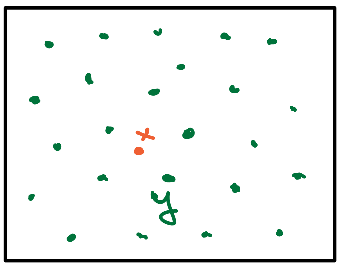
Recap

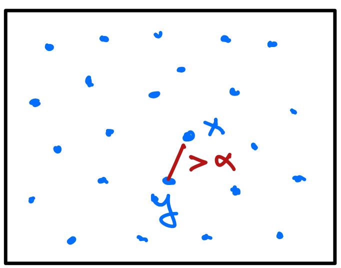
Sanov: $$-\frac{1}{n}\log P_{\bar{X}\sim Q^n}\left[P_{\bar{X}}\in\Pi\right] \longrightarrow \inf_{P\in\Pi} D(P\,\|\,Q)$$

Minimax: $$M_n(\Pi,\delta) \geq \Phi(\delta)\cdot\left(1-\frac{n\,\mathbb{E}_{V_1,V_2} D(P_{V_1}\,\|\,P_{V_2})}{\log|V|}\right)$$



- $c\cdot\frac{d}{n}$ (mean estimation)

- $c\cdot\frac{\log d}{n}$ (sparse mean)



I - Projections

$\mathcal{X}$ finite. Π : Set of distributions on $\mathcal{X}$

$$P^* = \underset{P \in \Pi}{\operatorname{argmin}} \; D(P \| Q)$$

Can assume $\mathrm{Supp}(Q) = \mathcal{X}$

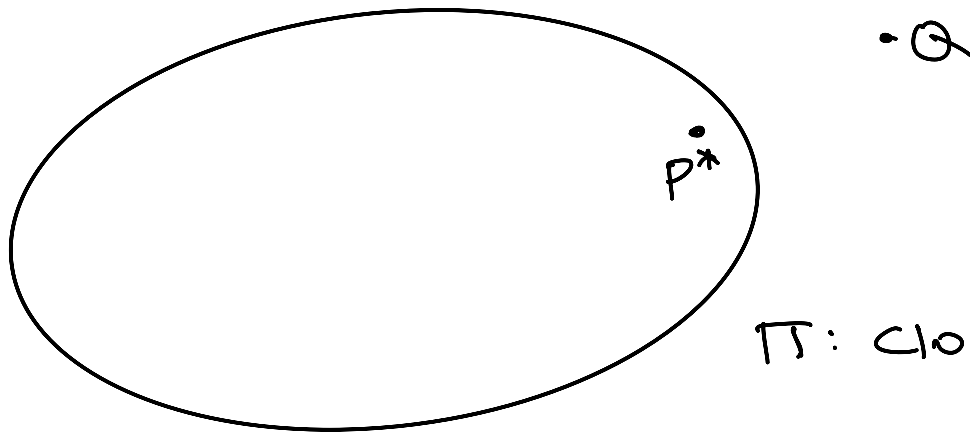
When $Q \equiv$ uniform, $D(P\|Q) = \sum_x P(x) \log \underbrace{\frac{1}{Q(x)}}_{|\mathcal{X}|} - H(P)$

$$= \log |\mathcal{X}| - H(P)$$

I-Projections for nice sets

$$P^* = \underset{P \in \Pi}{\operatorname{argmin}}\ D(P \| Q)$$

$$= \operatorname{Proj}_{\Pi}(Q)$$



Π: closed, convex

- $P^* = \underset{P \in \Pi}{\operatorname{argmin}}\ D(P \| Q)$ exists and is unique

