Three-way Interactions

- Computational Complexity and Practical Issues
  - Page 10 in iFor paper
  - FS has a cost of $O(nm)$ for each step
    - iForm two-way has at most $p + \frac{n(n+1)}{2}$ or $m \leq p + n(n+1)$ holds for any step
    - Overall Complexity $nO\left(n\big(p + n(n+1)\big)\right) = O(n^2 p + n^4)$
    - $D$ controls length of procedure, they tried $\frac{n}{4}, \frac{n}{3}, \frac{n}{2}$
- Theoretical Results
- Long-term concern about two-stage models because of theoretical validity
  - Hao and Zhang proved that two-stage model captures main effects in ultra-high dimensional situations
- Screening Consistency
  - Page 14

Naively, we can use any one-stage variable selection tool to fit (1.1) directly (as long as computation is feasible), ignoring the hierarchical structure. Though the model consistency or screening consistency result (Zhao & Yu, 2006; Wang, 2009; Fan & Lv, 2011) could be generalized to the context of interaction selection, the extension of earlier proofs is not straightforward due to heavy tails of interaction effects. Actually, all the existing proof technique would require some regularity conditions on the eigenvalues of $\Sigma_{(2)}$. Next, we establish the screening consistency of FS2 under conditions that are related only to $\Sigma_{(1)}$.

- C1
  - $X_{i1}, \ldots, X_{ip}$ are jointly and marginally standard normal

In this section we work on the total covariance matrix $\Sigma$ and show it is determined by the covariance of the matrix $\Sigma^{(1)}$ of main effects under the Gaussian assumption

For $X_j$ the main effects and $Z_{kl} = X_k X_l - E(X_k X_l)$ for $(k,l) \in \mathcal{P}_2$ the interactions and $W_{rst} = X_r X_s X_t - E(X_r X_s X_t)$ for $(r,s,t) \in \mathcal{P}_3$ for order 3 effects

**Lemma 1**

Under the normality condition (C1), for $\forall\ j,k,l,r,s,t$

1. $cov\left(X_j, Z_{kl}\right) = 0$
2. $cov\left(X_j, W_{rst}\right) = 0$
3. $cov(\ Z_{kl}, W_{rst}) = 0$

$$\Sigma = \begin{pmatrix} \Sigma^{(1)} & 0 & 0 \\ 0 & \Sigma^{(2)} & 0 \\ 0 & 0 & \Sigma^{(3)} \end{pmatrix}$$

Proof:

1. $cov\left(X_j, Z_{kl}\right) = cov\left(X_j,\ X_k X_l\right) = E\left(X_j X_k X_l\right) - E\left(X_j\right) E(X_k X_l) = 0$
2. $cov\left(X_j, W_{rst}\right) = cov\left(X_j,\ X_r X_s X_t\right) = E\left(X_j X_r X_s X_t\right) - E\left(X_j\right) E\left(X_r X_s X_t\right) = 0$
3. $cov(Z_{kl}, W_{rst}) = cov(X_k X_l,\ X_r X_s X_t) = E(X_k X_l X_r X_s X_t) - E(X_k X_l) E(X_r X_s X_t) = 0$

   This holds if the joint density of

   $X_1, \ldots., X_p\ is\ symmetric\ with\ repsec\ to\ the\ origin\ point\ 0$

**Lemma 2** (still need to extend to order-three interactions)

Generic Formula:

$$E\left\{\prod_{i=1}^{n} X_i^{a_i}\right\} = \sum_{I \in S_a} d_{a,I} \left(\prod_{i=1}^{n}\prod_{j=1}^{n} \varphi_{ij}^{I_{ij}}\right)\left(\prod_{j=1}^{n} \mu_j^{I_{aj}}\right)$$

$$d_{a,I} = \frac{\prod_{k=1}^{n} a_k!}{2^{M_I}\left(\prod_{i=1}^{n}\prod_{j=1}^{n} I_{ij}!\right)\left(\prod_{j=1}^{n} I_{aj}!\right)}$$

Under the normality condition (C1)

