Large deviations, moment problems and sum rules

Jan Nagel

TU Dortmund

40th Finnish Summer School on Probability and Statistics

Lammi, 23.05. - 27.05.

Overview

In this course, we will talk about (random) probability measures in different languages – in terms of *moments, orthogonal polynomials or densities*.

Sometimes, switching to another language is easier in order to formulate properties/convergence or distributions of random measures.

For large deviation statement in particular, it is helpful to have different descriptions at hand, which opens up a very interesting connections between random matrices and important identities in spectral theory called *sum rules*.

By proving large deviation in two different languages, we obtain a probabilistic proof and interpretation of a sum rule

$$\mathcal{I}_A(\mu) = \mathcal{I}_B(\mu).$$

Overview

1st lecture: Random moments and large deviations

2nd lecture: Large deviation for random matrices and sum rules

3rd lecture: A probabilistic proof of sum rules

Overview

1st lecture: Random moments and large deviations

2nd lecture: Large deviation for random matrices and sum rules

3rd lecture: A probabilistic proof of sum rules

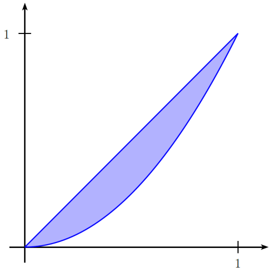
The classical moment problem

- Let $\mathcal{P}([0,1])$ be the set of all probability measures on $[0,1]$.
- For $\mu \in \mathcal{P}([0,1])$, let $m_k(\mu) = m_k = \int x^k \, d\mu(x)$ be the k-th moment.
- $\mu \in \mathcal{P}([0,1])$ is uniquely determined by its moment sequence.
- The moment space: $\mathcal{M}([0,1]) = \{(m_k(\mu))_{k \geq 1} : \mu \in \mathcal{P}([0,1])\}$
- Hausdorff moment problem (classical/deterministic): Is a given sequence $c = (c_k)_{k \geq 1}$ a moment sequence, i.e., is

$$c \in \mathcal{M}([0,1])?$$

The classical moment problem

- The n-th Moment space: $\mathcal{M}_n([0,1]) = \{(m_1(\mu), \dots, m_n(\mu)) : \mu \in \mathcal{P}([0,1])\}$
- $\mathcal{M}_n([0,1])$ is the convex hull of $\{(x, x^2, \dots, x^n) : x \in [0,1]\}$.



- Given $m_1 \in \mathcal{M}_1([0,1]) = [0,1]$, we have $(m_1, m_2) \in \mathcal{M}_2([0,1])$ iff

$$m_1^2 \leq m_2 \leq m_1.$$

- Given $(m_1, \dots, m_{n-1}) \in \mathcal{M}_{n-1}([0,1])$, there exists m_n^+, m_n^- (depending on first $n-1$ moments), such that:

$$(m_1, \dots, m_n) \in \mathcal{M}_n([0,1]) \quad \Leftrightarrow \quad m_n^- \leq m_n \leq m_n^+$$

Moreover, $m_n^- < m_n^+$, iff $(m_1, \dots, m_{n-1}) \in \mathrm{Int}(\mathcal{M}_{n-1}([0,1]))$.

Answer to the classical moment problem

- $(c_1, \ldots, c_n) \in \mathcal{M}_n([0,1])$ iff $L(P) = \sum_{i=0}^{n} a_i c_i \geq 0$ whenever $P(x) = a_n x^n + \cdots + a_0$ is a polynomial nonnegative on $[0,1]$.
- A polynomial P of degree $2m$ is nonnegative on $[0,1]$ iff there are polynomials B_m, C_{m-1} such that

$$P(x) = B_m(x)^2 + x(1-x)C_{m-1}(x)^2$$

- A polynomial P of degree $2m+1$ is nonnegative on $[0,1]$ iff there are polynomials B_m, C_m such that

$$P(x) = xB_m(x)^2 + (1-x)C_m(x)^2$$

- So $(c_1, \ldots, c_n) \in \mathcal{M}_n([0,1])$ iff

$$L(B_m^2), \quad L(x(1-x)C_{m-1}^2), \quad L(xB_m^2), \quad L((1-x)C_m^2)$$

are nonnegative, whenever L is applies to a polynomial of degree $\leq n$.

