26.04.2016

Remark (Final remark). *Entropy, relative entropy and so on can also be considered for RVs with (infinite???) values in countable sets, or for probability "vectors" $p = (p_1, p_2, p_3, \dots)$, but one has to require absolute convergence of*

$$H(p_1, p_2, \dots) = \sum_{n=1}^{\infty} p_n(-\log_2 p_n)$$

Same for

$$D(p||q) = \sum_{n=1}^{\infty} p_n \log_2 \frac{p_n}{q_n}$$

Example. *If $\sum_{n=1}^{\infty} np_n < \infty$, then $H(p) < \infty$. Use:*

- $x \mapsto -x\log_2 x$ *increasing on* $[0, \frac{1}{e}]$. add pic
- $-p_n \log_2 p_n \leq \max\{n, -\log_2 p_n\}p_n$

3) Asymptotic entropy

Definition. *(a) A* stochastic process *in discrete time is a sequence $(X_n)_{n\geq 1}$ of random variables.*

(b) Suppose that the X_n are discrete and take values in a finite (or countable) set $\mathcal{X}$. Then the asymptotic entropy *(entropy rate) of the stochastic process is*

$$h = \lim_{n\to\infty} \frac{H(X_1, \dots, X_n)}{n},$$

if the limit exists.

Question: For which types of stochastic processes does the limit exist? For which types of stochastic processes is there a formula for h?

Simplest case: $(X_n)_{n\in\mathbb{N}}$ are iid

$$H(X_1, \dots, X_n) = H(X_1) + \dots + H(X_n) = nH(X_1)$$
$$h = H(X_1)$$

Recall chain rule:

$$H(X_1, \dots, X_n) = \sum_{k=1}^{n} H(X_k|X_{k-1}, \dots, X_1)$$

And recall from Analysis 1: If (a_n) is a sequence of real numbers such that $\lim_{n\to\infty} a_n = a$ exists, then $\lim_{n\to\infty} \frac{1}{n}(a_1 + \dots + a_n)$ exists and

$$\lim_{n\to\infty} \frac{1}{n}(a_1 + \dots + a_n) = a$$

Consequence: If $h' = \lim_{n\to\infty} H(X_n|H_{n-1}, \dots, X_1)$ exists, then h exists and $h = h'$.

Definition. *The process* $(X_n)_{n\geq 1}$ *is called* stationary*, if for all* $n, l \in \mathbb{N}$*:*

$$p_{X_1,\dots,X_n} = p_{X_{l+1},\dots,X_{l+n}}$$

Recall: for all $x_1, \dots, x_n \in \mathcal{X} : \mathbb{P}[X_1 = x_1, \dots, X_n = x_n] = \mathbb{P}[X_{l+1} = x_{l+1}, \dots, X_{l+n} = x_{l+n}]$

Lemma 0.1. *If* (X_n) *is stationary, then* $h'(= h)$ *exists.*

Proof.

$$C \leq H(X_n|X_{n-1}, \dots, H_2, H_1) \leq H(X_n|X_{n-1}, \dots, H_2) = H(X_{n-1}|H_{n-2}, \dots, H_1)$$

□

Definition. *An* $\mathcal{X}$*-valued stochastic process* $(X_n)_{n\in\mathbb{N}_0}$ *is called a (*time-homogeneous*)* Markov chain *if for all* $x_0, x_1, \dots, x_{n-1}, x, y \in \mathcal{X}$

$$\begin{aligned}\mathbb{P}[X_{n+1} = y|X_n = x, X_{n--1} = x_{n-1}, \dots, X_0 = x_0] &= \mathbb{P}[X_{n+1} = y|X_n = x] \\ &= p(y|x) = p_{x,y}\end{aligned}$$

whenever $\mathbb{P}[X_{n+1} = y|X_n = x, X_{n--1} = x_{n-1}, \dots, X_0 = x_0] > 0$*. We call* $p(y|x)$ *transition probability form* x *to* y*. Transition matrix:*

$$P = (p(y|x))_{x,y\in\mathcal{X}}$$

where x *denotes the row and* y *denotes the column. So we can also write* $P_{x,y}$*.*

Remark. *Avoid in future courses:* $p(x, y)$*!*

Consider $\mathcal{X}, P)$ as a directed graph. The vertex set is $\mathcal{X}$ and the edges are $x \overset{p(y|x)>0}{\rightarrow} y$

Example (Weather in the land OZ)**.** *Weather evolves as follows: there are 3 types: bright (b), rainy (r), snowy (s) The rules are:*

- *If one day is bright, then the next day it will always turn bad, r and s with equal probability*
- *If one day is bad (r or s), in 50% of all cases, it will stay as it is, and only in 25% of cases it will turn bright.*

$\mathcal{X} = \{b, r, s\}$ add pic

$$P = \begin{array}{c c} & \begin{array}{ccc} b & r & s \end{array} \\ \begin{array}{c} b \\ r \\ s \end{array} & \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix} \end{array}$$

We also need the *initial distribution*

$$\mu(x) = \mu_x = \mathbb{P}[X_0 = x]$$

$\nu(x) = \nu_x$ (special)

$\mathbb{P}_\mu$ is the distribution, when initial distribution is μ.

$$\begin{aligned}
\mathbb{P}_\mu[X_0 = x_0, &X_1 = x_1, \dots, X_n = x_n] \\
&= \mathbb{P}_\mu[X_0 = x_0, \dots, X_{n-1} = x_{n-1}]\, \mathbb{P}[X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0] \\
&= \underbrace{\mathbb{P}[X_0 = x_0]}_{\mu_{X_0}} \underbrace{p_{x_0,x_1}}_{=p(x_1|x_0)} \cdots \underbrace{p_{x_{n-1},x_n}}_{p(x_n|x_{n-1})}
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}[X_2 = y | X_0 = x] &= \sum_{w \in \mathcal{X}} \mathbb{P}[X_2 = y, X_1 = w | X_0 = x] \\
&= \sum_{w \in \mathcal{X}} \underbrace{\mathbb{P}[X_1 = w | X_0 = x]}_{p_{x,w}} \underbrace{\mathbb{P}[X_2 = y | X_1 = w, X_0 = x]}_{p_{w,y}} \\
&= \sum_{w \in \mathcal{X}} p_{x,w} p_{w,y}
\end{aligned}$$

Lemma 0.2. $\mathbb{P}[X_{l+n} = y | X_l = x] = p^{(n)}_{x,y} = p^{(n)}(y|x)$ *where* $P^n = (p^{(n)}_{x,y})_{x,y \in \mathcal{X}}$
The (unconditional) distribution of X_n *is*

$$\mu P^n$$

(row vectors matrix)

$$\begin{aligned}
\mathbb{P}_\mu[X_n = y] &= \sum_{x \in \mathcal{X}} \mathbb{P}_\mu[X_n = y, X_0 = x] \\
&= \sum_{x \in \mathcal{X}} \mathbb{P}_\mu[X_0 = x]\, \mathbb{P}_\mu[X_n = y | X_0 = x] \\
&= \sum_{x \in \mathcal{X}} \mu_x p^{(n)}_{x,y}
\end{aligned}$$