

Probabilidade Bayesiana

Flávio Luiz Seixas¹

Instituto de Computação
fseixas@ic.uff.br, <http://www.ic.uff.br/~fseixas>

1 Paradigmas Frequentista e Bayesiano

O paradigma frequentista admite a probabilidade num contexto restrito a fenômenos que podem ser medidos por frequências relativas. O paradigma Bayesiano entende-se que a probabilidade é uma medida racional e condicional de incerteza. Uma medida do grau de plausibilidade de proposições quaisquer, as quais não precisam necessariamente estar associadas a fenômenos medidos por frequência relativa. Por exemplo, no paradigma Bayesiano admite-se falar da probabilidade de extinção de uma espécie, o que não seria admissível sob o paradigma frequentista.

A inferência estatística é o processo formal utilizado para fazer afirmações genéricas com base em informações parciais. Essas afirmações são probabilísticas pois se caracterizam por incluir componentes de incerteza.

Na perspectiva bayesiana, a inferência estatística sobre qualquer quantidade de interesse é descrita como a modificação que se processa nas incertezas à luz de novas evidências.

A conceituação frequentista admite falar em probabilidades somente no contexto de frequências relativas. Em contraste, na conceituação bayesiana, probabilidades quantificam as plausibilidades de proposições ou eventos. Ao atribuir plausibilidades diferenciadas a proposições, a formalização bayesiana de probabilidade estende a lógica dedutiva, restrita a classificar proposições em verdadeiras (probabilidade igual a 1) ou falsas (probabilidade igual a zero), para um conjunto de possibilidades entre estes dois extremos.

O rápido crescimento do uso do paradigma bayesiano em ciências aplicadas ao longo das últimas décadas foi facilitado pelo surgimento de vários programas para efetuar as computações estatísticas. Entre esses, destaca-se o R (programa de livre distribuições e de código aberto).

2 As Regras de Probabilidade

A probabilidade será um número real e função de dois argumentos: o evento incerto E e a premissa H . Utilizaremos o símbolo $Pr(E|H)$ lido como probabilidade de E dado que H é fato, ou a probabilidade de E condicionada ao fato H .

A Lei da convexidade

A probabilidade de um evento qualquer E , condicionado a H é um número real no intervalo $[0, 1]$

$$0 < Pr(E|H) < 1 \quad (1)$$

A Lei da adição

Se E_1 e E_2 são eventos exclusivos sob H , então a probabilidade da união lógica de $E_1 + E_2$ é igual a soma aritmética das suas probabilidades individuais condicionadas a H .

$$Pr(E_1 + E_2|H) = Pr(E_1|H) + Pr(E_2|H) \quad (2)$$

A Lei do produto

Se E_1 e E_2 são eventos quaisquer então a probabilidade do produto lógico $E_1 E_2$ condicionado a H é o produto da probabilidade de E_1 condicionado a H multiplicado pela probabilidade de E_2 condicionado a $E_1 H$

$$Pr(E_1 E_2|H) = Pr(E_1|H) \cdot Pr(E_2|E_1 H) \quad (3)$$

Nos casos em que estamos tratando de eventos independentes, a lei do produto pode ser reescrita como:

$$Pr(E_1 E_2|H) = Pr(E_1|H) \cdot Pr(E_2|H) \quad (4)$$

3 O Teorema de Bayes

Mutas propriedades do cálculo de probabilidades podem ser deduzidas a partir das três leis básicas indicadas na seção anterior. Depois teoremas adicionais merecem especial destaque, o Teorema da Probabilidade Total e o Teorema de Bayes.

Teorema da Probabilidade Total

Seja $E_j; j = 1, \dots, m$ um conjunto de m eventos exclusivos e exaustivos sob H , e seja A outro evento qualquer. Então $Pr(A|H)$ pode ser reescrito estendendo a conversa para a inclusão dos E_j .

$$Pr(A|H) = \sum_{j=1}^m Pr(A|E_j H) \cdot Pr(E_j|H) \quad (5)$$

Teorema de Bayes

Sejam E e F dois eventos quaisquer e $Pr(E|H) > 0$, então:

$$Pr(F|EH) = \frac{Pr(E|FH) \cdot Pr(F|H)}{Pr(E|H)} \quad (6)$$

3.1 Exemplo

Um estudo de uma mamografia no diagnóstico de câncer é apresentado na Tabela 1. Os dados foram obtidos experimentalmente sobre a efetividade do exame na detecção de um tumor de mama maligno ou benigno. Por exemplo, se um tumor é maligno Ca , a probabilidade de que o exame resulte positivo é $Pr(Pos|Ca) = 0,792$, ou seja, 79,2%. De forma similar temos $Pr(Neg|Ca') = 0,904$ como a probabilidade de que o exame resulte negativo se o tumor não é maligno (Ca'). Os percentuais para falsos positivos e falsos negativos são 9,6% e 20,8%, respectivamente.

