LISTA 1 PROBABILIDADES PRIMEIRO SEMESTRE 2020

ENTREGA: 31/03/2020 LISTA 1

Exercício 1. Leia com cuidado as notas de aula até o capítulo sobre a distribuição normal.

Exercício 2. (a) Suponha $A = A_1 + A_2 + \cdots A_n$, onde os A_i são mutuamente exclusivos. Estamos interessados na probabilidades condicionadas por A. Isto será útil no futuro. Mostre, usando as regras do produto e da soma que

$$P(B|AI) = \frac{\sum_{i=1}^{n} P(A_i|I)P(B|A_iI)}{\sum_{i=1}^{n} P(A_i|I)}$$

(b) Seja a asserção B definida pela soma lógica $B = \Sigma_{i=1}^{n} A_i$, mostre que

$$P(B|I) \leq \sum_{i=1}^{n} P(A_i|I)$$

Exercício 3. São feitas 3 perguntas a N pessoas. As respostas só podem ser "sim"ou "não". Considere dados:

- o número de pessoas que respondeu sim à terceira pergunta é N_c (e portanto $\bar{N}_c = N - N_c$, respondeu não) é conhecido;
- N_2 é o número de pessoas que respondeu sim a duas perguntas e não a uma pergunta não importando quais especificamente;
- N_3 é o número de pessoas que respondeu sim às três perguntas.
- N_{AB} é número que respondeu afirmativamente às duas primeiras perguntas

Quantas pessoas responderam que não às duas primeiras perguntas e sim à terceira?

Exercício 4. O resultado de um teste para uma doença pode ser positivo ou negativo. As asserções relevantes são

- D: "a paciente está doente"e $\bar{D}$: "a paciente está sã"
- t_+: "o resultado do teste foi positivo", t_-: "o resultado do teste foi negativo"

O teste não é perfeito. Suponha que tenhamos informação que o teste é bom: as probabilidades $P(t_+|D)$ e $P(t_-|\bar{D})$ são altas, por exemplo ambas são 0.9. Suponha que a paciente recebe a informação de que o teste deu positivo. Com que probabilidade ela está doente? Ou seja queremos saber $P(D|t_+)$, a probabilidade de verdadeiro positivo. A probabilidade $P(\bar{D}|t_-)$ é a de um verdadeiro negativo. As probabilidades de erro são $P(D|t_-)$ para um falso negativo e $P(\bar{D}|t_+)$ para um falso negativo.
Como proceder? A regra do produto e a consistência leva a regra de Bayes de inversão. A teoria não é suficente para responder isso porque falta informação. Isto não é uma falha da teoria. Em lugar de fazer suposições tácitas mas não explícitas a teoria indica o que mais é necessário saber para chegar a esse número; (a) Descubra que informação está faltando. (b) Faça uma estimativa para alguma doença específica e obtenha $P(D|t_+)$ a probabilidade de estar doente.

Exercício 5. Tentamos enviar um sinal através dos circuitos da figura 1 compostos por cabos e chaves $\{s_i\}$. A variável s_i toma valores 1 se fechada e 0 se aberta. Cada chave tem probabilidade

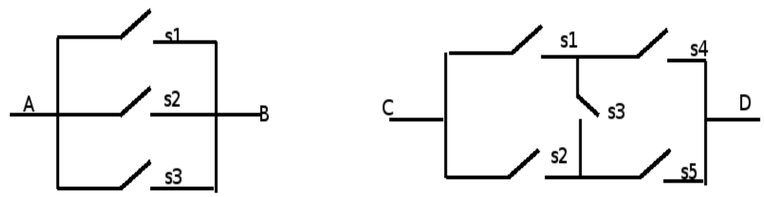

FIGURA 1. Circuitos em sistema de communicação

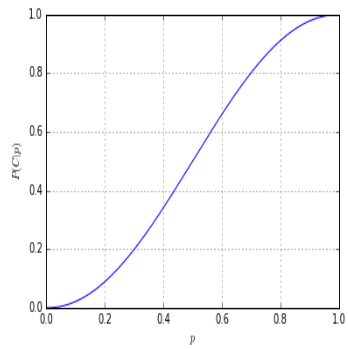

FIGURA 2. Probabilidade de passar o sinal de C a D

p de estar fechada independentemente de qualquer outra informação no problema. Isto significa que

$$P(s_i|p) = p\delta_{s_i 1} + (1-p)\delta_{s_i 0}$$

Encontre a probabilidade de passar um sinal

(1) de A a B na figura 1 à esquerda;
(2) de Ca D na figura 1 à direita.

Nota 1: Use as regras do produto e da soma. Faça os gráficos como função de p. A figura 2 mostra o resultado no segundo caso. Nota 2: Note que p não é uma distribuição de probabilidade mas um parâmetro. A distribuição de probabilidade toma o valor do parâmetro p quando $s_i = 1$.

Exercício 6. Considere um problema de "urnas"onde há cartas em lugar de bolas. Suponha que ninguém sabe fazer maço. Um baralho de 52 cartas é formado por cartas numeradas de 1 a 13 de 4 naipes (a)Três cartas são selecionadas de um baralho sem reposição. Encontre a probabilidade de não tirar um coração.
(b) Um jogador recebe 5 cartas. Qual é a probabilidade que três tenham o mesmo número?

