Statistics G8243
Thursday, February 26, 2009
Handout #8

(Some) Solutions for HW Set # 2

1. *Entropy of a sum.*

   (a) Here's the way most of you did the problem: Since $Z = X + Y$, $P(Z = z|X = x) = P(Y = z - x|X = x)$.

$$\begin{aligned} H(Z|X) &= \sum P(x)H(Z|X = x) \\ &= -\sum_x P(x) \sum_z P(Z = z|X = x) \log P(Z = z|X = x) \\ &= \sum_x P(x) \sum_y P(Y = z - x|X = x) \log P(Y = z - x|X = x) \\ &= \sum P(x)H(Y|X = x) \\ &= H(Y|X). \end{aligned}$$

   Here's another way, which is more cute. Note that (X, Y) and $(X + Y, X - Y)$ are of course in a 1-1 relationship. Then,

$$\begin{aligned} H(X) + H(Y|X) &= H(X, Y) \\ &= H(X + Y, X - Y) \\ &= H(X, X + Y, X - Y) \\ &= H(X) + H(X + Y|X) + H(X - Y|X + Y, X) \\ &= H(X) + H(X + Y|X), \end{aligned}$$

   and the result follows.

   (b) If X and Y are independent, then $H(Y|X) = H(Y)$. Since conditioning reduces entropy, $H(Z) \geq H(Z|X) = H(Y|X) = H(Y)$. Similarly we can show that $H(Z) \geq H(X)$.

   (c) Consider the following joint distribution for X and Y. Let

$$X = -Y = \begin{cases} 1 & \text{with probability } 1/2 \\ 0 & \text{with probability } 1/2 \end{cases}$$

   Then $H(X) = H(Y) = 1$, but $Z = 0$ with prob. 1 and hence $H(Z) = 0$.

   (d) We have

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

   because Z is a function of (X, Y) and $H(X, Y) = H(X) + H(Y|X) \leq H(X) + H(Y)$. We have equality iff (X, Y) is a function of Z and $H(Y) = H(Y|X)$, i.e., X and Y are independent.

2. *Inequalities.*

(a) Using the chain rule for conditional entropy,

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

with equality iff $H(Y|X,Z) = 0$, that is, when Y is a function of X and Z.

(b) Using the chain rule for entropy and the fact that conditioning reduces entropy,

$$\begin{aligned} H(X,Y,Z) - H(X,Y) &= H(Z|X,Y) \\ &\leq H(Z|X) \\ &= H(X,Z) - H(X), \end{aligned}$$

with equality iff Y and Z are conditionally independent given X.

3. *A compound Poisson approximation bound.* Let $X_1, X_2, \ldots, X_n$ be IID Bern(λ/n) random variables, let $Y_1, Y_2, ..., Y_n$ be an independent sequence of IID random variables with distribution $Q = (1/2, 1/2)$ on the set $\{1,2\}$, let $Z_1, Z_2, \ldots, Z_n$ are IID CPo$(\lambda/n, Q)$, If we write T_n for the sum of the Z_i, then P_{T_n} is the CPo(λ, Q) distribution. Following the same steps as those used for Poisson

$$\begin{aligned} D(P_{S_n} \| \mathrm{CPo}(\lambda, Q)) &= D(P_{S_n} \| P_{T_n}) \\ &\leq D(P_{X_1Y_1,\ldots,X_nY_n} \| P_{Z_1,\ldots,Z_n}) \\ &= nD(P_{X_1Y_1} \| P_{Z_1}). \end{aligned}$$

Now if $X \sim$ Bern(p), $Y \sim Q$ and $Z \sim$ CPo(p, Q) are independent, then it is easy to calculate that,

$$D(P_{XY} \| P_Z) \leq p^2 + (1-p)[p + \log(1-p)] - \frac{p}{2}\log(1 + p/4) \leq p^2,$$

and combining the last two bounds, taking $p = \lambda/n$ yields the result.

4. *Monotonicity.* Note that we can write,

$$\begin{aligned} \hat{P}_{n+1}(x) &= \frac{1}{n+1} \sum_{1 \leq i \leq n+1} \mathbb{I}_{\{X_i = x\}} \\ &= \frac{1}{n+1} \sum_{1 \leq j \leq n+1} \frac{1}{n} \sum_{i \neq j} \mathbb{I}_{\{X_i = x\}} \\ &= \frac{1}{n+1} \sum_{1 \leq j \leq n+1} \hat{P}_n^{(j)}, \end{aligned}$$

where $\hat{P}_n^{(j)}$ is the empirical distribution induced by $X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_n$. Then, by the convexity of the relative entropy,

$$D(\hat{P}_{n+1} \| Q) = D\Big(\frac{1}{n+1} \sum_{1 \leq j \leq n+1} \hat{P}_n^{(j)} \Big\| Q\Big) \leq \frac{1}{n+1} \sum_{1 \leq j \leq n+1} D(\hat{P}_n^{(j)} \| Q),$$

and since the $\hat{P}_n^{(j)}$ all have the same distribution, taking expectations yields,

$$E\big[D(\hat{P}_{n+1}\|Q)\big] \le E\big[D(\hat{P}_n\|Q)\big],$$

as claimed.

5. *Estimating the entropy.* The fact that $E[H(\hat{P}_n)] \le H(P)$ follows from the concavity of the entropy (using Jensen's inequality), upon noting that $E(\hat{P}_n) = P$.

6. *Size of type-class.* This is mostly straightforward computations.

7. *Hypothesis testing.* Stein Lemma says we should use the decision region,

$$B_n^* = B_n^*(\epsilon) = \{x_1^n \in A^n \ : \ 2^{D(P_1\|P_2)-\epsilon} \le \frac{P_1^n(x_1^n)}{P_2^n(x_1^n)} \le 2^{D(P_1\|P_2)+\epsilon}\}.$$

Under Q, the likelihood ratio of interest has,

$$\frac{1}{n}\log\frac{P_1^n(X_1^n)}{P_2^n(X_1^n)} \to R := \sum_x Q(x)\log\frac{P_1(x)}{P_2(x)} = D(Q\|P_2) - D(Q\|P_1),$$

in probability, as $n \to \infty$.

Now consider three cases: $R < D(P_1\|P_2) - \epsilon$, $D(P_1\|P_2) - \epsilon < R < D(P_1\|P_2) + \epsilon$, and $R > D(P_1\|P_2) + \epsilon$.