# AN APPROXIMATION FOR PERMUTATIONS

by

GRAHAM PYATT

NUMBER 9

April 1970

This paper is circulated for discussion purposes and its contents should be considered preliminary.

## Introduction

Both in numerical calculation and in the formal development of limit theorems in statistics the need often arises for an approximation to the number of permutations of X items which can be made from a collection of N items where $N > X$. This need is usually met by the use of Stirling's approximation for a factorial.

Some difficulties attend the use of Stirling's approximation and two of these led to the analysis presented in this paper. First, the derivations of Stirling's formula which are to be found in the text books are of a degree of difficulty which is usually beyond the grasp of students at the stage at which they are first exposed to the need for such an approximation in statistical work. In consequence, students are either asked to accept Stirling's formula as an act of faith, or they are presented with less elegant methods of derivation of the important limit theorems. Secondly, proofs of limit theorems, such as that of De Moivre /Laplace, based on Stirling's approximation are not of themselves easy to follow. Accordingly there is much to be said for a simpler way of deriving such theorems if one can be found.

The purpose of this paper is first to derive an approximation for permutations, i.e. for the ratio of two factorials, and then to use this approximation to derive (a) the binomial limit of a hypergeometric distribution, (b) the Poisson limit of the binomial, and (c) the normal approximation to the binomial. Following these three applications, the relationship between the approximation to be derived and Stirling's formula is briefly discussed.

## An approximation for $^{N!}/_{(N-X)!}$

The number of permutations of X items which can be made from N distinct items, where N > X, is N!/(n-X)!. This number arises often in statistics, and for large N is cumbersome to calculate. Accordingly there is a need in numerical work for an approximation to it. Moreover, in some contexts the limit of the number is of direct interest and here an approximation can be useful in deriving the limiting form of relationships in which it occurs.

Let P(N,x) be the number of permutations N!/(N-X)! and let Π(N,X) be an approximation to it. Π(N,X) is to be such that

$$\Pi(N,X) \sim P(N,X)$$

where ~ denotes that under conditions to be stated, the ratio of the two sides tends to one. It follows that

$$\log P(N,X) - \log \Pi(N,X) \rightarrow 0$$

To derive Π(N,X), we can write

$$P(N,X) = \frac{N!}{(N-X)!}$$

$$= N(N-1)(N-2) \ldots\ldots (N-(X-1))$$

$$= N^X (1 - \tfrac{1}{N})(1 - \tfrac{2}{N}) \ldots\ldots (1 - \tfrac{X-1}{N})$$

Hence the problem can be reduced to one of approximating the geometric mean of the terms $(1 - {}^1/_N)$, $(1 - {}^2/_N)$,...., $(1 - \frac{X-1}{N})$. And we can do this by noting that the goemetric mean can be approximated by the square root of the product of the first and last term of the series. In other words

$$\{ (1 - {}^1/_N)(1 - \tfrac{X-1}{N}) \}^{\frac{1}{2}} \doteqdot \{ (1 - \tfrac{1}{N})(1 - \tfrac{2}{N}) \ldots\ldots (1 - \tfrac{X-1}{N}) \}^{\frac{1}{X-1}}$$

Alternatively, we can approximate the geometric mean by taking the square root not of the first and last terms but of the term immediately before the first term and the term immediately after the last term, i.e. of 1 and $(1 - \frac{X}{N})$. This gives

$$\left(1 - \frac{X}{N}\right)^{\frac{1}{2}} \doteq \left\{ \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right) \ldots\ldots \left(1 - \frac{X-1}{N}\right) \right\}^{\frac{1}{X-1}}$$

and is the procedure we shall follow. From it we get

$$P(N,X) = N^X\left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right) \ldots\ldots \left(1 - \frac{X-1}{N}\right)$$

$$\doteq N^X\left(1 - \frac{X}{N}\right)^{\frac{X-1}{2}}$$

$$= N^{\frac{X+1}{2}} (N - X)^{\frac{X-1}{2}}$$

$$= \Pi(N,X)$$

The above defines $\Pi(N,X)$ but does not establish how good an approximation it is. In particular we have yet to derive the conditions, if any, under which

