Unidade II

5 OTIMIZAÇÃO DE REDES

A otimização de redes é uma subárea da pesquisa operacional que se concentra no design, planejamento e operação de redes de tal forma que se possa alcançar um determinado objetivo, como maximizar a eficiência ou minimizar os custos, sob certas restrições. Ela lida principalmente com problemas que podem ser representados por uma estrutura de rede, composta por nós e arcos. Esses problemas abrangem diversas aplicações, desde a logística e transporte até a comunicação de dados e distribuição de energia.

São conceitos-chave na otimização de redes:

- **Nó**: ponto individual na rede. Pode representar cidades, computadores, pontos de distribuição, entre outros.
- **Arco**: ligação entre dois nós. Pode representar estradas, cabos de comunicação, tubulações etc. Cada arco tem associado um custo, uma capacidade ou um tempo de viagem.
- **Caminho**: série de arcos que conecta dois nós.
- **Fluxo**: quantidade de produto, informação, tráfego etc. que circula pela rede.

A otimização de redes apresenta alguns problemas comuns:

- **Problema do caminho mais curto**: compreende determinar o caminho mais curto (ou de menor custo) entre dois nós de uma rede – por exemplo, encontrar a rota mais rápida entre duas cidades em um mapa rodoviário.
- **Problema de fluxo máximo**: abrange encontrar o maior fluxo que pode ser enviado de um nó de origem para um nó de destino em uma rede sem exceder as capacidades dos arcos – por exemplo, determinar a máxima quantidade de dados que pode ser enviada de um servidor para um cliente em uma rede de computadores.
- **Problema do transporte**: envolve determinar a forma mais econômica de transportar mercadorias de várias fontes para vários destinos, dadas as capacidades e demandas.
- **Problema de designação**: implica alocar recursos a tarefas para minimizar custos ou maximizar eficiência.
- **Problema de circulação**: visa determinar a distribuição de fluxo em uma rede para satisfazer determinadas demandas e restrições.

Nessa área, uma variedade de técnicas matemáticas e algorítmicas é utilizada para resolver problemas. Algumas das mais notáveis são:

- **Algoritmo de Dijkstra**: usado para resolver o problema do caminho mais curto (Goldbarg, 2014).

- **Algoritmo de Ford-Fulkerson**: empregado no problema de fluxo máximo.

- **Método Simplex especializado**: pode ser adaptado para tratar de problemas de transporte e designação na estrutura de rede.

A otimização de redes é uma ferramenta poderosa na pesquisa operacional, com aplicações em muitos setores da indústria e do dia a dia. Com a ascensão das redes de comunicação, especialmente da internet, a importância dessa área só tem crescido.

5.1 Noções básicas de redes e grafos

Na pesquisa operacional, a teoria de redes e grafos é um campo fundamental, utilizado para representar e analisar sistemas complexos em diversas áreas, como transporte, telecomunicações e logística.

Observação

Um **grafo** (G = (V, E)) consiste em um conjunto (V) de vértices (ou nós) e um conjunto (E) de arestas (ou arcos) que conectam pares de vértices. Um **vértice** (ou nó) representa uma entidade dentro do grafo. Uma **aresta** (ou arco) representa uma relação ou conexão entre dois vértices. Dois vértices são **adjacentes** se estiverem conectados por uma aresta.

Um grafo pode ser dirigido, não dirigido, ponderado ou completo:

- **Grafo dirigido**: as arestas têm uma direção associada, representando relações direcionadas.

- **Grafo não dirigido**: as arestas não têm direção, representando relações bidirecionais.

- **Grafo ponderado**: as arestas têm um peso associado, que pode representar distâncias, custos etc.

- **Grafo completo**: cada par de vértices distintos é conectado por uma aresta única.

Vamos ver algumas noções básicas de redes e grafos:

- **Grafo conexo**: um grafo é conexo se existir um caminho entre cada par de vértices (Goldbarg; Goldbarg, 2021).

- **Caminho**: sequência de arestas que conecta um vértice a outro sem repetir vértices ou arestas.
- **Ciclo**: caminho fechado, ou seja, que começa e termina no mesmo vértice, sem repetir outros vértices ou arestas.
- **Árvore**: grafo conexo, acíclico e não dirigido.
- **Floresta**: conjunto de árvores desconexas.
- **Grau de entrada**: número de arestas dirigidas que entram em um vértice.
- **Grau de saída**: número de arestas dirigidas que saem de um vértice.
- **Lista de adjacência**: cada vértice possui uma lista de seus vértices adjacentes.
- **Matriz de adjacência**: matriz bidimensional cujas células representam a presença (ou peso) de arestas entre pares de vértices.

A teoria de redes e grafos pode ser aplicada na pesquisa operacional para solucionar o problema do caminho mais curto, o problema do fluxo máximo em um grafo dirigido e ponderado ou o problema do caixeiro viajante (PCV), que consiste em encontrar o ciclo hamiltoniano de menor custo em um grafo completo ponderado.

Essa teoria fornece um framework para modelar e resolver problemas de decisão em muitas áreas, oferecendo técnicas matemáticas para análise de redes e desenvolvimento de algoritmos eficientes.

5.2 Aplicações e métodos de resolução

No contexto da otimização de redes na pesquisa operacional, diversos métodos de resolução e aplicações são empregados para melhorar o desempenho de redes e sistemas.

Entre essas aplicações estão os já mencionados problema do fluxo máximo (aplicável às áreas de telecomunicações e sistemas de distribuição) e problema do caminho mais curto (aplicável a redes de computadores e sistemas de transporte). Além dessas, há ainda algumas outras aplicações, como **otimização de rede de distribuição**, que objetiva minimizar o custo de distribuição de produtos e é aplicável às áreas de logística e cadeia de suprimentos, e **problema de roteamento de veículos**, que visa minimizar a distância total percorrida por uma frota de veículos e é aplicável à área de logística e transporte.

A seguir estão alguns dos métodos de resolução utilizados:

- Programação linear, frequentemente usada para modelar problemas de otimização de redes. Emprega os métodos Simplex e do ponto interior.

- Algoritmo de Dijkstra, usado para resolver problemas do caminho mais curto em grafos com arestas de peso não negativo utilizando o método do relaxamento de arestas e seleção gulosa.

- Algoritmo de Ford-Fulkerson, empregado para resolver problemas de fluxo máximo em redes aplicando o método do aumento de caminhos e corte mínimo.

- Programação dinâmica, usada em problemas cuja solução ótima pode ser construída a partir de soluções ótimas de subproblemas. Emprega o método da divisão do problema em subproblemas e armazenamento de soluções intermediárias.

- Meta-heurística, aplicada para prover soluções aproximadas quando métodos exatos são impraticáveis, utilizando algoritmos genéticos, busca tabu, e *simulated annealing*.

- Decomposição de Benders, empregada em problemas de grande escala com estrutura de blocos, decompondo-os em subproblemas menores e aplicando resolução iterativa.

5.3 Problema do caixeiro viajante (caminho mínimo)

O PCV é um problema clássico que imagina um vendedor (o caixeiro viajante) que precisa visitar várias localidades e quer encontrar a rota mais curta que passa por cada cidade uma única vez e retorna ao ponto de partida. É aplicado principalmente nas áreas de:

- **Logística**: empresas de entrega usam variações do PCV para planejar rotas de veículos de entrega.

- **Turismo**: agências de viagem o usam para otimizar itinerários.

- **Tecnologia**: redes de computadores o utilizam para encontrar caminhos eficientes entre seus nós.

- **Manufatura**: indústrias o aplicam para otimizar trajetos de ferramentas de corte em máquinas CNC.

As origens do PCV remontam ao século XIX, quando o matemático irlandês William R. Hamilton buscava maneiras de solucionar o problema dos ciclos hamiltonianos. Ao longo dos anos, muitos matemáticos e cientistas da computação contribuíram para o desenvolvimento de métodos para resolver o PCV.

Apesar de parecer simples, este é um problema considerado NP-difícil, o que significa que não se conhece uma maneira de resolvê-lo rapidamente para um número grande de vértices. Devido à sua complexidade, muitas abordagens práticas envolvem métodos aproximados e heurísticas que podem encontrar soluções boas (mas não necessariamente ótimas) em tempo razoável.

Observação

O PCV intriga e desafia matemáticos e cientistas da computação há décadas, e continua a ser um tópico fundamental de pesquisa. Ele pergunta se há um caminho mais curto que visita um conjunto de vértices e volta ao ponto de partida (Arenales; Armentano, 2006).

5.4 Algoritmo de Dijkstra para problemas com menor caminho entre dois nós

O famoso algoritmo de Dijkstra é utilizado para resolver o problema do caminho mais curto em um grafo dirigido ou não dirigido com arestas de peso não negativo. Ele encontra o caminho mais curto de um vértice inicial para todos os outros vértices no grafo.

O algoritmo foi concebido pelo cientista da computação Edsger W. Dijkstra em 1956 e publicado em 1959. Dijkstra o desenvolveu como parte de uma demonstração de um conceito de programação relacionado ao sistema operacional Armac.

Assim funciona o algoritmo:

- **Inicialização**: atribui um valor de distância zero a um nó inicial e infinito aos outros nós.
- **Relaxamento**: para cada vizinho do nó atual, calcula-se a distância do nó inicial a este vizinho. Se a nova distância calculada for menor que a já registrada para esse vizinho, atualiza-se a distância.
- **Seleção**: após visitar todos os vizinhos do nó atual, escolhe-se o nó não visitado com a menor distância registrada para ser o próximo nó atual.
- **Repetição**: repetem-se os passos de relaxamento e seleção até que todos os nós tenham sido visitados.

Esse algoritmo pode ser aplicado em diversas áreas. Nas redes de computadores, é utilizado para encontrar o caminho mais curto entre nós em redes de roteamento. Nos sistemas de transporte, auxilia no desenvolvimento de sistemas de navegação GPS, encontrando a rota mais rápida entre dois pontos. Nas redes sociais, pode ser usado para analisar relações entre usuários. Em jogos de computador, pode ser utilizado para calcular o caminho que um personagem deve seguir para alcançar um objetivo.

O algoritmo de Dijkstra funciona apenas com arestas de peso não negativo, pois com pesos negativos a escolha gulosa do próximo vértice com menor distância pode não ser ótima. Para grafos com arestas de peso negativo, um algoritmo alternativo, como o de Bellman-Ford, pode ser utilizado. Dijkstra é um exemplo de algoritmo guloso, pois faz escolhas localmente ótimas na esperança de encontrar uma solução globalmente ótima.

Esse algoritmo continua sendo fundamental em teoria dos grafos e ciência da computação, com aplicações práticas em diversas áreas da tecnologia, contribuindo significativamente para o campo da pesquisa operacional.

5.5 Algoritmo de Kruskal para otimização de redes

O algoritmo de Kruskal é um método popular e clássico na pesquisa operacional para encontrar a árvore geradora mínima (MST, *minimum spanning tree*) de um grafo conectado e ponderado. Esta árvore é uma sub-rede que conecta todos os vértices com o menor custo total possível e sem formar ciclos. O objetivo desse algoritmo é minimizar o custo total da rede garantindo que todos os pontos (nós ou vértices) estejam interligados.

Vejamos o passo a passo do algoritmo de Kruskal:

- **Inicialização**: comece com uma floresta, que consiste em conjuntos isolados de vértices, em que cada vértice representa uma árvore individual na floresta. Liste todas as arestas do grafo e organize-as em ordem crescente de peso (ou custo).

- **Iteração**: selecione a aresta com o menor peso e a remova da lista. Verifique os dois vértices da aresta. Se eles pertencerem a árvores diferentes na floresta, adicione esta aresta à solução e una as duas árvores para formar uma única árvore na floresta. Se os dois vértices pertencerem à mesma árvore, ignore esta aresta (pois adicioná-la formaria um ciclo) e passe para a próxima aresta de menor peso.

- **Terminação**: o algoritmo termina quando há apenas uma árvore na floresta, que é a MST desejada, ou quando todas as arestas foram consideradas. Se o grafo não for conectado, o algoritmo resultará em uma floresta geradora mínima, que é um conjunto de MSTs para cada componente conectado do grafo.

