ILLC Project Course in Information Theory

Crash course

13 January – 17 January 2014
12:00 to 14:00

Student presentations

27 January – 31 January 2014
12:00 to 14:00

Location

ILLC, room F1.15,
Science Park 107, Amsterdam

Materials

informationtheory.weebly.com

Contact

Mathias Winther Madsen
mathias.winther@gmail.com

Monday

Probability theory
Uncertainty and coding

Tuesday

The weak law of large numbers
The source coding theorem

Wednesday

Random processes
Arithmetic coding

Thursday

Divergence
Kelly Gambling

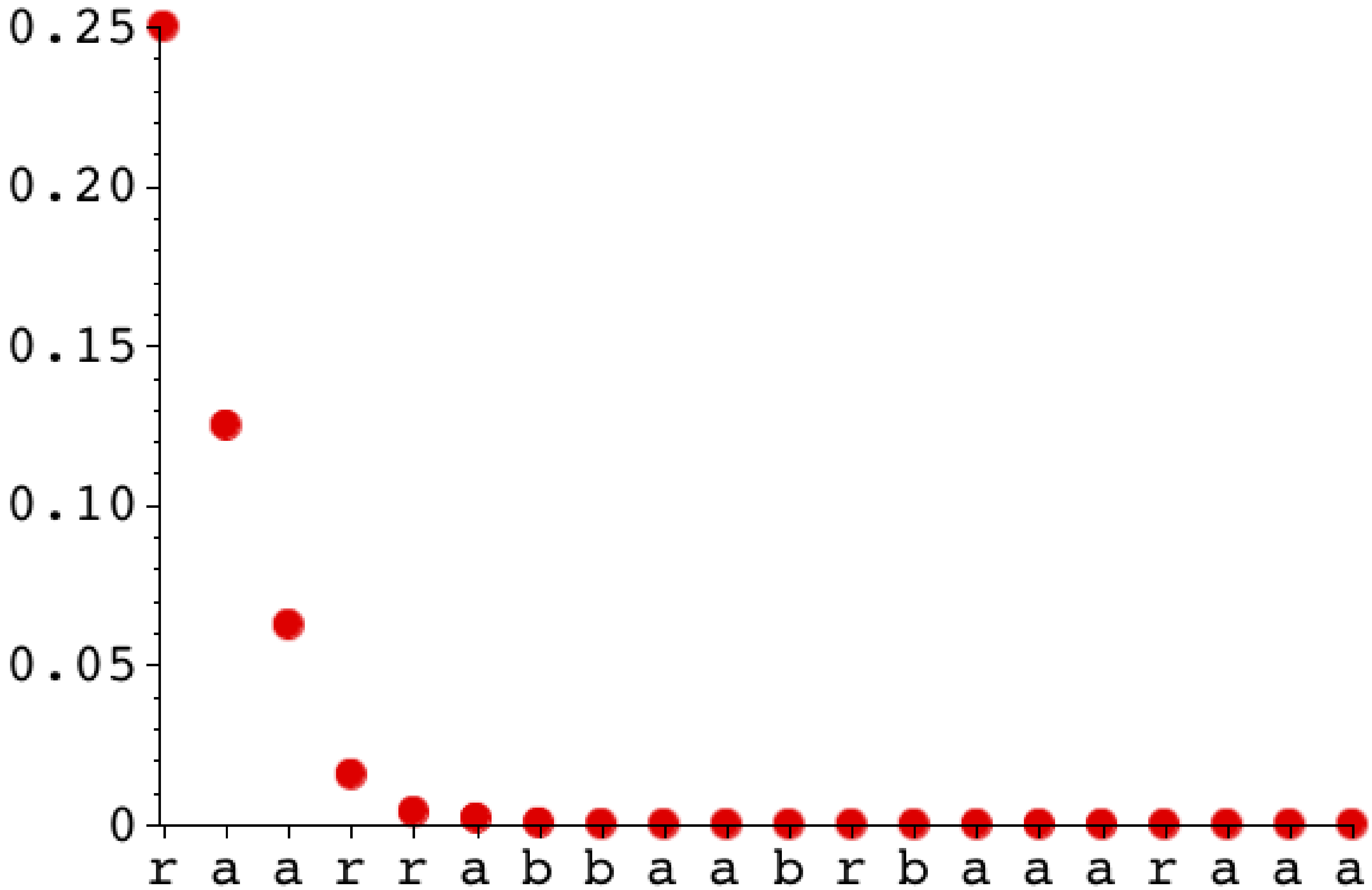
Friday

Kolmogorov Complexity
The limits of statistics

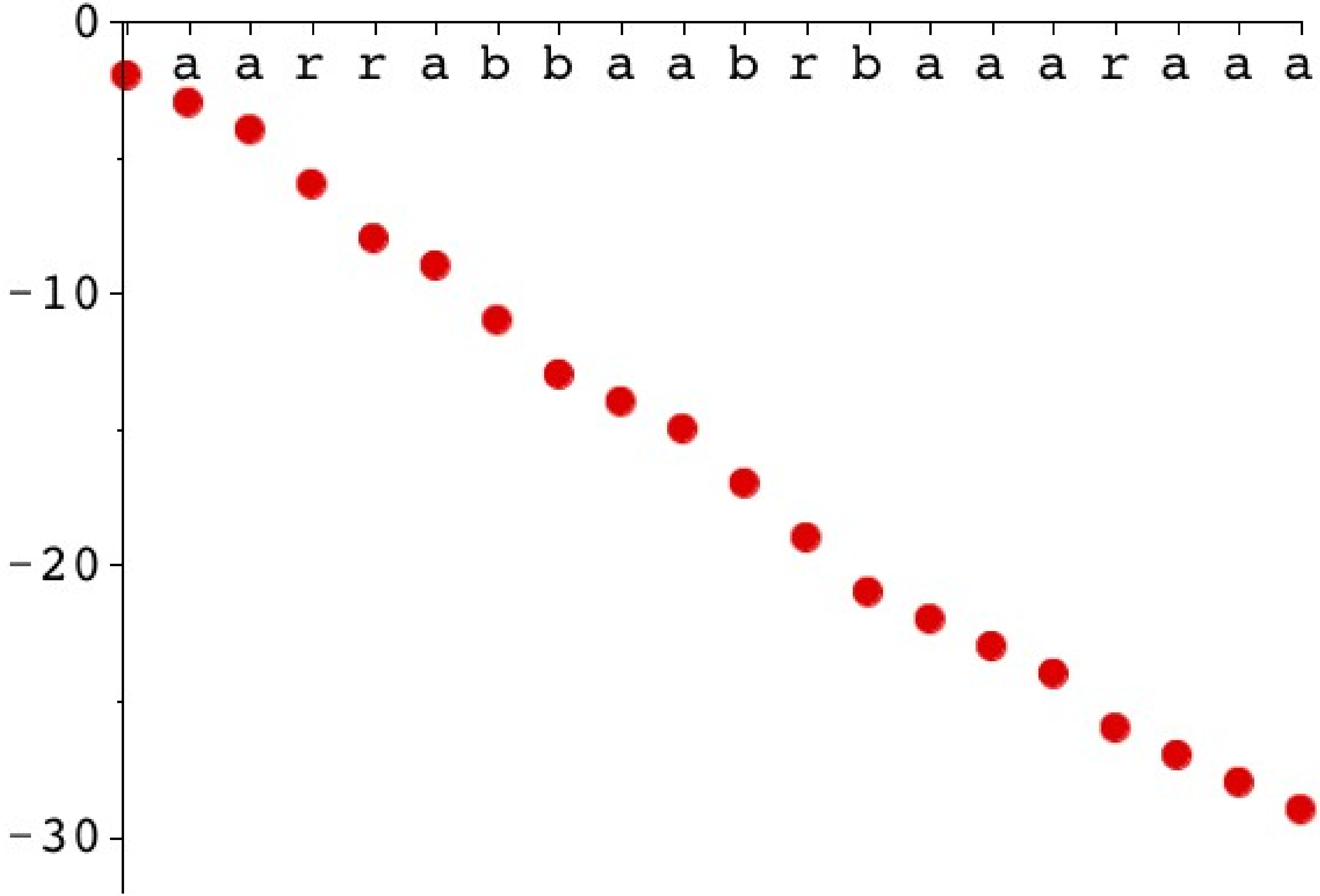
x	`'b'`	`'a'`	`'r'`
$\Pr(X=x)$	1/4	1/2	1/4

What is Pr(`'barrabarbara'`)?

