Entropy and concentration

$$\log E_P\left[\exp\left(\lambda\left(Z - E_P Z\right)\right)\right] = \sup_{Q \ll P}\left[\lambda\left(E_Q Z - E_P Z\right) - D(Q\|P)\right]$$

where if P, Q have densities p, q, then

$$D(Q\|P) = E_P\left[\Phi\left(\frac{q(X)}{p(X)}\right)\right] \text{ where } \Phi(x) = x\log x$$
$$= \mathrm{Ent}_P(Y) \text{ where } Y \text{ is s.t. } E_Q[Z] = E_P[YZ] \;\forall Z.$$

<u>Proof</u> :- When $Q \ll P$, $\exists Y$, $E_P[Y] = 1$ s.t. $\forall Z$, $E_Q[Z] = E_P[YZ]$.

Since $E_P[Y] = 1$, we have,

The assumption that P & Q have densitities is not required but we use it for simplicity.

$$\mathrm{Ent}_P[Y] \geqslant E_P(UY) \quad \text{if} \quad E[e^U] \leqslant 1.$$

Take $U = \lambda(Z - E_P[Z]) - \psi_{Z - E(Z)}(\lambda)$, so by defn. $E_P[e^U] = 1$.

So, we have

$$D(Q\|P) \geqslant \lambda(E_Q Z - E_P Z) - \psi_{Z - E[Z]}(\lambda)$$

or $\psi_{Z - E[Z]}(\lambda) \geqslant \lambda\left[E_Q Z - E_P Z\right] - D(Q\|P)$, when $Q \ll P$

Now consider

$$U = \lambda\left(Z - E_P(Z)\right) - \sup_{Q \ll P}\left(\lambda\left(E_Q Z - E_P Z\right) - D(Q\|P)\right)$$

Then for all Y with $E_P[Y] = 1$, we can define $Q \ll P$ by $E_Q[Z] = E_P[ZY]$ ($\forall$ r.v. Z). Then

$$\mathrm{Ent}_P(Y) = D(Q\|P), \text{ and}$$

$$E[UY] \leqslant D(Q\|P) \Rightarrow E[e^U] \leqslant 1$$

$$\Rightarrow \psi_{Z - E[Z]}(\lambda) \leqslant \sup_{Q \ll P}\left(\lambda(E_Q Z - E_P Z) - D(Q\|P)\right) \quad \blacksquare$$

Corollary: $\psi_{Z-E(Z)}(\lambda) \leq \frac{\nu\lambda^2}{2}\ \forall \lambda>0$ is equivalent to the statement that $\forall\ Q \ll P$,

$$E_Q Z - E_P Z \leq \sqrt{2\nu D(Q\|P)}.$$

Pf: From above we see that

$$\psi_{Z-E(Z)}(\lambda) \leq \frac{\nu\lambda^2}{2} \quad \forall \lambda>0$$

$$\iff \forall \lambda>0,\ Q\ll P \quad \lambda(E_Q Z - E_P(Z)) - D(Q\|P) \leq \frac{\nu\lambda^2}{2}$$

$$\iff \forall \lambda>0,\ Q\ll P \quad E_Q Z - E_P Z \leq \frac{\nu\lambda}{2} + \frac{D(Q\|P)}{\lambda}$$

$$\iff \forall Q\ll P \quad E_Q Z - E_P Z \leq \inf_{\lambda>0} \frac{\nu\lambda}{2} + \frac{D(Q\|P)}{\lambda}$$

$$\iff \forall Q\ll P \quad E_Q Z - E_P Z \leq \sqrt{2\nu D(Q\|P)}$$

Consider now the following situation. Let $Z = f(X_1, X_2, \ldots X_n)$, where X_i are independent r.v.'s taking values in Ω_i. Our goal is to prove concentration for Z. We denote by P_i the distribution of X_i, so that $P = \bigotimes_{i=1}^n P_i$ is the joint product distribution.

Suppose f satisfies a co-ordinatewise local Lipschitz condition.

$$f(\vec{y}) - f(\vec{x}) \leq \sum_i c_i(x)\, d_i(x_i, y_i)$$ where d_i is some distance fn. on Ω_i.

