Entropic Curves: Schrodinger Bridge

Logan Timmins

October 31 2019

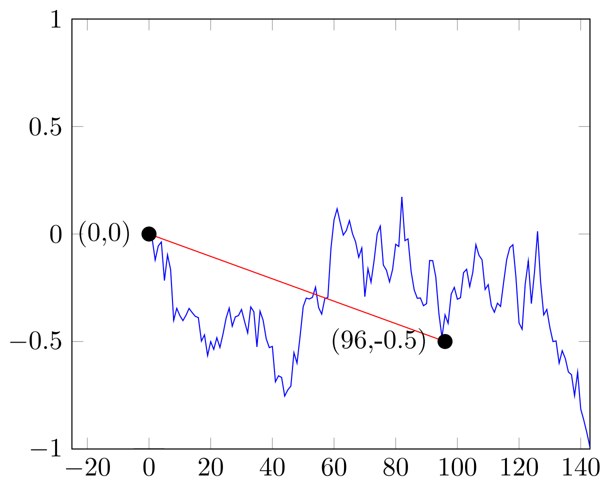


Figure 1. Distance of paths connecting points (0,0) to (96,-0.5). A **Brownian Bridge** (blue curve) and the shortest path (red).

A **Schrodinger Bridge** is a distribution of paths connecting x_0 to x_1 regarding more than 1 particle.

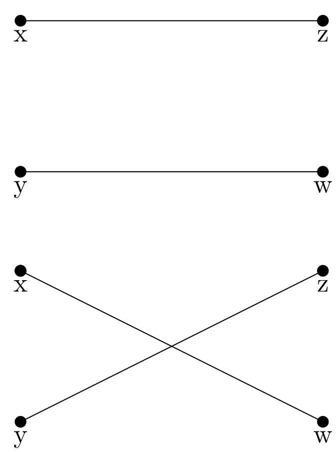


Figure 2. (Top) More likely paths for particles at points x and y to end at z and w, respectively. (Bottom) Potential, less likely paths for particles at points at x and y to end at w and z, respectively.

Consider a collection of indistinguishable particles undergoing Brownian motion with parameter ε. We observe:

$$x_0(t=0), x_1(t=0)...x_n(t=0)$$
$$x_0(t=1), x_1(t=1)...x_n(t=1)$$

When n=1, there's only one starting point and one end point. This distribution on paths is called a **Brownian Bridge**. When n is greater than 1, the distribution on paths is called a **Schrodinger Bridge**.

Two components of a Schrodinger Bridge

1. **Random Permutation Matrix (P)** describing which particle goes where.
2. Independent Brownian Bridges connecting each pair of particles.

Fact: Solving entropy-regularized optimal transport (OT) will give coupling π_ε^* connecting $\hat{\mu_0}$ to $\hat{\mu_1}$

$$\mathbb{E}P = \pi_\varepsilon^*$$

Positions:

$$x_0(t=0), x_1(t=0)...x_n(t=0) \Rightarrow \hat{\mu_0} = \sum_{i=1}^{n} \frac{1}{n}\delta_x i(0)$$

$$x_0(t=1), x_1(t=1)...x_n(t=1) \Rightarrow \hat{\mu_1} = \sum_{i=1}^{n} \frac{1}{n}\delta_x i(1)$$

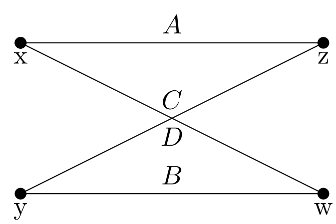


Figure 3. Potential paths from points x and y to z and w.

Let the path from x to z be called A, from y to w be B, x to w be C and y to z be D. Repeat this experiment n times and record the number of times each transition (A, B, C, D) is made. According to Brownian Motion:

$$\alpha = Prob(A) \propto e^{\frac{-||x-z||^2}{\varepsilon}}$$
$$\beta = Prob(B)$$
$$\gamma = Prob(C)$$
$$\delta = Prob(D)$$
$$Prob[Num(A) \approx pn, Num(B) \approx qn, Num(C) \approx rn, Num(D) \approx sn]$$
$$\approx \frac{n!}{(pn)!(qn)!(rn)!(sn)!}\alpha^{pn}\beta^{qn}\gamma^{rn}\delta^{sn}$$

By **Sterling's Formula:** $k! \approx exp[klogk - k]$ Therefore,

$$= exp[nlogn - nplognp - nqlognq - nrlognr - nslogns + nplog\alpha + nqlog\beta + nrlog\gamma + nslog\delta]$$

$$= exp[-n[plog\frac{p}{\alpha} + qlog\frac{q}{\beta} + rlog\frac{r}{\gamma} + slog\frac{s}{\delta}]$$

$$= exp[-nKL(\begin{bmatrix} p \\ q \\ r \\ s \end{bmatrix} \mid \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix})]$$

This calculation tells us that when n is large, we will only see proportions p, q, r and s that are as close to $\alpha, \beta, \gamma,$andδ as possible.

