Chapter 5. Multiple Random Variables

5.1: Joint Discrete Distributions

Slides (Google Drive) Alex Tsun Video (YouTube)

This chapter, especially Sections 5.1-5.6, are arguably the most difficult in this entire text. They might take more time to fully absorb, but you'll get it, so don't give up!

We are finally going to talk about what happens when we want the probability distribution of more than one random variable. This will be called the joint distribution of two or more random variables. In this section, we'll focus on joint discrete distributions, and in the next, joint continuous distributions. We'll also finally prove that variance the variance of the sum of independent RVs is the sum of the variances, an important fact that we've been using without proof! But first, we need to review what a Cartesian product of sets is.

5.1.1 Cartesian Products of Sets

Definition 5.1.1: Cartesian Product of Sets

Let A, B be sets. The Cartesian product of A and B is denoted:

$$A \times B = \{(a, b) : a \in A, b \in B\}$$

Further if A, B are finite sets, then $|A \times B| = |A| \cdot |B|$ by the product rule of counting.

Example(s)

Write each of the following in a notation that does not involve a Cartesian product:

1. $\{1, 2, 3\} \times \{4, 5\}$
2. $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$

Solution

1. Here, we have:

$$\{1, 2, 3\} \times \{4, 5\} = \{(1, 4), (1, 5), (2, 4), (2, 5), (3, 4), (3, 5)\}$$

   We have each of the elements of the first set paired with each of the elements of the second set. Note that $|\{1, 2, 3\}| = 3$, $|\{4, 5\}| = 2$, and $|\{1, 2, 3\} \times \{4, 5\}| = 6$.

2. This is the xy-plane (2D space), which is denoted:

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y) : x \in \mathbb{R}, y \in \mathbb{R}\}$$

$\square$

5.1.2 Joint PMFs and Expectation

We will now talk about how we can model the distribution of two or more random variables, using an example to start.

Suppose we roll two fair 4-sided die independently, one blue and one red. Let X be the value of the blue die and Y be the value of the red die. Note:

$$\Omega_X = \{1, 2, 3, 4\}$$
$$\Omega_Y = \{1, 2, 3, 4\}$$

Then we can also consider $\Omega_{X,Y}$, the joint range of X and Y. The joint range happens to be any combination of $\{1, 2, 3, 4\}$ for both rolls. This can be written as:

$$\Omega_{X,Y} = \Omega_X \times \Omega_Y$$

Further each of these will be equally likely (as shown in the table below):

X\Y	1	2	3	4
1	1/16	1/16	1/16	1/16
2	1/16	1/16	1/16	1/16
3	1/16	1/16	1/16	1/16
4	1/16	1/16	1/16	1/16

Above is a suitable way to write the joint probability mass function of X and Y, as it enumerates every probability of every pair of values. If we wanted to write it as a formula, $p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$ for $x, y \in \Omega_{X,Y}$ we have:

$$p_{X,Y}(x, y) = \begin{cases} \frac{1}{16}, & x, y \in \Omega_{X,Y} \\ 0, & \text{otherwise} \end{cases}$$

Note that either this piecewise function or the table above are valid ways to express the joint PMF.

Definition 5.1.2: Joint PMFs

Let X, Y be discrete random variables. The joint PMF of X and Y is:

$$p_{X,Y}(a, b) = \mathbb{P}(X = a, Y = b)$$

The joint range is the set of pairs (c, d) that have nonzero probability:

$$\Omega_{X,Y} = \{(c, d) : p_{X,Y}(c, d) > 0\} \subseteq \Omega_X \times \Omega_Y$$

Note that the probabilities in the table must sum to 1:

$$\sum_{(s,t) \in \Omega_{X,Y}} p_{X,Y}(s, t) = 1$$

Further, note that if $g : \mathbb{R}^2 \to \mathbb{R}$ is a function, then LOTUS extends to the multidimensional case:

$$\mathbb{E}\left[g(X, Y)\right] = \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} g(x, y) p_{X,Y}(x, y)$$

A lot of things are just the same as what we learned in Chapter 3, but extended! Note that the joint range above $\Omega_{X,Y}$ was always a subset of $\Omega_X \times \Omega_Y$, and they're not necessarily equal. Let's see an example of this.

