Order Thresholding

Michael Akritas

Joint work with
Ph.D. Student Min Hee Kim

May 31, 2008
Needles in a Haystack

Adaptive Neyman Truncation

Hard Thresholding

Order Thresholding

Motivation for Order Thresholding

L-Statistics

Solution of Problem 2

Ideas for Addressing Problems 3 and 4
Testing Problem 1
Testing Problem 1

- Let $X_1, \ldots, X_a$ be independent with $X_i \sim N(\mu_i, 1)$. 
Testing Problem 1

- Let $X_1, \ldots, X_a$ be independent with $X_i \sim N(\mu_i, 1)$.
- Want to test $H_0 : \mu_1 = \cdots = \mu_a = 0$ vs $H_a : H_0$ is not true.
Testing Problem 1

- Let $X_1, \ldots, X_a$ be independent with $X_i \sim N(\mu_i, 1)$.
- Want to test $H_0 : \mu_1 = \cdots = \mu_a = 0$ vs $H_a : H_0$ is not true.
- If $a$ is large, the test that rejects if

$$\sum_{i=1}^{a} X_i^2 > \chi^2_a(\alpha)$$

cannot detect alternatives of the order $||\mu||^2 = o(\sqrt{a})$ (Fan, 1996).
Testing Problem 1

- Let $X_1, \ldots, X_a$ be independent with $X_i \sim N(\mu_i, 1)$.
- Want to test $H_0 : \mu_1 = \cdots = \mu_a = 0$ vs $H_a : H_0$ is not true.
- If $a$ is large, the test that rejects if
  \[
  \sum_{i=1}^{a} X_i^2 > \chi_a^2(\alpha)
  \]
  cannot detect alternatives of the order $||\mu||^2 = o(\sqrt{a})$ (Fan, 1996).
- Is it possible to improve the power?
Testing Problem 2

Let $X_{ij}, i = 1, \ldots, a, j = 1, \ldots, n$, be independent with $X_i \sim N(\mu_i, \sigma^2)$, $\sigma$ unknown.

Test $H_0: \mu_1 = \cdots = \mu_a$ vs $H_a: H_0$ is not true. ($H_0$ does not require that $\mu_i = 0$).

If $a$ is large, the test that rejects if

$$F = \frac{\sum_{i=1}^{n} (X_i - \hat{\mu})^2}{\hat{\sigma}^2} > F_{a-1, na-1}(\alpha)$$

(2.1)

cannot detect alternatives of the order $||\mu||^2 = o(\sqrt{a})$ (Akritas and Papadatos, 2004).

Michael Akritas

Joint work with Ph.D. Student Min Hee Kim

Order Thresholding
Testing Problem 2

Let $X_{ij}, i = 1, \ldots, a, j = 1, \ldots, n$, be independent with $X_i \sim N(\mu_i, \sigma^2)$, $\sigma$ unknown.
Testing Problem 2

- Let $X_{ij}, i = 1, \ldots, a, j = 1, \ldots, n$, be independent with $X_i \sim N(\mu_i, \sigma^2)$, $\sigma$ unknown.
- Test $H_0 : \mu_1 = \cdots = \mu_a$ vs $H_a : H_0$ is not true. ($H_0$ does not require that $\mu_i = 0$.)

Michael Akritas] Joint work with Ph.D. Student Min Hee Kim Order Thresholding
Testing Problem 2

- Let $X_{ij}, i = 1, \ldots, a, j = 1, \ldots, n$, be independent with $X_i \sim N(\mu_i, \sigma^2)$, $\sigma$ unknown.

- Test $H_0 : \mu_1 = \cdots = \mu_a$ vs $H_a : H_0$ is not true. ($H_0$ does not require that $\mu_i = 0$.)

- If $a$ is large, the test that rejects if

$$F = \frac{MST}{MSE} = \sum_{i=1}^{a} \frac{n(\bar{X}_i - \hat{\mu})^2}{\hat{\sigma}^2} > F_{a-1, na-1}(\alpha)$$

(2.1)

cannot detect alternatives of the order $||\mu||^2 = o(\sqrt{a})$

(Akritas and Papadatos, 2004).
Testing Problems 3 and 4
Testing Problems 3 and 4

- Testing Problem 3: Remove Normality Assumption

Let $X_{ij}, i = 1, \ldots, a, j = 1, \ldots, n$, be independent with $X_i$ having mean $\mu_i$ and variance $\sigma_i^2$, both unknown.
Testing Problems 3 and 4

- Testing Problem 3: Remove Normality Assumption

  - Let $X_{ij}$, $i = 1, \ldots, a$, $j = 1, \ldots, n$, be independent with $X_i$ having mean $\mu_i$ and variance $\sigma^2$), both unknown.
Testing Problems 3 and 4

- Testing Problem 3: Remove Normality Assumption
  - Let $X_{ij}$, $i = 1, \ldots, a$, $j = 1, \ldots, n$, be independent with $X_i$ having mean $\mu_i$ and variance $\sigma^2$, both unknown.