Answer to the classical moment problem

- $(c_1, \dots, c_n) \in \mathcal{M}_n([0,1])$ iff

$$L(B_m^2), \quad L(x(1-x)C_{m-1}^2), \quad L(xB_m^2), \quad L((1-x)C_m^2)$$

  are nonnegative, whenever L is applies to a polynomial of degree $\leq n$.
- Define the Hankel matrices

$$\underline{H}_{2m} = \begin{pmatrix} c_0 & \dots & c_m \\ \vdots & & \vdots \\ c_m & \dots & c_{2m} \end{pmatrix} \quad \overline{H}_{2m} = \begin{pmatrix} c_1 - c_2 & \dots & c_m - c_{m+1} \\ \vdots & & \vdots \\ c_m - c_{m+1} & \dots & c_{2m-1} - c_{2m} \end{pmatrix}$$

$$\underline{H}_{2m+1} = \begin{pmatrix} c_1 & \dots & c_{m+1} \\ \vdots & & \vdots \\ c_{m+1} & \dots & c_{2m+1} \end{pmatrix} \quad \overline{H}_{2m+1} = \begin{pmatrix} c_0 - c_1 & \dots & c_m - c_{m+1} \\ \vdots & & \vdots \\ c_m - c_{m+1} & \dots & c_{2m} - c_{2m+1} \end{pmatrix}$$

- Then $(c_1, \dots, c_n) \in \mathcal{M}_n([0,1])$ iff $\underline{H}_k$ and $\overline{H}_k$ are nonnegative definite for all $k \leq n$.
  In particular, $(c_1, \dots) \in \mathcal{M}([0,1])$ iff $\underline{H}_k$ and $\overline{H}_k$ are nonnegative definite for all k.

The probabilistic moment problem

- Idea of Chang, Kemperman, Studden (1993): To understand the shape and structure of $\mathcal{M}_n([0,1])$, look at a typical point.
- Let $m^{(n)}$ be uniformly distributed on $\mathcal{M}_n([0,1])$, study the asymptotics of

$$(m_1^{(n)}, \ldots, m_k^{(n)}).$$

- Problem: moment space has complicated structure, strong dependency between moments

$$m_1 \leq m_2 \leq m_1^2, \qquad \frac{m_2^2}{m_1} \leq m_3 \leq m_2 - \frac{(m_1 - m_2)^2}{1 - m_1}$$

- Map to new coordinates, which are easier to handle...

Canonical moments

For $(m_1, \dots, m_n) \in \mathrm{Int}(\mathcal{M}_n([0,1]))$, define the *canonical moments*

$$u_k = \frac{m_k - m_k^-}{m_k^+ - m_k^-}.$$

That is, u_k is the relative position of m_k in the interval of possible k-th moments. Well-defined since $m_k^- < m_k^+$.

$$u_1 = m_1, \qquad u_2 = \frac{m_2 - m_1^2}{m_1 - m_1^2}, \quad \dots$$

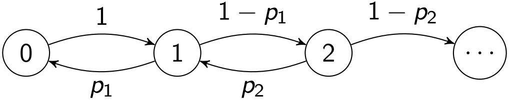
Nice properties:

- $u_k \in (0, 1)$
- The u_k are invariant under linear transformations of measure.
- μ is symmetric w.r.t. $\frac{1}{2}$ iff all odd canonical moments are $\frac{1}{2}$.

Canonical moments and random walks

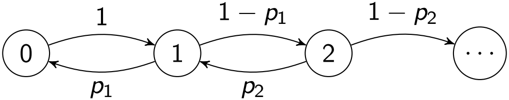
For $p_k \in [0,1], k \geq 1$, let $(X_n)_{n\geq 0}$ be the random walk on $\mathbb{N}_0$ starting at 0 and with transition probabilities

$$\mathbb{P}(X_{n+1} = y | X_n = x) = \begin{cases} p_k & \text{if } y = x-1, \\ 1 - p_k & \text{if } y = x+1. \end{cases}$$



- There exists a unique p.m. μ on $[0,1]$ such that $m_k = m_k(\mu) = \mathbb{P}(X_{2k} = 0 | X_0 = 0)$.
- 1-to-1 relation between $p_1, \ldots, p_n$ and $m_1, \ldots, m_n$.