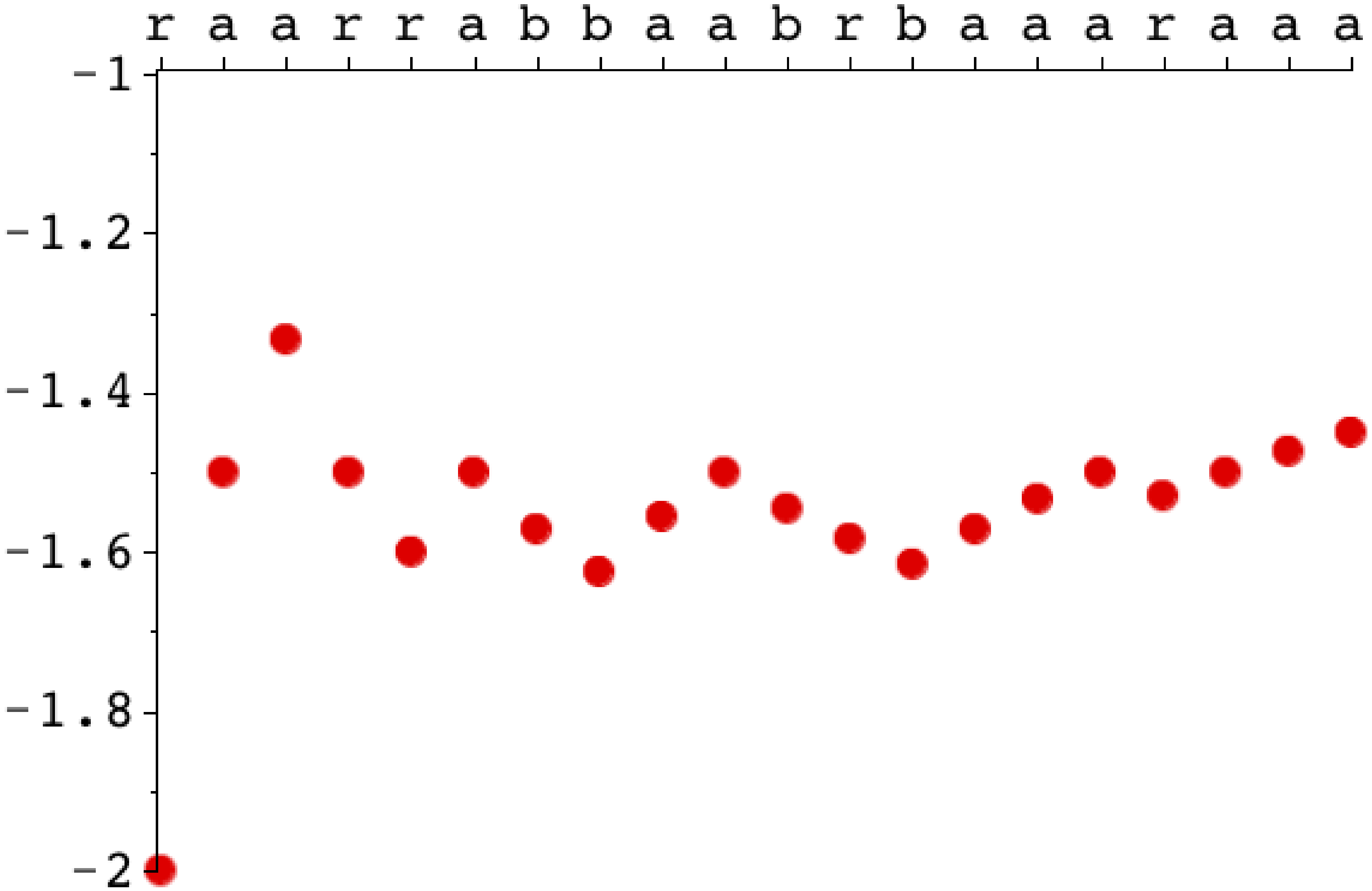
What is the most probable sequence of letters from this distribution?



