Teoria da Informação

Charles Casimiro Cavalcante

charles@gtel.ufc.br

Grupo de Pesquisa em Telecomunicações Sem Fio – GTEL
Programa de Pós-Graduação em Engenharia de Teleinformática
Universidade Federal do Ceará – UFC
http://www.gtel.ufc.br/∼charles

"A principal função de um sistema de comunicação é reproduzir, exatamente ou de forma aproximada, uma informação proveniente de outro ponto diferente."

Claude Shannon, 1948

Conteúdo do curso

1. Revisão de probabilidade
2. Informação e Entropia
3. Codificação de fontes
4. Codificação e capacidade de canal
5. Complexidade de Kolmogorov
6. Funções de otimização
7. *Independent Component Analysis*

Parte II

Informação e Entropia

O que é informação?

- Medida da quantidade de *incerteza* de um processo que ocorre com alguma probabilidade
- Definição de Shannon, 1948
- Ferramentas probabilísticas
- Contexto
  - Fonte discreta
  - Alfabeto finito: $\mathcal{A} = \{a_0,\ a_1,\ \cdots\ ,\ a_{K-1}\}$
  - Probabilidades: $\Pr(A = a_k) = p_k$ em que $\sum_{k=0}^{K-1} p_k = 1$

Definição

Informação

$$\begin{aligned}\mathcal{I}(a_k) &= \log_\alpha\left(\frac{1}{\Pr(a_k)}\right)\\ &= \log_\alpha\left(\frac{1}{p_k}\right)\\ &= -\log_\alpha(p_k)\end{aligned} \tag{23}$$

- Unidade da informação depende da base α, e.g.
  1. $\alpha = 2 \Rightarrow$ informação em bits
  2. $\alpha = e \Rightarrow$ informação em nats

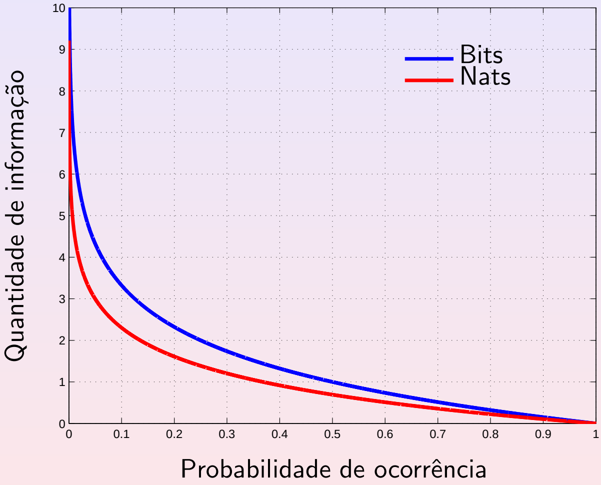
O que mede a informação?

- De uma forma mais informal, informação é a *surpresa* da ocorrência de um evento
- Quanto mais surpresa (incerteza) mais informação e, de forma contrária, quanto menos incerteza menos informação

Propriedades da informação

1. $\mathcal{I}(a_k) = 0$ se $p_k = 1$
2. $\mathcal{I}(a_k) \geq 0$ para $0 \leq p_k \leq 1$
   Nunca há *perda* de informação!
3. $\mathcal{I}(a_k) > \mathcal{I}(a_i)$ para $p_k < p_i$
4. $\mathcal{I}(a_k a_i) = \mathcal{I}(a_k) + \mathcal{I}(a_i)$ se a_k e a_i são estatisticamente independentes

Quantidade de informação



Informação pontual e informação média

- Pode-se desejar então calcular a quantidade média de informação de uma fonte A
- A essa média da informação denomina-se **entropia**

$$\begin{aligned} \mathcal{H}(A) &= \sum_{k=0}^{K-1} p_k \cdot \mathcal{I}(a_k) \\ &= -\sum_{k=0}^{K-1} p_k \cdot \log_{\alpha}(p_k) \end{aligned} \tag{24}$$

