Supplementary Material for the Paper: "Stochastic Bandit Based on Empirical Moments"

Junya Honda **Akimichi Takemura**

The University of Tokyo

In this document we summarize results on Tchebycheff systems and moment spaces and prove Theorem 3.

B Tchebycheff Systems and Moment Spaces

In this section all functions and measures are defined on $[a,b]\,(a<b)$ whereas they are on $[0,1]$ elsewhere. For any set of points $\{x_1,\cdots,x_l\}$, we always assume $a\le x_1<x_2<\cdots<x_l\le b$.

Definition 1. *Let $u_0(x),\cdots,u_d(x)$ denote continuous real-valued functions on $[a,b]$. These functions are called a Tchebycheff system (or T-system) if determinants*

$$\det\begin{pmatrix} u_0(x_0) & u_1(x_0) & \cdots & u_d(x_0)\\ u_0(x_1) & u_1(x_1) & \cdots & u_d(x_1)\\ \vdots & \vdots & & \vdots\\ u_0(x_d) & u_1(x_d) & \cdots & u_d(x_d)\end{pmatrix} \tag{15}$$

are positive for all $\{x_0,\cdots,x_d\}$.

A typical T-system is $u_i(x)=x^i\,(i=0,1,\cdots,d)$, where (15) is represented as the Vandermonde determinant

$$\det\begin{pmatrix} 1 & x_0 & \cdots & x_0^d\\ 1 & x_1 & \cdots & x_1^d\\ \vdots & \vdots & & \vdots\\ 1 & x_d & \cdots & x_d^d\end{pmatrix} = \prod_{0\le i<j\le d}(x_j-x_i)>0\,.$$

Let $Z(u)$ of a function $u(x)$ denote the number of distinct points $x\in[a,b]$ such that $u(x)=0$. Then T-systems are discriminated by the following proposition.

Proposition 3 (Karlin and Studden (1966), Chap. I, Theorem 4.1)**.** *If a system $\{u_i\}_{i=0}^d$ of continuous functions on $[a,b]$ satisfies $Z(u)\le d$ for all*

$$u(x)=\sum_{i=0}^d a_i u_i(x),\ \ \{a_i\}\in\mathbb{R}^{d+1}\setminus\{0^{d+1}\}\,,$$

then $(u_0,u_1,\cdots,u_{d-1},u_d)$ or $(u_0,u_1,\cdots,u_{d-1},-u_d)$ is a T-system.

Lemma 3. *For any p and $q>0$ satisfying $b<p/q$, $(1,x,\cdots,x^d,-\log(p-qx))$ is a T-system on $[a,b]$.*

Proof. Let $b'\in(b,\,p/q)$ be sufficiently close to p/q and consider function

$$u(x)=\sum_{m=0}^d a_m x^m + a_{d+1}\log(p-qx)$$

on $x\in[a,b']$. Since the derivative of $u(x)$ is written as

$$\frac{\mathrm{d}u(x)}{\mathrm{d}x}=\frac{(p-qx)\sum_{m=1}^d a_m x^{m-1}-a_{d+1}q}{p-qx}\,,$$

$u(x)$ has at most d extreme points in $[a,b']$. Therefore $Z(u)\le d+1$ and $(1,x,\cdots,x^d,\log(p-qx))$ or $(1,x,\cdots,x^d,-\log(p-qx))$ is a T-system on $[a,b']$ from Prop. 3.

