Homework Set #4

1. Let $a, b \geq 0$ and 
   
   \[ f(x) = (x + a) \log(x + b) - x \log x. \]
   
   (a) Show that $f(x)$ is increasing in $x > 0$.
   (b) Show that $f(x) \leq a \log(x + b) + b \log e$ for $x > 0$.

2. Let $c(x^n)$ be the number of complete phrases in the Lempel–Ziv parsing of a sequence $x^n$.
   (a) Show that 
   \[ c(x^n) \geq \left\lceil -1 + \frac{\sqrt{8n + 1}}{2} \right\rceil \geq \sqrt{2} n - 2. \]
   (b) Show that for any $\epsilon > 0$, if 
   \[ c \geq \frac{n+1}{\log(n+1) - (1+\epsilon) \log \log(n+1)}, \]
   then $c \log c > n$ for $n \geq 1$. Conclude that $c \log c \leq n$ implies 
   \[ c \leq \frac{n+1}{\log(n+1) - \log \log(n+1)}. \]
   (c) Let $\mathcal{X} = \{0, 1\}$ and $k^*$ be the largest $k$ such that $n \geq (k-1)2^{k+1} + 2$. Show that 
   \[ c(x^n) \leq 2^{k^*+2} - 3. \]
   (d) Using parts (b) and (c), show that 
   \[ c(x^n) \leq \frac{2(n+4)}{(\log(n+4) - 2) - \log(\log(n+4) - 2)}. \]
   (e) Generalize the result in (d) to the case $|\mathcal{X}| = m \geq 2$.

3. Polya’s urn. Suppose that we have an urn containing one ball labeled 0 and one ball labeled 1. We draw a ball at random from the urn and observe its label $X_1$. If $X_1 = 0$, we put the drawn ball plus another ball labeled 0 into the urn. If $X_1 = 1$, we put the drawn ball plus another ball labeled 1 into the urn. We then repeat this process. At the $n$-th stage, we draw a ball at random from the urn with $(n+1)$ balls, note its label $X_n$, and put the drawn ball plus another ball with the same label into the urn.
   (a) Show that $X = (X_n)_{n=1}^{\infty}$ is stationary.
(b) Show that $X$ is exchangeable, namely, the distribution of $(X_1, \ldots, X_n)$ is equal to that of $(X_{\sigma(1)}, \ldots, X_{\sigma(n)})$ for every permutation $\sigma$ and $n = 1, 2, \ldots$.

(c) Find the $n$-th order pmf $p(x^n)$.

(d) Find the conditional pmf $p(x_{n+1}|x^n)$ of $X_{n+1}$ given $\{X^n = x^n\}$.

(e) Find the entropy rate of $X$.

(f) Show that $X$ is not ergodic by proving that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = Z \text{ a.s.},$$

where $Z \sim U[0,1]$.

4. Show that the Lempel–Ziv probability measure $Q_{LZ}$ is mean universal (w.r.t. the class of all stationary ergodic processes).

5. **Lempel–Ziv process.** Let $X = (X_n)_{n=1}^\infty$ be the Lempel–Ziv process.

   (a) Show that $P(X_n = 1) = 1/2$, $n = 1, 2, \ldots$.

   (b) Given $\{X^n = x^n = (y^c, z)\}$, where $y^c$ is the sequence of complete phrases in $x^n$, find the conditional pmf $p(x_{n+1}|y^c, z)$ in terms of $y^c$ and $z$.

   (c) Let $N_1 = 0$ and $N_i, i = 2, 3, \ldots$, be the stopping time at which $X^{N_i}$ is parsed into exactly $i$ complete phrases, i.e., $X^{N_i} = Y^i$. Define

   $$U_i = X_{N_i+1}, \quad i = 1, 2, \ldots.$$ 

   Is $U = (U_n)_{n=1}^\infty$ stationary? Find the $n$-th order pmf $p(u^n)$. Compare the answer with Pólya’s urn model.

   (d) More generally, let $z \in X^k$ and let $N_i, i = 1, 2, \ldots$, be the stopping time at which $X^{N_i}$ is parsed into exactly $i$ complete phrases and the remainder $z$, i.e., $X^{N_i} = (Y^i, z)$. Define

   $$U_i = X_{N_i+1}, \quad i = 1, 2, \ldots.$$ 

   Find the $n$-th order pmf $p(u^n)$. 
