Tópicos Especiais em Qualidade

Processos estocásticos,
Distribuições de probabilidade e
Ajustamento de dados

Profa. Dra. Liane Werner

Introdução - Processo Estocástico

- Qualquer sistema real opera sempre em ambientes onde a incerteza impera, principalmente
- quando o sistema envolve, ações humanas (imprevisíveis) e/ou avarias de máquinas.
- Os modelos determinísticos certamente contribuem para a compreensão, a um nível básico, do comportamento dinâmico de um sistema.

Introdução - Processo Estocástico

- No entanto, por não poderem lidar com a incerteza, acabam por ser insuficientes nos processos de tomada de decisão.
- Assim, recorre-se a Processos Estocásticos como uma forma de tratar quantitativamente estes fenômenos,
- aproveitando certas características de regularidade que eles apresentam para serem descritos por modelos probabilísticos.

Introdução - Processo Estocástico

- Pode definir-se um Processo Estocástico como um conjunto de variáveis aleatórias
- indexadas a uma variável (geralmente a variável tempo), sendo representado por $\{X(t), t \in T\}$.
- $X=\{X(t_0), X(t_1), X(t_2), ..., X(t_n)\}$
- A variável aleatória $X(t)$ é definida em um espaço denominado de espaço de estado.

Introdução – Processo estocástico

- Estabelecendo o paralelismo com o caso determinístico, onde uma função f(t) toma valores bem definidos ao longo do tempo,
- um processo estocástico toma valores aleatórios ao longo do tempo.
- Aos valores que X(t) pode assumir chamam-se *estados* e ao seu conjunto X *espaço de estados*.

Introdução – Processo estocástico

- Como exemplos de processos estocásticos, pode-se citar:

1) X(t) representa o estado de uma máquina (ligada/desligada) no instante t;
2) X(t) representa o número de máquinas avariadas no fim do dia t;
3) X(t) representa o nível de estoque de um determinado artigo no fim do dia t;
4) X(t) representa a condição de funcionamento de um componente no instante t.

Introdução - Processo Estocástico

- Como pode ser constatado pelos exemplos, há casos em que o tempo é considerado de
- forma discreta (… no dia t) e outros em que é tomado de modo contínuo.
- A variável tempo é, por definição, uma variável contínua, a qual pode ser "discretizada" se os fenômenos forem observados em intervalos regulares.
- Outra constatação que se pode fazer é que,
- os "estados" tanto são valores que a variável X(t) pode assumir (núm. de clientes, núm. de máquinas, etc)
- como são estados (máquina avariada, funcionando, etc).

Mais exemplos...

1. Seja Xt o número de terremotos com magnitude maior que 5 que ocorrem na região de São Francisco no período de (0; t], onde 0 é o início do registro, por exemplo, 0:00 h do dia 01/01/1950. Neste caso temos um processo a tempo contínuo com espaço de estados discreto.
2. Seja (Xk; Yk) o número de nascimento e mortes, respectivamente, ocorridos no dia k em um hospital. Neste caso temos um processo a tempo discreto com espaço de estados discreto.

Mais exemplos...

3. Xt é a intensidade de um sinal a uma distância t da origem. Neste caso temos um processo a tempo contínuo com espaço de estados contínuo. Sendo que a distância é o “tempo”.
4. Temos duas caixas com um total de d bolas numeradas de 1 a d. Em cada experimento selecionamos uma bola ao acaso e a trocamos de caixa. Seja Xt o número de bolas na caixa 1 no instante t. Neste caso temos um processo a tempo discreto com espaço de estados discreto.

Processo Estocástico

Classificação dos Processos Estocásticos:

a) Em relação ao estado:

- Estado discreto (cadeia) – se X (t) é for um conjunto de estados finito ou contável
  (X = {0, 1, 2, …}, ou seja, o conjunto de inteiros não-negativos).
- Estado contínuo (processo) - X(t) caso contrário.

