Modelação e Simulação de Processos

Tornar o Modelo Probabilístico

Paulo Matos

1

Resumo

- Probabilidades e Estatística
  - Breve Revisão de Conceitos
- Tornar o modelo probabilístico
  - Usar os dados recolhidos directamente na simulação
  - Definir uma distribuição empírica com base nos dados recolhidos
  - Utilizar técnicas de inferência estatística para ajustar os dados a uma função de distribuição de probabilidades

2

Probabilidades e Estatística
Revisões

- Experimentar é um processo cujos resultados não são conhecidos com exactidão…
- Experiências aleatórias
  - Acontecimento: colecção de resultados de uma experiência
  - Acontecimento elementar: um resultado que não pode ser simplificado ou reduzido
  - Espaço de resultados: constituído por todos os acontecimentos elementares
    - S={Cara, Coroa}, ao atirar uma moeda
    - S={1,2,3,4,5,6}, ao lançar um dado

3

Probabilidades e Estatística
Revisões

- Variável aleatória
  - Uma variável aleatória é uma variável que toma um certo valor numérico, determinado pelo acaso, de cada vez que a experiência é realizada;
  - A variável aleatória associa números aos acontecimentos do espaço dos possíveis resultados;
  - Uma distribuição de probabilidade permite calcular a probabilidade correspondente a cada valor ou conjunto de valores da variável aleatória;
  - Uma variável aleatória discreta toma um nº finito ou infinito numerável de valores;
  - Uma variável aleatória contínua toma um nº infinito não numerável de valores, os quais podem ser associados com medidas numa escala contínua.

4

Probabilidades e Estatística
Revisões

- Probabilidade
  - P – denota a probabilidade
  - A,B e C – denotam acontecimentos específicos
  - P(A) – denota a probabilidade de ocorrer o acontecimento A
- Lei dos Grandes Números
  - Quando uma experiência é repetida um grande número de vezes, o valor da frequência relativa do acontecimento tende a aproximar –se do valor da verdadeira probabilidade
- Distribuições de Probabilidades
  - Descrevem o que provavelmente acontecerá em vez do que realmente aconteceu...

5

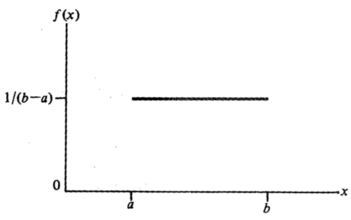
Probabilidades e Estatística
Revisões

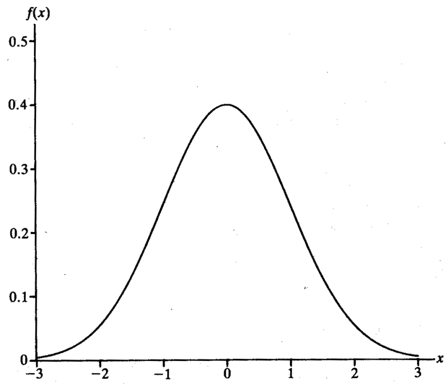
- Variáveis Aleatórias Contínuas
  - Função densidade de probabilidade
    - O gráfico f.d.p., ou curva da densidade, é um gráfico que traduz a distribuição de probabilidade de uma variável contínua
    - Todos os pontos sob a curva têm uma ordenada maior ou igual a zero;
    - A área total sob a curva tem de ser unitária;
    - As probabilidades obtêm-se a partir de áreas sob partes da curva.

6

Probabilidades e Estatística
Revisões

- Exemplos de curvas de densidade





7

Probabilidades e Estatística
Revisões

- Variáveis Contínuas - Função de Distribuição
  - F(x) da variável aleatória X, é definida para cada valor real x, como

    $F(x) = P(X <= x)$, para $-\infty < x < \infty$

    - P(X<=x) é a probabilidade associada ao evento {X<=x}
    - F(x) é a probabilidade de, ao realizar uma experiência, a variável aleatória X não ser superior a x
    - F(x) representa a probabilidade acumulada até x
  - F(x) tem as propriedades
    - 0 <= F(x) <= 1, para todo o x
    - F(x) é não decrescente (isto é, se $x_1 < x_2$, então $F(x_1) <= F(x_2)$)
    - $\lim_{x->-\infty} F(x) = 0$ e $\lim_{x->\infty} F(x) = 1$

8

Probabilidades e Estatística
Revisões

- Variáveis discretas
  - Variável aleatória toma um número finito de valores
  - A probabilidade de a variável ter um dos valores é dada por
    $p(x_i) = P(X = x_i)$ para i=1,2,3,…
    $$\sum_{i=1}^{\infty} p(x_i) = 1$$
  - Para I =[a,b], em que ***a*** e ***b*** são números reais tais que ***a*** <= x_i <= ***b,*** então
    $$P(X \in I) = \sum_{a \le x_i \le b} p(x_i)$$
  - A função de distribuição para a variável discreta X é
    $$F(x) = \sum_{x_i \le x} p(x_i), \quad \text{para} -\infty < \text{x} < \infty$$

9

Probabilidades e Estatística
Revisões

- Variáveis aleatórias
  - Valor médio ou valor esperado (μ_i ou $E(X_i)$) da variável aleatória X_i (em que i = 1,2, …, n)
    - É uma medida da tendência central

$$\mu_i = \begin{cases} \sum_{i=1}^{\infty} x_i p_{x_i}(x_i), \text{se } \mathrm{X_i} \text{ é discreta} \\ \int_{-\infty}^{\infty} x f_{x_i}(x) dx, \text{se } \mathrm{X_i} \text{ é contínua} \end{cases}$$

10

Probabilidades e Estatística
Revisões

- Variáveis aleatórias
  - Mediana
  - $x_{0,5}$ da variável X_i é definida como o mais pequeno valor de x para o qual $F_x(x) \geq 0,5$
  - Se X_i é uma variável aleatória contínua, então $F(x_{0,5}) = 0,5$
    - É uma medida da tendência central, preferível à média quando X_i pode tomar valores muito elevados ou muito baixos, pois os valores extremos podem afectar a média, ainda que sejam muito pouco prováveis de ocorrer

11

Probabilidades e Estatística
Revisões

- Variáveis aleatórias
  - Mediana

$$\hat{x}_{0,5}(n) = \begin{cases} X_{(n+1)/2} & \text{se } n \text{ ímpar} \\ \left[X_{(n/2)} + X((n/2)+1) \right] / 2 & \text{se } n \text{ par} \end{cases}$$

12

Probabilidades e Estatística
Revisões

- Variáveis aleatórias discretas – Exemplo
  - Considere uma variável aleatória discreta X que pode ter os valores 1, 2, 3, 4, 5 com probabilidades 0,2.
    - μ = 0,2*1 + 0,2*2 + 0,2*3 + 0,2*4 + 0,2*5 = 3
    - $X_{0,5} = 3$
  - Considere uma variável aleatória Y que pode tomar os valores 1, 2, 3, 4, 100, com probabilidades 0,2.
    - μ = 0,2*1 + 0,2*2 + 0,2*3 + 0,2*4 + 0,2*100 = 22
    - $X_{0,5} = 3$

13

Probabilidades e Estatística
Revisões

- Variáveis aleatórias
  - Variância (σ_i^2 ou Var(X_i)) da variável aleatória X_i (em que i = 1,2, …, n)

$$\sigma_i^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n-1}$$

    - Medida da dispersão de uma variável aleatória face ao seu valor médio
    - Quanto mais elevada for a variância, mais elevada será a probabilidade de a variável poder tomar valores afastados do seu valor médio
  - Desvio Padrão $\sigma_i = \sqrt{\sigma_i^2}$

14

Fontes de aleatoriedade

- Praticamente todos os sistemas têm uma, ou mais, fontes de aleatoriedade

Tipo de Sistema	Fontes de Aleatoriedade
Fabrico	Tempos de processamento, tempos entre avarias, tempos de reparação
Computação	Tempos entre a chegada de tarefas, tipos de tarefas, tempo de processamento requerido pelas tarefas
Comunicação	Tempos entre a chegada de mensagens, tipos de mensagens, tamanho das mensagens

- Representar cada fonte de aleatoriedade por uma distribuição de probabilidade adequada

15

Fontes de aleatoriedade

- Escolha de distribuição é determinante no sucesso do estudo de simulação
  - Exemplo de um sistema de espera com servidor único, por exemplo, uma máquina, do qual se sabe que os tempos entre as chegadas seguem a distribuição exponencial de valor médio 1 minuto, e do qual foram recolhidas 200 amostras sobre os tempos de serviço, não se sabendo a distribuição que seguem
  - Resultados obtidos aproximando as amostras pelas distribuições indicadas

Distribuição	Tempo Médio de Espera	Número médio de itens em Espera	Taxa de atrasos >= 20
Exponencial	6,71	6,78	0,064
Gamma	4,54	4,60	0,019
Weibull	4,36	4,41	0,013
Lognormal	7,19	7,30	0,078
Normal	6,04	6,13	0,045

16

Tornar o modelo estocástico

- Métodos de definir a distribuição de probabilidade com base em dados observados
  - 1) Usar os dados observados directamente na simulação
  - 2) Utilizar os dados na definição de uma distribuição empírica
  - 3) Utilizar técnicas de inferência estatística para ajustar os dados recolhidos a uma função de distribuição de probabilidades (curvas *standard*)