Canonical moments and random walks



- There exists a unique p.m. μ on $[0, 1]$ such that $m_k = m_k(\mu) = \mathbb{P}(X_{2k} = 0 | X_0 = 0)$.
- 1-to-1 relation between $p_1, \ldots, p_n$ and $m_1, \ldots, m_n$.
- Fixing $m_1, \ldots, m_{n-1}$ (or fixing $p_1, \ldots p_{n-1}$), $m_n = m_n^-$ is minimized for $p_n = 0$ and $m_n = m_n^+$ is maximized for $p_n = 1$.
- $m_n - m_n^- = (1 - p_1) \ldots (1 - p_{n-1}) p_n \ldots p_1$ is probability of single path going to n and back.
- Maximizing this (setting $p_n = 1$) gives

$$m_n^+ - m_n^- = (1 - p_1) \ldots (1 - p_{n-1}) p_{n-1} \ldots p_1.$$

- Therefore,

$$p_n = \frac{m_n - m_n^-}{m_n^+ - m_n^-},$$

  that is, the p_n are the canonical moments u_n of μ.

Distribution of canonical moments

- Let $(m_1^{(n)}, \dots, m_n^{(n)})$ be uniformly distributed on $\mathcal{M}_n([0,1])$.
- Density of canonical moments $(u_1^{(n)}, \dots, u_n^{(n)})$:

$$\frac{1}{\mathrm{Vol}\mathcal{M}_n([0,1])} \mathbb{1}_{\{u_k \in (0,1)\}} \det \left(\frac{\partial m_k}{\partial u_\ell} \right)_{k,\ell}$$

- Since m_k depends only on $u_1, \dots, u_k$, the matrix is lower triangular and

$$\begin{aligned} \det \left(\frac{\partial m_k}{\partial u_\ell} \right)_{k,\ell} &= \prod_{k=1}^{n} \frac{\partial m_k}{\partial u_k} = \prod_{k=1}^{n} (m_k - m_k^-) \\ &= \prod_{k=1}^{n} u_1(1-u_1) \dots u_{k-1}(1-u_{k-1}) \\ &= \prod_{k=1}^{n} [u_k(1-u_k)]^{n-k} \end{aligned}$$

- So the $u_1^{(n)}, \dots, u_n^{(n)}$ are independent and

$$u_k^{(n)} \sim \mathrm{Beta}(n-k+1, n-k+1).$$

Distribution of canonical moments

For $(m_1^{(n)}, \dots, m_n^{(n)})$ uniformly distributed on $\mathcal{M}_n([0,1])$, the canonical moments $u_1^{(n)}, \dots, u_n^{(n)}$ are independent and

$$u_k^{(n)} \sim \mathrm{Beta}(n-k+1, n-k+1).$$

Consequences:

- $\mathrm{Vol}\mathcal{M}_n([0,1]) = \prod_{k=1}^{n} \frac{\Gamma(n-k+1)^2}{\Gamma(2(n-k-1))} = 2^{-n^2(1+o(1))}$
- Each $u_k^{(n)}$ converges in probability to $\frac{1}{2}$. The measure with canonical moments all equal to $\frac{1}{2}$ is the arcsine measure on $[0,1]$,

$$d\mu_{\mathrm{arc}}(x) = \frac{1}{\pi\sqrt{x(1-x)}}dx$$

- The explicit distribution also allows to show CLT and a large deviation principle, exponentially fast concentration of measure.

Large deviations

Let $\mathcal{X}$ be a complete separable metric space with Borel σ-algebra, $\mathcal{I} : \mathcal{X} \to [0, \infty]$ and $a_n \to \infty$.
We say that a sequence $(X_n)_n$ of $\mathcal{X}$-valued r.v. satisfies the *large deviation principle* with speed a_n and rate function $\mathcal{I} : \mathcal{X} \to [0, \infty]$, if $\mathcal{I}$ is lower semicontinuous and

- for C closed: $\limsup_{n\to\infty} \frac{1}{a_n} \log \mathbb{P}(X_n \in C) \le - \inf_{x \in C} \mathcal{I}(x)$,
- for O open: $\liminf_{n\to\infty} \frac{1}{a_n} \log \mathbb{P}(X_n \in O) \ge - \inf_{x \in O} \mathcal{I}(x)$.