The determinant (15) for the system $(1,x,\cdots,x^d,\log(p-qx))$ is written as

$$\det\begin{pmatrix} 1 & x_0 & \cdots & x_0^d & \log(p-qx_0)\\ 1 & x_1 & \cdots & x_1^d & \log(p-qx_1)\\ \vdots & \vdots & & \vdots & \vdots\\ 1 & x_{d+1} & \cdots & x_{d+1}^d & \log(p-qx_{d+1})\end{pmatrix}$$
$$=\sum_{m=0}^{d+1}(-1)^{d+m+1}\left(\prod_{0\le i<j\le d+1:\,i,j\neq m}(x_j-x_i)\right)\cdot\log(p-qx_m)\,.$$

For the case that $x_{d+1}=b'$ with $b'\uparrow p/q$, $\log(p-qx_{d+1})$ goes to $-\infty$ and the sign of the determinant is controlled by the term involving $\log(p-qx_{d+1})$, which is written as

$$(-1)^{2d+2}\left(\prod_{0\le i<j\le d+1:\,i,j\neq d+1}(x_j-x_i)\right)\log(p-qx_{d+1})<0\,.$$

Then, $(1, x, \cdots, x^d, \log(p - qx))$ cannot be a T-system on $[a, b']$ for b' sufficiently close to p/q and therefore $(1, x, \cdots, x^d, -\log(p - qx))$ has to be a T-system on $[a, b']$. From the definition of T-system, it also is a T-system on $[a, b] \subset [a, b']$. □

Let $\mathcal{V}$ be the family of positive measures on $[a, b]$ and define a subset $\mathcal{V}(\tilde{\boldsymbol{M}})$ of $\mathcal{V}$ for a vector $\tilde{\boldsymbol{M}} = (M_0, M_1, \cdots, M_d)$ as

$$\mathcal{V}(\tilde{\boldsymbol{M}}) = \left\{\sigma \in \mathcal{V} : \forall m \in \{0, \cdots, d\}, \int_a^b x^m \mathrm{d}\sigma(x) = M_m\right\}. \tag{16}$$

The notion of *moment spaces* is essential to examine properties of T-systems.

Definition 2. *The moment space $\mathcal{M}_{d+1} \subset \mathbb{R}^{d+1}$ with respect to the T-system $\{u_i\}$ is given by*

$$\mathcal{M}_{d+1} \equiv \left\{\left(\int_a^b u_0(x)\mathrm{d}\sigma(x), \cdots, \int_a^b u_d(x)\mathrm{d}\sigma(x)\right) : \sigma \in \mathcal{V}\right\}.$$

Consider the case that $\tilde{\boldsymbol{M}} \in \mathcal{M}_{d+1}$ satisfies

$$M_m = \sum_{i=1}^{l} f_i u_m(x_i) \quad (m = 0, \cdots, d) \tag{17}$$

with $x_1, \cdots, x_l \in [a, b]$ and $f_1, \cdots, f_l > 0$ for any finite l. We call such an expression *representation* of $\tilde{\boldsymbol{M}}$. A representation of $\tilde{\boldsymbol{M}}$ corresponds uniquely to the measure

$$\sigma = \sum_{i=1}^{l} f_i \delta(x_i) \in \mathcal{V}$$

for the delta measure $\delta(x)$ at point x. We sometimes identify the measure σ with the representation of $\tilde{\boldsymbol{M}}$. The measure σ is a probability measure if $\sum_i f_i = 1$.

Similarly to the index of the measure given in Sect. 4, define the index of the representation (17) as the number of the points $(x_1, \cdots, x_l)$ under the special convention that the points a, b are counted as one half. A representation is called *principal* if its index is $(d + 1)/2$. Furthermore, a principal representation is *upper* if $(x_1, \cdots, x_l)$ contains b and *lower* otherwise.

For the proof of Theorem 3, it is necessary to study the nature on the set $\mathcal{V}(\tilde{\boldsymbol{M}})$. It differs according to whether $\tilde{\boldsymbol{M}}$ is a boundary point of $\mathcal{M}_{d+1}$ or an interior point of $\mathcal{M}_{d+1}$.