- Testing Problem 4: Unbalanced designs, heteroscedastic $X_i$'s.
The Adaptive Neyman Statistic

Motivated by Neyman (1937), Inglot et al. (1994), and Fan (1996) proposed the statistic
\[ T_{AN} = \max_{1 \leq m \leq n} \left\{ \left( \frac{2m}{m} \right)^{-1/2} \sum_{i=1}^{m} (X_i^2 - 1) \right\} \]
for Testing Problem 1.

Properly centered and standardized it converges in distribution to an extreme value distribution (exp\{−exp(−x)\}).

It has power 1 against alternatives
\[ \max_{1 \leq m \leq n} \left\{ \left( \frac{2m}{m} \right)^{-1/2} \sum_{i=1}^{m} \mu_i^2 \right\} - \sqrt{\log \log n} \rightarrow \infty. \]
The Adaptive Neyman Statistic

Motivated by Neyman (1937), Inglot et al. (1994), and Fan (1996) proposed the statistic

\[ T_{AN} = \max_{1 \leq m \leq n} \left\{ (2m)^{-1/2} \sum_{i=1}^{m} (X_i^2 - 1) \right\} \]

for Testing Problem 1.
The Adaptive Neyman Statistic

- Motivated by Neyman (1937), Inglot et al. (1994), and Fan (1996) proposed the statistic

\[ T_{AN} = \max_{1 \leq m \leq n} \left\{ (2m)^{-1/2} \sum_{i=1}^{m} (X_i^2 - 1) \right\} \]

for Testing Problem 1.

- Properly centered and standardized it converges in distribution to an extreme value distribution \((\exp\{-\exp(-x)\})\).
The Adaptive Neyman Statistic

- Motivated by Neyman (1937), Inglot et al. (1994), and Fan (1996) proposed the statistic

\[ T_{AN} = \max_{1 \leq m \leq n} \left\{ (2m)^{-1/2} \sum_{i=1}^{m} (X_i^2 - 1) \right\} \]

for Testing Problem 1.

- Properly centered and standardized it converges in distribution to an extreme value distribution \((\exp\{-\exp(-x)\})\).

- It has power 1 against alternatives

\[ \max_{1 \leq m \leq n} \left\{ (2m)^{-1/2} \sum_{i=1}^{m} \mu_i^2 \right\} - \sqrt{\log \log n} \to \infty. \]
Background/Applications

Introduced in the context of nonparametric function estimation using wavelets by Donoho and Johnstone (1994).

Johnstone and Silverman (2004) elaborate on the following additional applications of thresholding:

- Image processing,
- Model selection,
- Data mining.


Fan (1996) found that, for Testing Problem 1, hard thresholding outperforms soft thresholding and $T_{AN}$.

Beran (2004) considered a one-way ANOVA design, but from the estimation point of view.
Background/Applications

- Introduced in the context of nonparametric function estimation using wavelets by Donoho and Johnstone (1994).
Background/Applications

- Introduced in the context of nonparametric function estimation using wavelets by Donoho and Johnstone (1994).
- Johnstone and Silverman (2004) elaborate on the following additional applications of thresholding:
Background/Applications

- Introduced in the context of nonparametric function estimation using wavelets by Donoho and Johnstone (1994).
- Johnstone and Silverman (2004) elaborate on the following additional applications of thresholding:
  - Image processing, model selection,
Background/Applications

- Introduced in the context of nonparametric function estimation using wavelets by Donoho and Johnstone (1994).
- Johnstone and Silverman (2004) elaborate on the following additional applications of thresholding:
  - Image processing, model selection, data mining.
Background/Applications

- Introduced in the context of nonparametric function estimation using wavelets by Donoho and Johnstone (1994).
- Johnstone and Silverman (2004) elaborate on the following additional applications of thresholding:
  - Image processing, model selection, data mining.
Background/Applications

- Introduced in the context of nonparametric function estimation using wavelets by Donoho and Johnstone (1994).
- Johnstone and Silverman (2004) elaborate on the following additional applications of thresholding:
  - Image processing, model selection, data mining.
  - Fan (1996) found that, for Testing Problem 1, hard thresholding outperforms soft thresholding and $T_{AN}$. 