- A entropia mede a *quantidade de informação média por símbolo da fonte*

Propriedades da entropia

$$0 \leq \mathcal{H}(A) \leq \log_\alpha(K) \tag{25}$$

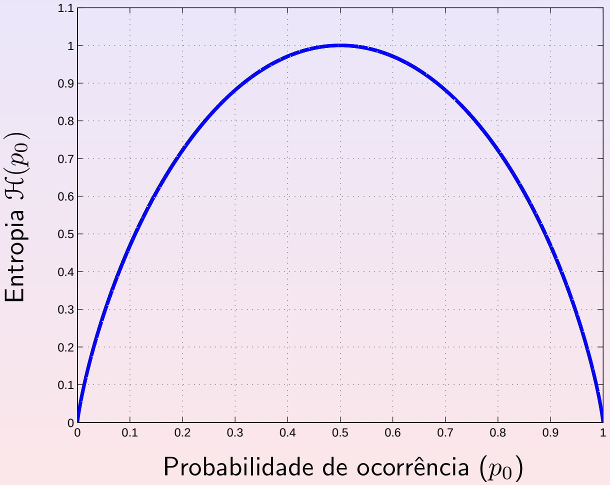
- $\mathcal{H}(A) = 0$ se e somente se a probabilidade de ocorrência p_k de um certo evento a_k for $p_k = 1$ e todas as demais forem iguais à zero. Neste ponto não existe nenhuma *incerteza* e conseqüentemente a entropia é mínima.
- $\mathcal{H}(A) = \log_\alpha(K)$ se e somente se as probabilidades de todos os eventos a_k forem iguais, ou seja, os eventos forem eqüiprováveis $\left(p_k = \frac{1}{K}\right)$.

Entropia de uma fonte binária

Seja uma fonte binária com p_0 e p_1 as probabilidades dos símbolos a_0 e a_1. A entropia é dada por:

$$\begin{aligned}\mathcal{H}(A) &= -p_0 \log_\alpha(p_0) - p_1 \log_\alpha(p_1)\\ &= -p_0 \log_\alpha(p_0) - (1-p_0)\log_\alpha(1-p_0)\end{aligned}$$

Entropia de uma fonte binária - gráfico $\mathcal{H}(p_0) \times p_0$



Entropia - observações

- A entropia pode ainda ser representada matematicamente como
$$\mathcal{H}(A) = -\mathbb{E}\left\{\log(p_A(a))\right\}, \tag{26}$$
em que $p_A(a)$ é a função de densidade de probabilidade de A.
- O que a entropia fornece é a de quanto de informação há, em média, num determinado símbolo de uma fonte. Isto será de grande interesse no projeto de codificadores de fonte.

- Até o momento foi vista a entropia de uma única variável aleatória
- Estendendo o conceito para duas variáveis tem-se novas definições
- **Entropia conjunta**

$$\begin{aligned}\mathcal{H}(A,B) &= -\sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} p(a,b)\log[p_{A,B}(a,b)] \\ &= -\mathbb{E}\left\{\log[p_{A,B}(a,b)]\right\}\end{aligned} \tag{27}$$

Fornece a quantidade de informação média na ocorrência de duas v.a.

Entropia condicional

- **Entropia condicional**

$$
\begin{aligned}
\mathcal{H}(A|B) &= \sum_{b\in\mathcal{B}} p_B(b)\cdot\mathcal{H}(A|B=b) \\
&= -\sum_{b\in\mathcal{B}} p_B(b) \sum_{a\in\mathcal{A}} p_{A|B}(a|b)\log[p_{A|B}(a|b)] \\
&= -\sum_{b\in\mathcal{B}}\sum_{a\in\mathcal{A}} p_{A,B}(a,b)\log[p_{A|B}(a|b)] \\
&= -\mathbb{E}\left\{\log[p_{A|B}(a|b)]\right\}
\end{aligned} \tag{28}
$$

Medida da quantidade média de informação de uma v.a. dada a ocorrência de outra

Entropia - Relações importantes

1. Regra da cadeia

$$\mathcal{H}(A, B) = \mathcal{H}(A) + \mathcal{H}(B|A) \tag{29}$$

A entropia de um par de variáveis é igual a entropia de uma mais a entropia condicional.

