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

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

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

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:[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 $X+Y\sim Pois(\lambda+\mu)$.

2.2 Definitions

A Poisson process with rate(intensity) $\lambda>0$ is an integer-valued stochastic process $\{X(t), t\geq 0\}$ for which there are independent increments; for $s\geq 0, t>0$, we have $X(s+t)-X(s)\sim Pois(\lambda t)$; $X(0)=0$.

Let $N((s,t])$ be a random variable counting the number of events occurring in the interval $(s,t]$, then $N((s,t])$ is a Poisson process of rate λ if: increments are independent; there is a positive constant λ such that $P(N((t,t+h])\geq 1)=\lambda h+o(h), h\to 0$; $(N((t,t+h])\geq 2)=o(h), h\to 0$.

Definition: Let $X(t)$ be a Poisson process with rate λ. Let W_n be the time of occurrence of n-th event. We set $W_0=0$. Define $S_n=W_{n+1}-W_n$.

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 $f_{W_1}(t)=\lambda e^{-\lambda t}$. All sojourn time $S_i\sim Exp(\lambda)$. Define Poisson process as interarrival time having independent exponential distribution with rate λ.

2.3 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 $W_1,\dots,W_n$ we have $f_{W_1,\dots,W_n|X(t)=n}(s)=\frac{n!}{t^n}$.

Theorem: Given that $X(t)=n$, the marginal distribution of n arrival/waiting times $W_1,\dots,W_n$ evaluates to $f_{W_k}(x)=\frac{n!}{(n-k)!(k-1)!}(\frac{x}{t})^{k-1}\frac{1}{t}(\frac{t-x}{t})^{n-k}$.

Theorem: Let $\{N_1(t):t\geq 0\},\dots,\{N_m(t):t\geq 0\}$ 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: 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: 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, $P(W_1^1<W_1^2)=\frac{\lambda_1}{\lambda_1+\lambda_2}$.

2.4 Variants

The Poisson process is **nonhomogeneous** if the rate λ is $\lambda(t)$. Same definition follows. In this case, the increment $X(s+t)-X(s)\sim Pois(\int_s^{s+t}\lambda(u)du)$. 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 and spliting theorem.

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 $E[X(t)]=\lambda tE[Y]$. $Var(X(t))=\lambda t(E[Y]^2+Var(Y))$

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 $(\lambda_1+\lambda_2, \frac{\lambda_1}{\lambda_1+\lambda_2}F_1+\frac{\lambda_2}{\lambda_1+\lambda_2}F_2)$.

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. 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 $P(N(t+s)-N(t)=m|N(t)=n)=\int_0^{+\infty}e^{-\lambda s}\frac{(\lambda s)^m}{m!}f_{L|N}(\lambda|n)d\lambda=\int_0^{+\infty}e^{-\lambda s}\frac{(\lambda s)^m}{m!}\frac{e^{-\lambda t}(\lambda t)^n g(\lambda)}{\int_0^{+\infty}e^{-ut}(ut)^n g(u)du}d\lambda$

3 Continuous time Markov chain

3.1 Specification

Definition: A stochastic process $X(t)$ such that for $s>u\geq 0, t\geq 0$, we have $P(X(s+t)=j|X(s)=i, X(u)=k)=P(X(s+t)=j|X(s)=i)$. We assume stationary increment, which implies that $P(X(t+s)=j|X(s)=i)=P(X(t)=j|X(0)=i)$.

To define a continuous time Markov chain, we need to define discrete state space S, time space $t\geq 0$ 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 $P_{ij}(t+s)=\sum_{k\in S}P_{ik}(t)P_{kj}(s)$. 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 $P(h)$. 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 ignore 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}.

3.2 Infinitesimal generator

Let $X(t)=i$, consider a small interval $(t,t+h)$. By exponential distribution, we have $P(\text{no jump})=e^{-v_ih}=1-v_ih+o(h)$. $P(\text{At least one jump})=1-e^{-v_ih}=v_ih+o(h)$. $P(\text{At least two jumps})=o(h)$.

