# 3 Entropy appears


| | | |
|---|---|---|
| 3a | Binomial LDP: The simplest case of Sanov's theorem | 12 |
| 3b | Multinomial LDP: Sanov's theorem | 15 |
| 3c | The simplest case of Cramer's theorem via Gibbs's conditioning | 16 |
| 3d | Back to the physical question | 19 |


## 3a Binomial LDP: The simplest case of Sanov's theorem

Tossing a fair coin $n$ times we get $k \in \{0, 1, \dots, n\}$ 'heads' with the probability $2^{-n}\binom{n}{k} = 2^{-n}\frac{n!}{k!(n-k)!}$. The frequency of heads is $k/n$. We consider the distribution $\mu_n$ of the frequency,

$$\mu_n \in P([0,1])\,, \qquad \int f \,\mathrm{d}\mu_n = \sum_{k=0}^{n} 2^{-n}\binom{n}{k} f\Big(\frac{k}{n}\Big)\,. \tag{3a1}$$

**3a2 Exercise.** Prove that

$$1 \le \frac{\|f\|_{L_n(\mu_n)}}{\max_{k=0,1,\dots,n}\Big(|f(k/n)| \cdot \big(2^{-n}\binom{n}{k}\big)^{1/n}\Big)} \le \underbrace{(n+1)^{1/n}}_{\to 1}\,.$$

**3a3 Exercise.** Prove that

$$\bigg(2^{-n}\binom{n}{k}\bigg)^{1/n} \sim \frac{1}{2}\exp\Big(-\frac{k}{n}\ln\frac{k}{n} - \frac{n-k}{n}\ln\frac{n-k}{n}\Big) = \Big(\frac{n}{2k}\Big)^{\frac{k}{n}}\Big(\frac{n}{2(n-k)}\Big)^{\frac{n-k}{n}}$$

as $n \to \infty$, uniformly in $k \in \{0, 1, \dots, n\}$ (here $0^0 = 1$ and $0 \ln 0 = 0$).

Hint: you do not need Stirling's formula; instead, note that $(n!)^{1/n} \sim n/\mathrm{e}$, since

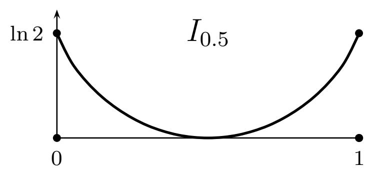
$$-\int_{1/n}^{1} \ln x \,\mathrm{d}x \le -\frac{1}{n}\Big(\ln\frac{1}{n} + \dots + \ln\frac{n}{n}\Big) \le -\int_0^1 \ln x \,\mathrm{d}x\,.$$

Further, $(k!)^{1/n} \sim (k/\mathrm{e})^{k/n}$; you may prove it separately for relatively small $k$ (say, $k \le \sqrt{n}$) and for other $k$.

**3a4 Exercise.** $(\mu_n)_n$ satisfies LDP with the rate function $I = I_{0.5}$ defined by

$$\text{(3a5)} \quad I_{0.5}(x) = x \ln x + (1-x) \ln(1-x) + \ln 2 = \\ = x \ln(2x) + (1-x) \ln((2(1-x)) \quad \text{for } 0 < x < 1 \,, \quad I(0) = I(1) = \ln 2 \,.$$

Prove it.




The expression $-x \ln x - (1-x) \ln(1-x)$ is well-known as the entropy of the distribution consisting of two atoms of masses $x$ and $1-x$.

See [5, Th. 1.3.1].