17

Tornar o modelo estocástico

- Usar os dados observados directamente na simulação
  - Utilizar tabelas ou histogramas com os valores retirados de experiências efectuadas sobre o sistema
  - Amostragem de Histogramas
    - Organizar as amostras recolhidas
    - Calcular a frequência de ocorrência de cada valor
    - Calcular a percentagem para cada valor
    - Construir o histograma H(x)
      - Representa, de forma discreta, a distribuição de probabilidades
    - Obter a distribuição de probabilidades acumuladas F(x)
      - Transformar a distribuição H(x) numa distribuição que se estenda pelo intervalo [0,1]
    - Gerar número aleatório entre [0,1] para obter o valor a utilizar

18

Tornar o modelo estocástico

- Amostragem de Histogramas - Exemplo
  - Considere, apresentados na tabela, os valores obtidos durante onze amostragens do número de automóveis que entraram por minuto num determinado parque de estacionamento

Número de Carros / min
17
20
12
17
17
10
17
20
11
12
17

19

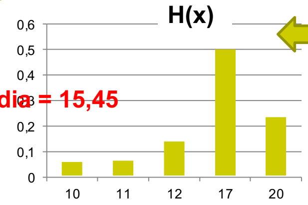
Tornar o modelo estocástico

- Amostragem de Histogramas

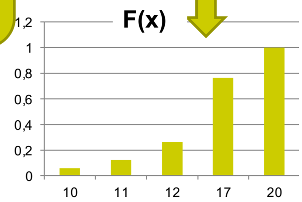
Número de Carros / min
17
20
12
17
17
10
17
20
11
12
17

Total = 170

Número de Carros / min Crescente	Total de Carros	Percentagem H(x)	Percentagem Acumulada F(x)
10	1*10	5,88	5,88
11	1*11	6,47	12,35
12	2*12	14,12	26,47
17	5*17	50,00	76,47
20	2*20	23,53	100,00



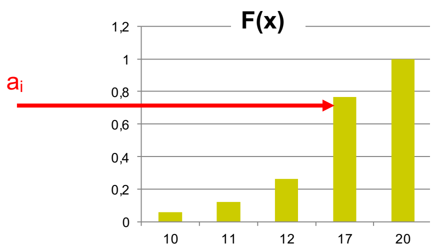
Média = 15,45



20

Tornar o modelo estocástico

- Face à distribuição de probabilidades do número de carros por minuto, como se pode introduzir tal comportamento no modelo?
  - Através da geração de números aleatórios



21

Tornar o modelo estocástico

- Usar os dados observados directamente na simulação
  - Só é possível reproduzir em simulação factos observados, sendo raro existirem dados em número suficiente para a quantidade de experiências de simulação a desenvolver
    - A definição de uma distribuição empírica com base em dados observados, permite ultrapassar esta limitação, já que permite, para dados contínuos, obter qualquer valor entre o valor mínimo e máximo observados
  - Contudo, o método é bastante útil para efectuar a validação do modelo por comparação dos resultados da simulação com os dados observados

22

Especificar Distribuição

- Distribuições Empíricas
  - Especificar directamente uma distribuição, a partir da qual serão geradas variáveis aleatórias
    - Nem sempre é possível ajustar os dados observados a uma distribuição *standard*
  - Observações individuais disponíveis
    - $X_1, X_2, X_3, \dots, X_n$
  - Observações individuais agrupadas sob a forma de intervalos ou histogramas
    - X_i agrupados em intervalos

23

Especificar Distribuição

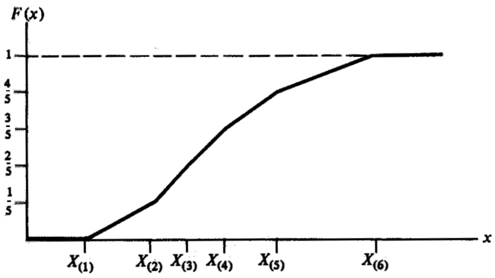
- Distribuições Empíricas
  - Observações individuais disponíveis
    - $X_1, X_2, X_3, \dots, X_n$
    - Ordenar os dados por ordem crescente, de modo a que $X_{(1)} <= X_{(2)} <= X_{(3)} <= X_{(n)}$
    - Definir a função de distribuição

$$F(x) = \begin{cases} 0 & , \text{se } x < X_{(1)} \\ \dfrac{i-1}{n-1} + \dfrac{x - X_{(i)}}{(n-1)(X_{(i+1)} - X_{(i)})} & , \text{se } X_{(i)} \leq x \leq X_{(i+1)} \\ 1 & , \text{se } X_{(n)} \leq x \end{cases}$$

24

Especificar Distribuição

- Distribuições Empíricas
  - Observações individuais disponíveis



- F(x) cresce mais acentuadamente nos intervalos em que há maior densidade de amostras recolhidas
- Para cada amostra X_i, $F(X_i) = \frac{(i-1)}{(n-1)}$, traduz a proporção de amostras que é inferior a X_i

25

Especificar Distribuição

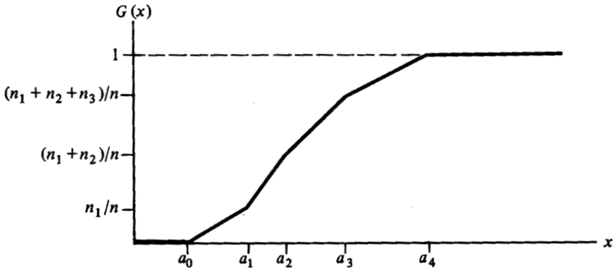
- Distribuições Empíricas
  - Observações individuais agrupadas sob a forma de intervalos ou histogramas
    - X_i agrupados em K intervalos adjacents $[a_0, a_1)$, $[a_1, a_2)$, …, $[a_{k-1}, a_k)$
    - Definir a função de distribuição

$$G(x) = \begin{cases} 0 & , \text{se } x < a_0 \\ G(a_{j-1}) + \dfrac{x - a_{(j-1)}}{a_j - a_{j-1}}[G(a_j) - G(a_{j-1})] & , \text{se } a_{j-1} \le x \le a_j \\ 1 & , \text{se } a_k \le x \end{cases}$$

26

Especificar Distribuição

- Distribuições Empíricas
  - Observações individuais agrupadas sob a forma de intervalos ou histogramas



    - G(x) cresce mais acentuadamente nos intervalos em que há maior densidade de amostras recolhidas
    - $G(a_j)$ é a proporção de X_i inferiores a a_j

27

Tornar o modelo estocástico

- Ajuste dos dados a uma distribuição de probabilidades é o mais recomendado
  - Distribuição empírica contém certas irregularidades, particularmente para amostras de pequena dimensão, que uma distribuição "teórica" suaviza, já que esta permite nivelar os dados
    - Contudo, nem sempre é possível obter aproximação por forma analítica
  - Permite, ao contrário das distribuições empíricas, onde normalmente não é possível gerar valores fora dos limites observados, o "estudo" de acontecimentos "raros", cuja influência no sistema é muitas vezes o objectivo do estudo de simulação
  - A distribuição "teórica" é uma forma compacta de representar o conjunto de dados
    - Numa distribuição empírica contínua, para ***n*** valores, será necessário fazer entrar no sistema e guardar ***2*n*** valores (dados e respectivas probabilidades cumulativas)

28

Distribuições de Probabilidade
Contínuas

- Parametrização de Distribuições Contínuas
  - Tipos básicos
    - Localização
      - Especifica o ponto, no eixo das abcissas, de localização da distribuição
      - Alterações implicam deslocar à esquerda ou direita a distribuição
    - Escala
      - Alterações fazem comprimir ou expandir a distribuição, sem alterar a sua forma
    - Forma
      - Altera propriedades da distribuição
      - Distribuição normal e exponencial não têm este tipo de parâmetro

29

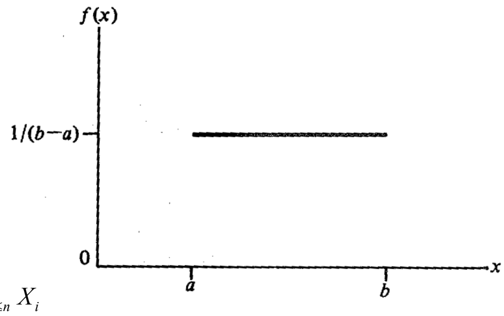
Distribuições de Probabilidade
Contínuas

- Uniforme - U(a,b)
  - Usada para modelar valores que se sabe estarem aleatoriamente distribuídos num intervalo
  - Distribuição U(0,1) é essencial na geração de valores a partir de todas as outras distribuições de probabilidade
  - Densidade $f(x) = \begin{cases} \dfrac{1}{b-a} & \text{se } a \le b \\ 0 & \text{outrosvalores} \end{cases}$
  - Distribuição $F(x) = \begin{cases} 0 & \text{se } x < a \\ \dfrac{x-a}{b-a} & \text{se } a \le x \le b \\ 1 & \text{se } b < x \end{cases}$
  - Parâmetros
    - ***a*** e ***b*** números reais, com ***a*** < ***b***
    - ***a*** é parâmetro de localização
    - ***b*** – ***a*** é factor de escala

