EIROPAS SOCIĀLAIS FONDS

LATVIJAS UNIVERSITĀTE
ANNO 1919

IEGULDĪJUMS TAVĀ NĀKOTNĒ

This work has been supported by the European Social Fund within the project «Support for Doctoral Studies at University of Latvia».

Bartlett correction for the empirical likelihood method for the two–sample mean problem

S. Vucāne[1]

[1]University of Latvia, Riga

28 th of May, 2011

Initial goal

Establish Bartlett correction for the empirical likelihood for the general two-sample problem

- mean difference
- quantile function difference
- probability-probability plots
- quantile-quantile plots
- ROC curves
- structural relationship models

Empirical likelihood method

Empirical likelihood method is the only nonparametric method that admits Bartlett adjustment and it was introduced by Art B. Owen in 1988.

$$L(F) = \prod_{i=1}^{n} P(X = X_i) = \prod_{i=1}^{n} p_i$$

Empirical likelihood (EL) for μ

- $X_1, \ldots, X_n$ iid with $EX_i = \mu_0 \in \mathbb{R}$.
- $g(X, \mu)$ such that $E\{g(X, \mu)\} = 0$ $(g(X_i, \mu) = X_i - \mu)$;
- Empirical likelihood for μ :

$$L(\mu) = \prod_{i=1}^{n} P(X = X_i) = \prod_{i=1}^{n} p_i$$

- $L(\mu)$ is maximized subject to constraints:

$$p_i \geq 0,\ \sum_i p_i = 1,\ \sum_i p_i g(X_i, \mu) = 0.$$

Empirical likelihood (EL) for μ

- Empirical likelihood ratio statistic for μ

$$R(\mu) = \frac{L(\mu)}{L(\hat{\mu})} = \prod_{i=1}^{n} \{1 + \lambda(\mu) g(X_i, \mu)\}^{-1},$$

Theorem (Owen, 1988)

$X_1, \ldots, X_n$ i.i.d. with $\mu_0 < \infty$. Then

$$W(\mu_0) = -2 \log R(\mu_0) \to_d \chi_1^2.$$

Simulation study

N(0,1)

	t $test$	B_{perc}	B_{norm}	B_{basic}	EL
$n = 20$	0.954	0.934	0.927	0.928	0.939
$n = 50$	0.959	0.950	0.950	0.950	0.945
$n = 100$	0.960	0.957	0.954	0.956	0.949

χ_1^2

	t $test$	B_{perc}	B_{norm}	B_{basic}	EL
$n = 20$	0.908	0.904	0.894	0.875	0.912
$n = 50$	0.931	0.930	0.923	0.916	0.933
$n = 100$	0.943	0.944	0.941	0.929	0.943

Development of Bartlett correction in one sample case

- Mean – Hall and LaScala (1990)
- Smooth function of mean – DiCiccio, Hall and Romano (1991)
- Quantiles – Chen un Hall (1993)
- Linear regression – Chen (1993, 1994)
- Nuisance parameters – Chen and Cui (1999)

Bartlett correction

Bartlett correction

Simple correction of $W(\mu_0)$ with its mean $\mathbb{E}\{W(\mu_0)\}$ reduces coverage error from $O(n^{-1})$ to $O(n^{-2})$.

Edgeworth series

Approximate a probability distribution in terms of its cumulants.

- $X_1, \ldots, X_n$ $i.i.d.$ with mean θ_0 and finite variance σ^2.
- $S_n = n^{1/2}(\hat{\theta} - \theta_0)/\sigma$, where $\hat{\theta} = \bar{X}$.

Edgeworth expansion for S_n

$$\mathbb{P}(S_n \leq x) = \Phi(x) + n^{-1/2} p_1(x)\phi(x) + n^{-1} p_2(x)\phi(x) + \ldots .$$

Derivation of Bartlett correction for the mean

Notation: $\alpha_k = \mathbb{E}(X^k)$ un $A_k = n^{-1}\sum_i X_i^k - \alpha_k$.

1. Solving $n^{-1}\sum_i (X_i - \mu)\{1 + \lambda(X_i - \mu)\}^{-1} = 0$ obtain expansion for $\lambda:$

$$\lambda = A_1 + \alpha_3 A_1^2 - A_1 A_2 + A_1 A_2^2 + A_1^2 A_3 + 2\alpha_3^2 A_1^3 - 3\alpha_3 A_1^2 A_2 - \alpha_4 A_1^3 + O_p(n^{-2}).$$

2. Obtain expansion of $n^{-1}W_0$:

$$n^{-1}W_0 = A_1^2 - A_2 A_1^2 + \frac{2}{3}\alpha_3 A_1^3 + A_2^2 A_1^2 + \frac{2}{3}A_3 A_1^3 - 2\alpha_3 A_2 A_1^3 + \alpha_3^2 A_1^4 - \frac{1}{2}\alpha_4 A_1^4 + O_p(n^{-5/2}).$$

