Empirical Likelihood: A Review

Stanislav Kolenikov
skolenik@unc.edu
117 New West, Cameron Ave,
University of North Carolina,
Chapel Hill, NC 27599-3260, US

Parametric Likelihood - I

Usual assumptions:

$$\dim \Theta < \infty, \tag{1}$$

$$\theta_1, \theta_2 \in \Theta, \theta_1 \neq \theta_2 \Rightarrow \text{not } f(x|\theta_1) = f(x|\theta_2) \text{ a.e.} \tag{2}$$

$$l(\theta|x) = \log f(x|\theta) \in C^3(\Theta);$$

$$\left| \frac{\partial^3}{\partial \theta^3} l(\theta|x) \right| < g(x), \quad E_\theta |g(x)| < \infty \, \forall \theta \tag{3}$$

$$s(\theta) = \frac{\partial}{\partial \theta} l(\theta|x) \Rightarrow \tag{4}$$

$$E s(\theta) s^T(\theta) = E\left(-\frac{\partial}{\partial \theta} s(\theta) \right) = I(\theta), \quad I(\theta) > 0 \tag{5}$$

Parametric Likelihood - II

ML estimate:

$$\hat\theta_{ML} = \arg\max_{\theta \in \Omega} l(\theta|x), \quad \text{or } (\textit{local max}) \; s(\hat\theta_{ML}) = 0 \tag{6}$$

Properties: consistent, asymptotically normal, asymptotically efficient, invariant

Wilks (1938) theorem: in testing $H_0 : \theta = \theta_0 \in \operatorname{int} \Theta$ *vs.* $H_1 : \theta \neq \theta_0$, $\dim \Theta = r$, Taylor expansion near θ_0 together with consistency give that

$$W = -2\Big(l(\theta_0|x) - \sup_{\theta \in \Theta \setminus \{\theta_0\}} l(\theta|x) \Big) \xrightarrow{d} \chi^2(r) \tag{7}$$

Confidence intervals are of the form $\Big\{ \theta \in \Theta | l(\theta|x) - l(\hat\theta|x) > c \Big\}$.

Empirical distribution function

Suppose $X_1, \dots, X_n \sim F$ for some unknown F. We then at least know that $\operatorname{supp} F \supseteq \{X_1, \dots, X_n\}$. What is the ML estimate of the distribution function in the class of functions that satisfy this property? I.e.

$$L(F) = \prod_{i=1}^n F\{x_i\} \rightarrow \begin{array}{c} \text{max over probability measures} \\ \text{with support on } [X_{(1)}, X_{(n)}] \end{array} \tag{8}$$

The answer is the CDF:

$$\hat F_{ML} = \frac{1}{n} \sum_i \delta_{x_i} \equiv F_n \tag{9}$$

Functionals of distributions

How do we test hypotheses when we have no parameters? (F_n is parameter-free).

Quite often, the parameter of the distribution is also its moment ($\theta = EX$) or a quantile ($P_\theta(X < \theta) = 1/2$).

Idea: use functionals of distributions: $\theta = T(F)$.

Examples:

mean: $\theta = EX = \int_{-\infty}^{+\infty} x\, dF(x)$

median: $\theta = a : -\int_{-\infty}^{a} dF(x) + \int_{a}^{+\infty} dF(x) = 0$

Empirical likelihood - I

Owen (1988): for a distribution function F, define the empirical likelihood ratio

$$R(F) = L(F)/L(F_n) \tag{10}$$

Then the likelihood ratio test statistic for $H_0 : T(F_0) = t$ is

$$W = -2 \sup\{\log R(F) | T(F) = t, \operatorname{supp} F \subset [X_{(1)}, X_{(n)}]\} \tag{11}$$

Also, confidence regions are

$$\mathcal{R}(c) = \{T(F) | R(F) \ge c\} \tag{12}$$

Empirical likelihood - II

How do "good" distribution functions under H_0 look like? It should be the case that candidate $F \ll F_n$. Then $F = \sum_i w_i \delta_{X_i}$ for some non-negative w_i's, $\sum_i w_i = 1$;

the likelihood function is $\tilde{L}(F, w) = \prod_{i=1}^{n} w_i$,

the likelihood ratio becomes $\tilde{R}(F, w) = \prod_{i=1}^{n} n w_i$.

