Lecture 5: Pseudo-maximization and Self-normalized Processes II

Lecturer: ***Víctor H. de la Peña*** *Scribes: Heyuan Yao*

5.1 Background on Pseudo-maximization (Method of Mixtures)

Let us carefully recall the canonical assumption in the last class. We assume that, for a pair of random variables A, B with $B > 0$,

$$\mathbb{E}\exp(\lambda A - \lambda^2 B^2/2) \leq 1, \tag{5.1}$$

holds

- for all real λ;
- for all $\lambda \geq 0$;
- for all $0 \leq \lambda < \lambda_0$, where $0 < \lambda < \infty$.

If the global minimizer $\hat{\lambda} := \frac{A}{B^2}$ lies in the regime we are interested, and is of course, deterministic, then by Chebyshev inequality, we have

$$\mathbb{P}\left(\frac{|A|}{|B|} > x\right) = \mathbb{P}\left(\frac{A^2}{2B^2} > \frac{x^2}{2}\right) \leq e^{-\frac{x^2}{2}}\mathbb{E}e^{\frac{A^2}{2B^2}} \leq e^{-\frac{x^2}{2}}.$$

a beautiful Gaussian bound.

Unfortunately, since $\frac{A}{B^2}$ is nonrandom, we need an alternative method for dealing with this maximization. And today, I introduce one celebrated approach: pseudo-maximization (method of mixtures). An informal framework of this method can be stated as follows:

(i)For $\lambda \in \Lambda$ a measurable set, we construct a finite (probability) measure[1] of λ, with distribution function F independent of A and B.

[1] If you construct a finite measure, you can always normalize it to obtain a probability law

(ii) Now we have that

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda A - \frac{-\lambda^2 B^2}{2}\right)\right] &= F(+\infty)^{-1}\int_{\mathbb{R}} \mathbb{E}\left[\exp\left(\lambda A - \frac{-\lambda^2 B^2}{2}\right)\right] dF \\ &= F(+\infty)^{-1}\mathbb{E}\left[\int_{\mathbb{R}} \exp\left(\lambda A - \frac{-\lambda^2 B^2}{2}\right) dF\right]\end{aligned}$$

by FUBINI.

To be effective for all possible pairs, the F chosen would need to be as uniform as possible so as to include the maximum value of $\exp(\lambda A - \lambda^2 B^2/2)$. And for (5.1) that holds for all real λ or $\lambda \geq 0$, since all finite measures vanish at infinity, we need to construct a measure decaying as slowly as we can manage.

5.2 Gaussian Bounds for $\frac{A}{\sqrt{B^2+\mathbb{E}^2 B}}$

One application of pseudo-maximization is to construct a Gaussian bound for $\frac{A}{\sqrt{B^2+\mathbb{E}^2 B}}$, with the theorem stated below.

Theorem 5.1 *Let A, B with $B > 0$ be random variables satisfying the canonical assumption (5.1) for all $\lambda \in \mathbb{R}$. Then*

$$\mathbb{P}\left(\frac{|A|}{\sqrt{B^2+\mathbb{E}^2 B}} \geq x\right) \leq \sqrt{2}\exp\left(-\frac{x^2}{4}\right). \tag{5.2}$$

Proof: Considering that $\lambda \in R$, we here let $\lambda \sim \mathcal{N}(0, \frac{1}{y^2})$. Multiplying both sides of (5.1) by $(2\pi^{-1/2})y\exp(-\lambda^2 y^2/2)$ and integrating over λ, we obtain that

