Hypothesis Tests

- null model and saturated model
  - ➢ null model: the smallest model we will entertain
    - Model for no relation between predictors and response
    - Usually, it means we fit a common mean μ for all y_x ← $y_x \sim 1$
      irrelevant to x
    - For some contingency table models, there will be additional parameters that represent row or column totals or other such constraints ⇒ null model has more than one parameter

      e.g. use Poisson to fit multinomial model. ⇒ case factor
  - ➢ saturated (full) model: the most complex model
    - Model in which data is explained exactly
    - Typically, k parameters for k data points ← distinct x_i's
    - It can be achieved by fitting a sufficiently high-order polynomial or treating quantitative predictors as qualitative predictors or adding enough interactions
    - The model tells us no more than the data itself and is usually uninformative

  - ➢ A statistical model S describes how we partition the data into systematic structure and random variation

    $y = X\beta + \varepsilon$
    $= \hat{y} + \hat{\varepsilon}$ ← random 隨
    observed; fitted ← systematic 規
    - Null model represents one extreme where the data is represented entirely as random variation
    - Saturated model represents the data as being entirely systematic
    - Model we want usually lie between these two extremes
- Deviance
  - ➢ **Q**: how to measure discrepancy between observed and fitted y?
  - ➢ Saturated model gives us a measure of how well *any* model could possibly fit ⇒ can consider the difference between the log-likelihood for the saturated and a model S of interest:

$$2\left(l(Y,\phi;Y) - l(\hat{\boldsymbol{\mu}},\phi;Y)\right)$$

LNp.7-6 ; $\hat{\boldsymbol{\mu}}$ estimated under S ; $H_0: S$, H_1: saturated ; $-2\log\Lambda$

(which has a rationale from likelihood-ratio test)
    - $l(Y,\phi;Y)$: the log-likelihood for the saturated model
    - $l(\hat{\boldsymbol{\mu}},\phi;Y)$: the log-likelihood for the model S

➢ Provided that the observations are independent and for an exponential family distribution with $a_i(\phi) = \phi/w_i$, ← LNp.7-1

$$2(l(Y,\phi;Y) - l(\hat{\boldsymbol{\mu}},\phi;Y)) = \leftarrow \text{LNp.7-6}$$

$$\sum_i 2w_i \left[y_i \left(\tilde{\theta}_{\mathbf{x}_i} - \hat{\theta}_{\mathbf{x}_i} \right) - b(\tilde{\theta}_{\mathbf{x}_i}) + b(\hat{\theta}_{\mathbf{x}_i}) \right] \Big/ \phi \equiv D(Y,\hat{\boldsymbol{\mu}})/\phi$$

where $\tilde{\theta}_{\mathbf{x}}$: the estimates of $\theta_{\mathbf{x}}$ under the saturated model

$\hat{\theta}_{\mathbf{x}}$: the estimates of $\theta_{\mathbf{x}}$ under S ← MLE

➢ $D(Y,\hat{\boldsymbol{\mu}})$ is called the *deviance* and $D(Y,\hat{\boldsymbol{\mu}})/\hat{\phi}$ is called the *scaled deviance* For binomial & Poisson GLMs, deviance = scaled deviance.

➢ Deviance for the common GLM

LNp.7-2. $\theta_x = \log(\mu_x)$ Saturated $\hat{\mu} = y$ $b(\theta) = e^{\theta} = \mu$

Family	deviance
Gaussian	$\sum_i (y_i - \hat{\mu}_{\mathbf{x}_i})^2 \longrightarrow RSS_S$
Poisson	$2\sum_i [y_i \log(y_i/\hat{\mu}_{\mathbf{x}_i}) - (y_i - \hat{\mu}_{\mathbf{x}_i})]$
Binomial	$2\sum_i [y_i \log(y_i/\hat{\mu}_{\mathbf{x}_i}) + (n - y_i)\log((n-y_i)/(n-\hat{\mu}_{\mathbf{x}_i}))]$
Gamma	$2\sum_i [-\log(y_i/\hat{\mu}_{\mathbf{x}_i}) + (y_i - \hat{\mu}_{\mathbf{x}_i})/\hat{\mu}_{\mathbf{x}_i}]$
Inverse Gaussian	$\sum_i (y_i - \hat{\mu}_{\mathbf{x}_i})^2/(\mu_{\mathbf{x}_i}^2 y_i)$

