

UNIVERSIDADE FEDERAL DE SANTA CATARINA

LIMITES SUPERIORES PARA A PROBABILIDADE DE ERRO E A DISTORÇÃO  
DE ERRO SOB O ESQUEMA GENERALIZADO COM DISTORÇÃO

José Ivanildo Coelho Dantas

Março - 1983

ESTA TESE FOI JULGADA ADEQUADA PARA A OBTENÇÃO DO TÍTULO

"MESTRE EM CIÊNCIAS"

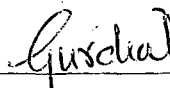
ESPECIALIDADE EM MATEMÁTICA, E APROVADA EM SUA FORMA FINAL PELO CURSO DE PÓS-GRADUAÇÃO EM MATEMÁTICA DA UNIVERSIDADE FEDERAL DE SANTA CATARINA.



---

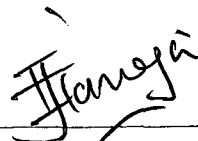
Prof. Inder Jeet Taneja, Ph.D.  
Coordenador

BANCA EXAMINADORA:



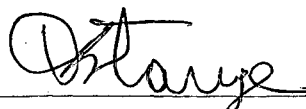
---

Prof. Gur Dial, Ph.D.  
Orientador



---

Prof. Inder Jeet Taneja, Ph.D.



---

Prof. Plinio Stange, Dr.

AGRADECIMENTOS

Ao Professor Gur Dial pela orientação segura e precisa, por sua disponibilidade e paciência, pelo entusiasmo que sempre procurou transmitir e pela sua amizade.

À D. Naninha, à Bentinha e aos meus irmãos pelas horas que deles tirei para poder realizar este trabalho.

Aos responsáveis pela minha formação, especialmente meus pais a quem muito devo.

Aos meus colegas do Curso de Pós-Graduação pelo carinho, incentivo e colaboração.

Estendo meus agradecimentos à Universidade Federal de Santa Catarina.

À minha esposa:

Sônia.

À minha filha :

Daniela.

### RESUMO

Neste trabalho propusemo-nos a determinar limites superiores para probabilidade de erro e distorção de erro usando o esquema generalizado com distorção.

No capítulo 1, apresentamos os resultados introdutórios que serão usados nos outros capítulos.

No capítulo 2, um esquema generalizado com distorção será introduzido e os limites superiores para probabilidade de erro serão obtidos. A função de confiança sob distorção será estudada.

No capítulo 3, distorção de erro será considerada e limites superiores para ela serão obtidos. A função de confiança será estudada.

ABSTRACT

In this work we propose to determine upper limits on probability of error and distortion due to error using the generalized scheme with distortion.

In chapter 1 we will present the introductory results which will be used in the following chapters.

In chapter 2 a generalized scheme with distortion will be introduced and upper limits to the probability of error will be obtained. The reliability function under distortion will be studied.

In chapter 3 distortion due to error will be considered and upper limits to it will be obtained. The reliability function will be studied.

SUMÁRIO

RESUMO .....	v
ABSTRACT .....	vi
CAPÍTULO 1 - RESULTADOS INTRODUTÓRIOS .....	1
1.1 - Introdução .....	1
1.2 - Preliminares .....	3
1.2.1 - Entropia de Shannon e informação mútua média .....	4
1.2.2 - Capacidade do canal e o teorema fundamental .....	6
1.2.3 - Teoria da taxa de distorção ....	7
1.2.4 - Esquemas de decodificação .....	9
1.2.5 - Probabilidade de erro .....	14
1.2.6 - Distorção de erro .....	16
1.2.7 - Limites sobre a probabilidade de erro e a distorção de erro .....	16
CAPÍTULO 2 - PROBABILIDADE DE ERRO SOBRE ESQUEMAS GENERALI- ZADOS ENVOLVENDO DISTORÇÃO .....	19
2.1 - Introdução .....	19
2.2 - Esquema de Decodificação Generalizada com Distorção .....	20
2.3 - Limites Sobre a Probabilidade de Erro Dentro do Esquema Generalizado .....	20
2.4 - Propriedades da Função Paramétrica de Confiança Sobre Distorção .....	28
CAPÍTULO 3 - LIMITES SUPERIORES PARA A DISTORÇÃO DE ERRO ..	42
3.1 - Introdução .....	42
3.2 - Propriedades da Função Confiança de Taxa de Distorção $E(R(D^*))$ .....	49
REFERÊNCIAS .....	57

## CAPÍTULO 1

### RESULTADOS INTRODUTÓRIOS

#### 1.1 - Introdução

Em 1948, Shannon na introdução do seu trabalho clássico, escreveu que o problema fundamental da comunicação é o da reprodução de um ponto exato ou aproximadamente na mensagem escolhida para um ponto qualquer. A fim de resolver este problema, ele criou um novo ramo da matemática aplicada, o qual é hoje chamado de teoria da informação. Passados 34 anos, a teoria da informação tem sido feita mais precisa, tem sido extendida e tem sido conduzida para um ponto onde ela é aplicada no prático sistema de comunicação. Como em toda teoria matemática, a teoria da informação trata somente de modelos e não de fonte e canais físicos.

É comum introduzir a teoria de informação como um ramo do estudo do "Sistema de Comunicação" cujo diagrama de bloco é dado como abaixo:



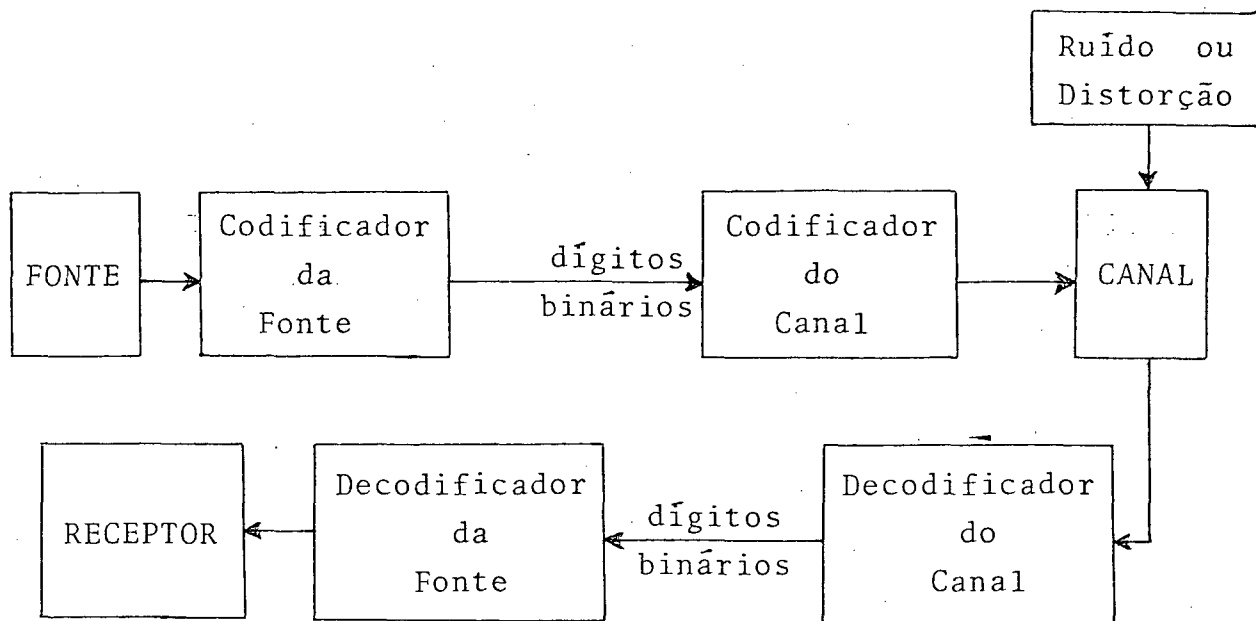


Figura 1

A fonte é o componente do sistema que gera a informação. Exemplos de fonte são: um autor de livros, um experimento científico, etc.. O codificador da fonte transforma a saída da fonte em dígitos binários. O codificador do canal é um meio de comunicação que está sujeito aos diferentes tipos de ruído. O decodificador do canal tenta transformar a saída do canal dentro do fluxo binário original e o decodificador da fonte tenta recriar o fluxo da fonte original. Esta separação de codificador (decodificador) dentro do codificador (decodificador) da fonte e canal tem óbvia vantagem do ponto de vista prático desde que os dados binários forneçam um tipo padrão de "interface" entre fontes e canais. De um ponto de vista conceptual esta separação é ainda mais importante desde que separe o problema da comunicação com ruído do problema da representação da fonte. Pela divisão de codificador e decodificador não estamos impondo quaisquer limitações sobre o sistema. Atualmente, um dos mais importantes

resultados da teoria, contudo, é que sob condições mais amplas tais limitações não são impostas (isto não mostra, contudo, que o codificador e o decodificador da figura é sempre o caminho mais econômico para se obter uma dada representação). De um ponto de vista prático a divisão de decodificador e codificador na figura é particularmente conveniente desde que faça o esquema do codificador e decodificador do canal realmente independente do codificador e decodificador da fonte usando dados binários como uma "interface". Isto facilita o uso de diferentes fontes sobre o mesmo canal.

O problema fundamental pode ser (Berger<sup>[2]</sup>) separado em dois, isto é

- i) Quanta informação será transmitida?
- ii) Que informação será transmitida?

O trabalho desenvolvido por Shannon<sup>[20]</sup> em 1948 está mais ligado ao primeiro problema, ou seja, o de selecionar codificadores de um conjunto de possíveis mensagens de tal maneira que elas possam ser transmitidas corretamente sobre um canal de comunicação com ruído. O segundo problema delineado acima permaneceu esquecido por algum tempo e é chamado "Teoria da Taxa de Distorção".

## 1.2 - Preliminares

No item anterior apresentamos nosso problema de uma maneira geral. Entretanto, precisamos formalizar as definições de certos entes aos quais já nos referimos e de outros que ainda serão citados tais como:

- Entropia de Shannon e informação mútua média;
- Capacidade do canal e o teorema fundamental;
- Teoria de taxa de distorção;
- Esquemas de decodificação.

### 1.2.1 - Entropia de Shannon e informação mútua média

Consideremos uma variável aleatória  $X$  discreta assumindo um número finito de valores

$$X = (x_1, \dots, x_I)$$

e associemos a esta variável aleatória uma distribuição de probabilidade

$$\underline{p} = (p(x_1), \dots, p(x_I)); p(x_i) \geq 0; \sum_{i=1}^I p(x_i) = 1 .$$

Então a entropia de Shannon<sup>[10]</sup> da distribuição de probabilidade  $\underline{p}$  é dada por

$$H(X) = H(p(x_1), \dots, p(x_I)) = - \sum_{i=1}^I p(x_i) \log p(x_i) ,$$

onde a base do logaritmo será "2" se estivermos usando a base binária na codificação das possíveis mensagens que a variável pode emitir. Como a base do sistema de codificação pode ser arbitrário, então a base do logaritmo no cálculo da entropia também o será. Na nossa dissertação utilizaremos sempre a base natural e a entropia será dada em nats.

Retornando ao nosso sistema discreto de comunicação

da Figura 1 observamos que as mensagens entram no canal codificadas em função do alfabeto de entrada  $X = (x_1, \dots, x_I)$  com uma distribuição  $p$  conhecida, saindo dele codificada em função do alfabeto de saída  $Y = (y_1, \dots, y_J)$  com uma distribuição  $q$ , digamos, desconhecida. Se conseguirmos a distribuição de  $Y/X$  (isto é,  $Y$  dado  $X$ ), ou seja, a matriz de transição do canal  $\{p(y_j/x_i)\}$   $i = 1, \dots, I$  e  $j = 1, \dots, J$ , automaticamente teremos condições de determinar as distribuições  $\{p(x_i, y_j)\}$ , da variável aleatória bidimensional  $(X, Y)$ ,  $\{q(y_j)\}$  da variável aleatória  $Y$ , e  $\{P(x_i/y_j)\}$  da variável aleatória  $X/Y$  como se segue:

$$p(x_i, y_j) = p(x_i) \cdot P(y_j/x_i) \quad \text{ou} \quad p(i, j) = p(i) \cdot P(j/i)$$

$$q(y_j) = \sum p(x_i, y_j) \quad \text{ou} \quad q(j) = \sum p(i, j)$$

$$P(x_i/y_j) = \frac{p(x_i, y_j)}{q(y_j)} \quad \text{ou} \quad P(i/j) = \frac{p(i, j)}{q(j)}$$

Podemos associar cinco diferentes entropias ao esquema, quais sejam:

i) Entropias marginais de  $X$  e  $Y$  por

$$H(X) = - \sum_{i=1}^I p(x_i) \log p(x_i) = - \sum_i p(i) \log p(i)$$

$$H(Y) = - \sum_{j=1}^J q(y_j) \log q(y_j) = - \sum_j q(j) \log q(j)$$

ii) Entropia conjunta de  $(X, Y)$

$$H(X, Y) = - \sum_{i=1}^I \sum_{j=1}^J p(x_i, y_j) \log p(x_i, y_j) =$$

$$= - \sum_i \sum_j p(i,j) \log p(i,j)$$

iii) Entropias condicionais X/Y e Y/X

$$H(X/Y) = - \sum_{i=1}^I \sum_{j=1}^J p(x_i, y_j) \log P(x_i/y_j) =$$

$$= - \sum_i \sum_j p(i,j) \log P(i/j)$$

$$H(Y/X) = - \sum_{i=1}^I \sum_{j=1}^J p(x_i, y_j) \log P(y_j/x_i) =$$

$$= - \sum_i \sum_j p(i,j) \log P(j/i).$$

A função do receptor é extrair, apesar do ruído, todas as informações possíveis sobre o sinal transmitido. A informação que Y providencia sobre X pode ser averiguada pela incerteza que Y remove sobre X, ou seja:

$$I(X,Y) = H(X) - H(X/Y) = H(Y) - H(Y/X).$$

Isto é simétrico em X e Y e é conhecido como informação mútua média. É fácil verificar que  $I(X,Y)$  é sempre não negativa (Ash<sup>[1]</sup>).

### 1.2.2 - Capacidade do canal e o teorema fundamental.

A capacidade do canal é, por definição, a taxa máxima sobre ele, ou seja:

$$C = \max_{p(x)} I(X,Y)$$

Como vimos acima, a maximização é feita com respeito a todas as possíveis escolhas das distribuições de entrada. A principal significância da capacidade aparece no teorema de codificação em canal com ruído de Shannon<sup>[10]</sup>.

