# 11 Entropy and information

**Exercises:** 11.1, 11.2, 11.3, 11.4, 11.5, 11.6, 11.7, 11.8, 11.9, 11.10, 11.11, 11.12, 11.13, 11.14, 11.15, 11.16, 11.17, 11.18, 11.19, 11.20, 11.21, 11.22, 11.23, 11.24, 11.25, 11.26.

## 11.1

Fair coin:

$$H(X) = 2\left(-\frac{1}{2}\log\frac{1}{2}\right) = 1.$$

Fair die:

$$H(X) = 6\left(-\frac{1}{6}\log\frac{1}{6}\right) = \log 6 \approx 2.585.$$

If they were unfair, the entropy would be smaller. Since there would be at least one outcome more probable than another, the average information gain after each toss would be smaller.

## 11.2

From property (1) we know that $p$ and $q$ are real values in the range $[0, 1]$, so there are always numbers $a$ and $b$ in the range $(-\infty, 0]$ such that $p = 2^a$ and $q = 2^b$. So we have $I(pq) = I(2^a 2^b) = I(2^{a+b})$, thus from property (3) we obtain that

$$I(2^{a+b}) = I(2^a) + I(2^b).$$

If we define the function $f(x) \equiv I(2^x)$, we get $f(a+b) = f(a) + f(b)$. Because of property (2), we know that $f$ must be continuous and smooth, and therefore, has a Taylor series form. For $a = b = 0$ we get $f(0) = 0$, and for $b = -a$ we get $f(-a) = -f(a)$, so it is clearly an odd function, hence

$$f(x) = \sum_{j=0}^{\infty} k_j x^{2j+1}.$$

Considering $x = a + b$ we obtain

$$f(a+b) = \sum_{j=0}^{\infty} k_j (a+b)^{2j+1} = k_0(a+b) + \sum_{j=1}^{\infty} k_j \sum_{n=0}^{2j+1} \binom{2j+1}{n} a^n b^{2j+1-n},$$

but using the property that $f$ must satisfy, we get

$$f(a) + f(b) = \sum_{j=0}^{\infty} k_j (a^{2j+1} + b^{2j+1}) = k_0(a+b) + \sum_{j=1}^{\infty} k_j (a^{2j+1} + b^{2j+1}).$$

Therefore, the only way for the property to be true is if $k_j = 0$ for all $j > 0$, meaning $f(x) = kx$ for some constant $k$. Now, since $f(x) = I(2^x)$, it is also true that $I(x) = f(\log x)$, and this gives us

$$I(p) = k \log p.$$

The average of this function over the values of the set $\{p_1, \cdots, p_n\}$ is precisely the Shannon entropy up to a multiplicative constant $k$

$$H(X) \equiv \langle I(p) \rangle = \sum_{i=1}^{n} p_i I(p_i) = k \sum_{j=1}^{n} p_i \log p_i.$$

## 11.3

$$\begin{aligned}
\frac{\mathrm{d}H_{\text{bin}}}{\mathrm{d}p} &= -\frac{\mathrm{d}}{\mathrm{d}p}\left[p \log(p)\right] - \frac{\mathrm{d}}{\mathrm{d}p}\left[(1-p)\log(1-p)\right] \\
&= -\log(p) - 1 + \log(1-p) + 1 \\
&= \log(1-p) - \log(p).
\end{aligned}$$

Considering $p \in [0, 1]$, $\mathrm{d}H_{\text{bin}}/\mathrm{d}p = 0$ only for $p = 1/2$. The second derivative yields

$$\frac{\mathrm{d}^2 H_{\text{bin}}}{\mathrm{d}p^2} = -\frac{1}{1-p} - \frac{1}{p} = -\frac{1}{p(1-p)} \quad \Longrightarrow \quad \frac{\mathrm{d}^2 H_{\text{bin}}}{\mathrm{d}p^2} < 0 \quad \forall\, p \in (0, 1),$$

corresponding to constant negative concavity, so $p = 1/2$ corresponds to a maximum.

## 11.4

The second derivative of the binary entropy is negative for all $p \in (0, 1)$ (see Exercise 11.3). It follows that for all, $p, x_1, x_2 \in [0, 1]$, it holds

$$H_{\text{bin}}(px_1 + (1-p)x_2) \geq pH_{\text{bin}}(x_1) + (1-p)H_{\text{bin}}(x_2),$$

with the innequality being strict for all values $p \in (0, 1)$. Since $H_{\text{bin}}(0) = H_{\text{bin}}(1) \equiv 0$, we get equality in the trivial cases: $x_1 = x_2$, or $p = 0$, or $p = 1$.