Exercício 7. (a) A variável aleatória X é distribuida uniformemente no intervalo $0 \leq x < 2\pi$. Fora desse intervalo a densidade de probabilidade é zero. A variável Y toma valores no intervalo $-1 \leq y \leq 1$ está relacionada com X por $Y = \sin X$. Encontre a densidade de probabilidade de y.
(b)A variável X tem densidade de probabilidade dada pela função $f(x)$. A variável Y é definida pela transformação $Y = f(X)$. Qual é a densidade de probabilidade de Y?
(c) Em Física 1 (ou antes) foi calculado o alcance $A(\theta, v_0)$ de um projétil, sob a ação de um campo gravitacional g uniforme num terreno plano, como função do ângulo de lançamento e da velocidade inicial de módulo v_0. Encontre a probabilidade de A, $P(A|I_1)$ sob a informação $I_1 : v_0$ é conhecido e θ é uniforme entre θ_1 e θ_2.
(d) $P(A|I_2)$ o mesmo do anterior onde I_2: θ é conhecido e v_0 é uniforme entre v_1 e v_2.

(e) $P(A|I_3)$ o mesmo do anterior onde I_3: θ é uniforme entre θ_1 e θ_2 e v_0 é uniforme entre v_1 e v_2.
(f) Refaça (c-e) com atrito...(brincadeira)

Exercício 8. Uma variável tem distribuição normal

$$P(x|\mu,\sigma) = N \exp -\frac{1}{2\sigma^2}(x-\mu)^2$$

(1) Encontre a normalização $N(\sigma)$
(2) Encontre os valores esperados $\mathbb{E}(x|\mu,\sigma)$ e $\mathbb{E}(x^2|\mu,\sigma)$.
(3) Para diferentes valores de $\mu = 0, 3$ e $\sigma = 1, 4$, desenhe a função $\phi(x|\mu,\sigma)$, a distribuição cumulativa de x, definida por

$$\phi(x|\mu,\sigma) = \int_{-\infty}^{x} P(x'|\mu,\sigma)dx'$$

(o esboço deve ser feito à mão)

Exercício 9. Duas variáveis que tomam valores nos reais tem distribuição conjunta normal

$$P(x,y|\rho) = N \exp -\frac{1}{2C}(x^2 - 2\rho xy + y^2)$$

onde ρ é um parâmetro positivo dado, entre 0 e 1.

(1) Encontre $C(\rho)$ para que as marginais sejam gaussianas padrão $P(x) = \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}x^2$,$P(y) = \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}y^2$.
(2) Encontre a normalização $N(\rho)$
(3) Encontre as probabilidades $P(x|y,\rho)$ e$P(y|x,\rho)$
(4) Encontre os valores esperados $\mathbb{E}(x|\rho)$,$\mathbb{E}(y|\rho)$ e $\mathbb{E}(xy|\rho)$. Interprete o significado de C e o de ρ.

Dica: Use as regras do produto e da soma.

Exercício 10. Um paradoxo é tipicamente o nome dado a uma situação onde uma resposta óbvia está errada. Um paradoxo é como uma ilusão visual que somente pode ser vista pela visão periférica mas some quando toda o esforço atencional lhe é dedicado.
O exemplo típico do paradoxo de Simpson [?] [?] é daquela droga que se mostrou prejudicial para grupos de homens e também para grupos de mulheres. O advogado da industria farmacéutica reconhece que isso é assim mas argumenta que é boa para humanos, como mostra usando os dados da tabela 1.

Tabela 1. *Tabela para o paradoxo de Simpson* [?] [?]

	Recuperado	Não recuperado	Fração Recuperada
Homens			
Receberam a droga	18	12	0.60
Não receberam a droga	7	3	**0.70**
Mulheres			
Receberam a droga	2	8	0.20
Não receberam a droga	9	21	**0.30**
Humanos (combinado)			
Receberam a droga	20	20	**0.50**
Não receberam a droga	16	24	0.40

Os dados foram colhidos num teste para avaliar a droga. Sugerem o seguinte comportamento por parte do médico. Se o sexo da pessoa é desconhecido então pode dar a droga pois $0.50 > 0.40$ a taxa de recuperação passou de 0.40 a 0.5 administrando a droga. Mas se o médico se torna

consciente do sexo do paciente, não deve administrá-la. Se for homen, a taxa cai de 0.70 para 0.60, se for mulher cai de 0.30 para 0.20. O leitor que não está insatisfeito como este resultado deve abandonar os estudos de probabilidade neste ponto e considerar o estudo de direito.
Pense sobre este problema, que será discutido em aula. Como sempre, devemos proceder de forma cautelosa. Primeiro, como em todo problema identifique as asserções relevantes:

- D toma valor V (verdade) se a droga foi consumida e F (falso) se não
- R toma valor V se o paciente se recuperou e F (falso) se não
- G toma valores M se o paciente for mulher e H se for homen.

Segundo, identifiquemos a questão a ser respondida. Discuta o uso da informação da tabela, em que contexto ocorreu a colheita da informação e em que contexto o tratamento seria usado. Neste caso queremos saber se a droga deve ser administrada ou não. Devemos atribuir valores a $P(R|DI)$ nas diferentes situações I.
Terceiro, identificar a informação que obtivemos no passado que será relevante para responder à pergunta no futuro.
Esta é a mais difícil. Entretanto a parte difícil é perceber que nela reside o problema. Uma vez trazida á tona para o debate, o mistério desaparece.

Exercício 11. Procure o significado do termo gerrymandering em inglés. Se aplica à divisão cuidadosa de regiões em distritos para fins de escolha distrital de representantes. Para distribuições espaciais não homogêneas a escolha pode ser feita de forma a ter um efeito Simpson como no exercício anterior. Pense sobre isso.