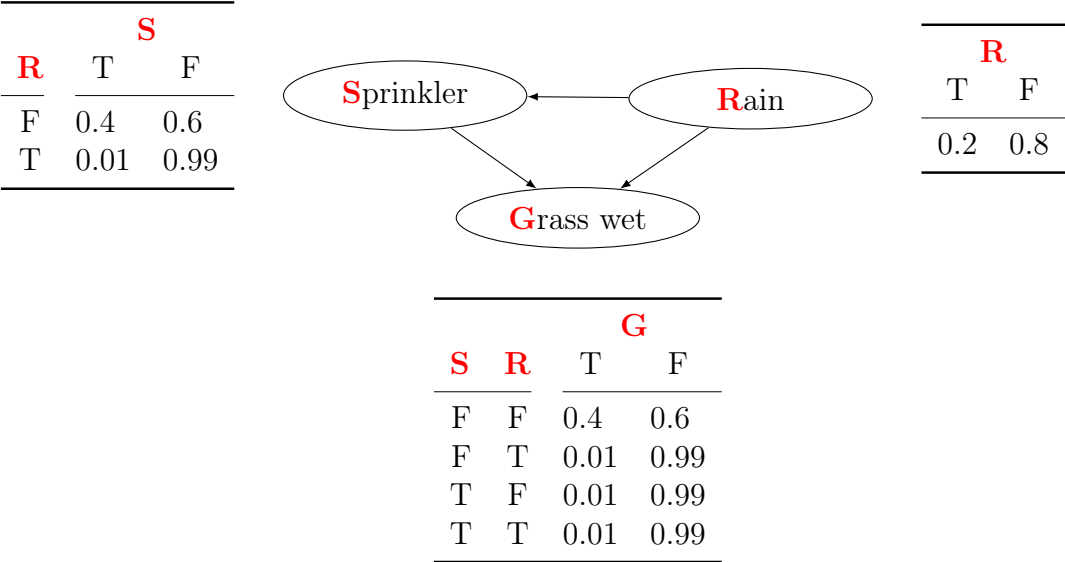


<p>CEFET/RJ          Inteligência Artificial (2018.1)          Prof. Eduardo Bezerra (ebezerra@cefet-rj.br)          Trabalho Prático 4</p>
---

Créditos: esse trabalho contém exercícios retirados do livro texto da disciplina (AIMA) e da disciplina CSEP 546: Data Mining.

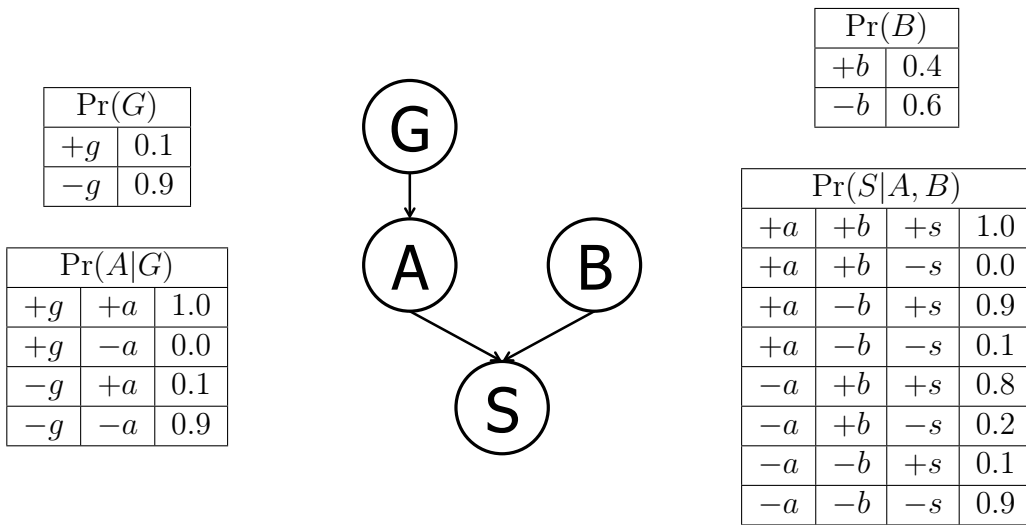
- Suponha que um exame médico é 99% preciso. Suponha também que aplicação desse exame em um paciente dá resultado positivo. A doença detectada pelo exame é rara: apenas 1 em cada 10.000 indivíduos da população tem a doença. Quais são as chances de o paciente ter realmente a doença, dado que o resultado do exame foi positivo?
- Considere a rede bayesiana fornecida abaixo. Forneça a distribuição conjunta sobre as três variáveis.



