High-dimensional Wilks Phenomenon under Sparsity

Wen-Xin Zhou
Princeton University

The 4th IMS-APRM, The Chinese University of Hong Kong
Outline

1. Statistical Motivation
2. Maximum Spurious Correlation (MSC) & Goodness Of Spurious Fit (GOSF)
3. Asymptotic Theory of MSC
4. Asymptotic Theory of GOSF
5. Bootstrap Approximation
Statistical Motivation

Spurious Discoveries
Over last two decades, many exciting data mining, statistical machine learning techniques have been developed. For example, LASSO (Tibshirani, 96) and SCAD (Fan & Li, 01) are introduced to associate a small set of covariates to a response from a large pool.

Spurious discoveries can easily arise in high-dimensional data analysis due to enormous possibilities of such selections. How can we know statistically our discoveries better than those by chance?

Can our discoveries be spurious?
False Discoveries

★ Over last two decades, many exciting data mining, statistical machine learning techniques have been developed.

Example: LASSO (Tibshirani, 96) and SCAD (Fan & Li, 01) are introduced to associate a small set of covariates to a response from a large pool.

★ Spurious discoveries can easily arise in high-dimensional data analysis due to enormous possibilities of such selections. How can we know statistically our discoveries better than those by chance?

Can our discoveries be spurious?
Over last two decades, many exciting data mining, statistical machine learning techniques have been developed.

**Example**: LASSO (Tibshirani, 96) and SCAD (Fan & Li, 01) are introduced to associate a small set of covariates to a response from a large pool.

Spurious discoveries can easily arise in high-dimensional data analysis due to enormous possibilities of such selections. How can we know statistically our discoveries better than those by chance?

**Can our discoveries be spurious?**
Big Dimension Big Spurious Correlation

★ \( Y, X \sim N(0, 1) \), \( \text{corr}(Y, X) = 0.5 \).
\[
\hat{\text{corr}}_n(Y, X) = 0.5378 \text{ based on 50 i.i.d. observations.}
\]

**Conclusion**: \( Y \) and \( X \) are reasonably correlated.

★ \( Y, X_1, \ldots, X_{1000} \sim \text{i.i.d. } N(0, 1) \). Based on 50 observations,
\[
|\hat{\text{corr}}_n(Y, X_{542})| = \max_{1 \leq j \leq 1000} |\hat{\text{corr}}_n(Y, X_j)| = 0.5046.
\]

**Conclusion**: \( Y \) and \( X_{542} \) are reasonably correlated?

Need methods to guard against spurious discoveries.

A new area “selective inference” arises *(Taylor & Tibshirani, 15)*.
Big Dimension Big Spurious Correlation

★ \( Y, X \sim N(0, 1), \text{corr}(Y, X) = 0.5 \). 
\[ \hat{\text{corr}}_n(Y, X) = 0.5378 \] based on 50 i.i.d. observations.

**Conclusion**: \( Y \) and \( X \) are reasonably correlated.

★ \( Y, X_1, \ldots, X_{1000} \sim i.i.d. \ N(0, 1) \). Based on 50 observations,

\[
|\hat{\text{corr}}_n(Y, X_{542})| = \max_{1 \leq j \leq 1000} |\hat{\text{corr}}_n(Y, X_j)| = 0.5046.
\]

**Conclusion**: \( Y \) and \( X_{542} \) are reasonably correlated?

Need methods to guard against spurious discoveries.

A new area “selective inference” arises (*Taylor & Tibshirani, 15*).
Big Dimension Big Spurious Correlation

\[ Y, X \sim N(0, 1), \text{corr}(Y, X) = 0.5. \]
\[ \hat{\text{corr}}_n(Y, X) = 0.5378 \text{ based on 50 i.i.d. observations.} \]

**Conclusion**: \( Y \) and \( X \) are reasonably correlated.

\[ Y, X_1, \ldots, X_{1000} \sim_{i.i.d.} N(0, 1). \text{ Based on 50 observations,} \]
\[ |\hat{\text{corr}}_n(Y, X_{542})| = \max_{1 \leq j \leq 1000} |\hat{\text{corr}}_n(Y, X_j)| = 0.5046. \]

**Conclusion**: \( Y \) and \( X_{542} \) are reasonably correlated?

- Need methods to guard against spurious discoveries.
- A new area “selective inference” arises (Taylor & Tibshirani, 15).
Big Dimension Big Spurious Correlation

★ $Y, X \sim N(0,1)$, $\text{corr}(Y, X) = 0.5$.