Processo Estocástico

Classificação dos Processos Estocásticos:

b) Em relação ao tempo:

- Tempo discreto – se t é finito ou enumerável {X(t) , t = 0, 1, 2, 3, …}..
- Tempo contínuo – caso contrário. {X(t) , t ≥ 0}.

Processo Estocástico

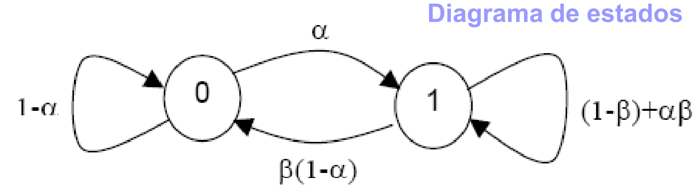
- Processos Estocásticos Estacionários – mantém seu comportamento invariante no tempo (a função distribuição da v.a. que o define não variar no tempo).
- Um processo estocástico diz-se Markoviano ou de Markov se for estacionário e
- gozar da propriedade de Markov ou da “perda de memória”.
- isto é, se o seu comportamento futuro apenas for condicionado pelo estado presente,
- independentemente dos estados visitados no passado (máquina funcionando ou não).

Cadeias de Markov em Tempo Discreto

- Uma Cadeia de Markov em tempo discreto fica completamente definida se conhecermos
- os estados $X = \{0, 1, 2, \ldots, s\}$ e as probabilidades de transição entre os estados em um período.
- Existem duas formas de representação:
- graficamente, através do “diagrama de estados”, ou
- através da matriz quadrada **P** que contém as probabilidades de transição em um período.

Cadeias de Markov em Tempo Discreto

Diagrama de estados



- $P_{00} = 1-\alpha$ O telefone se mantém desocupado
- $P_{01} = \alpha$ Telefone é ocupado com probabilidade α
- $P_{10} = \beta(1-\alpha)$ O telefone se torna livre se termina a chamada e não chega outra chamada
- $P_{11} = (1-\beta)+\alpha\beta$ O telefone continua ocupado com probabilidade (1-β) ou se fica livre e chega outra chamada

Cadeias de Markov em Tempo Discreto

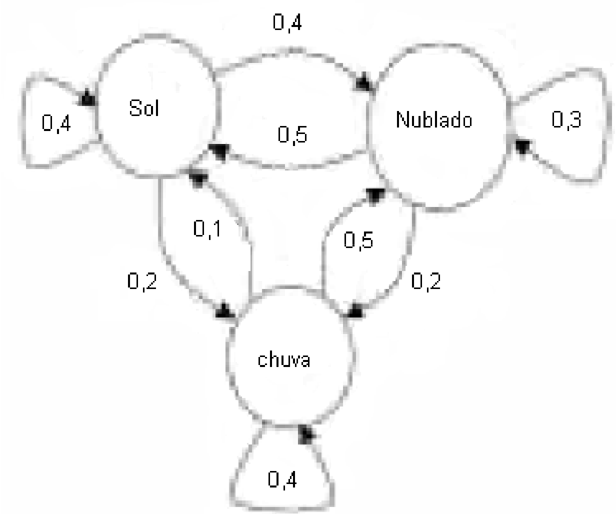
- Matriz P de probabilidade de transição:

$$P = \begin{array}{c|cc} & \text{desocupado} & \text{ocupado} \\ \hline \text{desocupado} & 1-\alpha & \alpha \\ \text{ocupado} & \beta\,(1-\alpha\,) & (1-\beta)+\alpha\beta \end{array}$$

O telefone continua ocupado com probabilidade (1-β) ou se fica livre e chega outra chamada

Exercício 1

- O tempo em uma área é classificado como “sol”, “nublado” ou “chuva” em um determinado dia. X_k é o estado do tempo no dia k, k=1,2,...
- Construa a matriz P de probabilidade de transição.



