King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 405

May 2009

The Probability of a Sample Outcome in Sampling Without Replacement

Anwar H. Joarder; A. Laradji and M.H. Omar

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

The Probability of a Sample Outcome in Sampling Without Replacement

Anwar H. Joarder; A. Laradji and M.H. Omar

Department of Mathematics and Statistics
King Fahd University of Petroleum and Minerals
Dhahran 31261, Saudi Arabia
Emails: anwarj@ kfupm.edu.sa, alaradji@kfupm.edu.sa, omarmh@kfupm.edu.sa

Abstract The probability of a sample outcome in sampling without replacement is shown to have a combinatorial form. Then it is used to calculate the probability of any number of successes in a given sample. The resulting form is equivalent to the well known mass function of hypergeometric distribution. Vandermonde's identity readily justifies the two forms of the mass function. The new form of the mass function embodies binomial coefficient showing much resemblance to that of binomial distribution. Some other related issues are discussed.

1. Introduction

Suppose that an urn contains K items of one kind (say defective) and $N-K$ items are of different kind (say non-defective). Let n items be drawn at random, without replacement, and X denote the number of defective items selected. The quantity $D_1D_2\cdots D_x D'_{x+1}\cdots D'_n$ denotes the x successive defectives and $n-x$ successive non-defective items. The sample space contains 2^n points. The probability of $D_1D_2\cdots D_x D'_{x+1}\cdots D'_n$ is expressed by truncated factorial by Joarder and Al-Sabah (2007). In this paper we show that it has a combinatorial form. We have used it for the probability of any number of successes which results in an equivalent but insightful form of the mass function of hypergeometric distribution. Since the combinatorial function is available in almost all calculators, this form is preferred to that presented by Joarder and Al-Sabah (2007). Vandermonde's identity readily justifies the equivalence of the two forms of the mass function. On the other hand, any of the two mass functions can also be used to prove Vandermonde's identity.

The new form of the mass function embodies binomial coefficient $\binom{n}{x}$ showing much resemblance to that of binomial distribution. That hypergeomeric mass function converges to that of binomial distribution will be more transparent to students. Some other related issues are discussed**.**

2. The Probability of x Successive Successes in n Trials

Lemma 2.1 Suppose that an urn contains K items of one kind (say defective) and $N-K$ items are of a different kind (say non-defective). Let n items be drawn at random, without replacement, and X denote the number of defective items selected. The probability of x successive successes in n trials is given by

$$P(D_1D_2\cdots D_x D'_{x+1}\cdots D'_n)=\frac{\binom{N-n}{K-x}}{\binom{N}{K}}, \qquad (2.1)$$

where $\max\{0, n-(N-K)\}\le x\le \min\{n, K\}$.

Proof. The probability of x successive successes in n trials is given by

$$\begin{aligned}
&P(D_1D_2\cdots D_x D'_{x+1}\cdots D'_m)\\
&=P(D_1)P(D_2\,|\,D_1)\cdots P(D_x\,|\,D_1D_2\cdots D_{x-1})P(D'_{x+1}\,|\,D_1D_2\cdots D_x)\cdots\\
&\times P(D'_n\,|\,D_1D_2\cdots D_{x-1}D_xD'_{x+1}\cdots D'_{n-1})\\
&=\frac{K+0}{K+(N-K)}\times\frac{(K-1)+0}{(K-1)+(N-K)}\times\cdots\times\frac{(K-x+1)+0}{(K-x+1)+(N-K)}\\
&\times\frac{0+(N-K)}{(K-x)+(N-K)}\frac{0+(N-K-1)}{(K-x)+(N-K-1)}\times\cdots\times\frac{0+[N-K-(n-x)+1]}{(K-x)+[N-K-(n-x)+1]},
\end{aligned}$$

which simplifies to

$$\begin{aligned}
&P(D_1D_2\cdots D_x D'_{x+1}\cdots D'_n)\\
&=\frac{K!}{(K-x)!}\times\frac{(N-x)!}{N!}\times\frac{(N-K)!}{(N-K-n+x)!}\times\frac{(N-n)!}{(N-x)!},
\end{aligned}$$

or,

$$P(D_1D_2\cdots D_x D'_{x+1}\cdots D'_n)=\frac{K!(N-K)!}{N!}\times\frac{(N-n)!}{(K-x)!(N-K-n+x)!}$$

which is equivalent to what we have in the theorem.

