King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 342

December 2005

Probability Issues in Without Replacement Sampling

Anwar H. Joarder and Walid S. Al-Sabah

DHAHRAN 31261 ● SAUDI ARABIA ● www.kfupm.edu.sa/math/ ● E-mail: mathdept@kfupm.edu.sa

Probability Issues in Without Replacement Sampling

Anwar H. Joarder and Walid S. Al-Sabah

Department of Mathematical Sciences
King Fahd University of Petroleum and Minerals
Dhahran 31261, Saudi Arabia
Emails: anwarj@ kfupm.edu.sa, walid@kfupm.edu.sa

Abstract Different methods scattered in different texts for calculating probabilities of events defined on a sample space generated by a sample drawn without replacement from a lot are branded. Their relative merits and limitations in application are pinpointed. An alternative representation of hypergeometric distribution appears to provide more insight into the problems usually encountered.

1. Introduction

Suppose that we are interested in the number of a kind of items say defectives in a sample of n units drawn from a lot containing $N = a + b$ balls, of which there are a ash colored balls and b blue colored balls. Let the sample be drawn in such a way that at each successive drawing, whatever units are left in the lot have the same chance of being selected. Let A_i denote the event of having an ash ball at i th draw, B_i denote the event of having a blue ball at i th draw. For a sample of size $n = 3$, the sample space would be $S = \{A_1A_2A_3,\ A_1A_2B_3,\ A_1B_2A_3,\ A_1B_2B_3,\ B_1A_2A_3,\ B_1A_2B_3,\ B_1B_2A_3,\ B_1B_2B_3\}$.

Let us assume that the items be drawn with replacement. Then the probability of $A_1A_2B_3$ is given by

$$\begin{aligned} P(A_1A_2B_3) &= P(A_1)P(A_2 \mid A_1)P(B_3 \mid A_1A_2) \\ &= \frac{a+0}{a+b}\ \frac{a+0}{a+b}\ \frac{0+b}{a+b} \qquad (1.1) \\ &= \left(\frac{a}{a+b}\right)^2 \frac{b}{a+b} \end{aligned}$$

Since each of the $\binom{3}{2} = 3$ sample points namely $\{A_1A_2B_3,\ A_1B_2A_3,\ B_1A_2A_3\}$ have the same probability, the probability of having 2 ash balls in a sample of size 3 drawn with replacement is given by

$$P(X = 2) = \binom{3}{2} \left(\frac{a}{a+b}\right)^2 \frac{b}{a+b} \qquad (1.2)$$

where X is the number of ash balls in the sample. For a sample of size n, the probability that all the balls in the sample are ash is given by $P(X = n) = \left(\frac{a}{a+b}\right)^n$.

In general let us have a sequence $A_1A_2\cdots A_x B_{x+1}\cdots B_n$ of x ash balls and $n-x$ blue balls. Then the probability of this sequence would be

$$P\left(\underbrace{A_1A_2\cdots A_x}_{x}\ \underbrace{B_{x+1}\cdots B_n}_{n-x}\right) = \frac{a+0}{a+b}\ \frac{a+0}{a+b}\cdots\frac{a+0}{a+b}\ \frac{0+b}{a+b}\ \frac{0+b}{a+b}\cdots\frac{0+b}{a+b}$$

i.e. $$P(A_1A_2\cdots A_x B_{x+1}\cdots B_n) = \left(\frac{a}{a+b}\right)^x \left(\frac{b}{a+b}\right)^{n-x} \qquad (1.3)$$

Since x ash balls from a ash balls and $n-x$ blue balls from b blue balls can happen in $\binom{n}{x}$ sequences, the probability of any outcome of this type is given by

$$P(X = x) = \binom{n}{x}\left(\frac{a}{a+b}\right)^x \left(\frac{b}{a+b}\right)^{n-x} \qquad (1.4)$$

Note that while (1.3) is the probability of a sample point, (1.4) is that of a compound event but each of the sample points in the compound event has the same probability as given by (1.3). Both the events in (1.3) and (1.4) are important in elementary courses in statistics.

