UNIVERSIDADE ESTADUAL NORTE FLUMINENSE
LCMAT

FILA M/M/∞

Processos Estocásticos

João Víttor Vieira Pinto
00119110369

Campos dos Goytacazes, 06 de Junho de 2024

Sumário

1 Introdução

A teoria de filas é uma ferramenta poderosa para modelar e analisar sistemas onde há variabilidade no tempo de chegada e serviço dos clientes. Em particular, para um site de live stream de pandas, a utilização do modelo M/M/∞ é apropriada devido à natureza imprevisível e altamente variável do tráfego de usuários. Este relatório focará na matemática subjacente ao modelo M/M/∞ e suas implicações para a gestão de desempenho de servidores.

2 Definição do Modelo M/M/∞

O modelo M/M/∞ é caracterizado pelas seguintes propriedades:

2.1 Chegadas (M)

- Memoryless Arrivals
- Processo de Poisson: As chegadas ocorrem aleatoriamente e independentemente umas das outras.
- Distribuição Exponencial: O tempo entre chegadas consecutivas é exponencialmente distribuído.
- λ: Número médio de chegadas por unidade de tempo.

2.2 Serviços (M)

- Memoryless Service
- Distribuição Exponencial: O tempo de serviço de cada cliente é exponencialmente distribuído.
- Propriedade de Sem Memória: A probabilidade de terminar o serviço não depende do tempo já gasto no atendimento.
- μ: Número médio de clientes atendidos por unidade de tempo.

2.3 Número Infinito de Servidores (∞)

- Servidores Infinitos: Não há limite no número de clientes que podem ser atendidos simultaneamente.
- Sem Tempo de Espera: Cada cliente é atendido imediatamente ao chegar no sistema.

2.4 Taxas

- λ: Taxa com que os clientes chegam ao sistema.
- μ: A taxa de partida de um único cliente.
- Taxa de Partida para k Clientes: Quando existem k clientes, a taxa de partida é $k\mu$.
- Taxa de Ocupação (ρ): A taxa de ocupação dos servidores, dada por $\rho = \frac{\lambda}{\mu}$.

3 Análise Matemática

3.1 Distribuição de Probabilidade

No modelo M/M/∞, a probabilidade de haver n usuários conectados ao sistema em um dado instante é dada por uma distribuição de Poisson com média $\frac{\lambda}{\mu}$.

3.2 Demonstrações

3.2.1 Probabilidade de Ocorrência do Estado n

Primeiro, temos a expressão para θ_n:

$$\theta_n = \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} = \frac{\lambda^n}{n! \mu^n}$$

Utilizando essa expressão, a distribuição limite P_n é dada por:

$$P_n = \lim_{t \to +\infty} P[X(t) = n] = \frac{\theta_n}{\sum_{n \geq 0} \theta_n}$$

Substituindo θ_n na expressão acima, temos:

$$P_n = \frac{\frac{\lambda^n}{n! \mu^n}}{\sum_{n \geq 0} \frac{\lambda^n}{n! \mu^n}}$$

Podemos simplificar essa expressão definindo $\rho = \frac{\lambda}{\mu}$.

$$P_n = \frac{\frac{\rho^n}{n!}}{\sum_{n \geq 0} \frac{\rho^n}{n!}}$$

A soma no denominador é a série de Taylor da função exponencial e^ρ:

$$\sum_{n \geq 0} \frac{\rho^n}{n!} = e^\rho$$

Portanto, a expressão de P_n se torna:

$$P_n = \frac{\rho^n}{n!} e^{-\rho}$$

3.2.2 Probabilidade de Ocorrência do Estado 0

Para $n = 0$ (ou seja, a probabilidade de que não haja nenhum cliente no sistema), temos:

$$P_0 = \frac{(\rho^0 / 0!)}{e^\rho}$$

Sabendo que $\rho^0 = 1$ e $0! = 1$, isso se simplifica para:

$$P_0 = \frac{1}{e^\rho}$$

Portanto:

$$P_0 = e^{-\rho}$$

3.2.3 Taxa de Ocupação (ρ)

A taxa de ocupação dos servidores em um dado instante é diretamente proporcional ao número médio de usuários conectados, dado que temos um número infinito de servidores disponíveis.

O número médio de clientes L em uma fila M/M/∞ é dado por:

$$L = \sum_{n=0}^{\infty} n \cdot P_n$$

Substituindo a fórmula da distribuição de Poisson:

$$L = \sum_{n=0}^{\infty} n \cdot \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} e^{-\frac{\lambda}{\mu}}$$

Reconhecendo que:

$$\sum_{n=0}^{\infty} \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} e^{-\frac{\lambda}{\mu}} = 1$$

E que para $n = 0, n \cdot P_n = 0$, temos:

$$L = \sum_{n=1}^{\infty} n \cdot \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} e^{-\frac{\lambda}{\mu}}$$