The sample space contains at most 2^n outcomes. If the x defective items occur in the first x trials, then the outcome is $D_1D_2\cdots D_xD'_{x+1}\cdots D'_n$**;** if the x defective items occur from the second trial to the $(x+1)$-th trial, then the outcome is $D'_1D_2\cdots D_{x+1}D'_{x+2}\cdots D'_n$**;** if the x defective items occur from the third trial to the $(x+2)$-th trial, then the outcome is $D'_1D'_2D_3\cdots D_{x+2}D'_{x+3}\cdots D'_n$**; …** if the x defective items occur in the last x trials, then the outcome is $D'_1D'_2\cdots D'_{n-x}D_{n-x+1}\cdots D_n$. Thus there are $\binom{n}{x}$ outcomes having x defectives and $(n-x)$ non-defectives out of at most 2^n outcomes. The motivation that led to the following theorem is also implicit in Joarder and Al-Sabah (2005) or Joarder and Al-Sabah (2007).

Theorem 2.1 Suppose that an urn contains K items of one kind (say defective) and $N-K$ items are of a different kind (say non-defective). Let n items be drawn at random, without replacement, and X denote the number of defective items selected. The probability of x successes in n trials is given by

$$P(X=x)=\frac{\binom{n}{x}\binom{N-n}{K-x}}{\binom{N}{K}} \tag{2.2}$$

where $\max\{0,n-(N-K)\}\le x\le \min\{n,K\}$.

Proof. Any sample outcome of n items that have exactly x defectives and $n-x$ non-defective items is given by (2.1). Since there are $\binom{n}{x}$ such outcomes, out of a maximum of 2^n outcomes in the sample space, we have

$$P(X=x)=\binom{n}{x}P(D_1D_2\cdots D_xD'_{x+1}\cdots D'_n). \tag{2.3}$$

Hence by (2.1), we have the mass function given by (2.2).

In this approach, we classify population into sampled units and non-sampled units as in the following table:

Sampled (n)	Non-sampled $(N-n)$	Total (N)
x	$K-x$	Defective (K)
$n-x$	$N-K-(n-x)$	Non-defective $(N-K)$

Note that the sample space can be written out whether the population items are distinguishable or indistinguishable. The sample outcomes are not equally likely or equi-probable. Thus this method produces a Random Sampling but not a Simple Random Sampling. The adjective simple refers to the equally likely outcomes.

Example 2.1 A random committee of size 3 is selected from 3 doctors and 2 nurses. What is the probability that there will be 2 doctors in the committee?

Solution: Let D_i $(i=1,2,3)$ be the event that in the i-th selection we have a doctor, and N_i $(i=1,2,3)$ be the event that in the i-th selection we have a nurse. Note that nether D_i nor N_i identifies the individual. This makes the individuals **indistinguishable**. Also let X be the number of doctors selected in the committee. The sample space of outcomes is given by $\{D_1D_2D_3,\ D_1D_2N_3,\ D_1N_2D_3,\ D_1N_2N_3,\ N_1D_2D_3,\ N_1D_2N_3,\ N_1N_2D_3\}$. Then

$$P(D_1D_2N_3)=P(D_1)P(D_2\,|\,D_1)P(N_3\,|\,D_1D_2)=\frac{3+0}{3+2}\times\frac{2+0}{2+2}\times\frac{0+2}{1+2}=\frac{12}{60}=\frac{1}{5}.$$

The event of interest is given by $\{D_1D_2N_3,\ D_1N_2D_3,\ N_1D_2D_3\}$ which has a probability

$P(X=2)=\binom{3}{2}P(D_1D_2N_3)=3\times\frac{1}{5}=0.60$. Alternatively, this can be directly done by

$$P(X=2)=\binom{n}{x}P(D_1D_2N_3)=\binom{n}{x}\frac{\binom{N-n}{K-x}}{\binom{N}{K}}=\binom{3}{2}\times\frac{\binom{5-3}{3-2}}{\binom{5}{2}}=3\times\frac{2}{10}=0.60.$$

Corollary 2.1 Suppose that an urn contains K items of one kind (say defective) and $N-K$ items are of a different kind (say non-defective). Let n items be drawn at random, without replacement, and X denote the number of defective items selected. As $N\to\infty$, and $p=K/N$, the limiting value of x successes in n trials is given by p^xq^{n-x}.

