Discrete Probabilistic Inverse Optimal Transport

ICML 2022 @ Baltimore

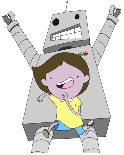
Wei-Ting Chiu[1,*], Pei Wang[1,*], Patrick Shafto[1,2]

iamwtchiu@gmail.com, peiwang425@gmail.com, patrick.shafto@gmail.com

[1] Department of Mathematics and Computer Science, Rutgers University Newark, NJ
[2] School of Mathematics, Institute for Advanced Study (IAS), Princeton NJ
[*] Equal contribution

July, 2022

RUTGERS UNIVERSITY | NEWARK



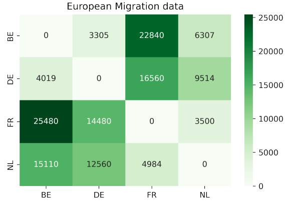
IAS INSTITUTE FOR ADVANCED STUDY

Table of Contents

Theory

Motivation

- Inverse OT is the problem of inferring the latent costs C from observed T.
  - It arises naturally in many domains, such as: cost of international migrations.

$T =$ 

$C = ?$ – Existing works.

Is C unique? – No !!

Space of all possible C? – Our work.

- Entropy regularized OT (EOT): a map $\Phi : (C, \mu, \nu) \to T$.
- Inverse OT is defined as $\Phi^{-1} : T \to \{(C, \mu, \nu)\}$ – not unique!!!.
- Goal: We aim to analyze the intrinsic properties of the entire cost space.

Probability Inverse Optimal Transport (PIOT)

Given a noisy observation T of T^*, PIOT is defined as:

$$P(C|T) = \int_{T^* \in (\mathbb{R}^*)^{m \times n}} P(C|T^*)P(T^*|T)dT^*, \quad (1)$$

where $P(C|T^*)$ is obtained through Bayes' Rule:

$$P(C|T^*) = \frac{P(T^*|C)P_0(C)}{P(T^*)}, \quad (2)$$

- Likelihood $P(T^*|C) = 1$ if $C \in \Phi^{-1}(T^*)$, otherwise $P(T^*|C) = 0$;

Support manifold $\Phi^{-1}(T)$ – No noise

Definition. Positive matrices A, B are cross ratio equivalent, $A \overset{c.r.}{\sim} B$, if:

$$r_{ijkl}(A) := \frac{a_{ik}a_{jl}}{a_{il}a_{jk}} = \frac{b_{ik}b_{jl}}{b_{il}b_{jk}} := r_{ijkl}(B). \quad (3)$$

Theorem. Given positive T of dimension $m \times n$, $\Phi^{-1}(T)$ is a hyperplane of dimension $m + n - 1$, consists cost matrices C satisfying $e^{-C} \overset{c.r.}{\sim} T$.

- Supp$[P(C|T)]$ for different priors, incomplete observations, submanifolds.

Support manifold - With noise

Lemma. Given T_1, T_2, hyperplanes $\Phi^{-1}(T_1)$, $\Phi^{-1}(T_2)$ are the same or parallel.

- The distance between IOT of T_1 and T_2 is then well-defined to be the Euclidean distance between $\Phi^{-1}(T_1)$ and $\Phi^{-1}(T_2)$.

Proposition. Given T, assume uniform noise on t_{11} with bounded size a, then

$$\text{supp}[P(C|T)] = \cup_{T' \in \mathbb{B}_a(T)} \Phi^{-1}(T') \text{– a collection of hyperplanes.}$$

Experiments

Approximate inference via MCMC

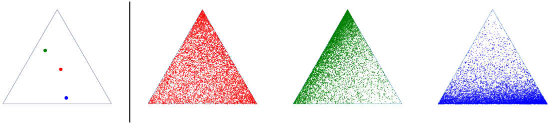
- Use MCMC method for inferring cost function C
- Advantages:
  - Flexibility on choice of prior $P_0(C)$ or $P_0(K)$
  - Generally inferring a set of possible C instead of a single one
- Procedure: given $K^{(0)} = T$; $K^{(i+1)} = \text{diag}D^r \ K^{(i)} \ \text{diag}D^c$
- Acceptance ratio a depending on the prior:
  - Prior on C: Metropolis (MetroMC): $a = \frac{P(-\ln(K^{(i+1)})|T)}{P(-\ln(K^{(i)}|T)} \overset{(*)}{=} \frac{P_0(C^{i+1},\beta)}{P_0(C^i,\beta)}$
  - Prior on K: Metropolis-Hasting (MHMC): $a = \min\left(1, \frac{P(K')}{P(K)} \frac{Q(K|K'))}{Q(K'|K)}\right)$
- Accept move if $a > u \in [0, 1)$.