30

Distribuições de Probabilidade
Contínuas

- Uniforme - U(a,b)
  - Gama
    - [a, b]
  - Média $\dfrac{a+b}{2}$
  - Variância $\dfrac{(b-a)^2}{12}$
  - MLE
    $\hat{a} = \min_{1\le i\le n} X_i, \hat{b} = \max_{1\le i\le n} X_i$
  - Comentários
    - U(0,1) é um caso especial da distribuição beta (com $\alpha_1 = \alpha_2 = 1$)
  - Se X ~U(0,1), e [x, x + Δx] é sub-intervalo de [0,1]



$$P(X \in [x, x+\Delta x]) = \int_x^{x+\Delta x} 1 d(x) = (x+\Delta x) - x = \Delta x$$

31

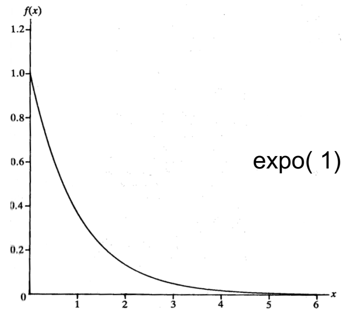
Distribuições de Probabilidade
Contínuas

- Exponencial - expo(β)
  - Usada para modelar tempos entre chegadas ao sistema
  - Densidade
    $$f(x) = \begin{cases} \dfrac{1}{\beta} e^{-x/\beta} & \text{se x} \ge 0 \\ 0 & \text{outros valores} \end{cases}$$
  - Distribuição
    $$F(x) = \begin{cases} 1 - e^{-x/\beta} & \text{se x} \ge 0 \\ 0 & \text{outros valores} \end{cases}$$
  - Parâmetros
    - Factor de escala β > 0
  - Gama
    - $[0, \infty[$

32

Distribuições de Probabilidade
Contínuas

- Exponencial - expo(β)
  - Média
    - β
  - Variância
    - β^2
  - MLE
    - $\beta = \overline{X}(n)$



  - Comentários
    - Caso especial da distribuição gamma e Weibull (α = 1 e factor de escala β, em ambas as situações)
    - É a única distribuição contínua sem propriedades de memorização
    - Variável aleatória X não tem propriedades de memorização se

$$P(X > t + s \mid X > t) = P(X > s) \quad \text{para todo t,s} > 0$$

33

Distribuições de Probabilidade
Contínuas

- Normal - N(μ, σ^2)
  - Simular erros de vários tipos; quantidades que traduzem a soma de um grande número de outras quantidades (teorema do Limite Central)
  - Densidade

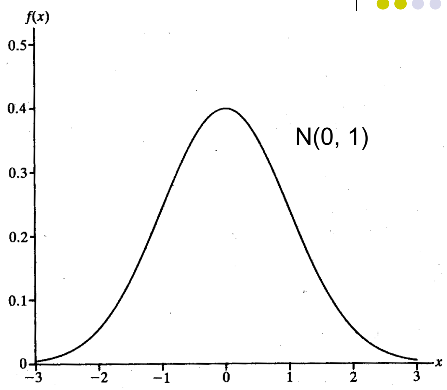
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \quad \text{, para qualquer número real } x$$

  - Parâmetros
    - μ є (-∞ , ∞) é parâmetro de localização
    - σ > 0 é factor de escala
  - Gama
    - (-∞, ∞)

34

Distribuições de Probabilidade
Contínuas

- Normal - N(μ, σ^2)
  - Média
    - μ
  - Variância
    - σ^2
  - MLE
    - $\mu = \overline{X}(n), \sigma = \left[\frac{n-1}{n} S^2(n)\right]^{1/2}$



  - Comentários
    - N(0,1) é designada distribuição standard
    - Se a distribuição normal for usada para gerar números aleatórios não negativos, a sua densidade deve ser truncada em x=0

35

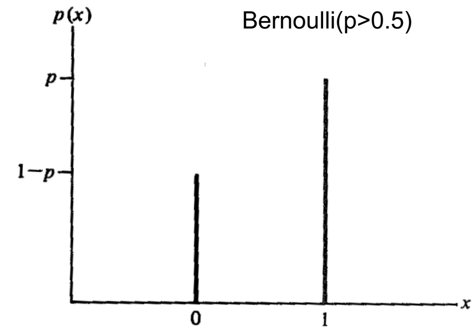
Distribuições de Probabilidade
Discretas

- Bernoulli – Bernoulli (p)
  - Ocorrências instantâneas com dois resultados possíveis; usada para gerar outras variáveis aleatórias discretas
  - Probabilidade $p(x) = \begin{cases} 1-p & \text{se } x = 0 \\ p & \text{se } x = 1 \\ 0 & \text{outros} \end{cases}$
  - Distribuição $F(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1-p & \text{se } 0 \le x \le 1 \\ 1 & \text{se } x \le 1 \end{cases}$
  - Parâmetros
    - p є (0,1)

36

Distribuições de Probabilidade
Discretas

- Bernoulli – Bernoulli (p)
  - Gama
    - { 0, 1}
  - Média
    - p
  - Variância
    - p(1-p)
  - MLE
    - $p = \overline{X}(n)$
  - Comentários
    - Variável aleatória X que segue a distribuição Bernoulli(p) pode traduzir o resultado de uma experiência que "falha" (X=0) ou tem "sucesso" (X=1)



37

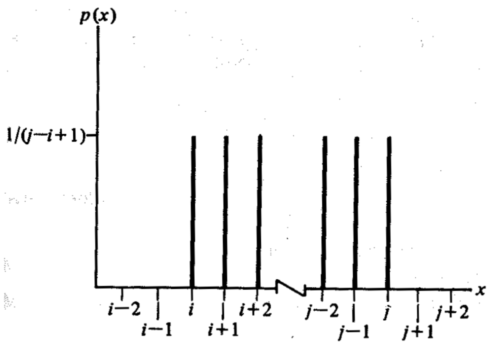
Distribuições de Probabilidade
Discretas

- Discreta Uniforme – DU (i, j)
  - Ocorrências aleatórias, todas com a mesma probabilidade
  - Usada como primeira abordagem, quando se sabe apenas que uma determinada quantidade varia entre dois valores inteiros
  - Probabilidade $p(x) = \begin{cases} \dfrac{1}{j-i+1} & \text{se } x \in \{i, i+1, ..., j\} \\ 0 & \text{outros valores} \end{cases}$
  - Distribuição $F(x) = \begin{cases} 0 & \text{se } x < i \\ \dfrac{\lfloor x \rfloor - i + 1}{j-i+1} & \text{se } i \le x \le j \\ 1 & \text{se } j < x \end{cases}$, em que $\lfloor x \rfloor$ representa o maior valor $\le$ x

38

Distribuições de Probabilidade
Discretas

- Discreta Uniforme – DU (i, j)
  - Parâmetros
    - i e j inteiros, com i ≤ j
    - i é parâmetro de localização
    - j – i factor de escala
  - Gama
    - { i, i+1, …, j}
  - Média
    - $\frac{i+j}{2}$
  - Variância
    - $\frac{(j-i+1)^2-1}{12}$
  - MLE
    - $\hat{i} = \min_{1\le k\le n} X_k, \hat{j} = \max_{1\le k\le n} Xk$
  - Comentários
    - DU(0,1) e Bernoulli(1/2) representam a mesma distribuição



39

Distribuições de Probabilidade
Discretas

- Poisson – Poisson (λ)
  - Número de eventos que ocorre, a taxa constante, num determinado intervalo de tempo
  - Por exemplo, número de itens num lote de tamanho aleatório, número de itens retirados de stock
  - Probabilidade

$$p(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & \text{se } x \in \{0,1,\ldots\} \\ 0 & \text{outros valores} \end{cases}$$

  - Distribuição

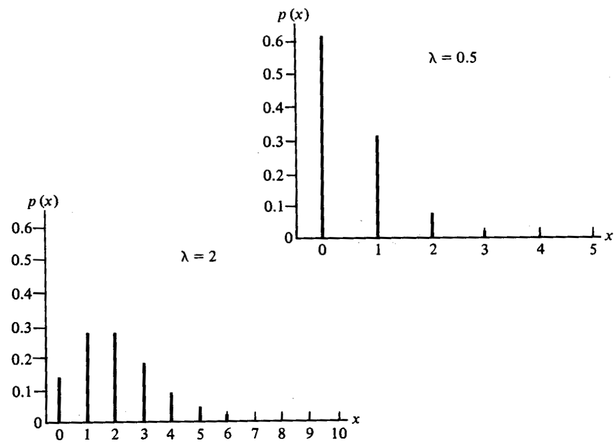
$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ e^{-\lambda}\displaystyle\sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!} & \text{se } 0 \le x \end{cases}$$

40

Distribuições de Probabilidade

Discretas

- Poisson – Poisson (λ)
  - Parâmetros
    - $\lambda > 0$
  - Gama
    - { 0,1, …}
  - Média
    - λ
  - Variância
    - λ
  - MLE
    - $\hat{\lambda} = \overline{X}(n)$



41

Tornar o modelo estocástico

- Ajustar os dados a uma distribuição de probabilidades standard
  - Actividade 1) -> Decidir qual das famílias de distribuições – exponencial, normal, etc. – se aproxima dos dados recolhidos, sem preocupações quanto aos parâmetros específicos que as caracterizam
  - Actividade 2) -> Estimar os valores dos parâmetros das distribuições seleccionadas na actividade 1)
  - Actividade 3) -> Determinar se, a distribuição(s) seleccionada(s), nas actividades anteriores, asseguram a representatividade adequada dos dados