A eficiência do algoritmo de Kruskal depende de como é implementado. Com estruturas de dados eficientes, como o Union-Find, a complexidade pode ser quase linear em relação ao número de arestas.

Na pesquisa operacional, o algoritmo de Kruskal é frequentemente usado em problemas de otimização de redes, como planejamento de redes de transporte ou de comunicação. A ideia é garantir a conectividade com o mínimo de custo.

Suas vantagens abrangem sua simplicidade e facilidade de implementação e sua adequação para grafos esparsos, enquanto sua principal desvantagem se deve ao fato de poder não ser o mais eficiente para grafos densos em comparação com outros algoritmos de MST, como o de Prim.

Em resumo, o algoritmo de Kruskal é uma ferramenta valiosa na pesquisa operacional para otimização de redes, garantindo uma solução eficiente para encontrar a MST em grafos conectados e ponderados.

Exemplo de aplicação

Uma empresa de tecnologia da informação deseja interligar seus escritórios em várias cidades através de uma rede de comunicação. O objetivo é garantir que cada escritório possa se comunicar com todos os outros diretamente ou através de intermediários. Para isso, a empresa considera estabelecer links diretos entre algumas cidades. O custo de estabelecer um link varia dependendo da distância e das infraestruturas existentes entre as cidades.

Dados

A seguir estão as cidades (escritórios) e os custos para estabelecer links diretos:

Tabela 8

Cidade	A	B	C	D	E
A	–	4	2	5	10
B	4	–	1	7	8
C	2	1	–	6	3
D	5	7	6	–	9
E	10	8	3	9	–

A empresa quer minimizar o custo total de estabelecer a rede enquanto garante a conectividade entre todos os escritórios.

Solução

Para solucionar este problema, podemos utilizar o algoritmo de Kruskal, que busca uma MST em um grafo conectado, aplicando os seguintes passos:

- Liste todas as arestas em ordem crescente de peso (ou custo).
- Comece com uma floresta (um conjunto de árvores), em que cada vértice é uma árvore separada.
- Combine as florestas em uma única MST, adicionando a aresta de menor peso que conecta duas árvores diferentes.

Aplicando o algoritmo de Kruskal:

- Arestas em ordem crescente de peso:

C-B (1), A-C (2), C-E (3), A-B (4), A-D (5), C-D (6), B-D (7), B-E (8), D-E (9), A-E (10)

- Comece com uma floresta com vértices {A}, {B}, {C}, {D}, {E}

- Arestas escolhidas na MST:

C-B (1): conecta {C} e {B}

A-C (2): conecta {A} e {C, B}

C-E (3): conecta {C, B, A} e {E}

A-D (5): conecta {C, B, A, E} e {D}

As outras arestas são descartadas, pois adicioná-las não traria nenhum benefício em termos de conectividade e apenas aumentaria o custo.

Resultado

A rede otimizada terá links diretos entre as cidades C-B, A-C, C-E e A-D, com um custo total de 1 + 2 + 3 + 5 = 11 unidades monetárias.

Esta solução garante que cada escritório está conectado diretamente ou através de intermediários com o menor custo possível.

6 TEORIA DA DECISÃO

A teoria da decisão é um campo interdisciplinar que oferece ferramentas para tomar decisões mais eficazes sob diferentes tipos de incertezas e restrições. Ela se sobrepõe significativamente com a pesquisa operacional, especialmente no que diz respeito à modelagem e análise de problemas complexos que envolvem escolhas.

Seus conceitos-chave são:

- **Tomadores de decisão**: indivíduos ou entidades que têm a responsabilidade de escolher entre várias alternativas.
- **Alternativas**: as diferentes opções disponíveis para escolha.
- **Critérios de decisão**: as métricas usadas para avaliar as diferentes alternativas. Esses critérios podem ser quantitativos (como custo e tempo) ou qualitativos (como satisfação do cliente ou reputação).
- **Matriz de pagamento**: representação tabular das recompensas (ou custos) associadas às várias combinações de alternativas e estados da natureza.
- **Risco e incerteza**: risco se refere às situações em que as probabilidades dos diferentes estados da natureza são conhecidas. Incerteza se refere às situações em que se desconhecem essas probabilidades.
- **Estratégias**: planos de ação que especificam qual alternativa seguir para cada possível estado da natureza.

A seguir são apresentados os métodos e modelos da teoria da decisão:

- **Árvores de decisão**: gráfico que representa as diferentes alternativas, os eventos incertos e seus respectivos pagamentos. Este modelo gráfico ajuda a visualizar e avaliar opções complexas.

- **Análise de sensibilidade**: estudo que examina como as mudanças nas entradas de um modelo de decisão afetam a decisão ótima. É particularmente útil quando há incerteza nos parâmetros do modelo.

- **Programação matemática**: técnicas como programação linear, programação inteira e programação não linear podem ser usadas para resolver problemas de decisão de otimização.

- **Teoria dos jogos**: subcampo que trata de situações em que a decisão de um jogador afeta e é afetada pelas decisões de outros jogadores. É utilizada, por exemplo, em problemas de leilão, negociação e competição de mercado.

- **Análise multicritério**: método usado quando há mais de um critério de decisão, e esses critérios muitas vezes são conflitantes. A ideia é encontrar uma solução que seja o melhor compromisso entre os diferentes critérios.

A teoria da decisão é frequentemente aplicada em conjunto com outras técnicas de pesquisa operacional para resolver problemas em áreas como gestão da cadeia de suprimentos, planejamento estratégico, política de preços, alocação de recursos e planejamento de operações em serviços de saúde, transporte e logística.

Saiba mais

A teoria da utilidade esperada, um componente crucial da teoria da decisão, foi formalizada por John von Neumann e Oskar Morgenstern no livro *Theory of games and economic behavior*, de 1944. Eles introduziram conceitos fundamentais para a economia, como o equilíbrio de Nash.

NEUMANN, J.; MORGENSTERN, O. *Theory of games and economic behavior*. Tóquio: Interbooks, 2021.

O livro indicado a seguir pode ser consultado para entender mais da teoria da decisão:

HILLIER, F. S. *et al. Introdução à ciência da gestão*: modelagem e estudos de caso com planilhas eletrônicas. 4. ed. Porto Alegre: AMGH, 2014.

6.1 Matriz de decisão

Na teoria da decisão, no contexto da pesquisa operacional, uma matriz de decisão, também chamada de método de pontuação ponderada, é uma ferramenta útil para analisar e facilitar a escolha entre várias estratégias ou opções disponíveis, especialmente quando há incertezas e riscos envolvidos. Uma matriz de decisão permite que os decisores comparem diferentes alternativas com base em vários critérios, de forma a escolher a alternativa mais adequada de acordo com os objetivos estabelecidos.

Uma matriz de decisão geralmente consiste em:

- **Linhas**: representando diferentes alternativas de ação ou estratégias de decisão.
- **Colunas**: representando diferentes estados da natureza, cenários ou resultados possíveis.
- **Células**: representando os payoffs, resultados ou valores associados a uma combinação específica de alternativa e estado da natureza.

Para usar uma matriz de decisão, percorra os seguintes passos:

- **Liste alternativas de ação**: identifique as possíveis estratégias ou opções de decisão.
- **Identifique estados da natureza**: liste os possíveis cenários ou condições futuras.
- **Determine payoffs**: atribua valores aos resultados possíveis para cada combinação de alternativa e estado da natureza.
- **Avalie alternativas**: analise os payoffs para determinar a melhor alternativa de ação.

Esse método é vantajoso por ser sistemático e estruturado (pois permite uma análise organizada das opções disponíveis), flexível (pode ser usado em uma variedade de situações de decisão e é facilmente adaptável) e por tornar a tomada de decisão mais transparente, já que todos os critérios e ponderações são explicitamente considerados. Já suas desvantagens envolvem sua subjetividade (a atribuição de pesos e pontuações pode variar entre diferentes avaliadores) e complexidade (em situações com muitos critérios ou alternativas, a técnica pode se tornar complexa e demorada).

Exemplo de aplicação

Suponha que uma empresa de software tem três alternativas: desenvolver um novo produto, melhorar um produto existente ou não fazer nada. Existem dois possíveis estados da natureza: o mercado pode ter uma alta demanda ou uma baixa demanda pelo produto.

Tabela 9

	Alta demanda	Baixa demanda
Desenvolver novo	100	50
Melhorar existente	50	10
Não fazer nada	0	0

Neste exemplo, os números na matriz representam os payoffs (lucros ou perdas) associados a cada combinação de alternativa e estado da natureza.

Ao usar uma matriz de decisão, os decisores também podem considerar:

- **Probabilidades**: se conhecidas, as probabilidades associadas a cada estado da natureza podem ser usadas para calcular o valor esperado de cada alternativa.
- **Atitudes de risco**: a escolha da alternativa pode depender da aversão, neutralidade ou preferência ao risco do decisor.
- **Restrições e preferências**: outros fatores, como restrições de recursos e preferências pessoais ou organizacionais, também podem influenciar a decisão final.

Em suma, uma matriz de decisão na pesquisa operacional fornece uma estrutura sistemática e lógica para comparar alternativas de ação sob incerteza e risco, ajudando os decisores a fazer escolhas informadas e bem ponderadas.

6.2 Decisão tomada sob risco

Na pesquisa operacional, a tomada de decisão sob risco é um contexto em que o tomador de decisão tem alguma informação sobre as probabilidades dos diferentes estados da natureza, os quais se referem a uma situação ou resultado possível que está fora do controle do tomador de decisão.

Os principais componentes dessa modalidade são:

- **Decisões**: as ações que o tomador de decisão pode escolher.
- **Estados da natureza**: os cenários que podem ocorrer no futuro, sobre os quais o tomador de decisão não tem controle.
- **Probabilidades**: estimativas quantitativas da chance de ocorrência de cada estado da natureza.
- **Payoffs (pagamentos)**: os resultados associados a cada combinação de decisão e estado da natureza.

Vejamos os métodos de tomada de decisão sob risco:

- **Critério maximax (otimista)**: seleciona a decisão com o maior payoff possível.

- **Critério maximin (pessimista)**: seleciona a decisão cujo pior payoff seja o melhor entre os piores payoffs possíveis.

- **Critério de Laplace (princípio da razão insuficiente)**: assume que todos os estados da natureza são igualmente prováveis e calcula a média de payoffs para cada decisão, escolhendo a decisão com a média mais alta.

- **Critério de Savage**: minimiza o arrependimento máximo, que é a diferença entre o payoff real e o melhor payoff possível em cada estado da natureza.

- **Valor esperado**: calcula o valor esperado de cada decisão e escolhe a decisão com o maior valor esperado.

Exemplo de aplicação

Um investidor precisa decidir sua próxima empreitada. Para isso, delineia os seguintes componentes:

- **Decisões**: investir em ações, títulos, ou imóveis.

- **Estados da natureza**: mercado em alta, mercado estável, mercado em baixa.

- **Probabilidades**: estimativas são associadas a cada estado da natureza, baseadas em dados históricos ou julgamento subjetivo.

- **Payoffs**: retornos financeiros associados a cada combinação de decisão e estado da natureza.

O tomador de decisão pode então usar os métodos já discutidos para escolher a melhor decisão com base nas probabilidades dos estados da natureza e nos payoffs associados a cada combinação de decisão e estado da natureza.

A tomada de decisão sob risco é crucial em várias áreas, incluindo finanças, em que as empresas têm de tomar decisões de investimento sob incerteza, e logística, em que as empresas têm de decidir sobre níveis de estoque considerando a variabilidade na demanda.

O principal desafio na tomada de decisão sob risco é a precisão da estimativa das probabilidades. Uma estimativa imprecisa pode levar a decisões subótimas e a significativas perdas financeiras.

6.3 Decisão tomada sob incerteza

A decisão sob incerteza na pesquisa operacional ocorre quando o decisor não tem informações suficientes sobre as probabilidades dos diferentes estados da natureza, ou seja, o tomador de decisão desconhece as chances de ocorrência de cada possível cenário. A ausência de conhecimento probabilístico faz com que o decisor use outros critérios para avaliar as opções.