Table 1. Resultados dos testes de câncer de mamas

Resultado do teste	Realidade	
	Ca (Tumor maligno)	Ca' (Tumor benigno)
Pos (Positivo)	0,792	0,096
Neg (Negativo)	0,208	0,904

Com essa tabela, fez-se a seguinte pergunta: "Suponha que uma paciente pertença a uma população (mesmo grupo etário, hábito alimentar, etc.) na qual a incidência geral de câncer de mama é de 1%. Detectado um nódulo no seio desta paciente, pede-se uma mamografia para avaliar a possibilidade de que se trate de um tumor maligno; o resultado é positivo. De posse deste conjunto de informações, qual é, em sua opinião, a probabilidade de tratar-se de um tumor maligno?"

Pelo Teorema de Bayes:

$$\begin{aligned}
 Pr(Ca|Pos) &= \frac{Pr(Pos|Ca) \cdot Pr(Ca)}{Pr(Pos)} \\
 &= \frac{Pr(Pos|Ca) \cdot Pr(Ca)}{Pr(Pos|Ca) \cdot Pr(Ca) + Pr(Pos|Ca') \cdot Pr(Ca')} \quad (7) \\
 &= \frac{0,792 \cdot 0,01}{0,792 \cdot 0,01 + 0,096 \cdot 0,99} = 0,077
 \end{aligned}$$

Observe que a acurácia retrospectiva do exame $Pr(Pos|Ca)$ é diferente de acurácia preditiva $Pr(Ca|Pos)$. Na prática, a importância atribuída à alta probabilidade de um resultado positivo quando o tumor é de fato maligno,

$Pr(Pos|Ca) = 0,792$ foi excessiva com relação a baixa probabilidade de incidência desse tipo de câncer $Pr(Ca) = 0,01$. No contexto mais geral de investigação científica, o Teorema de Bayes expressa o mecanismo pelo qual hipóteses científicas e evidências empíricas são integrados.

Seja F uma hipótese científica cuja probabilidade corrente é $Pr(F|H)$. Seja E a evidência contida nos dados, experimentais ou observacionais, e cuja probabilidade sob a premissa F é dada por $Pr(E|F, H)$. Então, se efetivamente foi observado E , o Teorema de Bayes permite calcular a probabilidade atualizada $Pr(F|E, H)$ da hipótese F . Portanto, a probabilidade atualizada de F é composta pela sua probabilidade a priori, modificada pela acresção das novas evidências E presentes nos dados.

4 Variáveis aleatórias

As variáveis aleatórias são de dois tipos. As discretas associadas a alguma contagem, e as contínuas que envolvem medições ou mesmo razões. A distinção é necessária para que as regras do cálculo das probabilidades sejam corretamente aplicadas a cada caso. Segue um resumo das propriedades e distinções dos principais modelos de probabilidades para variáveis aleatórias discretas e contínuas.

4.1 Variáveis Aleatórias Discretas

Uma variável aleatória discreta X é uma função que associa as proposições de interesse a um conjunto enumerável (ou categórico, que pode ser nominal ou ordinal), não necessariamente finito, de valores. Se, por exemplo, X caracteriza o número de amostras de sangue avaliadas antes do aparecimento de uma amostra que contém um vírus de interesse, então os possíveis valores para X que caracterizam um conjunto infinito $0, 1, 2, \dots$ já que não existe um limite superior para delimitá-lo.