We will only consider *good* rate functions, i.e., the level sets $\{x : \mathcal{I}(x) \le L\}$ are compact for all L.

As a consequence:

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{a_n} \log \mathbb{P}(X_n \in B_\varepsilon(x)) = -\mathcal{I}(x)$$

or informally,

$$\mathbb{P}(X_n \approx x) \approx e^{-a_n \mathcal{I}(x)}.$$

Large deviations

Some general principles:

- The rate function is unique.
- If $(X_n)_n$ and $(Y_n)_n$ are independent, each satisfying the LDP with rate $\mathcal{I}_X$, $\mathcal{I}_Y$, then $(X_n, Y_n)_n$ satisfies the LDP with rate $\mathcal{I}(x, y) = \mathcal{I}_X(x) + \mathcal{I}_Y(y)$.
- Dawson-Gärtner Theorem: If $X_n = (X_1^{(n)}, X_2^{(n)}, \dots)$ has values in $\mathcal{X}^{\mathbb{N}}$ and for each k, the projection $(\pi_k(X_n))_n$ onto the first k coordinates satisfies the LDP with rate $\mathcal{I}_k$, then $(X_n)_n$ satisfies the LDP (in product topology) with rate

$$\mathcal{I}(x) = \sup_k \mathcal{I}_k(\pi_k(x)).$$

- Contraction principle: If $(X_n)_n$ satisfies LDP in $\mathcal{X}$ with rate $\mathcal{I}_X$ and $f : \mathcal{X} \to \mathcal{Y}$ is continuous, then $(f(X_n))_n$ satisfies the LDP in $\mathcal{Y}$ with rate

$$\mathcal{I}_Y(y) = \inf\{\mathcal{I}_X(x) : f(x) = y\}.$$

Large deviation for canonical moments

- Easy situation: If X_n has a density on U (open)

$$C_n^{-1} F(x) \exp\{-nG(x)\}$$

with G continuous with compact level sets, $\inf_x G(x) = 0$ and $F \in L^1$ with $F \geq \varepsilon > 0$, then

$$\int_A C_n^{-1} F(x) \exp\{-nG(x)\}\, dx = \exp\{-n \inf_{x \in A} G(x)(1 + o(1))\}$$

and $(X_n)_n$ satisfies the LDP with speed n and rate function G.

- The canonical moment $u_k^{(n)}$ has a density of the form

$$C_n \exp\{-(n-k)(-\log(x - x^2) - \log(4))\}$$

so that $(u_k^{(n)})_n$ satisfies the LDP with speed n and rate function

$$H(x) = -\log(x - x^2) - \log(4).$$

Large deviation for canonical moments

- Each $(u_k^{(n)})_n$ satisfies the LDP with rate function

$$H(x) = -\log(x - x^2) - \log(4).$$

- Since the canonical moments are independent: $(u_1^{(n)}, \ldots, u_k^{(n)})_n$ satisfies LDP with rate function

$$H_k(u_1, \ldots, u_k) = \sum_{i=1}^{k} H(u_i).$$

- By the Dawson-Gärtner Theorem: The sequence $(u_1^{(n)}, \ldots)_n$ satisfies LDP with rate function

$$H_\infty(u_1, \ldots) = \sup_{k \geq 1} H_k(u_1, \ldots, u_k) = \sum_{i=1}^{\infty} H(u_i).$$

From random moments to random measures

- For $(m_1, \dots, m_{2n-1}) \in \mathrm{Int}\mathcal{M}_{2n-1}([0,1])$, there are infinitely many measures with these first $2n-1$ moments.
- There is a unique measure μ_n (called lower principal representation) corresponding to the moment sequence

$$(m_1, \dots, m_{2n-1}, m_{2n}^-, m_{2n+1}^-, \dots)$$

or equivalently to the sequence of canonical moments

$$(u_1, \dots, u_{2n-1}, 0, 0, \dots).$$

- Recall the random walk interpretation: μ_n satisfies

$$m_k(\mu_n) = \mathbb{P}(X_{2k} = 0 | X_0 = 0) = (P^{2k})_{1,1}$$

when P is the transition matrix of the random walk.