➢ Pearson's X^2 statistic

$$X^2 = \sum_i \frac{(y_i - \hat{\mu}_{\mathbf{x}_i})^2}{\underbrace{V}_{\text{variance function (LNp.7-5)}}(\hat{\mu}_{\mathbf{x}_i})} \to \text{square of scaled residual}$$

it is an alternative measure of discrepancy that is sometimes used in replace of the deviance

- Goodness-of-fit test: whether the current model S fit the data

$\hat{\mu} \longleftrightarrow y$ close enough?

➢ Given the model S is correct,

- $D(Y,\hat{\boldsymbol{\mu}}_S)/\phi \overset{a}{\sim} \chi^2_{df_S}$ ← k − (# of parameters in S)
- $X^2 \overset{a}{\sim} \chi^2_{df_S}$

➢ For Gaussian GLM, cannot use the test because do not know the value of the dispersion parameter ϕ → σ^2

(**Q**: why not replace ϕ by an estimate $\hat{\phi}$?) ← L = saturated model for saturated model $y = \hat{y} = \hat{\mu}$

➢ For the binomial and the Poisson, ϕ=1, so the test is practical

- Difference-in-(scaled) deviance test: compare two nested models $S \subset L$ (L not saturated)

➢ Given the model S is correct,

$$(D(Y,\hat{\boldsymbol{\mu}}_S) - D(Y,\hat{\boldsymbol{\mu}}_L))/\phi \overset{a}{\sim} \chi^2_{df_S - df_L}$$

- For the Gaussian model and other models where the dispersion ϕ is *not known*, this chi-square test cannot be directly used

ϕ is a parameter

  - We can insert an estimate of ϕ and

$$\frac{D(Y,\hat{\boldsymbol{\mu}}_S) - D(Y,\hat{\boldsymbol{\mu}}_L)}{\hat{\phi}_L} \overset{a}{\sim} F_{df_S - df_L, df_L}$$

where $\hat{\phi}_L = X_L^2/df_L$ and X_L^2 is the Pearson's X^2 statistic under the model L

$\hat{\sigma}_L^2$

  - For the Gaussian model, $\hat{\phi}_L = RSS_L/df_L$ and the resulting F-statistic has an exact F distribution under the model S
- Goodness-of-fit test: L=saturated model

• Notes:

- The null distribution in the goodness-of-fit and difference-in-deviance test is only asymptotically correct
- The approximation is better when comparing models than for the goodness of fit statistic

binomial, Poisson — overdispersion.

• Wald test for individual β_j:

  - $\hat{\beta}_j/se(\hat{\beta}_j) \overset{a}{\sim} N(0,1)$
  - $\hat{\beta}_j/(se(\hat{\beta}_j)\sqrt{\hat{\phi}}) \overset{a}{\sim} t_{df_L}$

• variable selection

$y \sim 1$ ... saturated

- stepwise methods

c.f. t-test in LM.

  - Can sequentially (forward or backward or a mix of both) apply difference-in-deviance test to compare nested models (in much the same manner as in standard regression models)
  - The usual concerns about the validity of multiple testing and missing good model carry over
- criterion-based methods

likelihood based method.

$$\text{AIC}_S = \text{Deviance}_S + 2p$$

(c.f. in LM, AIC = RSS + 2p)

$$\text{BIC}_S = \text{Deviance}_S + p\log(k)$$

grouped data / ungrouped data

where p is the number of parameters in the model S and k is the number of covariate classes

  - Choose the model with the smallest AIC or BIC
  - AIC will tend to pick a larger model than the BIC
  - AIC and BIC can be used to compare non-nested model

c.f.