exists: closed Π

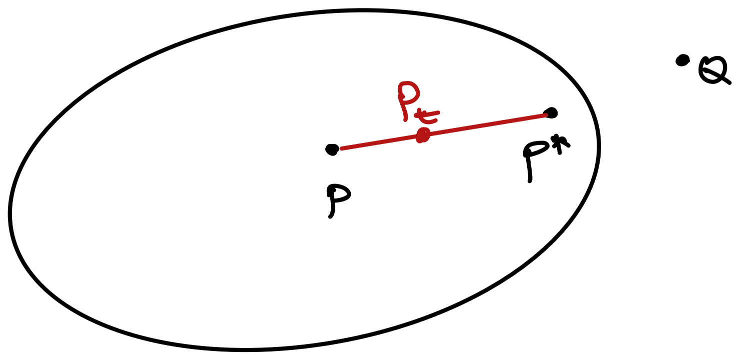
unique: Strict convexity of $D(P \| Q)$

<u>Proof:</u>

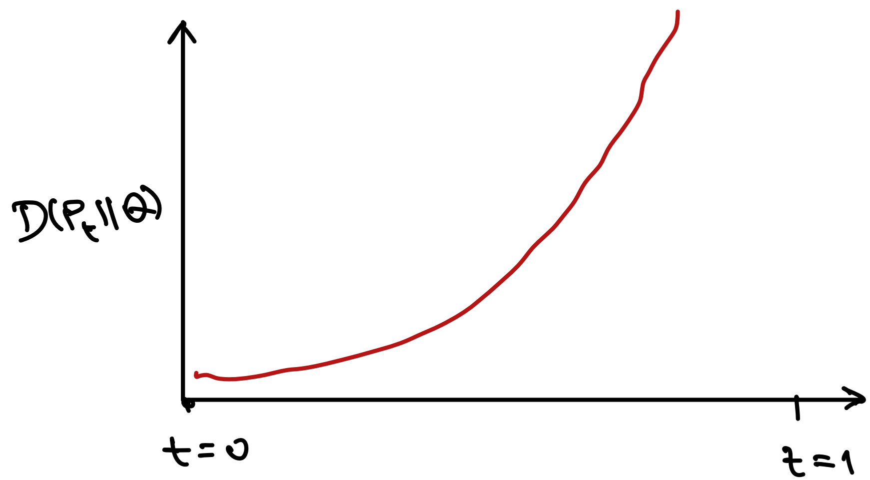
Better. But how much better?

- Let $P^* = \mathrm{Proj}_{\Pi}(Q)$. Then for all $P \in \Pi$
  - $\mathrm{Supp}(P) \subseteq \mathrm{Supp}(P^*)$
  - $D(P \| Q) \geq D(P \| P^*) + D(P^* \| Q)$

<u>Proof:</u>



$P_t = t \cdot P + (1-t) \cdot P^*$



<u>Ex:</u> $\frac{d}{dt} D(P_t \| Q) = \sum_{x \in \mathcal{X}} (p(x) - p^*(x)) \cdot \log \frac{P_t(x)}{q(x)}$

$$\frac{d}{dt} D(P_t \| Q) = \sum_{x \in \mathcal{X}} \underbrace{(p(x) - p^*(x))}_{>0} \cdot \log \underbrace{\frac{P_t(x)}{q(x)}}_{\to -\infty}$$

Say $p^*(x) = 0$, $P(x) > 0$ for some x

$$\frac{d}{dt} D(P_t \| Q) = \sum_{x \in \mathcal{X}} (p(x) - p^*(x)) \cdot \log \frac{P_t(x)}{q(x)} \Big|_{t=0} \geq 0$$

$$= \sum_x (p(x) - p^*(x)) \cdot \log \frac{p^*(x)}{q(x)}$$

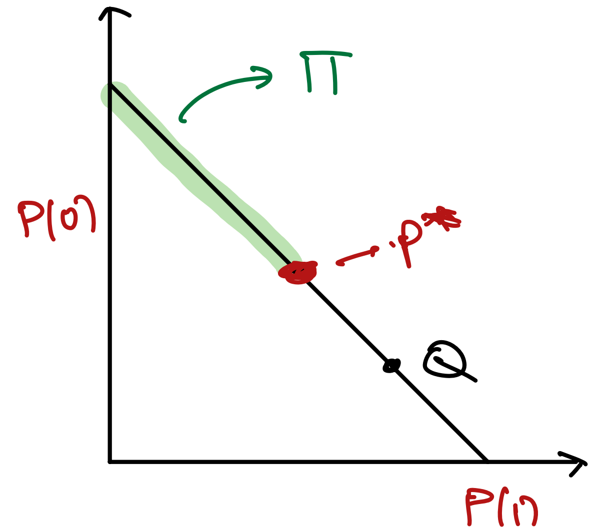
$$= \sum_x p(x) \log \frac{p^*(x)}{q(x)} \frac{p(x)}{p(x)} - D(P^* \| Q)$$

