6.3 Multinomial

Multinomial distribution is a generalization of the Binomial. The difference is:

- Binomial deals with trials of 2 categories: success or failure
- Multinomial deals with trials of multiple categories. i.e excellent, adequate, poor

Story 6.3.1 (Multinomial Distribution)

- There are n objects and k categories
- Each object belong to category j with probability p_j so that:
  - $p_j \geq 0$
  - $\sum_{j=1}^{k} p_j = 1$
- Let X_j be the number of objects belong to category j so that:

$$X_1 + \dots + X_k = n$$

Then $X = (X_1, \dots, X_k)$ is said to have Multinomial Distribution with parameters n and $p = (p_1, \dots, p_k)$.

Denoted as $X \sim \text{Mult}_k(n, p)$

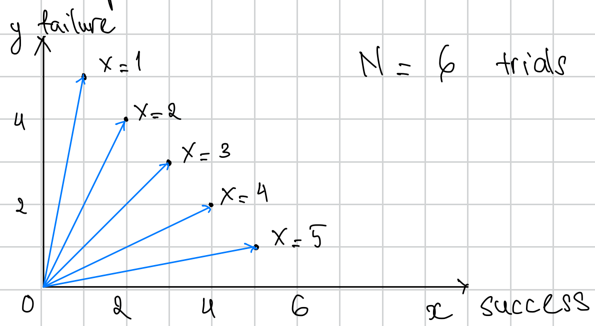
We call X an random vector because it is a vector of r.v.s

Creative way to visualize why it is a vector?

Lets consider the case of Binomial and then generalize the Multinomial case:

- We can understand r.v $X \sim \text{Bin}(n,p)$ as a collection of vectors in a 2 dimensional space since it has only 2 categories (success and failure) where each support is a point in the dimension. So:

$N = 6$ trials



- Generalize this, a r.v $X \sim \text{Mult}_k(n, p)$ can be thought of as a collection of vectors inside a k dimensional space where each support is a point in the dimension.

Theorem 6.3.2 (Multinomial joint PMF)

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

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

for $n_1, \dots, n_k$ satisfying $n_1 + \dots + n_k = n$

Proof:

- If $n_1, \dots, n_k$ don't add up to n, then the event $\{X_1 = n_1, \dots, X_k = n_k\}$ is impossible because either
  - existing object has to belong to 1 category
  - new object cannot appear out of nowhere

- If $n_1, \dots, n_k$ add up to n, then any particular way that n_1 objects belong to category 1, and n_2 objects belong to category 2, etc... has probability $p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$ and there are $\frac{n!}{n_1! \, n_2! \dots n_k!}$ ways to do this < divide for $n_j!$ since the order of objects inside the category doesn't matter >

So the PMF is $p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$

Theorem 6.3.3 (Multinomial Marginals)

$\boxed{\text{If } X \sim \text{Mult}_k(n, p) \text{ then } X_j \sim \text{Bin}(n, p_j)}$

Visualize:

If X is visualized as collection of vectors in k dimension, then X_j is a plane that is formed by squeezing other dimensions together.

Explanation:

Now let get the marginal distribution of X_j, the j^{th} component of X.

There are 2 ways to arrive at the formula:

1. Sum up all the joint PMF of X other than X_j
2. Using the story of Multinomial: X_j is number of objects belong to category j, where each of n objects independently belongs to category j with probability p_j. Define success as landing in category j, then we have n independent Bernoulli trials, so the marginal distribution of $X_j \sim \text{Bin}(n, p_j)$

Theorem 6.3.5 (Multinomial Lumping)

If $X \sim \text{Mult}_k(n, p)$, 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:

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

Example:

Suppose we randomly select n people from a country with 5 political parties. Let

- $X = (X_1, \ldots, X_5) \sim \text{Mult}_5(n, (p_1, \ldots, p_5))$ represent the political affiliations of the sample.
- X_j be number of people in party j.

- Suppose that party 1 and 2 are two dominant parties, while parties 3 through 5 are minority. And we want to lump together minority parties as "other", then:

$$Y = (X_1, X_2, X_3 + X_4 + X_5) \sim \text{Mult}_3(n, (p_1, p_2, p_3 + p_4 + p_5))$$

$$X_3 + X_4 + X_5 \sim \text{Bin}(n, p_3 + p_4 + p_5)$$

Theorem 6.3.6 (Covariance in Multinomial)

Let $(X_1, \ldots, X_k) \sim \text{Mult}_k(n, p)$, where $p = (p_1, \ldots, p_k)$.

$$\text{Cov}(X_i, X_j) = -n p_i p_j \quad , \text{ for } i \neq j$$

Proof:

- For concreteness, let $i = 1$ and $j = 2$
- Property 7 of covariance gives:

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) + 2\,\text{Cov}(X_1, X_2) \qquad (1)$$

- By lumping and marginal property of Multinomial:

$$X_1 \sim \text{Bin}(n, p_1)$$

$$X_2 \sim \text{Bin}(n, p_2)$$

$$X_1 + X_2 \sim \text{Bin}(n, p_1 + p_2)$$

- Equation (1) now becomes:

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

$$\Leftrightarrow \text{Cov}(X_1, X_2) = -n p_1 p_2$$

Conclusion:

- The components are negatively correlated, that make sense since $n_1 + n_2 = n$ means that knowing there are lots objects in category 1, then there aren't many objects in category 2.

6.4 Multivariate Normal

Definition 6.4.1 (Multivariate Normal Distribution)

- A random vector $X = (X_1, \dots, X_k)$ have a Multivariate Normal (MVN) distribution if every linear combination of the X_j has a Normal distribution. That is, we require:

$$t_1 X_1 + \dots + t_k X_k$$

to have a Normal distribution for any choice of constants $t_1, \dots, t_k$.

