Information Theory w/ Applications

Fall 2008 - Homework 1

1. Let $Z : \Omega \to \mathbb{R}$, then for $\tau > 0$, $a \in \mathbb{R}$

$$\mathbb{P}(Z \geq a) = \mathbb{P}(e^{\tau Z} \geq e^{\tau a})$$

$$= \mathbb{E}\left[\mathbb{1}_{e^{\tau Z} \geq e^{\tau a}} \right]$$

$$\leq \mathbb{E}\left[\frac{e^{\tau Z}}{e^{\tau a}} \right] = e^{-\tau a}\, \mathbb{E}\left[e^{\tau Z} \right].$$

2. a)
$$\begin{aligned}
H(X,Y,Z) &= H(X) + H(Y|X) + H(Z|X,Y) \\
&= H(X) + H(Z|X) + H(Y|X,Z) \\
&= H(Y) + H(X|Y) + H(Z|X,Y) \\
&= H(Y) + H(Z|Y) + H(X|Y,Z) \\
&= H(Z) + H(X|Z) + H(Y|X,Z) \\
&= H(Z) + H(Y|Z) + H(X|Y,Z)
\end{aligned}$$

b) $H(X) = H(\mathbb{P}_X)$

where $\mathbb{P}_X(0) = \mathbb{P}_X(1) = \frac{1}{2}$

$$\Rightarrow \quad H(X) = h(\tfrac{1}{2}) = 1 \text{ bit}$$

$H(Y|X=1) = 0$ b/c $P_Y(0|X=1) = 1$

$H(Y|X=0) = 1$ bit b/c $P_Y(0|X=0) = \frac{1}{2}$

so

$$H(Y|X) = \tfrac{1}{2}(0) + \tfrac{1}{2}(1) \text{ bit}$$

~~H~~ Now

$$\mathbb{P}(Z|X=a, Y=b) = \begin{cases} 1, & a=b=0 \text{ or } b=1 \\ \frac{1}{2}, & a=1,\ b=0 \\ 0, & \text{else} \end{cases}$$

gives

$$H(Z|X=0, Y=0) = 0$$

$$H(Z|X=0, Y=1) = 0$$

$$H(Z|X=1, Y=0) = 1 \text{ bit}$$

so

$$H(Z|X,Y) = \tfrac{1}{2}(1) \text{ bit}.$$

c) We conclude

$$H(X,Y,Z) = 1 + \frac{1}{2} + \frac{1}{2} \text{ a bit}$$

$$= 2 \text{ bit},$$

wich is (joint) entropy of prob. measure with 4 equal-prob. outcomes

$$H_4 = \sum_{j=1}^{4} \left(\frac{1}{4}\right) \log_2\left(\frac{1}{4}\right) = 2 \text{ bit}.$$

3. $$H(X|Z) = H(X,Y|Z) - H(Y|X,Z)$$

$$\leq H(X,Y|Z)$$

$$= H(X|Y,Z) + H(Y|Z)$$

$$\leq H(X|Y) + H(Y|Z).$$

4. a) $C_0(P,Q) = -\log \underbrace{\sum_{j=1}^{m} q_j^1}_{1} = 0$

$C_1(P,Q) = 0$ by symmetry

$C_\alpha(P,P) = -\log \underbrace{\sum_{j=1}^{m} p_j}_{1} = 0$

b) By Hölder's ineq.,

setting $p\alpha = q(1-\alpha) = 1 \Rightarrow \frac{1}{p} + \frac{1}{q} = 1$

gives

$$\sum_{j=1}^{m} p_j^{\alpha} q_j^{1-\alpha}$$

$$\leq \underbrace{\left(\sum_{j=1}^{m} p_j^{\overbrace{p\alpha}^{1}}\right)^{\frac{1}{p}}}_{1} \underbrace{\left(\sum_{j=1}^{m} q_j^{\overbrace{q(1-\alpha)}^{1}}\right)^{\frac{1}{q}}}_{1} = 1$$

so

$C_\alpha(P,Q) \geq 0$.

