(\omega_1, \dots, \omega_{200}) = -0.9 \quad \in (-1, 1)\,.$$

(Quite feasible.) Now, only 10 out of the 200 spins are +1, others are −1.

[1]So-called spin-$\frac{1}{2}$ particles.

You may ask: so what? We still do not need 10^{-44}. Right; but see the next example.

2a3 Example. Consider a system of n spins,[1] each taking on three values $-1, 0, +1$, described by $\Omega^n = \{-1, 0, 1\}^n$ (with the counting measure) and the one-particle Hamiltonian $h(\omega_1) = \omega_1$. Introduce another macroscopic observable $f^{(n)}$ where $f(\omega_1) = \omega_1^2$. We want to know the conditional distribution (according to 1b1) of $f^{(n)}$ given $h^{(n)} \approx -0.9$. This question is physically meaningful for quite large n (much more than 200, in fact, such as 10^{23}). Probabilistically, we want to know the conditional distribution, given the condition of exponentially small probability (much smaller than 10^{-44}). Such conditional probabilities are ratios of ridiculously small unconditional probabilities. . .[2]

2b Is the normal approximation helpful?

Consider for now $\Omega^n = \{-1, +1\}^n$ with the uniform distribution (the counting measure normalized by dividing by 2^n). The random variable $(\omega_1 + \dots + \omega_n)/\sqrt{n}$ is approximately normal standard (γ^1),

$$\mathbb{P}\Big(\frac{\omega_1 + \dots + \omega_n}{\sqrt{n}} = x\Big) \approx \frac{2}{\sqrt{n}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} \quad \text{for } x = \frac{-n}{\sqrt{n}}, \frac{-n+2}{\sqrt{n}}, \dots, \frac{n}{\sqrt{n}} \,.$$

Taking for example $n = 200$ and $x = \frac{-200+20}{\sqrt{200}} \approx -12.7$ we get

$$\mathbb{P}\Big(\frac{\omega_1 + \dots + \omega_{200}}{200} = -0.9\Big) \approx 3.7 \cdot 10^{-37}$$

instead of $1.4 \cdot 10^{-44}$. Quite bad!

Replacing -0.9 with -0.6 we get the normal approximation $1.3 \cdot 10^{-17}$ ($x \approx -8.5$) to the probability $1.3 \cdot 10^{-18}$. For -0.3: approximation $7.0 \cdot 10^{-6}$ ($x \approx -4.2$) to $6.3 \cdot 10^{-6}$. And for -0.15: approximation 0.00595 ($x \approx -2.1$) to 0.00596.

In fact, the normal approximation has a small *relative* error when[3] $x^4 \ll n$. For the event $(\omega_1 + \dots + \omega_n)/n = a$ we have $x = a\sqrt{n}$, thus, $x^4 \ll n$ when $a^4 \ll 1/n$. Rather good for coin tossing, but far not enough for statistical physics.

---

[1] So-called spin-1 particles.

[2] See 2j1 for the answer.

[3] Due to symmetry; for asymmetric distributions, $x^6 \ll n$ is required.

2c A non-normal approximation

For binomial probabilities, the normal approximation results from the Stirling formula

$$n! \sim \sqrt{2\pi n}\, n^n \mathrm{e}^{-n} \quad \text{as } n \to \infty\,.$$

In fact, straightforward application of Stirling formula leads to

$$\text{(2c1)}\quad \mathbb{P}\Big(\frac{\omega_1 + \dots + \omega_n}{n} = a \Big) \approx \\ \approx \frac{1}{\sqrt{1-a^2}} \frac{2}{\sqrt{n}} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{n}{2} \big((1-a)\ln(1-a) + (1+a)\ln(1+a) \big) \Big)$$

for $a = \frac{-n}{n}, \frac{-n+2}{n}, \dots, \frac{n}{n}$ (check it); the relative error is small when $(1-|a|)n$ is large. For example,

$$\mathbb{P}\Big(\frac{\omega_1 + \dots + \omega_{200}}{200} = -0.9 \Big) \approx 1.409 \cdot 10^{-44}$$

instead of $1.397 \cdot 10^{-44}$; quite good.

For small a we have

$$(1-a)\ln(1-a) + (1+a)\ln(1+a) = a^2 + O(a^4)$$

(check it), which gives the normal approximation if $na^4 \ll 1$ (check it). For larger a the non-normal approximation (2c1) works.

2d Large deviations: upper bound

Let Ω and μ be as in Sect. 1a,[1] and $f : \Omega \to \mathbb{R}$ a measurable function. For every $\lambda \in [0,\infty)$ and $a \in \mathbb{R}$,

$$\int \mathrm{e}^{\lambda f}\, \mathrm{d}\mu \ge \mathrm{e}^{\lambda a} \mu\{f \ge a\}$$

(think, why); of course, $\{f \ge a\}$ means $\{\omega : f(\omega) \ge a\}$. Thus,

$$\mu\{f \ge a\} \le \inf_{\lambda \ge 0} \frac{\int \mathrm{e}^{\lambda f}\, \mathrm{d}\mu}{\mathrm{e}^{\lambda a}}\,.$$

Similarly,

$$\mu\{f \le a\} \le \inf_{\lambda \le 0} \frac{\int \mathrm{e}^{\lambda f}\, \mathrm{d}\mu}{\mathrm{e}^{\lambda a}}\,.$$

[1] That is, μ is a finite or σ-finite positive measure on Ω.

We apply it to $f^{(n)}$, taking into account that

$$\int \mathrm{e}^{\lambda n f^{(n)}}\,\mathrm{d}\mu^n = \Big(\int \mathrm{e}^{\lambda f}\,\mathrm{d}\mu \Big)^n ,$$

and get

$$\mu^n\{f^{(n)} \ge a\} \le \bigg(\inf_{\lambda\ge 0} \frac{\int \mathrm{e}^{\lambda f}\,\mathrm{d}\mu}{\mathrm{e}^{\lambda a}} \bigg)^n , \quad \mu^n\{f^{(n)} \le a\} \le \bigg(\inf_{\lambda\le 0} \frac{\int \mathrm{e}^{\lambda f}\,\mathrm{d}\mu}{\mathrm{e}^{\lambda a}} \bigg)^n .$$

2d1 Exercise. The function $\Lambda : \mathbb{R} \to (-\infty, +\infty]$ defined by

$$\Lambda(\lambda) = \ln \int \mathrm{e}^{\lambda f}\,\mathrm{d}\mu$$

is convex and lower semicontinuous.

Prove it.

