Distribution	Probability Function	Mean	Variance	MGF
Binomial	$p(y) = \binom{n}{y} p^y (1-p)^{n-y}$	np	$np(1-p)$	$[pe^t + (1-p)]^n$
Geometric	$p(y) = p(1-p)^{y-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$
Hypergeometric	$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$	$\frac{nr}{N}$	$n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$	No closed form
Poisson	$p(y) = \frac{\lambda^y e^{-\lambda}}{y!}$	λ	λ	$e^{\lambda(e^t-1)}$
Negative binomial	$p(y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$\left(\frac{pe^t}{1-(1-p)e^t}\right)^r$
Uniform	$f(y) = \frac{1}{\theta_2-\theta_1}$	$\frac{\theta_1+\theta_2}{2}$	$\frac{(\theta_2-\theta_1)^2}{12}$	$\frac{e^{t\theta_2}-e^{t\theta_1}}{t(\theta_2-\theta_1)}$
Normal	$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{1}{2\sigma^2}\right)(y-\mu)^2}$	μ	σ^2	$e^{\mu t + \frac{t^2\sigma^2}{2}}$
Exponential	$f(y) = \beta e^{-\beta y}$	$\frac{1}{\beta}$	$\frac{1}{\beta^2}$	$\left(1-\frac{1}{\beta}t\right)^{-1}$
Gamma	$f(y) = \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right) y^{\alpha-1} e^{-\beta y}$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	$\left(1-\frac{1}{\beta}t\right)^{-\alpha}$
Chi-square	$f(y) = \frac{y^{\frac{\nu}{2}-1} e^{-\frac{y}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}$	ν	2ν	$(1-2t)^{-\frac{\nu}{2}}$
Beta	$f(y) = \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right) y^{\alpha-1}(1-y)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	No closed form

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{(ad)-(bc)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

$$\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x, Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$$

$$Pr(A) = \sum_k Pr(A \cap B_k)$$

where $B_k \cap B_j = \emptyset \ \forall k \neq j, Pr(\cup_j B_j) = 1$

$$Pr(B|A) = \frac{Pr(A|B)Pr(B)}{Pr(A|B)P(B) + Pr(A|\overline{B})Pr(\overline{B})}$$

$$Pr(B_i|A) = \frac{Pr(A|B_i)Pr(B_i)}{\sum_j Pr(A|B_j)Pr(B_j)}$$

$$E(Y) = \begin{cases} \sum_y y Pr(Y=y) & \text{discrete} \\ \int_{-\infty}^{\infty} y\ f_y(y)\ dy & \text{continuous} \end{cases}$$

$$E(Y) = \sum_{i=1}^{k} E(Y|A_i)P(A_i)$$

$$E(Y|X=x) = \begin{cases} \sum_y y Pr(Y=y|X=x) & \text{discrete} \\ \int_{-\infty}^{\infty} y\ f_y(y|x)\ dy & \text{continuous} \end{cases}$$

$$E(aY+b|X=x) = aE(Y|X=x) + b$$

$$E(g(y)|X=x) = \begin{cases} \sum_y g(y) Pr(Y=y|X=x) & \text{discrete} \\ \int_{-\infty}^{\infty} g(y)\ f_y(y|x)\ dy & \text{continuous} \end{cases}$$

$$E(Y|X=x) = E(Y) \text{ if } X \text{ indep } Y$$

$$E(Y) = E(E(Y|X))$$

$$Y = g(x) \implies E(Y|X=x) = g(x)$$

$$V(Y) = E(Y^2) - (E(Y))^2 = E[(Y-E(Y))^2]$$

$$V(Y|X=x) = E((Y-\mu_x)^2|X=x)$$

$$= \begin{cases} \sum_y (y-\mu_x)^2 Pr(Y=y|X=x) & \text{discrete} \\ \int_{-\infty}^{\infty} (y-\mu_x)^2 f_{Y|X}(y|x)\ dy & \text{cont} \end{cases}$$

$$V(Y) = E(V(Y|X)) + V(E(Y|X))$$

MATH447 Crib Sheet

Julian Lore

Probability

Law of Iterated Expectation

$$E_Y(Y) = E_X[E_{Y|X}(Y|X)]$$

Ex: # of coins to flip according to Poi(λ), coins have prob p heads, $1-p$ tails. Expected number of heads, $T = \sum_{i=1}^{n} x_i$ (indicator). N = # flips $\sim Poi(\lambda)$.

$E(T) = E_N(E_{T|N}(T|N)) = E_N(NP) = pE_n(N) = p\lambda$

$E(N\mu_x) = \mu_x\mu_N$ if x_i i.i.d and $\mu(N)$ indep of N

Law of Iterated Variance

$$V_Y(Y) = V_X(E_{Y|X}[Y|X]) + E_X(V_{Y|X}(Y|X))$$

Markov Chains

Stochastic process $\{X_t : t \in T\}$ where $Pr(x_t|x_{t-1}, \ldots, x_0) = Pr(x_t|x_{t-1})$.