We would therefore need to investigate [if we want to use the above corollary], for arbitrary $Q \ll P$

$$E_Q f - E_P f.$$

Now, let σ be a coupling between Q and P, meaning that if the pair (X, Y) is sampled according to σ then

$X \sim P$ and $Y \sim Q$.

Then

$$E_Q f - E_P f = \mathop{E}_{(X,Y)\sim G}[f(Y) - f(X)]$$

$$\leq \sum_{i=1}^{n} \mathop{E}_{(X,Y)\sim G}[c_i(X)\, d_i(x_i, y_i)]$$

$$\leq \sum_{i=1}^{n} E_P[c_i(X)^2]^{1/2} \mathop{E}_{(X,Y)\sim G}[d_i^2(x_i,y_i)]^{1/2}$$

$$\leq \left(\sum_{i=1}^{n} E_P[c_i(X)^2]\right)^{1/2} \left(\sum_{i=1}^{n} \mathop{E}_{(X,Y)\sim G}(d_i^2(x_i,y_i))\right)^{1/2}$$

So we need to bound

$$E_P\left[\sum_{i=1}^{n} c_i^2(X)\right] \leq \nu \text{ and } \min_{\substack{G:\text{coupling} \\ \text{of } P,Q}} E_G\left[\sum_{i=1}^{n} d_i^2(x_i, y_i)\right] \leq C\,D(Q\|P)$$

to get concentration.

Similarly, if

$$f(y) - f(x) \leq \sum_{i=1}^{n} c_i \, \mathbb{1}(x_i \neq y_i)$$

Then

$$E_Q f - E_Q f = \mathop{E}_{(X,Y)\sim G}[f(Y) - f(X)]$$

$$\leq \sum_{i=1}^{n} c_i \mathop{E}_{(X,Y)\sim G}[\mathbb{1}(x_i \neq y_i)]$$

$$\leq \left(\sum_{i=1}^{n} c_i^2\right)^{1/2} \left(\sum_{i=1}^{n} \mathop{P^2}_{(X,Y)\sim G}(X_i \neq Y_i)\right)^{1/2}.$$

So, if $\sum_{i=1}^{n} c_i^2 \leq \nu$ and $\min_{\substack{(X,Y)\sim G \\ G \text{ coupling} \\ \text{of } P,Q}} \sum_{i=1}^{n} P^2(X_i = Y_i) \leq C\,D(Q\|P)$ $\forall Q \ll P$,

then $f(X) \in \mathcal{G}\left(\frac{\nu C}{2}\right)$.

We are therefore led to the study of transportation inequalities: Given a fn $d_i : \Omega_i \times \Omega_i \to [0,\infty)$, and $P = P_1 \otimes P_2 \otimes \cdots \otimes P_n$ on $\Omega_1 \times \Omega_2 \times \cdots \Omega_n$

we want to prove inequalities of the form

$$\inf_{\substack{\mathcal{G} : \text{coupling} \\ \text{of } P, Q}} \sum_{i=1}^{n} \phi\left(E\left[d_i(x_i, y_i)\right]\right) \leq C\, D(Q\|P) \qquad \forall\, Q \ll P. \qquad — \circledast$$

Examples :- (1) Above we saw $\phi = x \mapsto x^2$, $d_i(x,y) = \mathbb{1}[x_i \neq y_i]$.

(2) Gaussian concentration:

Here we have $P = N(0, I_n)$ and $f : \mathbb{R}^n \to \mathbb{R}$ is L-Lipschitz.

$$E_Q f - E_P f = \underset{\substack{(X,Y) \sim \mathcal{G}, \\ \text{coupling of } P,Q}}{E} \left[f(Y) - f(X)\right]$$

$$\leq \underset{\substack{(X,Y) \sim \mathcal{G} \\ \text{Coupling of } P,Q}}{E} \left[L\left(\sum_{i=1}^{n}(X_i - Y_i)^2\right)^{1/2}\right]$$

$$\leq L\left(\sum_{i=1}^{n} \underset{\mathcal{G}}{E}\left[(X_i - Y_i)^2\right]\right)^{1/2}$$

So, we need to prove

$$\sum_{i=1}^{n} \underset{\mathcal{G}}{E}\left[(X_i - Y_i)^2\right] \leq 2\, D(Q\|P) \quad \forall\, Q \ll P$$

in order to get the Gaussian concentration theorem.

