1.68. *Coupon collector's problem.* We are interested now in the time it takes to collect a set of N baseball cards. Let T_k be the number of cards we have to buy before we have k that are distinct. Clearly, $T_1 = 1$. A little more thought reveals that if each time we get a card chosen at random from all N possibilities, then for $k \geq 1$, $T_{k+1} - T_k$ has a geometric distribution with success probability $(N-k)/N$. Use this to show that the mean time to collect a set of N baseball cards is $\approx N \log N$, while the variance is $\approx N^2 \sum_{k=1}^{\infty} 1/k^2$.

$X \sim \text{Geom}(p)$

$\mathbb{E}[X] = \frac{1}{p}$; $\text{Var}(X) = \frac{1-p}{p^2}$

$$\mathbb{E}[T_{k+1} - T_k] = \frac{1}{\frac{N-k}{N}} = \frac{N}{N-k}$$

$$\mathbb{E}[T_N] - \underbrace{\mathbb{E}[T_1]}_{1} = \mathbb{E}[T_N - T_{N-1}] + \mathbb{E}[T_{N-1} - T_{N-2}] + \mathbb{E}[T_{N-2} - T_{N-3}] + \dots + \mathbb{E}[T_2 - T_1]$$

$$\mathbb{E}[T_N] = 1 + \sum_{j=1}^{N-1} \frac{N}{N-j} = 1 + N \underbrace{\sum_{j=1}^{N-1} \frac{1}{j}}_{\sim \log N}.$$

T_k, $T_{k+1} - T_k$ are independent

$$\text{Var}(T_{k+1}) = \text{Var}(T_k) + \text{Var}(T_{k+1} - T_k)$$

Variance of Geom(p) is $\frac{1-p}{p^2}$

$$\mathrm{Var}(T_{k+1} - T_k) = \frac{1 - \frac{N-k}{N}}{\left(\frac{N-k}{N}\right)^2} = \frac{kN}{(N-k)^2}$$

telescopic sum

$$\mathrm{Var}(T_N) - \mathrm{Var}(T_1) = \mathrm{Var}(T_N - T_{N-1}) + \mathrm{Var}(T_{N-1} - T_{N-2}) + \dots + \mathrm{Var}(T_2 - T_1)$$

$$\mathrm{Var}(T_N) = \underbrace{\mathrm{Var}(T_1)}_{0} + \sum_{i=1}^{N-1} \frac{iN}{(N-i)^2} = \sum_{i=1}^{N} \frac{N(N-i)}{(N-(N-i))^2}$$

$$= \sum_{j=1}^{N} \frac{N(N-j)}{j^2} = \sum_{j=1}^{N} \left(\frac{N^2}{j^2} - \frac{Nj}{j^2}\right) = N^2 \sum_{j=1}^{N-1} \left(\frac{1}{j^2} - \frac{1}{jN}\right)$$

$$T_N - T_1 = (\underline{T_N} - T_{N-1}) + (\underline{T_{N-1}} - T_{N-2}) + \dots + (T_2 - \underline{T_1})$$

For every $\varepsilon > 0$ if N is large enough then $\sum_{j=1}^{N-1} \frac{1}{jN} < \varepsilon \cdot \sum_{j=1}^{N-1} \frac{1}{j^2}$. Namely,

$$\sum_{j=1}^{N-1} \frac{1}{jN} \approx \frac{1}{N} \cdot \sum_{j=1}^{N-1} \frac{1}{j} \sim \frac{1}{N} \log N \to 0.$$

1.69. *Algorthmic efficiency.* The simplex method minimizes linear functions by moving between extreme points of a polyhedral region so that each transition decreases the objective function. Suppose there are n extreme points and they are numbered in increasing order of their values. Consider the Markov chain in which $p(1,1) = 1$ and $p(i,j) = 1/(i-1)$ for $j < i$. In words, when we leave j we are equally likely to go to any of the extreme points with better value. (a) Use (1.25) to show that for $i > 1$

$$E_i T_1 = 1 + 1/2 + \cdots + 1/(i-1)$$

(b) Let $I_j = 1$ if the chain visits j on the way from n to 1. Show that for $j < n$

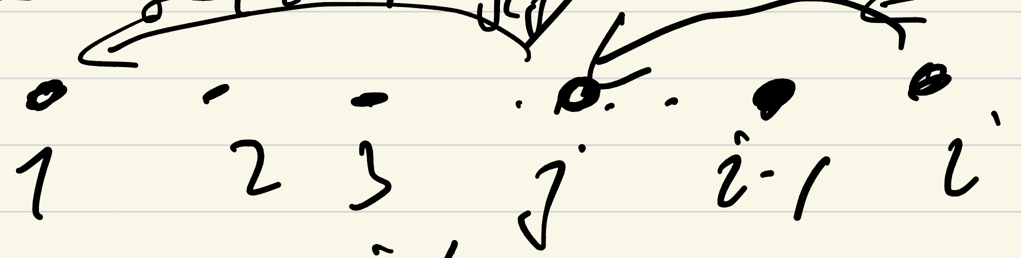
$$P(I_j = 1 | I_{j+1}, \dots I_n) = 1/j$$

to get another proof of the result and conclude that $I_1, \dots I_{n-1}$ are independent.