When the sampling is without replacement in the above situation, a close analogue of (1.3) or (1.4) is rarely found in text books. The investigation resulted in an alternative derivation of hypergeometric probability function presented in Section 3.2: A 'New' Representation. Since there are some other ways scattered in different textbooks, first we brand, compare and contrast them. It appears that the methods discussed in Section 2 may be preferred by readers not acquainted with conditional probabilities (cf. Barnett, 1998, 182).

2. Unconditional Probability Approach

2.1 Labeling Method

Let there be $a+b$ distinguishable items in the lot. They are labeled and n items are drawn without replacement. In case the items in the lot are indistinguishable, they may be labeled to make them distinguishable to use the method.

Suppose that there are three ash balls and two blue balls in an urn. If the balls are indistinguishable, we cannot use the **Labeling Method** unless we label the balls of the same kind. Let A^1, A^2 and A^3 be ash balls while B^1 and B^2 be blue. You want to draw 1 ball from them without replacement in succession. Then the sample space would be $S = \{A^1, A^2, A^3, B^1, B^2\}$. The probability that A^1 is selected in the sample would be $P(A^1) = 1/5$. In fact the probability that A^i is selected in the sample is

$1/5,\ (i=1,2,3)$ while the probability that B^i is selected in the sample is also $1/5,\ (i=1,2)$. But the probability that an ash ball is selected would be $P(A)=P(A^1,A^2,A^3)=3/5$ while the probability that a blue ball will be selected is $P(B)=P(B^1,B^2)=2/5$.

Example 2.1 Suppose that there are three ash balls and two blue balls in an urn. You want to draw a sample of 3 balls from them without replacement in succession. What is the probability that there will be two ash balls in the sample?

If the balls are indistinguishable, we cannot use the **Labeling Method** unless we label the balls of the same kind. Let A^1, A^2 and A^3 be ash balls while B^1 and B^2 be blue. Thus there will be a total of $\binom{5}{3}$ ways as one sees in the following sample space:

$$S=\{A^1A^2A^3,\ A^1A^2B^1,\ A^1A^2B^2,\ A^1A^3B^1,\ A^1A^3B^2,\ A^1B^1B^2,\ A^2B^1B^2, A^2A^3B^1,\ A^2A^3B^2,\ A^3B^1B^2\}.$$

Assuming that each sample point is equally likely (with probability 1/10), the probability that A^1 is selected in the sample is

$P\{A^1A^2A^3,\ A^1A^2B^1,\ A^1A^2B^2,\ A^1A^3B^1,\ A^1A^3B^2,\ A^1B^1B^2\}=6/10$

and the probability that B^1 is selected in the sample is

$P\{A^1A^2B^1,\ A^1A^3B^1,\ A^1B^1B^2,\ A^2B^1B^2, A^2A^3B^1,\ A^3B^1B^2\}=6/10$

Similarly, the probability that A^i is selected in the sample is $6/10,\ (i=1,2,3)$ and the probability that B^i is selected in the sample $6/10,\ (i=1,2)$. Thus, each element in the lot has the same chance of being selected in the sample.

Assuming that each sample point is equally likely, the probability of having two ash balls in the sample is given by

$P\{A^1A^2B^1,\ A^1A^2B^2,\ A^1A^3B^1,\ A^1A^3B^2,\ A^2A^3B^1,\ A^2A^3B^2\}=6/10$.

Let X be the number of ash balls in the sample. Then

x	1	2	3
$P(X=x)$	3/10	6/10	1/10

The following is a better example of the above situation as human beings are already distinguishable.

Example 2.2 Suppose that there are three males and two females in a family. You want to invite three of them for a dinner party. What is the probability that two males are invited?

Let M^1, M^2 and M^3 be males while F^1 and F^2 be females. Then the sample space would be

$$S = \{M^1M^2M^3,\ M^1M^2F^1,\ M^1M^2F^2,\ M^1M^3F^1,\ M^1M^3F^2,$$
$$M^1F^1F^2,\ M^2M^3F^1,\ M^2M^3F^2,\ M^2F^1F^2,\ M^3F^1F^2\}.$$

As in Example 2.1, assuming that each sample point is equally likely, the probability that the male M^i is selected in the sample is $6/10,\ (i = 1,2,3)$ and that the probability that female F^i is selected in the sample is $6/10,\ (i = 1,2)$. Thus, each element in the lot has the chance of being selected in the sample.