Subadditivity :- A simple but important observation is that when P is a product distribution one only needs to establish the inequality in $\circledast$ in one dimension (i.e., $n=1$).

Thm :- Assume ϕ convex, $d_i : \Omega_i \times \Omega_i \to [0,\infty)$, and $C > 0$. Suppose that $\forall\, Q_i \ll P_i$

$$\inf_{\substack{\mathcal{G} : \text{coupling} \\ \text{of } Q_i, P_i}} \phi\left(\underset{(X,Y)\sim\mathcal{G}}{E}\left[d_i(X,Y)\right]\right) \leq C\, D(Q_i \| P_i) \quad \forall\, 1 \leq i \leq n$$

then for $P = P_1 \otimes P_2 \otimes \cdots \otimes P_n$

$$\inf_{\substack{G:\text{ Coupling} \\ \text{of } Q, P}} \sum_{i=1}^{n} \phi\left(\mathop{E}_{(X,Y)\sim G}\left[d_i(X_i, Y_i) \right] \right) \leq C\, D(Q\|P)$$

<u>Pf sketch</u>:- The main idea is the following decomposition result for $D(\cdot\|\cdot)$. If P has density $p_1(x_1) \cdots p_n(x_n)$ and q has density $q(x_1, \ldots x_n)$, then,

$$D(Q\|P) = \int q(x_n)\, q(x_1, \ldots x_{n-1}/x_n) \log\left(\frac{q(x_1 \ldots x_{n-1}/x_n)}{p(x_1) \cdots p(x_{n-1})} \right) dx$$

$$+ D(Q_n \| P_n)$$

$$= E\left[D\left(Q(\ldots\ldots | X_n) \| P_1 \otimes P_2 \cdots \otimes P_{n-1}\right)\right] + D(Q_n\|P_n).$$

The minimizing coupling can then be constructed inductively, and the result can be proved via applications of Jensen's inequality to interchang $E[\cdot]$ and $\phi(\cdot)$.

Transport inequality (1)

$$\min_{\substack{\mathcal{C}:\ \text{coupling } (X,Y)\sim\mathcal{C} \\ \text{of } P,Q}} P[X \neq Y] = \int (p(x)-q(x))_+ \,\mu(dx)$$

$$= \frac{1}{2}\int |p(x)-q(x)|\,\mu(dx) .$$

Transport inequality (2)

$$\min_{\substack{\mathcal{C}:\ \text{coupling } (X,Y)\sim\mathcal{C} \\ \text{of } P,Q)}} E[(X-Y)^2] = E[(X-Z)^2] \quad \text{where}$$

$$Z = G^{-1}(\Phi(X))$$

$$\text{where} \quad G(x) := \Pr_{Z\sim Q}[Z \leq x]$$

$$T(x) = G^{-1}(\Phi(x))$$

$$g(x) = t(x)\,\phi(x). \qquad G(T(x)) = \Phi(x)$$

$$t(T(x))\,\phi(T(x))\,T'(x) = \phi(x)$$

$$T'(x) = \frac{\phi(x)}{t(T(x))\,\phi(T(x))} .$$

$$D(Q\|P) = E[\log t(Y)]$$

$$= E[\log t(T(X))]$$

$$= E\left[\log\frac{\phi(X)}{\phi(Y)} - \log T'(X)\right]$$

$$\geq \frac{1}{2}E[Y^2] - \frac{1}{2}E[X^2] + 1 - E[T'(X)]$$

$$= \frac{1}{2}E[Y^2] - \frac{1}{2}E[X^2] + 1 - \underbrace{E[X\,T(X)]}_{\text{Stein's characterization for the Gaussian}}$$

$$E[T'(X)] = E[X\,T(X)]$$

$\forall\ T$ that grows at most polynomially fast.