NEGATIVE MULTINOMIAL DISTRIBUTION

MASAAKI SIBUYA, ISAO YOSHIMURA
AND RYOICHI SHIMIZU

1. Summary and introduction

This paper reviews properties of the negative multinomial distribution and some related distributions. On the negative binomial distribution (NBn) much has been written and the contributions were summarized in two recent survey reports ([3], [9]). In the course of researches on the NBn its multivariate extension has been tried. The notion of the negative multinomial distribution (NMn) was first introduced in the model of the inverse sampling in multiple Bernoulli trials and accordingly the parameter "k" was limited to integral values. Later, in the papers discussing statistical theory of accident, absenteeism and contagion, the NMn was introduced under the name "multivariate negative binomial distribution" ([1], [2], [4], etc.). Among them Bates and Neyman's paper [4] was the first which treated the NMn systematically.

Surveying the properties of the NMn we remark the relations among distributions, which make clear the probabilistic structure of the individual distributions. We notice especially that the relation between the binomial distribution (Bn) and the NBn is quite similar to that between the multinomial distribution (Mn) and the NMn, so the name NMn is preferable to the multivariate NBn.

On the way of discussions a multivariate extension of Fisher's logarithmic series, the negative hypergeometric distribution (NHg) and its multivariate extension will be treated. Here, the name NHg is proposed, though this distribution has been discussed in literatures under different names.

2. Characteristics of the NMn

The NMn (r variate) is the probability distribution defined by

$$P(\boldsymbol{X}=\boldsymbol{x})=P(X_1=x_1, \cdots, X_r=x_r)$$

$$=(-k)^{(y)}\theta_0^{-k-y}\prod_{i=1}^{r}(\theta_i^{x_i}/x_i!) \tag{2.1}$$

$$=\frac{\Gamma(k+y)}{\Gamma(k)\prod_{i=1}^{r} x_i\,!}p_0^k\prod_{i=1}^{r} p_i^{x_i}\,,$$

where

$$x_i=0,\ 1,\ 2,\ \cdots;\qquad y=\sum_{i=1}^{r} x_i\,;$$

$$k>0;\qquad \theta_i<0,\ i=1,\ 2,\ \cdots,\ r,\qquad \theta_0=1-\sum_{i=1}^{r}\theta_i>1\,;$$

$$p_0=1/\theta_0>0,\qquad p_i=-\theta_i/\theta_0>0,\ i=1,\ 2,\ \cdots,\ r,\qquad \sum_{i=0}^{r} p_i=1\,,$$

and $a^{(b)}$ denotes the factorial product $a(a-1)\cdots(a-b+1)$. Note that the first expression is a typical term of the multinomial expansion of

$$(2.2)\qquad (\theta_0+\sum_{i=1}^{r}\theta_i)^{-k}=1\,.$$

The expression (2.1) is formally obtained by a suitable replacement of the parameter in the expression of the Mn,

$$(2.3)\qquad P(X=x)=m^{(y)}\rho_0^{m-y}\prod_{i=1}^{r}(\rho_i^{x_i}/x_i\,!)\,,$$

$$y=\sum_{i=1}^{r} x_i\leqq m\,;\ \rho_i>0\,,\ i=0,\ 1,\ \cdots,\ r,\ \sum_{i=0}^{r}\rho_i=1\,.$$

The NBn

$$(2.4)\qquad P(X=x)=\frac{\Gamma(k+x)}{\Gamma(k)x!}p^k(1-p)^x$$

is the univariate case of the NMn.

The probability generating function, the characteristic function, the moments, etc. are obtained from those of the Mn by changing the parameter as summarized in Table 1 [27]. See also Table 2 ([23], [24]). This suggests that the NMn has the nice properties of the Mn as will be seen below.

Let $J_0,\ J_1,\ \cdots,\ J_t$ be mutually disjoint subsets of $\{1,\ 2,\ \cdots,\ r\}$. Then the marginal distribution of the partial sums is again the NMn,

$$(2.5)\qquad P(\textstyle\sum_{J_1}X_i=y_1,\ \cdots,\ \sum_{J_t}X_i=y_t)$$

$$=\frac{\Gamma(k+\sum_{\nu=1}^{t} y_\nu)}{\Gamma(k)\prod_{\nu=1}^{t} y_\nu\,!}\cdot\frac{p_0^k\prod_{\nu=1}^{t}(\sum_{\nu=1}^{t}\sum_{J_\nu}p_i)y_\nu}{(p_0+\sum_{\nu=1}^{t}\sum_{J_\nu}p_i)^{k+\sum_{\nu=1}^{t}y_\nu}}\,.$$

The conditional distribution of the above partial sums given $X_{J_0}=x_{J_0}$ is the NMn,

$$(2.6)\quad \begin{aligned} &P(\textstyle\sum_{J_1}X_i=y_1,\ \cdots,\ \sum_{J_t}X_i=y_t \mid X_{J_0}=x_{J_0}) \\ &=P(\textstyle\sum_{J_1}X_i=y_1,\ \cdots,\ \sum_{J_t}X_i=y_t \mid \sum_{J_0}X_i=\sum_{J_0}x_i=y_0) \\ &=\frac{\Gamma(k+\sum_{\nu=0}^{t} y_\nu)}{\Gamma(k+y_0)\prod_{\nu=1}^{t} y_\nu !}\cdot\frac{(p_0+\sum_{J_0}p_i)^{k+y_0}\prod_{\nu=1}^{t}(\sum_{J_\nu}p_i)^{y_\nu}}{(p_0+\sum_{\nu=0}^{t}\sum_{J_\nu}p_i)^{k+\sum_{\nu=0}^{t}y_\nu}} \end{aligned}$$

Note that the expectation $E(\sum_{J_\nu}X_i)$, $\nu=1, \cdots, r$ is linear in x_i's, $i \in J_0$ (cf. Table 1).

