Modelos de Distribuição de Séries Temporais

Hiago O. B. Batista - 96704
Departamento de Engenharia Elétrica
Universidade Federal de Viçosa
Viçosa, Brasil
hiago.batista@ufv.br

Werikson F. O. Alves - 96708
Departamento de Engenharia Elétrica
Universidade Federal de Viçosa
Viçosa, Brasil
werikson.alves@ufv.br

Dyuliano S. Soares - 2022104711
Departamento de Engenharia Elétrica
Universidade Federal de Viçosa
Viçosa, Brasil
dyuliano.soares@ufv.br

***Resumo*—Uma distribuição estatística é uma função que define uma curva, sendo a área sob esta curva, a probabilidade de ocorrer o evento por ela correlacionado. Dessa forma, este relatório visa implementar uma rotina que gere uma classe com distribuição gaussiana e uma classe com distribuição uniforme tomando como variáveis de entradas os parâmetros de distribuição (média e variância) e a saída como os vetores de abscissa e ordenada, medindo ao final a distorção entre duas séries temporais.**

Palavras-chave: Modelos de distribuição; distribuição gaussiana; distribuição uniforme; séries temporais; distorções entre séries; *MatLab*.

I. Introdução

Uma distribuição estatística é uma função que define uma curva, sendo a área sob esta curva, a probabilidade de ocorrer o evento por ela correlacionado. Sendo assim, um modelo muito conhecido é a distribuição normal ou Gaussiana, a qual pode ser usado para representar o comportamento de diversos processos nas empresas e muitos fenômenos comuns, como, por exemplo, altura ou peso de uma população, a pressão sanguínea de um grupo de pessoas, o tempo que um grupo de estudantes gasta para realizar uma prova [1].

Esta distribuição é conhecida por possuir uma curva simétrica no formato de sino, onde o ápice da curva é o seu ponto médio, o que implica que a média, a mediana e a moda são todas coincidentes, sendo que a média refere-se ao centro da distribuição e o desvio padrão ao espalhamento (ou achatamento) da curva. Este modelo é muito utilizado para a análise de histogramas e serve também como base para a inferência estatística clássica, podendo também ser usada para aproximar distribuições discretas de probabilidade, como, por exemplo, a distribuição binomial [2].

Uma determinada variável aleatória que possui distribuição normal, deve seguir a função de densidade de probabilidade, dada pela Equação (1), em que sua distribuição depende apenas do valor do desvio padrão e do valor da média (μ), sendo μ e σ números reais com $\sigma > 0$. As Figuras 1 e 2 mostram o comportamento dessa distribuição para diferentes valores de média e variância.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-xm}{\sigma}\right)^2} \tag{1}$$

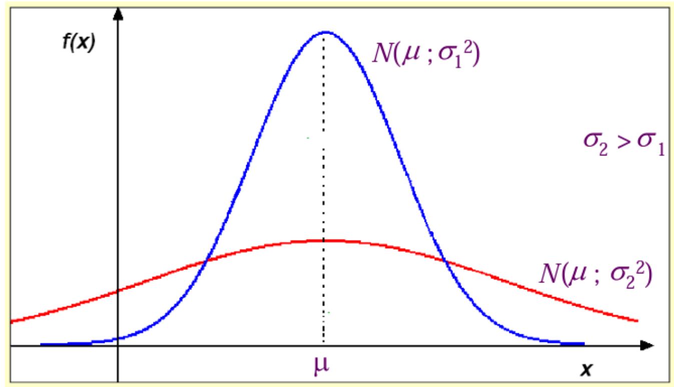

Figura 1: Exemplo de uma distribuição gaussiana para diferentes varianças.