Assuming that each sample point is equally likely, the probability that two males will be invited is given by
$$P\{M^1M^2F^1,\ M^1M^2F^2,\ M^1M^3F^1,\ M^1M^3F^2,\ M^2M^3F^1,\ M^2M^3F^2\} = 6/10.$$

The number 10 in the denominator of the above probability is the number of elements in the sample space. In case the sample size increases, it is difficult to label the sample points. This is why we do not have a general proof for this method, rather we have the following equivalent method.

2.2 Combinatorial Method

The Combinatorial Method, better known as Hypergeometric Method in this case, is a mathematical way of counting the groups or combinations where the items in the lot are distinguishable.

In Example 2.1, two of the three ash balls can be chosen in $\binom{3}{2} = 3$ ways, one of the two blue ball can be chosen from 2 blue balls in $\binom{2}{1} = 2$ ways. Hence the 2 ash balls and 1 blue ball can be chosen in $3 \times 2 = 6$ ways:

$$\left[\left\{A^1A^2, A^1A^3, A^2A^3\right\}\text{'}\times\text{'}\left\{B^1,\ B^2\right\}\right]$$
i.e. $\left\{A^1A^2B^1,\ A^1A^2B^2,\ A^1A^3B^1,\ A^1A^3B^2, A^2A^3B^1,\ A^2A^3B^2\right\}$.

Also any three ball can be chosen from 5 balls in $\binom{5}{3} = 10$ ways. If we consider the possibilities as equally likely, the probability of having 2 ash balls and 1 blue ball in a sample of size 3 is given by $6/10$.

A general proof is outlined here for some avid readers. The x ash balls can be chosen from a ash balls in the lot in $\binom{a}{x}$ ways, and the $(n-x)$ blue balls can be chosen from b blue balls in the lot in $\binom{b}{n-x}$ ways. Hence the x ash balls and $(n-x)$ blue balls can be chosen in the sample in ways $\binom{a}{x}\binom{b}{n-x}$ ways. Also n balls can be chosen from $a+b=N$ items in $\binom{a+b}{n}$ ways. If we consider the possibilities as equally likely, the probability of having $x\ (0 \le x \le n)$ ash balls and $(n-x)$ blue balls in a sample of size n is given by

$$P(X=x)=\binom{a}{x}\binom{b}{n-x}\div\binom{a+b}{n}, \quad x=0,1,\cdots,n$$

where $x \le a$ and $n-x \le b$, i.e., $\max\{0,\ n-b\} \le x \le \min\{a,\ n\}$ (Rohatgi, 1984, 336).

In Example 2.1, x, the number of ash balls in the sample, cannot exceed $a=3$, the number of ash balls in the lot, and $3-x$, the number of blue balls in the sample cannot exceed $b=2$, the number of blue balls in the lot i.e. $x \le 3$ and $3-x \le 2$, i.e. $x \le 3$ and $x \ge 1$, i.e. $1 \le x \le 3$.

x	1	2	3
$3-x$	2	1	0

Since the sample size is 3, we cannot have 0 ash balls since that needs 3 blue balls in the sample which is not possible as the lot has only two blue balls.

3. Conditional Probability Approach

3.1 Tree Diagram Method

In most elementary textbooks the Tree Diagram Method is used to calculate the probability of a simple event or any compound event. However, it is difficult to draw a tree diagram when the sample size increases. While calculating probability, we show a breakdown of items in the lot size and also in the sample size allowing us to solve any problem with relatively larger sample size in an efficient manner without having recourse to the Tree Diagram itself. What we discussed in the introduction is actually an example of Tree Diagram Method though the sampling method is with replacement.

In Example 2.1 if we assume all balls are identical (indistinguishable) except for color we cannot use the **Labeling Method** but we can use **Tree Diagram Method** provided sample size is not large.

Example 3.1 Refer to Example 2.1 with the assumption that all balls are identical (indistinguishable) except for color**.** What is the probability of having two ash balls in the sample?