2d2 Exercise. (a) The set $\{\Lambda < \infty\} = \{\lambda : \Lambda(\lambda) < \infty\}$ is an interval.

(b) Every interval can appear this way: (a,b), $[a,b]$, $[a,b)$, $(a,b]$, $(-\infty,b)$, $(-\infty,b]$, (a,∞), $[a,\infty)$, $(-\infty,\infty)$, $\{a\}$ and $\emptyset$.

(c) The restriction of Λ to $\{\Lambda < \infty\}$ is continuous.

(d) The restriction of Λ to $\{\Lambda < \infty\}$ is strictly convex, unless $f = \mathrm{const}$. Prove it.

2d3 Exercise. If $\Lambda(\lambda-\varepsilon) < \infty$ and $\Lambda(\lambda+\varepsilon) < \infty$ then

$$\mathrm{e}^{\Lambda(\lambda\pm\varepsilon)} = \sum_{n=0}^{\infty} \frac{(\pm\varepsilon)^n}{n!} \int f^n \mathrm{e}^{-\lambda f}\,\mathrm{d}\mu .$$

Prove it.

Thus, Λ is infinitely differentiable on $\mathrm{Int}\{\Lambda < \infty\}$ (the interior of the interval), and

$$\frac{\mathrm{d}^n}{\mathrm{d}\lambda^n} \mathrm{e}^{\Lambda(\lambda)} = \int f^n \mathrm{e}^{\lambda f}\,\mathrm{d}\mu \quad \text{for } \lambda \in \mathrm{Int}\{\Lambda < \infty\} . \tag{2d4}$$

In terms of Λ we have

$$\mu^n\{f^{(n)} \ge a\} \le \exp n \inf_{\lambda\ge 0} \big(\Lambda(\lambda) - a\lambda\big) , \quad \mu^n\{f^{(n)} \le a\} \le \exp n \inf_{\lambda\le 0} \big(\Lambda(\lambda) - a\lambda\big) .$$

If $a = \Lambda'(\lambda)$ for some $\lambda \in \mathrm{Int}\{\Lambda < \infty\}$ then $\inf_{\lambda_1\in\mathbb{R}}(\Lambda(\lambda_1) - a\lambda_1) = \Lambda(\lambda) - a\lambda = \Lambda(\lambda) - \lambda\Lambda'(\lambda)$ (think, why); thus,

(2d5)

$$\mu^n\{f^{(n)} \ge \Lambda'(\lambda)\} \le \exp n \big(\Lambda(\lambda) - \lambda\Lambda'(\lambda)\big) \quad \text{for } \lambda \in [0,\infty) \cap \mathrm{Int}\{\Lambda < \infty\} ,$$
$$\mu^n\{f^{(n)} \le \Lambda'(\lambda)\} \le \exp n \big(\Lambda(\lambda) - \lambda\Lambda'(\lambda)\big) \quad \text{for } \lambda \in (-\infty,0] \cap \mathrm{Int}\{\Lambda < \infty\} .$$

2d6 Exercise. Consider again $\Omega = \{-1, 1\}$ with the counting measure, and $f(\omega_1) = \omega_1$. Calculate Λ and simplify the inequalities, getting

$$\mu^n\{f^{(n)} \ge a\} \le 2^n \exp\Big(-\frac{n}{2}\big((1-a)\ln(1-a) + (1+a)\ln(1+a)\big)\Big) \text{ for } a \in [0, 1)\,,$$
$$\mu^n\{f^{(n)} \le a\} \le 2^n \exp\Big(-\frac{n}{2}\big((1-a)\ln(1-a) + (1+a)\ln(1+a)\big)\Big) \text{ for } a \in (-1, 0]\,.$$

Does it hold for $a = \pm 1$?

Compare it with (2c1).

2e Better bound via tilting

Let Ω, μ, f and Λ be as in Sect. 2d, $\operatorname{Int}\{\Lambda < \infty\} \ne \emptyset$, and $f \ne \text{const}$.[1]

Given $\lambda \in \operatorname{Int}\{\Lambda < \infty\}$, we introduce a measure well-known in the large deviations theory as tilted measure, in mathematics and physics as Gibbs measure, and in statistical physics as canonical ensemble:

$$\nu = \exp\big(\lambda f - \Lambda(\lambda)\big) \cdot \mu\,.$$

2e1 Exercise. Check that

$$\int d\nu = 1\,, \quad \int f\, d\nu = \Lambda'(\lambda)\,, \quad \int f^2\, d\nu = \Lambda''(\lambda) + \Lambda'^2(\lambda)\,,$$
$$\ln \int e^{\alpha f}\, d\nu = \Lambda(\lambda + \alpha) - \Lambda(\lambda) \quad \text{for } \alpha \in \mathbb{R}\,.$$

2e2 Exercise. If $\mu = \gamma^1$ (that is, $N(0, 1)$) and $f(x) = x$, then
(a) $\Lambda(\lambda) = \lambda^2/2$;
(b) ν is γ^1 shifted by λ (that is, $N(\lambda, 1)$);
(c) $\int_0^\infty e^{\lambda x}\, \gamma^1(dx) = e^{\lambda^2/2} \gamma^1\big([-\lambda, \infty)\big)$ for $\lambda \in \mathbb{R}$;
(d) $\int_0^\infty e^{-\lambda x}\, \gamma^1(dx) \le \min\big(\frac{1}{2}, \frac{1}{\sqrt{2\pi}\,\lambda}\big)$ for $\lambda > 0$.
Prove it.

W.r.t. the probability measure ν the function f may be thought of as a random variable with $\mathbb{E}\, f = \Lambda'(\lambda)$ and $\operatorname{Var} f = \Lambda''(\lambda) > 0$ (just because $f \ne \text{const}$). We note this fact also for subsequent use:

(2e3) $\quad \Lambda''(\lambda) > 0 \quad \text{for all } \lambda \in \operatorname{Int}\{\Lambda < \infty\} \quad \text{provided that } f \ne \text{const}\,.$

2e4 Exercise. Check that

$$\nu^n = \exp n\big(\lambda f^{(n)} - \Lambda(\lambda)\big) \cdot \mu^n\,.$$

[1] As before, it means $\operatorname{ess\,inf} f \ne \operatorname{ess\,sup} f$.

W.r.t. the probability measure ν^n the function $nf^{(n)}$ may be thought of as the sum of n independent copies of the random variable f;

$$\mathbb{E}\, f^{(n)} = \Lambda'(\lambda)\,, \quad \operatorname{Var} f^{(n)} = \frac{\Lambda''(\lambda)}{n}\,.$$

2e5 Exercise. For every $\varepsilon > 0$ there exists $\delta > 0$ such that for all n,

$$\nu^n\{|f^{(n)} - \Lambda'(\lambda)| > \varepsilon\} \le \mathrm{e}^{-\delta n}\,.$$

Prove it.