Proposition 4 (Karlin and Studden (1966), Chap. II, Theorem 2.1)**.** *$\tilde{\boldsymbol{M}} \in \mathcal{M}_{d+1}$ is a boundary point of $\mathcal{M}_{d+1}$ if and only if there exists a representation of $\tilde{\boldsymbol{M}}$ with index at most $d/2$. Moreover, if $\tilde{\boldsymbol{M}}$ is a boundary point of $\mathcal{M}_{d+1}$ then $\mathcal{V}(\tilde{\boldsymbol{M}})$ has a unique element.*

Proposition 5 (Karlin and Studden (1966), Chap. II, Corollary 3.1)**.** *If $\tilde{\boldsymbol{M}}$ is an interior point of $\mathcal{M}_{d+1}$ then there exist precisely one upper and one lower principal representations of $\tilde{\boldsymbol{M}}$.*

We use Prop. 5 implicitly in Prop. 6 below. Prop. 6 is the main result of this section.

Proposition 6 (Karlin and Studden (1966), Chap. III, Theorem 1.1)**.** *Assume $(u_0, u_1, \cdots, u_d)$ and $(u_0, u_1, \cdots, u_d, h)$ are T-systems and $\tilde{\boldsymbol{M}}$ is an interior point of $\mathcal{M}_{d+1}$. Then*

$$\max_{\sigma \in \mathcal{V}(\tilde{\boldsymbol{M}})} \int_a^b h(x)\mathrm{d}\sigma(x)$$

is attained uniquely by $\bar{\sigma}$, the upper principal representation of $\tilde{\boldsymbol{M}}$. Similarly,

$$\min_{\sigma \in \mathcal{V}(\tilde{\boldsymbol{M}})} \int_a^b h(x)\mathrm{d}\sigma(x)$$

is attained uniquely by $\underline{\sigma}$, the lower principal representation of $\tilde{\boldsymbol{M}}$.

C Proof of Theorem 3

We omit the proof of (ii) (iii) of Theorem 3 since it is obtained as a direct application of Props. 3 and 6. Theorem 3 (i) is proved by the results in the previous section and the basic result on the existence of saddle-points in the following. For a function $\varphi(x, y) : \mathcal{X} \times \mathcal{Y} \to [-\infty, +\infty]$, a point $(\bar{x}, \bar{y}) \in \mathcal{X} \times \mathcal{Y}$ is called a saddle-point if $\varphi(\bar{x}, y) \le \varphi(\bar{x}, \bar{y}) \le \varphi(x, \bar{y})$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. A necessary and sufficient condition for a saddle-point is

$$\sup_{y \in \mathcal{Y}} \varphi(\bar{x}, y) = \inf_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \varphi(x, y) = \sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \varphi(x, y) = \inf_{x \in \mathcal{X}} \varphi(x, \bar{y}).$$

Proposition 7 (Minimax Theorem (Neumann, 1928))**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be compact subsets of topological vector spaces $\mathcal{V}$ and $\mathcal{U}$, respectively. Let $\varphi(x, y) : \mathcal{X} \times \mathcal{Y} \to [-\infty, +\infty]$ be a function such that $\varphi(\cdot, y)$ is convex and lower-semicontinuous for any fixed y and $\varphi(x, \cdot)$ is concave and upper-semicontinuous for any fixed x. Then there exists a saddle point $(\bar{x}, \bar{y}) \in \mathcal{X} \times \mathcal{Y}$.*

In the following proof, we regard a probability measure F as an element of the family $\mathcal{V}$ of positive measures on

$[0,1]$ to exploit the results in the previous section. By letting $\tilde{\boldsymbol{M}} := (1, M_1, \cdots, M_d)$ for $\boldsymbol{M} = (M_1, \cdots, M_d)$, $D^{(d)}_{\min}(\boldsymbol{M}, \mu)$ is rewritten as

$$D_{\min}(\boldsymbol{M}, \mu) = \inf_{F \in \mathcal{V}(\tilde{\boldsymbol{M}})} \max_{0 \le \nu \le \frac{1}{1-\mu}} \mathrm{E}_F[\log(1-(X-\mu)\nu)] \,, \quad (18)$$

where $\mathcal{V}(\tilde{\boldsymbol{M}})$ is the set of positive measures with $0, 1, \cdots, d$-th moments equal to $\tilde{\boldsymbol{M}}$, which is formally defined in (16).

