Topics in Probability Theory **Fall 2023 (Georgia Institute of Technology)**

Lecture 4: Pseudo-maximization and Self-normalized Processes I

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

4.1 Background

Self-normalized statistic, literally, takes the form $\frac{A_n}{B_n}$ (resp., $\frac{A_t}{B_t}$ for continuous cases), where both A_n and B_n are functions of your observations $X_1, ..., X_n$ (resp., A_t, B_t the function of $(X_s)_{0\le s\le t}$). One of the bonuses of self-normalization is that you can obtain a statistic, which, with sample size n increasing, maintains a bounded tail probability. Some self-normalized statistics you may encounter are, for example, sample Gini coefficient $\tilde{G} = \frac{1}{2}\frac{\frac{1}{n^2}\sum 1\le i,j\le n|X_i - X_j|}{\bar{X}_n}$ of positively supported random variables, which almost surely takes value between 0 and 1, and sample squared coefficient of variation $\widehat{c_V}^2 := \frac{s_n^2}{\bar{X}_n^2}$, which cancels the scaling.

A celebrated self-normalized statistic you learned in your undergraduate time is Student's t-statistic, by W. Gosset. Recall that you have $\{X_i\}$ i.i.d normal $\mathcal{N}(\mu, \sigma^2)$, and the sample mean and sample variance, respectively defined by

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n} \qquad s_n^2 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n-1},$$

then the t-statistic is constructed: $T_n = \frac{\bar{X}_n - \mu}{s_n/\sqrt{n}} \sim t_{n-1}$, the t-distribution with degree of freedom n-1. In addition, if you denote by $Y_i = X_i - \mu,\ A_n = \sum_{i=1}^n Y_i,\ B_n^2 = \sum_{i=1}^n Y_i^2$, then you will find that,

$$T_n = \frac{\frac{A_n}{B_n}}{\sqrt{(n - (A_n/B_n)^2)/(n-1)}},$$

a function of the self-normalized statistic $\frac{A_n}{B_n}$.

One remarkable progression brought by t-statistic is that "Student", i.e., W. Gosset, did not assume the variance was finite. This was a big breakthrough and after that, scientists did not need to assume the variation of their data to test what their mean was. Several years later, B. Efron, the founder of the bootstrap, developed a self-normalized inequality for independent symmetric variables. And after Efron, there were some more developments in self-normalization for independent variables, and, for martingales or dependent variables. An example of self-normalized processes in dependent variables arose in the context of Maximum Likelihood Estimators (MLEs) for the parameter in auto-regressive (AR) processes.

4.2 Initiation of Canonical Assumption, from Auto-Regressive Models

Let us consider the following Auto-Regressive Gaussian process $(Y_i)_0^\infty$, such that

$$Y_i = \alpha Y_{i-1} + \epsilon_i, \qquad Y_0 = 0, \tag{4.1}$$

where $\alpha \neq 0$ is a fixed, unknown parameter and ϵ_i are independent standard normal random variables $\mathcal{N}(0,1)$.

To obtain the MLE of α, we establish our log-likelihood function

$$\begin{aligned} l(\alpha; Y_1, ..., Y_n) &= \log_\alpha f(Y_1, ..., Y_n) \\ &= \sum_{j=1}^{n} (Y_j - \alpha Y_{j-1})^2/2 - n\log(\sqrt{2\pi}). \end{aligned}$$

Taking the derivative w.r.t. α, equating to zero and solving for α, we obtain the MLE for α,

$$\hat{\alpha} = \frac{\sum_{j=1}^n Y_{j-1}Y_j}{\sum_{j=1}^n Y_{j-1}^2} = \frac{\sum_{j=1}^n Y_{j-1}(\alpha Y_{j-1} + \epsilon_j)}{\sum_{j=1}^n Y_{j-1}^2} = \alpha + \frac{\sum_{j=1}^n Y_{j-1}\epsilon_j}{\sum_{j=1}^n Y_{j-1}^2}. \tag{4.2}$$

You now find, without much surprise, that this MLE is a self-normalization, and so is $\hat{\alpha} - \alpha$:

$$\hat{\alpha} - \alpha = \frac{\sum_{j=1}^n Y_{j-1}\epsilon_j}{\sum_{j=1}^n Y_{j-1}^2}. \tag{4.3}$$

