The beta - mixture shrinkage prior for sparse covariances with near-minimax posterior convergence - proof of Lemma6

백승찬

June 9, 2023

1 Introduction

Lemma6

If $s_0^2(\log p)^3 = O(p^2 n)$ and $s_0^2 = O(p^{3-\epsilon})$ for some $\epsilon > 0$, then

$$R^{\gamma_{-1},\lambda_{-1}}_{\lambda_1,\lambda_1'} = -2\log\big(1 - J\epsilon_{np}^2\big) + R^{\gamma_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1',}$$

where $R^{\gamma_{-1},\lambda_{-1}}_{\lambda_1,\lambda_1'} = -\log\det\big(I_p - \Sigma^{-2}(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)\big)$ and $R^{\gamma_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1'}$ satisfies

$$\mathbb{E}_{(\lambda_1,\lambda_1')|J}\left[\mathbb{E}_{(\gamma_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')}\left(\exp\left(\frac{n}{2}R^{\gamma_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1'}\right)\right)\right] \le \frac{3}{2}$$

Proof.

- 먼저 용어에 대한 정의를 한다.
  - $r = \left\lfloor \frac{p}{2} \right\rfloor, \ \epsilon_{np} = \nu\sqrt{\frac{\log p}{n}}, \ \nu = \sqrt{\frac{\epsilon}{4}}$
  - Define parameter space :

$$B_1 := \left\{\Sigma(\theta) : \Sigma(\theta) = I_p + \epsilon_{np}\sum_{m=0}^{r}\gamma_m A_m(\lambda_m), \theta = (\gamma, \lambda) \in \boldsymbol{\theta}\right\}$$

  - $\Lambda := \Big\{\lambda = (\lambda_1, \ldots, \lambda_r)^\top : \lambda_m = (\lambda_{mi}) \in \{0,1\}^p, \|\lambda_m\|_0 = k, \sum_{i=1}^{p-r}\lambda_{mi} = 0, m \in \{1, \ldots, r\}, \quad \text{satisfying} \quad \max_{1\le i\le p}\sum_{m=1}^{r}\lambda_{mi} \le 2k, k = \lceil c_{np}/2\rceil - 1, c_{np} = \lceil s_0/p\rceil\Big\}$

- Consider $R^{\gamma_{-1},\lambda_{-1}}_{\lambda_1,\lambda_1'}$, then
  - 먼저 A를 정의한다 :

$$A = [I - (\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)]^{-1}(\Sigma_0^{-2} - I)(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)$$

  - $R^{\gamma_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1'} \stackrel{\text{def}}{=} -\log\det(I - A)$

– $R^{\gamma_{-1},\lambda_{-1}}_{\lambda_1,\lambda_1'} \stackrel{\text{def}}{=} -\log\det\big[I-(\Sigma_0-\Sigma_1)(\Sigma_0-\Sigma_2)-(\Sigma_0^{-2}-I)(\Sigma_0-\Sigma_1)(\Sigma_0-\Sigma_2)\big]$

- 이제 $R^{\gamma_{-1},\lambda_{-1}}_{\lambda_1,\lambda_1'}$에 대해서 다시 살펴보게 되면
  - Note that

$$\begin{aligned}
R^{\gamma_{-1},\lambda_{-1}}_{\lambda_1,\lambda_1'} &= -\log\det\big[I-(\Sigma_0-\Sigma_1)(\Sigma_0-\Sigma_2)-(\Sigma_0^{-2}-I)(\Sigma_0-\Sigma_1)(\Sigma_0-\Sigma_2)\big]\\
&= -\log\det[\{I-(\Sigma_0-\Sigma_1)(\Sigma_0-\Sigma_2)\}\{(I-A)\}]\\
&= -\log\det[I-(\Sigma_0-\Sigma_1)(\Sigma_0-\Sigma_2)]-\log\det[I-A]\\
&= -\log\det[I-(\Sigma_0-\Sigma_1)(\Sigma_0-\Sigma_2)]+R^{\gamma_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1'}
\end{aligned}$$