By the central limit theorem,

$$\nu^n\Big\{a \le \frac{f^{(n)} - \Lambda'(\lambda)}{\sqrt{\Lambda''(\lambda)/n}} \le b\Big\} \to \gamma^1\big([a,b]\big) \quad \text{as } n \to \infty$$

whenever $-\infty \le a < b \le \infty$. In terms of μ^n,

$$\int \mathbb{1}_{J_n}(f^{(n)}) \mathrm{e}^{n(\lambda f^{(n)} - \Lambda(\lambda))}\, \mathrm{d}\mu^n \to \gamma^1\big([a,b]\big)\,,$$

where $J_n = [\Lambda'(\lambda) + a\sqrt{\Lambda''(\lambda)/n}, \Lambda'(\lambda) + b\sqrt{\Lambda''(\lambda)/n}]$. Also,

(2e6)
$$\mu^n\{f^{(n)} \in J_n\} = \int \mathbb{1}_{J_n}(f^{(n)})\, \mathrm{d}\mu^n = \int \mathbb{1}_{J_n}(f^{(n)}) \exp n(\Lambda(\lambda) - \lambda f^{(n)})\, \mathrm{d}\nu^n =$$
$$= \mathrm{e}^{n\Lambda(\lambda)} \mathbb{E}\, \mathrm{e}^{-\lambda n f^{(n)}} \mathbb{1}_{J_n}(f^{(n)}) = \exp n(\Lambda(\lambda) - \lambda\Lambda'(\lambda)) \mathbb{E}\, \mathrm{e}^{-\lambda n(f^{(n)} - \Lambda'(\lambda))} \mathbb{1}_{J_n}(f^{(n)})\,.$$

In particular,

$$\mu^n\Big\{f^{(n)} \ge \Lambda'(\lambda)\Big\} = \exp n\big(\Lambda(\lambda) - \lambda\Lambda'(\lambda)\big) \mathbb{E}\, \mathrm{e}^{-\lambda n(f^{(n)} - \Lambda'(\lambda))} \mathbb{1}_{[0.\infty)}(f^{(n)} - \Lambda'(\lambda))\,,$$
$$\mu^n\Big\{f^{(n)} \le \Lambda'(\lambda)\Big\} = \exp n\big(\Lambda(\lambda) - \lambda\Lambda'(\lambda)\big) \mathbb{E}\, \mathrm{e}^{-\lambda n(f^{(n)} - \Lambda'(\lambda))} \mathbb{1}_{(-\infty,0]}(f^{(n)} - \Lambda'(\lambda))\,.$$

We reduce the expectation to probabilities:

$$\mathrm{e}^{-\lambda n x} \mathbb{1}_{[0,\infty)}(x) = \int_0^\infty \mathbb{1}_{[0,y]}(x) \lambda n \mathrm{e}^{-\lambda n y}\, \mathrm{d}y$$

(check it), therefore

$$\mathbb{E}\, \mathrm{e}^{-\lambda n(f^{(n)} - \Lambda'(\lambda))} \mathbb{1}_{[0,\infty)}(f^{(n)} - \Lambda'(\lambda)) = \int_0^\infty \nu_n\{0 \le f^{(n)} - \Lambda'(\lambda) \le y\} \lambda n \mathrm{e}^{-\lambda n y}\, \mathrm{d}y$$

(check it). It follows that

$$\left| \mathbb{E}\, \mathrm{e}^{-\lambda n (f^{(n)} - \Lambda'(\lambda))} 1\!\!1_{[0.\infty)}(f^{(n)} - \Lambda'(\lambda)) - \int_0^\infty \mathrm{e}^{-\lambda n \sqrt{\Lambda''(\lambda)/n}\, x} \gamma^1(\mathrm{d}x) \right| \le \varepsilon_n \, ,$$

where

$$\varepsilon_n = \sup_{x \ge 0} \left| \nu^n \Big\{ 0 \le f^{(n)} - \Lambda'(\lambda) \le \sqrt{\frac{\Lambda''(\lambda)}{n}}\, x \Big\} - \gamma^1([0,x]) \right| .$$

Using 2e2(d), for $\lambda > 0$,

$$\mu^n\{f^{(n)} \ge \Lambda'(\lambda)\} \le \exp n\big(\Lambda(\lambda) - \lambda\Lambda'(\lambda)\big) \cdot \left(\varepsilon_n + \min\Big(\frac12, \frac{1}{\sqrt{2\pi}\lambda\sqrt{n\Lambda''(\lambda)}} \Big) \right) .$$

The central limit theorem ensures[1] that $\varepsilon_n \to 0$, which gives

$$\mu^n\{f^{(n)} \ge \Lambda'(\lambda)\} = \exp n\big(\Lambda(\lambda) - \lambda\Lambda'(\lambda)\big) \cdot o(1) \quad \text{as } n \to \infty$$

for $\lambda \in (0,\infty) \cap \operatorname{Int}\{\Lambda < \infty\}$; and similarly,

$$\mu^n\{f^{(n)} \le \Lambda'(\lambda)\} = \exp n\big(\Lambda(\lambda) - \lambda\Lambda'(\lambda)\big) \cdot o(1) \quad \text{as } n \to \infty$$

for $\lambda \in (-\infty,0) \cap \operatorname{Int}\{\Lambda < \infty\}$.

However, a stronger result,

$$\varepsilon_n \le \frac{C}{\sqrt{n}}$$

is ensured by the Berry-Esseen theorem,[2] provided that

$$\int |f|^3 \, \mathrm{d}\nu < \infty \, ,$$

which is not a problem here, since $\big(\int |f|^3 \, \mathrm{d}\nu\big)^{1/3} \le \big(\int f^4 \, \mathrm{d}\nu\big)^{1/4} < \infty$. The constant C depends on the distribution (thus, on λ) and does not depend on n. We get the following.

2e7 Theorem. [3] For every $\lambda \in \operatorname{Int}\{\Lambda < \infty\} \setminus \{0\}$ there exists $C < \infty$ such that for all n,

$$\mu^n\Big\{ f^{(n)} \ge \Lambda'(\lambda) \Big\} \le \frac{C}{\sqrt{n}} \exp n\big(\Lambda(\lambda) - \lambda\Lambda'(\lambda)\big)$$

[1] In combination with a monotonicity argument.

