MA3238 Stochastic Processes I

- $(1-\frac{1}{n})^n \to e^{-1}$ as $n\to\infty$
- $\mathbb{N}_0 := \mathbb{N}\cup\{0\}$

- Sample space: set of all possible outcomes
- Event: subset of sample space
- For finite / countable sample space: prob. mass. function $P: S \to [0,1]$ with $\sum_{\omega\in S} P(\omega) = 1$
- Prob measure: $P(E) \in [0,1]$, $E \subseteq S$, s.t.
  - $P(\phi) = 0$
  - $P(S) = 1$
  - For any countable collection of events $E_1, E_2, \dots$ which are pairwise disjoint, $P(\bigcup_{i\in\mathbb{N}} E_i) = \sum_{i\in\mathbb{N}} P(E_i)$ (countable additivity)
- Basic properties:
  - $P(E^c) = 1 - P(E)$
  - If $E \subseteq F \subseteq S$, then $P(E) \le P(F)$
  - If $E, F \subseteq S$, then $P(E\cup F) = P(E) + P(F) - P(E\cap F)$
  - If $E_1, \dots, E_n \subseteq S$, then $P(E_1 \cup \dots \cup E_n) \le P(E_1) + \dots + P(E_n)$
- Conditional probability: $P(E|F) := \frac{P(E\cap F)}{P(F)}$
- Law of total probability: If $F_1, \dots, F_n$ is a partition of S, then for any event A, $P(A) = \sum_{i=1}^n P(A|F_i)P(F_i)$
- Peeling lemma: $P(E_1\cap\dots\cap E_n) = P(E_1)P(E_2|E_1)\cdots P(E_n|E_1\cap\dots\cap E_{n-1})$
- Independence of events:
  - For 2 events: $P(E\cap F) = P(E)P(F)$
    - equiv: $P(E|F) = P(E)$ and $P(F|E) = P(F)$
  - (Joint independence) For n events: $P(\bigcap_{i\in I} E_i) = \prod_{i\in I} P(E_i)$
    - equiv: $P(E_j | \bigcap_{i\in I\setminus\{j\}} E_i) = P(E_j)$
  - If E and F are independent, then E and F^c, E^c and F, and E^c and F^c are independent.
- Random variable: some function of the outcome of the experiment
- Probability space: (S, P) — S: sample space, P: probability measure
  - To avoid having a sample space, we use an abstract probability space (Ω, P) (master expt.). If the outcome is Z, then the outcome of any other expt is $X = f(Z)$ for deterministic $f: \Omega \to S$ (or: $X: \Omega\to S$)
    - often, $\Omega = [0,1]$ and P = unif. measure on $[0,1]$.
  - For a ran. var. $X: \Omega \to S$,
    - $P_X(A) := P(X\in A) := P(\omega\in\Omega : X(\omega)\in A)$ is a probability measure on S. (probability dist. of X)
  - For ran. vars. $X_i: \Omega \to S_i$, $i\in\mathbb{N}$,
    - joint distribution of $\vec{X} := (X_1, X_2, \dots)$ is $P_{\vec{X}}(A_1\times A_2\times\cdots) := P(\omega\in\Omega : X_1(\omega)\in A_1, \dots) = P(X_1\in A_1, X_2\in A_2, \dots)$
    - independence: $P(X_1\in A_1, \dots) = \prod_{i\in I} P(X_i\in A_i)$
- (cumulative) Distribution function (of $\mathbb{R}$-valued ran. var.):
  - $F_X: \mathbb{R}\to[0,1]$, $x \mapsto P(X\le x)$
  - $P(X\in(a,b]) = P(X\le b) - P(X\le a) = F_X(b) - F_X(a)$
  - it is a nondecreasing function
  - there is a bijection between F_X and P_X (for $\mathbb{R}$-valued ran. var.)
  - jump discontinuity: x where $P(X=x)$ (aka. the dist. has an atom at x)
  - continuous dist = no atoms
- Probability density function: $f: \mathbb{R}\to[0,\infty)$ s.t. $P(X\in(a,b]) = \int_a^b f(x)\,dx$
  - $F_X(x) = \int_{-\infty}^x f_X(x)\,dx$
  - $f_X(x) = \frac{d}{dx}F_X(x)$ (if f_X cts at x)
- Equiv condition for dist. functions (must satisfy all)
  1. $F(x) \le F(y)$ for all $x\le y$
  2. $\lim_{x\downarrow-\infty} F(x) = 0$ and $\lim_{x\uparrow\infty} F(x) = 1$
  3. For all $x\in\mathbb{R}$, $F(x) = \lim_{y\downarrow x} F(y)$ (i.e. right-continuity) and $F(x^-) := \lim_{y\uparrow x} F(y)$ exists (left-limit; this fact comes from math analysis); $P(X=x) = F(x) - F(x^-)$
- Countable additivity
  - $A_1 \subseteq A_2 \subseteq \dots \subseteq \Omega \implies P(\bigcup_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} P(A_n)$
  - $\Omega \supseteq A_1 \supseteq A_2 \supseteq \cdots \implies P(\bigcap_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} P(A_n)$ ← Pf: let $B_i = A_i \setminus A_{i-1}$ and use countable additivity.
- Thm for generating a distribution: Let $F: \mathbb{R}\to[0,1]$ be a distribution function, and $F^{\leftarrow}(y) := \inf\{x\in\mathbb{R} : F(x)\ge y\}$ (leftmost inverse). If Z is a uniform ran. var. on $[0,1]$ then $X := F^{\leftarrow}(Z)$ is a ran. var. with distrib. fn. F.

