Supplementary to "Stability and Deviation Optimal Risk Bounds with Convergence Rate $O(1/n)$"

Anonymous Author(s)
Affiliation
Address
email

1 Proof of the lower tail inequality for weakly self-bounded functions

Since the result is not presented in the literature in the form we need, we reproduce the standard argument. Let $Z = f(X_1, \ldots, X_n)$, $Z_i = f_i(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ are such that $Z \le Z_i$ almost surely, and the weakly self-bounding property holds, that is, $\sum_{i=1}^{n}(Z_i - Z)^2 \le aZ + b$. Let us first apply a modified logarithmic Sobolev inequality [1, Theorem 6.6]. We have

$$\lambda\mathbb{E}[Ze^{\lambda Z}] - \mathbb{E}[e^{\lambda Z}]\log\mathbb{E}[e^{\lambda Z}] \le \mathbb{E}\left(e^{\lambda Z}\sum_{i=1}^{n}\phi(-\lambda(Z - Z_i))\right),$$

where $\phi(x) = e^x - x - 1$. Since for $x \ge 0$, $\phi(-x) \le x^2/2$ and $Z - Z_i \le 0$ for all $i = 1, \ldots, n$, we have for any $\lambda \le 0$,

$$\lambda\mathbb{E}[Ze^{\lambda Z}] - \mathbb{E}[e^{\lambda Z}]\log\mathbb{E}[e^{\lambda Z}] \le \frac{\lambda^2}{2}\mathbb{E}\left(e^{\lambda Z}\sum_{i=1}^{n}(Z - Z_i)^2\right) \le \frac{\lambda^2}{2}(a\mathbb{E}[Ze^{\lambda Z}] + b\mathbb{E}[e^{\lambda Z}]).$$

Define $G(\lambda) = \log\mathbb{E}[e^{\lambda Z}]$, so that $G'(\lambda) = \mathbb{E}[Ze^{\lambda Z}]/\mathbb{E}[e^{\lambda Z}]$. Dividing both sides of the last display by $\mathbb{E}[e^{\lambda Z}]$, we obtain

$$\lambda G'(\lambda) - G(\lambda) \le \frac{\lambda^2}{2}(aG'(\lambda) + b), \qquad \lambda \le 0.$$

For $\lambda < 0$, we have

$$\left(\left(\frac{1}{\lambda} - \frac{a}{2}\right)G(\lambda)\right)' = \left(\frac{1}{\lambda} - \frac{a}{2}\right)G'(\lambda) - \frac{G(\lambda)}{\lambda^2} \le \frac{b}{2}.$$

Finally, we integrate this inequality. Observe that $G(0) = 0$ and we also have $G'(0) = \mathbb{E}Z$. Hence, as $\lambda \to -0$, by Taylor's theorem $(1/\lambda - a/2)G(\lambda) = (1/\lambda - a/2)(\lambda\mathbb{E}Z + o(\lambda)) = \mathbb{E}Z + o(1)$. Integrating the last display over the interval $[\lambda, 0]$, we obtain

$$\mathbb{E}Z - \left(\frac{1}{\lambda} - \frac{a}{2}\right)G(\lambda) \le (-\lambda)\frac{b}{2}.$$

After some rearrangements this leads to the following inequality (recall that $\lambda < 0$),

$$\log\mathbb{E}[e^{\lambda(Z-\mathbb{E}Z)}] = G(\lambda) - \lambda\mathbb{E}Z \le \frac{\lambda^2(a\mathbb{E}Z + b)}{2(1 - \lambda a/2)} \le \frac{\lambda^2}{2}(a\mathbb{E}Z + b).$$

It remains to use the Markov inequality (recall again that $\lambda < 0$) to show that

$$\mathbb{P}(Z < \mathbb{E}Z - t) = \mathbb{P}\left(\lambda(Z - \mathbb{E}Z) > -\lambda t\right) = \mathbb{P}\left(e^{\lambda(Z-\mathbb{E}Z+t)} > 1\right) \le \exp\left(t\lambda + \frac{\lambda^2}{2}(a\mathbb{E}Z + b)\right).$$

The latter exponent is minimized by $\lambda = -t/(a\mathbb{E}Z + b) < 0$ implying the statement. □

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

References

[1] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.