Definition: For any pair of states i,j define $q_{ij}=v_iP_{ij}$ as instantaneous transition rate.

By the definition, q_{ij} is determined by v and P. Note that $\sum_{j\in S}P_{ij}=1$, thus $v_i=\sum_{j\in S}v_iP_{ij}=\sum_{j\in S}q_{ij}, P_{ij}=\frac{q_{ij}}{v_i}=\frac{q_{ij}}{\sum_{j\in S}q_{ij}}$, so we could also determine v and P given q_{ij}.

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: $P(X(t+h)-X(t)=1|X(t)=k)=\lambda_kh+o(h); P(X(t+h)-X(t)=0|X(t)=k)=1-\lambda_k(h)+o(h); P(X(t+h)-X(t)<0)=0, h\to 0$.

3.4 Birth and death process

Definition: Consider a sequence of positive numbers $\{\lambda_0,\lambda_1,\dots\}$ and $\{\mu0,\mu1,\dots\}$. A birth and death process $X(t)$ is a Markov process where the possible values are non-negative integers and satisfies the following postulates: $P(X(t+h)-X(t)=1|X(t)=k)=\lambda_kh+o(h), i\geq 0, h\to 0; P(X(t+h)-X(t)=-1|X(t)=k)=\mu_kh+o(h), i\geq 1, h\to 0; P(X(t+h)-X(t)=0|X(t)=k)=1-(\lambda_k+\mu_k)h+o(h), i\geq 0, h\to 0$

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

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

3.5 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

$P_{ij}(t)=\sum_{n=0}^\infty\frac{(\lambda t)^n}{n!}e^{-\lambda t}[P^n]_{ij}$

If we truncate the first k terms as a numerical approximation, then the error is $\sum_{n=k}^\infty\frac{(\lambda t)^n}{n!}e^{-\lambda t}[P^n]_{ij}\leq\sum_{n=k}^\infty\frac{(\lambda t)^n}{n!}e^{-\lambda t}=P(N(t)\geq k)$

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

3.6 Absorbing state

Definition: For a continuous time Markov chain, if there is a state i such that for any $t>0, s\geq 0$, we have $P(X(t+s)=i|X(s)=i)\iff P_{ii}=1$, then it is an absorbing state.

Suppose 0 is an absorbing state, and we are interested in absorbing probability u_i starting from state i. We have that $u_i=\lim_{t\to\infty}P(X(t)=0|X(0)=i)$. If we consider the embedded chain Z_n, then 0 is also an absorbing state. We denote $v_i=\lim_{n\to\infty}P(Z_n=0|Z_0=i)$ and we could solve v_i by first step analysis. Since time till absorption does not matter, we have $u_i=v_i$.

If we let w_i denote the expected time until absorption when $X(0)=i$, then by first step analysis. $w_i=\mathbb{E}[\text{time to make first jump}]+\sum_{j\in S, j\neq i}P_{ij}w_j$.

3.7 Stationary distribution

Definition: A stationary distribution for a continuous time Markov chain is a distribution such that $\pi=\pi P(t), t\geq 0$.

Theorem:[Global balance equation] For a continuous time Markov chain $\{X(t)\}$ with the infinitesimal generator G, a distribution π is stationary if and only if $\pi G=0$. It can be written as $\sum_{j\neq i}\pi_iq_{ij}=v_j\pi_j$

for any $j \in S$. Intuitively, it suggests that for a distribution to be stationary, the long run rate into state j should be equal to the long run rate out of state j.

3.8 Limiting distribution

Definition: For a continuous time Markov chain $\{X(t)\}$ with state space S and transition probability matrix $P(t)$, and let $P_j = \lim_{t\to\infty} P_{ij}(t)$ for $j \in S$, then $\{P_j\}_{j\in S}$ is called the limiting distribution of $\{X(t)\}$.

Note that if a limiting distribution exists, it must be stationary. If we know that a limiting distribution exists, then finding stationary distribution is equivalent to finding limiting distribution. A sufficient condition for ergodic chain is that the chain only has one communication class, and is positive recurrent.

