Prof. Leonardo Araújo

Lista de Exercícios

TEORIA DA INFORMAÇÃO

Engenharia de Telecomunicações

Universidade Federal de São João del-Rei

Sumário

1 Entropia

1.1 Entropia Máxima e Mínima

1. Qual é o menor e o maior valor para $H(p_1, p_2, \ldots, p_n) = H(p)$? Encontre os vetores p de dimensão n que forneçam o máximo e o mínimo. Justifique sua resposta.

Resolução:

A entropia será máxima quando a distribuição da fonte for uniforme, ou seja, $p_1 = p_2 = \ldots = p_n = 1/n$, onde $n = |\mathcal{X}|$, a cardinalidade do alfabeto da v.a. $X \sim p$.

$$
\begin{aligned}
H(X) - \log n &= -\sum_x p(x)\log p(x) - \log n \underbrace{\sum_x p(x)}_{=1} \\
&= -\sum_x p(x)\log p(x) - \sum_x p(x)\log n \\
&= -\sum_x p(x)\log p(x) n \\
&= \frac{1}{\ln 2}\sum_x p(x)\ln\frac{1}{p(x)n} \\
&\leq \frac{1}{\ln 2}\sum_x p(x)\left[\frac{1}{p(x)n} - 1\right] \\
&= \frac{1}{\ln 2}\left[\underbrace{\sum_x \frac{1}{n}}_{\sum_{x\in\mathcal{X}}\frac{1}{n}=n\frac{1}{n}=1} - \underbrace{\sum_x p(x)}_{=1}\right] = 0 \qquad (1)
\end{aligned}
$$

Logo, $H(X) - \log n \leq 0$ e assim $H(X) \leq \log n = \log|\mathcal{X}|$. A igualdade ocorrerá quando todos os eventos forem equiprováveis, ou seja, quando X possuir distribuição uniforme.

O limite inferior da entropia é 0 (zero), ou seja, $H(X) \geq 0$, com igualdade sse p for uma distribuição tal que existe apenas um evento possível, isto é, todos os p_i's, exceto um, são nulos. A distribuição p é da forma $(1,0,0,\ldots,0)$, $(0,1,0,\ldots,0)$, ..., $(0,0,0,\ldots,1)$, assim teremos,

$$H(X) = -\sum_x p(x)\log p(x) = 0\log 0 + \ldots + 0\log 0 + 1\log 1 = 0. \qquad (2)$$

Para os demais casos, teremos $p_i > 0$, $\forall 1 \leq i \leq n$, sendo assim, para cada um dos termos de $H(X)$, teremos $-p_i \log p_i > 0$ e consequentemente $H(X) > 0$.

1.2 $H(X,Y|Z) = H(X|Z) + H(Y|X,Z)$

1. Mostre que

$$H(X,Y|Z) = H(X|Z) + H(Y|X,Z). \qquad (3)$$

Resolução:

Temos que

$$p(x,y|z) = p(x|z)p(y|x,z), \qquad (4)$$

logo, tomando $-\log$ de ambos os lados, temos

$$-\log p(x,y|z) = -\log p(x|z) - \log p(y|x,z). \qquad (5)$$

Vamos agora tomar o valor esperado em relação a distribuição conjunta,

$$\begin{aligned}
-E_{(x,y,z)}\left[\log p(x,y|z)\right] &= E_{(x,y,z)}\left[\log p(x|z)\right] - E_{(x,y,z)}\left[\log p(y|x,z)\right] \\
&= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\sum_{z\in\mathcal{Z}} p(x,y,z)\log p(x|z) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}\sum_{z\in\mathcal{Z}} p(x,y,z)\log p(y|x,z) \\
&= -\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} \log p(x|z)\sum_{y\in\mathcal{Y}} p(x,y,z) \\
&\quad -\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} p(x,z)\sum_{y\in\mathcal{Y}} p(y|x,z)\log p(y|x,z) \\
&= -\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} p(x,z)\log p(x|z) + \sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{Z}} p(x,z)H(Y|X=x,Z=z) \\
&= H(X|Z) + H(Y|X,Z) \qquad (6)
\end{aligned}$$

1.3 Sub-estimação do desvio padrão

1. Sejam $x_1, x_2, \ldots, x_n$ amostras retiradas identicamente e independentemente (i.i.d) de uma determinada distribuição. Considere $\overline{x}$ a média amostral e $\widehat{\sigma} = \sqrt{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2/n}$ o desvio padrão amostral. Mostre que $E\left[\widehat{\sigma}\right] \leq \sigma$, onde σ é o desvio padrão populacional, ou seja, $\sigma = \sqrt{E\left[(X-\mu)^2\right]}$ e $\mu = E\left[X\right]$. (dica: utilize a desigualdade de Jensen).

Resolução:

Demonstração. Note que a função $\sqrt{\cdot}$ é uma função concava. Iremos então aplicar a desigualdade de Jensen: se f é uma função concava, então $E\left[f(X)\right] \leq f\left(E\left[X\right]\right)$.

$$\begin{aligned}
E\left[\widehat{\sigma}\right] &= E\left[\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}\right] \\
&\quad \text{aplicando a desigualdade de Jensen para a função concava } \sqrt{\cdot} \\
&\leq \sqrt{E\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2\right]} = \sigma \qquad (7)
\end{aligned}$$

■

1.4 Entropia condicional nula

1. Mostre que se $H(Y|X) = 0$, então Y é uma função de X, isto é, para todo x com probabilidade $p(x) > 0$, existe apenas um único valor de y com $p(x,y) > 0$.

Resolução: Vamos realizar uma demonstração por contradição.

Demonstração. Assuma que existe x_0, y_1 e y_2 tais que $p(x_0,y_1) > 0$ e $p(x_0,y_2) > 0$. Então

$$p(x_0) \geq \underbrace{p(x_0,y_1) + p(x_0,y_2)}_{\text{parcela da marginal}} > 0 \qquad (8)$$

Teremos as seguintes relações para as probabilidades condicionais:

$$p(y_1|x_0) = \frac{p(x_0, y_1)}{p(x_0)} > 0 \tag{9}$$

$$p(y_2|x_0) = \frac{p(x_0, y_2)}{p(x_0)} > 0 \tag{10}$$

ambos não são iguais a 0 nem a 1.

$$\begin{aligned}
H(Y|X) &= -\sum_x p(x) \sum_y p(y|x) \log p(y|x) \\
&\quad \text{tomando apenas um termo } x = x_0 \\
&\geq -p(x_0) \sum_y p(y|x_0) \log p(y|x_0) \\
&\quad \text{tomando apenas dois termos } y = y_1 \text{ e } y = y_2 \\
&\geq -\underbrace{p(x_0)}_{>0} \left(\underbrace{p(y_1|x_0) \log p(y_1|x_0)}_{<0} + \underbrace{p(y_2|x_0) \log p(y_2|x_0)}_{<0} \right) \\
&> 0
\end{aligned} \tag{11}$$

Teremos igualdade apenas se $y_1 = y_2$, ou seja, Y é função de X.

■

1.5 Métrica e Distância

1. Uma função $f : S \times S \to \mathbb{R}$ é uma métrica em S se, para todo $x, y, z \in S$, as seguintes condições são satisfeitas:

   - $f(x, y) \geq 0$ (não negatividade)
   - $f(x, y) = f(y, x)$ (simetria)
   - $f(x, y) = 0$ se e somente se $x = y$ (identidade dos indiscerníveis)
   - $f(x, y) + f(y, z) \geq f(x, z)$ (desigualdade triangular).

   Uma função $f : S \times S \to \mathbb{R}$ é uma distância em S se, para todo $x, y \in S$, as seguintes condições são satisfeitas:

   - $f(x, y) \geq 0$ (não negatividade)
   - $f(x, y) = f(y, x)$ (simetria)
   - $f(x, x) = 0$.

   (a) Mostre através de um exemplo que a divergência de Kullback-Leibler (entropia relativa) não é uma distância e sequer uma métrica.

   (b) Mostre que $\rho(X, Y) = H(X|Y) + H(Y|X)$ possui as propriedades de uma métrica. Note que $\rho(X, Y)$ é o número de bits necessário para X e Y comunicarem seus valores um para o outro.

   (c) Verifique que $\rho(X, Y)$ também pode ser expresso como

$$\begin{aligned}
\rho(X, Y) &= H(X) + H(Y) - 2I(X; Y) \\
&= H(X, Y) - I(X; Y) \\
&= 2H(X, Y) - H(X) - H(Y)
\end{aligned} \tag{12}$$

Resolução:

(a) A divergência de Kullback-Leibler não é simétrica. Exemplo: considere $p = (\frac{1}{2}, \frac{1}{2})$ e $q = (\frac{1}{4}, \frac{3}{4})$. Teremos assim:

$$D(p||q) = \frac{1}{2}\log\frac{1/2}{3/4} + \frac{1}{2}\log\frac{1/2}{1/4} = 1 - \frac{1}{2}\log 3 = 0.2075\text{bits}, \tag{13}$$

$$D(q||p) = \frac{3}{4}\log\frac{3/4}{1/2} + \frac{1}{4}\log\frac{1/4}{1/2} = \frac{3}{4}\log 3 - 1 = 0.1887\text{bits}. \tag{14}$$

Podemos observar que neste exemplo temos $D(p||q) \neq D(q||p)$, ou seja, a divergência de Kullback-Leibler não é simétrica.

(b)
- (não-negatividade) Como $H(X|Y) \geq 0$ e $H(Y|X) \geq 0$, teremos que $\rho(X,Y) \geq 0$.
- (simetria) Segue pela definição: $\rho(X,Y) = H(X|Y) + H(Y|X) = \rho(Y,X)$.
- (identidade dos indiscerníveis) Se $X = Y$, então $H(X|Y) = H(Y|X) = 0$ e assim $\rho(X,Y) = 0$. Por outro lado, se $\rho(X,Y) = 0$, então $H(X|Y) + H(Y|X) = 0$, mas $H(X|Y) \geq 0$ e $H(Y|X) \geq 0$, logo devemos ter $H(X|Y) = H(Y|X) = 0$, então $X = Y$.
- (desigualdade triangular)

$$\begin{aligned}
H(X|Y) + H(Y|Z) &\geq \text{condicionar não aumenta a entropia} \\
&\quad H(X|Y,Z) + H(Y|Z) \\
&= \text{regra da cadeia} \\
&\quad H(X,Y|Z) \\
&= \text{regra da cadeia} \\
&\quad H(X|Z) + \underbrace{H(Y|X,Z)}_{\geq 0} \\
&\geq H(X|Z)
\end{aligned} \tag{15}$$

da mesma forma podemos mostrar que $H(Z|Y) + H(Y|X) \geq H(Z|X)$. Teremos assim

$$\begin{aligned}
H(X|Y) + H(Y|Z) + H(Z|Y) + H(Y|X) &\geq H(X|Z) + H(Z|X) \\
\rho(X,Y) + \rho(Y,Z) &\geq \rho(X,Z)
\end{aligned} \tag{16}$$

(c)

$$\begin{aligned}
\rho(X,Y) &= \underbrace{H(X|Y)}_{H(X)-I(X;Y)} + \underbrace{H(Y|X)}_{H(Y)-I(X;Y)} \\
&= H(X) + H(Y) - 2I(X;Y) \\
&\quad \text{utilizando } H(X,Y) = H(X) + H(Y) - I(X;Y) \\
&= H(X,Y) - I(X;Y) \\
&\quad \text{utilizando } I(X;Y) = H(X) + H(Y) - H(X,Y) \\
&= 2H(X,Y) - H(X) - H(Y)
\end{aligned} \tag{17}$$

1.6 Entropia Conjunta

1. Seja $p(x,y)$ dado por

X \ Y	0	1
0	1/3	1/3
1	0	1/3

Calcule

(a) $H(X)$, $H(Y)$

(b) $H(X|Y)$, $H(Y|X)$

(c) $H(X,Y)$

(d) $H(Y) - H(Y|X)$

(e) $I(X;Y)$

(f) Faça um diagrama de Venn para representar as grandezas calculadas acima.

Resolução:

(a) Para calcular as entropia, precisamos primeiramente calcular as marginais.
$P(X=0) = \frac{2}{3}$ e $P(X=1) = \frac{1}{3}$.

$$\begin{aligned} H(X) &= -\sum_x p(x)\log p(x) \\ &= -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} \\ &= 0.38998 + 0.52832 = 0.91830 \text{ bits.} \end{aligned} \tag{18}$$

$P(Y=0) = \frac{1}{3}$ e $P(Y=1) = \frac{2}{3}$.

$$H(Y) = 0.91830 \text{ bits.} \tag{19}$$

(b)

$$\begin{aligned} H(X|Y) &= \sum_{y\in\mathcal{Y}} p(y)H(X|Y=y) \\ &= \frac{1}{3}H(X|Y=0) + \frac{2}{3}H(X|Y=1) \\ &= \frac{1}{3}H(1,0) + \frac{2}{3}H\left(\frac{1}{2},\frac{1}{2}\right) \\ &= 0 + \frac{2}{3} = \frac{2}{3} \end{aligned} \tag{20}$$

$$\begin{aligned} H(Y|X) &= \sum_{x\in\mathcal{X}} p(x)H(Y|X=x) \\ &= \frac{2}{3}H(Y|X=0) + \frac{1}{3}H(Y|X=1) \\ &= \frac{2}{3}H\left(\frac{1}{2},\frac{1}{2}\right) + \frac{1}{3}H(1,0) \\ &= \frac{2}{3} \end{aligned} \tag{21}$$

(c)

$$H(X,Y) = H(X) + H(Y|X) = 0.91830 + 0.66666 = 1.5850 \text{ bits} \tag{22}$$

$$H(X,Y) = H(Y) + H(X|Y) = 0.91830 + 0.66666 = 1.5850 \text{ bits} \tag{23}$$

(d)

$$H(Y) - H(Y|X) = 0.91830 - 0.66666 = 0.25164 \text{ bits} \tag{24}$$

(e)

$$I(X;Y) = H(Y) - H(Y|X) = 0.25164 \text{ bits} \tag{25}$$

(f)

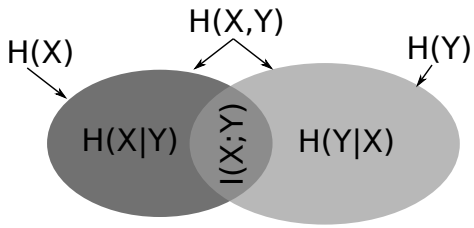


Figura 1: Diagrama de Venn.

1.7 $\ln x \leq x - 1$ e $\ln x \geq 1 - \frac{1}{x}$

1. Mostre as seguintes desigualdades abaixo, utilizando para tanto a expansão em série de Taylor.
   (a) $\ln x \leq x - 1$, para $x > 0$;
   (b) $\ln x \geq 1 - \frac{1}{x}$, para $x > 0$

Resolução:

(a) *Demonstração.* Considerando $f(x) = \ln x$ temos que

$$f^{(k)}(x) = (-1)^{k-1} \frac{(k-1)!}{x^k}. \tag{26}$$

A série de Taylor de uma função $f(\cdot)$ em torno de um ponto x_0 é dada por

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k. \tag{27}$$

Desta forma, a série de Taylor de $f(x) = \ln x$ em torno de $x_0 = 1$ será dada por

$$f(x) = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{(x-1)^k}{k}. \tag{28}$$

A função $f(\cdot)$ pode ser representada por

$$f(x) = f_n(x) + E_n(x), \tag{29}$$

onde f_n é a expansão em série de Taylor até o termo de ordem n e E_n o erro associado ao truncamento da série em n termos. Teremos que

$$E_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)^{(n+1)}, \tag{30}$$

onde $\xi \in [x, x_0]$.

Se tomarmos a aproximação apenas com o primeiro termo ($n = 1$), teremos

$$\ln x = (x - 1) + E_1(x), \tag{31}$$

onde $E_1(x) = (-1)(x-1)^2/2 < 0$, logo

$$\ln x \leq x - 1. \tag{32}$$

■

(b) *Demonstração.* Considere a expressão acima, dada na Equação (32) e aplique para $1/x$, obtendo assim:

$$\begin{aligned} \ln \frac{1}{x} &\leq \tfrac{1}{x} - 1 \\ -\ln x &\leq \tfrac{1}{x} - 1 \\ \ln x &\geq 1 - \tfrac{1}{x} \end{aligned} \tag{33}$$

■

1.8 Entropia da Soma

1. Seja X e Y variáveis aleatórias com valores $x_1, \ldots, x_r$ e $y_1, \ldots, y_s$ respectivamente. Seja $Z = X + Y$.
   (a) Mostre que $H(Z|X) = H(Y|X)$. Argumente que se X, Y são independentes, então $H(Y) \leq H(Z)$ e $H(X) \leq H(Z)$. Desta forma, a adição de variáveis aleatórias independentes traz incerteza.
   (b) Dê um exemplo de variáveis aleatória dependentes em que $H(X) > H(Z)$ e $H(Y) > H(Z)$.
   (c) Sob quais condições temos $H(Z) = H(X) + H(Y)$?

Resolução:

(a) *Demonstração.*

$$\begin{aligned} H(Z|X) &= \sum_x p(x) H(Z|X = x) \\ &= -\sum_x p(x) \sum_z p(Z = z|X = x) \log p(Z = z|X = x) \\ &= -\sum_x p(x) \sum_y p(Y = z - x|X = x) \log p(Y = z - x|X = x) \\ &= -\sum_x p(x) \sum_y p(Y = y|X = x) \log p(Y = y|X = x) \\ &= \sum_x p(x) H(Y|X = x) \\ &= H(Y|X) \end{aligned} \tag{34}$$

Se $X \perp\!\!\!\perp Y$, então $H(Y|X) = H(Y)$. Teremos assim

$$\begin{aligned} I(X;Z) &\geq 0 \\ H(Z) - H(Z|X) &\geq 0 \\ H(Z) - H(Y|X) &\geq 0 \\ H(Z) - H(Y) &\geq 0. \end{aligned} \tag{35}$$

Então, $H(Z) \geq H(Y)$. De forma semelhante podemos mostrar que $H(Z) \geq H(X)$. ■

(b) Considere

$$X = -Y = \begin{cases} 1 & \text{, com probabilidade } \frac{1}{2} \\ 0 & \text{, com probabilidade } \frac{1}{2} \end{cases} \tag{36}$$

Neste caso, teremos $H(X) = H(Y) = 1$ e $H(Z) = 0$, já que $Z = 0$ com probabilidade 1.

(c) Se Z for uma função de X e Y.

$$H(Z) \leq H(X,Y) \tag{37}$$

sendo que teremos igualdade se (X,Y) for uma função de Z.

$$H(X,Y) = H(X) + \underbrace{H(Y|X)}_{\leq H(Y)} \leq H(X) + H(Y) \tag{38}$$

com igualdade quando $X \perp\!\!\!\perp Y$.

Se (X,Y) for uma função de Z e $X \perp\!\!\!\perp Y$, teremos $H(Z) = H(X) + H(Y)$.

1.9 World Series

1. As Séries Mundiais (*World Series*) é uma série de sete jogos que termina assim que um dos times conseguir quatro vitórias. Seja X a variável aleatória que representa o resultado de uma Série Mundial entre os times A e B; possíveis valores para X são $AAAA$, $BABABAB$ e $BBBAAAA$. Seja Y o número de partidas disputadas, que varia de 4 a 7. Assumindo que A e B são times igualmente bons (mesma probabilidade de vencer) e que as partidas são independentes, calcule $H(X)$, $H(Y)$, $H(Y|X)$ e $H(X|Y)$.

Resolução: A probabilidade de um time ganhar uma partida qualquer é 1/2.

As partidas podem ter de 4 a 7 jogos.

- Existem 2 partidas de 4 jogos, são elas $AAAA$ e $BBBB$, todas com probabilidade 2^{-4}.
- Existem $8 = 2\binom{4}{3}$ particas de 5 jogos: $BAAAA$, $ABAAA$, ..., $ABBBB$, com probabilidade 2^{-5} (o último da sequência deve ser necessária mente o time que ganhou a série, logo sobraram 4 posições para dispor as demais 3 partidas do time vencedor).
- Exitem $20 = 2\binom{5}{3}$ particas de 6 jogos, com probabilidade 2^{-6}.
- Exitem $40 = 2\binom{6}{3}$ particas de 7 jogos, com probabilidade 2^{-7}.

$$
\begin{aligned}
H(X) &= -\sum p(x)\log p(x) \\
&= 2\frac{1}{16}\log 16 + 8\frac{1}{32}\log 32 + 20\frac{1}{64}\log 64 + 40\frac{1}{128}\log 128 \\
&= 5.81 \text{ bits.}
\end{aligned} \tag{39}
$$

$$
\begin{aligned}
H(Y) &= -\sum p(y)\log p(y) \\
&= \frac{1}{8}\log 8 + \frac{1}{4}\log 4 + \frac{5}{16}\log\frac{16}{5} + \frac{5}{16}\log\frac{16}{5} \\
&= 1.92 \text{ bits.}
\end{aligned} \tag{40}
$$

Y é uma função determinística de X, logo conhecendo X não resta incerteza quanto a Y, ou seja, $H(Y|X) = 0$.

Como $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$, teremos $H(X|Y) = H(X) - H(Y) = 3.89$ bits.

1.10 Uma Medida de Correlação

1. Sejam X_1 e X_2 identicamente distribuídos ($\sim p$), mas não necessariamente independentes. Considere

$$
\rho = 1 - \frac{H(X_2|X_1)}{H(X_1)}. \tag{41}
$$

   (a) Mostre que $\rho = I(X_1; X_2)/H(X_1)$.
   (b) Mostre que $0 \leq \rho \leq 1$.
   (c) Quando teremos $\rho = 0$?
   (d) Quando teremos $\rho = 1$?

Resolução:

(a) *Demonstração.*

$$
\begin{aligned}
\rho &= 1 - \frac{H(X_2|X_1)}{H(X_1)} \\
&= \frac{H(X_1) - H(X_2|X_1)}{H(X_1)} \\
&= \frac{H(X_2) - H(X_2|X_1)}{H(X_1)} \quad X_1 \text{ e } X_2 \sim p \\
&= \frac{I(X_1; X_2)}{H(X_1)}
\end{aligned} \tag{42}
$$

■

(b) *Demonstração.* Como $0 \leq H(X_2|X_1) \leq H(X_2) = H(X_1)$, teremos

$$
0 \leq \frac{H(X_2|X_1)}{H(X_1)} \leq 1, \tag{43}
$$

e assim $0 \leq \rho \leq 1$. ■

(c) Teremos $\rho = 0$ sse $X_2 \perp\!\!\!\perp X_1$, pois neste caso $H(X_2|X_1) = H(X_2)$. Como X_2 e X_1 possuem a mesma distribuição, teremos $H(X_1) = H(X_2)$ e assim $\rho = 0$.

(d) $\rho = 1$ sse $H(X_2|X_1) = 0$, ou seja, se X_2 for uma função de X_1 e X_1 for uma função de X_2, ou seja, uma função bijetiva.

1.11 Processamento de Dados

1. Considere que $X_1 \to X_2 \to X_3 \to \ldots \to X_n$ seja uma cadeia de Markov nesta ordem, isto é,

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2|x_1)\ldots p(x_n|x_{n-1}). \tag{44}$$

Simplifique $I(X_1; X_2; \ldots; X_n)$.

Resolução: Pela regra da cadeia da informação mútua, temos

$$I(X_1; X_2; \ldots; X_n) = I(X_1; X_2) + I(X_1; X_3|X_2) + \ldots + I(X_1; X_n|X_2; \ldots, X_{n-2}). \tag{45}$$

Pela desigualdade de Markov, o passado e o futuro são independentes, dado o presente. Desta forma, todos os termos da equação anterior são nulos, exceto o termo $I(X_1; X_2)$. Teremos assim

$$I(X_1; X_2; \ldots; X_n) = I(X_1; X_2). \tag{46}$$

1.12 Mistura aumenta entropia

1. Mostre que a entropia da distribuição de probabilidade $(p_1, \ldots, p_i, \ldots, p_j, \ldots, p_m)$ é menor do que a entropia da distribuição $(p_1, \ldots, \frac{p_i+p_j}{2}, \ldots, \frac{p_i+p_j}{2}, \ldots, p_m)$. Mostre que, em geral, qualquer transferência de probabilidade, que torne a distribuição mais uniforme, aumenta a entropia.

Resolução:

Para este problema iremos utilizar que a função logaritmo é concava para argumento maior que zero, ou seja,

$$\log(\alpha x_1 + (1-\alpha)x_2) \geq \alpha \log(x_1) + (1-\alpha)\log(x_2). \tag{47}$$

Temos

$$P_1 = (p_1, \ldots, p_i, \ldots, p_j, \ldots, p_m) \text{ e} \tag{48}$$

$$P_2 = (p_1, \ldots, \frac{p_i + p_j}{2}, \ldots, \frac{p_i + p_j}{2}, \ldots, p_m). \tag{49}$$

Queremos mostrar que $H(P_2) \geq H(P_1)$, ou seja, $H(P_2) - H(P_1) \geq 0$.

$$\begin{aligned}
H(P_2) - H(P_1) &= -p_1 \log p_1 - \ldots - \frac{p_i + p_j}{2} \log \left(\frac{p_i + p_j}{2} \right) \\
&\quad - \frac{p_i + p_j}{2} \log \left(\frac{p_i + p_j}{2} \right) - \ldots - p_m \log p_m \\
&\quad + p_1 \log p_1 + \ldots + p_i \log p_i + p_j \log p_j + \ldots + p_m \log p_m \\
&= -2(p_i + p_j) \log \left(\frac{p_i + p_j}{2} \right) + p_i \log p_i + p_j \log p_j \\
&\geq -2(p_i + p_j) \left(\frac{1}{2} \log p_i + \frac{1}{2} \log p_j \right) + p_i \log p_i + p_j \log p_j \\
&= -p_j \log p_i - p_i \log p_j \geq 0 \qquad (50)
\end{aligned}$$

1.13 Retirando elementos com e sem reposição

1. Uma urna contém r bolas vermelhas, w bolas brancas e b bolas pretas. Qual das opções possui maior entropia? 1) Retirar $k \geq 2$ bolas da urna com substituição ou 2) sem substituição?

   Bernoulli para ter sido o primeiro a utilizar o modelo de urnas no estudo das probabilidades. A inspiração de Bernoulli pode ter sido as loterias, eleições, ou jogos de azar, que envolvem sortear as bolas de um recipiente. Tem sido reconhecido que nas eleições na época medieval e renascentista de Veneza, incluindo a de Doge[1], muitas vezes incluíam a escolha dos eleitores por sorteio, usando bolas de cores diferentes, extraídas de uma urna.

Resolução: A resposta é simples, pois ao utilizar a estratégia de substituição, teremos que o número de possíveis escolhas em cada etapa será a mesma, ao passo que, ao adotar a estratégia sem substituição, a cada iteração teremos o número de possíveis escolhas reduzido e, por conseguinte, teremos menos entropia.

Para o caso 1) com substituição, teremos $H(X_i|X_{i-1}, \ldots, X_1) = H(X_i)$. Podemos calcular $H(X_i)$:

$$H(X_i) = H\left(\frac{r}{r+w+b}, \frac{w}{r+w+b}, \frac{b}{r+w+b} \right) \qquad (51)$$

Calcular a entropia condicional no caso 2) sem substituição é mais complicado.

$$\begin{aligned}
H(X_2|X_1) &= \sum_x p(x) H(Y|X = x) \\
&= \frac{r}{r+w+b} H(Y|X = \text{vermelha}) + \frac{w}{r+w+b} H(Y|X = \text{branca}) \\
&\quad + \frac{b}{r+w+b} H(Y|X = \text{preta}) \\
&= \frac{r}{r+w+b} H\left(\frac{r-1}{r+w+b-1}, \frac{w}{r+w+b-1}, \frac{b}{r+w+b-1} \right) + \\
&\quad \frac{w}{r+w+b} H\left(\frac{r}{r+w+b-1}, \frac{w-1}{r+w+b-1}, \frac{b}{r+w+b-1} \right) + \\
&\quad \frac{b}{r+w+b} H\left(\frac{r}{r+w+b-1}, \frac{w}{r+w+b-1}, \frac{b-1}{r+w+b-1} \right) \qquad (52)
\end{aligned}$$

[1] Doge é a denominação do chefe ou primeiro magistrado eleito, das antigas repúblicas de Gênova e Veneza.

Para calcular as entropias condicionais seguintes seria ainda bem mais complicado.

1.14 Desigualdade de Fano

1. Para uma dado canal de comunicação, X é emitido pela fonte e Y é recebido pelo receptor. É dada a seguinte probabilidade conjunta (X, Y):

X \ Y	1	2	3
1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$

Considere $\widehat{X}(Y)$ um estimador para X (baseado em Y) e $P_e = \Pr\{\widehat{X}(Y) \neq X\}$.

(a) Encontre o estimador $\widehat{X}(Y)$ que fornece a menor probabilidade de erro e calcule P_e associado.

(b) Avalie a desigualdade de Fano para este problema e compare os resultados.

Resolução:

(a) O melhor estimador é $\widehat{X}(Y) = Y$.

A probabilidade de erro para este estimador será dada por

$$\begin{aligned} P_e &= \Pr\{\widehat{X}(Y) \neq X\} = 1 - \Pr\{\widehat{X}(Y) = X\} \\ &= 1 - (\Pr\{Y = 1, X = 1\} + \Pr\{Y = 2, X = 2\} + \Pr\{Y = 3, X = 3\}) \\ &= 1 - \left(3 \times \frac{1}{6}\right) \\ &= \frac{1}{2} \end{aligned} \tag{53}$$

(b) Desigualdade de Fano:

$$P_e \geq \frac{H(X|Y) - 1}{\log(|\mathcal{X}| - 1)} \tag{54}$$

Precisamos calcular $H(X|Y)$. Para tanto precisamos calcular a marginal $p(y)$ e a condicional $p(x|y)$. Temos que $p(y) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ e $p(x|y) = \frac{p(x,y)}{p(y)}$:

X \ Y	1	2	3
1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
2	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
3	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

$$\begin{aligned} H(X|Y) &= \sum_y p(y) H(X|Y = y) \\ &= -\sum_y p(y) \sum_x p(x|y) \log p(x|y) \\ &= -3 \times \frac{1}{3} \left(2 \times \frac{1}{2} \log \frac{1}{2} + \frac{1}{4} \log \frac{1}{4}\right) \\ &= -\left(-1 - \frac{1}{2}\right) = \frac{3}{2} = 1.5 \end{aligned} \tag{55}$$

Teremos assim

$$P_e \geq \frac{H(X|Y) - 1}{\log\left(|\mathcal{X}| - 1\right)} = \frac{1.5 - 1}{\log(3-1)} = 0.5 \tag{56}$$

Verificamos assim que o estimador dado anteriormente realmente fornece a menor probabilidade de erro, o que é comprovado pela desigualdade de Fano.

1.15 Canal de comunicação

1. Seja um canal de comunicação caracterizado por $p(y|x)$ dado abaixo

X \ Y	0	1
0	1/4	3/4
1	2/3	1/3

e uma fonte com distribuição dada por $p(x) = (4/7, 3/7)$.

(a) Calcule os valores das entropias $H(X)$ e $H(Y)$.

(b) Calcule a informação mútua $I(X;Y)$.

(c) Calcule as entropias condicionais $H(X|Y)$ e $H(Y|X)$, e também a entropia conjunta $H(X,Y)$.

(d) Seja $\widehat{X}$ um estimador para X baseado em Y. Qual é o melhor estimador? Qual é a probabilidade de erro para este estimador? Compare com o limite imposto pela desigualdade de Fano.

Resolução:

(a) $H(X)$ pode ser calculado diretamente.

$$\begin{aligned}
H(X) &= -\sum_{x\in\mathcal{X}} p(x)\log p(x) && (57)\\
&= -\frac{4}{7}\log\frac{4}{7} - \frac{3}{7}\log\frac{3}{7} && (58)\\
&= \frac{4}{7}(\log 7 - \log 4) + \frac{3}{7}(\log 7 - \log 3) && (59)\\
&= \log 7 - \frac{8}{7} - \frac{3}{7}\log 3 && (60)\\
&= 0.98523 \text{ bits.} && (61)
\end{aligned}$$

Para calcular os demais itens iremos precisar da distribuição conjunta $p(x,y) = p(x)p(y|x) = p(y)p(x|y)$ dada por

X \ Y	0	1
0	1/7	3/7
1	2/7	1/7

e também da marginal $p(y) = (3/7, 4/7)$.

Como a distribuição de Y é apenas uma permutação da distribuição de X, teremos $H(Y) = H(X)$.

(b) Utilizando a distribuição conjunta calculada no item anterior, temos

$$\begin{aligned}
I(X;Y) &= \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\
&= \frac{1}{7}\log\left(\frac{1/7}{4/7\times 3/7}\right)+\frac{3}{7}\log\left(\frac{3/7}{4/7\times 4/7}\right)+\frac{2}{7}\log\left(\frac{2/7}{3/7\times 3/7}\right) \\
&\quad +\frac{1}{7}\log\left(\frac{1/7}{3/7\times 4/7}\right) \\
&= \frac{1}{7}\log\left(\frac{7}{12}\right)+\frac{3}{7}\log\left(\frac{21}{16}\right)+\frac{2}{7}\log\left(\frac{14}{9}\right)+\frac{1}{7}\log\left(\frac{7}{12}\right) \\
&= \log 7\left(\frac{1}{7}+\frac{3}{7}+\frac{2}{7}+\frac{1}{7}\right)+\log 3\left(-\frac{1}{7}+\frac{3}{7}-\frac{2}{7}\times 2-\frac{1}{7}\right) \\
&\quad +\left(-\frac{1}{7}\times 2-\frac{3}{7}\times 4+\frac{2}{7}-\frac{1}{7}\times 2\right) \\
&= \log 7-\frac{3}{7}\log 3-2 \\
&= 0.12809 \text{ bits.} \qquad (62)
\end{aligned}$$

(c)

$$\begin{aligned}
H(X,Y) &= H(X)+H(Y)-I(X;Y) \\
&= 0.98523+0.98523-0.12809=1.8424 \\
&= \log 7-\frac{8}{7}-\frac{3}{7}\log 3+\log 7-\frac{8}{7}-\frac{3}{7}\log 3-\left(\log 7-\frac{3}{7}\log 3-2\right) \\
&= \log 7-\frac{3}{7}\log 3-\frac{2}{7}=1.8424 \qquad (63)
\end{aligned}$$

$$\begin{aligned}
H(Y|X) &= H(X,Y)-H(X)=0.85717 \qquad (64) \\
&= \log 7-\frac{3}{7}\log 3-\frac{2}{7}-\left(\log 7-\frac{8}{7}-\frac{3}{7}\log 3\right) \qquad (65) \\
&= \frac{6}{7} \qquad (66)
\end{aligned}$$

$$H(X|Y)=H(X,Y)-H(Y)=\frac{6}{7}=0.85717 \qquad (67)$$

(d) O melhor estimador é $\widehat{X}=1-Y$, ou seja,

$$\widehat{X}=\begin{cases}0 & \text{quando } Y=1,\\ 1 & \text{quando } Y=0.\end{cases} \qquad (68)$$

A probabilidade de erro é dada por

$$\begin{aligned}
P_e &= \Pr\{\widehat{X}(Y)\neq X\} \\
&= \Pr\{Y=0,X=0\}+\Pr\{Y=1,X=1\} \\
&= 2\times\frac{1}{7}=\frac{2}{7} \qquad (69)
\end{aligned}$$

A desigualdade de Fano estabelece que

$$H(P_e) + P_e \log(|\mathcal{X}| - 1) \geq H(X|Y) \tag{70}$$

Vamos utilizar a forma fraca da desigualdade de Fano, nos valendo de que $H(P_e) \leq 1$ e $\log(|\mathcal{X}| - 1) \leq \log(|\mathcal{X}|)$. A forma fraca da desigualdade de Fano fornece o seguinte limite para a probabilidade de erro:

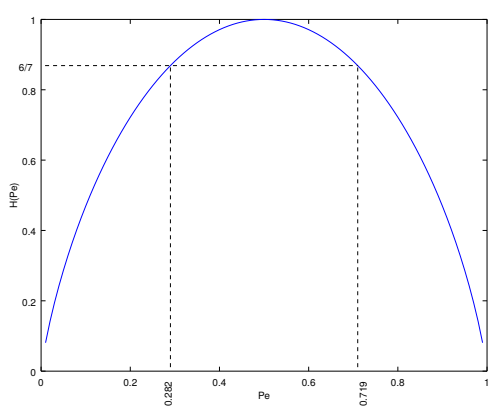
$$P_e \geq \frac{H(X|Y) - 1}{\log(|\mathcal{X}|)} = \frac{\frac{6}{7} - 1}{\log 2} = -\frac{1}{7}. \tag{71}$$

A desigualdade encontrada não agrega informação, pois já sabemos previamente que $P_e \geq 0$. Se analisarmos a desigualdade de Fano para o caso em que $\mathcal{X}$ é um alfabeto binário, teremos

$$\begin{aligned} H(P_e) + P_e \log(|\mathcal{X}| - 1) &\geq H(X|Y) \\ H(P_e) + P_e \log(2 - 1) &\geq H(X|Y) \\ H(P_e) &\geq H(X|Y) \\ H(P_e) &\geq \frac{6}{7} \end{aligned} \tag{72}$$

Podemos ainda encontrar limites para P_e analisando $H(P_e)$.

```
Pe = [0.01:0.01:0.99];
HPe = -Pe.*log2(Pe) -(1-Pe).*log2(1-Pe);
id1 = find(HPe > 6/7, 1, 'first');
id2 = find(HPe > 6/7, 1, 'last');
figure; plot(Pe,HPe);
xlabel('Pe'); ylabel('H(Pe)');
line([Pe(id1) Pe(id1)],[0 HPe(id1)],'LineStyle','--');
line([Pe(id2) Pe(id2)],[0 HPe(id2)],'LineStyle','--');
line([0 Pe(id2)],[HPe(id1)  HPe(id1)],'LineStyle','--');
p1 = 0.001; H1 = 0;
while H1 < 6/7, p1=p1+0.001; H1=-p1.*log2(p1)-(1-p1).*log2(1-p1); end;
p2 = 0.6; H2 = 1;
while H2 > 6/7, p2=p2+0.001; H2=-p2.*log2(p2)-(1-p2).*log2(1-p2); end;
text(p1,-0.1,num2str(p1),'rotation',90);
text(p2,-0.1,num2str(p2),'rotation',90);
text(-0.05,6/7+0.01,'6/7');
axis([0 1 0 1]);
print -dsvg errorentropy.svg;
system('inkscape errorentropy.svg --export-pdf=errorentropy.pdf');
```



Podemos observar então que $0.282 \leq P_e \leq 0.719$ satisfaz a condição $H(P_e) \geq 6/7$. O valor exato de P_e encontrado anteriormente, $P_e = 2/7 = 0.28571$ está neste intervalo.