- Discrete random variables (dist. is purely atomic):
  - Bernoulli(p) ($p \in (0,1)$): $P(X=1)=p$; $P(X=0)=1-p$; $E[X]=p$; $Var(X)=p(1-p)$
  - Binomial(n,p) ($n \in \mathbb{N}, p \in (0,1)$): $P(X=i) = \binom{n}{i} p^i (1-p)^{n-i}$; $E[X]=np$; $Var(X)=np(1-p)$
  - Geometric(p) ($p \in (0,1)$): $P(X=i) = (1-p)^{i-1} p$ ($i \in \{0,1,2,\dots\}$) ; memoryless ; $E[X]=\frac{1}{p}$; $Var[X] = \frac{1-p}{p^2}$; $P(X>i)=(1-p)^i$
  - Poisson(λ): $P(X=i) = e^{-\lambda}\frac{\lambda^i}{i!}$ ($i \in \{0,1,2,\dots\}$) ; $E[X]=Var(X)=\lambda$
- Continuous random variables (there is no x s.t. $P(X=x)>0$) → the ones we will look at additionally have a p.d.f. $f: \mathbb{R} \to [0,\infty)$ s.t. $\forall a<b$, $P(X \in (a,b]) = \int_a^b f(x)dx$
  - Uniform$[0,1]$: $f(x) = \mathbb{1}_{[0,1]}(x)$; $E[X]=\frac{1}{2}$; $Var(X)=\frac{1}{12}$
  - Exponential(λ) ($\lambda>0$): $f(x)=\lambda e^{-\lambda x}\mathbb{1}_{[0,\infty)}(x)$; mean $=\frac{1}{\lambda}$; memoryless ; $F(x)=1-e^{-\lambda x}$; $E[X]=\frac{1}{\lambda}$; $Var[X]=\frac{1}{\lambda^2}$
  - Normal(μ,σ^2) : $f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
- Memoryless: $P(X>t+t_0 \mid X>t_0) = P(X>t)$
- Neither discrete or cts ran. var. examples:
  - mixture of discrete & cts
  - cts ran var with no density (singularly cts) : there is a set $A \subseteq \mathbb{R}$ with $P(X\in A)=1$, s.t. the "size" of A w.r.t. the Lebesque measure on $\mathbb{R}$ is 0.
- Any ran. var. can be decomposed into a mixture of discrete, cts with p.d.f., and singular cts.
- Expectation: $E[X] := \begin{cases} \text{Discrete: } \sum_{x\in F} x P(X=x) \text{ (i.e. weighted average)} \\ \text{Continuous: } \int x \rho(x) dx \end{cases}$
  - Linearity: $\forall a,b \in \mathbb{R}$, $E[aX+bY]=aE[X]+bE[Y]$
  - If independent: $E[XY]=E[X]E[Y]$
- Integer discrete sum: $E[X] = \sum_{n=0}^{\infty} nP(X=n) = \sum_{n=1}^{\infty} P(X\geq n)$
