Lecture 11 ①

Review: The Entropy Method

* For $Y \geq 0$,

$$\mathrm{Ent}(Y) \stackrel{\text{def}}{=} \mathbb{E}[Y \log Y] - \mathbb{E}[Y] \log \mathbb{E}[Y]$$

1. Herbst's argument: $\dfrac{\mathrm{Ent}(e^{\lambda Z})}{\mathbb{E}[e^{\lambda Z}]} \leq \dfrac{\lambda^2 v}{2}, \ \forall \lambda > 0$

$$\Rightarrow \psi_{Z-\mathbb{E}Z}(\lambda) \leq \frac{\lambda^2 v}{2}, \ \forall \lambda > 0$$

2. Subadditivity of Ent: $\mathrm{Ent}(Y) \leq \sum_{i=1}^{n} \mathbb{E}\left[\mathrm{Ent}_{(i)}(Y)\right]$

3. log-Sobolev inequality and subaddivity are used to get Herbst's condition.

$\boxed{B}$ Step 2: Subadditivity of Ent

We note a connection between $\mathrm{Ent}(\cdot)$ and KL-divergence $D(Q \| P)$. Consider a rv $Y \geq 0$ on (Ω, Σ, P).

$$\mathrm{Ent}(Y) = \mathbb{E}[Y \log Y] - \mathbb{E}[Y \log \mathbb{E}[Y]]$$
$$= \mathbb{E}\left[Y \log \frac{Y}{\mathbb{E}[Y]}\right]$$

Consider $Q \ll P$ s.t. $\dfrac{dQ}{dP} = \dfrac{Y}{\mathbb{E}[Y]}$. Then,

$$D(Q \| P) = \mathbb{E}\left[\frac{Y}{\mathbb{E}[Y]} \log \frac{Y}{\mathbb{E}[Y]}\right] = \frac{\mathrm{Ent}(Y)}{\mathbb{E}[Y]}.$$

(Herbst's condition: $\boxed{D(P_\lambda \| P) \leq \frac{\lambda^2 v}{4}}$, where P_λ is the $e^{\lambda Z}/\mathbb{E}[e^{\lambda Z}]$-tilting of P)

<u>Lemma</u> (Han's inequality for divergence)

Consider two measures Q and P on $\mathcal{X}_1 \times \ldots \times \mathcal{X}_n$ with P being a product measure. Let

$$Q^{(i)}(x^{(i)}) = \sum_{y \in \mathcal{X}_i} Q(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n),$$

and similarly define $P^{(i)}$. Then,

$$D(Q \| P) \le \sum_{i=1}^{n} \left(D(Q \| P) - D(Q^{(i)} \| P^{(i)}) \right).$$

(a) <u>Proof for discrete X</u>

(Han's inequality) For discrete rvs $X_1, \ldots, X_n$,

$$H(X_1, \ldots, X_n) \le \frac{1}{n} \sum_{i=1}^{n} H(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n).$$

<u>Proof.</u> $H(X_1, \ldots, X_n) \le H(X^{i-1}, X_{i+1}^{n}) + H(X_i | X^{i-1})$

$$\Rightarrow n H(X_1, \ldots, X_n) \le \sum_{i=1}^{n} H(X^{i-1}, X_{i+1}^{n}) + H(X_1, \ldots, X_n). \quad \blacksquare$$

Thus, $$H_Q(X) \le \frac{1}{n} \sum_{i=1}^{n} H_Q(\underbrace{X^{(i)}}_{(X^{i-1}, X_{i+1}^{n})})$$

which is the same as

$$\sum_{x} Q(x) \log Q(x) \ge \frac{1}{n} \sum_{i=1}^{n} \sum_{x^{(i)}} Q^{(i)}(x^{(i)}) \log Q^{(i)}(x^{(i)})$$

Also, $\sum_{x} Q(x) \log P(x) = \sum_{x} Q(x) \log \prod_{i=1}^{n} P_i(x_i)$

$$= \mathbb{E}_Q\left[\sum_{i=1}^{n} \log P_i(X_i) \right]$$

$$= \mathbb{E}_Q\left[\frac{1}{n-1} \sum_{i=1}^{n} \log P^{(i)}(X^{(i)}) \right]$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} \mathbb{E}_{Q^{(i)}}\left[\log P^{(i)}(X^{(i)})\right]$$

Therefore,

$$D(Q\|P) = \sum_x Q(x) \log Q(x) - \sum_x Q(x) \log P(x)$$

$$\geq \frac{1}{n-1} \sum_{i=1}^{n} \left(\mathbb{E}_{Q^{(i)}}\left[\log \frac{Q^{(i)}(X^{(i)})}{P^{(i)}(X^{(i)})}\right]\right)$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} D\left(Q^{(i)} \| P^{(i)}\right)$$

which is the same as

$$D(Q\|P) \leq \sum_{i=1}^{n} \left(D(Q\|P) - D\left(Q^{(i)}\|P^{(i)}\right)\right). \qquad \blacksquare$$

Now, recall that $\mathrm{Ent}_{(i)}(Y) = \mathbb{E}_{(i)}[Y \log Y] - \mathbb{E}_{(i)}[Y] \log \mathbb{E}_{(i)}[Y]$.

