

PROBABILIDADE PROVA 2 - JUNHO 2018. INDIVIDUAL - ENTREGA: 06 DE JUNHO

PROBABILIDADES

PROVA 2

Instruções: a prova deve ser feita individualmente. Ela não visa medir se o grupo sabe. O objetivo é obter informação sobre o que cada aluno sabe. A nota V será dada por

$$V = \min(10, v_0 \sum_{i=1}^n v_i),$$

onde v_i é o valor da nota da questão i . A soma dos valores é maior que dez pontos. Quem não responder à pergunta 0 terá uma prova oral. Apresente as soluções cuidadosamente deixando claro as hipóteses feitas e mostrando o desenvolvimento matemático. Frases telegráficas ou equações soltas não são a forma adequada de comunicar os resultados. Não precisa fazer em Latex, provas manuscritas são bem-vindas. Provas em Word não o serão. Numere as páginas e os exercícios. Grampeie as folhas de forma ordenada por número de questão. As folhas de provas sem nome serão enviadas para reciclagem.

Exercício 0.

Responda: Você fez a prova de forma individual? Resposta SIM terá nota $v_0 = 1$ e a resposta NÃO nota $v_0 = 0$.

Exercício 1. Distribuição normal A variável aleatória X tem distribuição normal com média μ e desvio padrão σ .

- Z toma valores $z = ax^3 + bx^2$. Encontre o valor esperado de Z .
- No caso em que $\mu = 0$ e $\sigma = 1$ a variável Y toma valores $y = \cos x$. Encontre a densidade de probabilidade $P(Y)$ e sua média.
- Suponha novamente que $\mu = 0$ e $\sigma = 1$. Obtenha a distribuição de probabilidade para a variável aleatória $S_N = \sum_{i=1}^N x_i^2$, onde os x_i são amostras independentes e igualmente distribuídas (i.i.d) de X .

Valores: .5, .5, 1.

Exercício 2. Em uma experiência o operador tem controle de escolha do valor x e pode medir para cada escolha de x o valor de y . Uma corrida da experiência gerou o conjunto de dados $(x_i, y_i)_{i=1 \dots N}$. No que segue teremos informação que os valores de y_i são corrompidos por ruído gaussiano e independente com média nula e variância σ^2 .

- Há motivos teóricos para achar que o modelo $M_1 : y = f_1(x; \theta)$ é adequado. Encontre uma expressão que permita encontrar θ_{ML} o valor de θ de máxima verossimilhança.
- Encontre uma expressão para a incerteza da estimativa de θ . Justifique a escolha dessa estimativa.

Valores: 1., 1.

Exercício 3. Considere a seguinte informação $I =$ “Uma moeda é jogada para cima, bate no teto, no ventilador do teto, e cai no chão plano”. Há vários motivos para atribuir $p = 1/2$ à probabilidade que caia a cara para cima, isto é $p = P(s = 1|I) = 1/2$ e $q = P(s = -1|I) = 1/2$. Poderíamos considerar outra experiência I' onde p, q tem outros valores (entre zero e um). Consideremos as jogadas independentes, para duas jogadas i e j quaisquer $P(s_i|s_j I') = P(s_i|I')$.

Chame m o número de caras para cima, quando a moeda é jogada n vezes. A frequência de caras é definida por $f = m/n$

- (A) Mostre que a distribuição de m , é a distribuição binomial:

$$(1) \quad P(m|N = nI') = \frac{n!}{m!(n-m)!} p^m q^{n-m}$$

- (B) Calcule $\langle m \rangle$, $\langle m^2 \rangle$. [Dica: Use a expansão binomial de (i) $(p+q)^n$, (ii) $p \frac{\partial}{\partial p} p^m = mp^m$ e (iii) a normalização $p+q=1$; resposta: $\langle m \rangle = np$, $\langle m^2 \rangle = n^2 p^2 + np(1-p)$]
- (C) Refaça a dedução da desigualdade de Chebyshev para distribuições de variáveis que tomam valores discretos e mostre que para ϵ fixo, a probabilidade que a frequência f se afaste do *valor esperado* $\langle f \rangle = p$ por mais que ϵ , cai com $1/n$.
- (D) Discuta e pense: Então de que forma a frequência está ligada à probabilidade? A frequência *converge*, quando n cresce, para a probabilidade p . Toda convergência precisa ser definida em termos de uma distância, que vai para zero quando se toma algum limite. É fundamental entender que a distância aqui não é ϵ , mas é a **probabilidade** que f se afaste de p por mais de ϵ . Assim, a frequência f converge **em probabilidade** à probabilidade p .

A conclusão do exercício acima é fundamental. Como poderíamos definir probabilidades em termos de frequência, se para mostrar que a frequência está associada à probabilidade usamos o conceito de convergência em probabilidade? Discuta se é errado ou não definir um conceito usando esse conceito na definição.

Mas o exercício acima mostra porque pode parecer sedutor usar a frequência em lugar da probabilidade. Se tivermos informação I' sobre uma experiência e dados sobre uma sequência de experimentos nas condições I' podemos atribuir valor à probabilidade de forma mais segura. A frequência é informação que pode ser usado para atribuir um número à probabilidade, mas não é o único tipo de informação para fazer isso. **Valores:** 2.