$$P(N,X) \sim \Pi(N,X)$$

To establish the conditions, note that

$$P(N,X) = \frac{\left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right) \ldots\ldots \left(1 - \frac{X-1}{N}\right)}{\left(1 - \frac{X}{N}\right)^{\frac{X-1}{2}}} \Pi(N,X)$$

so that

$$\log P(N,X) - \log \Pi(N,X) = \sum_{i=1}^{X-1} \log\left(1 - \frac{i}{N}\right) - \frac{X-1}{2} \log\left(1 - \frac{X}{N}\right)$$

Since $0 < \frac{i}{N} < 1$ for $i = 1, \ldots\ldots, X - 1$, we can write $\log\left(1 - \frac{i}{N}\right)$ as

$$\log\left(1 - \frac{i}{N}\right) = -\frac{i}{N} - \frac{1}{2}\frac{i^2}{N^2} - \frac{1}{3}\frac{i^3}{N^3} - \ldots\ldots \text{ ad inf.}$$

$$= -\sum_{r=1}^{\infty} \frac{1}{r}\left(\frac{i}{N}\right)^r$$

Hence

$$\sum_{i=1}^{X-1} \log(1 - \frac{i}{N}) = - \sum_{i=1}^{X-1} \sum_{r=1}^{\infty} \frac{1}{r} (\frac{i}{N})^r$$

$$= - \sum_{r=1}^{\infty} \frac{1}{rN^r} \sum_{i=1}^{X-1} i^r$$

Similarly $\log (1 - {}^X/_N) = - \sum_{r=1}^{\infty} \frac{X^r}{rN^r}$

and so

$$\log P(N,X) - \log \Pi(N,X) = \sum_{r=1}^{\infty} \frac{1}{rN^r} \sum_{i=1}^{X-1} (\frac{X^r}{2} - i^r)$$

But

$$\sum_{i=1}^{X-1} \frac{X}{2} = \frac{X(X-1)}{2} = \sum_{i=1}^{X-1} i$$

hence

$$\log P(N,X) - \log \Pi(N,X) = \sum_{r=2}^{\infty} \frac{1}{rN^r} \sum_{i=1}^{X-1} (\frac{X^r}{2} - i^r)$$

From the above it follows that the error made in approximating P by $\Pi$ derives from the approximation of terms $i^r$ for $i = 1 \ldots X - 1$, on average, by ${X^r}/_2$ for values of $r \geq 2$.

The expression $\sum_{i=1}^{X-1} (\frac{X^r}{2} - i^r)$ for integer $r \geq 2$ is a polynomial in X of order r + 1. It yields

for $r = 2$, $\quad {(X-1)\, X\, (X+1)}/_6$

for $r = 3$, $\quad {(X-1)\, X^2 (X+1)}/_4$

and so on. More generally, we have

$$\log P(N,X) - \log \Pi(N,X) = \sum_{r=2}^{\infty} \frac{P(X:r+1)}{N^r}$$

where $P(X:r+1)$ denotes a polynomial in X of order r+1. Hence the approximation of P by Π holds in the limit as $\sum_{r=2}^{\infty} \frac{P(X:r+1)}{N^r}$ tends to zero. Further, since $N \geq X \geq 1$, the largest term in this expression is $X^3/N^2$. So if this term tends to zero, then the whole expression tends to zero, i.e.

$$\log \Pi(N,X) \to \log P(N,X) \quad \text{as} \quad \frac{X^3}{N^2} \to 0$$

Consequently

$$\Pi(N,X) \sim P(N,X)$$

as

$$X^3/N^2 \to 0$$

For X = 0 or 1 the approximation Π(N,X) involves no error for all N, as can easily be verified. For larger values of X there is an error which diminishes as N increases, as shown in the following table.