$\hat{\text{corr}}_n(Y, X) = 0.5378$ based on 50 i.i.d. observations.

**Conclusion**: $Y$ and $X$ are reasonably correlated.

★ $Y, X_1, \ldots, X_{1000} \sim \text{i.i.d. } N(0,1)$. Based on 50 observations,

$$|\hat{\text{corr}}_n(Y, X_{542})| = \max_{1 \leq j \leq 1000} |\hat{\text{corr}}_n(Y, X_j)| = 0.5046.$$ 

**Conclusion**: $Y$ and $X_{542}$ are reasonably correlated?

Need methods to guard against spurious discoveries.

A new area “selective inference” arises (*Taylor & Tibshirani, 15*).
Maximum Spurious Correlation
**Max. spurious corr.:** When $Y$ and $X$ are independent,

$$\hat{R}_n(s, p) = \max_{\beta \in \mathbb{R}^p : \|\beta\|_0 \leq s} \text{corr}_n(Y, \beta^T X)$$

represents the maximum spurious correlation.

**Question 1:** What is the distribution of $\hat{R}_n(s, p)$?
Multiple spurious corr.: $\hat{R}_n(s, p) = \max_{\|\beta\|_0 = s} \text{corr}_n(\beta^T X, \epsilon)$.

Example: $n = 50, p = 1000, s = 1, 2, 5, 10$

increment = order stat. of $\chi_1^2 / n$.  (Fan, Shao and Z., 15)
**An Alternative View**

**$R^2$ statistic**: For each $S \subseteq [p]$ with $|S| = s$, its $R^2$ statistic is

$$R^2_S = \max_{\theta \in \mathbb{R}^s} \text{corr}_n^2(Y, X_S^T \theta).$$

**Difficulty**: When the response is not a quantitative value, we need a new way to assess correlation.

**Measure of goodness of fit**: Likelihood ratio (LR) statistic.

**Generalization of $R^2$**: $- \log(1 - R^2_S) = \frac{2}{n} \{ \ell(\hat{\beta}_S) - \ell(0) \}$,

$$\ell(\hat{\beta}_S) = \log L(\hat{\beta}_S) \text{ and } \ell(0) = \log L(0)$$

are log-MLR of the fitted and null model.
**$R^2$ statistic**: For each $S \subseteq [p]$ with $|S| = s$, its $R^2$ statistic is

$$R^2_S = \max_{\theta \in \mathbb{R}^s} \text{corr}_n^2(Y, X_T^S \theta).$$

**Difficulty**: When the response is not a quantitative value, we need a new way to assess correlation.

**Measure of goodness of fit**: Likelihood ratio (LR) statistic.

**Generalization of $R^2$**: $-\log(1 - R^2_S) = \frac{2}{n} \{\ell(\hat{\beta}_S) - \ell(0)\},$

where $\ell(\hat{\beta}_S) = \log L(\hat{\beta}_S)$ and $\ell(0) = \log L(0)$

are log-MLR of the fitted and null model.
**An Alternative View**

**$R^2$ statistic**: For each $S \subseteq [p]$ with $|S| = s$, its $R^2$ statistic is

$$R^2_S = \max_{\theta \in \mathbb{R}^s} \text{corr}_n^2 (Y, X^T_S \theta).$$

**Difficulty**: When the response is not a quantitative value, we need a new way to assess correlation.

**Measure of goodness of fit**: Likelihood ratio (LR) statistic.

**Generalization of $R^2$**: 

$$-\log (1 - R^2_S) = \frac{2}{n} \{ \ell(\hat{\beta}_S) - \ell(0) \},$$

where 

$$\ell(\hat{\beta}_S) = \log L(\hat{\beta}_S) \quad \text{and} \quad \ell(0) = \log L(0)$$

are log-MLR of the **fitted** and **null** model.
Goodness of Spurious Fit

Beyond Linear Models
**Goodness of spurious fit**

**Quasi-likelihood**: \( L_n(\beta) \) from the sample \( \{(Y_i, X_i^T)\}_{i=1}^n \).

**Best subset**: \( \hat{\beta}(s) = \arg\min_{\beta \in \mathbb{R}^p : \|\beta\|_0 \leq s} L_n(\beta) \).

**Goodness Of Spurious Fit (GOSF)**: If \( Y \perp \perp X \),

\[
\mathcal{LR}_n(s, p) = L_n(0) - L_n(\hat{\beta}(s)).
\]