$$= \sum_x p(x) \log \frac{p(x)}{q(x)} - \sum_x p(x) \log \frac{p(x)}{p^*(x)} - D(P^* \| Q)$$

$$= D(P \| Q) - D(P \| P^*) - D(P^* \| Q)$$

Eg. $\mathcal{X} = \{0,1\}$ $\qquad \Pi = \{ P \mid P(1) \le 1/2 \}$

$$Q = \begin{cases} 1 & \text{wp } 3/4 \\ 0 & \text{wp } 1/4 \end{cases}$$



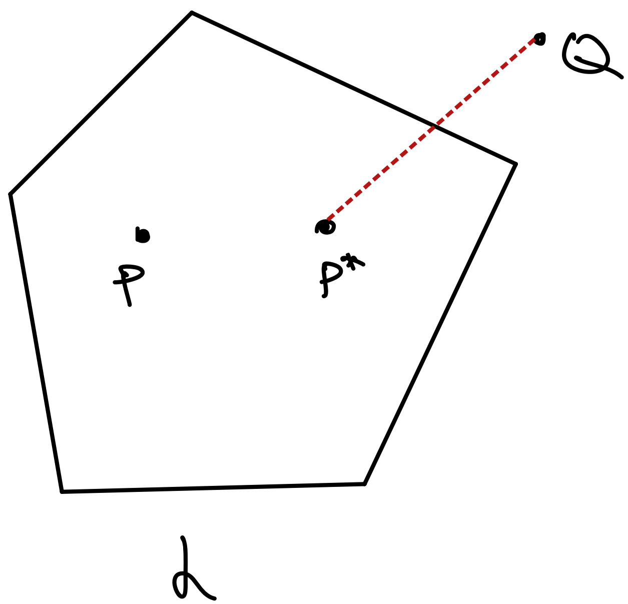
$D(P \| Q) =$

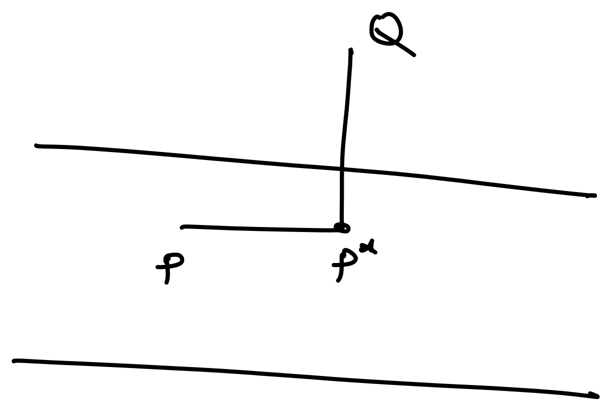
$P^* =$

I-Projections on linear families

$\left.\begin{matrix} f_1 \\ \vdots \\ f_k \end{matrix}\right\} \mathcal{X} \to \mathbb{R}$ $\qquad$ $\left.\begin{matrix} \alpha_1 \\ \vdots \\ \alpha_k \end{matrix}\right\} \in \mathbb{R}$

$$\mathcal{L} = \left\{ P \;\middle|\; \mathbb{E}_{X \sim P} f_i(X) = \alpha_i \;\; \forall i \in [k] \right\}$$





$$D(P \| Q) = D(P \| P^*) + D(P^* \| Q)$$

Pythagorean identity

- $\mathcal{L} = \{ P \mid \sum P(x) \cdot f_i(x) = \alpha_i \quad \forall i \in [k] \}$ $\quad \bigcup_{P \in \mathcal{L}} \text{Supp}(P) = \mathcal{X}$

Given: Some $P \in \mathcal{L}$, Q, $P^* = \text{Proj}_{\pi}(Q)$

- $\exists \beta > 0$ s.t. for $t \in [-\beta, 0]$ $\quad P_t = t \cdot P + (1-t) \cdot P^* \in \mathcal{L}$

