limit

$\boxed{0 \to 1}$

PSYCH 205 Final Exam

Likelihood $L = \prod_{i=1}^{n} p(y_i | M)$

- probability of observing the data given the model

for lm/glm: $L_m = \prod_{i=1}^{n} p(y_i | \hat{y}_i)$ or $\prod_{i=1}^{n} p(y_i | M(y | X_i))$] in case of lm/glm

probability of observing the response data y_i given predictors X_i and the model that predicts the mean of ~~y~~ given x.

$[p(y|\hat{y}) \sim N(\hat{y}, \sigma_\epsilon^2)]$ ⟺ has the following distribution

---

Deviance

$\boxed{0 \to \infty}$ limit

$$D = 2(\underbrace{\text{Log}_{LS}}_{\text{saturated model}} - \underbrace{\text{Log}_{LM}}_{\text{our model}})$$

(Limits) deviance is b/w $0 \to \infty$

- measure of goodness of fit lesser
- less deviance = less error

Likelihood Ratio Test

- compares goodness of fit of 2 nested models, ~~when 1~~ ~~...~~
- "how much more likely are the data under 1 model vs. the other?"

examples: anything w/ t, F, or χ^2 stat

- t-test
- ANOVA
- correlation

$\boxed{\text{NORMAL ~~UNIVAR~~ EQ FOR UNI + MULTIVARIATE}}$

UNIVARIATE

$$\boxed{P(y) = \frac{1}{\sqrt{2\pi}\,\sigma_y} e^{\frac{-\frac{1}{2}(y-\mu_y)^2}{\sigma_y^2}}}$$

MULTIVARIATE

$$\boxed{P(\vec{y}) = \frac{1}{\sqrt{2\pi}\,\det|C_y|^{1/2}} e^{-1/2\,\vec{y}^T C^{-1} \vec{y}}}$$

$\boxed{\text{COVARIANCE / VARIANCE MATRIX}}$

X = redundant info

$$\begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \end{array}
\begin{array}{c}
\begin{array}{ccc} X_1 & X_2 & X_3 \end{array} \\
\begin{pmatrix}
\sigma_{x_1}^2 & r_{12}\cancel{\sigma_{x_1}\sigma_{x_2}} & r_{13}\cancel{\sigma_{x_1}\sigma_{x_3}} \\
r_{12}\sigma_{x_1}\sigma_{x_2} & \sigma_{x_2}^2 & r_{23}\cancel{\sigma_{x_2}\sigma_{x_3}} \\
r_{13}\sigma_{x_1}\sigma_{x_3} & r_{23}\sigma_{x_2}\sigma_{x_3} & \sigma_{x_3}^2
\end{pmatrix}
\end{array} = C$$

(covariance matrix)

correlation matrix

$$C / \sigma_{x_i} = r$$

$$\begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \end{array}
\begin{array}{c}
\begin{array}{ccc} X_1 & X_2 & X_3 \end{array} \\
\begin{pmatrix}
1 & \cancel{r_{12}} & \cancel{r_{13}} \\
r_{12} & 1 & \cancel{r_{23}} \\
r_{31} & r_{32} & 1
\end{pmatrix}
\end{array} = r$$

correlation matrix

Derivation of bivariate normal EQ

summary: $\log L = \text{constant} - SSE$, ∴ to max L, minimize SSE

① $$\log L = \sum_{i=1}^{n} \log(p(y_i \mid \hat{y}))$$

1. plug in def. of $p(y \mid \hat{y})$

$$p(y \mid \hat{y}) = \frac{1}{\sqrt{2\pi}\,\sigma_E} e^{-\frac{1}{2}\frac{(y_i - \hat{y}_i)^2}{\sigma_E^2}}$$

② $$\log L = \sum_{i=1}^{n} \log\left[\frac{1}{\sqrt{2\pi}\,\sigma_E} e^{-\frac{1}{2}\frac{(y_i-\hat{y}_i)^2}{\sigma_E^2}}\right]$$

