About Yor's problem.

5. An equivalent measure: proofs

Boris Tsirelson (Tel Aviv University)

Deducing Theorem 103 from Lemma 215

301. Choose $\varepsilon_0, \varepsilon_1, \ldots \in (0,1)$ and $N_0, N_1, \ldots \in \mathbb{N}$ suct that

$$N_{k+1} \le 2N_k, \tag{302}$$

$$\prod_k \Bigl(1 - \frac{C}{\varepsilon_k N_k} \Bigr) > 0 \tag{303}$$

(C as in (216)), and

$$\sum_k \varepsilon_k^2 N_k < \infty. \tag{304}$$

(For example, $N_k = 2^k N_0$ and $\varepsilon_k = \theta^k \varepsilon_0$ with $\theta \in (1/2,\, 1/\sqrt{2})$ and $\varepsilon_0 N_0 > C$).

Apply Lemma 215 for $n = N_k$ and $\varepsilon = \varepsilon_k$. It gives us 2^{2N_k} measures from $\mathcal{P}_{\varepsilon_k}^{N_k}$; choose $2^{N_{k+1}}$ of them, and number them by N_{k+1} indices with two values each. So, we have $\mu_k(\xi_1,\ldots,\xi_{N_{k+1}}) \in \mathcal{P}_{\varepsilon_k}^{N_k}$ (each ξ_i being 0 or 1) such that

$$\mathrm{KR}^{N_k}(\mu_k(\xi'_1,\ldots,\xi'_{N_{k+1}}), \mu_k(\xi''_1,\ldots,\xi''_{N_{k+1}})) \ge 1 - \frac{C}{\varepsilon_k N_k} \tag{305}$$

unless $\xi'_1 = \xi''_1, \ldots, \xi'_{N_{k+1}} = \xi''_{N_{k+1}}$. We may consider μ_k as a Markov transition probability from $\{0,1\}^{N_{k+1}}$ to $\{0,1\}^{N_k}$.

306. Divide the set $\{\ldots,-2,-1,0\}$ into segments of length N_k; that is, take $M_k = N_0 + \ldots + N_k$ and $\Delta_{-k} = \{-M_k+1, -M_k+2, \ldots, -M_{k-1}\}$ (here $M_{-1} = 0$). We claim that there exists a random sequence $\{X_k\}_{k\le 0}$, each X_k taking values $0,1$ only, such that

$$\mathbb{P}\,\{\, X(\Delta_{-k}) \,|\, X(\Delta_{-(k+1)}) \,\} = \mu_k(X(\Delta_{-(k+1)}))(X(\Delta_{-k})), \tag{307}$$

(308) the distribution of $\{X_k\}_{k\le 0}$ is equivalent to the Bernoulli-1/2 measure.

Indeed, to construct the needed distribution, it is enough to prove convergence of the product for its density

$$\prod_k p_{-k} = \prod_k p_{-k}\,(\, x(\Delta_{-(k+1)}),\, x(\Delta_{-k})\,) \tag{309}$$

in L_2 on the Bernoulli-1/2 measure; here p_{-k} is the density of μ_k. From (214) we obtain

$$\mathbb{E}\,(p_{-k}^2 \mid x_{-\infty}^{-M_k}) \le (1+\varepsilon_k^2)^{N_k} \le \exp(\varepsilon_k^2 N_k).$$

Hence $\mathbb{E}p_{-k}^2 \le \exp(\varepsilon_k^2 N_k)$, and then*

$$\begin{aligned}\mathbb{E}\,(\,p_{-(k+1)}^2\, p_{-k}^2\,) = \mathbb{E}\,\mathbb{E}\,(\,p_{-(k+1)}^2 p_{-k}^2 \mid x_{-\infty}^{-M_k}\,) = \mathbb{E}\,(\,p_{-(k+1)}^2\;\mathbb{E}\,(p_{-k}^2 \mid x_{-\infty}^{-M_k}\,)\,) \le \\ \le \mathbb{E}\,(\,p_{-(k+1)}^2\;\exp(\varepsilon_k^2 N_k)\,) = \exp(\varepsilon_k^2 N_k)\,\mathbb{E}p_{-(k+1)}^2 \le \exp(\varepsilon_k^2 N_k + \varepsilon_{k+1}^2 N_{k+1}).\end{aligned}$$