3. Derive signed root of $n^{-1}W_0$.
$n^{-1}W_0 = R^2 + O_p(n^{-5/2}),\ R = R_1 + R_2 + R_3 + O_p(n^{-2})$ and

$$R_1 = A_1,\ R_2 = -\frac{1}{2}A_2A_1 + \frac{1}{3}\alpha_3 A_1^2,$$
$$R_3 = \frac{3}{8}A_2^2A_1 + \frac{1}{3}A_3A_1^2 - \frac{5}{6}\alpha_3A_2A_1^2 + \frac{4}{9}\alpha_3^2A_1^3 - \frac{1}{4}\alpha_4A_1^3.$$

4. Derive moments and cumulants of

$$n^{-1}W_0 = R_1^2 + 2R_1R_2 + 2R_1R_3 + R_2^2 + O_p(n^{-5/2}).$$

Johnson and Kotz showed, that sth cumulant of nR^2 is

$$\kappa_s = 2^{s-1}(s-1)!\{\mathbb{E}(nR^2)\}^s + O(n^{-3/2}).$$

And sth cumulant of $(nR^2)\{\mathbb{E}(nR^2)\}^{-1}$ is $2^{s-1}(s-1)!$, whitch is also $s-$th cumulant of χ_1^2.

Derivation of Bartlett correction for the mean

5. $\mathbb{P}\left[W_0\{\mathbb{E}(nR^2)\}^{-1} \leq z\right] = \mathbb{P}(\chi_1^2 \leq z) + O(n^{-2})$.

$$\begin{aligned}\mathbb{E}(nR^2) &= n\{\mathbb{E}(R_1^2) + 2\mathbb{E}(R_1R_2) + 2\mathbb{E}(R_1R_3) + \mathbb{E}(R_2^2)\} + O(n^{-2}) \\ &= 1 + n^{-1}\left(-\frac{1}{3}\alpha_3^2 + \frac{1}{2}\alpha_4\right) + O(n^{-2})\end{aligned}$$

6. If EL confidence interval for μ is defined as

$$I_\alpha = \{\mu : W(\mu) \leq c_\alpha\},$$

where c_α is such that $\mathbb{P}(\chi_1^2 \leq c_\alpha) = 1 - \alpha$, then EL confidence interval with Bartlett adjustment can be defined

$$I'_\alpha = \{\mu : W(\mu) \leq c_\alpha(1 + n^{-1}a)\},$$

Simulation study

		$N(0,1)$
$n = 10$	EL	0.8975
	$EL_{B_{theo}}$	0.9162
	$EL_{B_{est}}$	0.9118
$n = 20$	EL	0.9334
	$EL_{B_{theo}}$	0.9420
	$EL_{B_{est}}$	0.9410
$n = 50$	EL	0.9456
	$EL_{B_{theo}}$	0.9486
	$EL_{B_{est}}$	0.9482

		χ^2_1
$n = 10$	EL	0.8329
	$EL_{B_{theo}}$	0.8826
	$EL_{B_{est}}$	0.8480
$n = 20$	EL	0.8925
	$EL_{B_{theo}}$	0.9198
	$EL_{B_{est}}$	0.9042
$n = 50$	EL	0.9265
	$EL_{B_{theo}}$	0.9387
	$EL_{B_{est}}$	0.9328

Empirical likelihood for two sample case

- $X_1, \ldots, X_{n_1}$ are $i.i.d.$ from F_1 and $Y_1, \ldots, Y_{n_2}$ are $i.i.d.$ from F_2
- θ_0, $\Delta_0 \Longleftrightarrow w_1(X_i, \theta_0, \Delta_0, t)$, $w_2(Y_j, \theta_0, \Delta_0, t)$
- For mean difference:

$$\theta_0 = \int X dF_1(x),\ \Delta_0 = \int Y dF_2(y) - \int X dF_1(x),$$
$$w_1 = X - \theta_0,\ w_2 = Y - \theta_0 - \Delta_0$$

- $\mathbb{E}_{F_1} w_1(X, \theta_0, \Delta_0, t) = 0$ and $\mathbb{E}_{F_2} w_2(Y, \theta_0, \Delta_0, t) = 0.$

Empirical likelihood for two sample case

- Empirical likelihood ratio is defined as

$$R(\Delta, \theta) = \sup_{\theta, p, q} \prod_{i=1}^{n_1} (n_1 p_i) \prod_{j=1}^{n_2} (n_2 q_j).$$

- Empirical loglikelihood ratio is defined as

$$W(\Delta, \theta) = -2 \log R(\Delta, \theta) = 2 \sum_{i=1}^{n_1} \log(1 + \lambda_1(\theta) w_1(X_i, \theta, \Delta, t))$$

$$+ 2 \sum_{j=1}^{n_2} \log(1 + \lambda_2(\theta) w_2(Y_j, \theta, \Delta, t)).$$

Under certain conditions

$$-2 \log R(\Delta_0, \hat{\theta}) \to_d \chi_1^2.$$

Development of Bartlett correction in two sample case

Jing (1995)

Obtained Bartlett correction for EL for two sample mean difference.