  - Note that

$$(I-A)=I-[I-(\Sigma_0-\Sigma_1)(\Sigma_0-\Sigma_2)]^{-1}$$

위 식의 양변에 $[I-(\Sigma_0-\Sigma_1)(\Sigma_0-\Sigma_2)]$ 를 곱한 다음 양변에 $-\log\det$ 를 취해 주게되면

$$\begin{aligned}
&\Rightarrow -\log\det([I-(\Sigma_0-\Sigma_1)(\Sigma_0-\Sigma_2)](I-A))\\
&\qquad= -\log\det\big(I-(\Sigma_0-\Sigma_1)(\Sigma_0-\Sigma_2)-(\Sigma_0^{-2}-I)(\Sigma_0-\Sigma_1)(\Sigma_0-\Sigma_2)\big)
\end{aligned}$$

- 이제 다음과 같은 사실을 보이려고 한다.

$$-\log\det(I-(\Sigma_0-\Sigma_1))(\Sigma_0-\Sigma_2)=-2\log\big(1-J\epsilon_{np}^2\big)$$

  - 이때 $\Sigma_0,\Sigma_1,\Sigma_2$ 의 정의에 관해 설명하고자 한다.

(1)

$$\Sigma_0=\begin{pmatrix}1 & \mathbf{0}_{1\times(p-1)}\\ \mathbf{0}_{(p-1)\times1} & \mathbf{S}_{(p-1)\times(p-1)}\end{pmatrix}$$

(2)

$$\Sigma_1=\begin{pmatrix}1 & \mathbf{v}_{1\times(p-1)}\\ \mathbf{v}_{(p-1)\times1} & \mathbf{S}_{(p-1)\times(p-1)}\end{pmatrix}$$

(3)

$$\Sigma_2=\begin{pmatrix}1 & \mathbf{v}^*_{1\times(p-1)}\\ \mathbf{v}^*_{(p-1)\times1} & \mathbf{S}_{(p-1)\times(p-1)}\end{pmatrix}$$

(4)

$$\Sigma_1 - \Sigma_0 = \begin{pmatrix} 0 & \mathbf{v}_{1\times(p-1)} \\ \mathbf{v}_{(p-1)\times 1} & \mathbf{0}_{(p-1)\times(p-1)} \end{pmatrix}$$

(5)

$$\Sigma_2 - \Sigma_0 = \begin{pmatrix} 0 & \mathbf{v}^*_{1\times(p-1)} \\ \mathbf{v}^*_{(p-1)\times 1} & \mathbf{0}_{(p-1)\times(p-1)} \end{pmatrix}$$

(6)

$$\mathbf{v}_{1\times(p-1)} = (v_j)_{2\le j\le p} = \begin{cases} 0 & (\ 2 \le j \le p-r) \\ 0\ or\ \epsilon_{np} & (\ p-r+1 \le j \le p) \end{cases} with\ \ \|\mathbf{v}\|_0 = k$$

(7)

$$\mathbf{v}^*_{1\times(p-1)} = (v^*_j)_{2\le j\le p} = \begin{cases} 0 & (\ 2 \le j \le p-r) \\ 0\ or\ \epsilon_{np} & (\ p-r+1 \le j \le p) \end{cases} with\ \ \|\mathbf{v}^*\|_0 = k$$

(8)

$$(v_j) = \begin{cases} \epsilon_{np} & (\ p-r+1 \le j \le p-r+k) \\ 0 & (\ o.w.) \end{cases}$$

(9)

$$(v_j)^* = \begin{cases} \epsilon_{np} & (\ p-r+k-J+1 \le j \le p-r+2k-J) \\ 0 & (\ o.w.) \end{cases}$$

