STAT 601 Homework 2

Yifan Zhu

January 25, 2018

These two models have nested parameter spaces. Since the reduced model is nested in the full model, so with $Y_1 \sim \text{Binom}(p_1, n_1)$, $Y_2 \sim \text{Binom}(P_2, n_2)$ independent, we have

$$Y_1 + Y_2 \sim \text{Binom}(p, n_1 + n_2) \iff Y_1 \sim \text{Binom}(p, n_1), Y_2 \sim \text{Binom}(p, n_2)$$

"$\Leftarrow$" is easy to show. The MGF of Y_1 and Y_2 are $(1 - p + pe^t)^{n_1}$ and $(1 - p + pe^t)^{n_2}$ now. So MGF of $Y_1 + Y_2$ is $(1 - p + pe^t)^{n_1 + n_2}$, which is MGF of $\text{Binom}(p, n_1 + n_2)$.

For "$\Rightarrow$", since Y_1 and Y_2 are independent random variables from $\text{Binom}(p_1, n_1)$ and $\text{Binom}(p_2, n_2)$ based on the full model, then MGF of $Y_1 + Y_2$ is $(1 - p_1 + p_1 e^t)^{n_1}(1 - p_2 + p_2 e^t)^{n_2}$. And with $Y_1 + Y_2 \sim \text{Binom}(p, n_1 + n_2)$, the MGF is $(1 - p + pe^t)^{n_1 + n_2}$, so

$$(1 - p_1 + p_1 e^t)_1^n (1 - p_2 + p_2 e^t)^{n_2} = (1 - p + pe^t)^{n_1 + n_2}, \forall t \in \mathbb{R}, \Rightarrow p_1 = p_2 = p$$

Hence the reduced model is corresponding to the parameter space $\{(p_1, p_2) : p_1 = p_2 \,\&\, p_1, p_2 \in (0, 1)\}$, which is nested in the parameter space with respect to the full model $\{(p_1, p_2) : p_1, p_2 \in (0, 1)\}$.

Suppose we have $(y_{1,i}, y_{2,i})$, $i = 1, 2, \ldots, m$ from the joint distribution, then the loglikelihood for the full model is

$$\ell_{\text{full}} = \sum_{i=1}^m y_{1,i} \log\left(\frac{p_1}{1 - p_1}\right) + \sum_{i=1}^m y_{2,i} \log\left(\frac{p_2}{1 - p_2}\right) + m(n_1 \log(1 - p_1) + n_2 \log(1 - p_2)) + \sum_{i=1}^m \left(\log\binom{n_1}{n_2} + \log\binom{n_1}{n_2}\right)$$

And the loglikelihood for the reduced model is

$$\ell_{\text{reduced}} = \sum_{i=1}^m (y_{1,i} + y_{2,i}) \log\left(\frac{p}{1 - p}\right) + m(n_1 + n_2)(1 - p) + \sum_{i=1}^m \log\binom{n_1 + n_2}{y_{i,1} + y_{i_2}}$$

The estimation of parameters will be correct for eitehr dropping the constant or keeping the constant in both loglikelihoods. However, while testing will still be correct if we drop the constant in both loglikelihoods, it will not be correct if we keep the constant. As the constant cannot be canceled out in the LRT statistic $-2(\ell_{\text{reduced}} - \ell_{\text{full}})$, so the statistic is not correct.

The problem here is that we are using the loglikehood corresponding to the distribution of $Y_1 + Y_2$ in the reduced model. And when comparing the two log likelihoods in LRT, we should use the loglikehood from the joint distribuion. In this case, loglikelihood of the joint distribution (Y_1, Y_2) should be used for both ℓ_{full} and ℓ_{reduced}.