



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA
E COMPUTAÇÃO CIENTÍFICA



**Relatório de Pesquisa de Iniciação Científica:
Processos de renovação e aplicações**

Pedro Matos Pevide

Élcio Levensztayn

O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brasil. Processo nº 2023/13453-5.

Campinas

Setembro de 2025

Sumário

Introdução	1
1. Definição e Teoremas Limites	3
1.1. Processos de renovação	3
1.1.1. Exemplos	4
1.2. Distribuição de $N(t)$	5
1.3. Teoremas Limites	6
1.3.1. Exemplos	8
2. Processos de Renovação com Recompensa	9
2.1. Definição e resultado fundamental	9
2.1.1. Exemplos	10
3. Aparecimento de Padrões	13
3.1. Introdução	13
3.2. Padrões de Variáveis Aleatórias Discretas	13
3.2.1. Sobreposição de padrões	14
3.2.2. Exemplos	16
3.3. Tempo Esperado para uma Seguida Máxima de Valores Distintos	18
3.4. Seguidas Crescentes de Variáveis Aleatórias Contínuas	19
3.4.1. Exemplo	21
4. Filas	22
4.1. Introdução	22
4.2. Filas $GI/G/1$	23
4.3. Equações de Custo	24
4.3.1. Exemplo	25
4.4. Filas $M/G/1$	26
4.4.1. Definindo a Cadeia de Markov X_n	26
4.4.2. Período médio ocupado do servidor e tempo médio de espera na fila . .	27
4.4.3. Exemplo	27

5. Processos semimarkovianos	28
5.1. Definição	28
5.2. O processo $\{X_n, n \geq 0\}$ e sua matriz de probabilidades	29
5.2.1. Exemplo	30
5.3. Comportamento a Longo Prazo	30
5.3.1. Equações de Balanço Globais	30
5.3.2. Tempo Médio Entre Visitas	31
5.3.3. Exemplo	32
5.4. Distribuições de Ocupação	32
5.4.1. Exemplo	33
5.5. Taxas de Custo a Longo Prazo	33
5.5.1. Exemplo	34
6. Simulações	35
6.1. Introdução	35
6.2. Simulações Escolhidas e suas Implementações	35
6.3. Resultados: Filas Paralelas	36
6.4. Resultados: Patrimônio de uma Seguradora	40
Referências bibliográficas	42

Introdução

Esta pesquisa tem como objetivo apresentar a teoria dos processos de renovação, uma generalização dos processos de Poisson, e seus principais teoremas limites, bem como trazer diversos exemplos em que a teoria é aplicada de forma prática e realista na modelagem de problemas do cotidiano. Assim, reforça-se a sua utilidade em diversas áreas, ao contemplar, por exemplo, estimativas de durabilidade de aparelhos eletrônicos, estudo de filas e análise do desempenho de sistemas e máquinas de estados.

O presente texto é uma versão expandida do relatório final da bolsa PIBIC de Iniciação Científica do aluno Pedro Matos Pevide, sob a orientação de Élcio Levensztayn, projeto intitulado “Processos de renovação e aplicações”. Aqui desenvolvemos mais exemplos de aplicação da teoria e demonstrações adicionais ou mais detalhadas. Com enfoque primordialmente teórico, a pesquisa foi realizada, principalmente, a partir dos livros de Kulkarni (2011), Durrett (2016) e Ross (2023a), e complementada, quando necessário ou buscando maior clareza, com Mitov e Omey (2014). Na última seção, aborda-se o tópico de simulações, modelando em Python cenários onde a teoria se aplica, a fim de obter medidas a longo prazo dos processos e esclarecer comportamentos probabilísticos por meio de gráficos e listas, ampliando, assim, o escopo aplicado-computacional da pesquisa.

A Seção 1 contém a definição de um processo de renovação e introduz suas principais terminologias, notações e teoremas limites, construindo os fundamentos necessários para as próximas seções. Passando para a Seção 2, o conhecimento anterior é utilizado para desenvolver o conceito de renovação com recompensa, ilustrado em cenários envolvendo filas, ferrovias e compras.

Utilizando-se do caráter renovador dos processos abordados, a análise do aparecimento de padrões em sequências de eventos é o foco da Seção 3, contemplando tanto variáveis aleatórias discretas quanto contínuas, assim como sobreposição de padrões, tempos esperados para uma sequência máxima e seguidas crescentes de variáveis aleatórias.

O estudo detalhado de filas com um servidor, na Seção 4, contempla as filas $GI/G/1$, com o tratamento da famosa Fórmula de Little, e as filas $M/G/1$, com a definição de uma cadeia de Markov associada e a análise do período médio ocupado do servidor e do tempo médio de espera na fila. A Seção 5 é dedicada aos processos semimarkovianos, com a exposição dos conceitos básicos e teoremas relacionados ao comportamento a longo prazo.

Finalmente, a Seção 6, baseada em Ross (2023b), introduz o conceito de simulação de eventos discretos e apresenta dois cenários modelados em Python: filas paralelas em um estabelecimento e a evolução do patrimônio de uma seguradora. Por meio da simulação desses cenários, busca-se analisar seus comportamentos a longo prazo, com o estudo de características essenciais como o tempo médio de espera para uma fila estável e a probabilidade de falência da seguradora em função dos parâmetros do modelo.

1. Definição e Teoremas Limites

1.1. Processos de renovação

Considere um processo de contagem $\{N(t), t \geq 0\}$ e seja T_n o tempo entre o $(n - 1)$ -ésimo e o n -ésimo evento desse processo, com $n \geq 1$.

Definição 1.1. Se a sequência de variáveis aleatórias não negativas $\{T_1, T_2, \dots\}$ é independente e identicamente distribuída (i.i.d.), então o processo de contagem $\{N(t), t \geq 0\}$ é chamado de *processo de renovação*.

Assim, um processo de renovação é um processo de contagem tal que se o tempo até a ocorrência do primeiro evento possui uma distribuição F , então o tempo entre o primeiro e o segundo evento possui, independentemente do tempo do primeiro evento, a mesma distribuição F , e assim por diante. Quando um evento ocorre, dizemos que uma renovação aconteceu.

Para um processo de renovação com tempos de ocorrência entre eventos T_1, T_2, \dots , sejam

$$S_0 = 0, \quad S_n = \sum_{i=1}^n T_i.$$

Isto é, $S_1 = T_1$ é o tempo da primeira renovação; $S_2 = T_1 + T_2$ é o tempo até a primeira renovação mais o tempo entre a primeira e a segunda renovação, ou seja, S_2 é o tempo da segunda renovação. De forma geral, S_n define o tempo da n -ésima renovação. Veja-se a Figura 1.1.

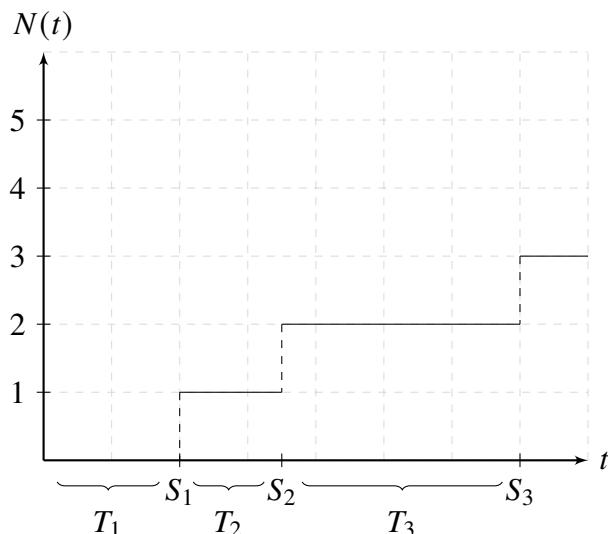


Figura 1.1: Representação de um processo de contagem genérico.

Denotaremos por F a distribuição do tempo entre eventos e, para evitar trivialidades, assumimos que $F(0) = P\{T_n = 0\} < 1$. Além disso, definimos

$$\mu = E[T_n], \quad n \geq 1$$

como o tempo médio entre renovações sucessivas. Segue da não negatividade de T_n e do fato de T_n não ser identicamente 0 que $\mu > 0$.

Vale notar que não é possível ocorrer um número infinito de renovações em um tempo finito. Para provar isso, considere a reescrita de $N(t)$ como

$$N(t) = \max\{n : S_n \leq t\}. \quad (1.1)$$

Agora, pela Lei Forte dos Grandes Números, temos que, com probabilidade 1,

$$\frac{S_n}{n} \longrightarrow \mu \quad \text{quando } n \longrightarrow \infty.$$

Como $\mu > 0$, isso implica que S_n deve tender a infinito quando n tende a infinito. Portanto, S_n pode ser menor ou igual a t para no máximo um número finito de valores de n e, dessa forma, pela equação (1.1), $N(t)$ é finito para todo t .

Contudo, ainda que $N(t) < \infty$ para cada t , também é verdade que, com probabilidade 1,

$$N(\infty) \equiv \lim_{t \rightarrow \infty} N(t) = \infty.$$

1.1.1. Exemplos

Exemplo 1.1. (Troca de Resistor). Considere um morador que deve trocar o resistor de seu chuveiro assim que ele para de funcionar. Considere, também, que o tempo para a troca do resistor é instantâneo. Definimos $N(t)$ como o número de resistores trocados durante os primeiros t meses de funcionamento do chuveiro. Além disso, a fim de evitar falhas não planejadas, o morador troca o resistor do chuveiro assim que ele completa 3 meses de vida. Naturalmente, caso ele falhe antes desses 3 meses, o morador também realiza a troca. Seja V_n o tempo de vida, em meses, do n -ésimo resistor. Supondo que $\{V_n, n \geq 1\}$ seja uma sequência de variáveis i.i.d. não negativas, o morador realiza a troca do primeiro resistor no tempo $T_1 = 3$, se $V_1 > 3$, ou no tempo $T_1 = V_1$, de tal forma que $T_1 = \min\{V_1, 3\}$. Analogamente, $T_2 = \min\{V_2, 3\}$ e, de forma geral, $T_n = \min\{V_n, 3\}$. Como $\{V_n, n \geq 1\}$ é i.i.d., os tempos entre as trocas $\{T_n, n \geq 1\}$ também são i.i.d. Portanto, $\{N(t), t \geq 0\}$ é um processo de renovação.

Exemplo 1.2. (Cadeia de Markov a tempo discreto). Seja X_n o estado de um sistema no tempo n e suponha que $\{X_n, n \geq 0\}$ seja uma Cadeia de Markov a tempo discreto com espaço de estados $\{1, 2, \dots, N\}$. Supondo que a cadeia comece em um estado fixado i , $1 \leq i \leq N$, e definindo S_n como a n -ésima vez que o processo visita o estado i , temos:

$$S_0 = 0,$$

$$S_n = \min\{k > S_{n-1} : X_k = i\}, \quad n \geq 1.$$

Nesse caso, $N(t)$ conta o número de visitas ao estado i até o tempo t . Considerando $T_n = S_n - S_{n-1}$ como sendo o intervalo entre a $(n-1)$ -ésima e a n -ésima visita do processo ao estado i e, lembrando que o processo começa nesse estado, temos que as visitas subsequentes acontecem entre os tempos T_1, T_2, \dots, T_n . Como essas variáveis são independentes e possuem mesma distribuição, concluímos que $\{T_n, n \geq 1\}$ é uma sequência de variáveis i.i.d. não negativas e, portanto, $\{N(t), t \geq 0\}$ é um processo de renovação.