- $D(P \| Q) = D(P \| P^*) + D(P^* \| Q)$

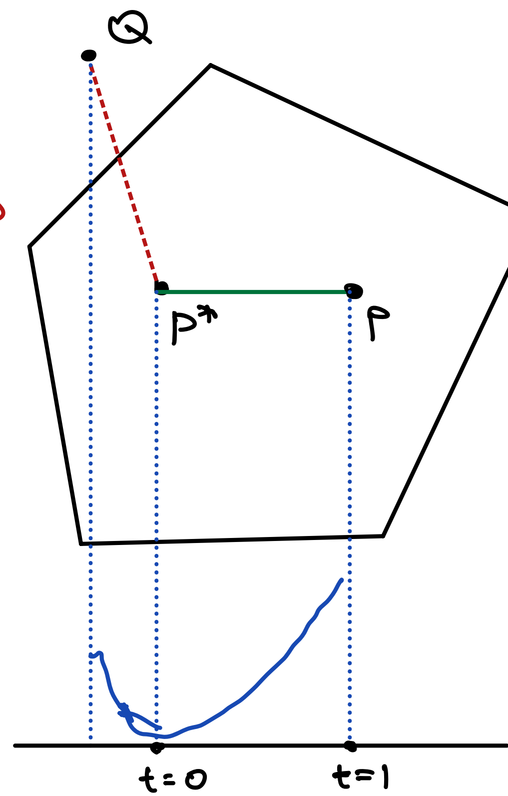
<u>Proof:</u>

$$\mathbb{E}_{x \sim P_t} f_i(x) = t \cdot \underbrace{\mathbb{E}_{x \sim P} f_i(x)}_{\alpha_i} + (1-t) \cdot \underbrace{\mathbb{E}_{x \sim P^*} f_i(x)}_{\alpha_i}$$

$$= \alpha_i \qquad \forall t$$

$$P_t(x) = \underbrace{P^*(x)}_{>0} + t\,(P(x) - P^*(x))$$

$$\geq 0 \quad \text{for} \quad t \geq - \min_x \frac{P^*(x)}{|P(x) - P^*(x)|}$$



Projections to linear families

$P^* = \mathrm{Proj}_{\mathcal{L}}(Q)$: $\exists\ \lambda_1 \dots \lambda_k$ s.t.

$$P^*(x) = C \cdot \underbrace{Q(x) \cdot e^{\lambda_1 \cdot f_1(x) + \dots + \lambda_k \cdot f_k(x)}}_{\in\ \mathcal{E}_Q(f_1, \dots, f_k)}$$

Computing for $k=1$

$$\min \quad D(P\|Q)$$

$$\text{s.t.} \quad \sum P(x)\cdot f(x) = \alpha \quad \Big\} \lambda_1$$

$$\sum P(x) = 1 \quad \Big\} \lambda_0$$

$$P(x) \geq 0 \quad \forall x$$

$$\Lambda(P,\lambda_0,\lambda_1) = D(P\|Q) + \lambda_0\Big(\underbrace{\sum P(x) - 1}\Big) + \lambda_1\Big(\sum P(x) f(x) - \alpha\Big)$$

$$\inf_{P\geq 0} \ \sup_{\lambda_0,\lambda_1\in\mathbb{R}} \Lambda(P,\lambda_0,\lambda_1) \geq \sup_{\lambda_0,\lambda_1\in\mathbb{R}} \ \inf_{P\geq 0} \Lambda(P,\lambda_0,\lambda_1)$$

= under mild conditions

$\exists P$ s.t. $P(x) > 0 \ \forall x$

[Boyd - Vandenberghe]
Convex Optimization

The dual problem

$$\sup_{\lambda_0, \lambda_1 \in \mathbb{R}} \ \inf_{P \geq 0} \ \Lambda(P, \lambda_0, \lambda_1)$$

$$\sup_{\lambda_0, \lambda_1} \ \inf_{P \geq 0} \ D(P \| Q) + \lambda_0 \left(\sum P(x) - 1 \right)$$

$$+ \lambda_1 \left(\sum P(x) \cdot f(x) - 1 \right)$$