Back to our example of the blue and red die rolls. Again, let X be the value of the blue die and Y be the value of the red die. Now, let $U = \min\{X, Y\}$ (the smaller of the two die rolls) and $V = \max\{X, Y\}$ (the larger of the two die rolls). Then:

$$\Omega_U = \{1, 2, 3, 4\}$$
$$\Omega_V = \{1, 2, 3, 4\}$$

because both random variables can take on any of the four values that appear on the dice (e.g., t is possible that the minimum is 4 if we roll $(4, 4)$ and the maximum to be 1 if we roll $(1, 1)$).

However, there is the constraint that the minimum value U is always at most the maximum value V. That is, the joint range would not include the pair $(u, v) = (4, 1)$ for example, since the probability that the minimum is 4 and the maximum is 1 is zero. We can write this formally as the subset of the Cartesian product subject to $u \leq v$:

$$\Omega_{U,V} = \{(u, v) \in \Omega_U \times \Omega_V : u \leq v\} \neq \Omega_U \times \Omega_V$$

This will just be all the ordered pairs of the values that can appear as U and V. Now, however these are not equally likely, as shown in the table below. Notice that any pair (u, v) with $u > v$ has zero probability, as promised. We'll explain how we got the other numbers under the table.

U\V	1	2	3	4
1	1/16	2/16	2/16	2/16
2	0	1/16	2/16	2/16
3	0	0	1/16	2/16
4	0	0	0	1/16

As discussed earlier, we can't have the case where $U > V$, so these are all 0. The cases where $U = V$ occurs when the blue and red die have the same value, each which occurs with probability of $\frac{1}{16}$ as shown earlier. For example, $p_{U,V}(2,2) = \mathbb{P}(U = 2, V = 2) = 1/16$ since only one of the 16 equally likely outcomes $(2,2)$ gives this result. The others in which $U < V$ each occur with probability $\frac{2}{16}$ because it could be the red die with the max and the blue die with the min, or the reverse. For example, $p_{U,V}(1,3) = \mathbb{P}(U = 1, V = 3) = 2/16$ because two of the 16 outcomes $(1,3)$ and $(3,1)$ would result in the min being 1 and the max being 3.

So for the joint PMF as a formula $p_{U,V}(u,v) = \mathbb{P}(U = u, V = v)$ for $u, v \in \Omega_{U,V}$ we have:

$$p_{U,V}(u,v) = \begin{cases} \frac{2}{16}, & u, v \in \Omega_U \times \Omega_V, \quad v > u \\ \frac{1}{16}, & u, v \in \Omega_U \times \Omega_V, \quad v = u \\ 0, & \text{otherwise} \end{cases}$$

Again, the piecewise function and the table are both valid ways to express the joint PMF, and you may choose whichever is easier for you. When the joint range is larger, it might be infeasible to use a table though!

5.1.3 Marginal PMFs

Now suppose we didn't care about both U and V, just U (the minimum value). That is, we wanted to solve for the PMF $p_U(u) = \mathbb{P}(U = u)$ for $u \in \Omega_U$. Intuitively, how would you do it? Take a look at the table version of their joint PMF above.

You might think the answer is 7/16, but how did you get that? Well, $\mathbb{P}(U = 1)$ would be the sum of the first row, since that is all the cases where $U = 1$. You computed

$$\mathbb{P}(U = 1) = \mathbb{P}(U = 1, V = 1) + \mathbb{P}(U = 1, V = 2) + \mathbb{P}(U = 1, V = 3) + \mathbb{P}(U = 1, V = 4) = \frac{1}{16} + \frac{2}{16} + \frac{2}{16} + \frac{2}{16} = \frac{7}{16}$$

Mathematically, we have

$$\mathbb{P}(U = u) = \sum_{v \in \Omega_V} \mathbb{P}(U = u, V = v)$$

Does this look like anything we learned before? It's just the law of total probability (intersection version) that we derived in 2.2, as the events $\{V = v\}_{v \in \Omega_V}$ partition the sample space (V takes on exactly one value)! We can refer to the table above sum each row (which corresponds to a value of u to find the probability of that value of u occurring). That gives us the following:

$$p_U(u) = \begin{cases} \frac{7}{16}, & u = 1 \\ \frac{5}{16}, & u = 2 \\ \frac{3}{16}, & u = 3 \\ \frac{1}{16}, & u = 4 \end{cases}$$