1.2. Distribuição de $N(t)$

Consideremos agora a seguinte relação:

$$N(t) \geq n \iff S_n \leq t. \quad (1.2)$$

Isto é, o número de renovações em um dado tempo t é maior ou igual a n se e somente se a n -ésima renovação ocorre antes do tempo t ou exatamente no tempo t . Da equação (1.2), obtemos

$$\begin{aligned} P\{N(t) = n\} &= P\{N(t) \geq n\} - P\{N(t) \geq n + 1\} \\ &= P\{S_n \leq t\} - P\{S_{n+1} \leq t\}. \end{aligned} \quad (1.3)$$

Como as variáveis aleatórias T_i , $i \geq 1$ são independentes e possuem distribuição comum F , segue que $S_n = \sum_{i=1}^n T_i$ possui distribuição F_n , isto é, a n -ésima convolução de F consigo mesma. Assim, pela equação (1.3), obtemos

$$P\{N(t) = n\} = F_n(t) - F_{n+1}(t).$$

A probabilidade $P\{N(t) = n\}$ também pode ser obtida condicionando-se em S_n :

$$P(N(t) = n) = \int_0^\infty P(N(t) = n | S_n = y) f_{S_n}(y) dy.$$

Pode-se perceber que, caso o n -ésimo evento ocorra em um tempo $y > t$, então menos de n eventos terão ocorrido no tempo t . Ao passo que, caso o n -ésimo evento ocorra em um tempo $y \leq t$, então exatamente n eventos ocorrem até o tempo t caso o próximo tempo entre eventos seja maior que $t - y$. Como buscamos garantir a ocorrência de n eventos no tempo t , exigimos que $T_{n+1} > t - y$. Dessa forma,

$$\begin{aligned} P(N(t) = n) &= \int_0^\infty P(T_{n+1} > t - y \mid S_n = y) f_{S_n}(y) dy \\ &= \int_0^t [1 - F(t - y)] f_{S_n}(y) dy \\ &= \int_0^t \bar{F}(t - y) f_{S_n}(y) dy, \end{aligned}$$

onde $\bar{F} = 1 - F$.

Por meio da equação (1.2), podemos obter $m(t)$, a esperança de $N(t)$, da seguinte forma:

$$\begin{aligned} m(t) &= E[N(t)] \\ &= \sum_{k=1}^{\infty} k P\{N(t) = k\} \\ &= \sum_{k=1}^{\infty} \sum_{n=1}^k P\{N(t) = k\} \\ &= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P\{N(t) = k\} \\ &= \sum_{n=1}^{\infty} P\{N(t) \geq n\} \\ &= \sum_{n=1}^{\infty} P\{S_n \leq t\} \\ &= \sum_{n=1}^{\infty} F_n(t). \end{aligned}$$

A função $m(t)$ também é chamada de *valor médio* ou *função de renovação*.

1.3. Teoremas Limites

Na Seção 1.1, foi visto que $N(t)$ tende a infinito com t tendendo a infinito. Contudo, nada foi dito sobre a velocidade com que $N(t)$ tende a infinito em relação ao parâmetro t , conforme este

também cresce. Pode-se demonstrar (Ross, 2023a, Proposition 7.1) que, com probabilidade 1,

$$\frac{N(t)}{t} \rightarrow \frac{1}{\mu} \text{ quando } t \rightarrow \infty.$$

Essa demonstração estabelece o seguinte teorema:

Teorema 1.1 (Taxa de renovação a longo prazo). *Seja $\{N(t), t \geq 0\}$ um processo de renovação gerado por uma sequência de variáveis aleatórias não negativas e i.i.d. $\{T_n, n \geq 1\}$ com média comum $0 < \mu \leq \infty$. Então,*

$$\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \frac{1}{\mu}$$

com probabilidade 1.

O Teorema 1.1 nos diz que a taxa média de renovação tende a $1/\mu$ quando $t \rightarrow \infty$. Já o teorema a seguir, de forma complementar, nos diz o que acontece com a taxa média *esperada* de renovação, isto é, o que acontece com $m(t)/t$ com $t \rightarrow \infty$.

Teorema 1.2 (Teorema Elementar da Renovação).

$$\frac{m(t)}{t} \rightarrow \frac{1}{\mu} \text{ quando } t \rightarrow \infty.$$

Quando $\mu = \infty$, $1/\mu$ é interpretado como 0.

Sua demonstração pode ser encontrada em Mitov e Omey (2014, p. 15–17).

Finalmente, temos um terceiro e muito importante teorema, o **Teorema Central do Limite para Processos de Renovação**, que diz que, para t grande, $N(t)$ possui distribuição aproximadamente normal com média t/μ e variância $t\sigma^2/\mu^3$, onde μ e $\sigma^2 \in (0, \infty)$ são, respectivamente, a média e a variância da distribuição do tempo entre renovações.

Teorema 1.3 (Mitov e Omey (2014), Theorem 1.3).

$$\lim_{t \rightarrow \infty} P \left\{ \frac{N(t) - t/\mu}{\sqrt{t\sigma^2/\mu^3}} \leq x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy.$$

Além disso, também é possível demonstrar que

$$\lim_{t \rightarrow \infty} \frac{\text{Var}(N(t))}{t} = \frac{\sigma^2}{\mu^3}.$$

1.3.1. Exemplos

Exemplo 1.3. (Troca de Resistor: Continuação). Considerando no Exemplo 1.1 que os tempos de duração dos resistores sigam uma distribuição uniforme $U(1, 5)$, podemos obter a taxa a longo prazo de troca do resistor. Lembrando que V_n é o tempo de vida do n -ésimo resistor e $T_n = \min\{V_n, 3\}$ é o tempo de troca do n -ésimo resistor, temos:

$$\begin{aligned}\mu &= E[T_n] \\ &= \int_1^3 x f_{V_n}(x) dx + \int_3^5 3 f_{V_n}(x) dx \\ &= \frac{1}{4} \int_1^3 x dx + \frac{1}{4} \int_3^5 3 dx \\ &= 2.5 \text{ meses.}\end{aligned}$$

Pelo Teorema 1.1, temos que a taxa de troca a longo prazo de resistores é de $1/2.5 = 0.4$. Logo, é necessário, a longo prazo, realizar uma troca de resistor a cada dois meses e meio.

Exemplo 1.4. (Fila de um banco). Considere um banco contando com apenas um atendente. Os clientes potenciais chegam ao banco conforme um processo de Poisson com taxa λ , mas só entram no banco caso o atendente esteja livre. Suponha que o tempo de atendimento do atendente seja uma variável aleatória com média μ_G . Consideremos como início do processo o momento em que o primeiro cliente entra no banco. Então, pela propriedade de falta de memória do processo de Poisson, o tempo médio entre clientes (renovações) é $\mu = \mu_G + 1/\lambda$. Assim, usando o Teorema 1.1, concluímos que a taxa a longo prazo com que os clientes entram no banco é

$$\frac{1}{\mu} = \frac{\lambda}{1 + \lambda \mu_G}.$$

Como os clientes chegam ao banco com taxa λ , temos que, a longo prazo, a proporção dos clientes potenciais que efetivamente entram no banco é

$$\left(\frac{\lambda}{1 + \lambda \mu_G} \right) \div \lambda = \frac{1}{1 + \lambda \mu_G}.$$

Exemplo 1.5. (Corretores de provas). Considere um professor universitário auxiliado por um assistente a corrigir provas não só de seus próprios alunos, mas também, por uma infeliz ocasião, de todos os alunos da universidade, de todos os cursos. Nesse cenário fantástico, ambos professor e assistente possuem um número basicamente infinito de provas a serem

corrigidas. O tempo necessário (em horas) para que o professor corrija uma prova é uma variável aleatória com distribuição gama e parâmetros $\alpha = 4$, $\lambda = 2$, enquanto o tempo de correção do assistente é uniformemente distribuído entre 0 e 4. Usando o Teorema 1.3, podemos aproximar a probabilidade de que, juntos, o professor e o assistente corrijam pelo menos 90 provas até o tempo $t = 100$. Sendo $N_i(t)$ o número de provas corrigidas pelo corretor i ($1 = \text{professor}$, $2 = \text{assistente}$), temos que $\{N_1(t), t \geq 0\}$ e $\{N_2(t), t \geq 0\}$ são processos de renovação independentes. Pelo Teorema Central do Limite para processos de renovação, $N_1(100)$ é aproximadamente normal com média 50 e variância 100/8, enquanto $N_2(100)$ é aproximadamente normal com média 50 e variância 100/6. Consequentemente, o número total de provas corrigidas no tempo $t = 100$, $N_T(100) = N_1(100) + N_2(100)$, é aproximadamente normal com média 100 e variância 175/6. Logo,

$$\begin{aligned} P\{N_1(100) + N_2(100) > 89.5\} &= P\left\{\frac{N_1(100) + N_2(100) - 100}{\sqrt{175/6}} > \frac{89.5 - 100}{\sqrt{175/6}}\right\} \\ &\approx \Phi(1.944) \\ &\approx 0.9741. \end{aligned}$$

2. Processos de Renovação com Recompensa

2.1. Definição e resultado fundamental

Considere um processo de renovação $\{N(t), t \geq 0\}$ com tempos de ocorrência entre eventos T_n , $n \geq 1$, e suponha que recebemos uma recompensa a cada vez que uma renovação ocorrer. Denotamos por R_n a recompensa obtida no tempo da n -ésima renovação. Supomos que os pares aleatórios (T_n, R_n) , $n \geq 1$, são independentes e identicamente distribuídos. Observamos que R_n pode depender de T_n . Definindo a seguinte função

$$R(t) = \sum_{n=1}^{N(t)} R_n,$$

então $R(t)$ representa o total de recompensas obtidas até o tempo t . Além disso, sejam

$$E[R] = E[R_n], \quad E[T] = E[T_n], \quad n \geq 1.$$

Teorema 2.1. Se $E[R] < \infty$ e $E[T] < \infty$, então:

$$(a) \text{ Com probabilidade 1, } \lim_{t \rightarrow \infty} \frac{R(t)}{t} = \frac{E[R]}{E[T]}.$$

$$(b) \lim_{t \rightarrow \infty} \frac{E[R(t)]}{t} = \frac{E[R]}{E[T]}.$$

Para provar (a), consideramos a seguinte escrita:

$$\frac{R(t)}{t} = \left(\frac{\sum_{n=1}^{N(t)} R_n}{N(t)} \right) \left(\frac{N(t)}{t} \right).$$

Pela Lei Forte dos Grandes Números, temos que

$$\frac{\sum_{n=1}^{N(t)} R_n}{N(t)} \rightarrow E[R] \quad \text{quando } t \rightarrow \infty.$$

E pelo Teorema 1.1,

$$\frac{N(t)}{t} \rightarrow \frac{1}{E[T]} \quad \text{quando } t \rightarrow \infty.$$

A prova do item (b) usa a Identidade de Wald e pode ser vista em [Mitov e Omey \(2014, Theorem 3.18\)](#).

2.1.1. Exemplos

Exemplo 2.1. (Fila de banco: continuação). Suponha que os clientes que entram no banco depositam uma quantia aleatória seguindo uma distribuição H com média μ_H . Lembramos que o tempo médio entre clientes, ou seja, a esperança do tempo entre renovações, é $\mu = \mu_G + 1/\lambda$. Assim, pelo Teorema 2.1, a taxa a longo prazo com que os depósitos acumulam no banco é dada por

$$\frac{\mu_H}{\mu} = \frac{\mu_H}{\mu_G + 1/\lambda}.$$

Considere, agora, $\lambda = 0.5$ e $\mu_G = 1$. Nesse caso, temos uma taxa de 1 cliente por 2 unidades de tempo e um tempo médio de atendimento de 1 unidade. Por praticidade, tomemos o minuto como unidade de tempo. Suponha, também, que cada cliente deposita, em média, 75 reais no banco. Nesse caso, temos que a taxa a longo prazo de depósitos é

$$\frac{75}{1+2} = 25 \text{ reais / minuto.}$$

Exemplo 2.2. (Estação de trem: Custos de operação).

Supondo que os passageiros de uma estação de trem cheguem lá de acordo com um processo de renovação com tempo entre chegadas μ . Sempre que existirem N passageiros esperando na estação, um trem parte. Se o custo de operação da estação é de nc reais por unidade de tempo quando há n passageiros esperando, sendo c o custo unitário, qual é o custo médio de operação da estação?