**Question 2**: What is the distribution of \( \mathcal{LR}_n(s, p) \)?
Examples

**Generalized linear models**: $L_n(\beta) = \sum_{i=1}^{n} \{ b(X_i^T \beta) - Y_i \cdot X_i^T \beta \}$, where $b$ is a model-dependent convex function.

**$L_1$ regression**: $L_n(\beta) = \sum_{i=1}^{n} |Y_i - X_i^T \beta|$.

**Support vector machine**: $L_n(\beta) = \sum_{i=1}^{n} (Y_i - X_i^T \beta)_+.$

**AdaBoost**: $L_n(\beta) = \sum_{i=1}^{n} \exp(- Y_i \cdot X_i^T \beta)$.
Asymptotic Theory of MSC
**Sampling Schemes and Conditions**

**Spurious corr.**: \( \hat{R}_n(s, p) = \max_{\beta \in \mathbb{R}^p: \|\beta\|_0 = s} \text{corr}_n(\varepsilon, \beta^T X) \).

**Data**: i.i.d. sample from \( \mathbb{E}\varepsilon^2 = \sigma^2, \mathbb{E}XX^T = \Sigma \).

**Scale invariance**: WLOG, \( \sigma^2 = 1, \text{diag}(\Sigma) = I_p \) (correlation).

**Cond. 1**: \( X \) and \( \varepsilon \) are sub-Gaussian.

\[
K_0 := \|\varepsilon\|_{\psi_2} < \infty, \quad K_1 := \sup_{\alpha \in \mathbb{S}_{p-1}} \|\alpha^T \Sigma^{-1/2} X\|_{\psi_2} < \infty.
\]

★ Sub-Gaussian norm: \( \|X\|_{\psi_2} = \sup_{q \geq 1} q^{-1/2} (\mathbb{E}|X|^q)^{1/q} \).
Spurious corr.: $\hat{R}_n(s, p) = \max_{\beta \in \mathbb{R}^p : \|\beta\|_0 = s} \hat{\text{corr}}_n(\varepsilon, \beta^T X)$.

Data: i.i.d. sample from $\mathbb{E}\varepsilon^2 = \sigma^2$, $\mathbb{E}XX^T = \Sigma$.

Scale invariance: WLOG, $\sigma^2 = 1$, $\text{diag}(\Sigma) = I_p$ (correlation).

Cond. 1: $X$ and $\varepsilon$ are sub-Gaussian.

$$K_0 := \|\varepsilon\|_{\psi^2} < \infty, \quad K_1 := \sup_{\alpha \in \mathbb{S}^{p-1}} \|\alpha^T \Sigma^{-1/2} X\|_{\psi^2} < \infty.$$  

★ Sub-Gaussian norm: $\|X\|_{\psi^2} = \sup_{q \geq 1} q^{-1/2}(\mathbb{E}|X|^q)^{1/q}$.  

Wen-Xin Zhou (Princeton University)  High-dimensional Wilks Phenomenon under Sparsity
Regularity condition on $\Sigma$

**$s$-sparse min & max eigenvalues**: (Bickel, Ritov & Tsybakov, 09)

$$
\phi_{\min}(s) = \min_{u \in \mathbb{R}^p : 1 \leq \|u\|_0 \leq s} \frac{u^T \Sigma u}{u^T u}, \quad \phi_{\max}(s) = \max_{u \in \mathbb{R}^p : 1 \leq \|u\|_0 \leq s} \frac{u^T \Sigma u}{u^T u}.
$$

