1/05/16 ALSM2 EXAM NOTES : POSTERIOR ODDS

Binomial and Posterior Odds

- Example village, dirty water, how many infected?

$n = 10$ people tested, $y = 7$ infected

$H_0 : \theta \leq 0.5$ $\quad H_1 : \theta > 0.5$

not endemic $\quad$ endemic

Binomial : $\binom{10}{7}\theta^7(1-\theta)^3 \quad \theta^*_b = y/n = 0.7$

Posterior distribution of θ given y : $P(\theta|y) = \dfrac{P(y|\theta).P(\theta)}{\int P(y|\theta)P(\theta)\,d\theta}$

"normalising constant" $p(y) \Rightarrow$

$\mathbb{R} = \int_{0.5}^{1} P(\theta|y)\,d\theta > 0.95$ then H is true

Belief ① : θ can be anywhere between 0 and 1

"Uniform probability for value of θ"

Has to integrate to 1

$$\mathbb{P} = \frac{\int_{0.5}^{1}\binom{10}{7}\theta^7(1-\theta)^3\,d\theta}{\int_0^1 \binom{10}{7}\theta^7(1-\theta)^3\,d\theta} \qquad P(\theta) = 1$$

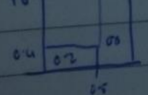
$P(\theta|y) \Rightarrow$ Checking that integration from 0.5 to 1 is 95% + to prove H_1

$\theta^7(1-\theta)^3 = \theta^7 - 3\theta^8 + 3\theta^9 - \theta^{10}$ then integrate

$$\frac{[\quad]_{0.5}^{1}}{[\quad]_0^1} = 0.88 = \theta^*$$

Reject H_1, this is not > 0.95

Belief ② expert say 80% sure infection is endemic



Compute the prior again (integral function)

$$\mathbb{P}_D = \frac{\int_{0.5}^{1} \binom{10}{7}\theta^7(1-\theta)^3(1.6)\,d\theta}{\underbrace{\int_0^{0.5} \binom{10}{7}\theta^7(1-\theta)^3(0.4)}_{<0.5} + \underbrace{\int_{0.5}^{1} \binom{10}{7}\theta^7(1-\theta)^3(1.6)}_{\geq 0.5}}$$

$$p(\theta) = \begin{cases} 1.6 & \geq 0.5 \\ 0.4 & < 0.5 \end{cases}$$

- Check if this is ≥ 0.95 to confirm H.
- Result is 0.9691
- Expert saves money by not needing extra tests or larger n size to gain significant results

05/16 ALSM 2 EXAM NOTES AIC-DEVIANCE

AIC

- A measure of goodness of fit defined by: $AIC = -2 \log L(\hat{\beta}) + 2p$
  - $p = \dim(\hat{\beta})$ = # of parameters estimated by model
  - $\hat{\beta}$ are the estimated parameters that maximise the likelihood or log likelihood
  - $\log L(\hat{\beta})$ is the maximum value of the log likelihood
- Best model is a trade off between one that maximises the likelihood with also having the minimum number of parameters
- Select model with lowest AIC

Deviance

Null Deviance = $2(\log L(\text{Saturated Model}) - \log L(\text{Null Model}))$

Residual Deviance = $2(\log L(\text{Saturated Model}) - \log L(\text{Proposed Model}))$

- Saturated model assumes each data point has it's own parameter (n parameters)
- Null model assumes 1 parameter
- Proposed model assumes you can explain your data point with p parameters plus an intercept = p+1
- Deviance is a measure of goodness of fit - lower numbers indicate a better fit
- Or rather a measure of badness of fit, higher numbers indicate worse model

observations - # predictors

- If model is good, it is approximately χ^2 with $(n-m, n=0)$ DF.

$D_{cal} \sim \chi^2(n-m, n=0)$

- If $D_{cal} < D_{0.95}$ it is a good model

05/16 ALSM 2 EXAM NOTES: MULTINOMIAL DISTRIBUTION

Odds Ration Confiden.

- Use OR value and standard errors
- When comparing two groups only 1 variable is changing i.e. lake, or sex or age groups
- Use the standard error for these variables
- In alligator example lake is only variable that changes lake 1 v lake 2

S.E lake 1 and y_4 is 0.79

S.E lake 2 and y_4 is 1.2

Pool the standard errors = $\sqrt{(0.79)^2+(1.2)^2} = 1.45$

Odds ration $\pm$ 2 * 1.45 = 95% CI

If 1 is in this interval, there may be no difference between groups (lake 1, lake 2)