From random moments to random measures

- A direct calculation shows: $m_k(\mu_n) = (P^{2k})_{1,1} = (J_n^k)_{1,1}$

$$J_n = \begin{pmatrix} b_1 & a_1 & & & \\ a_1 & \ddots & \ddots & & \\ & \ddots & & & a_{n-1} \\ & & & a_{n-1} & b_n \end{pmatrix}$$

  where (with $u_0 = 0$)

$$b_k = (1 - u_{2k-3})u_{2k-2} + (1 - u_{2k-2})u_{2k-1}$$
$$a_k = \sqrt{(1 - u_{2k-2})u_{2k-1}(1 - u_{2k-1})u_{2k}}$$

- We call the measure μ_n with moments $m_k(\mu_n) = (J_n^k)_{1,1}$ the spectral measure of J_n.
- A moment comparison shows:

$$\mu_n = \sum_{i=1}^{n} w_i \delta_{\lambda_i}$$

  where λ_i are the (distinct) eigenvalues of J_n with orthonormal eigenvectors v_i and $w_i = v_{i,1}^2$.

Relation between moments and measures

We have a continuous bijection between:

- a moment vector $(m_1, \ldots, m_{2n-1}) \in \mathrm{Int}\mathcal{M}_{2n-1}([0,1])$
- canonical moments $(u_1, \ldots, u_{2n-1}) \in (0,1)^{2n-1}$
- a spectral measure μ_n with n support points in $(0,1)$

This can be extended to infinite sequences, there is a 1-to-1 correspondence between:

- a moment vector $(m_1, m_2, \ldots) \in \mathrm{Int}\mathcal{M}([0,1])$
- canonical moments $(u_1, u_2, \ldots) \in (0,1)^{\mathbb{N}}$
- a spectral measure μ (of an infinite tridiagonal matrix J) with infinitely many support points

In particular, the infinite canonical moment sequence $(\frac{1}{2}, \frac{1}{2}, \ldots)$ gives the infinite matrix J with spectral measure the arcsine distribution μ_{arc}.

From convergence of moments to measures

The mapping from $(u_1, \dots, u_{2n-1}, 0, \dots)$ to the spectral measure μ_n is continuous, the LDP for the canonical moments can be transferred with the contraction principle.

Chang, Kemperman, Studden (1993); Gamboa, Lozada-Chang (2004):

For $(m_1^{(n)}, \dots, m_{2n-1}^{(n)})$ uniformly distributed on $\mathcal{M}_{2n-1}([0,1])$ with lower principal representation μ_n,

$$\mu_n \xrightarrow[n\to\infty]{\mathbb{P}} \mu_{\mathrm{arc}}$$

and $(\mu_n)_{n\geq 1}$ satisfies the LDP with speed n and rate function

$$\mathcal{H}(\mu) = \sum_{k=1}^{\infty} -\log(u_k(\mu) - u_k(\mu)^2) - \log(4).$$

Note that the rate function has unique minimizer μ_{arc} and $\mathcal{H}(\mu) = \infty$, if μ has finite support.

Rewriting the rate function

The rate function can be rewritten as

$$\begin{aligned}\mathcal{H}(\mu) &= \lim_{n\to\infty} \sum_{k=1}^{n} -\log(u_k(\mu) - u_k(\mu)^2) - \log(4) \\ &= \lim_{n\to\infty} -\log\left(4^n \prod_{k=1}^{n} u_k(\mu)(1-u_k(\mu))\right) \\ &= \lim_{n\to\infty} -\log\left(4^n(m_{n+1}^{+}(\mu) - m_{n+1}^{-}(\mu))\right)\end{aligned}$$