Percebe-se que o processo acima é um processo de renovação com recompensa, na medida em que um ciclo é completo sempre que um trem parte. Sendo o tempo médio do ciclo o tempo médio de chegada de N passageiros, temos:

$$E[\text{duração do ciclo}] = N\mu.$$

Assim, chamando de T_n o tempo entre o n -ésimo e o $(n+1)$ -ésimo tempo de chegada em um ciclo, o custo esperado do ciclo se torna:

$$E[\text{custo}] = E[cT_1 + 2cT_2 + \dots + (N-1)cT_{N-1}].$$

Isto é, enquanto só existe um passageiro aguardando, temos o custo c por segundo até o tempo T_1 , tempo em que o segundo passageiro chega, acarretando um novo custo operacional de $2c$, enquanto o terceiro passageiro não chega, e assim por diante. Como $E[T_n] = \mu$, a equação acima se reduz a

$$c\mu \underbrace{[1 + 2 + \dots + (N-1)]}_{\text{Soma de P.A.}} = c\mu \frac{N}{2}(N-1).$$

Portanto, o custo médio por ciclo é

$$\frac{c\mu N(N-1)}{2N\mu} = \frac{c(N-1)}{2}.$$

Consideremos, por exemplo, que, toda vez que um trem parte da estação, adicionam-se 6 reais ao custo total de operação. Supondo, também, um custo c de 2 reais e um tempo médio entre chegadas μ de 1 segundo, podemos ir um pouco além e considerar qual valor de N que minimiza o custo operacional da estação. Primeiramente, percebe-se que o novo custo médio por ciclo é de

$$\frac{6 + c\mu N(N-1)/2}{N\mu} = (N-1) + \frac{6}{N}.$$

Considerando agora que esse custo seja uma função contínua em N , temos que sua derivada é

$$\frac{d(\text{custo})}{dN} = 1 - \frac{6}{N^2}.$$

Igualando a derivada a 0, resulta em

$$N^* = \sqrt{6} \approx 2.45.$$

Ou seja, o valor ótimo inteiro de N é ou 2 ou 3. Nesse exemplo, percebe-se que tanto para $N = 2$ quanto para $N = 3$ tem-se o mesmo custo (igual a 4), implicando que ambos os valores são minimizadores do custo médio a longo prazo.

Exemplo 2.3. (Compra de celulares). Suponha que a vida útil de um celular seja uma variável contínua com distribuição V e função densidade de probabilidade v . O Sr. De Moura segue uma política de comprar um novo celular sempre que o seu aparelho atual parar de funcionar ou atingir uma idade de T anos. Considerando que um novo aparelho custe C_1 reais, que um custo adicional de C_2 reais seja incorrido quando seu aparelho para de funcionar e que o aparelho a ser trocado não possui valor de venda, qual é o custo médio a longo prazo do Sr. De Moura?

Inicialmente, percebe-se que o exemplo aborda um processo de renovação com recompensa (que nesse caso é um custo), uma vez que considerarmos como o fim de um ciclo o momento em que o Sr. De Moura comprar um novo celular. Assim, pelo Teorema 2.1 (a) dessa seção, temos que o custo médio equivale a

$$\frac{E[\text{custo durante um ciclo}]}{E[\text{duração de um ciclo}]}.$$

Sendo X o tempo de vida do celular durante um ciclo qualquer, o custo desse ciclo é dado por

$$\begin{aligned} C_1, & \quad \text{se } X > T \\ C_1 + C_2, & \quad \text{se } X \leq T. \end{aligned}$$

De forma que a esperança do custo total é

$$\begin{aligned} C_1 P\{X > T\} + (C_1 + C_2) P\{X \leq T\} &= C_1 \underbrace{(P\{X > T\} + P\{X \leq T\})}_{=1} + C_2 P\{X \leq T\} \\ &= C_1 + C_2 V(T). \end{aligned}$$

Quanto à duração de um ciclo, temos que

$$\begin{aligned} T, & \quad \text{se } X > T \\ X, & \quad \text{se } X \leq T. \end{aligned}$$

Portanto, a esperança do tamanho de um ciclo é:

$$\int_0^T xv(x)dx + \int_T^\infty Tv(x)dx = \int_0^T xv(x)dx + T[1 - V(T)].$$

Finalmente, temos que o custo médio a longo prazo é

$$\frac{C_1 + C_2 V(T)}{\int_0^T xv(x)dx + T[1 - V(T)]}.$$

Em um cenário onde C_1 seja 8 (mil) reais e C_2 seja 0.5 (mil) reais, considere que o Sr. De Moura seja um amante das tecnologias mais recentes, de forma que ele troque anualmente de celular para se manter sempre atualizado, portanto $T = 1$. Supondo uma distribuição uniforme $U(0, 5)$ para V , temos um custo médio a longo prazo de

$$\frac{8 + 0.5 * 0.2}{0.1 + 1[1 - 0.2]} = \frac{8.1}{0.9} = 9 \text{ mil reais.}$$

3. Aparecimento de Padrões

3.1. Introdução

Para esta seção, antes de descrevermos os aparecimentos de padrões propriamente, é conveniente definir o conceito de um processo de renovação *adiado* (ou *geral*). Um processo de contagem $\{N(t), t \geq 0\}$ com tempos entre eventos independentes T_1, T_2, \dots é dito *adiado* ou *geral* quando T_1 possui uma distribuição diferente do resto das variáveis i.i.d. T_2, T_3, \dots do processo. Ainda que esse tipo de processo seja distinto do previamente estudado, os teoremas limites sobre $N(t)$ permanecem válidos.

3.2. Padrões de Variáveis Aleatórias Discretas

Sejam X_1, X_2, \dots variáveis independentes com $P\{X_i = j\} = p(j)$, $j \geq 0$, e seja T o primeiro tempo em que o padrão x_1, \dots, x_m ocorre. Considerando como uma renovação o momento n , $n \geq m$, em que $(X_{n-m+1}, \dots, X_n) = (x_1, \dots, x_m)$ — isto é, quando a sequência apóis

os primeiros m termos for igual ao padrão procurado —, então $N(n)$, $n \geq 1$, é um processo de renovação adiado, com $N(n)$ representando o número de renovações até o tempo n . Conforme mencionamos anteriormente, sendo $N(n)$ um processo adiado, são válidos os teoremas limites de um processo não adiado, portanto

$$\frac{E[N(n)]}{n} \rightarrow \frac{1}{\mu} \quad \text{quando } n \rightarrow \infty, \quad (3.1)$$

$$\frac{\text{Var}(N(n))}{n} \rightarrow \frac{\sigma^2}{\mu^3} \quad \text{quando } n \rightarrow \infty.$$

Passemos, agora, para o cálculo explícito de μ e σ^2 . Seja $I(i) = 1$ se uma renovação ocorre no tempo i e 0 caso contrário, com $i \geq m$. Como $P\{I(i) = 1\} = P\{X_{i-m+1} = i_1, \dots, X_i = i_m\}$, temos que, por independência, isso é equivalente a $\prod_{i=1}^m p(x_i)$. Denotamos p o resultado desse produto. Sendo $I(i)$ uma v.a. binária, temos que ela tem distribuição de Bernoulli com parâmetro p . Assim,

$$N(n) = \sum_{i=m}^n I(i) \implies E[N(n)] = \sum_{i=m}^n E[I(i)] = (n - m + 1)p.$$

Dividindo por n e tomado $n \rightarrow \infty$, temos pela equação (3.1):

$$\underbrace{\lim_{n \rightarrow \infty} \frac{E[N(n)]}{n}}_{=1/\mu} = \lim_{n \rightarrow \infty} p \left(1 - \frac{m}{n} + \frac{1}{n}\right) = p \quad (3.2)$$

Ou seja, temos que $\mu = 1/p$. De forma similar, pode-se demonstrar (Ross, 2023a, p. 492) que a variância de $N(n)$ é

$$\sigma^2 = p^{-2}(1-p) + 2p^{-3} \sum_{j=1}^{m-1} \text{Cov}(I(m), I(m+j)). \quad (3.3)$$

Resta-nos, agora, definir o conceito de sobreposição nos padrões, com o intuito de separar o processo em dois grupos: com sobreposição e sem sobreposição.

3.2.1. Sobreposição de padrões

Denomina-se sobreposição o número máximo de itens ao final do padrão (x_1, \dots, x_m) que, definindo uma subsequência, podem ser usados no começo de um novo padrão. Por exemplo, o padrão $(1, 1, 3, 1, 1)$ possui sobreposição 2, uma vez que as subsequências (1) e $(1, 1)$ — último dígito e dois últimos dígitos do padrão, respectivamente — podem ser usadas para formar o

início do padrão. Já o padrão $(0, 1, 1, 0)$ possui sobreposição 1, já que a única subsequência possível de se utilizar para formar o padrão é (0) . Veja-se a Figura 3.1.

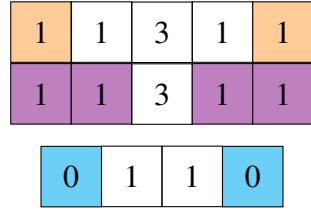


Figura 3.1: Padrões com suas possíveis subsequências coloridas.

Chamando o tamanho da sobreposição de k , $k > 0$, podemos defini-lo formalmente como:

$$k = \max\{j < m : (i_{m-j+1}, \dots, i_r) = (i_1, \dots, i_j)\}.$$

Naturalmente, $k = 0$ se, para todo $p = 1, \dots, m - 1$, temos $(i_{m-p+1}, \dots, i_m) \neq (i_1, \dots, i_p)$.

Supomos, agora, que o processo $N(n)$, $n \geq 1$, não possua sobreposição. Nesse caso temos um processo ordinário de renovação e T é o tempo entre dois padrões, com média μ e variância σ^2 . Pela equação (3.2):

$$E[T] = \mu = \frac{1}{p}.$$

Quanto à covariância, é possível perceber que, para $1 \leq j \leq m - 1$,

$$\text{Cov}(I(m), I(m + j)) = E[\underbrace{I(m) \cdot I(m + j)}_0] - E[I(m)] \cdot E[I(m + j)] = -p^2, \quad (3.4)$$

onde se conclui que $I(m) \cdot I(m + j) = 0$ uma vez que, não havendo sobreposição, deve haver uma distância de pelo menos m entre dois padrões. Utilizando (3.4) em (3.3), temos:

$$\text{Var}(T) = \sigma^2 = p^{-2}(1 - p) + 2p^{-3}(m - 1)p^2 = p^{-2} - (2m - 1)p^{-1}.$$

Agora, no caso em que $N(n)$ possui sobreposição de tamanho k , podemos escrever

$$T = T_{i_1, \dots, i_k} + T^*.$$

O primeiro termo, T_{i_1, \dots, i_k} , é o tempo até o aparecimento do padrão i_1, \dots, i_k , enquanto T^* , distribuído como um tempo entre eventos do processo de renovação, é o tempo adicional necessário, começando com i_1, \dots, i_k , para que o padrão i_1, \dots, i_m seja formado. Pela independência dessas

variáveis, tem-se

$$E[T] = E[T_{i_1, \dots, i_k}] + E[T^*], \quad (3.5)$$

$$\text{Var}(T) = \text{Var}(T_{i_1, \dots, i_k}) + \text{Var}(T^*). \quad (3.6)$$

De forma análoga ao processo sem sobreposição, podemos obter os valores de $E[T^*]$ e $\text{Var}(T^*)$ a partir de (3.2) e (3.3):

$$E[T^*] = \mu = \frac{1}{p},$$

$$\text{Var}(T^*) = \sigma^2 = p^{-2} - (2m-1)p^{-1} + 2p^{-3} \sum_{j=m-k}^{m-1} E[I(m) \cdot I(m+j)],$$

onde a variância é encontrada considerando que $I(m) \cdot I(m+j) = 0$ se $1 \leq j \leq m-k-1$ (duas renovações não podem ocorrer com distância menor ou igual a $m-k-1$).

3.2.2. Exemplos

Exemplo 3.1. (Lançamento de dados: padrão sem sobreposição). Suponha que estamos interessados em saber qual o número esperado de lançamentos de um dado honesto necessários para que o padrão $(1, 3, 2, 2)$ ocorra. Nesse caso, temos que $m = 6$ e $p = \frac{1}{1296}$. Usando (3.2), temos

$$E[T] = \frac{1}{p} = 1296 \text{ lançamentos.}$$

Exemplo 3.2. (Lançamento de moedas: padrão com sobreposição). Sejam c e r cara e coroa, respectivamente. Desejamos obter o número médio de lançamentos de uma moeda honesta até que o padrão (c, c, r, c, c) apareça. Como, nesse caso, há sobreposição com $k = 2$, temos que $T = T_{1, \dots, k} + T^* = T_{c,c} + T^*$. Dessa forma, a esperança é

$$E[T] = E[T^*] + E[T_{c,c}] = 32 + E[T_{c,c}].$$

Ainda resta calcular a esperança do tempo até aparecer o padrão (c, c) , o qual, por sua vez, também é um processo com sobreposição (nesse caso, $k = 1$). Daí,

$$E[T_{c,c}] = E[T_{c,c}^*] + E[T_c] = 4 + E[T_c].$$

Por fim, verifica-se de imediato que $E[T_c] = 2$. Assim,

$$E[T] = 38 \text{ lançamentos.}$$

Exemplo 3.3. (Padrão com sobreposição e probabilidades distintas). Considere uma variável aleatória X_n tal que $P\{X_n = i\} = p_i$. Deseja-se encontrar a esperança e a variância de T para o padrão $(4, 1, 0, 3, 2, 4, 1)$. Primeiro, pode-se perceber que $p = p_4^2 p_1^2 p_0 p_3 p_2$ e $k = 2$. Com isso,

$$E[I(7) \cdot I(12)] = p_4^2 p_1^3 p_0^2 p_3^2 p_2^2$$

$$E[I(7) \cdot I(13)] = 0$$

A segunda igualdade vale 0 uma vez que, caso a sequência seja completa no tempo $T = 7$, os primeiros dois termos de sua segunda ocorrência já estarão presentes. Assim, se a segunda sequência continuar corretamente, ela acaba em $T = 12$, enquanto no melhor cenário após um erro — isto é, quando um número diferente de 0 é produzido no oitavo termo, mas os seguintes são corretos — ela acaba somente em $T = 15$. Ou seja, em ambos os casos a sequência não pode acabar, também, em $T = 13$. Raciocínio semelhante é utilizado para justificar que, se o padrão ocorre em $T = 13$, ele não pode ter ocorrido em $T = 7$. O exemplo é ilustrado na Figura 3.2.