42

Fontes de aleatoriedade

- Escolha de distribuição é determinante no sucesso do estudo de simulação
  - Exemplo de um sistema de espera com servidor único, por exemplo, uma máquina, do qual se sabe que os tempos entre as chegadas seguem a distribuição exponencial de valor médio 1 minuto, e do qual foram recolhidas 200 amostras sobre os tempos de serviço, não se sabendo a distribuição que seguem
  - Resultados obtidos aproximando as amostras pelas distribuições indicadas

Distribuição	Tempo Médio de Espera	Número médio de itens em Espera	Taxa de atrasos >= 20
Exponencial	6,71	6,78	0,064
Gamma	4,54	4,60	0,019
Weibull	4,36	4,41	0,013
Lognormal	7,19	7,30	0,078
Normal	6,04	6,13	0,045

43

Tornar o modelo estocástico

- Ajustar os dados a uma distribuição de probabilidades
  - Actividade 1) ->Seleccionar famílias de distribuições que possam ser representativas dos dados existentes
  - Recorrer a conhecimento teórico
    - Por exemplo, quando a chegada de clientes a um serviço ocorre a taxa constante, e o número de clientes que chegam em intervalos de tempo disjuntos for independente, há fundamento teórico para utilizar a distribuição exponencial na sua representação
    - Tempos de serviço não devem, em princípio, ser gerados com base na distribuição normal, já que um valor negativo poderia ser obtido…
  - Heurísticas
    - Resumo Estatístico
    - Histogramas e linhas de gráficos
    - Análise de sectores

44

Tornar o modelo estocástico

- Seleccionar famílias de distribuições
  - Resumo Estatístico

Funções	Estimativa	Continua(C) / Discreta(D)	Comentários
Mínimo, Máximo	$X_{(1)}, X_{(n)}$	C, D	Medida da gama de valores
Média μ	$\overline{X}(n)$	C, D	Medida da tendência central
Mediana $x_{0,5}$	$\hat{x}_{0,5}(n)=\begin{cases} X_{(n+1)/2} & \text{se n impar} \\ \left[X_{(n/2)}+X((n/2)+1)\right]/2 & \text{se n par} \end{cases}$	C, D	Medida alternativa da tendência central
Variância σ^2	$S^2(n)$	C, D	Medida da variabilidade
Coeficiente de Variação, $cv=\sqrt{\sigma^2}/\mu$	$\widehat{cv}(n)=\frac{\sqrt{S^2(n)}}{\overline{X}(n)}$	C	Medida alternativa da variabilidade
Coeficiente de variação, $\tau=\sigma^2/\mu$	$\hat{\tau}(n)=\frac{S^2(n)}{\overline{X}(n)}$	D	Medida alternativa da variabilidade
Assimetria, $\nu=E[(X-\mu)^3]/(\sigma^2)^{3/2}$	$\hat{\nu}(n)=\frac{\sum_{i=1}^{n}[X_i-\overline{X}(n)]^3/n}{[S^2(n)]^{3/2}}$	C, D	Medida de simetria

45

Tornar o modelo estocástico

- Seleccionar famílias de distribuições - Resumo Estatístico
  - Comparar parâmetros estatísticos dos dados recolhidos com os valores que caracterizam as distribuições de probabilidade
  - Características das distribuições
    - Nas distribuições simétricas, a média é igual à mediana (aproximadamente igual no caso de distribuições discretas)
      - Por exemplo, a distribuição normal
  - Coeficiente de variação também fornece informação muito útil
    - Distribuição exponencial é =1
  - Assimetria
    - Para distribuições simétricas é = 0
      - por ex., distribuição normal
    - Para distribuições que se prolongam à direita é >0
      - Por ex., distribuição exponencial
    - Nas distribuições que se prolongam à esquerda é < 0

46

Tornar o modelo estocástico

- Seleccionar famílias de distribuições – Histogramas e Linhas de Gráficos
  - Obter estimativa gráfica da função de densidade associada à distribuição dos dados recolhidos
  - Analisar e comparar com as curvas das distribuições *standard*
  - Histogramas
    - Estimativa da função de densidade associada a conjunto de dados contínuos
  - Linhas de Gráficos
    - Estimativa da função de densidade associada a conjunto de dados discretos

47

Tornar o modelo estocástico

- Seleccionar famílias de distribuições – Histogramas e Linhas de Gráficos
  - Histogramas
    - Separar os dados em ***k*** intervalos adjacentes $[b_0, b_1)$, $[b_1, b_2)$, …, $[b_{k-1}, b_k)$, com a mesma largura $\Delta b = b_j - b_{j-1}$
    - Definir a função de densidade com base em

$$h(x) = \begin{cases} 0 & \text{se } x < b_0 \\ h_j & \text{se } b_{j-1} \le x \le b_j \\ 0 & \text{se } b_k \le x \end{cases}$$

      - h_j proporção de valores que pertencem ao intervalo
    - Comparar com a forma das curvas das distribuições teóricas (ignorar localização e factor de escala)

48

Tornar o modelo estocástico

- Seleccionar famílias de distribuições – Histogramas e Linhas de Gráficos
  - Histogramas
    - Aplicáveis a qualquer distribuição contínua
    - Permitem ter uma ideia visual relativa à semelhança com as funções de distribuição de referência
    - Problema em definir o número ***k*** de intervalos a utilizar
      - Usar regra de Sturge's, e escolher ***k*** de acordo com

$$k = \lfloor 1 + \log_2 n \rfloor = \lfloor 1 + 3.322 \log_{10} n \rfloor$$

      - Tentar vários valores de Δb até conseguir aspecto similar ao de uma das distribuições standard

Tornar o modelo estocástico

- Seleccionar famílias de distribuições – Histogramas e Linhas de Gráficos
  - Linhas de Gráficos
    - Estimativa da função de densidade associada a conjunto de dados discretos
    - Para cada valor x_i, sendo h_j a proporção de amostras iguais a x_i
      - Desenhar linhas verticais associadas aos valores de h_j
      - Comparar com as curvas das distribuições standard
    - Abordagem similar à dos histogramas, na sua construção, interpretação e utilização, mas mais atractiva na medida em que não envolve aspectos subjectivos, como as decisões quanto à largura e localização dos intervalos

Tornar o modelo estocástico

- Seleccionar famílias de distribuições – Análise de Sectores
  - Análise da amostra com vista a determinar se a função distribuição de probabilidade é simétrica, ou estará *enviesada* à esquerda ou direita
    - Aplicável a distribuições contínuas ou discretas
  - Para 0<***q***<1, o ***q***-quantil é o valor x_q para o qual $F(x_q) = q$
    - Particularmente importantes são: $x_{0,5}$(a mediana), $x_{0,25}$ e $x_{0,75}$, e $x_{0,125}$ e $x_{0,875}$

51

Tornar o modelo estocástico

- Seleccionar famílias de distribuições – Análise de Sectores
  - Estrutura da análise de sectores para amostras de $X_1, X_2, \ldots, X_n$

Quantile	Depth	Sample value(s)		Midpoint
Median	$i = (n+1)/2$	$X_{(i)}$		$X_{(i)}$
Quartiles	$j = (\lfloor i \rfloor + 1)/2$	$X_{(j)}$	$X_{(n-j+1)}$	$[X_{(j)} + X_{(n-j+1)}]/2$
Octiles	$k = (\lfloor j \rfloor + 1)/2$	$X_{(k)}$	$X_{(n-k+1)}$	$[X_{(k)} + X_{(n-k+1)}]/2$
Extremes	1	$X_{(1)}$	$X_{(n)}$	$[X_{(1)} + X_{(n)}]/2$

- Se a distribuição é simétrica os 4 valores do meio deverão ser aproximadamente idênticos
- Se está enviesada à direita (esquerda) então os 4 valores são crescentes (decrescentes)

52

Tornar o modelo estocástico

- Ajustar os dados a uma distribuição de probabilidades
  - Actividade 2) -> Estimar os valores dos parâmetros das distribuições seleccionadas na actividade 1)
    - Usar os dados para especificar valor numérico para parâmetro desconhecido -> estimar o parâmetro a partir dos dados
      - Por exemplo, a média de uma amostra , enquanto variável aleatória, $\overline{X}(n)$, é um estimador da média da população
    - Estimador é uma função numérica dos dados
      - Por exemplo, MLE para a distribuição exponencial

$$\beta = \overline{X}(n)$$

53

Tornar o modelo estocástico

- Ajustar os dados a uma distribuição de probabilidades
  - Actividade 3) -> Determinar se, a distribuição(s) seleccionada(s), nas actividades anteriores, assegura representatividade adequada dos dados
    - Se há várias que representam adequadamente os dados, deve determinar-se qual a que melhor se ajusta
    - Recorrer a heurísticas
      - Comparação de Frequências
    - Métodos de *goodness-of-fit*
      - Testes Qui-Quadrado