Suppose π is the limiting distribution for the continuous time chain. Suppose the embedded chain has a limiting distribution ψ which satisfies $\psi = \psi P$, then $\pi G = 0 \implies \pi_j v_j = \sum_i \pi_i q_{ij} = \sum_i \pi_i \frac{v_i q_{ij}}{v_i} = \sum_i \pi_i v_i P_{ij}$. Note that ψ_i, π_i, v_i are constants, thus we could fix constant C such that $\psi_i = C\pi_i v_i$ and hence $\psi_j = C\pi_j v_j = C\sum_i \pi_i v_i P_{ij} = \sum_i \psi_i P_{ij}$. By the fact that $\sum_i \psi_i = \sum_i C\pi_i v_i = 1$, we have $C = \frac{1}{\sum_i \pi_i v_i}$, and hence $\psi_i = \frac{\pi_i v_i}{\sum_j \pi_j v_j}$. Similarly, if ψ is the limiting distribution of the embedded chain, we have $\pi_i = \frac{\psi_i/v_i}{\sum_j \psi_i/v_i}$. Note that the existence of either ψ or π is not implied.

3.9 Time reversibility

As $t \to \infty$, $\{X(t)\}$ achieves a limiting distribution. We then set up $\{Y(t)\}$ by $Y(s) = X(t-s)$ for $s \in [0,t]$. Assuming t is large, so $\{Y(t)\}$ is long enough. The staying time in state i for $\{Y(t)\}$ also follows $Exp(v_i)$. Let the jump probability matrix for $\{Y(t)\}$ be Q, we have $Q_{ij} = \frac{\pi_j v_j P_{ji}}{\pi_i v_i}$. If $Q = P$, then the probability structure for $\{X(t)\}, \{Y(t)\}$ are the same.

Definition: We call a continuous time Markov chain time reversible in the sense that the process reversed in time has the same probabilistic structure as the original process.

Hence, we must have $Q_{ij} = P_{ij} \iff \frac{\pi_j v_j P_{ji}}{\pi_i v_i} = P_{ij} \iff \pi_j v_j P_{ji} = \pi_i v_i P_{ij} \iff \pi_j q_{ji} = \pi_i q_{ij}$ for $i, j \in S$. The set of equations is called local balance equations. Intuitively, it implies that the rate from state i to j is equal to the rate from state j to i. Note that if we could find π that satisfies the local balance equation, then π also satisfies the global balance equation, and thus is the limiting distribution.

Theorem: A time reversible chain with limiting probability π that is truncated to $A \subseteq S$ and remains irreducible is also time reversible and has the limiting probability π^A defined by $\pi^A_j = \frac{\pi_j}{\sum_{i\in A}\pi_i}$.

Theorem: If $\{X_i(t)\}_i$ are independent and time reversible continuous time Markov chain, then the vector process $(X_1(t), \dots, X_n(t))$ is also time reversible.

4 Martingale

4.1 Discrete time

Definition: Let $\{X_n\}$ be a discrete time stochastic process. We call it a martingale if $\mathbb{E}[|X_n|] < \infty$ and $\mathbb{E}[X_{n+1}|X_n, X_{n-1}, \dots, X_0] = X_n$.

Therefore, a Markov chain is a martingale if and only if $\mathbb{E}[X_{n+1}|X_n] = X_n$.

Recall that a stopping time T is a variable where the event $\{T = n\}$ depends on $X_0, \dots, X_n$ only.

Theorem:[Optional stopping theorem] Suppose $\{X_n\}$ is a martingale and T is a stopping time. Assume $P(T < \infty) = 1$ and $X_{\min(n,T)}$ is uniformly integrable, then $\mathbb{E}[X_T] = \mathbb{E}[X_0]$. One sufficient condition for uniform integrability is that $\mathbb{E}[|X_T|] < \infty$ and $\mathbb{E}[X_n|T > n]P(T > n) \to 0$ as $n \to \infty$.