The statement 3a4 suggests an approximation

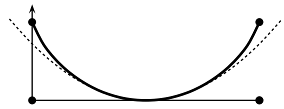
$$2^{-n} \binom{n}{k} \approx \exp\Big( - n I_{0.5}\Big(\frac{k}{n}\Big)\Big) = \Big(\frac{n}{2k}\Big)^k \Big(\frac{n}{2(n-k)}\Big)^{n-k} .$$

But on the other hand, the central limit theorem (or its special case, the De Moivre-Laplace theorem) suggests another approximation,

$$2^{-n} \binom{n}{k} \approx \sqrt{\frac{2}{\pi n}} \exp\Big( - \frac{(2k-n)^2}{2n} \Big) \approx \exp\Big( - n \cdot 2 \Big(\frac{k}{n} - \frac{1}{2}\Big)^2 \Big) .$$

Of course, $I_{0.5}(x) \ne 2(x-0.5)^2$. However,

$$\text{(3a6)} \qquad I_{0.5}(x) \sim 2(x-0.5)^2 \quad \text{as } x \to 0.5 \,,$$



since $I_{0.5}(0.5) = 0$, $I'_{0.5}(0.5) = 0$ and $I''_{0.5}(0.5) = 4$. Look at some numerics: for $n = 200$,

| $k$ | 100 | 115 | 130 | 145 | 160 | 175 | 190 |
|---|---|---|---|---|---|---|---|
| $2^{-n}\binom{n}{k}$ | $6 \cdot 10^{-2}$ | $6 \cdot 10^{-3}$ | $6 \cdot 10^{-6}$ | $5 \cdot 10^{-11}$ | $1 \cdot 10^{-18}$ | $3 \cdot 10^{-29}$ | $1 \cdot 10^{-44}$ |
| $\sqrt{\frac{2}{\pi n}} \exp(-\frac{(2k-n)^2}{2n})$ | $6 \cdot 10^{-2}$ | $6 \cdot 10^{-3}$ | $7 \cdot 10^{-6}$ | $9 \cdot 10^{-11}$ | $1 \cdot 10^{-17}$ | $2 \cdot 10^{-26}$ | $4 \cdot 10^{-37}$ |
| $\exp\big(-nI_{0.5}(\frac{k}{n})\big)$ | 1 | $1 \cdot 10^{-1}$ | $1 \cdot 10^{-4}$ | $8 \cdot 10^{-10}$ | $2 \cdot 10^{-17}$ | $3 \cdot 10^{-28}$ | $1 \cdot 10^{-43}$ |

Tossing an unfair coin $n$ times we get $k \in \{0,1,\dots,n\}$ 'heads' with the probability $\binom{n}{k} p^k (1-p)^{n-k}$; here $p \in (0,1)$ is a parameter of the coin. Similarly to (3a1),

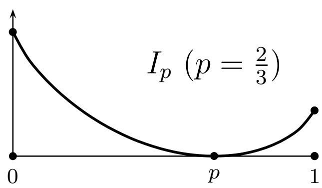
$$\mu_n^{(p)} \in P([0,1]) \,, \qquad \int f \, \mathrm{d}\mu_n^{(p)} = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} f\Big(\frac{k}{n}\Big) \,. \tag{3a7}$$

**3a8 Exercise.** $(\mu_n^{(p)})_n$ satisfies LDP with the rate function $I_p$ defined by

$$I_p(x) = x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p} \quad \text{for } 0 < x < 1 \,, \tag{3a9}$$
$$I_p(0) = -\ln(1-p), \; I_p(1) = -\ln p \,.$$

Prove it.

Hint: similar to 3a4.




The case $p = 0.5$ conforms to (3a5).

The expression (3a9) for $I_p(x)$ is well-known as the relative entropy of the distribution $(x, 1-x)$ w.r.t. the distribution $(p, 1-p)$; it may also be written as

$$I_p(x) = \Big( \frac{x}{p} \ln \frac{x}{p} \Big) \cdot p + \Big( \frac{1-x}{1-p} \ln \frac{1-x}{1-p} \Big) \cdot (1-p) \,.$$

Alternatively, we can derive LDP for $(\mu_n^{(p)})_n$ from the case $p = 0.5$ by means of 2c (change of measure). Indeed,

$$\frac{\mathrm{d}\mu_n^{(p)}}{\mathrm{d}\mu_n^{(0.5)}} \Big( \frac{k}{n} \Big) = c_n \Big( \frac{p}{1-p} \Big)^k \,, \quad c_n = (2(1-p))^n \,,$$

thus,

$$\frac{\mathrm{d}\mu_n^{(p)}}{\mathrm{d}\mu_n^{(0.5)}} (x) = c_n \mathrm{e}^{-nh(x)} \,, \quad h(x) = -x \ln \frac{p}{1-p} \,.$$

