**Quantum Information** 2018
**Sheet 1**

### 1. Shannon Entropy

Consider the 26 character long basic Latin alphabet (in capitals): $A, B, C, \ldots, X, YZ$. A block of text is generated by randomly choosing letters from this alphabet. All consonants are equally probable, as are all vowels, but each vowel is $k$ times more likely than each consonant.

a) What is the Shannon entropy when $k = 1$? Give the answer both in units of bits and the above basic Latin letters.

b) Do the same for $k = 10$ and $k = \infty$.

c) Consider a block of text encoded as a bit string with optimal compression. What is the Shannon entropy per bit for this encoding?

### 2. Composite Variables

Consider a random variable $X$ made up of two components, $Y$ and $Z$.

a) If the components are uncorrelated, such that $p(x) = p(y)p(z)$, show that

$$H[X] = H[Y] + H[Z].$$

b) Show that, in general

$$H[X] \leq H[Y] + H[Z].$$

### 3. Correlated Variables

Consider the asymptotic compression rate,

$$\lim_{N\to\infty} \frac{K_N[X]}{N},$$

where $K_N[X]$ is the amount of information required to store $N$ outcomes of the random variable $X$ after compression. If the outcomes of $X$ are i.i.d., this compression rate is equal to the Shannon entropy $H[X]$. If they are not i.i.d., the compression rate must be calculated differently.

a) Let's first consider the case that the variables are not independent. Instead, let's suppose that each odd numbered outcome occurs randomly with probability distribution $\{p(x)\}$, but each even numbered outcome is identical to the odd numbered outcome that preceded it. Explain why

$$\lim_{N\to\infty} \frac{K_N[X]}{N} = \frac{H[p(x)]}{2}.$$

b) Now suppose that the variables are not identically distributed. The first $n$ values are generated according to the probability distribution $\{p(x)\}$, while the next $N-n$ are generated according to $\{p'(x)\}$. Explain why

$$\lim_{N\to\infty} \frac{K_N[X]}{N} = \frac{nH[p(x)] + (N-n)H[p'(x)]}{N}.$$

c) Finally, consider the case that $x \in \{0,1\}$. For odd numbered outcomes $p(0) = p$ and $p(1) = 1-p$. Even numbered outcomes have the same result as the preceding odd numbered outcome with probability $p'$, but the opposite value otherwise. Explain why

$$\lim_{N\to\infty} \frac{K_N[X]}{N} = \frac{H(p) + H(p')}{2}.$$

**4. Surprise!**

Assume that the surprise $S(x)$ for a particular outcome $x$ of a random variable $X$ is a continuous function of the the probability $p(x)$. Also assume the conditions given in the slides:

- $S(x) = 0$ if $p(x) = 1$;
- $S(x) \to \infty$ as $p(x) \to 0$;
- $S(xy) = S(x) + S(y)$ if $p(xy) = p(x)p(y)$.

Show then that

$$S(x) = -\log_n p(x),$$

is the *unique* measure of surprise that satisfies all conditions (up to the choice of $n$).

You may also assume that $S(x)$ can be represented by a Taylor series.

**5. Insecurity of the Many-time Pad**

Consider an $n$ character message sent with an $n$ character key, as described in lecture. The messages sent are typical sentences written in English. In the lecture we assumed that a different key is be used for each message. However, suppose that the same key is used for a very large number of messages.

a) Show that this does not satsify the conditional derived in lecture for security.

b) Prove that it is insecure by finding a way to deduce the key from the collection of encoded messages.