

Código de Shannon-Fano-Elias

Ideia: Definir como palavras de código a representação binária da distribuição de probabilidade acumulada $F(x)$ truncada a $l(x)$ bits.

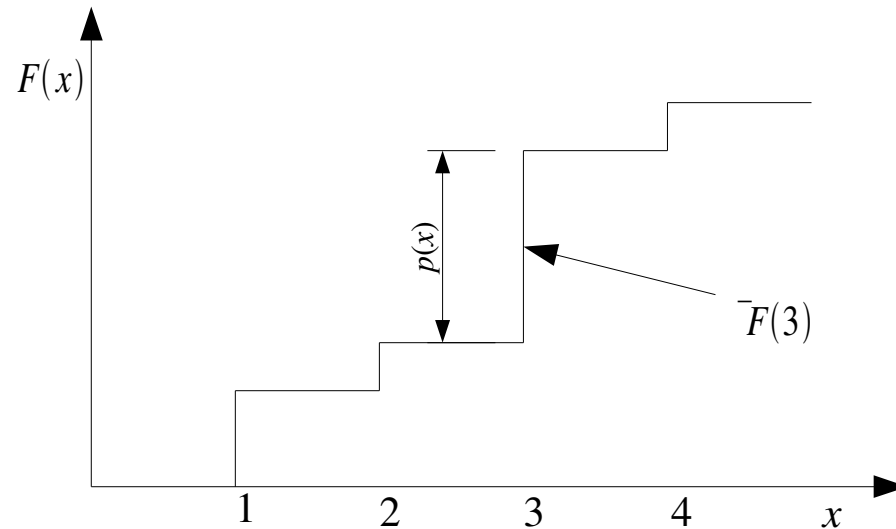
Vantagens:

- O cálculo das palavras de código é muito rápida a partir das probabilidades $p(x)$.
- Não é necessário manipular árvores.
- Não é necessário nenhuma ordenação especial das probabilidades

Desvantagens:

- Não é ótimo.
- É necessário conhecer as probabilidades à partida.

Define-se a distribuição acumulada modificada $\bar{F}(x) = \sum_{a < x} p(a) + \frac{1}{2} p(x)$



Usam-se como palavras de código os primeiros $l(x)$ bits de $\bar{F}(x)$ em que $l(x) = \lceil -\log p(x) \rceil + 1$

Note-se que o comprimento do código tem mais um bit, que tem como consequência que $L(C) < H(X) + 2$.

Exemplo:

x	$p(x)$	$F(x)$	$F'(x)$	$\overset{bin}{F'(x)}$	$l(x)$	$C(x)$
1	0.25	0.25	0.125	0.001	3	001
2	0.5	0.75	0.5	0.1	2	10
3	0.125	0.875	0.8125	0.1101	4	1101
4	0.125	1.0	0.9375	0.1111	4	1111

$$L(C) = 2.75 \text{ bits}$$

$$H(X) = 1.75 \text{ bits}$$

O código não é o melhor que se pode obter, mas desenhando o código para blocos de vários símbolos pode obter-se um comprimento médio arbitrariamente perto da entropia. O algoritmo de **codificação aritmética** explora esta vertente.