The Multinomial Distribution

April 15, 2020

1 Multinomial Random Variables

Recall if we have N independent trials with only 2 outcomes for each trial (usually called successes and failures) then the number of successes has a Binomial distribution. Note that if we know the number successes then we know the number of successes.

If we instead consider N independent trials with $k > 2$ possible outcomes then we should compute the joint distribution of the number of times each of the outcomes occurred. This motivates the following definitions: a multinomial experiment, the generalization of a Bernoulli experiment, and the multinomial distribution, the generalization of the binomial distribution.

Definition 1.1. *A* ***multinomial experiment*** *consists of n independent, identical trials, with the outcome of each trial falling into one of k classes. We label the classes $1,2,...,k$. The probability that the outcome of a single trial is the i^{th} outcome is denoted p_i.*

Note that for this definition to be consistent we require $\sum_{i=1}^{k} p_i = 1$.

Definition 1.2. *The* ***multinomial distribution*** *is the joint distribution of the number times the outcome of a multinomial experiment fell into each class.*

We then say $Y_1,Y_2,...,Y_k$ are jointly multinomially distributed if their joint pmf is

$$p_{Y_1,Y_2,...,Y_k}(y_1,y_2,...,y_k) = \binom{n}{y_1,y_2,...,y_k} p_1^{y_1} p_2^{y_2} ... p_k^{y_k}$$

for some positive integer n and $p_1,p_2,...,p_k$ such that $\sum_{i=1}^{k} p_i = 1$. Recall

$$\binom{n}{y_1,y_2,...,y_k} = \frac{n!}{y_1!y_2!...y_k!}.$$

Already in Chapter 2 for a fixed $y_1,y_2,...,y_k$ we saw how to compute this value. We can now study it in more detail, because we have developed better language and more tools.

The first thing we can note with the multinomial distribution is that the marginal distribution of Y_i is Binomial(n,p_i), because if we define a success to be outcome i, then Y_i is counting the number of successes.

We can then see the Y_i's are not independent, because their joint distribution is not equal to the product of the their marginal distribution. To get a measure of their dependence we should then compute their covariance.

Example 1.3. *Let n be a positive integer n and $p_1,p_2,\ldots,p_k$ such that $\sum_{i=1}^{k} p_i = 1$. Let $Y_1,Y_2,\ldots,Y_k$ be jointly multinomial random variables with parameters n and $p_1,p_2,\ldots,p_k$.*

Compute $\mathrm{Cov}(Y_i,Y_j)$.

Solution:

Below we will assume $i \neq j$, because the case $i=j$ is just the variance of a binomial random variable.

We should try something similar to how we computed the variance of a binomial random variable in the last section, because compute $\mathbb{E}[Y_iY_j]$ directly as weighted sum of the joint pmf looks difficult.

Let for s,t between 1 and n, let A_s equal one if the s^{th} trial has outcome i and let B_t equal one if the t^{th} trial has outcome j. Then we have $Y_i = \sum_{s=1}^{n} A_s$ and $Y_j = \sum_{t=1}^{n} B_t$. To compute their covariance we use the formula we derived in the last notes

$$\mathrm{Cov}(Y_i,Y_j) = \sum_{s=1}^{n}\sum_{t=1}^{n} \mathrm{Cov}[A_s B_t]$$

If $s \neq t$ then A_s and B_t are independent because the different trials are independent. If $s=t$ then

$$\mathrm{Cov}[A_s B_s] = \mathbb{E}[A_s B_s] - \mathbb{E}[A_s]\mathbb{E}[B_s] = 0 - p_i p_j$$

where we have used $A_s B_s = 0$ (note: no expectation) because on each trial at least one of them must be zero, and that $\mathbb{E}[A_s]$ is the probability that the i^{th} trial is outcome i.

So we have

$$\mathrm{Cov}(Y_i,Y_j) = -n p_i p_j.$$

Note that the covariance is negative. This should intuitively make sense. If Y_i is large, then it is not possible to also have a large number of the j outcomes, so Y_j must be smaller. Likewise if Y_i is small, then it is more likely more of outcome were of type j.