Solution: Let A_i be the event of having an ash ball in the $i(i=1,2)$ th selection without replacement and B_j be the event of having blue ball in the $j(j=1,2)$ th selection without replacement. Then by the **Tree Diagram Method** the sample space would be $S=\{A_1A_2A_3,\ A_1A_2B_3,\ A_1B_2A_3,\ A_1B_2B_3,\ B_1A_2A_3,\ B_1A_2B_3,\ B_1B_2A_3\}$ (cf. Lapin, 1997, 177). Since the sampling is without replacement, $B_1B_2B_3$ is not a possible sample point.

The total number of possibilities are $2\times 2\times 2-1=7$. The probability of having two ash balls in the sample is given by

$$P(A_1A_2B_3, A_1B_2A_3, B_1A_2A_3)$$
$$=P(A_1A_2B_3)+P(A_1B_2A_3)+P(B_1A_2A_3)$$
$$=P(A_1)P(A_2\,|\,A_1)P(B_3\,|\,A_1A_2)+P(A_1)P(B_2\,|\,A_1)P(A_3\,|\,A_1B_2)$$
$$+P(B_1)P)(A_2\,|\,B_1)P(A_3\,|\,B_1A_2)$$
$$=\frac{3+0}{3+2}\times\ \frac{2+0}{2+2}\ \times\frac{0+2}{1+2}+\frac{3+0}{3+2}\ \times\frac{0+2}{2+2}\ \times\frac{2+0}{2+1}+\frac{0+2}{3+2}\ \times\frac{3+0}{3+1}\ \times\frac{2+0}{2+1}$$
$$=\frac{6}{10}$$

The probability of having an ash ball at the ith draw is $P(A_i)=3/5$ while that of having a blue ball at the that ith draw is $P(B_i)=2/5$. One can check that $P(A_1)=P(A_1A_2A_3,\ A_1A_2B_3,\ A_1B_2A_3,\ A_1B_2B_3)=3/5$.

Example 3.2 Refer to Example 2.2. It is a natural example for the **Tree Diagram Method.** Let M_i be the event that in the $i(i=1,2)$ th selection without replacement we have a male and F_j be the event that in the $j(j=1,2)$ th selection we have a female. Then by the Tree Diagram Method the sample space is given by

$S=\{M_1M_2M_3,\ M_1M_2F_3,\ M_1F_2M_3,\ M_1F_2F_3,\ F_1M_2M_3,\ F_1M_2F_3,\ F_1F_2M_3\}$. Sample space does not include $F_1F_2F_3$ as we have only two females. As above, the probability that two males are invited in the sample of 3 persons is given by $P(M_1M_2F_3,\ M_1F_2M_3,\ F_1M_2M_3)=6/10$.

Probability of having an individual is in the sample is based on the relative frequency depending on the composition of different sexes.

3.2 A 'New' Representation

We now generalize the Tree Diagram method. With the notations in the introduction, the probability of selecting x ash balls from a ash balls and $n-x$ blue balls from b blue balls under sampling without replacement, we proceed as follows. Let A_i denote the event of having an ash ball at i th draw, B_i denote the event of having a blue ball at i th draw. For a sample of size $n=3$ without replacement, the sample space would be

$$S=\{A_1A_2A_3,\ A_1A_2B_3,\ A_1B_2A_3,\ A_1B_2B_3,\ B_1A_2A_3,\ B_1A_2B_3,\ B_1B_2A_3\}.$$

Then the probability of having 2 ash balls in the sample is given by

$$P(A_1A_2B_3) = P(A_1)P(A_2 \mid A_1)P(B_3 \mid A_1A_2)$$
$$= \frac{a+0}{a+b}\frac{(a-1)+0}{(a-1)+b}\;\frac{0+b}{(a-2)+b} \qquad (3.1)$$
$$= \frac{{}_aP_2}{{}_NP_2}\frac{b}{N-2}$$

where ${}_aP_x = a(a-1)\cdots(a-x+1)$. The equation (3.1) is an analogue for (1.1). Also since each of the $\binom{3}{2}=3$ sample points namely $\{A_1A_2A_3,\ A_1A_2B_3,\ B_1A_2A_3\}$ have the same probability, the probability of having two ash balls in a sample of size 3 is given by

$$P(X=2) = \binom{3}{2}\;\frac{{}_aP_2}{{}_NP_2}\frac{b}{N-2} \qquad (3.2)$$

where X is the number of ash balls in a sample of size 3. This is an analogue of (1.2)