One more example with $u = 4$ is:

$$\mathbb{P}(U = 4) = \mathbb{P}(U = 4, V = 1) + \mathbb{P}(U = 4, V = 2) + \mathbb{P}(U = 4, V = 3) + \mathbb{P}(U = 4, V = 4) = 0 + 0 + 0 + \frac{1}{16} = \frac{1}{16}$$

This brings us to the definition of marginal PMFs. The idea of these is: given a joint probability distribution, what is the distribution of just one of them (or a subset)? We get this by *marginalizing* (summing) out the other variables.

Definition 5.1.3: Marginal PMFs

Let X, Y be discrete random variables. The marginal PMF of X is:

$$p_X(a) = \sum_{b \in \Omega_Y} p_{X,Y}(a, b)$$

Similarly, the marginal PMF of Y is:

$$p_Y(d) = \sum_{c \in \Omega_X} p_{X,Y}(c, d)$$

(Extension) If Z is also a discrete random variable, then the marginal PMF of z is:

$$p_Z(z) = \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} p_{X,Y,Z}(x, y, z)$$

This follows from the law of total probability, and is just like taking the sum of a row in the example above.

Now if asked for $\mathbb{E}[U]$ for example, we actually don't need the joint PMF anymore. We've extracted the pertinent information in the form of $p_U(u)$, and compute $\mathbb{E}[U] = \sum_u u p_U(u)$ normally.

We'll do more examples right after the next section!

5.1.4 Independence

We'll now redefine independence of RVs in terms of the joint PMF. This is completely the same as the definition we gave earlier, just with the new notation we learned.

Definition 5.1.4: Independence (DRVs)

Discrete random variables X, Y are independent, written $X \perp Y$, if for all $x \in \Omega_X$ and $y \in \Omega_Y$:

$$p_{X,Y}(x, y) = p_X(x) p_Y(y)$$

Again, this just says that $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ for every x, y.

Theorem 5.1.1: Check for Independence (DRVs)

Recall the joint range $\Omega_{X,Y} = \{(x,y) : p_{X,Y}(x,y) > 0\} \subseteq \Omega_X \times \Omega_Y$ is always a subset of the Cartesian product of the individual ranges. A necessary but not sufficient condition for independence is that $\Omega_{X,Y} = \Omega_X \times \Omega_Y$. That is, if $\Omega_{X,Y} \neq \Omega_X \times \Omega_Y$, then X and Y cannot be independent, but if $\Omega_{X,Y} = \Omega_X \times \Omega_Y$, then we have to check the condition above.

This is because if there is some $(a,b) \in \Omega_X \times \Omega_Y$ but not in $\Omega_{X,Y}$, then $p_{X,Y}(a,b) = 0$ but $p_X(a) > 0$ and $p_Y(b) > 0$, violating independence. For example, suppose the joint PMF looks like:

$X \setminus Y$	8	9	Row Total $p_X(x)$
3	1/3	1/2	5/6
7	1/6	0	1/6
Col Total $p_Y(y)$	1/2	1/2	1

Also side note that the marginal distributions are named what they are, since we often write the row and column totals in the margins. The joint range $\Omega_{X,Y} \neq \Omega_X \times \Omega_Y$ since one of the entries is 0, and so $(7,9) \notin \Omega_{X,Y}$ but $(7,9) \in \Omega_X \times \Omega_Y$. This immediately tells us they cannot be independent - $p_X(7) > 0$ and $p_Y(9) > 0$, yet $p_{X,Y}(7,9) = 0$.

Example(s)

Suppose X, Y are jointly distributed with joint PMF:

$X \setminus Y$	6	9	Row Total
0	3/12	5/12	?
2	1/12	2/12	?
3	0	1/12	?
Col Total	?	?	1

1. Find the marginal probability mass functions $p_X(x)$ and $p_Y(y)$.
2. Find $\mathbb{E}[Y]$.
3. Are X and Y independent?
4. Find $\mathbb{E}\left[X^Y\right]$.

