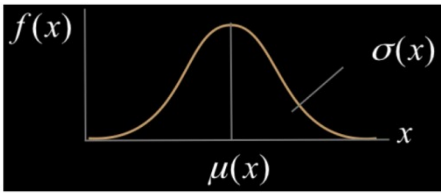
Conceitos iniciais para entendimento da Aula 6 de ECD – Inferência Bayesiana

A inferência é um conjunto de técnicas e ferramentas que permitem extrair informações sobre uma população a partir dos resultados obtidos com uma amostra (que seja representativa). As conclusões são acompanhadas de incerteza (grau de confiabilidade).

Parâmetro: medida usada para descrever numericamente uma característica da população.

Estimador: valor numérico sobre um determinado parâmetro obtido a partir de uma amostra. Também conhecido como **estatística de um parâmetro populacional**. O estimador é uma variável aleatória, ou seja, é uma função que associa a todo evento, em um espaço amostral, um único número real.



CUIDADO: nós estamos habituados com o termo variável como sendo uma estrutura que armazena um ou mais valores e a função como uma estrutura que retorna um valor após operar em um conjunto de valores. Conceitualmente a variável aleatória é diferente!

Distribuição da população:

Distribuição da amostra: Para cada uma das possíveis amostras, calcula-se um valor de estimador (média, variância, desvio padrão,...). O resultado é uma distribuição amostral desse estimador. Existem várias distribuições de probabilidades e elas podem ser associadas a eventos cotidianos. Exemplo:

- Distribuição de Poisson: usamos quando temos uma taxa associada: pessoa por hora, pacotes por segundo, ônibus por dia, contagem de ocorrências/taxa, etc.
- Distribuição Binomial: associada a quando temos uma probabilidade fixa de sucesso (p) e fracasso (1-p) e com eventos independentes (com reposição). A distribuição Binomial é uma generalização da distribuição de Bernoulli.
- Distribuição Uniforme: o valor da variável aleatória é o mesmo em um certo intervalo.
- Distribuição Exponencial:
- Distribuição t- Student: importante para inferência, pois usamos quando a amostra é pequena.
- Distribuição Chi-Quadrado: usada para testes de hipóteses e seleção de atributos em aprendizagem de Máquinas.

A inferência clássica (ou frequentista) é baseada na probabilidade clássica, onde a probabilidade de um evento ocorrer é a probabilidade é igual o número de vezes que o evento ocorreu (sua freqüência)/ dividido pelo número total de eventos.

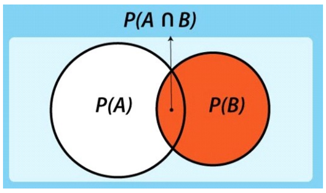
A inferência Bayesiana é baseada no **Teorema de Bayes** e permite que incorporemos um conhecimento que temos (ou um especialista ou alguma pessoa) já tenha do assunto (conhecimento à priori) que consideramos as suas distribuições (afinal é uma variável aleatória?), incluindo as informações contidas na base de dados para encontrar um valor da probabilidade à posteriori. Fazemos isso repetidas vezes (num ciclo) onde a nova probabilidade à priori é considerada a prioridade à posteriori obtida na iteração anterior.

O Teorema de Bayes é baseado:

- Na probabilidade condicional:

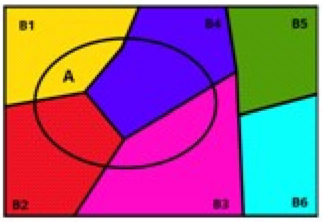
$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Dado que B já aconteceu, a única possibilidade de ocorrer A está na intersecção de A com B.



- No Teorema da Probabilidade Total: particionamos o espaço amostral considerando as partições como eventos mutuamente exclusivos (disjuntos) e concluímos que a probabilidade de ocorrer A é dada pelo somatório de todas as intersecções de A com cada peça da partição. Unindo com o conhecimento da probabilidade condicional, concluímos que:

$$\boldsymbol{P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots + P(B_r)P(A|B_r)}$$



Então, pelo Teorema de Bayes, temos:

Verossimilhança　Probabilidade à priori

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Probabilidade à posteriori　k: constante normalizadora

A probabilidade de um evento A ocorrer, dado que B já ocorreu (probabilidade à posteriori) é igual a probabilidade de acontecer A dado que B já ocorreu (Verossimilhança = distribuição conjunta de A e B – obtida com o conjunto de dados) vezes a probabilidade de ocorrer A (informação à priori ou distribuição marginal de A), dividido por P(B) (k: constante normalizadora que não depende de A – que pode ser omitida, pois quando a integramos ela se torna um valor numérico e que pode ser recuperada usando as informações do problema).

