Let $(X_n)_{n\in\mathbb{N}}$ be $\mathcal{X}$-valued random variables
as entropy: $h = \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, H_n)$, if the limit exists.

- If X_n are iid, then $h = H(X_1)$.
- If (X_n) is stationary $[p_{X_1,\ldots,X_n} = p_{X_{k+1},\ldots,X_{k+n}}]$, then the sequence $H(X_n|X_{n-1}, \ldots, X_1)$ is decreasing and has a limit $[h']$, and (chain rule): $h = h'$
- (time-homogeneous) Markov chains: transition matrix $P = (p(y|x))_{x,y\in\mathcal{X}}$

$$\mathbb{P}[X_{n+1} = y | X_n = x,\ \text{past}] = p_(y|x)$$

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

$$\mathbb{P}_\mu[X_n = y] = \sum_{x\in\mathcal{X}} \mu(x) p^{(n)}(y|x)$$

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

Distribution of X_n is μP^n.

When stationary? necessary: $\mu P = \mu$, so also $\mu P^l = \mu$ for all l.

sufficient?

$$\begin{aligned}\mathbb{P}_\mu[X_{l+1} = x_1, \ldots, X_{l+n} = x_n] &= \mathbb{P}_\mu[X_{l+1} = x_1] p(x_2|x_1) \cdots p(x_n|x_{n-1}) \\ &= \mu(x_1) p(x_2|x_1) \cdots p(x_n|x_{n-1})\end{aligned}$$

does not depend on l!

Notation change: ν will be stationary distribution, $\nu = (\nu(x))_{x\in\mathcal{X}}$ probability vector with $\nu P = \nu$

Lemma 0.1. *If $\mathcal{X}$ is finite, then stationary (invariant) distribution(s) exist.*

Proof. Take any initial distribution μ. Then

$$\mu_n = \frac{1}{n}(\mu + \mu P + \mu P^2 + \cdots + \mu P^{n-1})$$

is a probability vector. There is a subsequence (μ_{n_k}) which converges to a vector ν.

$$\begin{aligned}\mu_{n_k}(y) &\to \nu(y) \ \ \forall y \in \mathcal{X} \\ \mu_{n_k} &\to \nu \\ \mu_k P &\to \nu P\end{aligned}$$

$$\begin{aligned}\mu_n P - \mu_n &= \frac{1}{n}((\mu P + \mu P^2 + \cdots + \mu P^n) - (\mu + \mu P + \cdots + \mu P^{n-1})) \\ &= \frac{1}{n}(\mu P^n - \mu) \to 0 = (0, \ldots, 0)\end{aligned}$$

$\mu_{n_k} P - \mu_{n_k}$ converges to 0 and it converges to $\nu P - \nu$. Thus, $\nu P = \nu$. $\square$

Corollary 0.2. *If the Markov chain $(X_n)_{n\geq 0}$ starts with ν like above $h = h' = \lim_{n\to\infty} H(X_n|X_{n-1}, \ldots, X_0)$ exists.*

Think about $H(X_n|X_{n-1},\dots,X_0)$ as X_n is the future, X_{n-1} is the present and $X_{n-2},\dots,X_0$ is the past.

$$\begin{aligned}H(X_n|X_{n-1},\dots,X_0) &= H(X_n|X_{n-1}) = H(X_1|X_0)\\ &= \sum_x \mathbb{P}[X_0 = x]H(X_1|X_0 = x)\\ &= \sum_x \nu(x)H(p(\cdot|x))\end{aligned}$$

Then

$$h = \sum_x \nu(x)H(p(\cdot|x))$$

$$H(X_n|X_{n-1},\dots,X_0) = \frac{1}{n}\sum_{k=0}^{n} H(X_k|X_{k-1},\dots,X_0)-$$

Suppose MC starts with any μ

$$\begin{aligned}-\frac{1}{n}\sum_{k=0}^{n-1} H(X_k|X_{k-1}) &= \frac{1}{n}\sum_{k=0}^{n-1}\sum_x \mu P^{k-1}(x)H(p(x))\\ &= \sum_x \overbrace{(\underbrace{\frac{1}{n}\sum_{k=0}^{n-1}\mu P^k(x))}_{\to\nu(x)}}^{\mu_n(x)} H(p(\cdot|x))\end{aligned}$$

$$\begin{aligned}H(X_k|X_{k-1}) &= \sum_x \mathbb{P}[X_{k-1} = x]H(p(\cdot|x))\\ &= \sum_x \mu P^{k-1}(x)H(p(\cdot|x))\end{aligned}$$

Remark. *Suppose μ_n converges not ν, then (μ_n) has some other accumulation point $\mu_{n_l} \to \tilde{\nu} \implies \tilde{\nu} = \nu\lightning$.*

Theorem 0.3. *Suppose there is a* unique *stationary probability distribution ν, then $\mu_n \to \nu$, and*

$$\frac{1}{n}H(X_0,\dots,X_n) \to h = \sum_x \nu(x)H(p(\cdot|x))$$

for any initial μ.

