STAT 770 Sep. 9 Lecture Part A

LRT Theory for Multinomial Likelihoods

Reading for this lecture: Chap. 1 Agresti, plus proofs in Ch. 16, Sec. 16.3 through 16.3.4) on Asymptotics of LRT.

Material for this lecture segment:

(i) Finishing up with Confidence Intervals for unknown Binomial Proportion p

(ii) A bit of theory and computation related to the Yates Continuity Correction

(iii) General Wilks Theorem statement (about LRT) and application to Categorical Data problems

Confidence Intervals for p in Binomial Case

Wald: $\{p: \ (\widehat{p}-p)^2 \le \chi^2_{1,\alpha} \cdot \frac{\widehat{p}(1-\widehat{p})}{n}\} \ = \ \widehat{p} \pm \Phi^{-1}(1-\alpha/2)\,\frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\sqrt{n}}$

Score = Wilson, (inverted): $\{p: \ (\widehat{p}-p)^2 \le \chi^2_{1,\alpha} \cdot \frac{p(1-p)}{n}\}$

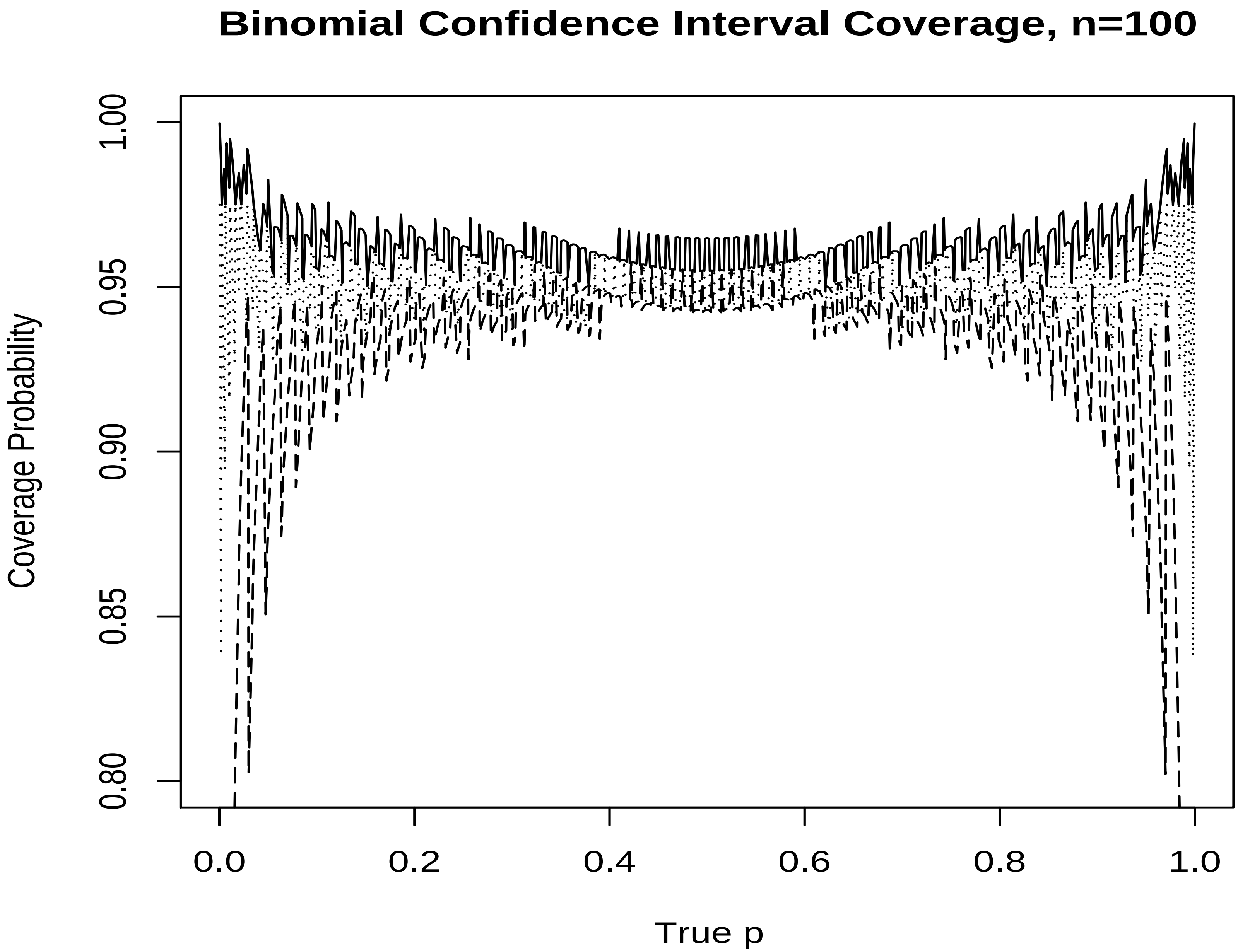
(solve quadratic in p: interval approximated by *Agresti-Coull* interval which is Wald with $n \mapsto n+2$, $N_1 \mapsto N_1+1$)

Clopper-Pearson $[p_L, p_U]$ **(which is very conservative)** inverts exact one-sided binomial tails tests such that

$$\texttt{pbinom}(N_1, n, p_U) \ = \ \alpha/2 \ = \ 1 - \texttt{pbinom}(N_1-1, n, p_L)$$

Issue in Sec. 1.4.2, 16.6.1 and Exercise (A): the common Wald CI has poor (low) coverage, bad for surprisingly large n !!