De modo geral, o valor $Pr(X = x_j|H)$ denota a massa de probabilidade no ponto x_j . Sobre o conjunto de todos os pontos plausíveis, ou seja, pontos com massa de probabilidade maior que zero, a Lei da adição garante que:

$$\sum_j Pr(X = x_j|H) = 1 \quad (8)$$

A média ou valor esperado de X , $E(X)$, e a sua variância, $V(X)$, são definidas por:

$$\begin{aligned} E(X) &= \sum_j x_j Pr(X = x_j|H) \\ V(X) &= \sum_j (x_j - E(X))^2 Pr(X = x_j|H) \end{aligned} \quad (9)$$

4.2 Variáveis Aleatórias Contínuas

Quando os possíveis valores de X foram um subconjunto que compreende pelo menos um intervalo da escala de números reais, a variável aleatória é denominada contínua. Variáveis aleatórias contínuas são caracterizadas pela sua função distribuição cumulativa $F(x)$ que é definida para todo o número real $x \in (-\infty, \infty)$.

$$F(x) = Pr(X \leq x|H) \quad (10)$$

A função densidade de probabilidade é definida por $f(x) = \frac{dF(x)}{dx}$. A integração da densidade de probabilidade sobre os números reais é igual a 1, ou seja: $\int_{-\infty}^{\infty} f(x) \cdot dx = 1$.

A média e a variância de X também são expressos em integrais:

$$\begin{aligned} E(X) &= \int x \cdot f(x) dx \\ V(X) &= \int (x - E(X))^2 \cdot f(x) dx \end{aligned} \quad (11)$$

5 Distribuições de Probabilidade

Denomina-se de distribuição de probabilidade de alguma variável aleatória a regra geral que define a função de massa de probabilidade (variável discreta) ou de densidade de probabilidade (variável contínua) para a variável de interesse.

5.1 Distribuições Discretas

Distribuição Uniforme Discreta Ud(a,b)

$$p(x) = \frac{1}{b - a + 1} \quad (12)$$

Distribuição Binomial Bin(n,θ)

$$p(x) = \binom{n}{x} \cdot \theta^x \cdot (1 - \theta)^{n-x} \quad (13)$$

Distribuição Hipergeométrica Hip(M, N, n)

$$p(x) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}} \quad (14)$$

Distribuição de Poisson $\text{Poi}(\mu)$

$$p(x) = \frac{e^\mu \mu^x}{x!} \quad (15)$$

Distribuição Binomial Negativa $\text{BinN}(\mathbf{a}, \theta)$

$$p(x) = \frac{(a+x-1)!}{x!(a-1)!} \cdot \theta^a (1-\theta)^x \quad (16)$$

5.2 Distribuição Contínua**Distribuição Uniforme $\text{U}(\mathbf{c}, \mathbf{d})$**

$$p(x) = \frac{1}{d-c} \quad (17)$$

Distribuição Beta (α, β)

$$p(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot x^{\alpha-1}(1-x)^{\beta-1} \quad (18)$$

Distribuição Beta Gerenalizada**Distribuição Exponencial****Distribuição Gama****Distribuição Gama Inversa****Distribuição Normal****Distribuição Student****5.3 Exemplo de códigos gerados no R**

```

1 # binomial
2 x <- 0:20
3 plot(x, dbinom(x, size=20, prob=.3), type='h')
4
5 # Hipergeometrica
6 m <- 10; n <- 7;
7 x <- 0:9
8 plot(x, dhyper(x, m, n, k), type='h')
9
10 # Poisson
11 m <- 1.8
12 x <- 0:9
13 plot(x, dpois(x, m), type='h')
14
15 # Binomial invertida
16 a <- 5
17 x <- 0:40
18 plot(x, dnbinom(x, a, prob=0.3), type='h')
```

6 Análise Bayesiana de Dados

A inferência bayesiana consiste na construção de uma distribuição de probabilidade posterior via o Teorema de Bayes. Essa distribuição resulta da combinação de informações prévias, sumarizadas em uma distribuição denominada priori, com dados estatísticos descritos por algum modelo probabilístico e resumidos na função de verossimilhança.

A distribuição posterior é a forma mais completa de expressar o estado do conhecimento sobre o fenômeno investigado. Toda pergunta específica é respondida a partir da análise da distribuição posterior. Ela contém toda a informação necessária para a inferência.

Além disso, o processo é dinâmico. A distribuição posterior de hoje pode se transformar na priori em estudos futuros, caracterizando o elemento cumulativo de aquisição de informações.

A distribuição de probabilidade posterior pode ser obtido com o cálculo direto do produto $Pr(\theta) \cdot Pr(X|\theta)$ para cada valor de θ . Ou, pode ser calculada usando famílias conjugadas de distribuições de probabilidade, ou recorrendo a procedimentos de simulação estocástica.