**$s$-sparse condition number**: $\gamma_s = \sqrt{\frac{\phi_{\max}(s)}{\phi_{\min}(s)}}$ well defined.

$s$ is allowed to diverge with $n$ and $p$. 
Asymptotics of $\hat{R}_n(s, p)$

**Notation:** $Z \sim N_p(0, \Sigma)$.

$\star \mathcal{F} = \{\alpha \in S^{p-1} : \|\alpha\|_0 = s\}$;

**Theorem:** (Berry-Esseen bound)

$$\sup_{t \geq 0} \left| \mathbb{P}\{\sqrt{n}\hat{R}_n(s, p) \leq t\} - \mathbb{P}\{R_0(s, p) \leq t\} \right| \lesssim \frac{s \{\log \frac{\gamma sp}{s} \vee \log n\}^{7/8}}{n^{1/8}},$$

$$R_0(s, p) = \sup_{\alpha \in \mathcal{F}} \frac{\alpha^T Z}{\sqrt{\alpha^T \Sigma \alpha}} = \max_{s \subset \{1, \ldots, p\}} \sqrt{\frac{Z^T \Sigma_{SS}^{-1} Z_S}{|S| = s}}.$$  

Gaussian approx. (Chernozhukov, Chetverikov & Kato, 14)
Asymptotics of $\hat{R}_n(s, p)$

**Notation:** $Z \sim N_p(0, \Sigma)$.

**Theorem:** (Berry-Esseen bound)

$$\sup_{t \geq 0} \left| \mathbb{P}\{\sqrt{n}\hat{R}_n(s, p) \leq t\} - \mathbb{P}\{R_0(s, p) \leq t\} \right| \lesssim \frac{s \{\log \frac{\gamma_s p}{s} \lor \log n\}^{7/8}}{n^{1/8}},$$

$$R_0(s, p) = \sup_{\alpha \in \mathcal{F}} \frac{\alpha^T Z}{\sqrt{\alpha^T \Sigma \alpha}} = \max_{S \subseteq \{1, \ldots, p\}, |S| = s} \sqrt{Z^T S \Sigma_S^{-1} Z_S}.$$

Gaussian approx. *(Chernozhukov, Chetverikov & Kato, 14)*
Isotropic case: $\Sigma = I_p$

**Corollary**: If $s \log p \ll n^{1/7}$, then

$$\sup_{t \geq 0} |\mathbb{P}\{n\hat{R}_n^2(s, p) \leq t\} - \mathbb{P}\{Z_{(1)}^2 + \cdots + Z_{(s)}^2 \leq t\}| \to 0.$$ 

$Z_{(1)}^2 \geq Z_{(2)}^2 \geq \cdots \geq Z_{(p)}^2$ are reverse-order stat of $\{Z_1^2, \ldots, Z_p^2\}$.

$\bowtie n\hat{R}_n^2(s, p) \xrightarrow{a} Z_{(1)}^2 + \cdots + Z_{(s)}^2 = sa_p + O_p(1), \ a_p = 2\log \frac{p}{\sqrt{\log p}}.$

$\bowtie n\hat{R}_n^2(1, p) - a_p \xrightarrow{d} \exp\{-\exp(-t/2)/\sqrt{\pi}\}$ (Gumbel).

$\bowtie$ indep relaxed to $\mathbb{E}(\epsilon X) = 0, \mathbb{E}(\epsilon^2|X) = \sigma^2, \mathbb{E}(\epsilon^4|X) < C.$
Isotropic case: $\Sigma = I_p$

**Corollary**: If $s \log p \ll n^{1/7}$, then

$$\sup_{t \geq 0} \left| \mathbb{P}\{n\hat{R}_n^2(s, p) \leq t\} - \mathbb{P}\{Z_{(1)}^2 + \cdots + Z_{(s)}^2 \leq t\} \right| \to 0.$$ 

$Z_{(1)}^2 \geq Z_{(2)}^2 \geq \cdots \geq Z_{(p)}^2$ are reverse-order stat of $\{Z_1^2, \ldots, Z_p^2\}$.

$\star$ $n\hat{R}_n^2(s, p) \overset{d}{\sim} Z_{(1)}^2 + \cdots + Z_{(s)}^2 = sa_p + O_p(1)$, $a_p = 2\log \frac{p}{\sqrt{\log p}}$.