See Picture for $n = 100$ on next slide (also on Course Web-page) to compare these intervals !!



Binomial Confidence Interval Coverage, n=100
Coverage Probability
True p
1.00
0.95
0.90
0.85
0.80
0.0
0.2
0.4
0.6
0.8
1.0

Binomial Distribution, CLT & Continuity Correction

Rely on (DeMoive-Laplace) CLT to approximate

$$\text{Binom}(n,p) \approx \mathcal{N}(np, np(1-p)).$$

But for moderate n, discrete Binomial is granular:

Slud (1977, Ann. Prob.) inequality from my thesis:

for $X \sim Binomial(n,p)$, if either $k \geq np$ and $p \leq 1/4$, or $np \leq k$,

then $$1 - B(k-1, n, p) = P(X \geq k) \geq 1 - \Phi\left(\frac{k-np}{\sqrt{np(1-p)}}\right)$$

Relation to Yates correction: Binomial prob. $\texttt{dbinom}(k,n,p)$ covers interval $(k-1/2,\ k+1/2)$, so better normal approx is

$$P(X \geq k) \approx 1 - \Phi\Big((k - 0.5 - np)/\sqrt{np(1-p)}\Big) \quad \text{for} \quad k \geq np + 1$$

Mid-P Idea: instead use $\{P(X \geq k) + P(X > k)\}/2$ or approx.

Illustrating Continuity Correction with R code

```
> ApproxTab = array(c(pbinom(8:20,40,1/3),
          pnorm( (8:20 - 40/3)/sqrt(40*2/9)),
          pnorm( (8:20  - 40/3 + 0.5)/sqrt(40*2/9)),
          0.5*(pnorm((8:20-40/3)/sqrt(40*2/9)) +
                  pnorm((9:21-40/3)/sqrt(40*2/9)))),
          c(13,4), dimnames=list(8:20,
                 c("Binom","Norm","Yates","NMidP")))
> round(100*t(ApproxTab),1)
        8    9   10   11   12   13   14   15   16   17   18   19   20
Binom 4.8  9.7 17.1 27.4 39.7 53.0 65.8 76.9 85.6 91.7 95.6 97.9 99.0
Norm  3.7  7.3 13.2 21.7 32.7 45.5 58.8 71.2 81.4 89.1 94.1 97.1 98.7
Yates 5.2  9.9 17.1 26.9 39.0 52.2 65.2 76.6 85.6 91.9 95.8 98.1 99.2
NMidP 5.5 10.2 17.4 27.2 39.1 52.2 65.0 76.3 85.3 91.6 95.6 97.9 99.1
```

General Statement of Wilks Theorem

Ch. 16, Sec. 4 & handouts (3) on Web-page for proofs.

Assume data *iid* governed by model $Y_i \sim f(x, \beta)$, $\beta \in \mathcal{U} \subset \mathbb{R}^d$, f twice cont. diff. in β, with $\int \|\nabla \log f(y, \beta)\|^2 f(y, \beta) dy < \infty$ and $I(\beta) = \int \nabla\nabla' \log(f(y, \beta))\, f(y, \beta) dy < \infty$ (MLE regularity cond'ns)

Let $\hat{\beta}$ maximize $L(\beta)$ on $\mathcal{U}$, $\beta = (\gamma, \lambda)$, $\gamma \in \mathbb{R}^q$ and

$\hat{\beta}_r =$ maximizer of $L(\beta)$ on $\{(\gamma, \lambda) \in \mathcal{U} : \gamma = \gamma_0\}$ (r=restricted)

restricted model has dimension $d - q$

Likelihood Ratio test statistic under hypothesis $\mathbf{H}_0 : \gamma = \gamma_0$,

$$\Lambda = -2 \log \left(L(\hat{\beta}_r)/L(\hat{\beta}) \right) \overset{\mathcal{D}}{\to} \chi^2_{d-(d-q)} = \chi^2_q \quad \text{as} \quad n \to \infty$$

Extensions exist to independent non-i.d. data

Application to Contingency Table Setting

Recall: $Y_a = (Z_a, X_a)$ Multinomial with probabilities $p_{z,c}$

$\theta = \{p_{z,c} : \ (z,c) \in \mathcal{K}\}, \quad \beta = (\theta_1, \ldots, \theta_d), \ d = |\mathcal{K}| - 1$

$$L(\beta; \underline{\mathbf{Y}}) \ = \ (\text{multinom. coeff.}) \cdot \textstyle\prod_{(z,c)\in\mathcal{K}} \ p_{z,c}^{N_{z,c}}$$

Lower dimensional model $p_{z,c} = \pi_{z,c}(\gamma_0, \lambda)$ is Null Hypothesis
(Many examples will follow !)

So LRT $\Lambda = G^2 = -2 \log \left[L(\{\pi_{z,c}(\gamma_0, \hat{\lambda}_r\}) \Big/ L(\{\hat{p}_{x,c}\}) \right]$

$$= 2 \sum_{(z,c)\in\mathcal{K}} N_{z,c} \log \Big(\frac{N_{z,c}/n}{\pi_{z,c}(\gamma_0, \hat{\lambda}_r)} \Big)$$