- Variance: $Var(X) := E[X^2]-E[X]^2 = E[(X-E[X])^2]$
  - $\forall a,b\in\mathbb{R}$, $Var(aX+b)=a^2Var(X)$
- Markov's ineq: $P(X\geq a) \leq \frac{E[X]}{a}$ $\quad \forall a>0, \forall X\geq 0$
- Chebyshev's ineq: $P(|X-E[X]|>a) \leq \frac{Var(X)}{a^2}$ $\quad \forall a>0$ (i.e. X is unlikely to be too far away from its mean)
- Covariance: $Cov(X,Y) := E[XY]-E[X]E[Y] = E[(X-E[X])(Y-E[Y])]$
- Correlation: $Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = Corr(Y,X) \in [-1,1]$
  - independent ⟹ $Cov=0=Corr$
  - uncorrelated ⟺ $Corr=0$
- Variance of sum: $Var(X_1+\dots+X_n) + \sum_{i=1}^n Var(X_i) + \underbrace{2\sum_{1\leq i<j\leq n} Cov(X_i,X_j)}$ if X_i's are pairwise independent, then this is zero.
- WLLN: Given $(X_i)_{i\in\mathbb{N}}$ of i.i.d. ran. var (real-valued) s.t. $\mu := E[X_i] \in \mathbb{R}$ and $\sigma := \sqrt{Var(X_i)} < \infty$ and let $S_n := \sum_{i=1}^n X_i$.
  Then $\frac{S_n}{n}$ converges in probability to μ,
  i.e. $\forall \varepsilon>0$, $P(|\frac{S_n}{n}-\mu|>\varepsilon) \to 0$ as $n\to\infty$
- SLLN: With probability 1, $\frac{X_1+X_2+\dots+X_n}{n} \to \mu$ as $n\to\infty$
- Central Limit Theorem (CLT): Let $(X_i)_{i\in\mathbb{N}}$ be i.i.d. ran. var. with mean $\mu\in\mathbb{R}$ and variance $\sigma^2\in(0,\infty)$.
  Let $S_n := \sum_{i=1}^n X_i$, and $W_n := \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Then W_n converges in distribution to Z.
  (i.e. $\forall a<b$, $P(W_n\in[a,b]) \xrightarrow[n\to\infty]{} P(Z\in[a,b]) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$
- Poisson Limit Theorem: For $n\in\mathbb{N}$, let $X_{n,1},\dots,X_{n,n}$ be i.i.d. Bernoulli ran. var. with $P(X_{n,i}=1)=\frac{\lambda}{n}$ for some $\lambda>0$.
  Then $S_n := \sum_{i=1}^n X_{n,i}$ converges in distribution to Poisson(λ).
  (i.e. $\forall k\in\mathbb{N}\cup\{0\}$, $P(S_n=k) \xrightarrow[n\to\infty]{} e^{-\lambda}\cdot\frac{\lambda^k}{k!}$)