1.16 Dados de Lurian

1. Lurian vai participar de um jogo de dados. Neste jogo, todos os dados devem possuir 6 lados. Lurian é muito 'esperta' e adiciona um pequeno peso a seu dado, de forma que o lado 6 passe a ter probabilidade maior do que os demais. Após utilizar este artifício, a função massa de probabilidade que rege o laçamento de seu dado será $p = (p_1, \ldots, p_5, p_6) = \left(\frac{1}{7}, \ldots, \frac{1}{7}, \frac{2}{7}\right)$. Desta forma, o lado de valor 6 terá o dobro da probabilidade que cada um dos demais.
   (a) Calcule a entropia associada a um lance do dado de Lurian.
   (b) Suponha que você queira gerar uma sequência de m lances de dados, mas seria necessário que o dado fosse honesto. Entretanto você dispõe apenas do dado fornecido por Lurian. Proponha uma maneira de utilizar o dado fornecido por Lurian e gerar uma sequência equivalente a uma sequencia proveniente de um dado honesto. Como podemos determinar a abordagem para resolver este problema que forneça o menor *overhead* [2] possível?

Resolução:

(a)

$$\begin{aligned} H(X) &= -\sum_{x \in \mathcal{X}} p(x) \log p(x) \\ &= -1/7 \times \log(1/7) - \ldots - 1/7 \times \log(1/7) - 2/7 \times \log(2/7) \\ &= (5/7) \times \log(7) + (2/7) \times \log(7) - (2/7) \\ &= \log(7) - (2/7) \approx 2.521 \end{aligned} \tag{73}$$

(b) Vamos mapear sequências de comprimento n fornecidas pelo dado de Lurian em sequências de comprimento m. De forma geral, devemos ter $n > m$, uma vez que a entropia associada ao dado de Lurian é menor que a entropia de um dado honesto. O *overhead* será mínimo quando $n = m + 1$ e tomando m maior possível, de forma a minimizar o *overhead* que será de $1/m$. Devemos então encontrar o maior m tal que $n = m + 1$ e a entropia associada a uma sequência de comprimento m para um dado honesto (mH_1, onde H_1 é a entropia de um dado honesto e estamos considerando que os lances de dados são independentes) deverá ser menor ou igual à entropia associada a uma sequência de comprimento n para o dado de Lurian (nH_2, onde estamos chamando de H_2 a entropia associada ao lance do dado de Lurian e considerando que os lances independentes).

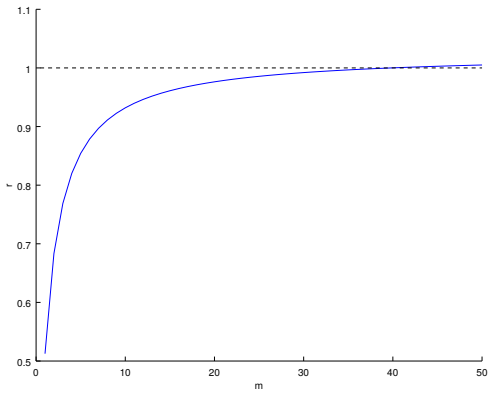
```
% Lurian
p = [1/7 1/7 1/7 1/7 1/7 2/7];
H2 = - sum( p.* log2(p) )

% Honesto
p = 1/6 * ones(1,6);
H1 = - sum( p.* log2(p) )

m=[1:50]; r = m*H1./((m+1)*H2);
h = figure;
plot(m,r);
```

[2] *overhead*: custo excedente por unidade.

```
line([0 50], [1 1], 'LineStyle','--');
xlabel('m'); ylabel('r');
set(gca, 'Box', 'off');
print -dsvg dicelurian.svg;
system('inkscape dicelurian.svg --export-pdf=dicelurian.pdf');
```



Podemos verificar então que o valor máximo de m é 39, para o qual a relação $mH_1 \leq nH_2$ é satisfeita. Para este caso ($m = 39$) teremos um *overhead* de apenas 2.5%.

1.17 Caça Níquel

1. Diante da crise econômica, o Palácio do Planalto está elaborando uma pesquisa ampla para "avaliar os impactos da eventual liberação de cassinos no Brasil e os possíveis modelos de exploração de jogos de azar" [3]. Você foi escolhido para planejar a máquina caça níquel tupiniquim. Esta máquina deverá funcionar da seguinte maneira: para jogar, você precisa inserir uma moeda de R$1,00 e poderá ter 3 (três) resultados distintos: 1) perder (não recebe de volta o R$1,00 depositado), ou seja, sai no prejuízo de R$1,00; 2) ganhar R$1,00 (apenas recebe de volta o que foi depositado), não sai no prejuízo, nem no lucro; 3) ganhar R$5,00 (recebe o R$1,00 depositado acrescido de uma bonificação de R$4,00), sai com um lucro de R$4,00. Esta máquina deverá ser projetada atendendo ainda a dois requisitos: a) o lucro esperado do jogador deverá ser nulo (R$0,00); b) os resultados produzidos pela máquina deverão apresentar incerteza máxima, ou seja, a maior aleatoriedade possível. Vamos chamar de $p = (p_1, p_2, p_3)$ a distribuição de massa sobre os possíveis resultados (1), (2) e (3), conforme acima descritos. Determine como devemos proceder para encontrar essa distribuição de massa de forma a satisfazer também os requisitos (a) e (b). Escreva as equações e relações pertinentes e proponha uma forma (algorítmica e algébrica) para encontrar esta distribuição de massa. Encontre algebricamente a distribuição de massa desejada.

Resolução: Devemos encontrar a distribuição que maximiza a entropia, sujeito ao valor esperado do lucro ser nulo. Vamos definir a v.a. X com sendo o lucro. Teremos então $p_1 = P(X = -1)$, $p_2 = P(X = 0)$ e $p_3 = P(X = 4)$.

Queremos que o lucro esperado seja nulo,

$$E[X] = p_1 \times (-1) + p_2 \times 0 + p_3 \times 4 = 0. \tag{74}$$

[3] http://g1.globo.com/politica/noticia/2016/01/governo-faz-estudo-sobre-impacto-da-liberacao-de-cassino-e-bingo-no-brasil.html

Assim podemos concluir que $p_1 = 4p_3$. Além disso, temos que $p_1 + p_2 + p_3 = 1$, logo teremos $p_2 = 1 - 5p_3$. Podemos escrever então

$$\begin{aligned} H(p) &= H(p_1, p_2, p_3) && (75)\\ &= H(4p_3, 1-5p_3, p_3) && (76)\\ &= -4p_3 \log 4p_3 - (1-5p_3)\log(1-5p_3) - p_3 \log p_3. && (77) \end{aligned}$$

Note que, devemos ter $p_i \in [0,1]$, $i = 1,2,3$. Podemos assim concluir que $p_3 \in [0, 0.2]$ para que os demais p_i também estejam no intervalo $[0,1]$.

Sabemos que $H(p)$ é uma função concava em p. Isto implica a existência de um ponto de máximo absoluto. Podemos encontrar este ponto de máximo através de um algoritmo de maximização (como o método do gradiente) ou algebricamente.

Para determinar algebricamente, devemos encontrar o valor de p_3 que faça $\partial H/\partial p_3 = 0$,

$$\begin{aligned} \frac{\partial H}{\partial p_3} &= -4\log 4p_3 - 4p_3 \frac{1}{4p_3} 4 - (-5)\log(1-5p_3) - (1-5p_3)\frac{1}{(1-5p_3)}(-5) - \log p_3 - p_3 \frac{1}{p_3} && (78)\\ &= -4\log 4p_3 + 5\log(1-5p_3) - \log p_3 = 0 && (79) \end{aligned}$$

ou seja, devemos ter

$$\begin{aligned} \log \frac{(1-5p_3)^5}{(4p_3)^4 p_3} &= 0 && (80)\\ \log \frac{(1-5p_3)^5}{4^4 p_3^5} &= 0 && (81)\\ \frac{(1-5p_3)^5}{4^4 p_3^5} &= 1 && (82)\\ \frac{1-5p_3}{4^{4/5} p_3} &= 1 && (83)\\ 1-5p_3 &= 2^{8/5} p_3 && (84)\\ p_3 &= \frac{1}{5+2^{8/5}} \approx 0.12451. && (85) \end{aligned}$$

Podemos agora determinar p_1 e p_2 utilizando as relações vistas anteriormente. Teremos assim

$$p_1 = \frac{4}{5+2^{8/5}} \quad (86) \qquad p_2 = \frac{2^{8/5}}{5+2^{8/5}} \quad (87) \qquad p_3 = \frac{1}{5+2^{8/5}} \quad (88)$$

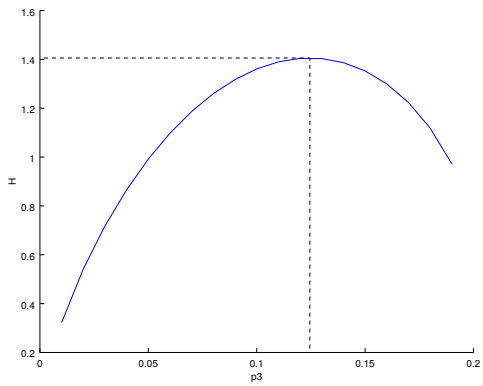
Abaixo é apresentado um código onde é utilizado o método do gradiente para encontrar numericamente a solução.

```
p = [0.01 : 0.01 : 0.19];
H = - 4*p.*log2(4*p) - (1-5*p).*log2(1- 5*p) - p.*log2(p);
figure; plot(p,H);
xlabel('p3'); ylabel('H');

% metodo gradiente
alfa = 0.01; eps = 1E-3;
p3 = 0.1;
for i=1:1E3,
    dH = -4*log2(4*p3) + 5*log2(1-5*p3) - log2(p3);
    p3 = p3 + alfa * dH;
    if abs(dH) < eps, break; end;
end;
```

```
H3 = - 4*p3.*log2(4*p3) - (1-5*p3).*log2(1- 5*p3) - p3.*log2(p3);
line([0 p3], [H3 H3], 'LineStyle','--');
line([p3 p3], [0.2 H3], 'LineStyle','--');

set(gca, 'Box', 'off');
print -dsvg niquel.svg;
system('inkscape niquel.svg --export-pdf=niquel.pdf');
% comparacao com o resultado analitico
p3 - 1 / (2^(8/5) + 5)
```



1.18 Entropia e Informação Mútua

1. Uma fonte produz símbolos aleatórios $X \in \mathcal{X} = \{a, b, c, d\}$ que serão enviados através de um canal de comunicação ruidoso. A saída do canal de comunicação é a variável aleatória $Y \in \mathcal{Y} = \{a, b, c, d\}$. A distribuição conjunta entre essas duas v.a.s é a seguinte

	$x = a$	$x = b$	$x = c$	$x = d$
$y = a$	1/8	1/16	1/16	1/4
$y = b$	1/16	1/8	1/16	0
$y = c$	1/32	1/32	1/16	0
$y = d$	1/32	1/32	1/16	0

   (a) Encontre a distribuição marginal de X e calcule a entropia marginal $H(X)$ em bits.

   (b) Encontre a distribuição marginal de Y e calcule a entropia marginal $H(Y)$ em bits.

   (c) Calcule a entropia conjunta $H(X, Y)$ em bits.

   (d) Calcule a entropia condicional $H(Y|X)$ em bits.

   (e) Calcule a informação mútua $I(X; Y)$ em bits.

Resolução:

(a) to-do

(b) to-do

(c) to-do

(d) to-do

(e) to-do

1.19 Entropia de um alfabeto

1. Uma fonte produz uma v.a. X em um alfabeto $\mathcal{X} = \{0, 1, 2, \ldots, 9, a, b, c, \ldots, z\}$, sendo que, com probabilidade de 1/3 teremos um número natural $\{0, 1, 2, \ldots, 9\}$; com probabilidade de 1/3 teremos uma vogal $\{a, e, i, o, u\}$; e com probabilidade de 1/3 teremos uma consoante $\{b, c, d, \ldots, z\}$. Todos os numerais são equiprováveis, assim como todas as vogais e todas as consoantes. Determine a entropia de X (não é necessário calcular os logaritmos de números que não sejam potência de 2).

Resolução:

$$\begin{aligned} H(X) &= H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) + \frac{1}{3}\left(H(\text{numerais}) + H(\text{vogais}) + H(\text{consoantes})\right) \\ &= \log 3 + \frac{1}{3}\left(\log 10 + \log 5 + \log 21\right) = \log 3 + \frac{1}{3}\left(\log 2 + 2\log 5 + \log 3 + \log 7\right) \\ &= \frac{1}{3}\left(1 + 4\log 3 + 2\log 5 + \log 7\right) \end{aligned} \tag{89}$$

Podemos também resolver da seguinte forma:

Calcular as probabilidades dos símbolos: $p_{\text{numeral}} = \frac{1}{30}$, $p_{\text{vogais}} = \frac{1}{15}$ e $p_{\text{consoantes}} = \frac{1}{63}$.

Teremos então

$$\begin{aligned} H(X) &= -\sum_{x \in \mathcal{X}} p(x) \log p(x) \\ &= 10 \times \left(-\frac{1}{30}\log\frac{1}{30}\right) + 5 \times \left(-\frac{1}{15}\log\frac{1}{15}\right) + 21 \times \left(\frac{1}{63}\log\frac{1}{63}\right) \\ &= \frac{1}{3}\log 30 + \frac{1}{3}\log 15 + \frac{1}{3}\log 63 \\ &= \frac{1}{3}(\log 3 + \log 10) + \frac{1}{3}(\log 3 + \log 5) + \frac{1}{3}(\log 3 + \log 21) \\ &= \log 3 + \frac{1}{3}(\log 10 + \log 5 + \log 21) = \frac{1}{3}\left(1 + 4\log 3 + 2\log 5 + \log 7\right) \end{aligned} \tag{90}$$

1.20 Quantizador

1. Na conversão analógico-digital é utilizado um quantizador Q. O quantizador mapeia os valores de sua entrada X em um conjunto finito de valores (níveis de quantização) $Y \in \{y_1, y_2, \ldots, y_N\}$. Como podemos determinar a informação mútua $I(X;Y)$ conhecendo a distribuição da saída Y? É possível utilizar $I(X;Y)$ para determinar a entropia de X? Justifique sua resposta.

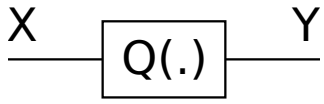


Figura 2: Quantizador.

Resolução: Como a saída Y é uma função de X, $Y = Q(X)$, teremos que a incerteza sobre Y dado X será nula, $H(Y|X) = 0$. Desta forma, teremos

$$\begin{aligned} I(X;Y) &= H(Y) - \underbrace{H(Y|X)}_{=0} \\ &= H(Y) \\ &= -\sum_{y\in\mathcal{Y}} p(y)\log p(y) \end{aligned} \tag{91}$$

Conhecendo a distribuição $p(y)$ poderemos calcular a informação mútua $I(X;Y)$.

Por outro lado, não podemos afirmar que X é uma função de Y (de forma geral, isto não ocorrerá), assim $H(X|Y)$ não será nula e não teremos como determinar $H(X)$ conhecendo apenas $p(y)$.

1.21 Entropia Cruzada

1. A entropia cruzada entre duas distribuições p e q, sobre um mesmo alfabeto $\mathcal{X}$, mede a número esperado de bits necessários para identificar um símbolo do alfabeto se for utilizado um esquema de codificação otimizado para a distribuição q, onde q é uma estimativa para a real distribuição dos símbolos, p. A entropia cruzada de p e q é definida então como

$$H(p,q) = E_p\left[-\log q\right]. \tag{92}$$

   (a) Mostre que $H(p,q) = H(p) + D(p||q)$.
   (b) Mostre que $H(p,q) \geq 0$.
   (c) Mostre que $H(p,u) = \log|\mathcal{X}|$, onde u é a distribuição uniforme.
   (d) Mostre que a entropia da distribuição p é menor ou igual à entropia cruzada de p e q (desigualdade de Gibbs).
   (e) Suponha que $|\mathcal{X}| = 4$, $p = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right)$ e q é uniforme. Calcule $H(p,q)$, $H(q,p)$, $H(p)$, $H(q)$, $D(p||q)$ e $D(q||p)$.

Resolução:

(a) *Demonstração.*

$$\begin{aligned} H(p,q) &= E_p\left[-\log q\right] \\ &= -\sum_{x\in\mathcal{X}} p(x)\log q(x) \\ &= -\sum_{x\in\mathcal{X}} p(x)\log\left(\frac{p(x)q(x)}{p(x)}\right) \\ &= -\sum_{x\in\mathcal{X}} p(x)\log p(x) + \sum_{x\in\mathcal{X}} p(x)\log\left(\frac{p(x)}{q(x)}\right) \\ &= H(p) + D(p||q) \quad \blacksquare \end{aligned} \tag{93}$$

■

(b) *Demonstração.*

$$H(p,q) = \underbrace{H(p)}_{\geq 0} + \underbrace{D(p||q)}_{\geq 0} \geq 0 \quad \blacksquare \tag{94}$$

■

(c) *Demonstração.*

$$\begin{aligned} H(p,u) &= \underbrace{H(p)}_{=\log|\mathcal{X}|-D(p||u)} + D(p||u) \\ &= \log|\mathcal{X}| - D(p||u) + D(p||u) \\ &= \log|\mathcal{X}| \quad \blacksquare \end{aligned} \tag{95}$$

■

(d) *Demonstração.*

$$\begin{aligned} H(p,q) &= H(p) + \underbrace{D(p||q)}_{\geq 0} \\ &\geq H(p) \quad \blacksquare \end{aligned} \tag{96}$$

■

Ou seja, teremos

$$-\sum_{x\in\mathcal{X}} p(x)\log q(x) \geq -\sum_{x\in\mathcal{X}} p(x)\log p(x) \tag{97}$$

Podemos também demonstrar utilizando a desigualdade de Jensen e sem utilizar a não-negatividade da divergência KL. Vamos supor $X \sim p$ e $Y \sim q$.

$$\begin{aligned} -\sum_{x\in\mathcal{X}} p(x)\log p(x) + \sum_{x\in\mathcal{X}} p(x)\log q(x) &= \sum_{x\in\mathcal{X}} p(x)\log\frac{q(x)}{p(x)} \\ &= E_p\left[\log\frac{Y}{X}\right] \\ &\quad \text{como log é concavo podemos utilizar Jensen} \\ &\leq \log E_p\left[\frac{Y}{X}\right] \\ &= \log\sum_{x\in\mathcal{X}} p(x)\frac{q(x)}{p(x)} \\ &= \log\sum_{x\in\mathcal{X}} q(x) = \log 1 = 0 \end{aligned} \tag{98}$$

Temos assim

$$\begin{aligned} -\sum_{x\in\mathcal{X}} p(x)\log p(x) &\leq -\sum_{x\in\mathcal{X}} p(x)\log q(x) \\ H(p) &\leq H(p,q) \quad \blacksquare \end{aligned} \tag{99}$$

(e) Como $q = u$, podemos utilizar o resultado do item anterior

$$H(p,u) = \log|\mathcal{X}| = \log 4 = 2. \tag{100}$$

Podemos também calcular da seguinte forma:

$$\begin{aligned} H(p,q) &= -\sum_{x\in\mathcal{X}} p(x)\log q(x) \\ &= -\frac{1}{2}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{4} \\ &= 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} = 2 \end{aligned} \tag{101}$$

$$
\begin{aligned}
H(q,p) &= -\sum_{x\in\mathcal{X}} q(x)\log p(x) \\
&= \frac{1}{4}\log 2 + \frac{1}{4}\log 4 + \frac{1}{4}\log 8 + \frac{1}{4}\log 8 \\
&= \frac{1}{4} + \frac{1}{2} + \frac{3}{4} + \frac{3}{4} = \frac{9}{4}
\end{aligned} \tag{102}
$$

$$
\begin{aligned}
H(p) &= -\sum_{x\in\mathcal{X}} p(x)\log p(x) \\
&= \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{8}\log 8 + \frac{1}{8}\log 8 \\
&= \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{3}{8} = \frac{7}{4}
\end{aligned} \tag{103}
$$

$$
H(q) = 4 \times \frac{1}{4}\log 4 = 2 \tag{104}
$$

$$
\begin{aligned}
D(p||q) &= H(p,q) - H(p) \\
&= 2 - \frac{7}{4} = \frac{1}{4}
\end{aligned} \tag{105}
$$

$$
\begin{aligned}
D(q||p) &= H(q,p) - H(q) && (106) \\
&= \frac{9}{4} - 2 = \frac{1}{4} && (107)
\end{aligned}
$$

1.22 Probabilidade de Erro

1. Considere a seguinte distribuição conjunta em (X, Y):

$X\backslash Y$	a	b	c
1	1/6	1/12	1/12
2	1/12	1/6	1/12
3	1/12	1/12	1/6

Seja $\hat{X}(Y)$ um estimador de X baseado em Y e $P_e = \Pr\{\hat{X}(Y) \neq X\}$.

(a) Encontre o estimador $\hat{X}(Y)$ com a menor probabilidade de erro. Calcule a P_e associada.

(b) Calcule o limite dado pela desigualdade de Fano e compare com o resultado anterior.

Resolução:

(a) Por inspeção, podemos verificar que o estimador com menor probabilidade de erro será

$$
\hat{X}(Y) = \begin{cases} 1, & y = a \\ 2, & y = b \\ 3, & y = c \end{cases} \tag{108}
$$

A probabilidade de erro é

$$\begin{aligned}
P_e &= \Pr\{X=1,Y=b\}+\Pr\{X=1,Y=c\}+\Pr\{X=2,Y=1\}+\Pr\{X=2,Y=c\}+ \\
P_e &= \Pr\{X=1,Y=b\}+\Pr\{X=1,Y=c\}+\Pr\{X=2,Y=1\}+\Pr\{X=2,Y=c\}+ \\
&\quad \Pr\{X=3,Y=1\}+\Pr\{X=3,Y=b\} && (109)\\
&= 6\times\frac{1}{12}=\frac{1}{2} && (110)
\end{aligned}$$

(b) Pela forma forte da desigualdade de Fano, sabemos que

$$P_e \geq \frac{H(X|Y)-1}{\log(|\mathcal{X}|-1)} \tag{111}$$

Precisamo calcular $H(X|Y)$.

$$\begin{aligned}
H(X|Y) &= H(X|Y=a)\Pr\{Y=a\}+H(X|Y=b)\Pr\{Y=b\}+H(X|Y=c)\Pr\{Y=c\} \\
&= H\left(\frac{1}{2},\frac{1}{4},\frac{1}{4}\right)\Pr\{Y=a\}+H\left(\frac{1}{4},\frac{1}{2},\frac{1}{4}\right)\Pr\{Y=b\}+H\left(\frac{1}{4},\frac{1}{4},\frac{1}{2}\right)\Pr\{Y=c\} \\
&= H\left(\frac{1}{2},\frac{1}{4},\frac{1}{4}\right)(\Pr\{Y=a\}+\Pr\{Y=b\}+\Pr\{Y=c\}) \\
&= H\left(\frac{1}{2},\frac{1}{4},\frac{1}{4}\right) \\
&= \frac{1}{2}\log 2+\frac{1}{4}\log 4+\frac{1}{4}\log 4=\frac{1}{2}+\frac{1}{2}+\frac{1}{2}=\frac{3}{2} && (112)
\end{aligned}$$

Teremos assim

$$P_e \geq \frac{3/2-1}{\log(3-1)}=\frac{1}{2} \tag{113}$$

Podemos observar que o limite dado pela desigualdade de Fano, neste exemplo, coincide com a probabilidade de erro calculada no item anterior.

1.23 Entropia de uma fonte discreta

1. Sobre a entropia de uma fonte discreta podemos afirmar que:

   A. A entropia máxima ocorre quando a distribuição da fonte for uniforme.

   B. A entropia mínima ocorre quando a distribuição da fonte for uniforme.

   C. A entropia mínima ocorre quando a distribuição da fonte for normal.

   D. A entropia máxima ocorre quando a distribuição da fonte for normal.

   E. A entropia máxima ocorre quando a distribuição da fonte for d-ádica.

2. Sobre a entropia de uma fonte discreta podemos afirmar que:

   A. a entropia nunca será negativa.

   B. a entropia nunca será nula.

   C. a entropia nunca será positiva.

   D. a entropia nunca será maior do que a informação mútua.

   E. a entropia nunca será maior do que a cardinalidade do alfabeto da fonte.

3. A entropia é uma função:

   A. estritamente côncava.

   B. estritamente convexa.

   C. nem côncava, nem convexa.

   D. linear nos parâmetros da distribuição da fonte.

   E. linear por partes.

4. A divergência (KL) entre uma dada função massa de probabilidade de uma v.a. e distribuição uniforme é igual a:

   A. logaritmo da cardinalidade do alfabeto menos a entropia da v.a..

   B. cardinalidade do alfabeto mais a entropia da v.a..

   C. informação mútua entre a v.a. e a distribuição uniforme.

   D. cardinalidade do alfabeto menos a entropia da distribuição uniforme.

   E. logaritmo da cardinalidade do alfabeto mais a entropia da distribuição uniforme.

1. Seja a entropia de uma v.a. binária $H(p) = -p \log p - (1-p)\log(1-p)$ Faça o que se pede:

   (a) Qual é o valor de $H(1/4)$? (considere $\log 3 \approx 1,585$).

   A. $0,811$

   B. $1/4$

   C. 1

   D. $1,584$

   E. $3,168$

   F. 0

   G. $1/2$

   (b) Qual é o valor máximo de $H(p)$ (em bits)?

   A. 1

   B. $1/2$

   C. $1/4$

   D. 2

   E. ∞

   F. $\log e$

   G. $2 \log e$

   (c) Qual é o valor mínimo de $H(p)$ (em bits)?

   A. 0

   B. -1

   C. -2

   D. 1

   E. $-\infty$

   F. $1/2$

   G. $\log e$

   (d) Qual é o valor esperado da entropia binária $\mathrm{E}[H(p)]$ se $p \sim \mathcal{U}(0,1)$ (em bits)?

   A. $1/2 \log e$

   B. 1

   C. $1/e - 1$

   D. $\log e - 1$

E. $1/2$
F. $1/4$
G. 0

Resolução:

(a)

$$\begin{aligned} H(1/4) &= -1/4 \log 1/4 - 3/4 \log 3/4 \\ &= 1/4 \times 2 - 3/4 \log 3 + 3/4 \times 2 \\ &= 1/2 - 3/4 \log 3 + 3/2 = 2 - 3/4 \log 3 \\ &= 2 - 0,75 \times 1,585 = 2 - 1,1887 = 0,8112 \end{aligned} \tag{114}$$

(b)

$$H(p) \leq \log |\mathcal{X}| = \log 2 = 1 \tag{115}$$

Logo o máximo é 1, que ocorre quando $p = 1/2$.

(c)

$$H(p) \geq 0 \tag{116}$$

Logo o mínimo é 0, que ocorre quando $p = 1$ ou quando $p = 0$.

(d)

$$\begin{aligned} \mathrm{E}_p H(p) &= \int_0^1 1 \times H(p) \, \mathrm{d}p \\ &= -\log e \int_0^1 (p \ln p + (1-p) \ln(1-p)) \, \mathrm{d}p \\ &= -2 \log e \int_0^1 p \ln p \, \mathrm{d}p \\ &= -2 \log e \left(\frac{p^2}{2} \ln p - \frac{p^2}{4} \right) \Bigg|_{p=0}^{p=1} \\ &= 2 \log e \frac{1}{4} = \frac{1}{2} \log e \end{aligned} \tag{117}$$

1.24 Canal ternário

1. Considere um canal de comunicação ternário, em que $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$. Sabemos que a probabilidade conjunta $p(x, y)$ quando $x = y$ é duas vezes maior que a probabilidade conjunta quando $x \neq y$. Considere um estimador $\hat{X}(Y) = Y$. Utilizando a desigualdade de Fano, encontre um limite para a probabilidade de erro para o estimador dado.

   A. $P_e \geq 1/2$
   B. $P_e \geq 0$
   C. $P_e \leq 1$
   D. $P_e \geq 1$
   E. $P_e \geq 1/4$
   F. $P_e \leq 3/4$
   G. $P_e \geq 3/4$

Resolução: A probabilidade conjunta (X, Y) é

X \ Y	1	2	3
1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$

Desigualdade de Fano:

$$P_e \geq \frac{H(X|Y) - 1}{\log\left(|\mathcal{X}| - 1\right)} \tag{118}$$

Precisamos calcular $H(X|Y)$. Para tanto precisamos calcular a marginal $p(y)$ e a condicional $p(x|y)$. Temos que $p(y) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ e $p(x|y) = \frac{p(x,y)}{p(y)}$:

X \ Y	1	2	3
1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
2	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
3	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

$$\begin{aligned} H(X|Y) &= \sum_y p(y) H(X|Y = y) \\ &= -\sum_y p(y) \sum_x p(x|y) \log p(x|y) \\ &= -3 \times \frac{1}{3}\left(\frac{1}{2}\log\frac{1}{2} + 2 \times \frac{1}{4}\log\frac{1}{4}\right) \\ &= (-1) \times \left(-\frac{1}{2} - 1\right) = \frac{3}{2} \end{aligned} \tag{119}$$

Teremos assim

$$P_e \geq \frac{H(X|Y) - 1}{\log\left(|\mathcal{X}| - 1\right)} = \frac{3/2 - 1}{\log(3 - 1)} = \frac{1}{2} \tag{120}$$

1.25 Divergência de Jensen-Shannon

1. A divergência de Jensen-Shannon é um forma de medir a similaridade entre duas distribuições, sendo utilizada, por exemplo, em técnicas de aprendizado de máquina na tarefa de discriminação de classes.

   A divergência de Jensen-Shannon entre duas distribuições p e q é definida por

   $$\mathrm{JSD}(p||q) = \frac{1}{2} D_{\mathrm{KL}}(p||m) + \frac{1}{2} D_{\mathrm{KL}}(q||m),$$

   onde m é a mistura das distribuições p e q, $m = 1/2(p + q)$, e D_{KL} é a divergência de Kullback-Leibler.

   (a) Sobre a divergência de Jensen-Shannon, podemos mostrar a seguinte relação entre $\mathrm{JSD}(p||q)$, $H(m)$, $H(p)$ e $H(q)$:

   A. $\mathrm{JSD}(p||q) = H(m) - \frac{1}{2}H(p) - \frac{1}{2}H(q)$

   B. $\mathrm{JSD}(p||q) = H(m) + \frac{1}{2}H(p) + \frac{1}{2}H(q)$

   C. $\mathrm{JSD}(p||q) = H(m) - H(p) - H(q)$

D. $\text{JSD}(p||q) = H(m) + H(p) + H(q)$
E. $\text{JSD}(p||q) = H(m) - 2H(p) - 2H(q)$
F. $\text{JSD}(p||q) = H(m) + 2H(p) + 2H(q)$
G. $\text{JSD}(p||q) = 2H(p) + 2H(q) - H(m)$

(b) Sobre a divergência de Jensen-Shannon, podemos mostrar seu limite superior e inferior

A. $0 \leq \text{JSD}(p||q) \leq 1$
B. $0 \leq \text{JSD}(p||q) \leq 2$
C. $0 \leq \text{JSD}(p||q) \leq 1/2$
D. $1/2 \leq \text{JSD}(p||q) \leq 2$
E. $1 \leq \text{JSD}(p||q) \leq 2$
F. $-1 \leq \text{JSD}(p||q) \leq 1$
G. $1/2 \leq \text{JSD}(p||q) \leq 3/2$

(dica) Para mostrar o limite superior de $\text{JSD}(p||q)$, encontre o limite superior de cada uma das parcelas da soma na Equação de definição da divergência de Jensen-Shannon.

Resolução:

(a) A relação entre $\text{JSD}(p||q)$ e $H(m)$, $H(p)$ e $H(q)$ é dada por

$$\begin{aligned}
\text{JSD}(p||q) &= \frac{1}{2}D_{\text{KL}}(p||m) + \frac{1}{2}D_{\text{KL}}(q||m) \\
&= \frac{1}{2}\sum p \log p/m + \frac{1}{2}\sum q \log q/m \\
&= \frac{1}{2}\underbrace{\sum p \log p}_{-H(p)} - \frac{1}{2}\sum p \log m + \frac{1}{2}\underbrace{\sum q \log q}_{-H(q)} - \frac{1}{2}\sum q \log m \\
&= -\frac{1}{2}H(p) - \frac{1}{2}H(q) - \sum m \log m = H(m) - \frac{1}{2}H(p) - \frac{1}{2}H(q) \qquad (121)
\end{aligned}$$

(b) O limite inferior de $\text{JSD}(p||q)$ é facilmente constato utilizando-se o limite a divergência de KL. Assim teremos $\text{JSD}(p||q) \geq 0$. Para encontrar o limite superior, iremos encontrar o limite de cada uma das parcelas que compõem $\text{JSD}(p||q)$.

$$\begin{aligned}
D_{\text{KL}}(p||m) &= \sum p \log p/m = \sum p \log 2p/(p+q) \\
&= \underbrace{\sum p \log \underbrace{p/(p+q)}_{\leq 1}}_{\leq 0} + \underbrace{\sum p \log 2}_{=1} \leq 1 \qquad (122)
\end{aligned}$$

Seguindo os mesmos passos podemos demonstrar que $D_{\text{KL}}(q||m) \leq 1$, e assim teremos

$$\text{JSD}(p||q) = \frac{1}{2}D_{\text{KL}}(p||m) + \frac{1}{2}D_{\text{KL}}(q||m) \leq 1 \qquad (123)$$

1.26 Máxima verossimilhança

1. A distribuição que minimiza a divergência de Kullbach-Leibler (KL) para a distribuição empírica é aquela que maximiza a verossimilhança. Considere um experimento de Bernoulli em que são observados n_1 ocorrências de um determinado símbolo (ou evento), por exemplo, $\{X = 1\}$, em uma sequência de N realizações deste experimento. Usando divergência de KL mostre como encontrar o parâmetro θ da

distribuição de Bernoulli, em que $\theta = Pr(X = 1)$, que maximiza a verossimilhança para a sequencia de N observações. Determine o valor de θ, da distribuição de Bernoulli, em função de n_1 e N que minimize a divergência com a distribuição empírica ($\hat{p} = \left(\frac{n_1}{N}, \frac{N-n_1}{N}\right)$).

Resolução:

$$D_{\mathrm{KL}}(\hat{p} \parallel p_\theta) = \sum_{x \in \mathcal{X}} \hat{p}(x) \log \frac{\hat{p}(x)}{p_\theta(x)} \tag{124}$$

onde a distribuição empírica é dada por $\hat{p} = \left(\frac{n_1}{N}, \frac{N-n_1}{N}\right)$ e a distribuição de Bernoulli é dada por $p_\theta = (\theta, 1-\theta)$. Teremos então:

$$D_{\mathrm{KL}}(\hat{p} \parallel p_\theta) = \frac{n_1}{N} \log \frac{n_1/N}{\theta} + \frac{N-n_1}{N} \log \frac{(N-n_1)N}{1-\theta} \tag{125}$$

Queremos determinar θ que minimize a D_{KL}:

$$\underset{\theta \in \Theta}{\operatorname{argmin}}\, D_{\mathrm{KL}}(\hat{p} \parallel p_\theta) \tag{126}$$

Como a divergência é convexa, teremos apenas um ponto de máximo. Basta então encontrar o ponto em que a derivada se anula.

$$\begin{aligned}
\frac{\mathrm{d}D_{\mathrm{KL}}}{\mathrm{d}\theta} &= \frac{n_1}{N} \frac{\theta}{n_1/N} \frac{-n_1/N}{\theta^2} + \frac{N-n_1}{N} \frac{1-\theta}{(N-n_1)/N} \frac{-(N-n_1)/N}{(1-\theta)^2} = 0 \\
\frac{n_1}{N} \frac{1}{\theta} &= \frac{N-n_1}{N} \frac{1}{1-\theta} \\
\frac{\theta}{1-\theta} &= \frac{n_1}{N-n_1} \\
\theta N &= n_1 \\
\theta &= \frac{n_1}{N}
\end{aligned} \tag{127}$$

1.27 Entropia em um jogo de dados

1. Suponha um jogo de dados em que são utilizados dois dados. Um dado com 6 (seis) lados e outro dado com 4 (quatro) lados. Seja X e Y as v.a.s associadas ao lançamento de cada um dos dados, respectivamente, e considere $\mathcal{X} = \{1, \ldots, 6\}$ e $\mathcal{Y} = \{1, \ldots, 4\}$. Sabe-se que os dados utilizados neste jogo são honestos. Participam do jogo 3 (três) jogadores $(0, 1, 2)$. Todos jogadores começam com 36 pontos. Em cada rodada lança-se ambos os dados ao mesmo tempos. A soma Z dos dados será o quantos pontos um determinado jogador irá perder. O jogador que perderá ponto é determinado por $W = Z \mod 3$. Responda/faça o que se pede a seguir.
   (a) Determine $H(Z)$.
   (b) Determine $H(W)$.
   (c) Determine $H(Z|W)$.
   (d) Algum dos jogadores é privilegiado neste jogo? Qual deles? Explique.

Resolução:

(a) A v.a. Z possui a seguinte distribuição:

$$Z \sim \left(\frac{1}{24}, \frac{2}{24}, \frac{3}{24}, \frac{4}{24}, \frac{4}{24}, \frac{4}{24}, \frac{3}{24}, \frac{2}{24}, \frac{1}{24}\right). \tag{128}$$

$$\begin{matrix} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ (\frac{1}{24} & \frac{2}{24} & \frac{3}{24} & \frac{4}{24} & \frac{4}{24} & \frac{4}{24} & \frac{3}{24} & \frac{2}{24} & \frac{1}{24}) \end{matrix} \tag{129}$$

Entropia de Z é dada por

$$\begin{aligned} H(Z) &= -\sum_z p(z)\log p(z) && (130)\\ &= 2\times\frac{1}{24}\log 24 + 2\times\frac{1}{12}\log 12 + 2\times\frac{1}{8}\log 8 + 3\times\frac{1}{6}\log 6 && (131)\\ &= \frac{1}{12}(3+\log 3) + \frac{1}{6}(2+\log 3) + \frac{1}{4}\times 3 + \frac{1}{2}(1+\log 3) && (132)\\ &= \frac{11}{6} + \frac{3}{4}\log 3. && (133) \end{aligned}$$

(b)

$$W \sim \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \tag{134}$$

$$H(W) = \log 3. \tag{135}$$

(c) Iremos utilizar que $H(W|Z) = 0$, pois W é uma função de Z.

$$\begin{aligned} H(Z|W) &= H(Z,W) - H(W) && (136)\\ &= H(W|Z) + H(Z) - H(W) && (137)\\ &= 0 + H(Z) - H(W) && (138)\\ &= \frac{11}{6} + \frac{3}{4}\log 3 - \log 3 && (139)\\ &= \frac{11}{6} - \frac{1}{4}\log 3. && (140) \end{aligned}$$

(d) O jogador associado a $w = 2$ é privilegiado pois, apesar de todos os jogadores possuírem a mesma probabilidade de perder ponto a cada jogada, o jogador $w = 2$ perde em média menos pontos que os demais. Para verificar, basta analisar $E[Z|W = w]$.

$$E[Z|W=0] = 3\times\frac{2}{8} + 6\times\frac{4}{8} + 9\times\frac{2}{8} = \frac{48}{8} = \frac{24}{4} \tag{141}$$

$$E[Z|W=1] = 4\times\frac{3}{8} + 7\times\frac{4}{8} + 10\times\frac{1}{8} = \frac{50}{8} = \frac{25}{4} \tag{142}$$

$$E[Z|W=2] = 2\times\frac{1}{8} + 5\times\frac{4}{8} + 8\times\frac{3}{8} = \frac{46}{8} = \frac{23}{4} \tag{143}$$

1.28 Média dos valores quadráticos

1. Mostre que a média dos valores quadráticos excede (ou é igual) o quadrado da média dos valores. Em qual situação haverá igualdade? Justifique todos os passos.