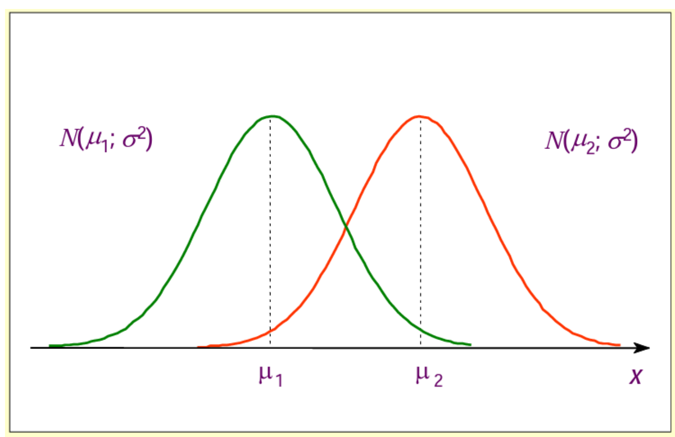

Figura 2: Exemplo de uma distribuição gaussiana para diferentes médias.

Outro modelo de distribuição muito conhecido é a distribuição uniforme, a qual é uma distribuição de probabilidades contínua em que a probabilidade de gerar um ponto em um intervalo contido no espaço amostral é proporcional ao tamanho do intervalo, ou seja, se a probabilidade de X assumir valores num subintervalo é a mesma que para qualquer outro subintervalo de mesmo comprimento, então, esta variável tem distribuição uniforme, sendo X uma variável aleatória contínua que assume valores no intervalo $[a,b]$.

O conceito da distribuição uniforme é importante para problemas de simulação estocástica, onde um valor entre zero e um é sorteado aleatoriamente para amostragem de uma função de distribuição acumulada condicional (método de Monte Carlo). Sendo assim, esta distribuição é usada, muitas vezes, quando não se conhece muitas informações a respeito de uma variável, entretanto, conhece-se seu intervalo de variação.

Uma variável X tem distribuição Uniforme Contínua quando sua função densidade de probabilidade é dada pela Equação (2), onde a e b são os parâmetros da distribuição uniforme, e fora do intervalo, a função densidade de probabilidade é zero. Esta distribuição tem valor médio ou esperança matemática, dada pela Equação (3) e variância dada pela Equação (4). A Figura 3 mostra um exemplo do comportamento dessa distribuição.

$$f(x) = \left\{ \begin{array}{cc} 0 & \texttt{para x < a} \\ \frac{x-a}{b-a} & \texttt{para a < x < b} \\ 0 & \texttt{para x > b} \end{array} \right\} \quad (2)$$

$$E(X) = \frac{a+b}{2} \quad (3)$$

$$Var(X) = \frac{(b-a)^2}{12} \quad (4)$$

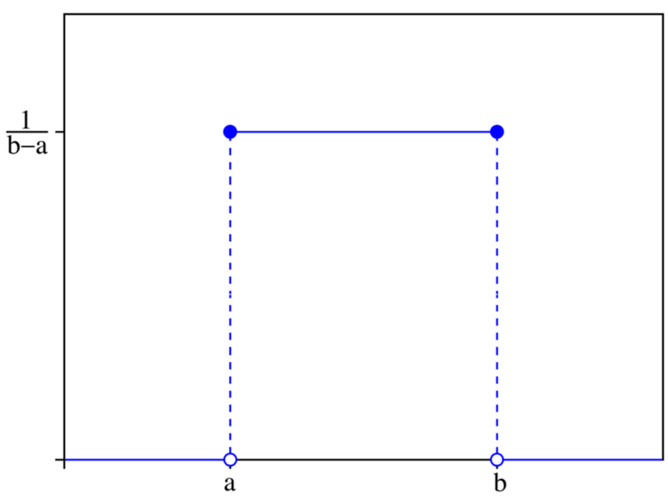

Figura 3: Exemplo de uma Distribuição uniforme.