6.1 Probabilidades posteriores para famílias conjugadas de distribuições

Seja \mathcal{F} uma família de distribuições para a verossimilhança $Pr(x|\theta)$ e \mathcal{P} uma família de distribuição para a priori $Pr(\theta)$. Dizemos que \mathcal{F} e \mathcal{P} são famílias conjugadas de distribuições se a distribuição posterior $Pr(\theta|X)$ também é um membro de \mathcal{P} .

Ver https://en.wikipedia.org/wiki/Conjugate_prior

6.2 Probabilidades posteriores e simulação estocástica

Embora as famílias conjugadas de distribuições simplifiquem consideravelmente as análises, elas por outro lado restringem as possibilidades de aplicações ao induzirem a escolha de um modelo muito mais pela sua conveniência matemática que pelo seu realismo. Em muitas situações de interesse, as famílias conjugadas de distribuições não se aplicam. Portanto, para ampliar a abrangência do uso da estatística bayesiana, precisamos usar métodos que tenham aplicabilidade irrestrita.

O princípio básico que fundamenta a relação entre simulação estocástica e distribuições posteriores está na dualidade entre a formulação algébrica de uma distribuição de probabilidade $Pr(\theta)$ e uma amostra simulada $\theta_1, \theta_2, \theta_3, \dots, \theta_n$ obtida dessa mesma distribuição.

Na análise estatística de distribuições posteriores quase sempre é necessário calcular quantidades $J_H(x)$ que se traduzem na razão entre integrais com a seguinte forma:

$$J_H(x) = \int H(\theta) \cdot Pr(\theta|x) d\theta = \frac{\int H(\theta) \cdot Pr(x|\theta) d\theta}{p(x)} \quad (19)$$

Os métodos utilizados para resolver esses problemas complexos estão fundamentados no princípio da Integração de Monte Carlo. Seja $\theta_1, \dots, \theta_n$ uma amostra de n variáveis aleatórias independentes e identicamente distribuídas com densidade $Pr(\theta|x)$. Então, para n suficientemente grande, segue pela Lei dos Grandes Números que:

$$\int H(\theta) \cdot Pr(\theta|x) d\theta \approx \frac{1}{n} \sum_{j=1}^n H(\theta_j) \quad (20)$$

A Lei dos Grandes Números sugere que para melhorar a aproximação, basta aumentar o tamanho da amostra. Vale lembrar que, para utilizar a integração de Monte Carlo é necessário dispor da amostra $\theta_1, \dots, \theta_n$ da distribuição posterior $Pr(\theta|x)$. Há no entanto maneiras de gerar amostras da distribuição posterior de forma indireta:

- Método de re-amostragem por importância (Sampling Importance Resampling - SIR).
- Método com Cadeias de Markov (Markov Chain Monte Carlo - MCMC).

6.3 Re-amostragem por Importância

Na impossibilidade de amostrar diretamente da distribuição posterior, a amostragem por importância usa outra densidade de probabilidade $g(\cdot)$ que seja semelhante a $Pr(\theta|x)$.

O peso é calculado a partir da razão:

$$w(\theta) = \frac{Pr(\theta|x)}{g(\theta)} = \frac{Pr(x|\theta) \cdot Pr(\theta)}{g(\theta)} \quad (21)$$

A esperança posterior de H :

$$E^p[H] = \frac{\int H(\theta) \cdot Pr(\theta|x) d\theta}{\int Pr(\theta|x) d\theta} = \frac{\int H(\theta) \cdot w(\theta) g(\theta) d\theta}{\int w(\theta) g(\theta) d\theta} \quad (22)$$

$$E^p[H] = \frac{E^g[H_w]}{E^g[w]}$$

Pesos normalizados:

$$K = \sum_{i=1}^n w(\theta_i) \quad (23)$$

$$w_i = \frac{w(\theta_i)}{K}$$

Integração de Monte Carlo:

$$E^p[H] \approx \sum_{i=1}^n H(\theta_i)w_i \quad (24)$$

Propriedades de interesse de $g(\cdot)$:

- propiciar que seja facilmente gerada uma amostra $\theta_1, \dots, \theta_n$ de valores independentes;
- ter caudas mais pesadas que as caudas de $Pr(\theta|x)$;
- ter forma semelhante a $Pr(\theta|x)$ evitando esforço de amostragem em regiões pouco plausíveis de θ .