Resolução: Iremos utilizar a desigualdade de Jensen, utilizando que $f(x) = x^2$ é uma função convexa. Suponha que tenhamos N valores, x_i, $i = 1, \ldots, N$.

Demonstração.

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} x_i^2 &= \sum_x \frac{1}{N} f(x) = \sum_x p(x) f(x) \\
&\geq f\left(\sum_x p(x) x\right) = \\
&= \left(\sum_{i=1}^{N} \frac{1}{N} x_i\right)^2 = \left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2
\end{aligned} \tag{144}$$

■

Haverá igualdade quando $x_i = 1$, $\forall i$, pois $x = 1$ é o único que satisfaz $x^2 = x$. Neste caso teremos $\frac{1}{N}\sum_{i=1}^{N} 1^2 = 1$ e $\left(\frac{1}{N}\sum_{i=1}^{N} 1\right)^2 = 1$.

1.29 One-time pad (OTP)

1. O One-time pad (OTP) é um esquema de criptografia simétrica que utiliza uma chave secreta tão longa quanto a mensagem e é usada apenas uma vez. A segurança do OTP baseia-se na propriedade de que, se a chave é verdadeiramente aleatória, não reutilizada e mantida em segredo, é matematicamente impossível derivar qualquer informação útil sobre a mensagem sem a chave.

   Cada bit da mensagem é combinado com o bit correspondente na chave usando a operação XOR (soma módulo 2). Para descriptografar, a mesma chave é usada novamente, e a operação XOR é aplicada. Como XOR é reversível (X XOR Y XOR $Y = X$), a mensagem original é recuperada.

   Seja X a mensagem da fonte e Y a chave OTP, a mensagem cifrada Z é dada por

$$Z = X \oplus Y, \tag{145}$$

   ou seja, Z é resultado do XOR entre X e a chave Y.

Resolução: Um fato fundamental é que o resultado do XOR de uma sequência de bits completamente aleatória com uma outra sequência de bits, com qualquer distribuição, produz outra sequência de bits completamente aleatória. Ou seja, se a distribuição de Y é uniforme, $P(Y = 1) = P(Y = 0) = \frac{1}{2}$, e X tem uma distribuição qualquer, $P(X = 1) = \theta$ e $P(X = 0) = 1 - \theta$, então, como podemos observar da tabela do XOR (ver abaixo),

$$P(Z = 1) = P(X = 1)P(Y = 0) + P(X = 0)P(Y = 1) \tag{146}$$

$$= \theta\frac{1}{2} + (1 - \theta)\frac{1}{2} = \frac{1}{2} \tag{147}$$

e da mesma forma, $P(Z = 0) = \frac{1}{2}$, ou seja, independente de qual seja a distribuição em X, a aleatoriedade máxima em Y garante que teremos também aleatoriedade máxima em Z.

X	Y	Z
0	0	0
1	0	1
0	1	1
1	1	0

Note ainda que, fixando Z, nada podemos saber sobre X, uma vez que Y é uniforme. Para $Z = 0$, X pode assumir os valores 0 ou 1 com igual probabilidade, já que $P(Y = 0) = P(Y = 1) = \frac{1}{2}$. O mesmo vale para $Z = 1$. Desta forma

$$\begin{aligned}
H(X|Z) = \sum_z p(z)H(X|Z = z) = P(Z = 0)H(X|Z = 0) + P(Z = 1)H(X|Z = 1) & \quad (148)\\
=(P(X = 0)P(Y = 0) + P(X = 1)P(Y = 1))H(X|Z = 0)+ & \quad (149)\\
(P(X = 1)P(Y = 0) + P(X = 0)P(Y = 1))H(X|Z = 1) & \quad (150)\\
=\frac{1}{2}H(X|Z = 0) + \frac{1}{2}H(X|Z = 1) & \quad (151)\\
=\frac{1}{2}H(X) + \frac{1}{2}H(X) = H(X). & \quad (152)
\end{aligned}$$

Conhecer assim o valor de Z nada nos diz sobre quem é X (o mesmo vale para Y, ou seja, Z não diz nada sobre Y). O OTP permanece seguro quando implementado corretamente com uma chave verdadeiramente aleatória e usada apenas uma vez.

2 Propriedade da Equipartição Assintótica

2.1 Propriedade da Equipartição Assintótica e Codificação de Fonte

1. Uma fonte discreta sem memória emite uma sequência de dígitos binários estatisticamente independentes com probabilidades $p(1) = 0.005$ e $p(0) = 0.995$. Os dígitos são tomados 100 a cada vez e uma palavra de código (*codeword*) binária é fornecido para cada sequência de 100 dígitos contendo três ou menos uns.
   (a) Assumindo que todas as palavras de código possuem o mesmo comprimento, encontre o menor comprimento necessário para fornecer palavras de código para todas as sequências com três ou menos uns.
   (b) Calcule a probabilidade de se observar uma sequência produzida pela fonte para a qual não foi associada nenhuma palavra de código.
   (c) Utilize a desigualdade de Chebyshev para encontrar o limite de se observar uma sequência da fonte para a qual nenhuma palavra código foi associada. Compare este limite com a probabilidade calculada no item anterior.

Resolução:

(a) O número de sequências binárias com 3 ou menos uns é

$$\binom{100}{0} + \binom{100}{1} + \binom{100}{2} + \binom{100}{3} = 1 + 100 + 4950 + 161700 = 166751. \quad (153)$$

Para codificar por uma palavra binárias as sequências de comprimento 100 com 3 ou menos uns, serão necessários $\lceil \log 166751 \rceil = 18$ bits. Note que $H(0.005) = 0.0454$ e 18 bits é consideravelmente maior que 4.5 bits de entropia para sequências de tamanho 100.

(b) A probabilidade de se observar uma sequência produzida pela fonte para a qual não foi associada nenhuma palavra de código é equivalente à soma da probabilidade das sequências com mais de

3 uns, ou então, um menos a probabilidade das sequências com 3 ou menos uns.

$$\begin{aligned}\sum_{i=4}^{100}\binom{100}{i}(0.005)^i(0.995)^{100-i} &= 1-\sum_{i=0}^{3}\binom{100}{i}(0.005)^i(0.995)^{100-i}\\ &= 1-(0.60577+0.30441+0.7572+0.01243)\\ &= 1-0.99833=0.00167. \end{aligned} \tag{154}$$

(c) No caso em que uma v.a. S_n é a soma de n v.a.s i.i.d. $X_1,\ldots,X_n$, a desigualdade de Chebyshev afirma que

$$\Pr(|S_n-n\mu|\geq\epsilon)\leq\frac{n\sigma^2}{\epsilon^2}, \tag{155}$$

onde μ e σ^2 são a média e variância de X_i, respectivamente. (Desta forma, $n\mu$ e $n\sigma^2$ são a média e variância de S_n). No problema em questão temos: $n=100$, $\mu=0.005$ e $\sigma^2=(0.005)(0.995)$. As sequência que não serão codificadas são aquelas em que $S_n\geq 4$. Note que teremos $S_n\geq 4$ se e somente se $|S_{100}-100(0.005)|\geq 3.5$. Devemos pois escolher $\epsilon=3.5$, assim

$$\Pr(S_{100}\geq 4)\leq\frac{100(0.005)(0.995)}{(3.5)^2}\approx 0.04061. \tag{156}$$

Verificamos assim que o limite é bem maior que o valor da probabilidade de 0.00167.

2.2 Tipos e Classes de Tipos

1. Sejam $\mathcal{X}=\{0,1\}$ e $\mathcal{Y}=\{a,b\}$ alfabetos de duas fontes. Sejam as distribuições da fonte para $\mathcal{X}$ e $\mathcal{Y}$ tais que $P(X=0)=1/2$ e $P(Y=a)=2/3$. Seja também $\mathcal{Z}=\mathcal{X}\times\mathcal{Y}$, i.e. $z\in\mathcal{Z}$ se $z=(x,y)$ onde $x\in\mathcal{X}$ e $y\in\mathcal{Y}$.

   (a) Considere qualquer sequência de tamanho 3 tirada de forma i.i.d. de $\mathcal{Z}$. Quantos tipos possíveis existem para sequências com este comprimento?

   (b) Considere a sequência $z=\{(0,a),(0,b),(0,a),(1,b),(1,a),(1,b)\}$. Calcule o tipo desta sequência. Qual é o tamanho da classe de tipo para o tipo encontrado para a sequência z? Considerando que as duas fontes são independentes, z será uma sequência típica? Por quê? Forneça argumentos com base na classe de tipo de z. Quão provável é a classe de tipo de z?

Resolução:

(a) Teremos

$$\mathcal{Z}=\{(0,a),(0,b),(1,a),(1,b)\}, \tag{157}$$

e desta forma $|\mathcal{Z}|=4$.

O número de tipos para sequências de comprimento n é dado por

$$|\mathcal{P}_n|=\binom{n+|\mathcal{Z}|-1}{|\mathcal{Z}|-1}. \tag{158}$$

Para $n=3$ e $|\mathcal{Z}|=4$ teremos

$$|\mathcal{P}_n|=\binom{3+4-1}{4-1}=\binom{6}{3}=\frac{6!}{3!\,3!}=\frac{6\times5\times4}{3\times2\times1}=20. \tag{159}$$

(b) O tipo da sequência $z = \{(0,a),(0,b),(0,a),(1,b),(1,a),(1,b)\}$ será

$$P_z = \left(\frac{2}{6}, \frac{1}{6}, \frac{1}{6}, \frac{2}{6}\right). \tag{160}$$

Para este tipo P_z, a classe de tipo terá o seguinte tamanho:

$$|T(P_z)| = \binom{6}{2\ 1\ 1\ 2} = \frac{6!}{2!\ 1!\ 1!\ 2!} = 6 \times 5 \times 3 \times 2 = 180. \tag{161}$$

A função massa de probabilidade Q para $Z \in \mathcal{Z}$ é $(\frac{1}{3}, \frac{1}{6}, \frac{1}{3}, \frac{1}{6})$. A divergência entre P_z e Q é dada por

$$\begin{aligned} D(P_z||Q) &= \sum p_z \log \frac{p_z}{q} \\ &= 2/6 \log \frac{2/6}{1/3} + 1/6 \log \frac{1/6}{1/6} + 1/6 \log \frac{1/6}{1/3} + 2/6 \log \frac{2/6}{1/6} \\ &= 2/6 \log 1 + 1/6 \log 1 + 1/6 \log 1/2 + 2/6 \log 2 = 0 + 0 - \frac{1}{6} + \frac{2}{6} = \frac{1}{6}. \end{aligned} \tag{162}$$

Utilizando a seguinte definição para conjunto típico,

$$T_Q^\epsilon = \{x_{1:n} : D(P_{x_{1:n}} \,||\, Q) \leq \epsilon\}, \tag{163}$$

onde ϵ usualmente é pequeno, poderemos concluir que as sequências do tipo P_z não são típicas. Em verdade, quando temos n pequeno, não podemos falar efetivamente em sequências típicas, umas vez que a tipicidade é um fenômeno assintótico para n grade suficiente, $n \to \infty$.

A probabilidade da sequência z, assim como de todas as sequências do mesmo tipo, será dada por

$$Q_z = \frac{1}{3} \times \frac{1}{3} \times \frac{1}{6} \times \frac{1}{3} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{5832}. \tag{164}$$

A probabilidade da classe de tipo para o tipo P_z será então

$$Q^n(T(P_z)) = Q_z|T(P_z)| = \frac{180}{5832} = \frac{5}{162} = 0.03086419753. \tag{165}$$

Note que a classe de tipo com maior probabilidade terá uma probabilidade 2 vezes maior que a probabilidade encontrada acima.

2.3 Estimação da pmf

1. Seja p a função massa de probabilidade (pmf) sob o conjunto $\mathcal{X} = \{a_1, \ldots, a_k\}$ de cardinalidade finita. Suponha que observamos uma sequência de n amostras $x_1, \ldots, x_n$, onde $x_i \in \mathcal{X}$, $1 \leq i \leq n$. As amostras são independentes e identicamente distribuídas (i.i.d.),

   Quando $n \gg k$, uma forma satisfatória para realizar esta estimação é o histograma empírico ou tipo. O tipo P é a proporção relativa de ocorrências de cada símbolo de $\mathcal{X}$, sendo definido por

$$P(X = a_i) = P_i = \frac{1}{n} \sum_{t=0}^{n} \mathbf{1}(x_t = a_i) \equiv \frac{n(a_i|x_{1:n})}{n} = \frac{n_i}{n}\ , i \in \{1, \ldots, k\}, \tag{166}$$

   onde $\mathbf{1}(\cdot)$ é a função indicadora e, desta forma, $n(a_i|x_{1:n})$ é o número de ocorrências do símbolo a_i na amostra $x_{1:n}$.

Por outro lado, quando n é pequeno, teremos problema em estimar as probabilidades. Laplace foi pioneiro neste assunto, desenvolvendo os estimadores Bayesianos (para $k = 2$) como forma alternativa ao tipo. Durante a II Guerra Mundial, enquanto trabalhava em sistemas para quebrar a criptografia alemã, Jack Good e Alan Turing propuseram um método para regularizar o tipo. No caso estudado, $k = 26$, o número de caracteres do alfabeto, e $n \approx 100$ a 1000.

O tipo, ou distribuição empírica, de uma amostra $x_{1:n}$ é dado por

$$P_{x_{1:n}} = \left(\frac{n_1}{n}, \ldots, \frac{n_k}{n}\right), \quad \text{onde} \quad n = \sum_{i=1}^{n} n_i, \tag{167}$$

e n_i, para $1 \leq i \leq k$, é o número de ocorrência de cada símbolo a_i na amostra. Vamos chamar de $\mathcal{P}^k$ o conjunto de pmfs sobre o conjunto de cardinalidade k e $\mathcal{P}_n^k$ o conjunto de tipos com denominador n sob um conjunto de cardinalidade k. A probabilidade, sob $p \in \mathcal{P}^k$, de observarmos uma determinada sequência $x_{1:n}$ é dada por

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{k} p_i^{n_i}, \tag{168}$$

onde p_i é a probabilidade do i-ésimo símbolo, dada pela pmf p e n_i é o número de observações do i-ésimo símbolo. Podemos ainda reescrever a equação da seguinte forma:

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{k} p_i^{n_i} = 2^{-n\left(D(P_{x_{1:n}}, p) + H(P_{x_{1:n}})\right)}, \tag{169}$$

onde $D(\cdot, \cdot)$ é a divergência de Kullback-Leibler entre duas distribuições e $H(\cdot)$ é a entropia de Shannon para uma dada distribuição.

(a) Demonstre a Equação 169.

(b) Mostre que é possível verificar, através da equação 169, que o tipo P é uma esteatítica suficiente para estimar p.

(c) Quando k é grande, podem existir diversos símbolos $1 \leq i \leq k$ para os quais $p_i \ll \frac{1}{n}$. Neste caso, com alta probabilidade, observaremos $n_i = 0$. Desta forma, eventos de baixa probabilidade tendem a ser subestimados e eventos de alta probabilidade tendem a ser sobrestimados por P. Uma manifestação disso é que o valor esperado da entropia do tipo subestima a entropia original da pmf. Mostre que isto é verdade ($E[H(P)] \leq H(p)$).

(d) Uma maneira de quantificar isso é dizer que aquilo que é observado deve ser mais provável sob a pmf subjacente do que qualquer outro evento. Se um determinado tipo é observado, queremos encontrar as prováveis pmf tais que a probabilidade de se observar o tipo observado P é maior do que a probabilidade de se encontrar um outro tipo qualquer P' sob uma pmf subjacente q. Considere $f(P, q)$ a probabilidade de se observar o tipo P sob a pmf q.
Escreva a equação para $f(P, q)$.

(e) Vamos chamar de conjunto de máxima verossimilhança $\mathcal{M}$ o conjunto formado pelas pmf's $q \in \mathcal{P}^k$ tal que, para um dado tipo P, a probabilidade de observarmos este tipo sob a pmf subjacente q é maior ou igual a probabilidade de observarmos qualquer outro tipo sob a mesma pmf subjacente, ou seja, $f(P, q) \geq f(Q, q)$, $\forall Q \in \mathcal{P}_n^k$.

$$\mathcal{M}(P) = \{p \in \mathcal{P}^k \ : \ f(P, q) \geq f(Q, q), \ \forall Q \in \mathcal{P}_n^k\}. \tag{170}$$

Segundo o critério exposto acima, qualquer uma das pmf's em $\mathcal{M}(P)$ seria igualmente boa para explicar a observação de um tipo P.

Com base em tudo o que foi exposto anteriormente, qual critério poderemos adotar para escolher uma melhor pmf dentre aquelas em $\mathcal{M}(P)$?

(f) Note que para uma pmf qualquer, deveremos ter $p_i \leq 1$, $\forall i \in \{1, \ldots, k\}$ e $\sum_{i=1}^{k} p_i = 1$. O conjunto de pontos em $\mathcal{R}^k$ que satisfaz estas propriedades é conhecido como simplex probabilístico (*probability simplex*). Os tipos possíveis $Q \in \mathcal{P}_n^k$ são pontos distintos dentro deste simplex probabilístico.

Este conjunto de pontos é finito? É possível estimar o seu tamanho para um dado alfabeto de cardinalidade k e sequências de comprimento n? Como podemos estimar? (se possível)

(g) O conjunto de máxima verossimilhança $\mathcal{M}(P)$ para um dado tipo será um subconjunto do simplex probabilístico compreendendo todos os pontos do simplex que estão mais próximos (segundo a divergência de Kullback-Leibler) de um determinado tipo do que de qualquer outro tipo. Estes conjuntos formarão um diagrama de Voronoi sobre o simplex probabilístico.

Mostre que $f(P, q) \doteq 2^{-D(P,q)}$, ou seja, a propriedade de observarmos um determinado tipo P sob a pmf subjacente q é igual até a primeira ordem de expoente à $2^{-D(P,q)}$, onde $D(P, q)$ é a divergência de Kullback-Leibler entre o tipo P e a pmf subjacente q.

A notação $a_n \doteq b_n$ significa que $\lim_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$.

Desta forma, para n suficientemente grande, o conjunto de máxima verossimilhança associado ao tipo P é dado por

$$\mathcal{M}(P) = \{p \in \mathcal{P}^k \ : \ D(P, q) \leq D(Q, q), \ \forall Q \in \mathcal{P}_n^k\}. \tag{171}$$

Este conjunto possui algumas propriedades uteis e interessantes. Seja $P = \left(\frac{n_1}{n}, \ldots, \frac{n_k}{n}\right)$ um tipo, os elementos $p = (p_1, \ldots, p_k) \in \mathcal{M}(P)$ devem satisfazer as seguintes propriedades:

$$\begin{aligned}
P \ll p \ \text{ i.e. } n_i > 0 \Rightarrow p_i > 0, &\qquad \forall 1 \leq i \leq k, \\
n_i < n_j \Rightarrow p_i \leq p_j, &\qquad \forall 1 \leq i, j \leq k, \\
\frac{n}{n+k} P_i \leq p_i \leq P_i + \frac{1}{n}, &\qquad \forall 1 \leq i \leq k, \\
\|p - P\|_1 = \sum_{i=1}^{k} |p_i - P_i| \leq \frac{2(k-1)}{n}, & \\
P \in \mathcal{M}(P), &
\end{aligned} \tag{172}$$

mas nenhum outro tipo com denominador n é um elemento de $\mathcal{M}(P)$. Se $x_1, \ldots, x_n$ são amostras independentes com mesma pmf $q \in \mathcal{P}^k$, então o conjunto de máxima verossimilhança definido pelo tipo P é tal que

$$\sup_{p \in \mathcal{M}(P)} \|p - q\|_1 \to 0 \text{ quando } n \to \infty \text{ com probabilidade 1.} \tag{173}$$

Toda pmf em $\mathcal{M}(P)$ satisfaz as propriedades acima e será considerada uma desejável estimativa da pmf subjacente, geradora das amostras $x_1, \ldots, x_n$, aquela que satisfizer um critério secundário já que admitimos $\mathcal{M}(P)$ como um conjunto admissível.

Seja $P = \left(\frac{n_1}{n}, \ldots, \frac{n_k}{n}\right)$ um tipo e $\mathcal{M}(P)$ o conjunto de máxima verossimilhança associado. Seja $q = (q_1 \ \ldots, q_k)$ uma pmf tal que $P \ll q$. Então existe um único elemento $p* \in \mathcal{M}(P)$ tal que

$$D(p*, q) = \min_{p \in \mathcal{M}(P)} D(p, q). \tag{174}$$

Isto pode ser mostrado pelo fato de $\mathcal{M}(P)$ ser um conjunto convexo e fechado na topologia Euclideana em $\mathcal{P}^k$ e pela convexidade de $p \to D(p, q)$. Quando $n \gg k$, teremos pouca influência de q na determinação de $p*$, visto que o conjunto $\mathcal{M}(P)$ terá pequeno raio. Quando $n \to 0$, a escolha de q irá influenciar de forma forte $p*$.

Resolução:

(a) *Demonstração.*

$$\begin{aligned}
p(x_1,\ldots,x_n) &= \prod_{i=1}^{n} p(x_i) = \prod_{a\in\mathcal{X}} p(a)^{n(a|x_{1:n})} \\
&= \prod_{a\in\mathcal{X}} p(a)^{nP_{x_{1:n}}(a)} = \prod_{a\in\mathcal{X}} 2^{\left\{nP_{x_{1:n}}(a)\log p(a)\right\}} \\
&= \prod_{a\in\mathcal{X}} 2^{n\left\{P_{x_{1:n}}(a)\log p(a) - P_{x_{1:n}}(a)\log P_{x_{1:n}}(a) + P_{x_{1:n}}(a)\log P_{x_{1:n}}(a)\right\}} \\
&= 2^{n\sum_{a\in\mathcal{X}}\left(-P_{x_{1:n}}(a)\log\frac{P_{x_{1:n}}(a)}{p(a)} + P_{x_{1:n}}(a)\log P_{x_{1:n}}(a)\right)} \\
&= 2^{-n\left(D(P_{x_{1:n}}||p) + H(P_{x_{1:n}})\right)} \qquad (175)
\end{aligned}$$

■

(b) Uma função $T(\cdot)$ é dita ser uma estatística suficiente em relação à família $\{f_\theta(x)\}$ se X é independente de θ dado $T(X)$ para qualquer distribuição em θ (i.e. $\theta \to T(X) \to X$ forma uma cadeia de Markov). Então

$$I(\theta; X) = I(\theta; T(X))\ \forall\theta \qquad (176)$$

Uma estatística suficiente preserva a informação mútua e reciprocamente

$$X \perp\!\!\!\perp \theta | T(X) \qquad (177)$$

Note que a probabilidade de uma sequência $x_{1:n}$ é a mesma para todas as sequências de um determinado tipo, conforme verificamos através da equação 169. Desta forma, a probabilidade de uma sequência específica é igual a 1 sobre o número de sequências de um determinado tipo,

$$p(x_{1:n}|P_{x_{1:n}}) = \frac{1}{|T(P_{x_{1:n}})|} = \begin{cases} \frac{1}{\binom{n}{n_1,n_2,\ldots,n_k}} & \text{se } x_{1:n} \in T(P_{x_{1:n}}) \\ 0 & \text{caso contrário,} \end{cases} \qquad (178)$$

ou seja, $X_{1:n} \perp\!\!\!\perp p | P_{x_{1:n}}$, e assim temos que o histograma empírico é uma estatística suficiente para p.

Observe também que P é o estimador de máxima verissimilhança (MLE) de p.

Vamos definier a função de verossimilhança,

$$\begin{aligned}
l'(p_1,\ldots,p_k|n_1,\ldots,n_k) &= p(x_{1:n}|P_{x_{1:n}}) \\
&= \binom{n}{n_1 \quad \ldots \quad n_k} \prod_{j=1}^{k} p_j^{n_j} \\
&= \frac{n!}{\prod_{i=1}^{k} n_i!} \prod_{j=1}^{k} p_j^{n_j}. \qquad (179)
\end{aligned}$$

A função de verossimilhança é uma função das probabilidades, dado o histograma empírico. Queremos encontrar qual é a distribuição p que maximiza a verossimilhança, dada uma observação com um determinado histograma empírico. Ao invés de maximizar diretamente a função de verossimilhança, iremos maximizar o logaritmo desta função, já que o logaritmo é

uma função monotonica crescente.

$$\begin{aligned} l(p_1,\ldots,p_k) &= \log l'(p_1,\ldots,p_k|n_1,\ldots,n_k) \\ &= \log\left(\frac{n!}{\prod_{i=1}^k n_i!}\prod_{j=1}^k p_j^{n_j}\right) \\ &= \log n! - \sum_{i=1}^k \log n_i! + \sum_{j=1}^k \log p_j^{n_j} \\ &= \log n! + \sum_{i=1}^k \log \frac{p_i^{n_i}}{n_i!}. \end{aligned} \tag{180}$$

Desejamos maximizar $l(p_1,\ldots,p_k)$ sujeito a $\sum_{i=1}^k p_i = 1$. Basta utilizar o método dos multiplicadores de Lagrange para incorporar a restrição à função que desejamos maximizar,

$$\begin{aligned} L(p_1,\ldots,p_k,\lambda) &= l(p_1,\ldots,p_k) + \lambda\left(1-\sum_{i=1}^k p_i\right) \\ &= \log n! + \sum_{i=1}^k \log \frac{p_i^{n_i}}{n_i!} + \lambda\left(1-\sum_{i=1}^k p_i\right). \end{aligned} \tag{181}$$

A derivada de L em relação a cada um dos parâmetros p_j será dada por

$$\begin{aligned} \frac{\partial L}{\partial p_j} &= \frac{n_j!}{p_j^{n_k}}\frac{n_j p_j^{n_j-1}}{n_j!} - \lambda \\ &= \frac{n_j}{p_j} - \lambda = 0, \end{aligned} \tag{182}$$

assim teremos $n_j = \lambda p_j$. Como $\sum_{i=1}^k n_i = n$, deveremos ter $\lambda = n$, e assim a estimativa para o parâmetro p_j que maximiza a verossimilhança será

$$\hat{p}_j = \frac{n_j}{N}. \tag{183}$$

(c)

$$E[H(P)] = -E\left[\sum_{i=1}^k P_i \log\frac{P_i}{p_i}p_i\right] = -E[D(P,p)] + H(p) \leq H(p). \tag{184}$$

Uma sequência observada deve ser bem provável; caso contrário não iriamos observá-la. Uma maneira de quantificar isso é dizer que aquilo que é observado deve ser mais provável sob a pmf subjacente do que qualquer outro evento. Se um determinado tipo é observado, queremos encontrar as prováveis pmf tais que a probabilidade de se observar o tipo observado P é maior do que a probabilidade de se encontrar um outro tipo qualquer P' sob uma pmf subjacente q. Considere $f(P,q)$ a probabilidade de se observar o tipo P sob a pmf q.

(d)

$$f(P,q) = \binom{n}{n_1\ldots n_k}\prod_{i=1}^k q_i^{n_i} \tag{185}$$

onde $\binom{n}{n_1\ldots n_k} = \frac{n!}{n_1!\ldots n_k!}$ é o coeficiente multinomial.

(e) Devemos usar o critério da máxima entropia, pois a entropia do tipo subestima a entropia da fonte, desta forma vamos selecionar aquela mais próxima de $H(p)$.

(f) O simplex probabilístico é um conjunto infinito de pontos. Já os tipos possíveis, fixando n e k, formam um conjunto finito cujo tamanho é limitado por $(n+1)^k$.

(g) Para qualquer $P \in \mathcal{P}_n$ e qualquer distribuição Q, a probabilidade da classe de tipo $T(P)$ sob Q^n é tal que $Q^n(T(P)) \doteq 2^{-nD(P||Q)}$. Especificamente, temos os limites

$$\frac{1}{(n+1)^{|\mathcal{X}|}} 2^{-nD(P||Q)} \leq Q^n(T(P)) \leq 2^{-nD(P||Q)} \tag{186}$$

$$\begin{aligned} Q^n(T(P)) &= \sum_{x_{1:n} \in T(P)} Q^n(x_{1:n}) = \sum_{x_{1:n} \in T(P)} 2^{-n(D(P||Q)+H(P))} \\ &= |T(P)| 2^{-n(D(P||Q)+H(P))} \end{aligned} \tag{187}$$

para completar a demonstração, devemos utilizar

$$\frac{1}{(n+1)^{|\mathcal{X}|}} 2^{nH(P)} \leq |T(P)| \leq 2^{nH(P)}. \tag{188}$$

O limite superior para o tamanho da classe de tipo pode ser obtido da seguinte forma:

$$\begin{aligned} 1 &\geq P^n(T(P)) = \sum_{x_{1:n} \in T(P)} P^n(x_{1:n}) = \sum_{x_{1:n} \in T(P)} 2^{-nH(P)} \\ &= |T(P)| 2^{-nH(P)}. \end{aligned} \tag{189}$$

O limite inferior pode ser obtido fazendo

$$\begin{aligned} 1 &= \sum_{Q \in \mathcal{P}_n} P^n(T(Q)) \leq \sum_{Q \in \mathcal{P}_n} \max_{R \in \mathcal{P}_n} P^n(T(R)) \\ &\quad \text{fazendo } P = \operatorname*{argmax}_{R \in \mathcal{P}_n} P^n(T(R)) \text{ teremos} \\ &= \sum_{Q \in \mathcal{P}_n} P^n(T(P)) \leq (n+1)^{|\mathcal{X}|} P^n(T(P)) \\ &= (n+1)^{|\mathcal{X}|} \sum_{x_{1:n} \in T(P)} P^n(x_{1:n}) \\ &= (n+1)^{|\mathcal{X}|} \sum_{x_{1:n} \in T(P)} 2^{-nH(P)} \\ &= (n+1)^{|\mathcal{X}|} |T(P)| 2^{-nH(P)} \end{aligned} \tag{190}$$

2.4 Jogo de moeda

1. Em um determinado jogo de moeda, uma moeda honesta é lançada 100 vezes.
   (a) Estime o limite superior da probabilidade de observarmos uma sequências com *i*) 60 caras, *ii*) 80 caras.
   (b) Mostre como podemos calcular explicitamente a probabilidade de obtermos uma sequência com *i*) 60 caras, *ii*) 80 caras (não é necessário realizar os cálculos).

(c) Mostre também como podemos calcular a probabilidade de obtermos uma sequência em que o número de caras esteja entre 40 e 60 (não é necessário realizar os cálculos).

(d) Escreva um algoritmo (Octave/Matlab ou pseudo-código) para calcular esta probabilidade.

(e) Qual é o problema em se realizar o cálculo explícito (neste caso, justificando a utilização do limite)?

Resolução:

(a) A probabilidade de uma sequência depende apenas do tipo. Um limite superior para a probabilidade da classe de tipo para um determinado tipo P, considerando uma distribuição subjacente Q, é dada por

$$Q^n(T(P)) \leq 2^{-nD(P||Q)}. \tag{191}$$

Considerando a moeda honesta, temos que $Q = (\frac{1}{2}, \frac{1}{2})$. Para o item a) temos o tipo $P_a = (0.6, 0.4)$ e para o item b) temos o tipo $P_b = (0.8, 0.2)$. Devemos então calcular as divergências

$$\begin{aligned} D(P_a||Q) &= \frac{3}{5}\log\left(\frac{3}{5}\times 2\right) + \frac{2}{5}\log\left(\frac{2}{5}\times 2\right) \\ &= \frac{3}{5}(1+\log 3 - \log 5) + \frac{2}{5}(2-\log 5) \\ &= \frac{7}{5} + \frac{3}{5}\log 3 - \log 5 \approx 0.029049 \end{aligned} \tag{192}$$

$$Q^n(T(P_a)) \leq 2^{-nD(P_a||Q)} \approx 2^{-2.9049} \approx 0.13352. \tag{193}$$

$$\begin{aligned} D(P_b||Q) &= \frac{4}{5}\log\left(\frac{4}{5}\times 2\right) + \frac{1}{5}\log\left(\frac{1}{5}\times 2\right) \\ &= \frac{4}{5}(3-\log 5) + \frac{1}{5}(1-\log 5) \\ &= \frac{13}{5} - \log 5 \approx 0.27807 \end{aligned} \tag{194}$$

$$Q^n(T(P_b)) \leq 2^{-nD(P_b||Q)} \approx 2^{-27.807} \approx 4.2585\times 10^{-9}. \tag{195}$$

```
function D = kldivergence(p,q)
D = sum(p.*log2(p./q));
endfunction

D = kldivergence([0.6 0.4],[0.5 0.5]);
2^(-100*D)
ans =  0.13351

D = kldivergence([0.8 0.2],[0.5 0.5]);
2^(-100*D)
ans =    4.2580e-09
```

(b) Explicitamente, a probabilidade da classe de tipo é igual à probabilidade de uma sequência deste dado tipo vezes o número de sequências deste dado tipo,

$$Q^n(T(P_a)) = \binom{100}{60} p^{60}(1-p)^{40} = \frac{100!}{60!40!}\left(\frac{1}{2}\right)^{100}. \tag{196}$$

De forma semelhante,

$$Q^n(T(P_a)) = \binom{100}{80} p^{80}(1-p)^{20} = \frac{100!}{80!20!}\left(\frac{1}{2}\right)^{100}. \tag{197}$$

(c) A probabilidade de obtermos uma sequência em que o número de caras esteja entre 40 e 60 é dada por

$$P_{40\leq k\leq 60} = \sum_{k=40}^{60} \binom{100}{k} p^k (1-p)^{100-k} = \sum_{k=40}^{60} \binom{100}{k} \left(\frac{1}{2}\right)^{100} \tag{198}$$

```
n=100; p=0; for k = 40:60,  p+=nchoosek(n,k)*(0.5)^n; end;
p =  0.96480
```

(d) Para realizar o cálculo explícito das probabilidades haverá problema de precisão numérica quando o valor de n for grande, além do alto custo computacional, não sendo possível realizar de forma prática os cálculos, justificando assim a utilização do limite para estes casos.

```
tic; n=10E4; p=0;
for k = 0.4*n:0.6*n,  p+=nchoosek(n,k)*(0.5)^n; end;
p, toc
p = NaN
Elapsed time is 14.2635 seconds.
```

2.5 Código genético do DNA

1. O código genético do DNA pode ser representado por uma sequência constituídas por letras de um alfabeto $\mathcal{X} = \{A, T, C, G\}$, que representam os 4 tipos de bases nitrogenadas (Adenina, Timina, Citosina e Guanina). Verificou-se que as sequências de comprimento 4 são relevantes nas funções regulatórias dos genes.

   Analisando estas sequências de comprimento 4 sob a ótica do método de tipos, considerando que as probabilidades das bases hidrogenadas são dadas por $q = (3/8, 3/8, 1/8, 1/8)$, respectivamente, e assumindo a independência na ocorrência dessas bases hidrogenadas, responda às questões abaixo.

   (a) Quantos tipos existem?

   (b) Quantas sequências do tipo P abaixo existem?

$$P_{x_{1:4}} = \left(\frac{3}{4}, \frac{1}{4}, 0, 0\right) \tag{199}$$

   (c) Qual é a probabilidade da classe de tipo para o tipo $P_{x_{1:4}} = (3/4, 1/4, 0, 0)$?

Resolução:

(a)

$$|\mathcal{P}_n| = \binom{n+|\mathcal{X}|-1}{|\mathcal{X}|-1} = \binom{7}{3} = \frac{7!}{3!4!} = 35 \tag{200}$$

(b) As sequências onde há 3 ocorrências de A e 1 ocorrência de T são apenas 4, a saber: TAAA, ATAA, AATA, AAAT.

Podemos também utilizar o coeficiente multinomial

$$\binom{4}{3,1,0,0} = \frac{4!}{3!1!0!0!} = 4. \tag{201}$$

(c)

$$\begin{aligned} Q^n(T(P)) &= |T(P)| q_A^3 q_T = 4(3/8)^3(3/8) && (202)\\ &= 4(3/8)^4 && (203) \end{aligned}$$

2.6 Tipicidade nas provas

1. Suponha que uma turma com 30 alunos realizou uma prova com 7 questões de múltipla escolha onde cada questão possuía 7 alternativas. Suponha que os alunos não tenham estudado e, sem saber resolver as questões, resolveram marcar aleatoriamente (sem preferência por nenhuma opção e sem nenhum critério para escolher as opções) as respostas

   (a) Calcule a probabilidade de que todos os alunos tenham acertado no mínimo 4 questões.

   (b) Foi observado que os alunos, sem apresentar a resolução de nenhuma das questões, acertaram no mínimo 4 questões. Utilizando os seu conhecimentos da Propriedade da Equipartição Assintótica, podemos concluir que esta observação pertence ou não ao conjunto típico? Por que? (considere apenas uma aproximação, uma vez que o conjunto típico surge quando observamos sequências arbitrariamente longas, ou seja, é um comportamento assintótico)

   (c) Podemos concluir que este grupo de aluno utilizou-se de algum meio ilícito para obter as respostas das questões? Qual é o grau de confiabilidade desta conclusão?

   (d) Explique o que é o conjunto típico e quais resultados seriam resultados típicos para esta prova, considerando que os alunos escolheram aleatoriamente as respostas. (Não é necessário realizar as contas, apenas deixe indicado como resolver a questão).

Resolução: Em cada questão o aluno pode acertar ou errar, então podemos ver o resultado de cada questão como uma realização de um experimento de Bernoulli com $p = 1/7$, uma vez que as opções de cada questão possuem a mesma probabilidade de serem escolhidas, ou seja, teremos uma probabilidade de $1/7$ para o acerto e $6/7$ para o erro. Um resultado dos 30 exames pode ser visto como uma sequência de comprimento $N = 30 \times 7 = 210$ (30 alunos e 7 questões por prova) de v.a. i.i.d., $X_{1:N}$, onde $X_i \sim \text{Bern}(p)$, $\forall i$, ou então como uma realização de uma v.a. Binomial, $X \sim \text{Bin}(N, k)$, onde N é o número de experimentos de Bernoulli e k o número de sucessos.

(a) Para calcular a probabilidade de que todos os alunos tenham acertado no mínimo 4 questões a esmo, devemos somar a probabilidade de todas as observações que satisfazem esta condição, ou seja, todas as sequências de N realizações do ensaio de Bernoulli onde ocorram $30 \times 4 = 120$ ou mais sucessos, ou seja, a probabilidade da v.a. binomial assumir um valor igual ou superior a 120.

$$\Pr(\#\text{ acertos} \geq 4) = \Pr(X \geq 120) = \sum_{k=120}^{N} \binom{N}{k} p^k (1-p)^{N-k} = 4.4 \times 10^{-47}. \tag{204}$$

(b) Esta observação não pertence ao conjunto típico. As sequências típicas são aquelas que apresentam aproximadamente $N \times p = 30$ sucessos e $N \times (1-p) = 180$ falhas. Como a variância de uma v.a. Binomial é $\text{Var}(X) = Np(1-p) = 180/7 \approx 25,7$, então o desvio padrão será $\sigma \approx 5$. Para uma distribuição Gaussiana, sabemos que ao tomar 3 desvios padrões acima e abaixo da média nos confere 99% de probabilidade. Para o caso Binomial não é muito diferente. Esperamos então que, ao tomar as os casos com 15 a 45 sucessos ($30-3\times5$ e $30+3\times5$) teremos aproximadamente toda a probabilidade. Desta forma, podemos considerar como observações típicas aquelas em que observamos de 15 a 45 sucessos. Uma observação com 120 sucessos ou mais está bem distante do que é típico.

