1 Probability preliminary

Binomial(n,p): $p(x) = \binom{n}{x}p^x(1-p)^{n-x}, \phi(t) = (pe^t+(1-p))^n, \mu = np, \sigma^2 = np(1-p)$

Poisson(λ): $p(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \phi(t) = e^{\lambda(e^t-1)}, \mu = \lambda, \sigma^2 = \lambda$

Geometric(p): $p(x) = p(1-p)^{x-1}, \phi(t) = \frac{pe^t}{1-(1-p)e^t}, \mu = \frac{1}{p}, \sigma^2 = \frac{1-p}{p^2}$

Uniform(a,b): $f(x) = \frac{1}{b-a}, x \in (a,b), f(x) = 0, x \notin (a,b), \phi(t) = \frac{e^{tb}-e^{ta}}{t(b-a)}, \mu = \frac{a+b}{2}, \sigma^2 = \frac{(b-a)^2}{12}$

Exponential(λ): $f(x) = \lambda e^{-\lambda x}, x > 0, f(x) = 0, x < 0, \phi(t) = \frac{\lambda}{\lambda - t}, \mu = \frac{1}{\lambda}, \sigma^2 = \frac{1}{\lambda^2}, F(x) = 1-e^{-\lambda x}$

Gamma(n,λ): $f(x) = \frac{\lambda e^{-\lambda x}(\lambda x)^{n-1}}{(n-1)!}, x \geq 0, f(x) = 0, x < 0, \phi(t) = (\frac{\lambda}{\lambda-t})^n, \mu = \frac{n}{\lambda}, \sigma^2 = \frac{n}{\lambda^2}$

Normal(μ,σ^2): $f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \phi(t) = e^{\mu t+\frac{\sigma^2t^2}{2}}, \mu = \mu, \sigma^2 = \sigma^2$

Conditional probability: $P(E|F) = \frac{P(EF)}{P(F)}, f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$

Expectation: $E[X] = \int_{-\infty}^{+\infty} xf(x)dx$

Variance: $Var(X) = E[(X-E[X])^2] = E[X^2] - E[X]^2$

Total probability: $p_X(x) = \sum_y p_{X|Y}(x|y)P_Y(y), f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dy$

Total expectation : $E[X] = E[E[X|Y]]$

Total variance: $Var(X) = E[Var(X|Y)] + Var(E[X|Y])$

Independence: $P(EF) = P(E)P(F) \iff P(E|F) = P(E)$

Bayes formula: Let $\{F_i\}_{i=1}^n$ be mutually exclusive events that forms a union of sample space S, then $E = \cup_{i=1}^n EF_i$, we have $P(E) = \sum_{i=1}^n P(E|F_i)P(F_i)$, thus $P(F_j|E) = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^n P(E|F_i)P(F_i)}$

Covariance: $Cov(X,Y) = E[(X-E[X])(Y-E[Y])] = E[XY]-E[X]E[Y]$. We have $Cov(X,X) = Var(X), Cov(cX,Y) = cCov(X,Y), Cov(X,Y+Z) = Cov(X,Y)+Cov(X,Z)$

MGF: $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$ for independent X,Y. $E[X^k] = \frac{d^k}{dx^k}M_X(t)|_{t=0}, M_{aX+b}(t) = e^{bt}M_X(at)$

Theorem 1 (Markov's inequality)**.** If X is a nonnegative random variable, for $a > 0$, we have $P(X \geq a) \leq \frac{E[X]}{a}$.

Theorem 2 (Cheybeshev's inequality)**.** If X is a random variable with mean μ and variance σ^2, then for $k > 0$, we have $P(|X-\mu| \geq k) \leq \frac{\sigma^2}{k^2}$

Theorem 3 (Strong law of large numbers)**.** Let (X_i) be a sequence of independent random variables having identical distribution, and let $E[X_i] = \mu$, then with probability 1, $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n X_i = \mu$

Theorem 4 (Central limit theorem)**.** Let (X_i) be a sequence of independent random variables having identical distribution, each with mean μ and variance σ^2, then $\lim_{n\to\infty}P(\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \leq a) = \frac{1}{2\pi}\int_{-\infty}^a e^{-\frac{x^2}{2}}dx$. Alternatively, we could directly approximate $\sum_{i=1}^n X_i$ by $N(n\mu, n\sigma^2)$.

2 Poisson process

2.1 Poisson distribution

If $X \sim Pois(\lambda)$ and $Y \sim Pois(\mu)$ are independent, then $X+Y \sim Pois(\lambda+\mu)$.
If $X \sim Pois(\lambda)$ and $Z|X \sim Binom(X,r)$, then $Z \sim Pois(\lambda r)$.