Proof. It may be checked that

$$\frac{\binom{N-n}{K-x}}{\binom{N}{K}}=\frac{K}{N}\times\frac{K-1}{N-1}\times\cdots\times\frac{K-x+1}{N-x+1}\times\frac{N-K}{N-x}\times\frac{N-K-1}{N-x-1}\times\cdots\times\frac{N-K-(n-x-1)}{N-n+1}.$$

The above can be written as

$$\frac{\binom{N-n}{K-x}}{\binom{N}{K}}=\frac{K}{N}\times\left(\frac{K-1}{N}\times\frac{N}{N-1}\right)\times\cdots\times\left(\frac{K-x+1}{N}\times\frac{N}{N-x+1}\right)$$
$$\times\left(\frac{N-K}{N}\times\frac{N}{N-x}\right)\times\left(\frac{N-K-1}{N}\times\frac{N}{N-x-1}\right)\times\cdots$$
$$\times\left(\frac{N-K-(n-x-1)}{N}\times\frac{N}{N-n+1}\right),$$

which can be reorganized as

$$\frac{\binom{N-n}{K-x}}{\binom{N}{K}}=p\left(p-\frac{1}{N}\right)\cdots\left(p-\frac{x-1}{N}\right)\left(\frac{N}{N-1}\times\frac{N}{N-1}\times\cdots\times\frac{N}{N-x+1}\right)$$
$$\times q\left(q-\frac{1}{N}\right)\cdots\left(q-\frac{n-x-1}{N}\right)\left(\frac{N}{N-x}\times\frac{N}{N-x-1}\times\cdots\times\frac{N}{N-n+1}\right)$$

which has a limit of $p^x q^{n-x}$ if $N \to \infty$, and $p = K/N$.

The well known hypergeometric mass function is presented in the following theorem with an alternative proof.

Theorem 2.2 Suppose that an urn contains K items of one kind (say defective) and $N-K$ items are of a different kind (say non-defective). Let n items be drawn at random, without replacement, and X denote the number of defective items selected. The probability of x successes in n trials is given by

$$P(X = x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}, \qquad (2.4)$$

where $\max\{0, n-(N-K)\} \le x \le \min\{n, K\}$.

Proof.

$$P(D_1 D_2 \cdots D_x D'_{x+1} \cdots D'_n) = \frac{\binom{N-n}{K-x}}{\binom{N}{K}}$$

$$= \frac{K!(N-K)!}{N!} \times \frac{(N-n)!}{(K-x)!(N-K-n+x)!}$$

$$= \frac{K!}{x!(K-x)!} \times \frac{(N-K)!}{(n-x)!(N-K-n+x)!} \times \frac{x!(n-x)!}{n!} \times \frac{n!(N-n)!}{N!}.$$

That is

$$P(D_1 D_2 \cdots D_x D'_{x+1} \cdots D'_n) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{n}{x}\binom{N}{n}}.$$

Since there are $\binom{n}{x}$ such outcomes, out of a maximum of 2^n outcomes in the sample space, $P(X = x)$ is given by (2.4).

A combinatorial proof of this theorem is available in most textbooks on statistics (e.g. Johnson , 2007) and discrete mathematics (e.g. Barnett, 1998). There are $\binom{K}{x}$ ways of choosing x of the K items (say defective items) and $\binom{N-K}{n-x}$ ways of choosing $(n-x)$ of the $(N-K)$ non-defective items, and hence there are $\binom{K}{x}\binom{N-K}{n-x}$ ways of choosing x defectives and $(n-x)$ non-defective items. Since there are $\binom{N}{n}$ ways of choosing n of the N elements, assuming $\binom{N}{n}$ sample points are equally likely, the probability of having x defective items in the sample is given by (2.4). Vandermonde's identity readily justifies that the two forms of the hypergeometric distributions given by (2.2) and (2.4) are equivalent.

