Tail bound inequalities and empirical likelihood for the mean

Sandra Vucane[1]

[1]University of Latvia, Riga

29 th of September, 2011

Motivation

Dealing with confidence intervals for mean for small samples sizes from highly skewed, heavy tailed probability distributions
Rosenblum and van der Laan (2009)

Claimed

normal-based methods ($z-$test, $t-$test) and nonparametric bootstrap give poor coverage

Provided

alternative method using exponential type inequalities

Idea: Compare the method proposed by Rosenblum and van der Laan with the empirical likelihood method.

Bernstein's inequality

- $X_1, \ldots, X_n$ independent, bounded r.v.
- $\mathbb{P}(|X_i - \mathbb{E}X_i| \leq W) = 1$
- $v \geq \sum_{i=1}^{n} \mathbb{D}(X_i)$

Bernstein's inequality (1927)

For all $x > 0$

$$\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)\right| > x/\sqrt{n}\right) \leq 2\exp\left(-\frac{nx^2}{2(v + W\sqrt{n}x/3)}\right)$$

Alternatives: Bennett's inequality (1962), Hoeffding's inequality (1963), Berry-Esseen inequality (1941, 1942)

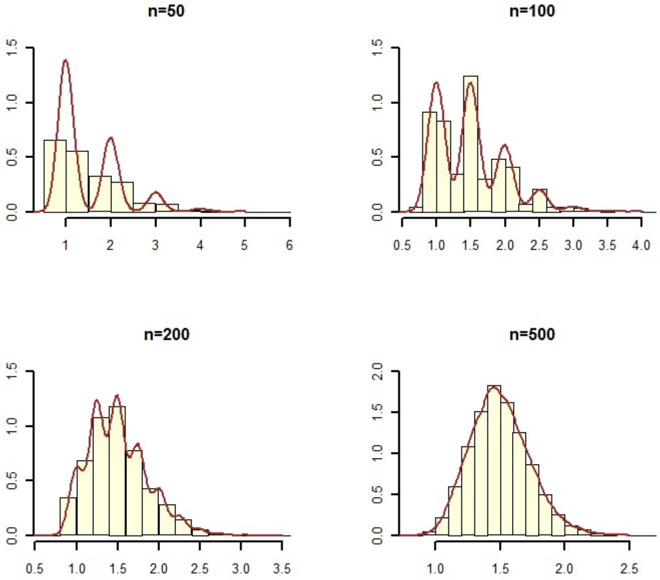
Example - standard confidence intervals perform poorly

Construction of random variable X is as follows:

- $X = A + Y$,
- A and Y are independent random variables,
- $Y \sim Pois(\lambda)$
- $\mathbb{P}(A = 0) = 1 - \delta$, $\mathbb{P}(A = t) = \delta$, where $\delta \in (0, 1)$
- $\mathbb{E}X = \mathbb{E}A + \mathbb{E}Y = t\delta + \lambda$,
- $\mathbb{D}X = \mathbb{D}A + \mathbb{D}Y = t^2\delta(1 - \delta) + \lambda$.

!!!For simulations we choose $\delta = 0.01, t = 50, \lambda = 1$.

Example - standard confidence intervals perform poorly



Example - standard confidence intervals perform poorly

Table: Coverage accuracy for the mean of X from Example, $\alpha = 0.05$

	t	B_{perc}	B_{norm}	B_{basic}	B_{stud}	$Bernst$
$n = 10$	0.696	0.644	0.567	0.593	0.760	0.908
$n = 20$	0.518	0.520	0.477	0.477	0.636	0.848
$n = 50$	0.429	0.428	0.423	0.426	0.450	0.668
$n = 100$	0.626	0.623	0.626	0.626	0.619	0.652
$n = 200$	0.879	0.878	0.879	0.866	0.873	0.881
$n = 500$	0.905	0.928	0.903	0.875	0.954	0.994

Empirical likelihood method (EL)

Empirical likelihood method was introduced by Art B. Owen in 1988. The idea - model data using distributions placing point masses on the data.

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

!!! Empirical likelihood method is the only nonparametric method that admits Bartlett adjustment.

Empirical likelihood

- $X_1, \dots, X_n$ 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

- 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

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

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

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

Example - standard confidence intervals perform poorly

	t	B_{perc}	B_{norm}	B_{basic}	B_{stud}	$Bernst$
$n = 10$	0.696	0.644	0.567	0.593	0.760	0.908
$n = 20$	0.518	0.520	0.477	0.477	0.636	0.848
$n = 50$	0.429	0.428	0.423	0.426	0.450	0.668
$n = 100$	0.626	0.623	0.626	0.626	0.619	0.652
$n = 200$	0.879	0.878	0.879	0.866	0.873	0.881
$n = 500$	0.905	0.928	0.903	0.875	0.954	0.994

