Random Matters

Adrian Sai-Wah TAM

October 3, 2005

Chapter 1

Probability

1.1 Probability Measure

- Sample space: Ω
- Events: $\mathscr{A} \subset \Omega$
- Probability measure: $P : \{\mathscr{A} : \mathscr{A} \subset \Omega\} \to [0,1]$

1.2 Probability models

1.2.1 Formulas and notations

- Expectation: $E\{X\} = \int_{-\infty}^{\infty} x f_X(x) dx$ for continuous and $E\{X\} = \sum_x x f_X(x)$ for discrete
  - $E\{h(X)\} = \int_{\Omega} h(x) f_X(x) dx$
- Variance: $\sigma_X^2 = E\{(X - E\{X\})^2\} = E\{X^2\} - (E\{X\})^2$
  - Known as Steiner's tyheorem
- Baye's formula: $\Pr[A_i|B] = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}$
- Total probability formula: $P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$
- Chebyshev's inequality: Given n experiments and the experimental mean is $\bar{X}$,

$$P(|\bar{X} - \mu| > t) \leq \frac{\sigma^2}{nt^2}$$

- Construction of random variables: Inverse transform method
  - Let y be a uniformly distributed random value in $[0,1]$
  - Commulative distribution function interested: $y = F(x) = \Pr[X \leq x]$
  - $x = F^{-1}(y)$
- Little o notation: $f(x) = o(x)$ if $\lim_{x \to 0} \frac{f(x)}{x} = 0$

1.2.2 Exponential distribution and Poisson formula

- Random event with average rate of occurance: λ
  - Average number of occurance in duration x: λx
  - Probability of having occurance in infinitesimal interval δt: $\lambda \delta t$
  - In time interval $(0,t]$, we partition the interval into n equal parts, each with length $\delta t = t/n$
  - Probability of having $k \leq n$ occurances in $(0,t]$: $p_k(t) = \binom{n}{k}(\lambda \delta t)^k (1-\lambda \delta t)^{n-k}$
  - Taking limit of $n \to \infty$: $p_k(t) = \lim_{n\to\infty} \binom{n}{k}(\lambda \delta t)^k (1-\lambda \delta t)^{n-k} = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$
- Probability of no occurance in $(0,t]$: $p_0(t) = \lambda e^{-\lambda t}$
  - Interarrival time distribution: $p_0(t) = \lambda e^{-\lambda t}$
  - c.d.f.: $P_0(t) = 1 - e^{-\lambda t}$
  - Mean interarrival time: $E(t) = 1/\lambda$
  - Variance: $\sigma_t^2 = 1/\lambda^2$

1.2.3 Erlang distribution

- Truncated Poisson distribution
  - Poisson arrivals is truncated by the capacity of outlets
  - Poisson arrivals, the waiting time for k more arrivals (i.e. the arrival time for the k-th customer):

$$p(t) = \frac{e^{-\lambda t}\lambda(\lambda t)^{k-1}}{(k-1)!} = \frac{e^{-\lambda t}\lambda^k t^{k-1}}{(k-1)!}$$

- Erlang-k distribution (E_k) is the distribution of the sum of k i.i.d. exponential random variables, i.e. convolution of k i.i.d. exponentials
  - Each random variable with mean λ
  - Mean of the sum (Erlang-k): k/λ
  - Variance: k/λ^2
- $E_{k-1,k}(\lambda)$ distribution
  - Mix of E_{k-1} and E_k
  - A random variable X is the sum of (1) $k-1$ i.i.d. exponential variables with probability p and (2) k i.i.d. exponential variables with probability $q = 1-p$:

$$p(t) = p\frac{e^{-\lambda t}\lambda^{k-1}t^{k-2}}{(k-2)!} + (1-p)\frac{e^{-\lambda t}\lambda^k t^{k-1}}{(k-1)!}$$

1.2.4 Other distributions

- Geometric distribution
  - pmf: $p(x) = (1-p)p^x$
  - Mean: $E(x) = p/(1-p)$
  - Variance: $\sigma_x^2 = p/(1-p)^2$
- Gamma distribution
  - General form of Erlang distribution (i.e. convolution of k i.i.d. exponentials where k is a positive real number)

- Probability density function:
$$f(x) = \frac{\left(\frac{x-\mu}{\beta}\right)\exp(\frac{x-\mu}{\beta})}{\beta\Gamma(\gamma)}$$
with $x \geq \mu$ and $\gamma, \beta \geq 0$

- Hyperexponential: Random select one of k exponentials (referred by H_k)
  - A random variable X is any one of the k exponential distributed random variables x_i $(i = 1, \ldots, k)$
  - $X = x_i$ with probability α_i
  - x_i is exponentially distributed with parameter λ_i
  - pdf of X: $p(t) = \sum_{i=1}^{k} \alpha_i \lambda_i e^{-\lambda_i t}$
  - Mean: $E(x) = \sum_{i=1}^{k} \alpha_i / \lambda_i$
  - Coefficient of variation: $C_X^2 = \frac{2\sum_{i=1}^{r} \alpha_i / \lambda_i^2}{(\sum_{i=1}^{r} \alpha_i/\lambda_i)^2} - 1 \geq 1$
    * The inequality can be proved by Cauchy-Schwarz inequality

- Hypoexponential
  - Sum of k exponential random variables each with parameter λ_i $(i = 1, \ldots, k)$
    * $X = \sum_{i=1}^{k} x_i$, with $p_{x_i}(t) = \lambda_i e^{-\lambda_i t}$
  - If $\lambda_i = \lambda$ for all i, it is the Erlang distribution

1.2.5 Important series formulas

- Geometric series: $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ with $x < 1$
- $\sum_{k=0}^{\infty} kx^k = \frac{x}{(1-x)^2}$ with $x < 1$

1.2.6 Markov Chains

- State space: S
- Transition probability: P_{ij}. Transition probability matrix $\mathbf{P} = (P_{ij})_{n \times n}$
- Classification of states:
  - Leads to: State i leads to state j if $\mathbf{P}_{ij}^n > 0$ for some n
  - Communicates: State i leads to state j and vice versa. Communicate is reflexive, symmetric and transitive.
  - Communicate is a class property
  - A Markov chain is irreducible if there is only one communicating class
  - State i is recurrent if $P_{ii}^n > 0$ for some n. Transient otherwise.
    * If a state is a communicating class is recurrent, all states are recurrent
    * If mean time to recurrent is ∞, we call it null recurrent
  - The period of a recurrent state i is the G.C.D. of $\{n : P_{ii}^n > 0\}$
  - Aperiodic: Period of 1
- Ergodic MC: aperiodic, positive recurrent, irreducible

1.2.7 Continuous-time Markov Chain

- A process $\{X(t) : t \geq 0\}$ takes on values of state space S such that $\Pr\{X(t+s) = j|X(t) = i\} = P_{ij}(s, s+t)$ is a CTMC
  - Memoryless
- CTMC with stationary transitions if $P_{ij}(s, s+t) = P_{ij}(0, t) = P_{ij}(t)$
- Chapman-Kolmogorov equations: $P_{ij}(s+t) = \sum_k P_{ik}(s)P_{kj}(t)$ or $\mathbf{P}(s+t) = \mathbf{P}(s)\mathbf{P}(t)$
- Time from entering state i to leaving state i (a.k.a. holding time) has exponential distribution with rate v_i
  - Instantaneous transition rate from i to j: $q_{ij} = P'_{ij}(0) = v_i P_{ij}$
  - Kolmogorov forward equations: $P'_{ij}(t) = \sum\limits_{k \neq j} q_{kj} P_{ik}(t) - v_i P_{ij}(t)$
  - Kolmogorov backward equations: $P'_{ij}(t) = \sum\limits_{k \neq j} q_{ik} P_{kj}(t) - v_i P_{ij}(t)$
  - Derivation:

$$\begin{aligned}
\mathbf{P}(t+h) &= \mathbf{P}(h)\mathbf{P}(t) \\
\frac{\mathbf{P}(t+h) - \mathbf{P}(t)}{h} &= \frac{\mathbf{P}(h)\mathbf{P}(t) - \mathbf{P}(t)}{h} = \frac{\mathbf{P}(h)\mathbf{P}(t) - \mathbf{P}(0)\mathbf{P}(t)}{h} \\
&= \frac{[\mathbf{P}(h) - \mathbf{P}(0)]\mathbf{P}(t)}{h} \\
\lim_{h \to 0} \frac{\mathbf{P}(t+h) - \mathbf{P}(t)}{h} &= \lim_{h \to 0} \frac{[\mathbf{P}(h) - \mathbf{P}(0)]\mathbf{P}(t)}{h} \\
\mathbf{P}'(t) &= \mathbf{P}'(0)\mathbf{P}(t) \\
\therefore\ P'_{ij}(0) &= \sum_k P'_{ik}(0) P_{kj}(t)
\end{aligned}$$

    or using $\mathbf{P}(t+h) = \mathbf{P}(t)\mathbf{P}(h)$ for forward equation

- If the CTMC is strongly recurrent, the transition probability approaches stationary distribution: $P_{ij}(t) \to \pi_j$
- Stationary initial probability: $v_j \pi_j = \sum\limits_{k \neq j} q_{kj} \pi_k$
- Transition probability: $\mathbf{P}(t) = e^{\mathbf{Q}t}$ with $Q_{ij} = \begin{cases} q_{ij} & i \neq j \\ -v_i & i = j \end{cases}$ and $e^{\mathbf{Q}t} = \sum\limits_{n=0}^{\infty} \frac{(\mathbf{Q}t)^n}{n!}$
  - $\mathbf{Q}$ is called the infinitesimal generator
  - Writing in symmetric matrix form: $\mathbf{Q} = \mathbf{M}\Lambda\mathbf{M}^{-1}$, then we have $\mathbf{P}(t) = \sum\limits_{n=0}^{\infty} (\mathbf{M}\Lambda\mathbf{M}^{-1})^n \frac{t^n}{n!} = \mathbf{M}\left(\sum\limits_{n=0}^{\infty} \frac{\Lambda^n t^n}{n!}\right)\mathbf{M}^{-1}$
  - $\pi\mathbf{Q} = \mathbf{0}$
  - Given the distribution at time $t = 0$ is $\pi(0)$, the distribution at time t is $\pi(t) = \pi(0)e^{\mathbf{Q}t}$
- Birth death process: $q_{ij} = \begin{cases} 0 & |i-j| > 1 \\ \lambda_i & j = i+1 \\ \mu_i & j = i-1 \end{cases}$ with birth rate λ_i and death rate μ_i
  - Linear birth and death process: $\lambda_n = n\lambda$ and $\mu_n = n\mu$ with $\lambda < \mu$ which has $\pi_n = \frac{(\lambda/\mu)^n}{1-(\lambda/\mu)}$.

1.2.8 Semi-Markov Process

- Definition
  - $\{Z_n\}$ is a DTMC with state space S and transition matrix $\mathbf{P}$
  - $\{Y_n^{(i)}\}$ $i \in S$ is several i.i.d. random sequences with c.d.f. $F_i(x) = \Pr\{Y_n^{(i)} \leq x\}$
  - If $\sum_{k=1}^{m} Y_k^{(Z_k)} \leq t < \sum_{k=1}^{m+1} Y_k^{(Z_k)}$, then $X(t) = Z_m$. $X(t)$ is the SMP.
- My view:
  - Z_n controls us to look at which $\{Y_n^{(i)}\}$
  - For every step n, we look at a (possibly) different $Y_n^{(i)}$, and accumulate the sum along n
  - We make up a sequence such as: $Y_0^{(1)}, Y_1^{(4)}, Y_2^{(3)}, \ldots, Y_k^{(2)}, Y_{k+1}^{(5)}, \ldots$
  - The sequence make up the time line by concatenation
  - At time t on the time line, we refer to last index of $Y_k^{(i)}$, i.e. i. Take it as the SMP
- If $\{Y_n^{(i)}\}$ are all exponential distributed with parameter λ_i, then $X(t)$ is a CTMC
- Assume $X(t)$ is a MC (or has a embedded MC) with transition probability λ_{ij}
  - $\lambda_{ij} = \lambda_i P_{ij}$
  - $P_{ij} = \dfrac{\lambda_{ij}}{\sum_{k \in S} \lambda_{ik}}$ with $i \neq j$, $P_{ii} = 0$
  - $\lambda_i = \sum_{j \in S} \lambda_{ij}$
- Stationary distribution:
  - Define $\pi'_j = \lim_{n \to \infty} \Pr\{Z_n = j\}$ and $\pi_j = \lim_{t \to \infty} \Pr\{X(t) = j\}$
  - $\pi_j = \dfrac{\pi'_j E\{Y^{(j)}\}}{\sum_{i \in S} \pi'_i E\{Y^{(i)}\}}$

1.2.9 References:

- [14], "Markov Process"
- [5], Lecture 19-20
- [2], Chapter 1, "Introduction"
- [11], Chapter 5, "The Exponential Distribution and the Poisson Process"

1.3 Wiener Process

1.3.1 Brownian Motion

1.3.1.1 Definition

- A symmetric random walk $X(t)$ with step size δx and take a step every δt time
  - $X(t) = (X_1 + X_2 + \cdots + X_n)\delta x$ where $n = \left\lfloor \dfrac{t}{\delta t} \right\rfloor$
  - $X_i = \begin{cases} 1 & \text{if step } i \text{ is upward} \\ -1 & \text{if step } i \text{ is downward} \end{cases}$
  - Mean: $E\{X(t)\} = 0$

- Variance: $Var\{X(t)\} = \left\lfloor \frac{t}{\delta t} \right\rfloor (\delta x)^2$
- If we set $\delta x = \sigma\sqrt{\delta t}$, and let $\delta t \to 0$, then $Var\{X(t)\} = \sigma^2 t$ and $X(t)$ is a Brownian motion

- A stochastic process $S(t)$ is a Brownian motion if
  1. $S(t)\ N(0,\sigma^2 t)$
  2. Independent increments: if $t_0 < t_1 < \cdots < t_n$, then $S(t_1)-S(t_0),\ S(t_2)-S(t_1),\ \ldots,\ S(t_n)-S(t_{n-1})$ are independent
  3. Increments are stationary regardless of s: $S(t+s)-S(s)\ N(0,\sigma^2 t)$
  4. $S(0) = 0$

1.3.1.2 Standard Brownian motion

- If $\sigma = 1$, the Brownian motion is called the standard Brownian motion. For any Brownian motion $X(t)$, $B(t) = X(t)/\sigma$ is a standard Brownian motion
  - If $X(t)$ is a standard Brownian motion, the p.d.f. of $X(t)$is given by: $f_t(x) = \frac{1}{\sqrt{2\pi t^2}} e^{-x^2/2t}$
  - P.d.f. of $X(t_1) = x_1,\ X(t_2) = x_2,\ \ldots,\ X(t_n) = x_n$ with $t_1 < t_2 < \cdots < t_n$:

$$\begin{aligned} f(x_1,x_2,\ldots,x_n) &= f_{t_1}(x_1)\, f_{t_2-t_1}(x_2-x_1)\cdots f_{t_n-t_{n-1}}(x_n-x_{n-1}) \\ &= \frac{\exp\left(-\frac{x_1^2}{2t_1} - \frac{(x_2-x_1)^2}{2(t_2-t_1)} - \cdots - \frac{(x_n-x_{n-1})^2}{2(t_n-t_{n-1})}\right)}{\sqrt{(2\pi)^n\left[t_1(t_2-t_1)\cdots(t_n-t_{n-1})\right]}} \end{aligned}$$

- Conditional probability with of $X(s)$ given $X(t) = B$ with $s < t$ (the future value):

$$\begin{aligned} f_{s|t}(x|B) &= \frac{f_s(x)\, f_{t-s}(B-x)}{f_t(B)} \\ &= \frac{\exp\left(-\frac{x^2}{2s} - \frac{(B-x)^2}{2(t-s)}\right)}{\sqrt{(2\pi)^2 s(t-s)}} \Bigg/ \frac{\exp\left(-\frac{B^2}{2t}\right)}{\sqrt{2\pi t}} \\ &= \sqrt{\frac{t}{2\pi s(t-s)}} \exp\left(-\frac{sB^2 - 2Bsx + tx^2}{2s(t-s)} + \frac{B^2}{2t}\right) \\ &= \sqrt{\frac{t}{2\pi s(t-s)}} \exp\left(-\frac{B^2 s/t - 2Bxs/t + x^2}{2s(t-s)/t} + \frac{B^2}{2t}\right) \\ &= \sqrt{\frac{t}{2\pi s(t-s)}} \exp\left(-\frac{x^2 - 2Bxs/t + (Bs/t)^2 - (Bs/t)^2 + B^2 s/t}{2s(t-s)/t} + \frac{B^2}{2t}\right) \\ &= \sqrt{\frac{t}{2\pi s(t-s)}} \exp\left(-\frac{(x-Bs/t)^2 - B^2 s(s-t)/t^2}{2s(t-s)/t} + \frac{B^2}{2t}\right) \\ &= \sqrt{\frac{t}{2\pi s(t-s)}} \exp\left(-\frac{(x-Bs/t)^2}{2s(t-s)/t} + \frac{B^2 s(s-t)/t^2}{2s(t-s)/t} + \frac{B^2}{2t}\right) \\ &= \sqrt{\frac{t}{2\pi s(t-s)}} \exp\left(-\frac{(x-Bs/t)^2}{2s(t-s)/t}\right) \end{aligned}$$

  - $E\{X(s)|X(t) = B\} = \frac{s}{t}B$
  - $Var\{X(s)|X(t) = B\} = \frac{s}{t}(t-s)$

- However, if $\tau < t$ (the past value) is given,
  - Expected increment is zero:

$$\begin{aligned} E\{S(t)|S(\tau)\} &= E\{S(t) - S(\tau) + S(\tau)|S(\tau)\} \\ &= E\{S(t) - S(\tau)|S(\tau)\} + E\{S(\tau)|S(\tau)\} \\ &= 0 + S(\tau) \\ &= S(\tau) \end{aligned}$$

– Variance increases with time:

$$\begin{aligned} Var\{S(t)|S(\tau)\} &= Var\{S(t) - S(\tau) + S(\tau)|S(\tau)\} \\ &= Var\{S(t) - S(\tau)|S(\tau)\} \\ &= Var\{S(t) - S(\tau)\} \\ &= t - \tau \end{aligned}$$

