The Source Coding Theorem

Mathias Winther Madsen
mathias.winther@gmail.com

Institute for Logic, Language, and Computation
University of Amsterdam

6 November 2015

The Convergence of Averages

Problem

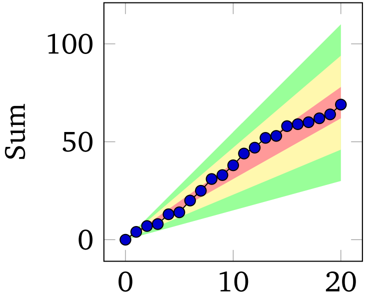
Which of the following is more probable?

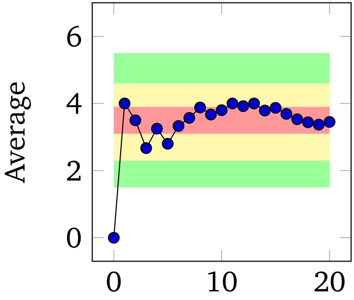
1. a sum of 4,000 in 1,000 dice rolls;
2. a sum of 4,000,000 in 1,000,000 dice rolls.

The Convergence of Averages

(Prelude to) The Weak Law of Large Numbers

$$E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = E[X], \qquad Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{Var[X]}{n}.$$





Sequence Probabilities

Problem

A source produces texts

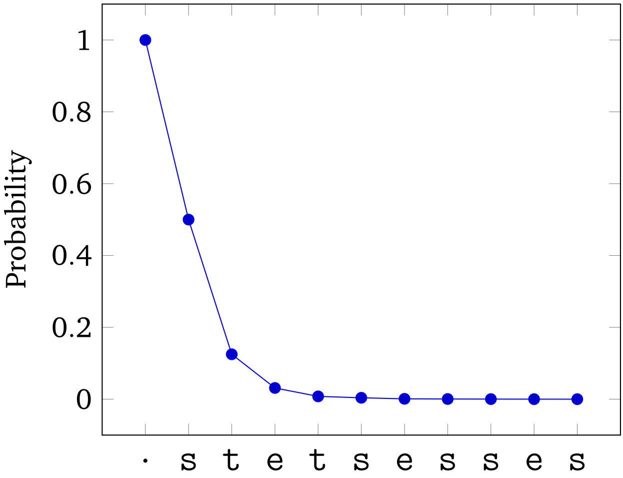
$$S = (X_1, X_2, \ldots, X_{10}),$$

by sampling letters from the distribution

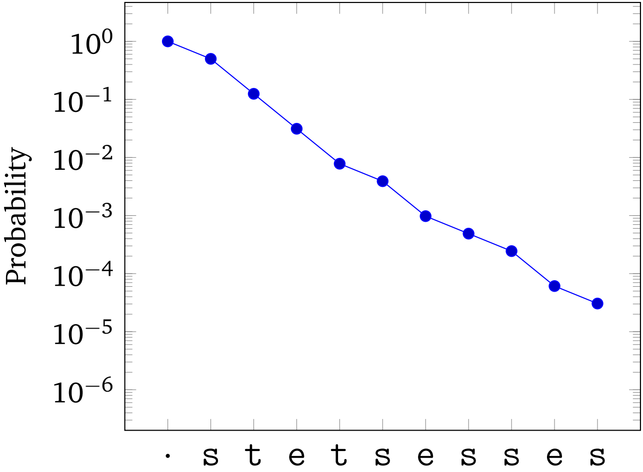
x	t	s	e
$P_X(x)$	$1/4$	$1/2$	$1/4$

1. What is $P_S(\texttt{stetsesses})$?
2. What is the most probable sequence?

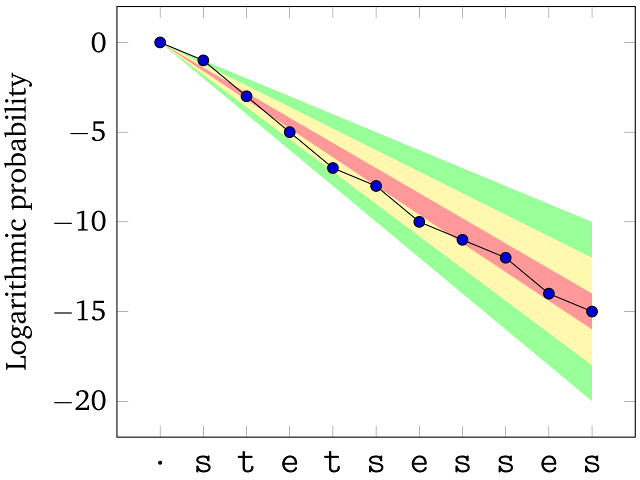
Sequence Probabilities



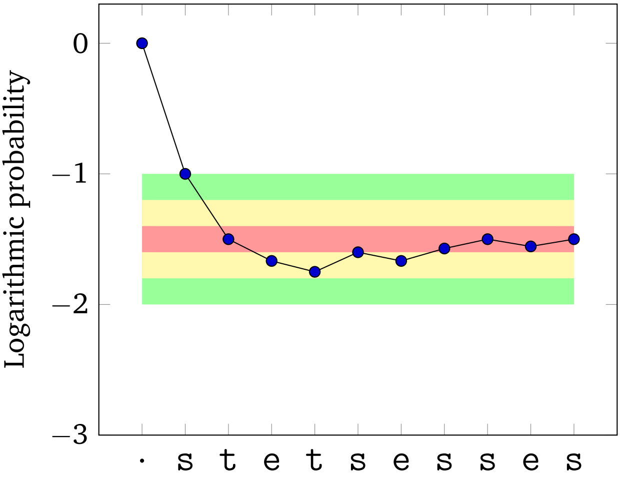
Sequence Probabilities



Sequence Probabilities



Sequence Probabilities



Typical Sequences

Definition

The **entropy** of a random variable X is

$$H = E\left[\log \frac{1}{P_X(X)}\right] = -E\left[\log P_X(X)\right].$$

Definition

An ε**-typical sequence** is a sequence $s = (x_1, x_2, \ldots, x_n)$ for which

$$\left|\frac{1}{n} \log \frac{1}{P_S(s)} - H\right| < \varepsilon.$$

Typical Sequences

The Law of Large Numbers

Eventually, almost all sequences are typical ($-\frac{1}{n}\log P_S(s) \approx H$).

The Asymptotic Equipartition Property

Eventually, everything has the same probability ($P_S(s) \approx 2^{-nH}$).

Shannon's Source Coding Theorem

Eventually, there are only 2^{Hn} sequences worth caring about.

Typical Sequences