2. Corolário da regra da cadeia

$$\mathcal{H}(A, B|C) = \mathcal{H}(A|C) + \mathcal{H}(B|A, C), \tag{30}$$

Entropia relativa

- **Entropia relativa:** é a medida de "distância" entre duas distribuições. Pode ser entendido como uma medida de *ineficiência* de assumir que uma v.a. tem distribuição $p(x)$ quando a verdadeira distribuição é $g(x)$.

$$\begin{aligned} D(p\|g) &= \sum_{x\in\mathcal{X}} p(x)\cdot\log\left(\frac{p(x)}{g(x)}\right) \\ &= \mathbb{E}_{p(x)}\left\{\log\left(\frac{p(x)}{g(x)}\right)\right\} \end{aligned} \tag{31}$$

- A Equação (31) é também conhecida como **Divergência de Kullback-Leibler (KLD)** ou ainda *entropia cruzada*

Propriedades

1. é sempre de valor positivo ou zero; KLD é zero para o caso específico de $p_x(\mathbf{x}) = g_x(\mathbf{x})$.
2. é invariante com relação às seguintes mudanças nos componentes do vetor $\mathbf{x}$;
   - permutação de ordem
   - escalonamento de amplitude
   - transformação monotônica não-linear
3. não é uma distância no espaço euclidiano pois $D(p\|g) \neq D(g\|p)$
4. é uma distância no espaço das distribuições de probabilidade (espaço de Riemann)

- **Definição:** para duas variáveis aleatórias A e B, a **informação mútua** é a entropia relativa entre a distribuição conjunta de A e B e o produto das distribuições marginais.

$$\begin{aligned}\mathcal{I}(A,B) &= \sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} p_{A,B}(a,b)\log\left(\frac{p_{A,B}(a,b)}{p_A(a)p_B(b)}\right)\\ &= D\left(p_{A,B}(a,b)\|p_A(a)p_B(b)\right)\\ &= \mathbb{E}_{A,B}\left\{\log\left(\frac{p_{A,B}(a,b)}{p_A(a)p_B(b)}\right)\right\}\end{aligned} \tag{32}$$

Informação mútua e entropia - relações importantes

1. Redução da incerteza de A devido ao conhecimento de B

$$\mathcal{I}(A,B) = \mathcal{H}(A) - \mathcal{H}(A|B) \tag{33}$$

2. Simetria da relação 1

$$\mathcal{I}(A,B) = \mathcal{H}(B) - \mathcal{H}(B|A) \tag{34}$$

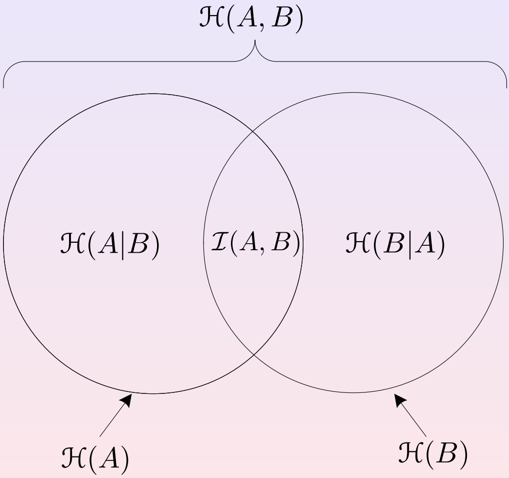
3. Soma de entropias

$$\mathcal{I}(A,B) = \mathcal{H}(A) + \mathcal{H}(B) - \mathcal{H}(A,B) \tag{35}$$

4. Auto-informação mútua

$$\mathcal{I}(A,A) = \mathcal{H}(A) - \mathcal{H}(A|A) = \mathcal{H}(A) \tag{36}$$

Informação mútua e entropia - relações importantes



Extensão de uma fonte discreta sem memória

- Utilização de *blocos* de dados, cada bloco com n símbolos da fonte
- Cada bloco pode ser entendido como sendo produzido por uma **fonte estendida**
- Alfabeto $\mathcal{A}^n$ com K^n blocos distintos, com K o número de símbolos na fonte original
- Considerando que os símbolos da fonte são estatisticamente independentes

$$P\left(s[A^n]\right) = \prod_{i=1}^{n} P\left(s_i[A]\right) \tag{37}$$

- Daí, podemos escrever então

$$\mathcal{H}(A^n) = n \cdot \mathcal{H}(A) \tag{38}$$

Definições importantes

- Informação condicional mútua de v.a. X e Y dado Z

$$\begin{aligned}\mathcal{I}(X,Y|Z) &= \mathcal{H}(X|Z) - \mathcal{H}(X|Y,Z) \\ &= \mathbb{E}_{p(x,y,z)}\left\{\log\left(\frac{p(X,Y|Z)}{P(X|Z)p(Y|Z)}\right)\right\}\end{aligned} \tag{39}$$