2.2 Definition by Poisson distribution

A Poisson process with rate(intensity) $\lambda > 0$ is an integer-valued stochastic process $\{X(t), t \geq 0\}$ for which for any time points $t_0 = 0 < t_1 < t_2 < \cdots < t_n$, the process increments $X_{t_i} - X_{t_{i-1}}, i = 1,2,\ldots,n$ are independent random variables; for $s \geq 0, t > 0$, we have $X(s+t) - X(s) \sim Pois(\lambda t)$; $X(0) = 0$.

2.3 Definition by rare events

Let ϵ_h be the total occurrences of an event within a time period h, we call this event an rare event if when $h \to 0$, we have $P(\epsilon_h = 0) = 1-\lambda h + o(h), P(\epsilon_h = 1) = \lambda h + o(h), P(\epsilon_h \geq 2) = o(h)$.

Theorem 5 (Law of rare events)**.** Let $\{\epsilon_i\}_{1\leq i\leq n}$ be independent Bernouli random variables where $P(\epsilon_i = 1) = p_i$. Let $S_n = \sum_{i=1}^n \epsilon_i$. The exact probability for S_n and Poisson probability with $\lambda = \sum_{i=1}^n p_i$ differ by at most $|P(S_n = k) - \frac{\lambda^k e^{-\lambda}}{k!}| \leq \sum_{i=1}^n p_i^2$.

Let $N((s,t])$ be a random variable counting the number of events occuring in the interval $(s,t]$, then $N((s,t])$ is a Poisson process of rate λ if: for any time $t_0 = 0 < t_1 < \cdots < t_n$, the process increments $N((t_{i-1},t_i])$ are independent; there is a positive constant λ such that the probability of at least one event happening in a time interval of length h is $P(N((t,t+h]) \geq 1) = \lambda h + o(h), h \to 0$; the probability of at least two events happening in a time interval of length h is $(N((t,t+h]) \geq 2) = o(h), h \to 0$.
The Poisson process is **nonhomogeneous** if the rate λ is not a constant but rather a time-dependent function $\lambda(t)$. Same definition follows. In this case, the increment $X(s+t) - X(s) \sim Pois(\int_s^{s+t}\lambda(u)du)$.

If we define $\Lambda(t) = \int_0^t \lambda(u)du$, and define $Y(s) = X(t)$ where $s = \Lambda(t)$, then $\{Y(s)\}_{s>0}$ is a Poisson process with rate $\lambda = 1$.

2.4 Definition by counting

Definition: Let $X(t)$ be a Poisson process with rate λ. Let W_n be the time of occurrence of n-th event. It is called the waiting time of n-th event. We set $W_0 = 0$. The time between two occurrences $S_n = W_{n+1} - W_n$ is called sojourn time, which measures the duration that the Poisson process stays in state n.
We know that $W_1 \sim Exp(\lambda)$. In general, W_n follows gamma distribution. We have PDF as $f_{W_n}(t) = \frac{\lambda^n t^{n-1}}{(n-1)!}e^{-\lambda t}$. In particular $f_{W_1}(t) = \lambda e^{-\lambda t}$. Recall that exponential distribution is memoryless, thus $W_1 - t|W_1 > t \sim Exp(\lambda)$. As a consequence, all sojourn time $S_i \sim Exp(\lambda)$.
Suppose $\{S_n : n \geq 0\}$ is a set of independent identical exponential random variable with parameter λ. Define a counting process by saying that the i-th event occurs at time $W_i = \sum_{j=0}^{i-1} S_j$. The resultant counting process will be a Poisson process with rate λ. This definition fails for nonhomogeneous case.

2.5 Properties of Poisson process

Suppose we know that $X(t) = 1$, by Bayes' formula, we have the conditional distribution of time of occurrence $f_{W_1|X(t)=1}(s) = \frac{1}{t}$. In general, given $X(t) = n$, the joint distribution of time of occurrence $W_1,\ldots,W_n$ we have $f_{W_1,\ldots,W_n|X(t)=n}(s) = \frac{n!}{t^n}$, which is the joint distribution of the ordered statistics of n independent uniform random variables over $(0,t)$.

Theorem 6. Given that $X(t) = n$, the n arrival/waiting times $W_1,\ldots,W_n$ have the same distribution as the order statistics corresponding to n independent random variables uniformly distributed on the interval $(0,t)$, which evaluates to $f_k(x) = \frac{n!}{(n-k)!(k-1)!}(\frac{x}{t})^{k-1}\frac{1}{t}(\frac{t-x}{t})^{n-k}$.