Simple example - I

Sample: 1, 7, 9

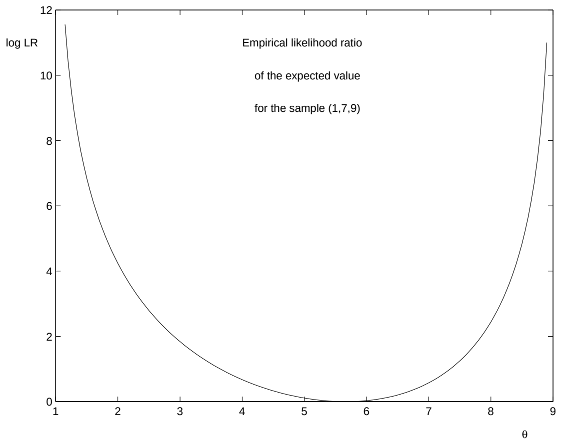
Mean: $\bar{X} = (1 + 7 + 9)/3 = 17/3 = 5.667$

Testing $\theta = EX = 5$: $w_1 = 0.4248$, $w_2 = 0.3009$, $w_3 = 0.2743$, $W = 0.1096$.

Testing $\theta = EX = 2$: $w_1 = 0.8544$, $w_2 = 0.0823$, $w_3 = 0.0633$, $W = 4.2383$.

What is the distribution of the empirical LR statistic? Should W be compared to $\chi^2(1)$?

Simple example - II



Empirical likelihood ratio in this simple example

Asymptotics - I

Owen (1988) shows the analogue of Wilks theorem for convergence of the empirical likelihood ratio for the population mean.

Theorem 1 *Let $X_1, X_2, \dots$ be independent random variables with nondegenerate distribution function F_0 s.t. $\int |x|^3 \, dF_0 < \infty$. For positive $c < 1$ let*

$$\mathcal{F}_{c,n} = \{F | R(F) \geq c, F \ll F_n\}, \tag{13}$$

and define $X_{U,n} = \sup_{\mathcal{F}_{c,n}} \int x \, dF$ and $X_{L,n} = \inf_{\mathcal{F}_{c,n}} \int x \, dF$. Then as $n \to \infty$,

$$\text{Prob}\{X_{L,n} \leq E[X] \leq X_{U,n}\} \to \text{Prob}\{\chi^2(1) \leq -2 \log c\}. \tag{14}$$

Note: # of ancillary parameters $\to \infty$.

Outline of the proof - I

1. $X_{U,n} = \sup \sum w_i X_i$, $X_{L,n} = \inf \sum w_i X_i$, $w_i \geq 0$, $\sum w_i = 1$, $\prod n w_i \geq c$.
2. Assume $E[X] = 0$ and define $G = \sum \log n w_i + \gamma(1 - \sum w_i) + n\lambda(0 - \sum w_i X_i)$. Then taking the derivatives gives $w_i = 1/(\gamma + n\lambda X_i)$ and further $\gamma = n$, which implies that the log likelihood ratio in question is $\log R_0 = -\sum \log(1 + \lambda_0 X_i)$ where λ_0 is the root of $0 = n^{-1} \sum X_i/(1 + \lambda X_i) \equiv g(\lambda)$, $\lambda_0 \in (-X_{(n)}^{-1}, -X_{(1)}^{-1})$.
3. $\lambda_0 = O_p(n^{-q})$ for some $q < 1/2$, as $n^{1/2} g(n^{-q}) \leq n^{1/2}\bar{X} - n^{1/2} S^2/n^q + n^{1/3}$ (where $S^2 = n^{-1} \sum X_i^2$) since the third moment is finite, and thus $\text{Prob}\{\max |X_i| > n^{1/3} \text{ i.o.}\} = 0$.

