Statistics 210B Lecture 15 Notes

Daniel Raban

March 8, 2022

1 Concentration of Sample Covariance of Gaussian Random Vectors

1.1 Eigenvalues of sample covariance of Gaussian random vectors

Last time, we started to talk about the eigenvalues of sample covariance matrices of Gaussian random vectors. We had $(x_i)_{\text{iid}}\sim N(0,\Sigma)$, where $\Sigma \in S^{d\times d}$ is a positive definite $d \times d$ matrix. We have

$$X = \begin{bmatrix} x_1^\top \\ \vdots \\ x_n^\top \end{bmatrix} \in \mathbb{R}^{n\times d}, \qquad \widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^n x_i x_i^\top = \frac{1}{n} X^\top X \in \S^{d\times d}.$$

We had the following theorem about the singular values of the random matrix.

Theorem 1.1.

1. $\mathbb{P}(\sigma_{\max}(X)/\sqrt{n} \ge \gamma_{\max}(\sqrt{\Sigma})(1+\tau) + \sqrt{\operatorname{tr}(\Sigma)/n}) \le e^{-nt^2/2}$.
2. $\mathbb{P}(\sigma_{\min}(X)/\sqrt{n} \le \gamma_{\min}(\sqrt{\Sigma})(1+\tau) - \sqrt{\operatorname{tr}(\Sigma)/n}) \le e^{-nt^2/2}$.

The proof strategy was the following:

Proof. For simplicity, take $\Sigma = 1$. We had three main steps:

(a) Concentration: $\mathbb{P}(|\sigma_k(X) - \mathbb{E}[\sigma_k(X)] \ge t) \le 2e^{-t^2/2}$.

(b) $\mathbb{E}[\sigma_{\max}(X)] \le \sqrt{n} + \sqrt{d}$.

(c) $\mathbb{E}[\sigma_{\min}(X)] \ge \sqrt{n} - \sqrt{d}$.

Now we will give the details.

To prove (a), we need to show that the singular values are Lipschitz. By Weyl's inequality,

$$|\sigma_k(x_1) - \sigma_k(x_2)| \le \|X_1 - X_2\|_{\text{op}} \le \|X_1 - X_2\|_F.$$

This impies that $\sigma_k(X)$ is 1-Lipschitz in $\|\cdot\|_F$, the Frobenius norm. Therefore, we get Gaussian concentration, i.e. $\sigma_k(X) - \mathbb{E}[\sigma_k(X)]$ is sG(1).

To prove (b), we wanted an upper bound of $\sigma_{\max}(X)$, using the variational formulation

$$\sigma_{\max} = \sup_{(u,v)\in S^{n-1}\times S^{d-1}} \underbrace{\langle u, Xv\rangle}_{Z_{u,v}}.$$

We introduced the following ineqaality

Lemma 1.1 (Sudakov-Fernique inequality)**.** *Let $\{Z_\theta\}_{\theta\in T}, \{Y_\theta\}_{\theta\in T}$ be two continuous Gaussian processes on a separable space T with $\mathbb{E}[Z_\theta] = \mathbb{E}[Y_\theta]$. If $\mathbb{E}[(Z_\theta - Z_{\theta'})^2] \le \mathbb{E}[(Y_\theta - Y_{\theta'})^2]$ for all $\theta, \theta' \in T$, then*

$$\mathbb{E}\left[\max_{\theta\in T} Z_\theta\right] \le \mathbb{E}\, squa\max_{\theta\in T} Y_\theta.$$

We will prove this later, but first, let's see how this helps us. Define $Z_{u,v} = \langle u, X_v\rangle$, where $X_{i,j} \overset{\text{iid}}{\sim} N(0,1)$, and define

$$Y_{u,v} = \sum_{i=1}^n u_i g_1 + \sum_{j=1}^d \nu_j g_i = \langle u, g\rangle + \langle v, h\rangle, \qquad \overset{\text{iid}}{\sim} N(0,1), h_j \overset{\text{iid}}{\sim} N(0,1).$$

We check the second moment conditions:

$$\mathbb{E}[Z_{u,v}Z_{u',v'}] = \mathbb{E}[\langle X, uv^\top\rangle\langle X, u'(v')^\top\rangle]$$

In the summations, all but the diagonal terms will vanish.

$$\begin{aligned} &= \langle u, v^\top, u'(v')^\top\rangle \\ &= \langle u, u'\rangle\langle v, v'\rangle. \end{aligned}$$