It is easy to prove that the probability of having an ash ball in a particular draw, say, in the 2nd draw, is given by $P(A_1A_2A_3,\ A_1A_2B_3,\ \ B_1A_2A_3,\ B_1A_2B_3) = a/N$. For a sample of size n the probability of having all ash balls in the sample is given by
$$P(X=n) = \binom{a}{n} \div \binom{a+b}{n}.$$

In general, let us have the sequence of $A_1A_2\cdots A_xB_{x+1}\cdots B_n$ of x ash balls and $n-x$ blue balls. Then the probability of this sequence would be

$$P(A_1A_2\cdots A_xB_{x+1}\cdots B_n)$$
$$= \frac{a+0}{a+b}\;\frac{(a-1)+0}{(a-1)+b}\cdots\frac{(a-x+1)+0}{(a-x+1)+b}$$
$$\times\frac{0+b}{(a-x)+b}\;\;\frac{0+(b-1)}{(a-x)+(b-1)}\cdots\frac{0+\big(b-(n-x)+1\big)}{(a-x)+(b-(n-x)+1)},$$

i.e. $P(A_1A_2\cdots A_xB_{x+1}\cdots B_n) = \dfrac{{}_aP_x}{{}_NP_x}\,\dfrac{{}_bP_{n-x}}{{}_{N-x}P_{n-x}}$ (3.3)

which is an analogue of (1.3). Since x ash balls from a ash balls and $n-x$ blue balls from b blue balls can happen in $\binom{n}{x}$ sequences, the probability of any outcome of this type is given by

$$P(X=x) = \binom{n}{x}\;\frac{{}_{a+0}P_x}{{}_{a+b}P_x}\;\;\frac{{}_{0+b}P_{n-x}}{{}_{(a-x)+b}P_{n-x}} = \binom{n}{x}\;\frac{{}_aP_x}{{}_NP_x}\;\;\frac{{}_bP_{n-x}}{{}_{N-x}P_{n-x}} \qquad (3.4)$$

where X is the number of ash balls in a sample of size n selected without replacement from the lot of size $a+b=N$. This is analogous to (1.4).

The limiting distribution of (3.4) is obvious. In case $a \to \infty, N \to \infty$ as $a/N \to p$, $(0 < p < 1)$, then $\frac{{}_aP_x}{{}_NP_x} = \frac{a(a-1)\cdots(a-x+1)}{N(N-1)\cdots(N-x+1)}$, the second term in (3.4) will tend to p^x while $\frac{{}_bP_{n-x}}{{}_{N-x}P_{n-x}} = \frac{b(b-1)\cdots(b-(n-x)+1)}{(N-n)(N-(n-1))\cdots(N-n+1)}$, the third term in (3.4) will tend to $(1-p)^{n-x}$. Thus (3.4) will have a binomial distribution in the limit. Since ${}_nP_x = \frac{n!}{(n-x)!}$ and $\binom{n}{x} = \frac{n!}{x!(n-x)!} = \frac{{}_nP_x}{x!}$, the above probability in (3.4) can be represented by

$$P(X = x) = \binom{n}{x} \frac{{}_aP_x}{{}_NP_x} \frac{{}_bP_{n-x}}{{}_{N-x}P_{n-x}}$$
$$= \frac{n!}{x!(n-x)!} \frac{{}_aP_x \; {}_bP_{n-x}}{{}_NP_n}$$
$$= \frac{{}_aP_x}{x!} \frac{{}_bP_{n-x}}{(n-x)!} \div \frac{{}_{a+b}P_n}{n!}$$

i.e. $P(X = x) = \binom{a}{x}\binom{b}{n-x} \div \binom{a+b}{n}$ (3.5)

which is the well known formula for the hypergeometric distribution (Johnson, 2005, 110) . We highlight below why the use of (3.4) instead of (3.5) would result in less mistakes by students.

Example 3.3 Suppose that a shipment of 9 digital voice recorders contains 4 that are defective. If n recorders are randomly chosen without replacement for inspection, what is the probability that

(i) the first two of $n = 3$ checked will be defective but the third one will be non-defective?
(ii) 2 of the $n = 3$ recorders will be defective?
(iii) 2 of the $n = 6$ recorders will be defective?

