

## 11.5: The Multinomial Distribution

### Basic Theory

#### Multinomial trials

A *multinomial trials process* is a sequence of independent, identically distributed random variables  $\mathbf{X} = (X_1, X_2, \dots)$  each taking  $k$  possible values. Thus, the multinomial trials process is a simple generalization of the Bernoulli trials process (which corresponds to  $k = 2$ ). For simplicity, we will denote the set of outcomes by  $\{1, 2, \dots, k\}$ , and we will denote the common probability density function of the trial variables by

$$p_i = \mathbb{P}(X_j = i), \quad i \in \{1, 2, \dots, k\} \quad (11.5.1)$$

Of course  $p_i > 0$  for each  $i$  and  $\sum_{i=1}^k p_i = 1$ . In statistical terms, the sequence  $\mathbf{X}$  is formed by sampling from the distribution.

As with our discussion of the binomial distribution, we are interested in the random variables that count the number of times each outcome occurred. Thus, let

$$Y_i = \# \{j \in \{1, 2, \dots, n\} : X_j = i\} = \sum_{j=1}^n \mathbf{1}(X_j = i), \quad i \in \{1, 2, \dots, k\} \quad (11.5.2)$$

Of course, these random variables also depend on the parameter  $n$  (the number of trials), but this parameter is fixed in our discussion so we suppress it to keep the notation simple. Note that  $\sum_{i=1}^k Y_i = n$  so if we know the values of  $k-1$  of the counting variables, we can find the value of the remaining variable.

Basic arguments using independence and combinatorics can be used to derive the joint, marginal, and conditional densities of the counting variables. In particular, recall the definition of the multinomial coefficient: for nonnegative integers  $(j_1, j_2, \dots, j_k)$  with  $\sum_{i=1}^k j_i = n$ ,

$$\binom{n}{j_1, j_2, \dots, j_k} = \frac{n!}{j_1! j_2! \cdots j_k!} \quad (11.5.3)$$

#### Joint Distribution

For nonnegative integers  $(j_1, j_2, \dots, j_k)$  with  $\sum_{i=1}^k j_i = n$ ,

$$\mathbb{P}(Y_1 = j_1, Y_2 = j_2, \dots, Y_k = j_k) = \binom{n}{j_1, j_2, \dots, j_k} p_1^{j_1} p_2^{j_2} \cdots p_k^{j_k} \quad (11.5.4)$$

**Proof**

By independence, any sequence of trials in which outcome  $i$  occurs exactly  $j_i$  times for  $i \in \{1, 2, \dots, k\}$  has probability  $p_1^{j_1} p_2^{j_2} \cdots p_k^{j_k}$ . The number of such sequences is the multinomial coefficient  $\binom{n}{j_1, j_2, \dots, j_k}$ . Thus, the result follows from the additive property of probability.

The distribution of  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_k)$  is called the *multinomial distribution* with parameters  $n$  and  $\mathbf{p} = (p_1, p_2, \dots, p_k)$ . We also say that  $(Y_1, Y_2, \dots, Y_{k-1})$  has this distribution (recall that the values of  $k-1$  of the counting variables determine the value of the remaining variable). Usually, it is clear from context which meaning of the term *multinomial distribution* is intended. Again, the ordinary binomial distribution corresponds to  $k = 2$ .

#### Marginal Distributions

For each  $i \in \{1, 2, \dots, k\}$ ,  $Y_i$  has the binomial distribution with parameters  $n$  and  $p_i$ :

$$\mathbb{P}(Y_i = j) = \binom{n}{j} p_i^j (1 - p_i)^{n-j}, \quad j \in \{0, 1, \dots, n\} \quad (11.5.5)$$

**Proof**

There is a simple probabilistic proof. If we think of each trial as resulting in outcome  $i$  or not, then clearly we have a sequence of  $n$  Bernoulli trials with success parameter  $p_i$ . Random variable  $Y_i$  is the number of successes in the  $n$  trials. The result could also be obtained by summing the joint probability density function in Exercise 1 over all of the other variables, but this would be much harder.

### Grouping

The multinomial distribution is preserved when the counting variables are combined. Specifically, suppose that  $(A_1, A_2, \dots, A_m)$  is a partition of the index set  $\{1, 2, \dots, k\}$  into nonempty subsets. For  $j \in \{1, 2, \dots, m\}$  let

$$Z_j = \sum_{i \in A_j} Y_i, \quad q_j = \sum_{i \in A_j} p_i \quad (11.5.6)$$

$\mathbf{Z} = (Z_1, Z_2, \dots, Z_m)$  has the multinomial distribution with parameters  $n$  and  $\mathbf{q} = (q_1, q_2, \dots, q_m)$ .

Proof

Again, there is a simple probabilistic proof. Each trial, independently of the others, results in an outcome in  $A_j$  with probability  $q_j$ . For each  $j$ ,  $Z_j$  counts the number of trials which result in an outcome in  $A_j$ . This result could also be derived from the joint probability density function in Exercise 1, but again, this would be a much harder proof.

