Hoeffding/Chernoff bound

- p = true bias of a coin
- $\widehat{p}$ = observed bias of a coin
- m = number of examples

$$Pr(p - \widehat{p} > \epsilon) \leq e^{-2\epsilon^2 m}$$

Sanov bound

- $D(q||p) = q \ln \frac{q}{p} + (1-q) \ln \frac{1-q}{1-p}$

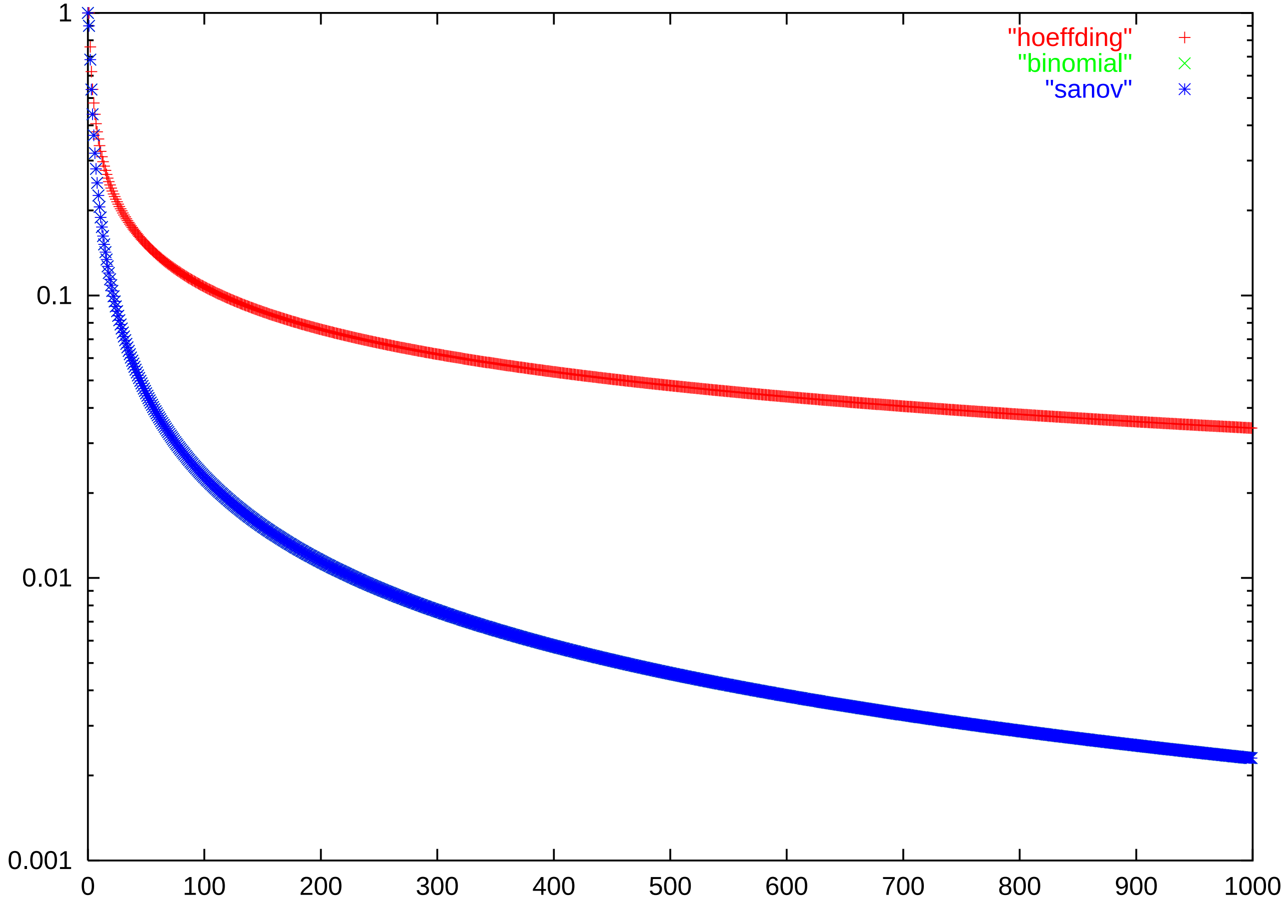
$$Pr(p - \widehat{p} > \epsilon) \leq e^{-mD(p-\epsilon||p)}$$