- If $t_1 X_1 + \dots + t_k X_k$ is a constant (such as when all $t_i = 0$), we consider it to have a Normal distribution, albeit a degenerate Normal with variance 0.
- An important special case is $k=2$, this distribution is called the Bivariate Normal (BVN).

Short definition:

The MVN is a continuous multivariate distribution that generalizes the Normal distribution into higher dimensions

The converse is false:

If $(X_1, \dots, X_k)$ is MVN, then individual X_j is Normal.

However, a bunch of Normal r.v.s $X_1, \dots, X_k$ such that $(X_1, \dots, X_k)$ is not Multivariate Normal

Visualize MVN linearly:

We can visualize $X = (X_1, \dots, X_k)$ is collection of vectors inside k-dimension space where each support is a vector, hence the linear combination

Example 6.4.2 (Non-example of MVN)

Here is an example of 2 r.v.s whose marginal distributions are Normal but whose joint distribution is not Bivariate Normal.

Let $X \sim N(0,1)$, and let $S = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$

be random sign independent of X. Then $Y = SX$ is a Standard Normal r.v, due to the symmetry of the Normal distribution.

However, (X, Y) is not Bivariate Normal because:

$$P(X + Y = 0) = P(S = -1) = \frac{1}{2}$$

< violates no point masses property of Normal distribution >

which implies that $X + Y$ can't be Normal since $X + Y$ is a linear combination of X and Y with $t_X = t_Y = 1$. So (X, Y) is not Bivariate Normal.

Example 6.4.3 (Actual MVN)

- For $Z, W \overset{\text{i.i.d}}{\sim} N(0,1)$, then:
  - (Z, W) is Bivariate Normal, because the sum of independent Normals is Normal
  - $(Z + 2W, 3Z + 5W)$ is also Bivariate Normal, since an arbitrary linear combination:

$$t_1(Z + 2W) + t_2(3Z + 5W)$$
$$= (t_1 + 3t_2) Z + (2t_1 + 5t_2) W \quad \text{< is Normal >}$$

$\Rightarrow$ This means we can take linear combinations of components of a Multivariate Normal and form another Multivariate Normal.

Theorem 6.4.4. (Subset of Multivariate Normal)

If (X_1, X_2, X_3) is Multivariate Normal, then so is subvector (X_1, X_2)

Proof:

- Since (X_1, X_2, X_3) is Multivariate Normal:

$\Rightarrow$ $(t_1 X_1 + t_2 X_2 + t_3 X_3)$ is Normal with any t_1, t_2, t_3

- Consider the case where $t_3 = 0$, then:

$(t_1 X_1 + t_2 X_2)$ is Normal

which means (X_1, X_2) is MVN.

Theorem 6.4.5 (Concatenate Multivariate Normal)

If $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_m)$ are MVN vectors with X independent of Y, then the concatenate random vector $W = (X_1, \dots, X_n, Y_1, \dots, Y_m)$ is Multivariate Normal.

Proof:

- Since X and Y are MVN:

$$\begin{cases} s_1 X_1 + \dots + s_n X_n \\ t_1 Y_1 + \dots + t_m Y_m \end{cases} \text{are Normals}$$

- Then:

$(s_1 X_1 + \dots + s_n X_n + t_1 Y_1 + \dots + t_m Y_m)$ is Normal since sum of Normals is Normal.

How to fully specify a Bivariate Normal distribution?

A Multivariate Normal distribution is fully specified by knowing 3 things:

- the mean of each component
- the variance of each component
- the covariance or correlation between any 2 components

Another way to say this is, for an MVN vector $(X_1, \dots, X_n)$ to be fully specified:

- the mean vector $(\mu_1, \dots, \mu_k)$ where $E(X_j) = \mu_j$
- the covariance matrix, which is the $k \times k$ matrix of covariances between components; row i, column j entry is $Cov(X_i, X_j)$

Example of fully specified MVN

Given MVN vector (X, Y), we need to know 5 parameters:

- the means $E(X)$, $E(Y)$
- the variances $Var(X)$, $Var(Y)$
- the correlation $Corr(X, Y)$

The joint PDF of Bivariate Normal (X, Y) with $N(0,1)$ marginals and correlation $\rho \in (-1, 1)$ is:

$$f_{X,Y}(x, y) = \frac{1}{2\pi\tau} \exp\left(-\frac{1}{2\tau^2}\left(x^2 + y^2 - 2\rho xy\right)\right)$$

with $\tau = \sqrt{1 - \rho^2}$

Theorem 6.4.7 (Special property of MVN)

Within an MVN random vector, uncorrelated implies independent.

In other words, given $X = (X_1, X_2) \sim$ MVN where X_1, X_2 are subvectors, and every components of X_1 is uncorrelated with every components of X_2, then X_1 and X_2 are independent.

In particular, if (X, Y) is Bivariate Normal and $Corr(X, Y) = 0$, then X and Y are independent.

Example 6.4.8 (Independence of sum and difference)

Let $X, Y \overset{i.i.d}{\sim} N(0,1)$. Find the joint distribution of $(X+Y, X-Y)$

Since $\begin{cases} (X+Y, X-Y) \text{ is Bivariate Normal} \\ Cov(X+Y, X-Y) = Var(X) - Cov(X,Y) + Cov(Y,X) - Var(Y) = 0 \end{cases}$

$\Rightarrow$ $(X+Y)$ is independent of $(X-Y)$

and $(X+Y), (X-Y) \overset{i.i.d}{\sim} N(0,2)$

Independence of sum and difference is a unique characteristic of the Normal.

That is, if X, Y are i.i.d and $(X+Y)$ is independent of $(X-Y)$, then X and Y must have Normal distributions (with the same variance)