STAT 770 Sep. 2 Lecture Part A

Theory for Multinomial Likelihoods & MLE's

Reading for today's lectures: Chap. 1 in Agresti's book, plus proofs in Ch. 16 on Asymptotics of MLE's, LRT's.

We review large-sample theory for MLEs and (in 2nd segment) LRTs for Multinomial Data and draw consequences for Tests and Confidence Intervals.

Start with Binomial & Multinomial Distributions, then review MLE theory.

From Last Lecture: Why Binomial and Multinomial ?

When Z_a are nonrandom: $N_{z,c} = \sum_{a=1}^{n} I_{[Z_a=z,\, X_a=c]} \sim \text{Binom}(n, p_{z,c})$
sum of iid binary r.v.'s, jointly distributed as Multinom$(n,\ \{p_{z,c}\}_{(z,c)})$
since each a belongs to only one $(z, c) = (Z_a, X_a)$, with prob. $p_{z,c}$.

Unconditional parameterization, where $\theta = \{p_{z,c}\}_{(z,c)}$
and $N_{z+} = \sum_{c \in C} N_{z,c}$ is a random outcome

Sometimes sample data (*stratified*) fixing $N_{z,+} \equiv n_z$, so that
$(N_{z,c},\ c \in C) \sim \text{Multinom}(n_z,\ \{p_{c|z}\}_{c \in C})$
where $p_{c|z} = P(X_a = c \,|\, Z_a = z) = p_{z,c} / \sum_{k \in C} p_{z,k}$
and $\theta = \{p_{c|z}\}_{(z,c) \in \mathcal{Z} \times C}$ is **conditional parameterization**

Basic Definitions for Likelihood and MLE

For discrete observed data $\underline{Y}$ (e.g. $= \{(X_a, Z_a)\}_{a=1}^n$ or $\{X_a\}_{a=1}^n$)

and parametric prob. mass function $p(\underline{y}, \theta) =$

$\prod_{a=1}^n P(Z_a = z_a, X_a = c_a \,|\, \theta)$ or $\prod_{a=1}^n P(X_a = c_a \,|Z_a = z_a,\, \theta)$

Likelihood Lik(θ; $\underline{Y}$) $= p(\underline{Y}, \theta)$ **as function of** θ

and **MLE** $=$ argmax$_{\theta \in \Theta}$ Lik($\theta, \underline{Y}$)

often unique, e.g. when prob. mass has exponential family form

Multinomial Likelihood

n sample size, random iid $(Z_a, X_a) \in \mathcal{Z} \times C = \mathcal{K})$

$X_a \in C$ indep., $p_{z,c} = P(Z_a = z,\, X_a = c)\,,\;\; (z,c) \in \mathcal{K}$

Reduced Data: $N_{z,c} = \sum_{a=1}^{n} I_{[Z_a=z, X_a=c]}\,,\;\; (z,c) \in \mathcal{K}$

Likelihood for (ordered) unit-level data: $\prod_{a=1}^{n} p_{z_a,c_a}^{I[Z_a=z_a, X_a=c_a]}$

Likelihood $L(\theta) \equiv L(\theta; \underline{N})$ for Multinomial $\{N_{z,c}\}$ data:

$$\binom{n}{N_{z,c},\ (z,c)\in\mathcal{K}} \prod_{(z,c)\in\mathcal{K}} p_{z,c}^{N_{z,c}} \;=\; n! \prod_{(z,c)\in\mathcal{K}} \left(p_{z,c}^{N_{z,c}} \Big/ (N_{z,c})! \right)$$

Other Forms of Same Likelihood

However the parameters $\theta = (p_{z,c},\, (z,c) \in \mathcal{K})$ are restricted, the previous 2 likelihoods are proportional, up to factors not depending on θ.

If Z_a variables are fixed along with $n_z = N_{z,+}$ and $p_{z,+}$, and X_a are indep. with $P(X_a = c \,|\, Z_a = z) = p_{c|z} = p_{z,c}/p_{z,+}$
then the Likelihood for $\theta' = \{p_{c|z}\}$ is again proportional, =

$$\prod_{z \in \mathcal{Z}} \left(n_z! \prod_{c \in C} \left[p_{c|z}^{N_{z,c}} \Big/ (N_{z,c})! \right] \right) = \prod_{z \in \mathcal{Z}} \left(\frac{n_z!}{p_{z+}^{n_z}} \prod_{c \in C} \frac{p_{z,c}^{N_{z,c}}}{(N_{z,c})!} \right)$$

Sampling Design, Conditioning & Poisson

Some extensions condition on Marginals, e.g. Fisher Exact Test fixes m_1, n_1, n in Multinomial

	X=0	1	Tot
Z=0	N_{00}	N_{01}	n_0
1	N_{10}	N_{11}	n_1
Tot	m_0	m_1	n

Useful **distributional fact**: Multinom(n, $\{p_{z,c}\}$) dist'n for $\{N_{z,c}\}$ is identical to the conditional joint distribution of independent $N_{z,c} \sim$ Poisson($\lambda p_{z,c}$) r.v.'s given $\sum_{(z,c)} N_{z,c} = n$.

(A good self-contained exercise for review, not to be handed in.)

With this fact, conditioning in multinomial-data setting can be viewed as further conditioning on indep. Poisson underlying data.

Parameter Spaces and Statistical Questions

In unconditional parameterization, Categorical Statistics is about Multinomial Data with parameters $\{p_{z,c}\}$: in interesting cases parameters are restricted/shared to reflect tabular and regression structure.

Examples: (a) $\log(p_{i,c}) = \alpha_c + \beta_i$ or $\log(p_{c|z}) = \alpha_c + \beta' z$

(b) multiway extensions, similar models with (z, c) interactions,

(c) extensions reflecting longitudinal c's, or other *link functions* relating $p_{z,c}$'s to $E(N_{z,c})$'s

Questions: Tests and Conf. Int's for parameter components, Predictons of $N_{z,c}$ (*Classification*)

Review: MLE Theory, Sec. 16.2

Collapse $(z, c) = k \in \mathcal{K}$, assume data probabilities $\theta = \{p_{z,c}\} = \{p_k\}_{k \in \mathcal{K}}$ are > 0 twice continuously differentiable functions of lower-dimensional parameter β in an open subset $\mathcal{U} \subset \mathbb{R}^d$

and $|\mathcal{K}| \times d$ matrix $\mathcal{J} \equiv (\frac{\partial \theta_k}{\partial \beta_j})$ has full rank d, **then** for large n

with probability $\to 1$ a consistent MLE $\hat{\beta}$ [a local maximizer of **locally concave**) $\log L(\, \theta(\beta)\,)$] exists and is unique on a sufficiently small neighborhood of the 'true' parameter β_0, and

$\sqrt{n}\,(\hat{\beta} - \beta_0)$ is asymptotically normally distributed with nonsingular variance matrix $V(\beta_0)$ about which we will say more later.

Note: there is a unique local solution $\tilde{\beta}$ of $\theta(\beta) = \{N_k/n\}_{k \in \mathcal{K}}$, and $\sqrt{n}(\hat{\beta} - \tilde{\beta}) \approx V \sqrt{n}\, \nabla_\beta \log L(\underline{N}/n)$ in probability.

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

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

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