4.2 General definition

Consider a measure space $(\Omega, \mathcal{F}, P)$, define the division $A_y = Y^{-1}(y) = \{\omega \in \Omega : Y(\omega) = y\}$. Then the σ-field generated by A_y denoted by $\mathcal{F}_y$ contains countable unions and intersections of elements in A_y. We have $\mathbb{E}[X|Y] = \mathbb{E}[X|\mathcal{F}_y]$.

Definition: Consider a series of σ-fields $\mathcal{F}_i$ on Ω, such a series is called a filtration if $\mathcal{F}_i \subset \mathcal{F}_{i+1}$. In particular, define $\mathcal{F}_n$ as the σ-field formed by $X_1, \dots, X_n$ as the canonical filtration.

Definition: A martingale is an ordered pair $(X_n, \mathcal{F}_n)$ where $\{\mathcal{F}_n\}$ is a filtration and $\{X_n\}$ is a stochastic process adapted to $\{\mathcal{F}_n\}$, where $\mathbb{E}[|X_n|] < \infty$ and $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[X_n|\mathcal{F}_n] = X_n$.

If we define $D_0 = 0$ and $D_n = X_n - X_{n-1}$, then $\mathbb{E}[D_{n+1}|\mathcal{F}_n] = X_n - X_n = 0$. For Markov chain, it is equivalent to $\mathbb{E}[D_{n+1}|X_n] = 0$. Consider a martingale $\{(X_n, \mathcal{F}_n)\}$ and a stochastic process $\{V_n\}$ where V_n is measurable on $\mathcal{F}_{n-1}$, then define $Y_n = \sum_{k=1}^n V_k D_k$, called the martingale transform of V with respect to X. If $\mathbb{E}[|V_n|^1]$ and $\mathbb{E}[|X_n|^p]$ exists for some $1 \le p, q < \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$, then Y_n is a martingale.

Definition: The random variable T is a stopping time for the filtration $\mathcal{F}_n$ if $\{T = n\} \in \mathcal{F}_n$ for all $n \ge 1$.

4.3 Continuous time

Definition: The ordered pair $(X(t), \mathcal{F}(t))$ where $t \in \mathbb{R}^{\ge 0}$ is a continuous time martingale if the σ-fields $\mathcal{F}_t$ form a filtration, and $X(t)$ is integrable and adapted on $\mathcal{F}(t)$ and $\mathbb{E}[X(t+h)|\mathcal{F}(t)] = X(t)$ holds for all $t \ge 0$ and $h > 0$.

If an integrable stochastic process with canonical filtration has independent increments, and $\mathbb{E}[X_t] = \mu$ for any t, then it is a martingale. Similarly, for stochastic process V_t, define the martingale transform $Y_t = \int_0^t V_s dX_s$, which is also a martingale.

Optional stopping theorem follows with an additional requirement of being right continuous: $\lim_{h\to 0} X(t+h) = X(t)$.

5 Renewal process

5.1 Definition

Definition: If a stochastic process $\{X_n\}$ has independent and identical sojourn time distribution F, then the counting process $N(t)$ is said to be a renewal process.

To figure out the distribution of $N(t)$, note that $N(t) \ge k \iff W_k = \sum_{i=1}^k X_i \le t \implies P(N(t) = k) = P(N(t) \ge k) - P(N(t) \ge k+1) = P(W_k \le t) - P(W_{k+1} \le t) = F_k(t) - F_{k+1}(t)$ By convolution formula, $F_k(t) = \int_0^t F_{k-1}(y)dF(y)$.

Definition: Let $M(t) = \mathbb{E}[N(t)]$ be the renewal function.

We condition on the time of first renewal $X_1 = x$, if $x > t$, clearly $N(t) = 0$. Otherwise, the process will restart itself at x. By law of total expectation, $M(t) = \mathbb{E}[\mathbb{E}[N(t)|X_1]] = \int_0^t \mathbb{E}[N(t)|X_1 = x]f(x)dx$. Since the process restarts itself at x, we have $\mathbb{E}[N(t)|X_1 = x] = \mathbb{E}[N(t-x)]+1 = M(t-x)+1$, thus $M(t) = \int_0^t (1 + M(t-x))f(x)dx = F(t) + \int_0^t M(t-x)f(x)dx$. We call the integral equation as the renewal equation. Take derivative on both sides to obtain a differential equation.

