k-Sample inference via Multimarginal Optimal Transport

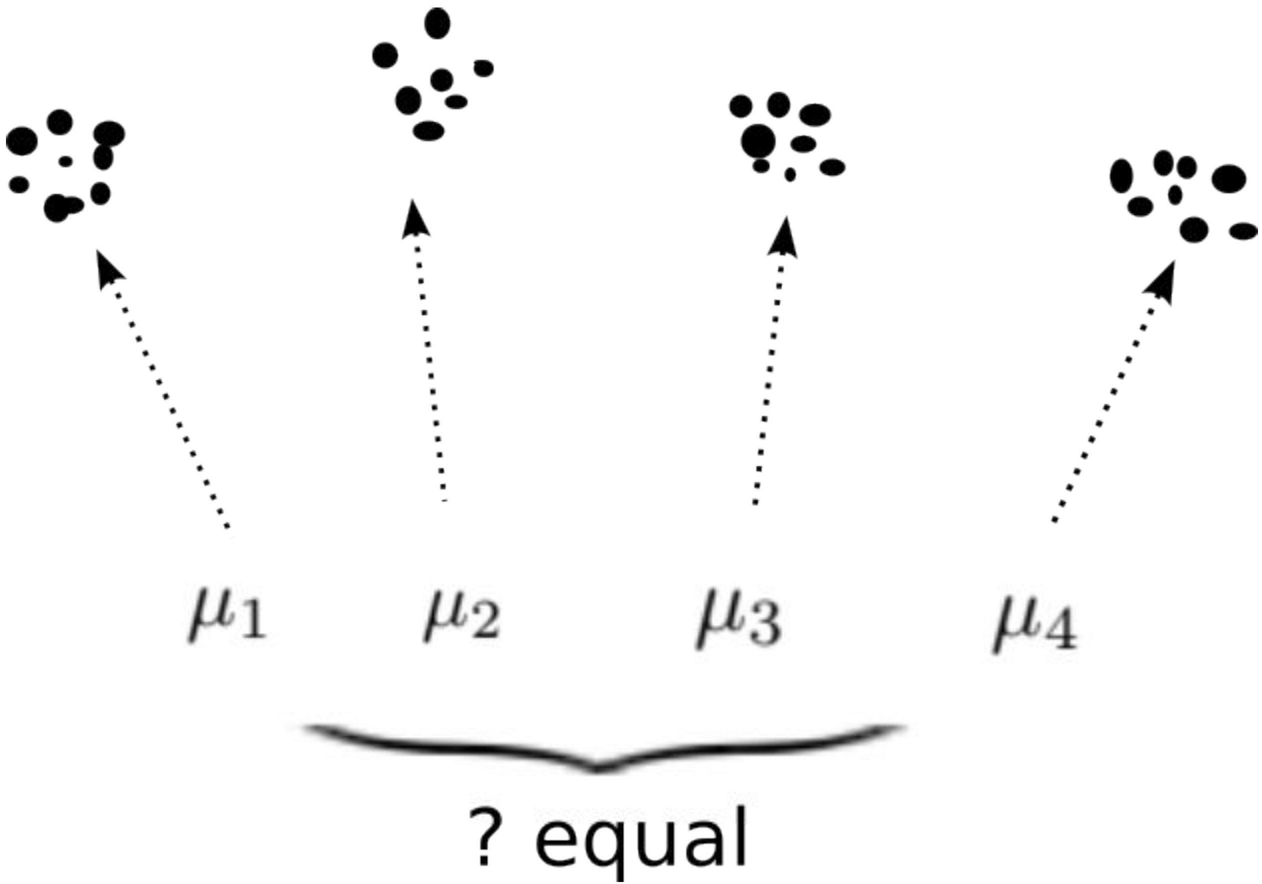
Natalia Kravtsova

Ohio State University/University of British Columbia

SIAM AN25

Optimal Transport in Natural and Data Sciences

k-sample problem

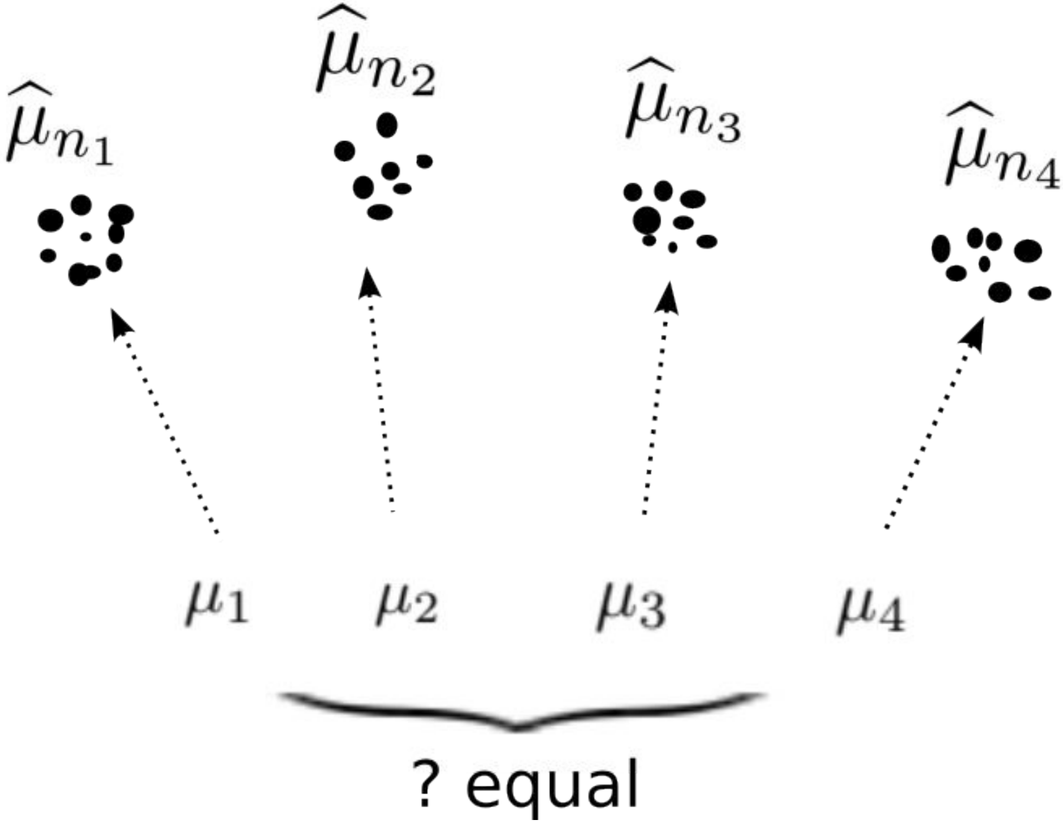


k-sample problem

Using the data,

test this

$$H_0 : \mu_1 = \cdots = \mu_k$$

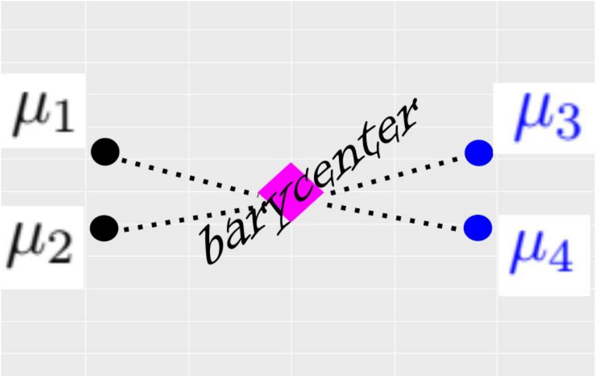


Proposed functional: $MOT(\mu)$

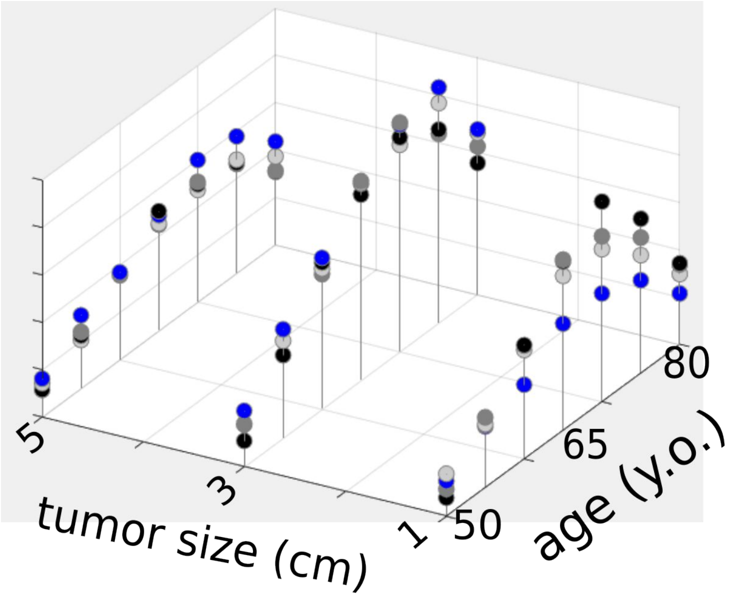