By Theorem 2c1, $(\mu_n^{(p)})_n$ satisfies LDP with the rate function $J = I_{0.5} + h - \lim_n \frac{1}{n} \ln c_n$;

$$J(x) = x \ln x + (1-x) \ln(1-x) + \ln 2 - x \ln p + x \ln(1-p) - \ln 2 - \ln(1-p) =$$
$$= x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p} = I_p(x) \,.$$

## 3b Multinomial LDP: Sanov's theorem

Throwing a fair die $n$ times we get an outcome $k = (k_1, \dots, k_6)$ (satisfying $k_1, \dots, k_6 \in \{0, 1, 2, \dots\}$, $k_1 + \dots + k_6 = n$) with the probability

$$6^{-n} \binom{n}{k_1, \dots, k_6} = 6^{-n} \frac{n!}{k_1! \dots k_6!} .$$

The frequencies $k_1/n, \dots, k_6/n$ may be treated as a (random) probability measure (well-known as the empirical measure or the empirical distribution),

$$\frac{1}{n} k \in P(\{1, \dots, 6\}) .$$

Similarly to (3a1), the distribution $\mu_n$ of the frequency is

(3b1) $$\mu_n \in P(P(\{1, \dots, 6\})) ,$$
$$\int f \, d\mu_n = \sum_{k_1, \dots, k_6} 6^{-n} \binom{n}{k_1, \dots, k_6} f\Big( \frac{k_1}{n}, \dots, \frac{k_6}{n} \Big) .$$

Do not be afraid of $P(P(\{1, \dots, 6\}))$; this is the set of probability measures on the 5-dimensional simplex $P(\{1, \dots, 6\}) = \{(x_1, \dots, x_6) : x_1, \dots, x_6 \ge 0, \, x_1 + \dots + x_6 = 1\}$.

**3b2 Exercise.** Prove that

$$\bigg( 6^{-n} \binom{n}{k_1, \dots, k_6} \bigg)^{1/n} \sim \frac{1}{6} \exp \Big( - \frac{k_1}{n} \ln \frac{k_1}{n} - \dots - \frac{k_6}{n} \ln \frac{k_6}{n} \Big) = \\ = \Big( \frac{n}{6k_1} \Big)^{\frac{k_1}{n}} \dots \Big( \frac{n}{6k_6} \Big)^{\frac{k_6}{n}}$$

as $n \to \infty$, uniformly in $k_1, \dots, k_6$.

Hint: similar to 3a3.

**3b3 Exercise.** $(\mu_n)_n$ satisfies LDP with the rate function (on the simplex)

$$I(x_1, \dots, x_6) = x_1 \ln x_1 + \dots + x_6 \ln x_6 + \ln 6 = x_1 \ln(6x_1) + \dots + x_6 \ln(6x_6) .$$

Prove it.

Hint: similar to 3a4.

An unfair die has a parameter $p \in P(\{1, \dots, 6\})$; $p = (p_1, \dots, p_6)$, $p_1, \dots, p_6 > 0$, $p_1 + \dots + p_6 = 1$. The probability of an outcome $k = (k_1, \dots, k_6)$ is

$$\binom{n}{k_1, \dots, k_6} p_1^{k_1} \dots p_6^{k_6} .$$

The distribution $\mu_n^{(p)}$ of the frequency is

$$\int f \, \mathrm{d}\mu_n^{(p)} = \sum_{k_1,\dots,k_6} \binom{n}{k_1,\dots,k_6} p_1^{k_1} \dots p_6^{k_6} f\Big(\frac{k_1}{n},\dots,\frac{k_6}{n}\Big) \, .$$

