Solutions: Homework Set # 3

Problem 1

A sequence of random variables $\{X_n\}$ **converges** toward X **in probability** if

$$\lim_{n\to\infty} \Pr[|X_n - X| \geq \varepsilon] = 0,$$

for any $\varepsilon > 0$. For example the Weak Law of Large Numbers implies that if $X_1, X_2, \ldots$ is a sequence of i.i.d. random variables, and $S_n = \frac{1}{n}\sum_{i=1}^n X_n$, then

$$\lim_{n\to\infty} \Pr[|S_n - \mathbb{E}[X]| \geq \varepsilon] = 0.$$

In other words, S_n converges to $\mathbb{E}[X]$ in probability and we denote this by $S_n \stackrel{\text{in probability}}{\longrightarrow} \mathbb{E}[X]$.

(a)

$$Y_n = \frac{1}{n}\log p\left(X_1, \ldots, X_n\right)$$

Since $\{X_i\}$ are independent $p\left(x_1, \ldots, x_n\right) = \prod_{i=1}^n p\left(x_i\right)$.
The random variables $\log p\left(x_i\right)$ are i.i.d. because X_i are i.i.d. So

$$Y_n = \frac{1}{n}\log\left(\prod_{i=1}^n p\left(x_i\right)\right) = \frac{1}{n}\sum_{i=1}^n \log p\left(x_i\right)$$

By the weak law of large numbers,

$$Y_n \stackrel{\text{in probability}}{\longrightarrow} \mathbb{E}\{\log p\left(X\right)\} = -H(X)$$

So, Y_n converges in probability to

$$\begin{aligned}
H\left(X\right) &= -\left[\frac{8}{23}\log\frac{8}{23} + \frac{6}{23}\log\frac{6}{23} + \frac{4}{23}\log\frac{4}{23} + \frac{2}{23}\log\frac{2}{23} + \frac{2}{23}\log\frac{2}{23} + \frac{1}{23}\log\frac{1}{23}\right] \\
&= -\left[\frac{1}{23}\left(8\log 8 + 8\log 23 + 6\log 6 + 6\log 23 + 4\log 4 + 4\log 23\right)\right] \\
&\quad - \left[\frac{1}{23}\left(4\log 2 + 4\log 23 + \log 1 + \log 23\right)\right] \\
&= -\frac{1}{23}\left(-23\log 23 + 24 + 6\log 6 + 8 + 4 + 0\right) \\
&= 2.28 \;\; \mathit{bits}
\end{aligned}$$

(b)

$$Z_n = \frac{1}{n}\sum_{i=1}^n X_i^2$$

Let $T_i = X_i^2$. We know $\{T_i\}$ are i.i.d. since $\{X_i\}$ are i.i.d. So again by the law of large numbers,

$$Z_n \overset{\text{in probability}}{\longrightarrow} \mathbb{E}\{T\} = \mathbb{E}\{X^2\}$$

$$\begin{aligned}\mathbb{E}\{X^2\} &= 0^2.\frac{8}{23} + 1^2.\frac{6}{23} + 2^2.\frac{4}{23} + 3^2.\frac{2}{23} + 4^2.\frac{2}{23} + 5^2.\frac{1}{23} \\ &= \frac{6}{23} + \frac{16}{23} + \frac{18}{23} + \frac{32}{23} + \frac{25}{23} = \frac{97}{23}\end{aligned}$$

(c)

$$Z = \mathbb{E}\{X^2\} = \frac{97}{23}$$

and

$$\left(\mathbb{E}\{X^2\}\right)^2 = \left[0.\frac{8}{23} + 1.\frac{6}{23} + 2.\frac{4}{23} + 3.\frac{2}{23} + 4.\frac{2}{23} + 5.\frac{1}{23}\right]^2 = \left(\frac{33}{23}\right)^2$$

We will show that $Z = \mathbb{E}\{X^2\} \geq \mathbb{E}^2\{X\}$. To see why this is so, let a random variable $V = (X - \mathbb{E}\{X\})^2 \geq 0$.

$$\begin{aligned}\mathbb{E}\{V\} &= \mathbb{E}\left\{(X - \mathbb{E}\{X\})^2\right\} \\ &= \mathbb{E}\{X^2 + \mathbb{E}^2\{X\} + 2X\mathbb{E}\{X\}\} \\ &= \mathbb{E}\{X^2\} + \mathbb{E}^2\{X\} - 2\mathbb{E}^2\{X\} \\ &= \mathbb{E}\{X^2\} - \mathbb{E}^2\{X\}\end{aligned}$$

But $\mathbb{E}\{V\} \geq 0$, since V can only take non-negative values. Thus $\mathbb{E}\{X^2\} \geq \mathbb{E}^2\{X\}$ in general.