Michael Akritas Joint work with Ph.D. Student Min Hee Kim Order Thresholding
Background/Applications

- Introduced in the context of nonparametric function estimation using wavelets by Donoho and Johnstone (1994).
- Johnstone and Silverman (2004) elaborate on the following additional applications of thresholding:
  - Image processing, model selection, data mining.
  - Fan (1996) found that, for Testing Problem 1, hard thresholding outperforms soft thresholding and $T_{AN}$.
- Beran (2004) considered a one-way ANOVA design, but from the estimation point of view.
The Hard Thresholding (HT) Statistic

For Testing Problem 1, the HT statistic is

\[ T_{HT} = n \sum_{j=1}^{\infty} X_j^2 I(|X_j| > \delta) \]

\[ \delta = \sqrt{\frac{2 \log(\frac{\pi}{2})}{n \log(n)}} \]

With

\[ b = \sqrt{\frac{2}{\pi}} \log(n) \]

\[ T_{HT} \] is centered and scaled by

\[ \mu_{HT} = b \delta (1 + \delta^2) \]

\[ \sigma_{HT} = b \delta^3 (1 + 3 \delta^2) \]

\[ T_{HT} - \mu_{HT} / \sigma_{HT} \xrightarrow{D} N(0,1) \]
The Hard Thresholding (HT) Statistic

- For Testing Problem 1, the HT statistic is

\[ T_{HT} = \sum_{j=1}^{n} X_j^2 I(|X_j| > \delta), \quad \delta = \sqrt{2 \log(n \log^{-2} n)} \]
The Hard Thresholding (HT) Statistic

- For Testing Problem 1, the HT statistic is

\[ T_{HT} = \sum_{j=1}^{n} X_j^2 I(|X_j| > \delta), \quad \delta = \sqrt{2 \log(n \log^{-2} n)} \]

- With \( b = \sqrt{2/\pi} \log^{-2} n \), \( T_{HT} \) is centered and scaled by

\[ \mu_{HT} = b\delta(1 + \delta^{-2}), \quad \sigma_{HT} = b\delta^3(1 + 3\delta^{-2}). \]
The Hard Thresholding (HT) Statistic

For Testing Problem 1, the HT statistic is

\[ T_{HT} = \sum_{j=1}^{n} X_j^2 I(|X_j| > \delta), \quad \delta = \sqrt{2 \log(n \log^{-2} n)} \]

With \( b = \sqrt{2/\pi} \log^{-2} n \), \( T_{HT} \) is centered and scaled by

\[ \mu_{HT} = b\delta(1 + \delta^{-2}), \quad \sigma_{HT} = b\delta^3(1 + 3\delta^{-2}). \]

\[ \frac{T_{HT} - \mu_{HT}}{\sigma_{HT}} \overset{D}{\rightarrow} N(0, 1) \]
Why Invent a Different Thresholding Method?
Why Invent a Different Thresholding Method?

- The theory for $T_{HT}$ is not generally applicable:

  - Centering and scaling are specific to normality and to $\delta$.
  - The choice of $\delta$ is specific to normality.
  - Even under normality, different $\delta$-values give better power against different alternatives (Johnstone and Silverman, 2004).
  - Small departures in the value of the thresholding parameter $\delta$ have a significant effect on the level of the test: $\delta - 5\delta - 4\delta - 3\delta - 2\delta - 1\delta = 0.0203, 0.0285, 0.0361, 0.0431, 0.0474, 0.0504$ for $a = 50$, $a = 150$, $a = 500$. 

Michael Akritas Joint work with Ph.D. Student Min Hee Kim Order Thresholding
Why Invent a Different Thresholding Method?

- The theory for $T_{HT}$ is not generally applicable:
  - Centering and scaling are specific to normality and to $\delta$.
Why Invent a Different Thresholding Method?

- The theory for $T_{HT}$ is not generally applicable:
  - Centering and scaling are specific to normality and to $\delta$.
  - The choice of $\delta$ is specific to normality.

Michael Akritas
Joint work with Ph.D. Student Min Hee Kim
Order Thresholding
Why Invent a Different Thresholding Method?

- The theory for $T_{HT}$ is not generally applicable:
  - Centering and scaling are specific to normality and to $\delta$.
  - The choice of $\delta$ is specific to normality.
- Even under normality, different $\delta$-values give better power against different alternatives (Johnstone and Silverman, 2004).
Why Invent a Different Thresholding Method?

- The theory for $T_{HT}$ is not generally applicable:
  - Centering and scaling are specific to normality and to $\delta$.
  - The choice of $\delta$ is specific to normality.

- Even under normality, different $\delta$-values give better power against different alternatives (Johnstone and Silverman, 2004).