- Combining Brownian motion: Let $X\ N(0, \sigma^2)$ and $Y\ N(0, \tau^2)$ are independent. $Z = X + Y$.
  – Conditional distribution of X given Z:

$$\begin{aligned} f_{X|Z}(x|z) &= \frac{f_{X,Z}(x,z)}{f_Z(z)} = \frac{f_{X,Y}(x,z-x)}{f_Z(z)} \\ &= \frac{f_X(x) f_Y(z-x)}{f_Z(z)} \end{aligned}$$

  where $Z\ N(0, \gamma^2)$ with $\gamma^2 = \sigma^2 + \tau^2$. Hence

$$\begin{aligned} f_{X|Z}(x,z) &= \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{x^2}{2\sigma^2}) \cdot \frac{1}{\sqrt{2\pi\tau^2}} \exp(-\frac{(z-x)^2}{2\tau^2})}{\frac{1}{\sqrt{2\pi\gamma^2}} \exp(-\frac{z^2}{2\gamma^2})} \\ &= \frac{1}{\sqrt{2\pi b^2}} \exp(-\frac{(x-a)^2}{2b^2}) \end{aligned}$$

  where $b = \dfrac{\tau\sigma}{\gamma}$ and $a = \dfrac{b^2 z}{\tau^2} = \dfrac{\sigma^2}{\sigma^2+\tau^2} z$. That is, given Z, the conditional distribution of $X\ N(a, b^2)$.
  – If $t < s$ such that $X = X(t)$, $Y = X(s) - X(t)$, and $Z = X + Y = X(s)$. Then $\sigma_X^2 = t$, $\sigma_Y^2 = s - t$, $\sigma_Z^2 = s - t + t = s$.
    * $a = \dfrac{t}{s} X(s)$
    * $b^2 = \dfrac{(s-t)t}{s}$
    * $E\{X|Z\} = E\{X(t)|X(s)\} = \dfrac{t}{s} X(s)$
    * $Var\{X|Z\} = E\{X(t)|X(s) = \dfrac{(s-t)t}{s}$

1.3.1.3 Hitting Time

- Hitting time: Given $X(0) = 0$ and $X(t)\ N(0,t)$. Let the first time of a particular motion to hit $a > 0$ is T_a,

$$\begin{aligned} \Pr\{X(t) \geq a\} &= \Pr\{X(t) \geq a | T_a \leq t\} \Pr\{T_a \leq t\} + \Pr\{X(t) \geq a | T_a > t\} \Pr\{T_a > t\} \\ &= \frac{1}{2} \Pr\{T_a \leq t\} + 0 \\ \therefore\ \Pr\{T_a \leq t\} &= 2 \Pr\{X(t) \geq a\} \\ &= \frac{2}{\sqrt{2\pi t}} \int_a^\infty e^{-x^2/2t} dx \\ &= \frac{2}{\sqrt{2\pi}} \int_{a/\sqrt{t}}^\infty e^{-x^2/2t} dx \end{aligned}$$

  – $\Pr\{X(t) \geq a | T_a \leq t\} = \frac{1}{2}$ because of the symmetric nature of Brownian motion
  – $\Pr\{X(t) \geq a | T_a > t\} = 0$ is obvious by definition of T_a
  – For general $a \in (-\infty, \infty)$, $\Pr\{T_a \leq t\} = \dfrac{2}{\sqrt{2\pi}} \displaystyle\int_{|a|/\sqrt{t}}^\infty e^{-x^2/2t} dx$
  – Similarly, we can have $\Pr\{\max\limits_{0 \leq s \leq t} X(s) \geq a\} = \Pr\{T_a \leq t\} = \dfrac{2}{\sqrt{2\pi}} \displaystyle\int_{|a|/\sqrt{t}}^\infty e^{-x^2/2t} dx$
  – The probability of hitting a before hitting $-b$ (with $a, b > 0$) can be analyzed by symmetric random walk with step size $\delta x \to 0$:

$$\Pr\{\text{hitting } a \text{ before hitting } -b\} = \frac{b}{a+b}$$

1.3.1.4 Box-Muller Method for Simulating $N(0,1)$

- Normal density function: $N(0,1) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$
- If two uniform random numbers are drawn from $[0,1]$,

$$y = \sqrt{-2\ln x_1}\cos(2\pi x_2)$$

  will give y $N(0,1)$

1.3.1.5 Brownian motion with drift

- Brownian motion with drift: $X(t) = \sigma B(t) + \mu t$ where $B(t)$ is a standard Brownian motion
  - $X(t)$ $N(\mu t, \sigma^2 t)$
  - $X(t)$still has stationary and independent increments, and $X(0) = 0$

1.3.2 Geometric Brownian motion

- If $\{X(t)\}$ is a Brownian motion with drift, then $Y(t) = e^{X(t)}$ is a geometric Brownian motion
- Geometric brownian motion: $Y(t) = e^{X(t)}$ where $X(t)$is a Brownian motion with drift
  - $X(t)$ $N(\mu t, \sigma^2 t)$, where μ is called the drift parameter and σ is called the volatility of $Y(t)$
  - Given all the past value up to $s < t$,

$$\begin{aligned} E\{Y(t)|Y(u):\ 0 \le u \le s\} &= E\{e^{X(t)}|X(u):\ 0 \le u \le s\} \\ &= E\{e^{X(s)+X(t)-X(s)}|X(u):\ 0 \le u \le s\} \\ &= e^{X(s)}E\{e^{X(t)-X(s)}|X(u):\ 0 \le u \le s\} \\ &= Y(s)E\{e^{X(t)-X(s)}\} \\ &= Y(s)e^{\mu(t-s)+(t-s)\sigma^2/2} \\ &= Y(s)e^{(t-s)(\mu+\sigma^2/2)} \end{aligned} \tag{1.1}$$

    where a result from the moment generating function $E\{e^{aW}\} = e^{aE\{W\}+a^2 Var\{W\}/2}$ is used
  - Ratio of a geometric Brownian motion $\{Y\}$ at $t+\tau$ and τ is lognormal distributed:

$$\log\left(\frac{Y(t+\tau)}{Y(\tau)}\right)\ N(\mu t, \sigma^2 t)$$

- A stochastic process $\{X(t)\}$ is a martingale if $E\{X(t)|X(u):\ 0 \le u \le s\} = X(s)$ whenever $s < t$.
  - A geometric brownian motion can be a martingale with $\mu = -\frac{\sigma^2}{2}$
- Use of geometric Brownian motion: Security price modeling
  Let $S(t)$ be the price of a security at t, then the expected price is given by

$$E[S(t)] = S(0)\exp\left(\frac{(\mu+\sigma^2)}{2}t\right)$$

  - $S(t)$ is a limit as the following:
    Price of the security will go up by the factor u with probability p and go down by the factor d with probability $1-p$ in every δt time units, where

$$\begin{aligned} u &= \exp(\sigma\sqrt{\delta t}) \\ d &= \exp(-\sigma\sqrt{\delta t}) \\ p &= \frac{1}{2}\left(1 + \frac{\mu}{\sigma}\sqrt{\delta t}\right) \end{aligned}$$

    As $\delta t \to 0$, $S(t)$ is obtained as previously mentioned

1.3.3 Pricing Stock Options

1.3.3.1 Arbitrage

- Option scenario: An option of price cy to buy y shares of a stock at price P at time t
  - At time 0, the option and $x$share of stock is purchased at price $P_0 < P$. Total cost is $P_0x + cy$
  - At time t, the stock price per share can either be $P_1 > P > P_0$ or $P_2 < P_0 < P$
    * If P_1: The total worth is $P_1x + (P_1 - P)y$
    * If P_2: The total worth is P_2x
    * Fix the worth at t:

$$\begin{aligned} P_1x + (P_1 - P)y &= P_2x \\ y &= -\frac{P_1 - P_2}{P_1 - P}x \end{aligned}$$

  - Gain at t:

$$\begin{aligned} P_2x - P_0x - cy &= (P_2 - P_0 + \frac{P_1 - P_2}{P_1 - P}c)x \\ &= \frac{PP_0 + P_1P_2 - PP_2 - P_0P_1 + cP_1 - cP_2}{P_1 - P}x \end{aligned}$$

    with careful choice of c and x, we can make the gain always non-negative. A sure win betting in this mannar is called an arbitrage.

- Arbitrage theorem:
  A set of possible outcomes $S = \{1, 2, \ldots, m\}$ and $n$wagers. Amount $x_i \in \mathbb{R}$ is bet on wager i and the return $x_ir_i(j)$ is earned if the outcome is j. A betting scheme $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ will have the return $R = \sum_{i=1}^{n} x_ir_i(j)$ on outcome j. Then either

  1. There exists a probability vector $\mathbf{p} = (p_1, p_2, \ldots, p_m)$ which $\sum_{j=1}^{m} p_jr_i(j) = 0$, $\forall i = 1, \ldots, n$, or
  2. There exists a betting scheme $x = (x_1, x_2, \ldots, x_n)$ which $\sum_{i=1}^{m} x_ir_i(j) > 0$, $\forall j = 1, \ldots, m$

1.3.3.2 Black-Scholes option pricing formula

- Present price of a stock is $X(0) = x_0$. Current interest rate is α (i.e. $100\alpha\%$).
- Let the price of a stock at t is $X(t)$. Then the present value of $X(t)$ is $e^{-\alpha t}X(t)$
- Option i costs c_i per share to allow a purchase of stock at time t_i for the price of K_i per share. $i = 1, \ldots, N$
- There is no sure win strategy, hence, by arbitrage theorem, there is a probability measure $\mathbf{P}$ over the set of outcomes under which all the wagers have zero expected return
- A wager observe the stock up to time s, then purchasing stock at s and selling it at t, with $0 \le s < t \le T$. The expected return to be zero.
  - Present value of purchasing: $e^{-\alpha s}X(s)$
  - Present value of selling: $e^{-\alpha t}X(t)$
  - Expected return is zero:

$$\begin{aligned} E_{\mathbf{P}}\left\{e^{-\alpha t}X(t) - e^{-\alpha s}X(s) :\ 0 \le u \le s\right\} &= 0 \\ E_{\mathbf{P}}\left\{e^{-\alpha t}X(t)|X(u) :\ 0 \le u \le s\right\} &= e^{-\alpha s}X(s) \end{aligned} \tag{1.2}$$

- The wager purchased an option for buying one share of the stock at t for the price K
  - At time t, the worth of the option is $(X(t) - K)^+$
  - Present value of the option: $e^{-\alpha t}(X(t) - K)^+$

- Cost of option at time 0 with no arbitrage possible: $c = E_{\mathbf{P}}\left\{e^{-\alpha t}\left(X(t) - K\right)^{+}\right\}$
- Condition for arbitrage possible: The is no $\mathbf{P}$ that satisfies both

$$E_{\mathbf{P}}\left\{e^{-\alpha t_i} X(t_i) | X(u):\ 0 \le u \le s\right\} = e^{-\alpha s} X(s) \tag{1.3}$$
$$E_{\mathbf{P}}\left\{e^{-\alpha t_i}\left(X(t_i) - K_i\right)^{+}\right\} = c_i \tag{1.4}$$

for some option i with present cost c_i and exercise price K_i at time t_i

- Suppose price $X(t) = x_0 e^{Y(t)}$ is a geometric Brownian motion with drift coefficient μ and variance parameter σ^2. If defining $\alpha = \mu + \sigma^2/2$, and using equation (1.1):

$$\begin{aligned} E\left\{X(t)|X(u):\ 0 \le u \le s\right\} &= X(s) e^{(t-s)(\mu+\sigma^2/2)} \\ &= X(s) e^{\alpha(t-s)} \\ \therefore \quad E\left\{e^{-\alpha t} X(t)|X(u):\ 0 \le u \le s\right\} &= e^{-\alpha s} X(s) \end{aligned}$$

  - If $\mathbf{P}$ is the probability measure governing $\{x_0 e^{Y(t)}:\ 0 \le t \le T\}$ where $Y(t)\ N(\mu t, \sigma^2 t)$, the equation (1.2) is satisfied
  - For no arbitrage, the price of an option is $c = E_{\mathbf{P}}\left\{e^{-\alpha t}\left(X(t) - K\right)^{+}\right\}$, i.e.

$$\begin{aligned} ce^{\alpha t} &= E_{\mathbf{P}}\left\{\left(X(t) - K\right)^{+}\right\} \\ &= \int_{-\infty}^{\infty} \left(x_0 e^y - K\right)^{+} p(y) dy \\ &= \int_{\log(K/x_0)}^{\infty} \left(x_0 e^y - K\right) \cdot \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-(y-\mu t)^2/2\sigma^2 t} dy \\ \text{Let} \quad w &= \frac{y - \mu t}{\sigma\sqrt{t}}, \\ \therefore\ dw &= \frac{1}{\sigma\sqrt{t}} dy \\ y &= \sigma\sqrt{t} w + \mu t \\ \text{Then} \quad ce^{\alpha t} &= \int_{\log(K/x_0)}^{\infty} \left(x_0 e^y - K\right) \cdot \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-(y-\mu t)^2/2\sigma^2 t} dy \\ &= \int_{\log(K/x_0)}^{\infty} \left(x_0 e^{\sigma\sqrt{t}w+\mu t} - K\right) \cdot \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-w^2/2} \sigma\sqrt{t} dw \\ &= \int_{\log(K/x_0)}^{\infty} \left(x_0 e^{\sigma\sqrt{t}w+\mu t} - K\right) \cdot \frac{1}{\sqrt{2\pi}} e^{-w^2/2} dw \\ &= \frac{x_0 e^{\mu t}}{\sqrt{2\pi}} \int_{w_0}^{\infty} e^{\sigma w\sqrt{t}} e^{-w^2/2} dw - \frac{K}{\sqrt{2\pi}} \int_{w_0}^{\infty} e^{-w^2/2} dw \\ \text{with} \quad w_0 &= \frac{\log(K/x_0) - \mu t}{\sigma\sqrt{t}}. \\ \text{Consider} \quad \frac{1}{\sqrt{2\pi}} \int_{w_0}^{\infty} e^{\sigma w\sqrt{t}} e^{-w^2/2} dw &= \frac{e^{t\sigma^2/2}}{\sqrt{2\pi}} \int_{w_0}^{\infty} e^{-(w-\sigma\sqrt{t})^2/2} dw \\ &= e^{t\sigma^2/2} \Pr\{N(\sigma\sqrt{t}, 1) \ge w_0\} \\ &= e^{t\sigma^2/2} \Pr\{N(0,1) \ge w_0 - \sigma\sqrt{t}\} \end{aligned}$$

  - Standard normal distribution function: $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)\right)$
    * $\Pr\{N(0,1) \ge z\} = \Pr\{N(0,1) \le -z\} = \Phi(-z)$
  - Hence we have

$$\begin{aligned} ce^{\alpha t} &= \frac{x_0 e^{\mu t}}{\sqrt{2\pi}} \int_{w_0}^{\infty} e^{\sigma w\sqrt{t}} e^{-w^2/2} dw - \frac{K}{\sqrt{2\pi}} \int_{w_0}^{\infty} e^{-w^2/2} dw \\ &= x_0 e^{\mu t} e^{\sigma^2 t/2} \Pr\{N(0,1) \ge w_0 - \sigma\sqrt{t}\} - K \Pr\{N(0,1) \ge w_0\} \\ &= x_0 e^{(\mu+\sigma^2)t/2} \Phi(\sigma\sqrt{t} - w_0) - K\Phi(-w_0) \end{aligned}$$

$$= x_0 e^{\alpha t}\Phi(\sigma\sqrt{t} - w_0) - K\Phi(-w_0)$$

$$\therefore \quad c = x_0\Phi\left(\sigma\sqrt{t} - w_0\right) - Ke^{-\alpha t}\Phi(-w_0)$$

$$\text{or} \quad cx_0\Phi\left(\frac{\sigma^2 t + \mu t - \log(K/x_0)}{\sigma\sqrt{t}}\right) - Ke^{-\alpha t}\Phi\left(\frac{\mu t - \log(K/x_0)}{\sigma\sqrt{t}}\right) \tag{1.5}$$

where (1.5) is called the Black-Scholes option cost valuation formula

* With $\alpha = \mu + \sigma^2/2$, the value of w_0 can be written as: (which has no μ involved explicitly)

$$w_0 = \frac{\log(K/x_0) - \alpha t + \sigma^2 t/2}{\sigma\sqrt{t}}$$

1.3.3.3 Obtaining σ^2

- In practice, we observe the price process $\{X(t) = e^{Y(t)}\}$ for time $t \in [0, s]$, with a fixed interval h
  - $N = \left\lfloor \frac{s}{h} \right\rfloor$ samples of $Y(t)$ obtained
  - Define $W_k = Y(kh) - Y((k-1)h)$, $k = 1, \ldots, N$
    They are i.i.d. normal random variables with variance $h\sigma^2$
  - Sample variance: $S^2 = \sum_{i=1}^{N} \frac{(W_i - \bar{W})^2}{N-1}$
  - $\frac{(N-1)S^2}{(\sigma^2 h)}$ has a χ-squared distribution with $N-1$ degrees of freedom, hence

$$E\left\{\frac{(N-1)S^2}{(\sigma^2 h)}\right\} = N - 1$$

$$Var\left\{\frac{(N-1)S^2}{(\sigma^2 h)}\right\} = 2(N-1)$$

    * For random variable x in chi-squared distribution with k degree of freedom, $E\{x\} = k$ and $Var\{x\} = 2k$
    * We have: $E\{S^2/h\} = \sigma^2$ and $Var\{S^2/h\} = 2\sigma^4/(N-1)$
      which means we can reduce the variance of S^2 by making h smaller (or equivalently, N larger) , hence S^2 can be a better estimate of σ^2
  - $\lim_{h\to 0} S^2 = \sigma^2$. This is the way to obtain σ^2 for equation (1.5)

1.3.3.4 Martingale

- A stochastic process $\{X(t) : \ t \geq 0\}$ is a martingale if $E\{X(t)|X(u) : \ 0 \leq u \leq s < t\} = X(s)$, hence $\{e^{-\alpha t}X(t) : \ t \geq 0\}$ is a martingale if $E\{e^{-\alpha t}X(t)|X(u) : \ 0 \leq u \leq s\} = e^{-\alpha s}X(s)$, which results in no arbitrage possibilities
  - For a martingale process $\{Z(t) : \ t \geq 0\}$ which governs the stock price, the cost of a option at time t with exercise price K should be:

$$c = E\left\{e^{-\alpha t}\left(e^{\alpha t}Z(t) - K\right)^+\right\}$$

$$= E\left\{\left(Z(t) - Ke^{-\alpha t}\right)^+\right\}$$

    for no arbitrage. C.f. equations (1.3) and (1.4)
  - Example of martingale:
    Let $\{N(t) : \ t \geq 0\}$ be a Poisson process with rate λ, and $Y_1, Y_2, \ldots$ is a sequence of independent random variables with common mean μ. Let

$$X(t) = x_0 \prod_{i=1}^{N(t)} Y_i$$

$$= x_0 \prod_{i=1}^{N(s)} Y_i \prod_{j=N(s)+1}^{N(t)} Y_j$$

for $s < t$. Then $E\{X(t)|X(u):\ 0 \le u \le s\} = X(s)E\left\{\prod_{j=N(s)+1}^{N(t)} Y_j\right\}$ and

$$\begin{aligned}
E\left\{\prod_{j=N(s)+1}^{N(t)} Y_j\right\} &= \sum_{n=0}^{\infty} \mu^n \cdot \frac{(\lambda(t-s))^n e^{-\lambda(t-s)}}{n!} \\
&= e^{-\lambda(t-s)} \sum_{n=0}^{\infty} \frac{(\lambda\mu(t-s))^n}{n!} \\
&= e^{-\lambda(t-s)} e^{\lambda\mu(t-s)} \\
&= e^{-\lambda(t-s)(1-\mu)} \\
\therefore \quad E\{X(t)|X(u):\ 0 \le u \le s\} &= e^{-\lambda(t-s)(1-\mu)} X(s) \\
E\{e^{-\lambda(\mu-1)t} X(t)|X(u):\ 0 \le u \le s\} &= e^{-\lambda(\mu-1)s} X(s)
\end{aligned}$$

which means $\alpha = \lambda(\mu - 1)$.