Pode-se simplificar essa soma ao notar que $n \cdot \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} = \frac{\left(\frac{\lambda}{\mu}\right) \cdot \left(\frac{\lambda}{\mu}\right)^{n-1}}{(n-1)!}$, então:

$$L = \left(\frac{\lambda}{\mu}\right) \sum_{n=1}^{\infty} \frac{\left(\frac{\lambda}{\mu}\right)^{n-1}}{(n-1)!} e^{-\frac{\lambda}{\mu}}$$

Reconhecendo que essa série é a série de Taylor da exponencial, deslocada de um índice:

$$L = \left(\frac{\lambda}{\mu}\right) \cdot \sum_{m=0}^{\infty} \frac{\left(\frac{\lambda}{\mu}\right)^m}{m!} e^{-\frac{\lambda}{\mu}}$$

Essa soma novamente equivale a 1, então:

$$L = \left(\frac{\lambda}{\mu}\right) \cdot 1 = \frac{\lambda}{\mu}$$

Portanto, a taxa de ocupação ρ em uma fila M/M/∞ é:

$$\rho = \frac{\lambda}{\mu}$$

3.2.4 Número Médio de Usuários Conectados

O número médio de usuários conectados ao sistema, L, é dado por:

$$L = \frac{\lambda}{\mu} = \rho$$

3.2.5 Tempo Médio de Conexão

O tempo médio que um usuário passa conectado ao site, W, é:

$$W = \frac{1}{\mu}$$

3.2.6 Tempo Médio de Espera na Fila e Número Médios de Clientes na Fila

A Lei de Little relaciona o número médio de clientes na fila (L_q) com a taxa média de chegada (λ) e o tempo médio de espera na fila (W_q):

$$L_q = \lambda W_q$$

Para $M/M/\infty$, sabemos que $L_q = 0$ porque não há fila. Aplicando a Lei de Little, temos:

$$0 = \lambda W_q$$

Para que essa equação seja verdadeira, W_q deve ser 0:

$$W_q = 0$$

3.3 Métricas de Desempenho

- O número médio de clientes no sistema (L_s) é igual à taxa de ocupação (ρ): $L_s = \rho$.
- O tempo esperado que cada cliente passa no sistema (W_s) é dado pela fórmula inversa da taxa de serviço (μ): $W_s = \frac{1}{\mu}$.
- W_s é inversamente proporcional à taxa de serviço. Quanto maior for a taxa de serviço, menor será o tempo esperado no sistema.
- Cada cliente que chega é atendido imediatamente por um dos infinitos servidores disponíveis. Portanto, não há clientes esperando na fila, ou seja, $L_q = 0$.
- Pelo mesmo motivo, como cada cliente é atendido imediatamente sem esperar, o tempo esperado que um cliente passa na fila (W_q) é zero. Não há tempo de espera antes de ser atendido.

3.3.1 Número Médio de Usuários Conectados

O número médio de usuários conectados ao sistema (L) é dado por:

$$L = \frac{\lambda}{\mu} \tag{1}$$

3.3.2 Tempo Médio de Conexão

O tempo médio que um usuário passa conectado ao site (W) é:

$$W = \frac{1}{\mu} \tag{2}$$

4 Algumas Probabilidades

- **Probabilidade de zero clientes no sistema (P_0)**:
  - Quando a taxa de ocupação é zero ($\rho = 0$), a probabilidade de haver zero clientes no sistema é 1.
  - À medida que a taxa de ocupação aumenta, essa probabilidade diminui.
- **Probabilidade de haver exatamente 1 cliente no sistema (P_1)**:
  - Quando $\rho = 0$, a probabilidade de existir 1 cliente é zero.
  - À medida que a taxa de ocupação aumenta, a probabilidade de existir exatamente 1 cliente aumenta inicialmente, mas nunca chega a 0.5, e depois começa a decrescer novamente.
- **Probabilidade de haver 2 ou mais clientes no sistema (P_n)**:
  - Semelhante ao caso anterior, começa em zero e aumenta com a taxa de ocupação, mas tende a um comportamento de diminuição após certo ponto.

5 Justificativa de Escolha do Modelo M/M/∞

A escolha do modelo M/M/∞ para um site de live stream é justificada pelos seguintes motivos:

- **Flexibilidade e Escalabilidade**: Com um número infinito de servidores, o sistema pode atender qualquer número de usuários simultaneamente sem deterioração na qualidade do serviço.
- **Simplificação Matemática**: As fórmulas resultantes são simples e permitem cálculos diretos e eficientes das métricas de desempenho.
- **Experiência do Usuário**: A ausência de filas garante que os usuários tenham uma experiência de visualização imediata e contínua.

6 Conclusão

O estudo e a aplicação do modelo M/M/∞ para um site de live stream demonstram a eficácia e a adequação deste modelo para cenários de alta variabilidade no tráfego de usuários. Através da análise matemática apresentada, foi possível compreender as principais métricas de desempenho, como o número médio de usuários conectados, o tempo médio de conexão e a taxa de ocupação dos servidores. Essas métricas são fundamentais para garantir uma experiência de usuário satisfatória, sem tempo de espera e com serviço contínuo.