Bayesian Inference for Optimal Transport with Stochastic Cost

Zejun Xie
谢泽君

2021. 1. 29

Outlines

Optimal Transport

$$\mathrm{OT}(\mu, \nu, C) = \min_{\Gamma \in \Pi(\mu,\nu)} \mathrm{OT}(C, \Gamma) \triangleq \min_{\Gamma \in \Pi(\mu,\nu)} \sum_{i,j} \Gamma_{ij} C_{ij} \tag{1}$$

- Locations of factories $x_i{}_{i=1}^n \subset \mathbb{R}^d$ and of outlets $y_j{}_{j=1}^m \subset \mathbb{R}^d$.
- Each of the factories produces μ amounts of goods, and the outlets have a demand of ν. $\mu, \nu \geq 0$ and $\sum_{i=1}^n \mu_i = \sum_{j=1}^m \nu_j = 1$.
- $\mu(x) = \sum_{i=1}^n \mu_i \delta_{x_i}(x), \nu(y) = \sum_{j=1}^m \nu_j \delta_{y_j}(y)$,
  where $\delta_x(y)$ stands for the Dirac delta function.
- The cost of transporting a unit amount of goods from x_i to y_j is $c(x_i, y_j)$, where $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^+$ is the cost function.
- $\Pi(\mu, \nu) \triangleq \left\{\Gamma : \sum_{j=1}^m \Gamma_{ij} = \mu_i, \sum_{i=1}^n \Gamma_{ij} = \nu_j\right\}$, AKA transport polytope. Its elements are transport plans, as Π_{ij} is the amount of mass transported from x_i to y_j.

Optimal Transport

$$\mathrm{OT}(\mu, \nu, C) = \min_{\Gamma \in \Pi(\mu,\nu)} \mathrm{OT}(C, \Gamma) \triangleq \min_{\Gamma \in \Pi(\mu,\nu)} \sum_{i,j} \Gamma_{ij} C_{ij}$$

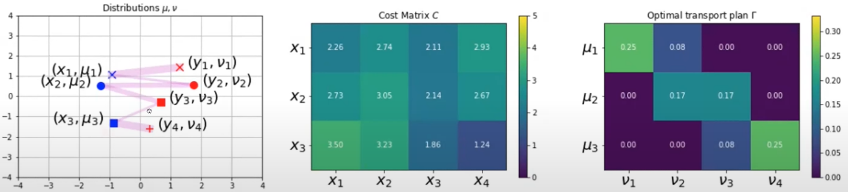


Figure 1: Left: Spatial distribution of μ and ν, $\sum_i \mu_i = \sum_j \nu_j = 1$. **Middle:** Cost matrix $C_{ij} = c(x_i, y_j) = \|x_i - y_j\|^2$. **Right:** Optimal transport plan $\Gamma^* = \arg\min_{\Gamma \in C(\mu,\nu)} \sum_{i,j} \Gamma_{ij} C_{ij}$.

*Typo: The horizontal coordinate should be y_i in Fig.1 Middle.

OT with Stochastic Cost

$$\Gamma \sim \underset{\Gamma \in \Pi(\mu,\nu)}{\arg\min} \mathrm{OT}(C, \Gamma), \quad C \sim \Pr(C) \tag{2}$$

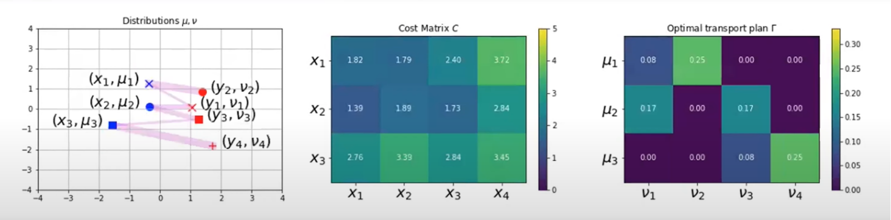
- Observe samples $C^k \sim \Pr(C), k = 1, \ldots, N$, from a stochastic cost C.
- In the rest of this work, the goal is to infer the distribution Γ inherits from C.

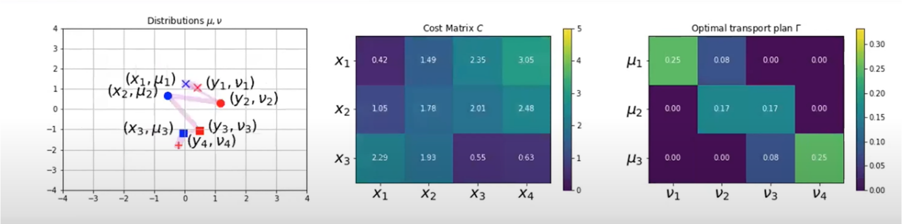
Application:

- Real life transportation has stochastic cost induced by the varying traffic conditions.
- Transporting goods between uncertain locations.
- Allow for uncertainty quantification in OT models.

OT with Stochastic Cost

$$\Gamma \sim \underset{\Gamma\in\Pi(\mu,\nu)}{\arg\min} \mathrm{OT}(C,\Gamma), \quad C \sim \Pr(C)$$





*Typo: The horizontal coordinate should be y_i in Middle.

