# Statistical Learning Theory, Exercise 4

Michael Hirsch

January 15, 2015

## Bounds on Membership Uncertainty

A sample size of $2t = 2 \times 10^4$ is drawn from some distribution, and this sample is then randomly split up into two half-samples of size $t = 10^4$

1. For any specific event $A$, these two half-samples define two frequencies, $f_1(A)$ and $f_2(A)$. Find an explicit upper bound on the probability that $|f_1(A) - f_2(A)| > 0.1$.

   In this scenario, we have two samples both of size $t = 10^4$, with the total population size, $T$, being unknown, as well as the number of successes in both the population and the sample.

   Further, we know that $E[s] = S/T \times 10^4$ in both samples.

   Since we can assume that $S$ is fixed, we have the equality:

$$Pr\{|f_1(A) - f_2(A)| > 0.1\} = Pr\{|f_1(A) - E[f_1(A)]| > 0.05\} \leq 2 \times exp(-2t(\frac{\varepsilon}{2})^2)$$

   □

2. We now make such a comparison for each $\Phi(3, 2 \times 10^4$ different sets. Find an explicit upper bound on the probability that $|f_1(A) - f_2(A)| > \varepsilon$ for at least one $A$.

   □