Theorem 7. Let $\{N_1(t) : t \geq 0\},\ldots,\{N_m(t) : t \geq 0\}$ be independent Poisson processes with rate λ_i respectively. Let $N(t) = \sum_{i=1}^m N_i(t), t \geq 0$, then $N(t)$ is a Poisson process with rate $\lambda = \sum_{i=1}^m \lambda_i$.

Theorem 8. Consider $\{N(t) : t \geq 0\}$ with rate λ and for each event having independent and identical distribution that this event is a type i event with probability p_i, then the processes $N_i(t)$ are all independent Poisson process with rate λp_i respectively.

Theorem 9. Let $X(t), Y(t)$ be two independent Poisson processes with rate λ_1, λ_2. Let W_n^1 denote the waiting time of n-th event of $X(t)$. Let W_m^2 denote the waiting time of m-th event of $Y(t)$. We have $P(W_n^1 < W_m^2) = \sum_{k=n}^{n+m-1}\binom{n+m-1}{k}(\frac{\lambda_1}{\lambda_1+\lambda_2})^k(\frac{\lambda_2}{\lambda_1+\lambda_2})^{n-m-1-k}$. In particular, when $m = n = 1$, $P(W_1^1 < W_1^2) = \frac{\lambda_1}{\lambda_1+\lambda_2}$.

2.6 Nonhomogeneous Poisson Process

In this case, $P(W_1 > t) = P(X(t) = 0) = e^{-\int_0^t \lambda(u)du}$, hence the density function is $f_{W_1}(t) = \lambda(t)e^{-\int_0^t \lambda(u)du}$, the conditional distribution is $P(W_1 < s|X(t) = 1) = \frac{\int_0^s \lambda(u)du}{\int_0^t \lambda(u)du}$.
We still have merging theorem.

Theorem 10. Let $N(t)$ be a nonhomogeneous Poisson process with rate $\lambda(t)$. Suppose for event at any point t, independent of what have occurred before t, the event was from N_k with probability $p_k(t)$, then each $N_k(t)$ is an independent nonhomogeneous Poisson process with rate $\lambda(t)p_k(t)$ respectively.

2.7 Compound Poisson process

Definition: A stochastic process $\{X(t) : t \geq 0\}$ is a compound Poisson process if it can be represented as $X(t) = \sum_{i=1}^{N(t)} Y_i$ where $N(t)$ is a Poisson process with rate λ, and Y_i follows identical and independent distribution of F.
We have $E[X(t)] = \lambda t E[Y]$. $Var(X(t)) = \lambda t(E[Y]^2 + Var(Y))$
If $X(t), Y(t)$ are two independent compound Poisson process with parameters (λ_1, F_1) and (λ_2, F_2) respectively, then $N(t) = X(t) + Y(t)$ is still a compound Poisson process with parameter $(\lambda_1+\lambda_2, \frac{\lambda_1}{\lambda_1+\lambda_2}F_1 + \frac{\lambda_2}{\lambda_1+\lambda_2}F_2)$.

2.8 Conditional Poisson process

Definition: Let $N(t)$ be a counting process defined as follows: (1)There is a positive random variable L with density function g. (2)Condition on $L = \lambda$, the counting process is a Poisson process with rate λ. Such a process is called a conditional Poisson process. This process still satisfies independent increments, but no longer a Poisson process. The distribution is $P(N(t+s) - N(s) = n) = \int_0^\infty e^{-\lambda t}\frac{(\lambda t)^n}{n!}g(\lambda)d\lambda$.
We have $E[N(t)] = E[L]t$. $Var(N(t)) = tE[L] + t^2Var(L)$.
Condition on $N(t) = n$, the updated distribution of L is $P(L \leq x|N(t) = n) = \frac{\int_0^x e^{-\lambda t}(\lambda t)^n g(\lambda)d\lambda}{\int_0^{+\infty} e^{-ut}(ut)^n g(u)du}$, where the updated PDF is $f_{L|N}(\lambda|n) = \frac{e^{-\lambda t}(\lambda t)^n g(\lambda)}{\int_0^{+\infty} e^{-ut}(ut)^n g(u)du}$, thus the posterior estimation of number of events on the following time interval will be $P(N(t+s) - N(t) = m|N(t) = n) = \int_0^{+\infty} e^{-\lambda s}\frac{(\lambda s)^m}{m!}f_{L|N}(\lambda|n)d\lambda = \int_0^{+\infty} e^{-\lambda s}\frac{(\lambda s)^m}{m!}\frac{e^{-\lambda t}(\lambda t)^n g(\lambda)}{\int_0^{+\infty} e^{-ut}(ut)^n g(u)du}d\lambda$

2.9 Multi-dimensional Poisson process

Definition: Let S be a subset of $\mathbb{R}^n$. Let $\mathcal{A}$ be the power set of S and for any set $A \in \mathcal{A}$, let $|A|$ denote the size of A, then $\{N(A) : A \in \mathcal{A}\}$ is a Poisson process with $\lambda > 0$ if: for each $A \in \mathcal{A}$, the random variable $N(A)$ has a Poisson distribution with parameter $\lambda|A|$; for every finite and disjoint collection of subsets $\{A_i\}$, the random variables $N(A_i)$ are independent.