Markov Property $X_n \perp X_0, \ldots, x_{n-2} | x_{n-1}$

Stochastic Matrix

Matrix with min vals ≥ 0, max vals ≤ 1, rows sum to 1. **State space** is set of vals for x_t. Transition prob matrix: $\mathbf{P} = \begin{array}{c} \\ \underbrace{\begin{array}{c} 0 \\ 1 \end{array}}_{x_t} \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1]x_{t+1} \end{array} \\ \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \end{array}$ $\mathbf{P}_{ij}$ = prob of going from $i \to j$ in one step. $\mathbf{P}_{ij}^n$ = prob of going from $i \to j$ in n steps.

Chapman-Kolmogorov Relationship

$P_{ij}^{m+n} = \sum_k P_{ik}^m P_{kj}^n$. For TH: $\Pr(X_{m+n} = j \mid X_0 = i) = \sum_k \Pr(X_m = k \mid X_0 = i)\Pr(X_{m+n} = j \mid X_m = k)$

Distribution $\Pr(X_n = j) = (aP^n)_j$, where a is the initial distribution.

Limiting Distribution

$\{X_t\}$ has a limiting distribution if $\lim_{n\to\infty}(P^n)_{ij} = \lambda_j$ for all i and j (not guaranteed). $\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ doesn't have a limiting distribution.

If $\vec{\lambda}$ is limiting distrib of $\mathbf{P}$, then $\vec{\lambda}\mathbf{P} = \vec{\lambda}$ (not a bi-implication). Still have $\vec{\pi}\mathbf{P} = \vec{\pi}$ for stationary distribution.

Limiting distrib gives you long-term expected proportion of time that the chain is in that state.

$$\lim_{n\to\infty} \mathbf{P}^n = \mathbf{\Lambda}$$

where $\mathbf{\Lambda}$ is a stochastic matrix with each row being λ.

Positive TPM

TPM whose entries are $> 0, \mathbf{P} > 0$. If not positive but $\exists n$ s.t. $\mathbf{P}^n > 0$ then $\mathbf{P}$ is a regular matrix. If $\mathbf{P}$ is regular, then $\exists$ **unique** $\vec{\pi}$ that is a stationary distrib for $\mathbf{P}$ and will also be a limiting distribution. $\mathbf{P}$ will not be regular if $\mathbf{P}^n$ and $\mathbf{P}^{n+1}$ have entries that are zero in both matrices.

How to find Stationary Dis

General: $\vec{\pi}\mathbf{P} = \vec{\pi}$, do the mult:

$$(\pi_1, \pi_2)\begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} = (\pi_1, \pi_2)$$

$$\implies \pi_1 P_{11} + \pi_2 P_{21} = \pi_1$$

$$\pi_1 P_{12} + \pi_2 P_{22} = \pi_2$$

NOTE: we solve by column, column 1 = π_1 etc.

Also have $\pi_1 + \pi_2 = 1$. Solve linear system with 1 redundant eq. If **stationary unique**, then can use/should use:

$$\vec{x} = (1, x_2, \ldots, x_k)$$

$$\vec{x}\mathbf{P} = \vec{x} \text{ (solve)}$$

$$\vec{\pi} = \left(\frac{1}{1 + x_2 + \ldots + x_k}\right)\vec{x}$$

Let W be $k \times k$ matrix. If $W\vec{v} = \lambda\vec{v}$ then $\vec{v}$ is a right eigenvector of W with eigenvalue λ. If we can construct matrix of eigenvectors $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_u}(V)$ where λ_i is the eigenvalue associated with $\vec{v_i}$ then $W = V\Lambda V \iff WV = V\Lambda$ (eigenvalue decomposition), where Λ is a matrix of 0s with $\lambda_1, \lambda_2, \ldots, \lambda_u$ on the diagonal. $\vec{\pi}\mathbf{P} = \vec{\pi} = 1 \cdot \vec{\pi} \implies \mathbf{P}^T\vec{\pi}^T = 1 \cdot \vec{\pi}^T$. Here $\vec{\pi}^T$ is an eigenvector of $\mathbf{P}^T$ corresponding to eigenvalue of 1.

Communication

If we can get from i to j, then j is accessible from i. For time-homogeneous $\mathbf{P}$, if $\exists n \geq 0$ s.t. $\mathbf{P}_{ij}^n > 0$, then j accessible from i.

If j accessible from i and i accessible from j, then i and j communicate. Communication is symmetric, reflexive ($\mathbf{P}_{ii}^0 = 1$), transitive.

If all states of chain communicate with each other, then the chain is irreducible.