Proof of Theorem 3 (i). Let

$$\mathcal{M}_{d+1} = \left\{ \left(\int_0^1 x^0 \mathrm{d}F, \cdots, \int_0^1 x^d \mathrm{d}F \right) : F \in \mathcal{V} \right\}$$

be the moment space with respect to the system $(1, x, \cdots, x^d)$. Since $\tilde{\boldsymbol{M}}$ is assumed to have a representation with the index larger than or equal to $(d+1)/2$, $\tilde{\boldsymbol{M}}$ cannot be a boundary point of $\mathcal{M}_{d+1}$ from Prop. 4. Therefore $\tilde{\boldsymbol{M}}$ is an interior point of $\mathcal{M}_{d+1}$.

Consider applying the minimax theorem to (18). First, $\mathcal{F} \supset \mathcal{V}(\tilde{\boldsymbol{M}})$ is compact with respect to the Lévy distance and $\mathrm{E}_F[\log(1-(X-\mu)\nu)]$ is linear in $F \in \mathcal{V}$ for any fixed ν. Next, $\mathrm{E}_F[\log(1-(X-\mu)\nu)]$ is upper-semicontinuous and concave in ν for any fixed F. Then we obtain from the minimax theorem that

$$D_{\min}(\boldsymbol{M}, \mu) = \max_{0 \le \nu \le \frac{1}{1-\mu}} \inf_{F \in \mathcal{V}(\tilde{\boldsymbol{M}})} \mathrm{E}_F[\log(1-(X-\mu)\bar{\nu})] \,.$$

Now we show that it suffices to consider the case $\nu < (1-\mu)^{-1}$. From Prop. 5, $\mathcal{V}(\tilde{\boldsymbol{M}})$ contains the upper principal representation $\bar{F}$ of $\tilde{\boldsymbol{M}}$, which has a positive weight at $x = 1$, i.e., $\bar{F}(\{1\}) > 0$. Therefore it holds for $\nu = (1-\mu)^{-1}$ from $\log 0 = -\infty$ that

$$\inf_{F \in \mathcal{V}(\boldsymbol{M})} \mathrm{E}_F[\log(1-(X-\mu)\nu)] \le \mathrm{E}_{\bar{F}}[\log(1-(X-\mu)\nu)] = -\infty$$

and therefore

$$D_{\min}(\boldsymbol{M}, \mu) = \max_{0 \le \nu < \frac{1}{1-\mu}} \inf_{F \in \mathcal{V}(\tilde{\boldsymbol{M}})} \mathrm{E}_F[\log(1-(X-\mu)\nu)] \,.$$

For $\nu < (1-\mu)^{-1}$, $(1, x, \cdots, x^d, -\log(1-(x-\mu)\nu))$ is a T-system on $[0,1]$ from Lemma 3 with $p := 1 + \mu\nu$ and $q := \nu$. Therefore, we obtain from Prop. 6 that

$$\begin{aligned}
&\max_{0 \le \nu < \frac{1}{1-\mu}} \inf_{F \in \mathcal{V}(\tilde{\boldsymbol{M}})} \mathrm{E}_F[\log(1-(X-\mu)\nu)] \\
&= \max_{0 \le \nu < \frac{1}{1-\mu}} \left\{ - \sup_{F \in \mathcal{V}(\tilde{\boldsymbol{M}})} \mathrm{E}_F[-\log(1-(X-\mu)\nu)] \right\} \\
&= \max_{0 \le \nu < \frac{1}{1-\mu}} \{ -\mathrm{E}_{\bar{F}}[-\log(1-(X-\mu)\nu)] \} \\
&= \max_{0 \le \nu < \frac{1}{1-\mu}} \mathrm{E}_{\bar{F}}[\log(1-(X-\mu)\nu)] = D_{\min}(\bar{F}, \mu) \,,
\end{aligned}$$

where $\bar{F}$ is the upper principal representation of $\tilde{\boldsymbol{M}}$. Since the upper principal representation is unique, it can be written as a unique solution of (8) and (9). □