$$\Pr = (1/4) \cdot (1/2) \cdot (1/2) \cdot (1/4) \cdot (1/4) \cdot \cdots$$



$$\log \Pr = -2 - 1 - 1 - 2 - 2 - 1 - 2 - 2 - \cdots$$



$$\overline{\log \Pr} = -2/1 - 3/2 - 4/3 - 6/4 - 8/5 - 9/6 \ \cdots$$

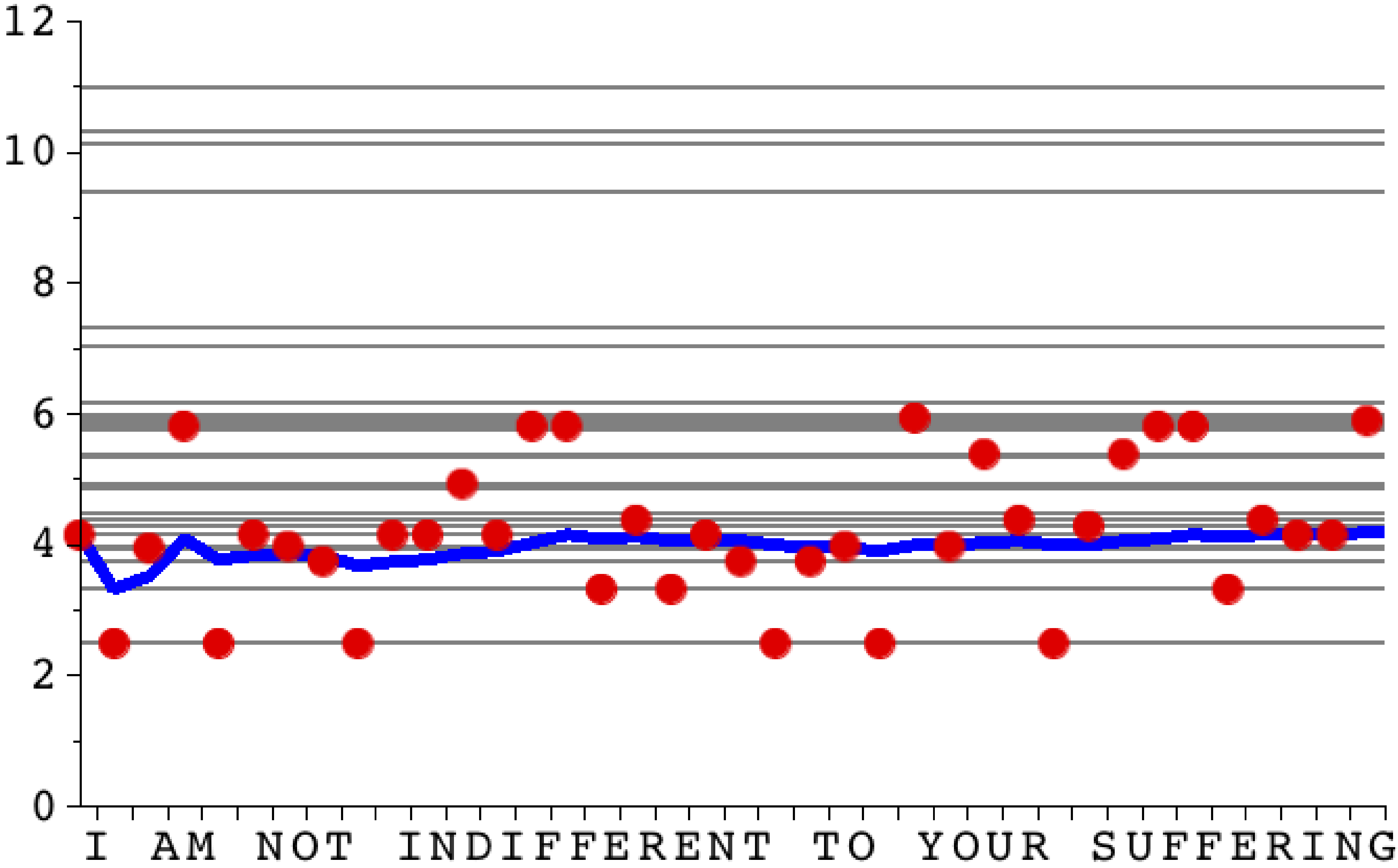
The Weak Law of Large Numbers

Let X_1, X_2, X_3, ... be a series of **independent** and **identically distributed** random variables with a (common) entropy $H(X)$.

Then for every $\varepsilon, \tau > 0$ there exists an N such that

$$\left| \log \frac{1}{\Pr(X_1, X_2, X_3, \ldots X_N)} - H(X) \right| < \varepsilon$$

with probability at least $1 - \tau$.



