DSO Screening Exam: 2016 In-Class Exam

Gregory Faletto

Exercise 1 (Probability/Mathematical Statistics). (a)

$$n^{-1} \log W_n = \frac{\log\left[(qr + (1-q)V_n)W_{n-1}\right]}{n} = \frac{\log\left[qr + (1-q)V_n\right]}{n} + \frac{\log W_{n-1}}{n}$$

Note that for q and r fixed, $\log[qr+(1-q)V_n]$ are i.i.d. random variables with mean $\mathbb{E}\left(\log[qr + (1-q)V_1]\right)$. Since $\mathbb{E}\left|\log[qr + (1-q)V_1]\right| < \infty$, by the Strong Law of Large Numbers

$$\frac{\log\left[qr + (1-q)V_n\right]}{n} \xrightarrow{a.s.} \mathbb{E}\log[qr + (1-q)V_1] = w(q).$$

So the result follows if $n^{-1} \log W_{n-1} \xrightarrow{a.s.} 0$. Note that

$$\frac{\log W_{n-1}}{n} = \frac{\log\left[(qr + (1-q)V_{n-1})W_{n-2}\right]}{n}$$

(b)

$$w(q) = \mathbb{E}\log[qr + (1-q)V_1] = \mathbb{E}\log[q(r - V_1) + V_1]$$

Let $t \in (0,1)$. Let $q_1, q_2 \in (0,1]$, and suppose without loss of generality $q_1 \leq q_2$. We wish to show that

$$w(tq_1 + (1-t)q_2) \geq tw(q_1) + (1-t)w(q_2) \tag{1}$$

$$\iff \mathbb{E}\log[(tq_1+(1-t)q_2)r+[1-(tq_1+(1-t)q_2)]V_1] \geq t\mathbb{E}\log[q_1r+(1-q_1)V_1]+(1-t)\mathbb{E}\log[q_2r+(1-q_2)V_1]$$

$$\iff \mathbb{E}\log[tq_1r+(1-t)q_2r+V_1-tq_1V_1-(1-t)q_2)V_1] \geq t\mathbb{E}\log[q_1r+(1-q_1)V_1]+(1-t)\mathbb{E}\log[q_2r+(1-q_2)V_1]$$

(c)

Exercise 2 (Mathematical statistics, Bayesian). (a)

(b)

Exercise 3 (Convergence; Wen says we don't need to worry about. From Trambak's exam; he took Analysis

(b)

Exercise 4 (Convex Optimization). (a)

(b)

Exercise 5 (High-Dimensional Statistics). Double-check solutions, comments from Jinchi grading homework

(a) Let

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_n^T \end{bmatrix} \in \mathbb{R}^{n \times p}$$

be the design matrix. Given that the mean is known to be 0, the covariance matrix is defined as

$$\boldsymbol{\Sigma} = \mathbb{E}\big(\boldsymbol{X}^T\boldsymbol{X}\big)$$

A natural unbiased estimator for $\boldsymbol{\Sigma}$ is the sample covariance

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\big(\boldsymbol{X}^T\boldsymbol{X}\big).$$

(b) We have different issues in each of these regimes.

(1) $p \leq n$ **and** p **is roughly of the same order as** n**:** there can be significant sampling error in estimating $\widehat{\boldsymbol{\Sigma}}$ in this regime. Fan et al. [2008] showed that under Frobenius norm this estimator has a very slow convergence rate even if $p < n$. Further, the expected value of its inverse is

$$\mathbb{E}\big(\widehat{\boldsymbol{\Sigma}}^{-1}\big) = \frac{n}{n-p-2}\boldsymbol{\Sigma}^{-1}$$

[Bai and Shi, 2011], so this bias can be quite large if $p \approx n$, even if $p < n$. (A better method for estimating $\boldsymbol{\Sigma}^{-1}$ directly is presented by Fan et al. [2008].)

(2) $p > n$ **or even** $p \gg n$**:** in that case $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\big(\boldsymbol{X}^T\boldsymbol{X}\big)$ will be rank-deficient and singular, even though the true covariance matrix will be nonsingular (and positive definite), so clearly $\widehat{\boldsymbol{\Sigma}}$ will not be an ideal estimate.

(c) Geman [1980] showed that in the case of $\boldsymbol{\Sigma} = I_p$,

$$\lambda_{\max}\big(\widehat{\boldsymbol{\Sigma}}\big) \xrightarrow{a.s.} (1+\gamma^{-1/2})^2 \text{ as } n/p \to \gamma \geq 1.$$

Further, numerical studies that that $\lambda_{\max}\big(\widehat{\boldsymbol{\Sigma}}\big)$ for $n = 100$ typically ranges between 1.2 - 1.5 for $p = 5$, between 2.6 and 3 for $p = 50$, and between 10 and 10.5 for $p = 500$. Of course, the correct maximum eigenvalue is 1 (since all eigenvalues of I_p are 1), so we see that covariance matrix estimation gets increasing unstable and inaccurate as $p \gg n$.

Regarding the limiting distribution of the largest eigenvalue $\lambda_{\max}\big(\widehat{\boldsymbol{\Sigma}}\big)$, Johnstone [2001] showed that

$$\frac{n\lambda_{\max}\big(\widehat{\boldsymbol{\Sigma}}\big) - \mu_{np}}{\sigma_{np}} \xrightarrow{D} \text{Tracy-Widom law of order 1 as } n/p \to \gamma \geq 1$$

where

$$\mu_{np} = (\sqrt{n-1} + \sqrt{p})^2, \qquad \sigma_{np} = (\sqrt{n-1} + \sqrt{p})(1/\sqrt{n-1} + 1/\sqrt{p})^{1/3}.$$

References

J. Bai and S. Shi. Estimating High Dimensional Covariance Matrices and its Applications. *Annals of Economics and Finance*, 12(2):199–215, 2011. URL `http://aeconf.com/articles/nov2011/aef120201.pdf`.

J. Fan, Y. Fan, and J. Lv. High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 147:186–197, 2008. doi: 10.1016/j.jeconom.2008.09.017. URL `www.elsevier.com/locate/jeconom`.

S. Geman. A Limit Theorem for the Norm of Random Matrices. *The Annals of Probability*, 8(2):252–261, 1980. URL `https://www-jstor-org.libproxy2.usc.edu/stable/pdf/2243269.pdf?refreqid=excelsior%3Aab0d3919c2193ca3a`

I. M. Johnstone. On The Distribution of the Largest Eigenvalue in Principal Components Analysis. *The Annals of Statistics*, 29(2):295–327, 2001. URL `https://projecteuclid.org/download/pdf_1/euclid.aos/1009210544`.