$cov\left(Z_{ij}, Z_{kl}\right) = \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}$ Bar & Dittrich, 1971

$$\begin{aligned} E\left(X_iX_jX_kX_l\right) &= E\left(X_iX_j\right)E(X_kX_l) + E(X_iX_k)E\left(X_jX_l\right) + E(X_iX_l)E\left(X_jX_k\right) \\ &\quad - 2E(X_i)E\left(X_j\right)E(X_k)E(X_l) = 0 + E(X_iX_k)E\left(X_jX_l\right) + E(X_iX_l)E\left(X_jX_k\right) - 0 \\ &= E(X_iX_k)E\left(X_jX_l\right) + E(X_iX_l)E\left(X_jX_k\right) = \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk} \end{aligned}$$

$$\begin{aligned} &E(X1X2X3X4X5X6) \\ &= \rho_{12}\rho_{34}\rho_{56} + \rho_{12}\rho_{35}\rho_{46} + \rho_{12}\rho_{36}\rho_{45} + \rho_{13}\rho_{24}\rho_{56} + \rho_{13}\rho_{25}\rho_{46} + \rho_{13}\rho_{26}\rho_{45} \\ &+ \rho_{14}\rho_{23}\rho_{56} + \rho_{14}\rho_{25}\rho_{36} + \rho_{14}\rho_{26}\rho_{35} + \rho_{15}\rho_{23}\rho_{46} + \rho_{15}\rho_{24}\rho_{36} + \rho_{15}\rho_{26}\rho_{34} \\ &+ \rho_{16}\rho_{23}\rho_{45} + \rho_{16}\rho_{24}\rho_{35} + \rho_{16}\rho_{25}\rho_{34} \\ &= 0 + 0 + 0 + 0 + \rho_{13}\rho_{25}\rho_{46} + \rho_{13}\rho_{26}\rho_{45} + 0 + \rho_{14}\rho_{25}\rho_{36} + \rho_{14}\rho_{26}\rho_{35} \\ &+ \rho_{15}\rho_{23}\rho_{46} + \rho_{15}\rho_{24}\rho_{36} + 0 + \rho_{16}\rho_{23}\rho_{45} + \rho_{16}\rho_{24}\rho_{35} + 0 \end{aligned}$$

(7.1)

Let $A = \left(A_{ij}\right)$ be an N x N matrix. In linear algebra, a K x K submatrix is called a principal submatrix if it is of the form $A_I = \left(A_{l_i,l_j}\right)$ where $\mathcal{I}$ is an index set $\mathcal{I} = \{1 \le l_1 < \cdots < l_K \le N\}$. Here with slight abuse of this conception, we allow arbitrary order of the index set $\mathcal{I}$. For example, let $\mathcal{I} = \{2,1\}$ and $A_{\mathcal{I}} = \begin{pmatrix} A_{22} & A_{21} \\ A_{12} & A_{11} \end{pmatrix}$ is still called a principal submatrix in this paper.

Based on the formula (7.1), we can decompose $\Sigma^{(2)}$ to a sum $\Sigma_1^{(2)} + \Sigma_2^{(2)}$. In fact, we have

**Lemma 3**

Both $\Sigma_1^{(2)}$ *and* $\Sigma_2^{(2)}$ *are principal submatrices of* $\Sigma^{(1)} \otimes \Sigma^{(1)}$

Proof: The Kronecker product (Laub, 2005) $\Sigma^{(1)} \otimes \Sigma^{(1)}$ is a $p^2 x\ p^2$ matrix whose rows and columns are both indexed by the set $\mathcal{P}_1 \times \mathcal{P}_1$. The entry corresponding to the index $(ij, kl) is\ \sigma_{ij}\sigma_{kl}$. By formula (7.1), both $\Sigma_1^{(2)}$ *and* $\Sigma_2^{(2)}$ *are* $\frac{p(p+1)}{2} \times \frac{p(p+1)}{2}$ *principal submatrices of* $\Sigma^{(1)} \otimes \Sigma^{(1)}$

**Lemma 4:** *Under C1 and C2a we have*

$$2\tau_{min} < \lambda_{min}(\Sigma) \leq \lambda_{max}(\Sigma) < \tau_{max}/2 \qquad (7.2)$$