[2] See W. Feller, An introduction to probability theory and its applications, vol. II, Sect. XVI.5.

[3] See A.Dembo and O.Zeitouni, Large deviations techniques and applications, Theorem 3.7.4 (Bahadur and Rao).

if $\lambda > 0$, and

$$\mu^n\Big\{f^{(n)} \le \Lambda'(\lambda)\Big\} \le \frac{C}{\sqrt{n}} \exp n\big(\Lambda(\lambda) - \lambda\Lambda'(\lambda)\big)$$

if $\lambda < 0$.

Compare it (again) with (2c1) (using 2d6).

2f Large deviations: lower bound

Let Ω, μ, f and Λ be as in Sect. 2e.

Consider intervals $J_n = [\Lambda'(\lambda) + a\sqrt{\Lambda''(\lambda)}/n, \Lambda'(\lambda) + b\sqrt{\Lambda''(\lambda)}/n]$ (of length $O(1/n)$, unlike Sect. 2e) for a given $\lambda \in \operatorname{Int}\{\Lambda < \infty\}$. We have

$$\Big|\nu^n\{f^{(n)} \in J_n\} - \gamma^1\Big(\Big[\frac{a}{\sqrt{n}}, \frac{b}{\sqrt{n}}\Big]\Big)\Big| \le \varepsilon_n \le \frac{C}{\sqrt{n}}$$

and $\gamma^1\big([\frac{a}{\sqrt{n}}, \frac{b}{\sqrt{n}}]\big) \sim \frac{1}{\sqrt{2\pi}} \frac{b-a}{\sqrt{n}}$ (think, why). Thus, for large n,

$$\nu^n\{f^{(n)} \in J_n\} \ge \frac{c}{\sqrt{n}}$$

provided that $\frac{b-a}{\sqrt{2\pi}} - C > c > 0$. By (2e6),

$$\mu^n\{f^{(n)} \in J_n\} = \exp n(\Lambda(\lambda) - \lambda\Lambda'(\lambda)) \mathbb{E}\, \mathrm{e}^{-\lambda n(f^{(n)} - \Lambda'(\lambda))} \mathbb{1}_{J_n}(f^{(n)})\,.$$

In particular, for $\lambda \ge 0$,

$$\begin{aligned} \mu^n\{0 \le f^{(n)} - \Lambda'(\lambda) \le b\sqrt{\Lambda''(\lambda)}/n\} \ge \\ \ge \mathrm{e}^{n(\Lambda(\lambda) - \lambda\Lambda'(\lambda))} \cdot \mathrm{e}^{-\lambda n b\sqrt{\Lambda''(\lambda)}/n} \nu^n\{f^{(n)} \in J_n\} \ge \\ \ge \mathrm{e}^{n(\Lambda(\lambda) - \lambda\Lambda'(\lambda))} \cdot \mathrm{e}^{-\lambda b\sqrt{\Lambda''(\lambda)}} \cdot \frac{c}{\sqrt{n}} \end{aligned}$$

provided that $\frac{b}{\sqrt{2\pi}} - C > c > 0$. Having C (dependent on the distribution, thus, on λ) we take $b > C\sqrt{2\pi}$ and get the following (the case $\lambda \le 0$ being similar).

2f1 Theorem. [1] For every $\lambda \in \operatorname{Int}\{\Lambda < \infty\}$ there exist $C < \infty$ and $c > 0$ such that for all n large enough,

$$\mu^n\Big\{\Lambda'(\lambda) \le f^{(n)} \le \Lambda'(\lambda) + \frac{C}{n}\Big\} \ge \frac{c}{\sqrt{n}} \exp n\big(\Lambda(\lambda) - \lambda\Lambda'(\lambda)\big) \quad \text{if } \lambda \ge 0\,,$$

$$\mu^n\Big\{\Lambda'(\lambda) - \frac{C}{n} \le f^{(n)} \le \Lambda'(\lambda)\Big\} \ge \frac{c}{\sqrt{n}} \exp n\big(\Lambda(\lambda) - \lambda\Lambda'(\lambda)\big) \quad \text{if } \lambda \le 0\,.$$

Compare it (once again) with (2c1) (using 2d6).

[1] See again the Bahadur-Rao theorem (footnote 3 on page 16).

2g Higher dimension: upper bound

In most cases it is enough to know that a measure is $\exp(-nI + o(n))$ for some known I. Of course, $\frac{1}{\sqrt{n}} = \exp o(n)$, and $n^\alpha = \exp o(n)$ for every α, and even $\mathrm{e}^{\pm\sqrt{n}} = \exp o(n)$.

Dimension 2 is treated here; other finite dimensions can be treated similarly.

Let Ω and μ be as before, and $f : \Omega \to \mathbb{R}^2$ a measurable function. We introduce $\Lambda : \mathbb{R}^2 \to (-\infty, \infty]$ by

$$\Lambda(\lambda) = \ln \int \mathrm{e}^{\langle \lambda, f \rangle} \, \mathrm{d}\mu$$

where $\langle \cdot, \cdot \rangle$ is the scalar product in $\mathbb{R}^2$.

2g1 Exercise. [1] Prove that Λ is convex and lower semicontinuous, and the set $\{\Lambda < \infty\} \subset \mathbb{R}^2$ is convex.

2g2 Exercise. Λ is infinitely differentiable on $\mathrm{Int}\{\Lambda < \infty\}$, and

$$\frac{\partial^{k+l}}{\partial \lambda_1^k \partial \lambda_2^l} \mathrm{e}^{\Lambda(\lambda)} = \int f_1^k f_2^l \mathrm{e}^{\langle \lambda, f \rangle} \, \mathrm{d}\mu$$

for $\lambda = (\lambda_1, \lambda_2) \in \mathrm{Int}\{\Lambda < \infty\}$; here $f(\omega) = (f_1(\omega), f_2(\omega))$.

Prove it.

We assume that $\{\Lambda < \infty\} \neq \emptyset$.

We introduce the so-called Fenchel-Legendre transform $\Lambda^* : \mathbb{R}^2 \to (-\infty, \infty]$ of Λ by

$$\Lambda^*(a) = \sup_{\lambda \in \mathbb{R}^2} \big(\langle \lambda, a \rangle - \Lambda(\lambda) \big) \, .$$

Being a supremum of linear functions, Λ^* is convex and lower semicontinuous.