Seus principais componentes são semelhantes aos da decisão tomada sob risco – decisões, estados da natureza e payoffs (suprimindo apenas as probabilidades) –, assim como seus métodos, que envolvem os critérios maximax, maximin e de Laplace, mas incluem também o critério de Hurwicz (compromisso), que decide com base em um coeficiente de otimismo que pondera o melhor e o pior payoff de cada decisão.

Exemplo de aplicação

Aproveitando os dados do exemplo anterior, temos:

- **Decisões**: investir em ações, títulos, ou imóveis.
- **Estados da natureza**: mercado em alta, mercado estável, mercado em baixa.
- **Payoffs**: retornos financeiros associados a cada combinação de decisão e estado da natureza.

Sem informações probabilísticas, o decisor aplica um dos critérios mencionados para tomar uma decisão com base nos possíveis payoffs.

Os métodos de decisão sob incerteza podem ser aplicados, por exemplo, na gestão de negócios, para decidir sobre investimentos, desenvolvimento de produtos ou estratégias de marketing quando as empresas não têm informações claras sobre o futuro do mercado. Esses métodos também podem ser aplicados na medicina: profissionais de saúde podem usá-los para escolher tratamentos quando as probabilidades de sucesso são desconhecidas.

O principal desafio ao tomar decisões sob incerteza é a falta de informações sobre as probabilidades dos estados da natureza, tornando a escolha muito dependente das preferências e aversão ao risco do tomador de decisão.

6.4 Critério maximax

Como já apresentamos, o critério maximax é considerado otimista, e é também conhecido como o critério do melhor dos melhores. Quando um tomador de decisão o utiliza, ele examina os melhores possíveis payoffs para cada alternativa de ação e escolhe a que oferece o maior payoff máximo. Como funciona:

- **Liste as decisões possíveis**: o primeiro passo é listar todas as decisões ou estratégias de ação que podem ser tomadas.

- **Determine os payoffs**: para cada decisão possível, liste os payoffs (resultados) associados a cada estado da natureza possível.

- **Identifique o melhor payoff**: para cada decisão, defina o melhor payoff possível.

- **Escolha a decisão com o maior payoff**: a decisão que tem o melhor dos melhores payoffs é escolhida.

Exemplo de aplicação

Suponha que um investidor esteja considerando três diferentes investimentos: A, B e C. Os payoffs possíveis associados a cada investimento, dependendo do estado da economia, podem ser representados da seguinte forma:

- **Investimento A**:
  - Economia boa: R$ 100.000,00.
  - Economia regular: R$ 50.000,00.
  - Economia ruim: R$ 20.000,00.
- **Investimento B**:
  - Economia boa: R$ 80.000,00.
  - Economia regular: R$ 40.000,00.
  - Economia ruim: R$ 30.000,00.
- **Investimento C**:
  - Economia boa: R$ 60.000,00.
  - Economia regular: R$ 45.000,00.
  - Economia ruim: R$ 35.000,00.

Usando o critério maximax, temos que:

- O melhor payoff para o investimento A é R$ 100.000.
- O melhor payoff para o investimento B é R$ 80.000.
- O melhor payoff para o investimento C é R$ 60.000.

Assim, usando o critério maximax, o investidor otimista escolheria o investimento A, pois este tem o maior payoff máximo entre todas as opções disponíveis.

A vantagem do critério maximax é ser simples e fácil de aplicar, sendo útil para tomadores de decisão que são naturalmente otimistas. Contudo, esse critério pode ser excessivamente otimista, especialmente quando os melhores payoffs são improváveis, podendo levar a decisões que não consideram adequadamente os riscos envolvidos.

O critério maximax pode ser uma abordagem eficaz quando se tem uma visão muito positiva do futuro, ou quando o tomador de decisão está disposto a assumir riscos para alcançar recompensas significativas. Porém, é crucial estar ciente das limitações deste critério e considerar se ele é o mais adequado ao contexto e à natureza do problema de decisão enfrentado.

6.5 Critério maximin

O critério maximin é considerado pessimista, e é o oposto do critério maximax. Quando se usa o critério maximin, considera-se o pior cenário possível para cada alternativa e seleciona-se a decisão cujo pior resultado seja o melhor dentre todos os piores resultados possíveis. Como funciona:

- **Liste as decisões possíveis**: relacione todas as decisões ou estratégias de ação possíveis.
- **Determine os payoffs**: para cada alternativa, os payoffs (resultados) associados a cada possível estado da natureza são identificados.
- **Identifique o pior payoff**: para cada alternativa, defina o pior payoff possível.
- **Escolha a alternativa com o melhor dos piores payoffs**: a decisão que tem o melhor dos piores payoffs é escolhida.

Exemplo de aplicação

Suponha que uma empresa está considerando três diferentes projetos de investimento: X, Y e Z. Os payoffs possíveis associados a cada projeto, dependendo do estado da economia, podem ser representados da seguinte forma:

- **Projeto X**:
  - Economia boa: R$ 100.000,00.
  - Economia regular: R$ 50.000,00.
  - Economia ruim: R$ 20.000,00.
- **Projeto Y**:
  - Economia boa: R$ 80.000,00.
  - Economia regular: R$ 40.000,00.
  - Economia ruim: R$ 30.000,00.
- **Projeto Z**:
  - Economia boa: R$ 60.000,00.
  - Economia regular: R$ 45.000,00.
  - Economia ruim: R$ 35.000,00.

Usando o critério maximin, temos:

- O pior payoff para o Projeto X é R$ 20.000.
- O pior payoff para o Projeto Y é R$ 30.000.
- O pior payoff para o Projeto Z é R$ 35.000.

Assim, usando o critério maximin, a empresa escolheria o Projeto Z, pois este tem o melhor payoff no pior cenário entre todas as opções disponíveis.

O critério maximin é uma abordagem conservadora e pode ser útil quando o tomador de decisão quer evitar riscos e minimizar possíveis perdas. Contudo, pode ser excessivamente conservador, especialmente quando os piores payoffs são muito improváveis, levando a decisões que podem não explorar oportunidades potencialmente lucrativas.

Esse critério pode ser uma abordagem eficaz quando se tem uma visão muito cautelosa do futuro, ou quando o tomador de decisão quer evitar riscos a todo custo. No entanto, é importante considerar o contexto e os trade-offs envolvidos e ponderar se uma abordagem mais equilibrada pode ser mais adequada, dependendo da situação específica e das preferências do tomador de decisão.

Exemplo de aplicação

Uma empresa de desenvolvimento de software que emprega práticas de Scrum e DevOps está considerando adotar uma nova ferramenta para facilitar a integração e entrega contínuas (CI/CD). A empresa identificou três opções de ferramentas: A, B e C, cada uma com seu conjunto de prós e contras. A empresa deseja tomar uma decisão baseada na maximização da utilidade considerando diversos critérios, como custo, facilidade de uso, integração com ferramentas existentes e suporte da comunidade.

Os critérios (com pesos relativos entre parênteses) são:

- Custo (0.4).
- Facilidade de uso (0.3).
- Integração (0.2).
- Suporte da comunidade (0.1).

As pontuações (quanto mais alta, melhor) para cada critério e para cada ferramenta, com base na avaliação de especialistas, são as seguintes:

Tabela 10

Ferramenta	Custo	Facilidade de uso	Integração	Suporte
A	7	8	9	7
B	8	7	8	6
C	6	8	7	8

Solução

Para tomar a decisão correta, podemos utilizar o método de pontuação ponderada (matriz de decisão). Este método leva em consideração os pesos relativos de cada critério para calcular uma pontuação total para cada alternativa.

$$\text{Pontuação Total da Ferramenta} =$$
$$\left(P_{\text{Custo}} \times W_{\text{Custo}}\right) +$$
$$\left(P_{\text{Facilidade}} \times W_{\text{Facilidade}}\right) +$$
$$\left(P_{\text{Integração}} \times W_{\text{Integração}}\right) +$$
$$\left(P_{\text{Suporte}} \times W_{\text{Suporte}}\right)$$

Em que:

- P é a pontuação dada para a ferramenta em cada critério.
- W é o peso de cada critério.

Cálculos

- Para a ferramenta A:

Pontuação total A = $(7 \times 0.4) + (8 \times 0.3) + (9 \times 0.2) + (7 \times 0.1)$

Pontuação total A = $2.8 + 2.4 + 1.8 + 0.7$

Pontuação total A = 7.7

- **Para a ferramenta B**:

Pontuação total B = $(8 \times 0.4) + (7 \times 0.3) + (8 \times 0.2) + (6 \times 0.1)$

Pontuação total B = $3.2 + 2.1 + 1.6 + 0.6$

Pontuação total B = 7.5

- **Para a ferramenta C**:

Pontuação total C = $(6 \times 0.4) + (8 \times 0.3) + (7 \times 0.2) + (8 \times 0.1)$

Pontuação total C = $2.4 + 2.4 + 1.4 + 0.8$

Pontuação total C = 7.0

Decisão

Com base no método de pontuação ponderada, a ferramenta A, com uma pontuação total de 7.7, parece ser a melhor escolha, pois maximiza a utilidade de acordo com os critérios estabelecidos e os pesos dados.

Lembrete

O método de pontuação ponderada é uma ferramenta valiosa para tomar decisões que envolvem múltiplos critérios, ajudando a organizar, quantificar e analisar as opções de maneira sistemática. Contudo, como toda ferramenta, deve ser usado com discernimento, levando em conta suas limitações. É importante lembrar que os pesos e pontuações são muitas vezes subjetivos e podem depender do contexto específico da empresa e da equipe.

7 TEORIA DAS FILAS

A teoria das filas é um ramo da pesquisa operacional que estuda como pessoas ou objetos se acumulam e aguardam para serem atendidos. Dada a prevalência de filas em muitos cenários da vida real – de supermercados e hospitais a sistemas computacionais –, entender e otimizar esse processo é crucial para melhorar a eficiência e a satisfação do cliente.

Saiba mais

A teoria das filas foi formalizada pelo engenheiro dinamarquês Agner K. Erlang em 1909, enquanto trabalhava para a Copenhagen Telephone Company. Ele desejava entender como as linhas telefônicas poderiam ser usadas de forma mais eficiente. A teoria das filas pode ser mais detalhadamente estudada nas obras a seguir:

FOGLIATTI, M. C.; MATTOS, N. M. C. *Teoria de filas*. Rio de Janeiro: Interciência, 2006.

PRADO, D. *Teoria das filas e da simulação*. 6. ed. Belo Horizonte: Falconi, 2017. v. 2.

Seus conceitos básicos são:

- **População e fonte de entrada**: refere-se à origem dos clientes, ou seja, de onde eles vêm. Essa população pode ser limitada (como um número fixo de aviões em um aeroporto) ou ilimitada (como clientes chegando a um supermercado).

- **Servidores**: representa o recurso ou entidade que atende os clientes. Pode haver um ou vários servidores – por exemplo, os caixas em um supermercado.

- **Mecanismo de atendimento**: define a maneira como os clientes são atendidos. Pode ser em ordem de chegada (Fifo), por prioridade etc.

- **Disciplina da fila**: refere-se à maneira como os clientes são selecionados para atendimento.

- **Capacidade do sistema**: indica o número máximo de clientes (na fila e em serviço) que o sistema pode acomodar.

Vejamos agora as características-chave da teoria das filas:

- **Taxa de chegada (λ)**: é a taxa média de clientes que chegam ao sistema por unidade de tempo.

- **Taxa de atendimento (μ)**: é a taxa média de clientes que um servidor pode atender por unidade de tempo.

- **Nível de utilização (ρ)**: é a fração de tempo em que o servidor está ocupado.

- **Número médio de clientes no sistema (L)**: é o cálculo que inclui clientes que estão esperando e sendo atendidos.

- **Tempo médio que um cliente gasta no sistema (W)**: é a soma do tempo que um cliente gasta esperando na fila e sendo atendido.