| X | N = 10 P(N,X) | N = 10 Π(N,X) | N = 20 P(N,X) | N = 20 Π(N,X) |
|---|---|---|---|---|
| 2 | 90 | 89.44 | 380 | 379.44 |
| 3 | 720 | 700 | 6,840 | 6,800 |
| 4 | 5,040 | 4,647.6 | 116,280 | 111,232 |
| 5 | 30,240 | 25,000 | 1,856,280 | 1,800,000 |
| 6 | 151,200 | 101,184 | $27,844\times10^3$ | $26,233\times10^3$ |
| 7 | " | " | $389,819\times10^3$ | $351,520\times10^3$ |
| 8 | " | " | $5,067,644\times10^3$ | $4,282,675\times10^3$ |

## Limited Theorems based on the Approximation

The order of convergence of the approximation $\Pi(N,X)$ is sufficient to establish several well known limit theorems in statistics. Three such proofs are given here.

### I Binomial limit of hypergeometric

A population of size N contains a proportion, p, of elements of type 1 and a proportion $q = 1-p$ of elements of type 2. A random sample, size X, is taken without replacement and found to contain $X_1$ elements of type 1 and $X_2$ elements of type 2.

$$\Pr(X_1,X_2) = \frac{\binom{pN}{X_1}\binom{qN}{X_2}}{\binom{N}{X}}$$

In the limit as $N \to \infty$ for fixed X the approximation $\Pi(N,X)$ is applicable since, with X fixed, $X^3/_{N^2} \to 0$. Hence by repeated application of the formula for $\Pi$ it is easily shown that

$$\Pr(X_1,X_2) \sim \frac{X!}{X_1!\,X_2!}\; p^{X_1} q^{X_2} \frac{\left(1-\frac{X_1}{pN}\right)^{\frac{X_1-1}{2}}\left(1-\frac{X_2}{qN}\right)^{\frac{X_2-1}{2}}}{\left(1-\frac{X}{N}\right)^{\frac{X-1}{2}}}$$

Now $\frac{X_1-1}{2}\log\left(1-\frac{X_1}{pN}\right) = -\frac{X_1-1}{2}\sum_{r=1}^{\infty}\frac{1}{r}\left(\frac{X_1}{pN}\right)^r$

Since terms of order $\frac{X^3}{N^2}$ tend to zero we have

$$\frac{X_1-1}{2}\log\left(1-\frac{X_1}{pN}\right) \to -\frac{X_1(X_1-1)}{2pN}$$

and likewise for other similar terms. Hence we get

$$\Pr(X_1,X_2) \sim \frac{X!}{X_1!\,X_2!}\; p^{X_1} q^{X_2} \exp\left\{\frac{1}{2N}\left(X(x-1) - \frac{X_1(X_1-1)}{p} - \frac{X_2(X_2-1)}{q}\right)\right\}$$

This is as far as we can get with convergence of terms of order $X^3/N^2$ to zero. If, however, we weaken the approximation to convergence to zero of terms of order $X^2/N$ then the exponential term in the above expression vanishes and we have the standard binomial approximation.

## II Poisson limit of the Binomial

If X is a random variable which is binomial with parameters N and p then

$$Pr(X) = \binom{N}{X} p^X q^{N-X}$$

Assuming that terms of order $\frac{X^3}{N^2}$ vanish we can apply the approximation $\Pi(N,X)$ to get

$$Pr(X) \sim \frac{\lambda^X}{X!} \left(1 - \frac{X}{N}\right)^{\frac{X-1}{2}} \left(1 - \frac{\lambda}{N}\right)^{N-X}$$

where $\lambda = N_p$ = constant, implying $p \to 0$ as $N \to 0$ .

Expanding the last two terms on the right-hand side without relaxing the order of convergence yields

$$Pr(X) \sim \frac{\lambda^X}{X!} \exp\left(- \lambda - \frac{1}{N}(\lambda^2 + X(X + \lambda - 1))\right)$$

It follows that the order of convergence must be relaxed to allow terms of order $X^2/N$ to vanish if we are to reach the standard result.

