# Unary coding

From Wikipedia, the free encyclopedia

**Unary coding**, sometimes called **thermometer code**, is an entropy encoding that represents a natural number, *n*, with *n* ones followed by a zero (if *natural number* is understood as *non-negative integer*) or with *n* − 1 ones followed by a zero (if *natural number* is understood as *strictly positive integer*). For example 5 is represented as 111110 or 11110. Some representations use *n* or *n* − 1 zeros followed by a one. The ones and zeros are interchangeable without loss of generality. Unary coding is both a Prefix-free code and a Self-synchronizing code.

| n (non-negative) | n (strictly positive) | Unary code | Alternative |
|---|---|---|---|
| 0 | 1 | 0 | 1 |
| 1 | 2 | 10 | 01 |
| 2 | 3 | 110 | 001 |
| 3 | 4 | 1110 | 0001 |
| 4 | 5 | 11110 | 00001 |
| 5 | 6 | 111110 | 000001 |
| 6 | 7 | 1111110 | 0000001 |
| 7 | 8 | 11111110 | 00000001 |
| 8 | 9 | 111111110 | 000000001 |
| 9 | 10 | 1111111110 | 0000000001 |

Unary coding is an optimally efficient encoding for the following discrete probability distribution

$$\operatorname{P}(n) = 2^{-n}$$

for $n = 1, 2, 3, ...$.

In symbol-by-symbol coding, it is optimal for any geometric distribution

$$\operatorname{P}(n) = (k-1)k^{-n}$$

for which *k* ≥ φ = 1.61803398879…, the golden ratio, or, more generally, for any discrete distribution for which

$$\operatorname{P}(n) \geq \operatorname{P}(n+1) + \operatorname{P}(n+2)$$

for $n = 1, 2, 3, ...$. Although it is the optimal symbol-by-symbol coding for such probability distributions, Golomb coding achieves better compression capability for the geometric distribution because it does not consider input symbols independently, but rather implicitly groups the inputs. For the same reason, arithmetic encoding performs better for general probability distributions, as in the last case above.

**Contents** [hide]

1 Unary code in use today
2 Unary coding in biological networks
3 Generalized unary coding
4 See also
5 References

## Unary code in use today [edit]

Examples of unary code uses include:

- In Golomb Rice code, unary encoding is used to encode the quotient part of the Golomb code word.
- In UTF-8, unary encoding is used in the leading byte of a multi-byte sequence to indicate the number of bytes in the sequence, so that the length of the sequence can be determined without examining the continuation bytes.
- Instantaneously trained neural networks use unary coding for efficient data representation.

## Unary coding in biological networks [edit]

New research has shown that unary coding is used in the neural circuits responsible for birdsong production.[1][2] The nucleus in the brain of the songbirds that plays a part in both the learning and the production of bird song is the HVC (high vocal center). This coding works as space coding which is an efficient strategy for biological circuits due to its inherent simplicity and robustness.

## Generalized unary coding [edit]

A generalized version of unary coding is able to represent numbers much more efficiently than standard unary coding.[3] Here's an example of generalized unary coding for integers from 1 through 15 that requires only 7 bits (where three bits are arbitrarily chosen in place of a single one in standard unary to show the number). Note that the representation is cyclic where one uses markers to represent higher integers in higher cycles.

| n | Unary code | Generalized unary |
|---|---|---|
| 0 | 0 | 0000000 |
| 1 | 10 | 0000111 |
| 2 | 110 | 0001110 |
| 3 | 1110 | 0011100 |
| 4 | 11110 | 0111000 |
| 5 | 111110 | 1110000 |
| 6 | 1111110 | 0010111 |
| 7 | 11111110 | 0101110 |
| 8 | 111111110 | 1011100 |
| 9 | 1111111110 | 0111001 |
| 10 | 11111111110 | 1110010 |
| 11 | 111111111110 | 0100111 |
| 12 | 1111111111110 | 1001110 |
| 13 | 11111111111110 | 0011101 |
| 14 | 111111111111110 | 0111010 |
| 15 | 1111111111111110 | 1110100 |

Generalized unary coding requires that the range of numbers to be represented be pre-specified because this range determines the number of bits that are needed.

## See also [edit]

- Unary numeral system

## References [edit]


1. ^ Fiete, I.R. and H.S. Seung, Neural network models of birdsong production, learning, and coding. New Encyclopediaof Neuroscience. Eds. L. Squire, T. Albright, F. Bloom, F. Gage, and N. Spitzer. Elsevier, 2007.
2. ^ Moore J.M. et al., Motor pathway convergence predicts syllable repertoire size in oscine birds. Proc. Nat. Acad. Sc. USA 108: 16440-16445, 2011.
3. ^ Kak, S., Generalized unary coding. Circuits, Systems and Signal Processing. 2015. http://link.springer.com/article/10.1007/s00034-015-0120-7#page-1

| | | |
|---|---|---|
| | Codec parts | A-law · μ-law · ACELP · ADPCM · CELP · DPCM · Fourier transform · LPC (LAR · LSP) · MDCT · Psychoacoustic model · WLPC |
| **Image** | Concepts | Chroma subsampling · Coding tree unit · Color space · Compression artifact · Image resolution · Macroblock · Pixel · PSNR · Quantization · Standard test image |
| | Methods | Chain code · DCT · EZW · Fractal · KLT · LP · RLE · SPIHT · Wavelet |
| **Video** | Concepts | Bit rate (average (ABR) · constant (CBR) · variable (VBR)) · Display resolution · Frame · Frame rate · Frame types · Interlace · Video characteristics · Video quality |
| | Codec parts | Lapped transform · DCT · Deblocking filter · Motion compensation |
| **Theory** | Entropy · Kolmogorov complexity · Lossy · Quantization · Rate–distortion · Redundancy · Timeline of information theory | |
| Compression formats · Compression software (codecs) | | |

Categories: Coding theory | Data compression | Lossless compression algorithms

This page was last modified on 20 August 2015, at 16:16.