

1. **Setup:** $X \sim \text{hyper}(A, B, n)$: number of red balls drawn from box with A red, B blue, n total draws *without replacement* (D&S notation)

Common alternate notation: $X \sim (n, N, m)$: n draws without replacement from population of size N, m of which have characteristic of interest

Reconciling two notations: $n = n$; $A + B = N$, $A = m$

$\text{spt}(X) = \{\max(0, n - A), \dots, \min(n, A)\}$ Why is this the support? Lower bound on support: $x \geq \max(0, n - B)$. If $n > B$ then you must draw at least $n - B$ red balls, since there are only B blue balls to draw. If $n < B$ (as will usually be the case) then this is not restrictive. $x \leq \min(n, A)$. Upper bound on support: $x \leq \min(n, A)$. x has to be less than or equal to n , the number of trials, because you can't draw more red balls than the number of trials. x has to be less than or equal to A , the number of red balls, because you can't draw more red balls than there are in the box.

2. **Dependence:** The hypergeometric distribution is a generalization of the binomial distribution to allow for the possibility of dependent trials. We can think of the binomial distribution (n independent trials of a bernoulli RV with probability p) in the following way: suppose you have a box with A red balls and B blue balls. Define "success" as drawing a red ball. Let $p = P(\text{success}) = \frac{A}{A+B}$. If you draw n times, *replacing the ball you choose each time*, then each draw will have probability p. Assuming you mix the balls sufficiently well, each draw will be independent. If you define your random variable X as the total number of red balls drawn, then $X \sim \text{bin}(n, p)$ (sum of n independent bernoulli trials, each with probability p).

Now, suppose you *do not* replace the ball that you draw. This will introduce some *dependence* between trials. Why? The first trial is identical to a bernoulli random variable with probability $\frac{A}{A+B}$:

$$X_1 \sim \text{bern}\left(\frac{A}{A+B}\right)$$

The second trial is also a bernoulli random variable, but the probability of success depends on the outcome of the first trial. If the first trial was a success (red ball drawn), then there are only $A - 1$ red balls left out of $A + B - 1$ total, so we have a bernoulli trial with probability $\frac{A-1}{A+B-1}$:

$$X_2|X_1 = 1 \sim \text{bern}\left(\frac{A-1}{A+B-1}\right)$$

However, if the first trial was a failure, then there are A red balls left out of $A + B - 1$ total, so we have a bernoulli trial with probability $\frac{A}{A + B - 1}$:

$$X_2|X_1 = 0 \sim \text{bern}\left(\frac{A}{A + B - 1}\right)$$

So we have *dependence* from one trial to the next.

3. PMF

(a) D&S notation

$$f_X(x) = \frac{\binom{A}{x} \binom{B}{n-x}}{\binom{A+B}{n}}$$

$\binom{A}{x}$ = number of ways to choose x red balls out of A

$\binom{B}{n-x}$ = number of ways to choose $n-x$ blue balls out of B

$\binom{A+B}{n}$ = number of ways to choose n balls out of $A+B$

(b) Alternate notation

$$f_X(x) = \frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}}$$

4. $E[X] = n \frac{A}{A+B}$

Derivation

$$E[X] = E[X_1] + \dots + E[X_n]$$

$E[X_1]$ is pretty easy

$$\begin{aligned} E[X_1] &= 1P(X_1 = 1) + 0P(X_1 = 0) \\ &= 1 \frac{A}{A+B} + 0 \left(1 - \frac{A}{A+B}\right) \\ &= \frac{A}{A+B} \end{aligned}$$

It is a suprising fact that $E[X_1] = \frac{A}{A+B}$ also! Why? Use the Law of Total Probability:¹

$$\begin{aligned}
 E[X_2] &= 1P(X_2 = 1) + 0P(X_2 = 0) \\
 &= P(X_2 = 1) \\
 &= P(X_2 = 1|X_1 = 1)P(X_1 = 1) + P(X_2 = 1|X_1 = 0)P(X_1 = 0) \\
 &= \frac{A-1}{A+B-1} \frac{A}{A+B} + \frac{A}{A+B-1} \frac{B}{A+B} \\
 &= \frac{A}{A+B} \text{ (wala)}
 \end{aligned}$$

Basically, if you got a red in the first trial, your probability of getting a red in the second goes down a bit, whereas if you got a blue in the first trial, your probability of getting a red in the second goes up a little bit, and when you take the weighted average (where the weights are just the probabilities of getting red or blue in the first trial – this is just the Law of Total Probability), it turns out to be the same as when you started.

It turns out that the same thing holds for all n trials, i.e.

$$E[X_1] = \dots = E[X_n] = \frac{A}{A+B}$$

so

$$\begin{aligned}
 E[X] &= E[X_1] + \dots + E[X_n] \\
 &= \frac{A}{A+B} + \dots + \frac{A}{A+B} \\
 &= n \frac{A}{A+B}
 \end{aligned}$$

A nice shorthand is to write

$$p = \frac{A}{A+B}$$

and have

$$E[X] = np$$

5. Variance:

$$V[X] = np(1-p) \frac{N-n}{N-1}$$

where

$$p = \frac{A}{A+B}$$

¹wala (pronounced “voila”) means “with a little algebra.”

as above, and

$$N = A + B$$

This is too tedious to prove here: check D&S if you don't believe me.