$$MOT(\mu) = \inf_{\pi \in \mathcal{C}(\mu_1, \ldots, \mu_k)} \int_{X \times \cdots \times X} c(x_1, \cdots, x_k)\, d\pi(x_1, \cdots, x_k)$$

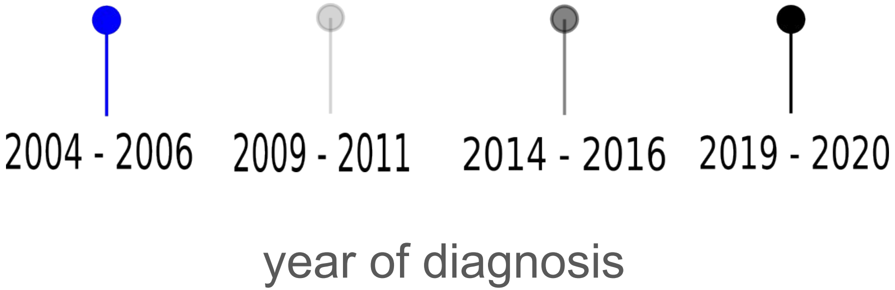
Equivalent

$$Barycenter(\mu) = \inf_{\nu \in \mathcal{P}^2(X)} \frac{1}{k} \sum_{i=1}^{k} W_2^2(\mu_i, \nu)$$



Assumption: finitely supported measures





data: https://seer.cancer.gov

$$MOT(\mu) = \inf_{\pi \in \mathcal{C}(\mu_1, \dots, \mu_k)} \int_{X \times \cdots \times X} c(x_1, \cdots, x_k) \, d\pi(x_1, \cdots, x_k)$$

is written $\min_{\pi \geq 0} \langle c, \pi \rangle$

such that $A\pi = \mu$

Observe that $$MOT(\mu) = 0 \iff H_0 \text{ is true}$$

Construct a test using behavior of random variable $MOT(\widehat{\mu}_n)$:

$$\rho_n \left(MOT(\widehat{\mu}_n) - \underbrace{MOT(\mu)}_{=0 \text{ under } H_0} \right) \xrightarrow{\text{in law}} L$$

The result

$$\rho_n \left(MOT(\widehat{\mu}_n) - MOT(\mu) \right) \xrightarrow{\text{in law}} L$$