1.3.4 Gaussian Process

- A Gaussian process is a process $\{X(t):\ t \ge 0\}$ with mean $\mu_X(t)$ and covariance function $C_X(t_1, t_2)$ such that $\{X(t_1), X(t_2), \ldots, X(t_n)\}$ is a multivariate normal distribution for any $t_1, t_2, \ldots, t_n$
  - For any vector $\mathbf{X} = (X(t_1), X(t_2), \ldots, X(t_n))$ formed by any sampling of $X(t)$, the p.d.f. is a multivariate normal distribution:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det \Sigma_X}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu_X)^T \Sigma_X^{-1} (\mathbf{x}-\mu_X)\right) \qquad \mathbf{x} \in \mathbb{R}^n$$

$$\text{where} \quad \mu_X = \begin{pmatrix} \mu_X(t_1) \\ \mu_X(t_2) \\ \vdots \\ \mu_X(t_n) \end{pmatrix}$$

$$\Sigma_X = \begin{pmatrix} C_X(t_1,t_1) & C_X(t_1,t_2) & \cdots & C_X(t_1,t_n) \\ C_X(t_2,t_1) & C_X(t_2,t_2) & \cdots & C_X(t_2,t_n) \\ \vdots & \vdots & & \vdots \\ C_X(t_n,t_1) & C_X(t_n,t_2) & \cdots & C_X(t_n,t_n) \end{pmatrix}$$

- Obviously, Brownian motion process is a Gaussian process
- If $X(t)$ is a standard Brownian motion process, we have

$$\begin{aligned}
C_X(s,t) &= Cov\{X(s), X(t)\} \\
&= Cov\{X(s), X(s) + X(t) - X(s)\} \\
&= Cov\{X(s), X(s)\} + Cov\{X(t) - X(s)\} \\
&= Cov\{X(s), X(s)\} \\
&= Var\{X(s)\} \\
&= s
\end{aligned}$$

  - For $s < 1$ conditioning on $X(1) = 0$, $E\{X(s)|X(1) = 0,\ s < 1\} = 0$. Hence for $s < t < 1$,

$$\begin{aligned}
Cov\{X(s), X(t)|X(1) = 0\} &= E\{X(s)X(t)|X(1) = 0\} \\
&= E\left\{E\{X(s)X(t)|X(t), X(1) = 0\}|X(1) = 0\right\} \\
&= E\left\{X(t)\ E\{X(s)|X(t)\}|X(1) = 0\right\} \\
&= E\left\{X(t)\frac{s}{t}X(t)|X(1) = 0\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{s}{t} E\left\{X^2(t) | X(1) = 0\right\} \\
&= \frac{s}{t} t(1-t) \\
&= s(1-t)
\end{aligned}
$$

where the conditional mean and variance of Brownian motion process is applied.

1.3.5 References

- [5], Lecture 21-23
- [10], Chapter 2, “Normal Random Variables”; Chapter 3, “Geometric Brownian Motion”
- [11], Chapter 10, “Brownian Motion and Stationary Processes”
- [1]

Chapter 2

Stochastic Calculus

2.1 Stochastic Integral

- Given $\{X(t) : t \geq 0\}$ is a standard Brownian motion, and f is a real function with continuous derivative defined on $[a,b]$, the stochastic integral is defined as Riemann sum:

$$\int_a^b f(t)dX(t) \triangleq \lim_{n\to\infty} \sum_{i=0}^{n} f(t_i)\left[X(t_{i+1}) - X(t_i)\right] \tag{2.1}$$

  with $a = t_0 < t_1 < \cdots < t_n < t_{n+1} = b$ is a partition of $[a,b]$ such that

$$\lim_{n\to\infty} \max_i (t_{i+1} - t_i) = 0$$

  - The definition of (2.1) is called the Reimann-Stieltjes integral or Lebesgue-Stieltjes integral of f with respect to X on $[a,b]$
  - Compare: Reimann integral of f with respect to t on $[a,b]$: $\int_a^b f(t)dt = \lim_{n\to\infty} \sum_{i=0}^{n} f(t_i)\delta t_i$

- Integration by parts formula in stochastic integral:

$$\sum_{i=0}^{n} f(t_i)\left[X(t_{i+1}) - X(t_i)\right] = f(b)X(b) - f(a)X(a) - \sum_{i=0}^{n} X(t_{i+1})\left[f(t_{i+1}) - f(t_i)\right]$$

$$\int_a^b f(t)dX(t) = f(b)X(b) - f(a)X(a) - \int_a^b X(t)df(t) \tag{2.2}$$

  which (2.2) is usually taken as the definition to calculate the stochastic integral.

  - Expectation is zero:

$$\begin{aligned} E\left\{\int_a^b f(t)dX(t)\right\} &= E\left\{f(b)X(b)\right\} - E\left\{f(a)X(a)\right\} - E\left\{\int_a^b X(t)df(t)\right\} \\ &= f(b)E\left\{X(b)\right\} - f(a)E\left\{X(a)\right\} - \int_a^b E\left\{X(t)\right\}df(t) \\ &= f(b)\cdot 0 - f(a)\cdot 0 - \int_a^b 0df(t) \\ &= 0 \end{aligned}$$

  - Variation:

$$\begin{aligned} Var\left\{\sum_{i=0}^{n} f(t_i)\left[X(t_{i+1}) - X(t_i)\right]\right\} &= \sum_{i=0}^{n} f^2(t_i)Var\left\{X(t_{i+1}) - X(t_i)\right\} \\ &= \sum_{i=0}^{n} f^2(t_i)(t_{i+1} - t_i) \\ \therefore \quad Var\left\{\int_a^b f(t)dX(t)\right\} &= \int_a^b f^2(t)dt \end{aligned}$$

- The process $\{dX(t): t \geq 0\}$ is called the white noise and $\int_a^b f(t)dX(t)$ is the white noise transformation as it can be imagined that a time varying function $f(t)$ travels through a white noise medium to yield the output at time b, $\int_a^b f(t)dX(t)$.
- Example: Particle in Brownian motion
  A particale moving with velocity $v(t)$ in viscous fluid. The retardation is $\beta v(t)$ and the accelaration due to Brownian motion is $\alpha X'(t)$ where $\{X(t): t \geq 0\}$ is a standard Brownian motion

$$\begin{aligned}
v'(t) &= -\beta v(t) + \alpha X'(t) \\
e^{\beta t}\left[v'(t) + \beta v(t)\right] &= \alpha e^{\beta t} X'(t) \\
\frac{d}{dt}\left[e^{\beta t} v(t)\right] &= \alpha e^{\beta t} X'(t) \\
\therefore \quad e^{\beta t} v(t) &= v(0) + \alpha \int_0^t e^{\beta s} X'(s) ds \\
v(t) &= v(0) e^{-\beta t} + \alpha \int_0^t e^{-\beta(t-s)} dX(s) \\
\therefore \quad v(t) &= v(0) e^{-\beta t} + \alpha \left[X(t) - \int_0^t X(s) \beta e^{-\beta(t-s)} ds\right]
\end{aligned}$$

  which the last line is using (2.2).

2.2 Itô Calculus

- A standard Wiener process/Brownian motion W_t
  - Infinitesimal increments dW_t in time dt has density: $\dfrac{1}{\sqrt{2\pi dt}} e^{-dW_t^2/2dt}$
  - Mean of dW_t: $\overline{dW_t} = E\{dW_t\} = 0$
  - Variance of $dW_t = E\left\{(dW_t - \overline{dW_t})^2\right\} = E\left\{(dW_t)^2\right\} = \overline{(dW_t)^2}$
    * But according to the density function, the variance is dt, hence

$$\overline{(dW_t)^2} = dt \tag{2.3}$$

  - A function of Wiener process: $f(t, W_t)$ has the differential

$$df(t, W_t) = f(t + dt, W_t + dW_t) - f(t, W_t) \tag{2.4}$$

- Taylor's expansion: [6]

$$f(x_1, \ldots, x_n) = \sum_{j=0}^{\infty} \left\{ \frac{1}{j!} \left[\sum_{k=1}^{n} (x_k - a_k) \frac{\partial}{\partial x'_k} \right]^j (x'_1, \ldots, x'_n) \right\}_{x'_1 = a_1, \ldots, x'_n = a_n}$$

$$\begin{aligned}
f(x + \delta x, y + \delta y) = f(x, y) &+ \left[\frac{\partial f(x,y)}{\partial x} \delta x + \frac{\partial f(x,y)}{\partial y} \delta y \right] \\
&+ \frac{1}{2!} \left[\frac{\partial^2 f(x,y)}{\partial x^2} (\delta x)^2 + 2 \frac{\partial^2 f(x,y)}{\partial x \partial y} \delta x \delta y + \frac{\partial^2 f(x,y)}{\partial y^2} (\delta y)^2 \right] \\
&+ \frac{1}{3!} \left[\frac{\partial^3 f(x,y)}{\partial x^3} (\delta x)^3 + 3 \frac{\partial^3 f(x,y)}{\partial x^2 \partial y} (\delta x)^2 \delta y + 3 \frac{\partial^3 f(x,y)}{\partial x \partial y^2} \delta x (\delta y)^2 + \frac{\partial^3 f(x,y)}{\partial y^3} (\delta y)^3 \right] \\
&+ \cdots + \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} \frac{\partial^n f(x,y)}{\partial x^k \partial y^{n-k}} (\delta x)^k (\delta y)^{n-k} + \cdots
\end{aligned}$$

  - Hence tyhe Taylor's expansion of (2.4):

$$\begin{aligned}
df(t, W_t) &= -f(t, W_t) + f(t + dt, W_t + dW_t) \\
&= -f(t, W_t) + f(t, W_t) + \frac{\partial f(t, W_t)}{\partial t} dt + \frac{\partial f(t, W_t)}{\delta W_t} dW_t
\end{aligned}$$

$$+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial t^2}(dt)^2+\frac{\partial^2 f(t,W_t)}{\partial t\partial W_t}dtdW_t+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial W_t^2}(dW_t)^2+\cdots$$
$$=\frac{\partial f(t,W_t)}{\partial t}dt+\frac{\partial f(t,W_t)}{\delta W_t}dW_t+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial t^2}(dt)^2+\frac{\partial^2 f(t,W_t)}{\partial t\partial W_t}dtdW_t+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial W_t^2}(dW_t)^2+\cdots$$

∗ Substituting (2.3), the mean behavior of $df(t,W_t)$ is therefore:

$$\begin{aligned}
df(t,W_t) &= \frac{\partial f(t,W_t)}{\partial t}dt+\frac{\partial f(t,W_t)}{\delta W_t}dW_t+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial t^2}(dt)^2+\frac{\partial^2 f(t,W_t)}{\partial t\partial W_t}dtdW_t+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial W_t^2}(dW_t)^2+\cdots\\
&= \frac{\partial f(t,W_t)}{\partial t}dt+\frac{\partial f(t,W_t)}{\delta W_t}dW_t+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial t^2}(dt)^2+\frac{\partial^2 f(t,W_t)}{\partial t\partial W_t}dtdW_t+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial W_t^2}dt+\cdots\\
&\approx \frac{\partial f(t,W_t)}{\partial t}dt+\frac{\partial f(t,W_t)}{\delta W_t}dW_t+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial W_t^2}dt\\
&= \left[\frac{\partial f(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial W_t^2}\right]dt+\frac{\partial f(t,W_t)}{\delta W_t}dW_t
\end{aligned}$$

- Chain rule:
  - Let $dx(t,W_t)=a(x,t)dt+b(x,t)dW_t$
  - For $f(x(t,W_t))$,

$$\begin{aligned}
df(x(t,W_t)) &= \left[a(x,t)\frac{\partial f(x)}{\partial x}+\frac{1}{2}b^2(x,t)\frac{\partial^2 f}{\partial x^2}\right]dt+b(x,t)\frac{\partial f}{\partial x}dW_t\\
\because\quad dx(t,W_t) &= \left[\frac{\partial x(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 x(t,W_t)}{\partial W_t^2}\right]dt+\frac{\partial x(t,W_t)}{\delta W_t}dW_t\\
\therefore\quad df(x(t,W_t)) &= \left[\left(\frac{\partial x(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 x(t,W_t)}{\partial W_t^2}\right)\frac{\partial f(x)}{\partial x}+\frac{1}{2}\left(\frac{\partial x(t,W_t)}{\delta W_t}\right)^2\frac{\partial^2 f(x)}{\partial x^2}\right]dt+\frac{\partial x(t,W_t)}{\delta W_t}\frac{\partial f(x)}{\partial x}dW_t
\end{aligned}$$

  - This is called the one-dimensional Itô's formula

- Traditional product rule:

$$\begin{aligned}
d(f(t)g(t)) &= f(t+dt)g(t+dt)-f(t)g(t)\\
&= [f(t+dt)-f(t)]\,g(t+dt)+f(t)\,[g(t+dt)-g(t)]\\
&= [f(t+dt)-f(t)]\,[g(t+dt)-g(t)]+[f(t+dt)-f(t)]\,g(t)+f(t)\,[g(t+dt)-g(t)]\\
&= df(t)dg(t)+g(t)df(t)+f(t)dg(t)\\
&= \frac{df(t)}{dt}\frac{dg(t)}{dt}(dt)^2+\frac{df(t)}{dt}g(t)dt+\frac{dg(t)}{dt}f(t)dt
\end{aligned}$$

  - If dt is infinitesimal, $(dt)^2=0$ and we get

$$d(f(t)g(t))=\frac{df(t)}{dt}g(t)dt+\frac{dg(t)}{dt}f(t)dt$$

- Product rule in Itô's calculus:

$$\begin{aligned}
d(f(t,W_t)g(t,W_t)) &= f(t+dt,W_t+dW_t)g(t+dt,W_t+dW_t)-f(t,W_t)g(t,W_t)\\
&= [f(t+dt,W_t+dW_t)-f(t,W_t)]\,g(t+dt,W_t+dW_t)+f(t,W_t)\,[g(t+dt,W_t+dW_t)-g(t,W_t)]\\
&= [f(t+dt,W_t+dW_t)-f(t,W_t)]\,[g(t+dt,W_t+dW_t)-g(t,W_t)]\\
&\quad+[f(t+dt,W_t+dW_t)-f(t,W_t)]\,g(t,W_t)+f(t,W_t)\,[g(t+dt,W_t+dW_t)-g(t,W_t)]\\
&= df(t,W_t)dg(t,W_t)+g(t,W_t)df(t,W_t)+f(t,W_t)dg(t,W_t)\\
\because\quad df(t,W_t) &= \left(\frac{\partial f(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial W_t^2}\right)dt+\frac{\partial f(t,W_t)}{\delta W_t}dW_t\\
dg(t,W_t) &= \left(\frac{\partial g(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 g(t,W_t)}{\partial W_t^2}\right)dt+\frac{\partial g(t,W_t)}{\delta W_t}dW_t\\
\therefore\quad d(f(t,W_t)g(t,W_t)) &= \left[\left(\frac{\partial f(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial W_t^2}\right)dt+\frac{\partial f(t,W_t)}{\delta W_t}dW_t\right]
\end{aligned}$$

$$
\begin{aligned}
&\left[\left(\frac{\partial g(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 g(t,W_t)}{\partial W_t^2}\right)dt+\frac{\partial g(t,W_t)}{\delta W_t}dW_t\right]\\
&+\left[\left(\frac{\partial f(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial W_t^2}\right)dt+\frac{\partial f(t,W_t)}{\delta W_t}dW_t\right]g(t,W_t)\\
&+\left[\left(\frac{\partial g(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 g(t,W_t)}{\partial W_t^2}\right)dt+\frac{\partial g(t,W_t)}{\delta W_t}dW_t\right]f(t,W_t)\\
=&\left(\frac{\partial f(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial W_t^2}\right)\left(\frac{\partial g(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 g(t,W_t)}{\partial W_t^2}\right)(dt)^2\\
&+\frac{\partial f(t,W_t)}{\delta W_t}\left(\frac{\partial g(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 g(t,W_t)}{\partial W_t^2}\right)dtdW_t\\
&+\left(\frac{\partial f(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial W_t^2}\right)\frac{\partial g(t,W_t)}{\delta W_t}dtdW_t+\frac{\partial f(t,W_t)}{\delta W_t}\frac{\partial g(t,W_t)}{\delta W_t}(dW_t)^2\\
&+\left[\left(\frac{\partial f(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial W_t^2}\right)dt+\frac{\partial f(t,W_t)}{\delta W_t}dW_t\right]g(t,W_t)\\
&+\left[\left(\frac{\partial g(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 g(t,W_t)}{\partial W_t^2}\right)dt+\frac{\partial g(t,W_t)}{\delta W_t}dW_t\right]f(t,W_t)\\
=&\frac{\partial f(t,W_t)}{\delta W_t}\frac{\partial g(t,W_t)}{\delta W_t}(dW_t)^2+\left[\left(\frac{\partial f(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial W_t^2}\right)dt+\frac{\partial f(t,W_t)}{\delta W_t}dW_t\right]g(t,W_t)\\
&+\left[\left(\frac{\partial g(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 g(t,W_t)}{\partial W_t^2}\right)dt+\frac{\partial g(t,W_t)}{\delta W_t}dW_t\right]f(t,W_t)\\
=&\frac{\partial f(t,W_t)}{\delta W_t}\frac{\partial g(t,W_t)}{\delta W_t}(dW_t)^2\\
&+\left[\left(\frac{\partial f(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial W_t^2}\right)g(t,W_t)+\left(\frac{\partial g(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 g(t,W_t)}{\partial W_t^2}\right)f(t,W_t)\right]dt\\
&+\left[\frac{\partial f(t,W_t)}{\delta W_t}g(t,W_t)+\frac{\partial g(t,W_t)}{\delta W_t}f(t,W_t)\right]dW_t
\end{aligned}
$$

which yields the following Itô's rule:

$$
\begin{aligned}
&d\left(f(t,W_t)g(t,W_t)\right)=\\
&\left[\left(\frac{\partial f(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 f(t,W_t)}{\partial W_t^2}\right)g(t,W_t)+\left(\frac{\partial g(t,W_t)}{\partial t}+\frac{1}{2}\frac{\partial^2 g(t,W_t)}{\partial W_t^2}\right)f(t,W_t)+\frac{\partial f(t,W_t)}{\delta W_t}\frac{\partial g(t,W_t)}{\delta W_t}\right]dt\\
&\qquad+\left[\frac{\partial f(t,W_t)}{\delta W_t}g(t,W_t)+\frac{\partial g(t,W_t)}{\delta W_t}f(t,W_t)\right]dW_t
\end{aligned}
$$