- Entropia relativa condicional

$$\begin{aligned}D\left(p(y|x)\|q(y|x)\right) &= \sum_x p(x)\sum_y p(y|x)\log\left(\frac{p(y|x)}{q(y|x)}\right) \\ &= \mathbb{E}_{X,Y}\left\{\log\left(\frac{p(y|x)}{q(y|x)}\right)\right\}\end{aligned} \tag{40}$$

- Entropia

$$\mathcal{H}(A) = -\int_{-\infty}^{\infty} p_A(a) \log\left(p_A(a)\right) \, da \tag{41}$$

- Divergência de Kullback-Leibler

$$D(p\|q) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) \, dx \tag{42}$$

Nota: Vamos estudar mais detalhes destas grandezas a seguir!

Entropia diferencial

- Embora a entropia definida por Shannon seja aplicada ao caso discreto, podemos expandir o conceito para variáveis contínuas
- A semelhança entre os casos discreto e contínuo é bastante grande, mas algumas diferenças são importantes e o uso de tal conceito merece cuidado
- Quando as variáveis são contínuas a entropia recebe o nome de **entropia diferencial**

Entropia diferencial - cont.

Definição

A entropia diferencial $\mathcal{H}(X)$ de uma variável aleatória contínua X com densidade de probabilidade $p_X(x)$ é definida por

$$\boxed{\mathcal{H}(X) = -\int_{\mathcal{S}} p_X(x) \cdot \log\left[p_X(x)\right] \, dx} \tag{43}$$

em que $\mathcal{S}$ é o conjunto suporte da v.a.

- Como no caso discreto, a entropia diferencial só depende da densidade de probabilidade, sendo por vezes escrita como $\mathcal{H}\left[p_X(x)\right]$ ao invés de $\mathcal{H}(X)$
- **Lembrete**: como em qualquer problema envolvendo integral ou densidade de probabilidade, nós precisamos garantir que elas existem.

Entropia diferencial - cont.

Exemplo - Distribuição uniforme

Seja uma v.a. distribuída uniformemente entre 0 e a, então sua densidade é $1/a$ entre 0 a a e 0 caso contrário. Então sua entropia diferencial é

$$\mathcal{H}(X) = -\int_0^a \frac{1}{a} \cdot \log\left[\frac{1}{a}\right]\, dx = \log(a) \qquad (44)$$

Note que, para $a < 1$, temos $\log(a) < 0$ e a entropia diferencial é negativa. Daí ao contrário da entropia discreta, a entropia diferencial pode ser negativa. Entretanto, $2^{\mathcal{H}(X)} = 2^{\log(a)} = a$ é o volume do conjunto suporte, o qual é sempre não-negativo, como esperado.

Entropia diferencial - cont.

Exemplo - Distribuição normal

Seja $X \sim \mathcal{N}(0, \sigma^2)$ em que denotamos $p_X(x) = \phi(x)$. Então, calculando a entropia diferencial em nats, temos

$$\begin{aligned}
\mathcal{H}\left[p_X(x)\right] &= -\int \phi(x) \ln[\phi(x)] \; dx \\
&= -\int \phi(x) \left[-\frac{x^2}{2\sigma^2} - \ln(\sqrt{2\pi\sigma^2}) \right] \; dx \\
&= \frac{E\{X^2\}}{2\sigma^2} + \frac{1}{2}\ln(2\pi\sigma^2) \\
&= \frac{1}{2} + \frac{1}{2}\ln(2\pi\sigma^2) \\
&= \frac{1}{2}\ln(e) + \frac{1}{2}\ln(2\pi\sigma^2) \\
&= \frac{1}{2}\ln(2\pi e\sigma^2) \quad \text{nats}
\end{aligned} \tag{45}$$