A partir disto, pode-se definir que séries temporais são um conjunto de observações de uma variável de interesse tomadas ao longo do tempo. E ao realizar a correlação que uma observação da série apresenta com relação às anteriores, é possível criar modelos e gerar previsões para os próximos valores que ela pode assumir. Para isto, é necessário o estudo do comportamento da série a fim de entender e determinar as suas principais características, realizando a modelagem mais adequada. Sendo assim, outra análise que se pode obter é sobre a distorção entre duas séries. Esta distorção pode ser calculada através da Equação 5, sendo o resultado obtido dado em dB.

$$d_{xy} = \log_{10}\left(\frac{E[(X(k)-Y(k))^2]}{E[X^2(k)]}\right) \quad (5)$$

II. Objetivos Gerais e Específicos

O objetivo desse trabalho consiste na implementação de uma rotina que gere uma classe com distribuição gaussiana e uma classe com distribuição uniforme tomando como variáveis de entradas os parâmetros de distribuição (média e variância) e a saída como os vetores de abscissa e ordenada. Além disso, este trabalho teve como objetivo, também, medir a distorção entre duas séries temporais, para isto, foi criada uma função baseada na Equação (5), sendo o valor de saída medido em decibéis (dB).

III. Materiais e Métodos

Com base nas informações apresentadas anteriormente, foram implementadas uma rotina e duas funções no MatLab, sendo a rotina para entrar com os parâmetros de entrada, e as funções para criar as classes e para calcular as distorções. Para isto, foram utilizadas algumas funções já existentes no *MatLab*, sendo estas apresentadas na Tabela I.

Tabela I: Comandos uteis utilizados.

Comandos	**Descrição**
Randn	Gera variáveis aleatórias com distribuição normal com média e desvio padrão, aproximadamente, nulos.
Rand	Gera variáveis aleatórias com distribuição uniforme com média e desvio padrão, aproximadamente, nulos.
Mean	Retorna a média de uma variável aleatória.
Sqrt	Retorna a raiz quadrada de do elemento.
Log10	Retorna o logaritmo comum do elemento.
Scatter	Cria um gráfico de dispersão com marcadores circulares nas coordenadas especificadas.

Conforme citado anteriormente, a rotina desenvolvida visa definir os parâmetros de entrada da função que irá criar as classes e calcular a distorção, tendo como parâmetros de entrada a média e a variância desejados, além da quantidade de dados a serem gerados para esta classe. O algoritmo 1 apresenta a estrutura da rotina desenvolvida.

A. Distribuição gaussiana

Para gerar a classe com distribuição gaussiana, foram gerados vários pontos aleatórios por meio da equação $y = ax + b$, onde x é gerado aleatoriamente através da função *randn()*, a é a raiz quadrada da variança e b é a média da classe. Com isto, utilizando esta equação para encontrar as coordenadas X e Y que irão compor a classe, é possível montar uma rotina que irá retornar uma classe gaussiana, conforme apresentado no Algoritmo 2, onde MX e MY representam as médias, VX e VY representam as varianças para cada uma das variáveis calculadas, obtendo assim, as coordenadas X e Y da classe. Ao final é gerado o gráfico contendo estes pontos e em seguida, calculada a distorção destas informações, segundo a Equação (5).

Algorithm 1 Código Principal.

Result: Gera o gráfico para as duas classes e calcula a distorção entre as series.

Parâmetros de entrada:

$Tipo$ ▷ Qual a classe desejada

MX ▷ Media desejada de X.

MY ▷ Media desejada de Y.

VX ▷ Variança desejada de X.

VY ▷ Variança desejada de Y.

QP ▷ Quantidade de pontos desejada

$Distribuicao(\text{"}Tipo\text{"}, MX, MY, VX, VY, QP);$

$Distribuicao(\text{"}Tipo\text{"}, MX, MY, VX, VY, QP);$

B. Distribuição uniforme

De forma semelhante, para a distribuição uniforme, foram gerados vários pontos aleatórios por meio da mesma equação $y = ax + b$, onde x, neste caso, é gerado aleatoriamente através da função *rand()*. Para rearranjar os valores obtidos pela função, foi realizada a multiplicação dos valores obtidos pela faixa de valores desejada, definida por $(b - a)$ e deslocando os valores por a, sendo a e b, os limites inferior e superior da faixa de valores. Estes valores são obtidos a partir das Equações (3) e (4).

