Chernoff and Hoeffding Bounds

Let us consider a sequence of independent random variables $X_1, X_2, \cdots, X_m$ such that

$$X_i = \begin{cases} 1 & with\ probability\ p \\ 0 & with\ probability\ 1-p \end{cases}$$

The sample mean is given by

$$\hat{p} = \frac{1}{m}\sum_{i=1}^{m} X_i .$$

Here, we want to get the upper bound of the probability

$\Pr\{\hat{p} > q\}$ for any $q > p$.

For any $\lambda > 0$,

$$\Pr\{\hat{p} > q\} = \Pr\{e^{m\lambda\hat{p}} > e^{m\lambda q}\}$$

since $e^{m\lambda x}$ is monotone increasing in x.

By taking the expectation of $e^{m\lambda x}$, we get

$$E[e^{m\lambda\hat{p}}] \geqq e^{m\lambda q}\Pr\{e^{m\lambda\hat{p}} > e^{m\lambda q}\} + 0 \cdot \Pr\{e^{m\lambda\hat{p}} \leqq e^{m\lambda q}\}.$$

This implies that

$\Pr\{\hat{p} > q\} \leqq e^{-m\lambda q}E[e^{m\lambda\hat{p}}]$ and

$$E[e^{m\lambda\hat{p}}] = E\left[e^{\lambda\sum_{i=1}^{m}X_i}\right] = E\left[\prod_{i=1}^{m}e^{\lambda X_i}\right] = \prod_{i=1}^{m}E[e^{\lambda X_i}]$$

$$= \prod_{i=1}^{m}[pe^{\lambda} + (1-p)e^0]$$

$$= [pe^{\lambda} + (1-p)]^m$$

Thus,

$$\Pr\{\hat{p} > q\} \leqq e^{-m\lambda q}[pe^{\lambda} + (1-p)]^m = e^{-m\lambda q + m\ln[pe^{\lambda} + (1-p)]}$$
$$\leqq e^{-mf_{\lambda}(p,q)}$$

where

$$f_{\lambda}(p,q) = \lambda q - \ln[pe^{\lambda} + (1-p)]$$

For the tightest bounds, $f_{\lambda}(p,q)$ should be maximized.

Here, the optimal value of λ is given as follows:

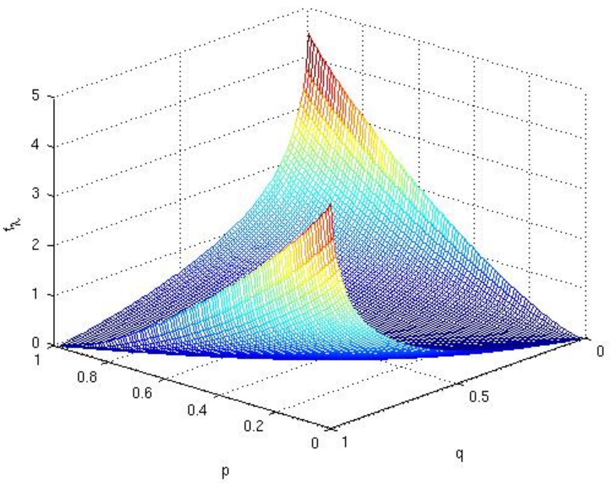
$$\frac{\partial f_{\lambda}}{\partial \lambda}|_{\lambda^*} = 0 \rightarrow \lambda^* = \ln\frac{q(1-p)}{p(1-q)}$$

That is,

$$f_{\lambda^*}(p,q) = q\ln\frac{q}{p} + (1-q)\ln\frac{1-q}{1-p}.$$

This implies that $f_{\lambda^*}(p,q)$ is the Kullback-Leibler (KL) distance between q and p, that is,

$$f_{\lambda^*}(p,q) = D(q\|p) = q\ln\frac{q}{p} + (1-q)\ln\frac{1-q}{1-p}.$$



Three important examples:

(1) $D(q\|p) \geqq 2(p-q)^2$

$\Pr\{\hat{p} > q\} \leqq e^{-2m(p-q)^2}$ (additive Chernoff bound)

Taking the union of two cases:

$\hat{p} > q > p$ and $\hat{p} < q < p$

$\Pr\{|\hat{p} - p| > \epsilon\} \leqq 2e^{-2m\epsilon^2}$ (Hoeffding bound)

(2) If $q \geqq p$, $D(q\|p) \geqq \frac{1}{3}\frac{(q-p)^2}{p}$

$\Pr\{\hat{p} > q\} \leqq e^{-\frac{1}{3}m\frac{(q-p)^2}{p}}$ (multiplicative Chernoff bound)

(3) If $q \leqq p$, $D(q\|p) \geqq \frac{1}{2}\frac{(q-p)^2}{p}$

$\Pr\{\hat{p} > q\} \leqq e^{-\frac{1}{2}m\frac{(p-q)^2}{p}}$ (multiplicative Chernoff bound)

Example: Hoeffding bound

Let p and $\hat{p}$ be the true and empirical errors. Then,

$$\Pr\{|\hat{p}-p|>\epsilon\} \leqq 2e^{-2m\epsilon^2}.$$

Assuming the finite hypothesis, that is, $|H|$ is bounded by some number. What is the sufficient number of samples to assure the best hypothesis?

For PAC learning,

$$\Pr\{|\hat{p}-p|>\epsilon\} \leqq 2|H|e^{-2m\epsilon^2} \leqq \delta.$$

$$2|H|e^{-2m\epsilon^2} \leqq \delta \rightarrow e^{-2m\epsilon^2} \leqq \frac{\delta}{2|H|} \rightarrow m \geqq \frac{1}{2\epsilon^2}\ln\frac{2|H|}{\delta}.$$