Functions $f^{(n)} : \Omega^n \to \mathbb{R}^2$ are defined as before; and still,

$$\int \exp n \langle \lambda, f^{(n)} \rangle \, \mathrm{d}\mu^n = \bigg(\int \exp \langle \lambda, f \rangle \, \mathrm{d}\mu \bigg)^n = \exp n\Lambda(\lambda) \, .$$

2g3 Exercise. Prove that $\mu^n\{f^{(n)} \in B\} \le \exp n\big(\Lambda(\lambda) - \inf_{x \in B} \langle \lambda, x \rangle\big)$ for every n, Borel set $B \subset \mathbb{R}^2$, and $\lambda \in \mathbb{R}^2$.

Denote $B_\delta(a) = \{x \in \mathbb{R}^2 : \|x - a\| < \delta\}$.

[1] Unlike 2d2(c), the restriction of Λ to $\{\Lambda < \infty\}$ need not be continuous. A sketch of a counterexample: it may happen that $\Lambda(\cos\varphi, \sin\varphi) = \sum_n \exp c_n(\cos(\varphi - \alpha_n) - 1)$; try $c_n = n^3$, $\alpha_n = 1/n$.

2g4 Exercise. Prove that for every $\lambda \in \mathbb{R}^2$, $\delta > 0$ and $a \in \mathbb{R}^2$,

$$\mu^n\{f^{(n)} \in B_\delta(a)\} \le \exp n\big(\Lambda(\lambda) - \langle \lambda, a\rangle + \delta\|\lambda\|\big)\,.$$

2g5 Exercise. For every $a \in \mathbb{R}^2$ and $C < \Lambda^*(a)$ there exists $\delta > 0$ such that for all n,

$$\mu^n\{f^{(n)} \in B_\delta(a)\} \le \mathrm{e}^{-nC}\,.$$

2g6 Exercise. For every compact set $K \subset \mathbb{R}^2$,

$$\mu^n\{f^{(n)} \in K\} \le \exp\Big(- n \min_K \Lambda^* + o(n) \Big)\,.$$

Prove it.

2h Higher dimension: lower bound

Let Ω, μ, f and Λ be as in Sect. 2g, and $\operatorname{Int}\{\Lambda < \infty\} \ne \emptyset$.

Given $\lambda \in \operatorname{Int}\{\Lambda < \infty\}$, we introduce the tilted measure

$$\nu = \exp\big(\langle \lambda, f\rangle - \Lambda(\lambda)\big) \cdot \mu\,.$$

2h1 Exercise. Check that

$$\int \mathrm{d}\nu = 1\,, \qquad \int f\,\mathrm{d}\nu = \operatorname{grad}\Lambda(\lambda)$$

and

$$\nu^n = \exp n\big(\langle \lambda, f^{(n)}\rangle - \Lambda(\lambda)\big) \cdot \mu^n\,.$$

As before, $\mathbb{E}\, f^{(n)} = \operatorname{grad}\Lambda(\lambda)$.

2h2 Exercise. Prove that for every $\delta > 0$,

$$\nu^n\{\|f^{(n)} - \operatorname{grad}\Lambda(\lambda)\| < \delta\} \to 1 \quad \text{as } n \to \infty\,.$$

2h3 Exercise. Prove that

$$\mu^n\{f^{(n)} \in B\} \ge \exp n\big(\Lambda(\lambda) - \sup_{x\in B}\langle \lambda, x\rangle\big)\nu^n\{f^{(n)} \in B\}$$

for every n, Borel set $B \subset \mathbb{R}^2$, and $\lambda \in \mathbb{R}^2$.

2h4 Exercise. Prove that for every $\lambda \in \operatorname{Int}\{\Lambda < \infty\}$ and $\delta > 0$,

$$\mu^n\{\|f^{(n)} - \operatorname{grad}\Lambda(\lambda)\| < \delta\} \ge (1-o(1)) \exp n\big(\Lambda(\lambda) - \langle \lambda, \operatorname{grad}\Lambda(\lambda)\rangle - \delta\|\lambda\|\big)\,.$$

2h5 Exercise. Let $G \subset \mathbb{R}^2$ be an open set. Prove that[1]

$$\mu^n\{f^{(n)} \in G\} \ge \exp\bigl(n \sup(\Lambda(\lambda) - \langle \lambda, \operatorname{grad}\Lambda(\lambda)\rangle) - o(n)\bigr)$$

where the supremum is taken over all $\lambda \in \operatorname{Int}\{\Lambda < \infty\}$ such that $\operatorname{grad}\Lambda(\lambda) \in G$.

2h6 Exercise. Prove that

$$\Lambda^*\bigl(\operatorname{grad}\Lambda(\lambda)\bigr) = \langle \lambda, \operatorname{grad}\Lambda(\lambda)\rangle - \Lambda(\lambda)$$

for every $\lambda \in \operatorname{Int}\{\Lambda < \infty\}$.

We define a set $T \subset \mathbb{R}^2$ by[2]

$$T = \bigl\{ \operatorname{grad}\Lambda(\lambda) : \lambda \in \operatorname{Int}\{\Lambda < \infty\} \bigr\} .$$

2h7 Exercise. Let $G \subset \mathbb{R}^2$ be an open set. Prove that

$$\mu^n\{f^{(n)} \in G\} \ge \exp\Bigl(-n \inf_{G\cap T} \Lambda^* - o(n) \Bigr) .$$

The same holds in $\mathbb{R}^n$ for all $n = 1, 2, 3, \dots$

2i Using conditions (1a1), (1a2)

We return to dimension one. Let Ω, μ, f and Λ be as in Sect. 2d, and $f \ne \text{const}$.

Assuming (1a1) or (1a2) we'll prove the so-called weak[3] large deviations principle (LDP) with the rate function Λ^*. It means the upper bound

$$\mu^n\{f^{(n)} \in K\} \le \exp\Bigl(-n \min_K \Lambda^* + o(n) \Bigr) \quad \text{for compact } K \subset \mathbb{R}$$

together with the lower bound

$$\mu^n\{f^{(n)} \in G\} \ge \exp\Bigl(-n \inf_G \Lambda^* - o(n) \Bigr) \quad \text{for open } G \subset \mathbb{R} .$$

To this end, by 2g6 and 2h7 (for dimension one) it suffices to show that

$$T = \operatorname{Int}\{\Lambda^* < \infty\} ,$$

[1] We admit that $o(n)$ may be infinite for a finite number of numbers n.
[2] "T" reminds of "tilting".
[3] "Weak" since the upper bound is stated for compact sets rather than all closed sets.

where $T = \big\{\Lambda'(\lambda) : \lambda \in \operatorname{Int}\{\Lambda < \infty\}\big\}$. The convex set $\{\Lambda^* < \infty\}$ need not be open, but anyway, the restriction of Λ^* to this set is continuous (due to convexity and lower semicontinuity); thus, $\inf_G \Lambda^* = \inf_{G\cap\{\Lambda^*<\infty\}} \Lambda^* = \inf_{G\cap\operatorname{Int}\{\Lambda^*<\infty\}} \Lambda^* = \inf_{G\cap T} \Lambda^*$.

