MMCT Mathematical background

Chris Walther Andersen

28th January 2023

* * *

1 Introduction

This document describes the mathematical background of the MMCT pyhton package, which can be used to do tests of multinomial data with monte carlo sampling. This document does not attempt to explain how to use the code. Instead look at the GitHub page for inspiration.

2 The multinomial distribution

The multinomial distribution is the natural extension of the binomial distribution, in which a random variable can fall in one of two cases with probability p and $1 - p$. In the multinomial distribution there are k different possibilities (bins) for the random variable, and accordingly each of the k bins are reached with probabilities $p_1, p_2, \ldots p_k$, such that

$$\sum_{i=1}^{k} p_i = 1. \tag{1}$$

We imagine drawing n times from probability distribution. We end up with x_1 draws in bin 1, x_2 draws in bin 2 and so on:

$$\sum_{i=1}^{k} x_i = n. \tag{2}$$

The probability density function (the probability of drawing exactly $x_1, x_2, \ldots, x_k$ given $p_1, p_2, \ldots, p_k$) is

$$f(x_1, \ldots, x_k; n, p_1, \ldots, p_k) = f(x_i; n, p_i) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}. \tag{3}$$

This probability will almost always be very small. For example, rolling a single D6 three times and getting 3, 4 and 5 eyes (in any order) is not unreasonable, however the probability is only a bit above 3%. To test whether the result is unreasonable we need to know if the observed outcome is *more* unlikely than other outcomes. For example rolling the D6 three times and getting 3 eyes each time has a probability of just 0.4%.

Bibliography