Exercício 4. Análise combinatorial. 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

$$** | *** || **$$

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 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

Bose e Einstein consideraram há quase 100 anos. Um sistema é feito de n partículas. Cada partícula i pode estar num estado descrito por r_i , com $\sum_{i=1,n} r_i = r$. Não importa qual é a partícula que está em cada estado. Queremos saber o número de configurações dados n e r . Se a energia E do sistema é proporcional a r , $E = \hbar\omega r$ e o número de átomos é n isto é um primeiro passo para estudar a termodinâmica de um sistema físico.

- Encontre o número de configurações para n e E dados.

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

Valores: 2.

Exercício 5. Suponha que as variáveis y_1 e y_2 tenham distribuição uniforme no intervalo $[0, 1]$ e sejam independentes.

- (1) A) Encontre a distribuição da variável $z = -\log y_1$. Dados y_1 e y_2 obtemos x_1 e x_2 a partir da transformação:

$$\begin{aligned}x_1 &= \sqrt{-2 \ln y_1} \cos 2\pi y_2 \\x_2 &= \sqrt{-2 \ln y_1} \sin 2\pi y_2\end{aligned}$$

- (2) b) Encontre a distribuição conjunta $P_X(x_1, x_2)$
- (3) c) Suponha que tenhamos informação que nos permite atribuir probabilidades a $P(a|I)$ e $P(m|I)$ as probabilidades de altura a e massa m respectivamente de membros de uma população. Suponha que a e m sejam independentes e encontre uma expressão para a probabilidade $P(b|I)$ da variável $b = a/m^{1/3}$, que representa uma característica geométrica do corpo. Suponha que a e m não sejam independentes, o que faltaria para encontrar $P(b|I)$?

Valores: 1.

Exercício 6. Sistema de Agentes e modelos de eleição. Uma sociedade de N agentes tem que escolher um entre dois candidatos A e B.

- (A) Uma pesquisa de opiniões feita com n eleitores tem como resultado n_A e n_B a favor de cada um dos candidatos. Supondo que $N \gg n$, qual é confiança que se pode ter sobre o uso do resultado da pesquisa como indicador do resultado da eleição caso não hajam mudanças de opiniões. Discuta se é razoável modelar a eleição como uma urna de Bernoulli.
- (B) Agora vamos fazer algo mais interessante, mas que só requer o uso das regras de soma e produto, que são bem estabelecidas e algumas suposições sobre o comportamento humano que podem ser muito discutíveis. O objetivo é mostrar ao aluno que tem as ferramentas para modelar situações muito mais interessantes. Continuando o item anterior suponha que as pessoas conversam entre si e mudam de opiniões. Suponha que no dia t as probabilidades de voto sejam $p_A(t)$ e $p_B(t)$ respectivamente. A dinâmica de mudanças de opiniões é bem complicada mas podemos fazer um modelo simples: Foquemos a atenção em um eleitor, este se reúne com mais dois e passa a ter a opinião da maioria. Ou seja do grupo de três, se dois ou três apoiam o candidato A em t , o eleitor em foco passa a apoiar A no dia $t+1$ e a apoiar B se somente 0 ou 1 dos membros do grupo de três apoiavam A. Suponha que as probabilidades de votar em um candidato de cada um dos membros do grupo cada um dos membros sejam independentes das dos outros membros do grupo. Para o grupo de três escreva como função de $p_A(t)$ as probabilidades da asserções
- (B1) $A_{01} = "0 \text{ ou } 1 \text{ membros apoiam A}"$

- (B2) $A_{23} = "2 \text{ ou } 3 \text{ apoiam A}"$ que são respectivamente $p_B(t+1) = f(p_B(t))$ e $p_A(t+1) = f(p_A(t))$ (note que por simetria são a mesma função com argumentos diferentes.) Faça o gráfico de $p_A(t+1) = f(p_A(t))$ contra $p_A(t)$. Inclua no gráfico a identidade (diagonal). O cruzamento de $f(p_A(t))$ com a diagonal indica $p_A(t+1) = p_A(t)$, chamado de ponto fixo.
- (C1) Identifique os pontos fixos p_A^* e p_B^* .
- (C2) Discuta a estabilidade dos pontos fixos. Isto é, se ao perturbar um pouco o ponto fixo: $p_A(t) = p_A^* + \Delta p_t$ este se afasta, no instante $t+1$ movendo-se na direção de outro ponto fixo (instável) ou se aproxima (estável) do ponto fixo. (Mais interessante ainda) Existem pessoas que ao interagir com o grupo decidem ser do contra, isto é adotam no próximo instante a posição minoritária. Suponha que a probabilidade de alguém ser do contra seja c , e é independente de qualquer outra coisa. Então com a notação anterior, e usando as regras da soma e do produto mostre que
- (D1) $p_A(t+1) = (1-c)f(p_A(t)) + cf(p_B(t))$
- (D2) Mostre que para $c > 1/6$ o valor $p_A = .5$ é o único ponto fixo estável. Portanto podemos esperar um resultado da eleição em que a sociedade está dividida em frações aproximadamente iguais.

Valores: 4.