Assume (1a1): $\mu(\Omega) < \infty$ and $\int e^{\lambda f}\, d\mu < \infty$ for all $\lambda \in \mathbb{R}$. Thus, $\operatorname{Int}\{\Lambda < \infty\} = \mathbb{R}$, and Λ' is strictly increasing (recall (2e3)), which implies existence of limits,

$$-\infty \le \Lambda'(-\infty) < \Lambda'(+\infty) \le +\infty\,.$$

2i1 Exercise. Prove that

$$T = \big(\Lambda'(-\infty), \Lambda'(+\infty)\big) \subset \{\Lambda^* < \infty\} \subset [\Lambda'(-\infty), \Lambda'(+\infty)]\,.$$

Thus, $T = \operatorname{Int}\{\Lambda^* < \infty\}$, and the LDP under (1a1) is verified. In addition, a relation to the bounds of f is shown below.

2i2 Exercise. Prove that

$$\frac{\Lambda(\lambda)}{\lambda} \to \Lambda'(-\infty) \text{ as } \lambda \to -\infty\,, \quad \text{and} \quad \frac{\Lambda(\lambda)}{\lambda} \to \Lambda'(+\infty) \text{ as } \lambda \to +\infty\,.$$

2i3 Exercise. Prove that

$$\operatorname{ess\,inf} f \le \Lambda'(-\infty) < \Lambda'(+\infty) \le \operatorname{ess\,sup} f\,.$$

If $a < \operatorname{ess\,sup} f$ then $\mu\{f \ge a\} > 0$; we have $\int e^{\lambda f}\, d\mu \ge e^{\lambda a}\mu\{f \ge a\}$ for $\lambda > 0$, thus $\Lambda(\lambda) \ge \lambda a + \ln \mu\{f \ge a\}$ and $\Lambda'(+\infty) \ge a$, which shows that $\Lambda'(+\infty) \ge \operatorname{ess\,sup} f$. Similarly, $\Lambda'(-\infty) \le \operatorname{ess\,inf} f$. We get under (1a1)

$$-\infty \le \operatorname{ess\,inf} f = \Lambda'(-\infty) < \Lambda'(+\infty) = \operatorname{ess\,sup} f \le +\infty\,.$$

Assume now (1a2): $\mu(\Omega) = \infty$ and $\int e^{\lambda f}\, d\mu < \infty$ for all $\lambda \in (-\infty, 0)$. Then $\{\Lambda < \infty\} = (-\infty, 0)$, since this convex set does not contain 0. We have

$$-\infty \le \Lambda'(-\infty) < \Lambda'(0-) = +\infty$$

(since $\Lambda'(0-) < \infty$ would imply $\Lambda(0) < \infty$).

2i4 Exercise. Prove that

$$T = \big(\Lambda'(-\infty), +\infty\big) \subset \{\Lambda^* < \infty\} \subset [\Lambda'(-\infty), +\infty)\,.$$

Also $\operatorname{ess\,sup} f = +\infty$ (since $\int e^{-\operatorname{ess\,sup} f}\, d\mu \le \int e^{-f}\, d\mu < \infty$).

2i5 Exercise. Prove that

$$\Lambda'(-\infty) \ge \operatorname{ess\,inf} f \, .$$

2i6 Exercise. Prove that

$$\Lambda'(-\infty) \le \operatorname{ess\,inf} f \, .$$

We get under (1a2)

$$-\infty \le \operatorname{ess\,inf} f = \Lambda'(-\infty) < \Lambda'(0-) = \operatorname{ess\,sup} f = +\infty \, .$$

In both cases,

$$T = \operatorname{Int}\{\Lambda^* < \infty\} = (\operatorname{ess\,inf} f, \operatorname{ess\,sup} f) \, .$$

2j The function φ, at last

Let $f, g : \Omega \to \mathbb{R}$ and $\varepsilon > 0$ be such that $f - \varepsilon g$ and $f + \varepsilon g$ satisfy (1a1) or (1a2), and $f \ne \text{const}$ (as required in Theorem 1b4). We combine them into $h : \Omega \to \mathbb{R}^2$, $h(\cdot) = \big(f(\cdot), g(\cdot) \big)$ and consider the corresponding $\Lambda_h : \mathbb{R}^2 \to (-\infty, +\infty]$ as in Sect. 2g.

Case $\mu(\Omega) < \infty$: the convex set $\{\Lambda_h < \infty\}$ contains two lines $\{(\lambda, \pm\varepsilon\lambda) : \lambda \in \mathbb{R}\}$, therefore it is the whole $\mathbb{R}^2$, and surely, $\mathbb{R} = \{(\lambda, 0) : \lambda \in \mathbb{R}\} \subset \operatorname{Int}\{\Lambda_h < \infty\} \subset \mathbb{R}^2$.

Case $\mu(\Omega) = \infty$: the convex set $\{\Lambda_h < \infty\}$ contains two rays $\{(\lambda, \pm\varepsilon\lambda) : \lambda \in (-\infty, 0)\}$, therefore it contains the sector $\{(\lambda_1, \lambda_2) : \lambda_1 < 0, |\lambda_2| \le \varepsilon|\lambda_1|\}$, and does not contain the origin; and surely, $\{(\lambda, 0) : \lambda \in (-\infty, 0)\} \subset \operatorname{Int}\{\Lambda_h < \infty\}$.

We have also the function $\Lambda_f : \mathbb{R} \to (-\infty, +\infty]$ corresponding to f as in 2i, and we know that

$$T_f = \big\{ \Lambda'_f(\lambda) : \lambda \in \operatorname{Int}\{\Lambda_f < \infty\} \big\} = (\operatorname{ess\,inf} f, \operatorname{ess\,sup} f) \, ,$$

since f satisfies (1a1) or (1a2). And clearly, Λ_f is the restriction of Λ_h to the line $\mathbb{R} = \{(\lambda, 0) : \lambda \in \mathbb{R}\} \subset \mathbb{R}^2$. Of course, $\operatorname{Int}\{\Lambda_f < \infty\}$ is either $\mathbb{R}$ (if $\mu(\Omega) < \infty$) or $(-\infty, 0)$ (if $\mu(\Omega) = \infty$). In both cases $\operatorname{Int}\{\Lambda_f < \infty\} \subset \operatorname{Int}\{\Lambda_h < \infty\}$ under the embedding $\mathbb{R} \to \mathbb{R}^2$, $\lambda \mapsto (\lambda, 0)$.