3 Continuous time Markov chain

3.1 Specification

Definition: For a stochastic process $X(t)$, if for $s > u \geq 0, t \geq 0$, we have $P(X(s+t) = j|X(s) = i, X(u) = k) = P(X(s+t) = j|X(s) = i)$ then we call the stochastic process satisfies Markovian property and is a continuous time Markov chain. We assume stationary, which implies that $P_{ij}(t,s) = P(X(t+s) = j|X(s) = i) = P(X(t) = j|X(0) = i) = P_{ij}(t)$.
To define a continuous time Markov chain, we need to define discrete state space S, continuous time space $T = t : t \geq 0$ and transition probability function matrix $P(t)$. $P(t)$ is defined such that the ij-entry is $P_{ij}(t)$.
$P(t)$ should have row sum 1 for all $t \in T$. By Markovian property, $P(t+s) = P(t)P(s) = P(s)P(t)$, thus we have $P_{ij}(t+s) = \sum_{k\in S} P_{ik}(t)P_{kj}(s)$, which is called the general form of Chapman-Kolmogorov equation.
We could discretize the continuous time Markov chain by defining equally spaced time points $t_k = kh$, and define $Y_n = X(t_n)$, then $\{Y_n\}_{n>0}$ is a stationary discrete time Markov chain. Y_n satisfies Markovian property, having state space S and transition probability matrix is $P(h)$. The Chapman-Kolmogorov equation of Y_n is a special case of that of $X(t)$. If one state is positively recurrent in Y_n, then in long run it will also be frequently visited in $X(t)$. If one

state in Y_n is absorbing, then it is also absorbing in $X(t)$.
The waiting time for any state i follows exponential distribution. The jump probability P_{ij} is a constant probability that only depends on i, j without dependence on time. Therefore, we could also specify a continuous time Markov chain by the state space S, the vector $v = (v_1, v_2, \dots)$ that contain the parameter of the waiting time distribution at state i, and P, where P_{ij} is the probability that the process jumps from state i to state j at the first transition. If i is absorbing, we define $P_{ii} = 1$. Otherwise, $P_{ii} = 0, \sum_{j\in S} P_{ij} = 1$.
Definition: For a continuous time Markov chain $X(t)$, if we consider the sequence of states visited, ignoring the amount of time spent in each state, then the corresponding sequence constitutes a discrete time Markov chain, we call this chain the embedded chain.
For the embedded chain, the state space remains the same, and the transition probability is simply P_{ij}. The embedded chain is different from discretized chain.

3.2 Infinitesimal generator

Let $X(t) = i$, consider a small interval $(t, t+h)$. By exponential distribution, we have $P(\text{no jump}) = e^{-v_i h} = 1 - v_i h + o(h)$. $P(\text{At least one jump}) = 1 - e^{-v_i h} = v_i h + o(h)$. $P(\text{At least two jumps}) = o(h)$.
Definition: For any pair of states i, j define $q_{ij} = v_i P_{ij}$, then it is called the instantaneous transition rate.
By the definition, q_{ij} is determined by v and P. Note that $\sum_{j\in S} P_{ij} = 1$, thus $v_i = \sum_{j\in S} v_i P_{ij} = \sum_{j\in S} q_{ij}, P_{ij} = \frac{q_{ij}}{v_i} = \frac{q_{ij}}{\sum_{j\in S} q_{ij}}$, so we could also determine v and P given q_{ij}.

Theorem 11. For a continuous-time Markov chain, $\lim_{h\to 0} \frac{P_{ii}(h)-P_{ii}(0)}{h} = \lim_{h\to 0} \frac{P_{ii}(h)-1}{h} = -v_i$. $\lim_{h\to 0} \frac{P_{ij}(h)-P_{ij}(0)}{h} = \lim_{h\to 0} \frac{P_{ij}(h)}{h} = q_{ij}$.