Liu, Zou and Zhang (2008)

Corrected mistake made by Jing and obtained correct version of Bartlett correction for EL for two sample mean difference.

Liu un Yu (2010)

Obtained Bartlett correction for adjusted EL for two sample mean difference.

Bartlett correction for two sample mean difference

Jing (1995) obtained $n^{-1}W$ as

$n^{-1}W = \Delta_1 + O_p(n^{-5/2})$.

Liu, Zou un Zhang (2008) obtained $n^{-1}W$ as

$n^{-1}W = \Delta_1 + \Delta_2^* + O_p(n^{-5/2})$.

During additional analysis is was obtained, that Δ_2^* should be replaced by Δ_2, where $\Delta_2 = \Delta_2^* + \delta$.

Simulation study for mean difference of $\exp(1)$ and $\exp(2)$

		$n_2 = 10$	$n_2 = 20$	$n_2 = 30$
$n_1 = 10$	EL	0.8862	0.8803	0.8757
	$EL_{B_{theo}}$	0.9186	0.9134	0.9167
	$EL_{B_{est}}$	0.9015	0.8962	0.8946
$n_1 = 20$	EL	0.9170	0.9163	0.9201
	$EL_{B_{theo}}$	0.9379	0.9378	0.9343
	$EL_{B_{est}}$	0.9305	0.9257	0.9280
$n_1 = 30$	EL	0.9200	0.9261	0.9289
	$EL_{B_{theo}}$	0.9389	0.9396	0.9430
	$EL_{B_{est}}$	0.9339	0.9378	0.9374

Simulation study for mean difference of χ_3^2 and $\exp(1)$

		$n_2 = 10$	$n_2 = 20$	$n_2 = 30$
$n_1 = 10$	EL	0.8877	0.9210	0.9284
	$EL_{B_{theo}}$	0.9183	0.9364	0.9412
	$EL_{B_{est}}$	0.9057	0.9352	0.9374
$n_1 = 20$	EL	0.8915	0.9213	0.9313
	$EL_{B_{theo}}$	0.9162	0.9358	0.9425
	$EL_{B_{est}}$	0.9056	0.9318	0.9403
$n_1 = 30$	EL	0.8838	0.9251	0.9342
	$EL_{B_{theo}}$	0.9119	0.9355	0.9427
	$EL_{B_{est}}$	0.9007	0.9286	0.9379

Achieved result in the establishment of Bartlett correction for EL for the general two-sample problem

$$
\begin{aligned}
W &= 2\tilde{v}_1\bar{w}_{12}^{-1}\bar{w}_{11} + (2\tilde{v}_1\bar{w}_{13}\bar{w}_{12}^{-3} - \tilde{v}_2\bar{w}_{12}^{-2})\bar{w}_{11}^2 \\
&+ \left(\frac{2}{3}\tilde{v}_3\bar{w}_{12}^{-3} - 2\tilde{v}_2\bar{w}_{13}\bar{w}_{12}^{-4} - 2\tilde{v}_1\bar{w}_{14}\bar{w}_{12}^{-4} + 4\tilde{v}_1\bar{w}_{13}^2\bar{w}_{12}^{-5}\right)\bar{w}_{11}^3 \\
&+ \left(2\tilde{v}_3\bar{w}_{12}^{-5}\bar{w}_{13} - \frac{1}{2}\tilde{v}_4\bar{w}_{12}^{-4} + 2\tilde{v}_2\bar{w}_{14}\bar{w}_{12}^{-5} - 5\tilde{v}_2\bar{w}_{13}^2\bar{w}_{12}^{-6}\right)\bar{w}_{11}^4 \\
&+ \left(\frac{2}{5}\tilde{v}_5\bar{w}_{12}^{-5} + 2\tilde{v}_2\bar{w}_{13}\bar{w}_{14}\bar{w}_{12}^{-7} - 2\tilde{v}_3\bar{w}_{14}\bar{w}_{12}^{-6} + 6\tilde{v}_3\bar{w}_{13}^2\bar{w}_{12}^{-7}\right)\bar{w}_{11}^5 \\
&- \left(4\tilde{v}_2\bar{w}_{13}^3\bar{w}_{12}^{-8} + 2\tilde{v}_4\bar{w}_{13}\bar{w}_{12}^{-6}\right)\bar{w}_{11}^5 + \sum_{k=5}^{j} R_{2k}\bar{w}_{11}^{k+1} \\
&+ (o_p(b) + O_p(\delta + l^{-1/2}))^{j+2},
\end{aligned}
$$

where $\tilde{v}_k = n_1\bar{w}_{1k} + n_2c^k\bar{w}_{2k}$, $\bar{w}_{1k} = n_1^{-1}\sum_{i=1}^{n_1} w_1^k$ and $\bar{w}_{2k} = n_2^{-1}\sum_{j=1}^{n_2} w_2^k$.

Thank you for your attention!