Definição

Seja um conjunto de N v.a. $X_1, X_2, \ldots, X_N$ com densidade $p_{\mathbf{X}}(\mathbf{x}) = p_{\mathbf{X}}(x_1, x_2, \ldots, x_N)$, a entropia diferencial é definida como

$$\boxed{\begin{aligned} \mathcal{H}\left[p_X(\mathbf{x})\right] &= -\int p_{\mathbf{X}}(\mathbf{x}) \cdot \log\left[p_{\mathbf{X}}(\mathbf{x})\right]\ d\mathbf{x} \\ &= -\int\int \cdots \int p_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N) \cdot \\ &\quad \cdot \log\left[p_{X_1,X_2,\ldots,X_N}(x_1, x_2, \ldots, x_N)\right]\ dx_1\ dx_2\ \ldots\ dx_N \end{aligned}} \tag{46}$$

Definição

Se X, Y têm uma função de densidade conjunta $p_{X,Y}(x, y)$, podemos definir a entropia diferencial condicional $\mathcal{H}(X|Y)$ como

$$\boxed{\mathcal{H}(X|Y) = -\int\int p_{X,Y}(x, y) \cdot \log\left[p_{X|Y}(x|y)\right]\ dx\ dy} \tag{47}$$

Uma vez que em geral $p_{X|Y}(x|y) = p_{X,Y}(x, y)/p_Y(y)$, podemos também escrever

$$\mathcal{H}(X|Y) = \mathcal{H}(X, Y) - \mathcal{H}(Y) \tag{48}$$

Deve-se entretanto garantir que nenhuma das entropias diferenciais seja infinita.

Regras da cadeia

Entropia

$$\mathcal{H}(A_1, A_2, \cdots, A_n) = \sum_{i=1}^{n} \mathcal{H}(A_i | A_{i-1}, A_{i-2}, \cdots, A_1) \tag{49}$$

Informação mútua

$$\mathcal{I}(A_1, A_2, \cdots, A_n; B) = \sum_{i=1}^{n} \mathcal{I}(A_i; B | A_{i-1}, A_{i-2}, \cdots, A_1) \tag{50}$$

Entropia relativa

$$D\left(p_{A,B}(a,b) \| q_{A,B}(a,b)\right) = D\left(p_A(a) \| q_A(a)\right) + D\left(p_{B|A}(b|a) \| q_{B|A}(b|a)\right) \tag{51}$$

1. $D(p\|g) \geq 0$
2. $\mathcal{I}(X, Y) \geq 0$ com igualdade se mantendo se e somente se X e Y são independentes
3. $\mathcal{H}(X|Y) \leq \mathcal{H}(X)$, com igualdade se mantendo se e somente se X e Y são independentes
4. $\mathcal{H}(X + c) = \mathcal{H}(X)$ - translação não altera entropia
5. $\mathcal{H}(cX) = \mathcal{H}(X) + \log(|c|)$
6. Para vetores e matrizes temos: $\mathcal{H}(\boldsymbol{C}\mathbf{X}) = \mathcal{H}(\mathbf{X}) + \log(|\boldsymbol{C}|)$, em que $|\boldsymbol{C}|$ é o determinante da matrix $\boldsymbol{C}$

Seja um vetor de N amostras aleatórias $\mathbf{X}$ formado de amostras independentes, ou seja,

$$p_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{N} p_{X_i}(x_i) \tag{52}$$

e seja um vetor $\mathbf{Y}$ definido em termos de $\mathbf{x}$ como $\mathbf{Y} = \boldsymbol{A}\mathbf{X}$, em que $\boldsymbol{A}$ é uma matriz não-diagonal. Seja $\widetilde{p}_{Y_i}(y_i)$ a densidade de probabilidade marginal de cada Y_i derivada a partir de $p_{\mathbf{Y}}(\mathbf{y})$. Então, a KLD entre $p_{\mathbf{X}}(\mathbf{x})$ e $p_{\mathbf{Y}}(\mathbf{y})$ admite a seguinte **decomposição Pitagórica**

$$\boxed{D(p_{\mathbf{Y}}||p_{\mathbf{X}}) = D(p_{\mathbf{Y}}||\widetilde{p}_{\mathbf{X}}) + D(\widetilde{p}_{\mathbf{Y}}||p_{\mathbf{X}})} \tag{53}$$

- Deriva da seguinte fórmula de função convexa

$$f\left(\lambda x_1 + (1-\lambda)x_2\right) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \quad (54)$$

- Exemplo de funções convexas: x^2, $|x|$, e^x, $x \log(x)$ para $x \geq 0$, etc
- *Inequação de Jensen*

$$\mathbb{E}\left\{f(X)\right\} \geq f\left(\mathbb{E}\left\{X\right\}\right) \quad (55)$$