4	1	0	3	2	4	1	0	3	2	4	1
4	1	0	3	2	4	1	X	4	1	0	3

Figura 3.2: Sobreposição sem erros: 12 termos. Falha na sobreposição: ao menos 15 termos.

Por (3.5) e (3.6), temos, respectivamente,

$$E[T] = 1/p + E[T_{4,1}],$$

$$\text{Var}[T] = p^{-2} - 13p^{-1} + 2p^{-1}(p_4 p_1)^{-1} + \text{Var}[T_{4,1}].$$

Como a subsequência $(4, 1)$ não possui sobreposição, resulta que

$$E[T_{4,1}] = 1/(p_4 p_1),$$

$$\text{Var}[T_{4,1}] = (p_4 p_1)^{-2} - 3(p_4 p_1)^{-1}.$$

Finalmente,

$$E[T] = \frac{1}{p} + \frac{1}{p_4 p_1},$$

$$\text{Var}[T] = p^{-2} - 13p^{-1} + 2p^{-1}(p_4 p_1)^{-1} + (p_4 p_1)^{-2} - 3(p_4 p_1)^{-1}.$$

3.3. Tempo Esperado para uma Seguida Máxima de Valores Distintos

Sejam $X_i, i \geq 1$, variáveis i.i.d. com chances equiprováveis de assumirem quaisquer dos valores $1, 2, \dots, m$. Observando essas variáveis em sequência, denota-se T o primeiro momento em que uma seguida de m valores consecutivos inclua todos os valores $1, \dots, m$. Formalmente,

$$T = \min\{n : X_{n-m+1}, \dots, X_n \text{ sejam todos distintos}\}$$

Podemos calcular $E[T]$ definindo um primeiro processo de renovação ocorrendo em T . Após essa primeira renovação, aguarda-se uma próxima renovação que não considere qualquer valor gerado até o tempo T (implicando um processo sem sobreposição), e assim por diante, de forma que a n -ésima renovação não utilize os valores da $(n-1)$ -ésima renovação para compor sua sequência. Se, por exemplo, $m = 2$ e a sequência observada é

$$\underbrace{1, 2}_{1}, \underbrace{1, 1}_{2}, \underbrace{1, 2}_{3}, \underbrace{2, 1}, \dots$$

então até o tempo 8 ocorreram 3 renovações: a primeira no tempo 2, a segunda no tempo 6 e a terceira no tempo 8. Como o uso de valores de uma renovação não é permitido para a formação da seguinte, não se pode dizer que houve uma renovação no tempo 3, apesar do aparecimento dos dois valores distintos (2 e 1) em sequência. Chama-se a sequência de m valores distintos que constituem uma renovação de *seguida de renovação*.

Pode-se transformar esse processo de renovação em um adiado supondo que uma recompensa de 1 seja oferecida no tempo $n, n \geq m$, se os valores X_{n-m+1}, \dots, X_n forem distintos. Logo, uma recompensa é obtida toda vez que m valores forem todos distintos (o que permite, agora, casos de sobreposição). Considerando a sequência apresentada acima, infere-se que as recompensas são obtidas nos tempos 2, 3, 6 e 8. Denotando R_i a recompensa obtida no tempo i , temos pelo Teorema 2.1 (a),

$$\lim_{n \rightarrow \infty} \frac{E[\sum_{i=1}^n R_i]}{n} = \frac{E[R]}{E[T]}, \quad (3.7)$$

onde R é a recompensa total obtida entre renovações. Chamando de A_i o conjunto dos primeiros i valores de uma sequência de renovação, e de B_i o conjunto dos primeiros i valores *após* uma sequência, temos:

$$\begin{aligned} E[R] &= 1 + \sum_{i=1}^{m-1} E[\text{recompensa obtida no tempo } i \text{ depois de uma renovação}] \\ &= 1 + \sum_{i=1}^{m-1} P\{A_i = B_i\} \\ &= 1 + \sum_{i=1}^{m-1} \frac{i!}{m^i} = \sum_{i=0}^{m-1} \frac{i!}{m^i} \end{aligned} \tag{3.8}$$

A penúltima igualdade é obtida lembrando que a probabilidade de $A_i = B_i$ resulta da razão entre o número de padrões onde os i itens são distintos e o número de padrões de tamanho i do espaço amostral, onde existem repetições. Este totaliza m^i padrões — já que para cada item de uma subsequência de tamanho i existem m possíveis valores —, enquanto que aquele totaliza $i!$ padrões, uma vez que repetições não são permitidas.

Agora, como para $i \geq m$,

$$E[R_i] = P\{X_{i-m+1}, \dots, X_i \text{ sejam todos distintos}\} = \frac{m!}{m^m}$$

conclui-se, a partir de (3.7), que

$$\frac{m!}{m^m} = \frac{E[R]}{E[T]}$$

Finalmente, com (3.8), temos

$$E[T] = \frac{m^m}{m!} \sum_{i=0}^{m-1} \frac{i!}{m^i}.$$

3.4. Seguidas Crescentes de Variáveis Aleatórias Contínuas

Seja X_1, X_2, \dots uma sequência de variáveis i.i.d. contínuas. Considere T o primeiro momento em que ocorre uma sequência de r valores consecutivos e crescentes. Formalmente,

$$T = \min\{n \geq r : X_{n-r+1} < X_{n-r+2} < \dots < X_n\}.$$

Podemos calcular $E[T]$ definindo um processo de renovação: considere a primeira renovação ocorrendo no tempo T e, a partir daí, usando somente os valores que surgem após T , tome como uma nova renovação o momento quando uma nova sequência de r valores crescentes consecutivos

surge, e assim por diante.

Como exemplo, considere $r = 3$ e a sequência $(69, 420, 13, 22, 12, 9, 7, 20, 0, 3, 2, 1)$. Nesse caso, temos 4 renovações: nos tempos 2, 4, 8 e 10. Chamando de $N(n)$ o número de renovações até o tempo n , temos pelo Teorema 1.2 que, quando $n \rightarrow \infty$,

$$\frac{E[N(n)]}{n} \rightarrow \frac{1}{E[T]}.$$

Para calcular $E[N(n)]$, define-se um processo estocástico cujo estado no tempo k , denominado S_k , equivale ao número de valores consecutivos crescentes no tempo k . Ou seja, para $1 \leq j \leq k$,

$$S_k = j \quad \text{se } X_{k-j} > X_{k-j+1} < \cdots < X_{k-1} < X_k,$$

onde $X_0 = \infty$. Pode-se perceber, também, que uma renovação ocorre no tempo k se e somente se $S_k = mr$, para algum $m \geq 1$. Por exemplo, considere a seguinte sequência com $r = 2$:

$$X_3 > X_4 < X_5 < X_6 > X_7 < X_8$$

Para esse caso, temos: $S_4 = 1$, $S_5 = 2$, $S_6 = 3$, $S_7 = 1$, $S_8 = 2$, com renovações ocorrendo nos tempos 3 e 6. Lembrando que, apesar da sequência $X_5 < X_6$ ter $r = 2$ e ser crescente, ela não conta como uma nova renovação já que utiliza um dos termos da renovação anterior $X_4 < X_5$. Para $k > j$,

$$\begin{aligned} P\{S_k = j\} &= P\{X_{k-j} > X_{k-j+1} < \cdots < X_{k-1} < X_k\} \\ &= P\{X_{k-j+1} < \cdots < X_{k-1} < X_k\} - P\{X_{k-j} < X_{k-j+1} < \cdots < X_{k-1} < X_k\} \\ &= \frac{1}{j!} - \frac{1}{(j+1)!} = \frac{j}{(j+1)!}, \end{aligned}$$

onde a penúltima igualdade é obtida percebendo que, se j variáveis são consideradas, temos um total de $j!$ possíveis combinações, das quais somente uma respeita o padrão crescente; o mesmo raciocínio vale para $j + 1$. Sendo essas variáveis contínuas, é sempre possível ordená-las em ordem estritamente crescente, uma vez que a probabilidade de existir dois números valores é 0. Utilizando a igualdade acima, nota-se que

$$\lim_{k \rightarrow \infty} P\{\text{renovação no tempo } k\} = \lim_{k \rightarrow \infty} \sum_{m=1}^{\infty} P\{S_k = mr\} = \sum_{m=1}^{\infty} \frac{mr}{(mr+1)!}.$$

Porém,

$$E[N(n)] = \sum_{k=1}^n P\{\text{renovação no tempo } k\}.$$

Ou seja, temos que

$$\lim_{n \rightarrow \infty} \frac{E[N(n)]}{n} = \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n P\{\text{renovação no tempo } k\}}{n}.$$

Pode-se mostrar que, para qualquer sequência de números $\{p_k\}_{k \geq 1}$ cujo limite $\lim_{k \rightarrow \infty} p_k$ existe, tem-se

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n p_k}{n} = \lim_{k \rightarrow \infty} p_k.$$

Com isso, concluímos que

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n P\{\text{renovação no tempo } k\}}{n} = \lim_{k \rightarrow \infty} P\{\text{renovação no tempo } k\} = \sum_{m=1}^{\infty} \frac{mr}{(mr + 1)!}.$$

Como o limite acima é igual a $1/E[T]$, obtemos:

$$E[T] = \frac{1}{\sum_{m=1}^{\infty} mr/(mr + 1)!}.$$

3.4.1. Exemplo

Exemplo 3.4. (Lançamento de moedas: tempo esperado para surgimento de um padrão). Denotando cara e coroa por c e r , respectivamente, busca-se calcular o tempo esperado $E[T]$ até que o padrão (c, c, c, r, c, c, c) apareça. Supomos que os lançamentos da moeda são independentes, e que cada lançamento resulta em cara com probabilidade p , ou coroa com probabilidade $q = 1 - p$.

Para esse exemplo, define-se um processo de renovação de forma que a primeira renovação ocorre assim que o padrão desejado aparecer pela primeira vez, recomeçando o processo imediatamente depois disso. Uma recompensa de 1 é obtida sempre que o padrão ocorre. Seja R a recompensa obtida entre renovações, tem-se

$$\begin{aligned} E[R] &= 1 + \sum_{i=1}^6 E[\text{recompensa obtida } i \text{ lançamentos após uma renovação}] \\ &= 1 + p^3 q + p^4 q + p^5 q. \end{aligned}$$

O resultado acima é obtido primeiramente percebendo que, para $i = 1, 2, 3$, ainda não há termos suficientes para formar uma segunda renovação e, portanto, a esperança é 0. A partir de $i = 4$,

contudo, temos os seguintes casos em que a recompensa pode ocorrer:

$$i = 4 : (c, c, c, r, \underbrace{c, c, c, \textcolor{red}{r}, \textcolor{teal}{c}, \textcolor{teal}{c}, \textcolor{teal}{c}}_{\text{recompensa}})$$

$$i = 5 : (c, c, c, r, c, \underbrace{c, c, \textcolor{teal}{c}, \textcolor{red}{r}, \textcolor{teal}{c}, \textcolor{teal}{c}, \textcolor{teal}{c}}_{\text{recompensa}})$$

$$i = 6 : (c, c, c, r, c, c, \underbrace{\textcolor{teal}{c}, \textcolor{teal}{c}, \textcolor{red}{r}, \textcolor{teal}{c}, \textcolor{teal}{c}, \textcolor{teal}{c}}_{\text{recompensa}})$$

Os lançamentos coloridos de verde-azulado acima são aqueles que precisam ocorrer para que a recompensa seja obtida no momento i desejado. Observando, por exemplo, o caso $i = 4$, verifica-se a necessidade de 3 caras e 1 coroa em sequência, resultado esse que possui probabilidade p^3q . A mesma lógica se aplica aos outros dois casos, e com isso fica esclarecido o resultado obtido na equação acima.