2.3 Stochastic Differential Equations

- Stochastic differential equations is of the form:

$$dX(t)=f(X(t))dt+\sum_{i=1}^{n}g_i(X(t))dN_i(t)$$

  where $\{X(t)\}$ is a stochastic process described by the stochastic differential equation, $N_i(t)$ are Poisson counters that drives $X(t)$, and $f(x)$, $g_i(x)$ are real-valued functions

  - Poisson counter $N_i(t)$ with $dN_i(t)=0$ if no event i occur at t and $dN_i(t)=1$ if event $i$occur at t

- Properties of stochastic differential equations

  1. If $h(t)=h(X(t))$ is a function of a stochastic process, then (without proof)

$$
\begin{aligned}
dh(t)&=\frac{dh(t)}{dt}\left(f\left(X(t)\right)dt+\sum_{i=1}^{n}g_i\left(X(t)\right)dN_i(t)\right)\\
&=\frac{dh(t)}{dt}f\left(X(t)\right)dt+\sum_{i=1}^{n}\frac{dh(t)}{dt}g_i\left(X(t)\right)dN_i(t)
\end{aligned}
$$

$$= \frac{dh(t)}{dt} f(X(t))\,dt + \sum_{i=1}^{n} \left[h(X(t) + g_i(X(t))) - h(X(t))\right] dN_i(t) \tag{2.5}$$

2. Let λ_i be the rate associated with $N_i(t)$, then

$$\frac{dE[X(t)]}{dt} = E[f(X(t))] + \sum_{i=1}^{n} \lambda_i E[g_i(X(t))] \tag{2.6}$$

- Example of using stochastic differential equations: Analyzing M/G/1 queue
  - Arrival is represented by Poisson counting process $\{N(t)\}$ with arrival rate λ, general service time X
  - Let $W(t)$ be the amount of work in the system (which can also be the queueing time of the customer arriving at t), then

$$\begin{aligned} dW(t) &= \begin{cases} -dt + XdN(t) & W(t) > 0 \\ XdN(t) & W(t) = 0 \end{cases} \\ &= -\mathbf{1}\{W(t) > 0\}dt + XdN(t) \end{aligned}$$

  - By (2.6), we have

$$\begin{aligned} \frac{dE[W(t)]}{dt} &= -E[\mathbf{1}\{W(t) > 0\}] + \lambda E[X] \\ &= -\Pr[W(t) > 0] + \lambda E[X] \end{aligned}$$

  - If the system is stable, $\rho \triangleq \lambda E[X] < 1$ and $dE[W(t)]/dt = 0$, hence

$$\begin{aligned} \frac{dE[W(t)]}{dt} &= 0 \\ \therefore \quad -\Pr[W(t) > 0] + \lambda E[X] &= 0 \\ \Pr[W(t) > 0] &= \lambda E[X] = \rho \end{aligned}$$

  - Similarly, we have:

$$\begin{aligned} dW^2(t) &= 2W(t)dW(t) \\ &= -2W(t)\mathbf{1}\{W(t) > 0\}dt + 2W(t)XdN(t) \\ &= -2W(t)\mathbf{1}\{W(t) > 0)dt + \left((W(t) + X)^2 - W^2(t)\right) dN(t) \qquad \text{(why?)} \\ \frac{dE[W^2(t)]}{dt} &= -2E[W(t)\mathbf{1}\{W(t) > 0\}] + \lambda \left(2E[W(t)X] + E[X^2]\right) \\ &= -2E[W(t)] + \lambda \left(2E[W(t)]E[X] + E[X^2]\right) \\ &= -2E[W(t)] + 2\rho E[W(t)] + \lambda E[X^2] \end{aligned}$$

  - In steady state, $dE[W^2(t)]/dt = 0$ which yields the Pollaczek-Khinchin formula

$$\begin{aligned} 0 &= -2E[W(t)] + 2\rho E[W(t)] + \lambda E[X^2] \\ 2(1-\rho)E[W(t)] &= \lambda E[X^2] \\ E[W(t)] = E[W_q] &= \frac{\lambda E[X^2]}{2(1-\rho)} \end{aligned}$$

2.4 References

- [4]
- [10], Chapter 7, “The Black-Scholes Formula”
- [11], Chapter 10, “Brownian Motion and Stationary Processes”
- [15]

Chapter 3

Stochastic Processes

3.1 Balance Equations

3.1.1 The queue

- Interarrival and service time distributions are both Markovian
- Probability measure: $p_n(t)$ defined as the probability that there are n units in the system at time t
- Considering $t \rightarrow t+\delta t$ where δt is really a short time
  - Can be any of the following:
    1. No arrival and no departure
    2. Only an arrival takes place
    3. Only a departure takes place
    4. Both departure and arrival occurs
  - Probability of an arrival occur in interval δt: $\lambda \delta t$
  - Probability of a departure occur in interval δt: $\mu \delta t$
  - Hence for $n \geq 0$,

$$\begin{aligned} p_n(t+\delta t) &= p_n(t)(1-\lambda_n \delta t)(1-\mu_n \delta t) \\ &\quad + p_n(t)(\lambda_n \delta t)(\mu_n \delta t) \\ &\quad + p_{n+1}(t)(\mu_{n+1} \delta t)(1-\lambda_{n+1} \delta t) \\ &\quad + p_{n-1}(t)(\lambda_{n-1} \delta t)(1-\mu_{n-1} \delta t) \\ &\quad + o(\delta t) \end{aligned}$$

  and for $n = 0$,

$$\begin{aligned} p_0(t+\delta t) &= p_0(t)(1-\lambda_0 \delta t) \\ &\quad + p_1(t)(\mu_1 \delta t)(1-\lambda_1 \delta t) \\ &\quad + o(\delta t) \end{aligned}$$

  - Take differentiation on $p_n(t)$:

$$\frac{d}{dt} p_n(t) = -(\lambda_n + \mu_n) p_n(t) + \lambda_{n-1} p_{n-1}(t) + \mu_{n+1} p_{n+1}(t)$$
$$\frac{d}{dt} p_0(t) = -\lambda_0 p_0(t) + \mu_1 p_1(t)$$

- Steady state probability is defined as $\frac{d}{dt} p_n(t) = 0$.
  - Hence the above differential equations become the balance equations

3.1.2 References

- [12] Sharma (1990), Chapter 1

3.2 Stochastic processes

3.2.1 Birth-Death Process

- $\{N(t) : t \geq 0\}$ is a birth-death process if

$$\Pr[N(t+\delta t) = k | N(t) = j] = \begin{cases} \lambda_j \delta t & k = j+1 \\ \mu_j \delta t & k = j-1 \\ 0 & |k-j| \geq 2 \end{cases} \qquad j,k = 0,1,\ldots$$

- Birth-death process is a Markovian model, as its state depends only on the previous state
- Balance equation:

$$(\lambda_j + \mu_j) p_j = \lambda_{j-1} p_{j-1} + \mu_{j+1} p_{j+1}$$
$$\lambda_0 p_0 = \mu_1 p_1$$

  – Solution: $p_j = \dfrac{\lambda_0 \lambda_1 \cdots \lambda_{j-1}}{\mu_1 \mu_2 \cdots \mu_j}$ and $p_0 = \left(1 + \displaystyle\sum_{j=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{j-1}}{\mu_1 \mu_2 \cdots \mu_j}\right)^{-1}$

3.2.2 Renewal process

- Renewal theory is describing the process of replacement. In a system, component N is on its duty cycle, and it will fail some time. Once it is failed, component $N+1$ will replace its role. Renewal theory is describing the partial sum:

$$S_N = X_1 + X_2 + \cdots + X_N$$

  where each of X_i is the random variable of the lifetime of component i.

- Number of renewals: $U(t) = \max\{N \geq 0 : S_N \leq t\}$ is the number of renewals in $[0,t]$
- If the lifetime of component is exponential, i.e. X_i has p.d.f $\lambda e^{-\lambda t}$, and the time of the k-th renewal is in Gamma distribution:

$$f_k(t) = e^{-\lambda t} \lambda^k \frac{t^{k-1}}{(k-1)!}$$

- Alternatively, the probability that there are exactly n renewals in time interval $[0,t)$ is in Poisson distribution:

$$\Pr_{[0,t)}\{N = n\} = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

- The expected number of renewals per unit time equals to mean lifetime:

$$\lim_{t \to \infty} \frac{U(t)}{t} = \lambda$$

3.2.3 Reference

- [11] Ross (2003), Chapter 5, “The Exponential Distribution and the Poisson Process”; Chapter 6, “Continuous-Time Markov Chains”
- [2] Akimaru and Konosuke (1999), Chapter 2, “Markovian Models”

Chapter 4

Queueing Theory

4.1 Kendall Notation

- To abbreviate the description of a general queueing system
- A/B/C
  - A, B: *interarrival* and service time distributions
  - C: number of channels or service counters
- A/B/C/D/E
  - D: Buffer size, i.e. system capacity. Tandem queue means D$\neq \infty$
  - E: Customer population
- Common symbols for distributions:
  - M: Markovian distribution, i.e. exponential
  - D: Deterministic distribution
  - E_k: Erlang-k distribution
  - H_k: Hyper-exponential of order k
  - G: General distribution
  - GI: General distribution with independent inter-arrival or service times (renewal)
  - $MMPP$: Markov modulated Poisson process (non-renewal)

4.2 Different Queues

4.2.1 M/M/1

- Birth-Death process with state-indepentent arrival (birth) rate λ and departure (death) rate μ
- Balance equation for system state (population in system):

$$\begin{aligned}
\lambda \pi_{n-1} &= \mu \pi_n \\
\pi_n &= \frac{\lambda}{\mu}\pi_{n-1} = \rho \pi_{n-1} \\
\therefore \quad \pi_n &= \rho^n \pi_0 \\
\sum_{n=0}^{\infty} \rho^n \pi_0 &= 1 \\
\pi_0 &= 1 - \rho \\
\pi_n &= \rho^n (1-\rho)
\end{aligned}$$

- Expected population in system: $E[N] = \sum_{k=0}^{n} k\pi_k = \rho/(1-\rho)$
  - Variance: $Var[N] = \rho/(1-\rho)^2$
- Expected waiting time in queue for the $(m+1)$-th user: $E[W_q|m] = mE[S] = m/\mu$
  - Expected waiting time: $E[W_q] = \sum_{k=0}^{\infty} \pi_k E[W_q|m=k] = \frac{1}{\mu}\sum_{k=0}^{\infty} k\pi_k = \frac{1}{\mu}\frac{\rho}{1-\rho}$
    * Variance: $Var[W_q] = \frac{2\rho-\rho^2}{\mu^2(1-\rho)^2}$
  - Probability of not waiting: $\Pr[W_q = 0] = 1-\rho$
  - Density function for waiting time t given k customers in front (Erlang-k distribution): $p(t) = \frac{e^{-\lambda t}\lambda^k t^{k-1}}{(k-1)!}$
  - Waiting time distribution:

$$\begin{aligned}
\Pr[W_q \le X] &= \Pr[W_q = 0] + \Pr[0 < W_q \le X] \\
&= (1-\rho) + \sum_{k=0}^{\infty} \left(\Pr[N=k]\Pr[E_k \le X]\right) \\
&= (1-\rho) + \sum_{k=0}^{\infty} \left((1-\rho)\rho^k \int_0^X \frac{e^{-\mu t}\mu^k t^{k-1}}{(k-1)!} dt\right) \\
&= 1 - \rho^{-(\mu-\lambda)X} \\
\Pr[W_q = t] &= \frac{d}{dt}\left(1 - \rho^{-(\mu-\lambda)t}\right) \\
&= \rho(\mu-\lambda)e^{-(\mu-\lambda)t}
\end{aligned}$$

  - Expected queue length (by Little's law): $E[N_q] = \lambda E[W_q] = \frac{\rho^2}{1-\rho}$
- Expected sojourn time in system for the $(m+1)$-th user: $E[W] = E[W_q] + \frac{1}{\mu} = \frac{1}{\mu}\frac{1}{1-\rho} = \frac{1}{\mu-\lambda}$
  - Variance: $Var[W] = \frac{1}{(\mu-\lambda)^2}$
  - Sojourn time distribution: $f_T(t) = (\mu-\lambda)e^{-(\mu-\lambda)t}$
  - c.d.f.: $F_T(t) = 1 - e^{-(\mu-\lambda)t}$
  - c.d.f. of waiting time: $F_W = 1 - \rho e^{-(\mu-\lambda)t}$
- PASTA: Poisson Arrivals See Time Average
  - Poisson arrival with rate λ
  - System state at time t given an arrival occur in $(t, t+\Delta t)$: $M(t)$
  - System state at time t: $N(t)$
  - PASTA:

$$\begin{aligned}
\Pr\{M(t)=n\} &= \Pr\{N(t)=n|\text{arrival in } (t,t+\Delta t)\} \\
&= \frac{\Pr\{N(t)=n\}\Pr\{\text{arrival in } (t,t+\Delta t)\}}{\Pr\{\text{arrival in } (t,t+\Delta t)\}} \\
&= \Pr\{N(t)=n\}
\end{aligned}$$

- Burke's Theorem: Departure process of M/M/1 queue is Poisson with rate λ independent of arrival process
  - Poisson input implies Poisson output

4.2.2 M/G/1 Queue with FCFS discipline

- Poisson arrival with λ and i.i.d. service time X with density function $f_X(x)$
- Analysis using Embedded Markov chain approach:
  - Let Y_n be the number of customers in the system immediately after the departure of customer n
  - Let A_n be the number of arrivals during the service time of customer n
  - Markov chain:

$$\begin{aligned} Y_{n+1} &= \begin{cases} Y_n - 1 + A_{n+1} & Y_n > 0 \\ A_{n+1} & Y_n = 0 \end{cases} \\ &= Y_n + A_{n+1} - U(Y_n) \end{aligned}$$

where $U(x)$ is the unit step function such that $U(x) = 1$ only if $x > 0$ and $U(x) = 0$ otherwise.

    * This markov chain is ergodic with period $\rho = \lambda E[X] < 1$

  - Steady value of Y_n: Taking limit $n \to \infty$

$$\begin{aligned} Y_{n+1} &= Y_n + A_{n+1} - U(Y_n) \\ \therefore \quad \lim_{n\to\infty} (E[Y_{n+1}]) &= \lim_{n\to\infty} (E[Y_n] + E[A_{n+1}] - E[U(Y_n)]) \\ E[Y] &= E[Y] + E[A] - E[U] \\ E[U] &= E[A] \\ &= \lambda \cdot \frac{1}{\mu} = \rho \end{aligned}$$

  - Expected value of Y^2:

$$\begin{aligned} E[Y_{n+1}^2] &= E[Y_n^2 + A_{n+1}^2 + U(Y_n)^2 + 2(Y_nA_{n+1} - Y_nU(Y_n) - A_{n+1}U(Y_n))] \\ &= E[Y_n^2] + E[A_{n+1}^2] + E[U(Y_n)^2] + 2(E[Y_nA_{n+1}] - E[Y_nU(Y_n)] - E[A_{n+1}U(Y_n)]) \\ E[Y^2] &= E[Y^2] + E[A^2] + E[U^2] + 2E[YA] - 2E[YU] - 2E[AU] \\ &= E[Y^2] + E[A^2] + E[U] + 2E[Y]E[A] - 2E[Y] - 2E[A]E[U] \\ 0 &= E[A^2] + 2E[Y](E[A] - 1) + E[U](1 - 2E[A]) \\ &= E[A^2] + 2E[Y](\rho - 1) + \rho(1 - 2\rho) \\ 2E[Y](1-\rho) &= E[A^2] + \rho - 2\rho^2 \\ E[Y] &= \frac{E[A^2] + \rho - 2\rho^2}{2(1-\rho)} = \frac{E[A^2] + 2\rho - 2\rho^2 - \rho}{2(1-\rho)} = \frac{E[A^2] + 2\rho(1-\rho) - \rho}{2(1-\rho)} \\ &= \rho + \frac{E[A^2] - \rho}{2(1-\rho)} \end{aligned}$$

    * $E[U^2] = E[U]$ as $1^2 = 1$ and $0^2 = 0$
    * Y and A are independent, hence $E[YA] = E[Y]E[A]$
    * $E[YU] = E[Y]$ as $U = 0$ only if $Y = 0$ and $U = 1$ if $Y \neq 0$
    * A is defined only if there is a user to serve, hence $E[AU] = E[A]E[U]$

  - Z-transform of $\Pr[A = n]$ is the Laplace transform of $f_X(t)$ with $s = \lambda(1-z)$:

$$\begin{aligned} A(z) &= \sum_{n=0}^{\infty} \Pr[A = n] z^n \\ &= \sum_{n=0}^{\infty} z^n \int_0^{\infty} \Pr[A = n | X = t] f_X(t) dt \\ &= \sum_{n=0}^{\infty} z^n \int_0^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} f_X(t) dt \\ &= \int_0^{\infty} \sum_{n=0}^{\infty} \frac{(\lambda t z)^n}{n!} e^{-\lambda t} f_X(t) dt \end{aligned}$$

$$
\begin{aligned}
&= \int_0^{\infty} e^{\lambda t z} e^{-\lambda t} f_X(t) dt \\
&= \int_0^{\infty} e^{-\lambda t(1-z)} f_X(t) dt \\
&= F_X(\lambda(1-z))
\end{aligned}
$$

– $E[A^2]$ can be obtained by $A(z)$:

$$
\begin{aligned}
E[A^2] &= A''(z)|_{z=1} + A'(z)|_{z=1} \\
&= \lambda^2 E[X^2] + \lambda E[X] \\
&= \lambda^2 E[X^2] + \rho
\end{aligned}
$$

– Substitute $E[A^2]$ to $E[Y]$ and obtain the Pollaczek-Khinchin formula:

$$
\begin{aligned}
E[Y] &= \rho + \frac{E[A^2] - \rho}{2(1-\rho)} \\
&= \rho + \frac{\lambda^2 E[X^2]}{2(1-\rho)}
\end{aligned}
$$

