Notes on PAC-Bayes and Online Bayesian Inference

Jungbin Jun

Department of Statistics, Inha University

1 Preliminaries

This note is largely based on several excellent references, including Alquier [2021, 2024], Chérief-Abdellatif et al. [2019].

Lemma 1 (Donsker-Varadhan's Variational Formula)**.** *For any measurable bounded function $h : \Theta \to \mathbb{R}$,*

$$\log \mathbb{E}_{\theta\sim\pi}[e^{h(\theta)}] = \sup_{p\in\mathcal{S}(\Theta)} \left[\mathbb{E}_{\theta\sim p}[h(\theta)] - KL(p\|\pi)\right]$$

2 Online Bayesian Inference

For Bayesian learning, thanks to Donsker-Varadhan's variational formula[e.g. See Corollary 2.3 of Alquier [2024]], we can view bayesian update as following:

$$\begin{aligned}
&p_{t+1}(\theta) \propto \exp(-\eta \sum_{i=1}^{t} \ell_t(\theta))\pi(\theta) \\
&\iff p_{t+1}(\theta) \propto \exp(-\eta\, \ell_t(\theta))p_t(\theta) \\
&\iff p_{t+1}(\theta) = \arg\min_{p\in\mathcal{S}(\Theta)} \left\{ \mathbb{E}_{\theta\sim p}\left[\sum_{i=1}^{t} \ell_i(\theta)\right] + \frac{KL(p\|\pi)}{\eta} \right\}.
\end{aligned}$$

For some reasons, the family $\mathcal{S}(\Theta)$ may be too general or computationally intractable. Thus, we restrict it to a tractable subfamily parameterized by ϕ, known as the variational family:

$$\mathcal{F} := \{q_\phi \in \mathcal{S}(\Theta) : \phi \in \Phi\}.$$

A similar argument applies here as in the case of $\mathcal{S}(\Theta)$. We simply replace $\mathcal{S}(\Theta)$ with $\mathcal{F}$ and perform a first-order expansion. Define

$$\bar{L}_t(\phi) := \mathbb{E}_{\theta\sim q_\phi}\left[\ell_t(\theta)\right].$$

Using a first-order approximation around ϕ_t, we have

$$\bar{L}_t(\phi) \approx \bar{L}_t(\phi_t) + (\phi - \phi_t)^\top \nabla_\phi \bar{L}_t(\phi_t).$$

Ignoring constants independent of ϕ, the update becomes

$$\phi_{t+1} = \arg\min_{\phi\in\Phi} \left\{ \phi^\top \nabla_\phi \bar{L}_t(\phi_t) + \frac{KL(q_\phi\|q_{\phi_t})}{\eta} \right\}.$$

which is known as Streaming Variational Bayes [Broderick et al., 2013].

2.1 Regret Bounds for Online Bayesian Inference

Since several papers, such as Chérief-Abdellatif et al. [2019], Alquier [2024], already provide very clear proofs, I will only describe my intuition and emphasize the key ideas behind the argument.
The proof heavily relies on two ingredients!

(i) the "convex" loss $\ell_t(\theta)$,

(ii) the posterior "mean" estimator $\hat{\theta}_t = \mathbb{E}_{\theta \sim p_t}(\theta)$.

For example, in online variational inference, by the convexity of $\ell_t(\theta)$, we have

$$\sum_{t=1}^{T} \ell_t(\hat{\theta}_t) - \inf_{\phi \in \Phi} \mathbb{E}_{\theta \sim q_\phi} \left[\sum_{t=1}^{T} \ell_t(\theta) \right] \leq \mathbb{E}_{\theta \sim p_t} \left[\sum_{t=1}^{T} \ell_t(\theta) \right] - \inf_{\phi \in \Phi} \mathbb{E}_{\theta \sim q_\phi} \left[\sum_{t=1}^{T} \ell_t(\theta) \right].$$

This shows that it suffices to prove that no regret with respect to the mean loss implies no regret with respect to the cumulative loss, giving a direct connection between the two perspectives!!

3 Still studying — notes will be added later

My plans for further reading are:

1. Connections between Online Conformal Prediction and PAC-Bayes. I find the idea quite interesting.

References

Pierre Alquier. Non-exponentially weighted aggregation: regret bounds for unbounded loss functions. In *International Conference on Machine Learning*, pages 207–218. PMLR, 2021.

Pierre Alquier. User-friendly introduction to pac-bayes bounds. *Foundations and Trends in Machine Learning*, 17(2):174–303, 2024.

Tamara Broderick, Nicholas Boyd, Andre Wibisono, Ashia C Wilson, and Michael I Jordan. Streaming variational bayes. *Advances in neural information processing systems*, 26, 2013.

Badr-Eddine Chérief-Abdellatif, Pierre Alquier, and Mohammad Emtiyaz Khan. A generalization bound for online variational inference. In *Asian conference on machine learning*, pages 662–677. PMLR, 2019.