Bayesian Bounds, and Large Deviations

Tom Kealy

December 3, 2014

This is a short note, outlining some ideas between the variational Bayesian framework and Large Deviations theory.

Following [1], we state a simple lemma (the compression lemma) which can simplify some bounds obtained in PAC Bayesian learning.

Lemma 0.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, on a metric space (H, μ). Let $\psi(h)$ be a measurable function on H, and P, Q be distributions on Ω. We have:*

$$\mathbb{E}_Q(\psi(h)) - \log \mathbb{E}_P(\exp \psi(h)) \leq D(Q||P) \tag{0.0.1}$$

futher

$$\sup_{\psi} \mathbb{E}_Q(\psi(h)) - \log \mathbb{E}_P(\exp \psi(h)) = D(Q||P) \tag{0.0.2}$$

This is an elementary version of the Donsker-Varadhan formula - it's true for reasons deeper than this, but the proof is correspondingly more difficult.

Proof. For any measurable $\psi(h)$, we have:

$$\begin{aligned}
\mathbb{E}\psi(h) &= \mathbb{E}_Q \log \frac{dQ}{dP} \exp \psi(h) \frac{dP}{dQ} && (0.0.3)\\
&= D(Q||P) + \mathbb{E}_Q \log \exp \psi(h) \frac{dP}{dQ} && (0.0.4)\\
&\leq D(Q||P) + \log \mathbb{E}_Q \exp \psi(h) \frac{dP}{dQ} && (0.0.5)\\
&= D(Q||P) + \log \mathbb{E}_P \exp \psi(h) && (0.0.6)
\end{aligned}$$

The supremum is achieved if we take:

$$\psi(h) = \log \frac{dQ}{dP} \tag{0.0.7}$$

□

References

[1] On Bayesian Bounds, Arindam Banerjee