Given $\sum_{\nu=1}^{t}\sum_{J_\nu}X_i=y$, the conditional distribution of the partial sums is the Mn (not the NMn),

$$(2.7)\quad \begin{aligned} &P(\textstyle\sum_{J_1}X_i=y_1,\ \cdots,\ \sum_{J_t}X_i=y_t \mid \sum_{\nu=1}^{t}\sum_{J_\nu}X_i=y) \\ &=\frac{y!}{\prod_{\nu=1}^{t} y_\nu !}\prod_{\nu=1}^{t}(\textstyle\sum_{J_\nu}p_i)^{y_\nu}/(\sum_{\nu=1}^{t}\sum_{J_\nu}p_i)^{y}\,. \end{aligned}$$

If $X_1, \cdots, X_n$ are independent NMn variates with common $(p_1, \cdots, p_r)$, then $\sum_{l=1}^{n} X_l=x$ has the NMn with parameter $(\sum_{l=1}^{n} k_l;\ p_1, \cdots, p_r)$.

Conversely, for any positive integer n and $k_l>0$, $l=1, \cdots, n$, such that $\sum_{l=1}^{n} k_l=k$, the NMn variate X with $(k;\ p_1, \cdots, p_r)$ can be expressed as the sum of independent X_l's with $(k_l;\ p_1, \cdots, p_r)$, $l=1, \cdots, n$. Thus, in particular, the NMn is infinitely divisible.

The family of NMn's does not include the joint distribution of r independent NBn's. Cor $(X_i, X_j)\to 0$ when (i) p_i and/or $p_j\to 0$ (in this case, the limit distribution degenerates on the set $x_i=0$ and/or $x_j=0$), or when (ii) $k\to\infty$ (in this case, as we shall see later, the limit distribution is the joint distribution of independent Poisson's). Cor $(X_i, X_j)\to 1$ when (i) $p_0\to 0$ (in this case, $E(X_l)$ and $V(X_l)\to\infty$, $l=1, 2, \cdots, r$), or when (ii) $k\to 0$ (in this case, the limit distribution degenerates on the origin).

3. Models inducing the NMn

a. Inverse sampling (waiting time) in multiple Bernoulli trials: We consider a sequence of independent trials, in each of which the event A_i occurs with probability p_i $(i=0, 1, 2, \cdots, r;\ \sum_{\nu=0}^{r} p_\nu=1)$. Let X_i be the frequency of A_i before the kth appearence of A_0. Then the distribution of $(X_1, \cdots, X_r)$ is the NMn.

b. Compounding of independent Poisson variates by a gamma distribution: Let $X_1, \cdots, X_r$ be independent Poisson variates,

$$P(X_1=x_1, \cdots, X_r=x_r)=\prod_{i=1}^{r} e^{-m\lambda_i}(m\lambda_i)^{x_i}/x_i\,! \,.$$

If the prior distribution of m is a gamma distribution,

$$f(m)=\frac{1}{a\Gamma(k)}\left(\frac{m}{a}\right)^{k-1}e^{-m/a}\,, \qquad a,\, k>0\,,$$

then

$$P(X=x)=\frac{\Gamma(\sum x_j+k)}{\Gamma(k)\prod x_i!}\left(\frac{1}{1+a\sum\lambda_i}\right)^{k}\prod\left(\frac{a\lambda_i}{1+a\sum\lambda_i}\right)^{x_i}.$$

c. Compounding the Mn by the NBn: Let the parameter of the Mn (2.3) be $(n;\, p_0, p_1, \cdots, p_r)$, and let n be distributed as the NBn (2.4) with parameter (k, ρ). Then the compound distribution is the NMn with parameter

$$\left(k;\ \frac{1}{1-p_0(1-\rho)}\,,\ \frac{(1-\rho)p_1}{1-p_0(1-\rho)}\,,\ \cdots,\ \frac{(1-\rho)p_r}{1-p_0(1-\rho)}\right)$$
$$0<\rho<1\,.$$

d. Generalization of the Poisson distribution by the "multivariate logarithmic series distribution": A multivariate extension of Fisher's logarithmic series is defined by

$$P(X_1=x_1, \cdots, X_r=x_r)=\alpha(y-1)\,!\prod_{i=1}^{r}(p_i^{x_i}/x_i!) \tag{3.1}$$
$$x_i=0, 1, 2, \cdots; \qquad y=\sum_{i=1}^{r}x_i=1, 2, \cdots;$$
$$p_0, p_1, \cdots, p_r>0, \qquad \sum_{i=0}^{r}p_i=1;$$
$$\alpha=-1/\log p_0>0\,.$$

As will be seen later, this is a limit distribution of the origin-truncated NMn, and its probability generating function is

$$-\alpha\log\,(1-\sum_{i=1}^{r}t_ip_i)\,. \tag{3.2}$$

The generalization of the Poisson distribution with mean m by (3.1) has the probability generating function

$$\exp\,\{m(-\alpha\log\,(1-\sum_{i=1}^{r}t_ip_i))-1\} \tag{3.3}$$

$$=e^{-m}(1-\sum_{i=1}^{r} t_i p_i)^{-\alpha m}$$

$$=p_0^k(1-\sum_{i=1}^{r} t_i p_i)^{-k}$$

$$k=\alpha\cdot m,\ p_0^k=e^{-m}\ ,$$

which is the probability generating function of the NMn.

e. Generalization of the NBn by the multi-point distribution: The generalization of the NBn with parameter (k, ρ) by the generalizer, $(r+1)$-point distribution,

$$P(X_1=x_1,\ \cdots,\ X_r=x_r)=p_0^{1-\sum_{i=1}^{r} x_i}\prod_{i=1}^{r} p_i^{x_i} \tag{3.4}$$

$$x_i=0,\qquad \sum_{i=1}^{r} x_i=0,\ 1\ ;$$

$$p_0,\ p_1,\ \cdots,\ p_r>0,\ \sum_{i=0}^{r} p_i=1\ ,$$

is the NMn. This process is equivalent to the compounding stated in 'c'.