Este teorema, que é o melhor resultado conseguido por

Shannon, estabelece que a transmissão de informação através de canais com ruído pode ser conseguida com probabilidade de erro arbitrariamente pequena quando a taxa de transmissão  $R$  é menor que a capacidade  $C$  do canal.

### 1.2.3 - Teoria da taxa de distorção.

Embora o princípio da teoria da taxa de distorção possa ser encontrado em Shannon<sup>[20]</sup>, trabalho de 1948, a principal mudança aconteceu em 1959<sup>[21]</sup>.

Considere uma fonte discreta sem memória a qual produz mensagens formadas de letras de um alfabeto  $U = \{u_1, \dots, u_I\}$ . As seqüências produzidas são transmitidas e reproduzidas pelo menos aproximadamente a um ponto receptor. Suponhamos, além disso, que as seqüências recebidas são formadas de elementos de um alfabeto  $V = \{v_1, \dots, v_J\}$ . O alfabeto  $V$  pode ser idêntico a  $U$  ou uma extensão de  $U$ . Para estudar a extensão da discordância e seus efeitos, Shannon introduziu a idéia de distorção entre as seqüências inicial e final de comunicação. A distorção pode ser considerada sobre palavras e letras.

Como primeira medida, que segue de Shannon, consideramos uma medida de distorção simples de letras. Seja  $d(u_i, v_j)$  a distorção quando  $u_i$  é transmitido e  $v_j$  é recebido. A quantidade  $d(u_i, v_j)$  pode ser considerada como uma perda, gasto ou prejuízo para esta alteração na letra  $u_i$  reproduzida assim como na letra  $v_j$ . Tomaremos esta quantidade igual a zero para cada reprodução correta. A distorção simples de letra  $d(u_i, v_j)$  é definida para todos os pares  $(u_i, v_j)$ ,  $i = 1, \dots, I$ ;  $j = 1, \dots, J$ . Se há um siste

ma de comunicação que reproduz  $u_i$  assim como  $v_j$  com probabilidade  $P_1(v_j/u_i)$ , então a distorção média obtida será

$$D = \sum_{i=1}^I \sum_{j=1}^J p(u_i) P_1(v_j/u_i) d(u_i, v_j).$$

A indicação da probabilidade condicional  $\{P_1(v_j/u_i)\}$  é dada por  $D^*$ , aceitável se, e somente, se  $D \leq D^*$ .

A função taxa de distorção da fonte relativa para a distorção média dada é definida como

$$R(D^*) = \min I(U, V), \\ \{P_1(v_j/u_i)\}$$

onde a minimização é feita com respeito a  $\{P_1(v_j/u_i)\}$  tal que  $D \leq D^*$ . Aqui  $D^*$  é denominado o nível de distorção.

Shannon<sup>[21]</sup> em 1959 definiu a função taxa de distorção de uma fonte de informação com respeito ao critério de fidelidade e estabeleceu o teorema fundamental que impregnava esta função com seu significado operacional. A função taxa de distorção  $R(D^*)$  é fundamental para todos os estudos da teoria da taxa de distorção. Ela representa a verdadeira taxa a que a fonte produz informação sujeita ao limite médio de tolerância  $D^*$  da usada. A taxa na qual uma fonte produz informação sujeita à necessidade de perfeita reprodução reduz a entropia da fonte e se a distorção média for tal que uma reprodução perfeita é determinada zero distorção, então  $R(0)$  é igual à entropia da fonte  $H(U)$ . Assim a função taxa de distorção é uma generalização da entropia sob as considerações qualitativas.

A teoria da taxa de distorção trata dos problemas "data de compressão" e "confiança de comunicação" permitindo distor

ção até certo nível. Nos sistemas de "data de compressão" somente a informação significativa é tirada da saída da fonte, eliminando o material irrelevante e redundante, possuindo um nível de fidelidade. Tal esquema é preciso para aumentar a taxa de transmissão.

O assunto da teoria de taxa de distorção já tem obtido importância e definido promessas para vastas aplicações de modelos de informação teórica na biologia e em outras ciências. As contribuições notáveis neste campo são as de Berger<sup>[2]</sup>, Blahut<sup>[3]</sup>, Gray<sup>[11]</sup>, Krich<sup>[14]</sup>, Leiner<sup>[15]</sup>, Omura<sup>[16]</sup>, Purseley<sup>[17]</sup>, Sakrison<sup>[19]</sup>, Shannon<sup>[20,21,22]</sup>, Wyner<sup>[28]</sup>, Zakai e Ziv<sup>[30]</sup> e outros.

#### 1.2.4 - Esquemas de decodificação.

Em muitos casos práticos, alguém que está recebendo uma mensagem fica com a responsabilidade de decidir, com base no sinal recebido, qual mensagem símbolo foi transmitida. O receptor fica então com um clássico problema de inferência estatística. Ele terá, com alguma base essencialmente subjetiva, que determinar a importância relativa dos vários tipos de erro que podem ocorrer. Somente então, em geral, pode ele fixar um esquema para decodificar, para cada símbolo recebido, qual mensagem símbolo ele terá, para melhor concluir o que foi emitido. Uma regra de decodificação pode ser definida como uma aplicação do conjunto de N-sequências de saída do canal no conjunto consistindo das mensagens emitidas. O objetivo da operação de decodificação é identificar a mensagem transmitida pela evidência verificada na N-sequência de saída. Para formular o problema mais precisamen-



te, definamos que o canal tem um alfabeto de entrada  $x_1, \dots, x_I$ , um alfabeto de saída  $y_1, \dots, y_j$  e uma matriz de transição  $\{P(y_j/x_i)\}$ . Suponhamos que alguma mensagem  $x$  é transmitida. O decodificador ou esquema de decisão é uma atribuição a cada símbolo de saída  $y_j$  de um símbolo de entrada  $x_i$ . Sejam  $X_N$  e  $Y_N$  denotando, respectivamente, o conjunto de todas as N-sequências de entrada e saída do canal. Vários critérios podem ser usados na decodificação e alguns são dados abaixo.

### i) Mínima Probabilidade de Erro

Seja  $P(\underline{y}/\underline{x}_m)$  a probabilidade de receber uma N-sequência  $\underline{y} \in Y_N$  dado que a N-sequência de entrada  $\underline{x}_m \in X_N$  é transmitida. A regra de decodificação de Mínima Probabilidade de Erro é justamente aquela que minimiza a probabilidade de decodificação para um dado conjunto de mensagens, conjunto de palavras código, e canal. Ela será então definida por: decodificar a N-sequência recebida  $\underline{y}$  em  $\underline{x}_m$ , para a qual

$$P(\underline{x}_{m'}/\underline{y}) > P(\underline{x}_m/\underline{y}), \text{ para todo } m' \neq m, 1 \leq m \leq M.$$

O esquema de decisão, que sempre escolhe  $\underline{x}$  cuja probabilidade condicional, no momento, é a maior, é sempre chamado de "Observador ideal".

### ii) Decodificação de Máxima-verossimilhança

O decodificador de Máxima-verossimilhança é um tipo alternativo de regra definido por: dado  $\underline{y}$ , escolher  $\underline{x}_m$ , tal que

$$P(\underline{y}/\underline{x}_{m'}) < P(\underline{y}/\underline{x}_m), \text{ para todo } m' \neq m, 1 \leq m \leq M.$$

A óbvia vantagem do decodificador de Máxima-verossimilhança é que ele pode ser usado quando as probabilidades das mensagens são desconhecidas a priori. Se as mensagens têm probabilidades a priori iguais, então a regra de decodificação de mínima probabilidade de erro é equivalente à de Máxima-verossimilhança.

### iii) Decodificação de Custo Mínimo

Um outro tipo de regra, usual quando custos desiguais são associados a diferentes espécies de erro, é a Decodificação de Custo Mínimo. Aqui  $y$  é decodificado numa mensagem  $m$  que minimiza o custo médio. Este tipo de regra de codificação é usado na teoria de codificação da fonte.

### iv) Decodificação em Lista

Algumas vezes é conveniente considerar em listagem, onde o decodificador, ao contrário da aplicação das seqüências recebidas em um simples inteiro, ou uma simples mensagem, o faz em uma lista de inteiros  $m$ ,  $1 \leq m \leq M$ , sendo  $M$  o número total de palavras código. Decodificação em Lista, ou em Rol, foi considerada primeiramente por Elias<sup>[6]</sup> para o canal simétrico binário. Shannon, Gallager e Berlekamp<sup>[22]</sup>, Ebert<sup>[5]</sup> e Forney<sup>[9]</sup> utilizaram o esquema de decodificação em Rol na obtenção dos limites sobre probabilidade de erro e de rasura.

### v) Decodificação Mutilada

Sharma e Raina<sup>[24]</sup> têm considerado o problema de Decodificação quando somente uma parte da mensagem transmitida é recebida e o receptor prepara sua decisão com base na palavra rece

bida incompleta, e têm obtido limites sobre probabilidade de erro. Para um método mais confiável de decisão é exigido que não mais que metade dos dígitos da mensagem transmitida sejam perdidos. Através da extensão de sua idéia de que o múltiplo da origem serve somente para traduzir um código, um limite superior sobre probabilidade média de erro de decodificação foi obtido por Sharma e Raina<sup>[25]</sup>.

Esse esquema de decodificação considera ambos "Mutilação" de dígitos e "erros aditivos". Através de Mutilação de um dígito queremos dizer que o dígito é desfigurado e não qualquer código de símbolos, e portanto está desprezado para o intento da decodificação. O termo "erro aditivo" é entendido no seu sentido usual da literatura.

vi) Decodificação de Máxima-verossimilhança com Distorção<sup>[23]</sup>

Seja  $d(x_k, y_j)$  um número não negativo, o qual pode ser considerado como perda ou distorção, com um par  $(x_k, y_j)$  de entrada e saída. Deve ser claro que a palavra distorção usada aqui é de forma conceptual diferente da introduzida por Shannon<sup>[21]</sup>. É de fato uma medida de mesma classe de perda, ou penalidade associada quando a saída é considerada com respeito à entrada. Uma distorção sobre o par  $(\underline{x}, \underline{y})$  das palavras entradas e saídas que são seqüências de comprimento  $N$  pode similarmente ser definida. Um método utilizado para definir a distorção  $d(\underline{x}, \underline{y})$ , sobre palavras, é em termos de distorções de letras simples das várias letras dele. Um método simples para fazer isto é tomar uma média de distorção de letras simples de modo que se  $\underline{x} = (x_1, \dots, x_N)$  e  $\underline{y} = (y_1, \dots, y_N)$  então

$$d(\underline{x}, \underline{y}) = N^{-1} \sum_{n=1}^N d(x_n, y_n) \quad (1.2.4.1)$$

O esquema de decodificação que consideraremos agora é uma generalização do esquema de decodificação de Máxima-verossimilhança.

O novo esquema para decodificação é chamado "Esquema de Máxima-verossimilhança com Distorção" e de acordo com isto o decodificador decodifica a seqüência de saída  $\underline{y}$  dentro da seqüência de entrada  $\underline{x}_m$  se

$$\frac{P(\underline{y}/\underline{x}_m)}{d(\underline{x}_m, \underline{y})} > \frac{P(\underline{y}/\underline{x}_{m'})}{d(\underline{x}_{m'}, \underline{y})}, \text{ para todo } m' \neq m, 1 \leq m' \leq M \quad (1.2.4.2)$$

Claramente, a decodificação de erro ocorre se enviado  $\underline{x}_m$ , a condição (1.2.4.2) não é satisfeita.

Para evitar evidentes dificuldades que surgem supondo valor zero de distorção no esquema de decodificação, tomaremos  $d(\underline{x}, \underline{y}) > 0$ . Querendo dizer por meio disto que para uma reprodução perfeita da mensagem transmitida, existe algum valor positivo associado, este valor mínimo pode ser interpretado como custo de reprodução. Também é claro que se ignorarmos a distorção, isto é, se tomarmos  $d(\underline{x}_m, \underline{y}) = d(\underline{x}_{m'}, \underline{y})$  para todo  $m$ , o esquema reduz-se ao esquema de Máxima-verossimilhança.

#### vii) Esquema Generalizado

Dados dois números positivos  $\alpha$  e  $\beta$ ,  $\alpha \geq \beta$ , o decodificador decodifica a seqüência de saída  $\underline{y}$  no inteiro  $m$  se

$$P(\underline{y}/\underline{x}_m)^\alpha > P(\underline{y}/\underline{x}_{m'})^\beta$$

para todo  $m' \neq m$ ,  $1 \leq m' \leq M$ .

Claramente quando  $\alpha = \beta$ , o esquema acima reduz-se ao esquema de decodificação de Máxima-verossimilhança. Este esquema depende somente das probabilidades condicionais e dos dois parâmetros. Se levarmos também em consideração a distorção, podemos considerar um esquema dependendo das probabilidades condicionais, distorções e dos dois parâmetros. Um tal esquema é considerado nos Capítulos 2 e 3.

#### 1.2.5 - Probabilidade de erro.

Como tomado por Shannon<sup>[20]</sup>, Gallager<sup>[10]</sup> e outros, consideraremos um canal discreto com alfabeto de entrada  $X = (x_1, \dots, x_K)$ , alfabeto de saída  $Y = (y_1, \dots, y_J)$  e matriz transição  $\{P(y_j/x_i)\}$ ,  $i = 1, \dots, k$ ,  $j = 1, \dots, J$ . Sejam  $\underline{X}_N$  e  $\underline{Y}_N$  os conjuntos de todas as seqüências de comprimento  $N$  que podem ser transmitidas e recebidas respectivamente sobre um dado canal. E seja  $P(\underline{y}/\underline{x})$  a probabilidade de receber  $\underline{y} \in \underline{Y}_N$  dado que  $\underline{x} \in \underline{X}_N$  foi transmitido.

Para um código bloco de comprimento  $N$  e taxa  $R$ , tal canal possui um conjunto de  $e^{NR}$  seqüências de entrada ou palavras código de comprimento  $N$  (Forney<sup>[9]</sup>, Gallager<sup>[10]</sup>, Fano<sup>[7]</sup>) como vemos abaixo

$$\underline{x}_m = \{x_{mi}\} = (x_{m1}, \dots, x_{mN}), \quad 1 \leq m \leq e^{NR},$$

onde  $R$  é a taxa de código em nats por símbolo do canal.

Um codificador é um esquema mecânico que admite um dos  $e^{NR}$  comandos de uma fonte de dados e gera a correspondente se-

seqüência de entrada a ser transmitida pelo canal. Um decodificador é um ente que observa uma seqüência de saída de comprimento  $N$ , processa esta seqüência e apresenta o resultado para usar na forma desejada. O evento no qual o estimador não é idêntico à palavra código de entrada é chamado um erro e a probabilidade deste evento é a probabilidade de erro.