This tells us that

$$\begin{aligned} \mathbb{E}[(Z_{u,v} - Z_{u',v'})^2] &= \underbrace{\mathbb{E}[Z_{u,v}^2]}_{=1} - 2\,\mathbb{E}[Z_{u,v}Z_{u,v'}] + \underbrace{\mathbb{E}[Z_{u',v'}^2]}_{=1} \\ &= 2 - 2\langle u, u'\rangle\langle v, v'\rangle. \end{aligned}$$

For Y, we have

$$\mathbb{E}[(Y_{u,v} - Y_{u',v'})^2] = \underbrace{\mathbb{E}[Y_{u,v}^2]}_{=1} - \underbrace{2\,\mathbb{E}[Y_{u,v}Y_{u'v'}]}_{=2(\langle u,u'\rangle+\langle v,v'\rangle)} + \underbrace{\mathbb{E}[Y_{u',v'}^2]}_{=1}$$

$$= 4 - 2(\langle u, u' \rangle + \langle v, v' \rangle).$$

Then

$$\mathbb{E}[(Y_{u,v} - Y_{u',v'})^2] - \mathbb{E}[(Z_{u,v} - Z_{u',v'})^2] = 2(1 - \langle u, u' \rangle)(1 - \langle v, v' \rangle) \geq 0$$

Now, applying the Sudakov-Fernique inequality gives

$$\begin{aligned}
\mathbb{E}\left[\max_{(u,v)\in S^{n-1}\times S^{d-1}} \langle u, Xv \rangle\right] &\leq \mathbb{E}\left[\max_{(u,v)\in S^{n-1}\times S^{d-1}} (\langle u, g \rangle + \langle v, h \rangle\right] \\
&= \mathbb{E}\left[\max_{(u,v)\in S^{n-1}\times S^{d-1}} \langle u, g \rangle\right] + \mathbb{E}\left[\max_{(u,v)\in S^{n-1}\times S^{d-1}} \langle v, h \rangle\right] \\
&= \mathbb{E}[\|g\|_2] + \mathbb{E}[\|h\|_2] \\
&\leq \mathbb{E}[\|g\|_2^2]^{1/2} + \mathbb{E}[\|h_2\|^2]^{1/2} \\
&= \sqrt{n} + \sqrt{d}.
\end{aligned}$$

For (c), we want to show a lower bound for $\sigma_{\min}(X)$. We want to show that $\sigma_{\min} \geq \sqrt{n} - \sqrt{d}$ (with $n \geq d$). We use the variational representation

$$\sigma_{\min}(X) = \min_{v\in S^{d-1}} \max_{u\in S^{n-1}} \underbrace{\langle u, Xv \rangle}_{Z_{u,v}}$$

Here is another Gaussian process inequality which is a sort of generalization of Sudakov-Fernique.

Theorem 1.2 (Gordon's inequality)**.** *Let $(Z_{s,,t})_{s\in S, t\in T}, (Y_{s,,t})_{s\in S, t\in T}$ be two Gaussian processes with $\mathbb{E}[Z_{s,t}] = \mathbb{E}[Y_{s,t}]$, and suppose that*

$$\begin{cases}
\mathbb{E}[(Z_{s,t_1} - Z_{s,t_2})^2] \geq \mathbb{E}[(Y_{s,t_1} - T_{s,t_2})^2] & \forall t_1, t_2 \in T, s \in S, \\
\mathbb{E}[(Z_{s_1,t_1} - Z_{s_2,t_2})^2] \leq \mathbb{E}[(Y_{s_1,t_1} - T_{s_2,t_2})^2] & \forall s_1 \neq s_2 \in S, t_1, t_2 \in T.
\end{cases}$$

Then

$$\mathbb{E}\left[\max_{s\in S} \min_{t\in T} Z_{s,t}\right] \leq \mathbb{E}\left[\max_{s\in S} \min_{t\in T} Y_{s,t}\right].$$

Take $Y_{u,v} = \langle g, u \rangle + \langle h, v \rangle$. Check that $Z_{u,v}$ and $Y_{u,v}$ satisfy the conditions in the theorem. Then

$$\begin{aligned}
-\mathbb{E}[\sigma_{\min(X)}] &= \mathbb{E}\left[\max_{v\in S^{d-1}} -\|Xv\|_2\right] \\
&= \mathbb{E}\left[\max_{v\in S^{d-1}} \min_{u\in S^{n-1}} \langle u, -Xv \rangle\right]
\end{aligned}$$

$$\leq \mathbb{E}\left[\max_{v\in S^{d-1}} \min_{u\in S^{n-1}} \langle g, u\rangle + \langle h, v\rangle\right]$$

where g, h are iid Gaussian random vectors.