f. Multiple inverse sampling without replacement: A lot consists of $m, n_1 \cdots, n_r$ items belonging to the 0th, the 1st, $\cdots$, the rth classes respectively. The items are drawn one by one without replacement until k items of the 0th class are observed. The joint distribution of the observed frequencies $X_1, \cdots, X_r$ of the 1st, $\cdots$, the rth classes is

$$\begin{aligned}
P(X_1&=x_1,\ \cdots,\ X_r=x_r) \\
&=\binom{m}{k-1}\prod_{i=1}^{r}\binom{n_i}{x_i}\Big/\binom{m+n}{k+y-1}\times\frac{m-(k-1)}{m+n-(k+y-1)} \\
&=\frac{(k+y-1)\,!\,(m+n-k-y)\,!\,m\,!}{(m+n)\,!\,(k-1)\,!\,(m-k)\,!}\prod_{i=1}^{r}\binom{n_i}{x_i}
\end{aligned} \tag{3.5}$$

$$y=\sum_{i=1}^{r} x_i,\quad n=\sum_{i=1}^{r} n_i\ .$$

If m and $n\to\infty$ in such a way that $m/(m+n)\to p_0$, $n_i/(m+n)\to p_i$, $i=1, 2, \cdots, r$, then the distribution (3.5) approaches to the NMn, as the problem reduces to the case of model 'a'.

g. An urn model: An urn contains $f_1, \cdots, f_r$ $(\sum_{i=1}^{r} f_i=g)$ balls of r different colours as well as f_0 white balls. We sample a ball and if it is coloured it is replaced with additional $c_1, \cdots, c_r$ balls of r colours, and if it is white it is replaced with additional $d=\sum_{i=1}^{r} c_i$ white balls, where $c_i/f_i=d/g=1/k=$const., $i=1, \cdots, r$. Let $X_0, X_1, \cdots, X_r$ $(\sum_{i=0}^{r} X_i=n)$ be the frequencies of the white and the coloured balls in a sequence of n such trials. Then

$$P(X_1=x_1, \cdots, X_r=x_r)$$

$$=\frac{n!}{\prod_{i=0}^{r} x_i!}\cdot\frac{\prod_{h=0}^{x_0-1}(f_0+hd)\prod_{j=0}^{y-1}(g+jd)\prod_{i=1}^{r}(f_i/g)^{x_i}}{\prod_{l=1}^{n-1}(f_0+g+ld)}$$

$$=\frac{n!}{(n-y)!}\cdot\frac{(f_0/d+n-y-1)^{(n-y)}}{((g+f_0)/d+n-1)^{(n)}}\cdot\frac{1}{\prod_{i=1}^{r} x_i!}\prod_{i=1}^{r}\left(\frac{f_i}{g}\right)^{x_i}$$

$$y=\sum_{i=1}^{r} x_i, \quad x_0+y=n .$$

If n and $f_0/d \to \infty$ in such a way that $n/(f_0/d+n)$ tends to a constant, then

$$\frac{n!}{(n-y)!}\frac{(f_0/d+n-y-1)^{(n-y)}}{((g+f_0)/d+n-1)^{(n)}} \sim \frac{ny(f_0/d)^{g/d}}{(f_0/d+n)^{y+g/d}} ,$$

hence

$$P(X_1=x_1, \cdots, X_r=x_r) \sim \frac{(k+y-1)^{(y)}}{\prod_{i=1}^{r} x_i!} p_0^k \prod_{i=1}^{r} p_i^{x_i} ,$$

where

$$p_0=\frac{f_0/d}{f_0/d+n} , \quad p_i=\frac{nf_i/g}{f_0/d+n} , \quad i=1, \cdots, r .$$

4. Limit distributions and a compound of the NMn

a. The NMn→ the independent Poisson ($k\to\infty$): Another expression of the NMn (2.1) is

$$P(X=x)=\frac{\Gamma(y+k)k^k}{\Gamma(k)(k+\sum_{i=1}^{r}\mu_i)^{k+y}}\prod_{i=1}^{r}\frac{\mu_i^{x_i}}{x_i!} \tag{4.1}$$

where

$$\mu_i=E(X_i)=kp_i/p_0 , \quad i=1, \cdots, r .$$

If $k\to\infty$, keeping $\mu_1, \cdots, \mu_r$ constant, we have

$$P(X=x) \sim \prod_{i=1}^{r} e^{-\mu_i}\mu^{x_i}/x_i! . \tag{4.2}$$

b. The origin-truncated NMn tends to the multivariate logarithmic series distribution ($k \to 0$): The origin-truncated NMn is defined by

$$P(X=x)=(1-p_0^k)^{-1}\frac{\Gamma(y+k)}{\Gamma(k)\prod_{i=1}^{r} x_i\,!}p_0^k \prod_{i=1}^{r} p_i^{x_i}\,. \tag{4.3}$$

If $k \to 0$,

$$\Gamma(y+k)/(k\Gamma(k)) \to (y-1)\,!$$

$$kp_0^k\,/\,(1-p_0^k) \to 1/\log(1/p_0)\,,$$

and therefore

$$P(X=x) \to \frac{-(y-1)\,!\prod_{i=1}^{r} p_i^{x_i}}{\log p_0 \prod_{i=1}^{r} x_i\,!}$$

which was introduced in the last section.

c. Compounding of the NMn by the multivariate beta distribution: The parameter $(p_0, p_1, \cdots, p_r)$ in (2.1) is supposed to be distributed according to the multivariate beta distribution (or the Dirichlet distribution [26]),

$$\frac{1}{B(\beta_0, \beta_1, \cdots, \beta_r)}\prod_{i=0}^{r} p^{\beta_i - 1} = \frac{\Gamma(\sum_{i=0}^{r} \beta_i)}{\prod_{i=0}^{r}\Gamma(\beta_i)}\prod_{i=0}^{r} p_i^{\beta_i - 1} \tag{4.4}$$

$$p_0, p_1, \cdots, p_r > 0, \quad \sum_{i=0}^{r} p_i = 1\,;$$

$$\beta_0, \beta_1, \cdots, \beta_r > 0\,.$$

The compound is (cf. [16]),

$$\begin{aligned} P(X=x) &= \frac{\Gamma(y+k)}{\Gamma(k)\prod_{i=1}^{r} x_i\,!} \cdot \frac{\Gamma(\sum_{i=0}^{r}\beta_i)\Gamma(k+\beta_0)\prod_{i=1}^{r}\Gamma(x_i+\beta_i)}{\Gamma(k+y+\sum_{i=0}^{r}\beta_i)\prod_{i=0}^{r}\Gamma(\beta_i)} \\ &= \frac{\Gamma(k+\beta_0)\Gamma(\sum_{i=0}^{r}\beta_i)}{(k+\sum_{i=0}^{r}\beta_i)\Gamma(\beta_0)} \cdot \frac{(y+k-1)^{(y)}\prod_{i=1}^{r}(\beta_i+x_i-1)^{(x_i)}}{(y+k+\sum_{i=0}^{r}\beta_i-1)^{(y)}\prod_{i=1}^{r} x_i\,!}\,. \end{aligned} \tag{4.5}$$

The right hand side of (4.5) can be a probability distribution for weaker conditions on the parameters, and the distributions may be called " multivariate generalized hypergeometric" (see [14], [20]), which include the usual multivariate hypergeometric and the multivariate negative hypergeometric distributions as special cases.

