6.3 Multinomial

Unit 6: Joint Distributions and Conditional Expectation

Adapted from Blitzstein-Hwang Chapters 7 and 9.

The Multinomial distribution is a generalization of the Binomial. Whereas the Binomial distribution counts the successes in a fixed number of trials that can only be categorized as success or failure, the Multinomial distribution keeps track of trials whose outcomes can fall into multiple categories, such as excellent, adequate, poor; or red, yellow, green, blue.

Story 6.3.1 (Multinomial distribution).

Each of n objects is independently placed into one of k categories. An object is placed into category j with probability p_j, where the p_j are nonnegative and $\sum_{j=1}^{k} p_j = 1$. Let X_1 be the number of objects in category 1, X_2 the number of objects in category 2, etc., so that $X_1 + \cdots + X_k = n$. Then $\mathbf{X} = (X_1, \ldots, X_k)$ is said to have the *Multinomial distribution* with parameters n and $\mathbf{p} = (p_1, \ldots, p_k)$. We write this as $\mathbf{X} \sim \mathrm{Mult}_k(n, \mathbf{p})$.

We call $\mathbf{X}$ a *random vector* because it is a vector of random variables. The joint PMF of $\mathbf{X}$ can be derived from the story.

Theorem 6.3.2 (Multinomial joint PMF).

If $\mathbf{X} \sim \mathrm{Mult}_k(n, \mathbf{p})$, then the joint PMF of $\mathbf{X}$ is

$$P(X_1 = n_1, \ldots, X_k = n_k) = \frac{n!}{n_1! n_2! \ldots n_k!} \cdot p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k},$$

for $n_1, \ldots, n_k$ satisfying $n_1 + \cdots + n_k = n$.

Proof

If $n_1, \ldots, n_k$ don't add up to n, then the event $\{X_1 = n_1, \ldots, X_k = n_k\}$ is impossible: every object has to go somewhere, and new objects can't appear out of nowhere. If $n_1, \ldots, n_k$ do add up to n, then any particular way of putting n_1 objects into category 1, n_2 objects into category 2, etc., has probability $p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}$, and there are

$$\frac{n!}{n_1!n_2!\dots n_k!}$$

ways to do this. So the joint PMF is as claimed.

Next let's find the marginal distribution of X_j, the jth component of $\mathbf{X}$. Were we to blindly apply the definition, we would have to sum the joint PMF over all components of $\mathbf{X}$ other than X_j. The prospect of $k-1$ summations is an unpleasant one, to say the least. Fortunately, we can avoid tedious calculations if we instead use the story of the Multinomial: X_j is the number of objects in category j, where each of the n objects independently belongs to category j with probability p_j. Define success as landing in category j. Then we just have n independent Bernoulli trials, so the marginal distribution of X_j is $\text{Bin}(n, p_j)$.

Theorem 6.3.3 (Multinomial marginals).
If $\mathbf{X} \sim \text{Mult}_k(n, \mathbf{p})$, then $X_j \sim \text{Bin}(n, p_j)$.

More generally, whenever we merge multiple categories together in a Multinomial random vector, we get another Multinomial random vector. For example, suppose we randomly sample n people in a country with 5 political parties. Let $\mathbf{X} = (X_1, \dots, X_5) \sim \text{Mult}_5(n, (p_1, \dots, p_5))$ represent the political party affiliations of the sample, i.e., let X_j be the number of people in the sample who support party j.

Suppose that parties 1 and 2 are the dominant parties, while parties 3 through 5 are minor third parties. If we decide that instead of keeping track of all 5 parties, we only want to count the number of people in party 1, party 2, or "other", then we can define a new random vector $\mathbf{Y} = (X_1, X_2, X_3 + X_4 + X_5)$, which lumps all the third parties into a single category. By the story of the Multinomial, $\mathbf{Y} \sim \text{Mult}_3(n, (p_1, p_2, p_3 + p_4 + p_5))$, which means $X_3 + X_4 + X_5 \sim \text{Bin}(n, p_3 + p_4 + p_5)$. The marginal distribution of X_j is an extreme case of lumping where the original k categories are collapsed into just two: "in category j" and "not in category j".