Exercício 2

- Mulheres adoram compras....
- A matriz ao lado se refere as probabilidades de compras (delas) em 4 lojas da capital.
- Construa o diagrama de estados.

Loja	G	Z	M	R
Gregory	0,40	0,30	0,22	0,03
Zara	0,25	0,60	0,50	0,30
Makenji	0,20	0,55	0,70	0,35
Renner	0,05	0,10	0,18	0,95

Dados fictícios

Cadeias de Markov em Tempo Contínuo

- Uma cadeia de Markov (estados discretos) em que a variável tempo é contínua
- Uma cadeia de Markov em tempo contínuo é um processo estocástico $\{X(t), t \geq 0\}$ com a propriedade de Markov, relembrando,
- a probabilidade de o processo estar no estado j num momento futuro depende
- apenas do estado presente e não dos estados visitados em qualquer momento passado.

Introdução a Simulação estocástica

- Simulação estocástica é a arte de gerar amostras de variáveis aleatórias
- em um ambiente computacional e usar estas amostras para a obtenção de um resultado.
- Ou sejam, " Markov-Chain Monte Carlo" (MCMC) é uma técnica que gera amostras aleatórias provenientes de uma distribuição de probabilidade específica,
- Simulando uma cadeia de Markov cujo o espaço estado inclua as estruturas de interesse.

Introdução a Simulação estocástica

- Mas se o processo estocástico em estudo toma valores aleatórios de uma distribuição de probabilidade f(t) ao longo do tempo.
- É necessário revisar as principais funções de distribuição de probabilidade.
  - Normal
  - Exponencial
  - Weibull
- Entre outras, como: gama, beta, lognormal, etc.

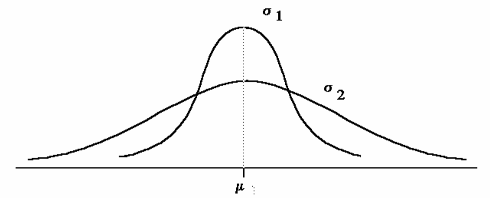
Distribuição de Probabilidade Normal

- *Características*:
  - A curva normal tem forma de sino;
  - É simétrica em relação à média;
  - Prolonga-se de $-\infty$ até $+\infty$;
  - Cada distribuição normal é especificada por seus parâmetros média (μ) que varia de $[-\infty,+\infty]$ e o desvio padrão (σ) que varia entre $[0, +\infty]$.
  - Existe uma curva normal distinta para cada combinação de μ,σ.

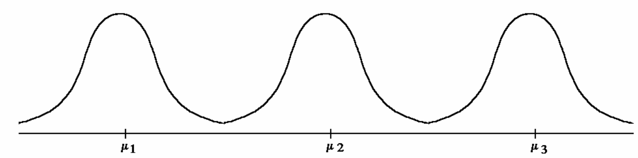
$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\mu)^2 / 2\sigma^2}$$

Distribuição de Probabilidade Normal

Mesmas médias e desvios-padrão diferentes



Mesmo desvio padrão e médias diferentes



Distribuição de Probabilidade Normal

- DISTRIBUIÇÃO NORMAL PADRÃO: é chamada escala z que significa o número de desvios a contar da média.
- As distribuições com $\mu \neq 0$ e/ou $\sigma \neq 1$, podem ser convertidas para a escala Z usando:

$$z = \frac{t-\mu}{\sigma}$$

z expressa a localização de unidades relativo a média usando o desvio padrão.

$P(T < a) = P(z < (a-\mu)/\sigma) = \phi\,(a-\mu)/\sigma)$ e

ϕ (.) é a distribuição normal acumulada e é tabelada.

Distribuição de Probabilidade Exponencial

- Se a probabilidade de um evento ocorrer num pequeno intervalo de tempo for muito pequena,
- mas estatisticamente independente da ocorrência de outros eventos,
- então o intervalo de tempo entre as ocorrências deste evento seguirá uma distribuição exponencial.
- Exemplos: intervalos de tempo entre
  - acidentes numa fábrica,
  - a chegada de pedidos numa empresa,
  - a ocorrência de falha em componentes eletrônicos.