The relations discussed in sections 2 - 4 and later sections are summarized in Figs. 1a and 1b. Note that the parallelism between univariate and multivariate cases breaks down at a point relating to the NHg.

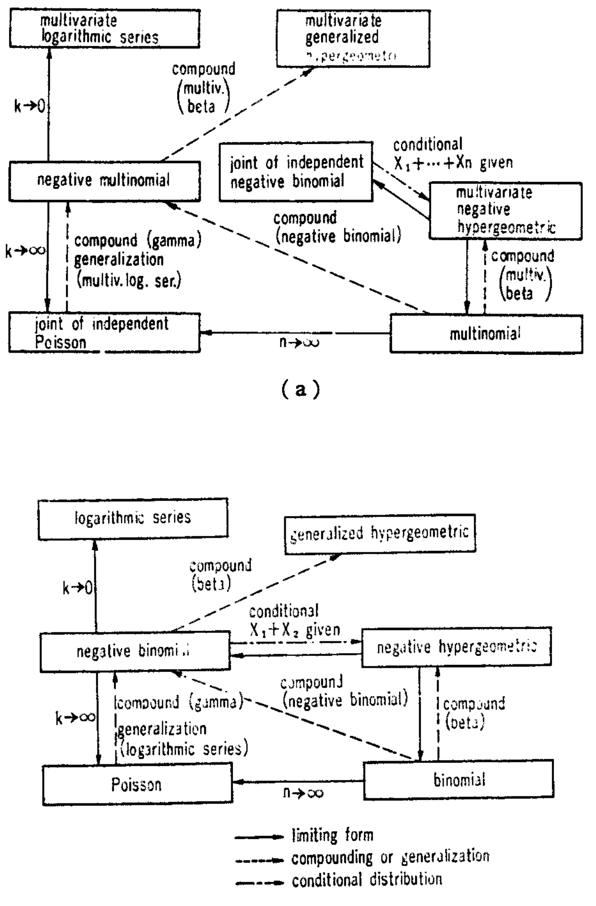

Fig. 1

5. Inference

When the parameter k is a known constant, the distribution (2.1) is a Koopman-type distribution or one of multivariate power series distributions. For a sample of size n, $X_1, \cdots, X_n$, the sum of observed vectors

$$T=(T_1, \cdots, T_r)=\sum_{\nu=1}^{n} X_\nu$$

is a complete sufficient statistic, and is again the NMn variate. Unbiased estimators based on T have the uniformly minimum variances (cf., for example, [18]).

a. Unbiased estimation (k known): For a sample of size n from (2.1) the unbiased estimator of $\prod_{i=1}^{r} p_i^{s_i}$ $(s_i=0, 1, 2, \cdots)$ based on T is

$$\varphi(T)=\begin{cases} \prod_{i=1}^{r} T_i^{(s_i)}/(T+nk-1)^{(s)}, & T_i \geqq s_i \quad i=1, \cdots, r, \\ 0, & \text{otherwise,} \end{cases} \tag{5.1}$$

where $s=\sum_{i=1}^{r} s_i$.

The unbiased estimator of any parametric function which can be expanded in a power series of p_i's is obtained by replacing each factor $\prod_{i=1}^{r} p_i^{s_i}$ in the series by its estimator (5.1).

For example, the unbiased estimator of the original probability (2.1) is

$$\varphi(T)=\begin{cases} \dfrac{(y+k-1)^{(y)}(T-y+(n-1)k-1)^{(T-y)}}{(T+nk-1)^{(T)}} \prod_{i=1}^{r}\binom{T_i}{x_i}, & T_i \geqq x_i\,. \\ 0, & \text{otherwise.} \end{cases}$$

b. Maximum likelihood estimation (k known): It is easy to see that the ML estimator of $\mu=(\mu_1, \cdots, \mu_r)=E(X)$ is

$$\hat{\mu}=(\hat{\mu}_1, \cdots, \hat{\mu}_r)=\frac{1}{n}\sum_{\nu=1}^{n} X_\nu$$

and

$$\operatorname{Var}(\hat{\mu}_i)=\frac{\mu_i(k+\mu_i)}{nk}, \quad \operatorname{Cov}(\hat{\mu}_i, \hat{\mu}_j)=\frac{\mu_i\mu_j}{nk}\,. \tag{5.2}$$

ML estimators of other parametric function and their asymptotic variances are derived from (5.2).

c. Maximum likelihood estimation (k unknown): Let Y_ν be the sum of components of each observation,

$$Y_\nu=\sum_{i=1}^{r} X_{\nu i} \qquad \nu=1, \cdots, n\,.$$

It has the NBn with (k, p_0). The log-likelihood function is written as

$$\log L = \text{const} + \sum_{\nu=1}^{n} \log \Gamma(y_\nu + k) - n \log \Gamma(k)$$
$$+ nk \log p_0 + \sum_{\nu=1}^{n} \sum_{i=1}^{r} x_{\nu i} \log p_i ,$$

which leads to the likelihood equations

$$\hat{\mu}_i = \bar{x}_i , \qquad i = 1, 2, \cdots, r ,$$

and

$$\frac{\partial \log L}{\partial k} = -n \log\left(1 + \frac{\bar{y}}{\hat{k}}\right) + \sum_{\nu=1}^{n} \left(\frac{1}{y_\nu + k - 1} + \frac{1}{y_\nu + k - 2} + \cdots + \frac{1}{k} \right)$$
$$= -n \log\left(1 + \frac{\bar{y}}{\hat{k}}\right) + \sum_{m=0}^{\infty} \frac{A_m}{\hat{k} + m} = 0 ,$$

where A_m denotes the number of observed y_ν which is larger than $m (= 0, 1, 2, \cdots)$. The last equation is the same as the ML equation of k of the NBn (k, p_0) based on $Y_1, \cdots, Y_n$. Sequential approximation by Newton's method starting from the moment estimate of k approaches to the solution rapidly and at the same time gives an estimate of the variance of the estimate [7].