- **Generating Functions:** If X is an $\mathbb{N}_0$-valued ran. var, then $G(s) := E[s^X] = \sum_{n=0}^{\infty} s^n P(X=n)$

  ($\mathbb{N}_0$-valued: i.e. discrete ran var with probability supported by whole numbers only)

  ($G(s)$: generating function of X.)

  - Properties:
    - $G(0) = P(X=0)$
    - $G(1) = 1$
    - G is increasing in $s \in [0,\infty)$
      (so $\exists s^* \in [1,\infty]$ s.t. $G(s) < \infty\ \forall s \in [0,s^*)$ and $G(s) = \infty\ \forall s \in (s^*,\infty)$)
    - G is a power series in s.
    - $G^{(k)}(s)\big|_{s=0} = k! \cdot P(X=k)$

      ($G^{(k)}$: k^{th} derivative of G)
    - If we let $s = e^t$, then $\Lambda(t) := G(e^t) = E[e^{tX}]$, then $\Lambda^{(k)}(t)\big|_{t=0} = E[X^k]$
    - $\Lambda(t) = \sum_{k=0}^{\infty} \frac{E[X^k]}{k!} t^k$

- **Convergence of generating functions:** If $(X_n)_{n\in\mathbb{N}}$ and X are ran. vars. with generating functions $(G_n)_{n\in\mathbb{N}}$ and G resp., and for all $s \in [0,a]$, $G_n(s) \xrightarrow[n\to\infty]{} G(s)$, then X_n converges in distribution to X.

  ($a > 0$)

  (i.e. $P(X_n = k) \xrightarrow[n\to\infty]{} P(X=k)$ for all $k \in \mathbb{N}_0$)

- **Limiting Poisson Limit Theorem:** For $n\in\mathbb{N}$, let $X_{n,1},\dots,X_{n,n}$ be indep. Berroulli ran. var. with mean $p_{n,1},\dots,p_{n,n}$ resp.

  If $\max_i p_{n,i} \xrightarrow[n\to\infty]{} 0$ and $\sum_i p_{n,i} \xrightarrow[n\to\infty]{} \lambda \in (0,\infty)$, then $S_n := \sum_{i=1}^{n} X_{n,i}$ converges in distribution to Poisson(λ).

  (i.e. $\forall k\in\mathbb{N}_0$, $P(S_n = k) \xrightarrow[n\to\infty]{} e^{-\lambda} \cdot \frac{\lambda^k}{k!}$)

- **Galton-Watson branching process:**
  - modelling population growth.
  - Z_n : size of population at time n
  - $Z_0 = 1$
  - Each of the Z_n individuals independently produces some offspring (using some common distribution)

    $f : \mathbb{N}_0 \to [0,1]$

    $G(s) := \sum_{n=0}^{\infty} s^n f(n)$

  Properties:
  - $E[Z_n] = Z_0 \mu_n = \mu^n$ where $\mu := \sum_{n=0}^{\infty} n \cdot f(n)$ is the mean number of offsprings per individual
  - If $G_n(s) := E[s^{Z_n}]$, then $G_n(s) = (\underbrace{G \circ \cdots \circ G}_{n \text{ fold composition}})(s)$
    - $G_n'(1) = E[Z_n]$
    - $G_n''(1) = E[Z_n(Z_n - 1)]$
  - Given $\mu := \sum_{n=0}^{\infty} n \cdot f(n)$ and $\sigma^2 := \sum_{n=0}^{\infty} n^2 \cdot f(n) - \mu^2$ of the offspring distribution,

    then $E[Z_n] = \mu^n$ and $Var(Z_n) = \begin{cases} n\sigma^2 & \text{if } \mu = 1 \\ \sigma^2(\mu^n - 1)\mu^{n-1}(\mu-1)^{-1} & \text{otherwise} \end{cases}$
  - Extinction probability: $\eta :=$ extinction probability (i.e. $P(Z_n = 0) \xrightarrow[n\to\infty]{} \eta$)

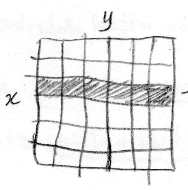
    Then η is the smallest non-negative root of equation $s = G(s)$.
    - If $\mu > 1$ then $\eta < 1$ (i.e. there is nonzero prob. that population never goes extinct)
    - If $\mu = 1$ and $f(1) \neq 1$ then $\eta = 1$
    - If $\mu < 1$ then $\eta = 1$.