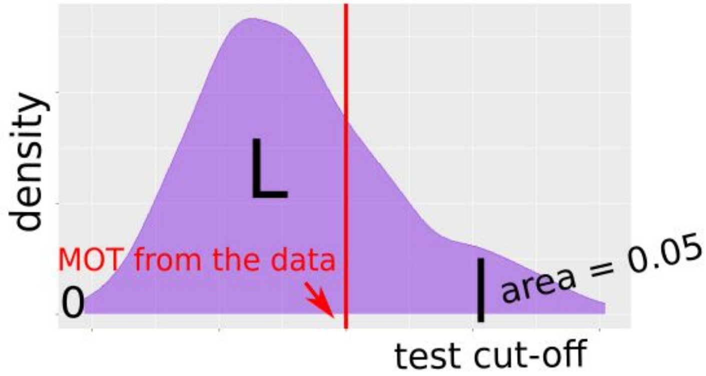
extends the result of *Sommerfeld and Munk (2018):*

$$r_n \left(W_2^2(\widehat{\mu}_{n_1}, \widehat{\mu}_{n_2}) - W_2^2(\mu_1, \mu_2) \right) \xrightarrow{\text{in law}} X$$

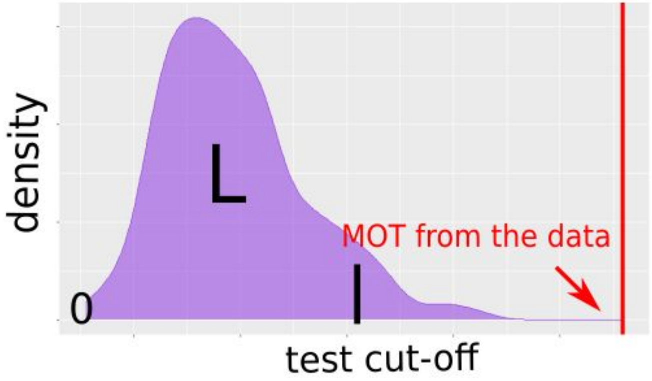
Sommerfeld & Munk (2018), J. R. Stat. Soc. Ser. B

Construct a test of $H_0 : \mu_1 = \ldots = \mu_k$ using $\rho_n \Big(MOT(\widehat{\mu}_n) - \underbrace{MOT(\mu)}_{=0 \text{ under } H_0} \Big) \xrightarrow{\text{in law}} L$