We now construct the filtration $\mathcal{F} := \sigma(Y_1, ..., Y_n; \epsilon_1, ...\epsilon_n)$, and therefore the numerator $\sum_{j=1}^n Y_{j-1}\epsilon_j =: A_n$ is a martingale w.r.t. $\mathcal{F}$. And the denominator

$$\sum_{j=1}^n Y_{j-1}^2 = \sum_{j=1}^n \mathbb{E}[Y_{j-1}^2 \epsilon_j^2 | \mathcal{F}_{j-1}] =: B_n^2 \tag{4.4}$$

is the conditional variance of A_n. Thus $\hat{\alpha} - \alpha = \frac{A_n}{B_n^2}$ is a process self-normalized by the conditional variance. Since ϵ_i's are $\mathcal{N}(0,1)$, then we have that for any $\lambda \in \mathbb{R}$,

$$M_n := exp\left(\lambda A_n - \frac{\lambda^2 B_n^2}{2}\right) \tag{4.5}$$

is an exponential martingale w.r.t. $\mathcal{F}_n$ (I leave this claim as an exercise left for you to verify). With optimal stopping theorem, you have that $\mathbb{E}(M_n) = \mathbb{E}(M_1) = 1$, for all $n \geq 1$, which leads that

$$exp\left(\lambda A_n - \frac{\lambda^2 B_n^2}{2}\right) \leq 1 \tag{4.6}$$

the canonical assumption in the next section.

4.3 Canonical Assumption and Exponential Bounds for Strong Law

So far, we have found an inequality (4.6). And such inequality, for a pair of random variables A, B with $B > 0$, taking the general form

$$\mathbb{E}\exp(\lambda A - \lambda^2 B^2/2) \leq 1, \tag{4.7}$$

frequently appears in probability theory and stochastic analysis. I list some self-normalized statistics $\frac{A}{B}$ that satisfies (4.7) for some λ in Appendix. There are three regimes of interest: (4.7) holds

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

In this lecture, we only focus on the first case, i.e., (4.7) holds for all real λ. One theorem that bounds the tail probability of A/B^2, with a constraint for B^2 is given below.

Theorem 4.1 *Under the canonical assumption for all real λ,*

$$\mathbb{P}\left(\frac{A}{B^2} > x, \frac{1}{B^2} \leq y\right) \leq \exp\left(-\frac{x^2}{2y}\right) \tag{4.8}$$

for all $x, y > 0$.

Proof: The key here is to **"keep" the indicator** when using MARKOV Inequality. In fact, for all measurable set $\boldsymbol{S}$,

$$\begin{aligned}
\mathbb{P}(A/B^2 > x, \boldsymbol{S}) &= \mathbb{P}(\exp(A) > \exp(xB^2), \boldsymbol{S}) \\
&\leq \inf_{\lambda>0} \mathbb{E}\left[\exp(\frac{\lambda}{2}A - \frac{\lambda}{2}xcB^2)\mathbb{I}_{\{A/B^2>x,\boldsymbol{S}\}}\right] \\
&= \inf_{\lambda>0} \mathbb{E}\left[\exp(\frac{\lambda}{2}A - \frac{\lambda^2}{4}B^2 - (\frac{\lambda}{2}x - \frac{\lambda^2}{4})B^2)\mathbb{I}_{\{A/B^2>x,\boldsymbol{S}\}}\right] \\
&\leq \sqrt{\mathbb{E}\left[\exp(\lambda A - \frac{\lambda^2}{2}B^2\right]} \sqrt{\mathbb{E}\left[-(\lambda x - \frac{\lambda^2}{2})B^2)\mathbb{I}_{\{A/B^2>x,\boldsymbol{S}\}}\right]}
\end{aligned}$$

by the Cauchy-Schwarz inequality. The first term in the last inequality is bounded by 1, by the canonical assumption. The value minimizing the second term is $\lambda = x$, and therefore

$$\mathbb{P}(A/B^2 > x, \boldsymbol{S}) \leq \sqrt{\mathbb{E}\left[\frac{-x^2B^2}{2}\right]}\mathbb{I}_{\{A/B^2>x,\boldsymbol{S}\}}.$$