First hitting time: $T_j = \min\{n > 0 : X_n = j\ if\ x_0 = j\}$

$f_j = Pr(T_j < \infty \mid X_0 = j) = 1 \iff$ state j is a recurrent state. If recurrent, expected number of returns is infinity (will become infinite sum of 1)

$f_j = Pr(T_j < \infty \mid X_0 = j) < 1 \iff$ state j is a transient state. If transient, expected returns is geometric $(1-f_j)$. $\sum_{n=0}^{\infty} P_{jj}^n = \frac{1}{1-f_j}$

Communication class: set of states who **all communicate with each other** and no one else. **State by itself looping** is a communication class. All states in a communication class are either **all recurrent or all transient**

Recurrent chain, irreducible (since all states recurrent)

Closed comm classes: C is closed $\iff \mathbf{P}_{ij} = 0\ \forall i \in C, j \notin C$. States is just a union of classes of transient and classes of recurrent states.

Finite irreducible MC $\implies$ all states recurrent.

Finite communication class closed only if it consists of all recurrent states.

Canonical decomposition of markov chain: $\begin{array}{c|ccc} & T & R_1 & R_2 \\ \hline T & * & * & * \\ R_1 & 0 & [P_1] & 0 \\ R_2 & 0 & 0 & [P_2] \end{array}$ Where T is the class of all transient states (may or may not communicate) and R_i are recurrent communication classes. P_i are irreducible mini tpm, mini markov chain on reduced state space.

How to find expected return time

$\mu_j = E[T_j|x_0 = j]$ (for finite irreducible markov chains, positive stationary distrib is unique)

1. Use stationary distrib. $\{x_0, \ldots, x_n\}$ finite (# of states) irreducible chain. Then $\mu_j < \infty$ and $\exists\vec{\pi}$ s.t. $\pi_j = \frac{1}{\mu_j}\ \forall j$. We have $\pi_j = \lim_{n\to\infty} \frac{1}{n}\sum_{m=0}^{n-1} \mathbf{P}_{ij}^m$ (limit of avg). This is NOT the same as $\lim_{n\to\infty} \mathbf{P}_{ij} = \lambda_j$, as this will not converge if there is no limiting distribution.
2. First step analysis. Find $e_x = E(T_a \mid x_0 = x)$ by finding e_k for all relevant states. Ex. $\mathbf{P} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$. Then $e_a = \frac{1}{2}(1) + 0 \cdot (1 + e_b) + \frac{1}{2}(1 + e_c), e_b = 1(1 + e_c), e_c = \frac{1}{4} \cdot 1 + \frac{1}{4}(1 + e_b) + \frac{1}{2}(1 + e_c)$. Get $e_a = \frac{7}{2} = \mu_a \implies \pi_a = \frac{2}{7}$
   Alternatively, just add 1 to every linear equation and ignore a (since wherever you transition to will add 1 to the amount of steps): $e_a = \frac{1}{2}e_c + 1 \mid e_b = e_c + 1 \mid e_c = \frac{1}{4}e_b + \frac{1}{2}e_c + 1$

Positive recurrent, recurrent j s.t. $E(T_j|x_0 = j) < \infty$

Null recurrent, recurrent j s.t. $E(T_j|x_0 = j) = \infty$

Periodicity

period of state i, $d(i)$ is the gcd of the set of possible return times to i. $d(i) = gcd\{n > 0 : \mathbf{P}_{ii}^n > 0\}$. If $d(i) = 1$, then i is aperiodic. If there is no return to i, $d(i) = \infty$. All states in a communication class have the same period. Markov chain is periodic if it is irreducible and all states have period > 1.

Otherwise, if irreducible and all states have period 1, then the chain is aperiodic. Chain is ergodic if irreducible, aperiodic and all states have finite return times. If chain is ergodic, then $\exists$ **unique** positive stationary distribution for the chain ($\pi_j = \lim_{n\to\infty}(\mathbf{P}^n)_{ij}\ \forall i, j$). Chain ergodic $\iff$ tpm regular.

A chain is time-reversible if $\pi_i\mathbf{P}_{ij} = \pi_j\mathbf{P}_{ji}\ \forall i, j$ (can't tell if I'm going forward or backwards) for stationary $\vec{\pi}$. These equations are called the detailed balance equations. $\pi_i\mathbf{P}_{ij} = Pr(x_0 = i)Pr(x_1 = j \mid x_0 = i) = Pr(x_0 = i, x_1 = j) \overset{\text{by time revers}}{=} Pr(x_0 = k, x_i = i) = Pr(x_0 = j)Pr(x_1 = i \mid x_0 = j) = \pi_j\mathbf{P}_{ji}$. Additionally $Pr(x_0 = i_0, x_1 = i_1, \ldots, x_n = i_n) = Pr(x_0 = i_n, x_1 = i_{n-1}, \ldots, x_n = i_0)$

State i is an absorbing state if $\mathbf{P}_{ii} = 1$. Markov chain is an absorbing chain if there is ≥ 1 absorbing state. For an absorbing chain with t transient states and k singleton absorbing states we have the canonical decomp: $\mathbf{P} = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}$, where Q is $t \times t$ matrix for transient states, 0 is $k \times t$ 0 matrix (can't go back to transient), R is $t \times k$ and I is $k \times k$ for absorbing. $\mathbf{P}^n = \begin{pmatrix} Q^n & (Q^{n-1} + Q^{n-2} + \ldots + Q + I)R \\ 0 & I \end{pmatrix}$

Lemma: square matrix A s.t. $\lim_{n\to\infty} A^n = 0$ then $\sum_{n=0}^{\infty} A^n = (I-A)^{-1}$. For absorbing above, $\lim_{n\to\infty} Q^n = 0 \implies \lim_{n\to\infty}\mathbf{P}^n = \begin{pmatrix} 0 & (I-Q)^{-1}R \\ 0 & I \end{pmatrix}$, where $F = (I-Q)^{-1}$ is the fundamental matrix of the absorbing chain (not a tpm). F_{ij} contains expected # of visits to j starting in i, where i and j are transient.

