DSO Screening Exam: 2016 In-Class Exam

Gregory Faletto

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

$$W_1 = (qr + (1-q)V_1)W_0 = qr + (1-q)V_1$$

$$W_2 = (qr + (1-q)V_2)W_1 = (qr + (1-q)V_2)(qr + (1-q)V_1)$$

$$W_3 = (qr + (1-q)V_3)W_2 = (qr + (1-q)V_3)(qr + (1-q)V_2)(qr + (1-q)V_1) = \prod_{i=1}^{3}(qr + (1-q)V_i)$$

$$\vdots$$

$$W_n = \prod_{i=1}^{n}(qr + (1-q)V_i)$$

$$\implies n^{-1}\log(W_n) = \frac{1}{n}\sum_{i=1}^{n}(qr + (1-q)V_i)$$

$\{qr + (1-q)V_i\}$ is a sequence of i.i.d. nonnegative random variables, so by the Strong Law of Large Numbers, the conclusion follows if $\mathbb{E}[|qr + (1-q)V_1|] = \mathbb{E}[qr + (1-q)V_1] = qr + (1-q)\mathbb{E}(V_1) < \infty$.

(b) Since $w(q)$ is twice differentiable, a sufficient condition for concavity is $w''(q) \leq 0$ for all $q \in (0, 1]$.

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

$$w'(q) = \frac{\partial}{\partial q}\mathbb{E}\log[q(r - V_1) + V_1] = \mathbb{E}\left(\frac{\partial}{\partial q}\log[q(r - V_1) + V_1]\right) = \mathbb{E}\left(\frac{r - V_1}{q(r - V_1) + V_1}\right) \tag{1}$$

$$w''(q) = \mathbb{E}\left(\frac{\partial}{\partial q}\frac{r - V_1}{q(r - V_1) + V_1}\right) = \mathbb{E}\left((r - V_1)\cdot\frac{\partial}{\partial q}\left[q(r - V_1) + V_1\right]^{-1}\right)$$

$$= \mathbb{E}\left((r - V_1)\cdot(-1)\left[q(r - V_1) + V_1\right]^{-2}\cdot(r - V_1)\right) = -\mathbb{E}\left(\left[\frac{(r - V_1)}{q(r - V_1) + V_1}\right]^2\right) \tag{2}$$

This is -1 times the expectation of a nonnegative random variable, so by Markov's Inequality we have

$$\mathbb{E}\left(\left[\frac{(r - V_1)}{q(r - V_1) + V_1}\right]^2\right) \geq 0 \iff w''(q) \leq 0 \qquad \forall q \in (0, 1],$$

proving concavity.

(c) By Jensen's Inequality and the concavity of $q \to \log[qr + (1-q)V_1]$, we have

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

$$\leq \log\left(qr + (1-q)r\right) = \log(r) = \mathbb{E}\log(r) = w(1).$$

where the third step used $\mathbb{E}(V_1) \leq r$ and the second-to-last step used the fact that r is non-random. For $q = 0$, we have

$$n^{-1}\log(W_n) = \frac{1}{n}\sum_{i=1}^{n} V_i.$$

$\{V_i\}$ is a sequence of i.i.d. nonnegative random variables, so by the Strong Law of Large Numbers, almost sure convergence applies if $\mathbb{E}[|V_1|] = \mathbb{E}[V_1] < \infty$. **But this is the same condition as in the case $q \in (0, 1]$.**

To show $w(q) \leq w(0)$, we will show that $w(q)$ is nonincreasing on $[0, 1]$. Recall from (1)

$$w'(q) = \mathbb{E}\left(\frac{r - V_1}{q(r - V_1) + V_1}\right)$$

We will show that (1) is upper-bounded by 0 on $[0, 1]$. Plugging in $q = 0$ yields

$$w'(0) = \mathbb{E}\left(\frac{r - V_1}{V_1}\right) = \mathbb{E}\left(rV_1^{-1} - 1\right) = r\mathbb{E}\left(V_1^{-1}\right) - 1 \leq 1 - 1 = 0.$$

where we used the assumption $\mathbb{E}\left(V_1^{-1}\right) \leq r$ on the second-to-last step. Recall that the second derivative (2) is continuous and nonpositive on $(0, 1]$. Therefore the first derivative (1) never exceeds $q(0) = 0$ on $[0, 1]$, so $w(q) \leq w(0)$.

Exercise 2 (**Mathematical statistics, Bayesian; don't think we need to worry about**). (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) Notice that the first constraint implies $x_1 \leq 0$ (since $x_1^2 + x_2^2 \geq 0$ for all $x_1, x_2 \in \mathbb{R}$). The second constraint implies $x_1 + x_2 = 0 \iff x_1 = -x_2$. So the feasible region is a ray (a one-dimensional cone) extending from the origin into quadrant 3 along the line $x_2 = -x_1$. It is somewhat ambiguous, but it seems that the point $(x_1, x_2) = (0, 0)$ is not feasible because then $h_1(x)$ ought to be undefined. If that is true, then the feasible region is not closed, so there is no reason the minimum needs to exist; indeed, it does not exist, because if $(0, 0)$ is feasible, that is precisely where the minimum would be.

In summary, if $(0, 0)$ is feasible (that is, if we define $0/0$ in such a way that $h_1(0, 0) \leq 0$), then the minimum is 0. If $(0, 0)$ is not feasible, then the minimum does not exist; the infimum is 0.

(b) Since f and g are convex, we have

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y), \qquad g(tx + (1-t)y) \leq tg(x) + (1-t)g(y)$$

We make use of these inequalities to show that $c_1 f + c_2 g$ satisfies (**??**) for any $x, y \in \mathbb{R}^n$ and any $t \in [0, 1]$:

$$[c_f + c_2 g](tx + (1-t)y) = f(tx + (1-t)y) + g(tx + (1-t)y) \leq tf(x) + (1-t)f(y) + tg(x) + (1-t)g(y)$$

$$= t[f(x) + g(x)] + (1-t)[f(x) + g(x)] = t[c_f + c_2 g](x) + (1-t)[c_f + c_2 g](y)$$

which proves the result. (Note that if the initial inequality is strict then strict convexity follows.)

Exercise 5 (High-Dimensional Statistics). Double-check solutions, comments from Jinchi grading homework 7

(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

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

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`.