5.2 Limiting theorems

Theorem: $\mu = \mathbb{E}[X_k]$, $P(\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu}) = 1$

Theorem: $\lim_{t\to\infty} \frac{M(t)}{t} = \frac{1}{\mu}$

Theorem:[Central limit theorem] Let $\mu = \mathbb{E}[X_k]$ and $\sigma^2 = Var(X_k)$, then $\frac{Var(N(t))}{t} \to \frac{\sigma^2}{\mu^3}$ when $t \to \infty$, and so we have $\lim_{t\to\infty} P(\frac{N(t) - t/\mu}{\sqrt{t\sigma^2\mu^3}} < x) = P(Z \le x)$. Approximately, $N(t) \sim N(\frac{t}{\mu}, \frac{t\sigma^2}{\mu^3})$

5.3 Variants

Definition: Consider a renewal process $N(t)$, having interarrival time X_n and suppose that each time a renewal occurs we receive a reward denoted by R_n. Assume R_n are identical and independent random variables and define $R(t) = \sum_{n=1}^{N(t)} R_n$ as the renewal reward process.

Theorem: If $\mathbb{E}[R_n] < \infty$ and $\mathbb{E}[X_n] < \infty$, then with probability 1 we have $\lim_{t\to\infty} \frac{R(t)}{t} = \frac{\mathbb{E}[R_n]}{\mathbb{E}[X_n]}$ and $\lim_{t\to\infty} \frac{\mathbb{E}[R(t)]}{t} = \frac{\mathbb{E}[R_n]}{\mathbb{E}[X_n]}$.

Intuitively, it suggests that long-run performance will converge to expected single-cycle performance.

Definition: If a stochastic process $X(t)$ has time points at which the process probabilistically restarts itself, then it is called a regenerative process. For example, for an ergodic MC which starts at state 0, then each time the process returns to 0, it probabilistically restarts.

Suppose $X(t)$ is a regenerative process with states in $\mathbb{N}$. We are interested in the long run proportion of time that $X(t) = i$, which is defined as $\frac{R(t)}{t} = \frac{\int_0^t \mathbf{1}(X(s)=i)ds}{t}$. We consider one cycle then $\frac{R(t)}{t} = \frac{\mathbb{E}[R(T_1)]}{\mathbb{E}[T_1]}$.

Suppose we start observing the process when the component in operation at $t = 0$ is not new, but all subsequent renewals are new. The waiting time of all renewals follow the same distribution, except the first one. The limiting theorems on $\frac{N(t)}{t}$ will not be affected since it is long time result.

Let Y_i be identical and independent with $P(Y_i = j) = p_j$, consider the sequence $\{(Y_{n-r+1}, Y_{n-r+2}, \dots, Y_n)\}$, we are interested in the event that $(Y_{n-r+1}, Y_{n-r+2}, \dots, Y_n) = (y_1, \dots, y_r)$ the pattern of interest. Every time when the event happens, we say a renewal occurs at time n, let $N(n)$ denote the number of renewals by time n. This is a delayed renewal process since for the first renewal, we do not have prior information about the past, hence the distribution for X_1 is different.

Consider a delayed renewal process, where X_2 onwards follow F and X_1 follows that $P(X_1 \le x) = \frac{1}{\mu}\int_0^x (1 - F(y))dy$. We have $M(t) = \frac{t}{\mu}$, and $N(t+s) - N(s)$ has same distribution with $N(t) - N(0)$. We call it stationary renewal process.

6 Brownian motion

Definition: A stochastic process $X(t)$ is called Brownian motion if $X(0) = 0$, $X(t)$ has staionary and independent increments and $X(t) \sim N(0, \sigma^2 t)$.

