

Lista de Exercícios - Probabilidades

Nestor Caticha

31 de março de 2017

Data de entrega 18 de abril.

Exercício 1. (a) Suponha $A = A_1 + A_2 + \dots + A_n$, 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 $B = \sum_{i=1}^n A_i$, mostre que

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

Exercício 2. 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 3. (a) X toma valores nos reais e sua densidade de probabilidade sob informação I é dada por $P(x|I) = f(x; \theta)$. Qual é a distribuição da variável Y definida por $Y = X^2$? (b) Suponha que f é (membro da família de distribuições) normal.

(c) Suponha que X toma valores no intervalo $[0, L]$ e sua distribuição é uniforme nesse intervalo. Qual é a distribuição de Y .

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 positivo.

Como proceder? A regra do produto e a consistência leva a regra de Bayes de inversão. A teoria não é suficiente para responder isso, mas ela 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.

Pense sobre o que as pessoas entendem pela informação: "baralho bem embaralhado" (não foi feito "maço"). Considere os baralhos mencionados a seguir bem embaralhados. Qual é a probabilidade ao retirar 3 cartas ao acaso de um baralho, que as 3 sejam de diamante, nos casos

1. Após a extração de cada carta, ela é retornada ao baralho, que é novamente embaralhado.
2. Após a extração a carta não retorna ao baralho.

Comece a retirar cartas de um baralho, uma a uma e retire-as do baralho. Encontre uma expressão que estima o número esperado de cartas até encontrar um diamante.

Exercício 6. Tentamos enviar um sinal através dos circuitos da figura 1 compostos por cabos e chaves $\{s_i\}$. Cada chave tem probabilidade p de estar fechada independentemente de qualquer outra informação no problema. Encontre a probabilidade de passar um sinal

1. de A a B na figura 1 à esquerda;
2. de C a D na figura 1 à direita.

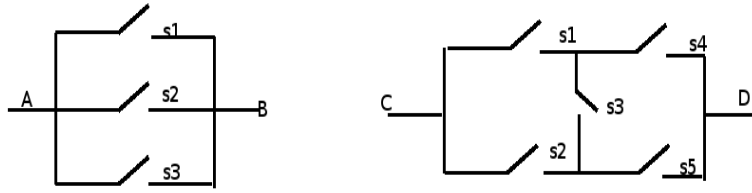


Figura 1: Circuitos em sistema de comunicação

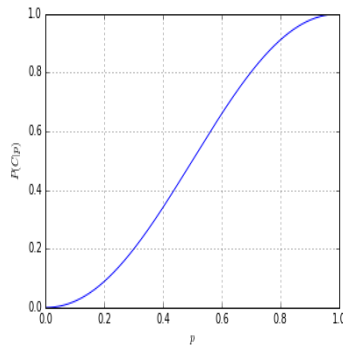


Figura 2: Probabilidade de passar o sinal de C a D

Exercício 7.

Considere um conjunto de símbolos $A = (a_1, a_2, \dots, a_n)$. Pode considerar A como um alfabeto ou uma lista de todas as palavras possíveis ou um dicionário sem as definições. Uma mensagem de tamanho r é composta por uma lista de r símbolos. No caso em que as palavras possam ser usadas com reposição o fato de uma palavra ter sido ou não usada não altera a possibilidade de uso futuro. Uma lista ordenada é uma mensagem em que a ordem das palavras importa. Uma lista desordenada é uma em que só o número de vezes em que cada palavra aparece importa.

- Mostre que o número de mensagens ordenadas com reposição de tamanho r é n^r .
- Com $r \leq n$, mostre que o número de mensagens ordenadas sem reposição é $\frac{n!}{(n-r)!}$.
- Com $r \leq n$, mostre que o número de mensagens desordenadas de tamanho r sem reposição é $\frac{n!}{r!(n-r)!}$.

