

Probabilidades

Advanced Institute for Artificial Intelligence

<https://advancedinstitute.ai>

Agenda

- ☐ Dispersão de Dados
- ☐ Introdução a Probabilidade
- ☐ Distribuição de Probabilidade
- ☐ Exemplos de Distribuição de Probabilidade
- ☐ Teste de Hipótese

Estimativa de densidade por Kernel (KDE - Kernel Density Estimate)

- ☐ forma não-paramétrica para estimar a Função densidade de probabilidade de uma variável aleatória.
- ☐ Possui a propriedade de estimar de forma continua de acordo com um kernel adequado (curva normal por exemplo)
- ☐ Opção de visualização a histogramas

Tendência central

- Média (μ): soma dos valores dividido pela quantidade de dados
- Mediana: valor situado no meio da amostra, quando a amostra está ordenada
- Moda: valor mais frequente na amostra

Dispersão

- Desvio médio: módulo da média aritmética dos desvios de cada elemento da série para a média da série
- Variância: mesmo conceito do desvio médio trocando módulo por elevar a diferença ao quadrado
- Desvio Padrão (σ): é representado como a raiz quadrada da variância

Interpretação da Dispersão

- ☐ Quanto mais uniforme forem os valores, mais próximo de zero estará o desvio padrão.
- ☐ Quando todos valores são iguais o desvio padrão é zero. Assim a amostra é perfeitamente uniforme.
- ☐ Quando estamos interessados em saber qual conjunto de valores possui uma maior regularidade podemos usar tanto a variância, como o desvio padrão.
- ☐ O desvio padrão é expresso na mesma unidade de medida das variáveis do conjunto.

Variância

- Variância simples: soma do desvio quadrado de cada valor em relação a média dividido por $n-1$ (população -1)
- Desvio padrão simples: raiz quadrada da variância simples
- Quando utilizamos a população completa é comum utilizar a população -1 para ajuste estatístico

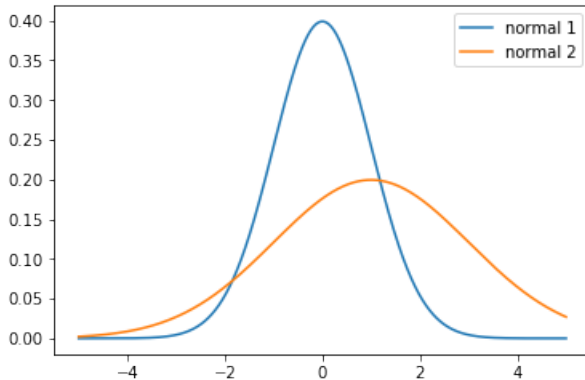
Distribuição

- Uma distribuição define uma curva (gráfico), e a área sob a curva define a probabilidade de acontecer um evento relacionado à distribuição.
- Matemáticos definiram muitas distribuições "famosas", que representam fenômenos importantes do mundo real.

Distribuição Normal

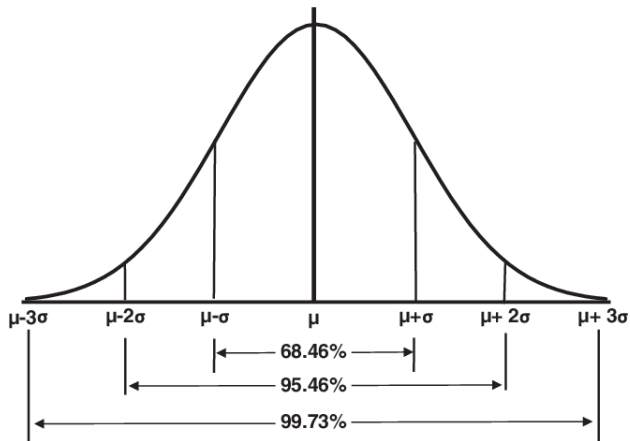
- ☐ É uma das mais importantes distribuições de probabilidade que caracteriza muitos fenômenos aleatórios
 - Fenômenos naturais
 - Altura
 - pressão sanguínea
- ☐ Desempenham papel importante nos métodos de inferência estatística
- ☐ A distribuição normal é uma variável aleatória contínua tem uma distribuição em forma de sino.

Distribuições Normais com diferentes valores de média e desvio padrão



Distribuições paramétricas

- Supondo que os dados seguem uma distribuição, os parâmetros da distribuição permitem calcular a probabilidade de um fenômeno (no caso da distribuição normal, os parâmetros são média e desvio padrão)
 - $\mu + \sigma$ e $\mu - \sigma$: 68%
 - $\mu + 2\sigma$ e $\mu - 2\sigma$: 95%
 - $\mu + 3\sigma$ e $\mu - 3\sigma$: 99%
- z-score: representa a distância entre uma dada medida e a média em termos de desvio padrão



Correlação

- Utiliza-se a covariância e o desvio padrão como base para definir métricas de correlação
 - A correlação perto de -1 é uma anti-correlação perfeita
 - A correlação perto de 0 indica que não há correlação
 - A correlação perto de 1 é uma correlação perfeita

