Order Thresholding

Michael Akritas

Joint work with
Ph.D. Student Min Hee Kim

May 31, 2008

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

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

cannot detect alternatives of the order $||\boldsymbol{\mu}||^2 = o(\sqrt{a})$ (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 $||\boldsymbol{\mu}||^2 = o(\sqrt{a})$ (Fan, 1996).
- Is it possible to improve the power?

Testing Problem 2

Testing Problem 2

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

Testing Problem 2

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

Testing Problem 2

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

$$F = \frac{MST}{MSE} = \sum_{i=1}^{a} \frac{n(\overline{X}_{i\cdot} - \widehat{\mu})^2}{\widehat{\sigma}^2} > F_{a-1,na-1}(\alpha) \qquad (2.1)$$

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

Testing Problems 3 and 4

Testing Problems 3 and 4

- Testing Problem 3: Remove Normality Assumption

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 μ_i and variance σ^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 μ_i and variance σ^2), both unknown.
- Testing Problem 4: Unbalanced designs, heteroscedastic X_i's.

The Adaptive Neyman Statistic

The Adaptive Neyman Statistic

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

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

for Testing Problem 1.

The Adaptive Neyman Statistic

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

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

  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

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

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

$$\max_{1 \le m \le n} \left\{ (2m)^{-1/2} \sum_{i=1}^{m} \mu_i^2 \right\} - \sqrt{\log \log n} \to \infty.$$

Background/Applications

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.
- Fan (1996), Fan and Lin (1998) consider applications of thresholding methods to testing problems.

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), Fan and Lin (1998) consider applications of thresholding methods to testing problems.
  - Fan (1996) found that, for Testing Problem 1, hard thresholding outperforms soft thresholding and T_{AN}.

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), Fan and Lin (1998) consider applications of thresholding methods to testing problems.
  - 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

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

$$\mu_{HT} = b\delta(1+\delta^{-2}), \; \sigma_{HT} = b\delta^3(1+3\delta^{-2}).$$

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

$$\mu_{HT} = b\delta(1+\delta^{-2}), \; \sigma_{HT} = b\delta^3(1+3\delta^{-2}).$$

- $\dfrac{T_{HT} - \mu_{HT}}{\sigma_{HT}} \xrightarrow{D} N(0,1)$

Why Invent a Different Thresholding Method?

Why Invent a Different Thresholding Method?

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

Why Invent a Different Thresholding Method?

- The theory for T_{HT} is not generally applicable:
    - Centering and scaling are specific to normality and to δ.

Why Invent a Different Thresholding Method?

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

Why Invent a Different Thresholding Method?

- The theory for T_{HT} is not generally applicable:
  - Centering and scaling are specific to normality and to δ.
  - The choice of δ is specific to normality.
- Even under normality, different δ-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 δ.
  - The choice of δ is specific to normality.
- Even under normality, different δ-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:

	$\delta - .5$	$\delta - .4$	$\delta - .3$	$\delta - .2$	$\delta - .1$	δ
$a = 50$	.0203	.0285	.0361	.0431	.0474	.0504
$a = 150$	.0271	.0315	.0365	.0426	.0462	.0519
$a = 500$	.0344	.0393	.0432	.0474	.0500	.0550

Order Thresholding: The Exponential Case

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 Exp(1) r.v.'s.

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)

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

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

In particular,

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

Corollary

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

where

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

Corollary

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

where

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

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?

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

When is S_a Asymptotically Normal?

- $\mathrm{Var}(S_a) \approx \frac{k_a}{a^2}(1 - \frac{k_a}{a}) + \frac{k_a + 1}{a^2}$
- $\max_j \mathrm{Var}(\alpha_{aj} V_j) =$
  $\max \left\{ \frac{k_a^a}{a^2(a-j+1)^2}, j = 1, \ldots, a - k_a + 1, \frac{1}{a^2} \right\} = \frac{1}{a^2}.$

When is S_a Asymptotically Normal?

- $\mathrm{Var}(S_a) \approx \frac{k_a}{a^2}(1 - \frac{k_a}{a}) + \frac{k_a + 1}{a^2}$
- $\max_j \ \mathrm{Var}(\alpha_{aj} V_j) =$
  $\max \left\{ \frac{k_a^a}{a^2(a-j+1)^2}, j = 1, \ldots, a - k_a + 1, \frac{1}{a^2} \right\} = \frac{1}{a^2}.$
- Thus,

$$\frac{\max_j \ \mathrm{Var}(\alpha_{aj} V_j)}{\mathrm{Var}(S_a)} \approx \frac{1}{k_a}$$

Exponential Case: Centering/Scaling is Easy!

Exponential Case: Centering/Scaling is Easy!

- Write

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

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

$$\mathrm{Var}(S_a) = \frac{1}{a^2}\sum_{j=1}^{a} \alpha_{aj}^2.$$

Order Thresholding: The General Case

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

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

The General Case Continued

The General Case Continued

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

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

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

The General Case Continued

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

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

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

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

The General Case Continued

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

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

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

- $\sigma_a^2 = \dfrac{1}{a^2}\displaystyle\sum_{i=1}^{a}\alpha_{ai}^2$, $\alpha_{ai} = \dfrac{1}{a-i+1}\displaystyle\sum_{j=i}^{a} c_{ai}\widetilde{H}'(\nu_{ai})$.
- **Solution to Problem 1:** Use as F the d.f. of χ_1^2, and check the conditions of C-G-J (1967).

