Activity: Hypergeometric Probabilities

Concepts: Applying Counting Techniques, Hypergeometric Probabilities

Prerequisites: Knowledge of basic counting techniques (permutations and combinations), the basic laws of probability, and computing probabilities using the equally likely principle.

Review of Counting Techniques:

Recall the *general product rule*: If a process consists of multiple stages (call the number of stages k) and if stage i can be completed in n_i ways regardless of which outcomes occur in earlier stages, then the process itself can be completed in $n_1 n_2 \cdots n_k$ ways.

Recall the number of *permutations* (where order matters) of n objects taken k at a time is: $n(n-1)(n-2)\cdots(n-k+1)$. Notice that there are k terms in this product. This can also be expressed as: $P_{k,n} = \frac{n!}{(n-k)!}$.

Recall the number of *combinations* (where order does not matter) of n objects taken k at a time is: $C_{k,n} = \binom{n}{k} = \frac{n!}{(n-k)!k!}$. This symbol $\binom{n}{k}$ is often pronounced "n choose k" because it is the number of ways of selecting a group of k objects from a collection of n objects.

Warm-Up Examples (Permutation vs. Combination):

1. A computer system has 20 circuits. How many ways are there to randomly select 6 circuits?
2. Three married couples purchase theater tickets, 6 seats in a row. How many ways are there for them to arrange themselves? How many ways can they be arranged if each individual couple does not want to be separated?
3. How many different five-card stud poker hands are there?
4. A radio station manager can choose between R&B, rock and roll, and classical music. How many different arrangements are there?
5. How many ways are there to have 3 R&B songs in the next 5 songs?
6. A gourmet hamburger stand has the following choices for hamburger orders: Type of burger (beef, turkey, veggie), type of bun (whole wheat, white), choice of cheese (none, American, cheddar, swiss, or provolone), choice of zero up to two condiments (from mayo, mustard, pesto, and special sauce), and choices of toppings (none to all of onion, tomato, lettuce, pickles, mushrooms, bell pepper, and artichokes). How many different burgers can be ordered at this stand?

© 2002 Rossman-Chance project, supported by NSF
Used and modified with permission by Lunsford-Espy-Rowell project, supported by NSF

<u>Scenario:</u> College Committee Formation I
A college professor found herself to be one of four women on a ten-person committee. The committee needed to choose a subcommittee of three representatives to speak for the committee to other groups, and three women were selected. You will investigate how unusual this would be if the three representatives had in fact been chosen at random.

(a) How many ways can the three-person committee be chosen from the ten-person committee?

(b) If the three representatives had been chosen at random (without regard to gender), then what could we say about the likeliness of each of the possible three-person subcommittees? What then is the probability for each of the possible subcommittees?

(c) How many of the outcomes for the subcommittee consist of an all female subcommittee? What is the probability that all three persons selected for the subcommittee would be women?

(d) Does this probability indicate that it would be very surprising to find all three representatives women if all possible subcommittees were equally likely? Explain.

(e) Assuming equal likeliness, what is the probability that two of the three persons selected would be women? What is the probability that one of the three persons would be a woman? What is the probability that none of the subcommittee members are women?

(f) What do you notice about the sum of the probabilities in (c), (d), and (e)? Explain why this makes sense.

(g) What is the probability that at least two women are chosen for the committee? What rules of probability are you using for this computation?

(h) Let's use a random variable to for this computation! Let X be the number of women chosen for the subcommittee. What are the possible values of X?

© 2002 Rossman-Chance project, supported by NSF
Used and modified with permission by Lunsford-Espy-Rowell project, supported by NSF

(i) Find the following probabilities: $P(X=0)$, $P(X=1)$, $P(X=2)$, $P(X=3)$. What is $\sum_{i=0}^{3} P(X=i)$? How do you think $P(X=i)$ compares to the relative frequency function, $f_X(i)$, for the discrete random variable X? Note: We will soon be use the term "probability mass function" in place of "relative frequency function" for discrete random variables.

Hypergeometric Probabilities:

The example we worked above deals with hypergeometric probabilities and random variables with the hypergeometric distribution. Let's approach the problem in a more general setting:

Suppose that a *population* of *N* objects consists of R "successes" and *N*- R "failures." Suppose that an unordered *sample* of *n* objects is to be chosen at random from the population. Let the random variable X denote the number of successes in the sample. We will try to find a general formula for $P(X=x)$.

(j) First, for the example above, what are N, R, $N-R$, and n and what do these parameters represent (men, women, etc.)?

(k) Now, how many ways are there to choose *n* objects from the population of *N* objects? Write your answer in terms of *n* and *N*.

(l) How many ways are there to select *x* successes from the population of R successes? Write your answer in terms of *x* and R.

(m) How many ways are there to select *n-x* failures from the population of *N*- R failures? Write your answer in terms of *n-x* and *N*- R.

(n) Now find a general expression for $P(X=x)$, i.e. produce an expression for the probability of obtaining *x* successes in the sample of *n* objects where of course x is a legitimate value of X. What are the possible values for the random variable X (be careful here!)?

