Lecture 14

Review

* Compression is equivalent to probability assignment upto 1 bit:

$$\bar{r}(k,n) = \min_{Q^{(n)}} \max_{P \in \mathcal{P}_k} D(P^n \| Q^{(n)})$$

$$r(k,n) = \min_{Q^{(n)}} \max_{\underline{x}} \max_{P \in \mathcal{P}_k} \log \frac{P^n(\underline{x})}{Q^{(n)}(\underline{x})}$$

$$= \min_{Q^{(n)}} \max_{P \in \mathcal{P}_k} \underbrace{\max_{\underline{x}} \log \frac{P^n(\underline{x})}{Q^{(n)}(\underline{x})}}_{D_{max}(P^n \| Q^{(n)})}$$

$$\geq \bar{r}(k,n)$$

* Bayesian method : uniform prior on $\mathcal{P}_k$ (Laplace prior)

$$Q^{(n)}(\underline{x}) = \prod_{i=1}^{n} Q^{(n)}(x_i | x^{i-1}) \text{ with } Q^{(n)}(x_i|x^{i-1}) = \frac{n(x_i|x^{i-1})+1}{i+k-1}$$

add-1 estimator

A Bayesian methods (continued) : Jeffrey's prior (Dirichlet prior for $k>2$)

k=2 case

Consider $\pi(q) \propto \frac{1}{\sqrt{q(1-q)}}$

As before, we need to evaluate

$$\int_0^1 q^{i-1/2}(1-q)^{l-i-1/2}\, dq = \frac{(i-\frac{1}{2})}{(l-i-\frac{1}{2}+1)} \cdots \frac{\frac{3}{2}}{(l-3/2)} \int_0^1 \sqrt{q}\,(1-q)^{l-3/2}\, dq$$

Thus, $Q_\pi(X_{l+1}=0 | X^l = x^l)$ equals

$$\frac{(\ell-i+1/2)}{(\ell-1/2)} \quad \frac{\int_0^1 \sqrt{q}\,(1-q)^{\ell-1/2}\,dq}{\int_0^1 \sqrt{q}\,(1-q)^{\ell-3/2}\,dq}.$$

Note that

$$I_{\ell+1} = \int_0^1 \sqrt{q}\,(1-q)^{\ell-1/2}\,dq = I_\ell - \frac{3}{2(\ell-1/2)} I_{\ell+1}$$

$$\Leftrightarrow \left(\frac{2\ell+2}{2\ell-1}\right) \cdot I_{\ell+1} = I_\ell$$

Therefore, $Q_\pi(X_{\ell+1}=0 \mid X^\ell = x^\ell) = \dfrac{\ell-i+1/2}{\ell+1}$

$$= \frac{n(0\mid x^\ell)+1/2}{n+1} \equiv \text{add-}\tfrac{1}{2} \text{ estimator}$$

(Krichevsky-Trofimov estimator)

Extension to a larger k: $Q_{KT}(x_i \mid x^{i-1}) = \dfrac{n(x_i\mid x^{i-1})+1/2}{i-1+\frac{k}{2}}$

$\boxed{B}$ <u>Using KT estimator for prob. assignment</u>

We restrict to the binary case $(k=2)$

$$Q_{KT}(\underline{x}) = Q_{KT}(x_1)\, Q_{KT}(x_2\mid x_1) \ldots Q_{KT}(x_n\mid x^{n-1})$$

$$= \frac{?}{1} \cdot \frac{?}{2} \cdot \frac{?}{3} \cdot \ldots \frac{?}{n}$$

$1/2$ $\quad \dfrac{n(x_2\mid x_1)+1/2}{2}$ $\quad \dfrac{n(x_n\mid x^{n-1})+1/2}{n}$

* If $\ell = n(1\mid \underline{x})$, the numerator is simply the sequence

$$\frac{1}{2}\cdot\left(1+\frac{1}{2}\right)\cdots\left(\ell-\tfrac{1}{2}\right)\cdot\left(1+\tfrac{1}{2}\right)\cdots\left(n-\ell-\tfrac{1}{2}\right).$$

Thus, $Q_{KT}(\underline{x}) = \frac{1}{n!} \cdot \frac{1}{2} \cdot \frac{3}{2} \cdot \dots \cdot \frac{(2l-1)}{2} \cdot \frac{3}{2} \cdot \frac{5}{2} \cdot \dots \left(\frac{2(n-l)-1}{2} \right)$ ③

$$= \frac{\Gamma(l + 1/2)\ \Gamma(n-l+1/2)}{\Gamma(n+1)} \qquad \left(\Gamma(\alpha) = (\alpha-1)(\alpha-2)\dots \right)$$

Stirling's approx.