(10)

$$\mathbf{S}_{(p-1)\times(p-1)} = (s_{ij})_{2\le i,j\le p} \text{ is uniquely determined by } (\gamma_{-1}, \lambda_{-1}),$$

$$\text{where } (\gamma_{-1}, \lambda_{-1}) = ((\gamma_2, \ldots, \gamma_r), (\lambda_2, \ldots, \lambda_r)), \text{and } (s_{ij}) = \begin{cases} 1 & (\ i = j) \\ \epsilon_{np} & (\ \gamma_i = \lambda_i(j) = 1) \\ 0 & (\ o.w.) \end{cases}$$

- 위의 정의를 바탕으로 $I - (\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)$ 의 특징과 그 성질을 알아보고자 한다.

- Define J to be of overlapping ϵ_{np}'s between Σ_1 and Σ_2 on the 1st row, $Q \triangleq (q_{ij})_{1 \le i,j \le p} = (\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)$

- Let index subset I_r and I_c in $2, \ldots, p$ with $\mathrm{Card}(I_r) = \mathrm{Card}(I_c) = k$, and $\mathrm{Card}(I_r \cap I_c) = J$, s.t. $(q_{ij}) = \begin{cases} J\epsilon_{np}^2 & (\ i = j = 1) \\ \epsilon_{np}^2 & (\ i \in I_r \ \ \& \ \ j \in I_c) \\ 0 & (\ o.w.) \end{cases}$

- Q가 어떤 원리로 구성되는지 살펴보기 위해 간단한 예시를 들어보고자 한다.

$$(\Sigma_0 - \Sigma_1) = (\Sigma_0 - \Sigma_2) = \begin{pmatrix} 0 & 0 & \epsilon_{np} & \epsilon_{np} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \epsilon_{np} & 0 & 0 & 0 & 0 \\ \epsilon_{np} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

- 이때 $(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)$ 은

$$(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2) = \begin{pmatrix} 2\epsilon_{np}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \epsilon_{np}^2 & \epsilon_{np}^2 & 0 \\ 0 & 0 & \epsilon_{np}^2 & \epsilon_{np}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

  와 같이 구성되는데 이때 $(1,1)$성분에서 2가 의미하는 것은 위에서 정의한 J와 같음을 알 수 있다. 즉 Q의 대각성분중 0이 아닌 것의 갯수와 일치하는 것이다. 또한 모든 가능한 $(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)$에 대하여 linearly independent vector는 단 2개밖에 존재하지 않으므로 $rank[(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)] = 2$ 임을 알 수 있다.

- 이러한 사실들을 토대로 $I - (\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)$ 의 characteristic polynomial 을 구해보고자 한다. 즉 $det[\lambda I - (\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)]$을 구하는 것인데 일반적인 경우의 값을 구하기 쉽지않아 특수한 경우에 한해서 구해보려고 한다. 즉 $J = k$ 인 경우만을 고려해보고자 한다. 이러한 경우

$$\det[\lambda I - (\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)] = \det\left[\begin{pmatrix} \lambda I_{p-k} - \begin{pmatrix} J\epsilon_{np}^2 & \mathbf{0} \\ \mathbf{0}^\top & \mathbf{O} \end{pmatrix} & \mathbf{0}_{(p-k)\times(k)} \\ \mathbf{0}_{(k)\times(p-k)} & \lambda I_k - \epsilon_{np}^2 \mathbf{1}_k \mathbf{1}_k^\top \end{pmatrix}\right]$$

$$= \det\left[\lambda I_{p-k} - \begin{pmatrix} J\epsilon_{np}^2 & \mathbf{0} \\ \mathbf{0}^\top & \mathbf{O} \end{pmatrix}\right] \det\left[\lambda I_k - \epsilon_{np}^2 \mathbf{1}_k \mathbf{1}_k^\top\right]$$

$$= (\lambda - J\epsilon_{np}^2)\lambda^{p-k-1}\det\left[\lambda I_k - \epsilon_{np}^2 \mathbf{1}_k \mathbf{1}_k^\top\right]$$