Proof: By Laub (2005) Theorem 13.12, the eigenvalues of $\Sigma^{(1)} \otimes \Sigma^{(1)}$ are $\lambda_i\lambda_j, 1 \leq i, j \leq p$, if the eigenvalues of $\Sigma^{(1)}$ are $\lambda_1, \ldots, \lambda_p$. Therefore under condition C2a we have

$$\tau_{min} < \lambda_{min}\left( \Sigma^{(1)} \otimes \Sigma^{(1)}\right) \leq \lambda_{max}\left( \Sigma^{(1)} \otimes \Sigma^{(1)}\right) < \tau_{max}/4$$

By Lemma 3, the eigenvalues of $\Sigma_1^{(2)}$ *and* $\Sigma_2^{(2)}$ are also bounded by $\tau_{min}$ *and* $\tau_{max}/4$, so

$$2\tau_{min} < \lambda_{min}\left(\Sigma^{(2)}\right) \leq \lambda_{max}\left(\Sigma^{(2)}\right) < \tau_{max}/2$$

It is straight forward to get (7.2)

**Appendix B. A Bernstein Inequality and Its Application**

**Lemma 5**

$Let\ W_1, \ldots, W_n\ be\ independent\ random\ variables\ with\ mean\ zero\ and\ variances\ bounded\ by\ \sigma^2 \geq 1.\ Assume\ for\ some\ 0 < \alpha < 1,$

$$E\left(|W_i|^{3(1-\alpha)} e^{t|W_i|^\alpha}\right) \leq \mathrm{A}, for\ all\ 1 \leq i \leq n, \qquad 0 \leq t \leq T$$

Then for $x > \left(\frac{2\mathrm{A}}{T^2}\right)^{\frac{1}{(1-\alpha)}}$,

$$\boldsymbol{P}\left(\left|\sum_{i=1}^{n} W_i\right| \geq x\right) \leq 2 * \exp\left\{-\frac{x}{2\left(n\sigma^2 + \frac{x^{2-\alpha}}{T}\right)}\right\} + \sum_{i=1}^{n} \boldsymbol{P}(|W_i| \geq x)$$

*Proof:* Let $W_i^* = W_i \cdot I_{(-\infty,x]}(W_i).\ Then$

$$\boldsymbol{P}\left(\sum_{i=1}^{n} W_i \geq x\right) \leq \boldsymbol{P}\left(\sum_{i=1}^{n} W_i^* \geq x\right) + \sum_{i=1}^{n} \boldsymbol{P}(W_i \geq x).$$

For $W_i^* \geq 0, we\ have$

$$e^{tW_i^*} \leq 1 + tW_i^* + \frac{t^2}{2} W_i^{*2} + \sum_{k=3}^{\infty} \frac{t^k}{k!} |W_i|^{k\alpha + 3(1-\alpha)} x^{(k-3)(1-\alpha)}$$

Note that (7.6) is true also for $W_i^* < 0$ because of the montonicity of function $f(u) = e^u - 1 - u - \frac{u^2}{2}$

It is easy to get $E|W_i|^{k\alpha+3(1-\alpha)} \leq \frac{k!\mathrm{A}}{T^k}$ from (7.3). Moreover, we have $E(W_i^*) \leq 0,\ Var(W_i^*) \leq \sigma^2$ from definition. Taking the expectation of (7.6)

$$E\left(e^{tW_i^*}\right) \leq 1 + \frac{t^2\sigma^2}{2} + \sum_{k=3}^{\infty} \frac{2\mathrm{A}}{T^2 x^{1-\alpha}} * \frac{1}{2} * \left(\frac{x^{1-\alpha}}{T}\right)^{k-2} t^k$$

$$\leq 1 + \frac{t^2\sigma^2}{2} + \frac{t^2}{2} \sum_{k=3}^{\infty} \left(\frac{tx^{1-\alpha}}{T}\right)^{k-2}$$

$$\leq 1 + \frac{t^2\sigma^2}{2\left(1 - \frac{t^{1-\alpha}}{T}\right)},$$

$$when\ \left|\frac{tx^{1-\alpha}}{T}\right| < 1$$

$$Let\ t = \frac{x}{n\sigma^2 + \frac{x^{2-\alpha}}{T}}.\ By\ the\ Markov\ inequality$$

$$P\left(\sum_{i=1}^{n} W_i^* \geq x\right) \leq e^{-tx} E\left(e^{t\sum_{i=1}^{n} W_i^*}\right)$$

$$\leq e^{-tx} \prod_{i=1}^{n} E\left(e^{tW_i^*}\right)$$

$$\leq e^{-tx} \left(1 + \frac{t^2\sigma^2}{2\left(1 - \frac{tx^{1-\alpha}}{T}\right)}\right)^n$$

$$\le \exp\left\{\frac{x^2}{n\sigma^2 + \frac{x^{2-\alpha}}{T}}\right\}\left(1 + \frac{x^2}{2n\left(n\sigma^2 + \frac{x^{2-\alpha}}{T}\right)}\right)^n$$

$$\le \exp\left\{\frac{x^2}{2\left(n\sigma^2 + \frac{x^{2-\alpha}}{T}\right)}\right\}$$