Bayesian Approach

- Choose a prior $\Pr(\Gamma)$, sample $C^k \sim \Pr(C), k = 1, \ldots, N$, and define likelihood of auxiliary optimality variables O_k.
-

$$\Pr\left(O_k = 1 \mid \mu, \nu, C^k, \Gamma\right) = \exp\left(-\mathrm{OT}\left(C^k, \Gamma\right)\right), \tag{3}$$

  where $O_k = 1$ if $\Gamma \in C(\mu, \nu)$ minimizes $\mathrm{OT}\left(C^k, \Gamma\right)$, and 0 else.
- The smaller $\mathrm{OT}\left(C^k, \Gamma\right)$ is, the more likely $O_k = 1$, i.e., the more likely Γ is the OT plan for C^k. As $C^k \geq 0, \mathrm{OT}\left(C^k, \mu, \nu\right) \geq 0$, and therefore $\mathrm{OT}\left(C^k, \Gamma\right) = 0$ implies $O_k = 1$.

Bayesian Approach: Condition 1

- Condition Γ to be optimal for **all** C_k, yieldig the posterior log-likelihood

$$Q_\forall(\Gamma) = -\log \mathrm{Pr}^{\forall}_{\mu,\nu}(\Gamma \mid C, O = 1) \tag{4}$$

$$= -\log \left(\mathrm{Pr}(\Gamma \mid \mu, \nu) \prod_{k=1}^{N} \mathrm{Pr}\left(O_k = 1 \mid \mu, \nu, C^k, \Gamma \right) \right) \tag{5}$$

$$= -\log \mathrm{Pr}(\Gamma) + \mathrm{OT}\left(\sum_{k=1}^{N} C^k, \Gamma \right) + \text{const.} \tag{6}$$

Maximum a posteriori(MAP) in this case is a solution to regularized OT with regularizer $-log\,\mathrm{Pr}(\Gamma)$.

Bayesian Approach: Condition 2

- Condition Γ to be optimal for **some** C_k, yieldig the posterior log-likelihood

$$\begin{aligned} Q_{\exists}(\Gamma) &= -\log \Pr^{\exists}_{\mu,\nu}(\Gamma \mid C, O = 1) && (7) \\ &= -\log\left(\Pr(\Gamma \mid \mu, \nu) \sum_{k=1}^{N} \Pr\left(O_k = 1 \mid \mu, \nu, C^k, \Gamma\right)\right) && (8) \\ &= -\log \Pr(\Gamma) - \log \sum_{k=1}^{N} \exp\left(-\mathrm{OT}\left(C^k, \Gamma\right)\right) + \text{ const.} && (9) \end{aligned}$$

Observe the log-sum-exp term, which can be viewed as a smooth minimum.

C1: MAP Estimation as Regularized OT

$$\Gamma_{\forall}^{*} = \arg\min_{\Gamma\in\Pi(\mu,\nu)} Q_{\forall}(\Gamma) = \arg\min_{\Gamma\in\Pi(\mu,\nu)} \left\{ -\log \Pr(\Gamma) + \mathrm{OT}\left(\sum_{k=1}^{N} C^{k}, \Gamma \right) \right\}, \tag{10}$$

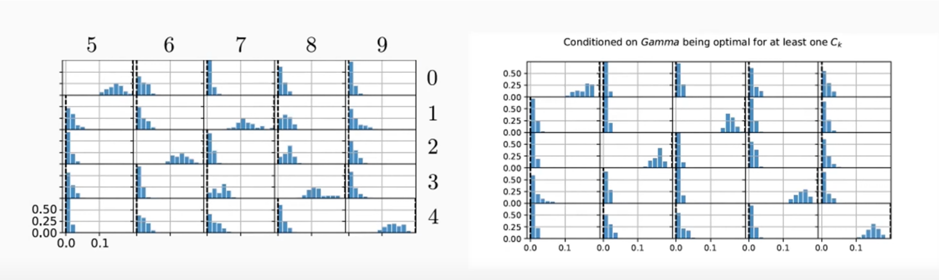
$$R(T) = -log \Pr(\Gamma). \tag{11}$$

- Constant Prior: Equivalent to origin OT.
- Entropy Prior: $R(\Gamma) = -\log \Pr(\Gamma) = -\epsilon H(\Gamma)$, $H(\Gamma) = -\sum_{ij} \Gamma_{ij} \log \Gamma_{ij}$. Thus, solving (10) corresponds precisely to solving the entropy-relaxed OT problem (Cuturi, 2013).
- Gaussian Prior: $\mathrm{vec}(\Gamma) \sim Pr(\bar{\Gamma}, \Sigma)$, $R(\Gamma) = \frac{1}{2}(\mathrm{vec}(\Gamma) - \bar{\Gamma})\Sigma^{-1}(\mathrm{vec}(\Gamma) - \bar{\Gamma})$. If $\bar{\Gamma} = 0$, the Gaussian prior results in quadratically regularized OT (Lorenz et al., 2019; Dessein et al., 2018)

Posterior Sampling

- MCMC
- HMC: a celebrated variant of MCMC, allowing for efficient sampling in high dimensions.

MNIST



- Demonstration of BayesOT between instances of digits 5-9 (columns) and 1-4 (rows). Each histogram shows the posterior of Γ_{ij}. **Left:** Γ conditioned to be optimal for all C^k. **Right:** Γ conditioned to be optimal for some C^k.
- Prior: Uniform Distribution. Cost: The squared Euclidean metric.
- Different results with different sampling, but same trend.

Discussion

- The increase in dimensions makes MCMC runtime too slow and HMC becomes an alternative to it.
- HMC is currently only works for smaller problem. In order to expand to a large data it is necessary to use better sampling methods such as Riemann Hamiltonian Monte Carlo (Ma et al., 2015).
- Possible future directions for BayesOT could include modelling the joint distribution (C, Γ) of the cost and the OT plan explicitly, which allows computing a posterior distribution for the total OT cost.