- Small departures in the value of the thresholding parameter have significant effect on the level of the test:

<table>
<thead>
<tr>
<th></th>
<th>$\delta - .5$</th>
<th>$\delta - .4$</th>
<th>$\delta - .3$</th>
<th>$\delta - .2$</th>
<th>$\delta - .1$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a = 50$</td>
<td>.0203</td>
<td>.0285</td>
<td>.0361</td>
<td>.0431</td>
<td>.0474</td>
<td>.0504</td>
</tr>
<tr>
<td>$a = 150$</td>
<td>.0271</td>
<td>.0315</td>
<td>.0365</td>
<td>.0426</td>
<td>.0462</td>
<td>.0519</td>
</tr>
<tr>
<td>$a = 500$</td>
<td>.0344</td>
<td>.0393</td>
<td>.0432</td>
<td>.0474</td>
<td>.0500</td>
<td>.0550</td>
</tr>
</tbody>
</table>
Order Thresholding: The Exponential Case

The idea is to look at the largest observations.

Let $V_1, \ldots, V_a$ be iid Exp(1) r.v.'s.

Let $V_1 < V_2 < \cdots < V_a$ be the order statistics.

We are interested in the asymptotic distribution of the $L$-statistic 

$$S_a = \frac{1}{a} \sum_{i=1}^{a} c_{ai} V_{i} = \frac{1}{a-k} \sum_{i=a-k+1}^{a} V_{i}$$


Michael Akritas Joint work with Ph.D. Student Min Hee Kim

Order Thresholding
Order Thresholding: The Exponential Case

- The idea is to look at the largest observations.
Order Thresholding: The Exponential Case

- The idea is to look at the largest observations.
- Let \( V_1, \ldots, V_a \) be iid \( \text{Exp}(1) \) r.v.’s.

\[ S_a = \frac{1}{a} \sum_{i=1}^{a} c_{ai} V_{ai} = \frac{1}{a} \sum_{i=a-k+1}^{a} V_{ai} \]

where \( c_{ai} = I(i > a - k) \).

Motivation for Order Thresholding

L-Statistics

Michael Akritas Joint work with Ph.D. Student Min Hee Kim Order Thresholding
Order Thresholding: The Exponential Case

- The idea is to look at the largest observations.
- Let $V_1, \ldots, V_a$ be iid Exp(1) r.v.’s.
- Let $V_{a1} < V_{a2} < \cdots < V_{aa}$ be the order statistics.
Order Thresholding: The Exponential Case

- The idea is to look at the largest observations.
- Let $V_1, \ldots, V_a$ be iid Exp(1) r.v.'s.
- Let $V_{a1} < V_{a2} < \cdots < V_{aa}$ be the order statistics.
- We are interested in the asymptotic distribution of the $L$-statistic

\[
S_a = \frac{1}{a} \sum_{i=1}^{a} c_{ai} V_{ai} = \frac{1}{a} \sum_{i=a-k_a+1}^{a} V_{ai}
\]

where $c_{ai} = I(i > a - k_a)$.
Chernoff, Gastwirth and Johns (1967)
Lemma

The $V_{ai}$, $1 \leq i \leq a$, may be represented in distribution as

$$V_{ai} \overset{D}{=} \frac{V_1}{a} + \frac{V_2}{a-1} + \cdots + \frac{V_i}{a-i+1} = \sum_{j=1}^{i} \frac{V_j}{a-j+1}.$$
Chernoff, Gastwirth and Johns (1967)

Lemma

The \( V_{ai} \), \( 1 \leq i \leq a \), may be represented in distribution as

\[
V_{ai} \overset{D}{=} \frac{V_1}{a} + \frac{V_2}{a-1} + \cdots + \frac{V_i}{a-i+1} = \sum_{j=1}^{i} \frac{V_j}{a-j+1}.
\]

In particular,

\[
\nu_{ai} = E(V_{ai}) = \sum_{j=1}^{i} \frac{1}{a-j+1}.
\]
Corollary

\[ S_a \overset{D}{=} \frac{1}{a} \sum_{i=a-k_a+1}^{a} \sum_{j=1}^{i} \frac{V_j}{a-j+1} = \frac{1}{a} \sum_{j=1}^{a} \alpha_{aj} V_j, \]

where

\[ \alpha_{aj} = \frac{j}{a-j+1} \sum_{i=j}^{a} c_{ai}, \quad \text{with} \quad c_{ai} = I(i > a - k_a). \]
Corollary

\[ S_a \overset{D}{=} \frac{1}{a} \sum_{i=a-k_a+1}^{a} \sum_{j=1}^{i} \frac{V_j}{a - j + 1} = \frac{1}{a} \sum_{j=1}^{a} \alpha_{aj} V_j, \]

where

\[ \alpha_{aj} = \frac{j}{a - j + 1} \sum_{i=j}^{a} c_{ai}, \quad \text{with} \quad c_{ai} = I(i > a - k_a). \]