54

Tornar o modelo estocástico

- Verificar representatividade dos dados
  - Heurística - Comparação de frequências
    - Dados Contínuos
      - Comparar histograma dos dados observados com os obtidos a partir da distribuição obtida por estimativa
      - Para ***k*** intervalos adjacentes $[b_0, b_1)$, $[b_1, b_2)$, …, $[b_{k-1}, b_k)$, com a mesma largura $\Delta b = b_j - b_{j-1}$
      - Com h_j a proporção de valores observados no intervalo $[b_{j-1}, b_j)$, e r_j a proporção de valores observados no mesmo intervalo, para a distribuição obtida por estimativa

$$r_j = \int_{b_{j-1}}^{b_j} \widehat{f}(x)dx$$

      - Desenhar histograma para h_j e para r_j para todos os intervalos j=1, 2,…, k
      - Se a distribuição estimada for uma boa representação dos dados, então h_j e r_j serão semelhantes

55

Tornar o modelo estocástico

- Verificar representatividade dos dados
  - Heurística - Comparação de frequências
    - Dados Discretos
      - Comparar linha de gráfico dos dados observados com os obtidos a partir da distribuição obtida por estimativa
      - Com h_j a proporção de valores observados que são iguais a x_j, e r_j a proporção de valores, obtidos através da distribuição obtida por estimativa, que seriam iguais a x_j

$$r_j = \widehat{p}(x_i)$$

      - Se a distribuição estimada for uma boa representação dos dados, então h_j e r_j serão semelhantes

56

Tornar o modelo estocástico

- Verificar representatividade dos dados
  - Métodos de *goodness-of-fit*
    - Testes estatísticos de hipóteses, usados para determinar se um conjunto de observações $X_1, X_2, \ldots, X_n$ representam amostras independentes de uma distribuição
    - Por exemplo
      - H_0 – As amostras X_i são variáveis aleatórias independentes e identicamente distribuídas da função de distribuição *F*
    - Usados, sobretudo para detectar diferenças razoáveis
    - Testes Qui-Quadrado
      - Teste utilizado no estudo da conformidade da distribuição amostral a um modelo ou distribuição populacional

57

Tornar o modelo estocástico

- Verificar representatividade dos dados
  - Métodos de *goodness-of-fit* - Testes Qui-Quadrado
    - Dividir a gama de valores da distribuição estimada em ***k*** intervalos adjacentes $[a_0, a_1), [a_1, a_2), \ldots, [a_{k-1}, a_k)$, sendo que a_0 pode ser $-\infty$ e/ou a_k pode ser ∞
    - Sendo N_j o número de X_i no intervalo $[a_{j-1}, a_j)$
      - Note-se que $\sum_{j=1}^{k} N_j = n$
    - Calcular a proporção de valores p_j que pertencem ao intervalo *j*
      - Contínuo $p_j = \int_{b_{j-1}}^{b_j} \hat{f}(x)dx$ Discreto $p_j = \sum_{a_{j-1} \le x_i \le a_j} \hat{p}(x_i)$
    - Teste Estatístico $\chi^2 = \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j}$

58

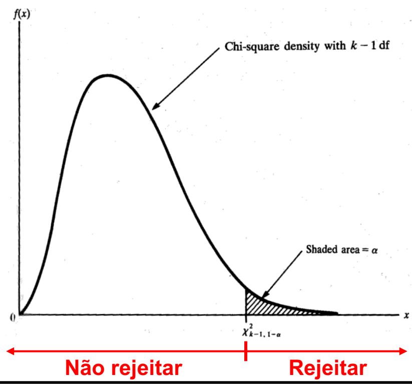
Tornar o modelo estocástico

- Verificar representatividade dos dados
  - Métodos de *goodness-of-fit* - Testes Qui-Quadrado
    - Teste Estatístico $\chi^2 = \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j}$
      - Rejeitar a hipótese se χ^2 for elevado
    - Se a hipótese é verdadeira, χ^2 converge (n -> ∞) para a distribuição qui-quadrado com k-1 graus de liberdade
    - A hipótese deve ser rejeitada, com grau de confiança 1-α se $\chi^2 > \chi^2_{k-1,1-\alpha}$ em que, $\chi^2_{k-1,1-\alpha}$ é o ponto crítico 1-α da distribuição qui-quadrado com k-1 graus de liberdade
      - Consultar tabela...

59

Tornar o modelo estocástico

- Verificar representatividade dos dados
  - Métodos de *goodness-of-fit* - Testes Qui-Quadrado
    - Valores de $\chi^2_{k-1,1-\alpha}$ tabelados, em que o teste é de nível *α* quando n-> ∞



60

Tornar o modelo estocástico

Critical points $\chi^2_{\nu,\gamma}$ for the chi-square distribution with ν df

$\gamma = P(Y_\nu \le \chi^2_{\nu,\gamma})$, where Y_ν has a chi-square distribution with ν df; for large ν, use the approximation for $\chi^2_{\nu,\gamma}$ in Sec. 7.4.1

	γ						
ν	0.250	0.500	0.750	0.900	0.950	0.975	0.990
1	0.102	0.455	1.323	2.706	3.841	5.024	6.635
2	0.575	1.386	2.773	4.605	5.991	7.378	9.210
3	1.213	2.366	4.108	6.251	7.815	9.348	11.345
4	1.923	3.357	5.385	7.779	9.488	11.143	13.277
5	2.675	4.351	6.626	9.236	11.070	12.833	15.086
6	3.455	5.348	7.841	10.645	12.592	14.449	16.812
7	4.255	6.346	9.037	12.017	14.067	16.013	18.475
8	5.071	7.344	10.219	13.362	15.507	17.535	20.090
9	5.899	8.343	11.389	14.684	16.919	19.023	21.666
10	6.737	9.342	12.549	15.987	18.307	20.483	23.209
11	7.584	10.341	13.701	17.275	19.675	21.920	24.725
12	8.438	11.340	14.845	18.549	21.026	23.337	26.217
13	9.299	12.340	15.984	19.812	22.362	24.736	27.688
14	10.165	13.339	17.117	21.064	23.685	26.119	29.141
15	11.037	14.339	18.245	22.307	24.996	27.488	30.578
16	11.912	15.338	19.369	23.542	26.296	28.845	32.000
17	12.792	16.338	20.489	24.769	27.587	30.191	33.409
18	13.675	17.338	21.605	25.989	28.869	31.526	34.805
19	14.562	18.338	22.718	27.204	30.144	32.852	36.191
20	15.452	19.337	23.828	28.412	31.410	34.170	37.566
21	16.344	20.337	24.935	29.615	32.671	35.479	38.932
22	17.240	21.337	26.039	30.813	33.924	36.781	40.289
23	18.137	22.337	27.141	32.007	35.172	38.076	41.638
24	19.037	23.337	28.241	33.196	36.415	39.364	42.980
25	19.939	24.337	29.339	34.382	37.652	40.646	44.314
26	20.843	25.336	30.435	35.563	38.885	41.923	45.642
27	21.749	26.336	31.528	36.741	40.113	43.195	46.963
28	22.657	27.336	32.620	37.916	41.337	44.461	48.278
29	23.567	28.336	33.711	39.087	42.557	45.722	49.588
30	24.478	29.336	34.800	40.256	43.773	46.979	50.892
40	33.660	39.335	45.616	51.805	55.758	59.342	63.691
50	42.942	49.335	56.334	63.167	67.505	71.420	76.154
75	66.417	74.334	82.858	91.061	96.217	100.839	106.393
100	90.133	99.334	109.141	118.498	124.342	129.561	135.807

Tornar o modelo estocástico

- Verificar representatividade dos dados
  - Métodos de *goodness-of-fit* - Testes Qui-Quadrado
    - Dificuldade ao recorrer a este método consiste em definir o número e tamanho dos intervalos
      - Não há indicações quanto à garantia de produzir bons resultados, contudo, sabe-se que alguma ambiguidade se consegue eliminar ao considerar $p_1=p_2=\ldots=p_k$, o que se designa de aproximação equi-provável
        - Pode ser mais complicado para distribuições contínuas na medida em que implica o recurso à inversa da função de distribuição
      - Yarnold provou que o teste qui-quadrado será válido se $k \geq 3$ e $np_j \geq 5$

Tornar o modelo estocástico

- Ajustar os dados a uma distribuição de probabilidades
  - Exemplo
    - Foram recolhidos dados, quanto ao padrão de chegada de veículos a um sistema de atendimento do tipo “drive –in”. Durante um período de 90 minutos registou-se a chegada de 220 carros, tendo-se anotado os tempos entre as chegadas X_i , de i=1, 2,…, 219
      - Actividade 1) -> Decidir qual das famílias de distribuições – exponencial, normal, etc. – se aproxima dos dados recolhidos
      - Actividade 2) -> Estimar os valores dos parâmetros das distribuições seleccionadas na actividade 1)
      - Actividade 3) -> Determinar se, a distribuição(s) seleccionada(s), nas actividades anteriores, assegura representatividade adequada dos dados

63

Tornar o modelo estocástico

- Ajustar os dados a uma distribuição de probabilidades
  - Exemplo – Dados Recolhidos

