A New Approach for Deriving Better Con dence Interval Supplemented with Point Estimators for Parameters of Some Most Used Discrete Exponential Family Distributions

Chia-Ying Wu
Adviser: Dr. Yu-Ling Tseng

Department of Applied Mathematics
National Dong Hwa University

June 28, 2011

Outline

1 Motivation

2 Binomial case

3 Other two discrete exponential family distribution: Poisson and negative binomial

Let $X_1, X_2, \ldots, X_n, \ldots$ be iid with $E(X_i) = \mu$ and $var(X_i) = \sigma^2 < \infty$.
Then $\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} N(0, \sigma^2)$, i.e $\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{D} N(0,1)$.

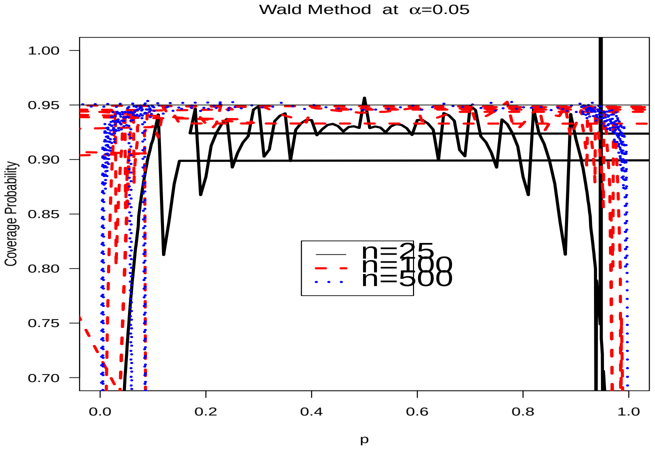
Suppose $X \sim Binomial(n, p)$ n : fixed, $0 < p < 1$

b) Con dence interval for p

Wald interval: One of the most used C.I.'s for p.

CI 362.83 57.9701 Tf 1f -.7828 Td [(W 362.53m 11-3631 Tf 1

Figure 1: Coverage probabilities of CI_W.



Q: Improve it?

Other C.I.s for p

Similar techniques used

- Wilson Score interval : It can be derived from

$$\left(\frac{\hat{p}-p}{\sqrt{p(1-p)/n}}\right)^2 = z^2_{\alpha/2}$$

A nite sample con dence interval for p.

Clopper-Pearson interval : Solving p in the equations.

$$\sum_{k=x}^{n} \binom{n}{k}$$

Figure 2: Comparison of the coverage probabilities for n=25.

Wilson Score interval

- use CLT & solve quadratic equation.
- Behaves satisfactorily except when p nears 0 or 1.

Clopper-Pearson interval

- Finite sample.
- Consistently conservative.

Wilson Score interval

- use CLT & solve quadratic equation.
- Behaves satisfactorily except when p nears 0 or 1.

Clopper-Pearson interval

- Finite sample.
- Consistently conservative.

⇒ Both of them do not o er supplemental point estimators!

Q:

Improve coverage probability?

New estimators for p, satisfactory?

d-tM

tn

Supposehthat T_n ishsequence of estimators of q , and

$$p \overline{n}(T_n$$

Figure 4: Comparison of the coverage probabilities for $n = 25$.

CI_o

- CI_o improves CI_W, especially when p nears 0 or 1.
- , $CI_o = [0, 1]$, when $\hat{p} = 0$ or 1.) meaningless and useless.

Restriction of CI_o

CI_o improves CI_W, especially when p nears 0 or 1.

BUT, $CI_o = [0, 1]$, when $\hat{p} = 0$ or 1. $\Rightarrow$ meaningless and useless.

$\Rightarrow$ modi ed CI_o

- $CI_{ZL} \Rightarrow$ conservative! Zhou, Li and Yang (2008)
- replacing $\hat{p}$ by $p = \frac{X+d}{n+2d}$, for $\frac{3}{32} \le d \le \frac{3}{16}$.

$\Rightarrow$ Our proposed con dence intervals for p:

$$CI_{TW} = \left[\frac{e^L}{1+e^L}, \frac{e^U}{1+e^U} \right] \equiv [L_{TW}, U_{TW}]$$

where $L = \text{logit}(p) - \frac{z_{\alpha/2}}{\sqrt{np(1-p)}}$ and $U = \text{logit}(p) + \frac{z_{\alpha/2}}{\sqrt{np(1-p)}}$.