In particular,

\[ \mu_a = E(S_a) = \frac{1}{a} \sum_{i=1}^{a} c_{ai} \nu_{ai} = \frac{1}{a} \sum_{j=1}^{a} \alpha_{aj}. \]
When is $S_a$ Asymptotically Normal?
When is $S_a$ Asymptotically Normal?

\[
\text{Var}(S_a) \approx \frac{k_a}{a^2} \left(1 - \frac{k_a}{a}\right) + \frac{k_a + 1}{a^2}
\]
When is $S_a$ Asymptotically Normal?

\[
\text{Var}(S_a) \approx \frac{k_a}{a^2} \left(1 - \frac{k_a}{a}\right) + \frac{k_a + 1}{a^2}
\]

\[
\max_j \text{Var}(\alpha_{aj} V_j) = \max \left\{ \frac{k_a^a}{a^2(a - j + 1)^2}, j = 1, \ldots, a - k_a + 1, \frac{1}{a^2} \right\} = \frac{1}{a^2}.
\]
When is $S_a$ Asymptotically Normal?

- $\text{Var}(S_a) \approx \frac{ka}{a^2}(1 - \frac{ka}{a}) + \frac{ka + 1}{a^2}$

- $\max_j \text{Var}(\alpha_{aj} V_j) = \max \left\{ \frac{ka}{a^2(a - j + 1)^2}, j = 1, \ldots, a - ka + 1, \frac{1}{a^2} \right\} = \frac{1}{a^2}$

- Thus,

$$\frac{\max_j \text{Var}(\alpha_{aj} V_j)}{\text{Var}(S_a)} \approx \frac{1}{ka}$$
Exponential Case: Centering/Scaling is Easy!
Exponential Case: Centering/Scaling is Easy!

Write

\[ S_a = \mu_a + Q_a, \quad \text{where} \quad Q_a = \frac{1}{a} \sum_{j=1}^{a} \alpha_j (V_j - 1) \]
Exponential Case: Centering/Scaling is Easy!

- Write

\[ S_a = \mu_a + Q_a, \quad \text{where} \quad Q_a = \frac{1}{a} \sum_{j=1}^{a} \alpha_{aj} (V_j - 1) \]

- The (exact) variance is

\[ \text{Var}(S_a) = \frac{1}{a^2} \sum_{j=1}^{a} \alpha_{aj}^2. \]
Order Thresholding: The General Case

Let $Y_1, \ldots, Y_a$ be iid $F$, and $Y_{a+1}, \ldots, Y_{aa}$ be the order statistics. Then, if $U_{a+1}, \ldots, U_{aa}$ are the uniform order statistics, $Y_{ai} \overset{D}{=} F^{-1}(U_{ai}) \overset{D}{=} F^{-1}(G(V_{ai}))$.

Thus, if $\tilde{H} = F^{-1} \circ G$, $c_{ai} = I(i > a-k)$,

$S_a = \sum_{i=a-k+1}^{a} Y_{ai} \overset{D}{=} \sum_{i=1}^{a} c_{ai} \tilde{H}(V_{ai})$. 

Michael Akritas, Joint work with Ph.D. Student Min Hee Kim
Order Thresholding: The General Case

- Let $Y_1, \ldots, Y_a$ be iid $F$, and $Y_{a1}, \ldots, Y_{aa}$ be the order statistics.
Order Thresholding: The General Case

- Let $Y_1, \ldots, Y_a$ be iid $F$, and $Y_{a1}, \ldots, Y_{aa}$ be the order statistics.
- Then, if $U_{a1}, \ldots, U_{aa}$ are the uniform order statistics,

$$Y_{ai} \overset{D}{=} F^{-1}(U_{ai}) \overset{D}{=} F^{-1}(G(V_{ai}))$$
Order Thresholding: The General Case

Let $Y_1, \ldots, Y_a$ be iid $F$, and $Y_{a1}, \ldots, Y_{aa}$ be the order statistics.