$$= (\lambda - J\epsilon_{np}^2)\lambda^{p-k-1}\det\left(1 - \epsilon_{np}^2 \mathbf{1}_k^\top (\lambda I_k)^{-1} \mathbf{1}_k\right)\det\left(\lambda I_k\right))$$

$$= (\lambda - J\epsilon_{np}^2)\lambda^{p-k-1}\left(1 - \frac{J}{\lambda}\epsilon_{np}^2\right)\lambda^k$$

$$= (\lambda - J\epsilon_{np}^2)^2\lambda^{p-2}$$

– Note that

$$\begin{aligned} R^{\gamma_{-1},\lambda_{-1}}_{\lambda_1,\lambda_1'} &= -\log\det[I - (\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)] + R^{\gamma_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1'} \\ &= -\log\left(1 - J\epsilon_{np}^2\right)^2 + R^{\gamma_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1'} \\ &= -2\log\left(1 - J\epsilon_{np}\right) + R^{\gamma_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1'} \end{aligned}$$

– Note that

$$\det[\lambda I - (\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)] \overset{\lambda=1}{=} \left(1 - J\epsilon_{np}^2\right)^2$$

- 이제 다음과 같은 사실을 보이려고 한다.

$$\mathbb{E}_{(\lambda_1,\lambda_1')|J}\left[\mathbb{E}_{(\gamma_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')}\left(\exp\left(\frac{n}{2}R^{\gamma_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1'}\right)\right)\right] \leq \frac{3}{2}$$

  - Recall : $A = [I - (\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)]^{-1}(\Sigma_0^{-2} - I)(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)$

  - It is important to observe that $rank(A) \leq 2$ due to the structure of $(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)$. Let ϱ be an eigenvalue of A. It is easy to see that $|\varrho| \leq \|A\|$.

  - We wish to find the upper bound for$\|A\|$. To proceed, first we can see that

$$\|A\| \leq \left\|[I - (\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)]^{-1}\right\|\left\|\Sigma_0^{-2} - I\right\|\|(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)\|.$$

  - From Tony Cai's "Optimal rates of convergence for sparse covariance matrix estimation" (22), we can see that

$$\|\Sigma_1 - \Sigma_0\| \leq \|\Sigma_1 - \Sigma_0\|_1 = k\epsilon_{np} \leq 2k\epsilon_{np} \leq c_{np}\epsilon_{np}^{1-q} \leq Mv^{1-q} < \frac{1}{3}$$

    Similarly, we can see that for $\|I - \Sigma_0\|$ and $\|(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)\|$,

$$\|I - \Sigma_0\| \leq \|I - \Sigma_0\|_1 = k\epsilon_{np} < \frac{1}{3},\ \|(\Sigma_0 - \Sigma_1)(\Sigma_0 - \Sigma_2)\| \leq \frac{1}{3} \times \frac{1}{3} < 1.$$

  - Note that : $|\log(1-x)| \leq 2|x|$, for $|x| < \frac{1}{6}$, $\quad R^{\gamma_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1'} \stackrel{\text{def}}{=} -\log\det(I - A)$

$$\begin{aligned}\Rightarrow R^{\gamma_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1'} &= -\log\det(I - A) \leq -2(1 - \det(I - A)) \\ &\leq |-2(1 - \det(I - A))| \leq |-2(\det(I) - \det(I - A)) \\ &\leq 2|\det(I) - \det(I - A)| = 2m\|I - (I - A)\| \\ &= 2m\|A\|, \quad \text{for some } m > 0.\end{aligned}$$

$$\Rightarrow R^{\gamma_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1'} \leq 4\|A\|, \quad \text{for } m = 2.$$

  - Note that : $\exp\left(\frac{n}{2}R^{\gamma_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1'}\right) \leq \exp\left(2n\|A\|\right)$,