6. Contingency table

In this section we consider the contingency table obtained by an inverse sampling [22]. Suppose that in an $a \times b$ table the (i, j) cell has the probability p_{ij} $(i = 1, \cdots, a;\ j = 1, \cdots, b)$, and that observations are continued until the count of the (1, 1) cell becomes k. As a result of such a sampling, the frequency X_{ij} of the (i, j) cell $((i, j) \neq (1, 1))$ is observed. The joint distribution of X_{ij}'s is the NMn,

(6.1) $$P(X_{ij} = x_{ij};\ i = 1, \cdots, a,\ j = 1, \cdots, b,\ (i, j) \neq (1, 1))$$
$$= \frac{k\Gamma(x_{..})}{\prod_{i=1}^{a} \prod_{j=1}^{b} x_{ij}!} \prod_{i=1}^{a} \prod_{j=1}^{b} p_{ij}^{x_{ij}} ,$$

where

$$x_{..} = \sum_{i=1}^{a} \sum_{j=1}^{b} x_{ij} \quad \text{and} \quad x_{11} = k .$$

Let the row-wise and the column-wise sums be $X_{i.}$ and $X_{.j}$ respectively. Then the maximum likelihood estimators of p_{ij}, $p_{i.} = \sum_{j=1}^{b} p_{ij}$ and $p_{.j} = \sum_{i=1}^{a} p_{ij}$ are

(6.2) $$\hat{p}_{ij}=X_{ij}/X_{..}, \quad \hat{p}_{i.}=X_{i.}/X_{..}, \quad \hat{p}_{.j}=X_{.j}/X_{..}\ .$$

Note that although each of $(X_{1.}-k,\ X_{2.},\ \cdots,\ X_{a.})$ and $(X_{.1}-k,\ X_{.2},\ \cdots,\ X_{.b})$ is an NMn variate, the joint distribution of them is not the NMn but a more complicated one, on which we shall not discuss here. Wishart [27] treated the distribution under the name "bivariate multimonial (Pascal case)" and Wiid [25] treated its conditional distribution.

Now the problem is to test the hypothesis of independence H_0: $p_{ij}=p_{i.}p_{.j}$ for all (i, j) against $p_{ij}\neq p_{i.}p_{.j}$ for some (i, j). The test procedures are obtained by the similar argument as that of Neyman's in the case of usual sampling [17]. First, the likelihood ratio principle leads to the critical region

(6.3) $$\lambda=2\Big(\sum_{i=1}^{a}\sum_{j=1}^{b} X_{ij}\log X_{ij}-\sum_{i=1}^{a} X_{i.}\log X_{i.}-\sum_{j=1}^{b} X_{.j}\log X_{.j}+X_{..}\log X_{..}\Big)>c$$

which is formally the same as the test of independence in usual contingency tables. The test statistic λ is distributed asymptotically as the χ^2 with d.f. $(a-1)(b-1)$.

Since the statistic

(6.4) $$Q=\sum_{(i,\ j)\neq(1,\ 1)}\frac{(X_{ij}-kp_{ij}/p_{11})^2}{kp_{ij}/p_{11}}-\frac{(X_{..}-k/p_{11})^2}{k/p_{11}}$$

is distributed asymptotically as χ^2 with d.f. ab-1, another critical region is defined by

$$Q'=\sum_{(i,\ j)\neq(1,\ 1)}\frac{(X_{ij}-kX_{i.}X_{.j}/X_{1.}X_{.1})^2}{kX_{i.}X_{.j}/X_{1.}X_{.1}}-\frac{(X_{..}-kX_{..}^2/X_{1.}X_{.1})^2}{kX_{..}^2/X_{1.}X_{.1}}$$
$$=\frac{X_{1.}X_{.1}}{kX_{..}}\Big(\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{X_{..}X_{ij}^2}{X_{i.}X_{.j}}-X_{..}\Big)>c'\ .$$

The asymptotic distributions of Q' and λ as well as the test statistics in the usual sampling case are the same.

Appendix

1. Negative hypergeometric distribution

The negative hypergeometric distribution (NHg) is defined by

(A.1) $$P(X=x)=\binom{-a}{x}\binom{-b}{n-x}\Big/\binom{-a-b}{n}$$
$$=\binom{a+x-1}{x}\binom{b+n-x-1}{n-x}\Big/\binom{a+b+n-1}{n}$$

$$=\frac{\Gamma(a+b)n!}{\Gamma(b)\Gamma(a+b+n)}\frac{(a+x-1)^{(x)}(b+n-x-1)^{(n-x)}}{x!}$$

$x=0, 1, 2, \cdots, n;\quad n$: positive integer; $\quad a, b>0.$

This is a special case of the generalized hypergeometric distribution ([14], [19], [20]) and has some nice properties contrasted with the hypergeometric, as is stated below. So it deserves the name NHg, which was used only by Hopkins [10].