Distribuição de Probabilidade Exponencial

- O modelo da distribuição Exponencial é o seguinte:

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

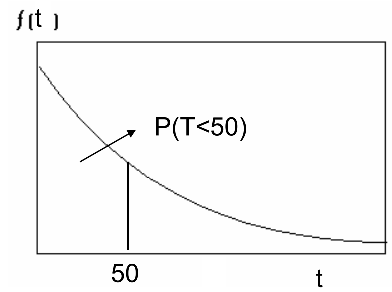
onde $\lambda > 0$ é uma constante (taxa de ocorrência).

- A probabilidade na exponencial pode ser obtida através da função acumulada, isto é:

$$F(a) = P\{T \leq a\} = \int_0^a \lambda e^{-\lambda t} dt = 1 - e^{-\lambda a} \qquad a \geq 0$$

- Sendo que $1/\lambda$ é o valor médio.

Distribuição de Probabilidade Exponencial



Exemplo: a vida útil de certo componente eletrônico é conhecido como exponencialmente distribuído com vida média de 100 horas.

Qual a probabilidade de um componente durar menos que 50 h?

$$P\{T \leq 50\} = F(50) = 1 - e^{-\frac{1}{100}50} = 0{,}393$$

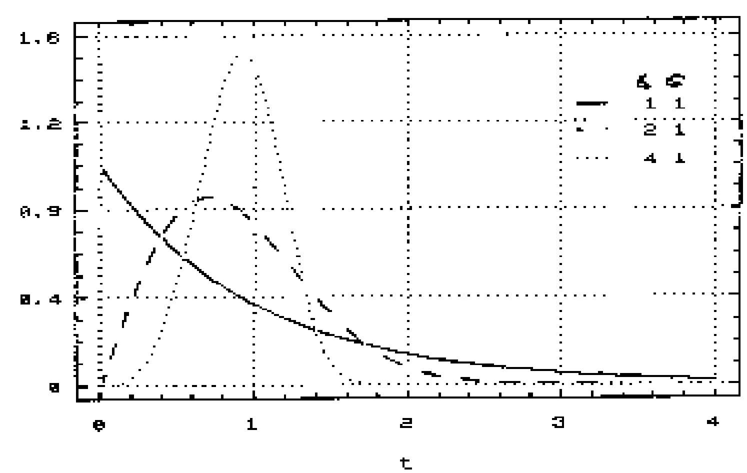
Distribuição de Probabilidade Weibull

- A distribuição de Weibull é um modelo para a distribuição dos mínimos de *n* valores provenientes de diversas distribuições iniciais à esquerda (como a exponencial, por exemplo).
- Exemplificando: Uma corrente formada por vários elos, se a tensão de ruptura dos elos forem exponenciais, então a tensão de ruptura da corrente será uma distribuição Weibull.

Distribuição de Probabilidade Weibull

$$f(t) = \frac{\gamma}{\theta} t^{\gamma - 1} e^{-\frac{t^{\gamma}}{\theta}}$$

onde γ = parâmetro de forma e θ = parâmetro de escala



Ajustamento de dados

- Para que realizar ajustamento no caso da simulação?

⇨ Pois em muitas situações não se conhece o comportamento das v.a. de entrada.

Ajustamento de dados – Verificando o ajustamento

- Como proceder para realizar o ajustamento de dados (*goodness of fit*)?
- Para algumas distribuições pode utilizar o papel de probabilidade, que é um método empírico.
- Aqui será abordado um teste estatístico (não-paramétrico) => teste de hipóteses
- conhecido por Kolmogorov-Smirnov para uma amostra,
- ou simplesmente k-s para uma amostra

Ajustamento de dados – testes k-s para uma amostra

- Este teste permite ao investigador, verificar se os dados observados apresentam o mesmo comportamento de uma distribuição candidata,
- através da comparação das funções distribuições acumuladas.
- Para aplicar este procedimento os dados devem ser contínuos e serem observados em escala, no mínimo, ordinal.