The function Λ'_f is strictly increasing by (2e3) and maps $\operatorname{Int}\{\Lambda_f < \infty\}$ onto $T_f = (\operatorname{ess\,inf} f, \operatorname{ess\,sup} f)$. For every $a \in (\operatorname{ess\,inf} f, \operatorname{ess\,sup} f)$ there exists one and only one $\lambda \in \operatorname{Int}\{\Lambda_f < \infty\}$ such that $\Lambda'_f(\lambda) = a$. We have

$$\operatorname{grad} \Lambda_h(\lambda, 0) = \big(a, \varphi(a) \big)$$

for a continuous $\varphi : (\operatorname{ess\,inf} f, \operatorname{ess\,sup} f) \to \mathbb{R}$, namely,

$$\varphi(a) = \frac{\mathrm{d}}{\mathrm{d}\lambda_2}\Big|_{\lambda_2=0} \Lambda\big((\Lambda'_f)^{-1}(a), \lambda_2\big) \, ,$$

not only a continuous function, but also an infinitely differentiable function. In terms of the tilted measure

$$\nu = \exp\big(\lambda f - \Lambda_f(\lambda)\big) \cdot \mu$$

we have (by 2h1) $\int h \, \mathrm{d}\nu = \operatorname{grad} \Lambda_h(\lambda, 0)$, that is,

$$a = \int f \, \mathrm{d}\nu \quad \text{and} \quad \varphi(a) = \int g \, \mathrm{d}\nu \, ;$$

an equivalent definition of φ.

2j1 Example. Continuing Example 2a3, consider $\Omega = \{-1, 0, 1\}$, $h(\omega) = \omega$, $f(\omega) = \omega^2$. The function $\varphi = [f|h]$ can be written out implicitly:

$$\varphi\Big(\frac{x^2-1}{x^2+x+1}\Big) = \frac{x^2+1}{x^2+x+1} \quad \text{for } x \in (0, \infty) \, ;$$

the tilted measure is

$$\nu(\{-1\}) = \frac{1}{x^2+x+1} \, , \quad \nu(\{0\}) = \frac{x}{x^2+x+1} \, , \quad \nu(\{1\}) = \frac{x^2}{x^2+x+1} \, .$$

In particular, $x \approx 0.09289$ gives $\varphi(-0.9) \approx 0.9157$. Thus, the conditional distribution of $f^{(n)}$ given $h^{(n)} \approx -0.9$ is concentrated near 0.9157.

2k Proving the theorem

Let f, g and φ be as in Sect. 2j. In order to prove Theorem 1b4 we have to prove that $\mathbb{P}\big(g^{(n)} \in (c,d) \,\big|\, f^{(n)} \in [a,b]\big) \to 1$ whenever $[a,b] \subset (\operatorname{ess\,inf} f, \operatorname{ess\,sup} f)$ and $(c,d) \subset \mathbb{R}$ satisfy $\varphi\big([a,b]\big) \subset (c,d)$. We'll prove a bit stronger statement:

$$\frac{\mu^n\{f^{(n)} \in [a,b] \text{ and } g^{(n)} \notin (c,d)\}}{\mu^n\{f^{(n)} \in (a,b)\}} \to 0$$

(the denominator being non-zero for large n).

2k1 Exercise. If $\frac{\mu^n\{f^{(n)} \in [a_1,b_1] \text{ and } g^{(n)} \notin (c,d)\}}{\mu^n\{f^{(n)} \in (a_1,b_1)\}} \to 0$, $\frac{\mu^n\{f^{(n)} \in [a_2,b_2] \text{ and } g^{(n)} \notin (c,d)\}}{\mu^n\{f^{(n)} \in (a_2,b_2)\}} \to 0$ and $b_1 = a_2$ then $\frac{\mu^n\{f^{(n)} \in [a_1,b_2] \text{ and } g^{(n)} \notin (c,d)\}}{\mu^n\{f^{(n)} \in (a_1,b_2)\}} \to 0$.

Prove it.

We use h, Λ_h, Λ_f introduced in Sect. 2j.

It may happen that $(\Lambda'_f)^{-1}(a) < 0 < (\Lambda'_f)^{-1}(b)$. In this case we split the interval $[a,b]$ in two and use 2k1. Thus, we restrict ourselves to the case $(\Lambda'_f)^{-1}(a) \ge 0$. (The other case, $(\Lambda'_f)^{-1}(b) \le 0$, is similar.)

We have $\lambda \in [0,\infty) \cap \operatorname{Int}\{\Lambda_f < \infty\}$ such that $\Lambda'_f(\lambda) = a$. Denote $r = \lambda a - \Lambda_f(\lambda)$, then $\Lambda^*_f(a) = r$. By the lower bound of the LDP (Sect. 2i),

$$\mu^n\{f^{(n)} \in (a,b)\} \ge \exp\Big(- n \inf_{(a,b)} \Lambda^*_f - o(n) \Big) \ge \exp\big(-nr - o(n) \big) \,,$$

since[1] $\inf_{(a,b)} \Lambda^*_f \le \Lambda^*_f(a) = r$. It is sufficient to prove that for some $\delta > 0$,

$$\mu^n\{f^{(n)} \in [a,b] \text{ and } g^{(n)} \notin (c,d)\} \le \exp\big(-n(r+\delta) \big) \,.$$

We have $(c,d) \ni \varphi(a) = \int g \, \mathrm{d}\nu$ where $\nu = \exp\big(\lambda f - \Lambda_f(\lambda)\big) \cdot \mu$ is the tilted measure. Also, $\int \mathrm{e}^{\alpha g} \, \mathrm{d}\nu < \infty$ for all α small enough (positive or negative), since $(\lambda, 0) \in \operatorname{Int}\{\Lambda_h < \infty\}$. By (2e5) (adapted a bit), for some $\delta > 0$,

$$\nu^n\{g^{(n)} \notin (c,d)\} \le \mathrm{e}^{-\delta n} \quad \text{for all } n \,.$$

Thus (using 2e4),

$$\begin{aligned} &\mu^n\{f^{(n)} \in [a,b] \text{ and } g^{(n)} \notin (c,d)\} = \\ &\quad = \int \mathbb{1}_{[a,b]}(f^{(n)}) \big(1 - \mathbb{1}_{(c,d)}(g^{(n)}) \big) \exp n \big(\Lambda_f(\lambda) - \lambda f^{(n)} \big) \, \mathrm{d}\nu_n \le \\ &\qquad \le \exp n \big(\Lambda_f(\lambda) - \lambda a \big) \int \big(1 - \mathbb{1}_{(c,d)}(g^{(n)}) \big) \, \mathrm{d}\nu_n \le \mathrm{e}^{-rn} \mathrm{e}^{-\delta n} \,, \end{aligned}$$

which completes the proof.