A entropia relativa à uniformidade é uma medida que utiliza a dispersão dos pesos para descrever o grau de proximidade entre as distribuições:

$$ERU = -\frac{\sum w_i \log(w_i)}{\log(n)} \quad (25)$$

O valor máximo de ERU é 1. A medida que a entropia entre os w_i aumenta $ERU < 1$. Para as distribuições unidimensionais, recomenda-se $ERU > 0,9$ (nos casos multidimensionais, valores mais baixos são aceitáveis).

6.4 Monte Carlo com Cadeia de Markov (MCMC)

O Monte Carlo com Cadeias de Markov foi formalizado por meio do artigo intitulado Monte Carlo Method em 1949. Pode ser descrito como método de simulação estatística de números aleatórios para desenvolver simulações.

As etapas para realizar o processo de Monte Carlo:

- a) Modelar o problema definindo uma FDP para representar o comportamento de cada uma das suas incertezas;
- b) Gerar valores pseudoaleatórios aderentes à FDP de cada incerteza do problema;
- c) Calcular o resultado determinístico substituindo as incertezas pelos valores gerados obtendo, assim, uma observação do problema;
- d) Repetir os passos b) e c) até se obter uma amostra com o tamanho desejado de realizações;
- e) Agregar e manipular os resultados da amostra de forma a obter uma estimativa da solução do problema.

7 Redes Bayesianas

Redes bayesianas são uma classe de modelos gráficos que permitem a representação de probabilidades condicionais entre um conjunto dado de variáveis aleatórias $X = X_1, X_2, \dots, X_p$ como um grafo acíclico direto (DAG) $G = (V, A)$. Cada nó $v_i \in V$ corresponde a uma variável aleatória X_i .

As redes Bayesianas podem ser utilizadas para:

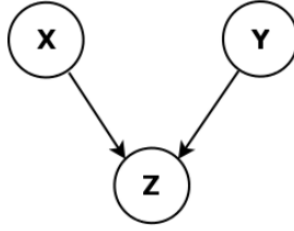


Fig. 1. Representação de probabilidades condicionais um grafo acíclico direto (DAG).

- Tomar decisões baseadas em probabilidades.
- Decidir quais evidências adicionais devem ser observadas, a fim de obter total conhecimento do domínio.
- Realizar uma análise sensível para entender quais aspectos do modelo tem maior impacto sobre determinadas variáveis.
- Explicar os resultados de uma inferência probabilística.

7.1 Propriedades das Redes Bayesianas

Um exemplo de DAG é mostrado na Figura 1. Nesse exemplo, a variável Z é condicionada às variáveis X e Y .

Assim, uma rede Bayesiana consiste:

- Um conjunto de variáveis e um conjunto de arcos ligando as variáveis;
- Cada variável possui um número limitado de estados mutuamente exclusivos;
- As variáveis e arcos formam um grafo dirigido e sem ciclos;
- Para cada variável A que possui como pais B_1, \dots, B_n , existe uma tabela de probabilidade condicional (TPC) $P(A|B_1, \dots, B_n)$.

Uma rede Bayesiana é a representação de um domínio caso a condição de Markov seja satisfeita. A condição de Markov é definida dessa forma: suponha a distribuição de probabilidade conjunto das variáveis aleatórias em um conjunto de nós V em um DAG $G = (V, E)$. Então, dizemos que (G, P) satisfazem a condição de Markov se cada variável $X \in V$, X é condicionalmente independente dos nós não descendentes dados os seus pais.

A condição de Markov afirma que as variáveis não-descendentes não fornecem informações adicionais sobre a variável em questão. Considerando f_X e Pa_X o conjunto de filhos e dos pais do nó X respectivamente, e ainda Pa_{F_X} como o conjunto dos pais dos descendentes diretos de X . O conjunto de nós formados pela união desses três conjuntos é denominado Markov Blanket (ver Figura 2).