4.2.3 Waiting time in M/G/1 queue using Little's Law

- Assume the average service time to be $\overline{X}$ and the customers are arrived in Poisson process with rate λ
- With respect to customer i, we define:
  - W_i: Waiting time in queue
  - R_i: Residual service time as seen by customer i. That is the time for the on-going service to complete by the epoch customer i arrives. $R_i = 0$ if the queue is empty.
  - X_i: Service time received
  - N_i: Number of customers already in queue upon the arrival of customer i
- Now we have:

$$
\begin{aligned}
W_i &= R_i + \sum_{j=i-N_i}^{i-1} X_j \\
E[W_i] &= E[R_i] + E\left\{ \sum_{j=i-N_i}^{i-1} X_j \right\} \\
&= E[R_i] + E\left\{ \sum_{j=i-N_i}^{i-1} E[X_j] \right\} \\
&= E[R_i] + \overline{X} E[N_i] \\
\therefore \quad W_q &= R + \overline{X} N_q
\end{aligned}
$$

- By Little's law,

$$
\begin{aligned}
N_q &= \lambda W_q \\
\therefore \quad W_q &= R + \overline{X}(\lambda W_q) \\
&= R + (\lambda \overline{X}) W_q \\
W_q &= \frac{R}{1 - \lambda X} = \frac{R}{1-\rho}
\end{aligned}
$$

4.2.4 Alternative way to derive Pollaczek-Khinchin formula in M/G/1 queue

- Let the residue service time at t to be $r(t)$ and there are $N(t)$ customers ever completed service by t
- Time-averaged residue service time:

$$\begin{aligned}
R = \lim_{t\to\infty} \frac{1}{t}\int_0^t r(\tau)d\tau &= \lim_{t\to\infty}\frac{1}{t}\sum_{n=1}^{N(t)}\frac{X_n^2}{2} \\
&= \lim_{t\to\infty}\left(\sum_{n=1}^{N(t)}\frac{N(t)}{t}\cdot\frac{X_n^2/2}{N(t)}\right) \\
&= \lim_{t\to\infty}\left(\frac{N(t)}{t}\cdot\frac{\sum_{n=1}^{N(t)}X_n^2}{N(t)}\cdot\frac{1}{2}\right) \\
&= \frac{1}{2}\cdot\lim_{t\to\infty}\frac{N(t)}{t}\cdot\lim_{t\to\infty}\frac{\sum_{n=1}^{N(t)}X_n^2}{N(t)} \\
\therefore \quad R &= \frac{1}{2}\lambda E[X^2]
\end{aligned}$$

- Hence the mean waiting time in queue:

$$\begin{aligned}
W_q &= \frac{R}{1-\rho} \\
&= \frac{\lambda E[X^2]}{2(1-\rho)}
\end{aligned}$$

  - Even if $\rho < 1$, $W_q = \infty$ is possible if $E[X^2] = \infty$, i.e. variance of service time is too large
  - Sojourn time: $W = \bar{X} + W_q = \bar{X} + \dfrac{\lambda E[X^2]}{2(1-\rho)}$
  - Queue length: $N_q = \lambda W_q = \dfrac{\lambda^2 E[X^2]}{2(1-\rho)}$
  - Population in system: $N = \lambda W = \lambda\bar{X} + \dfrac{\lambda^2 E[X^2]}{2(1-\rho)} = \rho + \dfrac{\lambda^2 E[X^2]}{2(1-\rho)}$

- P-K formula: If the service model is known and we can calculate $E[X^2]$, we can obtain W_q
  - M/M/1: $E[X^2] = \dfrac{2}{\mu} \quad \Rightarrow \quad W_q = \dfrac{\rho}{\mu(1-\rho)}$
  - M/D/1: $E[X^2] = \dfrac{1}{\mu^2} \quad \Rightarrow \quad W_q = \dfrac{\rho}{2\mu(1-\rho)}$

4.2.5 Mean-Value Analysis of M/G/1 FIFO queue

- Mean value analysis:

$$E[W_q] = E[N_q]E[X] + E[R]$$

  where W_q is the queueing time, N_q is the population in queue, X is the service time, and R is the remaining service time where an empty queue yields $R = 0$.

  - The equation means the expected waiting time for a newly arriving customer is the sum of the expected remaining service time of the customer in service, plus the expected service time to finish all customers in front

- If the service time is exponential, $E[R] = \rho\cdot\frac{1}{\mu}$. Using Little's Law: $E[N_q] = \lambda E[W_q]$,

$$\begin{aligned}
E[W_q] &= E[N_q]E[X] + E[R] \\
&= \lambda E[W_q]\frac{1}{\mu} + \rho\frac{1}{\mu}
\end{aligned}$$

$$(1-\rho)E[W_q] = \rho\frac{1}{\mu}$$
$$E[W_q] = \frac{\rho}{1-\rho}\frac{1}{\mu}$$

- If the service is generally distributed, $E[R] = \rho\left(\frac{E[X]}{2} + \frac{\sigma_X^2}{2E[X]}\right)$. Hence

$$\begin{aligned}
E[W_q] &= E[N_q]E[X] + E[R] \\
&= \lambda E[W_q]E[X] + \rho\left(\frac{E[X]}{2} + \frac{\sigma_X^2}{2E[X]}\right) \\
&= \rho E[W_q] + \rho\frac{E[X]}{2} + \frac{\lambda\sigma_X^2}{2} \\
(1-\rho)E[W_q] &= \frac{\rho E[X] + \lambda\sigma_X^2}{2} \\
E[W_q] &= \frac{\rho E[X] + \lambda\sigma_X^2}{2(1-\rho)}
\end{aligned}$$

  and the mean sojourn time is $E[W] = E[W_q] + E[X]$.

4.2.6 General queue: GI/G/c

- Multiserver queue with $c \geq 1$ identical servers
- Customers: General independent interarrival times with p.d.f. $A(t)$, arrival rate is λ
  - Expected service times: $E(S)$
  - Offered load: $\lambda E(S)$
  - Server utilization: $\rho = \lambda E(s)/c$
- Random variables:
  - $N(t)$: No. of customers in the system at time t, including those in service
  - $N_q(t)$: No. of customers in the queue at time t
  - D_n: Delay in queue of the n-th customer
  - R_n: Sojourn time of the n-th customer
  - $V(t)$: The workload at time t, i.e., the sum of service times of all customers in queue plus the sum of remaining service times of the customers in service at time t
- Definitions:
  - Probability of system state: $p_j = \lim_{t\to\infty} \Pr\{N(t) = j\}$
  - Waiting time in queue: $W_q(x) = \lim_{n\to\infty} \Pr\{D_n \leq x\}$, also, waiting time in system: $W(x) = \lim_{n\to\infty} \Pr\{R_n \leq x\}$
- Averages in the long-run:
  - $\lim_{t\to\infty} \frac{1}{t}\int_0^t N(u)du = E(N)$, and $\lim_{t\to\infty} \frac{1}{t}\int_0^t N_q(u)du = E(N_q)$
  - $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} D_k = E(W_q)$, and $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} R_k = E(W)$
  - $\lim_{t\to\infty} \frac{1}{t}\int_0^t V(u)du = E(V)$
- Little's Law:
  - $E(N_q) = \lambda E(W_q)$, and $E(N) = \lambda E(W)$
  - $E(\#\text{ busy servers}) = \lambda E(S)$

* If p_j is the probability of having j users in the system,

$$\begin{aligned}
E(\#\text{ busy servers}) &= \lambda E(S) \\
&= \sum_{j=0}^{c-1} j p_j + c \sum_{j=c}^{\infty} p_j \\
\lambda E(S) &= \sum_{j=0}^{c-1} j p_j + c\left(1 - \sum_{j=c}^{\infty} p_j\right)
\end{aligned}$$

- Expected amount of work in system: $E(V)$
  - Define function: $v(t)$ as the remaining amount of service to complete for a particular user
    * In waiting time, $v(t) = S$ as the amount of service S is never started
    * Receiving service for amount of time x, $v(t) = S - x$
    * Just completed service: $v(t) = S - S = 0$
  - $E(V) = \lambda E(\int_0^{W_q+S} v(t)dt)$
  - By Little's law, the amount of work in the system is the product of arrival rate and the amount of work due per user

$$\begin{aligned}
E(V) &= \lambda E(\int_0^{W_q+S} v(t)dt) \\
&= \lambda E(W_q S + \int_0^{S} (S-x)dt) \\
&= \lambda E(W_q)E(S) + E(\frac{1}{2}S^2)
\end{aligned}$$

  - Hence, $E(V) = \lambda E(W_q)E(S) + \frac{1}{2}E(S^2)$

4.2.7 References

- [14] Towsley (2002)
- [13] Tijms (1986), Chapter 4, Section 4.1
- [7] Nain (1998), Section 3
- [9] Prabhu (1997), Chapter 2, "Markovian Queueing Systems"; Chapter 3, "The Busy Period, Output and Queues in Series"

4.3 Markovian Delay Systems

4.3.1 M/M/c queue with infinite buffer: Delay system

- Service time: $E(S) = 1/\mu$
- Utilization: $\rho = \lambda E(S)/c$
- Time-average probabilities: (same as customer-average probabilities)

$$\begin{aligned}
\lambda p_{j-1} &= \min(j,c)\mu \quad j = 1,2,\ldots \\
\therefore\ p_j &= \begin{cases} \dfrac{(c\rho)^j}{j!} p_0 & 0 \le j \le c-1 \\ \dfrac{(c\rho)^j}{c!c^{j-c}} p_0 & j \ge c \end{cases} \\
p_0 &= \left\{ \sum_{k=0}^{c-1} \frac{(c\rho)^k}{k!} + \frac{(c\rho)^c}{c!(1-\rho)} \right\}^{-1}
\end{aligned}$$

- Delay probability: $\Pi_W = \sum_{j=c}^{\infty} p_j = \frac{(c\rho)^c}{c!(1-\rho)} p_0$

  - Erlang-C, a.k.a. Erlang's delay formula (blocked-calls-held system)

- Average queue size: $E(L_q) = \sum_{j=c}^{\infty}(j-c)p_j = \frac{(c\rho)^c\rho}{c!(1-\rho)^2} p_0 = \frac{\rho}{1-\rho}\Pi_W$

- If FIFO, the waiting-time distribution is drived as follows:

  - If $N \leq c$, $N_q = 0$
  - If $N > c$, service rate from the point of view of the system: $c\mu$
  - Distribution of service rate (inter-departure distribution): $S(x) = e^{-c\mu x}$
  - Probability that k users completed service in duration length x: $P(k) = \frac{e^{-c\mu x}(c\mu x)^k}{k!}$
  - Given there are j users in queue, for duration x, the probability that there are less than j users leave the system: $\sum_{k=0}^{j} \frac{e^{-c\mu x}(c\mu x)^k}{k!}$
  - Distribution of system population N: p_j, which corresponding to j users in the system and $j-c$ users in queue
  - $\Pr\{W_q > x\} = \sum_{j=c}^{\infty} p_j \sum_{k=0}^{j-c} \frac{e^{-c\mu x}(c\mu x)^k}{k!} = \Pi_W e^{-c\mu(1-\rho)x}$, for $x \geq 0$.
  - $\Pr\{W_q = 0\} = 1 - \Pi_W$
  - Average delay: $E(W_q) = \frac{(c\rho)^c}{c!c\mu(1-\rho)^2} p_0 = \frac{\Pi_W}{c\mu(1-\rho)}$
  - Average sojourn time: $E(W) = E(W_q) + \frac{1}{\mu} = \frac{\Pi_W}{c\mu(1-\rho)} + \frac{1}{\mu}$
  - Average queue length is provided by Little's formula, $E(N_q) = \lambda E(W_q)$ or average system population is provided by $E(N) = \lambda E(W)$

4.3.2 M/M/c/∞/N queue: Limited customer pool (need verify)

- Arrival at state $n \leq N$: $\lambda_n = (N-n)\lambda$

- Exponential service time with mean $E[X] = 1/\mu$

- System service rate at state n: $\mu_n = \begin{cases} n\mu & 0 \leq n \leq s \\ s\mu & s \leq n \end{cases}$

- Balance equations

$$((N-j)\lambda + \min(j,s)\mu)\, p_j = (N-j+1)\lambda p_{j-1} + \min(j+1,s)\mu p_{j+1}$$

$$\Longrightarrow \quad p_j = \begin{cases} \binom{n}{N} \rho^n p_0 & 0 \leq n \leq s \\ \frac{1}{s!}\rho^s \frac{n!}{s!s^{n-s}} \rho^n p_0 & s \leq n \leq N \end{cases}$$

$$p_0 = \left\{ \sum_{n=0}^{s-1} \binom{n}{N} \rho^n + \sum_{n=s}^{N} \binom{N}{n} \frac{n!}{s!s^{n-s}} \right\}^{-1}$$

with $\rho = \lambda/\mu$.

4.3.3 M/D/c queue

- Determinstic departure with service time D
- Utilization: $\rho = \lambda D/c$
- Derivation of steady probability
  - $p_k(t)$: Probability of having k users at time t
  - $p_0(t+D) = \sum_{k=0}^{c} p_k(t) e^{-\lambda D}$
    * Rationale: $\sum_k$(Prob. $k \leq c$ users in the system)(Prob. next arrival occur at $\delta t \geq D$)
  - $p_j(t+D) = \sum_{k=0}^{c} p_k(t) \frac{e^{-\lambda D}(\lambda D)^j}{j!} + \sum_{k=1}^{j} p_{c+k}(t) \frac{e^{-\lambda D}(\lambda D)^{j-k}}{(j-k)!}$
    * Rationale:
      1. If there are $k \leq c$ users, upon $t+D$, all the existing users will complete the service and leave, hence we need to have j arrivals in duration D
      2. If there are $c+k$ users, there will be c users leave and k remains, thus we need to have $j-k$ arrivals in duration D.
  - In the long run, $\lim_{t\to\infty} p_k(t) = p_k$,

$$p_0 = \left(\sum_{k=0}^{c} p_k\right) e^{-\lambda D}$$

$$p_j = \left(\sum_{k=0}^{c} p_k\right) \frac{e^{-\lambda D}(\lambda D)^j}{j!} + \sum_{k=1}^{j} p_{c+k} \frac{e^{-\lambda D}(\lambda D)^{j-k}}{(j-k)!}$$

- Z-transform:
  - Think $\{p_0,\ p_1,\ p_2,\ \dots\}$ as a sequence, the z-transform is $P(z) = \sum_{j=0}^{\infty} p_j z^j$
  - Partial sum starting with p_c: $P_q(z) = \sum_{j=c}^{\infty} p_j z^{j-c}$
  - Expected population: $E(N) = \sum_{j=0}^{\infty} j p_j = P'(1)$
  - Expected queue length: $E(N_q) = \sum_{j=c}^{\infty} (j-c) p_j = P_q'(1)$
- Derivation of $E(N)$ using z-transform:

$$p_j = \left(\sum_{k=0}^{c} p_k\right) \frac{e^{-\lambda D}(\lambda D)^j}{j!} + \sum_{k=1}^{j} p_{c+k} \frac{e^{-\lambda D}(\lambda D)^{j-k}}{(j-k)!}$$

$$\sum_{j=0}^{\infty} p_j z^j = P(z) = \sum_{j=0}^{\infty} \left\{ \left(\sum_{k=0}^{c} p_k\right) \frac{e^{-\lambda D}(\lambda D)^j}{j!} z^j + \sum_{k=1}^{j} p_{c+k} \frac{e^{-\lambda D}(\lambda D)^{j-k}}{(j-k)!} z^j \right\}$$

$$P(z) = \sum_{j=0}^{\infty} \left(e^{-\lambda D} \sum_{k=0}^{c} p_k \right) \frac{(\lambda D)^j}{j!} z^j + \sum_{j=0}^{\infty} \sum_{k=1}^{j} \left(p_{c+k} e^{-\lambda D} \right) \frac{(\lambda D)^{j-k}}{(j-k)!} z^j$$

$$= \left(e^{-\lambda D} \sum_{k=0}^{c} p_k \right) \sum_{j=0}^{\infty} \frac{(\lambda D z)^j}{j!} + \sum_{j=1}^{\infty} \sum_{k=1}^{j} \left(p_{c+k} e^{-\lambda D} z^k \right) \frac{(\lambda D z)^{j-k}}{(j-k)!} \tag{4.1}$$

$$= \left(e^{-\lambda D} \sum_{k=0}^{c} p_k \right) e^{\lambda D z} + \sum_{k=1}^{\infty} \sum_{j=k}^{\infty} \left(p_{c+k} e^{-\lambda D} z^k \right) \frac{(\lambda D z)^{j-k}}{(j-k)!} \tag{4.2}$$

$$
\begin{aligned}
&= e^{-\lambda D(1-z)} \sum_{k=0}^{c} p_k + \sum_{k=1}^{\infty} \left(p_{c+k} e^{-\lambda D} z^k \right) \sum_{j-k=0}^{\infty} \frac{(\lambda D z)^{j-k}}{(j-k)!} \\
&= e^{-\lambda D(1-z)} \sum_{k=0}^{c} p_k + \sum_{k=1}^{\infty} \left(p_{c+k} e^{-\lambda D} z^k \right) e^{\lambda D z} \\
&= e^{-\lambda D(1-z)} \sum_{k=0}^{c} p_k + \sum_{k=1}^{\infty} p_{c+k} e^{-\lambda D(1-z)} z^k \\
&= e^{-\lambda D(1-z)} \sum_{k=0}^{c} p_k + e^{-\lambda D(1-z)} \sum_{k=1}^{\infty} p_{c+k} z^k \\
&= e^{-\lambda D(1-z)} \left(\sum_{k=0}^{c} p_k + \sum_{k=1}^{\infty} p_{c+k} z^k \right) \\
P(z) z^c &= e^{-\lambda D(1-z)} \left(z^c \sum_{k=0}^{c} p_k + \sum_{k=1}^{\infty} p_{c+k} z^{c+k} \right) \\
&= e^{-\lambda D(1-z)} \left(z^c \sum_{k=0}^{c} p_k + \left(P(z) - \sum_{k=0}^{c} p_k z^k \right) \right) \\
P(z) z^c - P(z) e^{-\lambda D(1-z)} &= e^{-\lambda D(1-z)} \left(z^c \sum_{k=0}^{c} p_k - \sum_{k=0}^{c} p_k z^k \right) \\
P(z) &= \frac{e^{-\lambda D(1-z)} \left(z^c \sum_{k=0}^{c} p_k - \sum_{k=0}^{c} p_k z^k \right)}{z^c - e^{-\lambda D(1-z)}} \\
&= \frac{\sum_{k=0}^{c} p_k z^c - \sum_{k=0}^{c} p_k z^k}{z^c e^{\lambda D(1-z)} - 1}
\end{aligned}
$$