Solution

(a)

time to get ~~to~~ 1 $= g(i)$



with probab $\frac{1}{i-1}$

$$g(i) = 1 + \frac{1}{i-1} \cdot \sum_{j=1}^{i-1} g(j) \longrightarrow (i-1) g(i) = i - 1 + \sum_{j=1}^{i-1} g(j)$$

$$i\, g(i+1) = i + \sum_{j=1}^{i} g(j)$$

$$i\, g(i+1) - (i-1)\, g(i) = 1 + g(i) \Rightarrow g(i+1) = g(i) + \frac{1}{i}.$$

(b)

$$\mathbb{P}(I_j = 1 \mid I_{j+1} = i_{j+1}, \dots, I_{n-1} = i_{n-1}, I_n = 1)$$

Let $k := \min\{ l > j : i_l = 1 \}$.

$$\mathbb{P}(I_j = 1 \mid I_{j+1} = 0, \dots, I_{k-1} = 0, I_k = 1, I_{k+1} = i_{k+1}, \dots, I_n = i_n = 1\}.$$

$$= \mathbb{P}(I_j = 1 \mid I_{j+1} = 0, \dots, I_{k-1} = 0, I_k = 1) = \frac{1}{j}.$$

That is for all $(i_{j+1}, \dots, i_n)$ $\mathbb{P}(I_j = 1 \mid I_{j+1} = i_{j+1}, \dots, I_{n-1} = i_{n-1}, I_n = 1) = \frac{1}{j}$

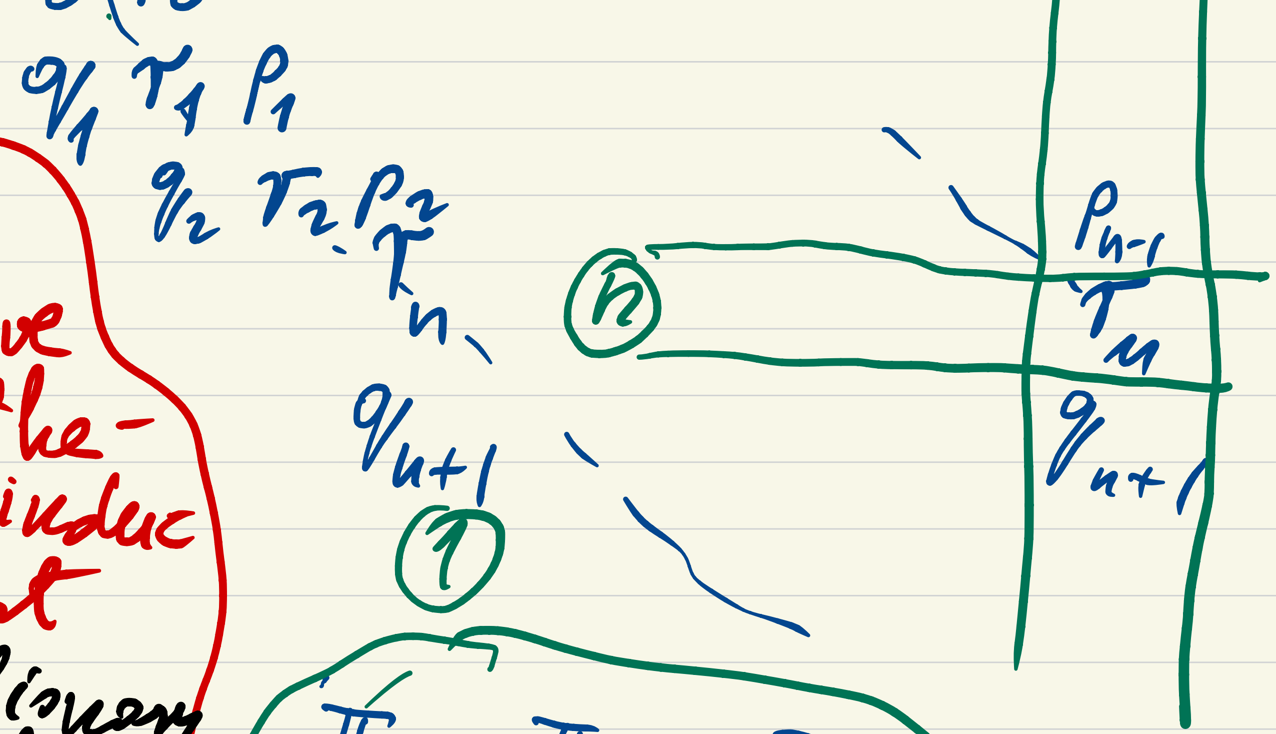
The Law of total probability:

$$\underline{\underline{\mathbb{P}(I_j = 1)}} = \sum_{i_{j+1} \dots i_{n-1}} \underbrace{\mathbb{P}(I_j = 1 \mid I_{j+1} = i_{j+1}, \dots, I_{n-1} = i_{n-1}, I_n = 1)}_{1/j} \cdot \mathbb{P}(\quad)$$

$$= \frac{1}{j} \sum_{i_{j+1} \dots i_{n-1}} \mathbb{P}(I_{j+1} = i_{j+1} \dots I_{n-1} = i_{n-1}, I_n = 1) = \underline{\underline{\frac{1}{j}}}.$$

$$\boxed{\pi_n = \pi_{n-1}\cdot p_{n-1} + \pi_n\cdot r_n + \pi_{n+1}\, q_{n+1}}$$

$$\pi_{n-1}\, p_{n-1} = \pi_n\, q_n$$



$$(p_n+q_n)\,\pi_n = \pi_{n-1}\, p_{n-1} + \pi_{n+1}\, q_{n+1}$$

2

$$\pi_1 = \underbrace{p_0\pi_0}_{\pi_1 q_1} + r_1\pi_1 + q_2\pi_2$$

$$(\underbrace{1-q_1-r_1}_{p_1})\,\pi_1 = q_2\pi_2$$

$$p_0\pi_0 = \pi_1 q_1$$

$$p_1\pi_1 = \pi_2 q_2$$

$$\vdots$$

$$\boxed{p_{n-1}\pi_{n-1} = \pi_n q_n}$$

assume that this holds

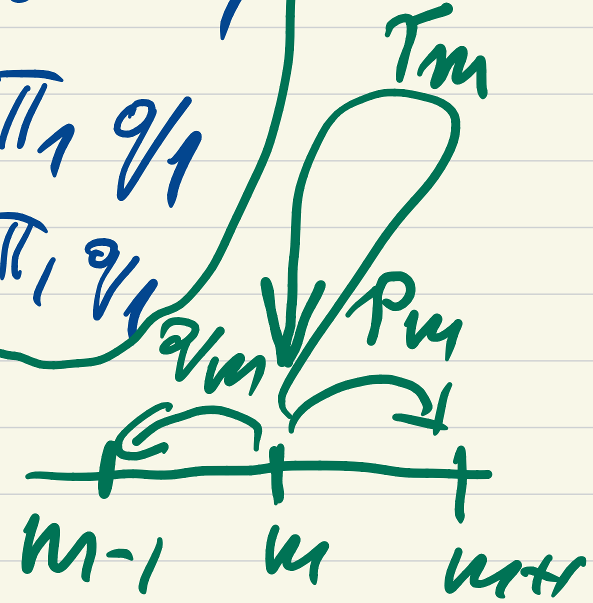
Here we prove by mathematical induction that if $\exists$ a stationary distr π for a death & birth chain then this π satisfies the detailed balance cond.