A probabilidade de erro  $P(e)$  depende do código, do canal e da estratégia usada pelo decodificador. Se o decodificador é determinístico então sua estratégia é descrita como uma aplicação do conjunto de todas as seqüências recebidas  $\underline{y}$  na palavra código  $\underline{x}_m$  e é especificada pela listagem do conjunto  $Y_m$  de seqüências  $\underline{y}$  que resultam no estimador decodificado de  $\underline{x}_m$ . Se supusermos que código, canal e decodificador são todos especificados, então a probabilidade de erro (Forney<sup>[9]</sup>) será dada por

$$P(e) = \sum_{m=1}^M p(\underline{x}_m) P(\underline{y} \in Y_m / \underline{x}_m \text{ é transmitido})$$

$$= \sum_{m=1}^M \sum_{\underline{y} \in Y_m} p(\underline{x}_m) P(\underline{y}/\underline{x}_m)$$

Suponhamos que, se o ruído é particularmente nocivo e o decodificador tem a opção de não decidir sobre todos os estimadores, então o resultado da saída que o receptor não estima é chamado uma rasura (Forney<sup>[9]</sup>) e a probabilidade deste evento é a probabilidade de rasura.

### 1.2.6 - Distorção de erro.

Um esquema envolvendo somente probabilidade poderia conduzir somente para analisar a probabilidade de erro verificando a confiança do método. Por outro lado um esquema que usa Máxima-verossimilhança com distorção também pode ser usado para determinar alguma outra quantidade na qual a maximização significaria uma decodificação eficiente.

Tal medida é o que chamamos de "Distorção de Erro" e a sua soma é determinada pela razão  $\frac{P(\underline{y}/\underline{x}_m)}{d(\underline{x}_m, \underline{y})}$  para palavras decodificadas falsamente na comunicação.

Temos apontado que o esquema de Máxima-verossimilhança com Distorção reduz o esquema de decodificação de Máxima-verossimilhança quando as distorções são ignoradas, isto é, se considerarmos que distorções ou custos são os mesmos a cada momento. E nisto, eventualmente, nossa distorção de erro, definida na secção seguinte, reduz a probabilidade de erro. Assim distorção de erro é em algum sentido uma generalização do conceito de probabilidade de erro. Este fato surge de acordo com o esquema de decodificação de Máxima-verossimilhança com Distorção e é provavelmente uma medida mais apropriada para minimizar e aperfeiçoar a credibilidade da comunicação.

### 1.2.7 - Limites sobre a probabilidade de erro e a distorção de erro.

A avaliação exata da probabilidade de erro é muito difícil de se executar, em geral. O problema da obtenção de limi-

tes sobre a probabilidade de erro aparece com o Teorema de Codificação, o qual determina que probabilidade de erro pequena pode ser obtida para taxas abaixo da capacidade do canal. Para canais discretos sem memória, a forma mais forte do Teorema foi provada por Fano<sup>[7]</sup> em 1961 juntamente com limites sobre a Máxima Probabilidade de Erro  $P_e$  para códigos bloco de comprimento  $N$  e para quaisquer taxas abaixo da capacidade entre os limites

$$e^{-N[E_L(R) + o(N)]} \leq P_e \leq 2e^{-N E(R)}, \quad (1.2.7.1)$$

onde  $E_L(R)$  e  $E(R)$  são funções positivas das probabilidades de transição do canal e da taxa  $R$ ;  $o(N)$  é uma função que tende a zero com  $N$  crescente.

Vários limites têm sido obtidos sobre a probabilidade de erro por Shannon<sup>[20]</sup>, Fano<sup>[7]</sup>, Feinstein<sup>[8]</sup>, Reiffen<sup>[18]</sup>, Gallager<sup>[10]</sup>, Jelinek<sup>[13]</sup>, Elias<sup>[6]</sup>, Wolfowitz<sup>[27]</sup> e outros.

Yudkin<sup>[29]</sup> em 1967 obteve limites para probabilidade de erro para determinados canais finitos.

Stiglitz<sup>[26]</sup> obteve o limite superior para probabilidade de erro para uma classe de canais desconhecidos.

Removendo a suposição de que o receptor já tem feito a tempo a informação requerida para codificar a versão recebida com ruído da palavra código transmitida, Chase<sup>[4]</sup> obteve formas do Teorema de Codificação juntamente com limites para probabilidade de erro.

A técnica usada na obtenção dos limites superiores para probabilidade de erro é o argumento da codificação ao acaso (Fano<sup>[7]</sup>), a qual é baseada sobre o mesmo argumento usado por Shannon em sua demonstração original do Teorema de Codificação.



É conveniente saber (ver Gallager<sup>[10]</sup>) que, se a informação é transmitida sobre um dado canal de uma dada fonte, e se a entropia da fonte por unidade de tempo é tão grande quanto a capacidade do canal por unidade de tempo, então a recepção arbitrariamente segura da fonte não é possível. Isto é chamado o inverso do Teorema de Decodificação.

Feinstein<sup>[8]</sup> demonstrou para canais discretos sem memória, que se  $R > C$ , a probabilidade de erro não pode aproximar-se de zero, mas é sempre limitada longe de zero. Isto é chamado o inverso fraco do Teorema de Decodificação desde que Wolfowitz<sup>[27]</sup> demonstrou o lado forte chamado inverso forte do Teorema de Decodificação o qual determina que a probabilidade de erro aproxima-se da unidade a medida que o bloco de comprimento  $N$  tende ao infinito.

Shannon, Gallager e Berlekamp<sup>[22]</sup> obtiveram limites inferiores para a mínima probabilidade de erro, que pode ser obtida através do uso da codificação bloco sobre canais discretos sem memória com ruído. Estes limites inferiores são do mesmo modo que os limites superiores, decrescentes exponencialmente com o comprimento do bloco.

No Capítulo 2, determinaremos o limite superior para a probabilidade de erro, usando o esquema generalizado de máxima-verossimilhança com distorção. A função confiança também será estudada.

No Capítulo 3, estudaremos o problema de distorção de erro e limite superior, usando o esquema introduzido no Capítulo 2. A função confiança com distorção também será estudada.

## CAPÍTULO 2

### PROBABILIDADE DE ERRO SOBRE ESQUEMAS GENERALIZADOS ENVOLVENDO DISTORÇÃO

#### 2.1 - Introdução

No Capítulo 1, seção 1.2.5, introduzimos o conceito de probabilidade de erro. Como vimos lá, consideramos um canal discreto sem memória com alfabeto de entrada  $X = (x_1, \dots, x_K)$ , alfabeto de saída  $Y = (y_1, \dots, y_J)$  e uma matriz de probabilidade  $\{P(y_j/x_k)\}$ ,  $k = 1, \dots, K$ ,  $j = 1, \dots, J$ . Denotaremos por  $X_N$  e  $Y_N$  o conjunto de todas as seqüências de entrada e saída de comprimento  $N$  que são respectivamente transmitidas e recebidas. Se  $R$  é a taxa de informação, então  $M$ , o número total de palavras código de comprimento  $N$ , é dado por  $M = e^{NR}$  (consideramos aqui a base natural).

Usando o esquema de Máxima-verossimilhança, Gallager<sup>[10]</sup> deu uma demonstração bem simples e elegante de Teorema fundamental da teoria da informação e deu também um limite superior para a probabilidade de erro. Trabalhos de outras pessoas nesta direção foram mencionados no Capítulo 1. Na seção 1.2.4 nos referimos ao trabalho de Sharma e Raina<sup>[25]</sup> que usaram o esquema generalizado de decodificação, isto é, dados dois números positivos  $\alpha$  e  $\beta$ ,  $\alpha \geq \beta$  o decodificador decodifica a seqüência recebida  $\underline{y} \in Y_N$  no inteiro  $m$  ( $x_m \in X_N$ ) se

$$P^\alpha(\underline{y}/x_m) > P^\beta(\underline{y}/x_{m'}), \text{ para todo } m' \neq m, 1 \leq m' \leq M.$$

Este estudo não leva em conta a qualidade da informação processada através do canal.

Na seção seguinte definiremos um esquema generalizado com distorção, e na seção 2.3 obteremos limites para a probabilidade de erro.

## 2.2 - Esquema de Decodificação Generalizada com Distorção

Dado dois números adequados  $\alpha$  e  $\beta$ ,  $\alpha \geq \beta$ , o decodificador decodifica a seqüência de saída  $y$  no inteiro  $m$  se

$$\left(\frac{P(\underline{y}/\underline{x}_m)}{d(\underline{x}_m, \underline{y})}\right)^\alpha > \left(\frac{P(\underline{y}/\underline{x}_{m'})}{d(\underline{x}_{m'}, \underline{y})}\right)^\beta \text{ para todo } m' \neq m, 1 \leq m \leq M \quad (2.2.1)$$

Claramente, quando  $\alpha = \beta$ , o esquema acima reduz-se ao esquema de decodificação de Máxima-Verossimilhança com distorção. Para evitar dificuldades que podem surgir se tomarmos o valor zero de distorção no esquema, tomaremos  $d(x, y) > 0$ . Na seção seguinte, obteremos o limite superior sobre a probabilidade de erro quando a decodificação é feita usando o esquema (2.2.1).

## 2.3 - Limites Sobre a Probabilidade de Erro Dentro do Esquema Generalizado

Denotaremos  $P_{em}$  a probabilidade de erro quando  $\underline{x}_m \in X_N$  é transmitido. Tomaremos uma situação na qual  $m$  não satisfaz

(2.2.1) como uma decodificação de erro. Consideraremos também que uma decodificação é feita se o inteiro decodificado é diferente do inteiro de entrada. Em vista desta definição podemos expressar  $P_{em}$  como

$$P_{em} = \sum_{\underline{y} \in Y_N} P(\underline{y}/\underline{x}_m) \Psi_m(\underline{y}) \quad (2.3.1)$$

onde

$$\Psi_m(\underline{y}) = \begin{cases} 1, & \text{se } \left(\frac{P(\underline{y}/\underline{x}_m)}{d(\underline{x}_m, \underline{y})}\right)^\alpha \leq \left(\frac{P(\underline{y}/\underline{x}_{m'})}{d(\underline{x}_{m'}, \underline{y})}\right)^\beta \\ & \text{para todo } m' \neq m \\ 0, & \text{caso contrário.} \end{cases} \quad (2.3.2)$$

Suponhamos que nenhuma distorção entre as letras excede um número  $\rho_0 > 0$ . Esta suposição é, de fato, natural na maioria das situações práticas. A probabilidade média de erro  $P_e$  é média de  $P_{em}$  sobre todas as palavras código. Um limite superior exponencial será obtido usando a técnica de Gallager<sup>[10]</sup>. Esta técnica em princípio, determina um limite superior apropriado para a função  $\Psi_m(\underline{y})$ . Também temos  $d(\underline{x}, \underline{y}) = \frac{1}{N} \sum_{n=1}^N d(x_n, y_n)$  onde  $d(x_n, y_n)$  é uma distorção entre as letras de  $\underline{y}_m$  (seqüência recebida) e  $\underline{x}_m$  (transmitida).

### Teorema 2.3.1:

Suponha que um canal discreto sem memória com um alfabeto de entrada  $(x_1, \dots, x_K)$  e um alfabeto de saída  $(y_1, \dots, y_J)$  é descrito pela probabilidade de transição  $\{P(y_j/x_k)\}$ ,  $j = 1, 2, \dots, J$ ,  $k = 1, 2, \dots, K$ . Então para algum bloco de comprimento  $N$  e uma fonte com  $M = \lfloor e^{NR} \rfloor$  palavras código, existe um código para o qual

a probabilidade de erro  $P_e$ , usando o esquema (2.2.1), é limitada por

$$P_e \leq \rho_0 \exp - N[-\rho R + E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})] \quad (2.3.4)$$

$$E_{\alpha, \beta}(\rho, \underline{p}, \underline{d}) = -\ln \sum_k \left( \sum_j p_k \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{(\beta/\alpha)(1 + (\beta/\alpha)\rho)^{-1}} \left( 1 + (\beta/\alpha)\rho \right) \quad (2.3.5)$$

onde  $\rho$  é um número arbitrário,  $0 \leq \rho \leq 1$ .

Observação:

Se a distorção é ignorada então temos o resultado<sup>[25]</sup>.

Demonstração:

Limitaremos superiormente a função  $\Psi_m(\underline{y})$ :

$$\Psi_m(\underline{y}) \leq \left( \frac{\sum_{m' \neq m} \left( \frac{P(\underline{y}/x_{m'})}{d(x_{m'}, \underline{y})} \right)^{\beta/(\alpha + \beta\rho)}}{\left( \frac{P(\underline{y}/x_m)}{d(x_m, \underline{y})} \right)^{\alpha/(\alpha + \beta\rho)}} \right)^{(\beta/\alpha)\rho}, \quad \rho > 0 \quad (2.3.6)$$

Substituindo (2.3.6) em (2.3.1), temos

$$P_{em} \leq \sum_{\underline{y} \in Y_N} d(x_m, \underline{y}) \left( \frac{P(\underline{y}/x_m)}{d(x_m, \underline{y})} \right)^{\alpha/(\alpha + \beta\rho)} \left( \sum_{m' \neq m} \left( \frac{P(\underline{y}/x_{m'})}{d(x_{m'}, \underline{y})} \right)^{\beta/(\alpha + \beta\rho)} \right)^{(\beta/\alpha)\rho}$$

$$\leq \sum_{\underline{y} \in Y_N} \rho_0 \left( \frac{P(\underline{y}/x_m)}{d(x_m, \underline{y})} \right)^{\beta/(\alpha + \beta\rho)} \left( \sum_{m' \neq m} \left( \frac{P(\underline{y}/x_{m'})}{d(x_{m'}, \underline{y})} \right)^{\beta/(\alpha + \beta\rho)} \right)^{(\beta/\alpha)\rho}$$

$$\text{Porque } d(\underline{x}_m, \underline{y}) = \frac{1}{N} \sum_{n=1}^N d(x_n, y_n) \leq \rho_0 \quad (2.3.7)$$