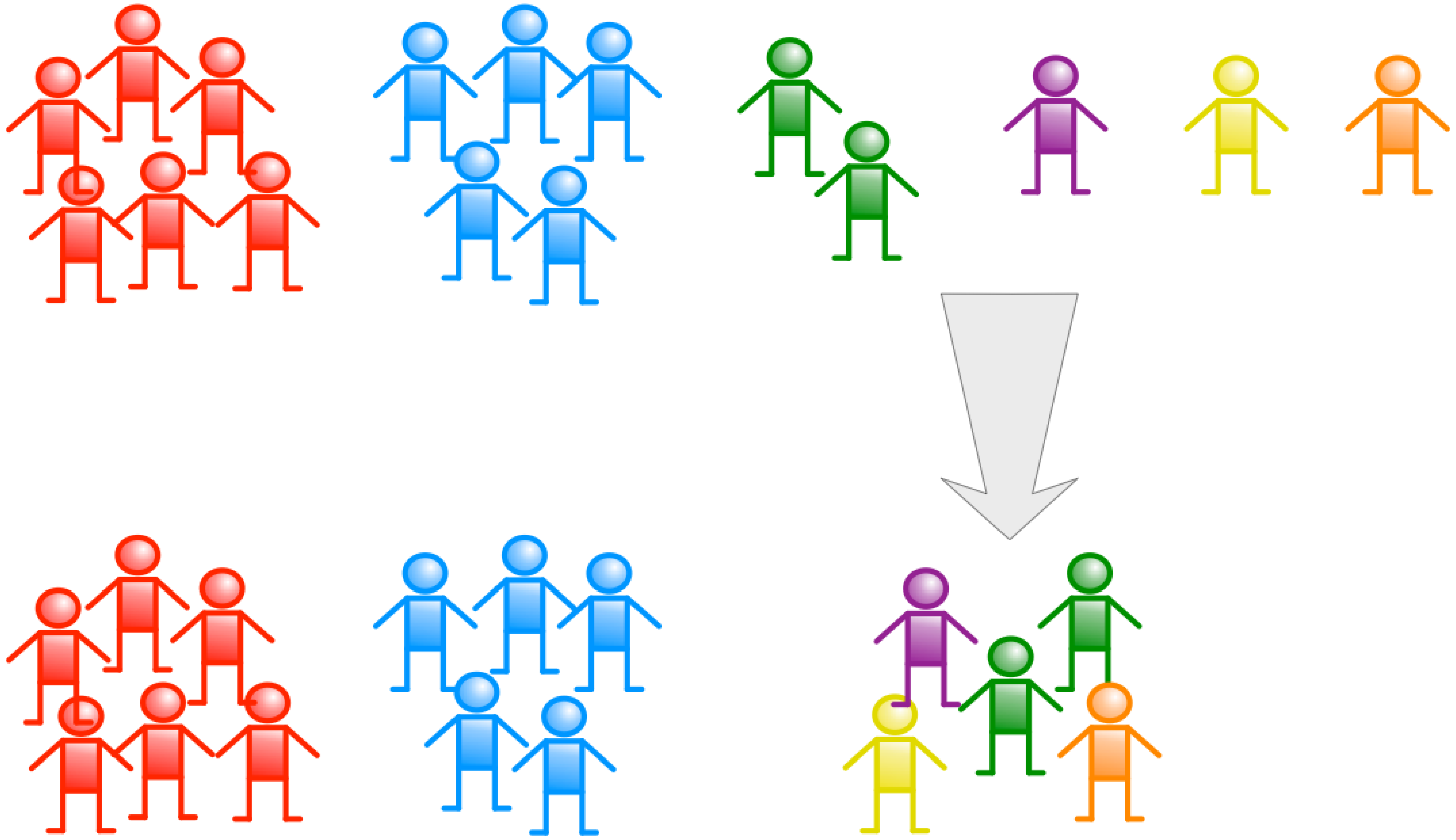

Figure 6.3.4: Lumping categories in a Multinomial random vector produces another Multinomial random vector.

View Larger Image

Image Description

This establishes another property of the Multinomial distribution:

THEOREM 6.3.5 (MULTINOMIAL LUMPING).

If $\mathbf{X} \sim \text{Mult}_k(n, \mathbf{p})$, then for any distinct i and j, $X_i + X_j \sim \text{Bin}(n, p_i + p_j)$. The random vector of counts obtained from merging categories i and j is still Multinomial. For example, merging categories 1 and 2 gives

$$(X_1 + X_2, X_3, \dots, X_k) \sim \text{Mult}_{k-1}(n, (p_1 + p_2, p_3, \dots, p_n)).$$

Finally, we know that components within a Multinomial random vector are dependent since they are constrained by $X_1 + \cdots + X_k = n$. To find the covariance between X_i and X_j, we can use the marginal and lumping properties we have just discussed.

THEOREM 6.3.6 (COVARIANCE IN A MULTINOMIAL).

Let $(X_1, \dots, X_k) \sim \text{Mult}_k(n, \mathbf{p})$, where $\mathbf{p} = (p_1, \dots, p_k)$. For $i \neq j$, $\text{Cov}(X_i, X_j) = -np_ip_j$.

Proof

For concreteness, let $i = 1$ and $j = 2$. Using the lumping property and the marginal distributions of a Multinomial, we know $X_1 + X_2 \sim \mathrm{Bin}(n, p_1 + p_2)$, $X_1 \sim \mathrm{Bin}(n, p_1)$, $X_2 \sim \mathrm{Bin}(n, p_2)$. Therefore

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2)$$

becomes

$$n(p_1 + p_2)(1 - (p_1 + p_2)) = np_1(1 - p_1) + np_2(1 - p_2) + 2\mathrm{Cov}(X_1, X_2).$$

Solving for $\mathrm{Cov}(X_1, X_2)$ gives $\mathrm{Cov}(X_1, X_2) = -np_1p_2$. By the same logic, for $i \neq j$, we have $\mathrm{Cov}(X_i, X_j) = -np_ip_j$. The components are negatively correlated, as we would expect: if we know there are a lot of objects in category i, then there aren't as many objects left over that could possibly be in category j.

Learn About Verified Certificates