Obviously, here the population is classified into defectives and non-defectives as in the following table:

Defective (K)	Non-defective $(N-K)$	Total (N)
x	$n-x$	Sampled (n)

The method requires that the population items be distinguishable. The method also guarantees that sample outcomes are equally likely or equiprobable. Thus this method produces a Simple Random Sampling where "simple" refers to the equally likely outcomes.

Example 2.2 A random committee of size 3 is selected from 3 doctors and 2 nurses. Suppose that the doctors and members can be identified well making the individuals distinguishable. What is the probability that there will be 2 doctors in the committee?

Solution: Suppose the doctors are labeled as D^1, D^2 and D^3, while the nurses are labeled as N^1 and N^2 to make the items in the population **distinguishable.** The sample space of outcomes is given by

$$\{D^1D^2D^3, D^1D^2N^1, D^1D^2N^2, D^1D^3N^1, D^1D^3N^2,$$
$$D^1N^1N^2, D^2D^3N^1, D^2D^3N^2, D^2N^1N^2, D^3N^1N^2\}$$

The event of having 2 doctors in the committee is given by

$$\{D^1D^2N^1, D^1D^2N^2, D^1D^3N^1, D^1D^3N^2, D^2D^3N^1, D^2D^3N^2\}$$

which has a probability of $6/10$. This can be directly done as

$$P(X=2)=\frac{\binom{3}{2}\binom{5-3}{3-2}}{\binom{5}{3}}=\frac{6}{10}.$$

Note that the event of interest produces a rectangular array of $\binom{K}{x}=\binom{3}{2}=3$ rows and $\binom{N-K}{n-x}=\binom{5-3}{3-2}=2$ columns as follows:

$D^1D^2N^1, D^1D^2N^2$
$D^1D^3N^1, D^1D^3N^2$
$D^2D^3N^1, D^2D^3N^2$

Also note that the number of sample points is $\binom{N}{n}=\binom{5}{3}=10.$ Every sample outcome has a probability of $1/10$.

Corollary 2.2 Suppose that an urn contains K items of one kind (say defective) and $N-K$ items of a different kind (say non-defective). Let n items be drawn at random, without replacement, and X denote the number of defective items selected. Then

a. $\binom{n}{x}\binom{N-n}{K-x}\binom{N}{n}=\binom{K}{x}\binom{N-K}{n-x}\binom{N}{K},$

b. $\binom{N}{n}\sum_{x=\max\{0,n-(N-K)\}}^{\min\{n,K\}}\binom{n}{x}\binom{N-n}{K-x}=\binom{N}{K}\sum_{x=\max\{0,n-(N-K)\}}^{\min\{n,K\}}\binom{K}{x}\binom{N-K}{n-x}$

c.

$$\sum_{x=\max\{0,n-(N-K)\}}^{\min\{n,K\}}\binom{n}{x}\binom{N-n}{K-x}=\binom{N}{K},\quad \sum_{x=\max\{0,n-(N-K)\}}^{\min\{n,K\}}\binom{K}{x}\binom{N-K}{n-x}=\binom{N}{n}$$

Proof. Part (a) is obvious. Summing the identity in part (a), we have

$$\sum_{x=\max\{0,n-(N-K)\}}^{\min\{n,K\}}\binom{n}{x}\binom{N-n}{K-x}\binom{N}{n}=\sum_{x=\max\{0,n-(N-K)\}}^{\min\{n,K\}}\binom{K}{x}\binom{N-K}{n-x}\binom{N}{K},$$

which is equivalent to part (b).

