Mehryar Mohri
Foundations of Machine Learning
Courant Institute of Mathematical Sciences
Homework assignment 1
Due: February 6, 2007

A. Senate Laws

For important questions, President Mouth relies on expert advice. He selects an appropriate advisor from a collection of $H = 2800$ experts.

1. Assume that laws are proposed in a random fashion independently and identically according to some distribution D determined by an unknown group of senators. Assume that President Mouth can find and select an expert senator out of H who has consistently voted with the majority for the last $m = 200$ laws. Give a bound on the chances that such a senator incorrectly predicts the global vote for a future law. What is the value of the bound with 95% confidence?

2. Assume now that President Mouth can find and select an expert senator out of H who has consistently voted with the majority for all but $m' = 20$ of the last $m = 200$ laws. What is the value of the new bound?

B. PAC Learning of Hyper-rectangles

1. An axis-aligned hyper-rectangle in $\mathbb{R}^n$ is a set of the form $[a_1, b_1] \times \ldots \times [a_n, b_n]$. Show that axis-aligned hyper-rectangles are PAC-learnable by extending the proof given in class for the case $n = 2$.

C. Bound Comparison

Let $X_1, \ldots, X_m$ be a sequence of random variables taking values in $[0, 1]$ with the same mean μ and variance $\sigma^2 < \infty$ and let $\overline{X} = \frac{1}{m} \sum_{i=1}^{m} X_i$.

1. For any $\epsilon > 0$, give a bound on $\Pr[|\overline{X} - \mu| > \epsilon]$ using Chebyshev's inequality, then Hoeffding's inequality. For what values of σ is Chebyshev's inequality tighter?

2. Assume that the random variables X_i take values in $\{0, 1\}$. Show that $\sigma^2 \leq \frac{1}{4}$. Use this to simplify Chebyshev's inequality. Choose $\epsilon =$

.05 and plot Chebyshev's inequality thereby modified and Hoeffding's inequality as a function of m [you can use your favorite program for plotting this, e.g., matlab, R].