© 2002 Rossman-Chance project, supported by NSF
Used and modified with permission by Lunsford-Espy-Rowell project, supported by NSF

You should have found that $P(X = x)$ or the probability of drawing x successes in a sample of size n from this population is: $\dfrac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}}$ where $x \le n$, $x \le R$, and $n - x \le N - R$. These are called *hypergeometric* probabilities. The discrete random variable X is said to have a *hypergeometric distribution*.

Return to considering the population of six men and four women, with a sample of size three to be drawn from that population.

(o) Use the above expression for hypergeometric probabilities to find the probability of selecting x women (i.e. find a general formula for $P(X = x)$ for the subcommittee. Be sure to indicate for what values of x your formula is valid!). Check your formula by computing $P(X = 2)$ and $P(X = 3)$. Do these agree with your calculations above?

<u>The Ball and Urn Model:</u>

Another way to think of hypergeometic probabilities is with the ball and urn model. Suppose that an urn contains N balls, R of which are red and the others green. Suppose that n balls are drawn from the urn at random. Let the random variable X denote the number of red balls in the sample.

(p) Relate the ball and urn model to the general hypergeometric probability model as described above (i.e. what is the population, what are successes, etc.).

(q) The Virtual Labs in Probability and Statistics (VLPS) has a Ball and Urn Experiment applet we can run. Go the VLPS website, click on applets at the bottom of the page and then click on the Ball and Urn Experiment (under Finite Sampling Models). Let’s look at the committee problem via this applet. Before we start adjusting the parameters for the Ball and Urn Applet, what are the parameters for the committee problem (i.e. what are the values of N, R, and n)?

(r) Use the parameter toolbar in the applet to set these values (notice how the picture of the distribution changes as you change these parameters). Note that you will probably need to “unpark” the tool bar by grabbing the grey area to the left with your mouse and dragging it out of the window. Set the parameters for the committee problem and then record the distribution values (i.e. in this discrete case the probabilities) for $X = 0,...,3$. (Note: In the

© 2002 Rossman-Chance project, supported by NSF
Used and modified with permission by Lunsford-Espy-Rowell project, supported by NSF

applet, Y is the name of the random variable that gives the number of red balls in the sample).

(s) Now find the probability that at least one woman is on the committee. Write this probability using the random variable X. How does this probability relate graphically to the distribution (in blue) shown on the applet?

(t) What is the complement of the event described in part (s) above in words? Find its probability and write it using the random variable X. How does this probability relate graphically to the distribution (in blue) shown on the applet?

(u) The beauty of the virtual labs is that we can simulate the random experiment quickly. Let's simulate the experiment 100 times (i.e. randomly chose a committee of size three from the ten person committee) and observe the empirical distribution, the empirical mean, and the empirical standard deviation. Note that the empirical distribution is shown in red.

(v) Compare your empirical statistics (mean and standard deviation) to the distribution mean and standard deviation (given in the box below the graphs of the distributions).

<u>Scenario:</u> Friendly Observers II

Reconsider the psychology experiment about whether an observer with a vested interest inhibits one's performance on a skilled task. To assess whether the experimental results (3/12 successes in group A vs. 8/12 successes in group B) were unlikely to occur just by random chance, you performed a simulation analysis via Minitab. Now you are ready to calculate this probability exactly using counting techniques, specifically hypergeometric probabilities. Let the random variable X denote the number of "winners" assigned to group A.

(w) What are the possible values for the random variable X? Specify the relevant hypergeometric parameters N, R, and n for calculating the general probability that X is equal to one of its relevant values. How will you compute the probability of at most three successes in group A just by chance? Also write this probability of at most three successes in group A in terms of the random variable X.

© 2002 Rossman-Chance project, supported by NSF
Used and modified with permission by Lunsford-Espy-Rowell project, supported by NSF

(x) Calculate the probability in (w), i.e. find the p-value of the experimental results. To verify your computation, use the VLPS to find the p-value. You have already used Minitab to get an empirical estimate of the p-value. Now use the Virtual Labs to run the experiment 1000 times and compute another empirical p-value. Complete the following table and comment on how closely your empirical estimates of this p-value via Minitab and VLPS came to the actual value.

Method	$P(X=0)$	$P(X=1)$	$P(X=2)$	$P(X=3)$	p-value	Mean of X	Standard Dev. of X
Probability Theory (computing the actual distribution)							
Minitab Simulation (computing empirical estimates)							
VLPS Simulation (also computing empirical estimates)							

(y) Finally, summarize whether this probability (i.e. the p-value) suggests that the experimental results are very unlikely to have occurred by chance alone if there were no effect of the observer with the vested interest. Use a probability cutoff of 0.05 or less as being very unlikely to have occurred by chance alone.

© 2002 Rossman-Chance project, supported by NSF
Used and modified with permission by Lunsford-Espy-Rowell project, supported by NSF