$$\begin{aligned}1 &\geq \int_{\mathbb{R}} \mathbb{E}\frac{y}{\sqrt{2\pi}}\exp\left(\lambda A - \frac{\lambda^2}{2}B^2\right)\exp\left(-\frac{\lambda^2 y^2}{2}\right)d\lambda \\ &= \mathbb{E}\left[\frac{y}{\sqrt{B^2+y^2}}\exp\left(\frac{A^2}{2(B^2+y^2)}\right)\int_{\mathbb{R}}\frac{\sqrt{B^2+y^2}}{\sqrt{2\pi}}\exp\{-\frac{B^2+y^2}{2}\left(\lambda^2 - 2\frac{A}{B^2+y^2} + \frac{A^2}{(B^2+y^2)^2}\right)\}d\lambda\right] \\ &\qquad\text{[Here we fix } A, B \text{ inside } \mathbb{E}(\cdot) \text{ and change the measure to } \lambda \sim \mathcal{N}(\frac{A}{B^2+y^2}, \frac{1}{B^2+y^2})] \\ &= \mathbb{E}\left[\frac{y}{\sqrt{B^2+y^2}}\exp\left(\frac{A^2}{2(B^2+y^2)}\right)\right]\end{aligned}$$

By CAUCHY-SCHWARZ inequality and the inequality above, we have that

$$\begin{aligned}\mathbb{E}\exp\left(\frac{A^2}{4(B^2+y^2)}\right) &\leq \sqrt{\mathbb{E}\frac{y\exp\left(\frac{A^2}{2(B^2+y^2)}\right)}{\sqrt{B^2+y^2}}\ \mathbb{E}\sqrt{\frac{B^2+y^2}{y^2}}} \\ &\leq \sqrt{\mathbb{E}\sqrt{\frac{B^2+y^2}{y^2}}}.\end{aligned}$$

Recall that B is nonnegative a.s., and you have that $\mathbb{E}\sqrt{\frac{B^2+y^2}{y^2}} \leq \mathbb{E}(\frac{B}{y}+1)$, which means that we can set $y = \mathbb{E}B$ so that

$$\mathbb{E}\exp\left(\frac{A^2}{4(B^2+y^2)}\right) \leq \sqrt{2}.$$

Finally, combining MARKOV's inequality, we have that

$$\mathbb{P}\left(\frac{|A|}{\sqrt{B^2+\mathbb{E}^2B}} \geq x\right) = \mathbb{P}\left(\frac{A^2}{4(B^2+\mathbb{E}^2B)} \geq \frac{x^2}{4}\right) \leq \sqrt{2}\exp\left(-\frac{x^2}{4}\right).$$

■

5.3 Boundary-crossing Problem

While the "young" inequality mentioned above was formally established in 2004, the concept of pseudo-maximization has been actively employed for over half a century. In the subsequent discussion, we delve into the analysis of specific boundary-crossing probabilities utilizing the method of mixtures. This method was first introduced by ROBBINS and SIEGMUND in 1970 and later refined by LAI in 1976, in Columbia.

We make the following refinement of the canonical assumption:

$$\{\exp(\lambda A_t - \lambda^2)B)t^2/2, t \geq 0\} \text{ is a supermartingale with mean } \leq 1, \tag{5.3}$$

for $0 \leq \lambda < \lambda_0$, with $A_0 = 0$.

Instead of constructing a probability measure, we let F be a finite positive measure on $(0, \lambda_0)$ and assume that $F(0, \lambda_0) > 0$. Without assuming the exact density (or mass) of F, let us first consider the following function $\Psi : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}^+$, such that

$$\Psi(u, v) = \int_{(0,\lambda_0)} e^{\lambda u - \frac{\lambda^2 v}{2}} dF(\lambda).$$

When $v > 0$ is fixed, $\Psi(\cdot, v)$, strictly increasing, maps onto $\mathbb{R}^+$, which induce that when we further fix $c > 0$, the equation

$$\Psi(u, v) = c$$

has a unique solution $u = \beta_F(v, c)$.

Moreover, you can verify that $\beta_F(v, c)$ is a concave function of v and

$$\lim_{v\to\infty} \frac{\beta_F(v, c)}{v} = \frac{b}{2},$$

where

$$b := \sup\{y > 0 : \int_0^y F(d\lambda) = 0\},$$

with sup over the empty set equal to zero.

