A Tight Quadratic Lower-Bound on the KL-Divergence

April 18, 2014

1 Motivation

We have the following concentration bound for bernoulli trials (Sanov's Bound): Let X `bernoulli`(p), and take n iid samples $X_1, ...X_n$ from X. Let $\hat{p} =_{def} \frac{1}{n}\sum_{i=1}^{n} X_i$ be the estimate for p based on these trials.

Then for any $q > p$, we have the following variant of Sanov's Bound:

$$P(\hat{p} > q) \leq e^{-n\mathtt{KL}(q\|p)}$$

If we have some lower bound $\alpha_p(q-p)^2 \leq \mathtt{KL}(q\|p)$, we can re-write the bound as:

$$P(\hat{p} > q) \leq e^{-n\alpha_p\epsilon^2}$$

where $\epsilon =_{def} (q-p) > 0$.

Let's say for $q < p$, we want $P(\hat{p} < q)$. Let $q' =_{def} 1-q$ and $p' =_{def} 1-p$, and $\hat{p'} = 1 - \hat{p}$. Then we have

$$P(\hat{p'} > q') \leq e^{-n\alpha_{p'}(q'-p')^2}$$

from above, which is equivalent to

$$P(\hat{p} < q) \leq e^{-n\alpha_{1-p}\epsilon^2}$$

Combining these, we have the following two-tailed bound:

$$P(|\hat{p}-p| > \epsilon) \leq e^{-n\alpha_{1-p}\epsilon^2} + e^{-n\alpha_p\epsilon^2} \leq 2e^{-n\mathtt{min}(\alpha_p,\alpha_{1-p})\epsilon^2}$$

2 Proposition

Define $\forall q \in (0,1), p \in (0,1)$

$$\mathtt{KL}(q\|p) = q\mathtt{log}\left(\frac{q}{p}\right) + (1-q)\mathtt{log}\left(\frac{1-q}{1-p}\right)$$

For each p, we want the largest α such that $\forall q > p$,

$$\alpha(q-p)^2 \leq \texttt{KL}(q\|p)$$

The α that achieves this for each p is:

$$\alpha = \begin{cases} \frac{\texttt{KL}(1-p\|p)}{(1-2p)^2} & p < 0.5 \\ \frac{1}{2p(1-p)} & p \geq 0.5 \end{cases}$$

3 Proof

3.1 Case: $p < 0.5$

For $p < 0.5$, we show that $\frac{d}{dq}\left[\texttt{KL}(q\|p) - \alpha(q-p)^2\right] = 0$ if and only if $q \in \{p, 0.5, 1-p\}$

Expanding/simplifying the derivative:

$$\frac{d}{dq}\left[\texttt{KL}(q\|p) - \alpha(q-p)^2\right]$$

$$= \texttt{log}\left(\frac{q}{1-q}\right) - \texttt{log}\left(\frac{p}{1-p}\right) - 2\alpha(q-p)$$

$$= \texttt{log}\left(\frac{q}{1-q}\right) - \texttt{log}\left(\frac{p}{1-p}\right) - 2\frac{\texttt{log}\left(\frac{1-p}{p}\right)}{1-2p}(q-p)$$

since $\alpha = \frac{\texttt{KL}(1-p\|p)}{(1-2p)^2} = \frac{(1-2p)\texttt{log}\left(\frac{1-p}{p}\right)}{(1-2p)^2} = \frac{\texttt{log}\left(\frac{1-p}{p}\right)}{1-2p}$

$$= \texttt{log}\left(\frac{q}{1-q}\right) - \texttt{log}\left(\frac{p}{1-p}\right) + 2\frac{\texttt{log}\left(\frac{p}{1-p}\right)}{1-2p}(q-p)$$

($\rightarrow$) PROVE THIS: there can be at most 3 roots.

($\leftarrow$)

$$\texttt{log}\left(\frac{q}{1-q}\right) - \texttt{log}\left(\frac{p}{1-p}\right) + 2\frac{\texttt{log}\left(\frac{p}{1-p}\right)}{1-2p}(q-p) =$$

$(q = p)$

$$\texttt{log}\left(\frac{p}{1-p}\right) - \texttt{log}\left(\frac{p}{1-p}\right) = 0$$

$(q = 1/2)$

$$\texttt{log}\left(\frac{1/2}{1/2}\right) - \texttt{log}\left(\frac{p}{1-p}\right) + 2\frac{\texttt{log}\left(\frac{p}{1-p}\right)}{1-2p}(1/2-p) =$$

$$-\texttt{log}\left(\frac{p}{1-p}\right) + \frac{\texttt{log}\left(\frac{p}{1-p}\right)}{1-2p}(1-2p) =$$

$$-\texttt{log}\left(\frac{p}{1-p}\right) + \texttt{log}\left(\frac{p}{1-p}\right) = 0$$

$(q = 1-p)$

$$\texttt{log}\left(\frac{1-p}{p}\right) - \texttt{log}\left(\frac{p}{1-p}\right) + 2\frac{\texttt{log}\left(\frac{p}{1-p}\right)}{1-2p}(1-2p)$$

$$= -2\texttt{log}\left(\frac{p}{1-p}\right) + 2\texttt{log}\left(\frac{p}{1-p}\right) = 0$$

This imples that the minimum of $\texttt{KL}(q\|p) - \alpha(q-p)^2$ must be at one of the points $q \in \{p, 0.5, 1-p, 1\}$. Since all of these points are nonnegative, we have $\texttt{KL}(q\|p) - \alpha(q-p)^2 \geq 0$ and $\texttt{KL}(q\|p) \geq \alpha(q-p)^2$.

3.2 Case: $p \geq 0.5$

$\alpha = \frac{1}{2p(1-p)}$ ***is sufficient.***

We want to show for $p > 0.5$,

$$\texttt{KL}(q\|p) \geq \frac{(q-p)^2}{p(1-p)}$$

At $q = p$, the values and derivatives at both sides are equal to 0. We will show that on $q > p$, the second derivative is strictly positive, and 0 $q = p$.

For $q > p > 1/2$, $q(1-q) < p(1-p)$. This is since $\frac{d}{dx}x(1-x) = 1 - 2x < 0$ for $x > 1/2$. Then

$$\frac{d^2}{dq^2}\texttt{KL}(q\|p) = \frac{1}{q(1-q)} > \frac{1}{p(1-p)} = \frac{d^2}{dq^2}\frac{(q-p)^2}{p(1-p)}$$

for $q > p > 1/2$

$\alpha = \frac{1}{2p(1-p)}$ ***is optimal.*** Take some $\alpha > \frac{1}{2p(1-p)}$. Then at $q = p$, the values/derivatives are equal, except the second derivative of the kl function is strictly larger than the second derivative of the quadratic function. Thus for small enough ϵ, $\texttt{KL}(p+\epsilon\|p) > \alpha(p+\epsilon-p)^2$

3.3 Useful Facts

$$\frac{d}{dq}\texttt{KL}(q\|p) = \texttt{log}\left(\frac{q}{1-q}\right) - \texttt{log}\left(\frac{p}{1-p}\right)$$

$$\frac{d^2}{dq^2}\texttt{KL}(q\|p) = \frac{1}{q(1-q)}$$

$$\texttt{KL}(1-p\|p) = (1-2p)\texttt{log}\left(\frac{1-p}{p}\right)$$