Conditional Expectation & Variance:

- Conditional expectation: $E[X|F] := \dfrac{E[X 1_F]}{P(F)}$
- Conditional variance: $Var(X|F) := E[X^2|F] - E[X|F]^2 = E[(X - E[X|F])^2|F]$
- Marginal distribution: Given two ran. vars. X, Y, $P(X\in A) = \sum_{y\in S} P(X\in A, Y=y) = \sum_{y\in S} P(X\in A|Y=y)\cdot P(Y=y)$ $= E[P(X\in A|Y)]$
- Marginal expectation: $E[X] = \sum_{y\in S} E[X 1_{\{Y=y\}}] = \sum_{y\in S} E[X|Y=y]P(Y=y)$ $= E[E[X|Y]]$
  - for $f(X)$: $E[f(X)] = E[E[f(X)|Y]]$
  - for $f(X) = 1_{X\in A}$: $P(X\in A) = E[P(X\in A|Y)]$
  - for $f(X,Y)$: $E[f(X,Y)] = E[E[f(X,Y)|Y]]$
  - $E[\phi(X_1,\dots,X_n)] = E[E[\cdots E[E[\phi(X_1,\dots,X_n)|X_1,\dots,X_{n-1}]|X_1,\dots,X_{n-2}]\cdots|X_1]]$

Markov Chain:

- A sequence of rand. var. $X := (X_n)_{n\in\mathbb{N}_0}$ satisfying the Markov property:
  For all $x_0,\dots,x_n$, $P(X_{n+1}\in\cdot|X_0=x_0,\dots,X_n=x_n) = P(X_{n+1}\in\cdot|X_n=x_n)$
  (i.e. conditioned on the present, the past is independent of the future)
  - state space S (i.e. possible values of X_i) is finite or countable
  - Time homogeneity: $\forall x\in S$, $P(X_{n+1}\in\cdot|X_n=x)$ does not depend on $n\in\mathbb{N}_0$
- Transition matrix: Π where $\Pi(x,y) = P(X_{n+1}=y|X_n=x)$
- Initial distribution (of X_0): $\mu(x_0) := P(X_0=x)$ $\quad (\mu := (\mu(x))_{x\in S})$
  - $P(X_0=x_0, X_1=x_1,\dots,X_n=x_n) = \mu(x_0)\Pi(x_0,x_1)\cdots\Pi(x_{n-1},x_n)$



→each row sums to 1.

- Stochastic matrix: any matrix $(\Pi(i,j))_{i,j\in S}$ satisfying $\Pi(i,j)\geq 0$ and $\sum_j \Pi(i,j) = 1$
  (S: finite or countable)

- n-step transition probabilities: $p_n(x,y) := P(X_n=y|X_0=x)$
  - $p_{m+n}(x,z) = \sum_{y\in S} p_m(x,y)p_n(y,z)$
  - $p_n(x,y) = \Pi^n(x,y)$
  - $P(X_n=y) = \sum_{x\in S}\mu(x)\Pi^n(x,y) = (\mu\Pi^n)(y)$
  - $(\Pi^n f)(x) = E[f(X_n)|X_0=x]$ $\quad (f := (f(y))_{y\in S})$
  - $\mu\Pi^n f = E[f(X_n)]$
- Eigenvalues of Π:
  - $|\lambda|\leq 1$ for all eigenvalues of Π (and hence also for Π^T)
  - All eigenvectors that are fully non-negative and sum to 1 have eigenvalue 1 (i.e. probability distribution-ness is preserved)
- Intercommunicating states: $x\sim y := P(X_m=y|X_0=x)>0$ and $P(X_n=x|X_0=y)>0$ for some $m,n\in\mathbb{N}_0$
  (i.e. can get from x to y and vice-versa with positive probability)
  - "$\sim$" is an equivalence relation
  - Irreducible Markov chain: there is a single equivalence class
- Return probability: $f_{xx} := P(T_x<\infty|X_0=x) = \sum_{n=1}^{\infty} P(T_x=n|X_0=x)$ (i.e. probability of starting from x and returning to x in finite time)
  $T_x := \min\{n\in\mathbb{N} : X_n = x\}$ (may be ∞)
  - recurrent state: $f_{xx}=1$
  - transient state: $f_{xx}<1$
  - $P(T_x=\infty|X_0=x) = 1 - P(T_x<\infty|X_0=x) = 1 - f_{xx}$