$\star$ $n\hat{R}_n^2(1, p) - a_p \overset{d}{\to} \exp\{-\exp(-t/2)/\sqrt{\pi}\}$ (Gumbel).

$\star$ indep relaxed to $\mathbb{E}(\varepsilon X) = 0$, $\mathbb{E}(\varepsilon^2 | X) = \sigma^2$, $\mathbb{E}(\varepsilon^4 | X) < C$. 

Wen-Xin Zhou (Princeton University)  High-dimensional Wilks Phenomenon under Sparsity
**Isotropic case:** $\Sigma = I_p$

**Corollary:** If $s \log p \ll n^{1/7}$, then

$$
\sup_{t \geq 0} |\mathbb{P}\{ n\hat{R}_n^2(s, p) \leq t \} - \mathbb{P}\{ Z_{(1)}^2 + \cdots + Z_{(s)}^2 \leq t \}| \to 0.
$$

$Z_{(1)}^2 \geq Z_{(2)}^2 \geq \cdots \geq Z_{(p)}^2$ are reverse-order stat of $\{ Z_1^2, \ldots, Z_p^2 \}$.

★ $n\hat{R}_n^2(s, p) \overset{a}{\sim} Z_{(1)}^2 + \cdots + Z_{(s)}^2 = sa_p + O_p(1)$, $a_p = 2 \log \frac{p}{\sqrt{\log p}}$.

★ $n\hat{R}_n^2(1, p) - a_p \overset{d}{\to} \exp\{-\exp(-t/2)/\sqrt{\pi}\}$ (Gumbel).

★ indep relaxed to $\mathbb{E}(\epsilon | X) = 0$, $\mathbb{E}(\epsilon^2 | X) = \sigma^2$, $\mathbb{E}(\epsilon^4 | X) < C$. 

Wen-Xin Zhou (Princeton University)
**Isotropic case:** $\Sigma = I_p$

**Corollary:** If $s \log p \ll n^{1/7}$, then

\[
\sup_{t \geq 0} |\mathbb{P}\{n\hat{R}_n^2(s, p) \leq t\} - \mathbb{P}\{Z_{(1)}^2 + \cdots + Z_{(s)}^2 \leq t\}| \to 0.
\]

$Z_{(1)}^2 \geq Z_{(2)}^2 \geq \cdots \geq Z_{(p)}^2$ are reverse-order stat of \{\(Z_1^2, \ldots, Z_p^2\}\).

\(\star\) $n\hat{R}_n^2(s, p) \overset{d}{\sim} Z_{(1)}^2 + \cdots + Z_{(s)}^2 = sa_p + O_\mathbb{P}(1)$, $a_p = 2 \log \frac{p}{\sqrt{\log p}}$.

\(\star\) $n\hat{R}_n^2(1, p) - a_p \overset{d}{\to} \exp\{-\exp(-t/2) / \sqrt{\pi}\}$ (**Gumbel**).

\(\star\) indep relaxed to $\mathbb{E}(\varepsilon X) = 0$, $\mathbb{E}(\varepsilon^2 | X) = \sigma^2$, $\mathbb{E}(\varepsilon^4 | X) < C$. 

Wen-Xin Zhou (Princeton University)  
High-dimensional Wilks Phenomenon under Sparsity
**Theorem**: If $\Sigma = I_p$ and $s^2 \log p \ll n^{1/7}$,

$$n\left(\hat{R}_1^2, \hat{R}_2^2 - \hat{R}_1^2, \ldots, \hat{R}_s^2 - \hat{R}_{s-1}^2\right)^T \xrightarrow{d} \left(Z_1^2, Z_2^2, \ldots, Z_s^2\right)^T.$$ 

$\hat{R}_k = \hat{R}_n(k, p)$.
Asymptotic Theory of GOSF
Model and data: i.i.d. sample from

\[ f(y|x; \beta^*) = \exp\left\{ yx^T \beta^* - b(x^T \beta^*) \right\} / \phi + c(y, \phi), \]

where \( \beta^* \) is reg. coef., \( \phi > 0 \) is dispersion. WLOG, \( \phi = 1 \).

Loss function: \( L_n(\beta) = \sum_{i=1}^{n} \{ b(X_i^T \beta) - Y_i \cdot X_i^T \beta \} \).