## III Normal approximation to the Binomial

If X is binomial with parameters N and p, then

$$Pr(X) = \binom{N}{X} p^X q^{N-X}$$

and

$$X^* = Np + p - \zeta$$

for some $\zeta$ such that $0 < \zeta < 1$ where $X^*$ is the modal value of X. Hence

$$1 \geq \frac{Pr(X)}{Pr(X^*)} = \frac{X^*!}{X!} \frac{(N-X^*)!}{(N-X)!} \left(\frac{p}{q}\right)^{X-X^*}$$

We wish to find the limit of this ratio as $\frac{(X-Np)^3}{N^2} \to 0$, in which event we have

$$\frac{(X^* - X)^3}{X^{*2}} \to 0, \quad \frac{(X^* - X)^3}{(N - X^*)^2} \to 0 \text{ and } \frac{X^* - X}{N} \to 0$$

so that the conditions for applying the approximation $\Pi(N,X)$ are satisfied. From this we get

$$\frac{Pr(X)}{Pr(X^*)} \sim \frac{X^{*\frac{X^*-X+1}{2}} \; X^{\frac{X^*-X-1}{2}} \; q^{X^*-X}}{(N - X)^{\frac{X^*-X+1}{2}} \; (N - X^*)^{\frac{X^*-X-1}{2}} \; p^{X^*-X}}$$

$$= \frac{\left(1 + \frac{P-\zeta}{Np}\right)^{\frac{X^*-X+1}{2}} \left(1 - \frac{(X^*-X) - (p-\zeta)}{Np}\right)^{\frac{X^*-X-1}{2}}}{\left(1 + \frac{(X^*-X) - (p-\zeta)}{Nq}\right)^{\frac{X^*-X+1}{2}} \left(1 - \frac{(p-\zeta)}{Nq}\right)^{\frac{X^*-X-1}{2}}}$$

Now expanding the logarithm of the right-hand side of the above and ignoring all terms which vanish under the condition stated, we get

$$\log \left\{ \frac{Pr(X)}{Pr(X^*)} \right\} \to - \frac{(X^*-X)^2}{2Npq}$$

which yields

$$Pr(X) \sim Pr(X^*) \, e^{-\frac{(X-Np)^2}{2Npq}}$$

$$= C \, e^{-\frac{(X-Np)^2}{2Npq}}$$

where C is a constant, i.e. a function of N and p which is independent of X.

It can now be shown that under the condition ${(X-Np)^3}/{N^2} \to 0$

$$e^{-\frac{(X-Np)^2}{2Npq}} \sim \int_{X-\frac{1}{2}}^{X+\frac{1}{2}} e^{-\frac{(y-Np)^2}{2Npq}} \, dy$$

Hence

$$Pr(X) \sim c \int_{X-\frac{1}{2}}^{X+\frac{1}{2}} e^{-\frac{(y-Np)^2}{2Npq}} \, dy$$

Since $\sum_{X=0}^{N} Pr(X) = 1$ it follows that c is the reciprocal of

$$\int_{-\frac{1}{2}}^{N+\frac{1}{2}} e^{-\frac{(y-Np)^2}{2Npq}} \, dy = (Npq)^{\frac{1}{2}} \int_{b}^{a} e^{-t^2/2} \, dt$$

where $a = \frac{Nq+\frac{1}{2}}{(Npq)^{\frac{1}{2}}}$, $b = -\frac{(Np+\frac{1}{2})}{(Npq)^{\frac{1}{2}}}$

Hence

$$\int_{-\frac{1}{2}}^{N+\frac{1}{2}} e^{-\frac{(y-Np)^2}{2Npq}} \, dy \sim (Npq)^{\frac{1}{2}} \int_{-\infty}^{\infty} e^{-t^2/2} \, dt$$

$$= (2\pi Npq)^{\frac{1}{2}}$$

Thus we reach the final result

$$Pr(X) \sim \frac{1}{(2\pi Npq)^{\frac{1}{2}}} \int_{X-\frac{1}{2}}^{X+\frac{1}{2}} e^{-\frac{(Y-Np)^2}{2Npq}} \, dy$$