Outline of the proof - II

4. Taylor expansion for $g^{-1}(\cdot)$: $\lambda_0 = g^{-1}(0) = g^{-1}(\bar{X}) + (0 - \bar{X})(g^{-1})'(\xi) = -\bar{X}(g^{-1})'(\xi)$, $|\xi| \leq |\bar{X}|$, $\eta = g^{-1}(\xi)$, $|\eta| \leq |\lambda_0|$, and then $\lambda_0 = r_0 \bar{X}/S^2$, $r_0 = -S^2/g'(\eta) \xrightarrow{p} 1$.
5. Taylor expansion for the likelihood ratio: $-2 \log R_0 = 2 \sum \log(1 + \lambda_0 X_i) = 2 \sum \left(\lambda_0 X_i - (\lambda_0 X_i)^2 + \eta_i\right) = 2n\bar{X}^2 r_0/S^2 - nS^2(\bar{X}^2 r_0/S^2)^2 + \sum \eta_i = (2r_0 - r_0^2) n\bar{X}^2/S^2 + \sum \eta_i$, $|\sum eta_i| \leq |\lambda_0|^3 \sum |X_i|^3 = O_p(n^{-1/2})$, $2r_0 - r_0^2 = 1 + o_p(1)$, and by the CLT, $n\bar{X}^2/S^2 \xrightarrow{d} \chi^2(1)$.

M-estimates

Theorem 1 can be extended to the M-estimates $\tau = T(F)$ that solve, for some "regular" ψ, the equation

$$\int \psi(X,\tau)\, F(dX) = 0 \tag{15}$$

Examples.

1. Mean: $\psi(x,t) = x - t$
2. Quantiles: $\psi(x,t) = \begin{cases} 1, & x \le t \\ -q/(1-q), & x > t \end{cases}$
3. Huber's robust location: $\psi(x,t) = \begin{cases} c, & x - t \ge c \\ x - t, & |x-t| \le c \\ -c, & x - t \le -c \end{cases}$

Asymptotics - II

Owen (1990) gives an extended version of Theorem 1:

Theorem 2 *If $X_1, X_2, \ldots \sim$ i.i.d. in $\mathbb{R}^p$, $\mu = E[X_1]$ is finite,* rk Cov $X_1 = q$, $C_{r,n} = \left\{ \int X\, dF | R(F) \ge r, F \ll F_n \right\}$. *Then*

$$\lim_{n\to\infty} \text{Prob}\,[\mu \in C_{r,n}] = \text{Prob}\,[\chi^2(q) \le -2\ln r] \tag{16}$$

Moreover, if $E\|X\|^4 < \infty$, then the rate of convergence is $O(n^{-1/2})$.

The solution is then given by the dual convex problem:

$$-\sum \ln(1 + \lambda'(X_i - \mu)) \to \min\{\lambda\} \tag{17}$$

$$0 = \sum_i \frac{X_i - \mu}{1 + \lambda'(X_i - \mu)} \tag{18}$$

$$F_\mu\{X_i\} = \frac{1}{n}\frac{1}{1 + \lambda'(x_i - \mu)} \tag{19}$$

Asymptotics - III

DiCiccio, Hall and Romano (1989) expand the latter expression in $\lambda'(X_i - \mu)$ to show that $\lambda = \bar{X} - \mu + \epsilon$ where $\epsilon = O(n^{-1}\log\log n)$ a.s. and $\epsilon = O(n^{-1})$ in probability if $E\|X\|^4 < \infty$. Furthermore, introducing

$$g(\nu) = \sum_{j,k} \nu^j \nu^k E[X^j X^k X] \quad \text{and} \quad \Delta = n^{-1}\sum_i X_i X_i^T - I,$$

they show that if $E\|X\|^6 < \infty$,

$$\lambda = \bar{X} - \mu + g(\bar{X} - \mu) - \Delta(\bar{X} - \mu) + \xi, \quad \xi = O_p(n^{-3/2}).$$