Solution:

(i) The probability is $P(D_1D_2D_3') = \frac{4+0}{4+5} \frac{3+0}{3+5} \frac{0+5}{2+5} = \frac{5}{42}$.

(ii) Since $4+5=9=N$, $2+1=3=n$, by the Representation (3.4), the probability that 2 of the 3 voice recorders will be defective is given by

$$P(X = 2) = P(D_1D_2D_3') + P(D_1D_2'D_3) + P(D_1'D_2D_3)$$
$$= \binom{3}{2} P(D_1D_2D_3') = \binom{3}{2} \times \frac{4+0}{4+5} \frac{3+0}{3+5} \frac{0+5}{2+5} = \frac{5}{14}$$

which is $P(X=x)=\binom{4}{2}\binom{5}{1}\div\binom{9}{3}=\frac{5}{14}$.

(iii) Since $4+5=9=N$, $2+4=6=n$, by the Representation (3.4), the probability that 2 of the 6 voice recorders will be defective is given by

$$P(X=2)=\binom{6}{2}P(D_1D_2D_3'D_4'D_5'D_6')=\binom{6}{2}\times\frac{{}^4P_2}{{}^9P_2}\times\frac{{}^5P_4}{{}^7P_4}=\frac{5}{14}$$

The students erroneously tend to use the hypergeometric probability function (3.5) to find the probability of a simple event in part (i) . Hence we recommend to use (3.4) which gives a better insight into the conditional probability. Note that the event in Example 3.3, part (i) is a simple event while that of others parts are compound.

4. Discussion

The sample space and probability for Example 2.1 by the Labeling Method (Combinatorial Method) are:

$$\{A^1A^2A^3,\ A^1A^2B^1,\ A^1A^2B^2,\ A^1A^3B^1,\ A^1A^3B^2, A^1B^1B^2,\ A^2B^1B^2, A^2A^3B^1,\ A^2A^3B^2,\ A^3B^1B^2\}$$

$$\left\{\frac{1}{10},\ \frac{1}{10},\ \frac{1}{10},\ \frac{1}{10},\ \frac{1}{10},\ \frac{1}{10},\ \frac{1}{10},\ \frac{1}{10},\ \frac{1}{10},\ \frac{1}{10}\right\}$$

while he sample space and probability for Example 2.1 by Tree Diagram Method (A 'New' Representation) revisited in Example 3.1:

$$\{A_1A_2A_3,\ A_1A_2B_3,\ A_1B_2A_3,\ A_1B_2B_3,\ B_1A_2A_3,\ B_1A_2B_3,\ B_1B_2A_3\},$$

$$\left\{\frac{1}{10},\ \frac{2}{10},\ \frac{2}{10},\ \frac{1}{10},\ \frac{2}{10},\ \frac{1}{10},\ \frac{1}{10}\right\}$$

x	Sample Points by Labeling Method / Combinatorial Method	Sample Points by Tree Diagram Method	$P(X=x)$
1	$A^1B^1B^2,\ A^2B^1B^2, A^3B^1B^2$	$A_1B_2B_3, B_1A_2B_3, B_1B_2A_3$	3/10
2	$A^1A^2B^1,\ A^1A^2B^2,\ A^1A^3B^1,$ $A^1A^3B^2, A^2A^3B^1,\ A^2A^3B^2$	$A_1A_2B_3,\ A_1B_2A_3,\ B_1A_2A_3$	6/10
3	$A^1A^2A^3$	$A_1A_2A_3$	1/10

The notations in the two sample spaces have quite different meanings. Also the number of sample points do differ. The sample points in the Labeling method (or Combinatorial Method) are equiprobable and hence an example of **simple** random sample without replacement. The elements in the sample space generated by the Tree Diagram method are not equiprobable and hence it is an example of random sampling without replacement.