Expected time to absorption from i

$a_i = \sum_{j\in T} F_{ij}$ or $\vec{a} = F\vec{1}$, where T is the set of transient states. i.e. expected time to absorption starting from state 1 is sum of row 1. Absorption time from $i = (F1)_i$. If we can only transition to ourselves or be absorbed, then expected number of visits to ourselves is expected time to absorption. Absorption probability: prob that from transient i, chain is absorbed in j is $(FR)_{ij}$.

Branching Processes

0 is an absorbing state, all nonzero states transient.

Offspring: Each member of pop produces offspring **independently**. Offspring distrib is **same** across children and time.

Pmf of offspring dis given by: $\vec{a} = (a_0, a_1, \ldots)$, where $a_k = \Pr(X_i = k)$, number of offspring produced by unit i.

Time is measured in "generations" ($t = 2 \implies$ generation 2).

Z_n =# units in gen n. $\{Z_n\}$ is a branching process. Z_n can be modeled as a Markov Chain. $Z_n \in \mathbb{N}$.

$\Pr(Z_{n+1} = i_{n+1} \mid Z_n = i_n, \ldots, Z_0 = i_0) \overset{MC}{=} \Pr(Z_{n+1} = i_{n+1} \mid Z_n = i_n) \overset{\text{time-homog}}{=} \Pr(Z_1 = i_{n+1} \mid Z_0 = i_n)$

If $a_0 = 0, Z_{n+1} > Z_n, \forall n$ (run off to ∞)

If $a_0 = 1, Z_1 = 0$ then $\forall n$ the pop is **extinct**

Assume $\mathbf{0 < a_0 < 1}$. 0 is an absorbing state.

Two possible outcomes: get absorbed $Z_n = 0$, extinct or process grows without bound.

Mean generation size

$Z_n = \sum_{i=1}^{Z_{n-1}} X_{n_i}$ where X_{n_i} =# of offspring for i^{th} member of gen $n-1$. Z_n is a sum of i.i.d r.v. $X_{n_1}, \ldots, X_{n_{Z_{n-1}}}$.

$E(X_i) = \mu = \sum_{k=0}^{\infty} k a_k \implies E(Z_n) = \mu^n E(Z_0)$. If $Z_0 = 1$ with probability 1 then $E(Z_n) = \mu^n$.

$$\lim_{n\to\infty} E(Z_n) = \lim_{n\to\infty} \mu^n = \begin{cases} 0 & \text{if } \mu < 1 \text{ (subcritical)} \\ 1 & \text{if } \mu = 1 \text{ (critical)} \\ \infty & \text{if } \mu > 1 \text{ (supercritical)} \end{cases}$$

$$V(Z_n) = \sigma^2 \mu^{n-1} \sum_{k=0}^{n-1} \mu^k = \begin{cases} n\sigma^2 & \mu = 1 \\ \sigma^2 \mu^{n-1} \frac{(\mu^n - 1)}{\mu - 1} & \mu \neq 1 \end{cases}$$

$$\lim_{n\to\infty} V(Z_n) = \begin{cases} 0 & \mu < 1 \\ \infty \text{ (increases linearly)} & \mu = 1 \\ \infty \text{ (increases exponentially)} & \mu > 1 \end{cases}$$

$$\Pr(Z_n = 0) = 1 - \mu^n$$

Probability of extinction for subcritical is 1 (take limit).

Probability Generation Function

$$G_X(s) = G(s) = E(s^X) = \sum_{k=0}^{\infty} s^k \Pr(X = k)$$

Power series with coeffs that sum to 1. $G(1) = 1$, series converges absolutely for $|s| \le 1$

$$G(0) = \Pr(X = 0), G(1) = 1$$

$$G'(0) = \frac{\partial}{\partial s} \sum_{k=0}^{\infty} s^k \Pr(X = k)\Big|_{s=0} = 0 + \sum_{k=1}^{\infty} k s^{k-1} \Pr(X = k)\Big|_{s=0} = \Pr(X = 1)$$

$$G^{(j)}(0) = j! \Pr(X = j) \implies \Pr(X = j) = \frac{G^{(j)}(0)}{j!}$$

If $Y = X_1 + \ldots + X_n$, i.e. sum of independent r.v.s, then $G_Y(s) = \prod_{i=1}^n G_{X_i}(s)$. If same dist, then $\prod_{i=1}^n G_{X_i}(s) = [G_X(s)]^n$
Moments:

$$G_X^{(1)}(1) = \sum_{k=0}^{\infty} k \Pr(X = k) = E(X)$$

$$G_X^{(2)}(s) = \sum_{k=0}^{\infty} k(k-1) \Pr(X = k) = E(X^2) - E(X)$$

$$V(X) = E(X^2) - (E(X))^2 = G^{(2)}(1) + G^{(1)}(1)(1 - G^{(1)}(1))$$
$$= G''(1) + G'(1) - G'(1)^2$$

If X and Y are r.v. such that $G_x(s) = G_y(s)$ then X and Y have same distrib.
If X and Y are indep, then $G_{X+Y}(s) = G_X(s) G_Y(s)$
For branching processes:

$$G_X(s) = \sum_{k=0}^{\infty} s^k a_k$$

$$G_n(s) = \sum_{k=0}^{\infty} s^k \Pr(Z_n = k), G_n(0) = \Pr(Z_n = 0)$$

If $Z_0 = 1$ with prob 1:

$$G_n(s) = G_{n-1}(G_X(s)) = G_X(G_X(G_X(\ldots(G_X(s))))) = G_X(G_{n-1}(s))$$

Probability of extinction is smallest root of $s = G_X(s)$. If $\mu \le 1$, extinction prob = 1.

Markov Chain Monte Carlo

Gibbs/Boltzmann distribution $\pi(\sigma) = \frac{e^{\beta E(\sigma)}}{\sum_\tau e^{-\beta E(\tau)}}$.
For Ising, $\beta = 0$, infinite temp, uniform. $\beta > 0$, more mass on low-energy, favoring similar spin neighbors. $\beta < 0$, more mass on high-energy.
$\pi(\sigma) \propto e^{-\beta E(\sigma)}$. How to avoid computing normalizing constant?
Let $X_0, X_1, \ldots$ be an **ergodic** MC with tpm **P**, where $\pi\mathbf{P} = \pi$ is the stationary (and limiting) distrib. Can we choose a **P** such that $\pi\mathbf{P} = \pi$?

Metropolis-Hastings Algorithm

$\pi = (\pi_1, \ldots, \pi_k)$. 1. Choose any irreducible tpm T s.t T and π have same state space (T should be easy to sample from). 2. Choose any starting state for X_0. For $n = 1, 2, \ldots$: 3. Propose to move from $X_{n-1} = i$ to $X_n = j$ according to T (i.e. choose j with probability T_{ij}). 4. "Accept" move with probability $a(i,j) = \min\left(1, \frac{\pi_j}{\pi_i} \times \frac{T_{ji}}{T_{ij}}\right)$. If $a(i,j) = 1$, then $X_n = j$. If $a(i,j) < 1$ then $X_n = j$ with prob $a(i,j)$, $X_n = i$ with prob $1 - a(i,j)$. Repeat for $n + 1$
X_n will converge to draw from stationary distrib.
Proposal distrib is T

Gibbs sampling, proposes to change 1 component of the target r.v. at a time (conditional on other r.vs fixed) but all proposals are accepted (original metropolis alg can change multiple components at once).
$\pi(\mathbf{x}) = \pi(x_1, \ldots, x_m)$ is a m-dimensional joint density. Identify conditional distribs by treating other conditioning variables as fixed constants (easier to get proportional expr).
Strong law of large numbers for MC, $X_0, X_1, \ldots$ be a Markov Chain with stationary dist π.

$$\frac{r(X_0) + r(X_1) + \ldots + r(X_n)}{n+1} \to pE_n(r(x))$$

Borel distribution, $0 < \mu < 1$, $E(X = \frac{1}{1-\mu})$, $\Pr(X = x) = \frac{e^{(-x\mu)}(x\mu)^{x-1}}{x!}$