- Deseja-se provar que $D(p\|q) \geq 0$, então tem-se

$$-D(p\|q) = -\sum_x p(x) \cdot \log\left(\frac{p(x)}{q(x)}\right) = \sum_x p(x) \cdot \log\left(\frac{q(x)}{p(x)}\right)$$

- Utilizando a inequação de Jensen

$$\sum_x p(x) \cdot \log\left(\frac{q(x)}{p(x)}\right) \leq \log\left(\sum_x p(x) \cdot \left(\frac{q(x)}{p(x)}\right)\right)$$

$$\begin{aligned}\log\left(\sum_x p(x) \cdot \left(\frac{q(x)}{p(x)}\right)\right) &= \log\left(\sum_x q(x)\right) \\ &= \log(1) \\ &= 0\end{aligned}$$

A entropia de Rényi, uma generalização da entropia de Shannon, é uma família de funcionais para quantificação da diversidade, incerteza ou aleatoriedade de um sistema.

Definição

A entropia de Rényi de ordem α, para $\alpha > 0$ é definida como

$$\boxed{\mathcal{H}_\alpha(X) = \frac{1}{1-\alpha}\log\left(\sum_{i=0}^{N-1} p_i^\alpha\right)} \tag{56}$$

em que p_i é a probabilidade do evento i.

Uma importante propriedade, é que se os eventos forem equiprováveis, então todas as entropias de Rényi (para qualquer α) são iguais para a distribuição com $\mathcal{H}_\alpha(X) = \log(N)$. Caso contrário, as entropias decrescem em função do α.

Alguns casos particulares

1. $\mathcal{H}_0(X) = \log(N)$ - é também chamada de **entropia de Hartley**
2. No limite quando $\alpha \to 1$ temos $\mathcal{H}_1(X) = -\sum_{i=0}^{N-1} p_i \log(p_i)$, que é a **entropia de Shannon**
3. Freqüentemente, a entropia de Rényi é dada para $\alpha = 2$ sendo
$$\mathcal{H}_2(X) = -\log\left(\sum_{i=0}^{N-1} p_i^2\right) \quad (57)$$
4. Para $\alpha \to \infty$ tem-se a **Min-entropia**, que é o menor valor de $\mathcal{H}_\infty(X)$ dada por
$$\mathcal{H}_\infty(X) = -\log\left(\sup_{i=1,\ldots,N} p_i\right) \quad (58)$$

Como a entropia de Rényi define o ganho de informação, há também uma medida para ganhos relativos de informação. Desta forma temos uma generalização da Divergência de Kullback-Leibler dada pela **Divergência Generalizada de Rényi** de ordem α

$$\boxed{D_\alpha(p||q) = \frac{1}{\alpha - 1} \log \left(\sum_{i=0}^{N-1} \frac{p_i^\alpha}{q_i^{\alpha-1}} \right)} \tag{59}$$

A exemplo da KLD, a divergência generalizada de Rényi é sempre não negativa.

Referência:

A. Rényi. "On measures of information and entropy". *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, 1960: 547-561.

Entropia usada na **termodinâmica**

$$\mathcal{H} = -k_B \sum_{\alpha} p_\alpha \log p_\alpha, \tag{60}$$

em que k_B é a constante de Boltzmann e p_α é a probabilidade do sistema estar no estado α.

A **entropia de Tsallis** é uma generalização da entropia de Boltzmann-Gibbs, que é a entropia da termodinâmica. Assim, a entropia de Tsallis é dada por

$$\mathcal{H}_q(p) = \frac{1}{q-1}\left(1 - \int p^q(x)\ dx\right) \tag{61}$$

ou, no caso discreto

$$\mathcal{H}_q(p) = \frac{1}{q-1}\left(1 - \sum p^q(x)\right) \tag{62}$$

Neste caso, p denota a densidade de probabilidade de interesse e q é um valor real. No limite quando $q \to 1$ obtém-se a entropia de Boltzmann-Gibbs

A entropia de von Neumann é utilizada para medir a informação média em densidade de *estados quânticos*. Um estado quântico representa as possibilidades de existência de uma partícula em diferentes estados, desta forma, representamos um estado quântico para dois estados possíveis, 0 e 1, por exemplo, como uma matriz de dimensão 2×2, ou seja,

$$\rho = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{63}$$

Com isso, define-se a entropia de von Neumann como

$$\boxed{\mathbb{H}(\rho) = \text{trace}\left[\rho \log(\rho)\right]} \tag{64}$$

a qual generaliza a entropia de Shannon.