A inequação (2.3.7) garante um limite para  $P_{em}$  para um código particular. O limite para  $P_{em}$  pode ser simplificado por averiguação sobre um conjunto apropriadamente escolhido de códigos. Pelo menos um código no conjunto terá uma probabilidade de erro que é tão pequena quanto o conjunto-média de probabilidade de erro. Definamos uma medida de probabilidade  $P(\underline{x})$  sobre o conjunto  $X_N$  de possíveis seqüências de entrada para o canal de modo que o conjunto de códigos é gerado escolhendo cada palavra código, independentemente, de acordo com a medida de probabilidade  $P(\underline{x})$ . Um código consistindo das palavras  $x_1, \dots, x_m$  terá a probabilidade  $\prod_{m=1}^M P(x_m)$  no conjunto. Usando uma barra para representar a média dos conjuntos de códigos, tomando  $\rho \leq 1$ , obtemos

$$\bar{P}_{em} \leq \rho_0 \sum_{\underline{y} \in Y_N} \left( \frac{P(\underline{y}/\underline{x}_m)}{d(\underline{x}_m, \underline{y})} \right)^{\beta/(\alpha + \beta\rho)} \left( \sum_{m' \neq m} \left( \frac{P(\underline{y}/\underline{x}_{m'})}{d(\underline{x}_{m'}, \underline{y})} \right)^{\beta/(\alpha + \beta\rho)} \right)^{(\beta/\alpha)\rho} \quad (2.3.8)$$

Imporemos agora a restrição adicional que  $\rho < 1$ . Então (ver os passos tomados em Gallager<sup>[10]</sup>), obtemos

$$\bar{P}_{em} \leq \rho_0 \sum_{\underline{y} \in Y_N} \left( \frac{P(\underline{y}/\underline{x}_m)}{d(\underline{x}_m, \underline{y})} \right)^{\beta/(\alpha + \beta\rho)} \left( \sum_{m' \neq m} \left( \frac{P(\underline{y}/\underline{x}_{m'})}{d(\underline{x}_{m'}, \underline{y})} \right)^{\beta/(\alpha + \beta\rho)} \right)^{(\beta/\alpha)\rho} \quad (2.3.9)$$

Mas, desde que as palavras código são escolhidas com a probabilidade  $p(\underline{x})$ ,

$$\left( \frac{P(\underline{y}/\underline{x}_m)}{d(\underline{x}_m, \underline{y})} \right)^{\beta/(\alpha + \beta\rho)} = \sum_{\underline{x} \in X_N} p(\underline{x}) \left( \frac{P(\underline{y}/\underline{x})}{d(\underline{x}, \underline{y})} \right)^{\beta/(\alpha + \beta\rho)} \quad (2.3.10)$$

Portanto, em vista de (2.3.10), (2.3.9) dá

$$\bar{P}_{em} \leq \rho_0 (M-1)^\rho \sum_{\underline{y} \in Y_N} \left[ \sum_{\underline{x} \in X_N} p(\underline{x}) \left( \frac{P(\underline{y}/\underline{x})}{d(\underline{x}, \underline{y})} \right)^{\beta/\alpha + \beta\rho} \right]^{1 + (\beta/\alpha)\rho} \quad (2.3.11)$$

$$= \rho_0 (M-1)^\rho \sum_{\underline{y} \in Y_N} \left[ \sum_{\underline{x} \in X_N} p(\underline{x}) \left( \frac{P(\underline{y}/\underline{x})}{d(\underline{x}, \underline{y})} \right)^{(\beta/\alpha)(1 + (\beta/\alpha)\rho)^{-1}} \right]^{1 + (\beta/\alpha)\rho}$$

para qualquer  $\rho$ ,  $0 \leq \rho \leq 1$  (2.3.12)

Este limite é válido para todas as escolhas de  $p(\underline{x})$  e todo  $\rho$ ,  $0 \leq \rho \leq 1$  e aplicável sobre qualquer canal discreto. Simplificaremos este quando o canal é sem memória. Isto é, se  $\underline{x} = (x_1, \dots, x_N)$  e  $\underline{y} = (y_1, \dots, y_N)$ , então

$$P(\underline{y}/\underline{x}) = \prod_{n=1}^N P(y_n/x_n) \quad (2.3.13)$$

Para todo  $\underline{x} \in X_N$ ,  $\underline{y} \in Y_N$  e todo  $N$ .

Consideraremos agora somente a classe de conjunto de códigos na qual cada letra de cada palavra código é escolhida independentemente de todas as outras letras com uma medida de probabilidade  $p(x)$ ;  $\underline{x} \in X$

$$P(\underline{x}) = \prod_{n=1}^N p(x_n) \quad (2.3.14)$$

Usando (2.3.13), (2.3.14) e a inequação de média aritmética e geométrica, isto é,  $d(\underline{x}, \underline{y}) = \frac{1}{N} \sum_{n=1}^N d(x_n, y_n) > \left( \prod_{n=1}^N d(x_n, y_n) \right)^{1/N}$ , temos

$$\bar{P}_{em} \leq \rho_0 (M-1)^\rho \sum_{\underline{y} \in Y_N} \left[ \sum_{\underline{x} \in X_N} \prod_{n=1}^N p(x_n) \left( \frac{P(y_n/x_n)}{d(x_n, y_n)^{\frac{1}{N}}} \right)^{(\beta/\alpha)(1 + (\beta/\alpha)\rho)^{-1}} \right]^{1 + (\beta/\alpha)\rho} \quad (2.3.15)$$

$$= \rho_0 (M-1)^\rho \sum_{\underline{y} \in Y_N} \left[ \prod_{n=1}^N \sum_{\underline{x} \in X_n} p(x_n) \left( \frac{P(y_n/x_n)}{d(x_n, y_n)^{\frac{1}{N}}} \right)^{(\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}} \right]^{1 + (\beta/\alpha)\rho} \quad (2.3.16)$$

O resultado (2.3.16) segue de (2.3.15), porque o termo do colchete em (2.3.16) é um produto de somas e é igual ao termo do colchete em (2.3.15) por uso comum da regra aritmética para multiplicação de produtos e somas. Tomando o produto fora do colchete em (2.3.16), aplicamos novamente a mesma regra e obtemos

$$\bar{P}_{em} \leq \rho_0 (M-1)^\rho \prod_{n=1}^N \sum_{y_n \in Y} \left[ \sum_{x_n \in X} p(x_n) \left( \frac{P(y_n/x_n)}{d(x_n, y_n)^{\frac{1}{N}}} \right)^{(\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}} \right]^{1 + (\beta/\alpha)\rho} \quad 0 \leq \rho \leq 1. \quad (2.3.17)$$

Simplificaremos agora a notação em (2.3.17) por observação de que  $X$  é o conjunto das letras de entrada  $x_1, \dots, x_K$  e  $Y$  é o conjunto das letras de saída  $y_1, \dots, y_J$ . Notando aqui que todos os termos no produto são idênticos, e incluindo o caso trivial  $\rho = 0$ , temos

$$\bar{P}_{em} \leq \rho_0 (M-1)^\rho \left[ \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{(\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}} \right)^{1 + (\beta/\alpha)\rho} \right]^N \quad 0 \leq \rho \leq 1 \quad (2.3.18)$$

Agora, se limitarmos superiormente  $M - 1$  por  $M = e^{NR}$ , (2.3.18) pode ser reescrito como

$$\bar{P}_{em} \leq \rho_0 \exp -N \{-\rho R - \ln \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{(\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}} \right)^{1 + (\beta/\alpha)\rho}\} \quad (2.3.19)$$



$$= \rho_0 \exp - N [- \rho R + E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})] \quad (2.3.20)$$

onde

$$E_{\alpha, \beta}(\rho, \underline{p}, \underline{d}) = - \ln \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) \left( \frac{P(y_j, x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha} (1 + (\beta/\alpha)\rho)^{-1} \right)^{1 + (\beta/\alpha)\rho}$$

Desde que o lado direito de (2.3.20) é independente de  $m$ , este é um limite para o conjunto probabilidade média de erro de decodificação e é independente das probabilidades com as quais as palavras são usadas. Assim, existe pelo menos um código no conjunto que precisa ter uma probabilidade de erro tão pequena como a média.

#### Corolário 2.3.1

Sob as condições do Teorema 2.3.1, existe um código para o qual

$$P_e \leq \rho_0 \exp [- N E_{\alpha, \beta}(\underline{R}, \underline{d})] \quad (2.3.21)$$

onde

$$E_{\alpha, \beta}(\underline{R}, \underline{d}) = \max_{\rho, \underline{p}} [- \rho R + E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})]. \quad (2.3.22)$$

Também pode ser obtida uma modificação suplementar do resultado para ter um limite para a probabilidade de erro que se aplique para cada palavra código separadamente melhor do que para a média.

#### Corolário 2.3.2

Sob as condições colocadas no teorema 2.3.1, existe

um código tal que para todo  $m$ ,  $1 \leq m \leq M$ , a probabilidade de erro, quando a  $m$ -ésima palavra código é transmitida, é limitada por

$$P_{em} \leq \rho_0 \cdot 4 \exp [-NE_{\alpha,\beta}(R,\underline{d})] \quad (2.3.23)$$

onde  $E_{\alpha,\beta}(R,\underline{d})$  é dada por (2.3.22).

Demonstração:

Escolhamos um código com  $M' = 2M$  palavras código o qual satisfaz o Corolário 2.3.1 quando a fonte usa  $2M$  palavras código com probabilidades iguais. Se removemos as  $M$  palavras no código para o qual  $P_{em}$  é maior, e desde que seja impossível para a metade das palavras no código terem uma probabilidade de erro tão grande quanto o dobro da média, as palavras código restantes devem satisfazer

$$P_{em} \leq 2 \rho_0 e^{-NE_{\alpha,\beta}(R',\underline{d})}, \quad (2.3.24)$$

onde  $R'$ , a taxa agora, é dado por

$$R' = \frac{\ln 2M}{N} = \frac{\ln M}{N} + \frac{\ln 2}{N} = M + \frac{\ln 2}{N}$$

Agora, desde que  $0 \leq \rho \leq 1$ , (2.3.22) dá

$$E_{\alpha,\beta}(R',\underline{d}) \geq E_{\alpha,\beta}(R,\underline{d}) - \frac{\ln 2}{N} \quad (2.3.25)$$

O resultado em (2.3.23) segue de (2.3.24) por uso de (2.3.25).

Chamaremos a função  $E_{\alpha,\beta}(R,\underline{d})$  como função confiança sobre distorção. Também será visto que  $(E_{\alpha,\beta}(R,\underline{d}), R, \underline{d})$  determina uma superfície.

No teorema seguinte, estudaremos as propriedades de  $E_{\alpha, \beta}(R, \underline{d})$ , e a função confiança sobre distorção.

#### 2.4 - Propriedades da Função Paramétrica de Confiança sobre Distorção

As propriedades de  $E_{\alpha, \beta}(R, \underline{d})$  dependem do comportamento da função  $E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})$ . No seguinte teorema analisaremos algumas das propriedades de  $E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})$  que são similares às da função de Gallager  $E_0(\rho, \underline{p})$  [10].

##### Teorema 2.4.1:

Considere um canal discreto sem memória com matriz de transição  $P(j/k)$ ,  $1 \leq j \leq J$ ,  $1 \leq k \leq K$ . Seja  $\underline{p} = (p_1, \dots, p_K)$  um vetor probabilidade sobre a entrada do canal. Supondo que existe pelo menos um  $j$  para o qual  $\frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}}$  muda com  $k$  para  $p_k \neq 0$ , temos

$$(a) \quad E_{\alpha, \beta}(\rho, \underline{p}, \underline{d}) \Big|_{\rho=0} = - \ln \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha} \right) \quad (2.4.1)$$

o que se reduz a zero se  $\beta = \alpha$ .

$$(b) \quad \rho > 0, \quad E_{\alpha, \beta}(\rho, \underline{p}, \underline{d}) > - \ln \sum_{j=1}^J \left( \sum_{k=1}^K p_k \left( \frac{P(y_j, x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha} \right) \quad (2.4.2)$$

para  $\rho > 0$

$$\begin{aligned}
 (c) \quad \frac{\partial E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})}{\partial \rho} \Big|_{\rho=0} &= \frac{\beta}{\alpha} \sum_{j=1}^J \sum_{k=1}^K \frac{p_k \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha}}{\sum_{j=1}^J \sum_{k=1}^K p_k \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha}} \\
 &\quad \times \ln \frac{\left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha}}{\sum_k p_k \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha}} \quad (2.4.3)
 \end{aligned}$$

o que se reduz à informação mútua se  $\beta = \alpha$  e  $d(x_k, y_j)$  é constante  $\forall_{k,j}$ .

$$(d) \quad \frac{\partial E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})}{\partial \rho} > - \ln \sum_{j=1}^J \sum_{k=1}^K p_k \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha}, \quad \rho > 0 \quad (2.4.4)$$

$$(e) \quad \frac{\partial^2 E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})}{\partial \rho^2} \leq 0 \quad \text{para} \quad \rho \geq 0 \quad (2.4.5)$$

com igualdade em (2.4.5) se, e somente, se as seguintes condições são satisfeitas:

$$(1) \quad \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \text{ é independente de } k, \text{ para } j, k \text{ tal que}$$

$$p(x_k)P(y_j/x_k) \neq 0;$$

$$(2) \quad \sum_{k: P(y_j/x_k) \neq 0} p(x_k) \text{ é independente de } j.$$

(f)  $E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})$  é uma função decrescente de  $\alpha$  quando  $\beta$  é mantido fixo, e atinge seu valor máximo  $E_0(\rho, \underline{p}, \underline{d})$  [10] quando  $\beta = \alpha$ .

Demonstração:

Para provar o teorema usamos o seguinte resultado (Galager<sup>[10]</sup>):

Seja  $a_1, \dots, a_L$  um conjunto de números não negativos e  $q_1, \dots, q_L$  um conjunto de probabilidades. Então

$$f(x) = \ln \left( \sum_{\ell} q_{\ell} a_{\ell}^{1/x} \right)^x \quad (2.4.6)$$

é não crescente com  $x > 0$  e é estritamente decrescente a não ser que os  $a_{\ell}$ , para os quais os  $q_{\ell} = 0$ , são todos iguais.  $f(x)$  é também convexa para baixo e é estritamente convexa para baixo a não ser que todos os  $a_{\ell}$  não zero, para os quais  $q_{\ell} = 0$ , são iguais.

Deste resultado segue que

$$\left( \sum_k p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{(\beta/\alpha)(1 + (\beta/\alpha)\rho)^{-1}} \right)^{1 + (\beta/\alpha)\rho}$$

é não crescente com respeito a  $\rho$ .