Finally, they expand the empirical likelihood ratio

$$l_E(\mu) = n\lambda^T(I + \Delta)\lambda - \frac{4}{3}\sum_i \left[(\bar{X} - \mu)^T X_i\right]^3 + R_n, \quad R_n = O_p(n^{-1}).$$

Asymptotics - IV

By getting a similar expansion for the parametric likelihoods, DiCiccio, Hall and Romano (1989) conclude that the two do not necessarily agree even to the first order, i.e. $O_p(n^{-1/2})$. They exemplify the point with a double exponential distribution (where the mean is not the location parameter, however). The confidence region coverage, however, have an error of order $O(n^{-1/2})$, i.e. the empirical likelihood provides right nominal coverage.

Bivariate example - I

Owen (1990) illustrates the multivariate extension with the genetic experiment on duck plumage data (see references in the article). Fig. 1 shows the empirical likelihood contours that used 20/9 times the $F_{2,9}$ distribution (small sample correction instead of the asymptotic $\chi^2(2)$). Fig. 2 shows parametric likelihood ratio contours for Hotelling's T^2 statistic based on $F_{2,9}$ distribution. Fig. 3 (not shown) attempts to construct the bootstrap based confidence regions.

Bivariate example - II

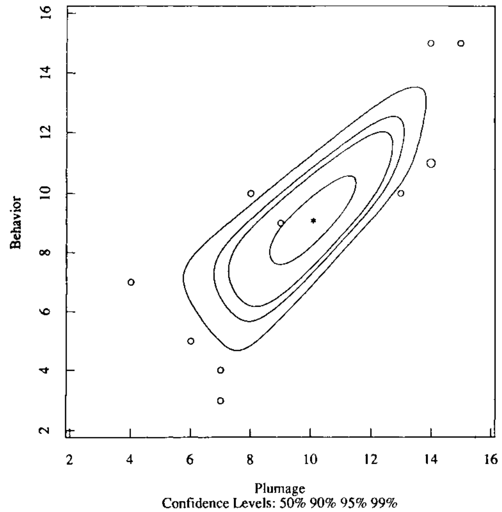


FIG. 1. *Empirical likelihood contours.*

Bivariate example - III

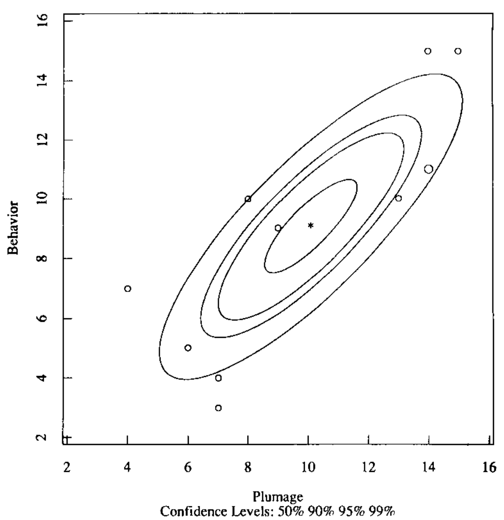


FIG. 2. *Normal likelihood contours.*

Regression - I

$$Y = X\beta_0 + \epsilon, \quad E\,\epsilon|X = 0, \quad \operatorname{Var}\epsilon|X = \sigma_i^2 \tag{20}$$

Stochastic regressors case is easier! Normal equations / estimating equations / moment conditions (econometrics):

$$Z_i = X_i'(Y_i - X_i\beta), \beta = \beta_0 \Rightarrow EZ_i = 0 \tag{21}$$

Empirical likelihood ratio test: whether $EZ_i = 0$.

$$R(\beta) = \max\{\sum \ln nw_i | w_i \geq 0, \sum w_i = 1, \sum w_i x_i'(y_i - x_i\beta)\} \tag{22}$$

Can also build CIs for a subset of r regressors — compare $R(\beta)$ to $\chi^2(r)$. Owen (1990): compare to scaled F distribution?