Também temos a entropia quântica relativa, que generaliza a entropia relativa (divergência de Kullback-Leibler) para estados quânticos, a qual é definida como

$$\boxed{\mathbb{D}(\rho||\varrho) = \mathsf{trace}\left[\rho \log(\rho) - \rho \log(\varrho)\right]} \tag{65}$$

em que ρ e ϱ são matrizes Hermitianas positivas com traço igual a 1.

A idéia da entropia quântica (relativa ou de von Neumann) é a de mensurar a informação "espalhada" nos diversos estados quânticos.

Referência:
T.T. Georgiou, "Relative entropy and the multivariable multidimensional moment problem", *IEEE Trans. on Information Theory*, vol. 52, No. 3, pp. 1052-1066, March 2006.

A idéia da **entropia espectral** é a de analisar a informação no espectro do sinal. Utilizando-se da entropia de Shannon, substitui-se a densidade de probabilidade pela densidade espectral. Assim, podemos definir como

$$\boxed{\mathcal{H}_{sp}(P) = -\sum_{i=f_l}^{f_h} P_i \log(P_i)} \tag{66}$$

em que a faixa $[f_l, f_h]$ define a faixa de freqüência de interesse.

O interesse é de medir a quantidade de informação média no espectro a partir da informação contida em cada uma das componentes de freqüência. Encontra aplicação em áreas de biomédicas, por exemplo.

Referência:

R. Ferenets, T. Lipping, A. Anier, V. Jäntti, S. Melto, and S.Hovilehto, "Comparison of Entropy and Complexity Measures for the Assessment of Depth of Sedation", *IEEE Trans. on Biomedical Engineering*, vol. 53, No. 6, pp. 1067-1077, June 2006.

A taxa de informação de uma fonte é determinada a partir de suas entropia e taxa de transmissão.

Sendo uma fonte de informação $\mathcal{S}$ que transmite r símbolos a cada segundo, cujos símbolos são variáveis aleatórias a_s pertencentes a um alfabeto $\mathcal{A} = \{a_s : 1 \leq s \leq S\}$ com entropia $\mathcal{H}(\mathcal{A})$, define-se a taxa de informação R como sendo determinada pela equação:

$$R = r \cdot \mathcal{H}(\mathcal{A}). \tag{67}$$

- Entropia e informação mútua são importantes características de processos aleatórios com possibilidade de aplicação em várias áreas.
- Um problema inerente ao tratamento por teoria da estimação é como estimar tais quantidades uma vez que elas são funções da densidade de probabilidade, a qual é difícil de estimar a partir dos dados.
- Como fazer estimativas dos dados a partir apenas de medidas?
- **Estimadores**!!

Baseado na expansão de Gram-Charlier

- Expansão de Gram-Charlier: aproximação polinomial da densidade de probabilidade em torno de uma pdf gaussiana

$$p_X(x) = p_G(x)\left(1 + \sum_{k=3}^{\infty} C_k \cdot \mathfrak{h}_i(x)\right), \tag{68}$$

- Os coeficientes C_k serão funções dos momentos/cumulantes de X e $\mathfrak{h}_i$ é o polinômio de Hermite de ordem k

Baseado na expansão de Gram-Charlier - cont.

- Sabendo que podemos escrever a entropia como (será visto em maiores detalhes mais adiante!)

$$\mathcal{H}(p_X) = \mathcal{H}(p_G) - \mathcal{N}_G(p_X) \tag{69}$$

em que $\mathcal{N}_G(p_X)$ é a chamada negentropia.

- Pode-se escrever então

$$\mathcal{H}(p_X) = \mathcal{H}(p_G) - \int_V p_X(\mathbf{v}) \log \left[\frac{p_X(\mathbf{v})}{p_G(\mathbf{v})} \right] \, d\mathbf{v} \tag{70}$$

Baseado na expansão de Gram-Charlier - cont.