- Considerando ainda a rede bayesiana do item anterior. Considere também que a tabela a seguir apresenta uma sequência de números aleatórios, que você pode usar para realizar inferência aproximada sobre a rede e responder a alguma consultas probabilísticas. Considere esses números fornecidos na seguinte ordem: de cima para baixo, e da esquerda para a direita.

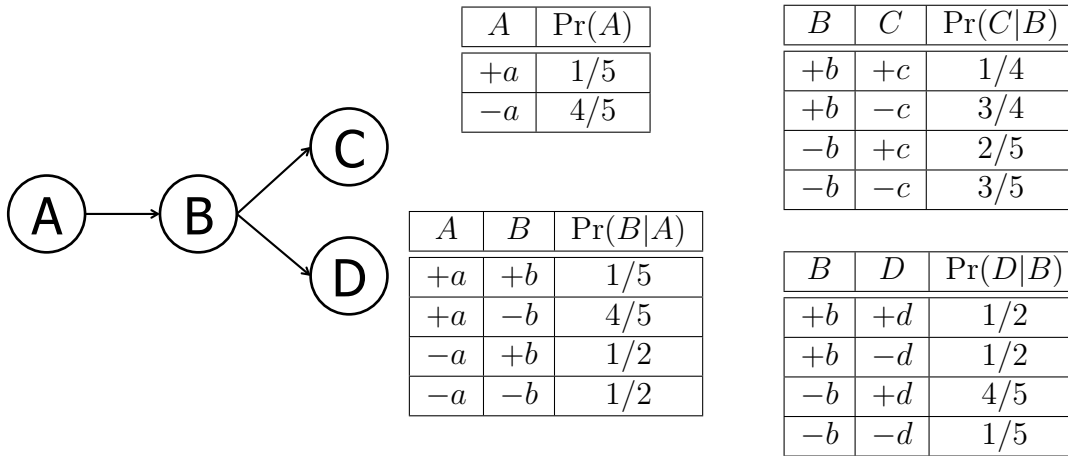
0.02	0.06	0.02	0.84	0.02	0.59
0.33	0.23	0.87	0.49	0.40	0.22
0.33	0.48	0.78	0.42	0.82	0.35

- (a) Quais são as ordenações lineares possíveis na rede bayesiana?
- (b) Usando Amostragem a Priori (*Prior Sampling*), compute aproximações para  $\Pr(R)$ ,  $\Pr(S)$ ,  $\Pr(G)$  e  $\Pr(G|S = True)$
- (c) Usando Amostragem por Rejeição (*Rejection Sampling*), compute aproximações para  $\Pr(R)$ ,  $\Pr(G|S = True)$  e  $\Pr(G)$
- (d) Usando Ponderação por Verossimilhança (*Likelihood Weighting*), compute aproximações para  $\Pr(R)$ ,  $\Pr(G|S = False)$  e  $\Pr(G)$
4. Suponha que um paciente possa ter um sintoma ( $S$ ) que pode ser causado por duas doenças diferentes ( $A$  e  $B$ ). Sabe-se que a variação do gene  $G$  desempenha um grande papel na manifestação da doença  $A$ . A rede bayesiana e tabelas de probabilidade condicionais correspondentes para esta situação são mostradas abaixo.



- (a) Compute a seguinte entrada para a distribuição oconjunta:  $\Pr(+g, +a, +b, +s) =$
- (b) Qual é a probabilidade de que um paciente tem a doença  $A$ ?  $\Pr(+a) =$
- (c) Qual é a probabilidade de que um paciente tem a doença  $A$  dado que el tem a doença  $B$ ?  $\Pr(+a|+b) =$

