Handout 11: Multinomial Distribution

Recall that the binomial distribution can be thought of as doing n flips of coin that lands heads with probability p and counting the number of resulting heads. The **multinomial distribution** is a generalization of this that can be conceptualized as rolling a k-sided die n times and counting the number of times that it lands on each side. As usual, one of the hardest things is picking a good notation. Let $x_1, \ldots, x_k$ represent a specific set of counts (these are integers that sum up to n) and $p_1, \ldots, p_k$ be the probabilities of landing on each side (positive values that sum up to 1). Using the counting theorems from probability theory, we can see that the likelihood function will have the following form:

$$\mathcal{L}(p_1, \ldots, p_k; x_1, \ldots, x_k) = \frac{n!}{x_1! \cdots x_k!} \times p_1^{x_1} \cdots p_k^{x_k} = \frac{n!}{\prod_j x_j} \times \prod_j p_j^{x_j}.$$

The multinomial is very useful in statistics because we can use it to model any distribution over a set of categories. The MLE estimators for the p_j's has a very nice form:[1]

$$\hat{p}_j = \frac{x_j}{n}, \quad j \in \{1, \ldots, k\}.$$

This just says that our best guess for the probability of being in category j is equal to the proportion of the data that was observed in category j.

The interesting thing happens when we consider the likelihood-ratio test for multinomial data. Let's consider testing the null hypothesis that the true probabilities are $\tilde{p}_1, \ldots, \tilde{p}_k$.[2] This gives, since we already know the MLE, the following value for G:

$$\Lambda = -2 \cdot \log \left[\frac{\frac{n!}{\prod_j x_j} \times \prod_j \tilde{p}_j^{x_j}}{\frac{n!}{\prod_j x_j} \times \prod_j \hat{p}_j^{x_j}} \right] = -2 \cdot \log \left[\prod_j \left(\frac{\tilde{p}}{\hat{p}} \right)^{x_j} \right]$$

Now, let's convert the null-hypothesis from probabilities into expected counts: $e_i = \tilde{p} \cdot n$. Also plugging in the form of the MLE, we then have:

$$\Lambda = -2 \cdot \sum_j x_j \cdot \log(e_j / x_j).$$

This specific application of the log-likelihood ratio test is often called the **G-test**; we will often use the letter G in place of Λ for this specific version of the test.

[1] The derivation is not too tricky, but requires using a constrained optimization technique such as Lagranian multipliers, which I do not think everyone has seen. The result is very intuitive, so we will skip the proof.

[2] Our usual notation used a zero in the subscript for the parameters of the null-hypothesis, but we already have subscripts for the different probabilities, which is why I am using a tilde instead.