## 11.5

$$\begin{aligned}
H(p(x,y)||p(x)p(y)) &= \sum_{x,y} p(x,y) \log\left[\frac{p(x,y)}{p(x)p(y)}\right] \\
&= \sum_{x,y} p(x,y) \log[p(x,y)] - \sum_{x,y} p(x,y) \log[p(x)p(y)] \\
&= \sum_{x,y} p(x,y) \log[p(x,y)] - \sum_{x} p(x) \log[p(x)] - \sum_{y} p(y) \log[p(y)] \\
&= H(p(x)) + H(p(y)) - H(p(x,y)).
\end{aligned}$$

Considering that $x$ are the possible outcomes of the random variable $X$, and $y$ the possible outcomes of the random variable $Y$, we have $H(p(x)) \equiv H(X)$, $H(p(y)) \equiv H(Y)$, and $H(p(x,y)) \equiv H(X,Y)$. From the non-negativity of the relative entropy, we get $H(X,Y) \leq H(X) + H(Y)$. If we consider equality, then $H(p(x,y)||p(x)p(y)) = 0$, which means $p(x,y) = p(x)p(y)$, hence $X$ and $Y$ are independent random variables. The converse is immediate.

## 11.6

Let us use the probability distribution given by $p(x|y)p(z|y)p(y)$. The relative entropy between this distribution and $p(x,y,z)$ will be

$$H(p(x,y,z)||p(x|y)p(z|y)p(y)) = \sum_{x,y,z} p(x,y,z) \log\left[\frac{p(x,y,z)}{p(x|y)p(z|y)p(y)}\right].$$

From Bayes' rule we can write $p(x|y) = p(x,y)/p(y)$, and $p(z|y) = p(y,z)/p(y)$, hence

$$\begin{aligned} H(p(x,y,z)||p(x|y)p(z|y)p(y)) &= \sum_{x,y,z} p(x,y,z) \log\left[\frac{p(x,y,z)p(y)}{p(x,y)p(y,z)}\right] \\ &= \sum_{x,y,z} p(x,y,z) \log[p(x,y,z)] + \sum_{y} p(y) \log[p(y)] \\ &\quad - \sum_{x,y} p(x,y) \log[p(x,y)] - \sum_{y,z} p(y,z) \log[p(y,z)] \\ &= -H(X,Y,Z) - H(Y) + H(X,Y) + H(Y,Z). \end{aligned}$$

Using the fact that $H(p(x,y,z)||p(x|y)p(z|y)p(y)) \geq 0$ we get strong subadditivity

$$H(X,Y,Z) + H(Y) \leq H(X,Y) + H(Y,Z).$$

If we consider equality, then $H(p(x,y,z)||p(x|y)p(z|y)p(y)) = 0$, which means (see Exercise A1.1)

$$p(x,y,z) = p(x|y)p(z|y)p(y) = p(z)p(y|z)p(x|y),$$

hence $Z \to Y \to X$ forms a Markov chain. The converse is immediate.

## 11.7

Let $n_Y$ be the number of possible outcomes for the variable $Y$. The uniform distribution $u(y)$ over $Y$ is then given by $u(y) = 1/n_Y$ for all $y$. The relative entropy $H(p(x,y)||p(x)u(y))$ yields

$$\begin{aligned} H(p(x,y)||p(x)u(y)) &= \sum_{x,y} p(x,y) \log\left[\frac{p(x,y)}{p(x)\frac{1}{n_Y}}\right] \\ &= \sum_{x,y} p(x,y) \log[p(x,y)] - \sum_{x} p(x) \log[p(x)] - \log\tfrac{1}{n_Y} \\ &= -H(X,Y) + H(X) + \log n_Y. \end{aligned}$$

By definition, we have $H(Y|X) = H(X,Y) - H(X)$, thus

$$H(Y|X) = \log n_Y - H(p(x,y)||p(x)u(y)).$$

Notice that $\log n_Y$ corresponds to the Shannon entropy of $Y$ when it is completely independent of $X$ and its probability distribution is the uniform distribution $u(y)$, which is where it has its maximum value, thus $0 \leq H(p(x,y)||p(x)u(y)) \leq \log n_Y$ and therefore $H(Y|X) \geq 0$. Equality will hold when $p(x,y)$ is the farthest possible from $p(x)u(y)$, which is when $Y$ is given by a deterministic function $f$ of $X$, that is, $p(x,y) = p(x)\delta(y - f(x))$. To verify that, we can explicitly calculate

$$H(p(x)\delta(y-f(x))||p(x)u(y)) = \sum_{x,y} p(x)\delta(y-f(x)) \log\left[\frac{\delta(y-f(x))}{\frac{1}{n_Y}}\right] = \sum_x p(x)\log n_Y = \log n_Y$$
$$\implies \quad H(Y|X) = 0.$$

