HW Set # 2

Homework due by Thursday February 19, at 2:40pm, at the *beginning* of class.

1. *Entropy of a sum.* Let X and Y be two (possibly dependent) random variables, with possibly different alphabets. Let $Z = X + Y$.

   (a) Show that $H(Z|X) = H(Y|X)$.

   (b) Show that if X, Y are independent, then $H(Y) \le H(Z)$ and $H(X) \le H(Z)$. Thus the addition of *independent* random variables adds uncertainty.

   (c) Give an example (of necessarily dependent random variables) in which $H(X) > H(Z)$ and $H(Y) > H(Z)$.

   (d) Under what conditions does $H(Z) = H(X) + H(Y)$?

2. *Inequalities.* Let X, Y and Z be possibly dependent random variables. Prove the following inequalities and find conditions for equality.

   (a) $H(X, Y|Z) \ge H(X|Z)$.

   (b) $H(X, Y, Z) - H(X, Y) \le H(X, Z) - H(X)$.

3. *A compound Poisson approximation bound.* Let $X_1, X_2, \ldots, X_n$ be IID Bern(λ/n) random variables, and 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 $\{1, 2\}$. Show that the distribution P_{S_n} of the sum $\sum_{i=1}^n X_i Y_i$ satisfies,

$$D(P_{S_n} \| \mathrm{CPo}(\lambda, Q)) \le \frac{(\log e)\lambda^2}{n},$$

   for all $n \ge 1$, where the compound Poisson distribution CPo$(\lambda, Q))$ is the distribution of the sum of Poisson(λ) many IID random variables $Y_i \sim Q$ as above. [*Hint.* Recall the infinite divisibility property of the compound Poisson distribution.]

4. *Monotonicity.* Let $\hat{P}_n$ denote the (random) empirical distribution, or type, of a random string X_1^n which is generated IID according to some distribution Q on A. Prove that $\hat{P}_n$ approaches to Q monotonically in terms of relative entropy, in that,

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

   for all n.

5. *Estimating the entropy.* Let $X_1, X_2, \ldots$ be IID random variables with distribution P given by $(1/8, 1/8, 1/4, 1/2)$ on $A = \{1, 2, 3, 4\}$.

(a) Use your favorite computer program to simulate about 50 such random variables X_i, and calculate the empirical distribution, or type $\hat{P}_n$ of your samples $X_1, \ldots, X_{50}$.

(b) Pretend you don't know the true distribution, and use the empirical distribution to come up with an estimate $\hat{H}$ for the entropy, by taking

$$\hat{H} = H(\hat{P}_n).$$

Compare your estimate with the true value of $H(X)$.

(c) Repeat steps (a) and (b) 10 or 20 times. Comment on the estimates you get. Do you see any systematic trends?

(d) Show that this entropy estimate is biased "downwards," that is, prove that for any distribution P,

$$E[H(\hat{P}_n)] \le H(P).$$

6. *Size of type-class.* In class we saw that

$$\frac{1}{(n+1)^m} 2^{nH(P)} \le |T(P)| \le 2^{nH(P)}.$$

Here you will show that $|T(P)|$ is in fact $\approx n^{-\frac{m-1}{2}} 2^{nH(P)}$. More precisely, assuming P is an n-type with all positive probabilities, show that

$$\log |T(P)| = nH(P) - \frac{m-1}{2} \log(2\pi n) - \frac{1}{2} \sum_{a \in A} \log P(a) - \frac{\vartheta(n,P)}{12} (\log e)^2 m$$

where $0 \le \vartheta(n,P) \le 1$.

Hint. Use Robbins' sharpening of Stirling's formula

$$\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n+\frac{1}{12(n+1)}} \le n! \le \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n+\frac{1}{12n}}$$

(see, e.g., Feller, vol. I, p. 54, or the handout given in class), noticing that $P(a) \ge \frac{1}{n}$ whenever $P(a) > 0$.