Therefore,

$$\boldsymbol{P}\left(\sum_{i=1}^{n} W_i \ge x\right) \le \boldsymbol{P}\left(\sum_{i=1}^{n} W_i^* \ge x\right) + \sum_{i=1}^{n} \boldsymbol{P}(W_i \ge x).$$

**Lemma 6**

*Under condition (C1) and (C2),* $for\ m = o\left(n^{\frac{1}{3}-\frac{1}{3}\xi}\right), \boldsymbol{\mathcal{M}} \subset \mathcal{P}_1,$

$$\boldsymbol{P}\left(\tau_{min} \le \min_{|\mathcal{M} \le m|}(\lambda_{min}(\widehat{\Sigma_{\mathcal{M}}})) \le \max_{|\mathcal{M} \le m|} \lambda(\widehat{\Sigma_{\mathcal{M}}}) \le \tau_{max}\right) \to 1.$$

Furthermore, under condition (C4), (7.8) holds for $m = O\left(n^{2\xi_0 + 4\xi_{min}}\right) = o\left(n^{\frac{1}{3}-\frac{1}{3}\xi}\right)$

**Lemma 7**

Let $W_1, \dots, W_n$ be independent random variables with zero mean and such that $E\left(e^{T_0|W_i^\alpha|}\right) \le \mathrm{A}_0\ for\ constants\ T_0 > 0,\ \mathrm{A}_0 > 0\ and\ 0 < \alpha < 1.$ Then for $\alpha\ sequence\ \alpha_n \to \infty\ with\ \alpha_n = o\left(n^{\frac{\sigma}{2(2-\alpha)}}\right), there\ exists\ constants\ c_1, c_2\ \ such\ that$

$$P\left(|W_1 + \cdots + W_n| \le \sqrt{n\alpha_n}\right) \le c_1 \exp(-c_2\alpha_n^2)$$

*Proof:*

The condition $E\left(e^{T_0|W_i^\alpha|}\right) \le \mathrm{A}_0$ implies $Var(W_i) \le \sigma^2$, $E\left(|W_i|^2 e^{T|W_i^\alpha|}\right) \le \mathrm{A}\ and\ E\left(|W_i|^{3(1-\alpha)} e^{T|W_i|^\alpha}\right) \le$ $\mathrm{A}$ for some constant $\sigma^2, T, and\ \mathrm{A}.\ By\ Lemma\ 5, we\ have$

$$\boldsymbol{P}\left(\left|\sum_{i=1}^{n} W_i\right| \geq x\right) \leq 2\exp\left\{-\frac{x^2}{2\left(n\sigma^2 + \frac{x^{2-\alpha}}{T}\right)}\right\} + \sum_{i=1}^{n} \boldsymbol{P}(|W_i| \geq x)$$

Let $x = \sqrt{n}\alpha_n,\ Then$

$$\exp\left\{-\frac{x^2}{2\left(n\sigma^2 + \frac{x^{2-\alpha}}{T}\right)}\right\} = \exp\left\{-\frac{n\alpha_n^2}{2\left(n\sigma^2 + \frac{n^{\frac{2-\alpha}{2}}\alpha_n^{2-\alpha}}{T}\right)}\right\} = \exp\left\{-\frac{\alpha_n^2}{2\sigma^2 + o(1)}\right\}$$

On the other hand, by the Markov Inequality

$$\boldsymbol{P}(|W_i| \geq x) = \boldsymbol{P}\left(W_i^2 e^{T|W_i|^\alpha} \leq x^2 e^{Tx^\alpha}\right) \leq \mathrm{A}x^{-2}\exp(-Tx^2) \leq \frac{\mathrm{A}}{n\alpha_n^2}\exp\left(-\frac{T\alpha_n^2}{o(1)}\right)$$

Hence, $\sum_{i=1}^{n} P(|W_i| \geq x) \ \leq \frac{\mathrm{A}}{n\alpha_n^2}\exp\left(-\frac{T\alpha_n^2}{o(1)}\right)$. And (7.9) is easily obtained.

