5.8.1 Random Vectors (RVTRs)

We will first introduce the concept of a random vector.

**Definition 5.8.1.1: Random Vectors**

Let $X_1, ..., X_n$ be arbitrary random variables, and stack them into a vector like such:

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$$

We call $X$ an $n$-dimensional random vector (rvtr).

Random vectors will be useful for our later distributions!

We define the expectation of a random vector just as we would hope, coordinate-wise:

$$\mathbb{E}[X] = \begin{bmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{bmatrix}$$

5.8.2 Motivation for the Multinomial Distribution

Suppose we have scenario where there are $r = 3$ outcomes, with probabilities $p_1, p_2, p_3$ respectively, such that $p_1 + p_2 + p_3 = 1$. Suppose we have $n = 7$ independent trials, and let $Y = (Y_1, Y_2, Y_3)$ be the rvtr of counts of each outcome. Suppose we define each $X_i$ as a one-hot vector (exactly one 1, and the rest 0) as below, so that $Y = \sum_{i=1}^n X_i$ (this is exactly like how adding indicators/Bernoulli’s gives us a Binomial):

<table>
<thead>
<tr>
<th></th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$X_4$</th>
<th>$X_5$</th>
<th>$X_6$</th>
<th>$X_7$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome 1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Outcome 2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Outcome 3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>
Now, what is the probability of this outcome - that \( (Y_1 = 2, Y_2 = 1, Y_3 = 4) \)? We get the following:

\[
p_{Y_1,Y_2,Y_3}(2,1,4) = \frac{7!}{2!1!4!} \cdot p_1^2 \cdot p_2 \cdot p_3^4 \quad [\text{recall from counting}]
\]

\[
= \binom{7}{2,1,4} \cdot p_1^2 \cdot p_2 \cdot p_3^4
\]

### 5.8.3 The Multinomial Distribution

Now let us define the Multinomial Distribution more generally.

**Definition 5.8.3.1: The Multinomial Distribution**

Suppose there are \( r \) outcomes, with probabilities \( \mathbf{p} = (p_1, p_2, \ldots, p_r) \) respectively, such that \( \sum_{i=1}^{r} p_i = 1 \). Suppose we have \( n \) independent trials, and let \( \mathbf{Y} = (Y_1, Y_2, \ldots, Y_r) \) be the rvtr of counts of each outcome. Then, we say:

\[
\mathbf{Y} \sim \text{Mult}_r(n, \mathbf{p})
\]

The joint PMF of \( \mathbf{Y} \) is:

\[
p_{Y_1,\ldots,Y_r}(k_1,\ldots,k_r) = \binom{n}{k_1,\ldots,k_r} \prod_{i=1}^{r} p_i^{k_i}, \quad k_1,\ldots,k_r \geq 0 \quad \text{and} \quad \sum_{i=1}^{r} k_i = n
\]

Notice that each \( Y_i \) is marginally \( \text{Bin}(n, p_i) \). Hence, \( E[Y_i] = np_i \) and \( \text{Var}(Y_i) = np_i(1 - p_i) \).

Then, we can specify the entire mean vector \( E[\mathbf{Y}] \) and covariance matrix:

\[
E[\mathbf{Y}] = np = \begin{bmatrix} np_1 \\ \vdots \\ np_r \end{bmatrix} \quad \text{Var}(Y_i) = np_i(1 - p_i) \quad \text{Cov}(Y_i,Y_j) = -np_i p_j
\]

Notice the covariance is negative, which makes sense because as the number of occurrences of \( Y_i \) increases, the number of occurrences of \( Y_j \) should decrease since they can not occur simultaneously.
Proof of Multinomial Covariance. Let $X_{ik}$ for $k = 1, \ldots, n$ be indicator/Bernoulli rvs of whether the $k^{th}$ trial resulted in outcome $i$, and similarly let $X_{j\ell}$’s to be indicators of whether the $\ell^{th}$ trial resulted in outcome $j$ for $\ell = 1, \ldots, n$.

Before we begin, we should argue that $\text{Cov}(X_{ik}, X_{j\ell}) = 0$ when $k \neq \ell$ since $k$ and $\ell$ are different trials and are independent.

Furthermore, $E[X_{ik}X_{jk}] = 0$ since it’s not possible that both outcome $i$ and $j$ occur at trial $k$.

$$\text{Cov}(X_i, X_j) = \text{Cov} \left( \sum_{k=1}^{n} X_{ik}, \sum_{\ell=1}^{n} X_{j\ell} \right) \quad \text{[indicators]}$$

$$= \sum_{k=1}^{n} \sum_{\ell=1}^{n} \text{Cov}(X_{ik}, X_{j\ell}) \quad \text{[bilinearity of covariance]}$$

$$= \sum_{k=1}^{n} \text{Cov}(X_{ik}, X_{jk}) \quad \text{[independent trials, cross terms are 0]}$$

$$= \sum_{k=1}^{n} E[X_{ik}X_{jk}] - E[X_{ik}]E[X_{jk}] \quad \text{[def of covariance]}$$

$$= \sum_{k=1}^{n} -p_i p_j \quad \text{[first expectation is 0]}$$

$$= -np_i p_j$$

\qed

5.8.4 Motivation for the Multivariate Hypergeometric (MVHG) Distribution

Suppose there are $r = 3$ political parties (Green, Democratic, Republican). The senate consists of $N = 100$ senators: $K_1 = 45$ Green party members, $K_2 = 20$ Democrats, and $K_3 = 35$ Republicans.

We want to choose a committee of $n = 10$ senators.