c) $$\frac{\partial}{\partial\alpha} C_\alpha(P,Q) = -\frac{1}{\sum_{i=1}^{m} p_i^\alpha q_i^{1-\alpha}} \left(\sum_{i=1}^{m} \left(p_i^\alpha (\log p_i)\, q_i^{1-\alpha} - p_i^\alpha \log(q_i)\, q_i^{1-\alpha} \right) \right)$$

$$\cancel{= -\left(\log\alpha + \log(1-\alpha)\right)}$$

$$= \mathbb{E}_\alpha\left[-(\log p_i + \log q_i)\right]$$

$$\frac{\partial^2}{\partial\alpha^2} C_\alpha(P,Q) = \frac{1}{\left(\sum_{i=1}^{m} p_i^\alpha q_i^{1-\alpha}\right)^2} \left(\sum_{i=1}^{m} \left(p_i^\alpha \log p_i\, q_i^{1-\alpha} - p_i^\alpha \log q_i\, q_i^{1-\alpha} \right) \right)^2$$

$$- \frac{1}{\sum_{i=1}^{m} p_i^\alpha q_i^{1-\alpha}} \sum_{i=1}^{m} \left(p_i^\alpha (\log p_i)^2 q_i^{1-\alpha} - 2 p_i^\alpha \log p_i \log q_i\, q_i^{1-\alpha} + p_i^\alpha (\log q_i)^2 q_i^{1-\alpha} \right)$$

By Cauchy-Schwarz,

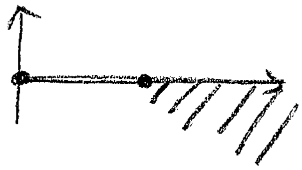
$$\langle \log p_i, \log q_i \rangle_\alpha \le \cancel{4} \|\log p_i\|_\alpha \|\log q_i\|_\alpha \le \tfrac{1}{2}\|\log p_i\|^2 + \tfrac{1}{2}\|\log q_i\|^2$$

where $\langle \cdot, \cdot \rangle = \sum_{i=1}^{m} p_i^\alpha q_i^{1-\alpha} (\cdot\,\cdot)$

$$\Rightarrow \sum_{i=1}^{m} (\cdots) \ge 0 \Rightarrow \frac{\partial^2}{\partial\alpha^2} C_\alpha(P,Q) \le 0$$

$$\Rightarrow C_\alpha(P,Q) \text{ concave.}$$

d) By concavity and $C_0(P,Q) = C_1(P,Q)$,



$C_\alpha(P,Q) \leq 0$ for $\alpha \geq 1$

e) By ~~concavity~~ Hölder, if ~~$\alpha \notin \{0,1\}$~~ $P \neq Q$ and $0 < \alpha < 1$,

then $C_\alpha(P,Q) \cancel{< 1} > 0$.

But since $C_1(P,Q) = 0$ and

$C_\alpha(P,Q)$ is concave, $C_\alpha(P,Q) < 0$

for $\alpha > 1$. Otherwise, $P = Q$

implies $C_\alpha(P,Q) = 0$.

f) Consider α^* such that $C_{\alpha^*}(P,Q)$ maximal,

then $C_{\alpha^*}(P,Q) = D(S_{\alpha^*} \| P) = D(S_{\alpha^*} \| Q)$.

From c) max. is unique and

$$\sum_{i=1}^{m} p_i^{\alpha} q_i^{1-\alpha} \underbrace{(\log p_i - \log q_i)}_{\log \frac{p_i}{q_i}} = 0$$

$$\Rightarrow \frac{1}{N} \sum_{i=1}^{n} p_i^{\alpha} q_i^{1-\alpha} \log p_i = \frac{1}{N} \sum_{i=1}^{m} p_i^{\alpha} q_i^{1-\alpha} \log q_i$$

where N is normalization factor.

Now subtracting ~~fro~~ both sides ~~from~~ from

$$\frac{1}{N} \sum_{i=1}^{m} p_i^{\alpha} q_i^{1-\alpha} \log \frac{p_i^{\alpha} q_i^{1-\alpha}}{N},$$

we obtain $D(S_{\alpha^*} \| P) = D(S_{\alpha^*} \| Q)$.

