ILLC Project Course in Statistical Learning Theory

Mathias Winther Madsen
mathias.winther@gmail.com

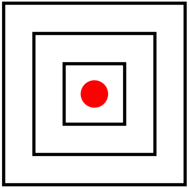
Institute for Logic, Language, and Computation
University of Amsterdam

January 2015

Symmetrization

The infinity norm

$$\|f - g\| \;:=\; \sup_{A \in S} |f(A) - g(A)|\,.$$



This satisfies the triangle inequality.

Uniform convergence

$$\lim_{t \to \infty} \Pr\{\|f - p\| > \varepsilon\} \;=\; 0$$

Predictive accuracy

$$\lim_{t \to \infty} \Pr\{\|f_1 - f_2\| > \varepsilon\} \;=\; 0$$

Symmetrization

Symmetrization

$$\Pr\{\|f_1 - f_2\| > \varepsilon\} \quad \leq \quad 1 - \left(1 - \Pr\left\{\|f - p\| > \frac{\varepsilon}{2}\right\}\right)^2$$

$$\Pr\{\|f_1 - f_2\| > \varepsilon\} \quad \geq \quad \Pr\{\|f - p\| > 2\varepsilon\} \times \left(1 - \frac{1}{\varepsilon^2 t}\right)$$

or equivalently,

$$\Pr\{\|f - p\| > \varepsilon\} \quad \geq \quad 1 - \sqrt{1 - \Pr\{\|f_1 - f_2\| > 2\varepsilon\}}$$

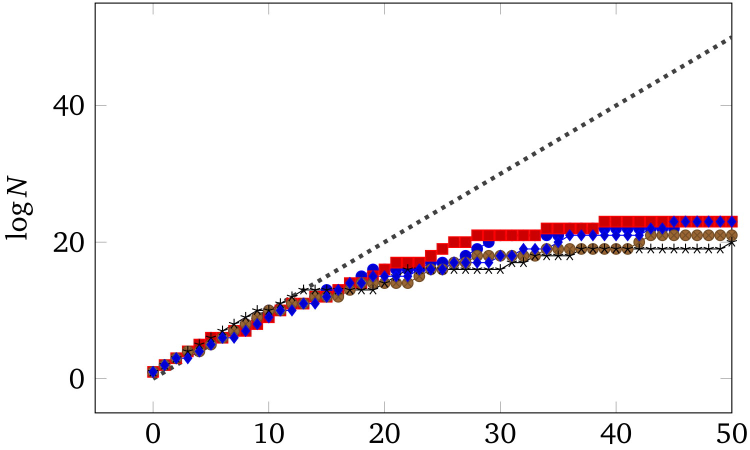
$$\Pr\{\|f - p\| > \varepsilon\} \quad \leq \quad \Pr\left\{\|f_1 - f_2\| > \frac{\varepsilon}{2}\right\} \times \left(\frac{\varepsilon^2 t}{\varepsilon^2 t - 4}\right)$$

Annealed Entropy

Suppose S contains all subsets of $\mathbb{N}$.

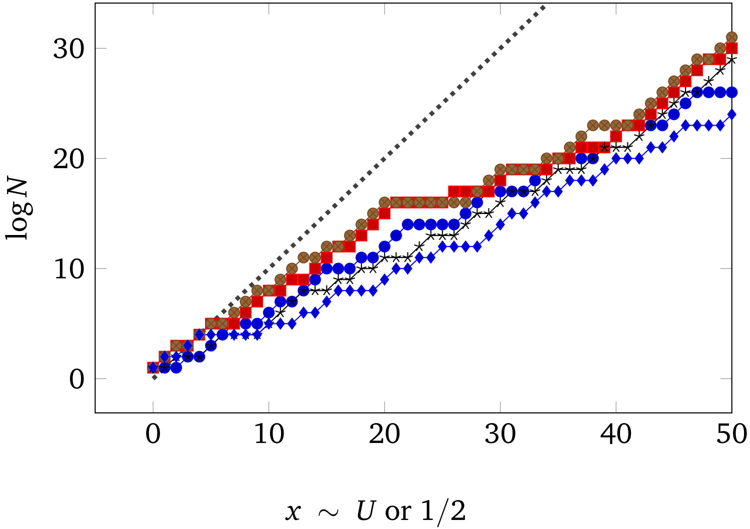
1. How does $N(t)$ grow?
2. How will $N(t \mid x)$ typically grow when x is a randomly sampled i.i.d. sequence?

Annealed Entropy



$x \sim$ Geometric(0.9).

Annealed Entropy



Annealed Entropy

Definition

The **annealed entropy rate** of a portfolio is

$$\bar{H} = \lim_{t \to \infty} \frac{E[\log N(t)]}{t}.$$

This depends on the underlying distribution.

Theorem

$$\Pr\left\{ \left| \frac{\log N(t)}{t} - \bar{H} \right| > \varepsilon \right\} \to 0, \qquad t \to \infty.$$

Necessary and Sufficient Conditions

Theorem

The frequencies of the sets in S will converge uniformly to their probabilities if and only if

$$\bar{H} = 0.$$

Proof.

We can spoil convergence at a rate of $\bar{H}$:

11111111 011010110010 00000000 101101110010

□