Let $Y = (Y_1, Y_2, Y_3)$ be the number of each party’s members in the committee (G, D, R in that order). What is the probability we get 1 Green party member, 6 Democrats, and 3 Republicans? It turns out is just the following:

$$p_{Y_1, Y_2, Y_3}(1, 6, 3) = \frac{{45 \choose 1} {20 \choose 6} {35 \choose 3}}{{100 \choose 10}}$$
5.8.5 The Multivariate Hypergeometric Distribution

Once again, let us define the MVHG Distribution more generally.

**Definition 5.8.5.1: The Multivariate Hypergeometric Distribution**

Suppose there are \( r \) different colors of balls in a bag, having \( K = (K_1, \ldots, K_r) \) balls of each color, \( 1 \leq i \leq r \). Let \( N = \sum_{i=1}^{r} K_i \) be the total number of balls in the bag, and suppose we draw \( n \) without replacement. Let \( Y = (Y_1, \ldots, Y_r) \) be the rvtr such that \( Y_i \) is the number of balls of color \( i \) we drew. We write that:

\[
Y \sim \text{MVHG}_r(N, K, n)
\]

The joint PMF of \( Y \) is:

\[
p_{Y_1, \ldots, Y_r}(k_1, \ldots k_r) = \prod_{i=1}^{r} \binom{K_i}{k_i} \binom{N - \sum_{j=1}^{r} k_j}{n - \sum_{j=1}^{r} k_j}
\]

Notice that each \( Y_i \) is marginally \( \text{HypGeo}(N, K_i, n) \), so \( E[Y_i] = n \frac{K_i}{N} \) and \( \text{Var}(Y_i) = n \frac{K_i}{N} \left( 1 - \frac{K_i}{N} \right) \). Then, we can specify the entire mean vector \( E[Y] \) and covariance matrix:

\[
E[Y] = n \frac{K}{N} = \begin{bmatrix} n \frac{K_1}{N} \\ \vdots \\ n \frac{K_r}{N} \end{bmatrix} \quad \text{Var}(Y_i) = n \frac{K_i}{N} \left( \frac{N - K_i}{N} \right) \left( \frac{N - n}{N - 1} \right) \quad \text{Cov}(Y_i, Y_j) = -n \frac{K_i K_j}{N^2} \left( \frac{N - n}{N - 1} \right)
\]

**Proof of Hypergeometric Variance.** We’ll prove the variance of a univariate Hypergeometric finally (the variance of \( Y_i \)), but leave the covariance matrix to you (can approach it similarly to the multinomial covariance matrix).

Let \( X \sim \text{HypGeo}(N, K, n) \) (univariate hypergeometric). For \( i = 1, \ldots, n \), let \( X_i \) be the indicator of whether or not we got a success on trial \( i \) (not independent indicators). Then, \( E[X_i] = \mathbb{P}(X_i = 1) = \frac{K}{N} \) for every trial \( i \), so \( E[X] = n \frac{K}{N} \) by linearity of expectation.

First, we have that since \( X_i \sim \text{Ber} \left( \frac{K}{N} \right) \):

\[
\text{Var}(X_i) = p(1 - p) = \frac{K}{N} \left( 1 - \frac{K}{N} \right)
\]

Second, for \( i \neq j \), \( E[X_i X_j] = \mathbb{P}(X_i X_j = 1) = \frac{K}{N} \cdot \frac{K - 1}{N - 1} \), so

\[
\text{Cov}(X_i, X_j) = E[X_i X_j] - E[X_i] E[X_j] = \frac{K}{N} \cdot \frac{K - 1}{N - 1} - \frac{K^2}{N^2}
\]
Finally,

\[ \text{Var}(X) = \text{Var} \left( \sum_{i=1}^{n} X_i \right) \quad \quad \text{[def of } X] \]

\[ = \text{Cov} \left( \sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j \right) \quad \quad \text{[covariance with self is variance]} \]

\[ = \sum_{i=1}^{n} \sum_{j=1}^{n} \text{Cov}(X_i, X_j) \quad \quad \text{[bilinearity of covariance]} \]

\[ = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{i<j} \text{Cov}(X_i, X_j) \quad \quad \text{[split diagonal]} \]

\[ = n \frac{K}{N} \left( 1 - \frac{K}{N} \right) + 2 \frac{n(n-1)}{2} \left( \frac{K}{N} - \frac{K-1}{N} - \frac{K^2}{N^2} \right) \]

\[ = n \frac{K}{N} \frac{N-K}{N} \cdot \frac{N-n}{N-1} \quad \quad \text{[algebra]} \]

\[ \square \]

5.8.6 Exercises

These won’t be very interesting since this could’ve been done in chapter 1 and 2!

1. Suppose you are fishing in a pond with 3 red fish, 4 green fish, and 5 blue fish.

   (a) You use a net to scoop up 6 of them. What is the probability you scooped up 2 of each?

   (b) You “catch and release” until you caught 6 fish (catch 1, throw it back, catch another, throw it back, etc.). What is the probability you caught 2 of each?

Solution:

(a) Let \((X_1, X_2, X_3)\) be how many red, green, and blue fish I caught respectively. Then, \(X \sim \text{MVHG}_3(N = 12, K = (3, 4, 5), n = 6)\), and

\[ P(X_1 = 2, X_2 = 2, X_3 = 2) = \frac{\binom{3}{2} \binom{4}{2} \binom{5}{2}}{\binom{12}{6}} \]

(b) Let \((X_1, X_2, X_3)\) be how many red, green, and blue fish I caught respectively. Then, \(X \sim \text{Mult}_3(n = 6, p = (3/12, 4/12, 5/12))\), and

\[ P(X_1 = 2, X_2 = 2, X_3 = 2) = \binom{6}{2} \left( \frac{3}{12} \right)^2 \binom{4}{2} \left( \frac{4}{12} \right)^2 \binom{5}{2} \left( \frac{5}{12} \right)^2 \]