- Note: Change of second summation in (4.1) is because $\sum_{k=1}^{0} = 0$
- Note: In (4.2), by enumeration,

$$
\begin{aligned}
\sum_{j=1}^{\infty} \sum_{k=1}^{j} T_{jk} &= T_{11} \ + \ T_{21} + T_{22} \ + \ T_{31} + T_{32} + T_{33} \ + \ \dots \\
&= T_{11} + T_{21} + T_{31} + \dots \ + \ T_{22} + T_{32} + \dots \ + \ T_{33} + T_{43} + \dots
\end{aligned}
$$

  hence we can have $\sum_{j=1}^{\infty} \sum_{k=1}^{j} T_{jk} \ \equiv \ \sum_{k=1}^{\infty} \sum_{j=k}^{\infty} T_{jk}$
- Hence the differentiation of z-transform:

$$
\begin{aligned}
P(z) &= \sum_{j=0}^{c-1} p_j z^j + z^c P_q(z) \\
P'(z) &= \sum_{j=1}^{c-1} p_j j z^{j-1} + c z^{c-1} P_q(z) + z^c P_q'(z) \\
\because \ P'(z) &= \frac{\left(z^c e^{\lambda D(1-z)} - 1 \right) \left(\sum_{k=0}^{c} p_k c z^{c-1} - \sum_{k=0}^{c} p_k k z^{k-1} \right) - \left(\sum_{k=0}^{c} p_k z^c - \sum_{k=0}^{c} p_k z^k \right) \left(c z^{c-1} e^{\lambda D(1-z)} - \lambda D e^{\lambda D(1-z)} \right)}{\left(z^c e^{\lambda D(1-z)} - 1 \right)^2} \\
&= \sum_{j=1}^{c-1} p_j j z^{j-1} + c z^{c-1} P_q(z) + z^c P_q'(z) \\
\therefore \ E(N_q) = P_q'(1) &= \frac{(c\rho)^2 - c(c-1) + \sum_{j=0}^{c-1} [c(c-1) - j(j-1)] \, p_j}{2c(1-\rho)}
\end{aligned}
$$

- Crommelin's formula [3] for W_q:

$$\Pr\{W_q \le x\} = \sum_{i=0}^{c-1} Q_i \sum_{j=0}^{m} e^{\lambda(x-mD)} \frac{[-\lambda(x-mD)]^{jc+c-1-i}}{(jc+c-1-i)!}$$
$$\Pr\{W_q \le x\} = \sum_{j=1}^{\infty} \sum_{i=0}^{c-1} Q_i e^{-\lambda[(m+j)D-x]} \frac{(\lambda[(m+j)D-x])^{(m+j)c+c-1-i}}{[(m+j)c+c-1-i]!}$$

  – where $Q_i = \sum_{j=0}^{i} p_j$, and m satisfies $mD \le x < (m+1)D$

4.3.4 References

- [13], Chapter 4, Sections 4.2 and 4.4
- [9], Chapter 2, "Markovian Queueing Systems"; Chapter 3, "The Busy Period, Output and Queues in Series"; Chapter 7, "The System M/G/1, Priority Systems"; Chapter 8, "The System GI/G/1, Imbedded Markov Chains"

4.4 Markovian Loss Systems

4.4.1 M/M/1/c

- Poisson arrival with rate λ
- Balance equation: Same as M/M/1, but with $n \le c$

$$\pi_n = \rho\pi_{n-1} = \rho^n \pi_0$$

  – If $\lambda \ne \mu$,

$$\sum_{k=0}^{c} \pi_k = \frac{1-\rho^{c+1}}{1-\rho}\pi_0 = 1$$
$$\therefore \quad \pi_0 = \frac{1-\rho}{1-\rho^{c+1}}$$
$$\pi_n = \frac{1-\rho}{1-\rho^{c+1}}\rho^n$$

  – If $\lambda = \mu$,

$$\sum_{k=0}^{c} \pi_k = \pi_0 \sum_{k=0}^{c} \rho^k = (c+1)\pi_0 = 1$$
$$\therefore \quad \pi_0 = \frac{1}{c+1}$$
$$\pi_n = \frac{1}{c+1}$$

- Expected number of users in the system:

  – If $\lambda \ne \mu$,

$$E[N] = \sum_{k=0}^{c} k\pi_k$$
$$= \sum_{k=0}^{c} \frac{1-\rho}{1-\rho^{c+1}} k\rho^k$$
$$= \frac{\rho}{1-\rho} - \frac{(c+1)\rho^{c+1}}{1-\rho^{c+1}}$$

- If $\lambda = \mu$,

$$
\begin{aligned}
E[N] &= \sum_{k=0}^{c} k\pi_k \\
&= \sum_{k=0}^{c} \frac{k}{c+1} \\
&= \frac{c}{2}
\end{aligned}
$$

- Blocking probability: π_c
- Throughput (i.e. real arrival): $(1-\pi_0)\mu = (1-\pi_c)\lambda$
- $E[T] = \dfrac{E[N]}{(1-\pi_c)\lambda}$

4.4.2 M/M/c/c

- Poisson arrival and exponential service time with no buffer
- Balance equation:

$$
\begin{aligned}
p_j(t+\delta t) &= (1-\lambda\delta t - j\mu\delta t)p_j(t) \; + \; \lambda\delta t p_{j-1}(t) \; + \; (j+1)\mu\delta t p_{j+1}(t) \\
&= p_j(t) + \left[-(\lambda + j\mu)p_j(t) + \lambda p_{j-1}(t) + (j+1)\mu p_{j+1}(t)\right]\delta t \\
\frac{p_j(t+\delta t) - p_j(t)}{\delta t} &= -(\lambda + j\mu)p_j(t) + \lambda p_{j-1}(t) + (j+1)\mu p_{j+1}(t) \\
\frac{d}{dt}p_j(t) &= -(\lambda + j\mu)p_j(t) + \lambda p_{j-1}(t) + (j+1)\mu p_{j+1}(t) \\
t \to \infty: \quad \frac{d}{dt}p_j(t) &= 0 \\
&= -(\lambda + j\mu)p_j + \lambda p_{j-1} + (j+1)\mu p_{j+1} \\
(\lambda + j\mu)p_j &= \lambda p_{j-1} + (j+1)\mu p_{j+1}
\end{aligned}
$$

- Recurrence formula:

$$
\begin{aligned}
\because \quad p_{-1} &= 0 \\
\therefore \quad \lambda p_0 &= \mu p_1 \\
p_1 &= (\lambda/\mu)p_0 \\
p_j &= \frac{\lambda/\mu}{j} p_{j-1} \\
p_j &= \frac{(\lambda/\mu)^j}{j!} p_0
\end{aligned}
$$

  - Defining $\rho = \lambda/\mu$ and normalizing $\sum_j p_j = 1$, we have:

$$
\begin{aligned}
p_0 &= \left(\sum_{k=0}^{c} \frac{\rho^k}{k!}\right)^{-1} \\
p_j &= \frac{\rho^j}{j!}\left[\sum_{k=0}^{c} \frac{\rho^k}{k!}\right]^{-1}
\end{aligned}
$$

  - The formula for p_c is known as the Erlang-B formula (blocked-calls-cleared system):

$$
p_c = \frac{\rho^c}{c!}\left[\sum_{k=0}^{c} \frac{\rho^k}{k!}\right]^{-1}
$$

  - If c is large, $\sum_{k=0}^{c} \frac{\rho^k}{k!} = e^{-\rho}$ hence Erlang distribution becomes Poisson distribution:

$$
\lim_{c\to\infty} \left(\frac{\rho^j}{j!} \bigg/ \sum_{k=0}^{c} \frac{\rho^k}{k!}\right) = \frac{\rho^j e^{-\rho}}{j!}
$$

4.4.3 M/M/s/c

- Only $c \geq s$ makes sense, otherwise it is only a M/M/c/c queue
- Buffer for at most $c-s$ customers, additional arrivals will be blocked
  - Effective arrival rate: $\begin{cases} \lambda & \text{if } 0 \leq n < c \\ 0 & \text{otherwise} \end{cases}$
  - Departure rate: $\begin{cases} n\mu & \text{if } 0 \leq n < s \\ s\mu & \text{otherwise} \end{cases}$
- By defining the balance equation, we have

$$\begin{aligned}
(\lambda + j\mu)p_j &= \lambda p_{j-1} + (j+1)\mu p_{j+1} && (j < s) \\
(\lambda + s\mu)p_j &= \lambda p_{j-1} + s\mu p_{j+1} && (s \leq j < c) \\
s\mu p_c &= \lambda p_{c-1} \\
\mu p_1 &= \lambda p_0
\end{aligned}$$

which yields the recurrence formula:

$$\begin{aligned}
p_j &= \begin{cases} p_0 \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n = p_0 \frac{\rho^n}{n!} & \text{if } 0 \leq n < s \\ p_0 \frac{1}{s!s^{n-s}}\left(\frac{\lambda}{\mu}\right)^n = p_0 \frac{\rho^n}{s!s^{n-s}} & \text{if } s \leq n < c \end{cases} \\
p_0 &= \left[\sum_{n=0}^{s-1} \frac{\rho^n}{n!} + \sum_{n=s}^{c} \frac{\rho^n}{s!s^{n-s}}\right]^{-1}
\end{aligned}$$

4.4.4 M/M/∞ queue

- Taking $c \to \infty$, M/M/c/c will become M/M/∞
- Balance equation:

$$\begin{aligned}
\lambda p_{j-1} &= j\mu p_j \\
p_j &= \frac{\lambda}{j\mu} p_{j-1} \\
\therefore \quad p_j &= \left(\frac{\lambda}{\mu}\right)^j \frac{1}{j!} p_0
\end{aligned}$$

- Solution:

$$\begin{aligned}
p_0 &= e^{-\lambda/\mu} = e^{-\rho} \\
p_n &= \frac{\rho^n}{n!} e^{-\rho}
\end{aligned}$$

- In steady state, the system population is in Poisson distribution with mean $E[N] = \lambda/\mu = \rho$
- Average sojourn time equals to average service time: $E[N] = \lambda E[S] \quad \Rightarrow \quad E[S] = 1/\mu$
- Waiting time is zero

4.4.5 Reference

- [2], Chapter 2
- [9], Chapter 4, "Erlangian Queueing Systems"

4.5 Queue with limited pool of customers

4.5.1 M/M/c/c/n queue (with Quasi-random Input)

- A queue with capacity c and finite sources of n inlets. Each inlet has arrival with rate λ.
- When there are k calls exist in the system, only $n-k$ inlets are idle, hence the effective arrival rate to the system is $(n-k)\lambda$
- Balance equations:

$$[(n-j)\lambda + j\mu]p_j = (n-j+1)\lambda p_{j-1} + (j+1)\mu p_{j+1}$$
$$n\lambda p_0 = \mu p_1$$
$$c\mu p_c = (n-c+1)\lambda p_{c-1}$$

  - Defining $\rho = \lambda/\mu$
  - Recurrence solution:

$$p_j = \frac{(n-j+1)\lambda}{j\mu} p_{j-1} = \frac{(n-j+1)\rho}{j} p_{k-1}$$
$$p_j = \binom{n}{j}\rho^j \left[\sum_{i=0}^{c}\binom{n}{i}\rho^i\right]^{-1} \tag{4.3}$$

  - The distribution p_j is known as the *Engset distribution*

- If $n \to \infty$, Engset distribution will converge to Erlang distribution:

$$\binom{n}{j}\rho^j = \frac{n(n-1)\cdots(n-j+1)}{n^j}\frac{(n\rho)^j}{j!} \to \frac{(n\rho)^j}{j!}$$

  - Effective arrival as $n \to \infty$: $n\lambda$
  - Service rate: μ
  - $n\lambda/\mu = n\rho$ is the effective system utilization

- If $n \le c$, Engset distribution p_j will become binomial
  - Denominator in (4.3): $\sum_{i=0}^{c}\binom{n}{i}\rho^i = \sum_{i=0}^{n}\binom{n}{i}\rho^i = (1+\rho)^n$
  - Steady probability: $p_j = \binom{n}{j}\rho^j(1+\rho)^n = \binom{n}{j}\left(\frac{\rho}{1+\rho}\right)^j\left(1-\frac{\rho}{1+\rho}\right)^{n-j}$

- Upon the arrival of a call arrival, the probability of having k calls exist in the system as seen by the arriving call is given by:

$$\begin{aligned}
\pi_k &= \Pr[k \text{ calls exist}|\text{arrival in } \delta t] \\
&= \frac{\Pr[k \text{ calls exist}]\Pr[\text{arrival in } \delta t|k \text{ calls exist}]}{\sum_{i=0}^{c}\Pr[i \text{ calls exist}]\Pr[\text{arrival in } \delta t|i \text{ calls exist}]} \\
&= \frac{\binom{n}{k}\rho^k p_0 \times (n-k)\lambda\delta t}{\sum_{i=0}^{c}\binom{n}{i}\rho^i p_0 \times (n-i)\lambda\delta t} \\
&= \frac{\binom{n-1}{k}\rho^k p_0 n\lambda\delta t}{\sum_{i=0}^{c}\binom{n-1}{i}\rho^i p_0 n\lambda\delta t} \\
&= \frac{\binom{n-1}{k}\rho^k}{\sum_{i=0}^{c}\binom{n-1}{i}\rho^i}
\end{aligned} \tag{4.4}$$

  - Let $\pi_k(n)$ be (4.4) to denote the dependence of number of inlets n

- $\pi_k(n) = p_k(n-1)$
- As $n \to \infty$, $\pi_k = p_k$, as PASTA expected

- If $n > c$, the probability of blocking calls is given by

$$B = \pi_c = \binom{n-1}{c}\rho^c \left[\sum_{i=0}^{c}\binom{n-1}{i}\rho^i\right]^{-1}$$

  - Engset loss formula, the probability of an arriving calls found blocked (call congestion probability)
  - p_c is the probability that an outside observer found the system is fully occupied (time congestion probability)
  - Expected number of calls in the system:

$$E[N] = \sum_{i=0}^{c} i p_i = p_0 n\rho \sum_{i=0}^{c-1}\binom{n-1}{i}\rho^i$$

  - Offered load:

$$\rho_{\text{offered}} = E[n-N]\rho = p_0 n\rho \sum_{i=0}^{c}\binom{n-1}{i}\rho^i$$

  - Blocking probability is also given by $B = \dfrac{\rho_{\text{offered}} - E[N]}{\rho_{\text{offered}}}$
  - Offered load in terms of blocking probability: $\rho_{\text{offered}} = \dfrac{n\rho}{1+\rho(1-B)}$

- Note: For the same parameters, variance of distribution decreases as the order Poisson > Erlang > Engset > binomial.

4.5.2 M/M/s/c/n

- Working as a M/M/s/c queue with the size of customer pool be n
- Balance equations:

$$\begin{aligned}
[(n-j)\lambda + j\mu]p_j &= (n-j+1)\lambda p_{j-1} + (j+1)\mu p_{j+1} && (j < s) \\
p_j &= (n-j+1)\lambda p_{j-1} + s\mu p_{j+1} && (s \le j < c) \\
s\mu p_c &= (n-c+1)\lambda p_{c-1} \\
n\lambda p_0 &= \mu p_1
\end{aligned}$$

- Solution:

$$p_j = \begin{cases} p_0\binom{n}{j}\left(\frac{\lambda}{\mu}\right)^j = p_0\binom{n}{j}\rho^j & \text{if } 0 \le j < s \\ p_0\binom{n}{j}\frac{j!}{s!s^{j-s}}\left(\frac{\lambda}{\mu}\right)^j = p_0\binom{n}{j}\frac{j!\rho^j}{s!s^{j-s}} & \text{if } s \le j < c \end{cases}$$

$$p_0 = \left[\sum_{j=0}^{s-1}\binom{n}{j}\rho^j + \sum_{j=s}^{c}\binom{n}{j}\frac{j!\rho^j}{s!s^{j-s}}\right]^{-1}$$

4.5.3 M/M/s/∞/n

- Infinite buffer: $c \to \infty$
- Balance equations:

$$\begin{aligned}
[(n-j)\lambda + j\mu]p_j &= (n-j+1)\lambda p_{j-1} + (j+1)\mu p_{j+1} && (j < s) \\
p_j &= (n-j+1)\lambda p_{j-1} + s\mu p_{j+1} && (s \le j < n) \\
n\lambda p_0 &= \mu p_1
\end{aligned}$$

- Solution:

$$p_j = \begin{cases} p_0\binom{n}{j}\left(\frac{\lambda}{\mu}\right)^j = p_0\binom{n}{j}\rho^j & \text{if } 0 \le j < s \\ p_0\binom{n}{j}\frac{j!}{s!s^{j-s}}\left(\frac{\lambda}{\mu}\right)^j = p_0\binom{n}{j}\frac{j!\rho^j}{s!s^{j-s}} & \text{if } s \le j < n \end{cases}$$

$$p_0 = \left[\sum_{j=0}^{s-1}\binom{n}{j}\rho^j + \sum_{j=s}^{n}\binom{n}{j}\frac{j!\rho^j}{s!s^{j-s}}\right]^{-1}$$

- Same effect as M/M/s/c/n queue with $c \geq n$

4.5.4 Reference

- [9], Chapter 7, "The System M/G/1, Priority Systems"; Chapter 8, "The System GI/G/1, Imbedded Markov Chains"

4.6 Multi-class Queue

4.6.1 Batch arrival: M$^{[X]}$/M/s/s loss system, with PBAS

- X: A random variable representing the batch size
- PBAS: Partial batch acceptance strategy
- Symbols:
  - p_j: Probability that j calls exist at any arbitrary instan
  - $P(z) = \sum_{j=0}^{\infty} z^j p_j$: Generating function of p_j
  - $b_i = \Pr[X = i]$: Probability of batch size i
  - $B(z) = \sum_{j=0}^{\infty} z^j b_j$: Generating function of b_i
  - $\phi_i = \sum_{j=i}^{\infty} b_j$: Probability of $X \geq i$
  - λ: Batch arrival rate
- Balance equation:

$$(\lambda + j\mu)p_j = \lambda\sum_{i=0}^{j} p_i b_{j-i} + (j+1)\mu p_{j+1} \qquad 0 \le j < s$$

$$s\mu p_s = \lambda\sum_{i=0}^{s} p_i\phi_{s-i}$$

  - Recurrence solution: $p_j = \dfrac{\lambda}{j\mu}\sum_{i=0}^{j-1} p_j\phi_{j-i}$

4.6.2 Processor Sharing Queue with Mixed traffic

- Two classes of jobs: $i = 1, 2$
- Fraction of jobs: α_i : $\alpha_1 + \alpha_2 = 1$
- Exponential service time with mean $1/\mu_i$ $(i = 1, 2)$
  - Assumed $1/\mu_1 > 1/\mu_2$

- Density function of service time, X:

$$f_X(t) = \alpha_1\mu_1 e^{-\mu_1 t} + \alpha_2\mu_2 e^{-\mu_2 t} \quad t \geq 0$$

  - Conditional probability:

$$\begin{aligned}\Pr[i=1|X \geq \tau] &= \frac{\alpha_1 e^{-\mu_1\tau}}{\alpha_1 e^{-\mu_1\tau} + \alpha_2 e^{-\mu_2\tau}} \\ &= \frac{\alpha_1 e^{(\mu_2-\mu_1)\tau}}{\alpha_1 e^{(\mu_2-\mu_1)\tau} + \alpha_2} \\ &> \frac{\alpha_1}{\alpha_1+\alpha_2} = \alpha_1\end{aligned}$$