⑤

- <u>Expected number of visits to starting state</u>: $G(x,x) := E\left[\sum_{n=0}^{\infty} \mathbb{1}\{X_n = x\} \,\middle|\, X_0 = x\right] = \sum_{n=0}^{\infty} \Pi^n(x,x) = \frac{1}{1-f_{xx}}$
  - $G(x,x) = \infty \iff x$ is recurrent

  if $f_{xx}=1$, then $G(x,x)=\infty$

- <u>Transience/Recurrence is a class property</u>: If $x \sim y$ then either they are both recurrent or both transient
  - <u>Transience/Recurrence of an irreducible Markov chain</u>
- All irreducible <u>finite</u> state Markov chains are recurrent
- <u>Path properties</u>: Given an irreducible countable state Markov chain X,
  - If X is <u>recurrent</u>: with probability 1, X visits each $y \in S$ infinitely often, i.e. $P(\forall y \in S, \sum_{n=0}^{\infty} \mathbb{1}\{X_n = y\} = \infty) = 1$
  - If X is <u>transient</u>: with probability 1, X visits each $y \in S$ finitely often, i.e. $P(\forall y \in S, \sum_{n=0}^{\infty} \mathbb{1}\{X_n = y\} < \infty) = 1$
- <u>Probability of escaping from a finite set</u>: Given an irreducible countable MC X, and $F \subset S$ a finite set of states, and $T_{F^c} := \min\{n \geq 0 : X_n \notin F\}$ (the first time X exits from F)

  Then $\exists\, C > 0$, $\rho \in (0,1)$ s.t. $\forall n \in \mathbb{N}_0$ and all initial distributions, $P(T_{F^c}(X) \geq n) \leq C\rho^n$
  - The optimal choice for ρ is the largest eigenvalue of the matrix $\Pi_F := (\Pi(x,y))_{x,y \in F}$
- <u>Pólya's theorem</u>: The simple symmetric random walk on $\mathbb{Z}^d$ is recurrent in dimensions $d=1$ and 2, and transient in $d \geq 3$.
- <u>Limiting distribution of transient Markov chain</u>: Given any irreducible transient MC:
  - $\forall x,y \in S$, $\Pi^n(x,y) \xrightarrow[n\to\infty]{} 0$
  - for any initial distribution, $\forall y \in S$, $P(X_n = y) \xrightarrow[n\to\infty]{} 0$
- <u>Stationary measure</u>: any $\mu: S \to [0,\infty)$ s.t. $\mu\Pi = \mu$ (i.e. $\sum_{x\in S} \mu(x)\Pi(x,y) = \mu(y)$)
  - <u>Stationary distribution</u>: stationary measure where $\sum_{x \in S} \mu(x) = 1$.
    - for finite S, we can find a stationary measure by solving $\mu(y) = \sum_{x \in S} \mu(x)\Pi(x,y)$, $\forall y \in S$
    - $\mu(x) = \frac{1}{E_x[T_1]}$

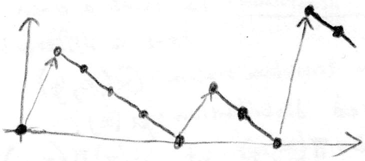
      expected return time to $x \in S$
- <u>Positive/Null recurrent</u>: Given $T_x := \min\{n \geq 1 : X_n = x\}$:
  - <u>Positive recurrent</u>: $E_x[T_x] < \infty$
  - <u>Null recurrent</u>: $E_x[T_x] = \infty$

  <u>Thm</u>: an irreducible MC is <u>positive recurrent</u> iff it has a <u>stationary distribution</u> (i.e. the sum of the stationary <u>measure</u> is finite)