5. Given acceptance region $\mathcal{A}_n'$

and

$$\alpha_n' = \mathbb{P}(X_1 \notin \mathcal{A}_n')$$

$$\beta_n' = \mathbb{P}(X_2 \in \mathcal{A}_n'),$$

we want to minimize the error probability

$$P(\theta \neq \phi(X^n)) = \pi_1 \alpha_n' + \pi_2 \beta_n'.$$

$$= \pi_1 \sum_{x \notin \mathcal{A}_n'} P_1(x) + \underbrace{\pi_2 \sum_{x \in \mathcal{A}_n'} P_2(x)}_{\pi_2 \left(1 - \sum_{x \notin \mathcal{A}'} P_2(x)\right)}$$

$$= \pi_1 \left(\sum_{x \notin \mathcal{A}_n'} \left(P_1(x) - \frac{\pi_2}{\pi_1} P_2(x)\right)\right) + \pi_2$$

so minimizer has acceptance region

$$\mathcal{A}_n = \{ x \in \mathbb{A}^n : \quad \pi_1 P_1(x) \geq \pi_2 P_2(x) \}$$

For given $x \in \mathbb{A}^n$, let $\mu_x(a) = \frac{1}{n} |\{ j : x_j = a \}|$, $a \in \mathbb{A}$

then

$$\frac{P_1(x)}{P_2(x)} \geq \frac{\pi_2}{\pi_1}$$

is equivalent to componentwise $(x = (x_1, x_2, \ldots x_n))$

$$\sum_{j=1}^{n} \log \frac{P_1(x_j)}{P_2(x_j)} \geq \log \frac{\pi_2}{\pi_1}$$

and

$$\sum_{a \in \mathbb{A}} \mu_x(a) \log \frac{P_1(a)}{P_2(a)} \geq \frac{1}{n} \log \frac{\pi_2}{\pi_1} \equiv \lambda_n \to 0 .$$

By law of large numbers, $\mu_x(a) \xrightarrow{n \to \infty} P_j(a)$ a.s. where $j = \theta$. Moreover, using the result of 2.31 b) we have

$$- \lim_{n \to \infty} \frac{1}{n} \log \alpha_n = D(R_0^\alpha \| P_1)$$

$$- \lim_{n \to \infty} \frac{1}{n} \log \beta_n = D(R_0^\alpha \| P_2)$$

where $\displaystyle\sum_{a \in \mathbb{A}} R_0^\alpha(a) \log \frac{P_1(a)}{P_2(a)} = 0$ and

$$R_0^\alpha(a) = \frac{P_1^\alpha(a) P_2^{1-\alpha}(a)}{\sum_{a} P_1^\alpha(a) P_2^{1-\alpha}(a)} \quad \text{for some } \alpha .$$

From

$$\mathbb{P}(\theta \neq \phi(X^n)) = \pi_1 \alpha_n + \pi_2 \beta_n$$

$$\leq 2 \max \{ \pi_1 \alpha_n, \pi_2 \beta_n \}$$

and

$$\mathbb{P}(\theta \neq \phi(X^n)) \geq \max \{ \pi_1 \alpha_n, \pi_2 \beta_n \},$$

the optimal test minimizes for $n \to \infty$

$$\max \{ \tfrac{1}{n} \log \alpha_n, \tfrac{1}{n} \log \beta_n \} .$$

Consequently, α is chosen to maximize

$$\min \{ D(R_0^\alpha \| P_1), D(R_0^\alpha \| P_2) \} .$$

Since $D(R_0^\alpha \| P_1)$ is monotonically decreasing,
$D(R_0^\alpha \| P_2)$ — " — increasing,
this max is achieved when $D(R_0^\alpha \| P_1) = D(R_0^\alpha \| P_2)$.
Now, by 2.28 f, we know then

$$D(R_0^\alpha \| P_1) = \max_\alpha C_\alpha(P_1, P_2) .$$