Além disso, desde que para pelo menos um  $j$ , para o qual

$\frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}}$  muda com  $k$  para  $p(x_k) \neq 0$ ; para esse  $j$ ,

$$\left( \sum_k p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{(\beta/\alpha)(1 + (\beta/\alpha)\rho)^{-1}} \right)^{1 + (\beta/\alpha)\rho}$$

é estritamente decrescente e por conseguinte  $E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})$  é estritamente crescente com  $\rho$ . Também do cálculo direto, obtemos

$$E_{\alpha, \beta}(\rho, \underline{p}, \underline{d}) \Big|_{\rho=0} = -\ln \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha} \right)$$

e conseqüentemente segue que para  $\rho > 0$

$$E_{\alpha, \beta}(\rho, \underline{p}, \underline{d}) > - \ln \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha} \right)$$

e

$$\frac{\partial E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})}{\partial \rho} > - \ln \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha} \right)$$

Logo, por diferenciação direta, obtemos

$$\begin{aligned} \frac{\partial E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})}{\partial \rho} \Big|_{\rho=0} &= \frac{\beta}{\alpha} \sum_{j=1}^J \sum_{k=1}^K \frac{p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha}}{\sum_{j=1}^J \sum_{k=1}^K p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha}} \\ &\quad \times \ln \frac{\left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha}}{\sum_{k=1}^K p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha}} \end{aligned}$$

Agora sejam  $\rho_1$  e  $\rho_2$  dois números positivos não iguais e  $\rho_3 = \lambda \rho_1 + (1 - \lambda) \rho_2$  onde  $0 < \lambda < 1$ . Então

$$1 + \left( \frac{\beta}{\alpha} \right) \rho_3 = \lambda \left( 1 + \left( \frac{\beta}{\alpha} \right) \rho_1 \right) + (1 - \lambda) \left( 1 + \left( \frac{\beta}{\alpha} \right) \rho_2 \right)$$

Usando o resultado que segue de (2.4.6), com substituições adequadas, obtemos

$$\left( \sum_k p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha} \right)^{-1} \left( 1 + \left( \frac{\beta}{\alpha} \right) \rho_3 \right)$$

$$\leq \left( \sum_k p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{(\beta/\alpha)(1 + (\beta/\alpha)\rho_1)^{-1}} \right)^\lambda (1 + (\beta/\alpha)\rho_1) \\ \times \left( \sum_k p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{(\beta/\alpha)(1 + (\beta/\alpha)\rho_2)^{-1}} \right)^{(1-\lambda)} (1 + (\beta/\alpha)\rho_2) \quad (2.4.7)$$

Agora, aplicando a inequação de Holder (ver Hardy<sup>[12]</sup>), a qual determina que se  $\{a_j\}$  e  $\{b_j\}$  são conjuntos de números não negativos, então

$$\sum_j a_j b_j \leq \left( \sum_j a_j^{\frac{1}{\lambda}} \right)^\lambda \left( \sum_j b_j^{\frac{1}{1-\lambda}} \right)^{(1-\lambda)}, \quad 0 < \lambda < 1,$$

Com igualdade somente se  $a_j$  e  $b_j$  são proporcionais, obtemos

$$\sum_j \left( \sum_k p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{(\beta/\alpha)(1 + (\beta/\alpha)\rho_3)^{-1}} \right)^{1 + (\beta/\alpha)\rho_3} \\ \leq \left[ \sum_j \left( \sum_k p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{(\beta/\alpha)(1 + (\beta/\alpha)\rho_1)^{-1}} \right)^{1 + (\beta/\alpha)\rho_2} \right]^\lambda \\ \times \left[ \sum_j \left( \sum_k p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{(\beta/\alpha)(1 + (\beta/\alpha)\rho_2)^{-1}} \right)^{1 + (\beta/\alpha)\rho_2} \right]^{1-\lambda} \quad (2.4.8)$$

Tomando sobre o logarítmo de (2.4.8) concluimos que  $E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})$  é convexa e assim  $\frac{\partial^2 E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})}{\partial \rho^2} \leq 0$ . A convexidade é estrita a não ser que ambas (2.4.7) e (2.4.8) sejam satisfeitas com igualdade. Será visto que existirá igualdade em (2.4.7) por causa da condição (1) do teorema e também as condições (1) e (2) do teorema garantem a igualdade em (2.4.8) (inequação de Holder).

Finalmente, para (f), se  $\alpha_2 > \alpha_1$ , e  $\beta$  é mantido fixo, temos

$$\left( \sum_k p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{1/N}} \right)^{\beta/(\alpha_1 + \beta\rho)} \right) < \left( \sum_k p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{1/N}} \right)^{\beta/(\alpha_2 + \beta\rho)} \right)$$

porque  $\frac{\beta}{\alpha_2 + \beta\rho} < \frac{\beta}{\alpha_1 + \beta\rho}$  (2.4.9)

ou

$$\left( \sum_k p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{1/N}} \right)^{\beta/(\alpha_1 + \beta\rho)} \right)^{1 + (\beta/\alpha_1)\rho} < \left( \sum_k p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{1/N}} \right)^{\beta/(\alpha_2 + \beta\rho)} \right)^{1 + (\beta/\alpha_1)\rho}$$

porque  $1 + \left(\frac{\beta}{\alpha_2}\right)\rho < 1 + \left(\frac{\beta}{\alpha_1}\right)\rho$  (2.4.10)

Somando sobre  $j$  e tomando o logarítmo, obtemos

$$E_{\alpha_1, \beta}(\rho, \underline{p}, \underline{d}) > E_{\alpha_2, \beta}(\rho, \underline{p}, \underline{d}) \text{ para } \alpha_2 > \alpha_1.$$

Que  $E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})$  toma seu valor máximo quando  $\alpha = \beta$  é agora óbvio. É um fato simples que quando  $\alpha = \beta$ ,  $E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})$  coincide com a função [10], isto é

$$E_{\alpha, \alpha}(\rho, \underline{p}, \underline{d}) = E_0(\rho, \underline{p}, \underline{d})$$

Isto completa a prova do teorema.

Definamos

$$E_{\alpha, \beta}(R, \underline{p}, \underline{d}) = \max_{0 \leq \rho \leq 1} [-\rho R + E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})] \quad (2.4.11)$$

Tomando a derivada parcial em relação a  $\rho$  da parte entre colchetes de (2.4.11) e igualando a zero, obtemos



$$R = \frac{\partial E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})}{\partial \rho} \quad (2.4.12)$$

Mostramos no último teorema que a derivada segunda de  $E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})$  é negativa. Entretanto, se algum  $\rho$  no intervalo  $0 \leq \rho \leq 1$  satisfaz (2.4.12), então este  $\rho$  precisa maximizar (2.4.11).

Ademais, de (2.4.7),  $\frac{\partial E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})}{\partial \rho}$  é não crescente com  $\rho$ , de maneira que uma solução para (2.4.12) existe, se  $R$  encontra-se no intervalo.

$$\frac{\partial E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})}{\partial \rho} \Big|_{\rho=1} \leq R \leq \frac{\beta}{\alpha} \frac{\sum_{j=1}^J \sum_{k=1}^K p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha}}{\sum_{j=1}^J \sum_{k=1}^K p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha}}$$

$$x \ln \frac{\left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha}}{\sum_{j=1}^J \sum_{k=1}^K p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{\beta/\alpha}} \quad (2.4.13)$$

Neste intervalo, podemos relatar  $E_{\alpha, \beta}(R, \underline{p}, \underline{d})$  e  $R$  parametricamente como função de  $\rho$ , dando assim

$$E_{\alpha, \beta}(R, \underline{p}, \underline{d}) = E_{\alpha, \beta}(\rho, \underline{p}, \underline{d}) - \rho \frac{\partial E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})}{\partial \rho} \quad (2.4.14)$$

$$R = \frac{\partial E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})}{\partial \rho}, \quad 0 \leq \rho \leq 1 \quad (2.4.15)$$

Para  $R < \frac{\partial E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})}{\partial \rho} \Big|_{\rho=1}$ , as equações paramétricas (2.4.14) e (2.4.15) não são válidas. Neste caso, a função

-  $\rho R + E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})$  cresce com  $\rho$  no intervalo  $0 \leq \rho \leq 1$ , e portanto o máximo ocorre para  $\rho = 1$ . Desta forma para

$$R < \left. \frac{\partial E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})}{\partial \rho} \right|_{\rho=1} \quad (2.4.16)$$

$$E_{\alpha, \beta}(R, \underline{p}, \underline{d}) = E_{\alpha, \beta}(1, \underline{p}, \underline{d}) - R \quad (2.4.17)$$

Agora voltamos nossa atenção para a presente maximização de  $E_{\alpha, \beta}(R, \underline{p}, \underline{d})$  sobre  $\underline{p}$ . Podemos reescrever (2.3.22) como

$$E_{\alpha, \beta}(R, \underline{d}) = \max_{0 \leq \rho \leq 1} [-\rho R + \max_p E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})] \quad (2.4.18)$$

Defina

$$F_{\alpha, \beta}(\rho, \underline{p}, \underline{d}) = \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right)^{(\beta/\alpha)(1 + (\beta/\alpha)\rho)^{-1}} 1 + (\beta/\alpha)\rho \quad (2.4.19)$$

De (2.3.5),  $E_{\alpha, \beta}(\rho, \underline{p}, \underline{d}) = -\ln F_{\alpha, \beta}(\rho, \underline{p}, \underline{d})$ , desta forma a minimização de  $F_{\alpha, \beta}(\rho, \underline{p}, \underline{d})$  sobre  $\underline{p}$  maximizará  $E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})$ .

#### Teorema 2.4.2:

Para algum  $\rho \geq 0$ ,  $F_{\alpha, \beta}(\rho, \underline{p}, \underline{d})$  é uma função convexa U de  $\underline{p}$  sobre a região onde  $\underline{p}$  é um vetor probabilidade. As condições necessárias e suficientes para que o vetor probabilidade  $\underline{p}$  maximize  $F_{\alpha, \beta}(\rho, \underline{p}, \underline{d})$  são:

$$\sum_{j=1}^J \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} (\beta/\alpha)(1 + (\beta/\alpha)\rho)^{-1} \gamma_j(\underline{d}) > \sum_{j=1}^J \frac{1 + (\beta/\alpha)\rho}{\gamma_j(\underline{d})} \quad (2.4.20)$$

para todo  $k$ .

com igualdade se  $p(x_k) \neq 0$ , onde

$$\gamma_j(d) = \left( \sum_{k=1}^K p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right) (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1} \right)^{1 + (\beta/\alpha)\rho} \quad (2.4.21)$$

Demonstração:

A função  $F_{\alpha, \beta}(\rho, \underline{p}, \underline{d})$  é dada por (2.4.19), a qual em (2.4.21) pode ser escrita como

$$F_{\alpha, \beta}(\rho, \underline{p}, \underline{d}) = \sum_{j=1}^J \gamma_j(d)^{1 + (\beta/\alpha)\rho} \quad (2.4.22)$$

Agora considere dois vetores probabilidades arbitrários

$\underline{p} = (p(x_1), \dots, p(x_K))$  e  $\underline{q} = (q(x_1), \dots, q(x_K))$  e

$$\gamma_j(d) = \sum_{k=1}^K p(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right) (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1} \quad (2.4.23)$$

e

$$\delta_j(d) = \sum_{k=1}^K q(x_k) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right) (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1} \quad (2.4.24)$$

Para algum  $\lambda$ ,  $0 < \lambda < 1$ , (ver Gallager<sup>[10]</sup>), temos

$$\begin{aligned} F_{\alpha, \beta}(\rho, \lambda \underline{p} + (1 - \lambda) \underline{q}, \underline{d}) &= \\ &= \sum_{j=1}^J \left[ \sum_{k=1}^K (\lambda p(x_k) + (1 - \lambda) q(x_k)) \left( \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right) (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1} \right]^{1 + (\beta/\alpha)\rho} \\ &= \sum_{j=1}^J [\lambda \gamma_j(d) + (1 - \lambda) \delta_j(d)]^{1 + (\beta/\alpha)\rho} \end{aligned} \quad (2.4.25)$$

Desde que  $\gamma_j(d)$  e  $\delta_j(d)$  devem ser não negativos, e desde que  $E^{1 + (\beta/\alpha)\rho}$  é uma função de  $x$  convexa para baixo para  $\rho \geq 0$ ,  $x \geq 0$ , podemos limitar superiormente o lado direito de (2.4.25) por

$$F_{\alpha,\beta}(\rho, \lambda \underline{p} + (1 - \lambda)\underline{q}, \underline{d}) \leq \sum_{j=1}^J [\lambda \gamma_j(\underline{d})^{1 + (\beta/\alpha)\rho} + (1 - \lambda)\delta_j(\underline{d})^{1 + (\beta/\alpha)\rho}] \\ \leq \lambda F_{\alpha,\beta}(\rho, \underline{p}, \underline{d}) + (1 - \lambda) F(\rho, \underline{q}, \underline{d}) \quad (2.4.26)$$

Deste modo  $F_{\alpha,\beta}(\rho, \underline{p}, \underline{d})$  é convexa para baixo com  $\underline{p}$  para  $\rho \geq 0$ .

Desde que  $E_{\alpha,\beta}(\rho, \underline{p}, \underline{d}) = -\ln F(\rho, \underline{p}, \underline{d})$ , segue que  $E_{\alpha,\beta}(\rho, \underline{p}, \underline{d})$  é convexa para cima com respeito a  $\underline{p}$ .