De forma mais rigorsa, devemos definir ϵ, por exemplo $\epsilon = 0.1$. Dado este valor de ϵ as sequências típicas serão aquelas que satisfazem

$$H(X) - \epsilon \leq -\frac{1}{n} \log p(x_{1:n}) \leq H + \epsilon. \tag{205}$$

Utilizando o seguinte código podemos determinar quais são estas sequências:

```
>> n=210; p=1/7; eps=0.1;
>> H=-p*log2(p)-(1-p)*log2(1-p);
>> for k=0:n,
       pn=p^k*(1-p)^(n-k);
       lpn=-(1/n)*log2(pn);
       if lpn>H-eps && lpn<H+eps, disp(k);endif;
   endfor;
```

e assim concluímos que as as sequências com $k \in [22, 38]$ pertencem ao conjunto típico.

(c) Sim. Não é factível observar tantos sucessos. Podemos afirmar com praticamente 100% de certeza que algo ilícito ocorreu (considerando os valores de .

Se considerarmos o conjunto típico $A_\epsilon^{(n)}$ como o conjunto das sequências com 5 a 45 sucessos, teremos

$$\Pr(A_\epsilon^{(n)}) = \sum_{k=15}^{45} \binom{N}{k} p^k (1-p)^{N-k} = 0.99768, \tag{206}$$

ou seja, a probabilidade do conjunto não-típcio é 0.002. Se usarmos 4 desvios padrões, ao invés de 3, como feito anteriormente, passaremos a ter $\Pr(A_\epsilon^{(n)}) = 0.99991$, e assim a probabilidade do conjunto não-típico será 0.00009.

Realizando os cálculos com $\epsilon = 0.1$, teremos

$$\Pr(A_\epsilon^{(n)}) = \sum_{k=22}^{38} \binom{N}{k} p^k (1-p)^{N-k} = 0.90735, \tag{207}$$

ou seja, podemos confirmar que $\Pr(A_\epsilon^{(n)}) > 1 - \epsilon$.

```
>> n=210; p=1/7; eps=0.1; PA=0;
>> H=-p*log2(p)-(1-p)*log2(1-p);
>> for k=0:n,
        pn=p^k*(1-p)^(n-k);
        lpn=-(1/n)*log2(pn);
        if lpn>H-eps && lpn<H+eps,
           PA+=nchoosek(n,k)*pn;
        endif;
   endfor;
```

(d) Conjunto típico é o menor conjunto que detém 'toda' a probabilidade. As sequências típicas são aquelas que possuem probabilidade de 2^{-nH}, onde n é o comprimento da sequência e H a entropia da fonte. No caso em questão, seriam resultados típicos observações com aproximadamente 30 sucessos. Como mostrado acima, considerando as sequências com 15 a 45 sucessos, este conjunto deterá mais de 99% da probabilidade. Se utilizarmos o $\epsilon = 0.1$, consideraremos as sequências com 22 a 38 sucessos com típicas, resultando em um conjunto com probabilidade maior que 90%.

A título de informação, para ganhar na megasena é necessário acertar um jogo de 6 números escolhido dentre os números de 1 a 60. O número de possíveis jogos na megasena é

$$\binom{60}{6} = 50063860 \tag{208}$$

Como todos os jogos são equiprováveis, a chance de ganhar na megasena com um jogo é de 1 em 50063860, ou seja, 1.99×10^{-8}. A probabilidade de observarmos 4 ou mais acertos nas questões dos 30 alunos é menor do que a probabilidade de ganhar 6 vezes na megasena.

```
>> Pmegasena = 1/50063860;
>> P=0;
>> for k=120:n,
       pn=p^k*(1-p)^(n-k);
       P+=nchoosek(n,k)*pn;
   endfor;
>> log(P)/log(Pmegasena)
ans =  6.0202
```

2.7 Tipo

1. Suponha uma fonte com alfabeto de tamanho $|\mathcal{X}| = 4$ e com distribuição $p = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right)$. Considere sequências de comprimento $n = 6$.
   (a) Quantos tipos existem?
   (b) Quantas sequências de comprimento 6 e do tipo P abaixo existem?

$$P_{x_{1:6}} = \left(\frac{3}{6}, \frac{2}{6}, \frac{1}{6}, 0\right) \tag{209}$$

   (c) Qual é a probabilidade de uma sequência deste tipo P?
   (d) Qual é a probabilidade da classe de tipo P?

Resolução:

(a) O número de tipos é dado por

$$\begin{aligned} |\mathcal{P}_n| &= \binom{n + |\mathcal{X}| - 1}{|\mathcal{X}| - 1} \\ &= \binom{6+4-1}{4-1} = \binom{9}{3} = \frac{9!}{3!6!} = \frac{9 \times 8 \times 7}{3 \times 2} = 84. \end{aligned} \tag{210}$$

(b) O número de sequências do tipo $P_{x_{1:6}}$ dado é o tamanho da classe de tipo deste tipo, dado pelo coeficiente polinomial

$$|T(P)| = \binom{n}{nP(a_1)\ nP(a_2)\ \ldots\ nP(a_n)} \tag{211}$$

$$= \binom{6}{3\ 2\ 1\ 0} \tag{212}$$

$$= \frac{6!}{3!2!1!0!} = 60. \tag{213}$$

(c) A probabilidade de uma sequência $x_{1:6}$ do tipo P é dada por

$$\begin{aligned} \Pr(x_{1:6}) &= p_1^{n_1} \times p_2^{n_2} \times p_3^{n_3} \times p_4^{n_4} \\ &= \left(\frac{1}{2}\right)^3 \times \left(\frac{1}{4}\right)^2 \times \left(\frac{1}{8}\right)^1 \times \left(\frac{1}{8}\right)^0 \times \\ &= \left(\frac{1}{2}\right)^{10} = 0.0009765625. \end{aligned} \tag{214}$$

A probabilidade da classe de tipo é dada pela soma das probabilidades de todas as sequências na classe de tipo. Como todas as sequências na classe de tipo possuem a mesma probabilidade, basta multiplicar o tamanho da classe de tipo pela probabilidade de uma sequência do tipo:

$$\Pr(T(P_{x_{1:6}})) = \Pr(x_{1:6}) \times |T(P)| = 60 \times \left(\frac{1}{2}\right)^{10} = 0.05859375. \tag{215}$$

2.8 Conjunto típico

1. Considere uma sequência i.i.d. de variáveis aleatórias $X_1, \ldots, X_n$ em um alfabeto ternário $\mathcal{X} = \{0, 1, 2\}$ com distribuição $p = (0.6, 0.3, 0.1)$.
   (a) Calcule $H(X)$.
   (b) Considere $n = 10$, $n = 50$, $n = 100$, $n = 200$, $n = 500$ e $n = 1000$ e $\epsilon = 0.1$. Quais sequências pertencem ao conjunto típico $A_\epsilon^{(n)}$? (Ou seja, quais tipos)? Quantos tipos existem no conjunto típico? Qual é a probabilidade do conjunto típico? Qual percentual do conjunto de todas as sequências está dentro do conjunto típico? Calcule os limites superiores e inferiores para: i) o tamanho do conjunto típico; e ii) a probabilidade do conjunto típico. Faça uma tabela para comparar os resultados para cada um dos valores de n. Verifique se o tamanho do conjunto típico está dentro dos limiares dados na teoria. (dica veja em `https://en.wikipedia.org/wiki/Multinomial_theorem` uma forma computacionalmente eficiente de calcular o coeficiente multinomial).
   (c) Para $n = 10$ e $n = 20$, faça um gráfico indicando o percentual das sequências do conjunto de todas as sequências que estão em cada uma das classes de tipo contidas no conjunto típico.

Resolução:

(a)
```
p = [0.6 0.3 0.1];
H = -sum(p.*log2(p));
H = 1.2955 bits
```

(b)
```
pxn = p1^k1 * p2^k2 * p3^k3  % onde p1, p2 e p3 são as
    ↪ probabilidades dos símbolos e k1, k2 e k3 o número de ocorrê
    ↪ ncia

eps=0.1; n=10;
count=0;
for k1=0:n, for k2=0:n-k1,
   pn=p(1)^k1*p(2)^k2*p(3)^(n-k1-k2);
   lpn=-(1/n)*log2(pn);
   if abs(lpn-H)<eps,
      count++;
      printf('%d %d %d\n',k1,k2,n-k1-k2);
   endif;
endfor; endfor;

count=0;
for k1=0:n, for k2=0:n-k1,
    k=[k1, k2, n-k1-k2];
    lpn=-(1/n)*(sum(k.*log2(p)));
    if abs(lpn-H)<eps,
```

```
            count++;
            printf('%d %d %d\n',k);
        endif;
    endfor; endfor;

    % https://en.wikipedia.org/wiki/Multinomial_theorem#
        ↪ Multinomial_coefficients
    function c=multicoeff (k),
        c=1;
        for i=1:length(k),
            c*=bincoeff(sum(k(1:i)),k(i));
        endfor;
    endfunction

    n=10; PA=0; count=0;
    for k1=0:n, for k2=0:n-k1,
        k=[k1, k2, n-k1-k2];
        lpn=-(1/n)*(sum(k.*log2(p)));
        pn=p(1)^k1*p(2)^k2*p(3)^(n-k1-k2);
        if abs(lpn-H)<eps,
            count++;
            printf('%d %d %d\n',k);
            PA+=pn*multicoeff(k);
        endif;
    endfor; endfor;

octave:81> H
H =  1.2955
octave:82> eps=0.1;
octave:83> n=500; PA=0; count=0;
for k1=0:n, for k2=0:n-k1, k=[k1, k2, n-k1-k2];
lpn=-(1/n)*(sum(k.*log2(p))); pn=p(1)^k1*p(2)^k2*p(3)^(n-k1-k2);
if abs(lpn-H)<eps, count++; PA+=pn*multicoeff(k); endif; endfor;
    ↪ endfor;
PA
PA =  0.99420
octave:84> n=200; PA=0; count=0; for k1=0:n, for k2=0:n-k1,
k=[k1, k2, n-k1-k2]; lpn=-(1/n)*(sum(k.*log2(p))); pn=p(1)^k1*p(2)^k2*
    ↪ p(3)^(n-k1-k2);
if abs(lpn-H)<eps, count++; PA+=pn*multicoeff(k); endif; endfor;
    ↪ endfor;
PA
PA =  0.91943
octave:85> n=100; PA=0; count=0; for k1=0:n, for k2=0:n-k1,
k=[k1, k2, n-k1-k2]; lpn=-(1/n)*(sum(k.*log2(p))); pn=p(1)^k1*p(2)^k2*
    ↪ p(3)^(n-k1-k2);
if abs(lpn-H)<eps, count++; PA+=pn*multicoeff(k); endif; endfor;
    ↪ endfor;
PA
PA =  0.78298
octave:86> n=10; PA=0; count=0; for k1=0:n, for k2=0:n-k1,
k=[k1, k2, n-k1-k2]; lpn=-(1/n)*(sum(k.*log2(p))); pn=p(1)^k1*p(2)^k2*
    ↪ p(3)^(n-k1-k2);
if abs(lpn-H)<eps, count++; PA+=pn*multicoeff(k); endif; endfor;
    ↪ endfor;
PA
```

```
PA =  0.21106
```

(c) to-do

2.9 Conjunto típico

1. Considere M a soma dos algarismos do seu número de matrícula e $m = 4 + M \mod 3$. Por exemplo, se o número de matrícula é 0123456789, teremos $M = 0+1+2+3+4+5+6+7+8+9 = 45$ e $m = 4+45 \mod 3 = 4 + 0 = 4$.

   Suponha $X_{1:n}$ uma sequência de comprimento n, de v.a. i.i.d., em um alfabeto $\mathcal{X} = \{1, 2, \ldots, m\}$, com distribuição p dada

$$p(X = x) = \begin{cases} \frac{1}{2} & x = 1, \\ \frac{1}{2^2} & x = 2, \\ \vdots & \\ \frac{1}{2^{m-2}} & x = m-2 \\ \frac{1}{2^{m-1}} & x = m-1 \text{ e } x = m. \end{cases} \tag{216}$$

   Seja $A_\epsilon^{(n)}$ o conjunto típico e considere $\epsilon = 0.1$ e $n = 16$.

   (a) Determine o limite superior e inferior para o tamanho do conjunto típico.

   (b) Qual é a sequência mais provável e qual é a menos provável?

   (c) Quantos tipos existem?

   (d) Qual é o tipo mais provável? Dê um exemplo de sequência deste tipo.

   (e) Qual é o limite superior e inferior para o tamanho da classe de tipo mais provável?

   (f) Qual é o tamanho da classe de tipo mais provável? Este valor está dentro dos limites encontrados no item anterior?

Resolução:

(a)
```
m = 4; n = 16; eps = 0.1;
p = 2.^(-[1:m]);
p(m)=p(m-1);

function H = entropy(p), id=find(p==0); p(id)=[];  H=-sum(p.*log2(p))
    ↪ ; endfunction
H = entropy(p)
L = ceil(2^(n*(H+eps)))
l = floor((1-eps)*2^(n*(H-eps)))
2^(n*H)
```

(b) Mais provável: sequência com apenas 1.
Menos provável: sequência em que x seja apenas m ou $m-1$.

(c) `nchoosek (n + m - 1, m - 1)`

(d) Aquele com a distribuição mais próximo da distribuição subjacente.

```
m=4 (1 (8), 2 (4), 3 (2), 4 (2))
     ex: [1,2,1,4,1,2,1,1,2,2,4,3,1,1,1,3]
[8, 4, 2, 2]
demais: (1 (8), 2 (4), 3 (2), 4 (1), 5 (1))
     ex: [1,2,1,4,1,2,1,1,2,2,5,3,1,1,1,3]
     [8, 4, 2, 1, 1], [8, 4, 2, 1, 1, 0], [8, 4, 2, 1, 1, 0, 0]
P = [8, 4, 2, 2]/n
P = [8, 4, 2, 1, 1]/n
```

(e)
```
HP = entropy(P)
LT = 2^(n*HP)
lT = 2^(n*HP)/((n+1)^m)
```

(f)
```
function c = multicoeff (n,m),
  %  Compute the multinomial coefficient using product of binomials
  %
  %       /           \
  %       |     n     |
  %  c =  | k1 ... km |
  %       \           /
  %
  %  c = multicoeff (k)
  %  where k is an array k = [k1 k2 ... km] and
  %  n is simplicity given by the sum: n = sum(k)
  %
  %  c = multicoeff (n,m)
  %  computes all multinomials at level n
  %  (m=2 for binomials, m=3 for trinomials, etc)

    if nargin < 2,
      k = n;
      c=1;
      for i=1:length(k),
        c*=bincoeff(sum(k(1:i)),k(i));
      endfor;
    else
      r = npermutek([0:n],m);
      id = find ( sum(r,2) == n);
      r = r(id,:);
      c = [];
      for i = 1:size(r,1),
        c(i) = multicoeff (r(i,:));
      endfor
      c = c';
    endif
  endfunction


  multicoeff (P*n)
```

3 Processos Estocásticos

3.1 Comportamento no limite

1. Sejam $X_1, X_2, \ldots$ tirados de forma i.i.d. de acordo com a seguinte distribuição:

$$X_i = \begin{cases} 1, & 1/2, \\ 2, & 1/4, \\ 3, & 1/4. \end{cases} \tag{217}$$

Encontre o comportamento limite para o produto

$$(X_1 X_2 \ldots X_n)^{1/n} \tag{218}$$

Resolução: Vamos definir

$$P_n = (X_1 X_2 \ldots X_n)^{\frac{1}{n}} = \left(\prod_{i=1}^{n} X_i \right)^{\frac{1}{n}}, \tag{219}$$

assim teremos

$$\log P_n = \frac{1}{n} \sum_{i=1}^{n} \log X_i \to \mathrm{E}[\log X], \tag{220}$$

pela lei forte dos grandes números. Assim, poderemos concluir que $P_n \to 2^{\mathrm{E}[\log X]}$.

Dada os valores que X_i assume e sua distribuição, podemos calcular

$$\mathrm{E}[\log X] = \frac{1}{2} \log 1 + \frac{1}{4} \log 2 + \frac{1}{4} \log 3 = \frac{1}{4} \log 6. \tag{221}$$

Desta forma, teremos $P_n \to 2^{\frac{1}{4} \log 6} = 6^{1/4} = 1.565$.

3.2 Taxa de entropia

1. Seja $X_1, X_2, \ldots$ variáveis aleatórias independentes e identicamente distribuídas de acordo com uma função massa probabilidade $p(x)$, $x \in \{1, 2, \ldots, m\}$. Então, $p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i)$. Sabemos que $-\frac{1}{n} \log p(X_1, X_2, \ldots, X_n) \to H(X)$ em probabilidade. Seja $q(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} q(x_i)$, onde q é outra função massa probabilidade em $\{1, 2, \ldots, m\}$.

   (a) Avalie $\lim -\frac{1}{n} \log q(X_1, X_2, \ldots, X_n)$ onde $X_1, X_2, \ldots$ são i.i.d. $\sim p(x)$.

   (b) Avalie agora o limite da razão de verossimilhança logarítmica $\frac{1}{n} \log \frac{q(X_1, \ldots, X_n)}{p(X_1, \ldots, X_n)}$, onde $X_1, X_2, \ldots$ são i.i.d. $\sim p(x)$. Desta forma, a vantagem favorecendo q é exponencialmente pequena quando p é verdadeiro.

Resolução:

(a)

$$\begin{aligned}
\lim -\frac{1}{n}\log q(X_1,X_2,\ldots,X_n) &= \lim -\frac{1}{n}\log\prod_{i=1}^{n} q(X_i) \\
&= \lim -\frac{1}{n}\sum_{i=1}^{n}\log q(X_i) \\
&= -\mathrm{E}[\log q(X)] \\
&= -\sum p(x)\log q(x) = \sum p(x)\log\frac{p(x)}{q(x)p(x)} \\
&= \sum p(x)\log\frac{p(x)}{q(x)} - \sum p(x)\log p(x) \\
&= D(p||q) + H(p) \qquad (222)
\end{aligned}$$

(b)

$$\begin{aligned}
\lim \frac{1}{n}\log\frac{q(X_1,X_2,\ldots,X_n)}{p(X_1,X_2,\ldots,X_n)} &= \lim \frac{1}{n}\log\frac{\prod_{i=1}^{n} q(X_i)}{\prod_{i=1}^{n} p(X_i)} \\
&= \lim \frac{1}{n}\sum_{i=1}^{n}\log\frac{q(X_i)}{p(X_i)} \\
&= \mathrm{E}\left[\log\frac{q(X)}{p(X)}\right] \\
&= \sum p(x)\log\frac{q(x)}{p(x)} \\
&= -\sum p(x)\log\frac{p(x)}{q(x)} \\
&= -D(p||q) \qquad (223)
\end{aligned}$$

3.3 Cadeia de Markov

1. Seja X_1 uniformemente distribuído sobre os estados $\{0,1,2\}$. Seja $\{X_i\}_1^\infty$ uma cadeia de Markov com matriz de transição P, ou seja, $P(X_{n+1}=j|X_n=i)=P_{ij}$, $i,j\in\{0,1,2\}$.

$$P=[P_{ij}]=\begin{bmatrix}\frac{1}{2} & \frac{1}{4} & \frac{1}{4}\\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4}\\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2}\end{bmatrix} \qquad (224)$$

   (a) $\{X_n\}$ é estacionária?

   (b) Calcule $\lim_{n\to\infty}\frac{1}{n}H(X_1,\ldots,X_n)$.

   (c) Considere agora um processo derivado $Z_1,Z_2,\ldots,Z_n$ onde

$$\begin{aligned} Z_1 &= X_1 \\ Z_i &= (X_i - X_{i-1}) \mod 3, \qquad i=2,\ldots,n \end{aligned} \qquad (225)$$

   Desta forma Z^n codifica as transições ao invés dos estados.
   Encontre $H(Z_1,Z_2,\ldots,Z_n)$.

   (d) Encontre $H(Z_n)$ e $H(X_n)$, para $n\geq 2$.

(e) Encontre $H(Z_n|Z_{n-1})$, para $n \geq 2$.

(f) Z_{n-1} e Z_n são independentes para $n \geq 2$?

Resolução:

(a) Seja μ_n a função massa de probabilidade no instante n. No instante $n = 1$ temos $\mu_1 = (1/3, 1/3, 1/3)$. No instante seguinte, teremos

$$\begin{aligned}
\mu_2 &= \mu_1 P && (226)\\
&= \left(\frac{1}{3}\times\frac{1}{2}+\frac{1}{3}\times\frac{1}{4}+\frac{1}{3}\times\frac{1}{4}, \frac{1}{3}\times\frac{1}{4}+\frac{1}{3}\times\frac{1}{2}+\frac{1}{3}\times\frac{1}{4}, \frac{1}{3}\times\frac{1}{4}+\frac{1}{3}\times\frac{1}{4}+\frac{1}{3}\times\frac{1}{2}\right)\\
&= \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) = \mu_1. && (227)
\end{aligned}$$

Da mesma forma podemos constatar que $\mu_n = (1/3, 1/3, 1/3)$ para todo n e assim $\{X_n\}$ é estacionário.

(b) Como $\{X_n\}$ é uma cadeia de Markov estacionária, teremos

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n) &= H(X_2|X_1)\\
&= \sum_{k=0}^{2} \Pr(X_1 = k) H(X_2|X_1 = k)\\
&= 3 \times \frac{1}{3} \times H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)\\
&= \frac{3}{2}. && (228)
\end{aligned}$$

(c) Como $X \in \{0, 1, 2\}$, teremos que existe uma função injetora de $(X_1, X_2, \ldots, X_n)$ em $(Z_1, Z_2, \ldots, Z_n)$ (eles são *one-to-one*). Desta forma, fazendo uso da regra da cadeia da entropia, podemos escrever

$$\begin{aligned}
H(Z_1, Z_2, \ldots, Z_n) &= H(X_1, X_2, \ldots, X_n)\\
&= \sum_{k=1}^{n} H(X_k|X_1, \ldots X_{k-1})\\
&= H(X_1) + \sum_{k=2}^{n} H(X_k|X_{k-1})\\
&= H(X_1) + (n-1) H(X_2|X_1)\\
&= \log 3 + (n-1) \times \frac{3}{2} && (229)
\end{aligned}$$

(d) Como $\{X_n\}$ é estacionário com $\mu_n = (1/3, 1/3, 1/3)$,

$$H(X_n) = H(X_1) = H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) = \log 3. \qquad (230)$$

Teremos $Z_n = 0$ quando $(X_n = 0, X_{n-1} = 0)$, $(X_n = 1, X_{n-1} = 1)$, $(X_n = 2, X_{n-1} = 2)$, assim $\Pr(Z_n = 0) = \frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{2} = \frac{1}{2}$.

Teremos $Z_n = 1$ quando $(X_n = 1, X_{n-1} = 0)$, $(X_n = 2, X_{n-1} = 1)$, $(X_n = 0, X_{n-1} = 2)$, assim $\Pr(Z_n = 1) = \frac{1}{3} \times \frac{1}{4} + \frac{1}{3} \times \frac{1}{4} + \frac{1}{3} \times \frac{1}{4} = \frac{1}{4}$.

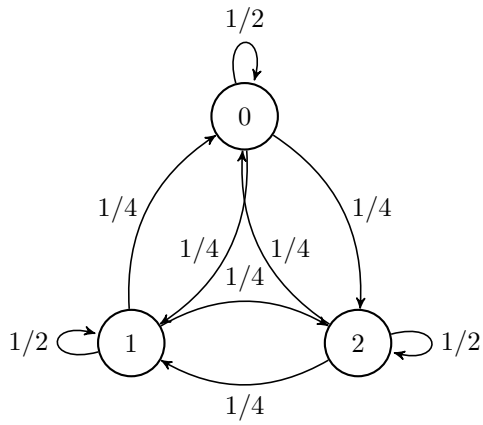
Teremos $Z_n = 2$ quando $(X_n = 2, X_{n-1} = 0)$, $(X_n = 0, X_{n-1} = 1)$, $(X_n = 1, X_{n-1} = 2)$, assim $\Pr(Z_n = 2) = \frac{1}{3} \times \frac{1}{4} + \frac{1}{3} \times \frac{1}{4} + \frac{1}{3} \times \frac{1}{4} = \frac{1}{4}$.

Para $n \geq 2$ temos

$$Z_n = \begin{cases} 0, & \frac{1}{2}; \\ 1, & \frac{1}{4}; \\ 2, & \frac{1}{4}. \end{cases} \tag{231}$$

Desta forma, $H(Z_n) = H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) = \frac{3}{2}$.

(e) Devido à simetria de P, $\Pr(Z_n|Z_{n-1}) = \Pr(Z_n)$ para $n \geq 2$. Pois, podemos verificar que Z_n depende apenas de X_n e X_{n-1}, sendo independente de X_{n-2}.



Desta forma, $H(Z_n|Z_{n-1}) = H(Z_n) = \frac{3}{2}$.

(f) Sim, pois $H(Z_n|Z_{n-1}) = H(Z_n)$.

3.4 Modelos de Ehrenfest

1. Paul Ehrenfest propôs um modelo simples para a troca de calor ou moléculas entre dois corpos isolados. Neste modelo da cadeia de Ehrenfest iremos considerar a troca de moléculas de gás entre dois compartimentos, rotulados por 0 e 1, conforme apresentado na Figura 3.

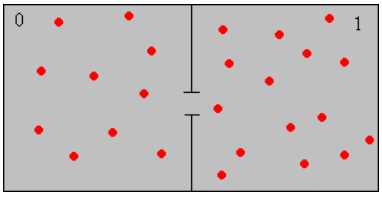


Figura 3: Modelo de Ehrenfest.

O número total de partículas no sistema é m. O estado do sistema em qualquer instante n pode ser descrito pelo número de partículas no compartimento 1. Iremos então denotar por X_n este número. Temos um processo estocástico gerando uma sequência de variáveis aleatórias $X_{1:n}$ que assumem valor em $\{0, 1, \ldots, m\}$.

Uma das m partículas no sistema é selecionada aleatoriamente, a cada instante, e esta partícula troca de

compartimento. Supondo que, no instante n, o compartimento 1 possua i bolas, o modelo de Ehrenfest nos fornece as seguintes probabilidades

$$\begin{aligned}\Pr(X_{n+1} = i+1 | X_n = i) &= \frac{m-i}{m}, \\ \Pr(X_{n+1} = i-1 | X_n = i) &= \frac{i}{m}.\end{aligned} \tag{232}$$

Podemos observar que neste processo, a probabilidade condicional é a mesma, se condicionarmos a todo histórico do processo até o instante n, ou seja,

$$\begin{aligned}\Pr(X_{n+1} = i+1 | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i) &= \Pr(X_{n+1} = i+1 | X_n = i), \\ \Pr(X_{n+1} = i-1 | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i) &= \Pr(X_{n+1} = i-1 | X_n = i).\end{aligned} \tag{233}$$

Esta propriedade é a propriedade de Markov, que diz que futuro e passado são independentes, dado o presente. Podemos verificar também que a distribuição condicional não depende de n, desta forma, temos um processo estocástico homogêneo. A cadeia de Markov poderá então ser descrita por uma matriz de transição fixa $P = [p_{ij}]_{ij}$.

Vamos considerar o caso em que $m = 4$ e faça o que se pede:

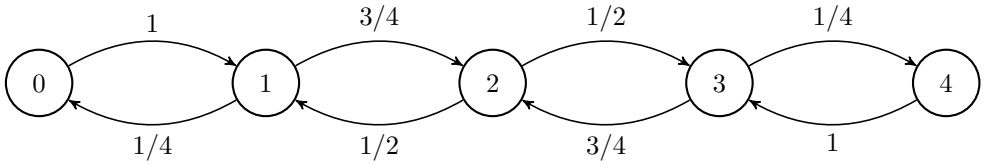
(a) Faça o grafo da cadeia de Markov.

(b) Encontre a matriz de transição.

(c) Calcule a distribuição de estado estacionário.

(d) Calcule a taxa de entropia do processo.

Resolução:



(a)

(b) Para $m = 4$, a matriz de transição ser

$$\mathbf{P} = \begin{array}{c} \\ \\ \\ \\ \\ \end{array}\begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/4 & 0 & 3/4 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \end{array}\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \tag{234}$$

A distribuição de estado estacionário é tal que $\mu^T P = \mu^T$, com $\sum_i \mu_i = 1$.

$$\begin{aligned}\mu^T P &= \mu^T \\ \mu^T (P - I) &= 0 \\ \mu^T Q &= 0\end{aligned} \tag{235}$$

Teremos aqui

$$Q = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 1/4 & -1 & 3/4 & 0 & 0 \\ 0 & 1/2 & -1 & 1/2 & 0 \\ 0 & 0 & 3/4 & -1 & 1/4 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}. \tag{236}$$

Note que, no sistema de equações acima, as colunas de P, e consequentemente as colunas de Q, são redundantes, de forma que uma dada coluna qualquer pode ser obtida através das demais pois sabemos que cada linha de P deve somar um e, da mesma forma, cada linha de Q deve somar zero. Podemos assim, sem prejuízo, substituir uma das colunas de Q de forma a incorporar a condição $\sum_i \mu_i = 1$. Basta para tanto substituir uma das colunas da matriz Q por uma coluna com uns e substituir um dos zeros no vetor de zeros, ao lado direito da equação, por um na posição respectiva. Podemos, por exemplo, escolher a última coluna de Q e assim teremos uma matriz

$$\tilde{Q} = \begin{pmatrix} -1 & 1 & 0 & 0 & 1 \\ 1/4 & -1 & 3/4 & 0 & 1 \\ 0 & 1/2 & -1 & 1/2 & 1 \\ 0 & 0 & 3/4 & -1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}, \tag{237}$$

e o novo sistema de equações será

$$\mu^T \tilde{Q} = (0, 0, 0, 1). \tag{238}$$

Observe que qualquer coluna de Q é determinada pelas demais colunas de Q, levando-se em consideração que a soma em cada linha de Q deve ser igual a 1, já que uma linha de Q representa a probabilidade de transição a um outro estado a partir do estado associado pela linha de Q em questão. Podemos então incorporar a condição $\sum_i \mu_i = 1$ substituindo como proposto anteriormente, ou seja, substituindo uma coluna de Q (por exemplo a j-ésima coluna) por uma coluna de uns, obtendo assim a matriz $\tilde{Q}$. Ao realizar a multiplicação de μ^T pela j-ésima coluna de $\tilde{Q}$, denotada por $\tilde{q}_{*,j}$, teremos

$$\mu^T \tilde{q}_{*,j} = \begin{bmatrix} \mu_1 & \dots & \mu_m \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \sum_i \mu_i = 1, \tag{239}$$

incorporando assim a condição de que μ deve pertencer ao simplex probabilístico.

Para solucionar a Equação 238, poderemos pós multiplicar ambos os lados pela matriz inversa de $\tilde{Q}$,

$$\mu^T = (0, 0, 0, 1)\tilde{Q}^{-1}. \tag{240}$$

```
P = [0 1 0 0 0; 1/4 0 3/4 0 0; 0 1/2 0 1/2 0; 0 0 3/4 0 1/4; 0 0 0 1
    ↪ 0];
Q = P - eye(size(P));
Q2 = Q;
Q2(:,5)=ones(5,1);
[0 0 0 0 1]*inv(Q2)
ans =
  0.062500   0.250000   0.375000   0.250000   0.062500
```

$$\mu^T = \left(\frac{1}{16}, \frac{1}{4}, \frac{6}{16}, \frac{1}{4}, \frac{1}{16}\right) \tag{241}$$

A taxa de entropia para um cadeia de Markov de primeira ordem estacionária será dada da

seguinte forma

$$
\begin{aligned}
H(\mathcal{X}) &= H'(\mathcal{X}) = \lim_{n\to\infty} H(X_n \mid X_{n-1}, \ldots, X_1) \\
&= \lim_{n\to\infty} H(X_n \mid X_{n-1}) \\
&\quad \text{dado que é Markov de 1a ordem} \\
&= H(X_2 \mid X_1) \quad \text{(estacionário)} \\
&= -\sum_{x_2,x_1} p(x_2, x_1) \log p(x_2 \mid x_1) = \sum_i \mu_i \left[-\sum_j p_{ij} \log p_{ij} \right] \\
&= \sum_i \mu_i H(\mathbf{p_i})
\end{aligned} \tag{242}
$$

onde μ é a distribuição estacionária, p_{ij} a probabilidade de transição de i para j (elementos da matriz P) e $\mathbf{p_i}$ a i-ésima linha da matriz P (as probabilidades de transição à partir do estado i).

Para o nosso exemplo, teremos

$$
\begin{aligned}
H(\mathcal{X}) &= \mu_1 H(\mathbf{p_1}) + \mu_2 H(\mathbf{p_2}) + \mu_3 H(\mathbf{p_3}) + \mu_4 H(\mathbf{p_4}) + \mu_5 H(\mathbf{p_5}) \\
&= \frac{1}{16} H(0,1,0,0,0) + \frac{1}{4} H(\frac{1}{4}, 0, \frac{3}{4}, 0, 0) + \frac{6}{16} H(0, \frac{1}{2}, 0, \frac{1}{2}, 0) + \frac{1}{4} H(0, 0, \frac{3}{4}, 0, \frac{1}{4}) \\
&\quad + \frac{1}{16} H(0,0,0,1,0) \\
&= 0 + 2 \times \frac{1}{4} \left(\frac{1}{4} \log 4 + \frac{3}{4} \log \frac{4}{3} \right) + \frac{6}{16} + 0 \\
&= \frac{1}{2} \left(\frac{1}{2} + \frac{3}{2} - \frac{3}{4} \log 3 \right) + \frac{6}{16} \\
&= 1 - \frac{3}{8} \log 3 + \frac{6}{16} = \frac{11}{8} - \frac{3}{8} \log 3 = 0.78064
\end{aligned} \tag{243}
$$

Note que tomando P^n, para n grande suficiente, verificaremos que P^n converge para uma matriz onde apenas duas linhas são distintas:

$$
\mu_1^T = \left(\frac{1}{8}, 0, \frac{3}{4}, 0, \frac{1}{8} \right) \quad \text{e} \quad \mu_2^T = \left(0, \frac{1}{2}, 0, \frac{1}{2}, 0 \right), \tag{244}
$$

sendo que P^n será da forma

$$
P^n = \begin{bmatrix} \mu_1^T \\ \mu_2^T \\ \mu_1^T \\ \mu_2^T \\ \mu_1^T \end{bmatrix} \tag{245}
$$

Observe que, neste caso, não haverá estado estacionário, mas a distribuição sobre os estados ficará oscilando entre μ_1^T e μ_2^T.

```
function y=plogp(p)
  y = zeros(size(p));
  id = find(p>0);
  y(id) = p(id).*log2(p(id));
endfunction;
```

```
function H=entropy(p)
   H = -sum(plogp(p));
endfunction;

function H=markovEntropyRate(P,mu)
  if nargin < 2,
     Q = P - eye(size(P));
     Q1 = Q;
     Q1(:,end)=ones(size(P,1),1);
     if det(Q1) == 0, % singular
        mu = [zeros(1,size(P,2)-1), 1] * pinv(Q1);
     else
        mu = [zeros(1,size(P,2)-1), 1] * inv(Q1);
     endif;
  endif;

  H=0;
  for i=1:size(P,1),
      H += mu(i)*(- sum( plogp( P(i,:) ) ) );
  endfor;
endfunction

P = [0 1 0 0 0; 1/4 0 3/4 0 0; 0 1/2 0 1/2 0; 0 0 3/4 0 1/4; 0 0 0 1
    ↪ 0];
H = markovEntropyRate(P);
```

Obverve novamente as equações:

$$\begin{aligned} \mu^T P &= \mu^T \\ \mu^T Q &= 0 \end{aligned} \tag{246}$$

Podemos verificar pela Equação 246 que μ deverá ser um autovetor à esquerda de P, com autovalor associado 1. De forma equivalente, através da Equação 246, deveremos ter que μ está no espaço nulo à esquerda de Q. Tomando a transposta da Equações 246

$$\begin{aligned} (\mu^T P)^T &= (\mu^T)^T \\ P^T \mu &= \mu, \end{aligned} \tag{247}$$

e da Equação 246 teremos

$$\begin{aligned} (\mu^T Q)^T &= 0^T \\ Q^T \mu &= 0, \end{aligned} \tag{248}$$

ou seja, μ é autovetor á direita de P^T, com autovalor associado 1, e de forma equivalente, é também um vetor no espaço nulo de Q^T.

Buscamos uma solução não trivial ($\mu \neq \mathbf{0}$). Desta forma, é necessário que a matriz Q seja singular, ou seja, não existe inversa. Isto correrá se, e somente se, seu determinante for nulo, $\det Q = 0$. Note que $\det Q = \det Q^T$, e o polinômio característico de P e P^T é o mesmo, e será aqui denotado por $X_P(\lambda)$.

$$X_P(\lambda) = \det(P - \lambda I) \tag{249}$$

Considerando que a matriz P é $n \times n$, poderemos escrever o polinômio característico como

$$X_P(\lambda) = \sum_{k=0}^{n} \lambda^{n-k} (-1)^k \operatorname{tr}(\Lambda^k P), \tag{250}$$

onde $\mathrm{tr}(\Lambda^k P)$ é o traço da k-ésima potência exterior de P. No caso 2×2 a equação característica será da forma

$$X_P(\lambda) = \lambda^2 - \mathrm{tr}(P)\lambda + \det(P). \tag{251}$$

Para o problema em questão, $\lambda = 1$ é um autovalor, que pode ocorrer com multiplicidade maior do que 1. Caso $\lambda = 1$ tenha multiplicidade 1, teremos uma cadeia de Markov ergódica. Este resultado é obtido a partir da utilização do teorema de Perron-Frobenius.

Para o exemplo desta questão, temos os seguintes autovalores

```
lambda = eig(P)
lambda =

   1.0000e+00
  -1.0000e+00
  -5.0000e-01
  -1.3754e-16
   5.0000e-01
```

3.5 Modelo Genético

1. O modelo genético simples de herança de traços (ou fenótipo) em animais consiste em considerar que estes traços são determinados por pares de genes. Neste modelo, cada um dos genes de um par pode ser de dois tipos G ou g. Um indivíduo pode ter uma das seguintes combinações: GG, Gg (que é geneticamente equivalente a gG) ou gg. Usualmente os tipos GG e Gg são indistinguíveis na aparência. Desta forma, dizemos que o gene G domina o gene g. Um indivíduo é chamado *dominante* se possuir os genes GG, *híbrido* se possuir os genes Gg e *recessivo* se possuir os genes gg.

   No acasalamento de dois animais, a prole herda um gene de cada par de cada um dos seus progenitores. A suposição básica em genética é que estes genes são escolhidos aleatoriamente e de forma independente um do outro. Esta suposição determina a probabilidade de ocorrência de cada um dos tipos de prole. A prole de dois pais puramente dominantes deverá ser dominante, a prole de dois pais recessivos deverá ser recessiva, e a prole de um dominante e um recessivo deverá ser híbrida.