6. Relationship of hypergeometric and binomial

Thinking back on what we know about the binomial distribution, let's compare the means and variances of the $\text{bin}(n, p)$ and the $\text{hyper}(n, N, m)$. Letting $p = \frac{m}{N}$, we have

	μ	σ^2
$\text{bin}(n, p)$	np	$np(1-p)$
$\text{hyper}(n, N, m)$	np	$np(1-p) \frac{N-n}{N-1}$

Note that the means are the same and the variances differ only by the term $\frac{N-n}{N-1}$. Note also that this term $\frac{N-n}{N-1} \rightarrow 1$ as N increases for fixed n (or for n growing more slowly than N), so the variances are approximately the same for large N and small n . This is a clue that as N increases, $\text{hyper}(n, N, m) \rightarrow \text{bin}(n, p)$, or, to put it another way, that the binomial distribution is a good approximation for the hypergeometric for large N . It is important to point out that this is only a clue – just because the mean and variances of two distributions converge does not at all mean that the two distributions converge, although in this case they do.

Thinking about this logically, this makes sense. The reason sampling without replacement matters – i.e. creates dependence – is that it keeps you from sampling that object again. But if the size of the population is very large, then the chance that that one object would be chosen again if you replaced it is very small, so this effect is relatively unimportant. Another way to see this is to compare the probabilities of success in the second trial after the first trial was successful or a failure:

$$\begin{aligned} P(X_2 = 1 | X_1 = 1) &= \frac{A-1}{A+B-1} \\ P(X_2 = 1 | X_1 = 0) &= \frac{A}{A+B-1} \end{aligned}$$

If A and B are large (and therefore $N = A + B$ is large), these are approximately the same.

7. Questions to ponder

(a) Suppose that every time we drew a red ball, we replaced it.

i. Would we still have dependence from one trial to the next?

- ii. Would we have restored this to a binomial random variable? If so, with what parameters? If not, can you think of another distribution we have seen that would describe this random variable?
- (b) Suppose we followed this procedure: if we draw red, we do not replace it; if we draw blue, we open up the box, put the blue back in, take out a red and set it aside (but do not count it towards our count of reds drawn). Then we close the box and shake it (the box) so that the balls are mixed up before drawing again.
- i. Would we still have dependence from one trial to the next?
 - ii. Would we have restored this to a binomial random variable? If so, with what parameters? If not, can you think of another distribution we have seen that would describe this random variable?

8. Problem set question

The pudú (*venadito*) is the world's smallest deer and a strong contender for the title of Cutest Animal Known to Man. It is indigenous to the Andean regions of Argentina and José's native country of Chile. Because of (a) habitat encroachment, (b) hungry predators undeterred by its general adorableness and (c) the fact that its main defense mechanisms are to look for cover or, if cover is unavailable, to trot in a zig-zag pattern in hopes that it can confuse its pursuer², it is seriously endangered.

Suppose we are interested in N , the size of the pudú population in a given region. It is not possible to determine what N is exactly, since the pudú (being only 18" high) is hard to spot and extremely skittish. We will have to come up with an indirect method of estimating its population.

Suppose that conservationists manage, at great effort and with much mirth, to catch m pudú, tag them and release them. Suppose that, having recovered from this indignity, the tagged pudú disperse into and are "thoroughly mixed with" the local population. That is, they are not clustered together any more than chance would dictate.

Now suppose the conservationists make a second catch, this time of n pudú. Let X be the number of *marked* pudú in this second catch. (Note: $0 \leq X \leq \min\{m, n\}$). Assume that the population, N , did not change in the time between the two catches and that, in the second catch, each pudú, tagged or untagged, was equally likely to be caught.

- (a) Argue that $X \sim \text{hyper}(n, N, m)$ (in the alternate notation) or, equivalently, that X

²One sub-species, the *pudú mephistopheles*, has half-inch horns. However, these horns are fairly soft and fuzzy and do not pose much of a threat to anything bigger than a medium-sized mouse.

has the following PMF:

$$P(X = x|n, N, m) = f_X(x|n, N, m) = \frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}}$$

- (b) Now, instead of thinking of $f_X(x|n, N, m)$ as a function of x for a given n, N, m , think of it as a function of N for a given x, n, m . One way to write this is

$$L(N|x, m, n) = f_X(x|n, N, m) = \frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}}$$

where we call L the *likelihood function*. One reasonable way to estimate of N is via the *method of maximum likelihood*, which is picking the N that maximizes the likelihood function $L(N|x, m, n)$ given the data (x, m, n) . What are we doing? Remember, $f_X(x|n, N, m)$ is a PMF, that is, it represents the probability of getting x , given n, N, m . n and m are data, so we are choosing the value of N that maximizes the probability of getting the data that we in fact got. We represent this *maximum likelihood estimate* by $\hat{N}_{MLE}$.

Prove that $\hat{N}_{MLE} = \left\lfloor \frac{mn}{x} \right\rfloor$, that is, the greatest integer less than or equal to $\frac{mn}{x}$

Hint: prove the following intermediate result:

$$\frac{L(N|x, m, n)}{L(N-1|x, m, n)} = \frac{(N-m)(N-n)}{N(N-m-n+x)}$$

Now show that this ratio is greater than one when $N < \frac{mn}{x}$ and less than one when $N > \frac{mn}{x}$

Why is this sufficient to prove that $\hat{N}_{MLE} = \left\lfloor \frac{mn}{x} \right\rfloor$?

If you would like to learn more about the pudú, click this link: http://www.bobharris.com/index.php?option=com_content&task=blogcategory&id=70&Itemid=99