Q : Comparing to others?

C.I.'s for p

Brown et. al (2001)

b) log(p) :

Pires and Amado (2008)

CI_{log} , $e^{(\log \hat{p}) \pm z_{a/2} \sqrt{\ldots}}$

The Agresti-Coull interval. Replaces $\hat{p}$ by $\tilde{p}$. Agresti and Coull (1998)

Let $\tilde{p} = \frac{\tilde{X}}{\tilde{n}}$, where $\tilde{X} = X + \frac{z^2_{a/2}}{2}$

Recommendable alternative intervals are :
CI_{TW}, CI_{CP}, $CI_{arcsin\ cc}$, CI_{AC}, CI_{WS}.

Other propositions :

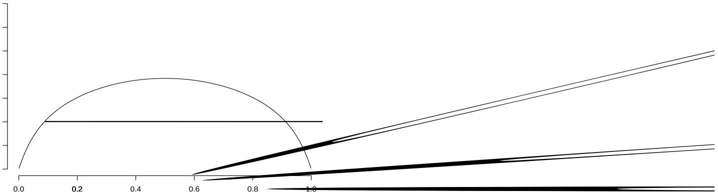
Average Width?

All of them o er supplemental point estimators!

- Take midpoint of each interval be it's supplemental point estimator.
- The midpoint of our proposed CI_{TW} : $\hat{p}_{TW}$

Compare them to $\bar{X}$ by MSE and IMSE.

Figure 7: Comparison of the average width of intervals,variance,bias and MSE of point estimators for p when n=25, a = 0.05.



Figureh8: Comparehtodi erent sample sizes at a = 0.05.

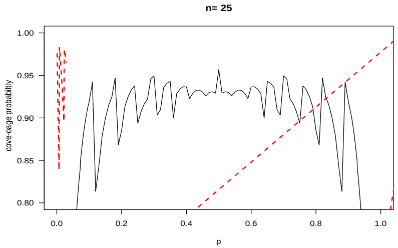


Figure 10: Compare to di erent con dence coe cients for $n = 25$.

Table 1: Comparison of IMSE of supplemented point estimators between CI_W, CI_{TW}, CI_{CP}, $CI_{arcsin\ cc}$, CI_{AC} and CI_{WS} for various a and n.

		IMSE						
a	n	CI_W	CI_{TW_1}	CI_{TW_2}	CI_{CP}	CI_{WS}	CI_{AC}	$CI_{arcsin\ cc}$
0.01	10	.0166	.0437	.0495	.0201	.0190	.0189	.0199
	25	.0066	.0173	.0229	.0074	.0077	.0077	.0071
	50	.0033	.0067	.0110	.0035	.0037	.0037	.0034
0.05	10	.0166	.0263	.0367	.0156	.0149	.0149	.0156
	25	.0066	.0089	.0140	.0064	.0064	.0064	.0064
	50	.0033	.0038	.0057	.0032	.0033	.0033	.0032
0.25	10	.0166	.0139	.0145	.0138	.0140	.0140	.0138
	25	.0066	.0061	.0062	.0061	.0062	.0062	.0061
	50	.0033	.0032	.0032	.0032	.0032	.0032	.0032

$CI_{TW} =$ h e

TW

TW

e

by straightforward but tedious derivations, we solve the inequality

$$n \quad \left(\frac{z}{B}\right)^2, \quad h(p),$$

where

$$B = \sqrt{p(1 \quad p)} \log \left(\sqrt{1 + 2E \quad \quad \quad \quad} \right)$$

$$\max h(p) = h(\tfrac{1}{2})$$

Take $n_{TW} = \lceil h(\tfrac{1}{2}) \rceil$

h E $p^2+(1 \quad p)^2$ i 2 E $p^2+(1 \quad p)^2$)

Figure 11: The function $h(p)$ when E 10% with various a.

The coverage probabilities of CI_w are too low.

The coverage probabilities of CI_{CP}, CI_{arcsin}

Figure 12: Flow chart of interval estimation for binomial case.