Diversos modelos foram desenvolvidos para analisar sistemas de filas, sendo o M/M/1 um dos mais básicos e amplamente estudados. A nomenclatura M/M/1 refere-se a:

- **M**: chegadas com distribuição de Poisson.

- **M**: atendimento com distribuição exponencial.

- **1**: um único servidor.

A teoria das filas é frequentemente utilizada na pesquisa operacional, por exemplo, para determinar o número ótimo de servidores em um call center; reduzir o tempo de espera em um hospital ou banco; melhorar a eficiência de sistemas de transporte; e otimizar o desempenho de sistemas computacionais e redes de comunicação.

Essa teoria oferece uma abordagem sistemática e matemática para entender e otimizar sistemas em que a espera é uma característica inerente. Em pesquisa operacional, é uma ferramenta fundamental para projetar sistemas mais eficientes e melhorar a experiência do cliente em diversos setores.

7.1 Introdução à teoria das filas

Filas de espera aparecem em diversos sistemas de produção, particularmente em sistemas de serviço, tais como bancos, supermercados, correios, postos de gasolina e sistemas de manufatura, com produtos aguardando processamento em máquinas ou estações de trabalho. Também aparecem em sistemas de transporte, como em aviões esperando para aterrissar em aeroportos, navios esperando para descarregar em portos e sistemas computacionais, por exemplo, com tarefas aguardando processamento ou pacotes de dados aguardando transmissão. Em geral, os usuários (clientes, produtos, veículos, tarefas) desses sistemas se deslocam até os servidores (caixas de atendimento, máquinas, pistas de aterrissagem, computadores) para obter algum tipo de serviço (sistemas usuários-para-servidores).

Mas também há casos em que os servidores se deslocam para atender os usuários (sistemas servidores-para-usuários), como no atendimento emergencial (ambulâncias, bombeiros, viaturas de polícia), em serviços de coleta e/ou entrega (caminhões de coleta de lixo, entregadores de alimentos e de remédios em domicílio) e em sistemas de manutenção (assistência técnica em domicílio).

Nestes casos, as filas de usuários ficam dispersas (espacialmente distribuídas), ao invés de concentradas fisicamente na frente dos servidores, como ocorre nos caixas de bancos, supermercados e correios.

A teoria de filas (ou teoria de congestão), inicialmente motivada por aplicações em sistemas telefônicos, é um ramo da pesquisa operacional que estuda as relações entre as demandas em um sistema e os atrasos sofridos pelos seus usuários. Filas se formam se a demanda excede a capacidade do sistema de fornecer o serviço em certo período. A teoria de filas auxilia no projeto e na operação dos sistemas para encontrar um balanceamento adequado entre os custos de oferecer serviços no sistema (por exemplo, custos operacionais, custos de capacidade) e os custos dos atrasos sofridos pelos usuários. Dependendo do sistema, os custos associados aos atrasos podem ser muito altos (como em um sistema de atendimento médico emergencial) ou nem tanto (como em um sistema de coleta de lixo).

A figura a seguir ilustra uma curva de trade-off (compromisso) entre a capacidade de um sistema (por exemplo, o número de caixas operando em um supermercado, o número de cabines abertas em uma praça de pedágio) e o tempo médio de espera dos usuários (os consumidores esperando em fila com suas compras, os veículos esperando em fila na frente das cabines). Note que o tempo de espera diminui com o aumento da capacidade e vice-versa, e, como os custos de capacidade em geral são altos, o ponto crucial é descobrir o equilíbrio ideal entre essas duas medidas para um dado sistema.

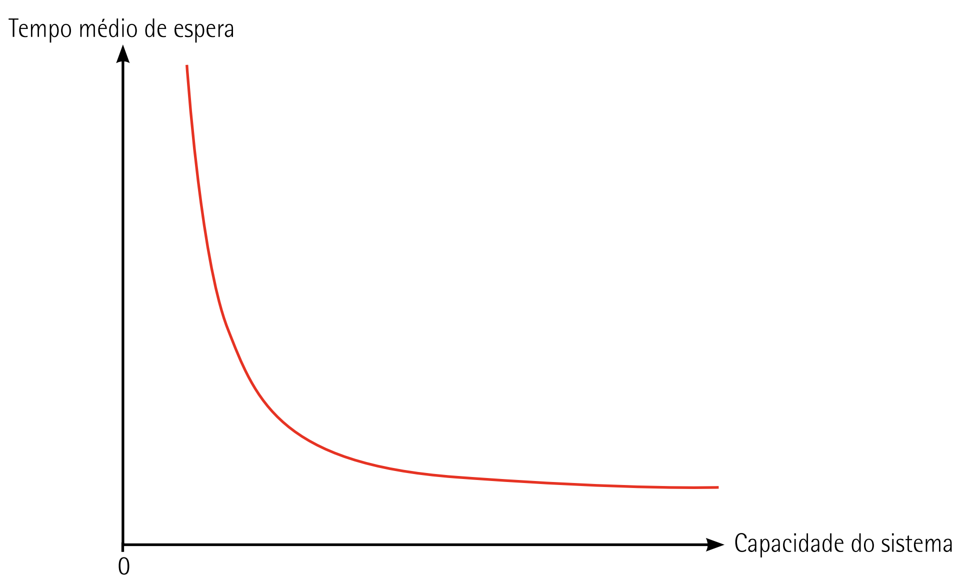

Figura 24

Os resultados da análise de sistemas de filas também podem ser usados em modelos de otimização – por exemplo, minimizando a soma dos custos de oferecer um nível de serviço no sistema e a soma dos custos dos atrasos ou perdas de usuários. A figura a seguir ilustra uma curva de custo total em função do nível de serviço do sistema – por exemplo, variando a capacidade do sistema. Note que, enquanto os custos operacionais de oferecer o serviço aumentam com o nível de serviço, os custos causados pelos atrasos sofridos pelos usuários diminuem.

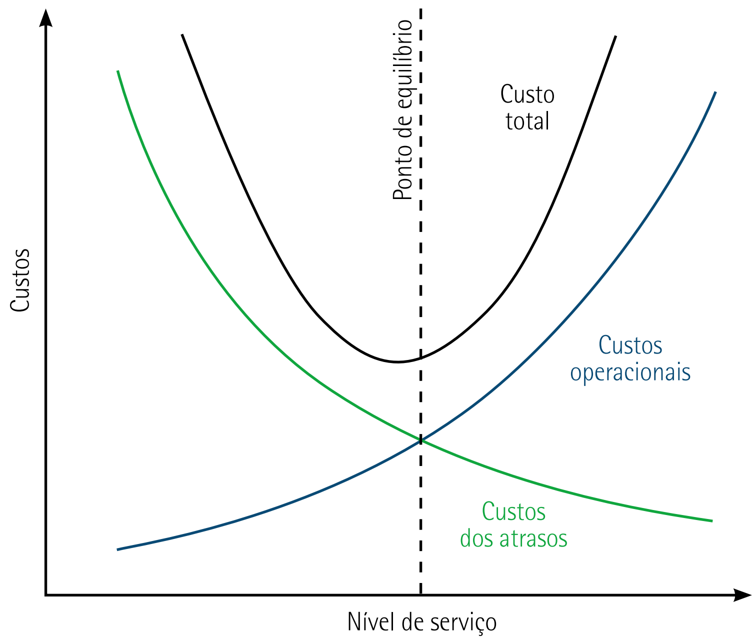

Figura 25

Uma dificuldade da análise desses modelos de minimização de custo total é obter uma boa estimativa dos custos unitários de espera dos usuários.

Como é a configuração de um bom sistema de filas? Essa questão depende de certos pontos de vista. Por exemplo, do ponto de vista dos usuários, certas medidas de desempenho de interesse são: o tempo médio de serviço, o tempo médio de espera em fila até receber o serviço, a probabilidade de o tempo de espera ser maior que um certo valor (digamos, 15 minutos), o número médio de usuários na fila, a probabilidade de o tamanho da fila ser maior que um certo valor (digamos, três usuários).

Do ponto de vista do gerente e dos operadores do sistema, outras medidas de interesse são: a utilização média do sistema, o número de servidores que ficam em média ocupados, o período médio em que os servidores permanecem continuamente ocupados atendendo usuários sem interrupção, a probabilidade de esse período ser maior que um certo valor (digamos, quatro horas). O foco da teoria de filas é a avaliação de medidas de desempenho do sistema em função de uma dada configuração. A análise de modelos de decisão em sistemas de filas depende de boas estimativas dessas medidas.

7.2 Definição e classificação de um sistema de filas

Na pesquisa operacional, a teoria das filas é um subcampo dedicado ao estudo de filas, ou, mais genericamente, ao estudo de modelos em que recursos são compartilhados de forma sequencial entre diversos usuários. O objetivo é analisar sistemas como esses para entender o seu comportamento e, assim, melhorar sua eficiência, aumentar sua capacidade, diminuir o tempo de espera e otimizar outros indicadores de desempenho.

Um sistema de filas é caracterizado por três componentes principais:

- **População de clientes**: define quem são os clientes e como eles chegam ao sistema. Podem ser finitos ou infinitos.
- **Mecanismo de serviço**: especifica como os clientes são atendidos – por exemplo, por um ou vários servidores –, e pode variar em termos de capacidade e disciplina (Fifo, Lifo, prioridades).
- **Política de filas**: determina a ordem pela qual os clientes serão atendidos na fila.

Esses sistemas podem ser classificados segundo a fonte de entrada, a disciplina da fila, o número de servidores, a capacidade do sistema e o tempo de chegada e serviço:

- **Segundo a fonte de entrada**:
  - **População infinita**: quando o número de possíveis clientes é muito grande, como num posto de gasolina em uma rodovia movimentada.
  - **População finita**: quando o número de possíveis clientes é limitado, como funcionários de uma empresa usando a mesma impressora.
- **Segundo a disciplina da fila**:
  - ***First in, first out* (Fifo)**: o primeiro que chega é o primeiro que sai.
  - ***Last in, first out* (Lifo)**: o último que chega é o primeiro que sai.
  - **Prioridade**: clientes com maior prioridade são atendidos primeiro.
- **Segundo o número de servidores**:
  - **Sistema de um único servidor**: apenas um servidor realiza todos os atendimentos.
  - **Sistema de múltiplos servidores**: mais de um servidor está disponível para atendimento.
- **Segundo a capacidade do sistema**:
  - **Capacidade ilimitada**: o número de clientes no sistema (na fila e em serviço) pode ser infinito.
  - **Capacidade limitada**: existe um limite para o número de clientes no sistema.
- **Segundo o tempo de chegada e serviço**:
  - **Determinístico**: o tempo entre chegadas e/ou o tempo de serviço são constantes.
  - **Estocástico**: o tempo entre chegadas e/ou o tempo de serviço seguem uma distribuição probabilística.

Cada sistema de filas pode ser uma combinação dessas categorias, e a análise precisa considerar todos esses fatores para modelar com precisão e eficiência o sistema em questão.

7.3 Processo de chegada e de serviço

Na teoria das filas em pesquisa operacional, o processo de chegada e o processo de serviço são duas partes fundamentais para entender e modelar um sistema de filas. Vamos conhecer as características e métricas de cada um a seguir.

Processo de chegada

- Características:
  - **Taxa de chegada (λ)**: define o número médio de clientes que chegam ao sistema por unidade de tempo. Pode ser constante ou variável.
  - **Distribuição de chegada**: a mais comum é a distribuição de Poisson, onde a probabilidade de n chegadas em um período t é dada pela seguinte fórmula:

$$P(n;t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

  - **Fonte de entrada**: pode ser uma população finita ou infinita. Em uma população infinita, a taxa de chegada não é afetada pelo número de clientes já no sistema.
  - **Disciplina de chegada**: a forma como os clientes chegam pode variar. Eles podem chegar individualmente ou em grupos (chamados lotes).
- Métricas:
  - **Tempo entre chegadas**: é o tempo que passa entre a chegada de um cliente e a chegada do próximo.

Processo de serviço

- Características:
  - **Taxa de serviço (μ)**: é o número médio de clientes que um servidor pode atender por unidade de tempo.
  - **Distribuição de serviço**: a distribuição exponencial é frequentemente usada para modelar o tempo de serviço, embora outras distribuições possam ser aplicadas, dependendo do contexto.