In the same way

$$\mathbb{E}\prod_{k=s}^{t} p_k^2 \le \exp\left(\sum_{k=s}^{t} \varepsilon_k^2 N_k\right). \tag{310}$$

Taking into account that $\mathbb{E}\prod_{k=s}^{t} p_k = 1$, we obtain

$$\mathbb{E}\left(\left(\prod_{k=s}^{t} p_k\right) - 1\right)^2 \le \exp\left(\sum_{k=s}^{t} \varepsilon_k^2 N_k\right) - 1,$$

and, by using the above trick with conditional expectation once more,

$$\mathbb{E}\left(\left(\left(\prod_{k=r}^{s-1} p_k\right) - 1\right)\cdot\prod_{k=s}^{t} p_k\right)^2 \le \left(\exp\left(\sum_{k=r}^{s-1} \varepsilon_k^2 N_k\right) - 1\right)\cdot\exp\left(\sum_{k=s}^{t} \varepsilon_k^2 N_k\right),$$

that is,

$$\mathbb{E}\left(\prod_{k=r}^{t} p_k - \prod_{k=s}^{t} p_k\right)^2 \le \exp\left(\sum_{k=r}^{t} \varepsilon_k^2 N_k\right) - \exp\left(\sum_{k=s}^{t} \varepsilon_k^2 N_k\right) \tag{311}$$

for any $r \le s \le t \le 0$. Now (304) implies convergence of (309).

312. Define ρ_0 as

$$\rho_0\,(\,\{x'_k\}_k,\,\{x''_k\}_k\,) = \max_{k\in\Delta_0} |x'_k - x''_k| \tag{313}$$

and consider the corresponding chain of metrics $\{\rho_k\}_k$ (see (209)). From (305) and (307) we obtain

$$\rho_{-M_0}\,(\,\{x'_k\}_k,\,\{x''_k\}_k\,) \ge \left(1 - \frac{C}{\varepsilon_0 N_0}\right)\max_{k\in\Delta_{-1}} |x'_k - x''_k|$$

* The rest of Item 306 is undoubtedly a well-known argument; now I am lazy to find a relevant reference.

for any $\{x'_k\}_k$, $\{x''_k\}_k$. Continuing in this way, we obtain

$$\rho_{-M_n}\left(\{x'_k\}_k,\, \{x''_k\}_k\right) \geq \left(\prod_{m=0}^{n}\left(1-\frac{C}{\varepsilon_m N_m}\right)\right) \max_{k\in\Delta_{-(n+1)}} |x'_k - x''_k|. \tag{314}$$

It follows that

$$\overline{\rho}_{-\infty} \geq \prod_m \left(1-\frac{C}{\varepsilon_m N_m}\right) > 0$$

by virtue of (303). So, Theorem 103 is deduced from Lemma 215.

Proof of Lemma 215

401. Another definition of the Kantorovich-Rubinstein metric, equivalent to one given in Item 7, follows:

$$\rho_{\mathrm{KR}}(\mu,\nu) = \sup_f \left|\int f\,d\mu - \int f\,d\nu\right|, \tag{402}$$

where supremum is taken over all $f : \mathcal{X} \to \mathbb{R}$ satisfying Lipschitz condition:

$$\forall x, y \in \mathcal{X} \quad |f(x) - f(y)| \leq \rho(x,y). \tag{403}$$

404. **Lemma.** Let each of two probability measures μ, ν be concentrated on a two-point set, μ on $\{a,b\}$ and ν on $\{c,d\}$, in a metric space. Suppose that

$$\begin{aligned} \rho(a,c) \leq 1, &\quad \rho(b,d) \leq 1,\\ \rho(a,d) = 1, &\quad \rho(b,c) = 1. \end{aligned}$$

Then

$$\begin{aligned} \rho_{\mathrm{KR}}(\mu,\nu) &= \frac{1}{2}(\mu\{a\}+\nu\{c\})\rho(a,c) + \frac{1}{2}(\mu\{b\}+\nu\{d\})\rho(b,d) + \\ &+ \left(1-\frac{1}{2}\rho(a,c)-\frac{1}{2}\rho(b,d)\right)|\mu\{a\}-\nu\{c\}|. \end{aligned} \tag{405}$$

406. Proof.* It is enough to consider the case when $\mu\{a\} \geq \nu\{c\}$. Take a Lipschitz function $f : \{a,b,c,d\} \to \mathbb{R}$ such that

$$\begin{aligned} f(c) - f(a) &= \rho(a,c),\\ f(d) - f(b) &= \rho(b,d),\\ f(d) - f(a) &= \rho(a,d) = 1. \end{aligned}$$

* A general method is known for calculating Kantorovich-Rubinstein distance for a finite number of points by solving a linear programming problem.