Binomial Equality

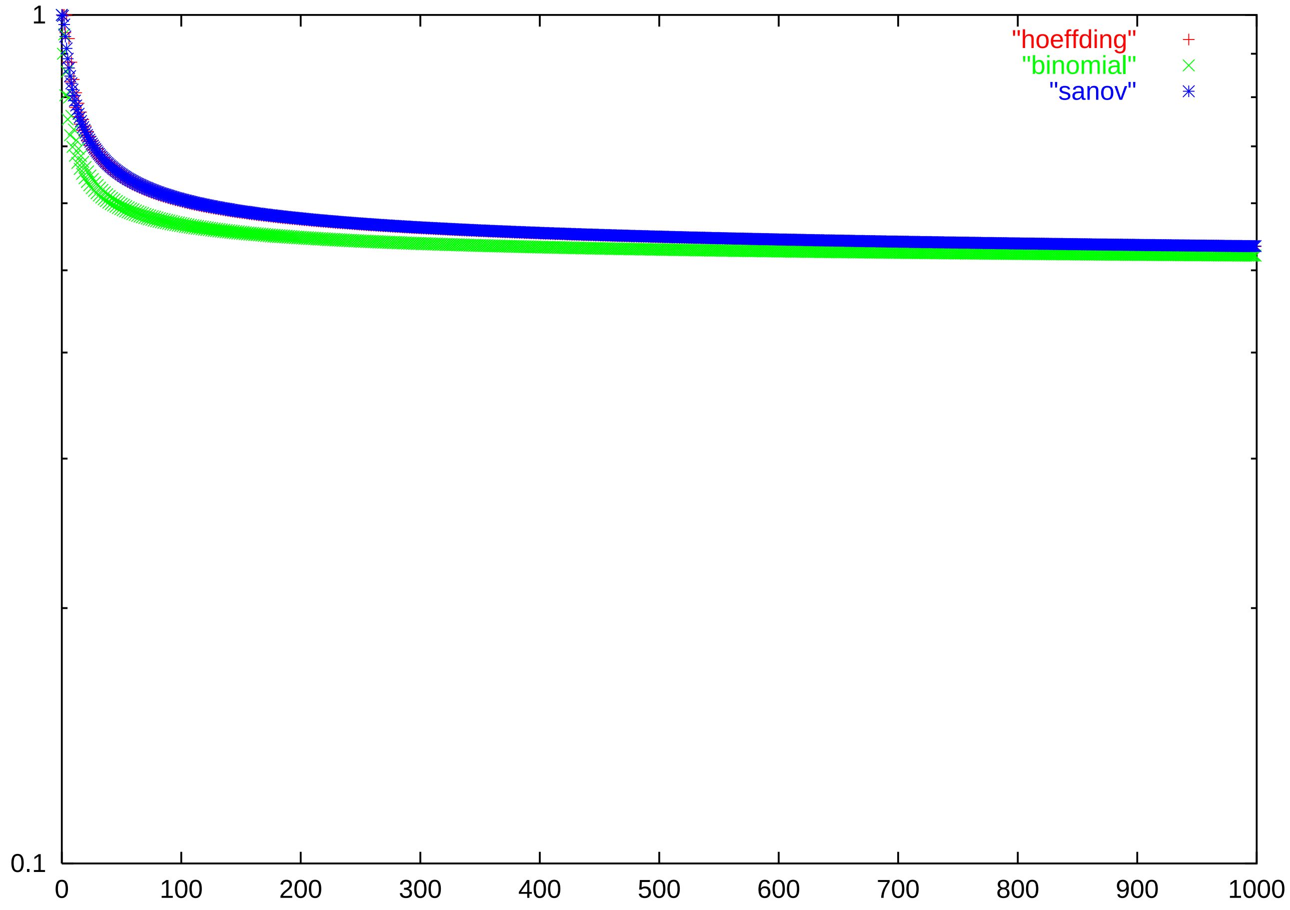
- $Bin(p, m, \widehat{p}) = \sum_{i=0}^{\widehat{p}} \left(\begin{array}{c} m \\ i \end{array} \right) p^i (1-p)^{m-i}$