Let us set $\boldsymbol{S} = \{\frac{1}{B^2} < y\}$ and Theorem 4.9 follows as we claimed. ■

Now let us go back to Section 4.2, and apply this bound with $y = \frac{1}{z}$ to (4.3), and yield that

$$\mathbb{P}\left(|\hat{\alpha} - \alpha| > x, \sum_{j=1}^{n} Y_{j-1}^2 \geq z\right) \leq 2\exp(\frac{-x^2z}{2}).$$

A variant of Theorem 4.9 generalizes a result of Lipster and Spokoiny in 1999.

Theorem 4.2 *Under the canonical assumption for all real λ,*

$$\mathbb{P}\left(\frac{|A|}{B} > x, b \leq B \leq bs\right) \leq 4\sqrt{e}x(1+\log s)\exp\left(-\frac{x^2}{2}\right) \tag{4.9}$$

for all $b > 0$, $s \geq 1$, and $x \geq 1$.

With a similar spirit, there are several exponential inequalities for martingales and supermartingales. Due to the time limit, I will not present the tons of progressions in class. But you can easily find them in the survey paper by my advisor M. J. Klass, my friend T.L. Lai, and me.

Appendix

Lemma 4.3 *Let W_t be a standard Brownian motion. Assume that T is a stopping time such that $T < \infty$ a.s.. Then for all $\lambda \in \mathbb{R}$,*

$$\mathbb{E}\exp(\lambda W_T - \lambda^2 T/2) \le 1.$$

Lemma 4.4 *Let M_t be a continuous, square-integrable martingale, with $M_0 = 0$. Then$\exp\{\lambda M_t - \lambda^2 \langle M \rangle_t / 2\}$ is a supermartingale for all $\lambda \in \mathbb{R}$, and therefore*

$$\mathbb{E}\exp(\lambda M_t - \lambda^2 \langle M \rangle_t / 2) \le 1.$$

The inequality is also valid if M_t is only assumed to be a continuous local martingale (by application of FATOU*'s lemma).*

Lemma 4.5 *Let $\{M_t : t \ge 0\}$ be a locally squared-integrable martingale, with $M_0 = 0$. Let $\{V_t\}$ be an increasing process, which is adapted, purely discontinuous and locally integrable; let $V^{(p)}$ be its dual predictable projection. Set $X_t = M_t + V_t$, $C_t = \sum_{s \le t}((\Delta X_s)^+)^2$, $D_t = \{\sum_{s \le t}((\Delta X_s)^-)^2\}_t^{(p)}$, $H_t = \langle M \rangle_t^c + C_t + D_t$. Then $\exp(X_t - V_t^{(p)} - 1/2 H_t)$ is a supermartingale, and hence*

$$\mathbb{E}\exp\left[\lambda(X_t - V_t^{(p)}) - \lambda^2/2 H_t\right] \le 1.$$

Lemma 4.6 *Let $\{d_i\}$ be a sequence of variables adapted to an increasing sequence of $\sigma-$fields $\{\mathcal{F}_i\}$. Assume that the d_i's are conditionally symmetric (i.e., $\mathcal{L}(d_i|\mathcal{F}_{i-1}) = \mathcal{L}(-d_i|\mathcal{F}_{i-1})$). Then $\exp(\lambda \sum_{i=1}^n d_i - \lambda^2 \sum_{i=1}^n d_i^2/2)$, $n \ge 1$, is a supermartingale with mean ≤ 1, for all $\lambda \in \mathbb{R}$.*

Lemma 4.7 *Let $\{d_n\}$ be a sequence of random variables adapted to an increasing sequence of $\sigma-$fields $\{\mathcal{F}_n\}$ such that $\mathbb{E}(d_n|\mathcal{F}_{n-1}) \le 0$ and $\dot{}n \le M$ a.s. for all n and some random positive constant M. Let $0 < \lambda_0 \le M^{-1}$, $A_n = \sum_{i=1}^n d_i$, $B_n^2 = (1 + \frac{1}{2}\lambda_0 M)\sum_{i=1}^n \mathbb{E}(d_i^2|\mathcal{F}_{i-1})$, $A_0 = B_0 = 0$. then $\{\exp(\lambda A_n - \frac{1}{2}B_n^2), \mathcal{F}_n, n \ge 0\}$ is a supermartingale for every $0 \le \lambda \le \lambda_0$.*