  - Note that : for any square matrix B, following statement is true :

$$\left(\sum_{m=1}^{\infty} B^m\right)^2 = \left(B + B^2 + \ldots\right)\left(B + B^2 + \ldots\right) = B^2 + 2B^3 + 3B^4 + \ldots = \sum_{m=0}^{\infty} mB^{m+1}$$

– From above statement, we can write the following :

$$\begin{aligned}
\Sigma_0^{-2} - I = (I-(I-\Sigma_0))^{-2} - I &= \left(I + \sum_{m=1}^{\infty}(I-\Sigma_0)^m\right)^2 - I \\
&= I + 2\sum_{m=1}^{\infty}(I-\Sigma_0)^m + \left(\sum_{m=1}^{\infty}(I-\Sigma_0)^m\right)^2 - I \\
&= 2\sum_{m=1}^{\infty}(I-\Sigma_0)^m + \left(\sum_{m=1}^{\infty}(I-\Sigma_0)^m\right)^2 \\
&= 2\sum_{m=0}^{\infty}(I-\Sigma_0)^{m+1} + \left(\sum_{m=1}^{\infty}(I-\Sigma_0)^m\right)^2 \\
&= 2\sum_{m=0}^{\infty}(I-\Sigma_0)^{m+1} + \sum_{m=0}^{\infty}m(I-\Sigma_0)^{m+1} \\
&= \left[\sum_{m=0}^{\infty}(m+2)(I-\Sigma_0)^m\right](I-\Sigma_0)\,.
\end{aligned}$$

– We can see that

$$\left\|\sum_{m=0}^{\infty}(m+2)(I-\Sigma_0)^m\right\| = \sum_{m=0}^{\infty}(m+2)\|I-\Sigma_0\|^m < \sum_{m=0}^{\infty}(m+2)\left(\frac{1}{3}\right)^m = \frac{13}{4} < 4.$$

– Define $A_* = (I-\Sigma_0)(\Sigma_0-\Sigma_1)(\Sigma_0-\Sigma_2)$. Then

$$\begin{aligned}
\|A\| &\le \left\|[I-(\Sigma_0-\Sigma_1)(\Sigma_0-\Sigma_2)]^{-1}\right\| \left\|\left[\sum_{m=0}^{\infty}(m+2)(I-\Sigma_0)^m\right]A_*\right\| \\
&\le \left\|[I-(\Sigma_0-\Sigma_1)(\Sigma_0-\Sigma_2)]^{-1}\right\| \left\|\sum_{m=0}^{\infty}(m+2)(I-\Sigma_0)^m\right\| \|A_*\| \\
&< \left\|[I-(\Sigma_0-\Sigma_1)(\Sigma_0-\Sigma_2)]^{-1}\right\| \cdot 4 \cdot \|A_*\| \\
&< 4 \cdot \frac{1}{1-\frac{1}{3}\cdot\frac{1}{3}} \cdot \|A_*\| = \frac{9}{2}\|A_*\| \le \frac{9}{2}\max\{\|A_*\|_1, \|A_*\|_\infty\}\,,
\end{aligned}$$

where $A_* = (a^*_{ij})_{1\le i,j\le p}$, and $\begin{cases} \|A_*\|_1 = \max\limits_{1\le m\le p}\sum\limits_{j=1}^{p}|a^*_{mj}| \\ \|A_*\|_\infty = \max\limits_{1\le m\le p}\sum\limits_{i=1}^{p}|a^*_{im}| \end{cases}$

– Summing up above results, we obtain following :

$$\exp\left(\frac{n}{2}R^{\gamma_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1'}\right) \le \exp(2n\|A\|) = \exp(9n\max\{\|A_*\|_1, \|A_*\|_\infty\})$$

which implies

$$\begin{aligned}
&\mathbb{E}_{(\lambda_1,\lambda_1')|J}\left[\mathbb{E}_{(\gamma_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')}\left(\exp\left(\frac{n}{2}R^{\gamma_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1'}\right)\right)\right] \\
&\le \mathbb{E}_{(\lambda_1,\lambda_1')|J}\left[\mathbb{E}_{(\gamma_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')}(\exp(9n\max\{\|A_*\|_1, \|A_*\|_\infty\}))\right]
\end{aligned}$$