The moments of (A.1) are easily obtained by changing the parameters from those of the hypergeometric distribution (Hg)

(A.2) $$P(X=x)=\binom{M}{x}\binom{N}{n-x}\Big/\binom{M+N}{n}$$

$x=\max(0, n-N), \cdots, \min(n, M);$

M, N, n : positive integers, $\quad n<M+N.$

	Hg	NBg
$E(X^{(r)})$	$\frac{M^{(r)}n^{(n)}}{(M+N)^{(r)}}$	$\frac{(a+r-1)^{(r)}n^{(r)}}{(a+b+r-1)^{(r)}}$
μ_1'	$\frac{Mn}{M+N}$	$\frac{an}{a+b}$
μ_2'	$\frac{Mn(Mn-n+N)}{(M+N)(M+N-1)}$	$\frac{an(an+n+b)}{(a+c)(a+b+1)}$
μ_2	$\frac{MNn(M+N-n)}{(M+N)^2(M+N-1)}$	$\frac{abn(a+b+n)}{(a+b)^2(a+b+1)}$

If $M, N\to\infty$, keeping $M/(M+N)=p$ constant in (A.2), and if $a, b\to\infty$, keeping $a/(a+b)=p$ constant in (A.1), then the both distributions approach to the binomial

(A.3) $$\binom{n}{x}p^x q^{n-x}$$

which has the variance $\sigma_b^2=p(1-p)/n$. Note that $\mu_2<\sigma_b^2$ for the Hg, while $\mu_2>\sigma_b^2$ for the NHg; these distributions approach to the binomial from "the opposite directions."

Now consider another limit process. If $n, N\to\infty$, keeping $n/N=p$ constant in (A.2), the distribution approaches to

(A.4) $$\binom{M}{x}p^x q^{M-x},$$

while if $n, b\to\infty$, keeping $n/b=-\theta>0$ constant in (A.1), the distribution approaches to the NBn

(A.5) $$\binom{-a}{x}\theta^x(1-\theta)^{-a-x}.$$

This difference is explained by the fact that (A.2) is invariant under the exchange of the parameters $(n, N)\leftrightarrow(M, M+N-n)$, while (A.1) has not such symmetry.

If X_1 and X_2 be independent Bn (resp. NBn) variates with X_ν having parameter (M_ν, p_ν) in (A.4) ((a_ν, θ_ν) in (A.5)), then the uniformly most powerful unbiased tests for the problems H_1: $p_1\leqq p_2$ $(\theta_1\leqq\theta_2)$ against $p_1>p_2$ $(\theta_1>\theta_2)$ and H_2: $p_1=p_2$ $(\theta_1=\theta_2)$ against $p_1\neq p_2$ $(\theta_1\neq\theta_2)$ are obtained in terms of X_1 given $X_1+X_2=t$ ([3] section 8, [15], 140–143). The conditional distribution for $p_1=p_2$ $(\theta_1=\theta_2)$ is

(A.6) $$P(X_1=x\mid X_1+X_2=t)=\binom{M_1}{x}\binom{M_2}{t-x}\Big/\binom{M_1+M_2}{t},$$

((A.7) $$\left.\binom{-a_1}{x}\binom{-a_2}{t-x}\Big/\binom{-a_1-a_2}{t}\right).$$

2. Models inducing the NHg [19]

a. Inverse sampling without replacement: A lot consists of m acceptable items and n defective ones. Suppose that items are drawn at random one by one and X defectives are observed before the ath acceptable one. Then we have

(A.8) $$P(X=x)=\binom{a+x-1}{x}\binom{m+n-a-x}{n-x}\Big/\binom{m+n}{m}$$

(the waiting-time distribution, hypergeometric case, [26]).

sampling with replacement : binomial	inverse sampling with replacement : negative binomial
sampling without replacement : hypergeometric	inverse sampling without replacement : negative hypergeometric

b. Compound binomial: The compound of the binomial distribution

$$\binom{n}{x}p^xq^{b-1}$$

by the compounder

$$\frac{1}{B(a, b)}p^{a-1}q^{b-1}$$

is the NHg (A.1). ([12], [21], etc.).

c. Exceedance: Let $(X_1, \cdots, X_m)$ and $(Y_1, \cdots, Y_n)$ be two random samples from the same population with a continuous distribution function, and S be the number of Y's which are larger than $X_{(d)}$, the dth order statistic of $(X_1, \cdots, X_m)$. Then,

$$\text{(A.9)} \qquad P(S=s)=\binom{m-d+s}{s}\binom{n+d-s-1}{n-s}\Big/\binom{m+n}{n}.$$

(See [5], [8], [19]).

d. An occupancy problem: If n indistinguishable balls are placed at random into m cells, then the probability that a group of d prescribed cells contains a total of exactly s balls is

$$P(S=s)=\binom{d+s-1}{s}\binom{m-d+n-s-1}{n-s}\Big/\binom{m+n-1}{n}.$$

The exceedances in 'c' is regarded as the balls which are placed into $m-d+1$ prescribed cells when n indistinguishable balls are placed at random into $m+1$ cells.

e. An urn model: An urn contains $b \cdot c$ coloured balls and $a \cdot c$ white ones. A ball is drawn at random and if it is coloured (white) it is replaced by additional c coloured (white) ones. Then, in n trials X coloured balls will be observed with probability (A.1). (A.1) is called the Pólya distribution in relation to this model.

3. Multivariate negative hypergeometric distribution

The r-variate NHg is defined as an extension of (A.1),

$$\text{(A.10)} \qquad P(X=x)=\binom{-a_0}{n-y}\prod_{i=1}^{r}\binom{-a_i}{x_i}\Big/\binom{-\sum_{i=0}^{r} a_i}{n}$$

$$=\frac{n!}{(n-y)\prod_{i=1}^{r} x_i!}\cdot\frac{\prod_{i=1}^{r}\Gamma(a_i+x_i)\Gamma(\sum_{i=0}^{r} a_i)\Gamma(a_0+n-y)}{\prod_{i=0}^{r}\Gamma(a_i)\Gamma(\sum_{i=0}^{r} a_i+n)},$$

where

$$a_i>0,\ i=0, 1, \cdots, r; \quad n: \text{positive integer},$$

$$x_i=0, 1, 2, \cdots; \quad \sum_{i=1}^{r} x_i=y=0, 1, \cdots, n.$$

The factorial moments are given by

$$E(\prod_{i=1}^{r} X_i^{(s_i)}) = \frac{n^{(s)}\prod_{i=1}^{r}(a_i+s_i-1)^{(s_i)}}{(\sum_{i=0}^{r} a_i+s-1)^{(s)}}, \qquad s=\sum_{i=1}^{r} s_i . \tag{A.11}$$

If $a_0, a_1, \cdots, a_r \to \infty$, fixing the ratios $a_i/\sum_{i=0}^{r} a_i = p_i$ constant, the limit distribution of (A.10) is the Mn (2.3), and if $a_0, n\to\infty$, fixing $a_0/(a_0+n)=p$, then the limit is the joint distribution of independent NBn.

$$\prod_{i=1}^{r} \frac{\Gamma(a_i+x_i)}{\Gamma(a_i)x_i\,!} p^{a_i} q^{x_i} . \tag{A.12}$$

Let $X_0, X_1, \cdots, X_r$ be independent variates with X_i having

$$\frac{\Gamma(a_i+x_i)}{\Gamma(a_i)x_i\,!} p^{a_i} q^{x_i} .$$

Then the conditional distribution of $(X_1, \cdots, X_r)$ given $\sum_{i=0}^{r} X_i = n$ is (A.10).

