STAT 770 Sep. 2 Lecture Part B
Forms of Tests and CIs for Multinomials

Lecture part A introduced MLEs in Binom & Multinom Examples

Recall in **Binomial Case:** $\hat{\theta} = N_1/n \overset{\mathcal{D}}{\approx} \mathcal{N}(p, \frac{p(1-p)}{n})$

Multinomial Case: $\hat{\theta} = \begin{pmatrix} \hat{\pi}_1 \\ \vdots \\ \hat{\pi}_K \end{pmatrix} = \left(\frac{1}{n}\sum_{a=1}^n I_{[X_a=k]},\ 1 \le k \le n\right)$

$\overset{\mathcal{D}}{\approx} \mathcal{N}\Big(\theta,\ n^{-1}\left\{\text{Diag}(\theta) - \theta\theta^{tr}\right\}\Big)$ by Multivariate CLT

because $E(I_{[X_a=k]}) = \pi_k$, $E(I_{[X_a=k]} I_{[X_a=j]}) = \pi_k\, I_{[k=j]}$

Three Classic Examples

(I) Binomial Proportion $X_a \sim \text{Bernoulli}(p),\ C = \{0,1\},\ \beta = p$
$\theta = (p, 1-p)$

$N_1 = \sum_{a=1}^{n} X_a \sim \text{Binom}(n,p),\quad L(\theta) \propto p^{N_1}(1-p)^{n-N_1},\quad \widehat{p} = N_1/n$

(II) Comparing Two Proportions. $X_a, Z_a \in \{0,1\}$ fixed.
$P(X_a = 1 \,|\, Z_a = z, \theta) = \pi_z,\ z = 0,1, \qquad \theta = (\pi_0, 1-\pi_0, \pi_1, 1-\pi_1)$
$L(\theta) \propto \prod_{z=0}^{1} \pi_z^{N_{z,1}}(1-\pi_z)^{N_{z,0}}$, MLEs $\widehat{\pi}_z = (N_{z,1}/N_{z,+},\ z = 0,1)$

(III) Multinomial Goodness of Fit. $X_a \in C = \{1,\ldots,K\}$
$\pi_k = P(X_a = k \,|\, \theta), \qquad \theta = (\pi_k,\ k = 1,\ldots,K), \qquad \widehat{\pi}_k = N_k/n$

Statistical tests and CIs based on $\widehat{\theta}$ and Likelihood Ratio Tests

Three Classic Forms of Hypothesis Tests

Setting: $\underline{Y} = \{(X_a, Z_a)\}$ iid. or $\{X_a\}$ indep & Z_a fixed

Model: $p(\underline{y}, \theta)$ prob. mass function, $\theta = \theta(\beta)$, $\beta \in \mathcal{U} \subset \mathbb{R}^d$

($d \leq K - 1$ in multinomial examples)

Null Hypothesis: $H_0 : \beta = (\gamma_0, \lambda)$, $\gamma \in \mathbb{R}^q$, $\lambda \in \mathbb{R}^{d-q}$, $q \leq d$

Wald Test based on $\widehat{\gamma} - \gamma_0$ within $\widehat{\beta} \overset{\mathcal{D}}{\approx} \mathcal{N}(\beta, V(\widehat{\beta})/n)$

Score Test ($q = d$) using $\nabla_\beta \log p(\underline{Y}, \theta_0) \overset{\mathcal{D}}{\approx} \mathcal{N}\Big(\mathbf{0},\, n\,(V(\beta_0))^{-1}\Big)$

LR Test ($q = d$) based on $\Lambda = -2 \log \Big(L(\theta_0)/L(\widehat{\theta})\Big) \overset{\mathcal{D}}{\approx} \chi^2_d$

More general $q < d$ versions of Score & LRT replace θ_0 in statistic by $\widehat{\theta}_r$, where r means restricted subject to $\gamma = \beta_0$

Tests of $H_0: \ \beta = p = p_0$ in Binomial Case

$(\ \theta = (p, 1-p),\ \beta = p\ )$

No Z_a's, $X_a \sim$ Bernoulli(p), $\hat{p} = N_1/n \sim \mathcal{N}(p,\, p(1-p)/n)$

(2-sided) Wald Test rejects when $|\hat{p} - p_0| \geq \Phi^{-1}(1-\alpha/2)\,\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$

or equivalently, when $\frac{n\,(\hat{p}-p_0)^2}{\hat{p}(1-\hat{p})} = \frac{(N_1 - np_0)^2}{N_1(1-N_1/n)} \geq \chi^2_{1,\alpha}$

Since $\nabla_p \log L(\theta, \underline{Y}) = \frac{N_1}{\theta} - \frac{n-N_1}{1-\theta} = \frac{N_1 - np}{p(1-p)} \sim \mathcal{N}\Big(0,\, \frac{n}{p(1-p)}\Big)$

Score Test rejects when $\frac{(N_1 - np_0)^2}{np_0(1-p_0)} \geq \chi^2_{1,\alpha}$

LRT rejects when $2 \cdot \log\Big(N_1 \log(\frac{\hat{p}}{p_0}) + (n - N_1) \log(\frac{1-\hat{p}}{1-p_0})\Big) \geq \chi^2_{1,\alpha}$

Confidence Intervals Related to Tests, $d = 1$

Generally with test statistic T a monotonic function of centered standardized estimate $|\tilde{\beta} - \beta_0|$ (in 2-sided testing), tests $T \geq c_\alpha$ suggest Confidence Interval: $|\tilde{\beta} - \theta_0| \leq \Phi^{-1}(1 - \alpha/2)\, \widehat{se}(\tilde{\beta})$

Inverted-Test CI defined differently, with $T = T(\beta_0)$ notationally linked to H_0 parameter, as $\{\beta \in \mathbb{R}^d : T(\beta) < c_\alpha\}$

Both CI ideas are based on reference distribution for $T(\beta)$ valid when data satisfy model with parameter β.

Need not be related to normal or symmetric d.f. , can instead use $\alpha/2$, $1 - \alpha/2$ quantiles for $\tilde{\beta}$.

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 $n \mapsto n+2$, $N_1 \mapsto N_1 + 1$)

LRT (inverted): $\left\{p: \ 2n \log\left(\widehat{p}\log(\frac{\widehat{p}}{p}) + (1-\widehat{p})\log(\frac{1-\widehat{p}}{1-p})\right) \le \chi^2_{1,\alpha}\right\}$

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 !!

Remarks about Moderate Sample-Size Behavior

(1) Part of the problem is the granularity of the discrete binomial for moderate n, as shown by the $n = 100$ picture on the Course web-page comparing coverage of the Wald, Wilson, and Clopper-Pearson CI's.

(2) Related issue is **Yates Continuity Correction:** binomial $N_1 \sim \text{Binom}(n, p)$ could be said to assign probability $\texttt{dbinom}(k, n, p)$ to interval $(k - 1/2, k + 1/2)$, so normal approx should assign $P(N_1 \geq k) \approx 1 - \Phi\Big((k - 0.5 - np)/\sqrt{np(1-p)}\Big)$ when $k \geq np + 1$.