Com isto, de forma semelhante à distribuição gaussiana, foi implementado uma função que irá retornar uma classe uniforme, conforme apresentado no Algoritmo 2, onde MX e MY representam as médias, VX e VY representam as varianças para cada uma das variáveis obtidas, obtendo assim as coordenadas X e Y da classe. Ao final foi gerado o gráfico contendo estes pontos e em seguida foi calculada a distorção destas informações, segundo a Equação (5), também.

C. Distorção entre séries temporais

Para realizar a distorção entre duas séries temporais foi implementada uma função baseada na Eq. 5, a qual pode ser vista no Algoritmo 3.

IV. Resultados e Discussões

Para obter os resultados, foram utilizados os valores de média, variância e número de amostras mostrados na Tabela II. A partir do que foi exposto anteriormente, foram gerados as duas classes, as quais são apresentadas nas Figuras 4 e 5.

Parâmetros	**Valores**
Tipo	"Gaussiana"ou "Uniforme"
MX	10
MY	20
VX	1
VY	2
QP	1000

Tabela II: Parâmetros utilizados

Algorithm 2 Distribuição Gaussiana e uniforme

Input: $(Tipo, MX, MY, VX, VY, QP)$ ▷ Parâmetros de entrada

Output: $Figura$ ▷ Gera o gráfico da classe gaussiana ou uniforme.

Function `Distribuicao`$(Tipo, MX, MY, VX, VY, QP)$**:**

- $Pontos = zeros(2, QP);$
- **if** $Tipo == \text{"}Gaussiana\text{"}$ **then**
  - **for** $i = 1 : QP$ **do**
    - $P1 = sqrt(VX) * randn() + MX;$
    - $P2 = sqrt(VY) * randn() + MY;$
    - $Pontos(:, i) = [P1; \quad P2];$
  - **end**
  - $scatter(Pontos(1, :), Pontos(2, :));$
  - $DistorcaodB(Pontos(1, :), Pontos(2, :))$
- **end**
- **if** $Tipo == \text{"}Uniforme\text{"}$ **then**
  - $X1 = (2 * MX - sqrt(12 * VX))/2;$ ▷ Coeficiente de menor valor da ditribuição de X
  - $X2 = (2 * MX + sqrt(12 * VX))/2;$ ▷ Coeficiente de maior valor da ditribuição de X
  - $Y1 = (2 * MY - sqrt(12 * VY))/2;$ ▷ Coeficiente de menor valor da ditribuição de Y
  - $Y2 = (2 * MY + sqrt(12 * VY))/2;$ ▷ Coeficiente de maior valor da ditribuição de Y
  - **for** $i = 1 : QP$ **do**
    - $P1 = X1 + (X2 - X1) * rand();$
    - $P2 = Y1 + (Y2 - Y1) * rand()$
    - $Pontos(:, i) = [P1; \quad P2];$
  - **end**
  - $scatter(Pontos(1, :), Pontos(2, :));$
  - $DistorcaodB(Pontos(1, :), Pontos(2, :))$
- **end**

End Function

Algorithm 3 Cálculo da distorção temporal

Input: (x, y) ▷ Parâmetros de entrada

Output: Dxy ▷ Valor da distorção da série temporal.

Function `Distorcao_dB`(x, y)**:**

- $Dxy = 10 \log_{10} \left(\frac{E[(X(k) - Y(k))^2]}{E[X^2(k)]} \right);$
- **return** Dxy

End Function

A. Distribuição gaussiana

A Figura 4 ilustra a distribuição gaussiana com valores médios de 10 e 20 na abscissa e ordenada, respectivamente. Perceba que nesta distribuição os pontos estão mais localizados próximo à média e sua dispersão é conforme a variância.