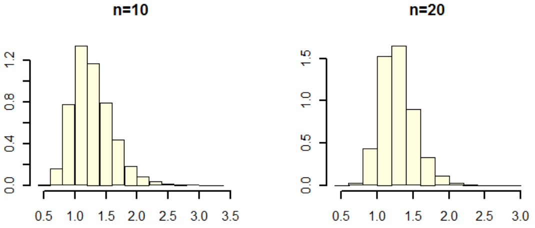
Example - standard confidence intervals perform poorly

	t	B_{perc}	B_{norm}	B_{basic}	B_{stud}	$Bernst$
$n = 10$	0.696	0.644	0.567	0.593	0.760	0.908
$n = 20$	0.518	0.520	0.477	0.477	0.636	0.848
$n = 50$	0.429	0.428	0.423	0.426	0.450	0.668
$n = 100$	0.626	0.623	0.626	0.626	0.619	0.652
$n = 200$	0.879	0.878	0.879	0.866	0.873	0.881
$n = 500$	0.905	0.928	0.903	0.875	0.954	0.994

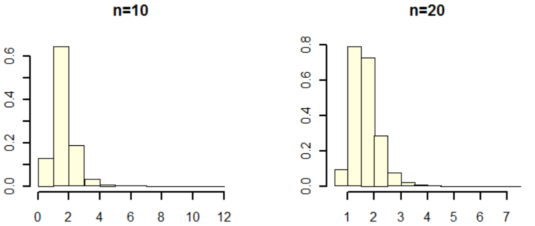
	t	EL	EL_B	EL_F	EL_{FB}	$Bernst$
$n = 10$	0.696	0.614	0.634	0.690	0.698	0.908
$n = 20$	0.518	0.522	0.546	0.570	0.581	0.848
$n = 50$	0.429	0.428	0.430	0.432	0.437	0.668
$n = 100$	0.626	0.611	0.613	0.611	0.614	0.652
$n = 200$	0.879	0.865	0.867	0.866	0.867	0.881
$n = 500$	0.905	0.933	0.944	0.935	0.945	0.994

Example: Lognormal distribution $\mathrm{LogN}(\mu, \sigma^2)$

$\boldsymbol{\sigma^2 = 0.5}$

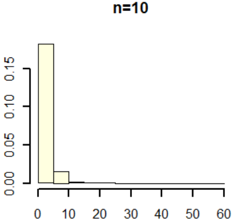


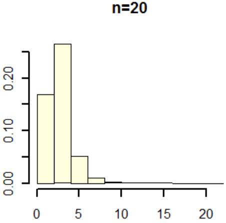
$\boldsymbol{\sigma^2 = 1}$



Example: Lognormal distribution LogN(μ, σ^2)

$\boldsymbol{\sigma^2 = 2}$





	σ^2	t	EL	EL_B	EL_F	EL_{FB}	B_{perc}	B_{norm}	B_{basic}	B_{stud}	$Bernst$
$n = 10$	0.5	0.894	0.856	0.871	0.893	0.902	0.858	0.852	0.839	0.931	0.978
	1.0	0.853	0.822	0.841	0.859	0.873	0.816	0.811	0.772	0.912	0.942
	2.0	0.757	0.743	0.763	0.791	0.808	0.739	0.719	0.675	0.877	0.876
	5.0	0.518	0.524	0.546	0.566	0.582	0.495	0.466	0.421	0.781	0.663
$n = 20$	0.5	0.922	0.916	0.928	0.935	0.941	0.904	0.896	0.886	0.943	0.986
	1.0	0.866	0.878	0.889	0.901	0.910	0.859	0.845	0.822	0.927	0.972
	2.0	0.785	0.802	0.820	0.826	0.839	0.786	0.760	0.732	0.898	0.926
	5.0	0.560	0.590	0.611	0.618	0.639	0.569	0.541	0.498	0.797	0.742

Summary for independent case

- Bernstein's method can be recommended for very sparse data usually with big variance
- Coverage accuracy for larger samples grow to 1, that makes the comparison more difficult
- For any distribution we can find such parameter W and v values that the coverage accuracy of Bernstein's method is 95%, but these parameter values usually don't satisfy assumptions

Dependent case

1. Empirical likelihood
   - Kitamura (1997) - empirical likelihood for weakly dependent processes
   - Kitamura (1997) - Bartlett correction for blockwise empirical likelihood of the smooth function
   - Nordman et al. (2007) - empirical likelihood for the mean of long-range dependent processes
2. Exponential type inequalities
   - Bosq (1996)
   - Dedecker, J., Doukhan, P., Lang, G., León R., J. R., Louhichi, S. and Prieur, C. (2007)