### Conditional Distribution

The multinomial distribution is also preserved when some of the counting variables are observed. Specifically, suppose that  $(A, B)$  is a partition of the index set  $\{1, 2, \dots, k\}$  into nonempty subsets. Suppose that  $(j_i : i \in B)$  is a sequence of nonnegative integers, indexed by  $B$  such that  $j = \sum_{i \in B} j_i \leq n$ . Let  $\mathbf{p} = \sum_{i \in A} p_i$ .

The conditional distribution of  $(Y_i : i \in A)$  given  $(Y_i = j_i : i \in B)$  is multinomial with parameters  $n - j$  and  $(p_i/p : i \in A)$ .

Proof

Again, there is a simple probabilistic argument and a harder analytic argument. If we know  $Y_i = j_i$  for  $i \in B$ , then there are  $n - j$  trials remaining, each of which, independently of the others, must result in an outcome in  $A$ . The conditional probability of a trial resulting in  $i \in A$  is  $p_i/p$ .

Combinations of the basic results involving [grouping](#) and [conditioning](#) can be used to compute any marginal or conditional distributions.

### Moments

We will compute the mean and variance of each counting variable, and the covariance and correlation of each pair of variables.

For  $i \in \{1, 2, \dots, k\}$ , the mean and variance of  $Y_i$  are

1.  $\mathbb{E}(Y_i) = np_i$
2.  $\text{var}(Y_i) = np_i(1 - p_i)$

Proof

Recall that  $Y_i$  has the binomial distribution with parameters  $n$  and  $p_i$ .

For distinct  $i, j \in \{1, 2, \dots, k\}$ ,

1.  $\text{cov}(Y_i, Y_j) = -np_i p_j$
2.  $\text{cor}(Y_i, Y_j) = -\sqrt{p_i p_j / [(1 - p_i)(1 - p_j)]}$

Proof

From the bi-linearity of the covariance operator, we have

$$\text{cov}(Y_i, Y_j) = \sum_{s=1}^n \sum_{t=1}^n \text{cov}[\mathbf{1}(X_s = i), \mathbf{1}(X_t = j)] \quad (11.5.7)$$

If  $s = t$ , the covariance of the indicator variables is  $-p_i p_j$ . If  $s \neq t$  the covariance is 0 by independence. Part (b) can be obtained from part (a) using the definition of correlation and the variances of  $Y_i$  and  $Y_j$  given [above](#).

From the [last result](#), note that the number of times outcome  $i$  occurs and the number of times outcome  $j$  occurs are negatively correlated, but the correlation does not depend on  $n$ .

If  $k = 2$ , then the number of times outcome 1 occurs and the number of times outcome 2 occurs are perfectly correlated.

Proof

This follows immediately from the result above on [covariance](#) since we must have  $i = 1$  and  $j = 2$ , and  $p_2 = 1 - p_1$ . Of course we can also argue this directly since  $Y_2 = n - Y_1$ .

## Examples and Applications

In the dice experiment, select the number of aces. For each die distribution, start with a single die and add dice one at a time, noting the shape of the probability density function and the size and location of the mean/standard deviation bar. When you get to 10 dice, run the simulation 1000 times and compare the relative frequency function to the probability density function, and the empirical moments to the distribution moments.

Suppose that we throw 10 standard, fair dice. Find the probability of each of the following events:

1. Scores 1 and 6 occur once each and the other scores occur twice each.
2. Scores 2 and 4 occur 3 times each.
3. There are 4 even scores and 6 odd scores.
4. Scores 1 and 3 occur twice each given that score 2 occurs once and score 5 three times.

Answer

1. 0.00375
2. 0.0178
3. 0.205
4. 0.0879

Suppose that we roll 4 ace-six flat dice (faces 1 and 6 have probability  $\frac{1}{4}$  each; faces 2, 3, 4, and 5 have probability  $\frac{1}{8}$  each). Find the joint probability density function of the number of times each score occurs.

Answer

$$f(u, v, w, x, y, z) = \binom{4}{u, v, w, x, y, z} \left(\frac{1}{4}\right)^{u+z} \left(\frac{1}{8}\right)^{v+w+x+y} \text{ for nonnegative integers } u, v, w, x, y, z \text{ that sum to 4}$$

In the dice experiment, select 4 ace-six flats. Run the experiment 500 times and compute the joint relative frequency function of the number times each score occurs. Compare the relative frequency function to the true probability density function.

Suppose that we roll 20 ace-six flat dice. Find the covariance and correlation of the number of 1's and the number of 2's.

Answer

covariance:  $-0.625$  correlation:  $-0.0386$

In the dice experiment, select 20 ace-six flat dice. Run the experiment 500 times, updating after each run. Compute the empirical covariance and correlation of the number of 1's and the number of 2's. Compare the results with the theoretical results computed [previously](#).

This page titled [11.5: The Multinomial Distribution](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist \(Random Services\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.