Examples

1. Exhibit MH alg to sample from binom with n, p. Use proposal distrib uniform on $\{0, 1, \ldots, n\}$. $\pi(y) \propto \frac{1}{y!(n-y)!}\left(\frac{p}{(1-p)}\right)^y$. $T_{ij} = \frac{1}{n+1}, \forall i, j$. $a(i,j) = \frac{j!(n-j)!}{i!(n-i)!}\left(\frac{p}{(1-p)}\right)^{j-i}$
2. Exhibit MH to sample for power-law. $\pi_i \propto i^s$, take proposal distrib as simple symmetric random walk with reflecting boundary at 1, i.e. always go from $1 \to 2$, otherwise, left or right with $\frac{1}{2}$ prob. $T_{ij} = \begin{cases} 1/2 & \text{if } j = i \pm 1, i > 1 \\ 1 & \text{if } i = 1 \text{ and } j = 2 \\ 0 & \text{ow} \end{cases}$.
Acceptance is $a(i, i+1) = \left(\frac{i}{i+1}\right)^s$ and $a(i+1, i) = \left(\frac{i+1}{i}\right)^s$ for $i \ge 2$ ($2, 1$ and $1, 2$ are special cases, can be computed).
3. Generate $Poi(\lambda)$ using simple symmetric random walk as proposal dist. $U \sim Unif[0,1]$. If walk is at state $k = 0$, move to $k = 1$ if $U < \lambda$, otherwise stay at $k = 0$. For $k \ge 1$, equal prob for $k-1$ or $k+1$. $a(k, k-1) = \frac{e^{-\lambda}\lambda^{k-1}/(k-1)!}{e^{-\lambda}\lambda^k/k!} = \frac{k}{\lambda}$. $a(k, k+1) = \frac{e^{-\lambda}\lambda^{k+1}/(k+1)!}{e^{-\lambda}\lambda^k/k!} = \frac{\lambda}{k+1}$
4. Random walk $T = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{pmatrix}$
5. $p(x,n) \propto \frac{e^{-3x}x^n}{n!}$. Sketch Gibbs.
Conditional distribs: $p(n \mid x) \propto p(n \mid x)p(x) \propto p(x,n) \propto \frac{e^{-3x}x^n}{n!} \propto \frac{x^n}{n!} \sim Poi(x)$.
$p(x \mid n) \propto p(x \mid n) \overbrace{p(n)}^{\text{constant wrt } x} \propto p(x,n) \propto \frac{e^{-3x}x^n}{n!} \propto e^{-3x}x^n \sim Gamma(n+1, 3)$
6. Gibbs for (X, Y) bivariate standard normal with correlation ρ. 1. Init $(x_0, y_0) = (0, 0)$. For $m = 1, \ldots$: 2. Gen x_m from $X \mid Y = y_{m-1}$, i.e. normal dist with mean ρy_{m-1}, var $1 - \rho^2$. 3. Gen y_m from $Y \mid X = x_m$, i.e. normal dist with mean ρx_m and var $1 - \rho^2$ 4. Repeat step 2

Counting Processes

Let N_t be # of events that occur in $[0, t]$. The collections $\{N_t : t \ge 0, t \in \mathbb{R}^+ \cup \{0\}\}$ is uncountable collection of discrete-valued r.v.s called a counting process.
More generally, counting process is collection of integer rvs s.t $0 \le s \le t \implies N_s \le N_t$

Poisson Process

3 definitions:

1. # of events in fixed intervals $[s, t]$. Poisson process with param λ is a counting process with: $N_0 = 0, \forall t > 0, N_t \sim Poi(\lambda t), \forall s, t > 0, N_{t+s} - N_s \sim N_t$, e.g. $\Pr(N_{t+s} - N_s = k) = \Pr(N_t = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$
For all $0 \le q < r \le s < t$, $N_t - N_s$ and $N_r - N_q$ are indep (time homog, independent increments)
2. Let $X_1, X_2, \ldots$ be seq of iid $Exp(\lambda)$ rvs. For $t \ge 0$, let $N_t = \max\{n : X_1 + X_2 + \ldots + X_n \le t\}$.
$S_n = X_1 + X_2 + \ldots + X_n$
S_k is k^{th} arrival time, X_k is interarrival time between $(k-1)^{th}$ and k^{th} arrival.
3. Counting process s.t: $N_0 = 0$, process has independent and stationary increments (interarrivals non-overlapping),
$\Pr(N_h = 0) = 1 - \lambda h + o(h), \Pr(N - h = 1) = \lambda h + o(h), \Pr(N_h > 1) = o(h)$

$E(N_t) = \lambda t, V(N_t) = \lambda t, \frac{E(N_t)}{time} = \lambda$ (rate of arrivals)
Translation process. Let $\{N_t : t \ge 0\}$ be a PP with rate λ. $\tilde{N}_t = N_{t+s} - N_s = \#$ events in $[s, t+s]$. Then $\left\{\tilde{N}_t : t \ge 0\right\}$ is also a PP with λ.
Let X be arrival time of first event. No arrivals in $[0, t] \iff X > t$.
$\Pr(N_t = 0) = e^{-\lambda t} \implies \Pr(X \le t) = e^{-\lambda t} \implies X \sim Exp(\lambda)$
Useful props
Memoryless process: $s \le t, s, t \ge 0$: If $X \sim Exp(\lambda)$, $\Pr(X > s + t \mid X > s) = \Pr(X > t)$
Minimum of exponential RVs: Let $M = \min(X_1, \ldots, X_n), X_i \overset{ind}{\sim} Exp(\lambda_i)$.
$\Pr(M > t) = e^{-t\sum_{i=1}^n \lambda_i} \implies M \sim Exp\left(\sum_{i=1}^n \lambda_i\right)$. $\Pr(M = X_k) = \frac{\lambda_k}{\lambda_1 + \ldots + \lambda_n}$
Maximum of exp RVs: Let $M = \max(X_1, \ldots, X_n), X_i \overset{ind}{\sim} Exp(\lambda_i)$.
$\Pr(M \le m) = \Pr(X_1 \le m) \cdot \ldots \cdot \Pr(X_n \le m) = (1 - e^{-m\lambda_1}) \cdot \ldots \cdot (1 - e^{-m\lambda_n})$
Arrival times: $S_n = \sum_{i=1}^n X_i, f_{S_n}(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}, s_n \sim Gamma(n, \lambda), E(S_n) = \frac{n}{\lambda}, V(S_n) = \frac{n}{\lambda^2}$
Little "oh": $f(h) = o(h)$ if $\lim_{h\to 0} \frac{f(h)}{h} = 0$. $f(h) = o(g(h))$ if $\lim_{h\to 0} \frac{f(h)}{g(h)} = 0$
If $f(h)$ and $g(h)$ are $o(h)$, then: $f(h) + g(h) = o(h), cf(h) = o(h)$. If $f(h) = o(1)$ then $f(h) \to 0$ as $h \to 0$.