B. Distribuição uniforme

A Figura 5 mostra a distribuição uniforme com valores médios também de 10 e 20. Porém, neste caso perceba que os pontos não estão tão próximos da média, mas sim conforme os limites dados de a e b, que neste caso foram de [5,15] à abscissa e [15,25] para a ordenada. Perceba que

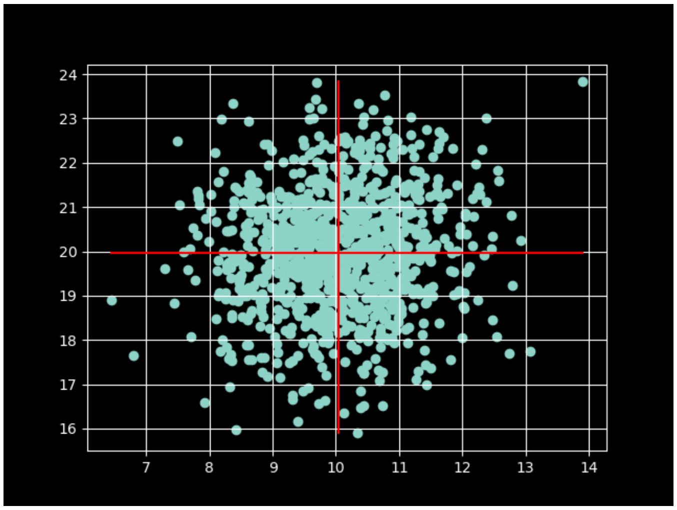

Figura 4: Classe Gaussiana.

diferentemente do caso da classe gaussiana, os pontos para essa distribuição são uniformemente espaçados.

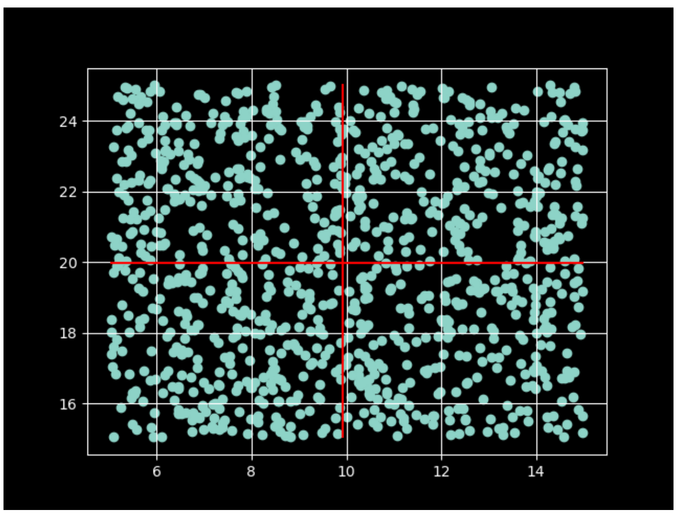

Figura 5: Classe uniforme.

C. Distorção entre séries temporais

A distorção da classe gaussiana foi avaliado em -0.028 enquanto da classe uniforme foi de 0.421. Isto significa que series temporais com distribuição gaussiana com valores médios próximos, variam menos que a classe uniforme, desde que o desvio padrão não seja alto.

V. Considerações Finais

Esse trabalho apresentou as etapas do processo de geração de distribuições uniformes e gaussianas, bem como o cálculo de distorção entre sinais. Os resultados obtidos para as distribuições, uniformes e gaussianas, correspondem com o esperado para o perfil de cada uma delas.

No caso da gaussiana, foi possível notar que os valores ficam mais focados em torno de seu valor médio, sendo que a uniforme é melhor distribuída dentro do intervalo. Em suma, a distribuição uniforme apresenta densidade de pontos quase constante em seu intervalo, sendo que a gaussiana tem maior densidade em locais próximos à média.

Referências

[1] Zsolt László Kovács. *Redes neurais artificiais.* Editora Livraria da Fisica, 2002.

[2] Leandro FLECK, Maria Hermínia Ferreira Tavares, Eduardo Eyng, AC Helmann, and MA de M Andrade. Redes neurais artificiais: Princípios básicos. *Revista Eletrônica Científica Inovação e Tecnologia*, 1(13):47–57, 2016.