\( \ell_0 \)-constrained LR: \( \mathcal{LR}_n(s, p) = nb(0) - \min_{\|\beta\|_0 \leq s} L_n(\beta) \).
**Conditions**

**Cond. 1:** $X$ is sub-Gaussian, $\mathbb{E}XX^T = \Sigma$ with $\text{diag}(\Sigma) = I_p$:

$$A_0 := \sup_{\alpha \in \mathbb{S}^{p-1}} \| \alpha^T (\Sigma^{-1/2} X) \|_{\psi_2} < \infty.$$ 

**Cond. 2:** $s$-sparse cond. no. $\gamma_s = \sqrt{\frac{\lambda_{\text{max}}(s)}{\lambda_{\text{min}}(s)}}$ well-defined,

$$\lambda_{\text{max}}(s) = \max_{\|\alpha\|_0 \leq s} \frac{\alpha^T \Sigma \alpha}{\alpha^T \alpha}, \quad \lambda_{\text{min}}(s) = \min_{\|\alpha\|_0 \leq s} \frac{\alpha^T \Sigma \alpha}{\alpha^T \alpha}.$$

**Cond. 3:** $\mathbb{E} e^{u(Y-\mu_Y)/\sigma_Y} \leq e^{a_0 u^2/2}$, $u \in \mathbb{R}$ for some $a_0 > 0$, where $\mu_Y = \mathbb{E} Y$ and $\sigma_Y^2 = \text{var}(Y)$.

**Cond. 4:** $\min_{|u| \leq 1} b''(u) \geq a_1$ and $\max_{|u| \leq 1} \left| b'''(u) \right| \leq A_1$. 

Wen-Xin Zhou (Princeton University)  
High-dimensional Wilks Phenomenon under Sparsity
Asymptotics of $L \mathcal{R}_n(s, p)$

$$R_0(s, p) = \max_{S \subset \{1, \ldots, p\}, |S| = s} \| \Sigma_S^{-1/2} Z_S \|_2 \text{ with } Z \sim N_p(0, \Sigma).$$

**Theorem**: (Berry-Esseen bound) Under Conditions 1–4,

$$\sup_{t \geq 0} |P\{2 L \mathcal{R}_n(s, p) \leq t\} - P\{R_0^2(s, p) \leq t\}| \lesssim \frac{\{s \log(\gamma_s pn)\}^{7/8}}{n^{1/8}} + \frac{\gamma_s^{1/2} \{s \log(\gamma_s pn)\}^2}{n^{1/2}}.$$ 

**Wilks theorem**: When $s = p$, $R_0^2(s, p) \sim \chi^2(p)$. 

Wen-Xin Zhou (Princeton University)  High-dimensional Wilks Phenomenon under Sparsity
Asymptotics of $L \mathcal{R}_n(s, p)$

\[ R_0(s, p) = \max_{S \subset \{1, \ldots, p\} \atop |S|=s} \| \Sigma^{1/2} S S^T \|_2 \quad \text{with} \quad \mathbf{Z} \sim N_p(0, \Sigma). \]

**Theorem**: (Berry-Esseen bound) Under Conditions 1–4,

\[
\sup_{t \geq 0} \left| \mathbb{P}\{ 2L \mathcal{R}_n(s, p) \leq t \} - \mathbb{P}\{ R_0^2(s, p) \leq t \} \right| \\
\lesssim \frac{\{s \log(\gamma_s pn)\}^{7/8}}{n^{1/8}} + \frac{\gamma_s^{1/2} \{s \log(\gamma_s pn)\}^2}{n^{1/2}}.
\]

**Wilks theorem**: When $s = p$, $R_0^2(s, p) \sim \chi^2(p)$. 
**Model and data**: i.i.d. sample from \((Y, X^T)\), \(Y = X^T\beta^* + \epsilon\).