Solution

1. Actually these can be found by filling in the row and column totals, since

$$p_X(x) = \sum_y p_{X,Y}(x,y), \qquad p_Y(y) = \sum_x p_{X,Y}(x,y)$$

For example, $\mathbb{P}(X=0) = p_X(0) = \sum_y p_{X,Y}(0,y) = p_{X,Y}(0,6) + p_{X,Y}(0,9) = 3/12 + 5/12 = 8/12$ is the sum of the first row.

$X \setminus Y$	6	9	Row Total $p_X(x)$
0	3/12	5/12	8/12
2	1/12	2/12	3/12
3	0	1/12	1/12
Col Total $p_Y(y)$	4/12	8/12	1

Hence,

$$p_X(x) = \begin{cases} 8/12 & x = 0 \\ 3/12 & x = 2 \\ 1/12 & x = 3 \end{cases}$$

$$p_Y(y) = \begin{cases} 4/12 & y = 6 \\ 8/12 & y = 9 \end{cases}$$

2. We can actually compute $\mathbb{E}[Y]$ just using p_Y now that we've eliminated/marginalized out X - we don't need the joint PMF anymore. We go back to the definition:

$$\mathbb{E}[Y] = \sum_y y p_Y(y) = 6 \cdot \frac{4}{12} + 9 \cdot \frac{8}{12} = 8$$

3. X, Y are independent, if for every table entry (x, y), we have $p_{X,Y}(x, y) = p_X(x)p_Y(y)$. However, notice $p_{X,Y}(3, 6) = 0$ but $p_X(3) > 0$ and $p_Y(6) > 0$. Hence we found an entry where this condition isn't true, so they cannot be independent. This is like the comment mentioned earlier: if $\Omega_{X,Y} \neq \Omega_X \times \Omega_Y$, they have no chance of being independent.

4. We use the LOTUS formula:

$$\mathbb{E}\left[X^Y\right] = \sum_x \sum_y x^y p_{X,Y}(x, y) = 0^6 \cdot \frac{3}{12} + 0^9 \cdot \frac{5}{12} + 2^6 \cdot \frac{1}{12} + 2^9 \cdot \frac{2}{12} + 3^6 \cdot 0 + 3^9 \cdot \frac{1}{12}$$

This just sums over all the entries in the table (x, y) and takes a weighted average of all values x^y weighted by $p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$.

$\square$

5.1.5 Variance Adds for Independent Random Variables

We will finally prove that variance adds for independent RVs. You are highly encouraged to read them because they give practice with expectations with joint distributions and LOTUS!

Lemma 5.1.1: Variance Adds for Independent RVs

If X, Y are independent random variables, denoted $X \perp Y$, then:

$$\mathsf{Var}(X + Y) = \mathsf{Var}(X) + \mathsf{Var}(Y)$$

If $a, b, c \in R$ are scalars, then:

$$\mathsf{Var}(aX + bY + c) = a^2\mathsf{Var}(X) + b^2\mathsf{Var}(Y)$$

Note this property relies on the fact that they are independent, whereas linearity of expectation always holds, regardless.

To prove this, we must first prove the following lemma:

Lemma 5.1.2: Expected Value of the Product of Independent Random Variables

If $X \perp Y$, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Proof of Lemma.

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_{x\in\Omega_X}\sum_{y\in\Omega_Y} xyp_{X,Y}(x,y) && \text{[LOTUS]}\\
&= \sum_{x\in\Omega_X}\sum_{y\in\Omega_Y} xyp_X(x)p_Y(y) && [X \perp Y\text{, so } p_{X,Y}(x,y) = p_X(x)p_Y(y)]\\
&= \sum_{x\in\Omega_X} xp_X(x) \sum_{y\in\Omega_Y} yp_Y(y)\\
&= \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