$$= \mathbb{E}\left[\max_{v\in S^{d-1}} \langle h, v\rangle\right] + \mathbb{E}\left[\min_{u\in S^{n-1}} \langle g, u\rangle\right]$$
$$= \underbrace{\mathbb{E}[\|h\|_2]}_{\approx\sqrt{d}} - \underbrace{\mathbb{E}[\|g\|_2]}_{\approx\sqrt{n}}.$$

So we get that

$$\mathbb{E}[\sigma_{\min}(X)] \geq \sqrt{n} - \sqrt{d}. \qquad \square$$

1.2 Proof of the Sudakov-Fernique inequality

Now we will prove the Sudakov-Fernique inequality using the Gaussian interpolation method. Here is a simpler version of the inequality for when the index set is finite.

Lemma 1.2 (Sudakov-Fernique inequality)**.** *Let $X, Y \in \mathbb{R}^n$ be two continuous Gaussian random vectors with $\mathbb{E}[X] = \mathbb{E}[Y]$. If $\mathbb{E}[(X_i - X_j^2] \leq \mathbb{E}[(Y_i - Y_j^2]$ for all i, j, then*

$$\mathbb{E}\left[\max_{i\in[n]} X_i\right] \leq \mathbb{E}\left[\max_{i\in[n]} Y_i\right].$$

Proof. Without loss of generality, we may take X, Y to be independent. Let $\mu = \mathbb{E}[X] = \mathbb{E}[Y]$, and define

$$\widetilde{X} = X - \mu, \widetilde{=} T - \mu, \in \mathbb{R}^n \qquad Z(\theta) = \cos\theta\widetilde{X} + \sin\theta\widetilde{+}\mu.$$

Fix $\beta > 0$, and define the soft max function $F_\beta : \mathbb{R}^n \to \mathbb{R}$ by $F_\beta(x) = \beta^{-1}\log(\sum_{i=1}^n e^{\beta x_i})$. The parameter β determines how soft this"soft max" function is; when $\beta \to \infty$, this will be the max function. For $\theta \in [0, \pi/2]$, let $\varphi(\theta) = \mathbb{E}[F_\beta(Z(\theta))]$. The idea is that $\varphi(0) \approx \mathbb{E}[\max_{i\in[n]} X_i]$ and $\varphi(\pi/2) \approx \mathbb{E}[\max_{i\in[n]} Y_i]$, and these will be exact as we let $\beta \to \infty$.

Using Fubini's theorem and the cain rule, we can calculate the derivative

$$\varphi'(\theta) = \mathbb{E}\left[\sum_{i=1}^n \partial_{x_i} F_\beta(Z(\theta))(-\sin\theta\widetilde{X}_i + \cos\theta\widetilde{Y})\right]$$

Using integration by parts or Stein's lemma,

$$\cos\theta\sin\theta\,\mathbb{E}\left[\sum_{i,j=1}^n \partial^2_{x_i,x_j} F_\beta(Z(\theta))\right](\mathbb{E}[\widetilde{Y};\widetilde{Y}_j] - \mathbb{E}[\widetilde{X}_i;\widetilde{Y}_j])$$

Define $p_i(x) = \partial_{x_i} F_\beta(x) = e^{\beta x_i}/\sum_{j=1}^n e^{\beta x_j}$, which is a probability distribution on $\mathbb{R}^n$. Using some algebra with p_i, we can show that $\varphi'(\theta) \geq 0$. This means that φ is increasing, so $\varphi(0) \leq \varphi(\pi/2)$. Then we let $\beta \to \infty$ to get the inequality. $\square$

The details of the algebra in the proof are contained in chapter 5 of Wainwright's book.