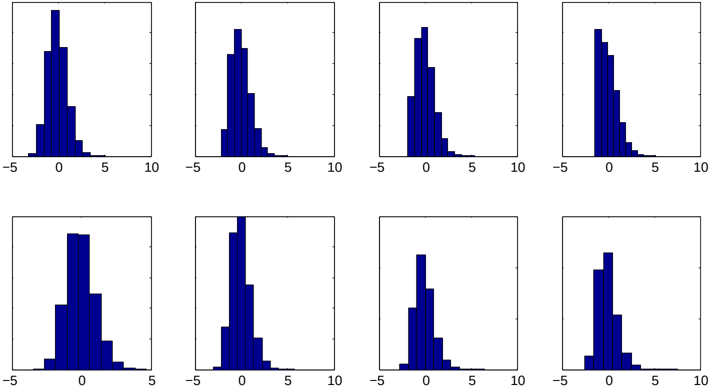

Figure: Top panel: Histograms of $T_H(\delta)$ for $\delta = 3.927$, 5.106, 5.672 and 6.665. Bottom panel: Histograms of $T_L(k)$ for $k = 10$, 5, 3, and 2.

- The one-way ANOVA statistic is (essentially)

$$\frac{\sigma^2}{MSE}\sum_{i=1}^{a}(\widetilde{Z}_i)^2, \text{ where } \widetilde{Z}_i = \frac{\sqrt{n}(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})}{\sigma}.$$

- The one-way ANOVA statistic is (essentially)

$$\frac{\sigma^2}{MSE}\sum_{i=1}^{a}(\widetilde{Z}_i)^2, \text{ where } \widetilde{Z}_i = \frac{\sqrt{n}(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})}{\sigma}.$$

- Add and subtract μ_0, the true (common under H_0) group mean, so that

$$\widetilde{Z}_i = Z_i + \frac{t}{\sqrt{a}}, \text{ where } t = -\frac{\sqrt{na}(\overline{X}_{\cdot\cdot} - \mu_0)}{\sigma} = O_p(1).$$

- The one-way ANOVA statistic is (essentially)

$$\frac{\sigma^2}{MSE}\sum_{i=1}^{a}(\widetilde{Z}_i)^2, \text{ where } \widetilde{Z}_i = \frac{\sqrt{n}(\overline{X}_{i\cdot} - \overline{X}_{\cdot\cdot})}{\sigma}.$$

- Add and subtract μ_0, the true (common under H_0) group mean, so that

$$\widetilde{Z}_i = Z_i + \frac{t}{\sqrt{a}}, \text{ where } t = -\frac{\sqrt{na}(\overline{X}_{\cdot\cdot} - \mu_0)}{\sigma} = O_p(1).$$

- Consider t fixed so that the $Z_{t;i}^2 \sim \chi_1^2(t^2/a)$ are iid, where $Z_{t;i} = Z_i + \frac{t}{\sqrt{a}}$, and verify the C-G-J (1967) conditions.

Solution to Problem 2 Continued

Solution to Problem 2 Continued

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

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

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

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

Solution to Problem 2 Continued

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

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

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

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

- Provided $k_a/a \to 0$, centering can be done as if μ_0 were known:

$$\frac{\mu_a^t(k_a) - \mu_a^0(k_a)}{\sigma_a^0(k_a)} \to 0, \text{ as } a \to \infty.$$

Table: Percentiles and $\hat{\alpha}$ for $a = 100$ and $n = 5$ case

	T_a^0			T_a^t		
	$\log a$	$a^{1/2}$	$a^{3/4}$	$\log a$	$a^{1/2}$	$a^{3/4}$
A95	1.645	1.645	1.645	1.645	1.645	1.645
S95	1.837	1.773	1.749	1.763	1.705	1.637
A90	1.282	1.282	1.282	1.282	1.282	1.282
S90	1.351	1.338	1.318	1.302	1.282	1.251
$\hat{\alpha}(.05)$	0.067	0.061	0.060	0.060	0.054	0.051
$\hat{\alpha}(.1)$	0.108	0.112	0.104	0.101	0.100	0.095

Simulations based on 3,000 runs.

Removing the Normality Assumption

Removing the Normality Assumption

- Three transformations:

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: $V = \Lambda(X) \sim \text{Exp}(1)$.

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
  $$\widehat{\Lambda}(t) = \int_0^t \frac{1}{1 - \widehat{F}(s-)} d\widehat{F}(s).$$

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
  $$\widehat{\Lambda}(t) = \int_0^t \frac{1}{1-\widehat{F}(s-)} d\widehat{F}(s).$$
- 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 $\overline{X}_{i\cdot}$ are not identically distributed even under H_0.

Unbalanced Designs, Heteroscedasticity

- In unbalanced designs the $\overline{X}_{i\cdot}$ are not identically distributed even under H_0.
- In heteroscedastic designs, each $\overline{X}_{i\cdot}$ needs to be scaled differently.

Unbalanced Designs, Heteroscedasticity

- In unbalanced designs the $\overline{X}_{i\cdot}$ are not identically distributed even under H_0.
- In heteroscedastic designs, each $\overline{X}_{i\cdot}$ 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 $\overline{X}_{i\cdot}$.