From the probability mass functions in (2.2) and (2.4), we have

$$\frac{\sum_{x=\max\{0,n-(N-K)\}}^{\min\{n,K\}}\binom{n}{x}\binom{N-n}{K-x}}{\binom{N}{K}}=1,$$

and

$$\frac{\sum_{x=\max\{0,n-(N-K)\}}^{\min\{n,K\}}\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}=1$$

respectively. Part (c) is then obvious.

The identity $\sum_{x\geq 0}\binom{K}{x}\binom{N-K}{n-x}=\binom{N}{n}$ is proved by equating the coefficients of y^K in the following identity $(1+y)^K(1+y)^L=(1+y)^{K+L}$ with x as index of summation and $L=N-K$. Similarly, the identity $\sum_{x\geq 0}\binom{n}{x}\binom{N-n}{K-x}=\binom{N}{K}$ is proved by equating the coefficients of y^n in the following identity $(1+y)^n(1+y)^m=(1+y)^{n+m}$ with x as index of summation and $m=N-n$. It is worth noting that the above identities are well known Vandermonde's identity.

3. Binomial and Hypergeometric Probabilities

Suppose that an urn contains K items of one kind (say defective) and $N-K$ items are of a different kind (say non-defective). Let n items be drawn at random, **with replacement**, and X denote the number of defective items selected. The probability that any item is defective at any draw is $p=K/N$ (say). Then with arguments similar to section 2, the probability of having x successive defectives and $(n-x)$ successive non-defectives is given by

$$\begin{aligned}P(D_1D_2\cdots D_xD'_{x+1}\cdots D'_n)&=P(D_1)P(D_2)\cdots P(D_x)P(D'_{x+1})\cdots P(D'_n)\\&=\frac{K}{N}\times\frac{K}{N}\times\cdots\frac{K}{N}\times\left(1-\frac{K}{N}\right)\times\left(1-\frac{K}{N}\right)\times\cdots\left(1-\frac{K}{N}\right)=p^xq^{n-x},\end{aligned}$$

so that

$$P(X=x)=\binom{n}{x}P(D_1D_2\cdots D_xD'_{x+1}\cdots D'_n)=\binom{n}{x}p^xq^{n-x}.$$

In case of sampling **without replacement**,

$$P(X = x) = \binom{n}{x} P(D_1 D_2 \cdots D_x D'_{x+1} \cdots D'_n)$$
$$= \binom{n}{x} P(D_1) P(D_2 \mid D_1) \cdots P(D_x \mid D_1 D_2 \cdots D_{x-1}) P(D'_{x+1} \mid D_1 D_2 \cdots D_x) \cdots$$
$$\times P(D'_n \mid D_1 D_2 \cdots D_{x-1} D_x D'_{x+1} \cdots D'_{n-1})$$

(see Theorem 2.2). Now if $N \to \infty$, and $p = K / N$, it has been proved in Corollary 2.1 that

$$P(D_1) P(D_2 \mid D_1) \cdots P(D_x \mid D_1 D_2 \cdots D_{x-1}) P(D'_{x+1} \mid D_1 D_2 \cdots D_x) \cdots$$
$$\times P(D'_n \mid D_1 D_2 \cdots D_{x-1} D_x D'_{x+1} \cdots D'_{n-1})$$
$$= \frac{\binom{N-n}{K-x}}{\binom{N}{K}}$$

has a limiting value of $p^x q^{n-x}$. This shows the equivalence of binomial and hypergeometric distribution in the limit. Though the fact is available in most textbooks on statistics, the factor $\binom{n}{x}$ in the hypergeometric mass function will be insightful to the students and instructors.

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.

Joarder, A.H. and Al-Sabah, W.S. (2005). Probability issues in without replacement sampling. *Technical Report No. 342* (December 2005), Department of Mathematical Sciences, KFUPM, Saudi Arabia.

Joarder, A.H. and Al-Sabah, W.S. (2007). Probability issues in without replacement sampling. *Interntaional Journal of Mathematical Education in Science and Technology*, 38(6), 823-831.

Johnson, R. (2007). *Miller and Freund's Probability and Statistics for Engineers. Pearson* Educational International, N.J, USA.

Devore, J.L (2008). *Probability and Statistics for Engineering and the Sciences*. Arnold, London, UK.