Visualizing support of $P(K|T)$

PIOT

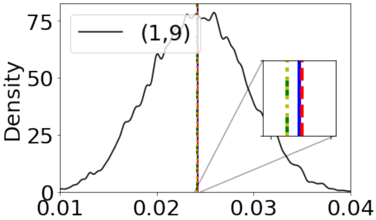
W.-T. Chiu,
P. Wang,
P. Shafto

- *Demonstrate effect of cross-ratio equivalence on* $P(K|T)$
- Prior over K is possible and suitable for visualization
- Column-wise Dirichlet prior for $P_0(K)$
  - Visualizing supp$[P(K|T)]$: plotting columns of T & columns of $P(K|T)$



Synthetic data: symmetric cost with noise

- *Accurate cost inference* with relative error $\leq 10^{-3}$;
- $C^g_{ij} = \left|\frac{(i-j)}{n}\right|^p + 10^{-4} * U_{[0,1)}$ (blue)[1]
- Median of the posterior distribution (red)
- Compare with Ma et al. (arXiv:2002.09650) (yellow & green)



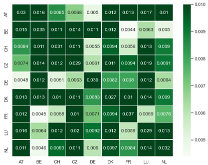
[1] The inset in the figure has a range of $C^g \pm 10^{-4}$.

Real-world data: European migration flow (1)

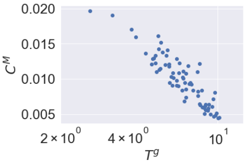
RUTGERS UNIVERSITY | NEWARK
IAS INSTITUTE FOR ADVANCED STUDY

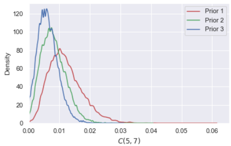
- Ground truth cost not available
- Inferred cost inversely proportional to log of migration flow
- Consider 3 priors, stronger prior gives sharper posterior
- *PIOT with soft constrain is less biased*

Ground truth T^g



Inferred cost C





Real-world data: European migration flow (2)

RUTGERS UNIVERSITY | NEWARK
IAS INSTITUTE FOR ADVANCED STUDY

- Evaluate predictability of PIOT on 2 cases: T with (1) noise (2) missing value
  - Procedure: T –[PIOT]→ $P(C|T)$ –[EOT]→ $P(T')$; $T^p = \overline{P(T')}$
  - NOISE: $T_{ij} = T^g_{ij} + \epsilon_{ij}$; MISSING VALUE: $T_{ij} \in [100, 25000)$
- *Accurate predictions on noisy/missing elements* → *PIOT infers cost well*

	Noise			Missing value		
	(3,6)	(4,5)	(5,7)	(3,6)	(4,5)	(5,7)
T^g	452	6635	16560	452	6635	16560
T^p	454	6593	16515	3699	7462	12349
$\|\Delta\|$	.45%	.87%	.25%	26%	6.6%	34%

Table: Comparison between the predicted coupling, T^p, with the ground truth coupling, T^g. $\Delta = (T^p - T^g)/\overline{T}^g$. Corresponding element is corrupted by a 4% Gaussian noise on the left column, and missing on the right column.

Conclusions

Conclusions

In this work we

- formulate PIOT and identify the support manifold for latent costs
- characterize the geometry of the support manifold
- analyze the support manifold for noisy and incomplete observations
- develop MCMC algorithms for posterior inference
- demonstrate PIOT through simulations on general examples

Thank you!

Acknowledgment

Research was sponsored by the Defense Advanced Research Projects Agency (DARPA) and the Army Research Office (ARO) and was accomplished under Cooperative Agreement Numbers W911NF-20- 2-0001 and HR00112020039 to P.S. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the DARPA or ARO, or the U.S. Government.