For $\dfrac{dQ}{dP} = \dfrac{Z}{\mathbb{E}[Z]}$, $\quad \dfrac{dQ_{(i)}}{dP_{(i)}} = \dfrac{\mathbb{E}_{(i)}[Z]}{\mathbb{E}[Z]}$,

$$D(Q\|P) = \frac{\mathrm{Ent}(Z)}{\mathbb{E}[Z]}$$

$$D(Q_{(i)}\|P_{(i)}) = \frac{1}{\mathbb{E}[Z]} \cdot \overbrace{\mathbb{E}\left[\mathbb{E}_{(i)}[Z] \log \frac{\mathbb{E}_{(i)}[Z]}{\mathbb{E}[Z]} \right]}^{\mathbb{E}[Z \log \mathbb{E}_{(i)}[Z]/\mathbb{E}[Z]]}$$

$$\Rightarrow D(Q\|P) - D(Q_{(i)}\|P_{(i)}) = \frac{1}{\mathbb{E}[Z]} \cdot \mathbb{E}\left[Z \log \frac{Z}{\mathbb{E}_{(i)}[Z]} \right]$$

$$= \frac{\mathbb{E}[\mathrm{Ent}_{(i)}(Z)]}{\mathbb{E}[Z]}.$$

Thus, Han's inequality for divergence $\Rightarrow$ subadditivity of Ent.

(b) Proof for general X

We need a new formula for $\mathrm{Ent}(Y)$.

Theorem (Variational formula for Ent)

(a) $$\mathrm{Ent}(Y) = \sup_{T \geq 0 : \mathbb{E}[T] < \infty} \mathbb{E}[Y \log (T/\mathbb{E}[T])]$$

(b) $$\mathrm{Ent}(Y) = \sup_{U \text{ s.t. } \mathbb{E}[e^U] = 1} \mathbb{E}[UY]$$

Proof. Key property of divergence:

$$D(Q \| P) \geq 0 \text{ with equality iff } P = Q$$

(a) $$\mathrm{Ent}(Y) = \mathbb{E}\left[Y \log \frac{Y}{\mathbb{E}[Y]}\right] = \mathbb{E}\left[Y \log \frac{T}{\mathbb{E}[T]}\right] + \mathbb{E}\left[Y \log \frac{Y}{\mathbb{E}[Y]} \cdot \frac{\mathbb{E}[T]}{T}\right]$$

Let Q_1 and Q_2 be defined s.t.

$$\frac{dQ_1}{dP} = \frac{Y}{\mathbb{E}[Y]}, \quad \frac{dQ_2}{dP} = \frac{T}{\mathbb{E}[T]}.$$

$$\mathrm{Ent}(Y) = \mathbb{E}[Y] \left[\mathbb{E}\left[\frac{dQ_1}{dP} \log \frac{dQ_2}{dP}\right] + \underbrace{D(Q_1 \| Q_2)}_{\substack{\geq 0 \text{ with equality} \\ \text{iff } Q_1 = Q_2 \\ Y = T \text{ a.s.}}}\right]$$

$$= \max_{T \geq 0 \text{ s.t. } \mathbb{E}[T] < \infty} \mathbb{E}[Y \log (T/\mathbb{E}[T])]$$

(b) Substitute: $U = \log \frac{T}{\mathbb{E}[T]}$. ∎

Connection with a much more famous formula

Gibbs variational principle: For X on (Ω, Σ, P),

$$\log \mathbb{E}_P[e^X] = \max_{Q \ll P} \left[\mathbb{E}_Q[X] - D(Q\|P) \right]$$

The variational formula we saw earlier can be rephrased as

$$D(Q\|P) = \max_{\substack{T \geq 0 \\ \mathbb{E}[T] < \infty}} \mathbb{E}\left[\frac{dQ}{dP} \log \frac{T}{\mathbb{E}[T]} \right]$$

$$= \max_{\substack{T \geq 0 \\ \mathbb{E}[T] < \infty}} \mathbb{E}_Q\left[\log \frac{T}{\mathbb{E}[T]} \right]$$

$$= \max_{Z \text{ s.t. } \mathbb{E}[e^Z] < \infty} \mathbb{E}_Q[Z] - \log \mathbb{E}_P[e^Z]$$

Variational formula for divergence

$$D(Q\|P) = \max_{Z \text{ s.t. } \mathbb{E}[e^Z] < \infty} \mathbb{E}_Q[Z] - \log \mathbb{E}_P[e^Z]$$

This implies Gibbs variational formula and vice-versa.