5. Considere novamente a rede bayesiana da questão 4.
- Qual é a probabilidade de que um paciente tem a doença  $A$  dado que ele tem o sintoma  $S$  e a doença  $B$ ?  $\Pr(+a | +s, +b) =$
  - Qual é a probabilidade de que um paciente tem a doença de carregar a variante do gene  $G$  dado que ele tem a doença  $A$ ?  $\Pr(+g | +a) =$
  - Qual é a probabilidade de que um paciente tem a doença de carregar a variante do gene  $G$  dado que ele tem a doença  $B$ ?  $\Pr(+g | +b) =$
6. Considere a rede bayesiana abaixo, e as CPTs correspondentes:



- Sua tarefa é estimar  $\Pr(+b | -a, -c, -d)$  usando amostragem. Abaixo são fornecidas algumas amostras que foram produzidas por meio de *prior sampling* (i.e., o estágio de rejeição na amostragem por rejeição não aconteceu ainda). Circule as amostras que seriam rejeitadas se fosse aplicada a amostragem por rejeição (*rejection sampling*):

-a	-b	+c	+d	-a	-b	-c	-d
+a	-b	-c	+d	-a	+b	+c	+d
-a	-b	+c	-d	+a	-b	-c	-d

- Usando as amostras acima, qual valor você estima para  $\Pr(+b | -a, -c, -d)$  usando amostragem por rejeição?
- Usando as amostras abaixo (que foram geradas usando ponderação por verossimilhança (*likelihood weighting*)), produza uma estimativa para  $\Pr(+b | -a, -c, -d)$  ponderação por verossimilhança, ou declare que a computação é impossível, se for o caso.

-a	-b	-c	-d
-a	+b	-c	-d
-a	-b	-c	-d

7. Suponha que você é o gerente de um projeto no qual um estagiário está trabalhando para lançar a nova versão de um produto de software. Suponha ainda o seguinte:

- Se o estagiário escreve um bom código e realiza testes suficientes, o lançamento é mais provável de ser bem sucedido.
- Se café estiver disponível, o estagiário escreve um código melhor.
- Se café estiver disponível, é mais provável que o estagiário escreva testes.
- Ter um prazo de lançamento apertado faz com que o estagiário escreva menos testes.

Agora, responda as seguintes questões.

- Desenhe a rede bayesiana de acordo com as informações acima, usando as variáveis a seguir, todas binárias: LancamentoComSucesso, CodigoTemQualidade, CoberturaTesteAdequada, PrazoApertado e CafeDisponivel.
  - De acordo com essa rede, se o estagiário escreve testes com cobertura adequada, o código tem mais probabilidade de ser bom? Justifique sua resposta.
  - De acordo com essa rede, a qualidade do código e o prazo são independentes? Justifique sua resposta.
  - De acordo com esta rede, a qualidade do código e o prazo são independentes, dado que foi realizado um lançamento bem-sucedido? Justifique sua resposta.
  - De acordo com essa rede, a qualidade do código e o prazo são independentes, dado que o estagiário não recebeu café? Justifique sua resposta.
8. Suponha que um modelo probabilístico seja definido por meio de 4 variáveis aleatórias binárias:  $a$ ,  $b$ ,  $c$  e  $d$ . Considere uma rede bayesiana correspondente a esse modelo e estruturada da seguinte forma:  $a \rightarrow b \rightarrow c \rightarrow d$ . A rede tem os seguintes parâmetros:

$$\begin{aligned}
 P(a) &= 0,6 \\
 P(b|a) &= 0,9 \\
 P(b|\neg a) &= 0,2 \\
 P(c|b) &= 0,3 \\
 P(c|\neg b) &= 0,7 \\
 P(d|c) &= 0,4 \\
 P(d|\neg c) &= 0,7
 \end{aligned}$$

Responda às seguintes perguntas:

- Calcule a probabilidade de  $P(d)$ . (dica: você precisa marginalizar as probabilidades de  $a$ ,  $b$  e  $c$ )
- Quantas somas você precisou calcular? Você pode fazer melhor com uma ordenação diferente de somatórios?
- Suponha que tenhamos uma rede bayesiana com uma estrutura semelhante, mas com  $n$  variáveis binárias em uma cadeia, ou seja,  $x_1 \rightarrow x_2 \rightarrow x_3 \rightarrow \dots \rightarrow x_n$ . Qual é o menor número de somas necessárias para calcular  $P(x_n)$ ?

(d) Qual é o pior número de somas necessárias para calcular  $P(x_n)$ ?