Ajustamento de dados – testes k-s para uma amostra

- Quando o teste k-s para uma amostra é aplicado, tem-se foco em duas distribuições acumuladas:
  - A distribuição acumulada hipotética (candidata)
  - A distribuição acumulada observada
- Para um dado X, F(X) é a probabilidade de que o valor da variável X é menor ou igual a um valor x. Isto é,

$$F(X) = P(X \leq x)$$

- Fo(X) é a distribuição acumulada hipotética.

Ajustamento de dados – testes k-s para uma amostra

- Hipóteses

  Seja Fo(X) a distribuição acumulada hipotética

  - H_0: F(X) = Fo(X) para todos os valores de x
  - H_1: F(X) ≠ Fo(X) para todos os valores de x

- Estatística do teste

Seja S(X) a distribuição de acumulada dos dados observados, isto é

S(X) = proporção de valores observados menores ou iguais a x.

$$S(X) = \frac{\text{número de valores observados menor ou igual a } x}{n}$$

Ajustamento de dados – testes k-s para uma amostra

- A estatística do teste para hipótese bilateral é:

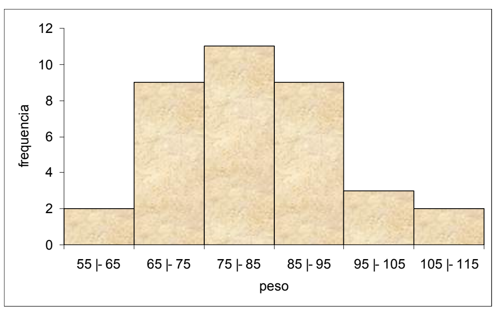
  D = máx | S(X) – Fo(X) | para todos os x.

- Regra de Decisão:
- Rejeita-se H_0 através da comparação do nível de significância α estabelecido e o valor do *p-value*. (*p-value* < α)
- Também pode-se comparar os valores do D calculado com os valores específicos do teste, que são tabelados.

Ajustamento de dados – testes k-s para uma amostra

- Os dados abaixo representam os pesos, em kg, de uma peça usinada.
- Verifique qual o modelo que se ajusta aos dados.

58	70	76	83	90	94
59	70	78	84	90	97
67	70	80	84	90	98
68	70	80	84	90	104
68	74	82	86	92	110
68	75	82	88	93	112



Parece uma normal...

Com média 82,33 kg e desvio padrão 13,09 kg

Ajustamento de dados – testes k-s para uma amostra

- Como tem-se indícios que é uma distribuição normal realiza-se o teste.
- Hipóteses
  - H_0: Apresenta distribuição normal com média 82,33 e desvio 13,09 [F(X) = Fo(X)]
  - H_1: Não apresenta distribuição normal com média 82,33 e desvio 13,09   [F(X) ≠ Fo(X)]