- **Número de servidores**: indica quantos atendentes estão disponíveis para servir os clientes. A configuração pode ser de servidor único ou de múltiplos servidores.

- **Disciplina de serviço**: similar à disciplina de chegada, indica a ordem em que os clientes são atendidos. Pode ser Fifo, Lifo ou baseada em prioridades.

- Métricas:

  - **Tempo de serviço**: é o tempo que leva para um servidor atender a um cliente.

  - **Utilização do sistema (ρ)**: é a fração de tempo em que o servidor está ocupado, geralmente calculada como $\rho = \frac{\lambda}{\mu}$.

Ambos os processos, de chegada e de serviço, são críticos para determinar métricas importantes do sistema de filas, como o tempo médio de espera na fila, o número médio de clientes no sistema, o tempo médio de um cliente no sistema, entre outros. A modelagem precisa desses processos é crucial para otimizar o sistema e melhorar a eficiência operacional.

7.4 Disciplina da fila

Este componente se refere à regra ou conjunto de regras que determinam a ordem segundo a qual os clientes são selecionados para o serviço a partir da fila. A disciplina da fila é um componente crítico no design e análise de sistemas de filas, pois pode influenciar significativamente seu desempenho e suas métricas.

As principais disciplinas da fila são:

- **Fifo *(first in, first out)* ou FCFS *(first come, first served)***: o cliente que entra na fila primeiro é atendido primeiro. É a disciplina mais comum, usada em situações onde a justiça ou igualdade é uma preocupação, como em caixas de supermercados ou filas de banco.

- **Lifo *(last in, first out)* ou LCFS *(last come, first served)***: o cliente que entra na fila mais recentemente é atendido primeiro. É menos comum que o Fifo, mas pode ser observado em algumas aplicações específicas, como em pilhas de armazenamento.

- **Siro *(service in random order)***: clientes são selecionados para atendimento em ordem totalmente aleatória. É raro em aplicações práticas, mas pode ser usado em simulações ou para modelagem teórica.

- **Atendimento com prioridade**: clientes são classificados com base em um certo nível de prioridade. O cliente com a maior prioridade é atendido primeiro. É comumente usado em situações de emergência ou em que alguns clientes têm um status ou necessidade mais urgente, como em um pronto-socorro.

- **Atendimento com reserva**: uma certa fração dos servidores é reservada para tipos específicos de clientes. Por exemplo, em um call center, alguns atendentes podem ser reservados para lidar com problemas técnicos, enquanto outros lidam com questões gerais.

- **Feedback**: após o atendimento, um cliente pode retornar à fila, possivelmente com uma prioridade diferente. Esta disciplina é utilizada em situações nas quais um cliente pode necessitar de múltiplos serviços ou revisões, como em sistemas computacionais em que tarefas podem retornar para filas após a primeira fase de processamento.

Ao modelar e analisar sistemas de filas, a disciplina da fila desempenha um papel importante, pois pode afetar várias métricas, como tempos de espera, número de clientes no sistema e tempo total no sistema. Além disso, a escolha da disciplina da fila pode ser influenciada por considerações práticas, como a natureza do sistema, as necessidades dos clientes e os objetivos do serviço.

7.5 Notação de Kendall

A notação de Kendall é amplamente utilizada em teoria das filas para descrever e classificar sistemas de filas de forma concisa. Esta notação é composta por três (ou às vezes mais) símbolos que são utilizados para sumarizar as características principais de um sistema de filas.

A notação básica de Kendall é A/B/C, onde:

- **A**: refere-se ao processo de chegada. Alguns dos símbolos mais comuns são:
  - **M**: chegadas com distribuição de Poisson (do inglês, *memoryless*).
  - **D**: chegadas determinísticas ou constantes.
  - **G**: chegadas gerais, onde não há um padrão específico.

- **B**: refere-se à distribuição do tempo de serviço. Assim como o processo de chegada, os símbolos mais comuns são:
  - **M**: tempo de serviço exponencialmente distribuído.
  - **D**: tempo de serviço determinístico ou constante.
  - **G**: tempo de serviço geral, sem uma distribuição específica.

- **C**: refere-se ao número de servidores no sistema.

Um exemplo clássico é o M/M/1.

Lembrete

O M/M/1 é um sistema de filas em que as chegadas são modeladas por um processo de Poisson, os tempos de serviço são exponencialmente distribuídos e há um único servidor.

Além da notação básica A/B/C, Kendall às vezes inclui mais símbolos para fornecer detalhes adicionais sobre a capacidade da fila, a disciplina da fila etc. Por exemplo, M/M/1/N/Fifo inclui detalhes de uma capacidade máxima de fila N e uma disciplina de fila Fifo.

A notação de Kendall é uma ferramenta valiosa porque permite aos pesquisadores e profissionais descrever rapidamente as características principais de um sistema de filas. Isso facilita a discussão, a pesquisa e a análise de diferentes sistemas e suas propriedades.

7.6 Sistema de filas e otimização

Os sistemas de filas e a otimização são duas áreas críticas na pesquisa operacional que se intersectam com frequência. Ambas visam melhorar o desempenho e a eficiência dos sistemas, embora a abordagem possa ser diferente.

O sistema de filas se preocupa com o estudo do tempo de espera, tanto na fila como dentro do sistema. No mundo real, os sistemas de filas são utilizados em várias indústrias, como manufatura, saúde, transporte e telecomunicações. O objetivo é analisar a estrutura de uma fila para melhorar a eficiência em termos de tempo de espera, alocação de recursos, capacidade e outros.

Já a otimização envolve a escolha da melhor solução de um conjunto de alternativas. Em pesquisa operacional, a otimização pode ser aplicada a uma ampla gama de problemas, como programação linear, problemas de transporte, problemas de designação, entre outros. O objetivo é maximizar ou minimizar uma função objetivo, sujeita a um conjunto de restrições.

Vejamos a seguir a relação entre sistemas de filas e otimização:

- **Definição da função objetivo**: em sistemas de filas, a função objetivo frequentemente se refere ao tempo de espera ou ao número de entidades no sistema. Na otimização, a função objetivo pode ser qualquer métrica que você deseja otimizar. Em muitos casos, você pode querer otimizar o desempenho de um sistema de filas.

- **Alocação de recursos**: em sistemas com múltiplos servidores ou canais, a pesquisa operacional pode ajudar a determinar a melhor alocação de recursos para minimizar o tempo de espera ou maximizar o throughput.

- **Análise de custos**: a otimização pode incorporar custos associados à operação de um sistema de filas, como custo de manutenção de servidores, e encontrar um equilíbrio entre custo e desempenho.

- **Simulação e modelagem**: técnicas de simulação são comuns em ambas as áreas. Em sistemas de filas, simulações podem ajudar a entender o comportamento do sistema sob diferentes condições. Em otimização, simulações podem ser usadas para testar diferentes estratégias de solução.

- **Decisões estratégicas**: ambas as áreas oferecem ferramentas para tomar decisões baseadas em dados. Por exemplo, um hospital pode usar teoria das filas para modelar o tempo de espera em uma sala de emergência e usar técnicas de otimização para decidir o número ótimo de médicos por turno.

- **Restrições**: tanto em otimização como em sistemas de filas, as restrições desempenham um papel crucial. Em filas, as restrições podem ser a capacidade da fila, o número de servidores etc. Em otimização, restrições podem ser orçamentárias, de tempo ou de recursos disponíveis.

- **Soluções de ponto a ponto**: às vezes, um problema de filas pode ser uma parte de um problema maior de otimização. Por exemplo, a otimização da rota de um veículo de entrega pode incluir considerações sobre o tempo de espera para o carregamento e descarregamento de mercadorias.

Dada esta relação intrínseca, é comum que profissionais de pesquisa operacional utilizem as ferramentas e métodos de ambas as áreas para criar sistemas mais eficientes e eficazes.

7.7 Medidas de desempenho de um sistema de filas

A teoria das filas, no contexto da pesquisa operacional, preocupa-se com a análise e modelagem de sistemas que envolvem o processo de espera. Medidas de desempenho são métricas que nos ajudam a quantificar a eficiência e a eficácia de um sistema de filas. Estas medidas são cruciais para entender o comportamento do sistema sob diversas condições e, assim, tomar decisões informadas sobre sua operação e otimização.

Aqui estão algumas das principais medidas de desempenho em sistemas de filas:

- **Taxa de chegada (λ)**: é a taxa média com que entidades (clientes, peças, chamadas etc.) chegam ao sistema. Em geral, assume-se que as chegadas seguem um processo de Poisson, embora isso possa variar de acordo com o sistema específico.

- **Taxa de serviço (μ)**: representa a taxa média com que os servidores atendem ou processam as entidades.

- **Número médio de entidades na fila (Lq)**: esta métrica reflete a quantidade média de entidades esperando na fila, sem contar aquelas já em serviço.

- **Tempo médio de espera na fila (Wq)**: é o tempo médio que uma entidade passa esperando na fila antes de ser atendida.

- **Número médio de entidades no sistema (L)**: esta é a quantidade total média de entidades no sistema, considerando tanto aquelas na fila quanto as em serviço.

- **Tempo médio no sistema (W)**: representa o tempo médio que uma entidade passa dentro do sistema, incluindo tanto o tempo de espera na fila quanto o tempo de atendimento.

- **Probabilidade de o servidor estar ocupado (ρ)**: reflete a fração do tempo em que o servidor (ou servidores) está ocupado atendendo entidades. É também conhecido como fator de utilização do sistema.

- **Probabilidade de um número específico de entidades no sistema (Pn)**: reflete a probabilidade de haver n entidades no sistema em um dado momento.

- **Probabilidade de o sistema estar vazio (P0)**: esta é a probabilidade de não haver entidades no sistema.

- **Tempo máximo de espera na fila**: embora menos comum do que as métricas médias, em alguns cenários é crucial conhecer o tempo máximo que uma entidade pode esperar na fila.

Estas medidas, em conjunto, proporcionam uma visão abrangente do desempenho de um sistema de filas. Elas são frequentemente utilizadas para comparar diferentes configurações de sistema, para avaliar o impacto das mudanças propostas ou para identificar gargalos e áreas de melhoria.

Ao avaliar essas métricas em diferentes cenários e combinações, os pesquisadores e profissionais podem otimizar o desempenho do sistema de filas, resultando em tempos de espera reduzidos, maior eficiência e melhor utilização dos recursos.

7.8 Fórmula de Little

A fórmula de Little é um resultado fundamental e surpreendentemente simples da teoria das filas. Ela desempenha um papel crucial na análise de sistemas de filas, relacionando três das suas métricas mais importantes: o número médio de entidades no sistema (L), a taxa média de chegada de entidades ao sistema (λ) e o tempo médio que uma entidade passa no sistema (W).

A fórmula de Little é assim expressa:

$$L = \lambda W$$

Onde:

L é o número médio de entidades (clientes, peças, chamadas etc.) no sistema (isso inclui tanto entidades em serviço quanto aquelas esperando na fila).

λ é a taxa média de chegada de entidades ao sistema.

W é o tempo médio que uma entidade passa dentro do sistema (incluindo tanto o tempo de espera quanto o tempo de serviço).

O que é notável na fórmula de Little é sua generalidade. Ela se aplica a praticamente qualquer sistema de filas, independentemente da natureza da distribuição de chegada ou da distribuição de serviço, desde que o sistema seja estável (ou seja, ele não explode com um número infinito de entidades com o tempo).

A fórmula de Little é incrivelmente útil porque, conhecendo duas das métricas indicadas, podemos facilmente calcular a terceira. Por exemplo, se soubermos a taxa de chegada e o tempo médio que as entidades passam no sistema, podemos calcular o número médio de entidades no sistema.

A equação de Little também nos fornece uma maneira de verificar a consistência dos resultados ao analisar sistemas de filas. Se, por algum motivo, as medidas obtidas ou calculadas para L, λ e W não satisfizerem a equação, isso pode indicar um erro nas medições ou na modelagem do sistema.

