COS 341, October 8, 1997
Handout Number 8

About Polling

Consider an upcoming election in a large metropolitan area, with a large voting population. Two candidates A and B are running. If you are a pollster, taking a random sample of size n (n is a moderate number compared with the population). How confident are you about the result you get? If your result has m votes for A, how do you decide on ϵ so that you can say that "the result is m/n for A, $(n-m)/n$ for B, with a margin of error of ϵ"?

You can fairly accurately model your polling as follows. Let $0 < p < 1$ be the true fraction of voters for A. Throw a biased (with bias p for $HEAD$ and $1-p$ for $TAIL$) coin n times. The HEADs are for A and TAILs are for B. Let X denote the random variable corresponding to the total number of HEADs.

You polling corresponds to making an observation of the value of X, say with result m. By Chebyshev's Inequality, m is likely to be within a few standard deviations of the expected value of X. That means in turn that pn (which is $E(X)$) is within a few standard deviation of m. That is, pn is within $\sigma(x)$ of m. Hence p is within $\sigma(X)/n$ of your polling result m/n.

It is easy to calculate $\sigma(X)$. Clearly, $p_k = \Pr\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}$. The generating function $F(x) = \sum_{0 \leq k \leq n} p_k x^k$, which is equal to $((1-p)+px)^n$ by the Binomial Theorem. It follows that $F'(x) = np((1-p)+px)^{n-1}$ and $F''(x) = n(n-1)p^2((1-p)+px)^{n-2}$. From this, we obtain $Var(X) = F''(1) + F'(1) - F'(1)^2 = np(1-p)$. This leads to $\sigma(X) = (Var(X))^{1/2} = (np(1-p))^{1/2}$. Now, $p(1-p) \leq 1/4$ (?? Can you prove it?). Thus, $\sigma(X) \leq \sqrt{n}/2$. Thus, from the discussions in the last paragraph, it is reasonable to say that your polling result has a margin of error about $1/(2\sqrt{n})$.