Grenander, Szegő (1958):

$$\lim_{n\to\infty} -\log\left(4^n(m_{n+1}^{+}(\mu) - m_{n+1}^{-}(\mu))\right) = \mathcal{K}(\mu_{\mathrm{arc}}|\mu)$$

where $\mathcal{K}(\mu_{\mathrm{arc}}|\mu) = \int \log \frac{\mathrm{d}\mu_{\mathrm{arc}}}{\mathrm{d}\mu} \mathrm{d}\mu_{\mathrm{arc}} = -\int \log(w)\, \mathrm{d}\mu_{\mathrm{arc}}$ is the Kullback-Leibler divergence, rewritten for $\mathrm{d}\mu(x) = w(x)\mathrm{d}\mu_{\mathrm{arc}}(x) + \mathrm{d}\mu_s(x)$.

A first sum rule

- $$\mathcal{H}(\mu) = \sum_{k=1}^{\infty} -\log(4(u_k(\mu) - u_k^2(\mu))) = \mathcal{K}(\mu_{\mathrm{arc}}|\mu)$$
- The rate function depends only on the part absolutely continuous w.r.t. μ_{arc}.
- For $\bar{m} \in \mathrm{Int}\mathcal{M}_\ell([0,1])$, let $\mathcal{S}(\bar{m})$ be the set of p.m. with those first moments. Then the rate function is minimized over $\mathcal{S}(\bar{m})$ by choosing $u_k = \frac{1}{2}$ for $k > \ell$, so that

$$\begin{aligned}\log(4^\ell(m_{\ell+1}^+ - m_{\ell+1}^-)) &= \sum_{k=1}^{\ell} -\log(4(\bar{u}_k - \bar{u}_k^2)) \\ &= \inf_{\mu \in \mathcal{S}(\bar{m})} \mathcal{H}(\mu) = \inf_{\mu \in \mathcal{S}(\bar{m})} \mathcal{K}(\mu_{\mathrm{arc}}|\mu).\end{aligned}$$

- The infimum is attained at a unique $\bar{\mu}$ with canonical moments $u_k = \frac{1}{2}$ for $k > \ell$. $\bar{\mu}$ is of Bernstein-Szegő-type,

$$d\bar{\mu}(x) = \frac{1}{Q(x)\sqrt{x(1-x)}}dx, \qquad Q \text{ polynomial.}$$

The constrained moment problem

- Fix $\bar{m} \in \mathrm{Int}\mathcal{M}_\ell([0,1])$ and let $\bar{\mu}_n$ have moments uniformly distributed on

$$\left\{ m^{(2n-1)} \in \mathcal{M}_{2n-1}([0,1]) : m_k^{(2n-1)} = \bar{m}_k \text{ for } k \leq \ell \right\}.$$

- Let φ be the continuous mapping, which sets the first ℓ moments to $\bar{m}$, then by the contraction principle, $(\bar{\mu}_n)_n$ satisfies the LDP in $\mathcal{S}(\bar{m})$ with rate function

$$\begin{aligned}
\bar{\mathcal{H}}(\mu) &= \inf_{\nu : \varphi(\nu) = \mu} \mathcal{H}(\nu) \\
&= \sum_{k=\ell+1}^{\infty} -\log(4(u_k(\mu) - u_k(\mu)^2)) \\
&= \sum_{k=1}^{\infty} -\log(4(u_k(\mu) - u_k(\mu)^2)) - \sum_{k=1}^{\ell} -\log(4(\bar{u}_k - \bar{u}_k^2)) \\
&= \mathcal{K}(\mu_{\mathrm{arc}} | \mu) - \mathcal{K}(\mu_{\mathrm{arc}} | \bar{\mu}).
\end{aligned}$$

- In particular, $\bar{\mu}_n$ converges exponentially fast to $\bar{\mu}$.

Thank you for your attention!

- H. Dette, W. Studden, *The theory of canonical moments with applications in statistics, probability, and analysis*, Wiley, 1997
- F. Chang, J. Kemperman, W. Studden, *A normal limit theorem for moment sequences*, The Annals of Probability 21, 1993
- F. Gamboa, L. Lozada-Chang, *Large deviations for random power moment problem*, The Annals of Probability 32, 2004
- H. Dette, D. Tomecki, M. Venker, *Random moment problems under constraints*, The Annals of Probability 48, 2020