#### Exemplo:

Considere um canal simétrico quadrado a  $x$  a com matriz transição dada por

$$\begin{array}{c} x_1 \\ x_2 \\ x_a \end{array} \begin{array}{c} y_1 \\ y_2 \\ y_a \end{array} \left[ \begin{array}{ccc} p_1 & p_2 & \cdots & p_a \\ p_a & p_1 & \cdots & p_{a-1} \\ p_2 & p_3 & \cdots & p_1 \end{array} \right]$$

na qual os símbolos reproduzidos e os símbolos produzidos possuem uma distorção cíclica simétrica dada por

$$\begin{array}{c} y_1 \qquad y_2 \qquad y_a \\ \begin{array}{l} x_1 \\ x_2 \\ x_a \end{array} \left[ \begin{array}{ccc} d_1 & d_2 & \dots & d_a \\ d_a & d_1 & \dots & d_{a-1} \\ d_2 & d_3 & \dots & d_1 \end{array} \right] \end{array}$$

Claramente, o vetor probabilidade de entrada que maximiza  $E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})$  é  $p(x_1) = \dots = p(x_a) = \frac{1}{a}$ . Para esta escolha de  $\underline{p}$ , temos

$$\begin{aligned} E_{\alpha, \beta}(\rho, \underline{p}, \underline{d}) &= -\ln \sum_{j=1}^a \left[ \sum_{k=1}^a p(x_k) \frac{P(y_j/x_k)}{d(x_k, y_j)^{\frac{1}{N}}} \right]^{(\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}} 1 + (\beta/\alpha)\rho \\ &= -\ln \sum_{j=1}^a \left[ \sum_{i=1}^a \frac{1}{a} \left( \frac{p_i}{d_i^{\frac{1}{N}}} \right) \right]^{(\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}} 1 + (\beta/\alpha)\rho \\ &= -\ln \frac{1}{a^{(\beta/\alpha)\rho}} \left[ \sum_{i=1}^a \left( \frac{p_i}{d_i^{\frac{1}{N}}} \right) \right]^{(\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}} 1 + (\beta/\alpha)\rho \\ &= (\beta/\alpha) \rho \ln a - (1 + (\beta/\alpha)\rho) \ln \sum_{i=1}^a \left( \frac{p_i}{d_i^{\frac{1}{N}}} \right) \left( \frac{\beta/\alpha}{1 + (\beta/\alpha)\rho} \right)^{-1} \end{aligned} \tag{2.4.27}$$

Agora diferenciaremos (2.4.27) e calcularemos as expressões paramétricas para expoente e taxa. Diferenciando (2.4.27) em relação a  $\rho$ , temos

$$\begin{aligned} R &= \frac{\partial E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})}{\partial \rho} \\ &= \frac{\beta}{\alpha} \ln a - \frac{\beta}{\alpha} \ln \sum_{i=1}^a \left( \frac{p_i}{d_i^{\frac{1}{N}}} \right) \left( \frac{\beta/\alpha}{1 + (\beta/\alpha)\rho} \right)^{-1} + \end{aligned}$$

$$\begin{aligned}
& + \sum_{i=1}^a \frac{\left(\frac{p_i}{d_i}\right)^{\frac{1}{N}} (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}}{\sum_{i=1}^a \left(\frac{p_i}{d_i}\right)^{\frac{1}{N}} (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}} \quad \times \\
& \times \frac{\beta}{\alpha} \ln \left(\frac{p_i}{d_i}\right)^{\frac{1}{N}} (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1} \\
& = \frac{\beta}{\alpha} \ln a + \frac{\beta}{\alpha} \sum_{i=1}^a \frac{\left(\frac{p_i}{d_i}\right)^{\frac{1}{N}} (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}}{\sum_{i=1}^a \left(\frac{p_i}{d_i}\right)^{\frac{1}{N}} (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}} \quad \times \\
& \times \ln \frac{\left(\frac{p_i}{d_i}\right)^{\frac{1}{N}} (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}}{\sum_{i=1}^a \left(\frac{p_i}{d_i}\right)^{\frac{1}{N}} (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}} \\
& = \frac{\beta}{\alpha} (\ln a - H(\delta)) \tag{2.4.28}
\end{aligned}$$

onde para  $\delta = (\delta_1, \dots, \delta_a)$ ,

$$\delta_i = \frac{\left(\frac{p_i}{d_i}\right)^{\frac{1}{N}} (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}}{\sum_{i=1}^a \left(\frac{p_i}{d_i}\right)^{\frac{1}{N}} (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}}, \quad i = 1, \dots, a \tag{2.4.29}$$

e  $H(\delta)$  é dado por  $H(\delta) = - \sum_{i=1}^a \delta_i \ln \delta_i$

Também

$$E_{\alpha, \beta}(R, \underline{p}, \underline{d}) = E_{\alpha, \beta}(\rho, \underline{p}, \underline{d}) - \rho \frac{\partial E_{\alpha, \beta}(\rho, \underline{p}, \underline{d})}{\partial \rho}$$

$$\begin{aligned}
&= (\beta/\alpha)\rho \ln a - (1 + (\beta/\alpha)\rho) \ln \sum_{i=1}^a \left(\frac{p_i}{d_i \frac{1}{N}}\right) (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1} - \rho [\beta/\alpha (\ln a - H(\delta))] \\
&= (\beta/\alpha)\rho H(\delta) - (1 + (\beta/\alpha)\rho) \cdot \ln \sum_{i=1}^a \left(\frac{p_i}{d_i \frac{1}{N}}\right) (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1} \\
&= (1 + (\beta/\alpha)\rho) H(\delta) - (1 + (\beta/\alpha)\rho) \ln \sum_{i=1}^a \left(\frac{p_i}{d_i \frac{1}{N}}\right) (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1} - H(\delta) \\
&= - (1 + (\beta/\alpha)\rho) \left[ \frac{\sum_{i=1}^a \left(\frac{p_i}{d_i \frac{1}{N}}\right) (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}}{\sum_{i=1}^a \left(\frac{p_i}{d_i \frac{1}{N}}\right) (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}} \right] \\
&\quad \times \left( \ln \frac{\left(\frac{p_i}{d_i \frac{1}{N}}\right) (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}}{\sum_{i=1}^a \left(\frac{p_i}{d_i \frac{1}{N}}\right) (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}} \right) + \\
&\quad + \ln \sum_{i=1}^a \left(\frac{p_i}{d_i \frac{1}{N}}\right) (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1} \Big] - H(\delta) \\
&= - \frac{\beta}{\alpha} \sum_{i=1}^a \delta_i \ln \frac{p_i}{d_i \frac{1}{N}} - H(\delta). \tag{2.4.30}
\end{aligned}$$

Estas equações são válidas para  $0 \leq \rho \leq 1$ , ou para

$$\frac{\beta}{\alpha} \ln a + \frac{\beta}{\alpha} \sum_{i=1}^a \frac{\left(\frac{p_i}{d_i \frac{1}{N}}\right) (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}}{\sum_{i=1}^a \left(\frac{p_i}{d_i \frac{1}{N}}\right) (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1}} \times$$

$$\begin{aligned}
 & \times \ln \frac{\left(\frac{p_i}{d_i} \frac{1}{N}\right) (\beta/\alpha) (1 + \beta/\alpha)^{-1}}{a \sum_{i=1}^a \left(\frac{p_i}{d_i} \frac{1}{N}\right) (\beta/\alpha) (1 + \beta/\alpha)^{-1}} \\
 \leq R & \leq \frac{\beta}{\alpha} \ln a + \frac{\beta}{\alpha} \sum_{i=1}^a \frac{\left(\frac{p_i}{d_i} \frac{1}{N}\right)^\beta}{a \sum_{i=1}^a \left(\frac{p_i}{d_i} \frac{1}{N}\right)^\beta} \cdot \ln \frac{\left(\frac{p_i}{d_i} \frac{1}{N}\right)^\beta}{a \sum_{i=1}^a \left(\frac{p_i}{d_i} \frac{1}{N}\right)^\beta} \quad (2.4.31)
 \end{aligned}$$

Para a taxa

$$\begin{aligned}
 R & < \frac{\beta}{\alpha} \ln a + \frac{\beta}{\alpha} \sum_{i=1}^a \frac{\left(\frac{p_i}{d_i} \frac{1}{N}\right) (\beta/\alpha) (1 + \beta/\alpha)^{-1}}{a \sum_{i=1}^a \left(\frac{p_i}{d_i} \frac{1}{N}\right) (\beta/\alpha) (1 + \beta/\alpha)^{-1}} \times \\
 & \times \ln \frac{\left(\frac{p_i}{d_i} \frac{1}{N}\right) (\beta/\alpha) (1 + \beta/\alpha)^{-1}}{a \sum_{i=1}^a \left(\frac{p_i}{d_i} \frac{1}{N}\right) (\beta/\alpha) (1 + \beta/\alpha)^{-1}} \quad (2.4.32)
 \end{aligned}$$

temos

$$\begin{aligned}
 E_{\alpha, \beta}(R, \underline{p}, \underline{d}) & = E_{\alpha, \beta}(1, \underline{p}, \underline{d}) - R \\
 & = \frac{\beta}{\alpha} \ln a - \left(1 + \frac{\beta}{\alpha}\right) \ln \sum_{i=1}^a \left(\frac{p_i}{d_i} \frac{1}{N}\right) (\beta/\alpha) (1 + \beta/\alpha)^{-1} - R \quad (2.4.33)
 \end{aligned}$$



## CAPÍTULO 3

### LIMITES SUPERIORES PARA A DISTORÇÃO DE ERRO

#### 3.1 - Introdução

Seja  $D_{em}$  a distorção de erro de decodificação quando  $\underline{x}_m$  é transmitido. Consideremos uma situação em que  $m$  não satisfaz

$$\left(\frac{P(\underline{y}/\underline{x}_m)}{d(\underline{x}_m, \underline{y})}\right)^\alpha > \left(\frac{P(\underline{y}/\underline{x}_{m'})}{d(\underline{x}_{m'}, \underline{y})}\right)^\beta \quad \forall m \neq m', 1 \leq m' \leq M, \alpha \geq \beta$$

como um erro de decodificação. Um erro de decodificação também ocorre se o número inteiro decodificado é diferente do número inteiro de entrada. Portanto, podemos expressar a distorção de erro  $D_{em}$  como

$$D_{em} = \sum_{\underline{y} \in Y_N} \frac{P(\underline{y}/\underline{x}_m)}{d(\underline{x}_m, \underline{y})} \psi_m(\underline{y}) \quad (3.1)$$

onde a função  $\psi_m(\underline{y})$  é definida por

$$\psi_m(\underline{y}) = \begin{cases} 1, & \text{se } \left(\frac{P(\underline{y}/\underline{x}_m)}{d(\underline{x}_m, \underline{y})}\right)^\alpha \leq \left(\frac{P(\underline{y}/\underline{x}_{m'})}{d(\underline{x}_{m'}, \underline{y})}\right)^\beta \quad \forall m' \neq m \\ 0, & \text{caso contrário.} \end{cases}$$

A média de  $D_{em}$  sobre todas as palavras código é então definida como "Distorção Média de Erro" ou resumidamente "Distorção de Erro". Denotaremos "Distorção de Erro" por  $D_e$ .

Nosso método de decodificação será útil somente se mos

trar resultados eficientes. Isto exige então que examinemos a natureza da distorção de erro. Precisamos obter um limite superior para ela e então garantir que a distorção de erro se aproxime de zero como palavra de comprimento  $N$  e assumamos valores grandes.

No teorema seguinte, obtemos um limite superior sobre  $D_e$  limitando superiormente de maneira conveniente a função  $\psi_m(\underline{y})$ .

Teorema 3.1.1:

Para algum  $D^* \geq 0$ , e para um bloco de comprimento  $N$ , existe um código com  $M$  palavras código onde  $M \leq \text{Exp } NR(D^*)$  para o qual a distorção média de erro é tal que

$$D_e \leq \exp - N[\rho R(D^*) - D^* + \frac{1}{N} + E_{\alpha, \beta}(\rho, p)], \quad 0 \leq \rho \leq 1 \quad (3.1.1)$$

onde

$$E_{\alpha, \beta}(\rho, p) = - \ln \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) \cdot x \right. \\ \left. \times P(y_j/x_k) \right)^{(\beta/\alpha)(1 + (\beta/\alpha)\rho)^{-1}} \cdot 1 + (\beta/\alpha)\rho \quad (3.1.2)$$

provido de uma única palavra distorção maior que  $(D)^{-1}$ .

Demonstração:

É fácil ver que

$$\psi_m(\underline{y}) \leq \left[ \frac{\sum_{m' \neq m} \left( \frac{P(\underline{y}/x_{m'})}{d(x_{m'}, \underline{y})} \right)^{\beta/(\alpha + \beta\rho)}}{\left( \frac{P(\underline{y}/x_m)}{d(x_m, \underline{y})} \right)^{\alpha/(\alpha + \beta\rho)}} \right]^{(\beta/\alpha)\rho}, \quad \rho \geq 0 \quad (3.2)$$

substituindo este valor de  $\psi_m(\underline{y})$  em (3.1), obtemos

$$D_{em} \leq \sum_{\underline{y} \in Y_N} \frac{P(\underline{y}/\underline{x}_m)}{d(\underline{x}_m, \underline{y})} \cdot \left[ \frac{\sum_{m' \neq m} \left( \frac{P(\underline{y}/\underline{x}_{m'})}{d(\underline{x}_{m'}, \underline{y})} \right)^{\beta/(\alpha + \beta\rho)}}{\left( \frac{P(\underline{y}/\underline{x}_m)}{d(\underline{x}_m, \underline{y})} \right)^{\alpha/(\alpha + \beta\rho)}} \right]^{(\beta/\alpha)\rho}$$

$$D_{em} \leq \sum_{\underline{y} \in Y_N} \left( \frac{P(\underline{y}/\underline{x}_m)}{d(\underline{x}_m, \underline{y})} \right)^{\beta/\alpha + \beta\rho} \left[ \sum_{m' \neq m} \left( \frac{P(\underline{y}/\underline{x}_{m'})}{d(\underline{x}_{m'}, \underline{y})} \right)^{\beta/(\alpha + \beta\rho)} \right]^{(\beta/\alpha)\rho} \quad (3.1.3)$$

Mas

$$\underline{d} \geq \frac{1}{\underline{D}}$$

por causa disto (3.1.3) dá

$$D_{em} \leq \underline{D} \sum_{\underline{y} \in Y_N} \left( P(\underline{y}/\underline{x}_m) \right)^{\beta/\alpha + \beta\rho} \left[ \sum_{m' \neq m} \left( P(\underline{y}/\underline{x}_{m'}) \right)^{\beta/(\alpha + \beta\rho)} \right]^{(\beta/\alpha)\rho} \quad (3.1.4)$$

A inequação (3.1.4) assegura um limite para  $D_{em}$  num código particular. Simplificaremos o limite em  $D_{em}$  pela averiguação, sobre uma escolha apropriada, do conjunto de códigos. Definamos uma medida de probabilidade no conjunto  $X_N$  das possíveis  $N$ -seqüências de entrada do canal. É evidente que pelo menos um código no conjunto terá a distorção de erro que será menor que a distorção média de erro. Usando uma barra para representar a média do conjunto de códigos temos

$$\overline{D_{em}} \leq \underline{D} \cdot \overline{\sum_{\underline{y} \in Y_N} \left( P(\underline{y}/\underline{x}_m) \right)^{\beta/\alpha + \beta\rho} \left[ \sum_{m' \neq m} \left( P(\underline{y}/\underline{x}_{m'}) \right)^{\beta/(\alpha + \beta\rho)} \right]^{(\beta/\alpha)\rho}} \quad (3.1.5)$$

$$\text{desde que } \overline{\left( P(\underline{y}/\underline{x}_m) \right)^{\beta/\alpha + \beta\rho}} = \sum_{\underline{x} \in X_N} p(\underline{x}) \left( P(\underline{y}/\underline{x}) \right)^{\beta/\alpha + \beta\rho}$$

$$\overline{D}_{em} \leq \underline{D}^{(M-1)^{\rho}} \sum_{\underline{y} \in Y_N} \left( \sum_{\underline{x} \in X_N} p(\underline{x}) (P(\underline{y}/\underline{x}))^{\beta/\alpha + \beta\rho} \right)^{1 + (\beta/\alpha)\rho} \quad (3.1.6)$$

mas desde que  $\underline{D} = ND \leq ND^*$ , temos

$$\overline{D}_{em} \leq (M-1)^{\rho} ND^* \sum_{\underline{y} \in Y_N} \left( \sum_{\underline{x} \in X_N} p(\underline{x}) (P(\underline{y}/\underline{x}))^{\beta/\alpha + \beta\rho} \right)^{1 + (\beta/\alpha)\rho} \quad (3.1.7)$$

Para simplificar (3.1.7), supomos que o canal é sem memória. Se  $\underline{x} = (x_1, \dots, x_N)$ ,  $\underline{y} = (y_1, \dots, y_N)$ , então temos

$$P(\underline{y}/\underline{x}) = \prod_{n=1}^N P(y_n/x_n) \quad (3.1.8)$$

para todo  $\underline{x} \in X_N$  e  $\underline{y} \in Y_N$  e todo  $N$ .

