5. The Multinomial Distribution

Basic Theory

Multinomial trials

A **multinomial trials process** is a sequence of independent, identically distributed random variables $\boldsymbol{X} = (X_1, X_2, ...)$ 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 $\{1, 2, ..., k\}$, and we will denote the common probability density function of the trial variables by

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

Of course $p_i > 0$ for each i and $\sum_{i=1}^{k} p_i = 1$. In statistical terms, the sequence $\boldsymbol{X}$ 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

$$Y_{n,i} = \#(\{j \in \{1, 2, ..., n\} : X_j = i\}), \quad i \in \{1, 2, ..., k\}$$

Note that $\sum_{i=1}^{k} Y_{n,i} = n$ so if we know the values of $k - 1$ of the counting variables, we can find the value of the remaining counting variable. As with any counting variable, we can express $Y_{n,i}$ as a sum of indicator variables:

1. Show that $Y_{n,i} = \sum_{j=1}^{n} \mathbf{1}(X_j = i)$.

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

$$\binom{n}{j_1, j_2, ..., j_k} = \frac{n!}{j_1! \, j_2! \cdots j_k!}$$

Joint Distribution

2. Show that for positive integers $(j_1, j_2, ..., j_k)$ with $\sum_{i=1}^{k} j_i = n$,

$$\mathbb{P}\left(Y_{n,1} = j_1,\ Y_{n,2} = j_2,\ \ldots,\ Y_{n,k} = j_k\right) = \binom{n}{j_1,\ j_2,\ \ldots,\ j_k} p_1^{\ j_1}\ p_2^{\ j_2} \cdots p_k^{\ j_k}$$

The distribution of $\boldsymbol{Y}_n = \left(Y_{n,1},\ Y_{n,2},\ \ldots,\ Y_{n,k}\right)$ is called the **multinomial distribution** with parameters n and $\boldsymbol{p} = (p_1,\ p_2,\ \ldots,\ p_k)$. We also say that $\left(Y_{n,1},\ Y_{n,2},\ \ldots,\ Y_{n,k-1}\right)$ 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

3. Show that $Y_{n,i}$ has the binomial distribution with parameters n and p_i:

$$\mathbb{P}\left(Y_{n,i} = j\right) = \binom{n}{j} p_i^{\ j} (1 - p_i)^{n-j},\ \ j \in \{0,\ 1,\ \ldots,\ n\}$$

a. Give a probabilistic proof, by defining an appropriate sequence of Bernoulli trials.
b. Give an analytic proof, using the joint probability density function.

Grouping

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

$$Z_{n,j} = \sum_{i \in A_j} Y_{n,i}, \quad q_j = \sum_{i \in A_j} p_i$$

4. Show that $\boldsymbol{Z}_n = \left(Z_{n,1},\ Z_{n,2},\ \ldots,\ Z_{n,m}\right)$ has the multinomial distribution with parameters n and $(q_1,\ q_2,\ \ldots,\ q_m)$.

a. Give a probabilistic proof, by defining an appropriate sequence of multinomial trials.
b. Give an analytic proof, using the joint probability density function.

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 $\{1,\ 2,\ \ldots,\ k\}$ into nonempty subsets. Suppose that $\{j_i : i \in B\}$ is a sequence of nonnegative integers, indexed by B such that

$$j = \sum_{i \in B} j_i \le n$$

Let

$$p = \sum_{i \in A} p_i$$

5. Show that the conditional distribution of $\{Y_{n,i} : i \in A\}$. given $\{Y_{n,i} = j_i : i \in B\}$. is multinomial with parameters $n - j$ and $\left\{\frac{p_i}{p} : i \in A\right\}$.

a. Give a probabilistic proof, by defining an appropriate sequence of Bernoulli trials.
b. Give an analytic proof, based on the joint probability density function.

Combinations of the basic results in Exercise 4 and Exercise 5 can be used to compute any marginal or conditional distributions.

Moments

We will compute the mean, variance, covariance, and correlation of the counting variables. Results from the binomial distribution and the representation in terms of indicator variables are the main tools.

6. Show that

a. $\mathbb{E}(Y_{n,i}) = n\, p_i$.
b. $\operatorname{var}(Y_{n,i}) = n\, p_i\,(1 - p_i)$

7. Show that for distinct i and j,

a. $\operatorname{cov}(Y_{n,i}, Y_{n,j}) = -n\, p_i\, p_j$.
b. $\operatorname{cor}(Y_{n,i}, Y_{n,j}) = -\sqrt{\frac{p_i\, p_j}{(1-p_i)\,(1-p_j)}}$.

From Exercise 7, 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 or k. Does this seem reasonable?

8. Use the result of Exercise 7 to show that if $k = 2$, then the number of times outcome 1 occurs and the number of times outcome 2 occurs are perfectly correlated. Does this seem reasonable?

Examples and Applications

9. 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 density function and the size and location of the mean/standard deviation bar. When you get to 10 dice, run the simulation with an update frequency of 10. Note the apparent convergence of the relative frequency function to the density function, and the empirical moments to the distribution moments.

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

a. scores 1 and 6 occur once each and the other scores occur twice each.
b. scores 2 and 4 occur 3 times each.
c. there are 4 even scores and 6 odd scores.
d. scores 1 and 3 occur twice each given that score 2 occurs once and score 5 three times.

11. 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.

12. In the dice experiment, select 4 ace-six flats. Run the experiment 500 times, updating after each run. Compute the joint relative frequency function of the number times each score occurs. Compare the relative frequency function with the true probability density function.

13. 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.

14. 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 in Problem 13.