

ESTIMATION OF THE MEAN OF A RANDOM VECTOR

OLIVIER CATONI

Olivier.Catoni@ensae.fr

<http://ocatoni.perso.math.cnrs.fr/>

CREST, CNRS — UMR 9194

Université Paris–Saclay

Nips Workshop 2017

*(Almost) 50 shades of Bayesian Learning:
PAC-Bayesian trends and insights*

December 9, 2017

*Joint work with
Ilaria Giulini*

Laboratoire de Probabilités et Modèles Aléatoires
Université Paris Diderot
`giulini@math.univ-paris-diderot.fr`

A good trade-off between simplicity and performance

Question

Given X_1, \dots, X_n , n independent copies of $X \in \mathbb{R}^d$, estimate $\mathbb{E}(X)$?

Thresholding the norm

- Consider the threshold function $\psi(t) = \min\{t, 1\}$, $t \in \mathbb{R}_+$.
- Put $Y_i = \frac{\psi(\lambda \|X_i\|)}{\lambda \|X_i\|} X_i$.
- Define the estimator $\hat{m} = \frac{1}{n} \sum_{i=1}^n Y_i$.

Ideas

- Remark that $0 \leq 1 - \frac{\psi(t)}{t} \leq \inf_{p \geq 1} \frac{t^p}{p+1} \left(\frac{p}{p+1} \right)^p$.
- Use the fact that $\lambda \|Y_i\| \leq 1$.

Convergence bounds

Assumptions and choices

- Assume that $\mathbb{E}(\|X\|^2) < \infty$.
- Assume that $\sup_{\theta \in \mathbb{S}_d} \mathbb{E}(\langle \theta, X - m \rangle^2) \leq v < \infty$, where $m = \mathbb{E}(X)$ and v is known.
- Choose $\lambda = 4 \sqrt{\frac{2 \log(\delta^{-1})}{1.2vn}}$,

Bounds

Proposition: With probability at least $1 - \delta$,

$$\|m - \hat{m}\| \leq \sqrt{\frac{2.4v \log(\delta^{-1})}{n}} + \sqrt{\frac{4 \max\{\mathbb{E}(\|X - m\|^2, v)\}}{n}}$$

where

$$+ \inf_{p \geq 1} \frac{C_p}{n^{p/2}} + \inf_{p \geq 2} \frac{C'_p}{n^{p/2}},$$

$$C_p = \frac{1}{p+1} \left(\frac{4p}{p+1} \right)^p \left(\frac{2 \log(\delta^{-1})}{1.2v} \right)^{p/2} \sup_{\theta \in \mathbb{S}_d} \mathbb{E}(\|X\|^p \langle \theta, X - m \rangle_-),$$

$$C'_p = \frac{1}{p+1} \left(\frac{4p}{p+1} \right)^p \left(\frac{2 \log(\delta^{-1})}{1.2v} \right)^{p/2} \mathbb{E}(\|X\|^p) \|m\| \\ \times \left(1 + \sqrt{\frac{0.6 \log(\delta^{-1})}{vn}} \|m\| \right).$$

Sketch

- Put $\tilde{m} = \mathbb{E}(Y)$.
- Decompose the directional error

$$\langle \theta, \hat{m} - m \rangle = \langle \theta, \tilde{m} - m \rangle + \frac{1}{n} \sum_{i=1}^n \langle \theta, Y_i - \tilde{m} \rangle.$$

Bounding the first term

$\langle \theta, \tilde{m} - m \rangle = \mathbb{E}[(\alpha - 1)\langle \theta, X \rangle] = \mathbb{E}[(\alpha - 1)\langle \theta, X - m \rangle] + \mathbb{E}(\alpha - 1)\langle \theta, m \rangle$,
where $\alpha = \frac{\psi(\lambda \|X\|)}{\lambda \|X\|}$, so that

$$\begin{aligned} \langle \theta, \tilde{m} - m \rangle &\leq \inf_{p \geq 1} \frac{\lambda^p}{p+1} \left(\frac{p}{p+1} \right)^p \mathbb{E}(\|X\|^p \langle \theta, X - m \rangle_-) \\ &\quad + \inf_{p \geq 2} \frac{\lambda^p}{p+1} \left(\frac{p}{p+1} \right)^p \mathbb{E}(\|X\|^p) \langle \theta, m \rangle_-. \end{aligned}$$