7. *Hypothesis testing.* Suppose we are in the setting of Stein's Lemma of testing IID data $X_1, X_2, \ldots$ for having distribution either P_1 or P_2. What would happen if we followed the asymptotically optimal procedure suggested by Stein's Lemma, but instead the data came from a third distribution Q?

8. *OPTIONAL. The Poisson as a maximum entropy distribution.* Let $B_n(\lambda)$ denote the class of all probability distributions that can be obtained as distributions of sums $\sum_{i=1}^n X_i$ of independent Bern(p_i) random variables, such that $p_1 + p_2 + \cdots + p_n = \lambda$. Also, let $B(\lambda)$ denote the union of all the $B_n(\lambda)$, $B(\lambda) = \cup_{n \ge 1} B_n(\lambda)$. Here you will show that the Poisson has the following maximum entropy property:

$$H(\text{Po}(\lambda)) = \sup_{P \in B(\lambda)} H(P).$$

(a) Suppose a function $f(x, y)$ is defined for x, y such that $x + y =$ a constant c, let $z = x - \frac{c}{2}$, and define a new function $g(z) = f(\frac{c}{2} + z, \frac{c}{2} - z)$. Show that if $g''(z) \leq 0$ then f is jointly concave in (x, y).

(b) Suppose that f is also symmetric in x, y, i.e., $f(x, y) = f(y, x)$. Show that f achieves its maximum at the point $(x, y) = (\frac{c}{2}, \frac{c}{2})$.

(c) Take $\lambda > 0$ fixed, let $S_n = \sum_{i=1}^{n} X_i$ where the X_i are independent Bernoulli RVs with parameters p_i such that $\sum_i p_i = \lambda$, and let $h(p_1, p_2, \ldots, p_n) = H(S_n)$. Using part (a), you will prove the following claim: *For any two $i \neq j$, h is a concave function of (p_i, p_j) when all the rest of the p's are kept fixed.*

   i. Without loss of generality (why?) take $i = 1$ and $j = 2$, so that $p_1 + p_2$ is equal to the constant $c = \lambda - p_3 - p_4 - \cdots - p_n$. Let $f(p_1, p_2) = h(p_1, p_2, \ldots, p_n)$, $z = p_1 - \frac{c}{2}$, and $g(z) = f(\frac{c}{2} + z, \frac{c}{2} - z)$. Define $u = z^2$, and explain why $\frac{d^2}{du^2} H(S_n) \leq 0$.

   ii. Express $\Pr(S_n = \ell)$ as the product of the symmetric functions s_ℓ^n (evaluated at suitable arguments) and a simple product involving the $p_i's$. Recall that the s_k^n are defined by

$$s_k^n(x_1, x_2, \ldots, x_n) = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} x_{i_1} x_{i_2} \cdots x_{i_k},$$

   for all positive x_i, and that they have the property that

$$s_k^n \, s_{k+2}^n \leq (s_{k+1}^n)^2.$$

   iii. Expressing $\frac{d}{du} H(S_n)$ as a telescoping sum show that

$$\frac{d}{du} H(S_n) = \sum_{\ell=0}^{n} \log \left(\frac{\Pr(S_n = \ell)\Pr(S_n = \ell + 2)}{\Pr(S_n = \ell + 2)^2} \right) \Pr(X_3 + X_4 + \cdots + X_n = \ell),$$

   and use part ii. to show that $\frac{d}{du} H(S_n) \leq 0$.

   iv. Show that we can expand $g''(z)$ as

$$g''(z) = \frac{d^2}{dz^2} H(S_n) = 2\frac{d}{du} H(S_n) + \left(\frac{du}{dz}\right)^2 \frac{d^2}{du^2} H(S_n),$$

   and use ii and iii to show that $g''(z) \leq 0$. Conclude that the claim in (c) is true.

(d) Show that

$$H\Big(\text{Bin}\Big(n, \frac{\lambda}{n}\Big)\Big) = \max_{P \in B_n(\lambda)} H(P).$$

(e) Take $n \to \infty$ in (d) and conclude that the Poisson has the stated maximum entropy property.