GAUSSIAN PROCESSES
EXERCISE SHEET 3: DISCRETE ENTROPY

Exercise 1.

Let X be geometric with parameter $p \in (0,1)$ on $\{1, 2, \dots\}$, i.e.

$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \qquad k \geq 1.$$

Write $q := 1 - p \in (0,1)$. By definition

$$H(X) = -\sum_{k=1}^{\infty} pq^{k-1} \log_2\big(pq^{k-1}\big).$$

Split the log and the sum:

$$H(X) = -\log_2 p \sum_{k=1}^{\infty} pq^{k-1} - \log_2 q \sum_{k=1}^{\infty} p(k-1)q^{k-1}.$$

The first sum equals 1. For the second sum note

$$\sum_{k=1}^{\infty} (k-1)q^{k-1} = \sum_{m=0}^{\infty} mq^m = \frac{q}{(1-q)^2} = \frac{q}{p^2},$$

so multiplying by the leading p gives $\frac{q}{p}$. Hence

$$H(X) = -\log_2 p - \frac{q}{p} \log_2 q.$$

Equivalently (multiplying numerator and denominator by p),

$$\boxed{H(X) = \frac{-p\log_2 p - q\log_2 q}{p} \;=\; -\log_2 p - \frac{1-p}{p}\log_2(1-p).}$$

This is the entropy of the geometric distribution. □

Exercise 2.

Let X, Y be discrete random variables with joint law supported on a countable set; denote joint probabilities by $p(\omega_x, \omega_y)$ etc.

(1) For any outcome ω_x we have $p(\omega_x) \in (0,1]$, so $-\log_2 p(\omega_x) \geq 0$. Taking expectation yields $H(X) \geq 0$.

If X and Y are independent then $p(\omega_x,\omega_y)=p(\omega_x)p(\omega_y)$, so

$$\begin{aligned} H(X,Y) &= -\sum_{\omega_x,\omega_y} p(\omega_x)p(\omega_y)\log_2\big(p(\omega_x)p(\omega_y)\big) \\ &= -\sum_{\omega_x,\omega_y} p(\omega_x)p(\omega_y)\log_2 p(\omega_x) - \sum_{\omega_x,\omega_y} p(\omega_x)p(\omega_y)\log_2 p(\omega_y) \\ &= -\sum_{\omega_x} p(\omega_x)\log_2 p(\omega_x)\cdot\sum_{\omega_y} p(\omega_y) \; - \; \sum_{\omega_y} p(\omega_y)\log_2 p(\omega_y)\cdot\sum_{\omega_x} p(\omega_x) \\ &= H(X)+H(Y). \end{aligned}$$

(2) Let $p(\omega_y \mid \omega_x)$ denote the conditional probability. Then

$$p(\omega_x,\omega_y)=p(\omega_x)p(\omega_y \mid \omega_x),$$

so

$$\begin{aligned} H(X,Y) &= -\sum_{\omega_x,\omega_y} p(\omega_x,\omega_y)\log_2 p(\omega_x,\omega_y) \\ &= -\sum_{\omega_x,\omega_y} p(\omega_x,\omega_y)\big(\log_2 p(\omega_x)+\log_2 p(\omega_y \mid \omega_x)\big) \\ &= -\sum_{\omega_x} p(\omega_x)\log_2 p(\omega_x)\sum_{\omega_y} p(\omega_y \mid \omega_x) \; - \; \sum_{\omega_x} p(\omega_x)\sum_{\omega_y} p(\omega_y \mid \omega_x)\log_2 p(\omega_y \mid \omega_x) \\ &= H(X)+\sum_{\omega_x} p(\omega_x)\Big(-\sum_{\omega_y} p(\omega_y \mid \omega_x)\log_2 p(\omega_y \mid \omega_x)\Big) \\ &= H(X)+H(Y \mid X). \end{aligned}$$

(3) From (2) we know that

$$H(X,Y)=H(X)+H(Y \mid X).$$

Thus to show $H(X,Y)\le H(X)+H(Y)$ it is enough to show $H(Y \mid X)\le H(Y)$. Write

$$H(Y \mid X)=\sum_{\omega_x} p(\omega_x)H\big(Y \mid X=\omega_x\big)=\sum_{\omega_x} p(\omega_x)\Big(-\sum_{\omega_y} p(\omega_y \mid \omega_x)\log_2 p(\omega_y \mid \omega_x)\Big).$$

Fix ω_y and consider the function $g(u)=-u\log_2 u$, which is strictly concave on $[0,1]$. By Jensen's inequality (applied to the convex combination with weights $p(\omega_x)$) we have for each ω_y,

$$\sum_{\omega_x} p(\omega_x)\big(-p(\omega_y \mid \omega_x)\log_2 p(\omega_y \mid \omega_x)\big) \le -\Big(\sum_{\omega_x} p(\omega_x)p(\omega_y \mid \omega_x)\Big)\log_2\Big(\sum_{\omega_x} p(\omega_x)p(\omega_y \mid \omega_x)\Big).$$

But $\sum_{\omega_x} p(\omega_x)p(\omega_y \mid \omega_x)=p(\omega_y)$. Summing this inequality over ω_y yields $H(Y \mid X)\le H(Y)$. Therefore

$$H(X,Y)=H(X)+H(Y \mid X)\le H(X)+H(Y).$$

Moreover, strict concavity of $-u\log_2 u$ implies equality holds in the Jensen step iff for every fixed ω_y the conditional probability $p(\omega_y \mid \omega_x)$ is almost surely constant in ω_x. That means $p(\omega_y \mid \omega_x)=p(\omega_y)$ for all ω_x,ω_y, i.e. X and Y are independent. Conversely, if X and Y are independent then equality holds as shown in part (1). Hence equality holds iff X and Y are independent. □