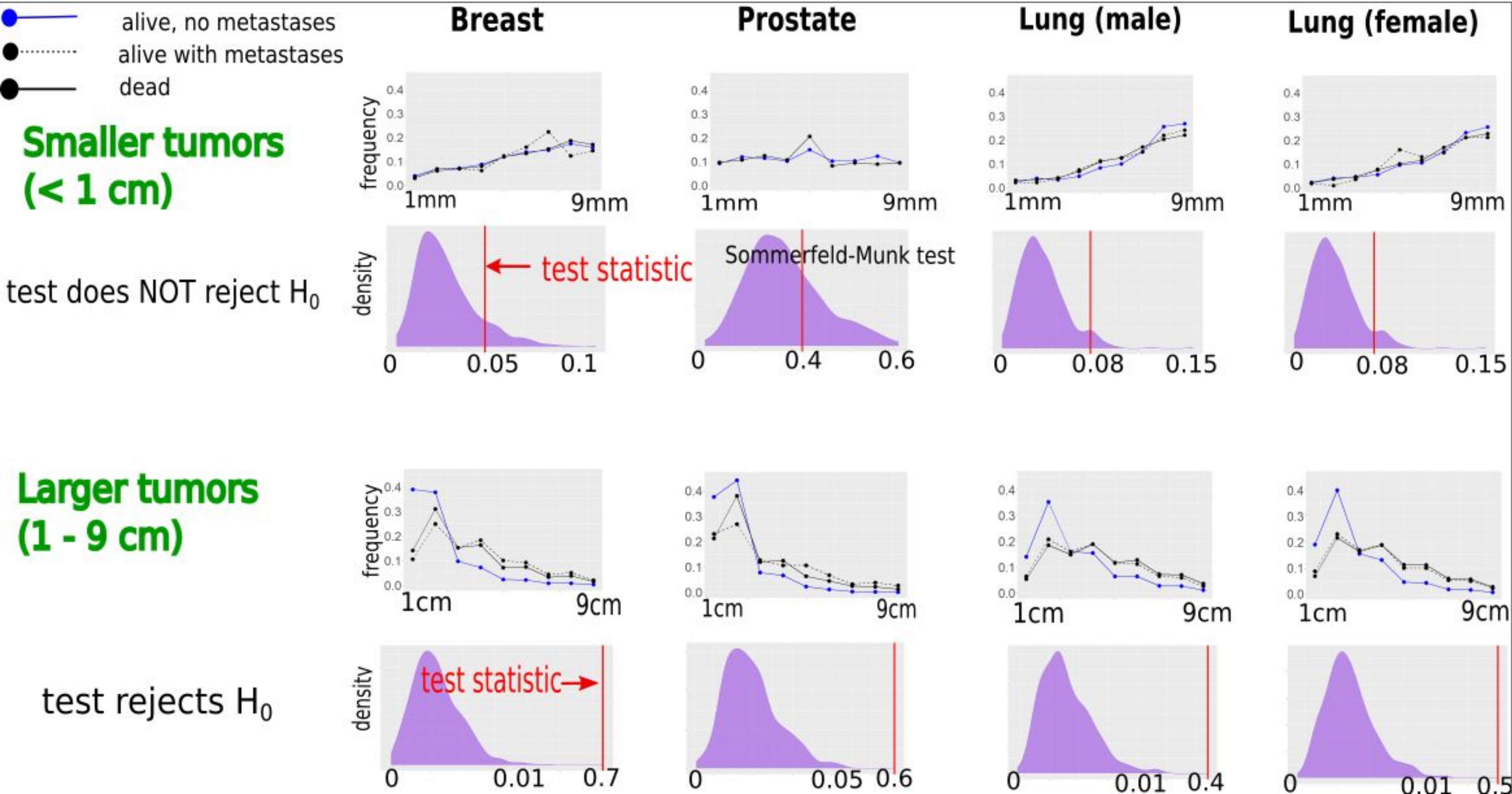
test does NOT reject H_0



test rejects H_0



data: https://seer.cancer.gov



To show $\rho_n \left(MOT(\widehat{\mu}_n) - MOT(\mu) \right) \xrightarrow{\text{in law}} L$

Proof idea:

$$\rho_n \left(\widehat{\mu}_n - \mu \right) \xrightarrow{\text{in law}} G$$

$$\implies \rho_n \left(MOT(\widehat{\mu}_n) - MOT(\mu) \right) \xrightarrow{\text{in law}} \underbrace{MOT'_\mu(G)} =: L$$

Delta Method

Directional derivative of *MOT* at μ in the direction of *G*

$$\rho_n \left(MOT(\widehat{\mu}_n) - MOT(\mu) \right) \xrightarrow{\text{in law}} L$$

$MOT(\mu)$ program (primal)

$$\min_{\pi \geq 0} \langle c, \pi \rangle$$

such that $A\pi = \mu$

How does *L* look?

$MOT(\mu)$ program (dual)

$$\max_{u:=(u_1,\ldots,u_k)} \sum_{i=1}^{k} \langle \mu_i, u_i \rangle$$

such that $A'u \leq c$

L program (~ dual)

$$\max_{u:=(u_1,\cdots,u_k)} \sum_{i=1}^{k} \langle G_i, u_i \rangle$$

such that $A'u \leq c$

and $\sum_{i=1}^{k} \langle \mu_i, u_i \rangle = MOT(\mu)$

Summary

- Proposed to use *MOT* for k-sample problem

- Provided limits

$$\rho_n \left(MOT(\widehat{\mu}_n) - MOT(\mu) \right) \xrightarrow{\text{in law}} L$$

- Used *L* to test

$$H_0 : \mu_1 = \cdots = \mu_k$$

Acknowledgements

NK thanks

- The organizers: **Prof. Gero Friesecke** and **Prof. Augusto Gerolin.**

- **Prof. Adriana Dawes** for advice and support via

  The National Institute of General Medical Science of the National Institutes of Health under award number R01GM132651 to Adriana Dawes.

- **Registration and travel support for this presentation was provided by the United States National Science Foundation.**

Thank you for your attention!

Questions?