Thus,

$$\frac{\mathrm{d}\mu_n^{(p)}}{\mathrm{d}\mu_n}\Big(\frac{k_1}{n},\dots,\frac{k_6}{n}\Big) = (6p_1)^{k_1} \dots (6p_6)^{k_6} \, .$$

Applying Theorem 2c1 (change of measure) for $c_n = 1$ and $h(x_1,\dots,x_6) = -x_1 \ln(6p_1) - \dots - x_6 \ln(6p_6)$, we get LDP for $(\mu_n^{(p)})_n$ with the rate function

$$I_p(x_1,\dots,x_6) = x_1 \ln \frac{x_1}{p_1} + \dots + x_6 \ln \frac{x_6}{p_6} \, .$$

The latter is well-known as the relative entropy, $H(x|p)$.

Replacing 6 with an arbitrary number we get Sanov's theorem.

**3b4 Theorem.** Let $A$ be a finite set and $p \in P(A)$ a probability measure on $A$. Define $\mu_n^{(p)} \in P(P(A))$ as the distribution of the empirical measure (in other words, frequencies) in a sample of size $n$ from the measure $p$. Then the sequence $(\mu_n^{(p)})_n$ satisfies LDP with the rate function $x \mapsto H(x|p)$.

Here $H(x|p)$ is the relative entropy,

$$H(x|p) = \sum_{a \in A} x_a \ln \frac{x_a}{p_a} \quad \text{for } x \in P(A) \, ;$$

by convention, $0 \ln \frac{0}{p_a} = 0$ (be $p_a$ positive or zero), and $x_a \ln \frac{x_a}{0} = +\infty$ for $x_a > 0$.

See [2, Th. 2.1.10], [5, Th. 1.4.3].

## 3c The simplest case of Cramer's theorem via Gibbs's conditioning

Let $X_1, X_2, \dots$ be independent, identically distributed random variables, each taking on the three values $-1, 0, 1$ with equal probabilities (1/3). We consider the distribution $\mu_n$ of the mean value $(X_1 + \dots + X_n)/n$;

$$\text{(3c1)} \quad \mu_n \in P([-1,1]) \, , \qquad \int f \, \mathrm{d}\mu_n = 3^{-n} \sum_{x_1,\dots,x_n \in \{-1,0,1\}} f\Big(\frac{x_1 + \dots + x_n}{n}\Big) \, .$$

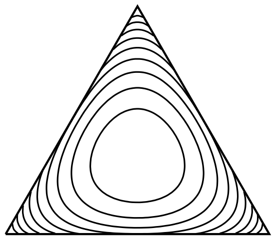
In order to use Sanov's theorem (and the contraction principle), we introduce the frequencies $\frac{k_-}{n}, \frac{k_0}{n}, \frac{k_+}{n}$, where

$$k_- = \#\{i : x_i = -1\}\,, \quad k_0 = \#\{i : x_i = 0\}\,, \quad k_+ = \#\{i : x_i = 1\}\,.$$

By Sanov's theorem, distributions $\nu_n$ of $(\frac{k_-}{n}, \frac{k_0}{n}, \frac{k_+}{n})$ satisfy LDP with the rate function

$$I_1(x_-, x_0, x_+) = x_- \ln(3x_-) + x_0 \ln(3x_0) + x_+ \ln(3x_+) = \ln 3 - H(x_-, x_0, x_+)\,,$$
$$H(x_-, x_0, x_+) = -x_- \ln x_- - x_0 \ln x_0 - x_+ \ln x_+$$

for $x_-, x_0, x_+ \ge 0$, $x_- + x_0 + x_+ = 1$. (As before, $0 \ln 0 = 0$.)