Ajustamento de dados – testes k-s para uma amostra

Estatística do teste: D = máx | S(X) – Fo(X) |

x	freq. obs	freq. acum.	S(x)	x	$z=(x-\mu)/\sigma$	Fo(X)	| S(x) -Fo(x) |
58	1	1	0,0278	58	-1,86	0,0315	0,0038
59	1	2	0,0556	59	-1,78	0,0374	0,0182
67	1	3	0,0833	67	-1,17	0,1208	0,0375
68	3	6	0,1667	68	-1,09	0,1368	0,0298
70	4	10	0,2778	70	-0,94	0,1731	0,1046
74	1	11	0,3056	74	-0,64	0,2623	[illegible]
75	1	12	0,3333	75	-0,56	0,2878	0,0456
76	1	13	0,3611	76	-0,48	0,3144	0,0468
78	1	14	0,3889	[illegible]	[illegible]	[illegible]	0,0185
80	2	16	0,4444	[illegible]	[illegible]	[illegible]	0,0151
82	2	18	0,5000	[illegible]	[illegible]	[illegible]	0,0101
83	1	19	0,5278	[illegible]	[illegible]	[illegible]	0,0074
84	3	22	0,6111	[illegible]	[illegible]	[illegible]	0,0604
86	1	23	0,6389	[illegible]	[illegible]	[illegible]	0,0285
88	1	24	0,6667	[illegible]	[illegible]	[illegible]	0,0009
90	4	28	0,7778	[illegible]	[illegible]	[illegible]	0,0567
92	1	29	0,8056	[illegible]	[illegible]	[illegible]	0,0356
93	1	30	0,8333	[illegible]	[illegible]	[illegible]	0,0409
94	1	31	0,8611	[illegible]	[illegible]	[illegible]	0,0475
97	1	32	0,8889	[illegible]	[illegible]	[illegible]	0,0201
98	1	33	0,9167	[illegible]	[illegible]	[illegible]	0,0323
104	1	34	0,9444	[illegible]	[illegible]	[illegible]	0,0066
110	1	35	0,9722	[illegible]	[illegible]	[illegible]	0,0105
112	1	36	1,0000	[illegible]	[illegible]	[illegible]	0,0117

D

Distribuição acumulada observada

Distribuição acumulada hipotética

Ajustamento de dados – testes k-s para uma amostra

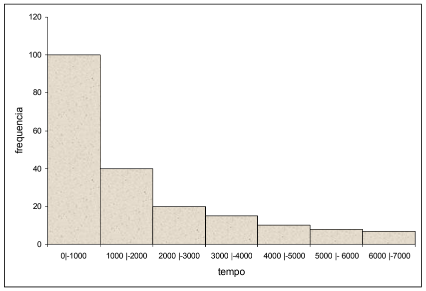
- Tomada de Decisão:
- Rejeita-se H_0 através da comparação do nível de significância α estabelecido e o valor do *p-value*.
- O *p-value* é 0,825 se α=0,05 a decisão a ser tomada é não rejeitar H_0.
- Assim os dados não fornecem evidências suficientes para garantir a conclusão que os pesos não tem distribuição normal.
- De onde passa-se a utilizar o modelo normal como modelo dos dados.

Ajustamento de dados – testes k-s para uma amostra no SPSS (arquivo TEQ2_exemplo.xls)

Ajustamento de dados – exercício de testes k-s para uma amostra no SPSS

- Os dados do tempo de bateria do exemplo, encontram-se no arquivo: TEQ2_tempo baterias.sav
- Verifique se esta variável tem distribuição exponencial ao nível de 5% de significância (α=0,05)

Solução Exercício



NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
tempo	200	1763,93	1690,653	8	6932

One-Sample Kolmogorov-Smirnov Test

		tempo
N		200
Exponential parameter.[a,b]	Mean	1763,93
Most Extreme Differences	Absolute	,068
	Positive	,068
	Negative	-,038
Kolmogorov-Smirnov Z		,956
Asymp. Sig. (2-tailed)		,320

a. Test Distribution is Exponential.

b. Calculated from data.

***p-value*=0,32 > α=0,05 => Não rejeita distribuição exponencial**

$1/\lambda = 1763{,}93 \quad => \lambda = 0{,}00057$

Simulação

- Descreve a situação;
- Verifica as variáveis envolvidas;
- Estabelece as distribuições de probabilidade para as variáveis envolvidas;
- Realiza a simulação;
- Verifica se as variáveis envolvidas se ajustam as distribuições estabelecidas para aumentar a confiança nos resultados obtidos;
- Utiliza os resultados da simulação.

Tópicos Especiais em Qualidade

Processos estocásticos,
Distribuições de probabilidade e
Ajustamento de dados

Profa. Dra. Liane Werner