Se uma rede Bayesiana satisfaz a condição de Markov, então a sua distribuição de probabilidade conjunto é igual ao produto das probabilidades condicionais de todos os nós dado os valores de seus pais.

$$Pr(X_1, \dots, X_n) = \prod_{i=1}^n Pr(X_i | pa(X_i)) \quad (26)$$

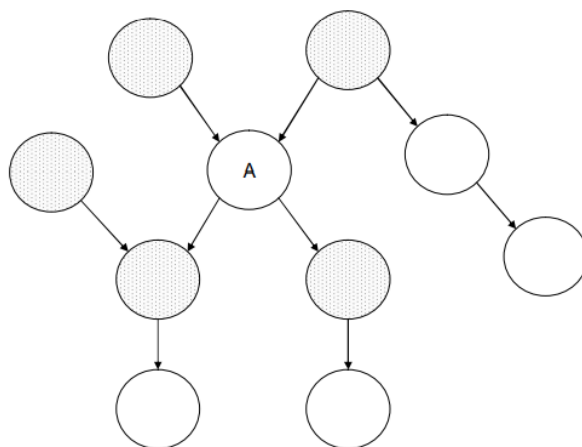


Fig. 2. Os nós preenchidos com cinza representam o conjunto de nós conforme critério Markov Blanket.

Através das propriedades markovianas, podemos considerar que a variável aleatória é independente de outra se não existe entre as variáveis analisadas um grupo específico de variáveis, podendo ser um grupo de evidências. Nesse caso, surge o conceito de d-separação. Para defini-la, consideremos alguns tipos de conexões. Seja X , Z , e Y variáveis de uma rede Bayesiana (ξ, V) , definimos alguns tipos de conexão:

1. Se $X \rightarrow Z \rightarrow Y$, temos um relacionamento head-to-tail;
2. Se $X \leftarrow Z \rightarrow Y$, temos um relacionamento tail-to-tail;
3. Se $X \rightarrow Z \leftarrow Y$, temos um relacionamento head-to-head.

Dada a rede Bayesiana (ξ, V) , representada na Figura 3. Podemos definir $A \subset V$, sendo X e $Y \in V - A$. Desta forma, para os casos 1 e 2, se consideramos que $Z \in A$, a variável Z bloqueará o caminho entre X e Y . Para o caso 3, se consideramos que Z e seus descendentes $\notin A$, a variável Z bloqueará o caminho entre X e Y . Se o caminho entre duas variáveis, ou conjunto de variáveis, é bloqueado, dizemos que essas variáveis, ou conjuntos, são d-separados.

7.2 Inferência em Redes Bayesianas

Pode-se extrair conhecimento da rede Bayesiana através de um processo de inferência. Existem vários métodos para realização de inferência. Inferências podem ser realizadas sobre redes Bayesianas em 04 maneiras distintas:

- Diagnósticos: partindo dos efeitos para as causas.
- Causa: partindo das causas para os efeitos.
- Intercausal: entre causas de um efeito comum.

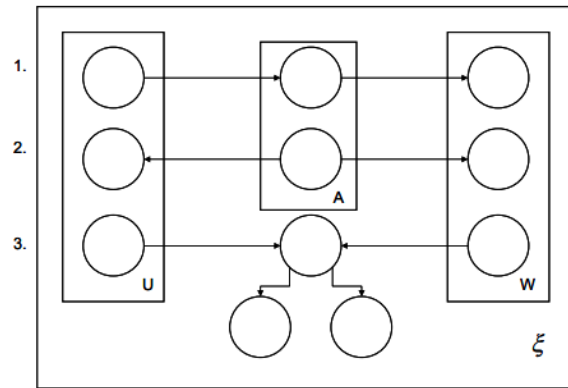


Fig. 3. A independência entre as variáveis aleatórias analisadas em relação ao tipo de conexão.