## 11.8

There are only four possible joint outcomes, which are given by: $(x=0, y=0, z=0)$, $(x=1, y=0, z=1)$, $(x=0, y=1, z=1)$, and $(x=1, y=1, z=0)$, all with equal probability of $1/4$. So we can calculate the entropies

$$H(X) = H(Y) = H(Z) = 2\left(-\frac{1}{2}\log\frac{1}{2}\right) = 1,$$
$$H(X,Y) = H(X,Z) = H(Y,Z) = H(X,Y,Z) = 4\left(-\frac{1}{4}\log\frac{1}{4}\right) = 2.$$

By definition, we obtain

$$H(X,Y:Z) = H(X,Y) + H(Z) - H(X,Y,Z) = 1,$$
$$H(X:Z) = H(X) + H(Z) - H(X,Z) = 0,$$
$$H(Y:Z) = H(Y) + H(Z) - H(Y,Z) = 0,$$

and thus conclude that $H(X,Y:Z) \nleq H(X:Z) + H(Y:Z)$.

## 11.9

There are only two possible joint outcomes, which are given by: $(x_1 = x_2 = y_1 = y_2 = 0)$, and $(x_1 = x_2 = y_1 = y_2 = 1)$, all with equal probability of $1/2$. So we calculate the entropies

$$H(X_i) = H(Y_i) = H(X_i,Y_i) = H(X_1,X_2) = H(Y_1,Y_2) = H(X_1,X_2,Y_1,Y_2) = 2\left(-\frac{1}{2}\log\frac{1}{2}\right) = 1,$$

for $i \in \{1,2\}$. By definition, we obtain

$$H(X_1:Y_1) = H(X_1) + H(Y_1) - H(X_1,Y_1) = 1,$$
$$H(X_2:Y_2) = H(X_2) + H(Y_2) - H(X_2,Y_2) = 1,$$

$$H(X_1, X_2 : Y_1, Y_2) = H(X_1, X_2) + H(Y_1, Y_2) - H(X_1, X_2, Y_1, Y_2) = 1,$$

and thus conclude that $H(X_1 : Y_1) + H(X_2 : Y_2) \not\leq H(X_1, X_2 : Y_1, Y_2)$.

## 11.10

If $X \to Y \to Z$ is a Markov chain, then the probability of $Y = y$ is conditioned on $X = x$, and the probability of $Z = z$ is conditioned on $Y = y$. In practice, this means that the probability of the joint result $(X, Y, Z) = (x, y, z)$ is given by $p(x)p(y|x)p(z|y)$. But from Bayes' rule (see Exercise A1.1)

$$p(x)p(y|x)p(z|y) = p(x|y)p(y)p(z|y) = p(x|y)p(y|z)p(z),$$

which is the probability of $(X, Y, Z) = (x, y, z)$ when $Z \to Y \to X$ is a Markov chain.

## 11.11

$$\rho = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}: \quad S(\rho) = -1 \times \log 1 - 0 \times \log 0 = 0.$$

$$\rho = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}: \quad \text{this is } \rho = |+\rangle\langle+| \text{ in the } X \text{ basis, so}$$

$$S(\rho) = -1 \times \log 1 - 0 \times \log 0 = 0.$$

$$\rho = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}: \quad \det(\rho - \lambda I) = \lambda^2 - \lambda + \frac{1}{9} = 0 \quad \Longrightarrow \quad \text{eigenvalues} = \left\{\frac{3-\sqrt{5}}{6}, \frac{3+\sqrt{5}}{6}\right\}, \text{ so}$$

$$S(\rho) = -\left(\frac{3-\sqrt{5}}{6}\right)\log\left(\frac{3-\sqrt{5}}{6}\right) - \left(\frac{3+\sqrt{5}}{6}\right)\log\left(\frac{3+\sqrt{5}}{6}\right) \approx 0.55.$$

## 11.12

$$\rho = p\,|0\rangle\langle 0| + (1-p)\,|+\rangle\langle+| = p\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \frac{1-p}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} 1+p & 1-p \\ 1-p & 1-p \end{bmatrix}.$$

$$\det(\rho - \lambda I) = \lambda^2 - \lambda + \frac{p(1-p)}{2} = 0$$

$$\Longrightarrow \quad \text{eigenvalues} = \left\{\frac{1-\sqrt{1-2p(1-p)}}{2}, \frac{1+\sqrt{1-2p(1-p)}}{2}\right\}.$$

If we define $q \equiv \frac{1-\sqrt{1-2p(1-p)}}{2}$, these eigenvalues can be rewritten as $\{q, 1-q\}$, so the von Neumann entropy will be

$$S(\rho) = -q\log(q) - (1-q)\log(1-q).$$

The Shannon entropy $H(p, 1-p)$ will be

$$H(p, 1-p) = -p\log(p) - (1-p)\log(1-p).$$

## 11.13

-

## 11.14

-

## 11.15

-

## 11.16

-

## 11.17

-

## 11.18

-

## 11.19

-

## 11.20

-

## 11.21

-

### 11.22

-

### 11.23

-

### 11.24

-

### 11.25

-

### 11.26

-