Regression - II

The case with fixed regressors is more difficult.

Theorem 3

$$\mu_4(x) = \int (Y - x\beta_0)^4 \, dF_x, \; n^{-2} \sum_i \|x_i\| \mu_4(x_i) \to 0,$$

$$\text{Prob}\big\{ \text{conv}\,\{x_i | Y_i - x_i\beta_0 > 0\} \cap \text{conv}\,\{x_i | Y_i - x_i\beta_0 < 0\} \neq \emptyset \big\} \to 1,$$

$$a < \sigma^2(x_i) < b\|x_i\|^\alpha, \; a, \, b > 0, \; \alpha \geq 0,$$

$$a < \min \lambda\{X'X/n\}, \; \frac{1}{n} \sum \|x_i\|^{2+\alpha} < b$$

$$\Rightarrow -2 \log R(\beta_0) \xrightarrow{p} \chi^2(p) \tag{23}$$

To the order $O_p(N^{-1/2})$, the empirical likelihood is equivalent to Huber / White / 1^{st} order linearization heteroskedasticity consistent covariance matrix.

Regression - III

Misspecification: $E(Z_i) = \mu_i$, $\text{Var}\, Z_i = V_i$ is of full rank (1D case: $V_i = \sigma_i^2$). We are testing whether $\bar{\mu} = 1/n \sum \mu_i$ takes a given value μ_0. The empirical likelihood test refers

$$\frac{n(\bar{Z} - \mu_0)^2}{1/n \, \sum (Z_i - \mu_0)^2}$$

to a χ^2 distribution. The variance estimate in the denominator is biased upward due to the model misspecification:

$$E1/n \, \sum (Z_i - \mu_0)^2 = 1/n \sum_i \sigma_i^2 + 1/n \, \sum (\mu_i - \mu_0)^2$$

Thus, Owen (1991) concludes, confidence sets for β_0 will not have consistent coverage, but will be conservative.

Small sample properties - I

Asymptotically, everything is standard normal, or χ^2 — what is the scope for empirical likelihood, then?

Hint: bias, skewness, heavy tails.

Owen (1990) argues that the empirical likelihood corrects the skewness as compared to Student's t, and notes that Johnson's t (Johnson 1978) corrects for both skewness and bias.

Edgeworth and Cornish-Fisher expansions - I

Edgeworth expansion: the one for distribution functions:

$$\text{Prob}\left\{ n^{1/2}(\hat{\theta} - \theta_0)/\sigma \leq x \right\} = \Phi(x) + n^{-1/2} p_1(x)\phi(x) +$$

$$+ n^{-1} p_2(x)\phi(x) + \ldots + n^{-j/2} p_j(x)\phi(x) + \ldots \tag{24}$$

Cornish-Fisher expansion: the one for quantiles:

$$\text{Prob}\left\{ n^{1/2}(\hat{\theta} - \theta_0)/\sigma \leq x \right\} = \alpha \Rightarrow$$

$$u_\alpha = z_\alpha + n^{-1/2} \tilde{p}_1(z_\alpha) + n^{-1} \tilde{p}_2(z_\alpha) + \ldots + n^{-j/2} \tilde{p}_j(z_\alpha) + \ldots \tag{25}$$

Edgeworth expansion - II

Sums of independent random variables: if $X_1, X_2, \dots$ are i.i.d., then $S_n = n^{1/2}(\bar{X} - E[X])/\operatorname{Var} X$ is asymptotically standard normal. If

$$\chi(t) = E\left[\exp\left(it\frac{X - E[X]}{(\operatorname{Var} X)^{1/2}}\right)\right] = \exp\left[\sum_j \kappa_j (it)^j/j!\right] \qquad (26)$$

is the characteristic function of the normalized distribution, where κ_j's are its cumulants, then

$$\chi_n(t) \equiv \chi_{S_n}(t) = \left[\chi(t/n^{1/2})\right]^n = \exp\{-\frac{t^2}{2} + n^{-1/2}\kappa_3\frac{(it)^3}{3!} + \dots\} =$$
$$= e^{-t^2/2}\{1 + n^{-1/2}r_1(it) + n^{-1}r_2(it) + \dots\}, \qquad (27)$$

where r_j is a polynomial of degree $3j$ depending on the cumulants up to order $j+2$.