Consideraremos agora somente a classe de conjuntos de códigos na qual cada letra de cada palavra código é escolhida independentemente de todas as outras letras com medida de probabilidade  $p(x)$ :  $\underline{x} \in X$ ,

$$p(\underline{x}) = \prod_{n=1}^N p(x_n), \quad (3.1.9)$$

usando 3.1.8 e 3.1.9 em 3.1.7, obtemos

$$\overline{D}_{em} \leq (M-1)^{\rho} ND^* \sum_{\underline{y} \in Y_N} \left( \sum_{\underline{x} \in X_N} \prod_{n=1}^N p(x_n) (P(y_n/x_n))^{\beta/\alpha + \beta\rho} \right)^{1 + (\beta/\alpha)\rho} \quad (3.1.10)$$

$$= (M-1)^{\rho} ND^* \sum_{\underline{y} \in Y_N} \left( \prod_{n=1}^N \sum_{x_n \in X_N} p(x_n) (P(y_n/x_n))^{\beta/\alpha + \beta\rho} \right)^{1 + (\beta/\alpha)\rho} \quad (3.1.11)$$

Vemos que o resultado 3.1.11 segue de 3.1.10 porque o termo entre colchetes em 3.1.11 é um produto de somas e é igual ao termo entre colchetes em 3.1.10 pela regra aritmética usual para multiplicar somas de produtos. Tomando então o produto fora dos colchetes em 3.1.11, aplicamos a mesma regra novamente e obtemos

$$\overline{D}_{em} \leq (M-1)^\rho ND^* \prod_{n=1}^N \sum_{\underline{y} \in Y_N} \left( \sum_{\underline{x} \in X_N} p(x_n) (P(y_n/x_n))^{\beta/\alpha + \beta\rho} \right)^{1 + (\beta/\alpha)\rho} \quad (3.1.12)$$

$$0 \leq \rho \leq 1$$

Simplificamos agora a notação de 3.1.12 por observação de que  $X$  é o conjunto de letras entradas  $x_1, \dots, x_K$  e  $Y$  é o conjunto de letras saídas  $y_1, \dots, y_J$ . Notando aqui que todos os termos do produto são idênticos, obtemos

$$\overline{D}_{em} \leq (M-1)^\rho ND^* \left[ \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) (P(y_j/x_k))^{\beta/\alpha + \beta\rho} \right)^{1 + (\beta/\alpha)\rho} \right]^N \quad (3.1.13)$$

$$0 \leq \rho \leq 1$$

Além disso, desde que temos  $M - 1 < M \leq \exp NR(D^*)$ , obtemos

$$\overline{D}_{em} \leq ND^* \exp [\rho NR(D^*)] \left[ \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) (P(y_j/x_k))^{\beta/\alpha + \beta\rho} \right)^{1 + (\beta/\alpha)\rho} \right]^N$$

$$= \exp [\rho NR(D^*) + \ln ND^* + N \ln \left[ \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) (P(y_j/x_k))^{\beta/\alpha + \beta\rho} \right)^{1 + (\beta/\alpha)\rho} \right]^N$$

$$= \exp [\rho NR(D^*) + \ln D^* - NE_{\alpha, \beta}(\rho, p)], \text{ usando (3.1.2)}. \quad (3.1.14)$$

Usando a inequação  $\ln x \leq x - 1$ , obtemos

$$\begin{aligned} \overline{D}_{em} &\leq \exp [\rho NR(D^*) + ND^* - 1 - NE_{\alpha, \beta}(\rho, \underline{p})] \\ &= \exp - N[-\rho R(D^*) - D^* + \frac{1}{N} + E_{\alpha, \beta}(\rho, \underline{p})] \end{aligned} \quad (3.1.15)$$

Desde que o lado direito de 3.1.15 é independente de  $m$  ele é um limite para o conjunto da distorção de erro e é independente das probabilidades com as quais as palavras código são usadas. Desde que pelo menos um código no conjunto terá uma distorção de erro menor que a média, temos provado o teorema.

O teorema 3.1.1 é válido para todos  $\rho$ ,  $0 \leq \rho \leq 1$ , e todos os vetores de probabilidade  $\underline{p} = (p(x_1), \dots, p(x_K))$ . Entretanto, podemos obter um limite mais refinado para  $D_e$  por minimização da função no colchete do termo do lado direito de 3.1.15 sobre  $\rho$  e  $\underline{p}$ . Isto nos dá o seguinte Corolário simples.

Corolário 3.1.1:

Sob as condições do teorema 3.1.1 existe um código para o qual

$$\overline{D}_{em} \leq \exp - NE_{\alpha, \beta}(R(D^*)) , \quad (3.1.16)$$

onde

$$E_{\alpha, \beta}(R(D^*)) = \max_{0 \leq \rho \leq 1, \underline{p}} [E_{\alpha, \beta}(R(D^*), \underline{p}, \rho)] \quad (3.1.17)$$

e

$$E_{\alpha, \beta}(R(D^*), \underline{p}, \rho) = E_{\alpha, \beta}(\rho, \underline{p}) - \rho R(D^*) - D^* + \frac{1}{N} \quad (3.1.18)$$

A maximização em 3.1.17 é tomada sobre  $\rho$ ,  $0 \leq \rho \leq 1$  e sobre to-

dos os vetores probabilidade  $\underline{P}$ .

A função  $E_{\alpha, \beta}(R(D^*))$  será chamada de taxa de confiança da função distorção.

Justamente como na dedução de Gallager<sup>[10]</sup> para um limite para a probabilidade de erro podemos obter um limite para distorção própria para erro que se aplica para cada palavra código separadamente antes que sobre a média. Isto é feito no seguinte Corolário.

Corolário 3.1.2:

Sob as condições do teorema 3.1.1 existe um código tal que, para todo  $m$ ,  $1 \leq m \leq M$ , a distorção própria para erro quando a  $m$ -ésima palavra código é transmitida, é limitada por

$$D_{em} \leq 4 e^{-NE_{\alpha, \beta}(R(D^*))} \quad (3.1.19)$$

onde  $E_{\alpha, \beta}(R(D^*))$  é dado por 3.1.17.

Demonstração:

Escolhamos um código com  $M' = 2M$  palavras código que satisfaça o Corolário 3.1.1 enquanto a fonte usa as  $2M$  palavras código com probabilidades iguais. Removemos as  $M$  palavras no código para as quais  $D_{em}$  é grande. É impossível sobre a metade das palavras no código termos uma distorção própria para erro maior que o dobro da média. Portanto, as palavras código restantes satisfarão

$$D_{em} \leq 2 e^{-N_{\alpha, \beta}(R'(D^*))} \quad (3.1.20)$$

Desde que  $R'(D^*) = \frac{\ln 2M}{N} = \frac{\ln 2}{N} + R(D^*)$ , porque  $M = (\exp N R(D^*))$ . Agora, desde que  $0 \leq \rho \leq 1$ , 3.1.17 nos dá

$$E_{\alpha, \beta}(R'(D^*)) \geq E_{\alpha, \beta}(R(D^*)) - \frac{\ln 2}{N}, \quad (3.1.21)$$

Usando (3.1.20) e (3.1.21), obtemos (3.1.19). Isto prova o Corolário.

Na secção seguinte estudaremos algumas das propriedades da função confiança de  $E_{\alpha, \beta}(R(D^*))$ .

### 3.2 - Propriedades da Função Confiança de Taxa de Distorção $E(R(D^*))$

A maximização de (3.1.18) sobre  $\rho$  e  $\underline{p}$  depende do comportamento da função  $E_{\alpha, \beta}(\rho, \underline{p})$ .

#### Teorema 3.2.1:

Considere um canal discreto sem memória com matriz de transição  $\{P(j/k)\}$ ,  $1 \leq j \leq J$ ,  $1 \leq k \leq K$ . Seja  $\underline{p} = (p_1, \dots, p_K)$  um vetor probabilidade sobre a entrada do canal. Supondo que  $P(j/k)$  muda com  $k$  para  $p(x_k) \neq 0$ , temos

$$(a) \quad E_{\alpha, \beta}(\rho, \underline{p}) \Big|_{\rho=0} = - \ln \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) P(j/k)^{\beta/\alpha} \right), \quad (3.2.1)$$

o que se reduz a zero se  $\beta = \alpha$ .



$$(b) \quad E_{\alpha, \beta}(\rho, \underline{p}) > - \ln \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) P(j/k)^{\beta/\alpha} \right) \quad (3.2.2)$$

para  $\rho > 0$

$$(c) \quad \frac{\partial E_{\alpha, \beta}(\rho, \underline{p})}{\partial \rho} \Big|_{\rho=0} = \frac{\beta}{\alpha} \frac{\sum_{j=1}^J \sum_{k=1}^K p(x_k) P(j/k)^{\beta/\alpha}}{\sum_{j=1}^J \sum_{k=1}^K p(x_k) P(j/k)^{\beta/\alpha}} \quad (3.2.3)$$

$$\times \ln \frac{P(j/k)^{\beta/\alpha}}{\sum_k p(x_k) P(j/k)^{\beta/\alpha}}$$

o que se reduz à informação mútua se  $\beta = \alpha$ .

$$(d) \quad \frac{\partial E_{\alpha, \beta}(\rho, \underline{p})}{\partial \rho} > - \ln \sum_{j=1}^J \sum_{k=1}^K p(x_k) P(j/k)^{\beta/\alpha} \quad (3.2.4)$$

$$(e) \quad \frac{\partial^2 E_{\alpha, \beta}(\rho, \underline{p})}{\partial \rho^2} \leq 0 \quad \text{para} \quad \rho \geq 0 \quad (3.2.5)$$

Com igualdade em (3.2.5) se, e somente, se as seguintes condições são satisfeitas:

(1)  $p(j/k)$  é independente de  $k$ , para  $j$ ,  $k$  tal que  $p(x_k) p(j/k) \neq 0$ ;

(2)  $\sum_{k:p(j/k)} p(x_k)$  é independente de  $j$ .

(f)  $E_{\alpha, \beta}(\rho, \underline{p})$  é uma função decrescente de  $\alpha$  quando  $\beta$  é mantido fixo e atinge seu máximo  $E_0(\rho, \underline{p})$  função de Gallager) [10] quando  $\alpha = \beta$ .

Demonstração:

Para provar o teorema usamos o seguinte resultado (Gallager<sup>[10]</sup>):

Seja  $a_1, \dots, a_L$  um conjunto de número não negativos e seja  $q_1, \dots, q_L$  um conjunto de probabilidades. Então

$$f(x) = \ln \left( \sum_{\ell} q_{\ell} a_{\ell}^{1/x} \right)^x \quad (3.2.6)$$

é não crescente com  $x > 0$  e é estritamente decrescente a não ser que os  $a_{\ell}$ , para os quais os  $q_{\ell} = 0$ , são todos iguais.  $f(x)$  é também convexa para baixo e é estritamente convexa para baixo a não ser que todos os  $a_{\ell}$  não zero, para os quais  $q_{\ell} = 0$ , são iguais.

Deste resultado segue que

$$\left( \sum_k p(x_k) P(j/k)^{(\beta/\alpha)(1 + (\beta/\alpha)\rho)^{-1}} \right)^{1 + (\beta/\alpha)\rho}$$

é não crescente com respeito a  $\rho$ .