Thinning

Let $\{N_t : t \ge 0\}$ be a PP with param λ. Each arrival, indep of other arrivals is marked as type k event with probability p_k, $k = 1, \ldots, n$ and $p_1 + \ldots + p_n = 1$. Let $N_t^{(k)}$ be # of type k events in $[0, t]$. Then $\left\{N_t^{(k)} : t \ge 0\right\}$ is a PP with λp_k and all $N_t^{(k)}$ are independent for different k.
Can separate PPs, e.g. PP of knee and ankle injuries can be separated into PP of knee injuries and PP of ankle injuries. superpositioning is the opposite, combining PPs. $N_t^{(1)}, \ldots, N_t^{(n)}$ are n indep PP with λ_i.
Then $N_t = N_t^{(1)} + \ldots + N_t^{(n)}$ is a PP with $\lambda = \lambda_1 + \ldots + \lambda_n$.

Uniform Distribution

Let $S_1, \ldots, S_n$ be arrival times of a PP(λ). Conditional on $N_t = n$, joint distrib of $(S_1, \ldots, S_n)$ is the same as distrib of order statistics of n iid $Unif[0, t]$ rvs, i.e. $\Pr(S_1, \ldots, S_n) = \frac{n!}{t^n}$ for $0 < S_1 < S_2 < \ldots < S_n < t$ i.e. distrib of $(U_{(1)}, U_{(2)}, \ldots, U_{(n)})$, where $U_i \overset{iid}{\sim} Unif[0, t]$
i.e. If we know that $N_5 = 1$, then S_1 is uniformly distributed over $[0, 5]$
$S_1 \mid N_t = 1 \sim Unif[0, t], S_2 \mid N_t = 1, N_s = 2 \sim Unif[t, s], t \ge s$
$E[S_2 \mid S_1] = S_1 + \frac{1}{\lambda}$ because of independent increments

Continuous Time Markov Chains (CTMC)

A continuous time stochastic process $\{X_t : t \ge 0\}$ with discrete state space S is a CTMC if
$\Pr(X_{t+s} = j \mid X_s = i, X_u = x_u) = \Pr(X_{t+s} = j \mid X_s = i), \forall s, t \ge 0, i, j, X_u \in S, 0 \le u < s$.
Time homogeneous iff $\Pr(X_{t+s} = j \mid X_s = i) = \Pr(X_t = j \mid X_0 = i), \forall s, t \ge 0$
Transition functions $\underline{P}(t)$ is a matrix function where $P_{ij}(t) = \Pr(X_t = j \mid X_0 = i)$.

Chapman-Kolmogorov Equations $\underline{P}(s + t) = \underline{P}(s)\underline{P}(t)$ i.e. $P_{ij}(s + t) = [\underline{P}(s)\underline{P}(t)]_{ij}, \forall i, j \in S, s, t \ge 0$
PP is a CTMC with:

$$\begin{array}{c|ccccc} & 0 & 1 & 2 & 3 & \ldots \\ \hline 0 & e^{-\lambda t} & \frac{e^{-\lambda t}\lambda t}{1!} & \ldots & & \\ 1 & 0 & e^{-\lambda t} & \frac{e^{-\lambda t}(\lambda t)^{2-1}}{(2-1)!} & & \\ 2 & 0 & 0 & e^{-\lambda t} & \ldots & \\ 3 & & & & & \\ \vdots & & & & & \end{array}$$