   Vamos considerar agora um processo de reprodução continuada, onde a cada etapa há o acasalamento de dois indivíduos produzindo uma nova prole. Para simplificar, vamos inicialmente considerar que um indivíduo qualquer se acasalará com um indivíduo híbrido. No processo de acasalamento de um híbrido com um dominante, a prole necessariamente terá um gene G proveniente do dominante e poderá receber um gene G ou um gene g do indivíduo híbrido, com igual probabilidade. Ou seja, a prole será do tipo GG com probabilidade 0.5, Gg com probabilidade 0.5 e gg com probabilidade 0. Considerando agora o acasalamento entre um híbrido e um recessivo, a prole poderá ser Gg com probabilidade 0.5 e gg com probabilidade 0.5. Por fim, no acasalamento de dois híbridos, a prole poderá ser GG com probabilidade 0.25, Gg com probabilidade 0.5 e gg com probabilidade 0.25.

   Este processo de acasalamento com um híbrido pode ser descrito através de uma cadeia de Markov, onde os estados são o gene da prole, podendo ser GG, Gg e gg. Este processo estocástico é homogêneo, então a cadeia de Markov poderá ser descrita pela seguinte matriz

$$\mathbf{P} = \begin{array}{c} \\ GG \\ Gg \\ gg \end{array} \begin{array}{c} \begin{array}{ccc} GG & Gg & gg \end{array} \\ \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.5 & 0.5 \end{pmatrix} \end{array}. \tag{252}$$

   Suponha que agora o processo continuado de acasalamento seja feito sempre com um indivíduo domi-

nante. Neste caso, a matriz de transição será

$$\mathbf{P} = \begin{array}{c} \\ GG \\ Gg \\ gg \end{array} \begin{array}{c} \begin{array}{ccc} GG & Gg & gg \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0.5 & 0.5 & 0 \\ 0 & 1 & 0 \end{pmatrix} \end{array}. \tag{253}$$

Vamos supor agora um modelo mais realista. Suponha que temos dois animais, de sexo oposto, e suponha que os traço seja independente do sexo. Vamos iniciar o processo com dois indivíduos de sexo oposto, este casal irá reproduzir. Vamos selecionar dois descendentes e acasalar estes dois, continuando indefinidamente o processo.

Agora, cada estado será representado por um par de indivíduos. Os estados do processo serão: $s_1 = (GG, GG)$, $s_2 = (GG, Gg)$, $s_3 = (GG, gg)$, $s_4 = (Gg, Gg)$, $s_5 = (Gg, gg)$ e $s_6 = (gg, gg)$.

(a) Calcule as probabilidades de transição e encontre a matriz de transição P.

(b) Represente a cadeia de Markov através de um grafo.

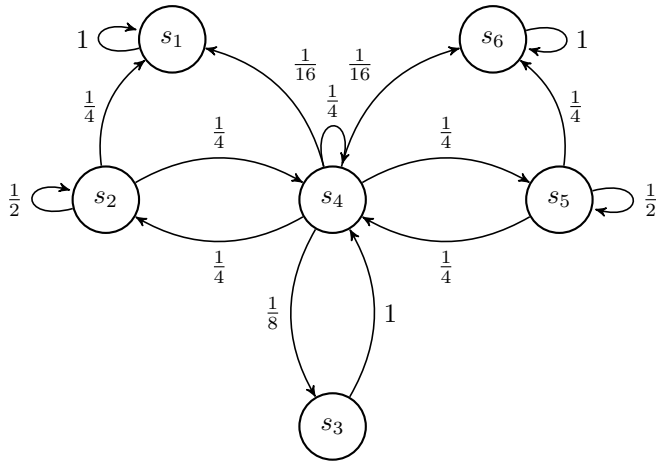
(c) Determine a distribuição de estado estacionário.

(d) Calcule a taxa de entropia do processo.

Resolução: Vamos ilustrar o cálculo das probabilidades de transição a partir do estado s_2. Quando o processo se encontra neste estado, um dos progenitores é do tipo GG e o outro é do tipo Gg. A probabilidade de gerar uma prole dominante é de 1/2. Desta forma, a probabilidade de transição para s_1 será de 1/4, a probabilidade de transição para s_2 será de 1/2 e a probabilidade de transição para s_4 será de 1/4.

A matriz de transição será da seguinte forma:

$$\mathbf{P} = \begin{array}{c} \\ GG,GG \\ GG,Gg \\ GG,gg \\ Gg,Gg \\ Gg,gg \\ gg,gg \end{array} \begin{array}{c} \begin{array}{cccccc} GG,GG & GG,Gg & GG,gg & Gg,Gg & Gg,gg & gg,gg \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.25 & 0.5 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0.0625 & 0.25 & 0.125 & 0.25 & 0.25 & 0.0625 \\ 0 & 0 & 0 & 0.25 & 0.5 & 0.25 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}. \tag{254}$$



```
P = [   1 0 0 0 0 0; ...
        0.25 0.5 0 0.25 0 0;
        0 0 0 1 0 0; ...
        0.0625 0.25 0.125 0.25 0.25 0.0625; ...
        0 0 0 0.25 0.5 0.25; ...
        0 0 0 0 0 1];
Q = P - eye(size(P));
Q1 = Q;
Q1(:,end)=ones(size(P,1),1);
det(Q1)
ans = 0
% faríamos como anteriormente,
% [zeros(1,size(P,2)-1), 1] * inv(Q1)
% mas a matriz Q eh singular
% vamos utilizar então a pseudo-inversa
[zeros(1,size(P,2)-1), 1] * pinv(Q1)
ans =

   0.50000   0.00000  -0.00000  -0.00000   0.00000   0.50000
% simulacao
mu=rand(1,6); mu=mu/sum(mu); for i=1:1E6, mu = mu*P; end; disp(mu)
   0.47487   0.00000   0.00000   0.00000   0.00000   0.52513

% utilizando a função criada no exercício anterior
H = markovEntropyRate(P);
% H =   -2.2898e-16
```

Podemos verificar que não existe solução única para este problema, o que fica explicito quando verificamos que a matriz $P - I$ é singular, não possuindo assim inversa. Qualquer solução da forma $\mu^T = (\mu_1, 0, 0, 0, 0, \mu_6)$ com $\mu_1 + \mu_6 = 1$ é uma solução válida, o que pode ser facilmente verificado realizando a multiplicação $\mu^T P$.

Note que, resolver $\mu^T P = \mu^T$ é equivalente a resolver $P^T \mu = \mu$, ou seja, buscamos o autovetor de P^T associado ao autovalor 1. Para este caso, o autovalor 1 possui multiplicidade 2, e assim observamos 2 autovetores associados ao autovalor 1.

```
[V, lambda] = eig(P')
V =

   1.00000   0.00000   0.00968   0.55780  -0.13608  -0.31623
   0.00000   0.00000  -0.26548  -0.32553   0.54433   0.63246
   0.00000   0.00000  -0.34752  -0.06217  -0.27217  -0.00000
   0.00000   0.00000   0.85912  -0.40238  -0.54433  -0.00000
   0.00000   0.00000  -0.26548  -0.32553   0.54433  -0.63246
   0.00000   1.00000   0.00968   0.55780  -0.13608   0.31623
lambda =

Diagonal Matrix

   1.00000         0         0         0         0         0
         0   1.00000         0         0         0         0
         0         0  -0.30902         0         0         0
         0         0         0   0.80902         0         0
         0         0         0         0   0.25000         0
         0         0         0         0         0   0.50000
```

Se denotarmos por $v_1 = (1,0,0,0,0,0)^T$ e $v_2 = (0,0,0,0,0,1)^T$, teremos que qualquer combinação da forma $\mu^T = \alpha v_1 + (1-\alpha)v_2$, para $0 \leq \alpha \leq 1$, será uma solução para o problema.

Conforme verificamos acima, na solução computacional, uma possível distribuição de estado estacionário é

$$\mu^T = \left(\frac{1}{2}, 0, 0, 0, 0, \frac{1}{2}\right). \tag{255}$$

Podemos calcular facilmente a taxa de entropia.

$$\begin{aligned} H(\mathcal{X}) &= \sum_i \mu_i H(\mathbf{p_i}) \\ &= \mu_1 H(1,0,0,0,0,0) + 0 + 0 + 0 + 0 + \mu_6 H(0,0,0,0,0,1) = 0. \end{aligned} \tag{256}$$

3.6 Taxa de entropia de uma cadeia de Markov de 1a ordem

1. Seja X_1 uniformemente distribuído sobre os estados $\{0,1,2\}$. Seja $\{X_i\}_1^\infty$ uma cadeia de Markov com matriz de transição P, ou seja, $P(X_{n+1} = j | X_n = i) = P_{ij}$, $i, j \in \{0,1,2\}$.

$$P = [P_{ij}] = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \tag{257}$$

   (a) $\{X_n\}$ é estacionária?

   (b) Calcule $\lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n)$.

Resolução:

(a) Como X_1 tem distribuição uniforme, a distribuição dos estados permanecerá a mesma nos instantes subsequentes, ou seja, teremos $\mu^T = \mu^T P$, o que pode ser facilmente verificado. Teremos então uma distribuição estacionária.

(b) A taxa de entropia para um cadeia de Markov de primeira ordem estacionária será dada da seguinte forma

$$\begin{aligned} H(\mathcal{X}) &= H'(\mathcal{X}) = \lim_{n\to\infty} H(X_n \mid X_{n-1}, \ldots, X_1) \\ &= \lim_{n\to\infty} H(X_n \mid X_{n-1}) \\ &\quad \text{dado que é Markov de 1a ordem} \\ &= H(X_2 \mid X_1) \quad \text{(estacionário)} \\ &= -\sum_{x_2, x_1} p(x_2, x_1) \log p(x_2 \mid x_1) = \sum_i \mu_i \left[-\sum_j p_{ij} \log p_{ij} \right] \\ &= \sum_i \mu_i H(\mathbf{p_i}) \end{aligned} \tag{258}$$

onde μ é a distribuição estacionária, p_{ij} a probabilidade de transição de i para j e $\mathbf{p_i}$ é a i-ésima linha da matriz P.

Teremos assim:

$$\begin{aligned} H(\mathcal{X}) &= \mu_1 H(\mathbf{p_1}) + \mu_2 H(\mathbf{p_2}) + \mu_3 H(\mathbf{p_3}) \\ &= H(\mathbf{p_1}) = -\tfrac{1}{2}\log\tfrac{1}{2} - \tfrac{1}{4}\log\tfrac{1}{4} - \tfrac{1}{4}\log\tfrac{1}{4} = \tfrac{3}{2}, \end{aligned} \tag{259}$$

onde utilizamos que a entropia das linhas são todas iguais, uma vez que são apenas permutações uma das outras e a distribuição de estado estacionário é uniforme.

3.7 Previsão do tempo

1. **Clima em Belo Horizonte** . Muitos belo-horizontinos acreditam que a melhor previsão para o tempo de amanhã é permanecer como está hoje. Assumindo-se que esta hipótese esteja correta, podemos modelar o tempo como uma cadeia de Markov. Para simplificar, vamos assumir que existam apenas dois estados: 'sol' e 'chuva'. Se o preditor ingenuo dos belo-horizontinos está correto 75% das vezes (independente do tempo de hoje ser 'sol' ou 'chuva'), então podemos modelar o tempo de Belo Horizonte como uma cadeia de Markov com dois estados $\mathcal{S} = \{s_1, s_2\}$ ($s_1 =$ 'sol' e $s_1 =$ 'chuva').

   Clima do Rio de Janeiro . No caso de Belo Horizonte, existe uma simetria perfeita entre 'sol' e 'chuva', de forma que 75% das vezes o clima de amanhã será igual ao de hoje, independente de como esteja o clima hoje. Este modelo pode ser realístico para Belo Horizonte, que é uma cidade localizada no interior do continente, entretanto, para o Rio de Janeiro, uma cidade litorânea, tempo de 'sol' é muito mais comum. Suponha então que, para o Rio, a probabilidade de ter 'sol' amanhã, dado que hoje é um dia de 'sol', é de 50%. Se hoje estiver com 'chuva', então a probabilidade de que amanhã faça 'sol' é de 90%.

   Para ambos modelos de previsão de tempo (Belo Horizonte e Rio de Janeiro) faça o que se pede abaixo.

   (a) Em ambos casos, podemos escrever uma matriz de transição para o modelo de previsão do tempo. Esta matriz será da forma
   $$P = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{pmatrix}.$$
   Qual é o valor de α e β para cada um dos dois modelos? Represente a cadeia de Markov através de um diagrama de transição.

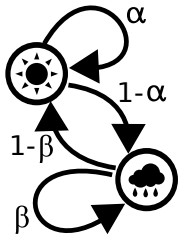
   (b) Em ambos casos a cadeia de Markov é invariante no tempo, possuindo assim uma distribuição estacionária $\mu = [\mu_1,\ \mu_2]$. Calcule μ_1 e μ_2.

   (c) Calcule a taxa de entropia para os dois modelos dados acima.

Resolução:

(a) **Belo Horizonte** : $\alpha = \beta = 0.75$.

Rio de Janeiro : $\alpha = 0.5$, $\beta = 0.1$.



(b)
$$\mu P = \mu \tag{260}$$

$$\begin{pmatrix} \mu_1 & \mu_2 \end{pmatrix} \begin{pmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{pmatrix} = \begin{pmatrix} \mu_1 & \mu_2 \end{pmatrix} \tag{261}$$

temos então

$$\begin{cases} \mu_1\alpha + \mu_2(1-\beta) = \mu_1 \\ \mu_1(1-\alpha) + \mu_2\beta = \mu_2 \\ \mu_1 + \mu_2 = 1 \end{cases} \tag{262}$$

e assim concluímos que

$$\mu_1 = \frac{1-\beta}{2-(\alpha+\beta)} \quad \text{e} \quad \mu_2 = \frac{1-\alpha}{2-(\alpha+\beta)} \tag{263}$$

Belo Horizonte : $\mu_1 = \mu_2 = 0.5$.

Rio de Janeiro : $\mu_1 = 0.64$, $\mu_2 = 0.36$.

(c)

$$\begin{aligned} H(\mathcal{X}) &= \sum_i \mu_i \left[-\sum_j p_{ij} \log p_{ij} \right] && (264) \\ &= \mu_1 \left[-\alpha \log \alpha - (1-\alpha)\log(1-\alpha) \right] + \mu_2 \left[-(1-\beta)\log(1-\beta) - \beta \log \beta \right] && (265) \\ &= \mu_1 H(\alpha) + \mu_2 H(\beta) && (266) \\ &= \frac{1-\beta}{2-(\alpha+\beta)} H(\alpha) + \frac{1-\alpha}{2-(\alpha+\beta)} H(\beta) && (267) \end{aligned}$$

Belo Horizonte : $H(\mathcal{X}) = 0.81$ bits.

Rio de Janeiro : $H(\mathcal{X}) = 0.81$ bits.

```
function H = entropy(p)
if length(p) == 1, p = [p, (1-p)]; end;
H = -sum(p.*log2(p));
endfunction

>> alpha=0.75; beta=0.75;
>> ((1 - beta)/(2 - (alpha+beta)))*entropy(alpha) + ...
((1 - alpha)/(2 - (alpha+beta)))*entropy(beta)
ans =  0.81128
>> alpha=0.5; beta=0.1;
>> ((1 - beta)/(2 - (alpha+beta)))*entropy(alpha) + ...
((1 - alpha)/(2 - (alpha+beta)))*entropy(beta)
ans =  0.81036
```

3.8 Caminhada do gato

1. A Figura 4 apresenta a planta de uma casa. Algumas portas da casa são de duplo-sentido, enquanto outras possuem sentido único, o que pode ser verificado na representação através das setas. Um gato caminha aleatoriamente pela casa. Quando o gato está em um cômodo, ele pode permanecer neste cômodo ou utilizar alguma porta para trocar de cômodo. Suponha que, a cada instante, o gato escolha entre as alternativas com igual probabilidade. Como existe um cachorro no quarto 3, o gato não permanece neste quarto, deixando-o imediatamente no próximo instante após ingressar neste quarto.
   (a) Encontre a matriz de transição e a cadeia de Markov para este problema (desenhe o grafo).
   (b) Qual é a distribuição em estado estacionário para o quarto no qual o gato está?
   (c) Calcule a taxa de entropia para a caminhada do gato.

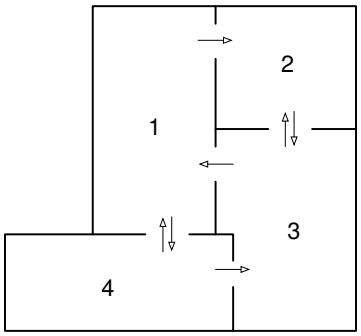


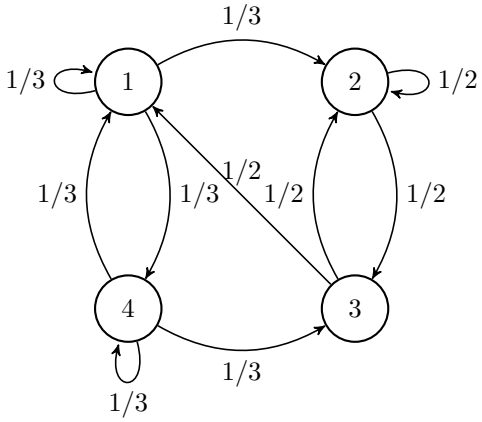
Figura 4: Caminhada do gato em uma casa.

Resolução:

(a) Poderemos considerar os quartos como estados em uma cadeia de Markov. Desta forma, a matriz de transição é dada por

$$P = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{pmatrix}. \tag{268}$$

A cadeia de Markov é ilustrada a seguir:



(b) A distribuição de estado estacionário é tal que $\mu^T P = \mu^T$, com $\sum_i \mu_i = 1$.

$$\begin{aligned} \mu^T P &= \mu^T & (269) \\ \mu^T (P - I) &= 0 & (270) \\ \mu^T Q &= 0 & (271) \end{aligned}$$

Teremos aqui

$$Q = \begin{pmatrix} -\frac{2}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & -\frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & -1 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & -\frac{2}{3} \end{pmatrix}. \tag{272}$$

Note que, no sistema de equações acima, podemos incorporar a condição $\sum_i \mu_i = 1$ bastando para tanto substituir uma das colunas da matriz Q por uma coluna com uns e substituir um dos zeros no vetor de zeros do lado direito da equação por um na posição respectiva. Podemos, por exemplo, escolher a última coluna de Q e assim teremos uma matriz

$$\tilde{Q} = \begin{pmatrix} -\frac{2}{3} & \frac{1}{3} & 0 & 1 \\ 0 & -\frac{1}{2} & \frac{1}{2} & 1 \\ \frac{1}{2} & \frac{1}{2} & -1 & 1 \\ \frac{1}{3} & 0 & \frac{1}{3} & 1 \end{pmatrix}, \tag{273}$$

e o novo sistema de equações será

$$\mu^T \tilde{Q} = (0, 0, 0, 1), \tag{274}$$

e a solução poderá ser obtida pós-multiplicando pela matriz inversa de $\tilde{Q}$,

$$\mu^T = (0, 0, 0, 1)\tilde{Q}^{-1}. \tag{275}$$

Calculando a inversa, encontramos

$$\tilde{Q}^{-1} = \begin{pmatrix} -\frac{21}{25} & -\frac{2}{5} & \frac{4}{25} & \frac{27}{25} \\ \frac{3}{5} & -2 & -\frac{2}{5} & \frac{9}{5} \\ \frac{3}{25} & -\frac{4}{5} & -\frac{22}{25} & \frac{39}{25} \\ \frac{6}{25} & \frac{2}{5} & \frac{6}{25} & \frac{3}{25} \end{pmatrix}, \tag{276}$$

e assim,

$$\mu^T = (\frac{6}{25}, \frac{2}{5}, \frac{6}{25}, \frac{3}{25}). \tag{277}$$

Para resolver o sistema, faremos

$$\mu^T \tilde{Q} = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} \tag{278}$$

$$\left(\mu^T \tilde{Q}\right)^T = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix}^T \tag{279}$$

$$\tilde{Q}^T \mu = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{280}$$

Iremos aplicar as operações elementares à matriz $\tilde{Q}^T$ aumentada para solucionar o sistema:

$$\begin{array}{c} l_1 \\ l_2 \\ l_3 \\ l_4 \end{array} \left(\begin{array}{cccc|c} -2/3 & 0 & 1/2 & 1/3 & 0 \\ 1/3 & -1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & -1 & 1/3 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{array}\right) \tag{281}$$

$$\begin{array}{c} -\frac{3}{2}l_1 \\ l_2 + \frac{1}{2}l_1 \\ 2l_3 \\ l_4 + \frac{3}{2}l_1 \end{array} \left(\begin{array}{cccc|c} 1 & 0 & -3/4 & -1/2 & 0 \\ 0 & -1/2 & 3/4 & 1/6 & 0 \\ 0 & 1 & -2 & 2/3 & 0 \\ 0 & 1 & 7/4 & 3/2 & 1 \end{array}\right) \tag{282}$$

$$\begin{array}{c} l_1 \\ -2l_2 \\ l_3 + 2l_2 \\ l_4 + 2l_2 \end{array} \left(\begin{array}{cccc|c} 1 & 0 & -3/4 & -1/2 & 0 \\ 0 & 1 & -3/2 & -1/3 & 0 \\ 0 & 0 & -1/2 & 1 & 0 \\ 0 & 0 & 13/4 & 11/6 & 1 \end{array}\right) \tag{283}$$

$$\begin{array}{c} l_1 \\ l_2 \\ -2l_3 \\ l_4 + \frac{13}{2} l_3 \end{array} \left(\begin{array}{cccc|c} 1 & 0 & -3/4 & -1/2 & 0 \\ 0 & 1 & -3/2 & -1/3 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 25/3 & 1 \end{array}\right) \tag{284}$$

Podemos assim concluir que:

$$\mu_4 = \frac{3}{25} \tag{285}$$

$$\mu_3 = 2\mu_4 = \frac{6}{25} \tag{286}$$

$$\mu_2 = \frac{3}{2}\mu_3 + \frac{1}{3}\mu_4 = \frac{2}{5} \tag{287}$$

$$\mu_1 = \frac{3}{4}\mu_3 + \frac{1}{2}\mu_4 = \frac{6}{25}. \tag{288}$$

e assim

$$\mu^T = (\frac{6}{25}, \frac{2}{5}, \frac{6}{25}, \frac{3}{25}). \tag{289}$$

(c) A taxa de entropia para um cadeia de Markov de primeira ordem estacionária será dada da seguinte forma

$$H(\mathcal{X}) = \sum_i \mu_i \left[-\sum_j p_{ij} \log p_{ij} \right] \tag{290}$$

$$= \sum_i \mu_i H(\mathbf{p}_i^T) \tag{291}$$

$$= \frac{6}{25} H\left(\frac{1}{3}, \frac{1}{3}, 0, \frac{1}{3}\right) + \frac{2}{5} H\left(0, \frac{1}{2}, \frac{1}{2}, 0\right) + \frac{6}{25} H\left(\frac{1}{2}, \frac{1}{2}, 0, 0\right) + \frac{3}{25} H\left(\frac{1}{3}, 0, \frac{1}{3}, \frac{1}{3}\right) \tag{292}$$

$$= \frac{6}{25} \log 3 + \frac{2}{5} + \frac{6}{25} + \frac{3}{25} \log 3 \tag{293}$$

$$= \frac{9}{25} \log 3 + \frac{16}{25} \approx 1.2106 \tag{294}$$

3.9 Eleições

1. Em uma eleição para presidente são registradas as intenções de votos para os candidatos. Suponha que a eleição conte com dois candidatos A e B. São realizadas diversas pesquisas ao longo do processo eleitoral. Sabe-se que os eleitores que declararam intenção de voto no candidato A, possuem 50% de chance de continuarem com a mesma intenção de voto na próxima pesquisa. Como o candidato B é mais persuasivo e/ou está menos implicado nos esquemas de propina (aparecendo menos nas notícias sobre corrupção), os eleitores que declararam votar em B, possuem 75% de chance de permanecer com a mesma intenção de voto na próxima pesquisa. Suponha que este processo perdure por um longo período.

   (a) Calcule a probabilidade de cada um dos candidatos vencer as eleições (se possível).

   (b) Calcule a entropia associada à série de pesquisas eleitorais.

Resolução:

(a) A matriz de transição para este problema é da forma

$$P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}. \tag{295}$$

A distribuição de estado estacionário deve satisfazer $\mu = \mu P$. Teremos então as equações

$$(1-a)\mu_1 + b\mu_2 = \mu_1 \tag{296}$$
$$a\mu_1 + (1-b)\mu_2 = \mu_2, \tag{297}$$

e assim constatamos que devemos ter $\mu_1 = \frac{b}{a}\mu_2$. Além disso, devemos ter $\mu_1 + \mu_2 = 1$, e assim podemos concluir que

$$\mu_1 = \frac{b}{a+b} \quad \text{e} \quad \mu_2 = \frac{a}{a+b}. \tag{298}$$

No caso em questão, temos $a = \frac{1}{2}$ e $b = \frac{1}{4}$, logo $\mu_1 = \frac{\mu_2}{2}$, $\mu_1 = \frac{1}{3}$ e $\mu_2 = \frac{2}{3}$. A probabilidade do candidato A vencer as eleições é de 33,3% e a probabilidade do candidato B vencer é de 66,6%.

(b) Como o processo estocástico em questão é uma cadeia de Markov de 1a ordem estacionária, teremos $H(\mathcal{X}) = H(X_2|X_1) = \sum_i \mu_i H(r_i)$, onde r_i representa a i-ésima linha da matriz de transição.

$$H(\mathcal{X}) = \mu_1 H(r_1) + \mu_2 H(r_2) = \frac{b}{a+b} H((1-a), a) + \frac{a}{a+b} H(b, (1-b)) \tag{299}$$
$$= \frac{1}{3} H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{3} H\left(\frac{1}{4}, \frac{3}{4}\right) \tag{300}$$
$$= \frac{1}{3} + \frac{2}{3}\left(\frac{1}{2} + \frac{3}{4}(2 - \log 3)\right) \tag{301}$$
$$= \frac{1}{3} + \frac{4}{3} - \frac{1}{2}\log 3 = \frac{5}{3} - \frac{1}{2}\log 3 \approx 0,87419. \tag{302}$$

2. Suponha agora que a eleição conte com três candidatos A, B e C. A probabilidade dos eleitores que declararam voto em A continuarem a preferir o candidato A é dada por $p_{A\to A} = 1/2$, a probabilidade de mudarem para o candidato B é $p_{A\to B} = 1/3$ e de mudarem para o candidato C é $p_{A\to C} = 1/6$. De forma semelhante, teremos $p_{B\to A} = 1/3$, $p_{B\to B} = 1/2$, $p_{B\to C} = 1/6$, $p_{C\to A} = 1/6$, $p_{C\to B} = 1/3$ e $p_{C\to C} = 1/2$.

   (a) Calcule a probabilidade de cada um dos candidatos vencer as eleições (se possível).

   (b) Calcule a entropia associada à série de pesquisas eleitorais.

Resolução:

(a) Para este caso teremos a seguinte matriz de transição:

$$P = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/3 & 1/2 & 1/6 \\ 1/6 & 1/3 & 1/2 \end{pmatrix}. \tag{303}$$

Iremos então resolver o sistema $\mu(P - I) = 0$, adicionando a condição $\sum_i \mu_i = 1$.

$$\left(\begin{array}{ccc|c} -1/2 & 1/3 & 1/6 & 0 \\ 1/3 & -1/2 & 1/3 & 0 \\ 1 & 1 & 1 & 1 \end{array}\right) \tag{304}$$

$$\left(\begin{array}{ccc|c} -1/2 & 1/3 & 1/6 & 0 \\ 0 & -5/18 & 4/9 & 0 \\ 0 & 5/3 & 4/3 & 1 \end{array}\right) \tag{305}$$

$$\left(\begin{array}{ccc|c} -1/2 & 1/3 & 1/6 & 0 \\ 0 & -5/18 & 4/9 & 0 \\ 0 & 0 & 4 & 1 \end{array}\right) \tag{306}$$

Logo, podemos concluir que $\mu_3 = 1/4$, $\mu_2 = 2/5$ e $\mu_1 = 7/20$, ou seja, o candidato A possui 35% de chance de vencer as eleições, B possui 40% e C possui 25%.

(b) Como o processo estocástico em questão é uma cadeia de Markov de 1a ordem estacionária, teremos $H(\mathcal{X}) = H(X_2|X_1) = \sum_i \mu_i H(r_i)$, onde r_i representa a i-ésima linha da matriz de transição (note que as linhas são permutações umas das outras).

$$\begin{aligned} H(\mathcal{X}) &= \mu_1 H(r_1) + \mu_2 H(r_2) + \mu_3 H(r_3) && (307) \\ &= H(r) = H\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right) && (308) \\ &= 1 + \frac{1}{3}\log 3 + \frac{1}{6}(1 + \log 3) = \frac{7}{6} + \frac{1}{2}\log 3 \approx 1,9591. && (309) \end{aligned}$$

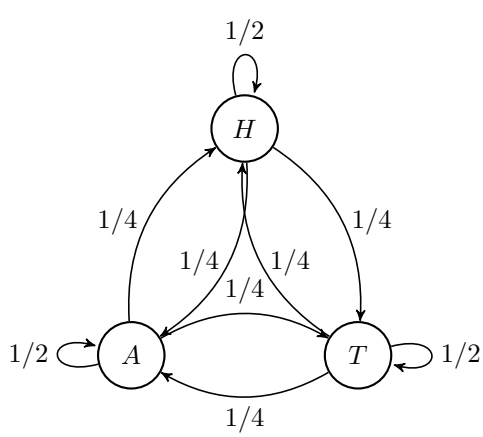
1. Foi observada uma língua em que são utilizados três símbolos: A, T, H. Foram observadas diversas sequências de símbolos produzidas por esta fonte e constatou-se que, ao longo das sequências, existe uma probabilidade $\frac{1}{2}$ de se repetir um dado símbolo. Caso não ocorra a repetição, haverá igual probabilidade do próximo símbolo ser cada um dos dois remanescentes.

   (a) Desenhe a cadeia de Markov de primeira ordem que modela este processo estocástico.

   (b) Como devemos proceder para determinar a distribuição de estado estacionário? Qual é a distribuição de estado estacionário?

   (c) Qual é a taxa de entropia deste processo?

Resolução:



(a)

(b) A matriz de transição para a cadeia de Markov $\{Y_n\}$ é

$$\mathbf{P} = \begin{array}{c} \\ A \\ T \\ H \end{array} \begin{array}{c} \begin{array}{ccc} A & T & H \end{array} \\ \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \end{array}. \tag{310}$$

Para encontrar a distribuição de estado estacionários, queremos encontrar $\mu^T P = \mu^T$, ou seja, $\mu^T(P - I) = 0$ e iremos denotar $Q = P - I$, assim teremos $\mu^T Q = 0$. Para incorporar a condução $\sum_i \mu_i = 1$, iremos substituir a última coluna de Q por uma coluna de uns e substituir o último elemento do vetor nulo à direita por 1. Teremos assim

$$\mu^T \tilde{Q} = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \tag{311}$$

$$\tilde{Q}^T \mu = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{312}$$

$$\begin{pmatrix} -1/2 & 1/4 & 1/4 \\ 1/4 & -1/2 & 1/4 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{313}$$

Devemos então resolver este sistema.

$$\left(\begin{array}{ccc|c} -1/2 & 1/4 & 1/4 & 0 \\ 1/4 & -1/2 & 1/4 & 0 \\ 1 & 1 & 1 & 1 \end{array}\right) \tag{314}$$

$$\left(\begin{array}{ccc|c} -1/2 & 1/4 & 1/4 & 0 \\ 0 & -3/8 & 3/8 & 0 \\ 0 & 3/2 & 3/2 & 1 \end{array}\right) \tag{315}$$

$$\left(\begin{array}{ccc|c} -1/2 & 1/4 & 1/4 & 0 \\ 0 & -3/8 & 3/8 & 0 \\ 0 & 0 & 3 & 1 \end{array}\right) \tag{316}$$

$$\tag{317}$$

Temos então que $\mu_3 = \frac{1}{3}$, $\mu_2 = \mu_3$, logo $\mu_2 = \frac{1}{3}$, e $\frac{1}{2}\mu_1 = \frac{1}{4}\mu_2 + \frac{1}{4}\mu_3$, logo $\mu_1 = \frac{1}{3}$. Assim, teremos

$$\mu^T = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}. \tag{318}$$

Pela simetria também podemos concluir que a distribuição de estado estacionário será $\mu = (1/31/31/3)$.

Resolvendo com auxílio do Octave, obtemos o mesmo resultados:

```
P = [0.5 0.25 0.25; 0.25 0.5 0.25; 0.25 0.25 0.5];
Q = P - eye(3);
Q2=Q; Q2(:,3)=ones(3,1);
[0 0 1]*inv(Q2)
ans =
   0.33333   0.33333   0.33333
```

(c) A taxa de entropia para um cadeia de Markov de primeira ordem estacionária será dada da

seguinte forma

$$
\begin{aligned}
H(\mathcal{X}) &= H'(\mathcal{X}) = \lim_{n\to\infty} H(X_n \mid X_{n-1}, \ldots, X_1) && (319)\\
&= \lim_{n\to\infty} H(X_n \mid X_{n-1}) && (320)\\
&\quad \text{dado que é Markov de 1a ordem} \\
&= H(X_2 \mid X_1) \quad \text{(estacionário)} && (321)\\
&= -\sum_{x_2,x_1} p(x_2, x_1) \log p(x_2 \mid x_1) = \sum_i \mu_i \left[-\sum_j p_{ij} \log p_{ij} \right] && (322)\\
&= \sum_i \mu_i H(\mathbf{p_i}) && (323)
\end{aligned}
$$

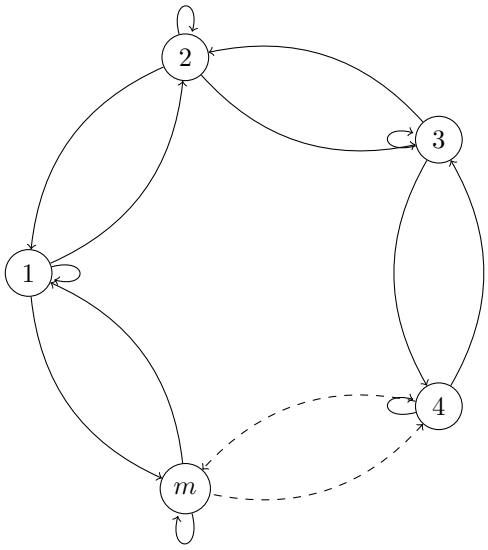
onde μ é a distribuição estacionária, p_{ij} a probabilidade de transição de i para j (elementos da matriz P) e $\mathbf{p_i}$ a i-ésima linha da matriz P (as probabilidades de transição à partir do estado i).

Para o nosso exemplo, teremos

$$
\begin{aligned}
H(\mathcal{X}) &= \mu_1 H(\mathbf{p_1}) + \mu_2 H(\mathbf{p_2}) + \mu_3 H(\mathbf{p_3}) = 3 \times \frac{1}{3} H(\mathbf{p_1}) && (324)\\
&= H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) && (325)\\
&= \frac{1}{2} + 2 \times \frac{1}{4} \times 2 = \frac{3}{2} \text{ bits.} && (326)
\end{aligned}
$$

3.10 Cadeia de Markov com m estados

1. Considere a cadeia de Markov com m estados ilustrada abaixo, onde todas as transições possuem a mesma probabilidade.



(a) Determine a distribuição de estado estacionário.

(b) Determine a taxa de entropia desta cadeia de Markov.

Resolução:

(a) `pela simetria: mu = 1/m*ones(1,m)`

(b)
```
% exemplo com m=5
m=5; p = [1/3, 1/3, 1/3, zeros(1,m-3)];
P = []; for i=-1:m-2, P(i+2,:)=shift(p,i); endfor
Q = P - eye(size(P)); Q1 = Q; Q1(:,end) = ones(size(P,1),1);
mu = [zeros(1, size(P,2)-1),1] * inv(Q1);

todas as linhas sao iguais
H = H(p_1) = log2(3) = 1.5850

H = \sum_i \mu_i H(p_i)
H=0; for i=1:length(mu), H+=mu(i)*entropy(P(i,:)); endfor
```

3.11 Caminhada do Pombo

1. Suponha que um pombo esteja andando em um tabuleiro unidimensional infinito. Este tabuleiro pode ser representado pelo conjunto dos números inteiros. Em cada instante o pombo dá um passo e muda de posição, podendo ir para frente (um passo em direção aos números positivos) ou para trás (um passo em direção aos números negativos). O pombo muda de direção com probabilidade $p = 0.25$. Suponha que o estado inicial seja $X_0 = 0$. O primeiro passo pode ser para frente ou para trás, com igual probabilidade. Um exemplo típico de caminhada deste pombo sobre o tabuleiro seria:

$$(X_0, X_1, X_2, \ldots) = (0, 1, 2, 3, 4, 3, 2, 1, 0, -1, -2, -1, 0, 1, 0, \ldots) \tag{327}$$

(a) Calcule o valor aproximado da entropia da sequência de comprimento $n + 1$, $H(X_0, X_1, \ldots, X_n)$, para $n = 1001$.

(b) Calcule a taxa de entropia para a caminhada do pombo.

(c) Qual é o número esperado de passos que o pombo dá em uma direção antes de mudar de direção? *dica:* $\sum_{k=0}^{\infty} kr^k = {}^{r}/_{(1-r)^2}$.

Resolução:

(a) Pela regra da cadeia temos

$$H(X_{0:n}) = \sum_{i=0}^{n} H(X_i|X_{0:i-1}) \tag{328}$$

$$= H(X_0) + H(X_1|X_0) + \sum_{i=2}^{n} H(X_i|X_{i-2:i-1}) \tag{329}$$

Onde utilizamos que a caminhada do pombo sobre o tabuleiro é uma cadeia de Markov de segunda ordem, visto que o próximo estado X_i depende apenas dos dois estados anteriores X_{i-1} e X_{i-2}, ou seja, $H(X_i|X_{0:i-1}) = H(X_i|X_{i-2:i-1})$ para $i \geq 2$.

O enunciado do problema fornece: $H(X_0) = 0$, pois o estado inicial é determinístico; $H(X_1|X_0) = H({}^{1}/_{2}) = 1$, já que o primeiro passo é igualmente provável para qualquer um dos dois lados; e

$H(X_i|X_{i-2:i-1}) = H(1/4, 3/4) = 2 - 3/4 \log 3$ pois a probabilidade de mudar de direção é dada $p = 0.25$.

Assim teremos

$$H(X_{0:n}) = 1 + (n-1)(2 - 3/4 \log 3) \tag{330}$$
$$H(X_{0:1001}) = 1 + 1000 \times 0.811 = 812. \tag{331}$$

(b) A taxa de entropia é dada por

$$\lim_{n\to\infty} \frac{H(X_{0:n})}{n+1} = \lim_{n\to\infty} \frac{1 + (n-1)(2 - 3/4 \log 3)}{n+1} \tag{332}$$
$$= (2 - 3/4 \log 3) = 0.811 \tag{333}$$

(c) Seja K o número de passos antes de inverter a direção. Queremos calcular $\mathrm{E}[K]$.

$$\mathrm{E}[K] = \sum_{k=1}^{\infty} k \Pr(K = k) \tag{334}$$
$$= \sum_{k=1}^{\infty} k (3/4)^{k-1} (1/4) \tag{335}$$
$$= \frac{1}{4} \frac{1}{(1 - 3/4)^2} = 4 \tag{336}$$

onde utilizamos que

$$\sum_{k=1}^{\infty} k r^{k-1} = \frac{1}{r} \sum_{k=1}^{\infty} k r^k \tag{337}$$
$$= \frac{1}{r} \frac{r}{(1-r)^2} = \frac{1}{(1-r)^2}. \tag{338}$$

4 Códigos

4.1 Unívocos e instantâneos

1. Quais dos seguintes códigos são univocamente decodificáveis e quais são instantâneos?
   (a) $C_1 = \{11, 01, 0\}$
   (b) $C_2 = \{00, 01, 100, 101, 11\}$
   (c) $C_3 = \{0, 10, 110, 1110, \ldots\}$
   (d) $C_4 = \{0, 00, 000, 0000\}$

Resolução:

(a) Este código é unicamente decodificável (livre de sufixo), mas não é instantâneo (livre de prefixo). Verifique que satisfaz a desigualdade de Kraft: $2^{-2} + 2^{-2} + 2^{-1} = 1 \leq 1$.