Como a recompensa esperada no tempo i é $E[R_i] = p^6q$, usando que

$$\lim_{n \rightarrow \infty} \frac{E[\sum_{i=1}^n R_i]}{n} = \frac{E[R]}{E[T]},$$

concluímos:

$$E[T] = \frac{1 + p^3q + p^4q + p^5q}{p^6q} = p^{-1} + p^{-2} + p^{-3} + p^{-6}q^{-1}.$$

4. Filas

4.1. Introdução

Nesta seção, serão estudados processos compostos pela ocorrência de eventos que possuem um tempo necessário de processamento a ser realizado por um processador (ou servidor). A taxa de ocorrência desses eventos e seus tempos de processamento são distribuídos aleatoriamente. Os eventos que se acumulam enquanto não são processados formam uma fila.

Intuitivamente, pode-se pensar numa fila como o objeto real encontrado em bancos, por exemplo. O evento é uma pessoa chegando ao estabelecimento, enquanto o servidor é um atendente de caixa.

4.2. Filas $GI/G/1$

O termo $GI/G/1$, notação para um tipo específico de fila, é composto por três partes:

- GI (*General Input*) nos diz que os tempos t_i entre as ocorrências dos eventos (ou chegadas sucessivas) são independentes e têm uma distribuição F com média $1/\lambda$.
- G (*General Service Times*) nos diz que o i -ésimo evento requer um tempo s_i de processamento (ou atendimento), onde as variáveis aleatórias s_1, s_2, \dots são independentes, com distribuição G e média $1/\mu$.
- O número 1 indica que o sistema conta com apenas um servidor.

Teorema 4.1 (Durrett, 2016, Theorem 3.5 e Example 3.6). *Suponha que $\lambda < \mu$. Se a fila começa com um número finito k , $k \geq 1$, de eventos que precisam ser processados, então, com probabilidade 1, a fila é esvaziada em um tempo finito. Além disso, a fração limite do tempo em que o servidor está ocupado é igual a λ/μ .*

A hipótese $\lambda < \mu$ é necessária e suficiente para que a fila $GI/G/1$ seja *recorrente positiva*. Em palavras, uma fila é recorrente positiva se, começando vazia, retorna a esse estado em um tempo com valor esperado finito. Para provar o Teorema 4.1, define-se $T_n = t_1 + \dots + t_n$ como o tempo de ocorrência do n -ésimo evento. Pela Lei Forte dos Grandes Números, temos

$$\lim_{n \rightarrow \infty} \frac{T_n}{n} \rightarrow \frac{1}{\lambda} \quad \text{quase certamente.}$$

Além disso, definem-se: Z_0 o tempo de processamento presente no tempo 0, S_n a soma dos tempos de processamento s_i de cada evento i após o tempo 0 e, finalmente, Z_n o tempo de processamento restante no sistema após o n -ésimo evento, depois do qual outros eventos param de ocorrer. Usando a Lei Forte dos Grandes Números e o fato de que a fila é recorrente positiva, temos

$$\lim_{n \rightarrow \infty} \frac{Z_0 + S_n - Z_n}{n} \rightarrow \frac{1}{\mu} \quad \text{quase certamente.}$$

Isto é, o tempo que o servidor passa processando os eventos ocorridos dividido por n tende a $1/\mu$ conforme n cresce. É no limite acima em que se utiliza a hipótese $\lambda < \mu$ para justificar que $Z_n/n \rightarrow 0$, conforme pode ser visto em Durrett (2016, p. 131).

Assim, concluímos que $(Z_0 + S_n - Z_n)/T_n \rightarrow \lambda/\mu$. Ou seja, o tempo que o servidor fica ocupado durante o intervalo $[0, T_n]$ converge para λ/μ . Além disso, como a fila é recorrente

positiva, ela retorna ao estado vazio em tempo finito.

4.3. Equações de Custo

Ainda abordando as filas $GI/G/I$, denota-se por X_s o número de eventos em um sistema no tempo s e por W_m o tempo que o m -ésimo evento passa no sistema. Em seguida, definem-se suas médias a longo prazo:

$$L = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t X_s ds, \quad W = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n W_m.$$

Por fim, seja $\lambda_a = \lim_{t \rightarrow \infty} N_a(t)/t$ a taxa média a longo prazo com que eventos entram no sistema, onde $N_a(t)$ é o número de eventos que ocorrem antes do tempo t e entram no sistema. Uma fórmula central em teoria de filas relaciona as três quantidades estabelecidas acima:

Teorema 4.2 (Fórmula de Little).

$$L = \lambda_a W.$$

A demonstração da fórmula foge ao escopo do texto, podendo ser encontrada no artigo de Little (1961). Contudo, visando a justificar intuitivamente a fórmula, apresentamos uma “prova” visual. Considerando um sistema em que eventos ocorreram e foram todos processados até o tempo t , a Figura 4.1 mostra uma trajetória possível do processo.

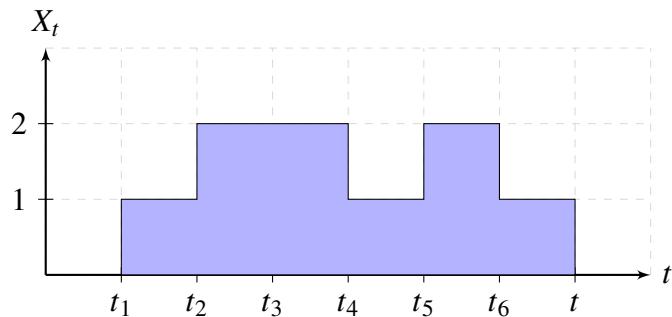


Figura 4.1: Exemplo de um ciclo de eventos em um sistema.

Nesse cenário, percebe-se que três eventos (chegadas) ocorreram: o primeiro no tempo t_1 , o segundo no tempo t_2 e o terceiro no tempo t_5 . Considere, além disso, que o primeiro evento foi completamente processado no tempo t_4 , o segundo no tempo t_6 e o último no tempo t . Assim, temos que o tempo médio W que um evento passa no sistema é

$$W = \frac{(t_4 - t_1) + (t_6 - t_2) + (t - t_5)}{3}.$$

Isto é, a soma do tempo total de cada um dos eventos dividido pelo número de eventos. Repare, contudo, que a equação acima pode ser reescrita como

$$W = \frac{(t - t_1) \cdot 1 + (t_4 - t_2) \cdot 1 + (t_6 - t_5) \cdot 1}{3} = \frac{\text{Área da figura}}{\text{nº de eventos}}.$$

Agora, pode-se também calcular o número médio L de eventos no sistema:

$$L = \frac{(t_2 - t_1) \cdot 1 + (t_4 - t_2) \cdot 2 + (t_5 - t_4) \cdot 1 + (t_6 - t_5) \cdot 2 + (t - t_6) \cdot 1}{t} = \frac{\text{Área da figura}}{t}.$$

Isolando a área na equação para W e substituindo em L , concluímos:

$$L = \frac{(\text{nº de eventos})}{t} \cdot W = \lambda W.$$

4.3.1. Exemplo

Exemplo 4.1. (Tempo de Espera em Fila e Constantes de Desempenho). Considere uma fila $GI/G/1$. Busca-se encontrar o tempo médio de espera que um cliente fica na fila antes de ser atendido. Chamamos esse tempo de W_Q , e assumimos que o tempo total médio W que um cliente passa no sistema é conhecido. Nota-se que o tempo médio de espera na fila de um cliente nada mais é do que o tempo médio total que ele passa no sistema subtraído do seu tempo médio de processamento $E[s_i]$:

$$W_Q = W - E[s_i].$$

Denota-se por L_Q o comprimento médio de uma fila em equilíbrio, onde não é contado, caso exista, o cliente em processamento. Pela Fórmula de Little, tem-se $L_Q = \lambda_a W_Q$.

Porém, o comprimento da fila é zero se o sistema está vazio ou, em caso contrário, é o número total de pessoas no sistema menos 1. Ou seja, a diferença entre o número total de pessoas no sistema e o comprimento da fila é a variável aleatória indicadora de que o sistema não está vazio. Portanto, chamando $\pi(0)$ a probabilidade de termos um sistema vazio, temos:

$$L_Q = L - [1 - \pi(0)].$$

Consequentemente,

$$\pi(0) = L_Q - L + 1 = \lambda_a W_Q - \lambda_a W + 1 = \lambda_a (W_Q - W) + 1 = 1 - \lambda_a E[s_i] = 1 - \frac{\lambda_a}{\mu}.$$

Como $\lambda_a = \lambda$ na fila $GI/G/1$, concluímos que $\pi(0) = 1 - \lambda/\mu$. Mas $\pi(0)$ é a proporção a longo

prazo do tempo que o sistema está vazio. Assim, podemos chamar $1 - \pi(0) = \lambda/\mu$ de *taxa de desempenho da fila*, uma vez que essa quantia representa a proporção a longo prazo do tempo em que o servidor está ocupado.

4.4. Filas $M/G/I$

Esse tipo de fila é um caso especial da fila $GI/G/I$, em que a letra M significa *Markovian input*. Dessa forma, consideramos que os eventos ocorrem segundo um processo de Poisson com taxa λ e, portanto, os tempos entre as chegadas não possuem memória. Quanto às outras características da fila, permanecem iguais.

4.4.1. Definindo a Cadeia de Markov X_n

Define-se X_n o número de eventos na fila quando o n -ésimo evento começa a ser processado. Seja a_k a probabilidade de k eventos entrarem no sistema durante um tempo de serviço. Obtém-se a_k por condicionamento na duração do tempo de serviço:

$$a_k = \int_0^\infty \frac{e^{-\lambda t} (\lambda t)^k}{k!} dG(t).$$

Isto é, para cada $k \geq 0$ fixado, pondera-se a função de probabilidade de ocorrência dos k eventos pela derivada de G . Utiliza-se para o cálculo de a_k a **integral de Lebesgue–Stieltjes**, a qual permite um tratamento unificado dos casos de distribuições discretas, contínuas ou mistas. Se a função de distribuição G é absolutamente contínua com densidade g , então g é a derivada de G (em quase todo ponto), e podemos escrever $dG(t) = g(t)dt$, tornando a equação mais familiar. Já o número esperado de eventos que entram no sistema durante um tempo de serviço é dado por

$$\sum_0^\infty k a_k = \int_0^\infty \lambda t dG(t) = \lambda \int_0^\infty t dG(t) = \lambda E[s_i] = \frac{\lambda}{\mu}.$$

A integral acima é obtida lembrando-se que o número médio de chegadas de eventos no sistema para um intervalo de serviço t é λt .

Para construir uma cadeia de Markov, considere ζ_1, ζ_2, \dots variáveis aleatórias i.i.d. tais que $P(\zeta_i = k) = a_k$. Assim, ζ_i representa o número de eventos que entram no sistema durante o i -ésimo tempo de processamento. Se $X_n > 0$, então

$$X_{n+1} = X_n + \zeta_n - 1.$$

Esse resultado é bem natural: somam-se aos eventos já em fila os ζ_n novos eventos que chegam durante o n -ésimo processamento, e se subtrai 1 na equação devido à saída do n -ésimo evento. Repare, contudo, que se $X_n = 0$ e $\zeta_n = 0$, teríamos $X_{n+1} = -1$. Para consertar isso, considera-se

$$X_{n+1} = (X_n + \zeta_n - 1)^+ = \max\{0, X_n + \zeta_n - 1\}.$$

4.4.2. Período médio ocupado do servidor e tempo médio de espera na fila

Supomos que $\lambda < \mu$, logo o processo é recorrente positivo. Como a fila retorna infinitas vezes ao estado vazio, o servidor alterna intervalos de duração B_n em que está ocupado (*busy*) e intervalos de duração I_n em que se encontra em descanso (*idle*). A falta de memória do processo de chegadas implica que I_n possui distribuição exponencial com taxa λ . Para encontrar a duração média do período ocupado do servidor, notamos que, pelo Teorema 2.1, a fração a longo prazo do tempo em que o servidor fica em descanso é

$$\pi(0) = \frac{1/\lambda}{1/\lambda + E[B_n]}.$$

Por outro lado, sabemos que $\pi(0) = 1 - \lambda/\mu$. Portanto,

$$\frac{1/\lambda}{1/\lambda + E[B_n]} = \frac{\mu - \lambda}{\mu} \implies E[B_n] = \frac{1}{\mu - \lambda}.$$

Finalmente, apresentamos um teorema sobre o tempo médio de espera na fila, sob a condição de que $\lambda < \mu$. A prova pode ser encontrada em Durrett (2016, p. 135).