2. use $\log(ab) = \log(a) + \log(b)$, ~~then $\log(e^x) = x$~~

③ $$\log L = \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}\,\sigma_E}\right) + \sum_{i=1}^{n} \log\left(e^{-\frac{1}{2}\frac{(y_i-\hat{y}_i)^2}{\sigma_E^2}}\right)$$

3. use $\log(e^x) = x$

④ $$= \sum_{i=1}^{n} \log\underbrace{\left(\frac{1}{\sqrt{2\pi}\,\sigma_E}\right)}_{\text{constant}} + \sum_{i=1}^{n}\left(-\frac{1}{2}\right)\left(\frac{(y_i-\hat{y}_i)^2}{\sigma_E^2}\right)$$

4. pull out constants from terms

⑤ $$= \underbrace{n\log\left[\frac{1}{\sqrt{2\pi}\,\sigma_E}\right] - \overbrace{\frac{1}{2\sigma_E^2}}^{\text{constant}}}_{\text{constant term}} \underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{SSE}$$

5.

SSE
↑ minimize to max $\log L$

⑥ $$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

max of $\log L$ = min of SSE

⑦ $$SSE = \sum_{i=1}^{n}(y_i - (b_0 + b_1 x_i))^2 \qquad \frac{d\,SSE}{d b_1} = 0$$

d. rescale to go through origin,

so $b_0 + b_1 x_1 = b_1 x_1$ rescale

9. $$SSE = \sum_{i=1}^{n} (y_i - b_1 x_i)^2$$ FOIL

$$= \sum_{i=1}^{n} (y_i^2 - 2 b_1 x_i y_i + b_1^2 x_i^2)$$

$$= \sum_{i=1}^{n} \frac{d}{db_1} (y_i^2 - 2 b_i x_i y_i + b_1^2 x_i^2)$$ take deriv wrt b_1

$$0 = \sum_{i=1}^{n} (-2 x_i y_i + 2 b_1 x_i^2)$$ set = 0

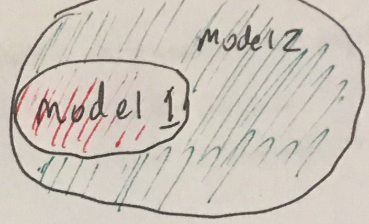
$$b_1 = \frac{\sum x_i y_i}{\sum x_i^2}$$ w/ mean = 0 solve for b_1

$$\boxed{b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}}$$ slope

$$\boxed{b_0 = \bar{y} - b_1 \bar{x}}$$ intercept

MODEL TESTING IN GLM / EXPECTATIONS UNDER NULL

for nested models



$$G = D_1 - D_2$$

dif in devainces ← G; D_1 ← deviance mod 1; D_2 ← deviance mod 2

$k_2 > k_1$

$D_2 < D_1$

→ $\boxed{\underline{\text{if}}\ M_1 \simeq M_2}$ ←

↓

[G has χ^2 distribution, with $k_2 - k_1$ df

$\chi^2(k_2 - k_1)$ = mean of G's distribution is $k_2 - k_1$]

Test Q

- if you fand deviance on cross validated data when $M_1 = M_2$, then $\underline{G = 0}$ (because neiter is actually betw

---

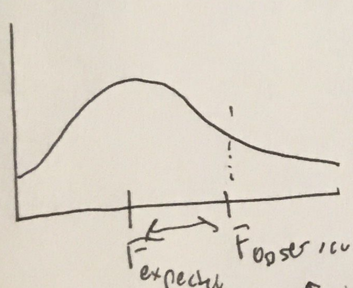
F-distribution (F)

under the null...

$$F = \frac{(SS_1 - SS_2)/(k_2 - k_1)}{SS_2/(n - k_2)}$$

where $k_2 > k_1$ and $M_1 = M_2$

(for nested models)