Students with little background in combinatorics, may use the Labeling Method provided the lot size or the sample size is small. Most readers would find the Tree Diagram Method suitable though it can be increasingly difficult as the sample size exceeds 3. The multiplication rule in the Tree Diagram Method, though seems obvious, gives good insight into joint probability. The combinatorial method works in situations whatever be the lot size or the sample size and whether the items are distinguishable or not. The Combinatorial method and the Tree Diagram Method are equally popular in elementary statistics courses, but the connection between the sample spaces seems to be irrelevant. Though the probability function in (3.4) and (3.5) are algebraically the same, the former is preferred for beginners as it provides insight into conditional probabilities, and can be used to find the probability of a simple event or a compound event as discussed in Example 3.3. Moreover the form in (3.4) is much analogous to the Binomial Probability function.

We recommend using the 'new' representation instead of Combinatorial Method as it can be used to calculate simple or compound events with much transparency.

An Appendix to Example 3.1

(i) Probability of selecting a sample

The probability of selecting a sample of $n = 3$ balls if order is regarded is

$=\frac{1}{5}\times\frac{1}{4}\times\frac{1}{3}$ if order is regarded, and $\frac{1}{5}\times\frac{1}{4}\times\frac{1}{3}\times 3! = \binom{5}{3}^{-1} = \frac{1}{10}$. Let the balls be denoted by $A^1A^2A^3B^1B^2$. The sample space of 3 balls is

$$\{A^1A^2A^3, A^1A^2B^1, A^1A^2B^2, A^2A^3B^1, A^2A^3B^2, A^1B^1B^2, A^2A^3B^1, A^2A^3B^2, A^2B^1B^2, A^3B^1B^2\}.$$

(ii) Probability that a ball is included in a particular draw

Then the probability that a ball A^1 is included in the 1st draw is $= P(A_1^1) = 1/5$. Then the probability that A^1 is included in the 2nd draw is

$= P\left(\bar{A}_1^1 A_2^1\right)$ since the sampling is without replacement

$= P\left(\bar{A}_1^1\right) P\left(A_2^1 \mid \bar{A}_1^1\right)$

$= \left(1-\frac{1}{5}\right)\frac{1}{4} = \frac{1}{5}$

Then the probability that A^1 is included in the 3rd draw is

$P\left(\bar{A}_1^1 \bar{A}_2^1 A_3^1\right)$ since the sampling is WOR

$= P\left(\bar{A}_1^1\right) P\left(\bar{A}_2^1 \mid \bar{A}_1^1\right) P\left(A_3^1 \mid \bar{A}_1^1 A_2^1\right)$

$= \left(1-\frac{1}{5}\right)\left(1-\frac{1}{4}\right)\frac{1}{3} = 1/5$

(iii) Probability that a ball is included in the sample

The probability that the ball A^1 is included in the sample is

$$P(A_1^1) + P\left(\bar{A}_1^1 A_2^1\right) + P\left(\bar{A}_1^1 \bar{A}_2^1 A_3^1\right)$$
$$= \frac{1}{5} + \frac{1}{5} + \frac{1}{5} = 3/5$$

Each ball has the same chance of being selected in the sample (SRSWOR).

(iv) Probability of selecting a sample (Cochran, 1977)

At the first draw the probability that one of the n specified units is selected is n/N. At the second draw the probability that one of the remaining $(n-1)$ specified units is drawn is $(n-1)/(N-1)$, and so on. Hence the probability that all n specified units are selected in n draws is

$$\frac{n}{N} \cdot \frac{n-1}{N-1} \cdot \frac{n-2}{N-2} \cdots \frac{n-(n+1)}{N-(n+1)} = \frac{n!(N-n)!}{N!} = \binom{N}{n}^{-1}$$

Acknowledgement

The authors gratefully acknowledge the excellent research support provided by King Fahd University of Petroleum and Minerals, Saudi Arabia.

References

Barnett, S. (1998). *Discrete Mathematics: Numbers and Beyond.* Pearson Education Limited. Essex, England.

Cochran, W.G. (1977). *Sampling Techniques.* John Wiley.

Johnson, R. (2005). *Miller and Freund's Probability and Statistics for Engineers.* Pearson Educational International, N.J, USA.

Lapin, L.L. (1997). *Modern Engineering Statistics.* International Thompson Publishing Co., N.Y., USA.

Rohatgi, V.K. (1984). *Statistical Inference.* Wiley.

File: c\ PedaStat\ HypergeometricDistpa.doc