O caso de mensagens desordenadas com tamanho r e com reposição é mais difícil. Cada símbolo a_i aparece um número de vezes r_i . Por motivos que ficarão mais claros ao estudar sistemas físicos, chamamos r_i o número de ocupação do estado i . Os valores de r_i são inteiros entre 0 e r e satisfazem $\sum_{i=1,n} r_i = r$. Por ser uma mensagem desordenada pode ser representada por um diagrama como no

exemplo abaixo

$$\mathbf{I} ** | *** || *** \mathbf{I}$$

que representa a mensagem feita de um dicionário de $n = 4$ símbolos, onde a_1 aparece duas vezes, a_2 três vezes, a_3 nenhuma e a_4 três vezes, ou seja $r_1 = 2, r_2 = 3, r_3 = 0, r_4 = 3$. O símbolo $\mathbf{I}$ aparece só para marcar o início e fim da mensagem e é o mesmo para qualquer mensagem. O problema foi reduzido a um problema com dois símbolos $b_1 = *$ e $b_2 = |$, onde b_1 aparece r vezes e b_2 aparece $n - 1$ vezes, ou seja um total de $r + n - 1$ símbolos e portanto o número total de mensagens desordenado sem reposição é dado por $\frac{(r+n-1)!}{r!(n-1)!}$. Você pode mandar a mensagem ou simplesmente o conjunto de números de ocupação.

Como exemplo considere o problema que Einstein e Bose consideraram há quase 100 anos. Um sistema é feito por um conjunto de n osciladores que podemos chamar $(a_1, a_2, \dots, a_n)$. Cada oscilador a_i pode estar num estado descrito pelo inteiro $r_i \geq 0$, com $\sum_{i=1, n} r_i = r$. Equivalente a se a palavra a_i foi usada r_i vezes. Não importa qual é o oscilador que está em cada estado. Queremos saber o número de configurações dados n e r . Neste problema a energia E do sistema é proporcional a r , $E = \hbar\omega r$ demos um primeiro passo para estudar a termodinâmica de um sistema físico a partir de considerações microscópicas.

- Encontre o número $\Omega(n, E)$ de configurações para n e E dados.

Falta ainda um passo muito grande que só daremos ao introduzir entropia, para descrever a termodinâmica.

Exercício 8. Para pensar: O Paradoxo de Simpson

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 indústria farmacêutica 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 groga 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 com 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 à tona para o debate, o mistério desaparece.

Exercício 9. (a) Problema de Linda 1. Amos Tversky and Daniel Kahneman colocaram a questão a seguir, chamada de Problema de Linda, sobre probabilidades. Considere as asserções a seguir:

- I : Linda tem 31 anos, é solteira, assertiva, e muito inteligente. Ela se formou em filosofia. Quando estudante, estava profundamente preocupada com questões de discriminação e justiça social, e também participou de manifestações anti-nucleares.
- A : Linda é bancária .
- B : Linda é bancária e participa do movimento feminista .

Responda rapidamente qual das duas asserções é mais provável?

(b) Problema de Linda 2. Não continue lendo até ter respondido à pergunta anterior.

Responda após pensar. O problema é atribuir números a $P(A|I)$ e $P(B|I)$. Qual é maior? Responda usando a regra do produto e use o fato que qualquer probabilidade tem uma cota superior 1. Este problema também é chamado de Falácia da conjunção. Introduza a asserção

- C : Linda é bancária e não participa do movimento feminista .

Qual seria o ordenamento das três probabilidades $P(A|I)$, $P(B|I)$ e $P(C|I)$? Procure alguém feminista e faça a pergunta, faça o mesmo com alguém machista. Divirta-se com a percepção que as pessoas são irracionais. O que você acha que as pessoas acham que respondem quando tem que ser rápidas? Note que muitas vezes ao fazer uma pergunta, quem responde está respondendo a uma pergunta parecida mas não exatamente aquela demandada.

(c) Problema de Linda 3. Mostre usando a regra do produto que $P(A|I) \geq P(B|I)$. Tente inferir o que as pessoas fazem quando acham que está certo que $P(A|I) \leq P(B|I)$. Encontre asserções $A'|I'$ e $B'|I'$ *parecidas* com $A|I$ e $B|I$ tal que seja razoável supor mais provável supor o ordenamento contrário.