0.01	0.06	0.12	0.23	0.38	0.53	0.88
0.01	0.07	0.12	0.23	0.38	0.53	0.88
0.01	0.07	0.12	0.24	0.38	0.54	0.90
0.01	0.07	0.13	0.25	0.39	0.54	0.93
0.01	0.07	0.13	0.25	0.40	0.55	0.93
0.01	0.07	0.14	0.25	0.40	0.55	0.95
0.01	0.07	0.14	0.25	0.41	0.56	0.97
0.01	0.07	0.14	0.25	0.41	0.57	1.03
0.02	0.07	0.14	0.26	0.43	0.57	1.05
0.02	0.07	0.15	0.26	0.43	0.60	1.05
0.03	0.07	0.15	0.26	0.43	0.61	1.06
0.03	0.08	0.15	0.26	0.44	0.61	1.09
0.03	0.08	0.15	0.26	0.45	0.63	1.10
0.04	0.08	0.15	0.27	0.45	0.63	1.11
0.04	0.08	0.15	0.28	0.46	0.64	1.12
0.04	0.09	0.17	0.28	0.47	0.65	1.17
0.04	0.09	0.18	0.29	0.47	0.65	1.18
0.04	0.10	0.19	0.29	0.47	0.65	1.24
0.04	0.10	0.19	0.30	0.48	0.69	1.24
0.05	0.10	0.19	0.31	0.49	0.69	1.28
0.05	0.10	0.20	0.31	0.49	0.70	1.33
0.05	0.10	0.21	0.32	0.49	0.72	1.38
0.05	0.10	0.21	0.35	0.49	0.72	1.44
0.05	0.10	0.21	0.35	0.50	0.72	1.51
0.05	0.10	0.21	0.35	0.50	0.74	1.72
0.05	0.10	0.21	0.36	0.50	0.75	1.83
0.05	0.11	0.22	0.36	0.51	0.76	1.96
0.05	0.11	0.22	0.36	0.51	0.77	
0.05	0.11	0.22	0.37	0.51	0.79	
0.06	0.11	0.23	0.37	0.52	0.84	
0.06	0.11	0.23	0.38	0.52	0.86	
0.06	0.12	0.23	0.38	0.53	0.87	

64

Tornar o modelo estocástico

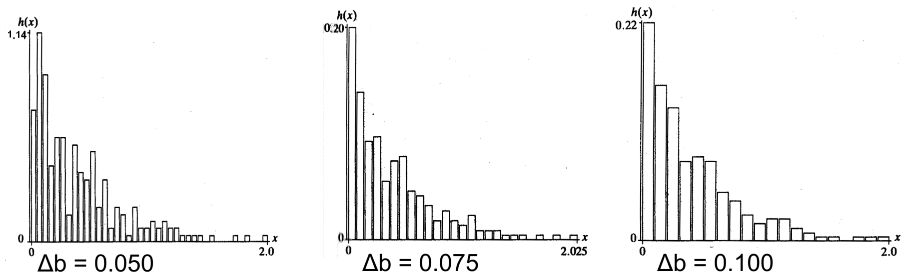
- Ajustar os dados a uma distribuição de probabilidades
  - Exemplo – Resumo Estatístico

Resumo Estatístico	Valor
Mínimo	0.010
Máximo	1.960
Média	0.399
Mediana	0.270
Variância	0.144
Coeficiente de Variação	0.953
Assimetria	1.458

  - Como $\bar{X}(219) = 0.399 > 0.270 = \hat{x}_{0.5}(219)$ e $\hat{v}(219) = 1.458$ sugere que a distribuição não é simétrica, estando enviesada à direita
  - Também $\widehat{cv}(219) = 0.953$, está próximo do valor teórico de 1, para a distribuição exponencial
  - Em termos teóricos, a aproximação seria feita por uma distribuição exponencial

Tornar o modelo estocástico

- Ajustar os dados a uma distribuição de probabilidades
  - Exemplo – Histogramas e Linhas de gráficos



  - Hipótese – tempos entre as chegadas seguem a distribuição exponencial

Tornar o modelo estocástico

- Ajustar os dados a uma distribuição de probabilidades
  - Actividade 2) -> Estimar os valores dos parâmetros das distribuições seleccionadas na actividade 1)
    - MLE = $\widehat{\beta} = \overline{X}(n)$
  - Exemplo – Resumo Estatístico
    - Hipótese – tempos entre as chegadas seguem a distribuição exponencial
      - $\widehat{\beta} = \overline{X}(219) = 0.399$

67

Tornar o modelo estocástico

- Verificar representatividade dos dados
  - Métodos de *goodness-of-fit*
    - Testes estatísticos de hipóteses, usados para determinar se um conjunto de observações X_1, X_2, …, X_n representam amostras independentes de uma distribuição
    - Por exemplo
      - H_0 – As amostras X_i são variáveis aleatórias independentes e identicamente distribuídas da função de distribuição *F*
    - Usados, sobretudo para detectar diferenças razoáveis
    - Testes Qui-Quadrado
      - Teste utilizado no estudo da conformidade da distribuição amostral a um modelo ou distribuição populacional

68

Tornar o modelo estocástico

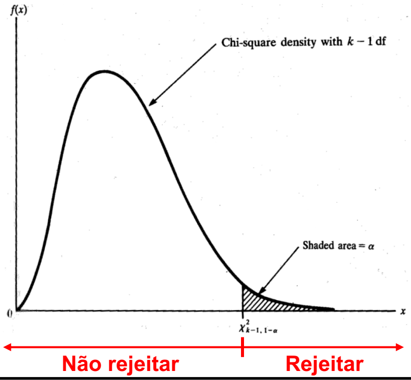
- Verificar representatividade dos dados
  - Métodos de *goodness-of-fit* - Testes Qui-Quadrado
    - Dividir a gama de valores da distribuição estimada em ***k*** intervalos adjacentes $[a_0, a_1), [a_1, a_2), \ldots, [a_{k-1}, a_k)$, sendo que a_0 pode ser $-\infty$ e/ou a_k pode ser ∞
    - Sendo N_j o número de X_i no intervalo $[a_{j-1}, a_j)$
      - Note-se que $\sum_{j=1}^{k} N_j = n$
    - Calcular a proporção de valores p_j que pertencem ao intervalo *j*
      - Contínuo $p_j = \int_{b_{j-1}}^{b_j} \widehat{f}(x)dx$ Discreto $p_j = \sum_{a_{j-1} \le x_i \le a_j} \widehat{p}(x_i)$
    - Teste Estatístico $\chi^2 = \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j}$

Tornar o modelo estocástico

- Verificar representatividade dos dados
  - Métodos de *goodness-of-fit* - Testes Qui-Quadrado
    - Teste Estatístico $\chi^2 = \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j}$
      - Rejeitar a hipótese se χ^2 for elevado
    - Se a hipótese é verdadeira, χ^2 converge (n -> ∞) para a distribuição qui-quadrado com k-1 graus de liberdade
    - A hipótese deve ser rejeitada, com grau de confiança 1-α se $\chi^2 > \chi^2_{k-1,1-\alpha}$ em que, $\chi^2_{k-1,1-\alpha}$ é o ponto crítico 1-α da distribuição qui-quadrado com k-1 graus de liberdade
      - Consultar tabela...

Tornar o modelo estocástico

- Verificar representatividade dos dados
  - Métodos de *goodness-of-fit* - Testes Qui-Quadrado
    - Valores de $\chi^2_{k-1,1-\alpha}$ tabelados, em que o teste é de nível α quando n-> ∞



71

Tornar o modelo estocástico

Critical points $\chi^2_{\nu,\gamma}$ for the chi-square distribution with ν df

$\gamma = P(Y_\nu \le \chi^2_{\nu,\gamma})$, where Y_ν has a chi-square distribution with ν df; for large ν, use the approximation for $\chi^2_{\nu,\gamma}$ in Sec. 7.4.1

	γ						
ν	0.250	0.500	0.750	0.900	0.950	0.975	0.990
1	0.102	0.455	1.323	2.706	3.841	5.024	6.635
2	0.575	1.386	2.773	4.605	5.991	7.378	9.210
3	1.213	2.366	4.108	6.251	7.815	9.348	11.345
4	1.923	3.357	5.385	7.779	9.488	11.143	13.277
5	2.675	4.351	6.626	9.236	11.070	12.833	15.086
6	3.455	5.348	7.841	10.645	12.592	14.449	16.812
7	4.255	6.346	9.037	12.017	14.067	16.013	18.475
8	5.071	7.344	10.219	13.362	15.507	17.535	20.090
9	5.899	8.343	11.389	14.684	16.919	19.023	21.666
10	6.737	9.342	12.549	15.987	18.307	20.483	23.209
11	7.584	10.341	13.701	17.275	19.675	21.920	24.725
12	8.438	11.340	14.845	18.549	21.026	23.337	26.217
13	9.299	12.340	15.984	19.812	22.362	24.736	27.688
14	10.165	13.339	17.117	21.064	23.685	26.119	29.141
15	11.037	14.339	18.245	22.307	24.996	27.488	30.578
16	11.912	15.338	19.369	23.542	26.296	28.845	32.000
17	12.792	16.338	20.489	24.769	27.587	30.191	33.409
18	13.675	17.338	21.605	25.989	28.869	31.526	34.805
19	14.562	18.338	22.718	27.204	30.144	32.852	36.191
20	15.452	19.337	23.828	28.412	31.410	34.170	37.566
21	16.344	20.337	24.935	29.615	32.671	35.479	38.932
22	17.240	21.337	26.039	30.813	33.924	36.781	40.289
23	18.137	22.337	27.141	32.007	35.172	38.076	41.638
24	19.037	23.337	28.241	33.196	36.415	39.364	42.980
25	19.939	24.337	29.339	34.382	37.652	40.646	44.314
26	20.843	25.336	30.435	35.563	38.885	41.923	45.642
27	21.749	26.336	31.528	36.741	40.113	43.195	46.963
28	22.657	27.336	32.620	37.916	41.337	44.461	48.278
29	23.567	28.336	33.711	39.087	42.557	45.722	49.588
30	24.478	29.336	34.800	40.256	43.773	46.979	50.892
40	33.660	39.335	45.616	51.805	55.758	59.342	63.691
50	42.942	49.335	56.334	63.167	67.505	71.420	76.154
75	66.417	74.334	82.858	91.061	96.217	100.839	106.393
100	90.133	99.334	109.141	118.498	124.342	129.561	135.807