1

$$\pi_0 = \pi_0 r_0 + \pi_1 q_1$$

$$(1-r_0)\,\pi_0 = \pi_1 q_1$$

$$p_0\pi_0 = \pi_1 q_1$$



$$\pi_n = \underbrace{p_{n-1}\pi_{n-1}}_{} + r_n\pi_n + q_{n+1}\pi_{n+1}$$

by assumption

$$\pi_n q_n + r_n\pi_n = q_{n+1}\pi_{n+1}$$

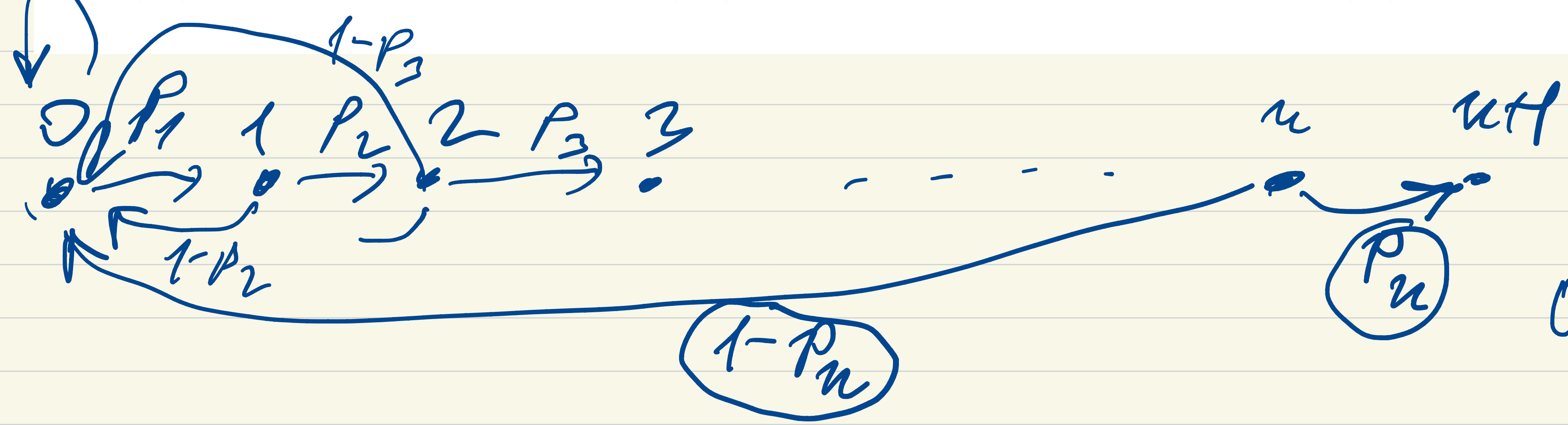
$$(\underbrace{1-q_n-r_n}_{p_n})\,\pi_n = q_{n+1}\pi_{n+1}$$

by mathematical induction

$$\boxed{p_n\pi_n = q_{n+1}\pi_{n+1}}$$

We are ready.

1.74. Consider the aging chain on $\{0, 1, 2, \ldots\}$ in which for any $n \geq 0$ the individual gets one day older from n to $n+1$ with probability p_n but dies and returns to age 0 with probability $1 - p_n$. Find conditions that guarantee that (a) 0 is recurrent, (b) positive recurrent. (c) Find the stationary distribution.