PAC-Bayesian inequality

Using the Laplace transform of a normal distribution

With probability at least $1 - \delta$, for any $\theta \in \mathbb{S}_d$,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \langle \theta, Y_i - \tilde{m} \rangle - \frac{\beta + 2 \log(\delta^{-1})}{2n\mu\lambda} \\ \leq \frac{1}{\mu\lambda} \log \left(\mathbb{E} \int \exp(\mu\lambda \langle \theta', Y - \tilde{m} \rangle) d\rho_\theta(\theta') \right) \\ = \frac{1}{\mu\lambda} \log \left[\mathbb{E} \left(\exp(\mu\lambda \langle \theta, Y - \tilde{m} \rangle + \frac{\mu^2 \lambda^2}{2\beta} \|Y - \tilde{m}\|^2) \right) \right] \end{aligned}$$

So we have to bound the exponential moments of a bounded r.v. !

Bounding the exponential of a bounded argument

Like in Bennett's bound,

With probability at least $1 - \delta$, for any $\theta \in \mathbb{S}_d$,

$$\begin{aligned} \langle \theta, \widehat{m} - \widetilde{m} \rangle &\leq g_2(2\mu) \frac{\mu\lambda}{2} \mathbb{E}(\langle \theta, Y - \widetilde{m} \rangle^2) \\ &\quad + \exp(2\mu) g_1\left(\frac{2\mu^2}{\beta}\right) \frac{\mu\lambda}{2\beta} \mathbb{E}(\|Y - \widetilde{m}\|^2) + \frac{\beta + 2\log(\delta^{-1})}{2\mu\lambda n}, \end{aligned}$$

where

$$g_2(t) = 2[\exp(t) - 1 - t]/t^2 \quad \text{and} \quad g_1(t) = [\exp(t) - 1]/t$$

are increasing functions equal to one at zero.

Comparing second moments

Since $X \mapsto Y$ is a contraction,

$$\mathbb{E}(\|Y - \tilde{m}\|^2) = \frac{1}{2} \mathbb{E}(\|Y_1 - Y_2\|^2) \leq \frac{1}{2} \mathbb{E}(\|X_1 - X_2\|^2) = \mathbb{E}(\|X - m\|^2).$$

Using convexity: putting $\alpha = \frac{\psi(\lambda\|X\|)}{\lambda\|X\|}$,

$$\begin{aligned} \mathbb{E}(\langle \theta, Y - \tilde{m} \rangle^2) &\leq \mathbb{E}(\langle \theta, Y - m \rangle^2) = \mathbb{E}[(\alpha \langle \theta, X - m \rangle - (1 - \alpha) \langle \theta, m \rangle)^2] \\ &\leq \mathbb{E}[\alpha \langle \theta, X - m \rangle^2 + (1 - \alpha) \langle \theta, m \rangle^2] \\ &\leq \mathbb{E}(\langle \theta, X - m \rangle^2) + \langle \theta, m \rangle^2 \inf_{p \geq 2} \frac{\lambda^p}{p + 1} \left(\frac{p}{p + 1} \right)^p \mathbb{E}(\|X\|^p). \end{aligned}$$

Putting all together

Let us put for short

$$a = g_2(2\mu), \quad b = \exp(2\mu)g_1\left(\frac{2\mu^2}{\beta}\right).$$

Lemma

With probability at least $1 - \delta$, for any $\theta \in \mathbb{S}_d$,

$$\begin{aligned} \langle \theta, \widehat{m} - m \rangle &\leq \frac{a\mu\lambda}{2} \mathbb{E}(\langle \theta, X - m \rangle^2) + \frac{b\mu\lambda}{2\beta} \mathbb{E}(\|X - m\|^2) \\ &\quad + \frac{\beta + 2\log(\delta^{-1})}{2\mu\lambda n} \\ &\quad + \inf_{p \geq 1} \frac{\lambda^p}{p+1} \left(\frac{p}{p+1}\right)^p \mathbb{E}(\|X\|^p \langle \theta, X - m \rangle_-) \\ &\quad + \inf_{p \geq 2} \frac{\lambda^p}{p+1} \left(\frac{p}{p+1}\right)^p \mathbb{E}(\|X\|^p) \left(\langle \theta, m \rangle_- + \frac{a\mu\lambda}{2} \langle \theta, m \rangle^2 \right). \end{aligned}$$