On the other hand,

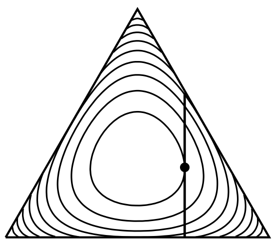
$$\frac{x_1 + \dots + x_n}{n} = \frac{k_+}{n} - \frac{k_-}{n}\,.$$

The contraction principle 2b1, applied to

$$F : \{(x_-, x_0, x_+) : x_-, x_0, x_+ \ge 0, x_- + x_0 + x_+ = 1\} \to [-1, 1]\,,$$
$$F(x_-, x_0, x_+) = x_+ - x_-\,,$$

tells us that $(\mu_n)_n$ satisfies LDP with the rate function

$$I_2(y) = \min\{I_1(x_-, x_0, x_+) : x_+ - x_- = y\}\,.$$



On the line $x_+ - x_- = y$ we have $x_+ = (1 - x_0 + y)/2$, $x_- = (1 - x_0 - y)/2$, thus,

$$\frac{\mathrm{d}}{\mathrm{d}x_0} I_1\Big(\frac{1 - x_0 - y}{2}, x_0, \frac{1 - x_0 + y}{2}\Big) =$$
$$-\frac{1}{2}\Big(1 + \ln \frac{1 - x_0 - y}{2}\Big) + (1 + \ln x_0) - \frac{1}{2}\Big(1 + \ln \frac{1 - x_0 + y}{2}\Big) = \ln x_0 - \frac{\ln x_- + \ln x_+}{2}\,.$$

The minimizer satisfies $x_0 = \sqrt{x_- x_+}$; that is, $x_-, x_0, x_+$ are a geometric progression. (The boundary values are local maxima, not minima.) We may write (recall 1d)

$$(x_-, x_0, x_+) = \frac{1}{\mathrm{e}^b + 1 + \mathrm{e}^{-b}} \cdot (\mathrm{e}^b, 1, \mathrm{e}^{-b})$$

where $b \in \mathbb{R}$ is determined by the equation

$$\frac{\mathrm{e}^b - \mathrm{e}^{-b}}{\mathrm{e}^b + 1 + \mathrm{e}^{-b}} = -y \tag{3c2}$$

(the left-hand side is strictly increasing in $b$, from $-1$ to $1$). We get

$$\begin{aligned} I_2(y) &= \ln 3 + x_- \ln x_- + x_0 \ln x_0 + x_+ \ln x_+ = \\ &= \ln 3 - \underbrace{(x_- + x_0 + x_+)}_{=1} \ln(\mathrm{e}^b + 1 + \mathrm{e}^{-b}) + \frac{b\mathrm{e}^b - b\mathrm{e}^{-b}}{\mathrm{e}^b + 1 + \mathrm{e}^{-b}} = \\ &= -by - \ln \frac{\mathrm{e}^b + 1 + \mathrm{e}^{-b}}{3} \, . \end{aligned}$$

The equation (3c2) may be written as

$$\frac{\mathrm{d}}{\mathrm{d}b} \Big( by + \ln \frac{\mathrm{e}^b + 1 + \mathrm{e}^{-b}}{3} \Big) = 0 \, ,$$

thus, $b$ is nothing but the minimizer of the (strictly convex) function $b \mapsto by + \ln \frac{\mathrm{e}^b+1+\mathrm{e}^{-b}}{3}$, which leads to another formula for $I_2$,

$$I_2(y) = \max_{b \in \mathbb{R}} \Big( -by - \ln \frac{\mathrm{e}^b + 1 + \mathrm{e}^{-b}}{3} \Big) \, . \tag{3c3}$$

Note that

$$\frac{\mathrm{e}^b + 1 + \mathrm{e}^{-b}}{3} = \mathbb{E}\, \mathrm{e}^{bX_1} = \big( \mathbb{E}\, \mathrm{e}^{b(X_1 + \dots + X_n)} \big)^{1/n} = \| f_b \|_{L_n(\mu_n)} \, ,$$

where $f_b(x) = \mathrm{e}^{bx}$ for $x \in [-1, 1]$. Therefore

$$\max_{x \in [-1,1]} \big( \mathrm{e}^{bx} \mathrm{e}^{-I_2(x)} \big) = \frac{\mathrm{e}^b + 1 + \mathrm{e}^{-b}}{3} \, ,$$

that is,

$$\min_{x \in [-1,1]} \big( I_2(x) - bx \big) = -\ln \frac{\mathrm{e}^b + 1 + \mathrm{e}^{-b}}{3} \, . \tag{3c4}$$

In fact, (3c3) can be deduced from (3c4), which is another way to (3c3) (assuming LD-convergence).