**Loss function**: 
\[ L_n(\beta) = \sum_{i=1}^{n} |Y_i - X_i^T\beta| \]

**\(\ell_0\)-constrained LR**: 
\[
\mathcal{L}R_n(s, p) = \sum_{i=1}^{n} |Y_i| - \min_{\beta \in \mathbb{R}^p : \|\beta\|_0 \leq s} L_n(\beta).
\]
**Cond 5:** $\mathbb{E}|\varepsilon|^\kappa < \infty$ for $\kappa > 1$. \exists \ a_2 < (\mathbb{E}|\varepsilon|)^{-1}, A_2, A_3$

\[
2 \max\{1 - F_\varepsilon(u), F_\varepsilon(-u)\} \leq (1 + a_2 u)^{-1}, \quad \text{for all } u \geq 0,
\]

\[
\max_u f_\varepsilon(u) \leq A_2, \quad \max \max_{|u| \leq 1} \{|f'_\varepsilon(u+)\}, |f'_\varepsilon(u-)\} \leq A_3,
\]

**Theorem:** Under Conditions 1, 2 & 5,

\[
\sup_{t \geq 0} \left| \mathbb{P}\left\{2 \mathcal{LR}_n(s, p) \leq t\right\} - \mathbb{P}\left\{R_0^2(s, p) \leq 2f_\varepsilon(0)t\right\} \right|
\]

\[
\lesssim n^{1-\kappa} + \frac{s \log(\gamma_s pn)}{n^{1/8}} + \frac{\gamma_s^{1/4} \{s \log(\gamma_s pn)\}^{3/2}}{n^{1/4}}.
\]
Bootstrap Approximation
Multiplier bootstrap approximation

**Challenges**: Dist of $R_0(s, p) = \sup_{\alpha \in \mathcal{F}} \frac{\alpha^T Z}{\sqrt{\alpha^T \Sigma \alpha}}$ is not analytic and depends on unknown $Z \sim N_p(0, \Sigma)$.

**Bootstrap approximation**: 
- Generate $\hat{Z} \sim N_p(0, \hat{\Sigma})$ via $\hat{Z} = n^{-1/2} \sum_{i=1}^{n} e_i(X_i - \bar{X}_n)$, where $e_1, \ldots, e_n \sim i.i.d. N(0, 1)$.
- Compute the bootstrap counterpart of $R_0(s, p)$:

$$R_n(s, p) = \max_{s \subset \{1, \ldots, p\}, |S| = s} \| \hat{\Sigma}^{-1/2} \hat{Z}_S \|_2.$$

**Validity**: If $s \log(\gamma_{spn}) = o(n^{1/5})$,

$$\sup_{t \geq 0} |\mathbb{P}\{R_0(s, p) \leq t\} - \mathbb{P}\{R_n(s, p) \leq t | X_1, \ldots, X_n\}| = o_{\mathbb{P}}(1).$$
Multiplier bootstrap approximation

**Challenges**: Dist of $R_0(s, p) = \sup_{\alpha \in \mathcal{F}} \frac{\alpha^T Z}{\sqrt{\alpha^T \Sigma \alpha}}$ is not analytic and depends on unknown $Z \sim N_p(0, \Sigma)$.

**Bootstrap approximation**:
- Generate $\hat{Z} \sim N_p(0, \hat{\Sigma})$ via $\hat{Z} = n^{-1/2} \sum_{i=1}^{n} e_i(X_i - \bar{X}_n)$, where $e_1, \ldots, e_n \sim i.i.d. N(0, 1)$.
- Compute the bootstrap counterpart of $R_0(s, p)$:
  $$R_n(s, p) = \max_{S \subset \{1, \ldots, p\} \mid |S| = s} \| \hat{\Sigma}_{SS}^{-1/2} \hat{Z}_S \|_2.$$

**Validity**: If $s \log(\gamma s \rho n) = o(n^{1/5})$,
  $$\sup_{t \geq 0} \left| \mathbb{P}\{R_0(s, p) \leq t\} - \mathbb{P}\{R_n(s, p) \leq t \mid X_1, \ldots, X_n\} \right| = o_{\mathbb{P}}(1).$$
★ Develop non-asymptotic, asymptotic and joint asymptotic theories for maximum spurious correlations (MSCs)

★ Define a generalized goodness of spurious fit, which extends maximum spurious correlation in linear models

★ Develop the asymptotic distribution of such Goodness Of Spurious Fit (GOSF) for GLIM and $L_1$ regression

★ Use and validate the multiplier bootstrap method to approximate the distributions of MSC and GOSF
Guarding from Spurious Discoveries in High Dimension. Preprint.