Holding time: Time spent in a state before transitioning. T_i = holding time for state i. For CTMC, T_i must always be exponentially distributed (exp is only memoryless continuous distribution).
$T_i \sim Exp(q_i)$, mostly $0 < q_i < \infty$. $q_i = 0 \implies i$ is absorbing. $q_i = \infty \implies i$ is explosive.
Embedded chain Let $Y_0, Y_1, \ldots$ be the r.v.s indicating the transition states. Y_j is the state transitions into on the j^{th} transaction. $Y_0 = X_0$. If $\{X_t : t \ge 0\}$ is a CTMC, then $\{Y_n : n = 0, 1, \ldots\}$ is a DTMC. $\tilde{P}$ be the tpm for $\{Y_n : n = 0, 1, \ldots\}$, then $\tilde{P}$ is a stochastic matrix but with diagonal elements all equal to 0.
$q_i = \sum_k q_{ik}, p_{ij} = \frac{q_{ij}}{q_i}$, where q_{ij} is alarm clock for i to j.
Alarm clock idea. Assume $X_t = i$. For $j \neq i$, set independnet alarm clock that goes off at a random time $\sim Exp(q_{ij})$ (rate version, $q_{ij} = \frac{1}{\mu}$). Chain transitions to state whose alarm goes off first. This is min of independent exp, so min has distrib $Exp\left(\sum_{j \neq i} q_{ij}\right)$.
$M = \min(\tilde{T}_1, \ldots, \tilde{T}_{i-1}, \tilde{T}_{i+1}, \ldots, \tilde{T}_k)$, where $\tilde{T}_i$ is random alarm time for state i.
Then $\Pr(M = \tilde{T}_\ell) = \frac{q_{i\ell}}{\sum_{j \neq i} q_{ij}} = \tilde{P}_{ij} = \Pr(Y_{n+1} = j \mid Y_n = i)$ with $\tilde{P}_{ii} = 0 \forall i$.

Generator Matrices

Matrix Q such that:

$$Q_{ij} = \begin{cases} P'_{ij}(0) = q_{ij} & i \neq j \\ -\sum_{j \neq i} Q_{ij} = -q_i & i = j \end{cases}$$

$q_{ij} = q_i p_{ij}$, where p_{ij} is i, j entry of embedded TPM.
Q is Not a stochastic matrix (rows sum to 0).

$$\underline{P}'(t) = \underline{P}(t)Q \text{ (forward Kolmogorov)}$$
$$\underline{P}'(t) = Q\underline{P}(t) \text{ (backward Kolmogorov)}$$

Or equivalently:

$$P'_{ij}(t) = \sum_k P_{ik}(t) q_{kj} = -P_{ij}(t) q_j + \sum_{k \neq j} P_{ik}(t) q_{kj}$$

$$P'_{ij}(t) = \sum_k q_{ik} P_{kj}(t) = -q_i P_{ij} + \sum_{k \neq i} q_{ik} P_{kj}(t)$$

Limiting Distribution π is a limiting distribution of a CTMC if $\forall i, j \in S \lim_{t\to\infty} P_{ij}(t) = \pi_j$ i.e. $\lim_{t\to\infty} \mathbf{P}(t) = \mathbf{\Pi}$, where $\mathbf{\Pi}$ is a matrix whose rows are π
π is a stationary distribution if:
$\pi = \pi P(t), \forall t \ge 0$ or $\pi_j = \sum_i \pi_i P_{ij}(t), \forall j, t \ge 0$
All CTMC aperiodic.
Finite irreducible CTMC has unique stationary distrib that is limiting distrib.
If π is a stationary distrib for $\{X_t : t \ge 0\}$, then $\pi Q = 0$ or $\sum_i \pi_i Q_{ij} = 0, \forall j$. To solve this you can use $\vec{x} = (1, x_2, \ldots, x_k)$ with $\vec{x}Q = 0$ and $\vec{\pi} = \left(\frac{1}{1 + x_2 + \ldots + x_k}\right)\vec{x}$ again.

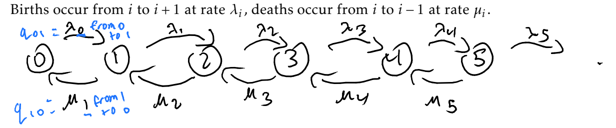
Time to Absorption

$$\text{Canonical } Q = \begin{array}{c|cc} & a & T \\ \hline a & 0 & 0 \\ T & * & V \end{array}$$

Expected time till absorption given that we started in i can be obtained from $a_i = \sum_j F_{ij}$, where F is the fundamental matrix $F = -V^{-1}$

Birth-Death Process

Births occur from i to $i + 1$ at rate λ_i, deaths occur from i to $i - 1$ at rate μ_i.



$$Q = \begin{array}{c|ccccc} & 0 & 1 & 2 & 3 & \ldots \\ \hline 0 & -\lambda_0 & \lambda_0 & 0 & 0 & \\ 1 & \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \\ 2 & 0 & \mu_2 & -(\mu_2 + \lambda_2) & \lambda_2 & 0 \\ 3 & 0 & 0 & \mu_3 & -(\mu_3 + \lambda_3) & \lambda_3 \\ \vdots & & & & & \end{array}$$

$$\pi_0 = \frac{1}{\sum_{k=0}^{\infty} \prod_{i=1}^k \frac{\lambda_{i-1}}{\mu_i}}, \pi_k = \pi_0 \prod_{i=1}^k \frac{\lambda_{i-1}}{\mu_i}$$

Under the condition that $\sum_{k=0}^{\infty} \prod_{i=1}^k \frac{\lambda_{i-1}}{\mu_i} < \infty$
$PP(\lambda) \implies$ rate λ. 1 per α min $\implies$ rate $\frac{1}{\alpha}$