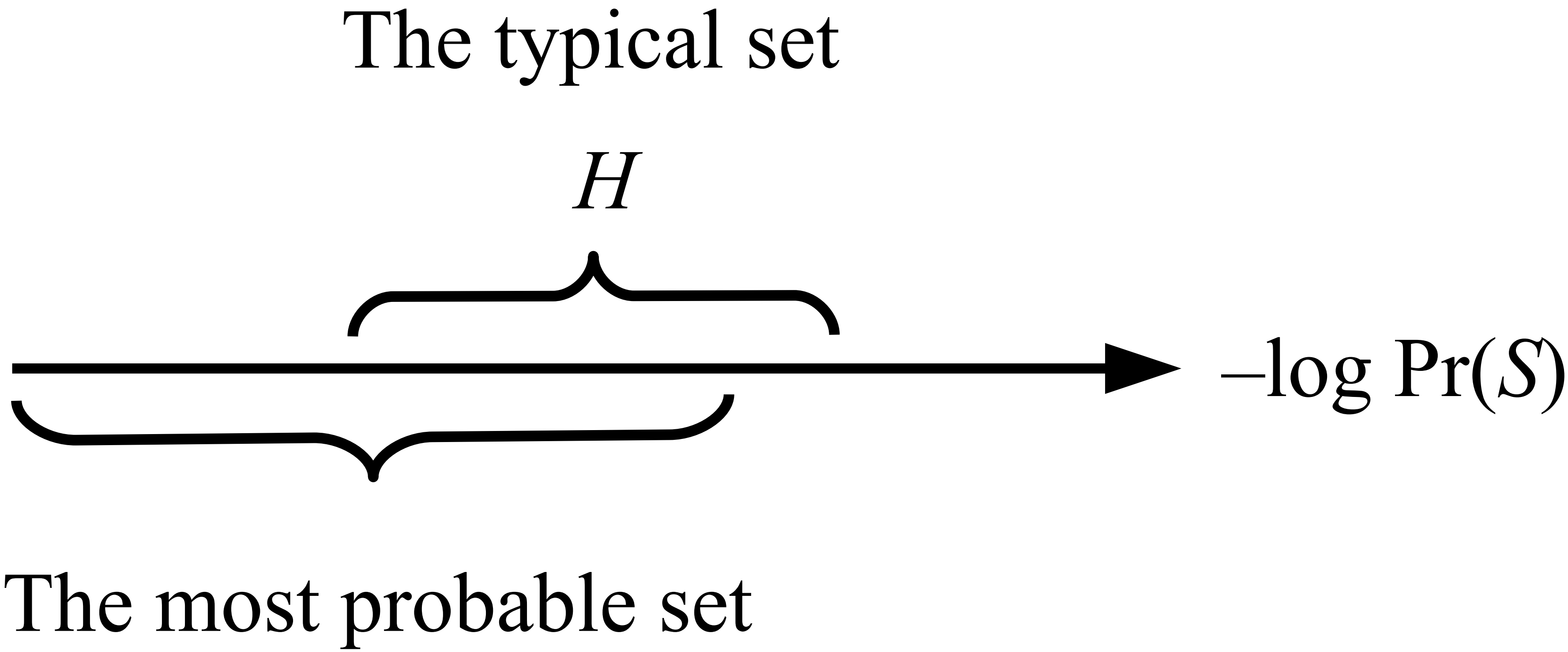
$H = 4.1$

An **typical sequence** $S = (x_1, x_2, \ldots x_n)$ is a sequence for which $| -\log \Pr(S) - H | < \varepsilon$.

The **typical set** is the set of typical sequences.

Observations:

- All typical sequences have probability "close" to 2^{-nH} (on a logarithmic scale).
- When n is large, a randomly sampled sequence be typical with probability slightly short of 1.
- There are consequently slightly more than 2^{nH} typical sequences.



The size of the typical set divided by the size of the most probable set is $\leq 2^{n\varepsilon}$, where ε measures the "radius" of the typical set, and n is a function of ε.

Source coding (Cover and Thomas, Exercise 3.7) An information source produces produces pixels $X_1, X_2, X_3, \ldots$ with $\Pr(X_i = \texttt{WHITE}) = 0.995$ and $\Pr(X_i = \texttt{BLACK}) = 0.005$.

You decide to brute-force encode outputs from this source, 100 pixels at a time, by means of a table of equally long codewords. You include all sequences with three or fewer black pixels in the table and accept there will be an error in the remaining cases.

1. Compute or estimate the number of codewords you will need for this encoding scheme.
2. How many bits do you need in order to encode that many sequences, and how does that compare to the theoretical minimum?
3. What are your options for improving this performance, theoretically and practically?
4. Find out how likely this encoding scheme is to encounter an untabulated sequence, either by using the Markov bound or by summing the binomial probabilities directly.

Probability threshold sets (Cover and Thomas, Exercise 3.5) Let $X_1, X_2, X_3, \ldots$ be a series of independent random variables drawn from a distribution X with entropy $H(X)$. Let further $C_n(\tau)$ be the set of all high-probability sequences $(x_1, x_2, \ldots, x_n)$ for which

$$\Pr(X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots, X_n = x_n) \geq 2^{-n\tau}.$$

1. What's the highest number of elements such a set can contain?
2. Sketch a graph of $\Pr C_n(\tau)$ as a function of τ for a large value of n, and for an extremely large value of n.