Problem 2

For a Markov chain, by the data processing theorem, we have

$$I(X_0; X_{n-1}) \geq I(X_0; X_n)$$

Therefore,

$$H(X_0) - H(X_0 \mid X_{n-1}) \geq H(X_0) - H(X_0 \mid X_n)$$

or $H(X_0 \mid X_n)$ increases with n.

Problem 3

We know that we can expand the mutual information as follows

$$\begin{aligned}I(\underbrace{X_1, X_2, \ldots, X_n}_{Y}; \underbrace{X_{n+1}, \ldots, X_{2n}}_{Z}) =& I(Y; Z) \\ =& H(Y) + H(Z) - H(Y, Z) \\ =& H(X_1, X_2, \ldots, X_n) + H(X_{n+1}, \ldots, X_{2n}) \\ &- H(X_1, X_2, \ldots, X_n, X_{n+1}, \ldots, X_{2n}) \\ =& H(X_1, X_2, \ldots, X_n) + H(X_{n+1}, \ldots, X_{2n}) \\ &- H(X_1, X_2, \ldots, X_{2n}).\end{aligned}$$

Because the process is stationary we have $H(X_{n+1},\ldots,X_{2n}) = H(X_1,X_2,\ldots,X_n)$ so we can write

$$I(X_1,X_2,\ldots,X_n;X_{n+1},\ldots,X_{2n}) = 2H(X_1,X_2,\ldots,X_n) - H(X_1,X_2,\ldots,X_{2n}).$$

Deviding both sides by $\frac{1}{2n}$ and taking the limit we have

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{2n}I(X_1,\ldots,X_n;X_{n+1},\ldots,X_{2n}) &= \lim_{n\to\infty}\frac{1}{2n}\left[2H(X_1,\ldots,X_n) - H(X_1,\ldots,X_{2n})\right] \\
&= \lim_{n\to\infty}\frac{1}{n}H(X_1,\ldots,X_n) - \lim_{n\to\infty}\frac{1}{2n}H(X_1,\ldots,X_{2n}) \\
&= H(\mathcal{X}) - H(\mathcal{X}) \\
&= 0.
\end{aligned}$$

Problem 4

(a) We have

$$\begin{aligned}
\frac{-1}{n}\log q(X_1,X_2,\ldots,X_n) &= \frac{-1}{n}\log\prod_{i=1}^{n} q(X_i) \\
&= \frac{-1}{n}\sum_{i=1}^{n}\log q(X_i).
\end{aligned}$$

According to L.L.N. (Law of Large Numbers) we can write

$$\begin{aligned}
\frac{-1}{n}\log q(X_1,X_2,\ldots,X_n) &\overset{\text{in probability}}{\longrightarrow} -E(\log q(X_1)) \\
&= \sum_{x=1}^{m} p(x)\log\frac{1}{q(x)}\frac{p(x)}{p(x)} \\
&= D(p\|q) + H(p).
\end{aligned}$$

(b)

$$\begin{aligned}
\frac{1}{n}\log\frac{q(X_1,X_2,...,X_n)}{p(X_1,X_2,...,X_n)} &= \frac{1}{n}\sum_{i=1}^{n}\log\frac{q(X_i)}{p(X_i)} \\
&\overset{\text{in probability}}{\longrightarrow} E[\log\frac{q(X_1)}{p(X_1)}] \\
&= -D(p\|q).
\end{aligned}$$

Thus,

$$\frac{q(x_1,X_2,...,X_n)}{p(x_1,X_2,...,X_n)} = 2^{-nD(p\|q)}.$$