$$Pr(p - \widehat{p} > \epsilon) = Bin(p, m, \widehat{p})$$

$\widehat{p} = 0$


"hoeffding"
"binomial"
"sanov"
1
0.1
0.01
0.001
0
100
200
300
400
500
600
700
800
900
1000

$\widehat{p} = 0.5$



1
0.1
0
100
200
300
400
500
600
700
800
900
1000
"hoeffding"
"binomial"
"sanov"

Basic PAC bound

- $e(h)$ = true error
- $\widehat{e}(h)$ = empirical error

$$Pr(\exists h \in H : e(h) - \widehat{e}(h) > \epsilon) \leq |H|e^{-m\epsilon^2}$$

Improvable to:

$$Pr(\exists h \in H : e(h) - \widehat{e}(h) > \epsilon) \leq |H|Bin(e(h), m, \widehat{e}(h))$$

Holdout Bound

- $\widehat{e}_{hold}(h)$ = empirical error on holdout set
- m_{hold} = the number of examples in the holdout set

$$Pr(\exists h \in H : e(h) - \widehat{e}_{hold}(h) > \epsilon) \leq Bin(e(h), m_{hold}, \widehat{e}_{hold}(h))$$

Microchoice Bound

- d = number of choice spaces
- C_i = ith choice space

$$Pr(\exists h \in H : e(h) - \hat{e}(h) > \epsilon) \leq Bin(e(h), m, \hat{e}(h)) \prod_{i=1}^{d} |C_i|$$

Distribution Dependent Bound

$$Pr(\exists h \in H, r \in [0...m] : e(h) - \frac{r}{m} > \epsilon) \leq m \sum_{e(h) \geq \frac{r}{m} + \epsilon} Bin(e(h), m, \frac{r}{m})$$

But don't know $e(h)$ so:

$$Pr(\exists h \in H, r \in [0...m] : e(h) - \frac{r}{m} > \epsilon) \leq \sqrt{8m \sum_{\hat{e}(h)} Bin(\max(\hat{e}(h), \frac{r}{m} + \epsilon), m, \frac{r}{m})}$$

But don't know $\hat{e}(h)$ so need to sample ... and bound sampling error.

Combined Bound

δ = probability of failure

Min over previous bounds except holdout with $\delta \to \frac{\delta}{3}$

Decision tree techniques:

- Pruning
- Super sampling (and drawbacks)

	m	$\widehat{e}_{ut}$	$\widehat{e}_{pt}$	k_u	k_p	e_{PAC}	e_{mcs}	e_{dd}	e_{comb}	$e_{holdout}$
shuttle	12	.33	.33	4	1	1	.82	.63	.69	0.95
yellow1	13	.67	.67	4	3	.76	.79	.75	.78	1
adult1	16	0	0	2	2	.58	.56	.44	.50	0.54
lung	26	0.5	0.5	7	1	1	.92	.75	.78	0.94
postop	72	0.5	0.5	28	0	1	.57	.27	.37	0.75
scale	500	0.2	.23	94	11	.88	.52	.37	.40	0.33
vote	348	.08	.02	20	1	1	.17	.07	.09	0.14
soy	246	.42	.52	27	11	1	.94	.87	.87	0.66
audio	160	.43	.68	87	3	1	.98	.89	.90	0.84
car	1383	.08	.10	85	37	.87	.32	.28	.29	0.16
krkp	2557	.005	.008	44	20	1	.10	.09	.10	0.05
shroom	6500	0	0	11	9	1	.017	.016	.017	.027
krk	22445	.84	.84	347	37	1	1	1	1	.85
nursery	10368	.006	.007	212	159	1	.12	.11	.11	.03