(b) Este código é instantâneo (livre de prefixo) e por conseguinte unicamente decodificável. O código satisfaz Kraft: $2^{-2} + 2^{-2} + 2^{-3} + 2^{-3} + 2^{-2} = 1 \leq 1$.

(c) Código instantâneo. Satisfaz Kraft: $2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} + \ldots = 1 \leq 1$.

(d) Este código não é unicamente decodificável, sequer instantâneo.

4.2 Códigos e entropia

1. Seja X uma variável aleatória com alfabeto $\mathcal{X} = \{1, 2, 3\}$ e distribuição

$$X = \begin{cases} 1, & \text{com probabilidade } 1/2; \\ 2, & \text{com probabilidade } 1/4; \\ 3, & \text{com probabilidade } 1/4. \end{cases} \tag{339}$$

Seja $X_1, X_2, \ldots$ independente identicamente distribuído seguindo esta distribuição e seja $Z_1 Z_2 Z_3 \ldots = C(X_1)C(X_2)\ldots$ a *string* formada pelos símbolos binários resultantes da concatenação das palavras de código, onde o código é $C = \{0, 10, 11\}$. Por exemplo, 122 torna-se 01010.

(a) Encontre a entropia de $H(\mathcal{X})$ e a taxa de entropia $H(\mathcal{Z})$ em bits por símbolo. Observe que Z não pode ser comprimido além deste ponto.

(b) Considere agora o código

$$C(x) = \begin{cases} 00, & \text{se } x = 1; \\ 10, & \text{se } x = 2; \\ 01, & \text{se } x = 3. \end{cases} \tag{340}$$

e encontre a taxa de entropia $H(\mathcal{Z})$.

(c) Considere agora o seguinte código

$$C(x) = \begin{cases} 00, & \text{se } x = 1; \\ 1, & \text{se } x = 2; \\ 01, & \text{se } x = 3. \end{cases} \tag{341}$$

e encontre a taxa de entropia $H(\mathcal{Z})$.

Resolução:

(a) Note que os X_is são i.i.d., assim

$$\begin{aligned} H(\mathcal{X}) &= \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n) && (342) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i) && (343) \\ &= \lim_{n\to\infty} \frac{1}{n} n H(X_1) && (344) \\ &= H(X_1) && (345) \\ &= \frac{1}{2}\log 2 + 2 \times \frac{1}{4}\log 4 = \frac{3}{2}. && (346) \end{aligned}$$

Podemos observar que o código $C(x)$ dado é ótimo para a distribuição em X dada. Além disso, como a distribuição é diádica, não há ganho em realizar a compressão em blocos. O processo resultante de $Z_1 Z_2 \ldots$ é binário $\sim$ Bernoulli(1/2). Desta forma, $H(\mathcal{Z}) = H(1/2) = 1$ bit.

(b) Iremos utilizar aqui o seguinte:

$$
\begin{aligned}
H(\mathcal{X}) &= \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n) && (347)\\
\lim_{n\to\infty} nH(\mathcal{X}) &= \lim_{n\to\infty} H(X_1, X_2, \ldots, X_n) && (348)\\
\lim_{n\to\infty} \frac{n}{2} H(\mathcal{X}) &= \lim_{n\to\infty} H(X_1, X_2, \ldots, X_{n/2}) && (349)
\end{aligned}
$$

Como todas as palavras do código possuem comprimento 2, teremos

$$
\begin{aligned}
H(\mathcal{Z}) &= \lim_{n\to\infty} \frac{1}{n} H(Z_1, Z_2, \ldots, Z_n) && (350)\\
&= \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_{n/2}) && (351)\\
&= \lim_{n\to\infty} \frac{1}{n}\frac{n}{2} H(\mathcal{X}) && (352)\\
&= \frac{H(\mathcal{X})}{2} && (353)\\
&= \frac{3}{4} && (354)
\end{aligned}
$$

(c) Se $H'(\mathcal{Z}) = \lim_{n\to\infty} H(Z_n|Z_1, \ldots, Z_{n-1})$ existe, então pelo lemma da média Cesaro teremos que $H(\mathcal{Z}) = \lim_{n\to\infty} H(Z_1, \ldots, Z_n)$ existe e $H'(\mathcal{Z}) = H(\mathcal{Z})$. É suficiente então mostrar que $H(Z_n|Z_1, \ldots, Z_{n-1})$ converge para um limite.

Vamos definir uma sequência de v.a. Y_n de forma que

$$
Y_n = \begin{cases} 1, & \text{se } Z_1, \ldots, Z_{n-1} \text{ é uma sequência completa de palavras,} \\ & \text{i.e. se } Z_n \text{ é o início de uma nova palavra} \\ 2, & \text{caso contrário.} \end{cases} \qquad (355)
$$

Note que Y_n é uma função de $(Z_1, \ldots, Z_{n-1})$. Note também que Z_n e $(Z_1, \ldots, Z_{n-1})$ são condicionalmente independentes dado Y_n. Em outras palavras, Y_n é uma estatística suficiente em $(Z_1, \ldots, Z_{n-1})$ sobre Z_n. Então,

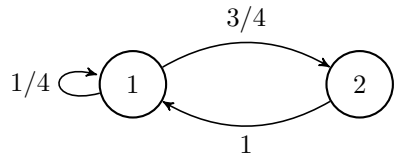
$$
\begin{aligned}
H(Z_n|Z_1, \ldots, Z_{n-1}) &= H(Z_n|Y_n) && (356)\\
&= \sum_y p(Y_n = y) H(Z_n|Y_n = y) && (357)\\
&= p(Y_n = 1) H(Z_n|Y_n = 1) + p(Y_n = 2) H(Z_n|Y_n = 2) && (358)\\
&= p(Y_n = 1) H(1/4) + p(Y_n = 2) H(1/3) && (359)
\end{aligned}
$$

onde acima utilizamos que $H(Z_n|Y_n = 1) = H(1/4)$ e $H(Z_n|Y_n = 2) = H(1/3)$, uma vez que podemos obter $p(z, y)$, conforme abaixo

z \ y	1	2
0	3/8	1/4
1	1/4	1/8

e assim teremos $p(z = 0|y = 1) = 3/4$, $p(z = 1|y = 1) = 1/4$, $p(z = 0|y = 2) = 2/3$ e $p(z = 1|y = 2) = 1/3$.

Embora $p(Y_n)$ mude com n, $p(Y_n)$ converge para uma distribuição estacionária única μ, já que Y_n é uma cadeia de Markov irredutível e aperiódica, conforme ilustrado abaixo:



A matriz de transição para a cadeia de Markov $\{Y_n\}$ é

$$\mathbf{P} = \begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{pmatrix} 1/4 & 3/4 \\ 1 & 0 \end{pmatrix} \end{array}. \tag{360}$$

Para encontrar a distribuição de estado estacionários, queremos encontrar $\mu^T P = \mu^T$, ou seja, $\mu^T(P - I) = 0$ e iremos denotar $Q = P - I$, assim teremos $\mu^T Q = 0$. Para incorporar a condução $\sum_i \mu_i = 1$, iremos substituir a última coluna de Q por uma coluna de uns e substituir o último elemento do vetor nulo à direita por 1. Teremos assim

$$\mu^T \tilde{Q} = \begin{pmatrix} 0 & 1 \end{pmatrix} \tag{361}$$

$$\tilde{Q}^T \mu = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{362}$$

$$\begin{pmatrix} -3/4 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{363}$$

Devemos então resolver este sistema.

$$\left(\begin{array}{cc|c} -3/4 & 1 & 0 \\ 1 & 1 & 1 \end{array}\right) \tag{364}$$

$$\left(\begin{array}{cc|c} -3/4 & 1 & 0 \\ 0 & 7/3 & 1 \end{array}\right) \tag{365}$$

$$\left(\begin{array}{cc|c} -3/4 & 1 & 0 \\ 0 & 1 & 3/7 \end{array}\right) \tag{366}$$

$$\left(\begin{array}{cc|c} -3/4 & 0 & -3/7 \\ 0 & 1 & 3/7 \end{array}\right) \tag{367}$$

$$\left(\begin{array}{cc|c} 1 & 0 & 4/7 \\ 0 & 1 & 3/7 \end{array}\right) \tag{368}$$

A análise mostra que $\mu = (4/7, 3/7)$. Desta forma,

$$\lim_{n\to\infty} H(Z_n|Z_1, \ldots, Z_{n-1}) = \frac{4}{7} H\left(\frac{1}{4}\right) + \frac{3}{7} H\left(\frac{1}{3}\right) \tag{369}$$

$$= \frac{6}{7} = H(\mathcal{Z}). \tag{370}$$

4.3 Código de Huffman

1. Considere a seguinte variável aleatória

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.49 & 0.26 & 0.12 & 0.04 & 0.04 & 0.03 & 0.02 \end{pmatrix} \tag{371}$$

(a) Encontre o código Huffman binário para X.

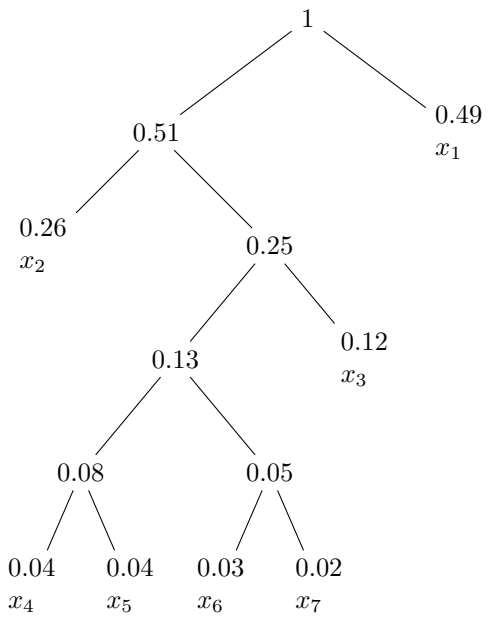
(b) Qual é o comprimento esperado para este código?

(c) Encontre o código Huffman ternário para X.

(d) Qual é o comprimento esperado para este código?

Resolução:

(a) O código de Huffman é apresentado a seguir:



símbolo	código	comprimento
x_1	1	1
x_2	00	2
x_3	011	3
x_4	01000	5
x_5	01001	5
x_6	01010	5
x_7	01011	5

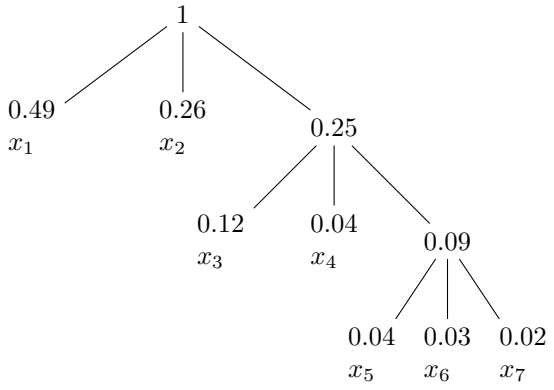
(b) Comprimento esperado:

$$L(C) = \sum_x p(x)l(x) = 2.02. \tag{372}$$

Entropia:

$$H(X) = -\sum_x p(x)\log p(x) = 2.0128\text{bits}. \tag{373}$$

(c) Para o código ternários, teremos:



símbolo	código	comprimento
x_1	0	1
x_2	1	1
x_3	20	2
x_4	21	2
x_5	220	3
x_6	221	3
x_7	222	3

(d) Comprimento esperado:

$$L(C) = \sum_x p(x)l(x) = 1.34. \tag{374}$$

Entropia:

$$H_3(X) = -\sum_x p(x)\log_3 p(x) = 1.2699\text{trits}. \tag{375}$$

4.4 huffman e shannon-fano-elias

1. Você deverá criar o códigos de Huffman e Shannon-Fano-Elias para comprimir um texto com as seguintes características. O texto possui apenas os símbolos 'a', 'b', 'c' e 'd'. Estes símbolos aparecem 50, 100, 25 e 25 vezes, respectivamente. Suponha que inicialmente o seu texto está codificado em ASCII (8 bits por símbolo). Faça o que se pede.
   (a) Calcule a entropia da fonte.
   (b) Calcule o tamanho original do texto.
   (c) Encontre o código de Huffman para este texto.
   (d) Encontre o código de Shannon-Fano-Elias para este texto.
   (e) Calcule a eficiência dos códigos.
   (f) Calcule o tamanho do arquivo codificado com cada um dos dois códigos.
   (g) Faça a árvore para ambos códigos.
   (h) Caracterize os códigos quanto aos seguintes aspectos: singularidade, univocidade, instantaneidade.
   (i) Verifique quais dos dois códigos satisfaz a desigualdade de Kraft.

Resolução:

(a) A estimativa da distribuição da fonte é

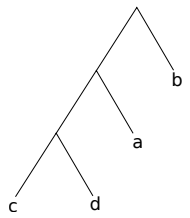
$$p = \begin{pmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{8} & \frac{1}{8} \end{pmatrix} \tag{376}$$

$$\begin{aligned} H &= -\sum_p p_i \log p_i && (377)\\ &= -\frac{1}{4}\log\frac{1}{4} - \frac{1}{2}\log\frac{1}{2} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{8}\log\frac{1}{8} && (378)\\ &= -\frac{1}{4}(-2) - \frac{1}{2}(-1) - 2\times\frac{1}{8}(-3) && (379)\\ &= \frac{1}{2} + \frac{1}{2} + \frac{3}{4} = \frac{7}{4} && (380) \end{aligned}$$

```
>> f=[50 100 25 25];
>> p=f./sum(f);
>> entropy(p)
ans =  1.7500
```

(b)

$$(50 + 100 + 25 + 25) \times 8 = 200 \times 8 = 1600 \text{ bits}. \tag{381}$$

(c) 


```
>> dict = huffmandict({'a','b','c','d'}, p)
dict =
{
  [1,1] =  0   1
  [1,2] =  1
  [1,3] =  0   0   0
  [1,4] =  0   0   1
}
```

$El_{\text{huffman}} = 0.5 + 0.25 \times 2 + 0.125 \times 3 + 0.125 \times 3 = 1.75$

(d)

x	$p(x)$	$F(x)$	$\overline{F}(x)$	$\overline{F}(x)$ (binário)	$l(x)$	palavra código
a	0.25	0.25	0.125	0.001	3	001
b	0.5	0.75	0.5	0.10	2	10
c	0.125	0.875	0.8125	0.1101	4	1101
d	0.125	1.0	0.9375	0.1111	4	1111

$El = 2.75$ bits

(e)

$$\text{eficiência}_{\text{huffman}} = \frac{1.75}{1.75} = 1 \tag{382}$$

$$\text{eficiência}_{\text{shannon-fano}} = \frac{1.75}{2.75} = \frac{7}{4} \times \frac{4}{11} = \frac{7}{11} = 0.64 \tag{383}$$

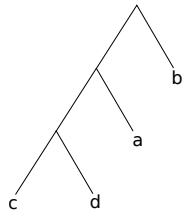
(f) Huffman

$$50 \times 2 + 100 \times 1 + 25 \times 3 + 25 \times 3 = 350 \text{ bits} \tag{384}$$

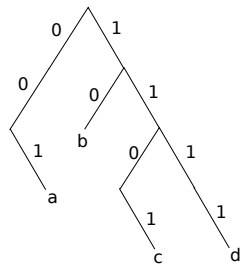
Shannon-Fano-Elias

$$50 \times 3 + 100 \times 2 + 25 \times 4 + 25 \times 4 = 550 \text{ bits} \tag{385}$$

Árvore de Huffman: $((a, (c, d)), b)$.



Árvore de Shannon-Fano-Elias:



(g) Ambos códigos são códigos de prefixo (instantâneos), por conseguinte, são singulares e unívocos.

(h) Ambos códigos são códigos de prefixo, devem portanto satisfazer a desigualdade de Kraft.

Vamos verificar.

O código de Huffman possui comprimentos $l = (2, 1, 3, 3)$.

$$\sum_i 2^{-l_i} = 2^{-2} + 2^{-1} + 2^{-3} + 2^{-3} \tag{386}$$

$$= \frac{1}{4} + \frac{1}{2} + \frac{1}{8} + \frac{1}{8} = 1 \leq 1. \tag{387}$$

O código de Shannon-Fano-Eliasn possui comprimentos $l = (3, 2, 4, 4)$.

$$\sum_i 2^{-l_i} = 2^{-3} + 2^{-2} + 2^{-4} + 2^{-4} \tag{388}$$

$$= \frac{1}{8} + \frac{1}{4} + \frac{1}{16} + \frac{1}{16} = \frac{1}{2} \leq 1. \tag{389}$$

4.5 Códigos binários

1. São dados os seguintes códigos binários: $C_I = \{0, 00, 10, 10\}$, $C_{II} = \{0, 00, 000, 0000\}$, $C_{III} = \{00, 01, 10, 000\}$, $C_{IV} = \{00, 01, 10, 11\}$ e $C_V = \{0, 10, 110, 111\}$. Caracterize os códigos segundo aos seguintes critérios, justificando sua resposta:
   (a) é um código singular ou não singular?;
   (b) é um código univocamente decodificável?;
   (c) satisfaz a desigualdade de Kraft?;
   (d) é um código de prefixo?
   (e) caso o código não seja de prefixo, é possível obter um código de prefixo com o mesmo comprimento esperado?

Resolução:

(a) O código $C_I = \{0, 00, 10, 10\}$ é singular, pois existe mais de um símbolo codificado com a mesma palavra (00 e 10). Este código não é por conseguinte univocamente decodificável e sequer de prefixo. Além disso, não satisfaz a desigualdade de Kraft: $\sum_i 2^{-l_i} = 2^{-1} + 2^{-2} + 2^{-2} + 2^{-2} = 5/4 > 1$. Não é possível obter um código de prefixo com o mesmo comprimento esperado.

(b) O código $C_{II} = \{0, 00, 000, 0000\}$ é não-singular, pois cada símbolo é codificado por uma palavra distinta. Este código não é univocamente decodificável, pois uma sequência de bits formada pelo resultado da codificação por um sequência de símbolos pode ser decodificada de mais de

uma maneira. Supondo que os símbolos sejam x_1, x_2, x_3, x_4, respectivamente, a sequência x_1x_2, por exemplo poderia ser decodificada como $x_1x_1x_1$ ou como x_3. Além disso, este código não satisfaz a condição de prefixo pois uma palavra é prefixo de outro (0 é prefixo de 00, 000 e 0000; 00 é prefixo de 000 e 0000; e 000 é prefixo de 0000). Este código satisfaz a desigualdade de Kraft: $\sum_i 2^{-l_i} = 2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} = \frac{15}{16} \leq 1$. É possível obter um código de prefixo com o mesmo comprimento esperado, pois os comprimentos satisfazem a desigualdade de Kraft, e na demonstração do teorema de Kraft vimos um algoritmo para criar tal código.

(c) O código $C_{III} = \{00, 01, 10, 000\}$ é não-singular, pois cada símbolo é codificado por uma palavra distinta. Não é univocamente decodificável. Por exemplo, a sequência $x_1x_1x_1$ será codificada como 000000, que poderá ser decodificada como $x_1x_1x_1$ ou x_4x_4. O código não satisfaz a condição de prefixo: 00 é prefixo de 000. Este código satisfaz a desigualdade de Kraft: $\sum_i 2^{-l_i} = 2^{-2}+2^{-2}+2^{-2}+2^{-3} = \frac{7}{8} \leq 1$. É possível obter um código de prefixo com o mesmo comprimento esperado, pois os comprimentos satisfazem a desigualdade de Kraft, e na demonstração do teorema de Kraft vimos um algoritmo para criar tal código.

(d) O código $C_{IV} = \{00, 01, 10, 11\}$ é não-singular, univocamente decodificável e satisfaz a condição de prefixo. Satisfaz, por conseguinte, a desigualdade de Kraft: $\sum_i 2^{-l_i} = 2^{-2}+2^{-2}+2^{-2}+2^{-2} = 1 \leq 1$.

(e) $C_V = \{0, 10, 110, 111\}$ é não-singular, univocamente decodificável e satisfaz a condição de prefixo. Satisfaz, por conseguinte, a desigualdade de Kraft: $\sum_i 2^{-l_i} = 2^{-1} + 2^{-2} + 2^{-3} + 2^{-3} = 1 \leq 1$.

4.6 Língua Mura

1. Foi encontrada uma nova língua da família linguística mura, falada em uma aldeia de pouco mais de 100 indígenas nas proximidades do rio Madeira, no Amazonas. Esta língua falada por um povo semi-nômade é bem parecida com o pirarrã. A língua é bem simples em seu repertório fonêmico, constituído de apenas 3 vogais e 3 consoantes: /i/, /a/, /u/, /p/, /t/, /k/. Linguistas construíram um corpus para esta língua e verificaram que os fonemas apresentam a seguinte distribuição $p = (6/21, 5/21, 4/21, 3/21, 2/21, 1/21)$. Verificou-se ainda que esta língua possui ainda duas formas escritas fonêmica: *a*) uma forma que utiliza um alfabeto com dois símbolos ($\sqcap$ e $\wedge$); *b*) e outra que utiliza um alfabeto com três símbolos ($\dagger$, $\ddagger$ e $\times$). Ambos códigos utilizados por esta tribo surpreenderam os cientistas ao verificarem que se tratavam de códigos ótimo de prefixo.

   (a) Encontre ambos os códigos utilizados e faça uma tabela para representar cada uma deles.

   (b) Calcule o comprimento esperado de cada um dos códigos.

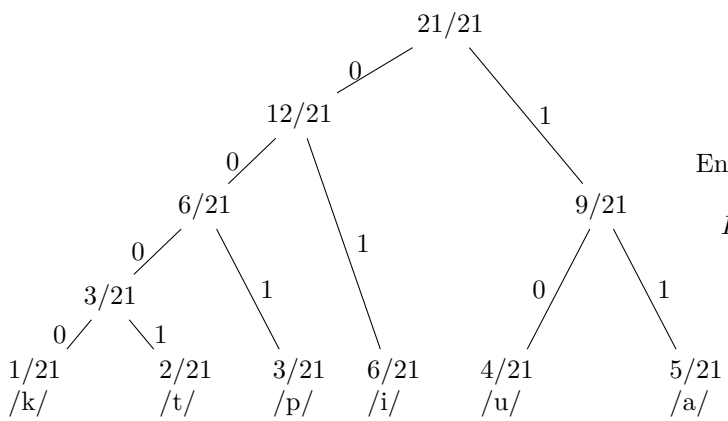
   (c) Seja X uma variável aleatória que representa os fonemas naquela língua, calcule a entropia de X.

   (d) Compare os valores encontrados e explique o que você pode observar através desta comparação. É possível melhorar a eficiência deste código? Explique como.

Resolução:

(a) Se as tribos utilizam códigos ótimo de prefixo, podemos concluir que trata-se de um código de Huffmann. Devemos então achar um código de Huffmann *a*) binário; *b*) e ternário.

x	$C(x)$	$l(x)$
/i/	01	2
/a/	11	2
/u/	10	2
/p/	001	3
/t/	0001	4
/k/	0000	4



Comprimento esperado:

$$L(C) = \sum_x p(x)l(x) = \frac{17}{7} = 2.43. \quad (390)$$

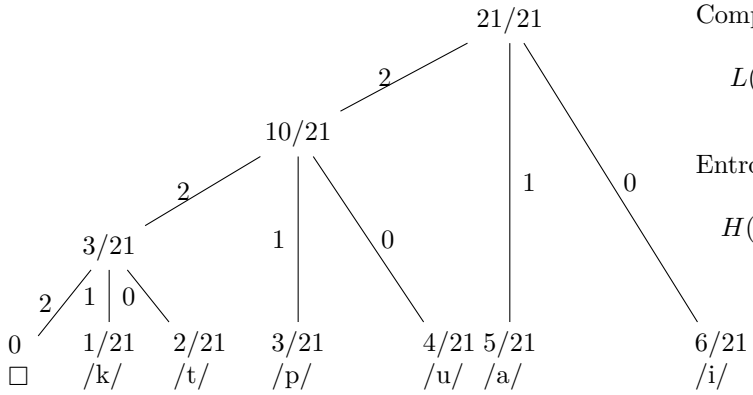
Entropia:

$$H(X) = -\sum_x p(x)\log p(x) = \log 7 - \frac{5}{21}\log 5 + \frac{12}{21}\log 3 - \frac{16}{21} = 2.3983 \text{ bits}. \quad (391)$$

No Octave:

```
pkg load communications
p = [6/21, 5/21, 4/21, 3/21, 2/21, 1/21];
h = huffmandict({'i','a','u','p','t','k'},p);
```

x	$C(x)$	$l(x)$
/i/	0	1
/a/	1	1
/u/	20	2
/p/	21	2
/t/	220	3
/k/	221	3



Comprimento esperado:

$$L(C) = \sum_x p(x)l(x) = 1.619. \quad (392)$$

Entropia:

$$H(X) = -\sum_x p(x)\log p(x) = 2.3983 \text{ bits} = \frac{1}{\log_2 3}H_2(X) \text{ trits} = 1.5132 \text{ trits}. \quad (393)$$

```
p = [1:6]/21;
```

```
l = [3, 3, 2, 2, 1, 1];

function H=entropy(p,b)
if nargin < 2, b=2; endif;
H = (-sum(p.*log2(p)))/(log2(b));
endfunction

>> entropy(p,2)
ans =  2.3983
>> entropy(p,3)
ans =  1.5132
```

Observe que em todos os casos temos $L > H$, respeitando assim o limite de Shannon ($L \geq H$). Apesar de L já ser próximo de H, é possível ainda aproximar mais do limite de Shannon, para tanto é necessário realizar a codificação em blocos ou em fluxo. O código de Huffmann é ótimo, dentre os códigos de símbolos, ou seja, fornece o menor L dentre os códigos para símbolos, desta forma os comprimentos das palavras serão inteiros. O código de Huffmann atingirá exatamente o limite quando a distribuição for d-ádica, o que não é o caso.

4.7 Comprimento esperado (Shannon e Huffman)

1. Suponha uma fonte que produza os símbolos do alfabeto $\mathcal{X} = \{A, B, C, D, E\}$ com distribuição $p = (2/5, 1/5, 1/5, 1/10, 1/10)$. Usando $\log 5 = 2,32$, calcule $H(X)$, L_s e L_H (entropia da fonte, comprimento esperado do código de Shannon e comprimento esperado do código Huffman, respectivamente).

Resolução: A entropia é dada por

$$\begin{aligned} H(X) &= -2/5 \log 2/5 - 2 \times 1/5 \log 1/5 - 2 \times 1/10 \log 1/10 && (394)\\ &= 2/5(\log 5 - 1) + 2/5 \log 5 + 1/5(\log 5 + 1) && (395)\\ &= \log 5(2/5 + 2/5 + 1/5) - 2/5 + 1/5 = \log 5 - 1/5 = 2,32 - 0,2 = 2,12. && (396) \end{aligned}$$

Os comprimentos do código de Shannon são dados por $\lceil -\log p(x) \rceil$. Teremos então os seguintes comprimentos:

$$l(A) = \lceil -\log 2/5 \rceil = \lceil \log 5 - 1 \rceil = \lceil 2,32 - 1 \rceil = \lceil 1,32 \rceil = 2 \quad (397)$$

$$l(B) = l(C) = \lceil -\log 1/5 \rceil = \lceil \log 5 \rceil = \lceil 2,32 \rceil = 3 \quad (398)$$

$$l(D) = l(E) = \lceil -\log 1/10 \rceil = \lceil \log 10 \rceil = \lceil \log 5 + 1 \rceil = \lceil 2,32 + 1 \rceil = \lceil 3,32 \rceil = 4 \quad (399)$$

assim o comprimento esperado do código de Shannon será

$$L_S = 2/5 \times 2 + 2 \times 1/5 \times 3 + 2 \times 1/10 \times 4 = 28/10 = 2,8 \quad (400)$$

O código de Huffman será $(A, (B, (C, (D, E))))$ e assim os comprimentos do código de Huffman serão $l(A) = 1$, $l(B) = 2$, $l(C) = 3$ e $l(D) = l(E) = 4$. O comprimento esperado deste código será

$$L_H = 2/5 \times 1 + 1/5 \times 2 + 1/5 \times 3 + 2 \times 1/10 \times 4 = 22/10 = 2,2 \quad (401)$$

4.8 Código para um alfabeto com 6 símbolos

1. Suponha os seguintes códigos para um alfabeto de 6 símbolos A, B, C, D, E, F.

símbolo	C-I	C-II
A	00	00
B	11	11
C	000	001
D	011	010
E	100	101
F	1011	0110

Faça o que se pede:

(a) Decodifique a seguinte sequencia binária 00011000, utilizando ambos códigos C-I e C-II.

(b) Caracterize os códigos quanto a: singular, univocamente decodificável, prefixo, satisfaz kraft.

(c) É possível obter um código de prefixo com o mesmo comprimento esperado que o código C-II? Justifique. Em caso afirmativo, encontre o código de prefixo.

(d) Utilize o algoritmo de Sardinas-Patterson para verificar quais códigos são decodificáveis.

Resolução:

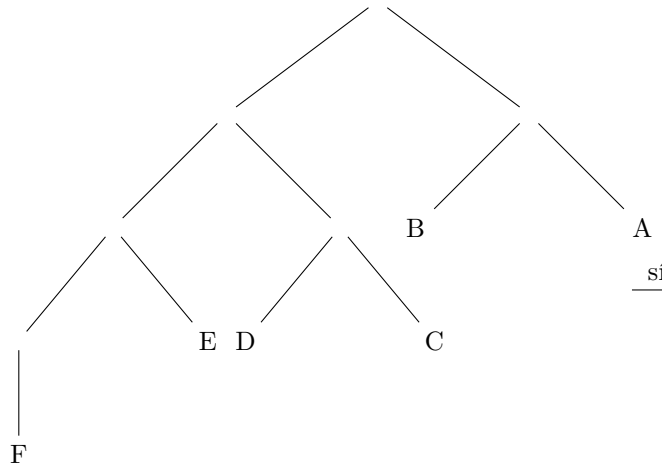
		00011000
(a)	C-I	ADC ou CBC
	C-II	AFA

		singular	univ. dec.	prefixo	kraft
(b)	C-I	não	não	não	sim
	C-II	não	sim	não	sim

(c) Como visto em aula, dados comprimentos que satisfazem Kraft, podemos utilizá-los para construir um código de prefixo. Desta forma obteremos um código com o mesmo comprimento esperado.

Vamos tomar então os comprimentos associados ao código C-II: 2, 2, 3, 3, 3, 4. Para cada um deles, iremos associar um nó na árvore binária na profundidade dada pelo valor do comprimento e remover todos os descendentes.

Um exemplo de árvore e código de prefixo que iremos obter é apresentado abaixo:



símbolo	código	comprimento
A	11	2
B	10	2
C	011	3
D	010	3
E	001	3
F	1111	4

(d) O algoritmo pode ser obtido na internet. Busque uma implementação e faça os testes.

4.9 Estimação e Codificação Aritmética

1. Suponha uma fonte i.i.d. com alfabeto $\mathcal{X} = \{a, b, c\}$. Foi feita uma breve observação de símbolos produzidos por esta fonte e observou-se 5 ocorrências do símbolo 'a' e 2 ocorrências do símbolo 'b'.
   (a) Utilize a regra de Laplace para estimar as probabilidades dos símbolos.
   (b) Explique qual é a razão de utilizarmos esta regra.
   (c) Suponha a seguinte sequência de 2 símbolos $x_{1:2} = ab$. Qual será a codificação binária dessa sequência dada pela codificação aritmética?

Resolução:

(a) Temos que $N_a = 5$, $N_b = 2$ e $N_c = 0$. Usando a regra de Laplace, teremos

$$p(a) = \frac{5+1}{10} = {}^{6}\!/_{10} \tag{402}$$

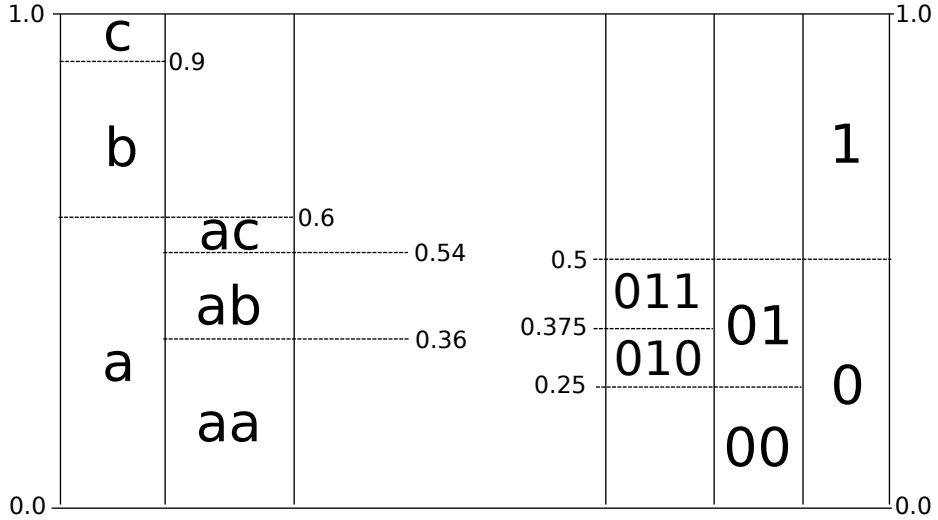
$$p(b) = \frac{2+1}{10} = {}^{3}\!/_{10} \tag{403}$$

$$p(c) = \frac{0+1}{10} = {}^{1}\!/_{10} \tag{404}$$

(b) Para que a estimativa da probabilidade de símbolos não observados não seja nula, iremos adicionar 1 à frequência de ocorrência de todos os símbolos (regra de Laplace). O fato de observamos um evento zero vezes ou uma única vez é apenas um acaso decorrente da amostragem. Não devemos supor que um símbolo tenha probabilidade nula pelo simples fato deste não ter ocorrido nenhuma vez em nossa amostra. A regra de Laplace equivale a supor incerteza sobre os parâmetros da distribuição da fonte de forma que os parâmetros possuam distribuição uniforme, o que levará à

$$p(s|x_{1:n}) = \frac{N(s|x_{1:n}) + 1}{\sum_{s'}(N(s'|x_{1:n}) + 1)} \tag{405}$$

(c)



A sequência será codificada por 011 ou 100, dependendo da convenção.

5 Canal de Comunicação

5.1 Capacidade de Canal

1. Determine a capacidade dos seguintes canais:

   (a) Dois canais binários simétricos.

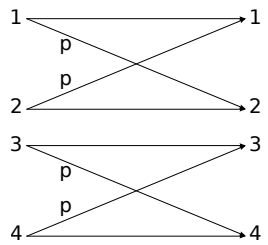


Figura 5: Dois canais binários simétricos.

   (b) Um canal binário simétrico e um único símbolo.

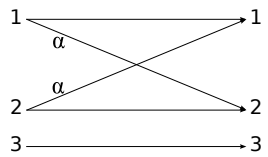


Figura 6: Um canal binário simétrico e um único símbolo.

   (c) Um canal binário simétrico e um canal ternário.

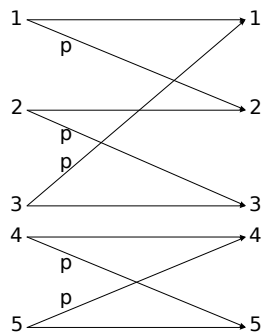


Figura 7: Um canal binário simétrico e um canal ternário.

   (d) Um canal ternário.

$$p(y|x) = \begin{bmatrix} 2/3 & 1/3 & 0 \\ 0 & 1/3 & 2/3 \end{bmatrix} \tag{406}$$

Resolução:

(a) Temos um canal com $\mathcal{X} = \mathcal{Y} = \{1, 2, 3, 4\}$, alfabeto de entrada e saída. A matriz de transição do canal é dada por

$$p(y|x) = \begin{pmatrix} 1-p & p & 0 & 0 \\ p & 1-p & 0 & 0 \\ 0 & 0 & 1-p & p \\ 0 & 0 & p & 1-p \end{pmatrix} \tag{407}$$

O canal em questão é simétrico, assim a capacidade poderá ser calculada da seguinte forma

$$\begin{aligned} C &= \log|\mathcal{Y}| - H(r) \\ &\quad \text{onde } r \text{ é uma linha da matriz de transição} \\ &= \log 4 - H(p, 1-p, 0, 0) \\ &= 2 - H(p). \end{aligned} \tag{408}$$

No exercício sobre o Canal da Soma, foi visto que a capacidade do canal poderá ser dada por

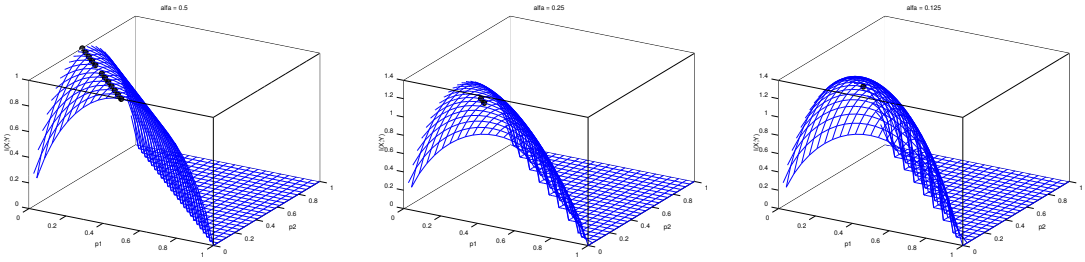
$$C = \log\left(2^{C_1} + 2^{C_2}\right), \tag{409}$$

onde C_1 e C_2 são as capacidades dos dois canais em paralelo que formam o canal original com capacidade C. No exercício em questão, temos $C_1 = C_2 = C'$ e assim

$$\begin{aligned} C &= \log\left(2^{C'+1}\right) \\ &= C' + 1 \\ &= 2 - H(p) \end{aligned} \tag{410}$$

onde $C' = 1 - H(p)$, a capacidade de uma canal binário simétrico.