Edgeworth expansion - III

In particular,

$$r_1(u) = \frac{1}{6}\kappa_3 u^3, \quad r_2(u) = \frac{1}{24}\kappa_4 u^4 + \frac{1}{72}\kappa_3^2 u^6. \qquad (28)$$

As $\chi_n(t) = \int e^{itx} dF_{S_n}(x) = \int e^{itx} d\left[\Phi(x) + \sum n^{-j/2}R_j(x)\right]$, the functions $R_j(\cdot)$ are the solutions of $\int e^{itx}\, dR_j(x) = r_j(it)e^{-t^2/2}$, which can be shown to be related to the derivatives of the normal CDF:

$$R_j(x) = r_j\left(-\frac{d}{dx}\right)\Phi(x) \qquad (29)$$

Then the first two terms of the Edgeworth expansion for the sample mean are,

$$R_1(x) = -\frac{1}{6}\kappa_3(x^2 - 1)\phi(x), \qquad (30)$$

$$R_2(x) = -x\left[\frac{1}{24}\kappa_4(x^2 - 3) + \frac{1}{72}\kappa_3^2(x^4 - 10x^2 + 15)\right] \qquad (31)$$

Cornish-Fisher expansion - IV

Owen (1990) expands the likelihood ratio (of the mean parameter):

$$-2\log R = t_\mu^2 + 2n^{-1/2}t_\mu^3\gamma_\mu/3 + o(n^{-1/2}) \qquad (32)$$

and concludes that the Cornish-Fisher expansion

$$CF(\sqrt{-2\log R}) = Z_1 - n^{-1/2}\gamma/6 - n^{-1/2}AZ_1Z_2 + o(n^{-1/2}),$$
$$Z_1, Z_2 \sim \text{i.i.d. } N(0,1)$$

indicates bias in the empirical likelihood ratio. The bias disappears for the two-sided CIs though. *C.f.* expansion for the Student's t:

$$CF(t) = Z_1 - n^{-1/2}\gamma/6 - n^{-1/2}\gamma Z_1^2/3 - n^{-1/2}AZ_1Z_2 + o(n^{-1/2}) \qquad (33)$$

Further Corrections - I

Empirical likelihood per se: corrects skewness; central CIs have coverage errors of $O(n^{-1})$, one-sided CIs, $O(n^{-1/2})$.

Hall (1990):

Bartlett correction reduces the central coverage errors to $O(n^{-2})$;

location scale modification reduces the one-sided coverage errors to $O(n^{-1})$;

location adjustment of order $O(n^{-1})$ makes the empirical likelihood confidence regions second order correct. That is, without the adjustment, they are of the first-order correct size, shape and orientation.

Corrections - II

If $\hat{\theta}$ is an estimator of the parameter θ_0 with $Q = \lim \text{Cov}[n^{1/2}\hat{\theta}]$, where $\hat{Q}$ is an estimate of the asymptotic covariance matrix (bootstrap, jackknife, unknowns replaced by the sample estimates), then the density of $\hat{\eta}_0 = \hat{Q}^{-1/2}(\hat{\theta} - \theta_0)$ can be approximated by $N(0, I)$. Hall (1990) shows that the empirical likelihood regions are rather based on $\hat{\xi}_0 = (Q^{1/2}\hat{Q}^{-1}Q^{1/2})^{1/2}Q^{-1/2}(\hat{\theta} - \theta_0)$. They agree to the second order to the pseudo-likelihood contours based on $\hat{\xi}_0 + n^{-1}\psi$ for a certain fixed ψ.