□

Proof of Variance Adds for Independent RVs. Now we have the following:

$$\begin{aligned}
\mathsf{Var}(X+Y) &= \mathbb{E}\left[(X+Y)^2\right] - (\mathbb{E}[X+Y])^2 && \text{[def of variance]}\\
&= \mathbb{E}\left[X^2 + 2XY + Y^2\right] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 && \text{[linearity of expectation]}\\
&= \mathbb{E}\left[X^2\right] + 2\mathbb{E}[XY] + \mathbb{E}\left[Y^2\right] - (\mathbb{E}[X])^2 - 2\mathbb{E}[X]\mathbb{E}[Y] - (\mathbb{E}[Y])^2 && \text{[linearity of expectation]}\\
&= \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 \quad + \mathbb{E}\left[Y^2\right] - (\mathbb{E}[Y])^2 \quad + 2\mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y] && \text{[rearranging]}\\
&= \left(\mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2\right) \quad + \left(\mathbb{E}\left[Y^2\right] - (\mathbb{E}[Y])^2\right) \quad + 2\left(\mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y]\right) && [\text{lemma, since } X \perp Y]\\
&= \mathsf{Var}(X) + \mathsf{Var}(Y) + 0 && \text{[def of variance]}
\end{aligned}$$

□

5.1.6 Reproving Linearity of Expectation

Proof of Linearity of Expectation. Let X, Y be (possibly dependent) random variables. We'll prove that $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

$$\begin{aligned}
\mathbb{E}[X+Y] &= \sum_x\sum_y (x+y)p_{X,Y}(x,y) && \text{[LOTUS]}\\
&= \sum_x\sum_y xp_{X,Y}(x,y) + \sum_x\sum_y yp_{X,Y}(x,y) && \text{[split sum]}\\
&= \sum_x x\sum_y p_{X,Y}(x,y) + \sum_y y\sum_x p_{X,Y}(x,y) && \text{[algebra]}\\
&= \sum_x xp_X(x) + \sum_y yp_Y(y) && \text{[def of marginal PMF]}\\
&= \mathbb{E}[X] + \mathbb{E}[Y] && \text{[def of expectation]}
\end{aligned}$$

□

5.1.7 Exercises

1. Suppose we flip a fair coin three times independently. Let X be the number of heads in the first two flips, and Y be the number of heads in the last two flips (there is overlap).

   (a) What distribution do X and Y have marginally, and what are their ranges?

   (b) What is $p_{X,Y}(x, y)$? Fill in this table below. You may want to fill in the marginal distributions first!

   (c) What is $\Omega_{X,Y}$, using your answer to (b)?

   (d) Write a formula for $\mathbb{E}[\cos(XY)]$.

   (e) Are X, Y independent?

   Solution:

   (a) Since X counts the number of heads in two independent flips of a fair coin, then $X \sim \text{Bin}(n = 2, p = 0.5)$. Y also has this distribution! Their ranges are $\Omega_X = \Omega_Y = \{0, 1, 2\}$.

   (b) First, fill in the marginal distributions, which should be $1/4, 1/2, 1/4$ for the probability that $X = 0$, $X = 1$, and $X = 2$ respectively (same for Y).
   First let's start with $p_{X,Y}(2, 2) = \mathbb{P}(X = 2, Y = 2)$. If $X = 2$, that means the first two flips must've been heads. If $Y = 2$, that means the last two flips must've been heads. So the probability that $X = 2, Y = 2$ is the probability of the single outcome HHH, which is $1/8$. Apply similar logic for $p_{X,Y}(0, 0) = \mathbb{P}(X = 0, Y = 0)$ which is the probability of TTT.

   Then, $p_{X,Y}(0, 2) = \mathbb{P}(X = 0, Y = 2)$. If $X = 0$ then the first two flips are tails. If $Y = 2$, the last two flips are heads. This is impossible, so $\mathbb{P}(X = 0, Y = 2) = 0$. Similarly, $\mathbb{P}(X = 2, Y = 0) = 0$ as well. Now use the constraints (the row totals and col totals) to fill in the rest! For example, the first row must sum to $1/4$, and we have two out of three of the entries $p_{X,Y}(0, 0)$ and $p_{X,Y}(0, 2)$, so $p_{X,Y}(0, 1) = 1/4 - 1/8 - 0 = 1/8$.

