Binomial distribution gives the probability of k successes in n independent trials where a success has probability p on each trial and $k \in [0, 1, 2, \ldots n]$:

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (1)$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (2)$$

In the study design, someone looked at 15 families. Each family is assumed to be independent, and each family has 12 children. The observed data $X = [x_1, x_2, \ldots x_{15}]$ is the count of the number of boys in each family. By assuming independence across families, we can calculate the likelihood as:

$$L(p) = \mathbb{P}(X \mid p, n) = \prod_{i=1}^{15} \mathbb{P}(x_i \mid p, n) \quad (3)$$

and we know that $n = 12$ for each family, so:

$$\begin{aligned} L(p) = \mathbb{P}(X \mid p, n = 12) &= \prod_{i=1}^{15} \mathbb{P}(x_i \mid p, n = 12) && (4) \\ &= \prod_{i=1}^{15} \binom{12}{k} p^{x_i} (1-p)^{12-x_i} && (5) \\ \ln L(p) = \ell(p) &= \sum_{i=1}^{15} \ln \left[\binom{12}{k} p^{x_i} (1-p)^{12-x_i} \right] && (6) \\ &= \sum_{i=1}^{15} \left[\ln \binom{12}{k} + \ln p^{x_i} + \ln (1-p)^{12-x_i} \right] && (7) \\ &= \sum_{i=1}^{15} \left[\ln \binom{12}{k} + x_i \ln p + (12 - x_i) \ln(1-p) \right] && (8) \end{aligned}$$

If we put in different values in the range $0 < p < 1$, we get different a log-likelihood for each value of the unknown parameter p (the probability that a child is male). Data shown JKK's spreadsheet had counts of boys from the 15 families. Because there were both boys and girls in the sample, $L(p = 0) = L(p = 1) = 0$, indicating that girls-only or boys-only models are completely incapable of producing data like the observed data.

Also of interest is the fact that the first term in the summation, $\ln \binom{12}{k}$, does not depend on p at all. So if we want to find the "best" value of p, we can ignore the first term of ℓ.

The trace plot of the ℓ as a function of p showed a peak for p just a bit greater than 0.5. This is the maximum likelihood estimate, $\hat{p}$, of p.

Some substition reveals:

$$\begin{aligned} L(p) &= \sum_{i=1}^{15} \left[\ln \binom{12}{k} + x_i \ln p + (12 - x_i) \ln(1-p) \right] && (9) \\ &= K + \sum_{i=1}^{15} \left[x_i \ln p + (12 - x_i) \ln(1-p) \right] && (10) \end{aligned}$$

where K is the constant (does not change when p changes) that we ignore.

Further simplifications using commutativity of addition, and the definition of the mean:

$$\bar{x} = \left(\sum_{i}^{15} x_i\right)/15 \tag{11}$$

$$\ln L(p) = K + \sum_{i=1}^{15} \left[x_i \ln p + (12 - x_i) \ln(1-p)\right] \tag{12}$$

$$= K + 15\bar{x} \ln[p] + 15(12 - \bar{x}) \ln[1-p] \tag{13}$$

Recall from calculus that the extreme values of a function occurs when the derivative is 0 or at the boundaries of the function. So we can differentiate with respect to p, to find the MLE of p:

$$\frac{\partial \ln L(p)}{\partial p} = \frac{15\bar{x}}{p} + \frac{-15(12-\bar{x})}{1-p} \tag{14}$$

$$\text{using: } \frac{\partial a \ln f(p)}{\partial p} = \left(\frac{\partial f(p)}{\partial p}\right)\left(\frac{a}{f(p)}\right) \tag{15}$$

Our traceplot showed that the endpoints were minima, and the point with a derivative of 0 is the maximum, so:

$$\frac{\partial \ln L(p)}{\partial p} = 0 \text{ when } p = \hat{p} \tag{16}$$

$$\frac{15\bar{x}}{\hat{p}} + \frac{-15(12-\bar{x})}{1-\hat{p}} = 0 \tag{17}$$

$$\frac{15\bar{x}}{\hat{p}} = \frac{15(12-\bar{x})}{1-\hat{p}} \tag{18}$$

$$\bar{x}(1-\hat{p}) = (12-\bar{x})\hat{p} \tag{19}$$

$$\bar{x} = 12\hat{p} \tag{20}$$

$$\hat{p} = \bar{x} \tag{21}$$

In other words, the MLE of the probability of male child is just the average number of male children per family divided by the family size (12). For the data shown in lecture $\bar{x} = 6.4$, so $\hat{p}$ was just a bit above 0.5.