A multivariate extension of inverse sampling without replacement induces the distribution treated in 'f' of section 3 and not (A.10). The compound of the Mn (2.3) by the multivariate beta distribution

$$\frac{1}{B(a_0, a_1, \cdots, a_r)} \prod_{i=0}^{r} p_i^{a_i-1}, \quad 0 \leqq p_i , \ \sum_{i=0}^{r} p_i = 1$$

is (A.10). (See Ishii and Hayakawa [12]).

Multivariate extension of the model of exceedance, the occupancy problem and the Pólya's urn model are straightforward.

THE INSTITUTE OF STATISTICAL MATHEMATICS
FACULTY OF ENGINEERING, UNIVERSITY OF TOKYO
THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] A. M. Adelstein, "Accident proneness: A criticism of the concept based upon an analysis of shunter's accidents," *J. Roy. Statist. Soc.*, (*A*), 115 (1952), 354-410.

[2] A. G. Arbous and H. S. Sichel, "New techniques for the analysis of absenteeism data," *Biometrika*, 41 (1954), 77-90.

[3] J. J. Bartko, "The negative binomial distribution: A review of properties and applications," *The Virginia Journal of Science*, 12 (1961), 18-37.

[4] G. F. Bates and J. Neyman, "Contributions to the theory of accident proneness, 'I. An optimistic model of the correlation between light and severe accidents' and 'II. True or false contagion'," *University of California Publications in Statistics*, 1 (1952), 215-275.

[5] B. Epstein, "Tables for the distribution of the number of exceedances," *Ann. Math. Statist.*, 25 (1954), 762-768.

[6] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 1, 2nd Ed., John Wiley, New York, 1957.

[7] R. A. Fisher, "Note on the efficient fitting of the negative binomial," *Biometrics*, 9 (1954), 197-199.

[8] E. J. Gumbel and H. von Schelling, "The distribution of the number of exceedances," *Ann. Math. Statist.*, 21 (1950), 247-262.

[9] J. Gurland, "Some applications of the negative binomial and other contagious distributions," *Amer. J. Pub. Health*, 49 (1959), 1388-1399.

[10] J. W. Hopkins, "An instance of negative hypergeometric in practice," *Bull. Int. Stat. Inst.*, 34 (1955), Book 4, 298-306.

[11] J. O. Irwin, "A distribution arising in the study of infectious diseases," *Biometrika*, 41 (1954), 266-268.

[12] G. Ishii and R. Hayakawa, "On the compound binomial distribution," *Ann. Inst. Stat. Math.*, 12 (1960), 69-80.

[13] N. I. Johnson, "Uniqueness of a result in the theory of accidental-proneness," *Biometrika*, 44 (1957), 530-531.

[14] C. D. Kemp and A. W. Kemp, "Generalized hypergeometric distributions," *J. Roy. Statist. Soc.*, (*B*), 18 (1956), 202-211.

[15] E. L. Lehmann, *Testing Statistical Hypotheses*, John Wiley, New York, 1959.

[16] J. E. Mosimann, "On the compound negative multinomial distribution and correlations among inversely sampled pollen counts," *Biometrika*, 50 (1963), 47-54.

[17] J. Neyman, "Contribution to the theory of the χ^2 test," *Proc. Berkeley Symposium on Math. Stat. and Prob.*, 1949, 239-273.

[18] G. P. Patil, "Minimum variance unbiased estimation and certain problems of additive number theory," *Ann. Math. Statist.*, 34 (1963), 1050-1056.

[19] K. Sarkadi, "On the distribution of the number of exceedances," *Ann. Math. Statist.*, 28 (1957), 1021-1023.

[20] K. Sarkadi, "Generalized hypergeometric distributions," *Publications of the Mathematical Institute of the Hungarian Academy of Science*, 11 (1957), 59-69.

[21] J. G. Skellam, "A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials," *J. Roy. Statist. Soc.*, (*B*) 10 (1948), 257-261.

[22] H. S. Steyn, "On χ^2-tests for contingency tables of negative multinomial types," *Statistica Neerlandica*, 13 (1959), 433-444.

[23] I. Yoshimura, "Unified system of cumulant recurrence relations," *Rep. Stat. Appl. Res., JUSE*, 11 (1964), 1-8.

[24] I. Yoshimura, "A complementary note on the multivariate moment recurrence equation," *Rep. Stat. Appl. Res., JUSE*, 11 (1964) 9-12.

[25] A. J. B. Wiid, "On the moments and regression equations of the fourfold negative and fourfold negative factorial binomial distributions," *Proc. Roy. Soc. Edin.*, (*A*), 65 (1957 58), 29 34.

[26] S. S. Wilks, *Mathematical Statistics*, John Wiley, New York, 1962.

[27] J. Wishart, "Cumulants of multivariate multinomial distributions," *Biometrika*, 36 (1949), 47-58.