Resumidamente, a Inferência Clássica faz suposições do modelo e estima os parâmetros (que são fixos e desconhecidos) através dos dados observados. Já a Inferência Bayesiana modela a incerteza que existe nos parâmetros com distribuições de probabilidade (distribuição dos dados à priori) acrescentando as informações obtidas com o conjunto de dados (Verossimilhança) para obter a probabilidade à posteriori (é uma atualização do conhecimento).

Podemos usar uma aproximação:

Distribuição à posteriori α Verossimilhança (dados) x distribuição à priori (conh. subjetivo)

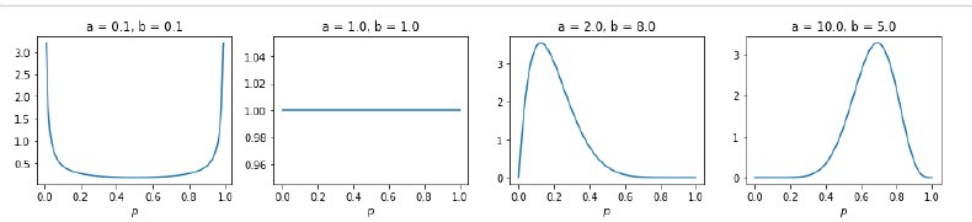
Obs: o símbolo α indica que é proporcional

As distribuições que podem ser assumidas para os dados à priori podem ser não informativas, ou seja, não adicionam informações para a obtenção dos novos parâmetros (como é o caso da distribuição uniforme, por exemplo), pode ser uma priori informativa (pensamos em uma distribuição que se adequa ao assunto e o pesquisador dá um chute inicial nos parâmetros) ou podemos pensar na priori conjugada (a priori e a posteriori terão a mesma distribuição e só os parâmetros serão diferentes). A vantagem de usar priori conjugada estão no ganho computacional.

Podemos usar para o conhecimento à priori a distribuição Beta para proporções, a Normal para médias, a Gama para taxa de falhas e a Dirichlet para uma variável aleatória categórica. Os núcleos das Verossimilhanças (distribuição dos dados) será Bernoulli, Normal, Poisson e Dirichlet, respectivamente. Usando a priori conjugada, podemos dizer que a posteriori terá a mesma distribuição da priori.

Priori	**Núcleo da Verossimilhança**	**Posteriori**
Beta	Bernoulli	Beta
Normal	Normal	**Normal**
Gama	Poisson	Gama
Dirichlet	Multinomial	Dirichlet

Para a distribuição Beta (para proporções), se consideramos os diferentes Parâmetros, teremos diferentes formatos da curva, portanto, se quisermos "dar um chute inicial" nos parâmetros, devemos considerar qual se adéque melhor a situação problema. Por exemplo, para a taxa de clientes inadimplentes de um banco, esperamos um número pequeno – poderíamos usar a distribuição Beta com parâmetros a=2 e b=8 ou valores próximos.



Quanto maior o número de amostras, menor é a importância da priori para atualizar o conhecimento (obter a nova posteriori iterativamente a partir da priori).

É importante ressaltar que para a Inferência Clássica usamos o Intervalo de Confiança (obtido com valores da amostra). Para um intervalo de confiança de 95% podemos afirmar que se fizermos um experimento 100 vezes, por exemplo, em 95 vezes os valores estarão dentro desse intervalo (considerando os valores fixos dos parâmetros que estimamos). Para a Inferência Bayesiana, usamos o intervalo de Credibilidade onde os parâmetros são variáveis aleatórias (valor obtido de uma distribuição de probabilidade) ele é obtido com dados amostrais e o conhecimento a priori.