It is easy to see that such f exists and is unique up to an additive constant. We have

$$\int f\,d\nu - \int f\,d\mu = f(c)\nu\{c\} + f(d)\nu\{d\} - f(a)\mu\{a\} - f(b)\mu\{b\} =$$
$$= (f(c)-f(a))\nu\{c\} + (f(d)-f(b))\mu\{b\} + (f(d)-f(a))(\mu\{a\}-\nu\{c\}),$$

because $\mu\{a\}-\nu\{c\} = \nu\{d\}-\mu\{b\}$. Hence

$$\begin{aligned}\rho_{\text{KR}}(\mu,\nu) &\ge (f(c)-f(a))\nu\{c\} + (f(d)-f(b))\mu\{b\} + (f(d)-f(a))(\mu\{a\}-\nu\{c\}) = \\ &= \rho(a,c)\nu\{c\} + \rho(b,d)\mu\{b\} + \mu\{a\} - \nu\{c\} = \\ &= \frac{1}{2}(\mu\{a\}+\nu\{c\})\rho(a,c) + \frac{1}{2}(\mu\{b\}+\nu\{d\})\rho(b,d) + \\ &\quad + \left(1 - \frac{1}{2}\rho(a,c) - \frac{1}{2}\rho(b,d)\right)|\mu\{a\}-\nu\{c\}|.\end{aligned}$$

An opposite inequality* can be obtained from the Item 7. Take

$$\xi(t) = \begin{cases} a, & \text{when } 0 < t < \mu\{a\}; \\ b, & \text{when } \mu\{a\} < t < 1; \end{cases}$$
$$\psi(t) = \begin{cases} c, & \text{when } 0 < t < \nu\{c\}; \\ d, & \text{when } \nu\{c\} < t < 1. \end{cases}$$

Then

$$\begin{aligned}\rho_{\text{KR}}(\mu,\nu) &\le \int_0^1 \rho(\xi(t),\psi(t))\,dt = \\ &= \int_0^{\nu\{c\}} \rho(a,c)\,dt + \int_{\nu\{c\}}^{\mu\{a\}} \rho(a,d)\,dt + \int_{\mu\{a\}}^1 \rho(b,d)\,dt = \\ &= \nu\{c\}\rho(a,c) + (\mu\{a\}-\nu\{c\})\rho(a,d) + (1-\mu\{a\})\rho(b,d) = \\ &= \frac{1}{2}(\mu\{a\}+\nu\{c\})\rho(a,c) + \frac{1}{2}(\mu\{b\}+\nu\{d\})\rho(b,d) + \\ &\quad + \left(1 - \frac{1}{2}\rho(a,c) - \frac{1}{2}\rho(b,d)\right)|\mu\{a\}-\nu\{c\}|,\end{aligned}$$

and Lemma 404 is proved.

407. Note some other forms of (405):

(408) $$\rho_{\text{KR}}(\mu,\nu) = \rho(a,c)\min(\mu\{a\},\nu\{c\}) + \rho(b,d)\min(\mu\{b\},\mu\{d\}) + |\mu\{a\}-\nu\{c\}|;$$

(409) $$1 - \rho_{\text{KR}}(\mu,\nu) = (1-\rho(a,c))\min(\mu\{a\},\nu\{c\}) + (1-\rho(b,d))\min(\mu\{b\},\mu\{d\}).$$

* In fact, the opposite inequality is not needed for our purpose.