Recall: $KL(\mathbb{P}|\mathbb{Q}) \geq 0$ or $= 0$ iff $\mathbb{P} = \mathbb{Q}$

Note: $\begin{bmatrix} p & r \\ q & s \end{bmatrix}$ Should be a coupling of $\hat{\mu_0} \to \hat{\mu_1}$ and $\begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$ shouldn't be a coupling of $\hat{\mu_0} \to \hat{\mu_1}$ because they don't have the right marginals. Now, we'll reformulate entropic optical transport:

$$\min_{\pi} \mathbb{E}_{\pi} c(x,y) - \varepsilon H(\pi)$$

$$(x,y) \sim \pi, x \sim \mathbb{P}, y \sim \mathbb{Q}$$

For **finitely** supported distributions:

$$\mathbb{P} = \sum_{i=1}^{n} \frac{\delta_x i}{n}$$

$$\mathbb{Q} = \sum_{i=1}^{n} \frac{\delta_y i}{n}$$

$$H[\pi] \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_{ij} log \pi_{ij}$$

For **continuous** distributions:

$$\min_{\pi \in U[\mathbb{P},\mathbb{Q}]} \int c[x,y] d\pi[x,y] + \varepsilon \int d\pi log \frac{\pi[x,y]}{\mathbb{P}[x]\mathbb{Q}[y]} \tag{1}$$

Notes:

- U$[\mathbb{P},\mathbb{Q}]$ is the set of couplings of $\mathbb{P}$ and $\mathbb{Q}$
- The log element in the above equation is equal to

$$KL[\pi|\mathbb{P}\mathbb{Q}]$$

Entropic Optimal Transport:

$$\min_{\pi \in U[\mathbb{P},\mathbb{Q}]} \int c[x,y] d\pi[x,y] + \varepsilon \int d\pi log \frac{\pi[x,y]}{\mathbb{P}[x]\mathbb{Q}[y]} \tag{2}$$

$$\text{where} \quad log \frac{\pi[x,y]}{\mathbb{P}[x]\mathbb{Q}[y]} = KL[\pi|\mathbb{P}\mathbb{Q}]$$

$$\pi \sim [x,y]$$

- In information theory, this is called **mutual information**

Claim: the above equation is equivalent to

$$\min_{\pi \in U[\mathbb{P},\mathbb{Q}]} KL[\pi|R] \tag{3}$$

$$\text{where} \quad dR[x,y] = e^{\frac{-c[x,y]}{\varepsilon}} d\mathbb{P}[x] d\mathbb{Q}[y]$$

- **Note:** if we have a state space, X then R is a measure of XX

Proof of equivalence:

$$\int c[x,y] d\pi[x,y] + \varepsilon \int d\pi log \frac{\pi[x,y]}{\mathbb{P}[x]\mathbb{Q}[y]}$$

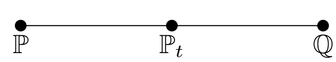
$$= \varepsilon \int [log[e^{\frac{c}{\varepsilon}}] + log \frac{\pi[x,y]}{\mathbb{P}[x]\mathbb{Q}[y]}] d\pi$$

$$= \varepsilon \frac{\pi}{\mathbb{P}\mathbb{Q}e^{\frac{-c}{\varepsilon}}} d\pi$$

$$= \varepsilon KL[\pi|R]$$

Dynamical Formulation of Optimal Transport

Without entropy: we saw that OT coupling gives a shortest path in W_z connecting $\mathbb{P}$ to $\mathbb{Q}$. We can interpolate along the constant speed geodesic $\mathbb{P}_t$ for t $\in$ (0,1).



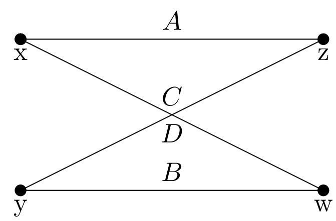
To get $\mathbb{P}_t$, take $\pi*$ and apply the map

$$[x, y] \mapsto [1-t]x + ty$$

$$[x, y] \sim \pi$$

$$\text{law}[[1-t]x + ty] = \mathbb{P}_t$$

With entropy:



1. ε low
   - 99% A and 1% C
2. ε medium
   - 90% A and 10% C
3. ε high
   - 50% A and 50% C