– But infact, :

$$\mathbb{E}_{(\lambda_1,\lambda_1')|J}\left[\mathbb{E}_{(\gamma_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')}\left(I\left\{\max\left\{\|A_*\|_1,\|A_*\|_\infty\right\}\geq 2tk\epsilon_{np}^3\right\}\right)\right]$$

$$=\mathbb{P}\left(\max\left\{\|A_*\|_1,\|A_*\|_\infty\right\}\geq 2tk\epsilon_{np}^3\right)$$

– So we wish to show that

$$\mathbb{P}\left(\sum_{j=1}^{p}|a_{mj}^*|\geq 2tk\epsilon_{np}^3\right)\leq\left(\frac{k^2}{p/8-1-k}\right)^t$$

which implies that

$$\mathbb{P}(\max\left\{\|A_*\|_1,\|A_*\|_\infty\right\}\geq 2tk\epsilon_{np}^3)\leq 2p\left(\frac{k^2}{p/8-1-k}\right)^t$$

– For each row m, define $E_m=\{1,2,\ldots,r\}\backslash\{1,m\}$. Note that for each column of λ_{E_m}, if the column sum of λ_{E_m} is less than or equal to $2k-2$, then the other two rows can still freely take values 0 or 1 in this column, because the total sum will still not exceed $2k$. Let $n_{\lambda_{E_m}}$ be the number of columns of λ_{E_m} with column sum at least $2k-1$, and define $p_{\lambda E_m}=r-n_{\lambda_{E_m}}$. Without loss of generality we assume that $k\geq 3$. Since $n_{\lambda_{E_m}}\cdot(2k-2)\geq r\cdot k$, the total number of 1's in the upper triangular matrix by the construction of the parameter setm we thus have $n_{\lambda_{E_m}}\geq r\cdot\dfrac{3}{4}$, which immediately implies $p_{\lambda_{E_m}}=r-n_{\lambda_{E_m}}\geq\dfrac{r}{4}\geq\dfrac{p}{8}-1$. Recall that the distribution of $(\gamma_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')$ is uniform over $\Theta^{-1}(\lambda_1,\lambda_1')$.

– Recall that J is the overlapping nonzero entries between the 1st rows of Σ_1 and Σ_2, i.e. $J=\lambda_1^\top\lambda_1'$. Then we can obtain the following results :

$$\mathbb{E}_J[I_{(J=t)}|\lambda_{E_m}]=\frac{\binom{k}{t}\binom{p_{\lambda_{E_m}}-k}{k-t}}{\binom{p_{\lambda_{E_m}}}{k}}=\left[\frac{k!}{(k-t)!}\right]^2\cdot\frac{[(p_{\lambda_{E_m}}-k)!]^2}{p_{\lambda_{E_m}}!(p_{\lambda_{E_m}}-2k+t)!}\cdot\frac{1}{t!}\leq\left(\frac{k^2}{p_{\lambda_{E_m}}-k}\right)^j$$

$$\Rightarrow\ \mathbb{E}_J[I_{(J=t)}]=\mathbb{E}_{\lambda_{E_m}}\left[\mathbb{E}_J(\,I_{(J=t)}|\lambda_{E_m}\,)\right]\leq\mathbb{E}_{\lambda_{E_m}}\left[\left(\frac{k^2}{p_{\lambda_{E_m}}-k}\right)^t\right]\leq\left(\frac{k^2}{p/8-1-k}\right)^t$$