Therefore, we have $\frac{dP_{ij}(t)}{dt}|_{t=0} = q_{ij}, i \neq j$, or $-v_i, i = j$.
Definition: The matrix G is called the infinitesimal generator where $G_{ii} = -v_i, G_{ij} = q_{ij}$

3.3 Pure birth process

Definition: Consider a sequence of positive numbers $\{\lambda_0, \lambda_1, \dots\}$. A pure birth process $X(t)$ is a Markov chain where the possible values are non-negative integers, and satisfies the following postulates: $P(X(t+h) - X(t) = 1|X(t) = k) = \lambda_k h + o(h); P(X(t+h) - X(t) = 0|X(t) = k) = 1 - \lambda_k(h) + o(h); P(X(t+h) - X(t) < 0) = 0, h \to 0$.

3.4 Birth and death process

Definition: Consider a sequence of positive numbers $\{\lambda_0, \lambda_1, \dots\}$ and $\{\mu 0, \mu 1, \dots\}$. A birth and death process $X(t)$ is a Markov process where the possible values are non-negative integers and satisfies the following postulates: $P(X(t+h) - X(t) = 1|X(t) = k) = \lambda_k h + o(h), i \geq 0, h \to 0; P(X(t+h) - X(t) = -1|X(t) = k) = \mu_k h + o(h), i \geq 1, h \to 0; P(X(t+h) - X(t) = 0|X(t) = k) = 1 - (\lambda_k + \mu_k)h + o(h), i \geq 0, h \to 0$
We consider a queuing system when customers arrive with a Poisson process with rate λ, and the single server has service time with exponential distribution having parameter μ. If we let $X(t)$ denote the number of customers in the system at time t, then it is a birth and death process with constant birth rate λ and constant death rate μ.
If we consider the system having s servers instead, then the process is still a birth and death process. The birth rate is still constant at λ, while the death rate μ_n is $n\mu$ when $n \leq s$, and $s\mu$ when $n > s$.

3.5 Kolmogorov equation

We have Chapman-Kolmogorov equation: $P(t+s) = P(t)P(s) = P(s)P(t)$, $P_{ij}(t+s) = \sum_{k\in S} P_{ik}(t)P_{kj}(s)$. By considering $P_{ij}(t+h) = \sum_{k\in S} P_{ik}(h)P_{kj}(t)$ and further considering $\lim_{h\to 0} \frac{P_{ij}(t+h)-P_{ij}(t)}{h}$, we obtain Kolmogorov's backward equation. If we consider $P_{ij}(t+h) = \sum_{k\in S} P_{ik}(t)P_{kj}(h)$, we obtain Kolmogorov's forward equations.

Theorem 12 (Kolmogorov's backward equations)**.** For all states i, j and time $t \geq 0$, we have $P'_{ij}(t) = \sum_{k\neq i} q_{ik}P_{kj}(t) - v_i P_{ij}(t) \iff P'(t) = GP(t)$

Theorem 13 (Kolmogorov's forward equations)**.** For all states i, j and time $t \geq 0$, we have $P'_{ij}(t) = \sum_{k\neq j} q_{kj}P_{ik}(t) - v_j P_{ij}(t) \iff P'(t) = P(t)G$

3.6 Uniformization

Suppose a continuous time Markov chain has same stay time distribution for all states, i.e.$v_i = \lambda$ for all $i \in S$. Let $N(t)$ be the number of jumps till time t, then it is a Poisson process with rate λ. Therefore, we could compute transition probability as

$$
\begin{aligned}
P_{ij}(t) &= \sum_{n=0}^{\infty} P(n \text{ jumps in}(0,t]) \\
&\quad * P(\text{jump from } i \text{ to } j \text{ by } n \text{ jumps}|n \text{ jumps in } (0,t]) \\
&= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t}[P^n]_{ij}
\end{aligned}
$$

If we truncate the first k terms as a numerical approximation, then the error is $\sum_{n=k}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t}[P^n]_{ij} \leq \sum_{n=k}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} = P(N(t) \geq k)$
For the case of general MC, where v_i is dependent on i, we perform uniformization. We add fake jumps to increase jumping rate up to the supremum of all existing jumping rates.
Suppose v is an upper bound of $\{v_i : i \in S\}$, we modify the jump probability as $P^*_{ij} = \frac{v_i}{v} P_{ij}$ if $i \neq j$ and $\frac{v-v_i}{v}$ if $i = j$.

Theorem 14. For a continuous time Markov chain $X(t)$ with rates v_i, if $v_i \leq v$, then $P_{ij}(t) = \sum_{n=0}^{\infty} \frac{(vt)^n}{n!} e^{-vt}[P^*]^n_{(ij)}$