$$\left| \log \frac{\Gamma(z)}{e^{-z} z^{z-1/2}} \right| \leq C$$

$$\Rightarrow Q_{KT}(\underline{x}) \geq c' \frac{e^{-n} \cdot (l + 1/2)^{l} (n-l+1/2)^{n-l}}{e^{-1} e^{-n} (n+1)^{n+1/2}}$$

$$\geq c'' \cdot \frac{l^{l} (n-l)^{n-l}}{n^{n+1/2}},$$

and so,

$$\log \frac{P^n(\underline{x})}{Q_{KT}(\underline{x})} \leq l \log p + (n-l) \log(1-p) - l \log \frac{l}{n} - (n-l) \log \frac{n-l}{n} + \frac{1}{2} \log n + O(1)$$

$$= -n\, D\left(\tfrac{l}{n} \middle\| p\right) + \frac{1}{2} \log n + O(1)$$

$$\leq \frac{1}{2} \log n + O(1)$$

Therefore,

$$r(2,n) \leq \frac{1}{2} \log n + O(1).$$

→ In general, we can show

$$r(k,n) \leq \frac{k-1}{2} \log n + O(1).$$

C The Shannon lower bound

$$\bar{\pi}(2,n) = \min_{Q} \max_{P \in \mathcal{P}_2} D(P^n \| Q)$$

$$= \min_{Q} \max_{\pi \in \mathcal{P}(\mathcal{P}_2)} \mathbb{E}_{P \sim \pi}\left[D(P^n \| Q)\right]$$

$$\geq \max_{\pi \in \mathcal{P}(\mathcal{P}_2)} \min_{Q} \mathbb{E}_{P \sim \pi}\left[D(P^n \| Q)\right]$$

Consider a channel $W: \mathcal{X} \to \mathcal{Y}$ with input $\mathcal{X} = \mathcal{P}_2$ and output $\mathcal{Y} = \{0,1\}^n$. Then, the right-side equals

$$\max_{P_X} \min_{Q \in \mathcal{P}(\mathcal{Y})} \mathbb{E}_{P_X}\left[D(W(\cdot|x) \| Q)\right]$$

$$= \max_{P_X} \min_{Q} D(W \| Q | P_X) = \max_{P_X} \min_{Q} D(W \| (W \circ P_X) | P_X) + D(W \circ P_X \| Q)$$

where $W \circ P_X(y) = \sum_x P_X(x) W(y|x)$.

$$= \max_{P_X} D(W \| (W \circ P_X) | P_X) = \max_{P_X} I(P_X ; W) \equiv C(W).$$

Choosing $P_X = \mathrm{unif}(\mathcal{X})$ $\left(\equiv \pi^* = \mathrm{unif}[0,1]\right)$,

$$C(W) \geq \mathbb{E}_{P \sim \mathrm{unif}[0,1]} D(P^n \| P_{\pi^*}^{(n)}).$$

Note now that $P_{\pi^*}^{(n)}(\underline{x})$ is the add-1 prob. assignment.

Therefore, for $n(1|\underline{x}) = \ell$,

$$P^{(n)}_{\pi^*}(\underline{x}) = \frac{\ell!\,(n-\ell)!}{2 \cdot 3 \cdot \ldots \cdot (n+1)}$$

$$\leq \frac{C}{\sqrt{n}} \cdot \left(\frac{\ell}{n}\right)^{\ell} \left(\frac{n-\ell}{n}\right)^{n-\ell} \cdot \sqrt{\frac{\ell(n-\ell)}{(n+1)^2}}$$

Thus,

$$\log \frac{P^n(\underline{x})}{P^{(n)}_{\pi^*}(\underline{x})} \geq -n\,D\left(\frac{\ell}{n} \,\Big\|\, p\right) - O(1) + \frac{1}{2}\log n$$

Next we use $D(\theta \,\|\, p) \leq \dfrac{\theta^2}{p} + \dfrac{(1-\theta)^2}{(1-p)} - 1$

$$\Rightarrow \mathbb{E}_p\left[D(L/n \,\|\, p)\right] \leq \frac{n^2p^2 + np(1-p)}{n^2 p} + \frac{n^2(1-p)^2 + np(1-p)}{n^2(1-p)} - 1$$

$$= \frac{1-p}{n} + \frac{p}{n} = \frac{1}{n}$$

Therefore,

$$D(P^n \,\|\, P^{(n)}_{\pi^*}) \geq \frac{1}{2}\log n - O(1) \Rightarrow \bar{r}(2,n) \geq \frac{1}{2}\log n - O(1).$$

We have shown:

$$r(2,n) = \bar{r}(2,n) = \frac{1}{2}\log n + O(1)$$