1.3 More on Gaussian comparison inequalities

Here are some comments on these Gaussian comparison inequalities, which are very useful in many cases. There is a more general statement of Gordon's inequality, which contains both an expectation version and a probabilistic version:

Theorem 1.3 (Gordon's inequality)**.** *Let S, T be finite sets (or separable sets with continuous processes). Let $(X_{s,t})_{s\in S,t\in T}, (Y_{s,t})_{s\in S,t\in T}$ be two Gaussian processes with $\mathbb{E}[X_{s,t}] = \mathbb{E}[Y_{s,t}] = 0$, and suppose that*

$$\begin{cases} \mathbb{E}[(X_{s,t_1} - X_{s,t_2})^2] \geq \mathbb{E}[(Y_{s,t_1} - T_{s,t_2})^2] & \forall t_1, t_2 \in T, s \in S, \\ \mathbb{E}[(X_{s_1,t_1} - X_{s_2,t_2})^2] \leq \mathbb{E}[(Y_{s_1,t_1} - T_{s_2,t_2})^2] & \forall s_1 \neq s_2 \in S, t_1, t_2 \in T. \end{cases}$$

Then

1. *For any deterministic function $Q(s,t)$,*

$$\mathbb{E}\left[\max_{s\in S}\min_{t\in T} X_{s,t} + Q(s,t)\right] \leq \mathbb{E}\left[\max_{s\in S}\min_{t\in T} Y_{s,t} + Q(s,t)\right].$$

2. *If we further have $\mathbb{E}[X_{s,t}^2] = \mathbb{E}[Y_{s,t}^2]$, then for all $\tau \in \mathbb{R}$ and functions $Q(s,t)$, we have*

$$\mathbb{P}\left(\min_{s\in S}\max_{t\in T}(X_{s,t} + Q(s,t)) \geq \tau\right) \leq \mathbb{P}\left(\min_{s\in S}\max_{t\in T}(Y_{s,t} + Q(s,t)) \geq \tau\right).$$

For the probabilistic version of the inequality, it is better to assume the mean is zero, but we do not need this for the expectation version.

This inequality can be used to derive the Gaussian contraction inequality: $\mathcal{G}(\phi(T)) \leq \mathcal{G}(T)$ if ϕ is 1-Lipshitz. We can also use it to prove the following.

Theorem 1.4 (Sudakov minorization)**.** *Let $\{X_\theta\}_{\theta\in T}$ be mean 0 Gaussian process on T. Then*

$$\mathbb{E}\left[\sup_{\theta\in T} X_\theta\right] \geq \sup_{\varepsilon>0} \frac{\varepsilon}{2}\sqrt{\log M(\varepsilon; T, \rho_X)},$$

where $M(\varepsilon; T, \rho_X)$ is the packing number of T with metric $\rho_X(\theta,\theta') = \sqrt{\mathrm{Var}(X_\theta - X_{\theta'})}$.

These applications are shown in chapter 5 of Wainwright's book.

1.4 Concentration of sub-Gaussian sample covariance

Now, we generalize our analysis to the case where x_i are sub-Gaussian random vectors, $\mathbb{E}[x_i x_i^\top] = \Sigma \in S^{d\times d}$ is a positive definite $d \times d$ matrix. Here,we still have

$$X = \begin{bmatrix} x_1^\top \\ \vdots \\ x_n^\top \end{bmatrix} \in \mathbb{R}^{n\times d}, \qquad \widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^n x_i x_i^\top = \frac{1}{n} X^\top X \in S^{d\times d}.$$

In this context, similar concentration results will hold.

Definition 1.1. We say a mean 0 random variable $x \in \mathbb{R}^d$ is **sub-Gaussian**(σ) if

$$\mathbb{E}[e^{\lambda\langle v,x\rangle}] \le e^{\lambda^2\|v\|_2^2\sigma^2/2} \qquad \forall \lambda \in \mathbb{R}, v \in \mathbb{R}^d.$$

Remark 1.1. This is not the same as saying that each entry of the vector is sub-Gaussian. But if we suppose $x \in \mathbb{R}^d$ with x_i independent $\mathrm{sG}(\sigma)$, then x is $\mathrm{sG}(\sigma)$:

$$\begin{aligned} \mathbb{E}\left[e^{\lambda\sum_{i=1}^n v_i x_i}\right] &= \prod_{i=1}^n \mathbb{E}[e^{\lambda v_i x_i}] \\ &\le \prod_{i=1}^n e^{\lambda^2 v_i^2\sigma^2/2} \\ &= e^{\lambda^2\|v\|_2^2\sigma^2/2}. \end{aligned}$$

Theorem 1.5. *Let $(x_i)_{i\in[n]}$ be independent mean zero $\mathrm{sG}(\sigma)$. Then with probability at least $1-\delta$, we have*

$$\|\widehat{\Sigma} - \Sigma\|_{\mathrm{op}} \le C\sigma^2\left(\sqrt{\frac{d+\log(1/\delta)}{n}} + \frac{d+\log(1/\delta)}{n}\right).$$

The upper bound is of the same order as the Gaussian case. The only difference is that we lose a universal constant C.