Likelihood ratio test, Wilk's thm, Goodness of fit etc.

Simple test $H_0 : \theta = \theta_0$ $H_a : \theta = \theta_1$
[as opposed to $\theta \in$ set]

Turns out that the rejection region of most powerful test is of the form

$$\frac{P(D|\theta_1)}{P(D|\theta_0)} \geq \eta$$

[Neyman-Pearson Lemma]

One could consider alternative hypotheses to be $\theta \neq \theta_0$ [composite case]

Makes sense to choose the rejection region to be

$$\frac{P(D|\hat{\theta})}{P(D|\theta_0)} \geq \eta$$

where $\hat{\theta}$ is the MLE

We saw that for large independent sample $2 \log P(D|\hat{\theta}) / P(D|\theta_0) \sim \chi^2_k$

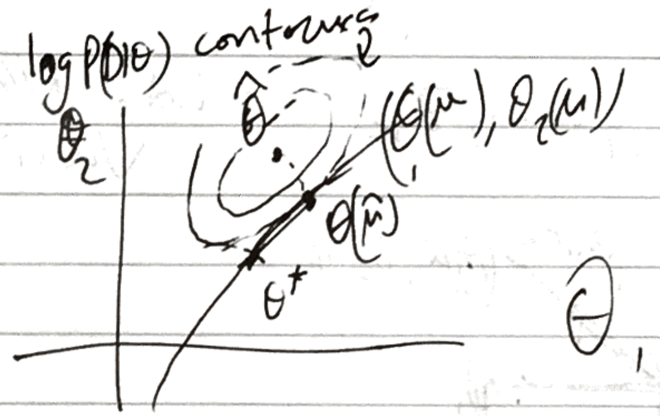
where K is the number of parameters

Interestingly, if we only optimize over a subclass of models $\theta(\mu)$ $\mu_1, [redacted] \dots, \mu_{K'}$ are parameters with $K' \leq K$

$$2\log\frac{P(D|\hat{\theta})}{P(D|\theta(\hat{\mu}))} \sim \chi^2_{K-K'}$$

← optimized over all θ

↖ " " μ

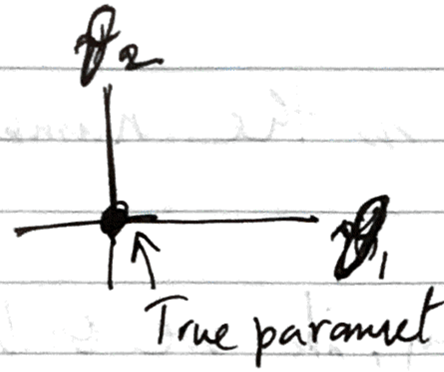


Wilk's Theorem!

Basically $\quad s(D,\theta^*) \approx \mathcal{I}_N(\theta^*)(\hat{\theta}-\theta^*)$

$$E[s(D,\theta^*)] = 0 \qquad E[s(D,\theta^*)\, s(D,\theta^*)] = \mathcal{I}_N(\theta^*)$$

Define $\quad z = I_N^{-1/2} s(D,\theta^*)$ and [redacted] $= I_N^{1/2}(\hat{\theta}-\theta^*)$



[redacted]

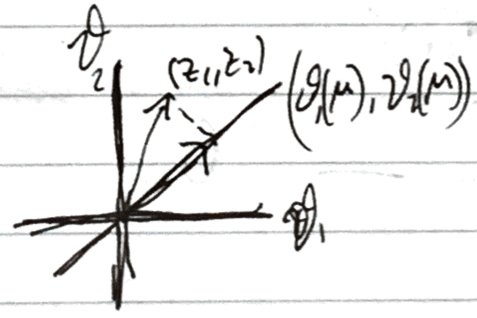
Prob distrn

$$\log \frac{P(D|\theta)}{P(D|\theta^*)} = S(D,\theta^*)(\theta-\theta^*) - \frac{1}{2}(\theta-\theta^*)^T I_N(\theta^*)(\theta-\theta^*)$$