Then we can obtain the following :

$$\Rightarrow\ \mathbb{P}\left(\sum_{j=1}^{p}|a_{mj}^*|\geq 2tk\epsilon_{np}^3\ \middle|\ \lambda_{E_m}\right)\leq\left(\frac{k^2}{p/8-1-k}\right)^t$$

which implies for every $t>2$,

$$\Rightarrow\ \mathbb{P}\left(\sum_{j=1}^{p}|a_{mj}^*|\geq 2tk\epsilon_{np}^3\lambda_{E_m}\right)\leq\left(\frac{k^2}{p/8-1-k}\right)^{t-1}$$

This implies : $\mathbb{P}(\max\{\|A_*\|_1, \|A_*\|_\infty\} \geq 2tk\epsilon_{np}^3) \leq 2p\left(\dfrac{k^2}{p/8-1-k}\right)^{t-1}$ for every $t>2$. so $\mathbb{E}_{(\lambda_1,\lambda_1')|J}\left[\mathbb{E}_{(\gamma_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')}\left(I\left\{\max\{\|A_*\|_1, \|A_*\|_\infty\} \geq 2tk\epsilon_{np}^3\right\}\right)\right] = \mathbb{P}\left(\max\{\|A_*\|_1, \|A_*\|_\infty\} \geq 2tk\epsilon_{np}^3\right) \leq 2p\left(\dfrac{k^2}{p/8-1-k}\right)^{t-1}$ for every $t>2$.

– Recall that

$$
\begin{aligned}
&\mathbb{E}_{(\lambda_1,\lambda_1')|J}\left[\mathbb{E}_{(\gamma_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')}\left(\exp\left(\frac{n}{2}R^{\gamma_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1'}\right)\right)\right] \\
&\leq \mathbb{E}_{(\lambda_1,\lambda_1')|J}\left[\mathbb{E}_{(\gamma_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')}(\exp(9n\max\{\|A_*\|_1, \|A_*\|_\infty\}))\right].
\end{aligned}
$$

– Note that for any r.v. $X \geq 0$ & constant $a \geq 0$, it is known that

$$
\begin{aligned}
\mathbb{E}[X] &= \int_{x\geq 0} P(X>x)\ dx = \int_{x\leq a} P(X>x)\ dx + \int_{x>a} P(X>x)\ dx \\
&= \int_{x\leq a}(1-F(x))\ dx + \int_{x>a} P(X>x)\ dx \\
&= \Big[(1-F(x))\Big]_0^a + \int_0^a xf(x)\ dx + \int_{x>a} P(X>x)\ dx \\
&= a(1-F(a)) + \int_0^a xf(x)\ dx + \int_{x>a} P(X>x)\ dx \\
&\leq a + \int_{x>a} P(X>x)\ dx
\end{aligned}
$$

– we can apply this fact to our objective, in other words, put $a = \exp\left\{2Cnk\epsilon_{np}^3\dfrac{1+2\epsilon}{\epsilon}\right\}$, since $k = \lceil c_{np}/2\rceil - 1, c_{np} = \lceil s_0/p\rceil, \epsilon_{np} = \nu\sqrt{\log p/n}, \nu = \sqrt{\epsilon/4}$. Then we could achieve the upper bound for a with the condition $s_0^2(\log p)^3 = O(p^2 n)$ as following : (Here, we put $9 = C,\ C>0$, for convenience, which doesn't affect the upper bound we are looking for.)