Lemma 4.8 *Let $\{d_n\}$ be a sequence of variables adapted to an increasing sequence of $\sigma-$fields $\{\mathcal{F}_n\}$ such that $\mathbb{E}(d_n|\mathcal{F}_{n-1}) = 0$ and $\sigma_n^2 = \mathbb{E}(d_n^2|\mathcal{F}_{n-1}) \le \infty$. Assume that there exists a positive constant M such that $\mathbb{E}(|d_n|^k|\mathcal{F}_{n-1}) \le (k!/2)\sigma_n^2 M^{k-2}$ a.s. or $\mathbb{P}(|d_n| \le M|\mathcal{F}_{n-1}) = 1$ a.s. for all $n \ge 1, k > 2$. Let $A_n = \sum_{i=1}^n d_i$, $V_n^2 = \sum_{i=1}^n \mathbb{E}(d_i^2|\mathcal{F}_{i-1})$, $A_0 = V_0 = 0$. Then $\{\exp(\lambda A_n - \frac{1}{2(1-M\lambda)}\lambda^2 V_n^2), \mathcal{F}_n, n \ge 0\}$ is a supermartingale for every $0 \le \lambda < \frac{1}{M}$.*

Lemma 4.9 *Let $\{d_n\}$ be a sequence of variables adapted to an increasing sequence of $\sigma-$fields $\{\mathcal{F}_n\}$ such that $\mathbb{E}(d_n|\mathcal{F}_{n-1}) \leq 0$ and $d_n \geq -M$ a.s. for all n and some nonrandom positive constant M. Let $A_n = \sum_{i=1}^n d_i$, $B_n^2 = 2C_\gamma \sum_{i=1}^n d_i^2$, $A_0 = B_0 = 0$, where $C_\gamma = -(\gamma_{\log}(1-\gamma))/\gamma^2$. Then $\{\exp(\lambda A_n - \frac{1}{2}B_n^2), \mathcal{F}_n, n \geq 0\}$ is a supermartingale for every $0 \leq \lambda \leq \gamma M^{-1}$.*

Lemma 4.10 *Let $\{\mathcal{F}_n\}$ be an increasing sequence of $\sigma-$fields and $\{y_n\}$ be $\mathcal{F}_n$-measurable random variables. Let $0 \leq \gamma_n < 1$ and $0 < \lambda_n \leq 1/C_{\gamma_n}$ be $\mathcal{F}_{n-1}$-measurable random variables, with C_γ given in Lemma 4.9. let $\mu_n - \mathbb{E}(y_n \mathcal{I}_{\{-\gamma_n \leq y_n \leq \lambda_n} | \mathcal{F}_{n-1}\})$. Then $\exp(\sum_{i=1}^n y_i - \mu_i - \lambda_i^{-1} y_i^2)$ is a supermartingale whose expectation is $\leq$ 1.*

Reference

de la Peña, V. H. (1999). A general class of exponential inequalities for martingales and ratios. *Ann. Probab.* 27 537–564.

de la Peña and Giné (1999). Decoupling: From Dependence to Independence. *Springer*, New York.

de la Peña, V. H., Klass, M. J. and Lai, T. L. (2000). Moment bounds for self-normalized martingales. In *High Dimensional Probability II* (E. Giné, D. M. Mason and J. A. Wellner, eds.) 1–11. Birkhauser, Boston.

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

Efron, B. and Tibshirani, R. (1993). An Introduction to the Bootstrap. *Chapman & Hall*, New York.

Giné, E. Göetze, F. and Mason, D. (1997). When is the Student t-statistic asymptotically standard normal? *Ann. Probab.* 25 1514–1531.

Karatzas, I. and Shreve, S. E. (1991). Brownian Motion and Stochastic Calculus, 2nd edition. *Springer*, New York.

Lipster, R. and Spokoiny, V. (2000). Deviation probability bound for martingales with applications to statistical estimation. *Statistics and Probability Letters* 46, 347–357.