Remark 1. We are interested in the case that $W_i = X_{ij}X_{ik}X_{il}, where\ X_{ij}, X_{ik}, X_{il}$ are joint normal and marginally standard normal. It is easy to see that $W_i$ satisfies $E\left(e^{\frac{1}{4}|W_i|^{\frac{2}{3}}}\right) \leq \sqrt{2}\ and\ Var(W_i) \leq 30$. Therefore, (7.9) holds for $c_1 = 3, c_2 = \frac{1}{61}\ when\ n$ is sufficiently large.

In order to show Theorem 2, we have to obtain an analogue of Lemma 6 for arbitrary submodel $\mathcal{M}$. We start from a generalization of Lemma A3 in Bickel & Levina (2008)

**Lemma 8**

Let $W_1, \dots, W_n$ be independent random variables with zero mean and such that $E\left(e^{T_0|W_i^\alpha|}\right) \le \mathrm{A}_0$ $for\ constants\ T_0 > 0,\ \mathrm{A}_0 > 0\ and\ 0 < \alpha < 1$. Then there exists constants $c_3, c_4, for\ 0 < \epsilon \le 1$

$$P(|W_1 + \cdots + W_n| \ge n\epsilon) \le c_3 \exp(-c_4 n^\alpha \epsilon^2)$$

*Proof:*

The condition $E\left(e^{T_0|W_i^\alpha|}\right) \le \mathrm{A}_0$ implies $Var(W_i) \le \sigma^2, E\left(|W_i|^2 e^{T|W_i^\alpha|}\right) \le \mathrm{A}\ and\ E\left(|W_i|^{3(1-\alpha)} e^{T|W_i^\alpha|}\right) \le \mathrm{A}\ for\ some\ constants\ \sigma^2, T\ and\ \mathrm{A}.\ When\ \alpha < 1, by\ Lemma\ 5,$

$$\boldsymbol{P}\left(\left|\sum_{i=1}^n W_i\right| \ge x\right) \le 2 * \exp\left\{-\frac{x}{2\left(n\sigma^2 + \frac{x^{2-\alpha}}{T}\right)}\right\} + \sum_{i=1}^n \boldsymbol{P}(|W_i| \ge x)$$

$Let\ x = n\epsilon.\ Then$

$$\exp\left\{-\frac{x^2}{2\left(n\sigma^2 + \frac{x^{2-\alpha}}{T}\right)}\right\} = \exp\left\{-\frac{n^2\epsilon^2}{2\left(n\sigma^2 + \frac{n^{2-\alpha}\epsilon^{2-\alpha}}{T}\right)}\right\}$$
$$= \exp\left\{-\frac{n^\alpha\epsilon^2}{2n^{\alpha-1}\sigma^2 + \frac{2\epsilon^{2-\alpha}}{T}}\right\}$$
$$\le \exp\left\{-\frac{n^\alpha\epsilon^2}{o(1) + \frac{2}{T}}\right\}$$

On the other hand, by the Markov inequality

$$\boldsymbol{P}(|W_i| \ge x) = \boldsymbol{P}\left(W_i^2 e^{T|W_i|^\alpha} + x^2 e^{Tx^\alpha}\right) \le \mathrm{A}x^{-2}\exp\{-Tx^\alpha\} \le \frac{\mathrm{A}}{n^2\epsilon^2}\exp\{-Tn^\alpha\epsilon^\alpha\}.$$

Hence, $\sum_{i=1}^n \boldsymbol{P}(|W_i| \ge x) \le \frac{\mathrm{A}}{n\epsilon^2}\exp\left\{-\frac{1}{2}Tn^\alpha\epsilon^\alpha\right\}\exp\left\{-\frac{1}{2}Tn^\alpha\epsilon^\alpha\right\} \le o(1)\exp\left\{-\frac{1}{2}Tn^\alpha\epsilon^\alpha\right\}$. And (7.10) is easily obtained.