72

Tornar o modelo estocástico

- Ajustar os dados a uma distribuição de probabilidades
  - Actividade 3) -> Determinar, se a distribuição seleccionada é de facto representativa dos dados
    - Teste de Qui-quadrado (rejeitar hipótese se $\chi^2 > \chi^2_{k-1,1-\alpha}$)
  - Exemplo
    - Hipótese – tempos entre as chegadas seguem a distribuição exponencial $\widehat{F}(x) = 1 - e^{-x/0.399}$, $\text{para } x \geq 0$
    - Definir *k*=20 intervalos, com $p_j = 1/k = 0.05$, para j=1, 2, …, 20
      - Satisfaz aproximação equi-provável e considerações $k \geq 3$ e $np_j \geq 5$
    - Encontrar a_j através da inversa de $\widehat{F}(x)$
      - Satisfazer $\widehat{F}(a_j) = j/20$
      - $a_0 = 0$, $a_{20} = \infty$; $a_j = -0.399 \ln(1-j/20)$, para j=1, 2, …, 19

73

Tornar o modelo estocástico

- Ajustar os dados a uma distribuição de probabilidades
  - Exemplo – Dados Recolhidos

0.01	0.06	0.12	0.23	0.38	0.53	0.88
0.01	0.07	0.12	0.23	0.38	0.53	0.88
0.01	0.07	0.12	0.24	0.38	0.54	0.90
0.01	0.07	0.13	0.25	0.39	0.54	0.93
0.01	0.07	0.13	0.25	0.40	0.55	0.93
0.01	0.07	0.14	0.25	0.40	0.55	0.95
0.01	0.07	0.14	0.25	0.41	0.56	0.97
0.01	0.07	0.14	0.25	0.41	0.57	1.03
0.02	0.07	0.14	0.26	0.43	0.57	1.05
0.02	0.07	0.15	0.26	0.43	0.60	1.05
0.03	0.07	0.15	0.26	0.43	0.61	1.06
0.03	0.08	0.15	0.26	0.44	0.61	1.09
0.03	0.08	0.15	0.26	0.45	0.63	1.10
0.04	0.08	0.15	0.27	0.45	0.63	1.11
0.04	0.08	0.15	0.28	0.46	0.64	1.12
0.04	0.09	0.17	0.28	0.47	0.65	1.17
0.04	0.09	0.18	0.29	0.47	0.65	1.18
0.04	0.10	0.19	0.29	0.47	0.65	1.24
0.04	0.10	0.19	0.30	0.48	0.69	1.24
0.05	0.10	0.19	0.31	0.49	0.69	1.28
0.05	0.10	0.20	0.31	0.49	0.70	1.33
0.05	0.10	0.21	0.32	0.49	0.72	1.38
0.05	0.10	0.21	0.35	0.49	0.72	1.44
0.05	0.10	0.21	0.35	0.50	0.72	1.51
0.05	0.10	0.21	0.35	0.50	0.74	1.72
0.05	0.10	0.21	0.36	0.50	0.75	1.83
0.05	0.11	0.22	0.36	0.51	0.76	1.96
0.05	0.11	0.22	0.36	0.51	0.77	
0.05	0.11	0.22	0.37	0.51	0.79	
0.06	0.11	0.23	0.37	0.52	0.84	
0.06	0.11	0.23	0.38	0.52	0.86	
0.06	0.12	0.23	0.38	0.53	0.87	

74

Tornar o modelo estocástico

- Ajustar os dados a uma distribuição de probabilidades
  - Exemplo – Teste de Qui-Quadrado

j	Interval	N_j	np_j	$\frac{(N_j - np_j)^2}{np_j}$
1	[0,0.020)	8	10.950	0.795
2	[0.020,0.042)	11	10.950	0.000
3	[0.042,0.065)	14	10.950	0.850
4	[0.065,0.089)	14	10.950	0.850
5	[0.089,0.115)	16	10.950	2.329
6	[0.115,0.142)	10	10.950	0.082
7	[0.142,0.172)	7	10.950	1.425
8	[0.172,0.204)	5	10.950	3.233
9	[0.204,0.239)	13	10.950	0.384
10	[0.239,0.277)	12	10.950	0.101
11	[0.277,0.319)	7	10.950	1.425
12	[0.319,0.366)	7	10.950	1.425
13	[0.366,0.419)	12	10.950	0.101
14	[0.419,0.480)	10	10.950	0.082
15	[0.480,0.553)	20	10.950	7.480
16	[0.553,0.642)	9	10.950	0.347
17	[0.642,0.757)	11	10.950	0.000
18	[0.757,0.919)	9	10.950	0.347
19	[0.919,1.195)	14	10.950	0.850
20	[1.195,∞)	10	10.950	0.082
				$\chi^2 = 22.188$

75

Tornar o modelo estocástico

Critical points $\chi^2_{\nu,\gamma}$ for the chi-square distribution with ν df

$\gamma = P(Y_\nu \le \chi^2_{\nu,\gamma})$, where Y_ν has a chi-square distribution with ν df; for large ν, use the approximation for $\chi^2_{\nu,\gamma}$ in Sec. 7.4.1

	γ						
ν	0.250	0.500	0.750	0.900	0.950	0.975	0.990
1	0.102	0.455	1.323	2.706	3.841	5.024	6.635
2	0.575	1.386	2.773	4.605	5.991	7.378	9.210
3	1.213	2.366	4.108	6.251	7.815	9.348	11.345
4	1.923	3.357	5.385	7.779	9.488	11.143	13.277
5	2.675	4.351	6.626	9.236	11.070	12.833	15.086
6	3.455	5.348	7.841	10.645	12.592	14.449	16.812
7	4.255	6.346	9.037	12.017	14.067	16.013	18.475
8	5.071	7.344	10.219	13.362	15.507	17.535	20.090
9	5.899	8.343	11.389	14.684	16.919	19.023	21.666
10	6.737	9.342	12.549	15.987	18.307	20.483	23.209
11	7.584	10.341	13.701	17.275	19.675	21.920	24.725
12	8.438	11.340	14.845	18.549	21.026	23.337	26.217
13	9.299	12.340	15.984	19.812	22.362	24.736	27.688
14	10.165	13.339	17.117	21.064	23.685	26.119	29.141
15	11.037	14.339	18.245	22.307	24.996	27.488	30.578
16	11.912	15.338	19.369	23.542	26.296	28.845	32.000
17	12.792	16.338	20.489	24.769	27.587	30.191	33.409
18	13.675	17.338	21.605	25.989	28.869	31.526	34.805
19	14.562	18.338	22.718	$\boxed{27.204}$	30.144	32.852	36.191
20	15.452	19.337	23.828	28.412	31.410	34.170	37.566
21	16.344	20.337	24.935	29.615	32.671	35.479	38.932
22	17.240	21.337	26.039	30.813	33.924	36.781	40.289
23	18.137	22.337	27.141	32.007	35.172	38.076	41.638
24	19.037	23.337	28.241	33.196	36.415	39.364	42.980
25	19.939	24.337	29.339	34.382	37.652	40.646	44.314
26	20.843	25.336	30.435	35.563	38.885	41.923	45.642
27	21.749	26.336	31.528	36.741	40.113	43.195	46.963
28	22.657	27.336	32.620	37.916	41.337	44.461	48.278
29	23.567	28.336	33.711	39.087	42.557	45.722	49.588
30	24.478	29.336	34.800	40.256	43.773	46.979	50.892
40	33.660	39.335	45.616	51.805	55.758	59.342	63.691
50	42.942	49.335	56.334	63.167	67.505	71.420	76.154
75	66.417	74.334	82.858	91.061	96.217	100.839	106.393
100	90.133	99.334	109.141	118.498	124.342	129.561	135.807

76

Tornar o modelo estocástico

- Ajustar os dados a uma distribuição de probabilidades
  - Actividade 3) -> Determinar, se a distribuição seleccionada é de facto representativa dos dados
    - Teste de Qui-quadrado (rejeitar hipótese se $\chi^2 > \chi^2_{k-1,1-\alpha}$)
  - Exemplo
    - Hipótese – tempos entre as chegadas seguem a distribuição exponencial $\widehat{F}(x) = 1 - e^{-x/0.399}, \ \text{para } x \geq 0$
    - Teste de Qui-Quadrado $\chi^2 = 22.188$, com $\chi^2_{19;0.90} = 27.204$
    - Logo, não há, à partida, razão para rejeitar a hipótese!