Table 1

	NMn (2.1) in terms of p's	NMn (2.1) in terms of θ's	Mn (2.3)
$p.g.f.$	$p_0^k(1-\sum_{j=1}^r p_j t_j)^{-k}$	$(1-\theta_0+\sum_{j=1}^r \theta_j t_j)^{-k}$	$(1-\rho_0+\sum_{j=1}^r \sigma_j t_j)^n$
$ch.f.$	$p_0^k(1-\sum_{j=1}^r p_j e^{it_j})^{-k}$	$(1-\theta_0+\sum_{j=1}^r \theta_j e^{it_j})^{-k}$	$(1-\rho_0+\sum_{j=1}^r \rho_j e^{it_j})^n$
$c.g.f.$	$k\{(\log p_0-\log(1-\sum_{j=1}^r p_j e^{t_j})\}$	$-k\log\{1-\theta_0+\sum_{j=1}^r \theta_j e^{t_j}\}$	$n\log\{1-\rho_0+\sum_{j=1}^r \rho_j e^{t_j}\}$
$f.m.g.f.$	$(1-\sum_{j=1}^r p_j t_j/p_0)^{-k}$	$(1-\sum_{j=1}^r \theta_j t_j)^{-k}$	$(1+\sum_{j=1}^r \rho_j t_j)^n$
factorial moment $\mu_{[s_1\cdots s_r]}$ $s_1+\cdots+s_r=s$	$(k+s-1)^{(s)}\prod_{j=1}^r p_j^{s_j}/p_0^s$	$(-k)^{(s)}\prod_{j=1}^r \theta_j^{s_j}$	$n^{(s)}\prod_{j=1}^r \rho_j^{s_j}$
cumulants		(put $\eta_i=1-\theta_i$)	(put $\tau_i=1-\rho_i$)
$\kappa_{\cdot 1\cdot\cdot}$	kp_i/p_0	$-k\theta_i$	$n\rho_i$
$\kappa_{\cdot 2\cdot\cdot}$	$kp_i(p_0+p_i)/p_0^2$	$-k\theta_i\eta_i$	$n\rho_i\tau_i$
$\kappa_{\cdot 11\cdot}$	kp_ip_j/p_0^2	$k\theta_i\theta_j$	$n\rho_i\rho_j$
$\kappa_{\cdot 3\cdot\cdot}$	$kp_i(p_0+p_1)(p_0+2p_i)/p_0^3$	$-k\theta_i\eta_i(\eta_i-\theta_i)$	$n\rho_i\tau_i(\tau_i-\rho_i)$
$\kappa_{\cdot 21\cdot}$	$kp_ip_j(p_0+2p_i)/p_0^3$	$k\theta_i\theta_j(\eta_i-\theta_i)$	$n\rho_i\rho_j(\tau_i-\rho_i)$
$\kappa_{\cdot 111\cdot}$	$-2kp_ip_jp_l/p_0^3$	$-2k\theta_i\theta_j\theta_l$	$2n\rho_i\rho_j\rho_l$
$\kappa_{\cdot 4\cdot\cdot}$	$kp_i(p_0+p_i)p_0^{-2}$ $\times(1-6p_i(p_0+p_i)p_0^{-2})$	$-k\theta_i\eta_i(1-6\theta_i\eta_i)$	$n\rho_i\tau_i(1-6\rho_i\tau_i)$
$\kappa_{\cdot 31\cdot}$	$kp_ip_jp_0^{-2}$ $\times(1-6p_i(p_0+p_i)p_0^{-2})$	$k\theta_i\theta_j(1-6\theta_i\eta_i)$	$n\rho_i\rho_j(1-6\rho_i\tau_i)$
$\kappa_{\cdot 22\cdot}$	$kp_ip_j\{(p_0+2p_i)(p_0+2p_j)$ $+p_ip_j\}p_0^{-4}$	$k\theta_i\theta_j\{(\eta_i-\theta_i)(\eta_j-\theta_j)$ $+2\theta_i\theta_j\}$	$n\rho_i\rho_j\{(\tau_i-\rho_i)(\tau_j-\rho_j)$ $+2\rho_i\rho_j\}$
$\kappa_{\cdot 211}$	$2kp_ip_jp_l(p_0+3p_i)/p_0^4$	$-2k\theta_i\theta_j\theta_l(\eta_i-2\theta_i)$	$2n\rho_i\rho_j\rho_l(\tau_i-2\rho_i)$
$\kappa_{\cdot 1111\cdot}$	$6kp_ip_jp_lp_m/p_0^4$	$6k\theta_i\theta_j\theta_l\theta_m$	$6n\rho_i\rho_j\rho_l\rho_m$

Table 2

$$Cor(X_i, X_j) = \sqrt{p_i p_j/(p_0+p_i)(p_i+p_j)} = \sqrt{\theta_i\theta_j/\eta_i\eta_j}$$
$$= \sqrt{\mu_i\mu_j/(\mu_i+k)(\mu_j+k)}$$

$$E(X_j \mid X_{J_0}=x_{J_0}) = E(X_j \mid \textstyle\sum_{J_0} X_i = y_0) = \frac{(k+y_0)p_j}{(p_0+\sum_{J_0} p_i)}$$

$$V(X_j \mid X_{J_0}=x_{J_0}) = V(X_j \mid \textstyle\sum_{J_0} X_i = y_0)$$
$$= \frac{(k+y_0)p_j(p_0+\sum_{J_0} p_i + p_j)}{(p_0+\sum_{J_0} p_i)^2}$$

recurrence formula for cumulants

$$\kappa_{s_1 \cdots s_{i+1} \cdots s_r} = p_i \frac{\partial}{\partial p_i} \kappa_{s_1 \cdots s_i \cdots s_r}$$
$$= \theta_i \left(\eta_i \frac{\partial}{\partial \theta_i} - \sum_{l \neq i, 0} \theta_l \frac{\partial}{\partial \theta_l} \right) \kappa_{s_1 \cdots s_i \cdots s_r}$$

recurrence formula for moments about $\boldsymbol{a} = (a_1, \cdots, a_2)$, which may be a parametric function

$$\mu_{s_1 \cdots s_{i+1} \cdots s_r}(\boldsymbol{a}) = p_i \frac{\partial \eta_{s_1 \cdots s_r}(\boldsymbol{a})}{\partial p_i} + \left(\frac{kp_i}{p_0} - a_i \right) \mu_{s_1 \cdots s_r}(\boldsymbol{a})$$
$$+ \sum_{j=1}^{r} p_i \frac{\partial a_l}{\partial p_i} s_l \mu_{s_1 \cdots s_j - 1 \cdots s_r}(\boldsymbol{a})$$