(b)




```
function H = entropy(p)
H = - sum( p.*log2(p) );
endfunction

a = 0.5;
p = linspace(0,1,25);
I = []; for i=1:length(p), for j=1:length(p),
        if(1-p(i)-p(j)>0),
                I(i,j) = entropy( ...
                [p(i)*(1-a)+p(j)*a, p(i)*a+p(j)*(1-a), 1-p(i)-p(j)]) -
                    ↪  ...
                entropy([a,1-a])*(p(i)+p(j));
        else,
                I(i,j) = 0;
        endif;
endfor; endfor;
figure; hold on;
mesh(p,p,I,'facecolor', 'none', 'edgecolor', 'b');
xlabel('p1'); ylabel('p2'); zlabel('I(X;Y)');
```

```
title(cstrcat('alfa = ',num2str(a)));
[id,jd] = find(I == max(max(I)));
for i=1:length(id),
        plot3(p(id(i)), p(jd(i)), I(id(i),jd(i)),'ok', ...
                'markerfacecolor','k','markersize',8);
endfor;
view(30, 30);
C = 0;
for k=1:length(id),
        printf('p = [%f, %f, %f]\n',p(id(k)),p(jd(k)),1-p(id(k))-p(jd(
            ↪ k)));
        C = max(C, I(id(k),jd(k)));
endfor;
printf('C = %f bits\n',C);
```

Temos que $C = \max_{p(x)} I(X;Y)$, onde X é a entrada do canal e Y a saída, onde $X \sim p$ e $Y \sim q$.

$$\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= H(Y) - \sum_x p(x) H(Y|X=x) \\
&= H(Y) - \left(p(1) \underbrace{H(Y|X=1)}_{H(\alpha)} + p(2) \underbrace{H(Y|X=2)}_{H(\alpha)} + p(3) \underbrace{H(Y|X=3)}_{0} \right) \\
&= H(Y) - H(\alpha)\,(p_1 + p_2)
\end{aligned} \tag{411}$$

A capacidade do canal será atingida quando a distribuição p for tal que maximize a informação mútua $I(X;Y)$. Como evidenciado na resolução numérica, este problema não terá necessariamente uma solução única. Temos que

$$\begin{aligned}
q_1 &= \Pr(y=1) = \Pr(x=1)(1-\alpha) + \Pr(x=2)\alpha \\
q_2 &= \Pr(y=2) = \Pr(x=1)\alpha + \Pr(x=2)(1-\alpha) \\
q_3 &= \Pr(y=3) = \Pr(x=3)
\end{aligned} \tag{412}$$

Fazendo $p_3 = 1 - p_1 - p_2$, teremos que a Equação 411, para α fixo, é função de p_1 e p_2. Para encontrar seu máximo, devemos encontrar o(s) ponto(s) em que as derivadas parciais em relação a p_1 e p_2 são nulas.

Entretanto, para resolver este problema é mais fácil considerar este canal como dois canais em paralelo: um canal binário simétrico com capacidade $C_1 = 1 - H(\alpha)$; e outro canal com capacidade $C_2 = 0$, já que para este canal unário, temos $I(X;Y) = H(X) - H(X|Y) = 0 - 0 = 0$. Utilizando agora o resultado já visto para o canal da soma, teremos

$$\begin{aligned}
C &= \log\left(2^{C_1} + 2^{C_2}\right) \\
&= \log\left(2^{1-H(\alpha)} + 1\right)
\end{aligned} \tag{413}$$

(c) Como já visto anteriormente, podemos ver este canal como um canal da soma de dois outros canais. Teremos assim

$$C = \log\left(2^{C_1} + 2^{C_2}\right), \tag{414}$$

onde C_1 é a capacidade do canal superior, que pode ser visto como uma máquina de escrever com ruído, e C_2 é o canal binário simétrico.

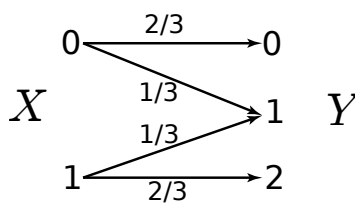
$$\begin{aligned}
C_1 &= \max_{p(x)} I(X;Y) \\
&= \max_{p(x)} \left(H(Y) - \underbrace{H(Y|X)}_{\sum_x p(x) H(Y|X=x) = H(p)} \right) \\
&= \max_{p(x)} \left(H(Y) - H(p) \right) \\
&= \log 3 - H(p) \qquad (415)
\end{aligned}$$

$$C_2 = 1 - H(p) \qquad (416)$$

Podemos agora determinar C.

$$\begin{aligned}
C &= \log \left(2^{C_1} + 2^{C_2} \right) \\
&= \log \left(2^{\log 3 - H(p)} + 2^{1-H(p)} \right) \\
&= \log \left(2^{-H(p)} \left(2 + 2^{\log 3} \right) \right) \\
&= \log \left(2^{-H(p)} \times 5 \right) \\
&= \log 5 - H(p) \qquad (417)
\end{aligned}$$

(d) Este canal é ilustrado abaixo



Este canal é um canal binário com apagamento, que já foi visto anteriormente. A capacidade deste canal será $C = 1 - \frac{1}{3} = \frac{2}{3}$.

5.2 O estatístico e o canal

1. É dado um canal de comunicação com as probabilidades de transição $p(y|x)$ e a capacidade de canal $C = \max_{p(x)} I(X;Y)$. Um estatístico resolveu ajudar e propôs processar a saída através de uma determinada função $\tilde{Y} = g(Y)$. Ele alega que isto irá melhorar a capacidade do canal estritamente.
   (a) Mostre que ele está errado.
   (b) Sob quais condições ele não irá estritamente diminuir a capacidade?

Resolução:

(a) Ao adotar $\tilde{Y} = g(Y)$, teremos $X \to Y \to \tilde{Y}$, formando uma cadeia de Markov. Poderemos assim aplicar a desigualdade de processamento de dados

$$I(X;Y) \geq I(X;\tilde{Y}). \tag{418}$$

Seja $\tilde{p}(x)$ a distribuição em x que maximiza $I(X;\tilde{Y})$, fornecendo a capacidade do canal $\tilde{C}$ entre X e $\tilde{Y}$, teremos:

$$\begin{aligned} C &= \max_{p(x)} I(X;Y) && (419)\\ &\geq I(X;Y)_{p(x)=\tilde{p}(x)} (\text{onde } \tilde{p} \text{ é a dist. que max. } I(X;\tilde{Y})) && (420)\\ &\geq I(X;\tilde{Y})_{p(x)=\tilde{p}(x)} \text{ (utilizando a desigualdade de proc. de dados)} && (421)\\ &= \max_{p(x)} I(X;\tilde{Y}) = \tilde{C}. && (422) \end{aligned}$$

Logo, qualquer processamento subsequente sobre Y que for realizado não irá aumentar a capacidade do canal.

(b) Para que não ocorra uma diminuição da capacidade de canal, deveremos ter igualdades na sequência de desigualdades acima. Para isso será necessário ter igualdade na desigualdade de processamento de dados, ou seja, $I(X;Y) = I(X;\tilde{Y})$. Isto ocorrerá quando, dada a distribuição $\tilde{p}(x)$ que maximiza $I(X;\tilde{Y})$, tivermos a seguinte cadeia de Markov $X \to \tilde{Y} \to Y$.

5.3 Soma módulo

1. Considere o canal discreto sem memória $Y = (X + Z) \mod 11$, onde

$$Z = \begin{pmatrix} 1, & 2, & 3 \\ 1/3, & 1/3, & 1/3 \end{pmatrix} \tag{423}$$

e $X \in \{0, 1, \ldots, 10\}$. Considere Z independente de X.

(a) Calcule a capacidade do canal.

(b) Qual é $p^*(x)$ que maximiza?

Resolução:

(a) Sabemos que $I(X;Y) = H(Y) - H(Y|X)$. E a capacidade de canal é $C = \max_{p(x)} I(X;Y)$. Temos que

$$H(Y|X) = H(Z|X) = H(Z) = \log 3. \tag{424}$$

Assim,

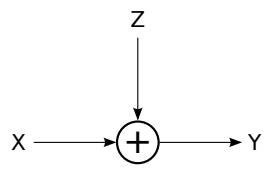
$$\begin{aligned} C &= \max_{p(x)} I(X;Y) \\ &= \max_{p(x)} H(Y) - \log 3 \\ &= \log 11 - \log 3 = \log \frac{11}{3} && (425) \end{aligned}$$

onde utilizamos que a informação mútua será máxima quando Y possuir distribuição uniforme e, consequentemente, X também terá distribuição uniforme, uma vez que $Y = (X+Z) \mod 11$ e Z possui distribuição uniforme.

(b) Conforme dado acima, a distribuição uniforme é aquela que maximiza a informação mútua entra X e Y, $p^*(x) = \left(\frac{1}{11}, \frac{1}{11}, \ldots, \frac{1}{11}\right)$.

5.4 Canal soma ruído

1. Encontre a capacidade de canal do seguinte canal discreto sem memória: onde $Pr\{Z = 0\} = Pr\{Z =$



$a\} = \frac{1}{2}$. Temos que o alfabeto de x é $\mathcal{X} = \{0, 1\}$ e o alfabeto de z é $\mathcal{Z} = \{0, a\}$. Assuma que Z seja independente de X. Observe que a capacidade de canal depende do valor a.

Resolução:

$$\begin{aligned} C &= \max_{p(x)} I(X;Y) \\ &= \max_{p(x)} H(Y) - H(Y|X) \\ &= \max_{p(x)} H(Y) - H(Z) \end{aligned} \tag{426}$$

onde utilizamos que

$$H(Y|X) = H(Z|X) = H(Z) = \begin{cases} 1, & a \neq 0 \\ 0, & a = 0 \end{cases} . \tag{427}$$

Podemos também utilizar

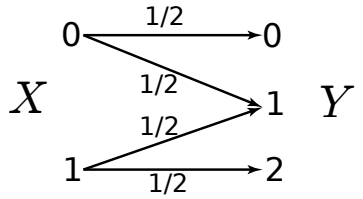
$$C = \max_{p(x)} I(X;Y) \tag{428}$$

$$= \max_{p(x)} H(X) - H(X|Y) \tag{429}$$

Será necessário agora determinar o alfabeto de Y. Note que $\mathcal{Y}$ dependerá do valor de a.

para $a = 0$: neste caso, $\mathcal{Y} = \{0, 1\}$. Neste caso, $Y = X$ e, por conseguinte, $H(Y) = H(X)$. Sabemos que $H(X) \leq 1$, logo teremos $C = 1$ bits por utilização do canal.

para $a = 1$: neste caso, $\mathcal{Y} = \{0, 1, 2\}$. O canal se comporta como um canal binário com apagamento.



Como visto em aula, a capacidade deste canal será $C = 1 - \alpha$, onde $\alpha = 1/2$, logo $C = 1/2$ bit por transmissão.

para $a = -1$: neste caso, $\mathcal{Y} = \{-1, 0, 1\}$. Caso similar ao caso em que $a = 1$. Teremos novamente $C = 1/2$ bit por transmissão.

para $a \neq 0, \pm 1$: neste caso, $\mathcal{Y} = \{0, 1, a, 1+a\}$. Neste caso, conhecendo Y sabemos quem foi o X enviado, assim $H(X|Y) = 0$.

$$\begin{aligned} C &= \max_{p(x)} I(X;Y) \\ &= \max_{p(x)} H(X) - \underbrace{H(X|Y)}_{=0} = \max_{p(x)} H(X) = 1 \end{aligned} \tag{430}$$

A capacidade será atingida quando $X \sim$ uniforme.

5.5 Máquina de escrever

1. Considere uma máquina de escrever com 26 teclas.
   (a) Se ao pressionar uma tecla temos como resultado a impressão da letra associada, qual é a capacidade C em bits?
   (b) Agora suponha que ao pressionar uma tecla podemos ter como resultado a impressão da letra associada ou da letra vizinha (com igual probabilidade). Desta forma $A \to A$ ou B, ..., $Z \to Z$ ou A. Qual é a capacidade agora?
   (c) Qual é o código de maior taxa, com blocos de comprimento unitário, que você consegue encontrar que alcança probabilidade zero de erro para o canal do item anterior?

Resolução:

(a)

$$\begin{aligned} C &= \max_{p(x)} I(X;Y) \\ &= \max_{p(x)} H(X) - \underbrace{H(X|Y)}_{=0} = \max_{p(x)} H(X) = \log 26 = 1 + \log 13 \end{aligned} \tag{431}$$

(b)

$$\begin{aligned} C &= \max_{p(x)} I(X;Y) \\ &= \max_{p(x)} H(X) - \underbrace{H(X|Y)}_{=H(\frac{1}{2})} = \max_{p(x)} H(X) - 1 = \log 26 - 1 = \log 13 \end{aligned} \tag{432}$$

(c) Podemos utilizar um código de blocos de comprimento unitário utilizando as letras alternadas, isto é, A, C, E, G, etc. Desta forma não haverá confusão e a taxa deste código será

$$R = \frac{\log(\text{num. de palavras})}{\text{tamanho do bloco}} = \frac{\log 13}{1} = \log 13. \tag{433}$$

Desta forma, iremos alcançar a capacidade do canal.

5.6 Capacidade dada matriz

1. Mostre que a capacidade do canal com probabilidade de transmissão dada pela matriz

$$P_{y|x} = \begin{bmatrix} 2/3 & 1/3 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/3 & 2/3 \end{bmatrix} \tag{434}$$

é alcançada por uma distribuição que coloca probabilidade zero em um dos símbolos de entrada. Qual é a capacidade deste canal? Por qual razão esta letra não é utilizada?

Resolução:

Seja $X \sim p = (p_1, p_2, p_3)$. As probabilidades dos 3 símbolos de saída serão dadas por $q = (q_1, q_2, q_3)$, onde $q_1 = \frac{2}{3}p_1 + \frac{1}{3}p_2$, $q_2 = \frac{1}{3}$ e $q_3 = \frac{1}{3}p_2 + \frac{2}{3}p_3$. Desta forma, teremos

$$\begin{aligned} I(X;Y) &= \underbrace{H(Y)}_{H\left(\frac{2}{3}p_1+\frac{1}{3}p_2, \frac{1}{3}, \frac{1}{3}p_2+\frac{2}{3}p_3\right)} - \underbrace{H(Y|X)}_{\sum_x p(x)H(Y|X=x)} \\ &= H\left(\frac{2}{3}p_1 + \frac{1}{3}p_2, \frac{1}{3}, \frac{1}{3}p_2 + \frac{2}{3}p_3\right) - p_1 H\left(\frac{2}{3}, \frac{1}{3}\right) - p_2 \log 3 - p_3 H\left(\frac{2}{3}, \frac{1}{3}\right) \\ &\quad \text{utilizando que } p_2 = 1 - p_1 - p_3 \\ &= H\left(\frac{1}{3} + \frac{1}{3}(p_1 - p_3), \frac{1}{3}, \frac{1}{3} - \frac{1}{3}(p_1 - p_3)\right) - (p_1 + p_3) H\left(\frac{2}{3}, \frac{1}{3}\right) - (1 - p_1 - p_3)\log 3 \end{aligned} \tag{435}$$

Se fixarmos $p_1 + p_3$, o segundo e terceiro termo da equação acima ficarão fixos, assim a informação mútua será maximizada quando o primeiro termo for maximizado, ou seja, quando $p_1 = p_3$. Neste caso, teremos

$$\begin{aligned} I(X;Y) &= \underbrace{H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)}_{=\log 3} - (p_1 + p_3) \underbrace{H\left(\frac{2}{3}, \frac{1}{3}\right)}_{=\log 3 - \frac{2}{3}} - (1 - p_1 - p_3)\log 3 \\ &= \log 3 - (p_1 + p_3)\left(\log 3 - \frac{2}{3}\right) - (1 - p_1 - p_3)\log 3 \\ &= (p_1 + p_3)\left(\log 3 - \log 3 + \frac{2}{3}\right) = (p_1 + p_3)\frac{2}{3} \end{aligned} \tag{436}$$

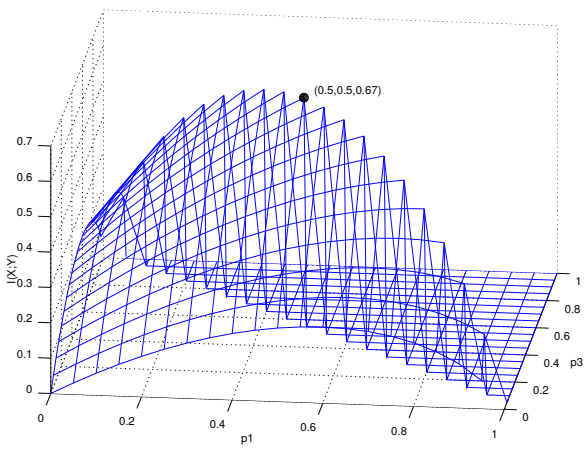
A informação mútua será máxima quando maximizarmos $p_1 + p_3$, sujeito a $p_1 = p_3$, ou seja, quando $p_1 = p_3 = \frac{1}{2}$ e por conseguinte $p_2 = 0$. Teremos assim $C = \frac{2}{3}$, que será atingida com a distribuição $p = (\frac{1}{2}, 0, \frac{1}{2})$.

```
function H=entropy(p)
H = (-sum(p.*log2(p)));
endfunction

p1=linspace(0,1,21);
p3=linspace(0,1,21);
I=[];
for i=1:length(p1), for j=1:length(p3),
  if(p1(i)+p3(j)<=1),
    I(i,j) = entropy( [1/3+1/3*(p1(i)-p3(j)), 1/3, 1/3 - 1/3*(p1(i)-p3(j)
        ↪ ) ] ) - ...
```

```
        (p1(i)+p3(j))*entropy([2/3,1/3]) - (1 - p1(i) - p3(j)) * log2(3);
  endif;
endfor; endfor;

[i,j] = find(I==(max(max(I))));
figure; mesh(p1,p3,I,'facecolor', 'none', 'edgecolor', 'b');
xlabel('p1'); ylabel('p3'); zlabel('I(X;Y)');
hold on; plot3(p1(i),p3(j),I(i,j),'ok','markerfacecolor','k','markersize'
    ↪ ,8);
text(p1(i)+0.02,p3(j)+0.02,I(i,j)+0.02, ...
   cstrcat('(',num2str(p1(i)),',',num2str(p3(j)),',',num2str(I(i,j),'%.2f
       ↪ '),')'));
print figure1.pdf
```



5.7 Capacidade dada matriz 2

1. Determine a capacidade do canal descrito pelo matriz de transição abaixo.

$$P_{y|x} = \begin{bmatrix} p & 0 & 1-p & 0 \\ 0 & q & 0 & 1-q \\ 1-p & 0 & p & 0 \\ 0 & 1-q & 0 & q \end{bmatrix} \tag{437}$$

Resolução: Note que este canal de comunicação é equivalente ao canal dado abaixo onde trocamos os papeis dos símbolos 2 e 3:

$$P_{y|x} = \begin{array}{c} \begin{array}{cccc} 1 & 3 & 2 & 4 \end{array} \\ \begin{pmatrix} p & 1-p & 0 & 0 \\ 1-p & p & 0 & 0 \\ 0 & 0 & q & 1-q \\ 0 & 0 & 1-q & q \end{pmatrix} \end{array} \begin{array}{c} \\ 1 \\ 3 \\ 2 \\ 4 \end{array} \tag{438}$$

Este canal, como já vimos anteriormente, é constituído pela soma de dois canais binários simétricos,

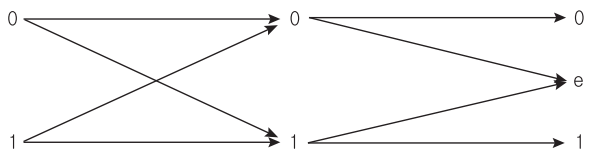
com capacidade $C_1 = 1 - H(p)$ e $C_2 = 1 - H(q)$. Combinados, estes canais terão capacidade

$$\begin{aligned} C &= \log \sum_{i=1}^{2} \left(2^{C_i}\right) \\ &= \log \left(2^{C_1} + 2^{C_2}\right) \\ &= \log \left(2^{1-H(p)} + 2^{1-H(q)}\right) \end{aligned} \tag{439}$$

5.8 Canal BSC em cascata com apagamento

1. Suponha que um canal binário simétrico de capacidade C_1 seja imediatamente seguido de um canal binário com apagamento e capacidade C_2. Determine a capacidade C do canal resultante.

Resolução:



Seja $C_1 = 1 - H(p)$ a capacidade do canal binário simétrico, com parâmetro p, e $C_2 = 1 - \alpha$ a capacidade do canal binário com apagamento, com parâmetro α. Seja $\tilde{Y}$ a saída do canal final (cascata dos dois canais), e seja Y a saída do canal binário simétrico. A regra de transição para o canal final em cascata é dado por

$$p(\tilde{y}|x) = \sum_{y=0,1} p(\tilde{y}|y)p(y|x) \tag{440}$$

para cada par $(x, \tilde{y})$.

Temos então a seguinte matriz de transição

$$p(\tilde{y}|x) = \begin{array}{c} \\ \\ \end{array} \begin{array}{c} \begin{array}{ccc} 0 & e & 1 \end{array} \\ \begin{pmatrix} (1-p)(1-\alpha) & \alpha & p(1-\alpha) \\ p(1-\alpha) & \alpha & (1-p)(1-\alpha) \end{pmatrix} \end{array} \begin{array}{c} \\ 0 \\ 1 \end{array} \tag{441}$$

Note que as linhas desta matriz são permutação uma das outra, mas a soma de cada coluna é diferente. Por tanto, esta matriz não é fracamente simétrica[4].

Seja $X \sim \text{Bern}(\pi)$ a entrada do canal, então

$$\begin{aligned} C &= \max_{p(x)} I(X; \tilde{Y}) \\ &= \max_{\pi} \left(H(\tilde{Y}) - H(\tilde{Y}|X)\right) \\ &= \max_{\pi} \left(H(\tilde{Y})\right) - H(\tilde{Y}|X) \end{aligned} \tag{442}$$

Dada a matriz de transição, podemos calcular $H(\tilde{Y}|X)$.

$$\begin{aligned}
H(\tilde{Y}|X) &= H\left((1-p)(1-\alpha), \alpha, p(1-\alpha)\right) \\
&= (1-p)(1-\alpha)\log\left((1-p)(1-\alpha)\right) - \alpha\log\alpha - p(1-\alpha)\log\left(p(1-\alpha)\right) \\
&= (1-\alpha)\left((1-p)\log(1-p) - p\log p\right) - (1-\alpha)\log(1-\alpha)\underbrace{\left((1-p)+p\right)}_{=1} - \alpha\log\alpha \\
&= (1-\alpha)H(p) + H(\alpha) \qquad (443)
\end{aligned}$$

Ainda precisamos analisar $H(\tilde{Y})$ para calcular C. Precisamos então determinar $p(\tilde{y})$.

$$\begin{aligned}
p(\tilde{y}=0) &= p(\tilde{y}=0, x=0) + p(\tilde{y}=0, x=1) \\
&= p(\tilde{y}=0|x=0)p(x=0) + p(\tilde{y}=0|x=1)p(x=1) \\
&= (1-p)(1-\alpha)\pi + p(1-\alpha)(1-\pi) \\
&= (1-\alpha)\left((1-p)\pi + p(1-\pi)\right) \qquad (444) \\
p(\tilde{y}=e) &= p(\tilde{y}=e, x=0) + p(\tilde{y}=e, x=1) \\
&= p(\tilde{y}=e|x=0)p(x=0) + p(\tilde{y}=e|x=1)p(x=1) \\
&= \alpha\pi + \alpha(1-\pi) = \alpha \qquad (445) \\
p(\tilde{y}=1) &= p(\tilde{y}=1, x=0) + p(\tilde{y}=1, x=1) \\
&= p(\tilde{y}=1|x=0)p(x=0) + p(\tilde{y}=1|x=1)p(x=1) \\
&= p(1-\alpha)\pi + (1-p)(1-\alpha)(1-\pi) \\
&= (1-\alpha)\left(p\pi + (1-p)(1-\pi)\right) \qquad (446)
\end{aligned}$$

e assim teremos que

$$H(\tilde{Y}) = H\left((1-\alpha)(\pi(1-p) + p(1-\pi)), \alpha, (1-\alpha)(p\pi + (1-p)(1-\pi))\right). \qquad (447)$$

Para maximizar $H(\tilde{Y})$, basta fazer $p(\tilde{y}=0) = p(\tilde{y}=1)$, uma vez que $p(\tilde{y}=e)$ depende apenas de α.

$$\begin{aligned}
(1-\alpha)\left((1-p)\pi + p(1-\pi)\right) &= (1-\alpha)\left(p\pi + (1-p)(1-\pi)\right) \\
\pi - p\pi + p - p\pi &= p\pi + 1 - \pi - p + p\pi \\
2\pi(1-2p) &= 1-2p \\
\pi &= \frac{1}{2} \qquad (448)
\end{aligned}$$

Podemos então calcular o valor máximo de $H(\tilde{Y})$.

$$\begin{aligned}
H(\tilde{Y}) &= H\left((1-\alpha)(\frac{1}{2} - \frac{1}{2}p + \frac{1}{2}p), \alpha, (1-\alpha)(\frac{1}{2}p + \frac{1}{2} - \frac{1}{2}p)\right) \\
&= H\left((1-\alpha)\frac{1}{2}, \alpha, (1-\alpha)\frac{1}{2}\right) \\
&= -2\frac{1}{2}(1-\alpha)\log\left(\frac{1}{2}(1-\alpha)\right) - \alpha\log\alpha \\
&= -(1-\alpha)\log(1-\alpha) + (1-\alpha)\log 2 - \alpha\log\alpha \\
&= (1-\alpha) + H(\alpha) \qquad (449)
\end{aligned}$$

A capacidade do canal então será

$$\begin{aligned} C &= \max_{\pi} \Big(H(\tilde{Y}) \Big) - H(\tilde{Y}|X) \\ &= (1-\alpha) + H(\alpha) - ((1-\alpha)H(p) + H(\alpha)) \\ &= \underbrace{(1-\alpha)}_{C_2} \underbrace{(1-H(p))}_{C_1} \\ &= C_1 C_2 \end{aligned} \tag{450}$$

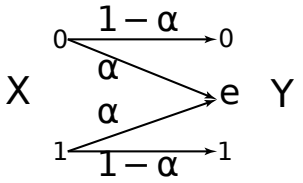
5.9 Canal de pombos

1. Suponha que você seja o comandante de um exército sitiado em um forte e a única forma de comunicação possível com seus aliados é através do envio de pombos correio carregando mensagens. Considere que cada pombo seja capaz de levar apenas um símbolo de 8 bits. Os pombos são enviados um a cada 5 minutos e levam 3 minutos para voar até seu destino.
   (a) Supondo que todos os pombos chegam ao seu destino com segurança, qual é a capacidade deste link estabelecido em bits/horas?
   (b) Suponha agora que o inimigo tente abater os pombos e que uma fração de α pombos serão efetivamente abatidos. Como os pombos são enviados a uma taxa constante, o seu aliado saberá quando um pombo foi abatido. Qual será agora a capacidade deste link?
   (c) Suponha agora que o inimigo seja mais astuto e envie um pombo com mensagem falsa toda vez que abater um pombo. A mensagem falsa será um símbolo de 8 bits escolhido aleatoriamente com distribuição uniforme. Qual será a capacidade deste link em bits/hora? Calcule o valor para $\alpha = 1/2$.

Resolução:

(a) A cada 60 minutos teremos 12 pombos que serão enviados e chegarão ao destino, totalizando assim o envio de $12 \times 8 = 96$ bits/hora.

(b) Este canal funciona como um canal binário com apagamento.



A capacidade deste canal é dada por $C = 1-\alpha$. Como estamos enviando 8 bits simultâneamente, a capacidade do canal de pombos será $8 \times (1-\alpha)$. São enviados 12 pombos por hora, logo teremos que a capacidade do canal por hora será de $12 \times 8 \times (1-\alpha) = 96(1-\alpha)$ bits por hora.

(c) Sabemos que existe uma probabilidade de α de que o pombo seja abatido, logo existe uma probabilidade de $(1-\alpha)$ de que o pombo chegue sem ser abatido. Quando um pombo é abatido será enviado um outro em seu lugar. A probabilidade deste novo pombo carregar a mesma mensagem que aquele abatido é de $1/256$. Concluímos assim que existe uma probabilidade de $(1-\alpha) + \alpha/256$ de que a mensagem seja recebida sem erro. Cada caso de erro ocorrerá com probabilidade $\alpha/256$. A matriz de transição do canal (de dimensões 256×256) será então:

$$p(y|x) = \begin{pmatrix} (1-\alpha)+\alpha/256 & \alpha/256 & \dots & \alpha/256 \\ \alpha/256 & (1-\alpha)+\alpha/256 & \dots & \alpha/256 \\ \vdots & & \ddots & \vdots \\ \alpha/256 & \alpha/256 & \dots & (1-\alpha)+\alpha/256 \end{pmatrix} \tag{451}$$

Temos um canal simétrico. A sua capacidade é dada por

$$C = \log|\mathcal{Y}| - H(r) \tag{452}$$

onde temos $|\mathcal{Y}| = 2^8 = 256$ e $r = ((1-\alpha)+\alpha/256, \alpha/256, \dots, \alpha/256)$.
Teremos assim

$$C = 8 - H\left((1-\alpha)+\alpha/256, \alpha/256, \dots, \alpha/256\right). \tag{453}$$

Como o canal é utilizado 12 vezes por hora, teremos

$$C = 12 \times \left(8 - H\left((1-\alpha)+\alpha/256, \alpha/256, \dots, \alpha/256\right)\right). \tag{454}$$

Para $\alpha = 1/2$ teremos

$$\begin{aligned} C &= 8 - H\left(1/2 + 1/512, 1/512, \dots, 1/512\right) \\ &= 8 + (257/512)\log(257/512) + (255/512)\log(1/512) \\ &= 8 + (257/512)(\log 257 - 9) + (255/512)(-9) \\ &= 3.0184 \text{ bits}, \end{aligned} \tag{455}$$

e assim

$$C = 12 \times 3.0184 = 36.221 \text{ bits/hora}. \tag{456}$$

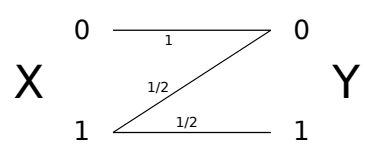
5.10 Canal Z

1. Considere o canal Z com entrada e saída binária e probabilidades de transição dadas pela matriz

$$Q = \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix} \qquad x, y \in \{0, 1\}. \tag{457}$$

Determine a capacidade deste canal e a distribuição maximizadora na entrada.

Resolução:



$$I(X;Y) = H(Y) - H(Y|X) \tag{458}$$

$$H(Y|X) = \sum_x p(x)H(Y|X=x) = \Pr(X=0)\times 0 + \Pr(X=1)\times 1 = p \tag{459}$$

$$H(Y) = H\left(\Pr(Y=1)\right) = H\left(\frac{1}{2}Pr(X=1)\right) = H\left(\frac{p}{2}\right) \tag{460}$$

Teremos assim:

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) = H\left(\frac{p}{2}\right) - p \\ &= -\frac{p}{2}\log\frac{p}{2} - \left(1-\frac{p}{2}\right)\log\left(1-\frac{p}{2}\right) - p \end{aligned} \tag{461}$$

Queremos $\frac{\partial I}{\partial p} = 0$, ou seja,

$$\begin{aligned} \frac{\partial I}{\partial p} &= -\frac{1}{2}\log\frac{p}{2} - \frac{p}{2}\frac{2}{p}\frac{1}{2} + \frac{1}{2}\log\left(1-\frac{p}{2}\right) - \left(1-\frac{p}{2}\right)\frac{1}{1-\frac{p}{2}}\left(-\frac{1}{2}\right) - 1 \\ &= \frac{1}{2}\log\frac{1-\frac{p}{2}}{\frac{p}{2}} - 1 \end{aligned} \tag{462}$$

Fazendo $\frac{\partial I}{\partial p} = 0$, teremos

$$\begin{aligned} \frac{1}{2}\log\frac{1-\frac{p}{2}}{\frac{p}{2}} - 1 &= 0 \\ \log\frac{1-\frac{p}{2}}{\frac{p}{2}} &= 2 \\ \frac{1-\frac{p}{2}}{\frac{p}{2}} &= 4 \\ 1-\frac{p}{2} &= 2p \\ p &= \frac{2}{5} \end{aligned} \tag{463}$$

Teremos então

$$\begin{aligned} C &= \max_p I(X;Y) = H\left(\frac{1}{5}\right) - \frac{2}{5} \\ &= \frac{1}{5}\log 5 - \frac{4}{5}\log\frac{4}{5} - \frac{2}{5} = \frac{1}{5}\log 5 + \frac{4}{5}\log 5 - \frac{4}{5}\times 2 - \frac{2}{5} \\ &= \log 5 - 2 \end{aligned} \tag{464}$$

A distribuição que maximiza a informação mútua é $\left(\frac{3}{5}, \frac{2}{5}\right)$.

5.11 Canal de Comunicação

1. Suponha o canal de comunicação discreto sem memória ilustrado na Figura 9.

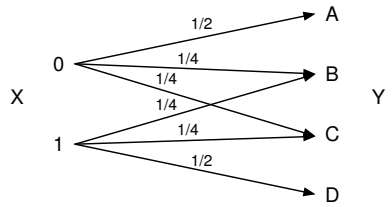


Figura 8: Canal de comunicação discreto sem memória.

(a) Qual á a capacidade deste canal e qual é a distribuição em X que alcança esta capacidade?

(b) Assuma que $\Pr(X = 0) = 1/6$, $\Pr(X = 1) = 5/6$ e que a fonte não possui memória. Encontre o código ótimo para comprimir a saída em Y. Qual é a comprimento médio das palavras do código? Calcule a entropia de Y e compare com o comprimento médio encontrado para o código ótimo.

Resolução:

(a) Para este canal teremos a seguinte matriz de transmissão:

$$p(y|x) = \begin{pmatrix} 1/2 & 1/4 & 1/4 & 0 \\ 0 & 1/4 & 1/4 & 1/2 \end{pmatrix} \tag{465}$$

Podemos verificar que as linhas são permutação uma da outra e a soma das colunas desta matriz é sempre 1/2, logo trata-se de um canal simétrico fraco e assim poderemos calcular sua capacidade através da equação

$$C = \log |\mathcal{Y}| - H(r) \tag{466}$$

onde r é uma linha da matriz de transmissão. Teremos assim

$$C = \log 4 - H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0\right) = 2 - \frac{3}{2} = \frac{1}{2} \text{ bit.} \tag{467}$$

Como o canal é simétrico fraco, teremos que

$$\begin{aligned} H(Y|X) &= \sum_x p(x) H(Y|X = x) \\ &= \sum_x p(x) H(r) \\ &= H(r) \end{aligned} \tag{468}$$

e desta forma, não depende da distribuição da fonte $p(x)$. Para encontrar a distribuição $p(x)$ que maximiza $I(X;Y)$, devemos então encontrar $p(x)$ que maximiza $H(Y)$, pois teremos

$$\begin{aligned} C &= \max_{p(x)} I(X;Y) \\ &= \max_{p(x)} H(Y) - H(Y|X) \\ &= \max_{p(x)} H(Y) - H(r). \end{aligned} \tag{469}$$

Como temos

$$\begin{aligned} \Pr(Y = A) &= \frac{1}{2} p_0 \\ \Pr(Y = B) &= \frac{1}{4} p_0 + \frac{1}{4} p_1 \\ \Pr(Y = C) &= \frac{1}{4} p_0 + \frac{1}{4} p_1 \\ \Pr(Y = D) &= \frac{1}{2} p_1, \end{aligned} \tag{470}$$

$H(Y)$ será maximizado quando $p_0 = p_1 = 1/2$.

Utilizaremos aqui que

$$\begin{aligned} C &= \max_{p(x)} I(X;Y) \\ &= \max_{p(x)} H(Y) - H(Y|X). \end{aligned} \tag{471}$$

Onde temos que

$$\begin{aligned}
H(Y|X) &= \sum_x p(x)H(Y|X=x) \\
&= p_0 H(Y|X=0) + p_1 H(Y|X=1) \\
&= p_0 H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) + p_1 H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right) \\
&= \underbrace{(p_0 + p_1)}_{=1} H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) \\
&= \frac{3}{2}.
\end{aligned} \tag{472}$$

Temos então que

$$C = \max_{p(x)} H(Y) - \frac{3}{2}. \tag{473}$$

Devemos agora analisar $H(Y)$, para tanto, usaremos $p(x)$ e $p(y|x)$.

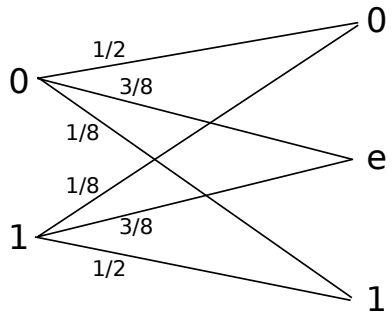
$$\begin{aligned}
\Pr(Y=A) &= \frac{1}{2}p_0 \\
\Pr(Y=B) &= \frac{1}{4}p_0 + \frac{1}{4}p_1 \\
\Pr(Y=C) &= \frac{1}{4}p_0 + \frac{1}{4}p_1 \\
\Pr(Y=D) &= \frac{1}{2}p_1.
\end{aligned} \tag{474}$$

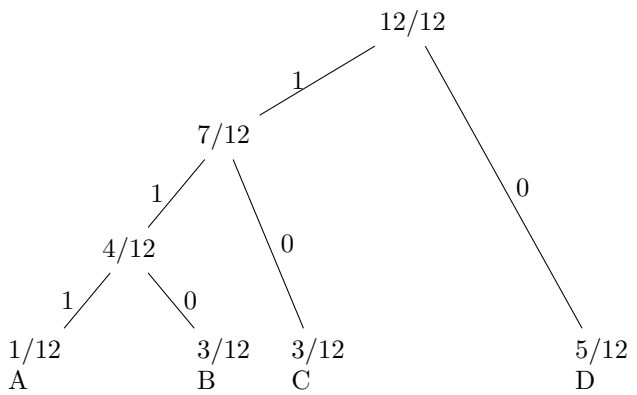
$H(Y)$ será maximizado quando $p_0 = p_1 = 1/2$, assim teremos $H(Y) = H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) = 2$ bits. Logo, teremos

$$C = 2 - \frac{3}{2} = \frac{1}{2} \text{ bit}. \tag{475}$$

(b) Dada a distribuição em X, devemos calcular a distribuição em Y para então criar o código ótimo, um código de Huffman.

$$\begin{aligned}
\Pr(Y=A) &= \frac{1}{2}p_0 = \frac{1}{12} \\
\Pr(Y=B) &= \frac{1}{4}p_0 + \frac{1}{4}p_1 = \frac{1}{4} \\
\Pr(Y=C) &= \frac{1}{4}p_0 + \frac{1}{4}p_1 = \frac{1}{4} \\
\Pr(Y=D) &= \frac{1}{2}p_1 = \frac{5}{12}
\end{aligned} \tag{476}$$





x	$C(x)$	$l(x)$
A	111	3
B	110	3
C	10	2
D	0	1

Comprimento esperado:

$$\begin{aligned} L(C) &= \sum_x p(y)l(y) = \\ &= \frac{3}{12} + \frac{9}{12} + \frac{6}{12} + \frac{5}{12} \\ &\approx 1.916. \end{aligned} \tag{477}$$

Entropia:

$$\begin{aligned} H(Y) &= -\sum_y p(y)\log p(y) \\ &= \frac{1}{12}\log 12 + \frac{3}{12}\log 4 + \\ &\frac{1}{4}\log 4 + \frac{5}{12}\log\frac{12}{5} \\ &= 2 + \frac{1}{2}\log 3 - \frac{5}{12}\log 5 \\ &\approx 1.825 \text{ bits.} \end{aligned} \tag{478}$$

Logo, temos $H(Y) \leq L(C)$, respeitando assim o teorema de Shannon.

