Martin Löf test (deficiency of randomness)

$d : \Sigma^* \to [0,\infty)$ such that (1) d is lowersemicomputable, (2) for all k, for all sufficiently large n,

$$\sum_{\substack{x \in \Sigma^n, \\ d(x) > k}} \Pr(x) < 2^{-k}.$$

The ML tests we consider.

(1) for WLLN

(2) for compressibility: $d_0(x) = |x| - K(x \mid |x|)$.

Martin-Löf test for weak law of large numbers.

A "random" string will half zeroes (approx.)

Formalize:

Empirical Distribution of a string $z \in \{0,1\}^*$.

Let $|z| = n$.

Let $N(z,0)$ be the number of 0s in the string,

& $N(z,1)$ " 1s in the string.

Note: $N(z,0) + N(z,1) = n$.

Hence $\left(\frac{N(z,0)}{n}, \frac{N(z,1)}{n}\right)$ is a probability distribution on $\{0,1\}$.

If z obeys WLLN, then the empirical distribution will be "approximately" $(\frac{1}{2}, \frac{1}{2})$. Otherwise, it could be very "far away from" $(\frac{1}{2}, \frac{1}{2})$.

The MLtest is KL divergence of the empirical distribution of the given string from $(\frac{1}{2}, \frac{1}{2})$.

Define for any string y, the empirical distribution of y induced by x by

$$P_x(y) = \underbrace{p_x}_{\frac{N(x,0)}{|x|}}^{N(y,0)} (1-p_x)^{N(y,1)}$$

p_x is a probability distribution on $\{0,1\}^n$, $|x|=n$.

$$\sum_{y \in \{0,1\}^n} p_x(y) = \sum_{y \in \{0,1\}^n} p_x^{N(y,0)} (1-p_x)^{N(y,1)}$$

$$= \sum_{k=0}^{n} \sum_{\substack{y \in \{0,1\}^n \\ N(y,0)=k}} p_x^{N(y,0)} (1-p_x)^{N(y,1)}$$

$$= \sum_{k=0}^{n} \binom{n}{k} p_x^{k} (1-p_x)^{n-k}$$

$$= (p_x + 1 - p_x)^n = 1.$$

The ML test is:

$$d(x) = \underbrace{\log\left(\frac{P_x(x)}{2^n}\right)}_{\text{“} D(P_x \,\|\, (1/2, 1/2)) \text{”}} - \underbrace{\log(n+1)}_{\text{“normalisation factor”}}$$

(1) d is computable

(2) See Notes: $\Pr\left[x \in \{0,1\}^n \mid d(x) > k\right] < 1/2^k$.

d_0 tests for incompressibility $= |x| - K(x \mid |x|)$

d tests for deviation from WLLN $= \log\left(\frac{P_x(x)}{2^n}\right) - \log(n+1)$.

If we show that $\forall z$ $d_0(z) \geq d(z)$, then if a string is incompressible, it obeys WLLN.

$x \longmapsto x.$

Consider strings x such that $d(x) > |x| - k$. There are at most 2^k such strings, since d is a ML test. Then

$$E = \{ (x, y, m) \mid y = |x| \text{ and } d(x) > |x| - k \}.$$

We can show E is a computably enumerable set since d is lower semicomputable

Fix y, m. Then E has at most $O(1) \cdot 2^m$ strings of the form (x, y, m).

Using E & the index of x in E, we can show that $K(x|y) \approx K(x|y, m)$

$< m + O(1)$.

$\Rightarrow K(x \mid |x|) < m + O(1) \Rightarrow d_0(x) > m + O(1).$