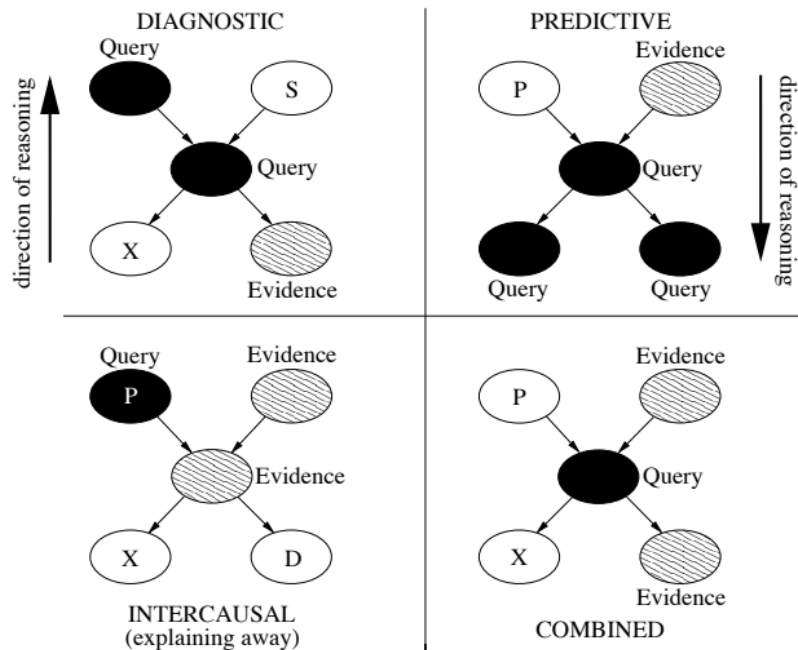


Fig. 4. Tipos de raciocínio presentes em uma rede Bayesiana.

– Mistas: combinação de dois ou mais tipos descritos acima.

A figura 5 ilustra essas 04 formas de raciocínio.

7.3 Estimação de Parâmetros das Redes Bayesianas

A estimação de parâmetros é a parte final da modelagem por redes bayesianas. Após a seleção das variáveis ter sido feita e a estrutura da rede ter sido fixada, precisamos agora conhecer as probabilidades condicionais de cada variável aleatória, dado os valores de seus antecessores. Para essa tarefa existem várias abordagens. Vamos descrever duas: estimação por máxima verossimilhança (MAP), e estimação por aprendizado sequencial, também conhecida como método bayesiano.

Considerando uma base de dados D completa, e considerando meta-independência global e meta-independência local.

blábláblá...

7.4 Decisão com Redes Bayesianas

blábláblá...

8 Exemplo 1

Um paciente tem sofrido de falta de ar (dispneia) e visita o médico, preocupado por ter câncer de pulmão. O médico sabe que outras doenças, como tuberculose e bronquite, são possíveis causas, além do câncer de pulmão. Ela também sabe que outras informações relevantes incluem se o paciente é fumante ou não (aumentando as chances de câncer e bronquite) e a que tipo de poluição do ar ele foi exposto. Um raio-X positivo indicaria tuberculose ou câncer de pulmão.

8.1 Definindo os nós e valores

Primeiro, o engenheiro do conhecimento deve identificar as variáveis de interesse. Isso envolve responder à pergunta: quais são os nós a representar e quais valores eles podem assumir, ou em que estado eles podem estar? Por enquanto, consideraremos apenas os nós que assumem valores discretos. Os valores devem ser mutuamente exclusivos e exaustivos, o que significa que a variável deve assumir exatamente um desses valores por vez. Os tipos comuns de nós discretos incluem:

- Nós booleanos, que representam proposições, assumindo os valores binários verdadeiro (T) e falso (F). Em um domínio de diagnóstico médico, o nó Câncer representaria a proposição de que um paciente tem câncer.
- Valores categóricos. Por exemplo, um nó de poluição pode representar a exposição de um paciente à poluição e assumir os valores baixo, médio, alto.
- Valores numéricos. Por exemplo, um nó chamado Idade pode representar a idade de um paciente e ter valores possíveis de 1 a 120.

Table 2. Escolhas preliminares de nós e valores para o exemplo do câncer de pulmão.

<i>Variável</i>	<i>Tipo</i>	<i>Valores</i>
Poluição	Binário	{Baixa, Alta}
Fumante	Booleano	{Verdadeiro, Falso}
Câncer	Booleano	{Verdadeiro, Falso}
Dispneia	Booleano	{Verdadeiro, Falso}
Raio-X	Binário	{Positivo, Negativo}

Estratégias podem ser usadas para representar a idade exata de um paciente pode ser agrupar os pacientes em grupos de diferentes idades, como bebê, criança, adolescente, jovem, meia-idade, velho. O truque é escolher valores que representem o domínio de forma eficiente, mas com detalhes suficientes para realizar o raciocínio necessário. A Tabela 2 lista as variáveis usadas para análise do caso do câncer de pulmão.