Then, if $U_{a1}, \ldots, U_{aa}$ are the uniform order statistics,

$$Y_{ai} \overset{D}{=} F^{-1}(U_{ai}) \overset{D}{=} F^{-1}(G(V_{ai}))$$

Thus, if $\tilde{H} = F^{-1} \circ G$, $c_{ai} = I(i > a - k_a)$,

$$S_a = \frac{1}{a} \sum_{i=a-k_a+1}^{a} Y_{ai} \overset{D}{=} \frac{1}{a} \sum_{i=1}^{a} c_{ai} \tilde{H}(V_{ai})$$
The General Case Continued
The General Case Continued

Expanding $\tilde{H}(V_{ai}) - \tilde{H}(\nu_{ai})$, it follows

$$S_a = \mu_a + Q_a + R_a,$$

where

$$\mu_a = \frac{1}{a} \sum_{i=1}^{a} c_{ai} \tilde{H}(\nu_{ai}),$$

$$Q_a = \frac{1}{a} \sum_{j=1}^{a} \alpha_{aj}(V_j - 1),$$

and

$$R_a = o_p(a^{-1/2}\sigma_a),$$

where
The General Case Continued

- Expanding $\tilde{H}(V_{ai}) - \tilde{H}(\nu_{ai})$, it follows

$$S_a = \mu_a + Q_a + R_a, \text{ where } \mu_a = \frac{1}{a} \sum_{i=1}^{a} c_{ai} \tilde{H}(\nu_{ai}),$$

$$Q_a = \frac{1}{a} \sum_{j=1}^{a} \alpha_{aj}(V_j - 1), \text{ and } R_a = o_p(a^{-1/2}\sigma_a), \text{ where}$$

$$\sigma_a^2 = \frac{1}{a^2} \sum_{i=1}^{a} \alpha_{ai}^2, \quad \alpha_{ai} = \frac{1}{a - i + 1} \sum_{j=i}^{a} c_{ai} \tilde{H}'(\nu_{ai}).$$
The General Case Continued

- Expanding $\tilde{H}(V_{ai}) - \tilde{H}(\nu_{ai})$, it follows

$$S_a = \mu_a + Q_a + R_a, \text{ where } \mu_a = \frac{1}{a} \sum_{i=1}^{a} c_{ai} \tilde{H}(\nu_{ai}),$$

$$Q_a = \frac{1}{a} \sum_{j=1}^{a} \alpha_{aj}(V_j - 1), \text{ and } R_a = o_p(a^{-1/2}\sigma_a), \text{ where}$$

$$\sigma_a^2 = \frac{1}{a^2} \sum_{i=1}^{a} \alpha_{ai}^2, \quad \alpha_{ai} = \frac{1}{a - i + 1} \sum_{j=i}^{a} c_{ai} \tilde{H}'(\nu_{ai}).$$

- **Solution to Problem 1**: Use as $F$ the d.f. of $\chi^2_1$, and check the conditions of C-G-J (1967).
Motivation for Order Thresholding

L-Statistics

Figure: Top panel: Histograms of $T_H(\delta)$ for $\delta = 3.927, 5.106, 5.672$ and $6.665$. Bottom panel: Histograms of $T_L(k)$ for $k = 10, 5, 3, \text{ and } 2$. 

Michael Akritas[J.5cm] Joint work with Ph.D. Student Min Hee Kim Order Thresholding
The one-way ANOVA statistic is (essentially) \( \sigma^2 \text{MSE} \), where \( \tilde{Z}_i = \sqrt{n} \left( X_i \cdot - X_{\cdot \cdot} \right) / \sigma \).

Add and subtract \( \mu_0 \), the true (common under \( H_0 \)) group mean, so that \( \tilde{Z}_i = Z_i + t \sqrt{n} / \sigma \), where \( t = -\sqrt{n} \sigma \left( X_{\cdot \cdot} - \mu_0 \right) / \sigma = O_p(1) \).

Consider \( t \) fixed so that the \( Z^2_t; i \sim \chi^2_1 \left( t^2 / n \right) \) are iid, where \( Z_t; i = Z_i + t \sqrt{n} \).
The one-way ANOVA statistic is (essentially)

\[
\frac{\sigma^2}{MSE} \sum_{i=1}^{a} (\tilde{Z}_i)^2, \text{ where } \tilde{Z}_i = \frac{\sqrt{n}(X_{i.} - \bar{X}_{..})}{\sigma}.
\]
The one-way ANOVA statistic is (essentially)
\[
\frac{\sigma^2}{\hat{MSE}} \sum_{i=1}^{a} (\tilde{Z}_i)^2, \text{ where } \tilde{Z}_i = \frac{\sqrt{n}(X_{i.} - \bar{X}_{..})}{\sigma}.
\]
Add and subtract $\mu_0$, the true (common under $H_0$) group mean, so that
\[
\tilde{Z}_i = Z_i + \frac{t}{\sqrt{a}}, \text{ where } t = -\frac{\sqrt{na}(\bar{X}_{..} - \mu_0)}{\sigma} = O_p(1).
\]
The one-way ANOVA statistic is (essentially)

\[
\frac{\sigma^2}{\text{MSE}} \sum_{i=1}^{a} (\tilde{Z}_i)^2, \quad \text{where} \quad \tilde{Z}_i = \frac{\sqrt{n}(X_{i.} - \overline{X}_{..})}{\sigma}.
\]