The Robbins-Siegmund (R-S) boundaries $\beta_F(v, c)$ can be used to analyse the boundary crossing probability

$$\mathbb{P}(A_t \geq g(B_t) \text{ for some } t \geq 0)$$

when $g(B_t) = \beta_F(B_t^2, c)$ for some F and $C > 0$. This probability equals

$$\begin{aligned}
\mathbb{P}(A_t \geq g(B_t) \text{ for some } t \geq 0) &= \mathbb{P}(\Psi(A_t, B_t^2) \geq c \text{ for some } t \geq 0) \\
&= \mathbb{P}\left(\sup_t \int_{(0,\lambda_0)} e^{\lambda A_t - \frac{\lambda^2 B_t^2}{2}} dF(\lambda) \geq c\right) \\
&\leq \frac{1}{c}\mathbb{E}\left[\int_{(0,\lambda_0)} e^{\lambda A_t - \frac{\lambda^2 B_t^2}{2}} dF(\lambda)\right] && \text{[by Doob's Inequality]} \\
&= \frac{1}{c}\int_{(0,\lambda_0)} \mathbb{E}\left[e^{\lambda A_t - \frac{\lambda^2 B_t^2}{2}}\right] dF(\lambda) && \text{[using pseudo-maximization]} \\
&\leq \frac{1}{c}\int_{(0,\lambda_0)} dF(\lambda) && \text{[by canonical assumption]} \\
&= \frac{F((0,\lambda_0))}{c}.
\end{aligned}$$

Let us denote by $\log_2(\cdot) = \log\log(\cdot)$ and $\log_3(\cdot) = \log\log\log(\cdot)$. For $\delta > 0$, $\lambda \in (0, e^{-e})$ we assume $F \ll \lambda$ (the Lebesgue measure not the variable), and

$$dF(\lambda) = \frac{1}{\lambda(\log\frac{1}{\lambda}(\log_2\frac{1}{\lambda})^{1+\delta})}$$

As shown in example 4 of Robbins and Siegmund, for fixed c,

$$\beta_F(v,c) = \sqrt{2v[\log_2 v + (\frac{3}{2}+\delta)\log_3 v + \log(\frac{c}{2\sqrt{\pi}}) + o(1)]},$$

as $v \to \infty$. With this choice of F, the probability $\mathbb{P}(A_t \geq g(B_t)$ for some $t \geq 0)$ is bounded by $F(0, e^{-e})/c$ for all $c > 0$. Given $\epsilon > 0$, take c large enough so that $F(0, e^{-e})/c < \epsilon$. Since ϵ can be arbitrarily small and since for fixed c, $\beta_F(v,c) \sim \sqrt{2v \log\log v}$ as $v \to \infty$,

$$\limsup \frac{A_t}{2B_t^2 \log\log B_t^2} \leq 1,$$

on the set $\{\lim_{t\to\infty} B_t = \infty\}$.

Reference

Chen, X. (1999). The law of the iterated logarithm for functionals of Harris recurrent Markov chains: Self normalization. *J. Theoret. Probab.* 12 421–445.

de la Peña, V. H., Klass, M. J. and Lai, T. L. (2007) Pseudo-maximization and self-normalized processes. *Probability Survey* 4 172-192.

Erdós, P. (1942). On the law of the iterated logarithm. *Ann. Math.* 43 419–436.

Giné, E. and Mason, D. M. (1998). On the LIL for self-normalized sums of i.i.d. random variables. *J. Theoret. Probab.* 11 (2) 351–370

Griffin, P. and Kuelbs, J. (1989). Self-normalized laws of the iterated logarithm. *Ann. Probab.* 17 1571–1601.

Griffin, P. and Kuelbs, J. (1991). Some extensions of the LIL via self-normalization. *Ann. Probab.* 19 380–395.

Lai, T. L. (1976). Boundary crossing probabilities for sample sums and confi- dence sequences. *Ann. Probab.* 4 299–312.

Robbins, H. (1970). Statistical methods related to the law of the iterated loga- rithm. *Ann. Math. Statist.* 41 1379–1409.

Robbins, H. and Siegmund, D. (1970). Boundary crossing probabilities for the Wiener process and sample sums. *Ann. Math. Statist.* 41 1410–1429.

Stout, W. F. (1973). Maximal inequalities and the law of the iterated logarithm. *Ann. Probab.* 1 322–328.