  - If $x \sim y$ then they are both positive recurrent or null recurrent
    - So we can call irreducible MCs either positive or null recurrent
  - For all irreducible recurrent MCs, $\nu(y) := E_x\left[\sum_{n=0}^{T_x - 1} \mathbb{1}\{X_n = y\}\right]$ $\forall y \in S$ is a stationary measure, and if it is positive recurrent, then we can normalise ν to a stationary distribution.
  - <u>Stationary measure is unique for recurrent MCs</u>
- <u>Period</u> of an irreducible MC: $\gcd\{n \in \mathbb{N} : \Pi^n(x,x) > 0\}$ (it is the same for any $x \in S$)
  - <u>Periodic</u>: period > 2
  - <u>Aperiodic</u>: period $= 1$
- <u>Convergence for aperiodic MCs</u>: If X is an aperiodic positive-recurrent MC, then: (need to show irreducible too)

  regardless of stationary distribution, X_n converges in distribution to μ (i.e. $P(X_n = y) \xrightarrow[n\to\infty]{} \mu(y)$ $\forall y \in S$)
- For null recurrent MCs, $\Pi^n(x,y) \xrightarrow[n\to\infty]{} 0$ $\forall x,y \in S$, and for any initial distribution, $P(X_n = y) \xrightarrow[n\to\infty]{} 0$ $\forall y \in S$

- **Discrete Renewal Process**: $(T_n)_{n\in\mathbb{N}_0}$ where $T_0=0$ and $(T_n-T_{n-1})_{n\in\mathbb{N}}$ are i.i.d. $\mathbb{N}\cup\{\infty\}$-valued ran. var. (with p.m.f. $f(k)=P(T_1=k)$ for $k\in\mathbb{N}\cup\{\infty\}$)
  - Interpretation: the sequence of times we need to change a light bulb
  - Counting the number of renewals: $N_n=\max\{i\in\mathbb{N}_0 : T_i\le n\}$
  - Markov chain: $X_0:=0$
    $\Pi(n,n-1)=1 \quad \forall n\in\mathbb{N}$
    $\Pi(0,k-1)=f(k) \quad \forall k\in\mathbb{N}\cup\{\infty\}$
    $\Pi(\infty,\infty)=1$



  - $P(n\in\{T_1,T_2,\dots\}) \underset{n\to\infty}{=} P(X_n=0)$
  - if $f(\infty)>0$: 0 is transient, so $P_0(X_n=0)\xrightarrow[n\to\infty]{}0$
  - if $f(\infty)=0$: X is a recurrent MC, and $E_0[T_0]=E[T_1]=\sum_{n=1}^{\infty} nf(n)$
    - $E_0[T_0]$: expected first return time to 0
    - $=\infty$: null recurrent, $P_0(X_n=0)\xrightarrow[n\to\infty]{}0$
    - $<\infty$: positive recurrent, if aperiodic then $P_0(X_n=0)\xrightarrow[n\to\infty]{}\frac{1}{E_0[T_0]}=\frac{1}{\sum_n nf(n)}$

- **Reversible measure**: $\nu: S\to[0,\infty)$ where $\nu(x)\Pi(x,y)=\nu(y)\Pi(y,x)$ (the detailed balance condition)
  - Markov chain is reversible: a reversible measure exists (and this is a stationary measure)
    - If X_0 has distribution ν, then $(X_0,\dots,X_n)$ has the same distribution as $(X_n,\dots,X_0)$, $\forall n\in\mathbb{N}$. (known as time reversibility)
    - If $x=x_0,x_1,\dots,x_{n-1},x_n=x$ is a cycle, then: $\Pi(x_0,x_1)\Pi(x_1,x_2)\cdots\Pi(x_{n-1},x_n)$
      $=\Pi(x_n,x_{n-1})\Pi(x_{n-1},x_{n-2})\cdots\Pi(x_1,x_0)$
      (Loop condition) (product in either direction is equal)
    - Loop condition is necessary & sufficient for reversibility
      - if there are only trivial loops in the state space (in the sense that every step must be retraced backword to form a loop) then the loop condition is satisfied.
    - Random walk on an electrical network: Given a graph $G=(V,E)$, where for each $uv\in E$, there is a "conductance" $C(u,v)=C(v,u)>0$
      and $\Pi(x,y)=\frac{C(x,y)}{\sum_{z:z\sim x} C(x,z)}$ for all $y\sim x$
      Then the measure $\nu(x)=\sum_{y:y\sim x} C(x,y)$ is reversible.
      - Where $C(\cdot,\cdot)=1$, this is the random walk on the graph G