Add and subtract \( \mu_0 \), the true (common under \( H_0 \)) group mean, so that

\[
\tilde{Z}_i = Z_i + \frac{t}{\sqrt{a}}, \quad \text{where} \quad t = -\frac{\sqrt{na(\overline{X}_{..} - \mu_0)}}{\sigma} = O_p(1).
\]

Consider \( t \) fixed so that the \( Z_{t;i} \sim \chi^2_1(t^2/a) \) are iid, where

\[
Z_{t;i} = Z_i + \frac{t}{\sqrt{a}},
\]

and verify the C-G-J (1967) conditions.
Thus, if \( S_{ta} = a - 1 \sum_{i=1}^{c} aiZ_t; ai, T_{ta}(ka) \) t = \( S_{ta} - \mu_{ta}(ka) \sigma_{ta}(ka) \rightarrow N(0,1) \).

If \( F_{ta}; a \) is the d.f. of \( T_{ta}(ka) \), it can be shown that
\[
\sup_{-\infty < x < \infty} |F_{ta}(x) - \Phi(x)| \rightarrow 0, 
\] as \( a \rightarrow \infty \).

Provided \( ka/a \rightarrow 0 \), centering can be done as if \( \mu_0 \) were known:
\[
\mu_{ta}(ka) - \mu_0 a(ka) \sigma_0 a(ka) \rightarrow 0, 
\] as \( a \rightarrow \infty \).
Thus, if $S_a^t = a^{-1} \sum_{i=1}^a c_{ai} Z_{t,ai}$,

$$T_a(k_a)^t = \frac{S_a^t - \mu_a^t(k_a)}{\sigma_a^t(k_a)} \xrightarrow{D} N(0,1).$$
Solution to Problem 2 Continued

► Thus, if \( S_a^t = a^{-1} \sum_{i=1}^{a} c_{ai} Z_{t;ai} \),

\[
T_a(k_a)^t = \frac{S_a^t - \mu_a^t(k_a)}{\sigma_a^t(k_a)} \xrightarrow{D} N(0, 1).
\]

► If \( F_{t;a} \) is the d.f. of \( T_a(k_a)^t \), it can be shown that

\[
\sup_{-M \leq t \leq M, -\infty < x < \infty} |F_{t;a}(x) - \Phi(x)| \rightarrow 0, \text{ as } a \rightarrow \infty.
\]
Solution to Problem 2 Continued

Thus, if $S_a^t = a^{-1} \sum_{i=1}^{a} c_{ai} Z_{t;ai}$,

$$T_a(k_a)^t = \frac{S_a^t - \mu_a^t(k_a)}{\sigma_a^t(k_a)} \overset{D}{\rightarrow} \mathcal{N}(0, 1).$$

If $F_{t;a}$ is the d.f. of $T_a(k_a)^t$, it can be shown that

$$\sup_{-M \leq t \leq M, -\infty < x < \infty} |F_{t;a}(x) - \Phi(x)| \rightarrow 0, \text{ as } a \rightarrow \infty.$$  

Provided $k_a/a \rightarrow 0$, centering can be done as if $\mu_0$ were known:

$$\frac{\mu_a^t(k_a) - \mu_0^0(k_a)}{\sigma_0^0(k_a)} \rightarrow 0, \text{ as } a \rightarrow \infty.$$
Table: Percentiles and $\hat{\alpha}$ for $a = 100$ and $n = 5$ case

<table>
<thead>
<tr>
<th></th>
<th>$\log a$</th>
<th>$T^0_a$</th>
<th>$a^{3/4}$</th>
<th>$\log a$</th>
<th>$T^t_a$</th>
<th>$a^{3/4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A95</td>
<td>1.645</td>
<td>1.645</td>
<td>1.645</td>
<td>1.645</td>
<td>1.645</td>
<td>1.645</td>
</tr>
<tr>
<td>S95</td>
<td>1.837</td>
<td>1.773</td>
<td>1.749</td>
<td>1.763</td>
<td>1.705</td>
<td>1.637</td>
</tr>
<tr>
<td>A90</td>
<td>1.282</td>
<td>1.282</td>
<td>1.282</td>
<td>1.282</td>
<td>1.282</td>
<td>1.282</td>
</tr>
<tr>
<td>S90</td>
<td>1.351</td>
<td>1.338</td>
<td>1.318</td>
<td>1.302</td>
<td>1.282</td>
<td>1.251</td>
</tr>
<tr>
<td>$\hat{\alpha}(0.05)$</td>
<td>0.067</td>
<td>0.061</td>
<td>0.060</td>
<td>0.060</td>
<td>0.054</td>
<td>0.051</td>
</tr>
<tr>
<td>$\hat{\alpha}(0.1)$</td>
<td>0.108</td>
<td>0.112</td>
<td>0.104</td>
<td>0.101</td>
<td>0.100</td>
<td>0.095</td>
</tr>
</tbody>
</table>