5.12 Canal com apagamento 2

1. Considere o canal discreto sem memória representado na figura abaixo.
   (a) Qual é a capacidade do canal?
   (b) Qual é a distribuição de X que alcança a capacidade do canal? $(p = (p_0, p_1))$

Resolução:

A capacidade de canal é dada por

$$\begin{aligned}
C &= \max_{p(x)} I(X;Y) \\
&= \max_{p(x)} H(Y) - H(Y|X) \\
&= \max_{p(x)} H(Y) - H\left(\frac{1}{2}, \frac{3}{8}, \frac{1}{8}\right) \\
&= H\left(\frac{5}{16}, \frac{6}{16}, \frac{5}{16}\right) - H\left(\frac{1}{2}, \frac{3}{8}, \frac{1}{8}\right) \\
&= \frac{5}{16}(4 - \log 5) + \frac{3}{8}(3 - \log 3) + \frac{5}{16}(4 - \log 5) - \frac{1}{2} - \frac{3}{8}(3 - \log 3) - \frac{3}{8} \\
&= \frac{13}{8} - \frac{5}{8}\log 5. \qquad (479)
\end{aligned}$$

Utilizamos a simetria para identificar que a distribuição $p(x)$ que maximiza a $H(Y)$ é $\left(\frac{1}{2}, \frac{1}{2}\right)$, levando à distribuição $\left(\frac{5}{16}, \frac{6}{16}, \frac{5}{16}\right)$ na saída.

5.13 Canal de Comunicação 3

1. Considere um canal de comunicação conforme ilustrado abaixo.

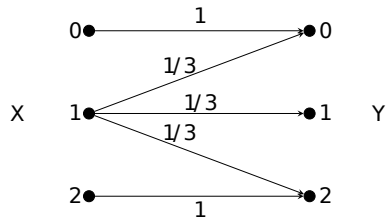


Figura 9: Canal de comunicação.

Para determinar a capacidade de um canal, devemos encontrar a distribuição sobre X que otimiza a informação mútua entre a entrada e saída. Pela simetria do canal em questão, e pelo fato de que a informação mútua é côncava nas probabilidades, podemos concluir que para otimizar a transmissão de informação no canal, devemos requerer que a distribuição em X seja tal que os símbolos 0 e 2 possuam a mesma probabilidade p.

(a) Determine a capacidade deste canal.

(b) Determine a distribuição em X que atinge esta capacidade.

(Não é necessário mostrar que a derivada segunda é negativa no ponto de distribuição encontrado.)

(Lembrete: $\log_b(x) = \log_a(x)/\log_a(b)$.)

Resolução: O problema nos fornece que $\Pr(X = 0) = \Pr(X = 2) = p$ e, por conseguinte, $\Pr(X = 1) = 1 - 2p$.

A capacidade do canal é dada por:

$$
\begin{aligned}
C &= \max_{p(x)} I(X;Y) = \max_{p(x)} \left(H(Y) - H(Y|X)\right) \\
&= \max_{p(x)} \left(H(Y) - \left(p \underbrace{H(Y|X=0)}_{=0} + (1-2p)H(Y|X=1) + p \underbrace{H(Y|X=2)}_{=0} \right) \right) \\
&= \max_{p(x)} \left(H(Y) - (1-2p) H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \right) = \max_{p(x)} \left(H(Y) - (1-2p)\log 3\right). \qquad (480)
\end{aligned}
$$

Para calcular $H(Y)$ devemos determinar a distribuição em Y. Podemos verificar que teremos $\Pr(Y = 0) = \Pr(Y = 2) = p + \frac{1}{3}(1-2p) = \frac{1+p}{3}$ e $\Pr(Y = 1) = \frac{1-2p}{3}$. Assim, poderemos expressar $H(Y)$ da seguinte forma:

$$
\begin{aligned}
H(Y) &= H\left(\frac{1+p}{3}, \frac{1-2p}{3}, \frac{1+p}{3}\right) \\
&= -2\frac{1+p}{3}\log\frac{1+p}{3} - \frac{1-2p}{3}\log\frac{1-2p}{3}. \qquad (481)
\end{aligned}
$$

Teremos assim

$$
C = \max_{p(x)} \left(\underbrace{-2\frac{1+p}{3}\log\frac{1+p}{3} - \frac{1-2p}{3}\log\frac{1-2p}{3} - (1-2p)\log 3}_{I_p(X;Y)} \right). \qquad (482)
$$

Para determinar o ponto de máximo, devemos encontrar o ponto em que a derivada é igual a zero. Para tanto, iremos primeiramente reescrever $I_p(X;Y)$ em nats:

$$
I_p(X;Y) = \frac{-2(1+p)}{3\ln 2}\ln\frac{1+p}{3} - \frac{1-2p}{3\ln 2}\ln\frac{1-2p}{3} - \frac{(1-2p)}{\ln 2}\ln 3 \quad \text{(nats)}. \qquad (483)
$$

E assim, a derivada será

$$
\begin{aligned}
\frac{\partial}{\partial p} I_p(X;Y) &= -\frac{2}{3\ln 2}\ln\frac{1+p}{3} - \frac{2(1+p)}{3\ln 2}\frac{3}{1+p} + \frac{2}{3\ln 2}\ln\frac{1-2p}{3} + \frac{2(1-2p)}{3\ln 2}\frac{3}{1-2p} + \frac{2}{\ln 2}\ln 3 \\
&= \frac{2}{3\ln 2}\left(\ln\frac{1-2p}{3} - \ln\frac{1+p}{3}\right) - \frac{2}{\ln 2} + \frac{2}{\ln 2} + \frac{2}{\ln 2}\ln 3 \\
&= \frac{2}{3\ln 2}\ln\frac{1-2p}{1+p} + \frac{2}{\ln 2}\ln 3 \quad \text{(nats)} \\
&= \frac{2}{3}\log\frac{1-2p}{1+p} + 2\log 3 \quad \text{(bits)} \qquad (484)
\end{aligned}
$$

Queremos $\frac{\partial}{\partial p} I_p(X;Y) = 0$, ou seja,

$$\begin{aligned}
\frac{2}{3}\log\frac{1-2p}{1+p} &= -2\log 3 \\
\log\frac{1-2p}{1+p} &= -3\log 3 \\
\frac{1-2p}{1+p} &= 2^{-3\log 3} \\
\frac{1-2p}{1+p} &= \frac{1}{27} \\
27-54p &= 1+p \\
-55p &= -26 \\
p &= \frac{26}{55}
\end{aligned} \tag{485}$$

Note que a derivada segunda de $I_p(X;Y)$ é

$$\begin{aligned}
\frac{\partial^2}{\partial p^2} I_p(X;Y) &= \frac{2}{3\ln 2}\frac{1+p}{1-2p}\frac{-2(1+p)-(1-2p)}{(1+p)^2} \\
&= \frac{2}{3\ln 2}\frac{1+p}{1-2p}\frac{-3}{(1+p)^2} \\
&= -\frac{2}{\ln 2}\frac{1}{1-2p}\frac{1}{1+p} = \frac{2}{\ln 2}\frac{1}{2p^2+p-1}
\end{aligned} \tag{486}$$

Como $0 \leq p \leq \frac{1}{2}$, teremos a derivada segunda sempre negativa, o que era de se esperar, já que $I_p(X;Y)$ é uma função côncava em p. Por conseguinte, o ponto encontrada é um ponto de máximo.

A capacidade do canal será dada por

$$C = H\left(\frac{1+26/55}{3}, \frac{1-2\times 26/55}{3}, \frac{1+26/55}{3}\right) - (1-2\times 26/55)\log 3 \tag{487}$$

$$= H\left(\frac{27}{55}, \frac{1}{55}, \frac{27}{55}\right) - \frac{3}{55}\log 3 \approx 1,0265. \tag{488}$$

5.14 Canais em cascata

1. Considere o canal de comunicação apresentado na figura abaixo, um canal constituído pela concatenação de um canal binário simétrico, com capacidade C_1, e um canal binário com apagamento, com capacidade C_2. Mostre que a capacidade do canal resultante será $C = C_1C_2$. (dica: pela concavidade e simetria da entropia é possível inferir de forma simples qual é a distribuição em X que leva ao máximo da informação mútua).

Resolução:

$$C = \max_{p(x)} I(X;\tilde{Y}) = \max_{p(x)} \Big(H(\tilde{Y}) - H(\tilde{Y}|X)\Big) \tag{489}$$

Vamos considerar $X \sim \mathrm{Ber}(\pi)$.

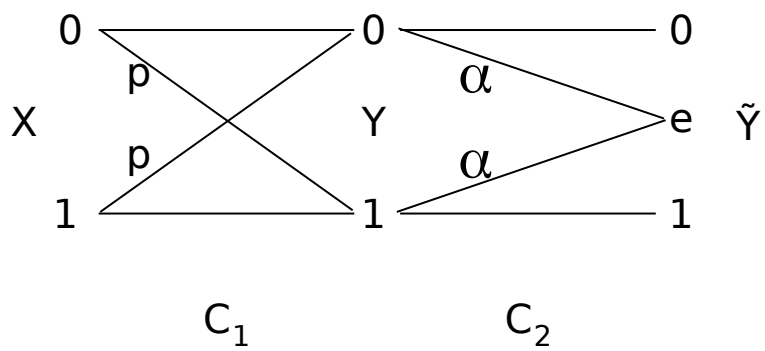


Figura 10: Canal de comunicação.

Para determinar $H(\tilde{Y})$ devemos calcular a distribuição em $\tilde{Y}$.

$$\begin{aligned}
\Pr(\tilde{Y}=0) &= (1-\alpha)(\pi p + (1-\pi)(1-p)) \\
\Pr(\tilde{Y}=e) &= \alpha \\
\Pr(\tilde{Y}=1) &= (1-\alpha)((1-\pi)p + \pi(1-p)),
\end{aligned} \tag{490}$$

assim, teremos

$$H(\tilde{Y}) = H\left((1-\alpha)(\pi p + (1-\pi)(1-p)), \alpha, (1-\alpha)((1-\pi)p + \pi(1-p))\right). \tag{491}$$

Para determinar $H(\tilde{Y}|X)$ devemos determinar as seguintes probabilidades condicionais:

$$\begin{aligned}
\Pr(\tilde{Y}=0|X=0) &= (1-p)(1-\alpha) \\
\Pr(\tilde{Y}=e|X=0) &= \alpha \\
\Pr(\tilde{Y}=1|X=0) &= p(1-\alpha) \\
\Pr(\tilde{Y}=0|X=1) &= p(1-\alpha) \\
\Pr(\tilde{Y}=e|X=1) &= \alpha \\
\Pr(\tilde{Y}=1|X=1) &= (1-p)(1-\alpha).
\end{aligned} \tag{492}$$

Assim teremos

$$\begin{aligned}
H(\tilde{Y}|X=0) &= H\left((1-p)(1-\alpha), \alpha, p(1-\alpha)\right), \\
H(\tilde{Y}|X=1) &= H\left(p(1-\alpha), \alpha, (1-p)(1-\alpha)\right) \text{ e} \\
H(\tilde{Y}|X) &= (1-\pi)H(\tilde{Y}|X=0) + \pi H(\tilde{Y}|X=1).
\end{aligned} \tag{493}$$

Pela concavidade e simetria de H podemos concluir que a informação mútua será máxima quando $\pi = \frac{1}{2}$. Teremos assim

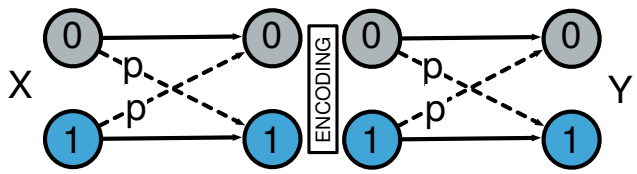
$$\begin{aligned}
H(\tilde{Y})\Big|_{\pi=\frac{1}{2}} &= H\left(\frac{1}{2}(1-\alpha), \alpha, \frac{1}{2}(1-\alpha)\right) \\
H(\tilde{Y}|X)\Big|_{\pi=\frac{1}{2}} &= H\left((1-p)(1-\alpha), \alpha, p(1-\alpha)\right).
\end{aligned} \tag{494}$$

Teremos então

$$
\begin{aligned}
C &= H\left(\frac{1}{2}(1-\alpha), \alpha, \frac{1}{2}(1-\alpha)\right) - H\left((1-p)(1-\alpha), \alpha, p(1-\alpha)\right) \\
&= -(1-\alpha)\log\frac{(1-\alpha)}{2} - \alpha\log\alpha + \\
&(1-p)(1-\alpha)\log\left((1-p)(1-\alpha)\right) + \alpha\log\alpha + p(1-\alpha)\log\left(p(1-\alpha)\right) \\
&= (1-\alpha)\left(1 + p\log p + (1-p)\log(1-p)\right) \\
&= (1-\alpha)\left(1 - H(p)\right) = C_1 C_2 \quad \blacksquare
\end{aligned} \tag{495}
$$

5.15 Canal binário simétrico em cascata

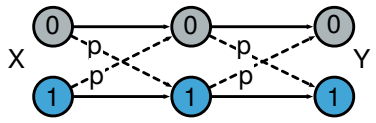
1. Considere os canais binários simétricos conectados em cascata com um codificador entre eles, conforme ilustrado abaixo. Calcule a capacidade do canal entre X e Y.



Qual é a capacidade deste canal?

Resolução: Os símbolos são re-codificados após o primeiro canal, desta forma a capacidade torna-se $C = \min(C_1, C_2)$, onde C_1 e C_2 são as capacidades do primeiro e segundo canais binários, respectivamente. Como neste caso os canais são iguais, temos $C_1 = C_2 = 1 - H(p)$. Logo, a capacidade do canal será $C = 1 - H(p)$, sendo este valor alcançado quando a distribuição de X for uniforme.

2. Considere um canal formado por dois canais binários simétricos em casacata, conforme ilustrado abaixo. Calcule a capacidade deste canal.



Resolução: O canal total terá a seguinte matriz de transmissão:

$$
p(y|x) = \begin{pmatrix} (1-p)^2 + p^2 & 2p(1-p) \\ 2p(1-p) & (1-p)^2 + p^2 \end{pmatrix} \tag{496}
$$

A capacidade do canal é dada por:

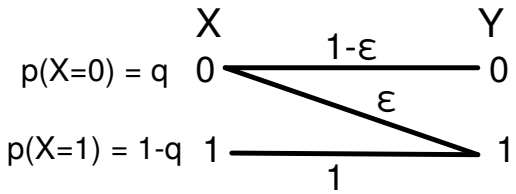
$$\begin{aligned}
C &= \max_{p(x)} I(X;Y) = \max_{p(x)} \left(H(Y) - H(Y|X)\right) \\
&= 1 - H\left((1-p)^2 + p^2, 2p(1-p)\right) = 1 - H\left(2p(1-p)\right) \\
&= 1 + \left((1-p)^2 + p^2\right) \log\left((1-p)^2 + p^2\right) + 2p(1-p)\log\left(2p(1-p)\right) \\
&= \left(1 - H(p)\right)^2
\end{aligned} \tag{497}$$

Faça a verificação numérica do último passo para $p \in [0,1]$.

Ou confira o resultado no WolframAlpha.

5.16 Canal Z (genérico)

1. Considere o canal Z, com capacidade C_Z, ilustrado abaixo. Considere a seguinte distribuição de entrada: $p(X=0) = q$, $p(X=1) = 1-q$. Encontre a equação que deverá ser otimizada para encontrarmos a capacidade deste canal (simplifique a equação ao máximo, mas não é necessário solucioná-la).



obs.: as características de erro em sistemas ópticos e memória de alguns semicondutores podem ser modeladas através de um canal z.

Resolução: Devemos encontrar q que maximiza $I(X;Y) = H(Y) - H(Y|X)$. Primeiramente iremos encontrar uma expressão para $I(X;Y)$ em termos de q e ϵ. Vamos tomar a derivada com relação a q e achar o q que faz com que esta derivada seja igual a zero.

Note que $p(Y=0) = q(1-\epsilon)$ e, por conseguinte, $p(Y=1) = 1 - q(1-\epsilon)$, assim, $H(Y) = H(q(1-\epsilon))$.

$$\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= H(q(1-\epsilon)) + (1-q)\underbrace{H(Y|X=1)}_{=0} + q\underbrace{H(Y|X=0)}_{=H(\epsilon)} \\
&= H(q(1-\epsilon)) + qH(\epsilon) \\
&= -q(1-\epsilon)\log q(1-\epsilon) - (1-q(1-\epsilon))\log(1-q(1-\epsilon)) + \\
&\quad qH(\epsilon)
\end{aligned} \tag{498}$$

Devemos tomar $\frac{dI(X;Y)}{dq} = 0$ e achar q que satisfaz esta equação.

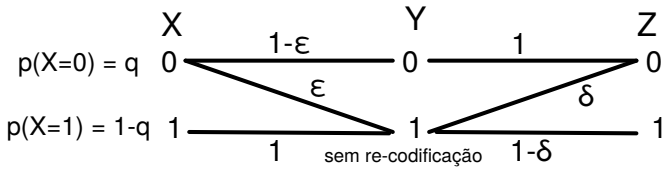
$$C \approx 1 - \frac{1}{2}H(\epsilon) \tag{499}$$

que será alcançada quando

$$q = \frac{1}{(1-\epsilon)(1 + 2^{H(\epsilon)/(1-\epsilon)})} \tag{500}$$

5.17 Canais Z em cascata

1. Considere a cascata de dois canais Z, conforme ilustrado na figura abaixo. A saída de um canal é inserida



diretamente no canal seguinte, sem recodificação. Encontre as probabilidades de transição do canal final criado pela cascata dos 2 canais Z.

(a) Encontre as probabilidades de transição para o canal final (entre X e Z).

(b) Encontre o valor de δ (em termos de ϵ) que faz com que o canal final XZ seja simétrico e encontre a capacidade deste canal C_{XZ}.

(c) Considere agora que, em Y, seja possível decodificar e recodificar a sequência recebida. Qual é a capacidade do sistema agora?

Resolução:

(a) O canal efetivo entre X e Z é descrito pelas probabilidades

$$p(Z=0|X=0) = (1-\epsilon) + \epsilon\delta \tag{501}$$

$$p(Z=1|X=0) = \epsilon(1-\delta) \tag{502}$$

$$p(Z=0|X=1) = \delta \tag{503}$$

$$p(Z=1|X=1) = 1-\delta \tag{504}$$

$$p(z|x) = \begin{pmatrix} (1-\epsilon)+\epsilon\delta & \epsilon(1-\delta) \\ \delta & 1-\delta \end{pmatrix} \tag{505}$$

(b) Devemos ter $p(Z=1|X=0) = p(Z=0|X=1)$, ou seja,

$$\epsilon = \frac{\delta}{1-\delta} \tag{506}$$

Note que esta escolha levará também a $p(Z=1|X=1) = p(Z=0|X=0)$.

Como o canal é simétrico, teremos

$$\begin{aligned} C &= \log|\mathcal{Z}| - H(r) \\ &= \log 2 - H(\delta) \\ &= 1 - H(\delta). \end{aligned} \tag{507}$$

onde r é uma linha da matriz de transição, ou seja, $r = [\delta \quad (1-\delta)]$, utilizando $\epsilon = \delta/(1-\delta)$.

(c)

$$C = \min(C_Z, C_{YZ}). \tag{508}$$

onde C_Z é a capacidade do canal X—Y e C_{YZ} é a capacidade do canal Y—Z.

6 Código de Hamming

6.1 Código de Hamming (7, 4, 3)

1. Código de Hamming $(7, 4, 3)$

$$x_4 = (x_1 + x_2 + x_3) \mod 2 \tag{509}$$

$$x_5 = (x_0 + x_2 + x_3) \mod 2 \tag{510}$$

$$x_6 = (x_0 + x_1 + x_3) \mod 2 \tag{511}$$

$$H = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}. \tag{512}$$

Considere os números a e b, o penúltimo e o último algarismos da sua matrícula, respectivamente. O dado a ser enviado é a representação binária com 4 bits de $((10a+b) \mod 16)$ (exemplo: se a matrícula fosse 114350047, teríamos $a = 4$ e $b = 7$, assim $47 \mod 16 = 15$ e o dado em binário seria 1111).

(a) Escreva sequência x que representa o código que será enviado utilizando a codificação de Hamming (7,4,3).

(b) Considere um ruído aditivo que trocará o k-ésimo bit transmitido, onde k será determinado por $k = ((a + b) \mod 7)$, ou sejam, $z = z_0 z_1 \ldots z_k \ldots z_7$, onde todos bits, exceto o bit z_k, são iguais a zero. (para o exemplo anterior, $k = 11 \mod 7 = 4$ e assim o ruído seria $z = 0000100$). Qual será o sinal recebido? Mostre como o decodificador poderá corrigir o erro calculando a síndrome.

(c) Qual é o tamanho do *codebook* para o código de Hamming (7,4,3)? Este conjunto é fechado em relação à soma? Mostre um exemplo.

(d) O que ocorrerá se houverem 2 bits errados? E se houverem 3 bits errados?

Resolução:

(a) Para o exemplo em que $a = 4$ e $b = 7$ teríamos $x = 1111111$.

(b) $y = x + z = 1111011$

A síndrome é $s = \mathbf{H}y = (100)^T$, que corresponde a quinta coluna de $\mathbf{H}$. Desta forma determinamos que $z = 0000100$, o que está correto. Podemos encontrar x fazendo $x = y + z = 1111011 + 0000100 = 1111111$.

(c) O codebook possui 16 palavras e é um conjunto fechado em relação à soma.

Exemplo: $w_1 = 0010110$ e $w_2 = 1011010$, $w_1 + w_2 = 1001100$ que satisfaz as equações 509, 510 e 511 e portanto é um código de Hamming (7,4,3).

(d) Se houverem 2 bits errados o código recebido será não satisfará ao mesmo tempo as equações 509, 510 e 511. Será possível determinar a existência de um erro, mas não será possível corrigir corretamente.

Se houverem 3 bits errados, a palavra recebida será um código de Hamming válido e assim o erro passará despercebido.

6.2 Código de Hamming Estendido

1. O Código de Hamming $(7, 4, 3)$ pode ser facilmente estendido para o código $(8, 4, 4)$, bastando para tanto, acrescentar um bit de paridade extra conforme a Figura 11, onde x_4, x_5, x_6 e x_7 são os bits de paridade, sendo x_7 o bit extra acrescentado ao código estendido.

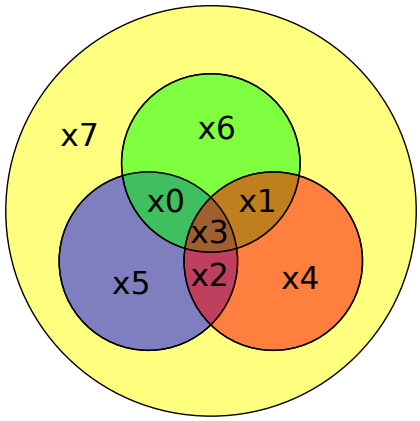


Figura 11: Código de Hamming estendido $(8, 4, 4)$.

(a) Apresente a matriz geradora para o código de Hamming estendido $(8, 4, 4)$.

(b) Apresente a matriz de verificação de paridade para este código.

(c) Qual é a distância mínima entre palavras no código de Hamming estendido $(8, 4, 4)$? Justifique.

(d) Qual é o peso do código? Justifique.

(e) Faça um exemplo de codificação. Considere o seu número de matrícula, contendo N algarismos, na forma $a_{N-1}a_{N-2}\ldots a_1a_0$. Considere como bits de dados: $x_0 = a_{N-2} \mod 2$, $x_1 = a_{N-3} \mod 2$, $x_2 = a_1 \mod 2$, $x_3 = a_0 \mod 2$. Determine $x = x_0, x_1, \ldots, x_7$, a sequência de bits codificada. Suponha que x será corrompido pelo ruído z, onde $z = z_0, z_1, \ldots, z_7$. Faça $z_i = 1$, $i = a_0 \mod 8$, e a_0 é o algarismo menos significativo do seu número de matrícula. A sequência corrompida pelo ruído será chamado de y. É possível determinar x a partir de y? Em caso afirmativo, mostre como.

(f) O que ocorrerá quando apenas um bit de x for trocado? Será possível detectar erro? Será possível corrigir o erro? Justifique.

(g) O que ocorrerá quando dois bits de x forem trocados? Será possível detectar erro? Será possível corrigir os erros? Justifique.

Resolução:

(a) Os primeiros 7 bits são iguais àqueles gerados pelo Código de Hamming $(7, 4, 3)$. O último bit será dado pela equação

$$\begin{aligned} x_7 &= (x_0 + x_1 + x_2 + x_3 + x_4 + x_5 + x_6) \mod 2 \\ &= (3x_0 + 3x_1 + 3x_2 + 4x_3) \mod 2 \\ &= (x_0 + x_1 + x_2) \mod 2. \end{aligned} \tag{513}$$

Logo, a matriz geradora será

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ \hdashline 1 & 1 & 1 & 0 \end{pmatrix}. \tag{514}$$

(b) Para o código de Hamming estendido serão 4 equações de verificação de paridade. Três delas são as mesmas dadas pelo Código de Hamming $(7, 4, 3)$. A última equação deverá realizar a verificação de paridade de todos os bits (ou alternativamente, x_0, x_1, x_2, x_7). Esta última equações para verificação de paridade será

$$x_0 + x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 = 0, \tag{515}$$

ou, alternativamente,

$$x_0 + x_1 + x_2 + x_7 = 0. \tag{516}$$

A matriz de verificação de paridade será

$$H = \left(\begin{array}{ccccccc:c} 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ \hdashline 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right). \tag{517}$$

ou, alternativamente,

$$H = \left(\begin{array}{ccccccc:c} 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ \hdashline 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{array}\right). \tag{518}$$

(c) A distância mínima do código de Hamming estendido $(8, 4, 4)$ é de 4. Sendo duas palavras do código $v_1, v_2 \in C$, então $H(v_1 - v_2) = 0$, pois o código é fechado em relação à soma e subtração. Suponha que v_1 e v_2 se difiram apenas na posição i, então $H(v_1 - v_2) = h_i$, onde h_i é a i-ésima coluna de H. Como não existe coluna nula em H, temos um contradição. Logo, não é possível que duas palavras do código v_1 e v_2 se difiram em apenas uma posição. Vamos supor agora que elas se difiram em 2 posições, i e j. Neste caso, $H(v_1 - v_2) = h_i + h_j$, ou seja, a combinação das colunas i e j de H. Entretanto, as colunas de H são distintas, logo não é possível que a soma de duas colunas seja nula. Mais uma vez, por contradição, verificamos que a distância mínima entre duas palavras do código dado não pode ser 2. Até aqui, seguimos os mesmos argumentos que aqueles dados para o código de Hamming $(7, 4, 3)$. Vamos agora analisar o caso em que v_1 e v_2 se difiram em três posições i, j e k. Teremos agora $H(v_1 - v_2) = h_i + h_j + h_k$, e temos que $v_1, v_2 \in C$, então $H(v_1 - v_2) = 0$. Note que o último elemento de cada coluna de H é 1, logo qualquer soma de 3 colunas de H terá em seu último elemento o valor 1 e, por conseguinte, não poderemos ter o resultado nulo. Por contradição, mais uma vez, concluímos que a distância mínima entre duas palavras no código de Hamming estendido $(8, 4, 4)$ não pode ser 3. Para uma distância 4 é possível fazer uma combinação de 4 colunas de H igualando a zero, como por exemplo, as colunas 1, 2, 3 e 8, somadas resultarão em zero.

(d) O peso do código é 4.

1. Não podemos ter peso 1 pois as palavras com peso 1 não estão no espaço nulo de $\mathbf{H}$. Suponha que a palavra x seja não nula apenas na posição i. Então $\mathbf{H}x$ será igual à i-ésima coluna de $\mathbf{H}$. Mas não existe coluna nula em $\mathbf{H}$, desta forma não é possível ter $\mathbf{H}x = 0$ com palavras de peso 1.
2. Suponha que o peso do código seja 2, com elementos não nulos nas posições i e j. Neste caso, por ser uma palavra do código, devemos ter $\mathbf{H}x = 0$, assim a soma da i-ésima e j-ésima colunas de $\mathbf{H}$ deverá ser igual a zero. Como as colunas de $\mathbf{H}$ são todas diferentes, a soma de quaisquer duas colunas é não nula. Desta forma não é possível que o código tenha peso 2.

3. Peso 3 também não é possível. Note que a última linha de **H** é toda igual a 1, logo ao somar 3 colunas de **H**, no último índice estaremos somando 1 três vezes, e terá como resultado 1. Desta forma, será impossível obter como resultado o 0 desejado.
4. Peso 4 será possível.

(e) Vamos supor aqui que o número de matrícula é 12345678. Neste caso, teremos $x_0 = 0$, $x_1 = 1$, $x_2 = 0$ e $x_3 = 1$, podemos assim calcular os bits de baridade e obter o vetor $x = [01010101]$. Para o exemplo dado, teremos $z = [10000000]$, logo $y = x + z = [11010101]$. Para determinar x a partir de y devemos calcular a síndrome, $s = Hy = H(x + z) = Hz$.

$$s = Hy = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}. \tag{519}$$

Como s não é nula, sabemos que ocorreu erro. s é igual à primeira coluna de H, logo z é um vetor todo nulo, exceto em sua primeira posição, $z = [10000000]$. Após obter z, podemos achar x, dado que $y = x+z$. Teremos então $x = [01010101]$. Quando o erro for ocasionado pela troca de apena um bit, poderemos corrigi-lo.

(f) Quando apenas um bit for trocado, será possível detectar e corrigir o erro, conforme exemplificado no item anterior. Se o erro for em apenas um bit, a síndrome será uma das colunas de H. Bastará então identificar qual coluna para saber qual bit foi trocado.

(g) Se dois bits forem trocados, será possível detectar o erro, pois a síndrome será a combinação de duas colunas de H, que não será nula. Entretanto, não será possível corrigir, pois para uma mesma síndrome existem diferentes combinações de duas colunas de H que levam a um mesmo resultado. Por exemplo, se o erro trocar os dois primeiros bits, $z = [11000000]$, teremos $y = x + z = [10010101]$.

$$s = Hy = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \tag{520}$$

que pode ser obtido pela diferentes combinações de duas colunas de H: $(1, 2)$, $(3, 8)$, $(4, 7)$ e $(5, 6)$.

6.3 Código de Hamming ternário

1. Considere o código de Hamming ternário (4,2,2) em que as palavras possuem comprimento 4, sendo formadas por 2 símbolos de dados e 2 símbolos de paridade. A matriz geradora deste código e a matriz de verificação de paridade são dadas a seguir,

$$G^T = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \end{pmatrix}, \tag{521}$$

$$H = \begin{pmatrix} 2 & 2 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{pmatrix}. \tag{522}$$

Escreva as equações para os símbolos de paridade e as equações de verificação de paridade. Faça 3 exemplos de utilização deste código (com casos em que a decodificação será bem sucedida, mesmo na presença de ruído, e casos em que a decodificação falhará): defina os símbolos de dados; calcule os símbolos de paridade; suponhas um ruído aditivo (que altere um ou dois símbolos); some o ruído à palavra gerada pelo código; calcule a síndrome; faça a decodificação, evidenciando os casos em que é possível detectar e corrigir erro, casos em que é possível apenas detectar erro, e casos em que o erro passa desapercebido.

Resolução: Seja x_1 e x_2 os símbolos de dados, $x = [x_1x_2x_3x_4]$ será a palavra código, em que os símbolos x_3 e x_4 são dados pelas seguintes equações:

$$x_3 = x1 + x_2 \pmod 3 \quad \text{e} \tag{523}$$

$$x_4 = x1 + 2x_2 \pmod 3. \tag{524}$$

As equações de verificação de paridade são:

$$2x_1 + 2x_2 + x_3 \pmod 3 = 0 \quad \text{e} \tag{525}$$

$$2x_1 + x_2 + x_4 \pmod 3 = 0. \tag{526}$$

exemplo 1
Suponha que $x_1 = 0$ e $x_2 = 1$. A palavra de código será $x = [0112]$.

- Considere o ruído aditivo $r = [1000]$. Neste caso, teremos $y = x + r \pmod 3 = [1112]$. A síndrome será calculada por $s = Hy = [22]$. Podemos verificar que s é igual à primeira coluna de H. Logo podemos corrigir o erro: $\hat{x} = y - [1000] = [0112] = x$.
- Considere o ruído aditivo $r = [2000]$. Teremos $y = x + r \pmod 3 = [2112]$. A síndrome será $s = Hy = [12]$. Não há coluna em H igual à s, logo é possível detectar o erro, mas não é possível corrigi-lo.
- Considere o ruído aditivo $r = [0100]$. Teremos $y = x + r \pmod 3 = [0212]$. A síndrome será $s = Hy = [21]$. Podemos verificar que s é igual à segunda coluna de H. Faremos $\hat{x} = y - [0100] = [0112] = x$.
- Considere o ruído aditivo $r = [0200]$. Teremos $y = x + r \pmod 3 = [0012]$. A síndrome será $s = Hy = [12]$. Não há coluna em H igual à s, logo é possível detectar o erro, mas não é possível corrigi-lo.

exemplo 2
Suponha que $x_1 = 2$ e $x_2 = 1$. A palavra de código será $x = [2101]$.

- Considere o ruído aditivo $r = [0001]$. Neste caso, teremos $y = x + r \pmod 3 = [2102]$. A síndrome será calculada por $s = Hy = [01]$. Podemos verificar que s é igual à ultima coluna de H. Logo podemos corrigir o erro: $\hat{x} = y - [0001] = [2101] = x$.
- Considere o ruído aditivo $r = [1001]$. Teremos $y = x + r \pmod 3 = [0102]$. A síndrome será $s = Hy = [10]$. Encontramos s igual à terceira coluna em H. Ao aplicar a correção teremos: $\hat{x} = y - [0010] = [0122]$. Fornecendo assim $\hat{x}_1 = 0$ e $\hat{x}_2 = 1$, o que está errado. Este erro passa desapercebido.
- Considere o ruído aditivo $r = [0200]$. Teremos $y = x + r \pmod 3 = [2001]$. A síndrome será $s = Hy = [12]$. Não há coluna em H igual à s, logo é possível detectar o erro, mas não é possível corrigi-lo.

6.4 Hamming

1. Um código de Hamming utiliza palavras de comprimento igual a 7 bits. Suponha que esta mensagem, codificada com este código, será enviada através de um canal binário simétrico com probabilidade de troca de bit $p = 0.1$. Faça o que se pede abaixo.
   (a) Determine a taxa deste código.

(b) Calcule a probabilidade de comunicação sem erro.

(c) Calcule a probabilidade de detecção de erro, dado que a comunicação teve erro.

(d) Faça exemplos de codificação, decodificação e decodificação com correção de erro, para ilustrar todos os possíveis casos de operação.

Resolução:

(a) Trata-se do código de Hamming (7,4,3), utiliza 4 bits de dados e 3 bits de paridade.

(b) A taxa do código é $4/7 \approx 0,571$.

(c) A comunicação será sem erro se a transmissão for sem erro, ou se for possível corrigir o erro. Para o código de Hamming (7,4,3) só é possível detectar e corrigir no máximo um único bit trocado na sequência. Para que a transmissão seja sem erro, nenhum bit pode ser trocado. Isto ocorrerá com probabilidade $(1-p)^7 \approx 0,47830$. Para que ocorra apenas um erro, a probabilidade será dada por $7 \times p(1-p)^6 \approx 0,37201$. Desta forma o percentual de comunicação sem erro será $(1+6p)(1-p)^6 \approx 0,47830 + 0,37201 = 0,85031$.

(d) Ocorrerá algum erro com probabilidade $1 - 0,47830 = 0,52170$. O erro será detectado se for de um número ímpar de bits trocados. Desta forma, a probabilidade dos casos em que é detectado erro é dada por

$$\binom{7}{1}p(1-p)^6 + \binom{7}{3}p^3(1-p)^4 + \binom{7}{5}p^5(1-p)^2 + \binom{7}{7}p^7 \approx 0,39513, \tag{527}$$

logo, dado que ouve erro na comunicação, a probabilidade de corrigir o erro é de $0,39513/0,52170 = 0,75739$

(e) Ver exemplos apresentados em aula.

6.5 Hamming Unicode

1. No Unicode, a letra 'a' possui o ponto de representação `U+0061`. Este, por sua vez, é codificado no UTF-8 como a sequência de 8 bits a seguir: 01100001.

   (a) Utilizando o código de Hamming (7,4,3) codifique esta sequência de bits que representa o caractere 'a' para serem transmitidos através de um canal ruidoso.

   (b) Suponha que um ruído aditivo esteja presente no canal, acarretando a troca dos bits nas posições 3 e 12 (considerando a sequência completa de bits formada após a codificação de Hamming). Utilize a decodificação do código de Hamming para detectar erro na transmissão e corrigi-lo. Mostre como.

   (c) ($A \to B \to C$) Suponha que o receptor B trabalha com uma extremidade de bit (*bit endianness*)[5] diferente daquela do transmissor A. O receptor B recebe a sequência de bits, enviados pelo transmissor A, através de um canal com ruído, e aplica a verificação e correção de erros do código de Hamming. Em seguida transmite os dados, através de um canal sem ruído, para um terceiro C que utiliza a mesma extremidade que o transmissor original A da mensagem. Como esta última transmissão foi através de um canal sem ruído, o último receptor C não precisa se preocupar com detecção e correção de erros. A correção de erros realizada pelo receptor intermediário B danificará os dados, impossibilitando que o receptor final C decodifique corretamente a mensagem? Justifique.

[5] Ordenação dos bits, convenção adotada para identificar a posição dos bits em um número binário. LSB 0: o bit menos significativo está no primeira posição. MSB 0: o bit mais significativo está na primeira posição.

Resolução:

(a) Os dados 0110|0001 darão origem a duas palavras do código de Hamming (7,4,3): 0110011 e 0001111 (usando as matrizes vistas em aula) ou 1100110 e 1101001 (usando as matrizes criadas pelo algoritmo). Estas são dadas pela multiplicação $x = Gp$, onde p são os bits de dados, x a palavra de hamming e G a matriz geradora.

(b) O sinal recebido será dado por $y = x + r$, onde r é o ruído aditivo.

Para realizar a verificação e correção de erros devemos calcular a síndrome: $s = Hy$. Se a síndrome for nula, não há erro, caso contrário, devemos encontrar a coluna de H igual à síndrome. Para corrigir o erro bastará trocar o bit na posição correspondente àquela onde encontramos s em H.

Como no exercício em questão temos apenas um erro em cada palavra de Hamming, ao proceder a correção este erro será corrigido.

(c) O receptor C receberá e decodificará a mensagem sem erro pois B irá corrigir os erros existentes. Como B trabalha com a extremidade de bit inversa, ele irá ter a matriz de verificação de paridade invertida, que será utilizada com o vetor de bits invertido. Isto equivale a trocar a ordem das equações de verificação de paridade, o que não muda em nada o funcionamento do sistema. Desta forma, tudo funcionará normalmente, porém na ordem inversa. Para C isso será transparente e não fará diferença.

Se B não usar a matriz invertida em relação a A, ou seja, B estaria utilizando uma matriz invertida em relação ao código de Hamming, teremos erro na decodificação, exceto nos casos em que a palavra transmitida seja um palíndromo.