**presenter:** Elchanan Mossel

**title:** U.C. Berkeley

**problem:**

For every sequence $x \in \{0,1\}^n$ define a distribution $P_x$ over all binary sequences of length at most $n$. $P_x$ is defined as follows. For $y \in \{0,1\}^r$ with $r \leq n$ let:

$$P_x(y) = p^r(1-p)^{n-r}\left|\left\{(i_1,\ldots,i_r) : 1 \leq i_1 < \ldots < i_r \leq n, \quad x(i_1) = y(1),\ldots,x(i_r) = y(r)\right\}\right|.$$

In words - $P_x$ is the distribution over strings obtained from $x$ by keeping each bit with probability $p$ and deleting it with probability $1-p$, where different deletions are independent.

Question: What is the asymptotics of

$$d(n) = \min\left\{d_{TV}(P_x, P_y) : x, y \in \{0,1\}^n, \quad x \neq y\right\}?$$

Here $d_{TV}$ is the total variation (or $L_1$) distance.

**background:**

I made this conjecture while trying to improve results by Thomas Holenstein, Michael Mitzenmacher, Rina Panigrahy and Udi Wieder from 2008 titled: Trace reconstruction with constant deletion probability and related results.

HMPU were interested in the algorithmic problem of reconstructing $x$ from samples of $P_x$.

If $d(n)$ grows polynomially with $n$ this will imply that statistically it is possible to recover $P_x$ from $poly(n)$ samples for all $x$. However it will not solve the algorithmic problem. The results of HMPU imply in particular that

$$\exp(-n^{0.5+o(1)}) \leq d(n) \leq O(n^{-1/2}).$$