(*) • Suppose y_i's are truly binomial and n_i's are relatively large — asymptotic Thms hold

↳ check * in LNp.3-1

cf. ① R^2 ② Test for lack-of-fit ($\hat{\sigma}^2$) in LM

➢ Test for goodness-of-fit (or lack-of-fit, LM, LNp.6-6~10) — $E(\chi^2_{k-s}) = k-s$, critical value is larger if s smaller

If the model S is correct (H_0), under H_0: $D_S \overset{a}{\sim} \chi^2_{k-s}$ ← asym. null dist.

Saturated model: H_1 — # of covariate classes = dim(satu. model) — # of parameters in S = dfs ↔(cf.) df of a model in LM = dim(residuals)

⇒ can use deviance to test whether the model is an adequate fit (i.e., not rejected)

Q: In LM, need model-free information of σ^2 in testing lack-of-fit. why do not need it here?

- For binomial y_i: know mean ⇒ var known ($n_i\hat{p}_i$ estimated)
- For normal y_i: know mean ⇏ σ^2 unknown ($\hat{\mu}_i$ estimated)

Note. $y_i \overset{?}{=} \hat{y}_i$'s. But, not mean S is the correct model

➢ The chi-square (null distribution) is only an approximation that becomes more accurate as the n_i's increase [often suggest $n_i \geq 5$]

H_0, $H_0 \cup H_1$

• Use deviances to compare two models S and L, where S nested in L (not necessary to be saturated)

RSS↓ cf. D↓ as adding more effects into model

➢ Larger model L: deviance D_L and df_L ($=k-l$) — L ↔ Sat.

➢ Smaller model S: deviance D_S and df_S ($=k-s$) — S ↔ Sat.

➢ To test H_0: S (say, $A\beta=c$) vs. H_1: $L \backslash S$, the LR test statistics is

$$0 \leq D_S - D_L = -2\log\left[\frac{\mathcal{L}(\hat{\beta}_S)}{\mathcal{L}(\hat{\beta}_L)}\right] = -2\log\left[\frac{\mathcal{L}(\hat{\beta}_{sat})}{\mathcal{L}(\hat{\beta}_L)} \times \frac{\mathcal{L}(\hat{\beta}_S)}{\mathcal{L}(\hat{\beta}_{sat})}\right]$$

reject if large enough; ≤ 1; LNp.3-8

which is asymptotically distributed as $\chi^2_{df_S - df_L}$ under H_0 — $l-s$

when (*) hold

• In terms of the accuracy of null distribution approximation, nested test is generally better than goodness-of-fit test

D_S: one deviance; $D_S - D_L$: 2 deviances

• (Wald test) alternative test for — convergence speed ↑ ↔ sample size ↓

S ⊂ L ⊂ Sat. (less parameters → more parameters)

H_0: $\beta_i=0$ vs. H_1: $\beta_i \neq 0$

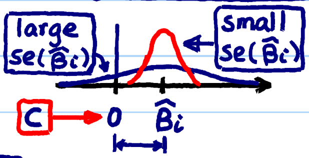
Note. Its result also depends on what other effects are in the model, e.g., testing H_0: $\beta_1=0$ when others = {x_2; x_2, x_3} could have different conclusions

ith: $A=(0,\ldots,0,1,0,\ldots,0)$

L: $X_L\beta_L = \beta_0 + \cdots + \beta_i x_i + \cdots + \beta_p x_p$
S: $X_S\beta_S = \beta_0 + \cdots + \cancel{\beta_i x_i} + \cdots + \beta_p x_p$

test statistics: $\dfrac{\hat{\beta}_i - c}{se(\hat{\beta}_i)}$ (MLE under L; $c=0$) $\equiv z$-value $\overset{a}{\sim} N(0,1)$ (under H_0)

reject if $|z|$ large; $|\cdot| = \sqrt{\text{Wald test stat.}}$; $T \sim N(0,1) \Rightarrow T^2 \sim \chi^2_1$



➢ Recall. Consider the MLE $\hat{\Theta}$ for estimating Θ

check $\widehat{cov}(\hat{\beta})$ in LNp.3-7

▪ Asymptotic normality: $\hat{\theta} \overset{a}{\sim} N(\theta, I(\theta)^{-1})$

▪ Linear transformation: $A\hat{\theta} \overset{a}{\sim} N(A\theta, A I(\theta)^{-1} A^T)$ — Fisher information matrix

▪ Wald test statistic for H_0: $A\Theta=c$ vs. H_1: $A\Theta \neq c$ is:

$$(A\hat{\theta}-c)^T [A I(\hat{\theta})^{-1} A^T]^{-1} (A\hat{\theta}-c) \overset{a}{\sim} \chi^2_{rank(A)} \quad (\text{under } H_0)$$

or replaced by a consistent est'or of $cov(\hat{\theta})$; $-1/2$, $-1/2$

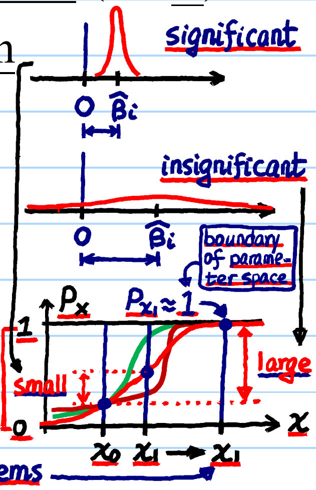
➢ Can be generalized to H_0: $\beta_i=c$ or H_0: $\beta=c$ or H_0: $A\beta=c$

➢ In contrast to normal linear model, these two test statistics (deviance-based and Wald) are *not identical*

In LM, $\dfrac{(RSS_\omega - RSS_\Omega)/(df_\omega - df_\Omega)}{RSS_\Omega/df_\Omega} \sim F_{1,df_\Omega} \xleftrightarrow{\text{equivalent}} \hat{\beta}_i/se(\hat{\beta}_i) \sim t_{df_\Omega}$; $F=t^2$

why not N(0,1)? $t_{df} \to N(0,1)$ as $df \to \infty$

- Brief comparison
  - ➢ Deviance-based test requires finding the MLEs for both models S and L, while Wald test needs just one MLE (for L)
    $\mathcal{L}(\hat{\beta}_S)$ $\mathcal{L}(\hat{\beta}_L)$ $\hat{\beta}_L$
    $\Rightarrow$ deviance-based test requires more computation
  - Wald test
  - ➢ Hauck-Donner effect (Hauck and Donner, 1977): for *sparse data* (i.e., many n_i's=1 or small), the standard errors can be *over-estimated* and so the z-value is too small than it should be and the significance of an effect could be missed
    $se(\hat{\beta}_i)$
    especially when β_i far away from 0
    $\Rightarrow$ deviance-based test is preferred unless computation is too burdensome
    has no such problem (why?) $\Rightarrow y_i \leftrightarrow \hat{y}_i$
    upper bound
    lower bound
    Check estimation problems (LN p.3-30)



- $100(1-\alpha)\%$ confidence interval for β_i
  - ➢ Duality between confidence interval and tests
    can use C.I. to do tests
    any constant
    $H_0: \beta_i = \beta_i^*$ vs. $H_1: \beta_i \neq \beta_i^*$
  - ➢ Approach 1: (from Wald test)
    acceptance region
    $$\hat{\beta}_i \pm z(\alpha/2) \times se(\hat{\beta}_i)$$
    estimate $\pm$ (critical value) $\times$ se(estimate)
    $$\left|\frac{\hat{\beta}_i - \beta_i^*}{se(\hat{\beta}_i)}\right| < z\left(\frac{\alpha}{2}\right)$$