  - $\Pr[X < t|i=1] = 1 - e^{-\mu_1 t}$, $\therefore \Pr[X > t|i=1] = e^{-\mu_1 t}$

- Processor sharing queue: If n jobs in queue, each receives service with rate μ/n simultaneously, where μ is a system parameter of the queue

- State space: (N_1, N_2) where N_1, N_2 are number of jobs of class 1 and class 2 respectively

  - $p_{ij} = \Pr[N_1 = i,\ N_2 = j]$

- Balance equation:

$$\begin{aligned}\left(\lambda + \mathbf{1}\{i>0\}\frac{i}{i+j}\mu_1 + \mathbf{1}\{j>0\}\frac{j}{i+j}\mu_2\right)p_{ij} &= \alpha_1\lambda\mathbf{1}\{i>0\}p_{i-1,j} \\ &\quad +\alpha_2\lambda\mathbf{1}\{j>0\}p_{i,j-1} \\ &\quad +\frac{i+1}{i+j+1}\mu_1 p_{i+1,j} \\ &\quad +\frac{j+1}{i+j+1}\mu_2 p_{i,j+1}\end{aligned}$$

  with solution:

$$p_{ij} = \left(\frac{\alpha_1\lambda}{\mu_1}\right)^i \left(\frac{\alpha_2\lambda}{\mu_2}\right)^j \frac{(i+j)!}{i!\,j!} p_{00}$$

- The processor sharing queue is same as M/M/1 queue:

$$\begin{aligned}\Pr[N=n] &= \sum_{i=0}^{n} p_{i,n-i} \\ &= (\lambda E[X])^n p_{00}\end{aligned}$$

4.6.3 General Processor Sharing Queue

- Poisson arrival with rate λ
- General i. i. d. service times X with density function $f_X(x)$, c.d.f. $F_X(x)$ and mean $E[X]$
- Service rate: $1/n$ for each customer if there are n customers in the system
- System population at time t: $N(t)$
- State of the system: $\mathbf{X}(t) = (X_1(t),\ X_2(t),\ \ldots,\ X_{N(t)}(t))$ where $X_i(t)$ is the remaining service time of i-th customer in the system
- Probability function: $f_n(t, x_1, \ldots, x_n) = \Pr[N(t) = n,\ \mathbf{X}(t) = (x_1, \ldots, x_n)]$
- Balance equation: $f_n(t+\delta t, x_1, \ldots, x_n)$ in terms of $f_m(t, x_1, \ldots, x_m)$

$$\begin{aligned}f_n(t+\delta t, x_1, \ldots, x_n) &= (1-\lambda\delta t) f_n(t, x_1 + \frac{\delta t}{n}, \ldots, x_n + \frac{\delta}{n}) \\ &\quad + \sum_{i=0}^{n} \int_0^{\frac{\delta t}{n+1}} f_{n+1}(t, x_1 + \frac{\delta t}{n+1}, \ldots, x_{i-1} + \frac{\delta t}{n+1}, y, x_i + \frac{\delta t}{n+1}, \ldots, x_n + \frac{\delta t}{n+1}) dy \\ &\quad + \lambda\delta t \sum_{i=0}^{n} f_X(x_i) f_{n-1}(t, x_1 + \frac{\delta t}{n-1}, \ldots, x_{i-1} + \frac{\delta t}{n-1}, x_{i+1} + \frac{\delta t}{n-1}, \ldots, x_n + \frac{\delta t}{n-1})\end{aligned}$$

– Approximation by using derivatives: $f_n(t, x_1 + \frac{\delta t}{n}, \ldots, x_n + \frac{\delta}{n}) = f_n(t, x_1, \ldots, x_n) + \sum_{i=1}^{n} \frac{\partial f_n(t, x_1, \ldots, x_n)}{\partial x_i} \frac{\delta t}{n} + o(\delta t)$

– Simplifying integral:

$$\int_0^{\frac{\delta t}{n+1}} f_{n+1}(t, x_1 + \frac{\delta t}{n+1}, \ldots, x_{i-1} + \frac{\delta t}{n+1}, y, x_i + \frac{\delta t}{n+1}, \ldots, x_n + \frac{\delta t}{n+1}) dy = f_{n+1}(t, x_1, \ldots, x_{i-1}, 0, x_i, \ldots, x_n) \frac{\delta t}{n+1} + o(\delta t)$$

– $\delta t\, f_{n-1}(t, x_1 + \frac{\delta t}{n-1}, \ldots, x_{i-1} + \frac{\delta t}{n-1}, x_{i+1} + \frac{\delta t}{n-1}, \ldots, x_n + \frac{\delta t}{n-1}) = \delta t\, f_{n-1}(t, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) + o(\delta t)$

– Hence the balance equation above can be rewritten as:

$$\begin{aligned} f_n(t+\delta t, x_1, \ldots, x_n) &= (1-\lambda\delta t) f_n(t, x_1, \ldots, x_n) + (1-\lambda\delta t) \sum_{i=1}^{n} \frac{\delta t}{n} \frac{\partial f_n}{\partial x_i} \\ &\quad + \sum_{i=0}^{n} f_{n+1}(t, x_1, \ldots, x_{i-1}, 0, x_i, \ldots, x_n) \frac{\delta t}{n+1} \\ &\quad + \lambda \delta t \sum_{i=0}^{n} f_X(x_i) f_{n-1}(t, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \\ &\quad + o(\delta t) \end{aligned}$$

– Partial differential equations:

$$\begin{aligned} \frac{\partial f_n}{\partial t} &= -\lambda f_n(t, x_1, \ldots, x_n) + (1-\lambda\delta t) \sum_{i=1}^{n} \frac{1}{n} \frac{\partial f_n}{\partial x_i} \\ &\quad + \sum_{i=0}^{n} f_{n+1}(t, x_1, \ldots, x_{i-1}, 0, x_i, \ldots, x_n) \frac{1}{n+1} \\ &\quad + \lambda \sum_{i=0}^{n} f_X(x_i) f_{n-1}(t, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \\ &= 0 \end{aligned}$$

where the partial derivative w.r.t. time t should be zero because the function f_n should converge in steady state

– Solution for the partial differential equation above is:

$$f_n(x_1, \ldots, x_n) = (1-\rho)\lambda^n \prod_{i=1}^{n} (1 - F_X(x_i))$$

* We can verify that: $f_{n+1}(x_1, \ldots, x_{i-1}, 0, x_i, \ldots, x_{n+1}) = \lambda f_n(x_1, \ldots, x_n)$
* $\frac{\partial f_n}{\partial x_i} = -(1-\rho)\lambda^n f_X(x_i) \prod_{j \neq i} (1 - F_X(x_j)) = \lambda f_X(x_i) f_{n-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$

– Probability of $N = n$ in steady state:

$$\begin{aligned} \Pr[N = n] &= (1-\rho)\lambda^n \prod_{i=1}^{n} \int_0^\infty (1 - F_X(x_i)) dx_i \\ &= (1-\rho)\lambda^n \prod_{i=1}^{n} E[X] \\ &= (1-\rho)\rho^n \end{aligned}$$

4.6.4 M/G/1 Non-preemptive Priority Service Queue

- Classes: $i = 1, 2, \ldots, n$ where $i < j$ means class i gets priority over j
- Service time of class i: X_i
- Poisson arrival with rate λ_i for class i. Further, $\lambda = \sum_{i=1}^{n} \lambda_i$

- $E[X] = \sum_{i=1}^{n} \frac{\lambda_i}{\lambda} E[X_i]$
- $E[X^2] = \sum_{i=1}^{n} \frac{\lambda_i}{\lambda} E[X_i^2]$
- $\rho_i = \lambda_i E[X_i]$ is the probability of class i customer in service
- $\rho = \sum_{i=1}^{n} \rho_i < 1$

- Queueing time of a class i customer is the sum of:
  1. Remaining service time of the customer in service (if any)
  2. Service time of class 1, 2, ..., i customers already in queue
  3. Service time of class 1, 2, ..., $i-1$ customers who arrive while this customer is waiting in queue
- At any time instant, the residue life time R of the customer currently in service is given by: (c.f. section 4.2.4)

$$\begin{aligned}
E[R] &= \lim_{t\to\infty} \frac{1}{t} \int_0^t R(\tau) d\tau = \lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{n} \sum_{k=1}^{N_i(t)} \int_{t_0}^{t_0+X_{i,k}} (X_{i,k} - \tau) d\tau \\
&= \lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{n} \sum_{k=1}^{N_i(t)} \frac{X_{i,k}^2}{2} \\
&= \lim_{t\to\infty} \sum_{i=1}^{n} \sum_{k=1}^{N_i(t)} \frac{N_i(t)}{t} \frac{X_{i,k}^2}{N_i(t)} \frac{1}{2} \\
&= \lim_{t\to\infty} \sum_{i=1}^{n} \frac{N_i(t)}{t} \frac{\sum_{k=1}^{N_i(t)} X_{i,k}^2}{N_i(t)} \frac{1}{2} \\
&= \frac{1}{2} \sum_{i=1}^{n} \left(\lim_{t\to\infty} \frac{N_i(t)}{t} \right) \left(\lim_{t\to\infty} \frac{\sum_{k=1}^{N_i(t)} X_{i,k}^2}{N_i(t)} \right) \\
&= \frac{1}{2} \sum_{i=1}^{n} \lambda_i E[X_i^2] \\
&= \frac{1}{2} \lambda E[X^2]
\end{aligned}$$

- Equation:

$$\begin{aligned}
E[W_q^{(i)}] &= E[R] + \sum_{k=1}^{i} E[N_q^{(k)}] E[X_k] + \sum_{k=1}^{i-1} E[W_q^{(i)}] \lambda_k E[X_k] \\
&= E[R] + \sum_{k=1}^{i} \lambda_k E[W_q^{(k)}] E[X_k] + E[W_q^{(i)}] \sum_{k=1}^{i-1} \lambda_k E[X_k] \\
&= E[R] + \sum_{k=1}^{i} \rho_k E[W_q^{(k)}] + E[W_q^{(i)}] \sum_{k=1}^{i-1} \rho_k \\
\left(1 - \sum_{k=1}^{i} \rho_k\right) E[W_q^{(i)}] &= E[R] + \sum_{k=1}^{i} \rho_k E[W_q^{(k)}] \\
\implies \quad E[W_q^{(i)}] &= \frac{E[R]}{\left(1 - \sum_{k=1}^{i} \rho_k\right)\left(1 - \sum_{k=1}^{i-1} \rho_k\right)} \qquad (4.5) \\
&= \frac{\lambda E[X^2]}{2\left(1 - \sum_{k=1}^{i} \rho_k\right)\left(1 - \sum_{k=1}^{i-1} \rho_k\right)} \\
\text{and} \quad E[W^{(i)}] &= E[W_q^{(i)}] + E[X_i]
\end{aligned}$$

4.6.5 M/G/1 Preemptive Resume Priority Queue

- Parameters: same as those in M/G/1 non-preemtive priority queue
- Sojourn time of a class i customer is the sum of:
  1. Time to clear all class 1, 2, ..., i customers already in the system upon arrival (referred as T_i)
  2. Time to clear all preemptive class 1, 2, ..., $i-1$ customers who arrive before this customer completes
  3. Service time of this customer
- Equation:

$$E[W^{(i)}] = E[T_i] + \sum_{k=1}^{i-1} \lambda_k E[W^{(i)}] E[X_k] + E[X_i]$$

$$\left(1 - \sum_{k=1}^{i-1} \rho_k\right) E[W^{(i)}] = E[T_i] + E[X_i]$$

$$E[W^{(i)}] = \frac{E[T_i] + E[X_i]}{1 - \sum_{k=1}^{i-1} \rho_k}$$

  – $E[T_i]$ is given by:

$$E[T_i] = \sum_{k=1}^{i} \frac{\lambda_k E[X_k^2]}{2(1-\rho_k)} = \frac{\sum_{k=1}^{i} \lambda_k E[X_k^2]}{2\left(1 - \sum_{k=1}^{i} \rho_k\right)}$$

  – Hence the sojourn time:

$$E[W^{(i)}] = \frac{\sum_{k=1}^{i} \lambda_k E[X_k^2]}{2\left(1 - \sum_{k=1}^{i} \rho_k\right)\left(1 - \sum_{k=1}^{i-1} \rho_k\right)} + \frac{E[X_i]}{1 - \sum_{k=1}^{i-1} \rho_k}$$

4.6.6 Reference

- [2], Chapter 4
- [14]
- [7], Section 3
- [9], Chapter 5, "Priority Systems"; Chapter 7, "The System M/G/1, Priority Systems"

4.7 Matrix-Geometric Technique

4.7.1 M/M/1 with different arrival rates

- Arrival rate: λ with the system is non-empty and λ' with the system is empty
- Balance equation:

$$\lambda' p_0 = \mu p_1$$
$$(\lambda + \mu) p_1 = \lambda' p_0 + \mu p_2$$
$$(\lambda + \mu) p_j = \lambda p_{j-1} + \mu p_{j+1} \quad j = 2, 3, \ldots$$

  yields the solution: $p_j = \rho^{j-1} p_1$

- Infinitesimal generator:

$$Q = \begin{pmatrix} -\lambda' & \lambda' & 0 & 0 & \cdots \\ \mu & -\lambda-\mu & \lambda & 0 & \cdots \\ 0 & \mu & -\lambda-\mu & \lambda & \cdots \\ 0 & 0 & \mu & -\lambda-\mu & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

- Probabilities p_j satisfies:

$$p_0 + \sum_{j=1}^{\infty} p_j = 1$$

$$p_0 + p_1 \sum_{j=1}^{\infty} \rho^{j-1} = 1$$

$$p_0 + \frac{p_1}{1-\rho} = 1$$

Hence p_0 and p_1 should satisfies:

$$\begin{pmatrix} 1 & \lambda' \\ \frac{1}{1-\rho} & -(\lambda+\mu)+\rho\lambda \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

- Expected number of customers in queue:

$$\begin{aligned} E[N_q] &= \sum_{j=1}^{\infty} (j-1) p_j \\ &= p_1 \sum_{j=1}^{\infty} (j-1)\rho^{j-1} \\ &= p_1 \frac{\rho}{(1-\rho)^2} \end{aligned}$$

4.7.2 Hyperexponential queues: M/H$_2$/1

- H_r density function: $f_X(t) = \sum_{k=1}^{r} \alpha_k \mu_k e^{-\mu_k t}$ with $\sum_k \alpha_k = 1$
- Consider a queue with arrival rates λ' if queue empty and λ if queue non-empty; H_2 service time distribution with parameters $\alpha, \mu_1, \bar{\alpha}, \mu_2$
  - Service class: 1, 2 for referring μ_1 and $\mu_2$respectively, a.k.a. exponential stage
  - Define system state as (n,s) where n is the number of users in the system and s is the service class of the job in service
  - Queue empty: $(0,0)$
- Infinitesimal generator:

$$Q = \begin{pmatrix} -\lambda' & \lambda'\alpha & \lambda'\bar{\alpha} & 0 & 0 & 0 & 0 & \cdots \\ \mu_1 & -\lambda-\mu_1 & 0 & \lambda & 0 & 0 & 0 & \cdots \\ \mu_2 & 0 & -\lambda-\mu_2 & 0 & \lambda & 0 & 0 & \cdots \\ 0 & \alpha\mu_1 & \bar{\alpha}\mu_1 & -\lambda-\mu_1 & 0 & \lambda & 0 & \cdots \\ 0 & \alpha\mu_2 & \bar{\alpha}\mu_2 & 0 & -\lambda-\mu_2 & 0 & \lambda & \cdots \\ 0 & 0 & 0 & \alpha\mu_1 & \bar{\alpha}\mu_1 & -\lambda-\mu_1 & 0 & \cdots \\ 0 & 0 & 0 & \alpha\mu_2 & \bar{\alpha}\mu_2 & 0 & -\lambda-\mu_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} B_{00} & B_{01} & 0 & 0 & \cdots \\ B_{10} & A_1 & A_0 & 0 & \cdots \\ 0 & A_2 & A_1 & A_0 & \cdots \\ 0 & 0 & A_2 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

where $B_{00} = (-\lambda')$, $B_{01} = \begin{pmatrix} \lambda'\alpha & \lambda'\bar{\alpha} \end{pmatrix}$, $B_{10} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $A_0 = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$, $A_1 = \begin{pmatrix} -\lambda-\mu_1 & 0 \\ 0 & -\lambda-\mu_2 \end{pmatrix}$, $A_2 = \begin{pmatrix} \alpha\mu_1 & \bar{\alpha}\mu_1 \\ \alpha\mu_2 & \bar{\alpha}\mu_2 \end{pmatrix}$

– Balance equation:

$$\begin{aligned} p_0 B_{00} + p_1 B_{10} &= 0 \\ p_0 B_{01} + p_1 A_1 + p_2 A_2 &= 0 \\ p_{j-1} A_0 + p_j A_1 + p_{j+1} A_2 &= 0 \end{aligned}$$

for $j \geq 2$

- Solution: $p_j = p_1 R^{j-1}$
  - $p = \begin{pmatrix} p_0 & p_1 & p_2 & \cdots \end{pmatrix}$; $pQ = 0$
  - Gives $A_0 + RA_1 + R^2 A_2 = 0$
  - Incorporate with normalization condition,

$$\begin{pmatrix} p_0 & p_1 \end{pmatrix} \begin{pmatrix} 1 & B_{01} \\ (I-R)^{-1}e & A_1 + RA_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \end{pmatrix}$$

- Iterative algorithm for solving R:
  1. Set $R_0 = \mathbf{0}$
  2. $R_{n+1} = (-A_0 - R_n A_2) A_1^{-1}$
  3. If the system is ergodic, $\lim_{n \to \infty} R_n = R$

- Expected number of customers in queue:

$$\begin{aligned} E[N_q] &= \sum_{j=1}^{\infty} (j-1) p_j e \\ &= p_1 \sum_{j=1}^{\infty} (j-1) R^{j-1} e \\ &= p_1 R (I-R)^{-2} e \end{aligned}$$

4.7.3 General QBD queues

- General quasi birth death queues:

$$Q = \begin{pmatrix} B_{00} & B_{01} & 0 & 0 & 0 & \cdots \\ B_{10} & B_{11} & A_0 & 0 & 0 & \cdots \\ B_{20} & B_{21} & A_1 & A_0 & 0 & \cdots \\ B_{30} & B_{31} & A_2 & A_1 & A_0 & \cdots \\ B_{40} & B_{41} & A_3 & A_2 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

  where B_{00} is $m' \times m'$, $B_{01}$is $m' \times m$, B_{n0} is $m \times m'$, others are $m \times m$

  - State space: (i, j) where i is the level and j is the phase
    * m' phases in level 0 and m phases in other levels

- A long vector: $p = \begin{bmatrix} p_0 & p_1 & p_2 & \cdots \end{bmatrix}$ where $p_0 = \begin{bmatrix} p_{01} & p_{02} & \cdots & p_{0m'} \end{bmatrix}$ and $p_i = \begin{bmatrix} p_{i1} & p_{i2} & \cdots & p_{im} \end{bmatrix}$, then

$$pQ = 0$$

  with balance equation

$$\begin{aligned} \sum_{k=0}^{\infty} p_{j-1+k} A_k = 0 \quad j \geq 2 \\ \Longrightarrow \qquad p_j = p_1 R^{j-1} \end{aligned}$$

  where $\sum_{k=0}^{\infty} R^k A_k = 0$.