- **Continuous time Markov chains**: $(X_t)_{t\ge0}$ where $t\in[0,\infty)$ and $X_t\in S$ $\forall t$.
  - Markov property: $P((X_s)_{s\ge t}\in\cdot\,|\,(X_s)_{0\le s\le t})=P((X_s)_{s\ge t}\in\cdot\,|\,X_t)$
  - $T_x:=\min\{s\ge0 : X_s\ne x\}$ given $X_0=x$.
  - T_x must be memoryless, so it must follow an exponential distribution
  - So $P_x(T_x\ge t)=e^{-\lambda_x t}$ for some $\lambda_x>0$ for all $t\ge0$.
  - If λ_x is uniform over all $x\in S$, then we can simulate it with a discrete-time Markov chain and an exponential timer with mean λ.
  - X is irreducible $\Leftrightarrow$ the discrete-time analogue is irreducible
  - X is irreducible $\Leftrightarrow$ $\Pi_t(x,y):=P_x(X_t=y)>0$ for all $x,y\in S$ and $t>0$
  - X is recurrent/transient $\Leftrightarrow$ the discrete-time analogue is recurrent/transient
  - If the discrete-time analogue is transient or null-recurrent then $\lim_{t\to\infty}\Pi_t(x,y)=0$
  - If the discrete-time analogue is positive-recurrent then $\lim_{t\to\infty}\Pi_t(x,y)=\mu(x,y)$
    - stationary distribution

- Monte Carlo algorithms: use repeated sampling to estimate a probability distribution of interest.
- Markov Chain Monte Carlo: estimate a probability distribution by simulating a Markov chain.
  - Ergodic theorem: If X is an irreducible Markov chain with countable state space S and stationary distribution μ, then for any initial distribution, $S_n := \frac{1}{n}\sum_{i=1}^{n} f(X_i) \xrightarrow[n\to\infty]{} E_\mu[f(Y)]$

    Y is a ran. var. with distribution μ.
- Metropolis-Hastings algorithm: To make a given Markov chain (irreducible) have a different stationary distribution:
  - Given transition matrix: $Q(x,y)$
  - Desired distribution: $\mu(x)$.
  - Want: $\Pi(x,y)$ s.t. $\mu(x)\Pi(x,y) = \mu(y)\Pi(y,x)$ (detailed balance of Π) and $\Pi(x,y) = Q(x,y)\,h(x,y)$ (accept probability) $\forall y \neq x$

    $\Pi(x,x) = 1 - \sum_{y\neq x}\Pi(x,y)$
  - $h(x,y) := \begin{cases} 1 & \text{if } \mu(y)Q(y,x) \geq \mu(x)Q(x,y) \\ \dfrac{\mu(y)Q(y,x)}{\mu(x)Q(x,y)} & \text{otherwise} \end{cases}$

  (i.e. $h(x,y) := \min\left\{1, \dfrac{\mu(y)Q(y,x)}{\mu(x)Q(x,y)}\right\}$ and $h(y,x) := \min\left\{1, \dfrac{\mu(x)Q(x,y)}{\mu(y)Q(y,x)}\right\}$)

  Pf: by observing that $\dfrac{h(x,y)}{h(y,x)} = \dfrac{\mu(y)Q(y,x)}{\mu(x)Q(x,y)}$
- Gibbs sampling:
  - To find the joint distribution μ of a collection $\vec{W} := (W_1, \dots, W_N)$ of ran. var. (not necessarily independent)
  - Steps: given $\vec{X}_n := (X_{n,1}, \dots, X_{n,N})$:
    - pick $i \in \{1, \dots, N\}$ uniformly randomly
    - $P(X_{n+1,i} = z) = P(W_i = z \mid W_j = X_{n,j} \;\; \forall j \neq i)$

      $= \dfrac{\mu((X_{n,1}, \dots, X_{n,i-1}, z, X_{n,i+1}, \dots, X_{n,N}))}{\sum_w \mu((X_{n,1}, \dots, X_{n,i-1}, w, X_{n,i+1}, \dots, X_{n,N}))}$
  - Then Π satisfies detailed balance