Probabilidade

- Pode ser usada como uma maneira de quantificar a incerteza associada a eventos escolhidos em um universo de eventos
 - Exemplo: rolar um dado
- O universo consiste em todos os resultados possíveis. qualquer subconjunto desses resultados é um evento
 - Dado: universo números de 1 a 6
 - exemplo de evento: dado retorna 1, dado retorna número maior que 3, jogando 10 vezes, sempre retorna um número maior que 2

Eventos

- $P(E)$ probabilidade do evento E
- Quando analisamos probabilidade de dois ou mais eventos, consideramos que tais eventos podem ser independentes ou dependentes
- Eventos independentes
 - dois eventos E e F são independentes se a probabilidade de ambos acontecerem é o produto das probabilidades de que cada um aconteça
 - $P(E, F) = P(E) * P(F)$
- Eventos dependentes
 - $P(E|F) = P(E, F) / P(F)$ probabilidade do evento E ocorrer caso F ocorra

Nascimento de duas crianças (pode ocorrer menino ou menina)

G	G
B	B
G	B
B	G

- ☐ Probabilidade da mais velha ser meninas: $2/4$
- ☐ Probabilidade de duas meninas: $1/4$
- ☐ Probabilidade de uma das duas ser menina: $3/4$

Variáveis Aleatórias

- ☐ Para uma variável (Contínua ou Discreta), uma distribuição de probabilidades determina as propriedades de cada evento
- ☐ A soma de todas as probabilidades de todos os eventos possíveis deve ser igual a 1
- ☐ Se repetirmos um experimento com uma distribuição de probabilidade associada várias vezes, a probabilidade encontrada seguirá a distribuição de de uma variável aleatória
- ☐ Essa distribuição serve então como modelo padrão para uma população

A distribuição de probabilidade pode ser representada também usando fórmulas

- Função de densidade de probabilidade (PDF)

- Para cada valor a função retorna uma probabilidade associada a esse valor

- Função de distribuição acumulada

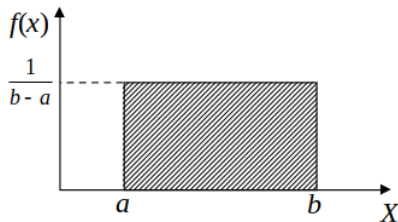
- Descreve como probabilidades são associadas a intervalos de valores de uma variável aleatória.
- Calculada como a integral da função de densidade de probabilidade

- Diversos tipos de Distribuições de probabilidades possuem características adequadas para modelar fenômenos específicos
 - Binomial: distribuições binárias, probabilidade de algo ocorrer ou não ocorrer
 - Uniforme: eventos possuem probabilidades estatisticamente muito próximas
 - Poisson: modela número de eventos que podem ocorrer em um período de tempo ou área

- As distribuições conhecidas e formalizadas possuem fórmulas para gerar a distribuição de probabilidade
- As bibliotecas numpy e scipy por exemplo, permitem gerar tais distribuições com base em parâmetros iniciais
- Tais funções permitem reproduzir fenômenos ou recriar cenários quando a distribuição de probabilidades é conhecida

Probabilidades

- Distribuição Uniforme
- Uma variável aleatória X que distribuição Uniforme no intervalo $[a, b]$
- PDF da função abaixo
- CDF da função abaixo: é a probabilidade acumulativa
 - probabilidade de dar um valor próximo de b é maior do que entre valor no meio do intervalo $[a, b]$, pois a área é maior



Teste de Hipótese

- ☐ Hipótese é uma assertiva que podemos provar ou refutar
- ☐ Refutar normalmente leva a um avanço do conhecimento
- ☐ Em estatística tenta-se por padrão provar que uma assertiva está errada, para concluir alguma coisa a partir de levantamento estatísticos

Teste de Hipótese

- Hipótese estatística: suposição quanto ao valor de um parâmetro populacional, ou quanto à natureza da distribuição de probabilidade de uma variável populacional.
- teste de Hipótese: É uma regra de decisão para aceitar ou rejeitar uma hipótese estatística com base nos elementos amostrais

Tipos de Hipótese

- ☐ hipótese nula H_0 é hipótese estatística a ser testada
- ☐ H_1 é a hipótese alternativa.
- ☐ H_0 é uma assertiva de como a realidade deveria ser, se nossa suposição estivesse errada
- ☐ A hipótese nula expressa uma igualdade
- ☐ hipótese alternativa é dada por uma desigualdade

- ❑ O teste de hipótese retorna um valor de significância chamado P-valor (*P-value*)
- ❑ O P-valor é definido como a probabilidade de obter um resultado igual ou mais extremo do que realmente foi observado.
- ❑ Quanto menor o p-valor, maior a significância, pois indica a investigador de que a hipótese em consideração pode não explicar adequadamente a observação.
- ❑ P-valor representa o critério para rejeitar ou não-rejeitar a hipótese nula
- ❑ Um valor padrão para P-valor é 0.05