Lecture 4 – Feb 1

back to the gender ratio of families w 12 children in Saxony

$$L(p) = \mathbb{P}(X \mid p) = \prod_{i=1}^{6115} \mathbb{P}(x_i \mid p) \tag{1}$$

$$\ell(p) = \ln[\mathbb{P}(X \mid p)] = \sum_{i=1}^{6115} \ln[\mathbb{P}(x_i \mid p)] \tag{2}$$

$$\mathbb{P}(x_i \mid p) = \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i} \tag{3}$$

If we have m_j families with j girls, then we can capitalize on the fact that every family with the same # of girls will have the same likelihood component:

$$\sum_{i=1}^{6115} \ln[\mathbb{P}(x_i \mid p)] = C + \sum_{j=0}^{12} m_j \left(j \ln[p] + (12-j)\ln[1-p]\right) \tag{4}$$

as shown previously, $\hat{p} = \frac{\bar{x}}{12}$.

For the real data, $\hat{p} \approx 5.769$, but we also noted that real data had more extreme gender ratios: fatter tailed distribution compared to the expected number.

What if p varies across families?

Let's consider saying that $p \sim \text{Beta}(\alpha, \beta)$. The Beta places probability densities over the range 0 to 1. If X follows a Beta, $\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}$, and if f is the probability density for the Beta:

$$f(p \mid \alpha, \beta) = cp^{\alpha-1}(1-p)^{\beta-1} \tag{5}$$

where c is a constant that is tedious to calculate.

Now we can integrate out over the unknown p of each family using the continuous version of the law of total probaility:

$$\mathbb{P}(x_i) = \int_0^1 f(p \mid \alpha, \beta)\mathbb{P}(x_i \mid p)dp \tag{6}$$

$$= \frac{\binom{n}{x_i} B[\alpha + x_i, \beta + n - x_i]}{B[\alpha, \beta]} \tag{7}$$

$$\tag{8}$$

where the B is the beta function:

$$B[a, b] = \int_0^1 x^{a-1}(1-x)^{b-1}dx \tag{9}$$

Contour plots of the lnL surface indicate that the MLE of α and β are probably close to each other and large. A trace plot of the lnL for the special case when $\alpha = \beta$ has a peak around 30. JKK used Mathematica and numerical optimization to find $\hat{\alpha} \approx 31$ and $\hat{\beta} \approx 34$, and the $\ln L$ improved by 41 over the model with just $\hat{p}$.