Definition. *$(\mathcal{X}, P)$ is called irreducible if the associated directed graph is strongly connected. Vertex set $\mathcal{X}$, edges are between all x, y with $p(y|x) > 0$.*

Equivalently: For all x, y there exists an $n = n_{x,y}$ such that $p^{(n)}(y|x) > 0$.

Theorem 0.4. *If P is irreducible, then it has a* unique *stationary probability distribution [$\nu P = \nu$]*

Proof blabla: $(\mathcal{X}_1, P_1), (\mathcal{X}_2, P_2), \mathcal{X} = \mathcal{X}_1 \uplus \mathcal{X}_2$ Then

$$P = \begin{array}{c c} & \begin{array}{cc} \mathcal{X}_1 & \mathcal{X}_2 \end{array} \\ \begin{array}{c} \mathcal{X}_1 \\ \mathcal{X}_2 \end{array} & \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix} \end{array}$$

Example. *Simple random walk on a finite, connected graph. Let $(\mathcal{X}, E)$ be a non-directed graph with no multiple edges and no loops. Let $x \sim y$ denote neighbours if there exists an edge between x and y. Then*

$$\deg(x) = |\{y : y \sim x\}|$$

add pic

$$p(y|x) = \begin{cases} \frac{1}{\deg(x)} & \textit{if } y \sim x \\ 0 & \textit{if } y \nsim x \end{cases}$$

$$\begin{aligned} \deg(x)p(y|x) &= \begin{cases} 1 & y \sim x \\ 0 & y \nsim x \end{cases} \\ &= \deg(y)p(x|y) \end{aligned}$$

$$\deg(x) = \sum_y \deg(y)$$

Consider the row vector deg*:*

$$\deg = \deg \cdot P$$

$$\nu(x) = \frac{\deg(x)}{\sum_w \deg(x)} = \frac{\deg(x)}{2|E|}$$

$$h = \sum_x \frac{\deg(x)}{2|E|} \log_2 \deg(x)$$

Proof. First, look at solutions of $Pf = f$ where $f : \mathcal{X} \to \mathbb{R}$ seen as column vectors.

$$Pf(x) = \sum_y p_{x,y} f(y)$$

Step 1: $P\mathbb{1} = \mathbb{1}$, and if $Pf = f$ then $f \equiv$ constant $\equiv c\mathbb{1}$ Because: Let x_0 be such that $f(x_0) = M = \max f$

$$\begin{aligned} P^n f &= f : \\ \sum_y p^{(n)}_{x_0,y} f(x_0) f(x_0) &= \sum_y p^{(n)}_{x_0,y}(x) f(y) \\ \sum_y \underbrace{p^{(n)}_{x_0,y}}_{\geq 0} \underbrace{(f(x_0) - f(y))}_{\geq 0} &= 0 \end{aligned}$$

So for all n, y:

$$p_{x_0,y}^{(n)}(f(x_0) - f(y)) = 0$$

Let $y \in \mathcal{X}$. Then by irreducibility, there exists an n such that

$$p_{x_0,y}^{(n)} > 0 \implies f(y) = f(x_0)$$

Step 2: Let $\nu P = \nu$, where ν is a probability vector. $\nu P^n = \nu$,

$$\nu(y) = \sum_x \nu(x) p_{x,y}^{(n)} \geq \nu(x_0) p_{x_0,y}^{(n)} > 0$$

Then there exists a x_0 such that $\nu(x_0) > 0$. By irreducibility, there exists an n such that

$$p_{x_0,y}^{(n)} > 0$$

So $\nu(y) > 0$ for all y.

$$\nu(y) = \sum_x \nu(x) p_{x,y}$$

$$\sum_x \underbrace{\frac{\nu(x) p_{x,y}}{\nu(y)}}_{\hat{p}_{y,x} = \frac{\nu(x) p_{x,<}}{\nu(y)}} = 1$$

$$\hat{p}_{x,y} = \frac{\nu(y) p_{y,x}}{\nu(x)} \to \hat{p}$$

stochastic irreducible

Step 3: Suppose $\mu P = \mu$. Then define $f(x) \coloneqq \frac{\mu(x)}{\nu(x)}$.

$$\begin{aligned}
\hat{P} f(x) = \sum_y \hat{p}_{x,y} f(y) &= \sum_y \frac{\nu(x) p_{y,x}}{\nu(x)} \frac{\mu(y)}{\nu(y)} \\
&= \frac{1}{\nu(x)} \underbrace{\sum_y \mu(y) p_{y,x}}_{\mu P(x) = \mu(x)} \\
&= \frac{\mu(x)}{\nu(x)} = f(x)
\end{aligned}$$

Then by Step 1 follows that $f(x) = c$ for all x.

$$\sum_x \mu(x) = c \cdot \nu(x) \ \ \forall x \implies c = 1, \mu = \nu$$

$\square$