Simulations based on 3,000 runs.
Removing the Normality Assumption
Removing the Normality Assumption

- Three transformations:
  - The probability transformation: $U = F(X) \sim \text{Uniform}(0,1)$.
  - The hazard transformation: $V = \Lambda(X) \sim \text{Exp}(1)$.
  - The Box-Cox transformation.
  - The empirical probability transformation: $\hat{F}(s) = \frac{1}{n} \sum_{i=1}^{n} 1_{X_i \leq s}$.
  - The empirical hazard transformation: $\hat{\Lambda}(t) = \int_0^t \frac{1}{\hat{F}(s)} \, ds$.
  - An alternative way to transform to (approximately) exponential r.v.'s is to apply the $-\log(1-u)$ transformation after the empirical hazard transformation.

Michael Akritas [Joint work with Ph.D. Student Min Hee Kim]
Removing the Normality Assumption

- Three transformations:
  - The probability transformation: \( U = F(X) \sim \text{Uniform}(0, 1) \).
Removing the Normality Assumption

- Three transformations:
  - The probability transformation: $U = F(X) \sim \text{Uniform}(0, 1)$.
  - The hazard transformation: $V = \Lambda(X) \sim \text{Exp}(1)$. 

Michael Akritas Joint work with Ph.D. Student Min Hee Kim
Removing the Normality Assumption

- Three transformations:
  - The probability transformation: \( U = F(X) \sim \text{Uniform}(0, 1) \).
  - The hazard transformation: \( V = \Lambda(X) \sim \text{Exp}(1) \).
  - The Box-Cox transformation.
Removing the Normality Assumption

- Three transformations:
  - The probability transformation: $U = F(X) \sim \text{Uniform}(0, 1)$.
  - The hazard transformation: $V = \Lambda(X) \sim \text{Exp}(1)$.
  - The Box-Cox transformation.

- The rank of an observation divided by the sample size is the *empirical probability transformation*.
Removing the Normality Assumption

- Three transformations:
  - The probability transformation: \( U = F(X) \sim \text{Uniform}(0, 1) \).
  - The hazard transformation: \( V = \Lambda(X) \sim \text{Exp}(1) \).
  - The Box-Cox transformation.

- The rank of an observation divided by the sample size is the empirical probability transformation.

- The empirical hazard transformation uses
  \[
  \hat{\Lambda}(t) = \int_0^t \frac{1}{1 - \hat{F}(s-)} d\hat{F}(s).
  \]
Removing the Normality Assumption

- Three transformations:
  - The probability transformation: \( U = F(X) \sim \text{Uniform}(0, 1) \).
  - The hazard transformation: \( V = \Lambda(X) \sim \text{Exp}(1) \).
  - The Box-Cox transformation.

- The rank of an observation divided by the sample size is the empirical probability transformation.

- The empirical hazard transformation uses
  \[
  \hat{\Lambda}(t) = \int_0^t \frac{1}{1 - \hat{F}(s)} d\hat{F}(s).
  \]

- An alternative way to transform to (approximately) exponential r.v.’s is to apply the \(- \log(1 - u)\) transformation after the empirical hazard transformation.
Unbalanced Designs, Heteroscedasticity
Unbalanced Designs, Heteroscedasticity

- In unbalanced designs the $\bar{X}_i$’s are not identically distributed even under $H_0$. 
Unbalanced Designs, Heteroscedasticity

- In unbalanced designs the $\overline{X}_i$ are not identically distributed even under $H_0$.
- In heteroscedastic designs, each $\overline{X}_i$ needs to be scaled differently.
Unbalanced Designs, Heteroscedasticity

- In unbalanced designs the $\bar{X}_i$ are not identically distributed even under $H_0$.
- In heteroscedastic designs, each $\bar{X}_i$ needs to be scaled differently.
- If the group sample sizes are also large, we can use double asymptotic arguments ($n \to \infty$ as $a \to \infty$) and rely on the approximate normality of the $\bar{X}_i$. 

Michael Akritas[.5cm] Joint work with Ph.D. Student Min Hee Kim Order Thresholding