Multinomial

The multinomial is an example of a *parametric* distribution for multiple random variables. The multinomial is a gentle introduction to joint distributions. It is a extension of the binomial. In both cases, you have n independent experiments. In a binomial each outcome is a "success" or "not success". In a multinomial there can be more than two outcomes (multi). A great analogy for the multinomial is: we are going to roll an m sided dice n times. We care about reporting the number of outcomes of each side of your dice.

Here is the formal definition of the multinomial. Say you perform n independent trials of an experiment where each trial results in one of m outcomes, with respective probabilities: $p_1, p_2, \ldots, p_m$ (constrained so that $\sum_i p_i = 1$). Define X_i to be the number of trials with outcome i. A multinomial distribution is a closed form function that answers the question: What is the probability that there are c_i trials with outcome i. Mathematically:

$$\begin{aligned} P(X_1 = c_1, X_2 = c_2, \ldots, X_m = c_m) &= \binom{n}{c_1, c_2, \ldots, c_m} \cdot p_1^{c_1} \cdot p_2^{c_2} \ldots p_m^{c_m} \\ &= \binom{n}{c_1, c_2, \ldots, c_m} \cdot \prod_i p_i^{c_i} \end{aligned}$$

This is our first joint random variable model! We can express it in a card, much like we would for random variables:

Multinomial Joint Distribution

Description:	Number of outcomes of each possible outcome type in n identical, independent experiments. Each experiment can result in one of m different outcomes.
Parameters:	$p_1, \ldots, p_m$ where each $p_i \in [0, 1]$ is the probability of outcome type i in one experiment. $n \in \{0, 1, \ldots\}$, the number of experiments
Support:	$c_i \in \{0, 1, \ldots, n\}$, for each outcome i. It must be the case that $\sum_i c_i = n$
PMF equation:	$P(X_1 = c_1, X_2 = c_2, \ldots, X_m = c_m) = \binom{n}{c_1, c_2, \ldots, c_m} \prod_i p_i^{c_i}$

Examples

Standard Dice Example: A 6-sided die is rolled 7 times. What is the probability that you roll: 1 one, 1 two, 0 threes, 2 fours, 0 fives, 3 sixes (disregarding order).

$$\begin{aligned} &\mathrm{P}(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 2, X_5 = 0, X_6 = 3) \\ &\quad = \frac{7!}{2!3!}\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^0\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^0\left(\frac{1}{6}\right)^3 \\ &\quad = 420\left(\frac{1}{6}\right)^7 \end{aligned}$$

Weather Example:
Each day the weather in Bayeslandia can be {Sunny, Cloudy, Rainy} where $p_{\text{sunny}} = 0.7$, $p_{\text{cloudy}} = 0.2$ and $p_{\text{rainy}} = 0.1$. Assume each day is independent of one another. What is the probability that over the next 7 days we have 5 sunny days, 1 cloudy day and 1 rainy days?

$$\begin{aligned} &\mathrm{P}(X_{\text{sunny}} = 6, X_{\text{rainy}} = 1, X_{\text{cloudy}} = 0) \\ &\quad = \frac{7!}{5!1!1!}(0.7)^5 \cdot (0.2)^1 \cdot (0.1)^1 \\ &\quad \approx 0.14 \end{aligned}$$

How does that compare to the probability that every day is sunny?

$$\begin{aligned}
\text{P}(X_{\text{sunny}} = 7, X_{\text{rainy}} = 0, X_{\text{cloudy}} = 0) \\
&= \frac{7!}{7!1!}(0.7)^7 \cdot (0.2)^0 \cdot (0.1)^0 \\
&\approx 0.08
\end{aligned}$$

The multinomial is especially popular because of its use as a model of language. For a full example see the Federalist Paper Authorship example.

Deriving Joint Probability

A way to deeper understand the multinomial is to derive the joint probability function for a particular multinomial. Consider the multinomial from the previous example. In that multinomial with $n = 7$ outcomes where each outcome can be one of three values $\{S, C, R\}$ where S stands for Sunny, C stands for Cloudy and R stands for Rainy, and days are independent. $p_s = 0.7$, $p_c = 0.2$, $p_r = 0.1$. We are going to derive the probability that out of the $n = 7$ days, 5 are sunny, 1 is cloudy and 1 is rainy.

Like our derivation for the binomial, we are going to consider all of the possible weeks with 5 sunny days, 1 rainy day and 1 cloudy day.

```
('S', 'S', 'S', 'S', 'S', 'C', 'R')
('S', 'S', 'S', 'S', 'S', 'R', 'C')
('S', 'S', 'S', 'S', 'C', 'S', 'R')
('S', 'S', 'S', 'S', 'C', 'R', 'S')
('S', 'S', 'S', 'S', 'R', 'S', 'C')
('S', 'S', 'S', 'S', 'R', 'C', 'S')
('S', 'S', 'S', 'C', 'S', 'S', 'R')
('S', 'S', 'S', 'C', 'S', 'R', 'S')
('S', 'S', 'S', 'C', 'R', 'S', 'S')
('S', 'S', 'S', 'R', 'S', 'S', 'C')
('S', 'S', 'S', 'R', 'S', 'C', 'S')
('S', 'S', 'S', 'R', 'C', 'S', 'S')
('S', 'S', 'C', 'S', 'S', 'S', 'R')
('S', 'S', 'C', 'S', 'S', 'R', 'S')
('S', 'S', 'C', 'S', 'R', 'S', 'S')
('S', 'S', 'C', 'R', 'S', 'S', 'S')
('S', 'S', 'R', 'S', 'S', 'S', 'C')
('S', 'S', 'R', 'S', 'S', 'C', 'S')
```

First, note that each outcome for assignments to the weeks are mutually exclusive. Then note that the probability of any one outcome will be $(p_S)^5 \cdot p_C \cdot p_R$. The number of unique weeks with the chosen count of outcomes can be derived using the rule for Permutations with Indistinct Objects. There are 7 objects, 5 are indistinct from one another:

$$\binom{7}{5,1,1} = \frac{7!}{5!1!1!} = 7 * 6 = 42$$

Since the outcomes are mutually exclusive, we are going to be adding the probability of each case to itself $\frac{7!}{5!1!1!}$ times. Putting this all together we get the multinomial joint function for this particular case:

$$\begin{aligned}
\text{P}(X_{\text{sunny}} = 5, X_{\text{rainy}} = 1, X_{\text{cloudy}} = 1) \\
&= \frac{7!}{5!1!1!}(0.7)^5 \cdot (0.2)^1 \cdot (0.1)^1 \\
&\approx 0.14
\end{aligned}$$