↑ if F observed is far from F expected: null likely is φ true

Chi-square distribution (χ^2)

- one parameter, mean (μ), which is M

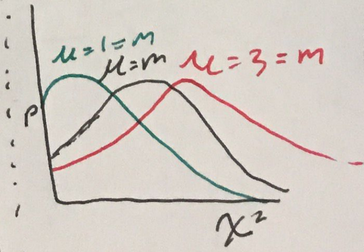
$$\chi^2 = \sum_{i=1}^{m} z_i^{2}$$

m ← sample size; z_i^2 ← square of z score

where

$Z \sim N(0, 1)$ — $\mu = 0$ = mean; $\sigma^2 = 1$ = variance

↑ z scores are normally distr. w/ mean = 0, variance = 1

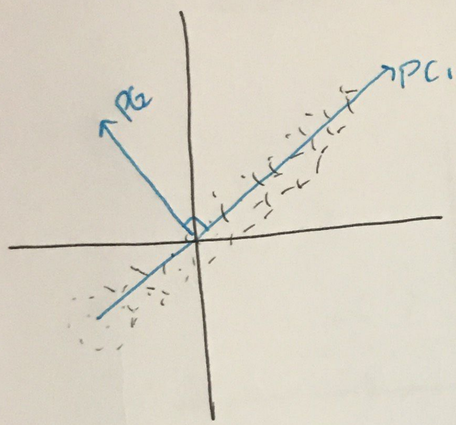


[Basically, if you are drawing residuals (z) from the null, they will follow this distribution. If they aren't, one model is better than the other! (null)]

PCA / ICA / ~~LDA~~

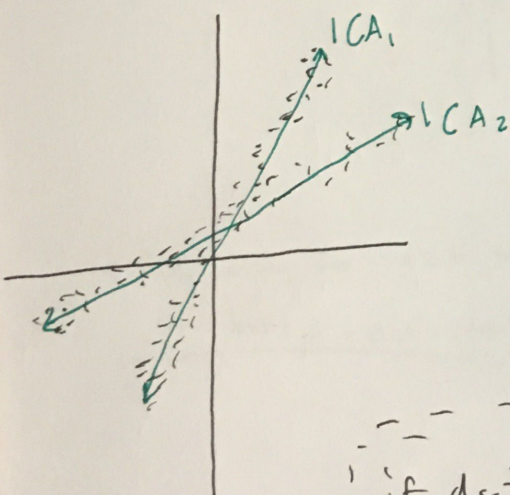
PCA
Principal component analysis

- ID orthogonal ~~independent~~ components that explain ↑ variability
- components must be orthogonal
- components are rank ordered by % of variability explained



ICA
independent components analysis

- ind. components are ∅ orthogonal
- ∅ ranked by % of variability explained
- can account for latent variables beyond just correlation
-