$X \setminus Y$	0	1	2	Row Total $p_X(x)$
0	1/8	1/8	0	1/4
1	1/8	1/4	1/8	1/2
2	0	1/8	1/8	1/4
Col Total $p_Y(y)$	1/4	1/2	1/4	1

   (c) From the previous part, we can see that the joint range is everything in the Cartesian product except $(0, 2)$ and $(2, 0)$, so $\Omega_{X,Y} = (\Omega_X \times \Omega_Y) \setminus \{(0, 2), (2, 0)\}$.

   (d) By LOTUS extended to multiple variables,

$$\mathbb{E}[\cos(XY)] = \sum_x \sum_y \cos(xy) p_{X,Y}(x, y)$$

   (e) No, the joint range is not equal to the Cartesian product. This immediately makes independence impossible. The intuitive reason is that, since $(0, 2) \notin \Omega_{X,Y}$ for example, if we know $X = 0$, then Y cannot be 2. Formally, there exists a pair $(x, y) \in \Omega_X \times \Omega_Y$ (namely $(x, y) = (0, 2)$) such that $p_{X,Y}(0, 2) = 0$ but $p_X(0) > 0$ and $p_Y(2) > 0$. Hence, $p_{X,Y}(0, 2) \neq p_X(0)p_Y(2)$, which violates independence.

2. Suppose radioactive particles at Area 51 are emitted at an average rate of λ per second. You want to measure how many particles are emitted, but your geiger-counter (device that measures radioactivity) fails to record each particle independently with some small probability p. Let X be the number of particles emitted, and Y be the number of particles observed (by your geiger-counter).

(a) Describe the joint range $\Omega_{X,Y}$ using set notation.

(b) Write a formula (not a table) for $p_{X,Y}(x, y)$.

(c) Write a formula for $p_Y(y)$.

Solution:

(a) $X \sim \text{Poi}(\lambda)$ can be any nonnegative integer $\{0, 1, 2, \ldots\}$, and Y must be between 0 and X. Hence, the joint range is $\Omega_{X,Y} = \{(x, y) \in \mathbb{Z}^2 : 0 \leq y \leq x\}$, where $\mathbb{Z}$ denotes the set of integers.

(b) We know the Poisson PMF is

$$p_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

and that the distribution of Y given $X = x$ is binomial $(Y \mid X = x) \sim \text{Bin}(x, 1 - p)$. This is because, given that $X = xx$ particles were emitted, we observe each one independently with probability $1 - p$. Hence,

$$\mathbb{P}(Y = y \mid X = x) = \binom{x}{y}(1-p)^y p^{x-y}$$

By the chain rule (or definition of conditional probability),

$$p_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y \mid X = x) = e^{-\lambda}\frac{\lambda^x}{x!} \cdot \binom{x}{y}(1-p)^y p^{x-y}$$

for $(x, y) \in \Omega_{X,Y}$.

(c) We are asked to find the probability that we observe $Y = y$ particles. To make this concrete, let's say we want $p_Y(5) = \mathbb{P}(Y = 5)$. Then there is some chance of this if $X = 5$ (observing 100% of particles), or $X = 6$, or $X = 7$, etc. Hence, for any $y \in \Omega_Y$,

$$p_Y(y) = \sum_{x \in \Omega_X} p_{X,Y}(x, y) = \sum_{x=y}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!} \cdot \binom{x}{y}(1-p)^y p^{x-y}$$

That is, we sum over all cases where $x \geq y$.

3. Suppose there are N marbles in a bag, composed of r different colors. Suppose there are K_1 of color 1, K_2 of color 2,..., K_r of color r, where $\sum_{i=1}^{r} K_i = N$. We reach in and draw n *without replacement*. Let $(X_1, \ldots, X_r)$ be a random vector where X_i is the count of how many marbles of color i we drew. What is $p_{X_1,\ldots,X_r}(k_1, \ldots, k_r)$ for valid values of $k_1, \ldots, k_r$? We say the random vector $(X_1, \ldots, X_r) \sim MVHG(N, K_1, \ldots, K_r, n)$ has a multivariate hypergeometric distribution!

Solution:

$$p_{X_1,\ldots,X_r}(k_1, \ldots, k_r) = \frac{\binom{K_1}{k_1} \cdots \binom{K_r}{k_r}}{\binom{N}{n}} = \frac{\prod_{i=1}^{r}\binom{K_i}{k_i}}{\binom{N}{n}}$$