77

Tornar o modelo estocástico

- Seleccionar distribuição de probabilidade na ausência de dados
  - E se não é possível recolher e analisar dados a partir do sistema real?
  - E se o sistema em estudo não existe?
    - Aproximações heurísticas
      - Identificar o intervalo [a,b], de modo a que P(X<a ou X>b) ≈ 0
      - Aproximação triangular
      - Aproximação pela distribuição beta
    - Conhecimento teórico

78

Modelação e Simulação de Processos

Tornar o Modelo Probabilístico

Paulo Matos

79

Probabilidades e Estatística

Revisões

- Variáveis aleatórias
  - Se X e Y são variáveis aleatórias contínuas, então

$$P(X \in A, Y \in B) = \int_B \int_A f(x,y)dxdy$$

- X e Y são independentes se

$$f(x,y) = f_x(x) f_y(y) \quad \text{,para todo o x, y}$$

- em que

$$f_x(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

$$f_y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

80

Probabilidades e Estatística
Revisões

- Variáveis aleatórias
  - Em simulação é habitual existirem várias variáveis aleatórias em simultâneo…
  - Se X e Y são variáveis aleatórias discretas, então

    p(x,y)=P(X=x, Y=y) para qualquer x, y
    - em que, p(x,y) é a função de probabilidade conjunta de X e Y
    - X e Y são independentes se $p(x,y) = p_x(x)*p_y(y)$, para qualquer x e y
    - em que

$$p_x(x) = \sum_{\text{todo o y}} \mathrm{p(x,y)}$$

$$p_y(y) = \sum_{\text{todo o x}} \mathrm{p(x,y)}$$

81

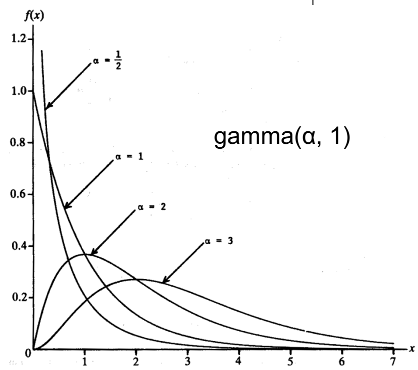
Distribuições de Probabilidade
Contínuas

- Gamma - gamma(α, β)
  - Simular tempo para completar tarefas, por exemplo, atendimento de um cliente, reparação de uma máquina,etc.
  - Densidade $f(x) = \begin{cases} \dfrac{\beta^{-\alpha} x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)} & se\, x > 0 \\ 0 & \text{outrosvalores} \end{cases}$
  - Função gamma $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$
  - Distribuição $F(x) = \begin{cases} 1 - e^{-x/\beta} \sum_{j=0}^{\alpha-1} \dfrac{(x/\beta)^j}{j!} & se\, x > 0 \\ 0 & \text{outros valores} \end{cases}$
  - Parâmetros
    - α > 0 é parâmetro de “forma”
    - β > 0 é factor de escala

82

Distribuições de Probabilidade
Contínuas

- Gamma - gamma(α, β)
  - Gama
    - [0, ∞)
  - Média
    - $\alpha\beta$
  - Variância
    - $\alpha\beta^2$



$$\lim_{x\to 0} f(x) = \begin{cases} \infty & \text{se } \alpha < 1 \\ \dfrac{1}{\beta} & \text{se } \alpha = 1 \\ 0 & \text{se } \alpha > 1 \end{cases}$$

  - Comentários
    - expo(β) e gamma(1,β) representam a mesma distribuição
    - Para ***m*** inteiro positivo, a distribuição gamma(m,β) é chamada distribuição de m-Erlang(β)

83

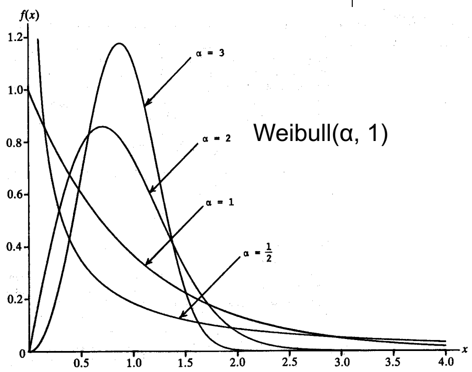
Distribuições de Probabilidade
Contínuas

- Weibull - Weibull(α, β)
  - Simular tempo para completar tarefas, tempos para a falha de equipamento,etc.
  - Densidade $f(x) = \begin{cases} \alpha\beta^{-\alpha}x^{\alpha-1}e^{-(x/\beta)^{\alpha}} & se\ x > 0 \\ 0 & \text{outrosvalores} \end{cases}$
  - Distribuição $F(x) = \begin{cases} 1 - e^{-(x/\beta)^{\alpha}} & se\ x > 0 \\ 0 & \text{outros valores} \end{cases}$
  - Parâmetros
    - α > 0 é parâmetro de “forma”
    - β > 0 é factor de escala

84

Distribuições de Probabilidade
Contínuas

- Weibull - Weibull(α, β)
  - Gama
    - [0, ∞)
  - Média
    - $\frac{\beta}{\alpha}\Gamma\left(\frac{1}{\alpha}\right)$
  - Variância
    - $\frac{\beta}{\alpha}\left\{2\Gamma\left(\frac{2}{\alpha}\right)-\frac{1}{\alpha}\left[\Gamma\left(\frac{1}{\alpha}\right)\right]^2\right\}$



$$\lim_{x\to 0} f(x)=\begin{cases}\infty & \text{se } \alpha<1\\ \frac{1}{\beta} & \text{se } \alpha=1\\ 0 & \text{se } \alpha>1\end{cases}$$

  - Comentários
    - expo(β) e Weibull (1,β) representam a mesma distribuição

85

Distribuições de Probabilidade
Contínuas

- Lognormal - LN(μ, σ²)
  - Simular tempos para desempenhar tarefas; quantidades que traduzem a soma de um grande número de outras quantidades (teorema do Limite Central)
  - Densidade

$$f(x)=\begin{cases}\frac{1}{x\sqrt{2\pi\sigma^2}}\exp\frac{-(\ln x-\mu)^2}{2\sigma^2} & \text{se } x>0\\ 0 & \text{outros}\end{cases}$$

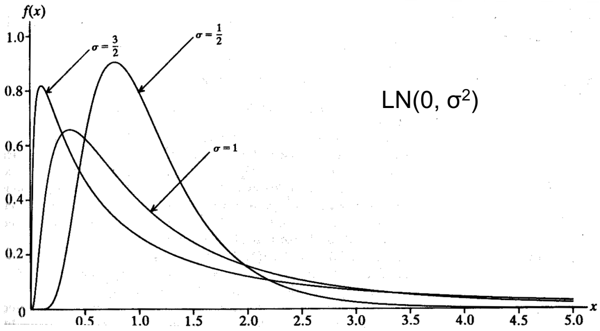
  - Parâmetros
    - σ > 0 é parâmetro de “forma”
    - μ ϵ (-∞, ∞) é factor de escala
  - Gama
    - [0, ∞)

86

Distribuições de Probabilidade
Contínuas

- Lognormal - LN(μ, σ²)
  - Média
    - $e^{\mu+\sigma^2/2}$
  - Variância
    - $e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$
  - MLE

$$\hat{\mu}=\frac{\sum_{i=1}^{n}\ln X_i}{n},\quad \hat{\sigma}=\left[\frac{\sum_{i=1}^{n}(\ln X_i-\hat{\mu})^2}{n}\right]^{1/2}$$



  - Comentários
    - $\lim_{x\to 0} f(x)=0$

87

Tornar o modelo estocástico

- Verificar representatividade dos dados
  - Métodos de *goodness-of-fit* - Testes Qui-Quadrado
    - Teste Estatístico $\chi^2=\sum_{j=1}^{k}\frac{(N_j-np_j)^2}{np_j}$
      - Rejeitar a hipótese se χ^2 for elevado
    - Se todos os parâmetros da distribuição forem conhecidos
      - Rejeitar hipótese, para nível *α*, se $\chi^2>\chi^2_{k-1,1-\alpha}$, em que $\chi^2_{k-1,1-\alpha}$ é o ponto crítico *1- α* para a distribuição qui-quadrado com k-1 df
    - Se *m* parâmetros forem estimados
      $\chi^2_{k-m-1,1-\alpha}\le\chi^2_{1-\alpha}\le\chi^2_{k-1,1-\alpha}$
      - Rejeitar se $\chi^2>\chi^2_{k-1,1-\alpha}$
      - Não rejeitar se $\chi^2<\chi^2_{k-m-1,1-\alpha}$

88

Probabilidades e Estatística

Revisões

- Variáveis aleatórias discretas – Exemplo
  - Uma empresa que vende um único produto, pretende um estudo que permita auxiliar na decisão de saber que quantidades manter em stock para os próximos meses. Os dados conhecidos sobre a procura são :

$$D = \begin{cases} 1, & \text{com probabilidade } 1/6 \\ 2, & \text{com probabilidade } 1/3 \\ 3, & \text{com probabilidade } 1/3 \\ 4, & \text{com probabilidade } 1/6 \end{cases}$$

  - $\mu = 1(1/6) + 2(1/3) + 3(1/3) + 4(1/6) = 5/2$
  - $E(X^2) = 1^2(1/6) + 2^2(1/3) + 3^2(1/3) + 4^2(1/6) = 43/6$
  - $Var(X) = E(X^2) - \mu^2 = 43/6 – (5/2)^2 = 11/12$

89