Exemplo de aplicação

Suponha um caixa de banco onde clientes chegam a uma taxa média de cinco clientes por hora ($\lambda = 5$) e passam em média 10 minutos (ou 1/6 de uma hora) no sistema, seja esperando na fila ou sendo atendidos ($W = 1/6$). Usando a fórmula de Little, podemos calcular o número médio de clientes no sistema (L) como:

$$L = \lambda W = 5 * \left(\frac{1}{6}\right) = \frac{5}{6}$$

Isso significa que, em média, haverá 5/6 (ou aproximadamente 0,83) clientes no sistema (no caixa ou esperando na fila).

A fórmula de Little fornece uma ferramenta robusta para a análise de sistemas de filas na pesquisa operacional e é uma das relações mais amplamente aplicadas e verificadas na teoria das filas.

7.9 A relação das distribuições exponencial e de Poisson

Na teoria das filas, as distribuições exponencial e de Poisson são frequentemente empregadas para modelar os processos de chegada e serviço em um sistema de filas. A relação entre essas duas distribuições é intrínseca e crucial para entender como as filas operam e para otimizar sistemas baseados em filas. Vamos examinar mais detalhadamente cada um desses aspectos.

A **distribuição de Poisson** é usada para modelar o número de chegadas de clientes (ou entidades) ao sistema em um intervalo de tempo fixo. Em um sistema M/M/1, por exemplo, as chegadas de clientes seguem um processo de Poisson, o que significa que a probabilidade de ter n chegadas em um intervalo de tempo t é dada pela fórmula a seguir:

$$P(n) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

Onde λ é a taxa média de chegada de clientes por unidade de tempo e e é a base do logaritmo natural.

A **distribuição exponencial**, por outro lado, é usada para modelar o tempo entre chegadas consecutivas (conhecido como tempo entre chegadas) e o tempo que um servidor leva para atender um cliente (conhecido como tempo de serviço).

O tempo entre chegadas (T) em um processo de Poisson é exponencialmente distribuído com uma taxa média λ, e sua função densidade de probabilidade é:

$$f(t) = \lambda e^{-\lambda t},\ t \geq 0$$

O mesmo se aplica ao tempo de serviço, mas com uma taxa média μ (em vez de λ), que representa o número médio de clientes que um servidor pode atender em uma unidade de tempo.

A relação essencial entre as distribuições de Poisson e exponencial é que em um processo de Poisson os tempos entre chegadas consecutivas são independentes e exponencialmente distribuídos. Ou seja, se as chegadas seguem um processo de Poisson com taxa λ, o tempo até a próxima chegada é exponencialmente distribuído com o mesmo λ.

Esta relação é fundamental na teoria das filas porque permite modelar e analisar sistemas complexos de forma matematicamente tratável. A escolha dessas distribuições simplifica o cálculo das medidas de desempenho do sistema, como o número médio de clientes na fila, o tempo médio de espera e outros.

Além disso, as propriedades da memória "sem memória" da distribuição exponencial fazem dela uma escolha conveniente para modelar sistemas em que o passado e o futuro são independentes – um cenário frequentemente encontrado em sistemas de filas do mundo real.

Em resumo, as distribuições de Poisson e exponencial fornecem uma estrutura robusta para modelar e analisar sistemas de filas, permitindo aos profissionais de pesquisa operacional e engenheiros otimizar o desempenho e a eficiência de diversos tipos de sistemas.

7.10 Modelo de nascimento e morte

O modelo de nascimento e morte é uma categoria especial de modelos na teoria das filas. É amplamente utilizado em pesquisa operacional para analisar sistemas onde a população muda ao longo do tempo devido a chegadas (nascimentos) e partidas (mortes). Este modelo é especialmente útil para sistemas que têm taxas de chegada e serviço dependentes do estado, o que significa que essas taxas podem variar dependendo do número atual de entidades (clientes, tarefas etc.) no sistema.

Na estrutura básica do modelo de nascimento e morte, consideramos um sistema com N estados, onde N representa o número de entidades no sistema. Cada estado n tem uma taxa de chegada λ_n e uma taxa de serviço (ou morte) μ_n .

- **Taxas de chegada (λ)**: indicam o número de novas entidades que entram no sistema por unidade de tempo. No estado n, a taxa de chegada é μ_n.

- **Taxas de serviço (ou morte) (μ)**: indicam o número de entidades que saem do sistema após receberem serviço por unidade de tempo. No estado n, a taxa de serviço é μ_n .

Para encontrar a probabilidade estável P_n de ter o sistema com n entidades, usam-se as **equações de equilíbrio**. Para um sistema em estado de equilíbrio, a taxa de fluxo para dentro e para fora de cada estado deve ser a mesma:

$$\lambda_n P_n = P_{n+1\ n+1}$$

Estas equações podem ser resolvidas para encontrar as probabilidades estáveis P_n, fundamentais para calcular métricas como o número médio de entidades no sistema, o tempo médio de espera, entre outros.

O modelo de nascimento e morte é usado em uma variedade de aplicações em pesquisa operacional, tais como:

- sistemas de atendimento ao cliente;

- redes de telecomunicações;

- sistemas de manufatura;

- logística e cadeias de suprimentos.

Exemplo de aplicação

Suponhamos um pequeno sistema de atendimento ao cliente com um único servidor. As chegadas de clientes seguem um processo de Poisson com taxa $\lambda = 4$ clientes por hora, e o tempo de serviço é exponencialmente distribuído com uma taxa média $\mu_n = 5$ clientes por hora.

Para tal sistema, o estado n teria uma taxa de chegada $\lambda = 4$ e uma taxa de serviço $\mu_n = 5$, independentemente do número de clientes no sistema (uma simplificação neste caso específico).

As equações de equilíbrio seriam usadas para encontrar as probabilidades estáveis e, a partir delas, calcular métricas como o número médio de clientes na fila, o tempo médio de espera na fila etc.

Em resumo, o modelo de nascimento e morte oferece uma estrutura poderosa para analisar e otimizar sistemas em pesquisa operacional, especialmente quando as taxas de chegada e serviço dependem do estado.

7.11 Modelos de fila com um e múltiplos servidores

Modelos de fila são fundamentais na pesquisa operacional para analisar sistemas que envolvem tempos de espera. Eles são comumente usados para analisar sistemas como linhas de produção, redes de computadores, atendimento ao cliente etc. A natureza do servidor – se é único ou múltiplo – desempenha um papel crucial na determinação das características de um sistema de fila, como tempo de espera, comprimento da fila e eficiência.

O modelo de fila com um servidor (M/M/1) é o mais básico e amplamente estudado na teoria das filas. Nele, temos as seguintes características:

- **Chegada**: taxa média de chegada de clientes na fila. As chegadas seguem um processo de Poisson com taxa λ.

- **Serviço**: capacidade média do servidor de atender clientes por unidade de tempo. O tempo de serviço é exponencialmente distribuído com taxa média μ.

- **Servidor**: entidade que atende os clientes. No modelo de fila com um servidor, há, como o nome indica, um único servidor.

- **Clientes**: entidades que requerem serviço.

As principais medidas de desempenho do modelo de fila com um servidor são:

- Número médio de clientes no sistema (L):

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

- **Tempo médio que um cliente passa no sistema (W)**:

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

- **Probabilidade de que o servidor esteja ocupado (ou de utilização) (ρ)**:

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

Além dessas métricas, é importante ainda determinar:

- **Número médio de clientes na fila (L_q)**:

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)}$$

- **Tempo médio que um cliente passa na fila (W_q .)**:

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)}$$

Já no modelo de fila com múltiplos servidores (M/M/c), temos:

- **Chegada**: as chegas seguem um processo de Poisson com taxa λ.
- **Serviço**: o tempo de serviço é exponencialmente distribuído com taxa média μ.
- **Servidores**: existem c servidores idênticos e cada um serve um cliente por vez.

Entre as principais medidas de desempenho do modelo de fila com múltiplos servidores, a utilização ρ é dada por:

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

O cálculo do número médio de clientes no sistema e do tempo médio que um cliente passa no sistema é mais complexo neste cenário, envolvendo fórmulas que consideram o número de servidores e a taxa de utilização.

Exemplo de aplicação

Vamos considerar um consultório médico.

Imagine um médico em um consultório atendendo pacientes. Os pacientes chegam a uma taxa média de três por hora e o médico leva, em média, 15 minutos para atender cada paciente (taxa de serviço de quatro por hora). Este é um exemplo de modelo M/M/1, isto é, **modelo de fila com um servidor**.

Agora, imagine que, devido ao aumento da demanda, o consultório decidiu empregar dois médicos. Os pacientes continuam chegando à mesma taxa, mas agora existem dois servidores (médicos) disponíveis para atendimento. Se cada médico ainda leva, em média15 minutos para atender um paciente, este é um exemplo do modelo M/M/2, ou sa, **modelo de fila com múltiplos servidores**.

Estes modelos, embora simplificados, são fundamentais para a teoria das filas, pois oferecem insights sobre como diferentes sistemas respondem às variações na demanda, na capacidade de serviço e no número de servidores. Eles fornecem as ferramentas básicas para projetar e otimizar sistemas de filas em diversas aplicações, desde consultórios médicos até grandes centrais de atendimento.

Exemplo de aplicação

Suponha um sistema de atendimento ao cliente com um servidor onde clientes chegam à taxa de dois clientes por hora ($\lambda = 2$) e o servidor atende à taxa de três clientes por hora ($\mu = 3$).

- **Utilização do servidor**:

$$\rho = \frac{2}{3}$$

- **Número médio de clientes no sistema**:

$$L = \frac{2}{3-2} = 2 \text{ clientes}$$

- **Tempo médio que umliente passa no sistema**:

$$W = \frac{1}{3-2} = 1 \text{ hora}$$

Lembrete

O modelo de fila com um servidor é comumente representado pela notação de Kendall como M/M/1, em que o primeiro M representa o processo de chegada (markoviano ou de Poisson), o segundo M representa o processo de serviço (markoviano ou exponencial) e o 1 representa o número de servidores.

No modelo M/M/1, a taxa média de chegadas é λ, a taxa média de serviço é μ, o sistema tem um único servidor e o tempo de chegada entre os clientes e o tempo de serviço têm distribuição exponencial.

7.12 Capacidade de servidor

Na teoria de filas, a capacidade de um servidor frequentemente se refere à taxa de serviço desse servidor (denotada pelo símbolo μ), ou seja, o número médio de clientes que é capaz de atenderor unidade de tempo, que pode ser em hora, minuto, segundo etc.

Vejamos alguns detalhes importantes sobre a capacidade do servidor:

- **Taxa de serviço (μ)**: é o inverso do tempo médio de serviço. Se um servidor leva em média 5 minutos para atender um cliente, a taxa de serviço é $(\frac{1}{5})$ clientes por minuto. Essa taxa determina quanto tempo, em média, um cliente passa no sistema e na fila, impactando diretamente as métricas de desempenho do sistema de filas. No contexto de um sistema M/M/1, por exemplo, se $(\mu > \lambda)$, o sistema é estável, mas se $(\mu < \lambda)$, o sistema é instável, e a fila crescerá indefinidamente.

- **Capacidade física**: em alguns contextos, a capacidade do servidor pode se referir à quantidade máxima de clientes que o servidor pode acomodar. Isso é especialmente relevante em modelos de fila com capacidade limitada, em que o número de clientes no sistema não pode exceder um valor específico.

- **Número de servidores**: a capacidade total de um sistema de filas também é influenciada pelo número de servidores no sistema. Um sistema com mais servidores pode atender mais clientes por unidade de tempo, diminuindo os tempode espera.

Exemplo de aplicação

Se temos um servidor com uma taxa de serviço de $(\mu = 4)$ clientes por hora, isso significa que, em média, o servidor pode atender quatro clientes a cada hora, ou um cliente a cada 15 minutos.

Compreender a capacidade de um servidor e como ela interage com a taxa de chegada de clientes é crucial para analisar e otimizar sistemas de filas. Isso permite, por exemplo, adequar o número de servidores às necessidades do sistema, minimizar tempos de espera e otimizar recursos.