Além disso, desde que para pelo menos um  $j$ ,  $P(j/k)$  muda com  $p(x_k) \neq 0$ ; para esse  $j$

$$\left( \sum_k p(x_k) P(j/k)^{(\beta/\alpha)(1 + (\beta/\alpha)\rho)^{-1}} \right)^{1 + (\beta/\alpha)\rho}$$

é estritamente decrescente. Assim,  $E_{\alpha, \beta}(\rho, \underline{p})$  é estritamente crescente com  $\rho$ . Também do cálculo direto, obtemos

$$E_{\alpha, \beta}(\rho, \underline{p}) \Big|_{\rho=0} = - \ln \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) P(j/k)^{\beta/\alpha} \right)$$

e conseqüentemente segue que para  $\rho > 0$

$$E_{\alpha, \beta}(\rho, p) > - \ln \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) P(j/k)^{\beta/\alpha} \right)$$

e

$$\frac{\partial E_{\alpha, \beta}(\rho, p)}{\partial \rho} > - \ln \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) P(j/k)^{\beta/\alpha} \right)$$

Logo, por diferenciação direta, obtemos

$$\begin{aligned} \frac{\partial E_{\alpha, \beta}(\rho, p)}{\partial \rho} \Big|_{\rho=0} &= \frac{\beta}{\alpha} \sum_{j=1}^J \sum_{k=1}^K \frac{p(x_k) P(j/k)^{\beta/\alpha}}{\sum_{j=1}^J \sum_{k=1}^K p(x_k) P(j/k)^{\beta/\alpha}} \\ &\quad \times \ln \frac{P(j/k)^{\beta/\alpha}}{\sum_{k=1}^K p(x_k) P(j/k)^{\beta/\alpha}} \end{aligned}$$

Agora sejam  $\rho_1$  e  $\rho_2$  dois números positivos não iguais e  $\rho_3 = \lambda \rho_1 + (1 - \lambda) \rho_2$  onde  $0 < \lambda < 1$ . Então

$$1 + \left(\frac{\beta}{\alpha}\right) \rho_3 = \lambda \left(1 + \left(\frac{\beta}{\alpha}\right) \rho_1\right) + (1 - \lambda) \left(1 + \left(\frac{\beta}{\alpha}\right) \rho_2\right)$$

Usando o resultado que segue de (3.2.6), com substituições adequadas, obtemos

$$\begin{aligned} &\left( \sum_k p(x_k) P(j/k)^{(\beta/\alpha)(1 + (\beta/\alpha)\rho_3)^{-1}} \right)^{1 + (\beta/\alpha)\rho_3} \\ &\leq \left( \sum_k p(x_k) P(j/k)^{(\beta/\alpha)(1 + (\beta/\alpha)\rho_1)^{-1}} \right)^{\lambda(1 + (\beta/\alpha)\rho_1)} \times \\ &\quad \times \left( \sum_k p(x_k) P(j/k)^{(\beta/\alpha)(1 + (\beta/\alpha)\rho_2)^{-1}} \right)^{(1 - \lambda)(1 + (\beta/\alpha)\rho_2)} \end{aligned}$$

(3.2.7)

Agora, aplicando a inequação de Holder, obtemos

$$\begin{aligned} & \sum_j \left( \sum_k p(x_k) P(j/k) \right)^{(\beta/\alpha) (1 + (\beta/\alpha)\rho_3)^{-1}} \left( 1 + (\beta/\alpha)\rho_3 \right) \\ & \leq \left[ \sum_j \left( \sum_k p(x_k) P(j/k) \right)^{(\beta/\alpha) (1 + (\beta/\alpha)\rho_1)^{-1}} \left( 1 + (\beta/\alpha)\rho_1 \right) \right]^\lambda \\ & \times \left[ \sum_j \left( \sum_k p(x_k) P(j/k) \right)^{(\beta/\alpha) (1 + (\beta/\alpha)\rho_2)^{-1}} \left( 1 + (\beta/\alpha)\rho_2 \right) \right]^{1-\lambda} \end{aligned} \quad (3.2.8)$$

Tomando sobre o logaritmo de (3.2.8) concluímos que  $E_{\alpha, \beta}(\rho, \underline{p})$  é convexa  $\cap$  e assim  $\frac{\partial^2 E_{\alpha, \beta}(\rho, \underline{p})}{\partial \rho^2} \leq 0$ . A convexidade é estrita a não ser que ambas (3.2.7) e (3.2.8) sejam satisfeitas com igualdade. Será visto que existirá igualdade em (3.2.7) por causa da condição (1) do teorema e também as condições (1) e (2) do teorema garantem a igualdade em (3.2.8) (inequação de Holder).

Finalmente, para (f), se  $\alpha_2 > \alpha_1$ , e  $\beta$  é mantido fixo, temos

$$\left( \sum_k p(x_k) P(j/k) \right)^{\beta/(\alpha_1 + \beta\rho)} < \left( \sum_k p(x_k) P(j/k) \right)^{\beta/(\alpha_2 + \beta\rho)}$$

$$\text{porque } \frac{\beta}{\alpha_2 + \beta\rho} < \frac{\beta}{\alpha_1 + \beta\rho} \quad (3.2.9)$$

$$\begin{aligned} \text{ou } \left( \sum_k p(x_k) P(j/k) \right)^{\beta/(\alpha_1 + \beta\rho)} \left( 1 + (\beta/\alpha_1)\rho \right) & < \left( \sum_k p(x_k) P(j/k) \right)^{\beta/(\alpha_2 + \beta\rho)} \left( 1 + (\beta/\alpha_1)\rho \right) \\ & < \left( \sum_k p(x_k) P(j/k) \right)^{\beta/(\alpha_3 + \beta\rho)} \left( 1 + (\beta/\alpha_2)\rho \right) \end{aligned}$$

$$\text{porque } 1 + \left(\frac{\beta}{\alpha_2}\right)\rho < 1 + \left(\frac{\beta}{\alpha_1}\right)\rho \quad (3.2.10)$$

Somando sobre  $j$  e tomando o logaritmo, obtemos

$$E_{\alpha_1, \beta}(\rho, \underline{p}) > E_{\alpha_2, \beta}(\rho, \underline{p}) \quad \text{para} \quad \alpha_2 > \alpha_1$$

Que  $E_{\alpha, \beta}(\rho, \underline{p})$  toma seu valor máximo quando  $\alpha = \beta$ , é agora óbvio. É um fato simples que quando  $\alpha = \beta$ ,  $E_{\alpha, \beta}(\rho, \underline{p})$  coincide com a função [10], isto é

$$E_{\alpha, \alpha}(\rho, \underline{p}) = E_0(\rho, \underline{p})$$

Definamos

$$E_{\alpha, \beta}(R(D^*), \underline{p}) = \max_{0 \leq \rho \leq 1} [-\rho R(D^*) - D^* + \frac{1}{N} + E_{\alpha, \beta}(\rho, \underline{p})] \quad (3.2.11)$$

Tomando a derivada parcial em relação a  $\rho$  da parte entre colchetes de (3.2.11) e igualando a zero, obtemos

$$R(D^*) = \frac{\partial E_{\alpha, \beta}(\rho, \underline{p})}{\partial \rho} \quad (2.3.12)$$

Mostramos no último teorema que a derivada segunda de  $E_{\alpha, \beta}(\rho, \underline{p})$  é negativa. Entretanto, se algum  $\rho$  no intervalo  $0 \leq \rho \leq 1$  satisfaz (3.2.12), então este  $\rho$  precisa maximizar (3.2.11).

Ademais, de (3.2.7),  $\frac{\partial E_{\alpha, \beta}(\rho, \underline{p})}{\partial \rho}$  é não crescente com  $\rho$ , de maneira que uma solução para (3.2.12) existe, se  $R$  encontra-se no intervalo

$$\frac{\partial E_{\alpha, \beta}(\rho, \underline{p})}{\partial \rho} \Big|_{\rho=1} \leq R(D^*) \leq \frac{\beta}{\alpha} \frac{\sum_{j=1}^J \sum_{k=1}^K p(x_k) P(j/k)^{\beta/\alpha}}{\sum_{j=1}^J \sum_{k=1}^K p(x_k) P(j/k)^{\beta/\alpha}}$$

$$x \ln \frac{P(j/k)^{\beta/\alpha}}{\sum_k p(x_k) P(j/k)^{\beta/\alpha}} \quad (3.2.13)$$

Neste intervalo, podemos relatar  $E_{\alpha, \beta}(R(D^*), \underline{p})$  e  $R(D^*)$  parametricamente como função de  $\rho$ , dando assim

$$E_{\alpha, \beta}(R(D^*), \underline{p}) = E_{\alpha, \beta}(\rho, \underline{p}) - \rho \frac{E_{\alpha, \beta}(\rho, \underline{p})}{\partial \rho} - D^* + \frac{1}{N} \quad (3.2.14)$$

$$R(D^*) = \frac{\partial E_{\alpha, \beta}(\rho, \underline{p})}{\partial \rho}, \quad 0 \leq \rho \leq 1 \quad (3.2.15)$$

Para  $R(D^*) < \left. \frac{\partial E_{\alpha, \beta}(\rho, \underline{p})}{\partial \rho} \right|_{\rho=1}$ , as equações paramétricas (3.2.14) e (3.2.15) não são válidas. Neste caso, a função  $E_{\alpha, \beta}(\rho, \underline{p}) - \rho R(D^*) - D^* + \frac{1}{N}$  cresce com  $\rho$  no espaço  $0 \leq \rho \leq 1$ , e portanto o máximo ocorre para  $\rho = 1$ . Desta forma para

$$R(D^*) < \left. \frac{\partial E_{\alpha, \beta}(\rho, \underline{p})}{\partial \rho} \right|_{\rho=1} \quad (3.2.16)$$

$$E_{\alpha, \beta}(R(D^*), \underline{p}) = E_{\alpha, \beta}(1, \underline{p}) - R(D^*) - D^* + \frac{1}{N} \quad (3.2.17)$$

Agora voltamos nossa atenção para a presente maximização de  $E_{\alpha, \beta}(R(D^*), \underline{p})$  sobre  $\underline{p}$ . Podemos reescrever (3.1.17) como

$$E_{\alpha, \beta}(R(D^*)) = \max_{0 \leq \rho \leq 1} \left[ -\rho R(D^*) - D^* + \frac{1}{N} + \max_{\rho} E_{\alpha, \beta}(\rho, \underline{p}) \right] \quad (3.2.18)$$

Defina

$$E_{\alpha, \beta}(\rho, \underline{p}) = \sum_{j=1}^J \left( \sum_{k=1}^K p(x_k) P(j/k)^{(\beta/\alpha)} (1 + (\beta/\alpha)\rho)^{-1} \right)^{1 + (\beta/\alpha)\rho} \quad (3.2.19)$$

De (3.1.2)  $E_{\alpha, \beta}(\rho, \underline{p}) = -\ln F_{\alpha, \beta}(\rho, \underline{p})$ . Desta forma a minimização

de  $F_{\alpha, \beta}(\rho, \underline{p})$  sobre  $\underline{p}$  maximizará  $E_{\alpha, \beta}(\rho, \underline{p})$ .

Teorema 3.2.2:

Para algum  $\rho \geq 0$ ,  $F_{\alpha, \beta}(\rho, \underline{p})$  é uma função convexa  $\cup$  de  $\underline{p}$  sobre a região onde  $\underline{p}$  é um vetor probabilidade. As condições necessárias e suficientes para que o vetor probabilidade  $\underline{p}$  maximize  $F_{\alpha, \beta}(\rho, \underline{p})$  são

$$\sum_{j=1}^J P(j/k) (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1} \gamma_j^{(\beta/\alpha)\rho} \geq \sum_{j=1}^J \gamma_j^{1 + (\beta/\alpha)\rho} \quad (3.2.20)$$

para todo  $K$

com igualdade se  $p(x_k) \neq 0$ , onde

$$\gamma_j = \left( \sum_{k=1}^K p(x_k) P(j/k) (\beta/\alpha) (1 + (\beta/\alpha)\rho)^{-1} \right) \quad (3.2.21)$$

A prova pode ser obtida por modificação adequada do método todo seguido no Capítulo 2 no teorema 2.4.2.

## REFERÊNCIAS

1. ASH, R.B. (1965). Information Theory. Interscience, New York.
2. BERGER, T. (1971). Rate Distortion Theory: a Mathematical Basis for Data Compression. Prentice Hall, New Jersey.
3. BLAHUT, R.E. (1972). An Hypothesis Testing Approach To Information Theory, Ph.D. Dissertation, Cornell University, Ithaca, New York.
4. CHASE, D. (1970). Coding Theorems for the Nonsynchronized Channels, IEEE Trans. Inf. Theory, 16, 55-75.
5. EBERT, P.M. (1968). An Extension of Rate Distortion Theory To Confusion Matrices, IEEE Trans. Inform. Theory, 14, 6-11.
6. ELIAS, P. (1957). List Decoding for Noisy Channels, MIT Research Lab. of Electronics, Tech. Rept. 335.
7. FANO, R.M. (1961). Transmission of Information, The MIT Press, Cambridge, Mass. USA.
8. FEINSTEIN, A. (1958). Foundations of Information Theory, McGraw Hill, New York.
9. FORNEY, G.D. (1968). Exponential Error Bounds for Erasure, list and decision feedback schemes, IEEE Trans. Inform. Theory, 4, 206-220.
10. GALLAGER, R.G. (1968). Information Theory and Reliable Communication, New York; Wiley.
11. GRAY, R.M. (1969). Information Rates of Autoregressive Sources, Ph.D. dissertation, University of Southern California, Los Angeles, USA.
12. HARDY, G.H.; LITTLEWOOD, J.E. and POLYA, G. (1934). Inequalities, Cambridge Univ. Press, London.
13. JELINEK, F. (1968). Probabilistic Information Theory, McGraw Hill, New York.
14. KRICH, S.I. (1972). Coding for a time varying fidelity criterion, Ph.D. dissertation, University of Southern Califór



nia, Los Angeles, USA.

15. LEINER, B.M. and GRAY, R.M. (1973). Bounds on Rate Distortion Functions for Stationary Sources and Context Dependent Fidelity Criterion, IEEE Trans. Inf. Theory, 19, 706-709.
16. OMURA, J.K. (1973). A Coding Theorem for Discrete Time Sources, IEEE Trans. Inf. Theory, 19, 490-498.
17. PURSLEY, M.B. (1974). Coding Theorems for Non-ergodic Sources and Sources with Unknown Parameters, USCEF Rept 466, Electronics Science Lab., Univ. of Southern California, Los Angeles, USA.
18. REIFFEN, B. (1966). A Per Letter Converse to the Channel Coding Theorem, IEEE Trans. Infor. Theory, 12, 475-480.
19. SAKRISON, D.J. (1969). The Rate Distortion Function for a Class of Sources, Information and Control 15, 165-195.
20. SHANNON, C.E. (1948). A Mathematical Theory of Communication, BSTJ, 27, 379-423, 623-656.
21. SHANNON, C.E. (1959). A Coding Theorem for a Discrete Source with a Fidelity Criterion, Information and Decision Processes, R.E. Machel, ed. McGraw Hill, New York.
22. SHANNON, C.E.; GALLAGER, R.G. and BERLEKAMP, E.R. (1967). Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels part I and II, Inform. and Contr., 13, 65-103 (part I) 522-552 (part II).
23. SHARMA, B.D. and GUR DIAL (1974). Bounds on Distortion Due to Error for Rates Below and above the Channel Capacity, Inform. and Control, 26, 272-279.
24. SHARMA, B.D. and RAINA, N. (1980). Coding Theorem for Partial Received Information, Inform. Sciences, 20, 181-189.
25. SHARMA, B.D. and RAINA, N. (1978). Coding Theorem for a Generalized Maximum Likelihood Decoding Scheme 9, 1091-1102, Indian Journal of Pure and applied Meths.
26. STIGLITZ, I.G. (1966). Coding for a Class of Unknown Channels, IEEE Trans. Inform. Theory, 12, 189-195.

27. WOLFOWITZ, J. (1964). Coding Theorems of Information Theory, Springer-Verlag and Prentice Hall, Englewood.
28. WYNER, A.D. (1970). Another Look at the Coding Theorem of Information Theory, Proceedings of the IEEE 58, 894-913.
29. YUDKIN, H. (1967). On the Exponential Error Bound and Capacity for Finite State Channels, International Symposium of Information Theory, San Remo, Italy.
30. ZAKAI, M. and ZIV, J. (1975). A Generalization of the Rate Distortion Theory and Applications, CISM, Udine, Italy.