$$
\begin{aligned}
a &= \exp\left\{2Cnk\epsilon_{np}^3\frac{1+2\epsilon}{\epsilon}\right\} \\
&= \exp\left(2Cn\left\{\left\lceil\frac{\lceil s_0/p\rceil}{2}\right\rceil - 1\right\}\right)\left(\frac{\epsilon}{4}\right)\left(\frac{\log p}{n}\right)^{3/2}\left(\frac{1+2\epsilon}{\epsilon}\right)\sqrt{\frac{\epsilon}{4}} \\
&\leq \exp\left(Cn\left\{\frac{s_o}{2p}+\frac{1}{2}\right\}\left(\frac{\log p}{n}\right)^{3/2}\left(\frac{1+2\epsilon}{4}\right)\sqrt{\epsilon}\right) \\
&= \exp\left(\frac{1}{2}C\left(\frac{1+2\epsilon}{4}\right)\sqrt{\epsilon}\left[\left(\frac{s_o}{p}+1\right)^2\frac{(\log p)^3}{n}\right]^{1/2}\right) \\
&\asymp \exp\left(\frac{1}{2}C\left(\frac{1+2\epsilon}{4}\right)\sqrt{\epsilon}\right) \asymp e^0 = 1 \\
&< \frac{3}{2},\ \ \text{for sufficiently small}\ \ \epsilon > 0.
\end{aligned}
$$

– Now, from our finding,

$$\mathbb{E}_{(\lambda_1,\lambda_1')|J}\left[\mathbb{E}_{(\gamma_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')}(\exp(Cn\max\{\|A_*\|_1,\|A_*\|_\infty\}))\right]$$

$$\leq a+\int_{x>a}\mathbb{E}_{(\lambda_1,\lambda_1')|J}\left[\mathbb{E}_{(\gamma_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')}\left(I\left\{\max\{\|A_*\|_1,\|A_*\|_\infty\}\geq 2tk\epsilon_{np}^3\right\}\right)\right]\ dx$$

$$\leq \frac{3}{2}+\int_{t\geq(1+2\epsilon)/\epsilon}2Cnk\epsilon_{np}^3\exp\left(2Ctnk\epsilon_{np}^3\right)2p\left(\frac{k^2}{p/8-1-k}\right)^{t-1}\ dt$$

$$\leq \frac{3}{2}+\int_{t\geq(1+2\epsilon)/\epsilon}\exp\left\{\log(2p)-(t-1)\log\left(\frac{p/8-1-k}{k^2}\right)+2C(t+1)nk\epsilon_{np}^3\right\}\ dt.$$

Thus, we complete the proof if we show that the second term of last inequality is of order $o(1)$. Note that :

$$\begin{aligned}
(t-1)\log\left(\frac{p/8-1-k}{k^2}\right) &\geq \left(1+\frac{1}{\epsilon}\right)\log\left(\frac{p/8-1-k}{k^2}\right)\\
&= \left(1+\frac{1}{\epsilon}\right)\log\left(\frac{p/8-1-(s_0/2p+1/2)}{(s_0/2p+1/2)^2}\right)\\
&\geq \left(1+\frac{1}{\epsilon}\right)\log\left(\frac{p/8-1-(s_0/p)}{(s_0/p)^2}\right)+C'\\
&= \left(1+\frac{1}{\epsilon}\right)\log\left(\frac{p^3/8-p^3-ps_0}{s_0^2}\right)+C'\\
&= \left(1+\frac{1}{\epsilon}\right)\log\left(\frac{p^\epsilon p^{3-\epsilon}}{s_0^2}\left(\frac{1}{8}-\frac{1}{p}-\frac{s_0}{p^2}\right)\right)+C'\\
&\geq \left(1+\frac{1}{\epsilon}\right)\log(p^\epsilon)+C''\\
&= (1+\epsilon)\log(p)+C'',
\end{aligned}$$

for any $t>(1+2\epsilon)/\epsilon$ and some constants $C'^{>0}$ and $C''>0$. The third inequality follows from the assumption $s_0^2=O(p^{3-\epsilon})$. Therefore, it implies that the second term of last inequality is of order $o(1)$, which gives the desired result:

$$\mathbb{E}_{(\lambda_1,\lambda_1')|J}\left[\mathbb{E}_{(\gamma_{-1},\lambda_{-1})|(\lambda_1,\lambda_1')}\left(\exp\left(\frac{n}{2}R^{\gamma_{-1},\lambda_{-1}}_{1,\lambda_1,\lambda_1'}\right)\right)\right]\leq\frac{3}{2}$$

□