Exemplo de aplicação

Imagine uma empresa que desenvolveu uma plataforma de e-commerce e oferece suporte técnico para seus clientes (outras empresas que vendem produtos on-line). Sua equipe de suporte tem dois principais níveis:

- **Nível 1**: lida com problemas comuns e dúvidas simples, geralmente resolvidos rapidamente.
- **Nível 2**: lida com problemas mais complexos, que exigem investigação e possivelmente alguma programação.

Descrição do problema

Cada cliente que entra em contato com o suporte é inicialmente atendido por um técnico de nível 1. Se o técnico de nível 1 não consegue resolver o problema, o caso é elevado para o nível 2.

Vamos assumir que:

- Os técnicos de nível 1 atendem em média 20 clientes por hora ($\mu_1 = 20$).
- Os técnicos de nível 2 atendem em média 5 clientes por hora ($\mu_2 = 5$).
- Chegam em média 10 clientes por hora ao sistema ($\lambda = 10$).

Se o técnico de nível 1 não consegue resolver o problema, o cliente é enviado para a fila de nível 2. Queremos calcular:

- A taxa de utilização dos técnicos nos dois níveis.
- O número médio de clientes nas filas.
- O tempo médio de um cliente no sistema.

Solução

1) Taxa de utilização (ρ) dos técnicos

A taxa de utilização (ρ) é calculada como a razão entre a taxa de chegada e a taxa de serviço. Vamos calcular isso para ambos os níveis.

$\rho = \dfrac{\lambda}{\mu}$ para nível 1.

$\rho_2 = \dfrac{\lambda}{\mu}$ para nível 2.

Substituindo os valores fornecidos:

$\rho_1 = \dfrac{10}{20} = 0{,}50$ (50% de utilização no nível 1).

$\rho_2 = \dfrac{10}{5} = 2$ (200% de utilização no nível 2, o que não é possível e indica um gargalo).

2) Número médio de clientes na fila (Lq)

Para sistemas M/M/1, a fórmula de Little $L = \lambda W$ pode ser expandida para obter o número médio de clientes na fila (L_q) com base na taxa de utilização (ρ).

$$L_q = \frac{\rho^2}{1-\rho}$$

- **Nível 1**:

$$L_{q1} = \frac{\rho^2}{1-\rho} = \frac{0{,}5^2}{1-0{,}5} = 0{,}50$$

- **Nível 2**:

$$L_{q2} = \frac{\rho^2}{1-\rho} = \frac{2^2}{1-2} = -4$$

Note que essa fórmula não é aplicável para (ρ2) porque (ρ2 > 1), o que significa que o sistema está sobrecarregado. Isso indica que, na prática, os clientes estão esperando indefinidamente, ou sendo rejeitados/desistindo.

3) Tempo médio de um cliente no sistema

O tempo médio de espera na fila (W_q) para sistemas M/M/1 é dado por:

$$W_q = \frac{L_q}{\lambda}$$

- **Nível 1**:

$$W_{q1} = \frac{L_{q1}}{\lambda} = \frac{0,5}{10} = 0,05 \text{ hora/cliente ou 3 minutos/cliente}$$

- **Nível 2**:

W_{q2} não pode ser calculado devido à superutilização.

O tempo médio que um cliente passa no sistema é a soma do tempo de espera na fila e o tempo de serviço:

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

- **Nível 1**:

$$W_1 = W_{q1} + \frac{1}{\mu 1} = 0,05 + 1/20 = 0,1 \text{ cliente/hora ou 6 minutos no sistema.}$$

- **Nível 2**:

Não é possível calcular W_2 devido à superutilização.

Os cálculos revelam que o sistema de suporte técnico está funcionando bem no nível 1, mas tem um grave problema no nível 2, com uma taxa de utilização de 200%. Isso necessitaria de uma investigação e reestruturação adicional, possivelmente contratando mais técnicos para o nível 2 ou aprimorando os existentes para resolver problemas mais rapidamente.

8 SIMULAÇÃO

A teoria da simulação é um ramo da pesquisa operacional que envolve a modelagem de sistemas complexos para prever seu comportamento em diferentes cenários. Em vez de solucionar problemas matematicamente, a simulação permite que os analistas configurem um modelo computacional do sistema em questão e realizem experimentos para observar o que acontece.

Saiba mais

Assim como muitas das técnicas em pesquisa operacional (Hillier *et al.*, 2012), a simulação tem suas raízes em aplicações militares. Foi extensivamente utilizada durante a Segunda Guerra Mundial para simular batalhas aéreas e terrestres. Consulte a obra a seguir para se aprofundar mais nos estudos da simulação:

SILVA, E. M. *et al. Pesquisa operacional*: programação linear, simulação. 5. ed. Rio de Janeiro: Atlas, 2017.

Vejamos a seguir os fundamentos da simulação:

- **Entidades**: componentes individuais que interagem dentro do sistema. Por exemplo, em uma simulação de tráfego, carros, semáforos e pedestres seriam as entidades.

- **Atributos**: características que definem as propriedades das entidades. No contexto de uma simulação de hospital, atributos podem incluir o tipo de tratamento necessário para cada paciente, o tempo de espera etc.

- **Eventos**: ações que ocorrem em pontos específicos no tempo e que alteram o estado das entidades e do sistema como um todo.

- **Horizonte de tempo**: duração total do estudo de simulação. Pode ser finito ou infinito.

- **Estado do sistema**: representação das variáveis de interesse em determinado momento.

- **Relógio de simulação**: variável que avança de forma incremental ou em saltos para representar o passar do tempo no sistema simulado.

Existem vários tipos de simulação, aplicáveis a diversos cenários:

- **Simulação de eventos discretos (DES)**: foca em sistemas nos quais as mudanças ocorrem em pontos específicos no tempo. É comum em simulações de fábricas e sistemas de transporte.

- **Simulação de Monte Carlo**: utiliza o acaso e variáveis aleatórias para modelar eventos incertos. É bastante utilizada em finanças e engenharia de risco.

- **Simulação contínua**: aplica-se a sistemas nos quais as mudanças ocorrem continuamente ao longo do tempo, como em simulações climáticas.

Uma simulação depende de diversas etapas. Na etapa **definição do problema**, deve-se entender o que se quer alcançar com a simulação; na etapa **construção do modelo**, deve-se representar o sistema de forma abstrata, normalmente com o auxílio de software de simulação; na **coleta de dados**, busca-se obter dados para as entradas do modelo; na etapa de **execução**, deve-se rodar o modelo e coletar os resultados; na **análise**, interpretam-se os resultados, frequentemente usando técnicas estatísticas; na etapa de **validação e verificação** se garante que o modelo é uma representação precisa do sistema real e que os resultados são confiáveis; e na etapa de **documentação** se criam registros detalhados do modelo, dos dados e das conclusões para futura referência ou para uso por outros pesquisadores.

As aplicações da simulação na pesquisa operacional abrangem a otimização de processos industriais; a análise de sistemas de saúde; o estudo de cadeias de suprimentos; a modelagem de redes de computadores; e o planejamento de operações logísticas.

Entre suas vantagens estão a flexibilidade para modelar sistemas complexos e a capacidade de testar uma grande variedade de cenários; contudo, sua dependência da qualidade dos dados de entrada e o possivelmente alto custo computacional e de tempo podem limitá-la.

Em suma, a simulação é uma ferramenta poderosa em pesquisa operacional para analisar sistemas complexos em que soluções analíticas são difíceis ou impossíveis de se encontrar. Ela permite aos decisores entender melhor o comportamento do sistema sob diferentes condições e tomar decisões mais informadas.

8.1 O método Monte Carlo

O método Monte Carlo é uma técnica de simulação estatística usada para modelar a probabilidade de diferentes resultados em um processo que não pode ser facilmente previsto devido à intervenção de variáveis aleatórias. Embora originalmente desenvolvido para problemas na física estatística, ele pode ser útil em uma variedade de campos, como pesquisa operacional, finanças, engenharia e muitas outras ciências aplicadas.

Observação

Técnica de simulação numérica amplamente utilizada, esse método foi nomeado em homenagem ao cassino de Monte Carlo porque usa números aleatórios e probabilidades, de modo semelhante aos jogos de azar (Arenales; Armentano, 2006; Prado, 2017).

O método envolve a geração de números aleatórios para simular a incerteza de parâmetros e variáveis de entrada e, em seguida, calcular os resultados correspondentes para formar uma distribuição de possíveis resultados.

Seus principais componentes são:

- **Variáveis de entrada**: as variáveis que você não pode controlar no sistema, como demanda do cliente, tempo de falha de máquinas etc.
- **Modelo estocástico**: um modelo matemático ou computacional que incorpora as variáveis de entrada.
- **Variáveis de saída**: as métricas que você está tentando entender ou otimizar, como custo, lucro, tempo de espera etc.

As etapas básicas do método Monte Carlo são:

- **Definição do modelo**: em que se identificam as variáveis aleatórias e os parâmetros do sistema.
- **Geração de entradas aleatórias**: em que se utilizam geradores de números aleatórios para criar uma série de entradas com base em distribuições de probabilidade conhecidas.
- **Simulação**: em que se usa o modelo estocástico para executar um experimento usando as entradas aleatórias.
- **Coleta de dados**: em que os resultados são coletados e analisados.
- **Análise estatística**: em que se realizam análises para entender a distribuição dos resultados e outras estatísticas descritivas.
- **Validação**: em que se valida o modelo para garantir que ele é uma representação precisa do sistema real.

Em pesquisa operacional, o método Monte Carlo pode ser aplicado na otimização de cadeia de suprimentos, para avaliar diferentes cenários de estoque e demanda; na gestão de riscos, para avaliar riscos financeiros em projetos de grande escala; no planejamento de recursos, para alocar recursos de modo a maximizar o lucro ou minimizar o custo sob incerteza; e na teoria das filas, para estudar o comportamento de sistemas de espera, como call centers, hospitais etc.

Suas vantagens envolvem sua flexibilidade para modelar ambientes complexos e incertos e a possibilidade de usá-lo para avaliar uma ampla gama de cenários. No entanto, o método Monte Carlo requer alto poder de computação para grandes simulações, e os resultados são apenas tão bons quanto os dados e o modelo, o que significa que dados ruins podem levar a conclusões imprecisas.

O método Monte Carlo foi criado para lidar com as variáveis desconhecidas e a complexidade nos sistemas, e é amplamente usado em pesquisa operacional para tomar decisões mais informadas sob condições de incerteza. Sua maior força reside na sua capacidade de transformar problemas complexos e estocásticos em experimentos simulados que podem ser analisados estatisticamente.

8.2 Casos interessantes de simulação

A simulação é uma ferramenta que permite aos tomadores de decisão observar os possíveis resultados de suas escolhas em um ambiente controlado, antes de implementar essas decisões no mundo real. Na pesquisa operacional, a simulação tem sido aplicada a uma variedade de problemas complexos e em diversas áreas. Eis alguns exemplos interessantes:

- Em aeroportos, é possível simular fluxos de passageiros e bagagens para projetar terminais mais eficientes, otimizar a alocação de portões e reduzir tempos de espera.
- Nos hospitais, pode-se simular o fluxo de pacientes para reduzir tempos de espera, otimizar a ocupação de leitos e melhorar a eficiência das salas de cirurgia.
- Na manufatura, simula-se a linha de produção para identificar gargalos, otimizar a sequência de produção e avaliar o impacto de falhas de equipamentos.
- Em logística e cadeia de suprimentos, é possível simular redes de distribuição para otimizar rotas de transporte, gerenciar estoques e avaliar estratégias sob diferentes cenários, como greves ou interrupções devido a desastres naturais.
- Nos sistemas financeiros, a simulação do mercado pode testar algoritmos de negociação, avaliar riscos e testar estratégias de investimento sob diferentes cenários econômicos.
- No planejamento urbano, pode-se simular fluxos de tráfego para otimizar a sincronização de semáforos, planejar novas vias ou avaliar os impactos de grandes eventos ou construções.
- Nas telecomunicações, é possível simular redes para avaliar sua capacidade, otimizar o posicionamento de torres e antenas e prever o desempenho sob diferentes volumes de tráfego.
- Na área militar, a simulação de cenários de combate permite treinar comandantes, testar estratégias e avaliar a eficácia de novos equipamentos ou táticas.
- Nos sistemas de energia, pode-se simular a geração e distribuição de energia para otimizar a produção, prever picos de demanda e avaliar a introdução de fontes renováveis.
- Na defesa civil, é possível simular evacuações em casos de desastres como furacões ou inundações, planejando rotas eficientes e garantindo que as pessoas possam se mover para locais seguros a tempo.