Relation to Stirling's Approximation

We have seen that for many purposes the approximation of $P(N,X)$ by $\Pi(N,X)$ is sufficiently powerful. If, however, we want a more powerful approximation this can be obtained by recognising explicitly the remainder terms. Thus

$$\log P(N,X) = \log \Pi(N,X) + \sum_{r=2}^{\infty} \frac{1}{rN^r} \sum_{i=1}^{X-1} \left(\frac{X^r}{2} - i^r\right)$$

$$= \frac{X+1}{2} \log N + \frac{X-1}{2} \log (N-X)$$

$$+ \frac{(X-1)\ X\ (X+1)}{12N^2} + \frac{(X-1)\ X^2\ (X+1)}{12N^3}$$

$$+ \sum_{r=4}^{\infty} \frac{1}{rN^r} \sum_{i=1}^{X-1} \left(\frac{X^r}{2} - i^r\right)$$

Now we can proceed either by including more terms than are contained in $\Pi$ in our approximation, or by seeking an approximation to the remainder terms as such. In particular we can seek an approximation for

$$\frac{(X-1)\ X\ (X+1)}{12N^2} + \frac{(X-1)\ X^2\ (X+1)}{12N^3}$$

Such an approximation is in fact given by

$$\log \left\{ e^{-X} \left(1 - \frac{X}{N}\right)^{\frac{X}{2} - N} \right\} = -X + \left(N - \frac{X}{2}\right) \sum_{r=1}^{\infty} \frac{X^r}{rN^r}$$

$$= \frac{X^3}{12N^2} + \frac{X^4}{12N^3} - \frac{X^5}{8N^4}$$

$$+ \left(N - \frac{X}{2}\right) \sum_{r=5}^{\infty} \frac{X^r}{rN^r}$$

which incidentally demonstrates that $(1 - \frac{X}{N})^{N - \frac{X}{2}}$ approximates $e^{-X}$

as $X^3/N^2 \rightarrow 0$ and is therefore a closer approximation than $(1 - \frac{X}{N})^N$.

From this we have

$$\frac{(X-1)\ X\ (X+1)}{12N^2} + \frac{(X-1)\ X^2\ (X+1)}{12N^2}$$

$$= \frac{X^3}{12N^2} - \frac{X}{12N^2} + \frac{X^4}{12N^2} - \frac{X^2}{12N^2}$$

$$= \log \{ e^{-X} (1 - \frac{X}{N})^{\frac{X}{2} - N} \} - \frac{X}{12N^2} - \frac{X^2}{12N^2}$$

$$+ \frac{X^5}{8N^4} - (N - \frac{X}{2}) \sum_{r=5}^{\infty} \frac{X^r}{rN^r}$$

So $$\frac{(X-1)\ X\ (X+1)}{12N^2} + \frac{(X-1)\ X^2 (X+1)}{12N^3} \rightarrow \log \{ e^{-X} (1 - \frac{X}{N})^{\frac{X}{2} - N} \}$$

as $\frac{X^5}{N^4} \rightarrow 0$. But under the same condition,

$$\log P(N,X) \rightarrow \log \Pi(N,X) + \frac{(X-1)\ X\ (X+1)}{12N^2} + \frac{(X-1)\ X^2\ (X+1)}{12N^3}$$

Hence

$$\log P(n,X) \rightarrow \log \Pi(N,X) + \log \{ e^{-X} (1 - \frac{X}{N})^{\frac{X}{2} - N} \}$$

$$= \log \{ \frac{N^{N+\frac{1}{2}}\ e^{-X}}{(N-X)^{N-X+\frac{1}{2}}} \}$$

and

$$P(N,X) \sim \frac{N^{N+\frac{1}{2}}\ e^{-X}}{(N-X)^{N-X+\frac{1}{2}}}$$

assuming that terms of order $X^5/N^4$ vanish. This then is a closer approximation to $P(N,X)$ than $\Pi(N,X)$. It is easily shown to be the approximation that would result from substituting Stirling's approximation, in the form

$$R! = (2\Pi)^{\frac{1}{2}} \; R^{R+\frac{1}{2}} \; e^{-R}$$

into the formula $P(N,X) = N! / (N-X)!$