Então

$$\begin{aligned}
\mathcal{H}(p_X) &\approx \mathcal{H}(p_G) - \int_V p_G(\mathbf{v})\,(1+Z(\mathbf{v}))\log\left[1+Z(\mathbf{v})\right]\,d\mathbf{v} \\
&\approx \mathcal{H}(p_G) - \int_V p_G(\mathbf{v})\left[Z(\mathbf{v})+Z^2(\mathbf{v})\right]\,d\mathbf{v}
\end{aligned}$$

$$= \mathcal{H}(p_G) - \frac{1}{12}\left(\sum_{i=1}^{d}\left(\kappa^{i,i,i}\right)^2 + 3\sum_{\substack{i,j=1\\ i\neq j}}^{d}\left(\kappa^{i,i,j}\right)^2 + \frac{1}{6}\sum_{\substack{i,j=1\\ i\neq j\\ j\neq k}}^{d}\left(\kappa^{i,j,k}\right)^2\right) \tag{71}$$

em que $Z(\mathbf{v}) = \frac{1}{3!}\sum_{i,j,k}\kappa^{i,j,k}\mathfrak{h}_{ijk}(\mathbf{v})$, $\mathfrak{h}_{ijk}$ é o polinômio de ordem ijk e $\kappa^{i,i,i}$ é o momento de terceira ordem.

Baseado na expansão de Gram-Charlier - cont.

- É importante mencionar que foi feita uma aproximação em série, de segunda ordem, na expansão de Edgeworth (expansão de Gram-Charlier ordenada pela ordem de importância dos seus termos).
- Os momentos podem então ser estimados a partir das amostras
- Os polinômios de Hermite têm forma fechada e pode ser calculados a partir dos dados.
- Entretanto, as expansões de Edgeworth e Gram-Charlier só podem aproxima funções que são **“próximas”** a uma gaussiana

Baseado na estimação de Parzen

- O estimador de Parzen para pdfs aproxima a densidade (qualquer!) por um somatório de funções *kernel*. Dentre as classes de funções que podem ser consideradas *kernel* a função gaussiana é a mais conhecida
- Assim podemos ter

$$p_X(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N}\mathcal{K}(\mathbf{x}-\mathbf{x}_i,\sigma\mathbf{I}) \tag{72}$$

- Sabe-se ainda que, para dois *kernels* vale

$$\int_{\mathbf{x}}\mathcal{K}(\mathbf{x}-\mathbf{x}_i,\sigma_1\mathbf{I})\mathcal{K}(\mathbf{x}-\mathbf{x}_j,\sigma_2\mathbf{I}) = \mathcal{K}\left(\mathbf{x}_i-\mathbf{x}_j,(\sigma_1+\sigma_2)\mathbf{I}\right) \tag{73}$$

Baseado na estimação de Parzen - cont.

- Assim, teremos o seguinte aproximador para a entropia de Shannon

$$\begin{aligned}\mathcal{H}(p_X) &= -\int_{\mathbf{x}} p_X(\mathbf{x}) \cdot \log\left(p_X(\mathbf{x})\right)\ d\mathbf{x} \\ &= -\int_{\mathbf{x}} \sum_{i=1}^{N} \mathcal{K}(\mathbf{x}-\mathbf{x}_i) \cdot \log\left[\sum_{i=1}^{N} \mathcal{K}(\mathbf{x}-\mathbf{x}_i)\right]\end{aligned} \tag{74}$$

Baseado na estimação de Parzen - cont.

- Para a entropia de Rényi temos para o caso de $\alpha = 2$:

$$\begin{aligned}\mathcal{H}(p_X) &= -\log\left[\int_{\mathbf{x}} p_X^2(\mathbf{x})\ d\mathbf{x}\right] \\ &\approx -\log\left[\frac{1}{N^2}\int_{\mathbf{x}}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathcal{K}(\mathbf{x}-\mathbf{x}_i,\sigma\mathbf{I})\mathcal{K}(\mathbf{x}-\mathbf{x}_j,\sigma\mathbf{I})\ d\mathbf{x}\right] \\ &\approx -\log\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathcal{K}(\mathbf{x}_i-\mathbf{x}_j,2\sigma\mathbf{I})\right]\end{aligned} \tag{75}$$

- É então possível fazer desenvolvimento similares para a entropia condicional, relativa e informação mútua