+ higher order ter

$$= Z\cdot\vartheta - \frac{1}{2}\vartheta^2$$

$Z_1, \dots, Z_K$ i.i.d. $N(0,1)$ variable



$$Z\cdot\nabla_\mu\vartheta = [redacted]\ \hat{\vartheta}\cdot\nabla_\mu\vartheta$$

$$Z = Z_\parallel + Z_\perp \qquad Z_\perp\cdot\nabla_\mu\vartheta = 0$$

$$Z_\parallel = \hat{\vartheta}$$

Trivial example: $P(D|\theta) = C\,e^{-\sum_i (x_i-\mu_i)^2/2}$

and $\mu_i = \mu$ subspace $2\log\frac{P(D|\hat{\theta})}{P(D|\theta(\hat{\mu}))} = 2\left(\log\frac{1}{C} - \sum(x_i-\bar{x})\right)^2 = \sum_i (x_i-\bar{x})^2$

Goodness of fit

If we have prob $\theta_1 \dots \theta_k$ in a multinomial distribution, and N ~~obs~~ observations, we expect $E_i = N\theta_i$ observations in the ith category. If we observe only O_i [redacted], then for large N,

$$\sum_i \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_\nu$$

when $\nu = N-1$ [redacted]

Some cases ~~[redacted]~~ e_i are fit from observations O_i [redacted].

In those cases $\nu = N - \#$ indep't constraints satisfied by $O_i \dots$

$= N-1-\#$ param's estimated

For example for checking independence

	Success	Failure
cond 1	A	B
cond 2	C	D

[redacted]

d.f. $= rc-1-(r-1)-(c-1) = (r-1)(c-1)$ for 2x2 d.f. = 1

Reason: $N! \prod_{i=1}^{k} \frac{\theta_i^{O_i}}{O_i!}$

log with Stirling

$N\log N - \cancel{N} + \sum_i O_i \log \theta_i$

$- \sum_i O_i \log O_i + \underbrace{\sum \cancel{O_i}}_{N} + O(\log N))$

Write $N \log N = \sum_i O_i \log N$

$\sum_i O_i \log N\theta_i - \sum O_i \log O_i$

$= \sum_i O_i \log \frac{E_i}{O_i} + \cdots$

Call $O_i = E_i + \delta O_i$

and expand

$\sum_i (E_i + \delta O_i) \log \frac{E_i}{E_i + \delta O_i} \quad \sum_i E_i \left(1 + \frac{\delta O_i}{E_i}\right)\left(-\frac{\delta O_i}{E_i} + \frac{1}{2}\frac{\delta O_i^2}{E_i^2}\right)$

$= - \underbrace{\sum \delta O_i}_{0} - \frac{1}{2} \sum \frac{\delta O_i^2}{E_i} + \cdots$

$$\overset{\text{Adj } (K-1) \text{ param}}{} \qquad 2 \log \frac{P(D|\hat{E}_i)}{P(D|E_i(\hat{\mu}))} = [redacted] \, 2\left[O - \frac{1}{2} \sum_i \frac{\delta O_i^2}{E_i(\hat{\mu})} \right]$$

adjust k' paramete

$\hat{E}_i = O_i$ $\qquad$ $df = K - 1 - k'$

<u>Multitesting</u>, <u>Bonferroni Correction</u>, <u>FDR</u> . . .

High throuput expts $\quad P(y_i = 1 | D) > \tau$
where $\mathcal{D} = \{x_i\}_{i=1}^N$ [Ex: many gene expression modulations]

6000 yeast genes, $\alpha = 1\%$ [redacted] $\Rightarrow \simeq 60$ positives even we have just noise

Prob of 1 False Positive when there is no signal.
$1 - (1-\tau)^N$. Controlling this is too stringent:
$1-(1-\tau)^N =$ [redacted] $\alpha = 10\%$. for $N = 6000 \Rightarrow$ [redacted] $\tau = 1.76 \times 10^{-5}$

[If $\alpha = 1 - (1-\tau)^N \simeq N\tau \Rightarrow \tau = \frac{\alpha}{N} = \frac{0.1}{6000} = 1.67 \times 10^{-5}$]
Too many real signals missed!

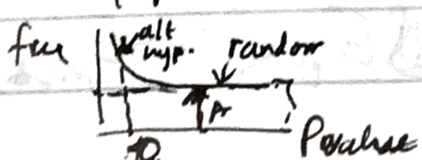
Instead, control False Discovery Rate

$$FDR = \frac{FP}{FP + TP} \leq \alpha$$

$$= \frac{FD(\tau, \mathcal{D})}{N(\tau, \mathcal{D})}$$

	Truth is	
conclusion	H_1	H_0
H_1	TP	FP
H_0	FN	TN

Adjust τ to achieve α.



$$FDR \simeq \frac{A \, P_c}{N(P < P_c)}$$