Teorema 4.3 (Fórmula de Pollaczek–Khintchine).

$$W_Q = \frac{\lambda E[s_i^2/2]}{1 - \lambda E[s_i]}.$$

4.4.3. Exemplo

Exemplo 4.2. (Linha de Produção). Um fabricante de computadores possui uma linha de montagem em que conjuntos das peças necessárias são deixadas para serem agregadas em cada um dos computadores disponibilizados pela empresa. O processo é realizado automaticamente por uma máquina, e o tempo de montagem possui média de 3 minutos e desvio padrão de 4 minutos. Sabendo que os conjuntos de peças chegam à linha de montagem com taxa $1/5$ por minuto, busca-se encontrar: a fração a longo prazo do tempo em que a linha de produção

permanece vazia, o tempo médio total no sistema de cada conjunto de peças e o comprimento médio da fila.

Para o primeiro item, vale lembrar que a fração a longo prazo do tempo com a linha vazia nada mais é do que $\pi(0)$. Neste exemplo, tem-se $\lambda = 1/5$ e $\mu = 1/3$, logo:

$$\pi(0) = 1 - \frac{\lambda}{\mu} = 1 - \frac{1/5}{1/3} = \frac{2}{5}.$$

Para a segunda pergunta, sabe-se que $W = W_Q + E[s_i]$. Como $E[s_i^2] = E[s_i]^2 + \text{Var}(s_i) = 3^2 + 4^2 = 25$, temos, pela Fórmula de Pollaczek–Khintchine,

$$W_Q = \frac{\lambda E[s_i^2/2]}{1 - \lambda E[s_i]} = \frac{\frac{1}{5} \cdot \frac{25}{2}}{1 - \frac{3}{5}} = \frac{25}{4} = 6.25 \text{ minutos.}$$

Portanto, $W = 6.25 + 3 = 9.25$ minutos.

Quanto ao comprimento médio da fila, usamos a Fórmula de Little:

$$L = \lambda W = \frac{1}{5} \cdot 9.25 = 1.85.$$

5. Processos semimarkovianos

5.1. Definição

Consideremos um sistema com espaço de estados $\{1, 2, \dots, N\}$. Supomos que o sistema entra no estado inicial X_0 no tempo $S_0 = 0$, e lá permanece por um tempo aleatório não negativo, até o momento em que salta para um novo estado X_1 no tempo S_1 . Em seguida, após algum tempo em X_1 (que pode ser o mesmo que X_0), o sistema salta para um novo estado X_2 (também pode ser o mesmo que X_1) no tempo S_2 , e assim sucessivamente, realizando infinitos saltos. Temos, então, que S_n é o tempo do n -ésimo salto do sistema, enquanto X_n é o n -ésimo estado visitado. Denotando por $X(t)$ o estado do sistema no tempo t , vemos que $X(S_n) = X_n$ para $n \geq 0$.

Definição 5.1. O processo estocástico $\{X(t), t \geq 0\}$ é denominado *processo semimarkoviano* quando apresenta a Propriedade de Markov em todos os seus tempos de transição S_n . Ou seja, a evolução do processo a partir do tempo $t = S_n$ em diante depende da história do processo até o tempo S_n somente por meio de X_n .

Matematicamente, se $\{X(t), t \geq 0\}$ é um processo semimarkoviano, então o processo $\{X(t + S_n), t \geq 0\}$, dado seu histórico $\{X(t), 0 \leq t \leq S_n\}$ e $X(S_n) = i$, é independente de

$\{X(t), 0 \leq t < S_n\}$ e é probabilisticamente idêntico a $\{X(t), t \geq 0\}$ dado que $X_0 = i$. Perceba que esse tipo de processo é chamado de semimarkoviano, uma vez que apresenta a Propriedade de Markov apenas em seus tempos de transição, em vez de em todos os tempos.

5.2. O processo $\{X_n, n \geq 0\}$ e sua matriz de probabilidades

Seja $\{X(t), t \geq 0\}$ um processo semimarkoviano. Então, $\{X_n, n \geq 0\}$ é uma cadeia de Markov a tempo discreto (CMTD) com espaço de estados $\{1, 2, \dots, N\}$, uma vez que o seu estado futuro depende somente do seu estado atual. Essa cadeia é chamada de CMTD *imersa* do processo semimarkoviano e possui uma matriz $P = [p_{i,j}]$ de probabilidades de transição, onde

$$p_{i,j} = P(X_{n+1} = j | X_n = i), \quad 1 \leq i, j \leq N.$$

Além disso, definimos

$$w_i = E[S_1 | X_0 = i], \quad 1 \leq i \leq N,$$

o tempo médio que o sistema fica no estado i . Note que, como o processo possui a Propriedade de Markov nos tempos de transição, esse tempo médio no estado i é w_i em qualquer visita a esse estado. Assim, definimos o seguinte vetor de tempos médios de permanência:

$$w = [w_1, w_2, \dots, w_N].$$

A Figura 5.1 apresenta a trajetória de um processo semimarkoviano genérico.

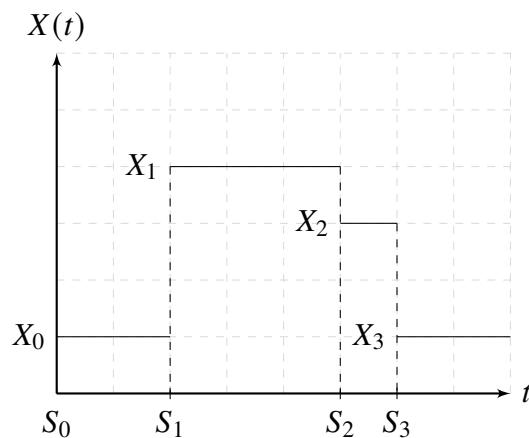


Figura 5.1: Representação de um processo semimarkoviano.

5.2.1. Exemplo

Exemplo 5.1. (*Videogame: Máquina de Estados*). Uma máquina de estados é uma forma de descrever o *estado* (ou ação) de um objeto a cada momento no tempo. No contexto de desenvolvimento de jogos eletrônicos, pode-se considerar o movimento de um inimigo em um jogo de ação como uma máquina de dois estados: ou ele se encontra parado (0), ou em movimento (1). Nesse caso, como são dois os estados possíveis para o inimigo, diz-se que seu movimento é uma *máquina de dois estados*. Naturalmente, pode-se expandir o número de estados para obter uma máquina de n processos.

Sejam M_n a n -ésima duração do intervalo de tempo em que o inimigo permanece em movimento, e seja I_n a n -ésima duração do intervalo de tempo em que permanece parado. Supomos que $\{(M_n, I_n), n \geq 1\}$ é uma sequência i.i.d. de variáveis aleatórias bivariadas, e que M_n é independente de I_n . Se $X(t)$ é o estado do inimigo no tempo t , podemos dizer que $\{X(t), t \geq 0\}$ é um processo semimarkoviano?

Note que, se o inimigo começa no estado $X_0 = 1$, isto é, se movendo, ele permanece nesse estado por um tempo M_1 até mudar para o estado $X_1 = 0$, parado. Daí, ele permanece parado por um tempo I_1 , independente do ocorrido até então, e retorna para o estado $X_2 = 1$, e o processo continua indefinidamente. Assim, $\{X(t), t \geq 0\}$ é um processo semimarkoviano com espaço de estados $\{0, 1\}$, e sua CMTD possui a seguinte matriz de probabilidades de transição:

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Além disso, note que

$$w_0 = E[S_1 | X_0 = 0] = E[M],$$

$$w_1 = E[S_1 | X_0 = 1] = E[I].$$

5.3. Comportamento a Longo Prazo

5.3.1. Equações de Balanço Globais

Seja $\{X_n, n \geq 0\}$ uma CMTD com espaço de estados $S = \{1, 2, \dots, N\}$ e matriz de probabilidades de transição P . Define-se uma *distribuição limite* ou *de estado estacionário* o

vetor de limites $\lim_{n \rightarrow \infty} P(X_n = j), j \in S$, quando eles existem. Matematicamente, tem-se

$$\pi = [\pi_1, \pi_2, \dots, \pi_N] \text{ com } \pi_j = \lim_{n \rightarrow \infty} P(X_n = j), j \in S.$$

Em Kulkarni (2011, p. 24), demonstra-se que, se existe uma distribuição limite π , então

$$\pi_j = \sum_{i=1}^N \pi_i p_{i,j}, \quad j \in S. \quad (5.1)$$

Além disso, tem-se

$$\sum_{j=1}^N \pi_j = 1.$$

Note que as equações (5.1) podem ser reescritas em forma matricial como $\pi = \pi P$, recebendo o nome de *equações de balanço globais*. Um vetor π que satisfaz (5.1) é uma *solução* das equações de balanço globais.

5.3.2. Tempo Médio Entre Visitas

Seja $\{X(t), t \geq 0\}$ um processo semimarkoviano com espaço de estados $\{1, 2, \dots, N\}$, matriz de transição P e vetor de tempos médios de permanência w . Desejamos obter o *tempo médio entre visitas*, ou seja, o tempo esperado entre duas visitas consecutivas a um estado j . Seja Y_j o primeiro momento em que o processo entra em um estado j . Matematicamente, tem-se

$$Y_j = \min\{S_n : n > 0, X_n = j\}.$$

Observe que, mesmo quando o processo já começa no estado j , ainda é necessário considerar o tempo até que esse estado seja novamente visitado no futuro (por isso, $n > 0$). Seja também

$$m_{i,j} = E[Y_j | X_0 = i], \quad 1 \leq i, j \leq N$$

o tempo médio da primeira passagem de um estado i para um estado j . Quando $i = j$, tem-se exatamente o tempo médio entre visitas do estado j .

Teorema 5.1 (Kulkarni, 2011, Theorem 5.4). *Os tempos médios de primeira passagem $m_{i,j}$ satisfazem*

$$m_{i,j} = w_i + \sum_{k=1, k \neq j}^N p_{i,k} m_{k,j}, \quad 1 \leq i, j \leq N.$$

O Teorema 5.1 permite o cálculo de $m_{j,j}$, mas não de forma eficiente: é necessário resolver vários conjuntos de N equações com N incógnitas, de modo a cobrir todos os valores de j . Sob algumas condições, o próximo resultado possibilita o cálculo de $m_{j,j}$ resolvendo um único conjunto de N equações com N incógnitas.

Teorema 5.2 (Kulkarni, 2011, Theorem 5.5). *Suponha que a CMTD imersa $\{X_n, n \geq 0\}$ é irreductível, e seja $\pi = [\pi_1, \pi_2, \dots, \pi_N]$ uma solução não nula para as equações de balanço globais. Então, os tempos médios entre visitas são dados por*

$$m_{j,j} = \frac{\sum_{i=1}^N \pi_i w_i}{\pi_j}, \quad 1 \leq j \leq N.$$

Enfatizamos que, no Teorema 5.2, qualquer solução não nula π das equações de balanço globais funciona; não precisamos de uma solução padronizada. A condição de irreductibilidade da cadeia imersa garante que $\pi_i > 0$ para todo i .

5.3.3. Exemplo

Exemplo 5.2. (*Videogame*: Continuação). Busca-se saber, agora, qual o tempo médio entre dois momentos em que o inimigo está se movendo no jogo. Para isso, note que $\pi = [\pi_0, \pi_1] = [1, 1]$ é uma solução das equações de balanço globais, considerando a matriz P já descrita na primeira parte do exemplo. Além disso, vale lembrar que $w_0 = E[I]$ e $w_1 = E[M]$. Logo, pelo Teorema 5.2,

$$m_{1,1} = \frac{\pi_0 w_0 + \pi_1 w_1}{\pi_1} = E[I] + E[M].$$

Esse resultado é natural, pois, para o inimigo saltar de um estado de movimento para outro também de movimento, é necessário, antes, que visite o estado parado.

