Hoeffding's Inequality

Suppose that we have a biased coin with probability of heads p and probability of tails $1-p$. If we flip the coin N times, assuming each flip is independent, how many times do we have to flip it to have a good sense of the value of p? To connect this problem to machine learning and neuroscience – how challenging is it to *learn* the value of p? A very simple answer is provided by **Hoeffding's inequality**. Let $X_i = 1$ if the i^{th} flip is heads, and $X_i = 0$ if it is tails. Let us *estimate* p with $\widehat{p}$, where

$$\widehat{p} \equiv \frac{1}{N}\sum_{i=1}^{N} X_i.$$

The inequality we're looking for states that, *regardless of* p, we can estimate p with exponentially vanishing error:

$$\mathrm{P}(|\widehat{p}-p| > \epsilon) \le 2\mathrm{e}^{-2\epsilon^2 N}.$$

This problem will show you how to prove this inequality.[1]

(a) Explain why, for any real h:

$$\mathrm{P}(\widehat{p}-p > \epsilon) \le \mathrm{e}^{-(p+\epsilon)hN}\prod_{i=1}^{N}\left\langle \mathrm{e}^{hX_i}\right\rangle.$$

(b) Show that

$$\mathrm{e}^{-(p+\epsilon)hN}\prod_{i=1}^{N}\left\langle \mathrm{e}^{hX_i}\right\rangle \le \left[\left(\frac{p}{p+\epsilon}\right)^{p+\epsilon}\left(\frac{1-p}{1-p-\epsilon}\right)^{1-p-\epsilon}\right]^N.$$

(c) Show that

$$\mathrm{P}(\widehat{p}-p > \epsilon) \le \mathrm{e}^{-2\epsilon^2 N}.$$

(d) By checking the probability that $\widehat{p}$ is smaller than p, prove Hoeffding's inequality.

(e) Is it harder to learn the value of p to high accuracy or high precision? Heuristically explain why.

[1] Thanks to Arthur Safira for suggesting this problem.