1 26 letters : 21 consonants
5 vowels

$P(x \in C) = \frac{1}{21+5k}$ $P(x \in V) = \frac{k}{21+5k}$

For $k=1$ $P(x) = \frac{1}{26}$

a) $H = \log 26 = \begin{cases} \log_{26} 26 = 1 \text{ latin character} \\ \log_2 26 = 4.7 \text{ bits} \end{cases}$

b) For $k=10$ $P(x \in C) = \frac{1}{71}$ $P(x \in V) = \frac{10}{71}$

$H = \frac{21}{71} \log 71 + \frac{50}{71} \log 7.1$

$= \begin{cases} 0.81 \text{ latin characters} \\ 3.81 \text{ bits} \end{cases}$

For $k = \infty$ $P(x \in C) = 0$ $P(x \in V) = \frac{1}{5}$

$H = \log 5 = \begin{cases} 0.49 \text{ latin characters} \\ 2.32 \text{ bits} \end{cases}$

c) Optimally compressed $\Rightarrow$ fully random

$H[B] = \log 2 = 1$ bit

2 $p(x) = p(y)\,p(z)$

$$-H[X] = \sum_{x} p(x) \log p(x)$$

$$= \sum_{y,z} p(y)\,p(z) \log p(y)\,p(z)$$

$$= \sum_{y,z} p(y)\,p(z)\left[\log p(y) + \log p(z)\right]$$

$$= \underbrace{\left[\sum_{z} p(z)\right]}_{=1} \sum_{y} p(y) \log p(y)$$

$$+ \underbrace{\left[\sum_{y} p(y)\right]}_{=1} \sum_{x} p(x) \log p(x)$$

$$= -H[Y] - H[Z]$$

$$\therefore\ H[X] = H[Y] + H[Z]$$

as required

3

a) $K_{all} = K_{odd} + K_{even}$

clearly $\frac{K_{odd}}{N/2} \to S[p(x)]$

K_{even} = information required to say "same as preceding odd"

$= O(1)$

$\therefore \frac{K_{all}}{N} \to \frac{S[p(x)]}{2}$

b) $K_{all} = K_{first\ n} + K_{rest}$

Clearly $\frac{K_{first\ n}}{n} \to S[p(x)]$

$\frac{K_{rest}}{N-n} \to S[p'(x)]$ $\therefore$ the required relation

c) Equivalent to $p(1)=p$ for odd and $p(0)=p'$ for even, since we can losslessly transform the string to this. Given this and the answers to (a) and (b), the required relation follows.

4. I'll use p_x for $p(x)$, etc.

Let's take the derivative of $S(p_x)$

$$\frac{d}{dp_x} S(p_x) = \lim_{dp_x \to 0} \frac{S(p_x + dp_x) - S(p_x)}{dp_x}$$

Note that $p_x + dp_x = p_x(1 + dp_x/p_x)$

$$\therefore S(p_x + dp_x) = S(p_x) + S(1 + dp_x/p_x)$$

$$\therefore \frac{d}{dp_x} S(p_x) = \lim_{dp_x \to 0} \frac{S(1 + dp_x/p_x)}{dp_x}$$

Writing $S(1+dp_x/p_x)$ as a Taylor series in terms of dp_x/p_x

$$\lim_{dp_x \to 0} \frac{S(1+dp_x/p_x)}{dp_x}$$

$$= \lim_{dp_x \to 0} \frac{c_0 + c_1\, dp_x/p_x + c_2 (dp_x/p_x)^2}{dp_x}$$

A non-zero c_0 causes this to diverge, so $c_0 = 0$. The limit then gives

$$\frac{d}{dp_x} S(p_x) = \frac{c_1}{p_x}$$

From here it should be straightforward.