Estes são apenas alguns exemplos, mas a simulação, como ferramenta de pesquisa operacional, pode ser aplicada a quase qualquer sistema ou processo, tornando-se inestimável para a tomada de decisões informadas e a otimização de resultados.

Saiba mais

Para os amantes de ficção científica, o livro *A hipótese da simulação*, de Rizwan Virk (2021), pode fornecer uma panorama amplo sobre temas relacionados a simulação, como realidade virtual, realidade aumentada, computação quântica e inteligência artificial, que são explorados de maneira instigante.

VIRK, R. *A hipótese da simulação*: conheça a teoria por trás da série Matrix. Alvorada: Citadel, 2021.

Exemplo de aplicação

Você é gerente de um pequeno supermercado e está considerando alterar a configuração dos caixas. Atualmente, você tem três caixas, e os clientes escolhem aleatoriamente para qual fila ir. Você está considerando mudar para um único caixa com uma única fila. Use simulação para determinar qual configuração resulta em tempos de espera mais curtos para os clientes.

Solução

O problema propõe uma simulação para avaliar duas configurações de caixa em um supermercado:

- **Configuração 1**: três caixas com três filas separadas, em que os clientes escolhem uma fila para entrar.

- **Configuração 2**: um caixa com uma única fila, de onde os clientes são servidos.

O objetivo é determinar qual configuração resulta em tempos de espera mais curtos para os clientes.

Vejamos o passo a passo da simulação:

- **Passo 1**: definição das variáveis.

  – Taxa de chegada dos clientes.

  – Taxa de atendimento dos caixas.

- **Passo 2**: construção do modelo de simulação.

  – **Configuração 1**: simule a chegada de clientes em cada uma das três filas, e como os caixas os atendem.

  – **Configuração 2**: simule a chegada dos clientes na única fila e como são atendidos pelo caixa único.

- **Passo 3**: execução da simulação.
  - Determine um período de simulação (por exemplo, uma hora, um dia etc.).
  - Execute a simulação para ambas as configurações.
- **Passo 4**: análise dos resultados.
  - Calcule o tempo médio de espera na fila, o tempo médio de serviço e a taxa de utilização do caixa para cada configuração.

Execução e resultados simplificados

Valores são necessários para que se possa realizar uma simulação. Sendo assim, vamos detalhar um pouco os cálculos para ambos os cenários propostos no problema. Note que estamos utilizando médias e assumindo um fluxo constante; na prática, a simulação real permitiria variabilidade em todos esses parâmetros.

- Fórmulas úteis:
  - Tempo médio de espera na fila (W_q): $\frac{1}{\mu - \lambda}$
  - Taxa de utilização do caixa (U): $\frac{\lambda}{\mu}$

Configuração 1: três caixas com filas separadas

Taxa de chegada (λ): 10 clientes/hora em cada fila.

Taxa de serviço (μ): 12 clientes/hora por caixa.

Cálculos:

$$W_{q1} = \frac{1}{12 \text{ clientes / hora} - 10 \text{ clientes / hora}} = 0{,}5 \text{ hora/cliente}$$

$$U_1 = \frac{10 \text{ clientes / hora}}{12 \text{ clientes / hora}} = 0{,}833 = 83{,}3\%$$

Configuração 2: um caixa com uma única fila

Taxa de chegada (λ): 30 clientes/hora em uma fila (somando as três filas).

Taxa de serviço (μ): 12 clientes/hora por caixa.

Cálculos:

$$W_{q2} = \frac{1}{12\text{ clientes}/\text{hora} - 30\text{ clientes}/\text{hora}} = -\ 0{,}05\text{ hora/cliente}$$

$$U_2 = \frac{30\text{ clientes}/\text{hora}}{12\text{ clientes}/\text{hora}} = 2{,}5 = 250\%$$

Análise e conclusões

Configuração 1: cada caixa tem um tempo médio de espera na fila de 6 minutos e uma taxa de utilização de 83,3%.

Configuração 2: a análise direta sugere um tempo médio de espera negativo e uma utilização superior a 100%, o que na prática é impossível e indica que essa configuração não é viável com um único caixa atendendo.

Nota: a configuração 2 apresentou uma taxa de utilização maior que 100%, o que indica que um único caixa não será suficiente para atender a uma taxa de chegada de 30 clientes por hora. Portanto, para tal cenário seriam necessários mais caixas para atender a essa demanda.

Observação

Estes são cálculos simplificados, e uma simulação real teria mais variações e detalhes a serem considerados, como a variabilidade nas taxas de chegada e serviço. A análise também pode ser aprofundada ao considerar outras métricas, como o número médio de clientes na fila ou o tempo total no sistema.

Resumo

Nesta unidade foram vistos alguns problemas da pesquisa operacional mais específicos que se valem dos métodos apresentados na unidade I. Todos eles foram abordados de forma bastante simplificada, pois este livro-texto tem o objetivo de ser um texto introdutório, porém abrangente.

Aprendemos sobre otimização de rede, cujos métodos são utilizados em muitas áreas da indústria e geralmente visam escolher um caminho mais curto, mais barato ou mais rápido. Também discutirmos a teoria de decisão, que envolve uma grande gama de problemas, como a escolha para compra de equipamentos ou que investimento seria mais lucrativo ou de menor risco.

Abordamos igualmente a teoria das filas, que serve de suporte para diversos ramos da indústria, da troca de mensagens em um sistema computacional a filas em um supermercado, e está diretamente envolvida com a teoria da simulação, usada para testar valores e cenários para orientar a tomada de decisão.

Exercícios

Questão 1. A pesquisa operacional abrange uma série de técnicas matemáticas e computacionais utilizadas para otimizar a tomada de decisões em diversos cenários. Entre essas técnicas, os algoritmos de otimização de redes se concentram no design, no planejamento e na operação de redes de tal forma que se possa alcançar determinado objetivo, como maximizar a eficiência ou minimizar os custos, sob certas restrições. A otimização de redes lida, principalmente, com problemas que podem ser representados por uma estrutura de rede, composta por nós e arcos.

Nesse contexto, avalie as asserções a seguir e a relação proposta entre elas.

I – Problemas de fluxo máximo ou de circulação podem ser otimizados por meio de algoritmos de otimização de redes.

porque

II – O algoritmo de Dijkstra pode ser utilizado para resolver problemas de caminho mais curto.

Assinale a alternativa correta:

A) As asserções I e II são verdadeiras, e a asserção II justifica a I.

B) As asserções I e II são verdadeiras, mas a asserção II não justifica a I.

C) A asserção I é verdadeira, e a II é falsa.

D) A asserção I é falsa, e a II é verdadeira.

E) As asserções I e II são falsas.

Resposta correta: alternativa B.

Análise das asserções

I – Asserção verdadeira.

Justificativa: os algoritmos de otimização de redes são usados para otimizar problemas como o caminho mais curto entre dois nós de uma rede, o fluxo máximo de um nó de origem para um nó de destino, transporte, designação e circulação, entre outros.

II – Asserção verdadeira.

Justificativa: o algoritmo de Dijkstra é um famoso algoritmo utilizado para resolver o problema do caminho mais curto em um grafo dirigido ou não dirigido com arestas de peso não negativo. Ele encontra o caminho mais curto de um vértice inicial para todos os outros vértices no grafo.

Relação entre as asserções: as asserções I e II são verdadeiras, mas a asserção I fala de possíveis aplicações para algoritmos de otimização de redes, enquanto a asserção II fala do objetivo do algoritmo de Dijkstra. Logo, não há uma relação de causa e consequência entre elas.

Questão 2. A teoria da decisão é um campo interdisciplinar que oferece ferramentas para tomar decisões mais eficazes sob diferentes tipos de incertezas e restrições.

A respeito dos conceitos-chave da teoria da decisão, avalie as afirmativas a seguir.

I – Os tomadores de decisão são os indivíduos ou entidades responsáveis por fazer uma escolha entre diversas possibilidades.

II – Os critérios de decisão correspondem a uma representação tabular das recompensas associadas às várias combinações de alternativas e de estados da natureza.

III – As estratégias são as situações em que as probabilidades dos diversos estados da natureza são conhecidas.

É correto o que se afirma em:

A) I, apenas.

B) III, apenas.

C) I e II, apenas.

D) II e III, apenas.

E) I, II e III.

Resposta correta: alternativa A.

Análise das afirmativas

I – Afirmativa correta.

Justificativa: entre os conceitos-chave da teoria da decisão, os tomadores de decisão são os indivíduos ou as entidades que têm a responsabilidade de decidir entre várias alternativas disponíveis.

II – Afirmativa incorreta.

Justificativa: os critérios de decisão correspondem às métricas utilizadas para avaliar as diferentes alternativas disponíveis. Esses critérios podem ser quantitativos ou qualitativos.

III – Afirmativa incorreta.

Justificativa: as estratégias correspondem aos planos de ação que especificam qual alternativa seguir para cada possível estado da natureza.

REFERÊNCIAS

Textuais

ARENALES, M.; ARMENTANO, V. *Pesquisa operacional*. Rio de Janeiro: Elsevier, 2006.

BRONSON, R. *Pesquisa operacional*. São Paulo: McGraw-Hill, 1985.

COLIN, E. C. *Pesquisa operacional*: 170 aplicações em estratégia, finanças, logística, produção, marketing e vendas. 2. ed. Rio de Janeiro: Atlas, 2017.

DUARTE JÚNIOR, A. M. *Programação linear*. Curitiba: Appris, 2023.

FOGLIATTI, M. C.; MATTOS, N. M. C. *Teoria de filas*. Rio de Janeiro: Interciência, 2006.

GOLDBARG, E.; GOLDBARG, M. *Grafos*. Rio de Janeiro: GEN, 2021.

GOLDBARG, M. *Programação linear e fluxos em redes*. Rio de Janeiro: GEN, 2014.

HILLIER, F. S. *et al. Introdução à ciência da gestão*: modelagem e estudos de caso com planilhas eletrônicas. 4. ed. Porto Alegre: AMGH, 2014.

HILLIER, F. S. *et al. Introdução à pesquisa operacional*. 9. ed. Porto Alegre: AMGH, 2012.

MARTIN, D.; BIRBECK, M.; KAY, M. *Professional XML*. Rio de Janeiro: Ciência Moderna, 2001.

MOREIRA, D. *Pesquisa operacional*: curso introdutório. 2. ed. São Paulo: Cengage Learning, 2010.

NASCIMENTO, S. V. *Pesquisa operacional e análise de investimentos*: suas aplicações na indústria e nos serviços: com utilização do software Lindo. Rio de Janeiro: Ciência Moderna, 2012.

PRADO, D. *Teoria das filas e da simulação*. 6. ed. Belo Horizonte: Falconi, 2017. v. 2.

PROBLEMA. *In: Dicionário Houaiss da língua portuguesa*. Rio de Janeiro: Objetiva, 2009.

SILVA, A. M. *Pesquisa operacional aplicada à logística*. Rio de Janeiro: Alta Books, 2023.

SILVA, E. M. *et al. Pesquisa operacional*: programação linear, simulação. 5. ed. Rio de Janeiro: Atlas, 2017.

SITUAÇÃO. *In: Dicionário Houaiss da língua portuguesa*. Rio de Janeiro: Objetiva, 2009.

SMITH, B. *JSON básico*: conheça o formato de dados preferido da web. São Paulo: Novatec, 2015.

VIRK, R. *A hipótese da simulação*: conheça a teoria por trás da série Matrix. Alvorada: Citadel, 2021.

UNIP
UNIVERSIDADE PAULISTA

Informações:
www.sepi.unip.br ou 0800 010 9000