When $\alpha = 1, E\left(e^{T_0|W_i^\alpha|}\right) \le \mathrm{A}_0$ implies $\sum_{k=0}^\infty \frac{1}{k!} T_0^k E\left(|W_i|^k\right) \le \mathrm{A}_0$. So $E\left(|W_i|^k\right) \le \frac{1}{2}k!\left(\frac{1}{T_0}\right)^{k-2}\frac{2\mathrm{A}_0}{T_0^2}\ for\ k\ \ge 2$. By Bernstein's Inequality, Lemma 2.2.11 in van der Vaart Wellner (1996), we have

$$\boldsymbol{P}\left(\left|\sum_{i=1}^n W_i\right| \ge n\epsilon\right) \le 2\exp\left\{-\frac{n^2\epsilon^2}{2\left(\frac{2n\mathrm{A}_0}{T_0^2} + \frac{n\epsilon}{T_0}\right)}\right\} \le 2\exp\left\{\frac{n\epsilon^2}{\frac{4A_0}{T_0^2} + \frac{2}{T_0}}\right\}$$

**Lemma 9**

Under condition (C1) and (C2), for $0 < \epsilon < 1, we\ have$

$$P\left(\left|\sum_{i=1}^{n} X_{si}X_{sj} - \sigma_{ij}\right| \geq n\epsilon\right) \leq C_1 \exp(-C_2 n\epsilon^2)$$

$$P\left(\left|\sum_{i=1}^{n} X_{si}X_{sj}X_{sk} - 0\right| \geq n\epsilon\right) \leq C_3 \exp\left(-C_4 n^{\frac{2}{3}}\epsilon^2\right)$$

$$P\left(\left|\sum_{i=1}^{n} X_{si}X_{sj}X_{sk}X_{sl} - \sigma_{ij}\sigma_{kl} - \sigma_{ik}\sigma_{jl} - \sigma_{il}\sigma_{jk}\right| \geq n\epsilon\right) \leq C_5 \exp\left(-C_6 n^{\frac{1}{2}}\epsilon^2\right)$$

Where $C_1, \ldots, C_6\ are\ constants.$

*Proof:*

We show the last inequality here. The first two are similar.

$$Let\ W_s = X_{si}X_{sj}X_{sk}X_{sl} - \sigma_{ij}\sigma_{kl} - \sigma_{ik}\sigma_{jl} - \sigma_{il}\sigma_{jk})$$

$$E\left(e^{\frac{1}{4}|W_s|^{\frac{1}{2}}}\right) = E\left(e^{\frac{1}{4}|X_{si}X_{sj}X_{sk}X_{sl} - \sigma_{ij}\sigma_{kl} - \sigma_{ik}\sigma_{jl} - \sigma_{il}\sigma_{jk}|^{\frac{1}{2}}}\right)$$

$$(\because (a+b)^{\frac{1}{2}} \leq a^{\frac{1}{2}} + b^{\frac{1}{2}}) \leq E\left(e^{\frac{1}{4}|X_{si}X_{sj}X_{sk}X_{sl} - \sigma_{ij}\sigma_{kl} - \sigma_{ik}\sigma_{jl} - \sigma_{il}\sigma_{jk}|^{\frac{1}{2}}}\right)$$

$$(\because |\sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk} \leq 3) \leq e^{\frac{\sqrt{3}}{4}} E\left(e^{\frac{1}{4}|X_{si}X_{sj}X_{sk}X_{sl}|^{\frac{1}{2}}}\right)$$

$$(\because |abcd \leq \frac{a^2+b^2+c^2+d^2}{4}) \leq e^{\frac{\sqrt{3}}{4}} E\left(e^{\frac{1}{4}\frac{X_{si}^2+X_{sj}^2+X_{sk}^2+X_{sl}^2}{4}}\right)$$

$$(again\ abcd \leq \frac{a^2+b^2+c^2+d^2}{4}) \leq e^{\frac{\sqrt{3}}{4}} E\left(\left[e^{\frac{X_{si}^2}{4}} + e^{\frac{X_{sj}^2}{4}} + e^{\frac{X_{sk}^2}{4}} + e^{\frac{X_{sl}^2}{4}}\right]/4\right)$$

$$= \sqrt{2}e^{\frac{\sqrt{3}}{4}}$$

The inequality follows directly from the last lemma.