8.2 Estrutura

A estrutura ou topologia da rede deve capturar relacionamentos qualitativos entre as variáveis. Em particular, dois nós devem ser conectados diretamente se um afetar ou causar o outro, com o arco indicando a direção do efeito. Portanto, em nosso exemplo de diagnóstico médico, podemos perguntar quais fatores afetam a chance de um paciente ter câncer? Se a resposta for “Poluição e fumo”, devemos adicionar arcos de Poluição e Fumante a Câncer. Da mesma forma, ter câncer afetará a respiração do paciente e as chances de ter um resultado de raios-X positivo. Portanto, adicionamos arcos de Câncer a Dispneia e Raio-X. A estrutura resultante é mostrada na Figura 5. É importante notar que esta é apenas uma estrutura possível para o problema; olhamos para estruturas de rede alternativas em

8.3 Exemplo de código em R e inferência Bayesiana

A listagem abaixo mostra o código escrito em R. O código utiliza o pacote `bnlearn` e `gRain`. A Rede Bayesiana é carregada no formato `bif` pelo arquivo `cancer.bif`. As probabilidades marginais de cada nó é representada na Figura 6

São declaradas as evidências `Pollution = "high"` e `Xray = "positive"`, ou $Pr(Pollution == high) = 1$ e $Pr(Xray == positive | Cancer == true) = 1$. Consultando a probabilidade posterior antes e depois da declaração das evidências, obtém-se os resultados:

```
Cancer
  True  False
0.01   0.99

Cancer: Pollution = high, Xray = positive.
  True  False
0.12   0.88
```

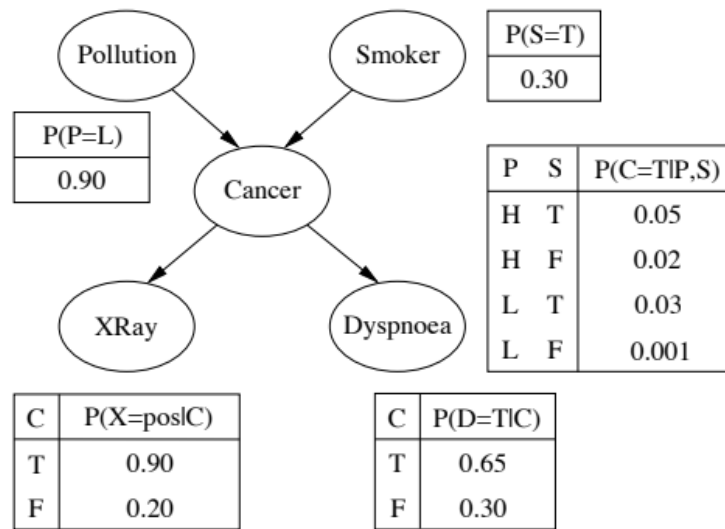


Fig. 5. Uma Rede Bayesiana para o problema do câncer do pulmão.

```

1 library(bnlearn)
2 library(gRain)
3
4 dag = read.bif('cancer.bif', debug = FALSE)
5 graphviz.chart(dag)
6
7 dag.junction = compile(as.grain(dag))
8 querygrain(dag.junction, nodes = c("Cancer"))
9 gin.find <- setFinding(dag.junction, nodes = c("Pollution", "
    Xray"), states = c("high", "positive"))
10 querygrain(gin.find, nodes = c("Cancer"))
    
```

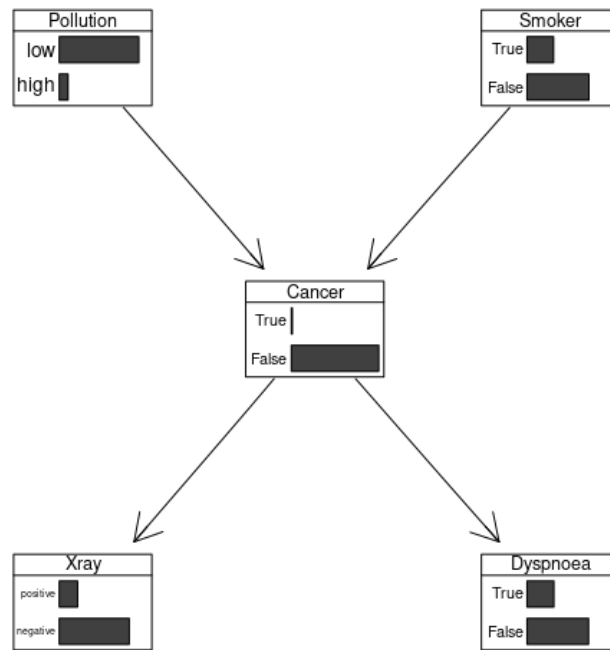


Fig. 6. Probabilidades marginais da rede Bayesiana.