Work in progress: comparison for weakly dependent processes

Empirical likelihood for weakly dependent processes

- $\{X_t\}$ $\mathbb{R}^d-$valued stationary stochastic process that satisfy strong mixing condition
$$\alpha_X(k) \to 0 \text{ as } k \to \infty;$$
- Parameter of interest $\theta_0 \in \Theta \subset \mathbb{R}^p$, $r \geq p$;
- Estimating equation $f : \mathbb{R}^d \times \Theta \to \mathbb{R}^r$ and $E\{f(X_t, \theta_0)\} = 0$;
- $M-$ length of block, $L-$block number $\Rightarrow Q = [(N-M)/L] + 1$ and $A_N = QM/N$;
- $T_i(\theta) = \phi_M(B_i, \theta)$, where B_i is the ith block of observations and mapping $\phi_M : \mathbb{R}^{rM} \times \Theta \to \mathbb{R}^r$ has following form
$$\phi_M(B_i, \theta) = \sum_{i=1}^{M} f(X_{(i-1)L+n}, \Theta)/M$$

Empirical likelihood for weakly dependent processes

- Empirical likelihood for θ :

$$L(\theta) = \prod_{i=1}^{Q} p_i$$

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

$$p_i > 0, \ \sum_i p_i = 1, \ \sum_i p_i T_i(\theta) = 0.$$

- Blockwise empirical log-likelihood ratio for θ

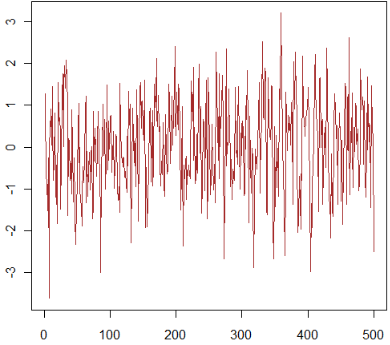
$$-2\log\left(L(\theta)/L_F\right) = 2\sum_{i=1}^{Q} \log\left(1 + \gamma_N(\theta)' T_i(\theta)\right).$$

Theorem (Kitamura, 1997)

$$LR_0 = 2{A_N}^{-1} \sum_{i=1}^{Q} \log\left(1 + \gamma_N(\theta)' T_i(\theta)\right) \rightarrow_d \chi^2_{r-p}.$$

Simulations

$$AR(1) \text{ with } X_t = 0.3X_{t-1} + \varepsilon_t$$



N/M	5	10	20	50
200	0.973	0.923	0.852	0.735
500	0.975	0.949	0.909	0.862
1000	0.981	0.948	0.931	0.888

Exponential type inequalities

Bosq (1996)

Let $(X_t, t \in \mathbb{Z})$ be a zero-mean real-valued process such that $\sup_{1 \le t \le n} \|X_t\|_\infty \le b$. Then for each integer $q \in [1, \frac{n}{2}]$ and each $\varepsilon > 0$

$$\mathbb{P}(|S_n| > n\varepsilon) \le 4 \exp\left(-\frac{\varepsilon^2 q}{8b^2}\right) + 22\left(1 + \frac{4b}{\varepsilon}\right)^{1/2} q\alpha\left(\left[\frac{n}{2q}\right]\right).$$

Doukhan (1995)

Let $(X_t, t \in \mathbb{N})$ be a zero-mean real-valued process such that $\exists \sigma^2 \in \mathbb{R}_+$, $\forall n, m \in \mathbb{N} : \frac{1}{m}(X_n + \ldots + X_{n+m})^2 \le \sigma^2$ and $\forall t \in \mathbb{N} : |X_t| \le M$. Then for each $\varepsilon > 0$, $\theta = \frac{\varepsilon^2}{4}$ and $q \le \frac{n}{1+\theta}$

$$\mathbb{P}(|S_n| \ge x) \le 4 \exp\left(-\frac{(1-\varepsilon)x^2}{2(n\sigma^2 + qMx/3)}\right) + 2^{n\frac{\beta(\lceil q\theta \rceil - 1)}{q}}.$$

!!!Problem - estimation of mixing coefficients

McDonald et al (2011)-method for estimating beta-mixing coefficients

Exponential type inequalities

Bennett-type inequality

Let $(X_i)_{i>0}$ such that $\|X_i\|_\infty$, and $\mathcal{M}_i = \sigma(X_k, 1 \leq k \leq i)$. Let $S_k = \sum_{i=1}^{k}(X_i - \mathbb{E}(X_i))$ and $\bar{S}_n = max_{1\leq k\leq n}|S_k|$. Let q be some positive integer, v_q some nonnegative number such that

$$v_q \geq \|X_{q[n/q]+1} + \ldots + X_n\|_2^2 + \sum_{i=1}^{[n/q]} \|X_{(i-1)q+1} + \ldots + X_{iq}\|_2^2$$

and $h(x) = (1+x)\log(1+x) - x$. Then for $\lambda > 0$,

$$\mathbb{P}(|S_n| \geq 3\lambda) \leq 4\exp\left(-\frac{v_q}{(qM)^2} h\left(\frac{\lambda q M}{v_q}\right)\right) + \frac{n}{\lambda}\tau_{1,q}(q+1).$$

!!!Problem - estimation of mixing coefficient

Thank you for your attention!