5.4. Distribuições de Ocupação

Seja $M_j(t)$ o tempo total, aleatório, que um processo semimarkoviano fica em um estado j durante o intervalo de tempo $[0, t]$. Se, para qualquer j , o limite

$$p_j = \lim_{t \rightarrow \infty} \frac{M_j(t)}{t}$$

existe com probabilidade 1, então denominamos o vetor $[p_1, p_2, \dots, p_N]$ a *distribuição de ocupação* do processo. O próximo resultado apresenta condições sob as quais a distribuição de

ocupação existe e como é obtida.

Teorema 5.3 (Kulkarni, 2011, Theorem 5.6). *Seja $\{X(t), t \geq 0\}$ um processo semimarkoviano com CMTD imersa irredutível $\{X_n, n \geq 0\}$ com matriz de probabilidades de transição P . Então, a distribuição de ocupação do processo existe, é independente do estado inicial do processo, e dada por*

$$p_j = \frac{\pi_j w_j}{\sum_{i=1}^N \pi_i w_i}, \quad 1 \leq j \leq N,$$

onde $[\pi_1, \pi_2, \dots, \pi_N]$ é uma solução positiva das equações de balanço globais e w_i , $1 \leq i \leq N$, é o tempo médio de permanência do sistema no estado i .

5.4.1. Exemplo

Exemplo 5.3. (*Videogame*: Continuação). Podemos, agora, obter a fração a longo prazo do tempo em que o inimigo passa se movimentando. O estado do inimigo é descrito por um processo semimarkoviano $\{X(t), t \geq 0\}$ de dois estados, e sua CMTD imersa possui matriz P irredutível. Logo, o Teorema 5.3 pode ser aplicado, de forma que a proporção a longo prazo do tempo em que o inimigo está se deslocando é:

$$p_1 = \frac{\pi_1 w_1}{\pi_0 w_0 + \pi_1 w_1} = \frac{E[M]}{E[I] + E[M]}.$$

5.5. Taxas de Custo a Longo Prazo

Seja $\{X(t), t \geq 0\}$ um processo semimarkoviano com espaço de estados $\{1, 2, \dots, N\}$. Suponha que o processo incorre em um custo à taxa $c(i)$ para cada unidade de tempo que permanece no estado i . Seja $C(T)$ o custo total do sistema até o tempo T . Caso o limite

$$g = \lim_{T \rightarrow \infty} \frac{C(T)}{T}$$

exista com probabilidade 1, então ele é chamado de *taxa de custo a longo prazo*.

Teorema 5.4 (Kulkarni, 2011, Theorem 5.7). *Suponha que as hipóteses do Teorema 5.3 sejam satisfeitas, e considere $[p_1, p_2, \dots, p_N]$ a distribuição de ocupação do processo semimarkoviano. Então, a taxa de custo a longo prazo existe, é independente do estado inicial do processo, e dada por*

$$g = \sum_{i=1}^N p_i c(i).$$

Para provar o Teorema 5.4, denotamos por $M_i(T)$ o tempo total que o processo passa no estado i durante o intervalo $[0, T]$. Como o processo possui uma taxa de custo $c(i)$ por unidade de tempo durante o estado i , temos:

$$C(T) = \sum_{i=1}^N c(i) M_i(T).$$

Portanto, pelo Teorema 5.3,

$$\lim_{T \rightarrow \infty} \frac{C(T)}{T} = \lim_{T \rightarrow \infty} \sum_{i=1}^N c(i) \frac{M_i(T)}{T} = \sum_{i=1}^N c(i) p_i.$$

5.5.1. Exemplo

Exemplo 5.4. (*Videogame*: Continuação). O jogador em controle do personagem principal do jogo, a fim de sobreviver e escapar do inimigo já mencionado, deve realizar ações com o seu personagem que incluem, dentre várias outras coisas, movimentá-lo. Suponha que o personagem possua um sistema de energia que se esgota conforme ele se move, mas que pode ser recuperado ao permanecer imóvel por algum tempo. Seja A o gasto enérgico por unidade de tempo do personagem enquanto se move, e seja B a energia recuperada, também por unidade de tempo, enquanto permanece parado. Nesse cenário, qual é a taxa a longo prazo de gasto enérgico do personagem principal?

Note, primeiramente, que o movimento do personagem pode ser descrito por um processo semimarkoviano idêntico ao do seu inimigo. Assim, tem-se

$$p_0 = \frac{E[I]}{E[I] + E[M]}, \quad p_1 = \frac{E[M]}{E[I] + E[M]}.$$

Além disso, perceba que $c(0) = -B$ e $c(1) = A$. Isto é, enquanto o personagem está parado, ele tem um gasto enérgico negativo, já que passa a recuperá-la; ao se movimentar, possui um gasto positivo. Substituindo esses valores no Teorema 5.4, obtém-se

$$g = -Bp_0 + Ap_1 = \frac{AE[M] - BE[I]}{E[I] + E[M]}.$$

6. Simulações

6.1. Introdução

Modelos probabilísticos podem ser computacionalmente simulados ao gerar seus mecanismos estocásticos e observar sua evolução temporal. Seguindo a metodologia de Ross (2023b), esta seção aborda especificamente as *simulações de eventos discretos*. Como o nome sugere, simulações de eventos discretos envolvem sistemas cujos estados em cada período de tempo são discretos. Isto é, a cada momento, um processo está em um, e apenas um, estado dentre todos os possíveis.

Os principais elementos de uma simulação discreta são suas variáveis e seus eventos, que serão registrados ao longo da simulação, para análise posterior. De forma geral, definem-se três tipos de variáveis para as simulações: de *tempo*, *contadoras* e de *estado do sistema*. Quando um evento ocorre, os valores dessas variáveis são atualizados, e quaisquer dados desejados são coletados. Por meio de uma lista de eventos, determina-se qual o próximo evento que ocorre na simulação, bem como os eventos após esse último. Por meio da coleta de dados, pode-se acompanhar a evolução temporal do sistema simulado.

6.2. Simulações Escolhidas e suas Implementações

Para este trabalho, duas simulações foram implementadas na linguagem de programação Python, a partir dos algoritmos descritos no Capítulo 7 de Ross (2023b):

- 1. Filas em Paralelo:** Dois servidores de um sistema trabalham paralelamente, atendendo clientes que lá entram, até um tempo limite em que o sistema fecha, impedindo a chegada de novos clientes. Os clientes são servidos por ordem de chegada pelo servidor que esteja livre primeiro (no caso em que ambos estejam livres, prioriza-se o primeiro servidor), formando uma fila caso contrário. Caso o sistema atinja o tempo limite ainda com clientes, todos eles são processados antes do fim da simulação.

Nas simulações, consideramos o **modelo de fila markoviana**, o qual assume que os tempos entre as chegadas dos clientes e os tempos de atendimentos dos servidores são variáveis aleatórias independentes, com distribuição exponencial.

Os parâmetros a serem inseridos no modelo são: Tempo Limite do Sistema Aberto (T),

Taxa de Chegada de Clientes (λ_c) e Tempos Médios de Atendimento dos Servidores 1 e 2 (μ_{s_1}, μ_{s_2}).

2. **Patrimônio de uma Seguradora:** Uma seguradora possui um patrimônio inicial p_0 e um número inicial n_0 de clientes. Cada cliente paga à seguradora uma quantia c por unidade de tempo, referente à contratação do seguro. A cada período de tempo, um dentre três eventos ocorre com sua respectiva probabilidade: um cliente aciona a seguradora, um cliente se desvincula dela, ou um novo cliente a contrata. O patrimônio da seguradora aumenta conforme os segurados pagam \$ c por unidade de tempo que permanecem com o serviço, ao passo que acionamentos causam sua redução. Dessa forma, é possível, dado um tempo T máximo de simulação, registrar as flutuações do patrimônio da seguradora ao longo do tempo.

Para esse modelo, além dos parâmetros descritos acima, tem-se: Taxa de Acionamentos por Período de Tempo (λ), Custo Médio por Acionamento (μ_{ac}), Desvio Padrão do Custo por Acionamento (σ_{ac}), Taxa de Contratação do Seguro (v), Taxa de Evasão por Cliente (e).

No cenário das filas, uma aplicação do algoritmo (detalhada na próxima seção) é a de estimar, em função dos parâmetros fornecidos, o tempo médio a longo prazo que os servidores passam no sistema até que o último cliente tenha sido atendido, e também o tempo médio que cada cliente permanece na fila antes de seu atendimento. Já no caso da seguradora, pode-se estimar, por exemplo, a probabilidade da seguradora não falir, conhecidos os parâmetros. Em ambos os casos, trata-se de uma análise a longo prazo de sistemas de renovação, de forma que seus funcionamentos e resultados podem ser analisados matematicamente pela teoria estudada.

Para ambas as simulações, geramos gráficos e saídas no terminal que fornecem informações úteis para a análise dos sistemas e que serão explicitados nas seções seguintes.

6.3. Resultados: Filas Paralelas

Conforme a teoria de filas, há três casos:

- Se $\lambda_c < 1/\mu_{s_1} + 1/\mu_{s_2}$, então o processo é *recorrente positivo*: começando no estado 0, retorna infinitas vezes a 0, e o tempo esperado até retornar ao estado 0 é finito. O sistema tem uma única distribuição estacionária, e a fila é dita *estável*.

- Se $\lambda_c = 1/\mu_{s_1} + 1/\mu_{s_2}$, então o processo é *recorrente nulo*: começando no estado 0, retorna infinitas vezes a 0, porém o tempo esperado até retornar ao estado 0 é infinito. Não existe distribuição estacionária, e a fila *não é estável*.
- Se $\lambda_c > 1/\mu_{s_1} + 1/\mu_{s_2}$, então o processo é *transitório*: cada estado é visitado somente um número finito de vezes. Assim, o número de clientes no sistema cresce para o infinito com o passar do tempo. Não existe distribuição estacionária, e a fila *não é estável*.

Cumpre ressaltar que somente no primeiro caso é possível calcular medidas teóricas de desempenho da fila sob equilíbrio.

Nas simulações, o algoritmo de filas paralelas apresentou o comportamento previsto pela teoria. Para exemplificar, considere uma simulação com os seguintes parâmetros: $T = 480$ minutos (8 horas), $\lambda_c = 1/5$ (chegada de 1 cliente em média a cada 5 minutos), e tempos médios de atendimento dos servidores $\mu_{s_1} = \mu_{s_2} = 10$ minutos / cliente. Nesse caso, a fila tem comportamento recorrente, já que $1/\mu_{s_1} + 1/\mu_{s_2} = \lambda_c$. Com efeito, uma única rodada dessa simulação retornou as informações abaixo, aqui com dicionários e listas limitados aos seus primeiros 5 itens, para fins de visualização.

Listing 1: Rodada Única: $T = 480$, $\lambda_c = 1/5$, $\mu_{s_1} = 10$, $\mu_{s_2} = 10$

```

1 Tempo total: 527.45 minutos
2 Numero de clientes ainda na loja ao fechar: 10
3 Numero de clientes em cada tempo de atualizacao: [1, 2, 3, 2, 3, 4]
4
5 Tempos de Chegada: {'Cliente 1': 1.122, 'Cliente 2': 4.381, 'Cliente 3':
   5.319, 'Cliente 4': 8.386, 'Cliente 5': 9.338}
6 Momentos de Atendimento: {'Cliente 1': 1.122, 'Cliente 2': 4.381, 'Cliente
   3': 5.907, 'Cliente 4': 10.648, 'Cliente 5': 31.321}
7 Tempos de Partida: {'Cliente 1': 31.321, 'Cliente 2': 5.907, 'Cliente 3':
   10.648, 'Cliente 4': 41.958, 'Cliente 5': 36.092}
8 Relacao Cliente/Servidor: {'Cliente 1': 1, 'Cliente 2': 2, 'Cliente 3': 2,
   'Cliente 4': 2, 'Cliente 5': 1}

```

Os dados anteriores confirmam que as taxas de serviço dos servidores não foram suficientes para evitar a formação de fila. Notamos que, ainda que a loja tenha fechado exatamente no tempo $T = 480$, ainda restavam 10 clientes no estabelecimento, e o último cliente teve seu atendimento concluído apenas 47 minutos após o fechamento da loja. Isso evidencia o comportamento não

estável da fila com esses parâmetros.