410. Introduce a probability measure M_ε^n on the set $\mathcal{P}_\varepsilon^n$ as the uniform distribution on the finite set $\operatorname{ex}(\mathcal{P}_\varepsilon^n)$ of all extremal points of $\mathcal{P}_\varepsilon^n$. That is, M_ε^n-distributed random element of $\mathcal{P}_\varepsilon^n$ may be written as the following distribution for $(X_1, \dots, X_n)$:

$$\mathbb{P}\,(\, X_k = 1 \mid X_1 = x_1,\, \dots,\, X_{k-1} = x_{k-1}\,) = \frac{1}{2}(1 + \varepsilon\tau(x_1, \dots, x_{k-1})), \tag{411}$$

where each $\tau(x_1, \dots, x_{k-1})$ is a random variable taking the values ± 1 with probability $1/2$ each, all such variables being independent for all corteges $(x_1, \dots, x_{k-1})$ of all lengthes $0, \dots, n-1$. We intend to estimate the expression

$$F_n(\lambda, \varepsilon) = \iint \exp\left(\lambda(1 - \mathrm{KR}^n(\mu, \nu))\right)\, M_\varepsilon^n(d\mu) M_\varepsilon^n(d\nu); \tag{412}$$

we will write also

$$F_n(\lambda, \varepsilon) = \mathbb{E} \exp\left(\lambda(1 - \mathrm{KR}^n(\mu, \nu))\right), \tag{413}$$

treating μ, ν as independent random element of $\mathcal{P}_\varepsilon^n$ distributed according to M_ε^n.

414. **Lemma.**

$$F_{n+1}(\lambda, \varepsilon) = F_n\left(\frac{1-\varepsilon}{2}\lambda, \varepsilon\right) \cdot \frac{1}{2}\left(F_n\left(\frac{1-\varepsilon}{2}\lambda, \varepsilon\right) + F_n\left(\frac{1+\varepsilon}{2}\lambda, \varepsilon\right)\right).$$

Proof. Consider $\mathrm{KR}^{n+1}(\mu, \nu)$ as $\rho_0(0,1)$ following to Item 212; so,

$$\rho_0(0,1) = \rho_{1,\mathrm{KR}}(\mu_1, \nu_1),$$

where μ_1 is concentrated on two points $a = (0,0)$ and $b = (0,1)$,

$$\mu_1\{a\} = \frac{1}{2}(1 - \varepsilon\tau_\mu), \quad \mu_1\{b\} = \frac{1}{2}(1 + \varepsilon\tau_\mu),$$

τ_μ takes the values ± 1 with probability $1/2$ each. Similarly, ν_1 is concentrated on two points $c = (1,0)$ and $d = (1,1)$,

$$\nu_1\{c\} = \frac{1}{2}(1 - \varepsilon\tau_\nu), \quad \nu_1\{d\} = \frac{1}{2}(1 + \varepsilon\tau_\nu).$$

It follows from the construction of M_ε^{n+1} that the four random variables

$$\tau_\mu, \quad \tau_\nu, \quad \rho_1(a,c), \quad \rho_1(b,d)$$

are independent, and

$$\begin{aligned}
\mathbb{E}\exp(\lambda(1 - \rho_1(a,c))) &= F_n(\lambda, \varepsilon), \\
\mathbb{E}\exp(\lambda(1 - \rho_1(b,d))) &= F_n(\lambda, \varepsilon), \\
\mathbb{E}\exp(\lambda(1 - \rho_0(0,1))) &= F_{n+1}(\lambda, \varepsilon).
\end{aligned}$$

It is easy to see that $\rho_1(a,d)=1$, $\rho_1(b,c)=1$. Lemma 404 in the form (409) gives

$$1-\rho_0(0,1)=(1-\rho_1(a,c))\cdot\frac{1}{2}(1-\varepsilon\max(\tau_\mu,\tau_\nu))+(1-\rho_1(b,d))\cdot\frac{1}{2}(1+\varepsilon\min(\tau_\mu,\tau_\nu)).$$

Multiplying by λ, taking exponent function and averaging, we obtain the contribution $F_n(\frac{1-\varepsilon}{2}\lambda,\varepsilon)\cdot F_n(\frac{1+\varepsilon}{2}\lambda,\varepsilon)$ from the case $\tau_\mu=\tau_\nu$, and $F_n^2(\frac{1-\varepsilon}{2}\lambda,\varepsilon)$ from the case $\tau_\mu\neq\tau_\nu$. So, Lemma 414 is proved.