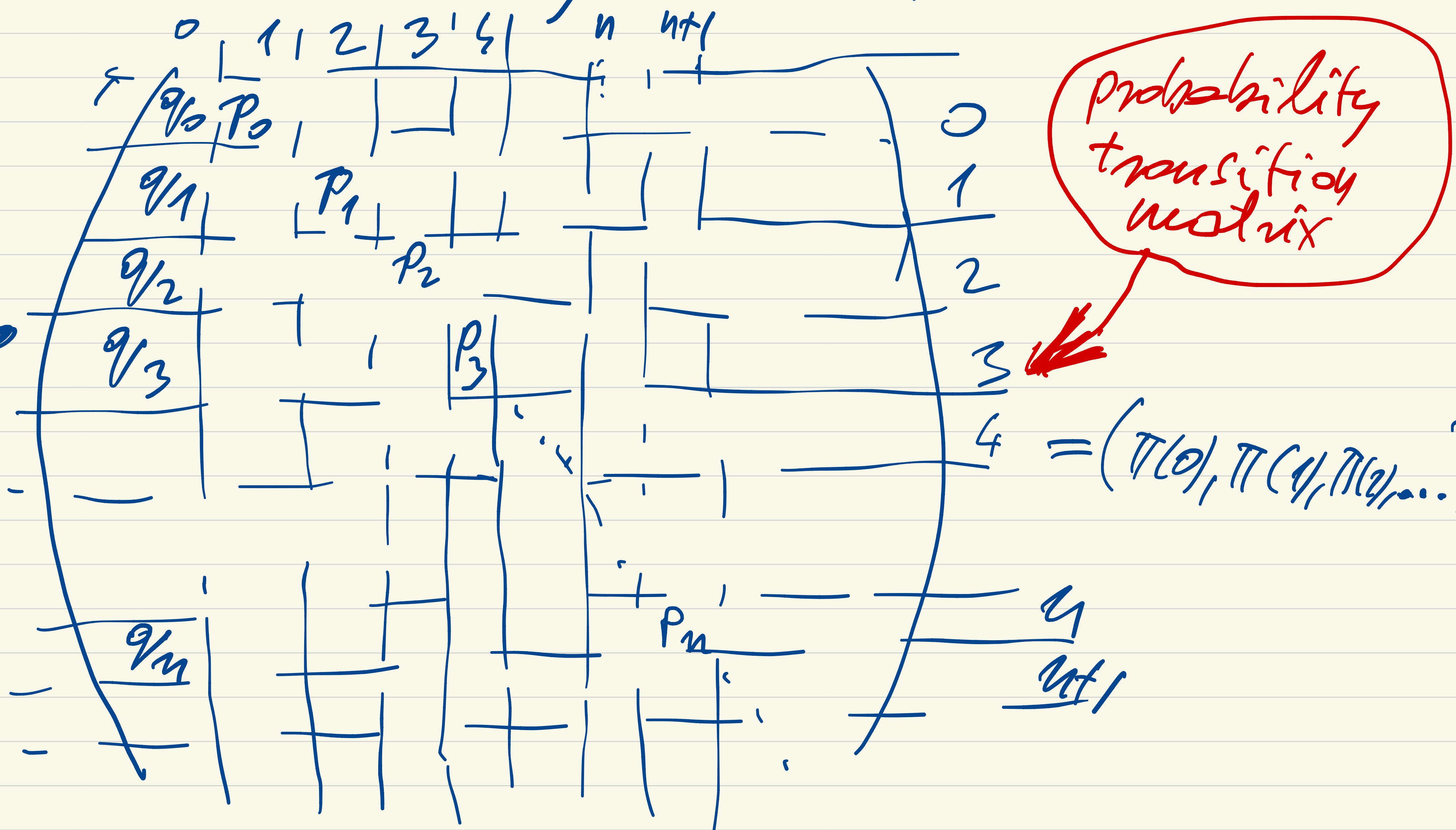


The chain is irreducible.
0 is recurrent $\Leftrightarrow \forall n$ recurrent.

Let $A_n := \{$the chain jumps from n to $n+1\}$. Then $p_n = \mathbb{P}(A_n)$.
Let T be the event that starting from 0 we never return to 0. Then $T = \bigcap_{n=0}^{\infty} A_n$. Clearly $\{A_n\}_{n=0}^{\infty}$ are independent from the Markov property. Hence $\mathbb{P}(T) = \prod_{n=0}^{\infty} \mathbb{P}(A_n) = \prod_{n=0}^{\infty} p_n$. By definition 0 is transient $\Leftrightarrow 0 < \mathbb{P}(T) = \prod_{n=0}^{\infty} p_n$. We know that this is equivalent to $\sum_{n=0}^{\infty} (1-p_n) < \infty$. So (a) has been answered.

(b) File A, Slide 60, Theorem 4.5 says: if $(X_n)_{n=1}^{\infty}$ is an irreducible Markov Chain & $\exists$ a stationary distr. π then $\pi(i) = \frac{1}{\mathbb{E}_i[T_i]}$, where $T_i \geq 1$ is the time of the first visit to i. That is if there exists a stationary distr. $\pi = (\pi(0), \pi(1), \dots)$ then $\mathbb{E}_0[T_0] < \infty$. Now we try to conditions under which stationary distribution π exists.

$(\pi(0), \pi(1), \pi(2), \dots) \cdot$



$= (\pi(0), \pi(1), \pi(2), \dots)$

This equation implies that

$\pi(1) = \pi(0)p_0$; $\pi(2) = \pi(1)p_1$, $\pi(3) = \pi(2)p_2, \ldots,$ $\pi(n+1) = \pi(n)p_n, \ldots$ - -

$$\boxed{\pi(n+1) = \pi(0) \cdot \prod_{k=0}^{n} p_k .}$$

$\Downarrow$ We need $\sum_{n=0}^{\infty} \pi(n) = 1$

$$\pi(0) \cdot \sum_{n=0}^{\infty} \prod_{k=0}^{n} p_k = 1 \Rightarrow \pi(0) = \frac{1}{\sum_{n=0}^{\infty} \prod_{k=0}^{n} p_k}$$

IF $\sum_{n=0}^{\infty} \prod_{k=0}^{n} p_k < \infty$

So, if

(*) $\sum_{n=0}^{\infty} \prod_{k=0}^{n} p_k < \infty \Rightarrow \exists$ stat. distr. $\pi(n) = \pi(0) \cdot \prod_{k=0}^{n} p_k$.

If $\sum_{n=0}^{\infty} \prod_{k=0}^{n} p_k = \infty$ & $\lim_{n \to \infty} \prod_{k=0}^{n} p_k = 0$ then 0 recurrent

If $\lim_{n \to \infty} \prod_{k=0}^{n} p_k > 0$ then the chain is transient.

This answers part (b) & (c).