Dois gráficos são gerados ao final das simulações: *Número de Clientes no Sistema em Função do Tempo* e *Número de Clientes Atendido por Servidor*, úteis, respectivamente, para visualizar a evolução temporal do sistema e sua distribuição de tarefas. A Figura 6.1 mostra o gráfico do número de clientes em função do tempo para a rodada mencionada acima.

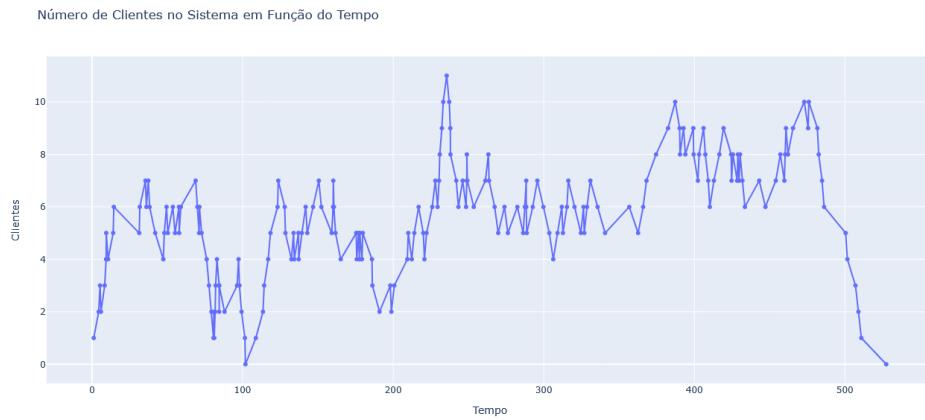


Figura 6.1: Gráfico da evolução temporal do número de clientes no sistema.

Somente pela análise desse gráfico, não se pode inferir que a fila seja recorrente, sobretudo pelo tempo relativamente pequeno de funcionamento do sistema. Contudo, seu caráter recorrente pode ser empiricamente confirmado ao considerarmos uma nova rodada com os mesmos parâmetros, mas com tempo de funcionamento bem maior, por exemplo, $T = 50000$. Dessa forma, conforme ilustrado na Figura 6.2, o padrão de comportamento a longo prazo aparece e as renovações (retornos a 0) se tornam visíveis.

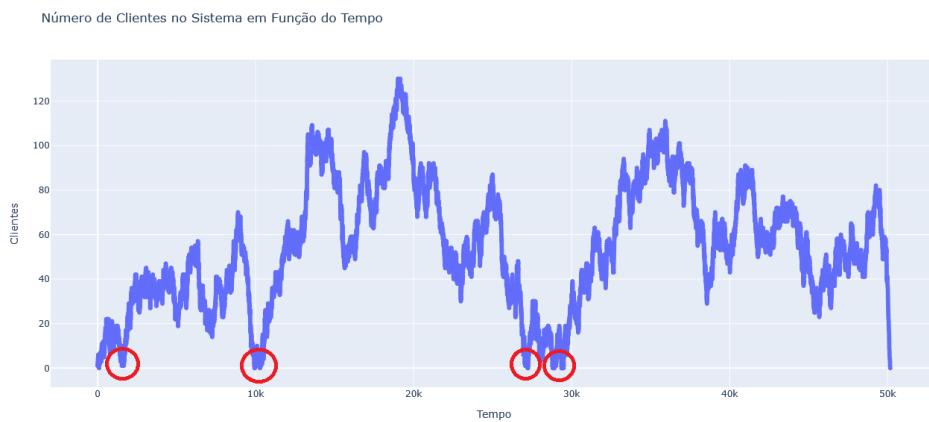


Figura 6.2: Gráfico da evolução temporal a longo prazo do número de clientes no sistema, com renovações circuladas em vermelho.

Note que, a partir do registro dos momentos em que cada cliente foi atendido e seu tempo de chegada, é possível saber seus tempos de espera até serem atendidos, subtraindo um valor do outro. Esse cálculo é usado para obter o tempo médio de espera em fila no sistema, rodando múltiplas simulações e, ao fim de todas elas, calculando a média dos tempos médios de espera de cada rodada. Como exemplo, considere dois cenários com seus respectivos parâmetros e resultados descritos abaixo, cada um simulado 5000 vezes.

Listing 2: Média de 5000 simulações: $T = 480$, $\lambda_c = 1/5$, $\mu_{s_1} = 5$, $\mu_{s_2} = 5$

1	Tempo medio de trabalho dos servidores pos fechamento: 3.38 min
2	Numero Medio de Clientes do Servidor 1: 56.336
3	Numero Medio de Clientes do Servidor 2: 39.842
4	Tempo Medio de Espera na Fila: 1.573 min

Listing 3: Média de 5000 simulações: $T = 480$, $\lambda_c = 1/5$, $\mu_{s_1} = 10$, $\mu_{s_2} = 10$

1	Tempo medio de trabalho dos servidores pos fechamento: 59.66 min
2	Numero Medio de Clientes do Servidor 1: 48.889
3	Numero Medio de Clientes do Servidor 2: 46.846
4	Tempo Medio de Espera na Fila: 30.585 min

No primeiro cenário, temos $\lambda_c < 1/\mu_{s_1} + 1/\mu_{s_2}$, garantindo uma fila estável. Isso é evidenciado por um tempo médio pequeno de espera na fila por cliente e por um tempo médio residual de serviço (depois que o sistema fecha) também pequeno. Por outro lado, o segundo cenário considera o caso crítico com $\lambda_c = 1/\mu_{s_1} + 1/\mu_{s_2}$, no qual a fila não é estável. As simulações apresentam fortes indicativos da falta de estabilidade desse sistema: o tempo médio de espera na fila por cliente é grande, e em média há quase 1 hora de trabalho extra dos servidores após o fechamento do sistema.

Curiosamente, é possível notar que, no primeiro cenário, o servidor 1 atende mais clientes do que o servidor 2, ainda que ambos possuam o mesmo tempo médio de atendimento. Isso se dá pelo fato de um cliente priorizar o servidor 1 em casos onde o sistema se encontra vazio. Essa situação ocorre com frequência, uma vez que os atendimentos dos servidores são suficientemente rápidos. Já no segundo cenário (caso crítico), os servidores ficam ambos ocupados com mais frequência, e a divisão de trabalho é mais equilibrada por existirem menos momentos com o sistema vazio.

6.4. Resultados: Patrimônio de uma Seguradora

Assim como na simulação das filas, o comportamento probabilístico esperado pôde ser verificado de forma bem sucedida no cenário da seguradora. Para esse modelo, os gráficos se mostram ainda mais importantes para uma análise detalhada e otimizada da evolução do sistema. Eles não só evitam a leitura e interpretação de muitos dados, como auxiliam na visualização de períodos de queda ou de crescimento econômico. Ao final de uma simulação, o algoritmo gera dois gráficos: *Patrimônio da Seguradora em Função do Tempo* e *Número de Clientes da Seguradora em Função do Tempo*. Vejam-se as Figuras 6.3 e 6.4.



Figura 6.3: Evolução temporal do patrimônio da seguradora.



Figura 6.4: Evolução temporal do número de clientes da seguradora.

Na simulação ilustrada na Figura 6.3, note que a seguradora permaneceu o período completo de um ano sem falir, apesar de apresentar períodos de queda patrimonial acentuada em

certos intervalos. Para casos em que a seguradora vai à falência, o terminal acusa o ocorrido e retorna o mesmo gráfico patrimonial, sinalizando o patrimônio final menor ou igual a zero.

A visualização dos dados no gráfico facilita o estudo da intensidade das flutuações do capital da seguradora nessa simulação, que foi realizada com os seguintes parâmetros:

- | | |
|--|--|
| <ul style="list-style-type: none"> • $T = 365$ dias • $n_0 = 20$ clientes • $p_0 = 100000$ reais • $c = 60 \text{ dia}^{-1}$ • $\lambda = 12/365 \text{ acionamentos} \cdot \text{dia}^{-1}$ | <ul style="list-style-type: none"> • $\mu_{ac} = 3000$ reais • $\sigma_{ac} = 1$ reais • $\nu = 2/30 \text{ clientes} \cdot \text{dia}^{-1}$ • $e = 1/1095 \text{ dia}^{-1}$ |
|--|--|

Já o segundo gráfico (Figura 6.4), gerado com os mesmos parâmetros, evidencia a tendência de crescimento do número de clientes da seguradora, o que condiz com as taxas fornecidas de permanência e contratação.

Por se tratar de uma simulação única, não é possível avaliar se o sucesso da seguradora com os parâmetros estabelecidos é de fato a norma ou fruto de uma rodada sortuda. Para responder essa questão, foram realizadas 50 mil rodadas com esses parâmetros, registrando os resultados e, por fim, retornando a razão entre os sucessos e o número de iterações, que nada mais é do que uma estimativa da probabilidade de que a seguradora não chegue à falência no primeiro ano:

Listing 4: Média de 50 mil iterações, com os parâmetros acima.

¹ Probabilidade da Seguradora nao falir: 0.997

² Patrimonio Medio da Seguradora no Final do Periodo: 101770.933

Ou seja, o sucesso da seguradora no cenário construído é, de fato, a norma. No entanto, pode-se perceber que seu patrimônio, em média, cresce pouco no período de um ano, apresentando um crescimento de aproximadamente 1,8%. Dessa forma, infere-se, também, que a rodada exemplo anterior foi sortuda quanto ao patrimônio acumulado, já que fechou o ano com um patrimônio de mais de 130 mil reais.

Curiosamente, percebe-se que esse sistema é extremamente sensível ao valor médio dos acionamentos: observe, abaixo, o resultado retornado pelo algoritmo quando mantemos os mesmos parâmetros, mas tomado $\mu_{ac} = 3500$ (apenas 500 reais a mais por acionamento).

Listing 5: Média de 50 mil simulações, $\mu_{ac} = 3500$ reais.

¹	Probabilidade da Seguradora não falir: 0.443
²	Patrimônio Médio da Seguradora no Final do Período: 15460.914

A partir de 3500 reais, a probabilidade de sucesso da seguradora se torna menor do que 50%. Para $\mu_{ac} = 4000$ reais, os resultados das simulações indicam uma probabilidade de falência de 99%. Esse exemplo de sensibilidade a variações numéricas pode levar à questão de encontrar para qual valor x de um dos parâmetros, fixando-se todos os outros, obtém-se uma probabilidade p de sucesso.

Para isso, basta registrar as probabilidades de sucesso da seguradora para cada configuração de parâmetros, configurações essas distintas, à medida que um dos parâmetros é gradativamente incrementado, até que um valor específico do parâmetro alvo resulte na probabilidade desejada. Por exemplo, sabendo que $\mu_{ac} = 3500$ causa uma probabilidade de sucesso de 44% e que, como visto acima, $\mu_{ac} = 4000$ reais causam uma falência praticamente garantida, o valor alvo para uma probabilidade de sucesso de 25% deve estar entre esses dois valores. Realizando incrementos de 10 reais em μ_{ac} e simulando 50 mil rodadas com cada um dos seus valores, tem-se que $\mu_{ac}^* = 3620$ reais é o primeiro valor que resulta em uma probabilidade menor ou igual a 0.25 (nesse caso, 0.241).

Referências bibliográficas

- R. Durrett. *Essentials of stochastic processes*. Springer Texts in Statistics. Springer, Cham, 3 ed., 2016. DOI: [10.1007/978-3-319-45614-0](https://doi.org/10.1007/978-3-319-45614-0).
- V. G. Kulkarni. *Introduction to modeling and analysis of stochastic systems*. Springer Texts in Statistics. Springer, New York, 2 ed., 2011. DOI: [10.1007/978-1-4419-1772-0](https://doi.org/10.1007/978-1-4419-1772-0).
- J. D. C. Little. A Proof for the Queuing Formula: $L = \lambda W$. *Operations Research* 9(3), 383–387. DOI: [10.1287/opre.9.3.383](https://doi.org/10.1287/opre.9.3.383)
- K. V. Mitov; E. Omey. *Renewal processes*. SpringerBriefs in Statistics. Springer, Cham, 2014. DOI: [10.1007/978-3-319-05855-9](https://doi.org/10.1007/978-3-319-05855-9).
- S. M. Ross. *Introduction to probability models*. Academic Press, London, 13 ed., 2023a. DOI: [10.1016/C2021-0-03471-4](https://doi.org/10.1016/C2021-0-03471-4).

S. M. Ross. *Simulation*. Academic Press, New York, 6 ed., 2023b. DOI: [10.1016/C2020-0-00043-5](https://doi.org/10.1016/C2020-0-00043-5).