415. **Lemma.**

$$F_n(\lambda,\varepsilon)\le\exp\left(\left(1-\frac{\varepsilon}{3}\right)^n\lambda\right)$$
$$\text{for}\quad 0\le\lambda\le\frac{1}{\varepsilon\left(1-\frac{\varepsilon}{3}\right)^n}\quad\text{and}\quad 0<\varepsilon\le\frac{1}{2}.$$

Proof. First, $F_0(\lambda,\varepsilon)=e^\lambda$. Further, suppose the inequality holds for n and prove it for $n+1$. Due to Lemma 414 it is enough to show that

$$\exp\left(\left(1-\frac{\varepsilon}{3}\right)^n\cdot\frac{1-\varepsilon}{2}\lambda\right)\cdot\frac{1}{2}\left(\exp\left(\left(1-\frac{\varepsilon}{3}\right)^n\cdot\frac{1-\varepsilon}{2}\lambda\right)+\exp\left(\left(1-\frac{\varepsilon}{3}\right)^n\cdot\frac{1+\varepsilon}{2}\lambda\right)\right)\le$$
$$\le\exp\left(\left(1-\frac{\varepsilon}{3}\right)^{n+1}\lambda\right)\quad\text{for}\quad 0\le\lambda\le\frac{1}{\varepsilon\left(1-\frac{\varepsilon}{3}\right)^{n+1}}.$$

That is,

$$\frac{1}{2}\exp\left(-\left(1-\frac{\varepsilon}{3}\right)^n\cdot\frac{2\varepsilon}{3}\lambda\right)+\frac{1}{2}\exp\left(\left(1-\frac{\varepsilon}{3}\right)^n\cdot\frac{\varepsilon}{3}\lambda\right)\le 1.$$

The left-hand side is convex, so it is enough to check the inequality for $\lambda=0$ and $\lambda=\varepsilon^{-1}\left(1-\frac{\varepsilon}{3}\right)^{-n}$. But $\exp(-2/3)+\exp(1/3)\le 2$.

416. It follows from Lemma 415 that

$$\mathbb{P}\left\{\mathrm{KR}^n(\mu,\nu)\le 1-\left(1-\frac{\varepsilon}{3}\right)^n a\right\}\le\exp\left(-\frac{a-1}{\varepsilon}\right)$$

for any $a>1$, $0<\varepsilon\le 1/2$; here, as before, μ and ν are independent random elements of $\mathcal{P}_\varepsilon^n$ having distribution M_ε^n.

417. Now we are prepared to complete our proof of Lemma 215. Choose the absolute constant C suct that

$$\frac{C}{2a}\exp\left(\frac{a}{3}\right)-1\ge a\cdot 4\ln 2 \tag{418}$$

for any $a\ge C$. For given n and ε we have find $N=2^{2n}$ measures $\mu_1,\dots,\mu_N\in\mathcal{P}_\varepsilon^n$ such that $\mathrm{KR}^n(\mu_i,\mu_j)\ge 1-\frac{C}{\varepsilon n}$ for $i\neq j$. Suppose that $\varepsilon n\ge C$; otherwise we have nothing to prove. We impose an additional condition $\varepsilon \le 1/2$ and shell prove a bit more: $\mathrm{KR}^n(\mu_i, \mu_j) \ge 1 - \frac{C}{2\varepsilon n}$. It is enough, because we may use $\varepsilon/2$ instead of the given ε, if it exceeds $1/2$.

Choose $\mu_1, \dots, \mu_N \in \mathcal{P}_\varepsilon^n$ at random, independently, according to the distribution M_ε^n, and the needed inequality will be satisfied with a positive probability. To prove this fact, it is enough to show that

$$\mathbb{P}\left\{\mathrm{KR}^n(\mu_i, \mu_j) \le 1 - \frac{C}{2\varepsilon n}\right\} \le \frac{1}{N^2}.$$

Put $a = C(2\varepsilon n)^{-1}(1 - \frac{\varepsilon}{3})^{-n}$ in (416); then $a \ge C(2\varepsilon n)^{-1}\exp(\varepsilon n/3)$, and we obtain

$$\mathbb{P}\left\{\mathrm{KR}^n(\mu_i, \mu_j) \le 1 - \frac{C}{2\varepsilon n}\right\} \le \exp\left(-\frac{a-1}{\varepsilon}\right).$$

It remains to show that $\exp(-\frac{a-1}{\varepsilon}) \le N^{-2}$, that is, $\frac{a-1}{\varepsilon} \ge 2\ln N$, or

$$\frac{C}{2\varepsilon n}\exp\left(\frac{\varepsilon n}{3}\right) - 1 \ge \varepsilon n \cdot 4\ln 2,$$

what follows from (418). So, Lemma 215 is proved.