Corrections - III

(i) $\mu_0 = E(X)$ (assume 0), $\text{Cov}\, X = \Sigma_0$ (assume I), $\theta : \mathbb{R}^r \to \mathbb{R}^s$, so $\theta = \theta(\mu)$

(ii) $\Theta = \partial\theta/\partial\mu|_{\mu=\mu_0}$, $Q = \Theta\Sigma_0\Theta^T$, $\hat{Q} = \hat{\Theta}\hat{\Sigma}_0\hat{\Theta}^T$, $R = \Theta^T Q^{-1}\Theta$,

(iii) $\alpha^{jkl} = E(X^j X^k X^l)$, $\psi^u = \sum_{j,k}\Big[-\frac{1}{2}(Q^{-\frac{1}{2}}\Theta)^{uj}R^{kl}\alpha^{jkl} + {}$
$+\frac{1}{2}\big(Q^{-\frac{1}{2}}(\theta_{jk}R^{jk} - \theta_{jj})\big)^u - (Q^{-\frac{1}{2}}\Theta)^{uj}(\Theta^T Q^{-1}\theta_{jk})^k\Big]$

(iv) $\hat{\xi} = \hat{\xi}_0 + n^{-1}\psi \Rightarrow g(y) = \phi(y)(1 + n^{-1/2}q(y) + O(n^{-1}))$ is the density of $\hat{\xi}$; $q(y)$ is a cubic polynomial in y.

(v) $T(x) = nx^T x - n2q(x) + s\ln(2\pi/n) + O_p(n^{-1}) \Rightarrow$ $l_\theta(\theta(\nu)) = T(\hat{\xi}(\nu)) + s\ln(n/2\pi) + O_p(n^{-1})$

Corrections - IV

Mentioned earlier: first order correct coverage of the CIs:

$$\text{Prob}[\theta_0 \in \mathcal{R}(x)] = \text{Prob}[\chi^2(s) \le x] + O(n^{-1}), \qquad (34)$$

Bartlett correction: define b by

$$E[l_\theta(\mu_0)] = s(1 + n^{-1}b) + O(n^{-2}). \qquad (35)$$

If $\hat{b}$ is a $\sqrt{n}$-consistent estimator of b (say obtained by replacing unknowns by their sample analogues), then the Bartlett-corrected confidence region is

$$\mathcal{R}_B(x) = \big\{\theta(\nu) | l_\theta(\nu) \le x(1 + n^{-1}\hat{b})\big\} \qquad (36)$$

The result is the second order correct coverage

$$\text{Prob}[\theta_0 \in \mathcal{R}_B(x)] = \text{Prob}[\chi^2(s) \le x] + O(n^{-2}). \qquad (37)$$

Conclusion

- Empirical likelihood is a non-parametric method of inference on distribution functionals.
- It is applicable for many settings including means, M-estimates, moments, and smooth functions of moments (correlation, regression).
- Empirical likelihood improves upon the normal approximation by adjusting for skewness / third order properties of the population distribution.
- Further improvement of coverage is through location and / or Bartlett corrections.

References

DiCiccio, T. J., P. Hall, and J. P. Romano (1989). Comparison of Parametric and Empirical Likelihood Functions. *Biometrika*, **76** (3), 465–476.

Hall, P. (1990). Pseudo-Likelihood Theory for Empirical Likelihood. *Ann. Stat.*, **18** (1), 121–140.

Johnson, N. J. (1978). Modified t tests and confidence intervals for asymmetrical populations. *JASA*, **73**, 536–544.

Owen, A. (1988). Empirical Likelihood Ratio Confidence Intervals for a Single Functional. *Biometrika*, **75** (2), 237–249.

Owen, A. (1990). Empirical Likelihood Ratio Confidence Regions. *Ann. Stat.*, **18** (1), 90–120.

Owen, A. (1991). Empirical Likelihood for Linear Models. *Ann. Stat.*, **19** (4), 1725–1747.

References - II

Cox, D. R., and D. V. Hinkley (1974). *Theoretical Statistics*. Chapman and Hall, London.

Hall, P. (1992) *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, NY.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Univ. Press, NY.