2l Equivalence of ensembles

The probability measure

$$B \mapsto \mathbb{P}\big(B \,\big|\, h^{(n)} \in [E, E+\Delta E] \big) = \frac{\mu^n(B \cap \{h^{(n)} \in [E, E+\Delta E]\})}{\mu^n\{h^{(n)} \in [E, E+\Delta E]\}}$$

for a large n and small ΔE is well-known in statistical physics as the microcanonical ensemble,[2] provided that h is the Hamiltonian.

[1] In fact, $\inf_{(a,b)} \Lambda^*_f = \Lambda^*_f(a)$, since Λ^*_f is increasing on $[a,b]$.

[2] "… the *basic postulate of equilibrium statistical mechanics* … expresses the fact that we know very little about the microscopic state of the system: we only assume that its energy lies in a narrow interval $(E, E+\Delta E)$. … each of these states is equally probable … This is the famous principle of equal *a priori* probabilities. … The microcanonical ensemble is of prime theoretical importance … however … proves to be a mathematically difficult and unflexible tool." R. Balescu, "Equilibrium and nonequilibrium statistical mechanics", 1975, Sect. 4.2 "The microcanonical ensemble".

The tilted measure

$$\nu^n = \exp n\big(\lambda h^{(n)} - \Lambda(\lambda)\big) \cdot \mu^n \,,$$

where λ is chosen so that $\Lambda'(\lambda) = E$, is well-known in statistical physics as the canonical ensemble,[1] traditionally written as[2]

$$\nu^n(\mathrm{d}\omega_1 \dots \mathrm{d}\omega_n) = \frac{1}{Z(\beta)} \mathrm{e}^{-\beta h(\omega_1) - \dots - \beta h(\omega_n)} \mu(\mathrm{d}\omega_1) \dots \mu(\mathrm{d}\omega_n)$$

where $\beta = -\lambda$ is called the inverse temperature, and $Z(\beta) = \mathrm{e}^{n\Lambda(\lambda)} = \int \mathrm{e}^{-\beta h(\omega_1) - \dots - \beta h(\omega_n)} \mu(\mathrm{d}\omega_1) \dots \mu(\mathrm{d}\omega_n)$ is called the partition function.[3]

Let h and g satisfy the conditions of Theorem 1b4.

For every $\varepsilon > 0$ there exists $\Delta E > 0$ such that

$$\mathbb{P}\big(g^{(n)} \in (\varphi(E) - \varepsilon, \varphi(E) + \varepsilon) \,\big|\, h^{(n)} \in [E, E + \Delta E] \big) \to 1$$

as $n \to \infty$. On the other hand, for every $\varepsilon > 0$ there exists $\delta > 0$ such that for all n,

$$\nu^n \{ g^{(n)} \in (\varphi(E) - \varepsilon, \varphi(E) + \varepsilon) \} \ge 1 - \mathrm{e}^{-\delta n} \,.$$

Thus, a macroscopic observable $g^{(n)}$ concentrates around the same value $\varphi(E)$ in both ensembles, microcanonical and canonical. This phenomenon is well-known in statistical physics as equivalence of ensembles.[4]

2m Hints to exercises

2d1: (a) Hölder inequality; (b) Fatou's lemma.

2d2 (b): try $\Omega = \mathbb{R}$, $f(x) = x$, $\mu(\mathrm{d}x) = p(x)\,\mathrm{d}x$, $p(x) \sim x^\alpha \mathrm{e}^{-\gamma x}$ and $p(-x) \sim x^\beta \mathrm{e}^{-\delta x}$ for $x \to +\infty$. (c): use 2d1. (d): strict Hölder inequality.

2d3: $|\sum_{n=0}^N \mathrm{e}^{\lambda f} \frac{(\pm \varepsilon f)^n}{n!}| \le \mathrm{e}^{\lambda f} \mathrm{e}^{\varepsilon |f|} \le \mathrm{e}^{(\lambda - \varepsilon) f} + \mathrm{e}^{(\lambda + \varepsilon) f}$; use the dominated convergence theorem.

2d6: Transform $\frac{\mathrm{e}^\lambda - \mathrm{e}^{-\lambda}}{\mathrm{e}^\lambda + \mathrm{e}^{-\lambda}} = a$ into $\mathrm{e}^\lambda = \sqrt{\frac{1+a}{1-a}}$.

[1] "It was introduced for the first time by J.W. Gibbs (in the classical case) around 1900." Balescu, Sect. 4.3 "The canonical ensemble", page 119.

[2] We restrict ourselves to ideal systems (recall Sect. 1a).

[3] "It is one of the most important quantities of equilibrium statistical mechanics." Balescu, Sect. 4.3, page 118.

[4] "This result is very important in practice. It allows us, in many cases, to use in a given problem interchangeably one or the other ensemble, the choice being motivated by practical convenience in the calculations." Balescu, Sect. 4.6 "Equivalence of the equilibrium ensembles: fluctuations."

2e5: Use (2d5) and 2e1: $\nu^n\{f^{(n)} \ge \Lambda'(\lambda+\alpha)\} \le \exp n\big(\Lambda(\lambda+\alpha) - \Lambda(\lambda) - \alpha\Lambda'(\lambda+\alpha)\big)$ for small $\alpha > 0$.

2g1: similar to 2d1.

2g2: similar to 2d4.

2g3: recall Sect. 2d.

2g6: show that $\limsup \frac{1}{n} \ln \mu^n\{f^{(n)} \in K\} \le -\min_K \Lambda^*$; to this end, given $C < \min_K \Lambda^*$, cover K by a finite number of disks $B_\delta(a)$ as in 2g5.

2i1: consider $\lambda a - \Lambda(\lambda)$ as $\lambda \to \pm\infty$.

2i5: $\int e^{\lambda f}\,d\mu \le e^{(\lambda+1)\operatorname{ess\,inf} f} \int e^{-f}\,d\mu$ for $\lambda < -1$.

2i6: Similarly to the case of (1a1).

Index