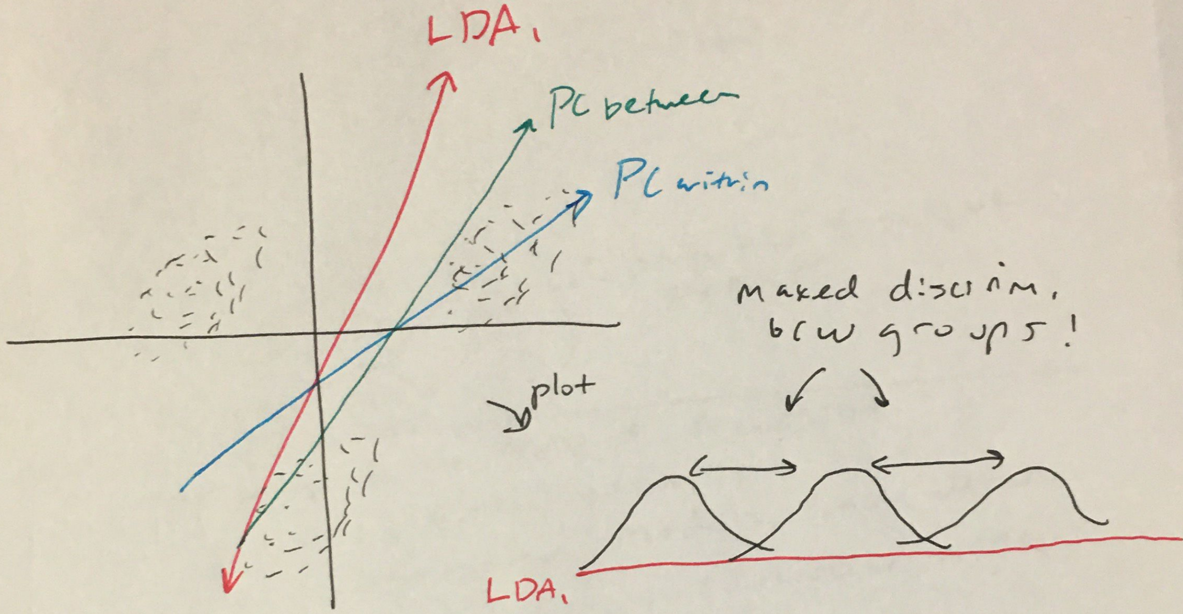
if data are normal, then

PCA = ICA

$\boxed{\text{LDA}}$ - linear discriminant analysis

- identify axes which discriminate b/w groups in data

$$\text{LDA} = \text{PCA}\left(\frac{\text{Covariance between}}{\text{Covariance within}}\right)$$



- The way to test an LDA is cross validation on held-out data

$\boxed{\text{\# of params in cov. matrix + mixture of Gaussians models!}}$

$$\text{cov of 2D} = \begin{array}{cc} & \begin{array}{cc} x_1 & x_2 \end{array} \\ \begin{array}{c} x_1 \\ x_2 \end{array} & \begin{pmatrix} 11 & \cancel{12} \\ 12 & 22 \end{pmatrix} \end{array} \quad \therefore \text{ 2x2 has 3 unique parameter}$$

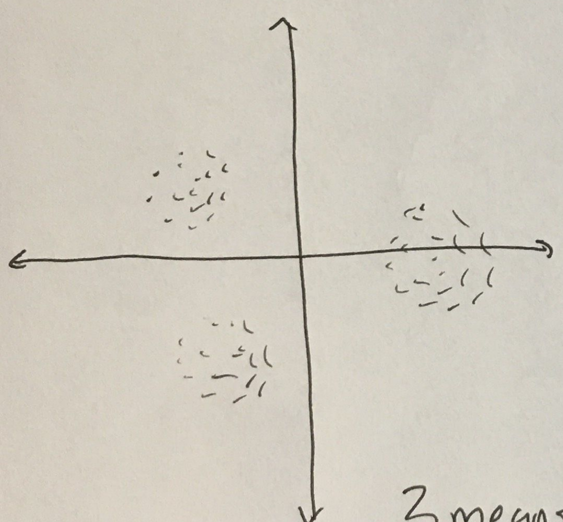
unique

$$\text{cov of 3D} = \begin{array}{cc} & \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} & \begin{pmatrix} 11 & \cancel{12} & \cancel{13} \\ 12 & 22 & \cancel{23} \\ 13 & 32 & 33 \end{pmatrix} \end{array} \quad \therefore \text{ 3x3 has 6 unique params}$$

---

applying to a mix of gaussians..

- 2 dimensions (x, y)
- 3 gaussian models, each defined by a mean + covariance matrix
- ~~how m~~
- how many unique params?



(3D×3D cov)

3 means × 2 dim. + 3 cov mat × 6 params

↓ 6 from means

↓ 18 from C matrix

$\boxed{\text{24 unique params}}$