Consider a k-dimensional vector $X = (X_1, \dots, X_k)^T$ with mean vector μ and covariance matrix Σ. We say X is multivariate normal distributed if joint density is defined by $f(x_1, \dots, x_k) = (2\pi)^{-\frac{k}{2}}|\Sigma|^{-\frac{1}{2}}\exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))$. Each entry follows a marginal normal distribution $X_i \sim N(\mu_i, \Sigma_{ii})$. If $X_1, \dots, X_k$ are independent normal random variables then they jointly follow multivariate normal distribution. For any dimension compatible vector or matrix A, the affine combination $AX + b$ follows $N(A\mu + b, A\Sigma A^T)$. If k-dimensional X is partitioned as $(X_1, X_2)^T$, then accordingly we partition $\mu = (\mu_1, \mu_2)^T$ and $\Sigma = (\Sigma_{11}, \Sigma_{12}, \Sigma_{21}, \Sigma_{22})$, then the conditional distribution for $X_1|X_2$ is that $X_1|X_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$.

Consider a Brownian motion with parameter σ, then for any single time point t, $X(t) \sim N(0, \sigma^2 t)$, for multiple time points $t_1, \dots, t_n$, $X \sim N(\mathbf{0}, \sigma^2\Sigma)$, where $\Sigma_{ij} = \min(t_i, t_j)$.

Definition: Standard Brownian motion is a Brownian motion with $\sigma = 1$.

6.1 Properties and variants

Define $T_\alpha = \min\{t \ge 0 : X(t) = \alpha\}$, by independent normal increment, we have $P(T_\alpha \le t) = 2\Phi(-\frac{|\alpha|}{\sigma\sqrt{t}})$.

Define $M(t) = \max_{s\in[0,t]} X(s)$, then $P(M(t) \ge \alpha) = P(T_\alpha \le t)$, hence the density function is $f_{M(t)}(\alpha) = \frac{2}{\sqrt{2\pi\sigma^2 t}}\exp(-\frac{a^2}{2\sigma^2 t})$

Definition: A Brownian motion with drift parameter μ and variance parameter σ has $X(t) \sim N(\mu t, \sigma^2 t)$. Everything same as non-drift Brownian motion except mean becomes μt_i for $X(t_i)$ in joint distribution.

Suppose $X(t)$ is a Brownian motion with drift μ and variance σ^2, define $Y(t) = e^{X(t)}$. $\mathbb{E}[Y(t)] = \mathbb{E}[e^{X(t)}] = e^{\mu t + \frac{\sigma^2 t}{2}}$, the moment generating function of $X(t)$ evaluated at 1. $Var(Y(t)) = e^{2\mu t + \sigma^2 t}(e^{\sigma^2 t} - 1)$. $Cov(Y(s), Y(t)) = e^{\mu(t+s)+\sigma^2(t+s)/2}(e^{\sigma^2 s} - 1)$.

6.2 Gaussian process

Definition: A stochastic process is called Gaussian process if any finite collection of $X(t_i)$ is jointly Gaussian. It is specified by mean vector and covariance function.

Definition: Let $B(t)$ be a standard Brownian motion. We call $\{B(t)|B(1) = 0, t \in [0,1]\}$ a Brownian bridge.

Brownian bridge is a Gaussian process with mean 0 and covariance $s(1-t)$ for $s \le t$. The increment is stationary but not independent.

Theorem: If $B(t)$ is a standard Brownian motion, then $X(t) = B(t) - tB(1)$ for $t \in [0,1]$ is a Brownian bridge.

Definition: Let $B(t)$ be a standard Brownian motion. Define $X(t) = \int_0^t B(s)ds$, it is called an integrated Brownian motion.

We have $X(t)$ is a Gaussian process. $\mathbb{E}[X(t)] = 0$. $Cov(X(s), X(t)) = s^2(\frac{t}{2} - \frac{s}{6})$. $Var(X(t)) = \frac{t^3}{3}$. The increment fails to be either stationary or independent.