- Denote $M^*$to be the matrix M with leftmost column removed, then we have

$$
\begin{pmatrix} p_0 & p_1 \end{pmatrix} \begin{pmatrix} B_{00} & B_{01} \\ \sum_{k=1}^{\infty} R^{k-1} B_{k0} & \sum_{k=1}^{\infty} R^{k-1} B_{k1} \end{pmatrix} = 0
$$

$$
\begin{pmatrix} p_0 & p_1 \end{pmatrix} \begin{pmatrix} 1 & B_{00}^* & B_{01} \\ (I-R)^{-1} e & \left[\sum_{k=1}^{\infty} R^{k-1} B_{k0}\right]^* & \sum_{k=1}^{\infty} R^{k-1} B_{k1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \end{pmatrix}
$$

4.7.3.1 Markov Modulated Poisson Process (MMPP)

- A k-state CTMC $\{X(t)\}$ with transition rates of $i \to j$ denoted with a_{ij}
- If in state $X(t) = n$, the arrival rate to the queue is λ_n
  - Arrival process is Poisson with rate $\lambda_{X(t)}$
  - The CTMC $\{X(t)\}$ modulates the arrival
- The queue with arrival rate $\lambda_{X(t)}$ and exponential service times with parameter μ
  - System state: (n, s) with n jobs in the system and $X(t) = s$ is the state of the modulating MC
  - Infinitesimal generator:

$$
Q = \begin{pmatrix} B_{00} & A_0 & 0 & 0 & \cdots \\ A_2 & A_1 & A_0 & 0 & \cdots \\ 0 & A_2 & A_1 & A_0 & \cdots \\ 0 & 0 & A_2 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}
$$

where $B_{00} = \begin{pmatrix} -\alpha_1 & a_{12} & \cdots & a_{1k} \\ a_{21} & -\alpha_2 & \cdots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & -\alpha_k \end{pmatrix}$, $A_0 = \mathrm{diag}(\ \lambda_1 \ \cdots \ \lambda_k\ )$, $A_1 = \begin{pmatrix} -\alpha_1 - \mu & a_{12} & \cdots & a_{1k} \\ a_{21} & -\alpha_2 - \mu & \cdots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & -\alpha_k - \mu \end{pmatrix}$,
$A_2 = \mathrm{diag}(\ \mu \ \cdots \ \mu\ )$, $\alpha_i = \lambda_i + \sum_{j \neq i} a_{ij}$

- It can be solved by Matrix-Geometric techniques

4.7.4 Reference

- [14], “More General Systems”
- [8]

Chapter 5

Queueing Networks

5.1 Queueing Networks

5.1.1 Definitions

- Open and closed system:

  Closed system Total number of customers in the system at all time is a constant, i.e. no external arrival or leaving the system

  Open system Customers arrive from an external source and may leave the system

- Markovian Network
  - A network with N nodes,
  - Customer arrive from external to (randomly) node j as a Poisson process with rate λ_j
  - Service offered (exponentially) at node j with rate $\mu_j(n_j)$ where n_j is the number of customers at this node. In particular, $\mu_j(0) = 0$
  - Upon completion, the customer transit from node j to node k with probability p_{jk}, or leaves the system with probability $q_j = 1 - \sum_k p_{jk}$
    - Transition is independent of history (hence Markovian)
    - Switching probability: p_{jk} where $j \neq k$
    - Probability of instantaneous feedback: p_{jj}
    - The stochastic matrix $(p_{jk})_{N\times N}$ is irreducible and aperiodic
  - Queue discipline is FCFS

5.1.2 Markovian Network

5.1.2.1 Open Markovian network

- Construct the queue length vector: $\mathbf{n} = (n_1, n_2, \cdots, n_N)$, the transition and transition rates are:

$$\begin{aligned}
(n_1,\cdots,n_j,\cdots,n_N) &\to (n_1,\cdots,n_j+1,\cdots,n_N) &&: \quad \lambda_j \\
(n_1,\cdots,n_j,\cdots,n_N) &\to (n_1,\cdots,n_j-1,\cdots,n_N) &&: \quad \mu_j(m_j)q_j \\
(n_1,\cdots,n_j,\cdots,n_k,\cdots,n_N) &\to (n_1,\cdots,n_j-1,\cdots,n_k+1,\cdots,n_N) &&: \quad \mu_j(m_j)p_{jk}
\end{aligned}$$

- Defining $q(\mathbf{n},\mathbf{m})$ be the rate of transition from queue length vector $\mathbf{n}$ to vector $\mathbf{m}$, the total rate of leaving state $\mathbf{n}$ is therefore:

$$r(\mathbf{n}) = \sum_{\mathbf{m}} q(\mathbf{n},\mathbf{m})$$

$$
\begin{aligned}
&= \sum_{j=1}^{N} \lambda_j + \sum_{j=1}^{N} \mu_j(m_j) q_j + \sum_{j=1}^{N} \sum_{k \neq j} \mu_j(m_j) p_{jk} \\
&= \sum_{j=1}^{N} \lambda_j + \sum_{j=1}^{N} \mu_j(m_j) \left(1 - \sum_{k=1}^{N} p_{jk}\right) + \sum_{j=1}^{N} \sum_{k \neq j} \mu_j(m_j) p_{jk} \\
&= \sum_{j=1}^{N} [\lambda_j + \mu_j(m_j)(1 - p_{jj})]
\end{aligned}
$$

and the balance equation is

$$
r(\mathbf{n})p(\mathbf{n}) = \sum_{\mathbf{m}} p(\mathbf{m}) q(\mathbf{m}, \mathbf{n})
$$

or equivalently:

$$
\begin{aligned}
p(\mathbf{n}) \left(\sum_k \lambda_k + \sum_k \mu_k(n_k) \right) &= \sum_k \lambda_k p(\cdots, n_k - 1, \cdots) \\
&+ \sum_k q_k \mu_k(n_k + 1) p(\cdots, n_k + 1, \cdots) \\
&+ \sum_k \sum_j p_{kj} \mu_k(n_k + 1) p(\cdots, n_k + 1, \cdots, n_j - 1, \cdots)
\end{aligned}
$$

- The effective arrival rate α_k to node j satisfies: (which is known as the "traffic equation")

$$
\alpha_k = \lambda_k + \sum_{j=1}^{N} \alpha_j p_{jk}
$$

- With k_j being is the normalization constant to make $\sum_n p_j(n) = 1$, the stationary distribution of queue length is:

$$
p(n_1, \cdots, n_N) = p_1(n_1) p_2(n_2) \cdots p_N(n_N)
$$

with the stationary distribution of the individual queue as:

$$
p_j(n_j) = k_j \frac{(\alpha_j)^{n_j}}{\mu_j(1)\mu_j(2)\cdots\mu_j(n_j)}
$$

5.1.2.2 Closed Markovian network

- For closed networks, there are finite and constant number of customers M in the system
- Due to no external arrival and departures, $\lambda_j = q_j = 0$, and the balance equation becomes:

$$
r(\mathbf{n})p(\mathbf{n}) = \sum_{\mathbf{m}} p(\mathbf{m}) q(\mathbf{m}, \mathbf{n})
$$

$$
\sum_{j=1}^{N} \mu_j(n_j)(1 - p_{jj}) p(\cdots, n_j, \cdots, n_k, \cdots) = \sum_{j=1}^{N} \sum_{k \neq j} \mu_j(n_j + 1) p_{jk} p(\cdots, n_j + 1, \cdots, n_k - 1, \cdots)
$$

- The effective arrival rate α_j to node j satisfies: (due to $\lambda_k \equiv 0$)

$$
\alpha_k = \sum_{j=1}^{N} \alpha_j p_{jk}
$$

- With k being is the normalization constant to make $\sum_{\mathbf{m}} p(\mathbf{m}) = 1$, the stationary distribution of queue length is:

$$
p(n_1, \cdots, n_N) = k \prod_{j=1}^{N} \frac{(\alpha_j)^{n_j}}{\mu_j(1)\mu_j(2)\cdots\mu_j(n_j)}
$$

5.1.3 Quasi-reversibility

For an open Markovian network, the departure at every node is independent Poisson process. Hence the departure $\hat{Q}(t)$ is a time-reverse of queue length process $Q(t)$, this feature is known as "quasi-reversibility" as $\hat{Q}(t) = Q(-t)$ corresponds to a hypothetical queueing network.

An open Markovian network $(\lambda, \mu, \mathbf{P})$ with:

- $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_N)$
- $\mu = (\mu_1, \mu_2, \cdots, \mu_N)$
- $\mathbf{P} = (P_{jk})_{N\times N}$

which gives queue length process $\{Q(t)\}$ has the time-reversed process $\{\hat{Q}(t)\}$ that correspond to the network $(\hat{\lambda}, \hat{\mu}, \hat{\mathbf{P}})$ where:

- $\hat{\lambda}_j = \alpha_j q_j$
- $\hat{\mu}_j = \mu_j$
- $\hat{P}_{jk} = \frac{\alpha_k}{\alpha_j} P_{jk}$

5.1.4 Jackson Network

- The difficulty with analysis of network is that the inter-arrival time after traversing the first queue are correlated with the queue length and not necessary Poisson
  - Jackson's theorem: The correlation is eliminated and randomization is used to divide traffic among different routes, such that the network can be analyzed
  - Result from J. R. Jackson in 1957
  - The significance of Jackson's theorem is the independence among the number of customers at distinct queues at a given time, even if the overall arrival is not a Poisson process.

5.1.4.1 Open Jackson Network

- Jackson network is a special case of open Markovian network $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \mathbf{P})$ such that each node of the network is a M/M/s queue with s_k identical servers at node k
  - There are total N nodes
  - The actual arrival rate at each node is same as Markovian network: $\alpha_k = \lambda_k + \sum_{j=1}^{N} \alpha_j p_{jk}$
  - Utilization factor of each queue: $\rho_k = \dfrac{\alpha_k}{s_k \mu_k}$
- The solution of the balance equation is: (by Jackson's theorem)

$$p(n_1, \cdots, n_N) = p_1(n_1) p_2(n_2) \cdots p_N(n_N) = \prod_{k=1}^{N} p_k(n_k)$$

  where:

$$\rho_k = \frac{\alpha_k}{s_k \mu_k} < 1$$

$$p_k(n_k) = \begin{cases} p_k(0) \dfrac{1}{n_k!} \left(\dfrac{\alpha_k}{\mu_k}\right)^{n_k} & 0 \le n_k \le s_k \\ p_k(s_k) \rho_k^{n_k - s_k} = p_k(0) \dfrac{1}{n_k! s_k^{n_k - s_k}} \left(\dfrac{\alpha_k}{\mu_k}\right)^{n_k} & n_k \ge s_k \end{cases}$$

$$p_k(0) = \left[\sum_{n_k=0}^{s_k - 1} \frac{1}{n_k!} \left(\frac{\alpha_k}{\mu_k}\right)^{n_k} + \frac{1}{s_k!(1-\rho_k)} \left(\frac{\alpha_k}{\mu_k}\right)^{s_k} \right]^{-1}$$

- If the network is comprises of only M/M/1 queues, i.e. $s_k = 1$ for all k, Jackson's Theorem becomes:

$$p(n_1, \cdots, n_N) = \prod_{k=1}^{N} p_k(n_k) = \prod_{k=1}^{N} \left(1 - \frac{\lambda_i}{\mu_i}\right)\left(\frac{\lambda_i}{\mu_i}\right)^{n_k}$$

- The solution for $p(\mathbf{n})$ is called a "product-form" solution

5.1.4.2 Closed Jackson Network

- Setting $\lambda_k = 0$ in open Jackson network, we have the arrival rate at node k becomes $\alpha_k = \sum_{j=1}^{N} \alpha_j p_{jk}$
  - Fixed M users in the network at any time
  - The network has N nodes
- The solution of the balance equation is: (by Jackson's theorem)

$$p(\mathbf{n}) = \frac{1}{G_{N,M}} \prod_{k=1}^{N} \left(\frac{(\alpha_k)^{n_k}}{\prod_{u=1}^{n_k} \mu_k(u)} \right)$$

where:

$$S_{\mathbf{n}} = \left\{ \mathbf{n} : \ (n_1, n_2, \dots, n_N) \in \{0, 1, \dots, M\}^N \ \wedge \ \sum_{k=1}^{N} n_k = M \right\}$$

$$G_{N,M} = \sum_{\mathbf{n} \in S_{\mathbf{n}}} \prod_{k=1}^{N} \left(\frac{(\alpha_k)^{n_k}}{\prod_{u=1}^{n_k} \mu_k(u)} \right)$$

$$\mu_k(u) = \begin{cases} u\mu_k & 0 \le u < s_k \\ s_k\mu_k & u \ge s_k \end{cases}$$

G is a normalization constant and $S_{\mathbf{n}}$ is a set

  - If the network is comprises of only M/M/1 queues, i.e. $s_k = 1$ for all k, it becomes:

$$p(\mathbf{n}) = \frac{1}{G} \prod_{k=1}^{N} \left(\frac{\alpha_k}{\mu_k} \right)^{n_k}$$

where:

$$S_{\mathbf{n}} = \left\{ \mathbf{n} : \ (n_1, n_2, \dots, n_N) \in \{0, 1, \dots, M\}^N \ \wedge \ \sum_{k=1}^{N} n_k = M \right\}$$

$$G_{N,M} = \sum_{\mathbf{n} \in S_{\mathbf{n}}} \prod_{k=1}^{N} \left(\frac{\alpha_k}{\mu_k} \right)^{n_k}$$

- In closed Jackson network, the computation of the normalization constant G is tedious due to the size of the set $S_{\mathbf{n}}$
  - J. Buzen game a convolution algorithm for the computation in 1973

5.1.4.3 Buzen's Convolution Algorithm

- Given $M = 0$, then the size of $S_{\mathbf{n}}$ is 1 and hence $G_{N,M} = G_{N,0} = 1$
- Given $N = 1$, then the Jackson's network is a single queue with

$$G_{N,M} = G_{1,M} = \sum_{\mathbf{n} \in S_{\mathbf{n}}} \frac{(\alpha_1)^{n_1}}{\prod_{u=1}^{n_1} \mu_1(u)} = \frac{(\alpha_1)^M}{\prod_{u=1}^{M} \mu_1(u)}$$

- For any general N, M pair, we have:

$$\begin{aligned}
G_{N,M} &= \sum_{\mathbf{n}\in S_{\mathbf{n}}} \prod_{k=1}^{N} \left(\frac{(\alpha_k)^{n_k}}{\prod_{u=1}^{n_k} \mu_k(u)} \right) \\
&= \sum_{i=0}^{M} \sum_{\substack{\mathbf{n}\in S_{\mathbf{n}} \\ n_N=i}} \prod_{k=1}^{N} \left(\frac{(\alpha_k)^{n_k}}{\prod_{u=1}^{n_k} \mu_k(u)} \right) \\
&= \sum_{i=0}^{M} \sum_{\substack{\mathbf{n}\in S_{\mathbf{n}} \\ n_N=i}} \left[\prod_{k=1}^{N-1} \left(\frac{(\alpha_k)^{n_k}}{\prod_{u=1}^{n_k} \mu_k(u)} \right) \right] \left(\frac{(\alpha_N)^{i}}{\prod_{u=1}^{i} \mu_N(u)} \right) \\
&= \sum_{i=0}^{M} \left(\frac{(\alpha_N)^{i}}{\prod_{u=1}^{i} \mu_N(u)} \right) G_{N-1,M-i}
\end{aligned}$$

- In the closed Jackson Network with only M/M/1 queues,

$$\begin{aligned}
G_{N,M} &= \sum_{i=0}^{M} \left(\frac{\alpha_N}{\mu_N} \right)^{i} G_{N-1,M-i} \\
G_{1,M} &= \left(\frac{\alpha_N}{\mu_N} \right)^{M} \\
G_{N,0} &= 1
\end{aligned}$$

5.1.5 Multi-class Jackson Network

- Class definition:

5.1.6 BCMP Network

5.1.7 Reference

- [9], Chapter 6, “Queueing Networks”
- [7], Section 4, “Single Class Queueing Networks”

Bibliography

[1] B. Aazhang. Gaussian process. The Connexions Project, Rice University, July 2002. http://cnx.rice.edu/content/m10238/latest/?format=pdf.

[2] H. Akimaru and K. Kawashima. *Teletraffic: Theory and Applications*. Telecommunication Networks and Computer Systems. Springer-Verlag, 2nd edition, 1999.

[3] C. D. Crommelin. Delay probability formulae when the holding times are constant. *Post Office Electrical Engineer's Journal*, 25:41–55, 1932.

[4] L. C. Evans. An introduction to stochastic differential equations. Lecture Notes. Department of Mathematics, U.C. Berkeley. version 1.2, http://math.berkeley.edu/ evans/.

[5] R. Lockhart. Course notes of STAT380: Stochastic processes. Department of Statistics & Actuarial Science, Simon Fraser University, Fall 2000. http://www.stat.sfu.ca/ lockhart/richard/380/00_3/.

[6] MathWorld. Taylor series. http://mathworld.wolfram.com/TaylorSeries.html.

[7] P. Nain. Basic elements of queueing theory: Application to the modelling of computer systems. Lecture notes. University of Massachusetts, Amherst, MA, Jan. 1998.

[8] R. Nelson. Matrix geometric solutions in markov models: a mathematical tutorial. Technical report, IBM, 1991.

[9] N. U. Prabhu. *Foundations of Queueing Theory*. Kluwer Academic, 1997.

[10] S. M. Ross. *An Elementary Introduction to Mathematical Finance: Options and Other Topics*. Cambridge University Press, 2nd edition, 2003.

[11] S. M. Ross. *Introduction to Probability Models*. Academic Press, 8th edition, 2003.

[12] O. P. Sharma. *Markovian Queues*. Ellis Horwood, Chichester, West Sussex, UK, 1990.

[13] H. C. Tijms. *Stochastic Modeling and Analysis: A Computational Approach*. John Wiley & Sons, 1986.

[14] D. Towsley. Course notes of CMPSCI 691R: Performance evaluation. Department of Computer Science, University of Massachusetts at Amherst, Fall 2002. Available http://www-net.cs.umass.edu/pe2002/.

[15] Y. M. Wong. A brief introduction to stochastic differential equations. http://members.shaw.ca/yinwong01/intro_SDEs.html, Mar. 2005.