See also [4, Sect. 4], [5, Sect. VIII.3], [1, Kullback's lemma on page 30], [3, Exercise 3.3.12] and [2, Sect. 2.2].

## 3d Back to the physical question

We return to the physical question of 1a. On the configuration space $\{-1,0,1\}^n$ we have two probability measures, the uniform distribution $U_n$ and the so-called Gibbs measure $G_n$;

$$\int f\,\mathrm{d}U_n = 3^{-n} \sum_{x\in\{-1,0,1\}^n} f(x)\,,$$

$$\int f\,\mathrm{d}G_n = \mathrm{const}_n \cdot \int f \exp(-\beta H_n)\,\mathrm{d}U_n = \frac{\int f \mathrm{e}^{-\beta H_n}\,\mathrm{d}U_n}{\int \mathrm{e}^{-\beta H_n}\,\mathrm{d}U_n}\,;$$

here (as in Sect. 1), $\beta = \frac{1}{k_{\mathrm{B}}T}$ is the inverse temperature, and $H_n$ is the Hamiltonian; recall that

$$H_n(s_1,\dots,s_n) = nf\Big(\frac{s_1+\dots+s_n}{n}\Big)\,,$$

where $f : [-1,1] \to \mathbb{R}$ is a given smooth function (not depending on $n$).

Accordingly, on $[-1,1]$ we have two probability measures, $\mu_n$ (recall (3c1)) and $\nu_n$,

$$\frac{\mathrm{d}\nu_n}{\mathrm{d}\mu_n} = \frac{\mathrm{e}^{-n\beta f}}{\int \mathrm{e}^{-n\beta f}\,\mathrm{d}\mu_n}\,.$$

They are the images of $U_n$ and $G_n$ respectively, under the map $(s_1,\dots,s_n) \mapsto (s_1+\dots+s_n)/n$.

By 3c, $(\mu_n)_n$ satisfies LDP with the rate function $I_2$ (recall (3c3)). By Theorem 2c1 (change of measure), $(\nu_n)_n$ satisfies LDP with the rate function

$$I = (I_2 + \beta f) - \min_{[-1,1]}(I_2+\beta f)\,.$$

By 2a20, $\nu_n$ concentrate near zeros of $I$ (in the sense that $\nu_n\big(\{x : I(x) \le \varepsilon\}\big) \to 1$ as $n \to \infty$), that is, minima of $I_2 + \beta f$. Assuming that $I_2+\beta f$ has a unique minimum at some $x_\beta \in [-1,1]$ we conclude that $\nu_n$ concentrate near $x_\beta$ (that is, $\nu_n\big([x_\beta-\varepsilon, x_\beta+\varepsilon]\big) \to 1$ as $n\to\infty$, for every $\varepsilon > 0$). Thus, for large $n$, with high probability, $(s_1+\dots+s_n)/n$ is close to $x_\beta$, therefore the energy per particle $f\big(\frac{s_1+\dots+s_n}{n}\big)$ is close to $f(x_\beta)$.

It remains to note that the entropy $S$ of 1d is $-I_2(x)+\ln 3$, thus $x_\beta$ in 1d is the same as $x_\beta$ here. The 'physical approach' of 1d conforms to the theory of large deviations.

# References

[1] J.A. Bucklev, *Large deviation techniques in decision, simulation, and estimation,* Wiley, 1990.

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

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

[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] R.S. Ellis, *Entropy, large deviations, and statistical mechanics,* Springer, 1985.

## Index

entropy, 13

relative entropy, 16

$b$, 18
$H$, 17
$H(x|p)$, 16
$I_1$, 17
$I_2$, 17
$I_p$, 14
$I_{0.5}$, 13
$k_-, k_0, k_+$, 17
$\mu_n$, 12, 15, 16
$\mu_n^{(p)}$, 14, 16
$P(P(\dots))$, 15