A COMPOUND BALL AND URN PROBABILITY PROBLEM

BU-37-M D. S. Robson April, 1953

We shall examine some characteristics of a sequential ball and urn experiment. An urn contains m black balls and u white balls. A random sample of Z_1 balls is drawn without replacement; if the number X_1 of black balls in the sample is equal to or greater than some preassigned positive integer k then the experiment terminates; otherwise, the Z_1 balls are replaced and a random sample of Z_2 balls is drawn. If $X_1 + X_2 \geq k$, the process terminates; otherwise, the Z_2 balls are replaced and a sample of Z_3 balls is drawn. The process thus continues until the total number of black balls drawn equals or exceeds the integer k. We shall assume, in addition, that the sample sizes Z_1, Z_2, Z_3, ... are independent and identically distributed binomial chance variables,

$$b(Z_i;\ m+u,\ p) \ = \ C_{Z_i}^{m+u} p^{Z_i} q^{m+u-Z_i} .$$

The three chance variables generated by this procedure which are to be considered here are (i) the number N of trials required to terminate the experiment, (ii) the total number $X = \sum_{i=1}^{N} X_i$ of black balls drawn and (iii) the total number $Y = \sum_{i=1}^{N}(Z_i - X_i)$ of white balls drawn.

The Distribution of N

The event $N > n$ is equivalent to the event $X_1 + X_2 + \ldots + X_n < k$, and since X_1, X_2, ... are independent and identically distributed binomial variables with

$$b(x_i;\ m,p) \ = \ C_{x_i}^{m} p^{x_i} q^{m-x_i}$$

then

$$P(X_1 + X_2 + \ldots + X_n < k) \ = \ \sum_{t=0}^{k-1} b(t;\ mn,p),$$

or

$$P(N \leq n) = 1 - P(X_1 + X_2 + \dots + X_n < k) = 1 - \sum_{t=0}^{k-1} b(t;\, mn,\, p).$$

Hence, the distribution of N is obtained as

$$P(N = n) = P(N \leq n) - P(N \leq n-1) = \sum_{t=0}^{k-1} b(t;\, m(n-1), p) - b(t; mn,\, p).$$

The expected number of trials is, therefore,

$$E(N) = \sum_{n=1}^{\infty} \sum_{t=0}^{k-1} n \left[b(t;\, m(n-1),\, p) - b(t;\, mn,\, p) \right]$$

$$= \sum_{n=1}^{\infty} \sum_{t=0}^{k-1} b(t;\, m(n-1),\, p)$$

$$= \sum_{t=0}^{k-1} \left[\sum_{n=1}^{\infty} b(t;\, m(n-1),\, p) \right]$$

$$= \sum_{n=0}^{\infty} q^{mn} + \sum_{n=1}^{\infty} C_1^{mn}\, pq^{mn-1} + \dots + \sum_{n=1}^{\infty} C_{k-1}^{mn}\, p^{k-1}\, q^{mn-k+1}$$

$$= \sum_{n=0}^{\infty} q^{mn} + \frac{p}{1!} \sum_{n=1}^{\infty} (mn)_1\, q^{mn-1} + \dots + \frac{p^{k-1}}{(k-1)!} \sum_{n=1}^{\infty} (mn)_{k-1} q^{mn-k+1}$$

$$= \sum_{n=0}^{\infty} q^{mn} + \frac{p}{1!} \frac{d}{dq} \left[\sum_{n=0}^{\infty} q^{mn} \right] + \dots + \frac{p^{k-1}}{(k-1)!} \frac{d^{k-1}}{dq^{k-1}} \left[\sum_{n=0}^{\infty} q^{mn} \right]$$

$$= \sum_{t=0}^{k-1} \frac{p^t}{t!} \frac{d^t Q_m}{dq^m}$$

where

$$Q_m = \sum_{n=0}^{\infty} q^{mn} = \frac{1}{1-q^m} \ .$$

The variance of this distribution is found as follows:

$$E(N^2) = \sum_{n=1}^{\infty} \sum_{t=0}^{k-1} n^2 \left[b(t;\, m(n-1),\, p) - b(t;\, mn,\, p) \right]$$

$$= \sum_{n=1}^{\infty} \sum_{t=0}^{k-1} (2n-1)\, b(t;\, m(n-1),\, p)$$

$$= 2 \sum_{t=0}^{k-1} \sum_{n=0}^{\infty} nb(t;\, mn,\, p) + E(N)$$

$$= E(N) + 2 \left[\sum_{n=1}^{\infty} nq^{mn} + \sum_{n=1}^{\infty} n\, C_1^{mn}\, pq^{mn-1} + \dots + \sum_{n=1}^{\infty} n\, C_{k-1}^{mn} p^{k-1} q^{mn-k+1} \right.$$

$$= E(N) + 2\left[\sum_{n=1}^{\infty} nq^{mn} + \frac{p}{1!}\sum_{n=1}^{\infty} n(mn)_1 q^{mn-1} + \dots \right.$$

$$\left. + \frac{p^{k-1}}{(k-1)!}\sum_{n=1}^{\infty} n(mn)_{k-1} q^{mn-k+1}\right]$$

$$= E(N) + 2\left[\sum_{n=1}^{\infty} nq^{mn} + \frac{p}{1!}\sum_{n=1}^{\infty}\frac{d(nq^{mn})}{dq} + \dots + \frac{p^{k-1}}{(k-1)!}\sum_{n=1}^{\infty}\frac{d^{k-1}(nq^{mn})}{dq^{k-1}}\right]$$

where

$$\sum_{n=1}^{\infty} nq^{mn} = \frac{q}{m}\sum_{n=1}^{\infty} mnq^{mn-1} = \frac{q}{m}\frac{d}{dq}\left(\sum_{n=0}^{\infty} q^{mn}\right) = \frac{q}{m}\frac{dQ_m}{dq} .$$

Hence,

$$E(N^2) = E(N) + \frac{2}{m}\sum_{t=0}^{k-1}\frac{p^t}{t!}\frac{d^t}{dq^t}\left(q\frac{dQ_m}{dq}\right)$$

$$= E(N) + \frac{2}{m}\left[qQ_m' + p(qQ_m'' + Q_m') + \frac{p^2}{2!}(qQ_m''' + 2Q_m'') + \dots \right.$$

$$\left. + \frac{p^{k-1}}{(k-1)!}\left(qQ_m^{(k)} + (k-1)Q_m^{(k-1)}\right)\right]$$

$$= E(N) + \frac{2}{m}\left[Q_m' + \frac{p}{1!}Q_m'' + \dots + \frac{p^{k-2}}{(k-1)!}Q_m^{(k-1)} + \frac{p^{k-1}}{(k-1)!}qQ_m^{(k)}\right]$$

$$= E(N) + \frac{2}{m}\left[Q_m' + \frac{p}{1!}Q_m'' + \dots + \frac{p^{k-1}}{(k-1)!}Q_m^{(k)} - \frac{p^k}{(k-1)!}Q_m^{(k)}\right]$$

$$= E(N) + \frac{2}{m}\left[\frac{d\ E(N)}{dq} - \frac{p^k}{(k-1)!}\frac{d^kQ_m}{dq^k}\right]$$

$$\text{var}(N) = \frac{2}{m}\left[q\frac{d\ E(N)}{dq} + p\frac{d\ E(N)}{dp}\right] - E(N)\left[E(N) - 1\right].$$

We also note that

$$E\left(\frac{1}{N}\right) = \sum_{t=0}^{k-1}\left[\sum_{n=1}^{\infty}\frac{b(t;\ m(n-1),\ p)}{n} - \sum_{n=1}^{\infty}\frac{b(t;\ mn,\ p)}{n}\right]$$

$$= \sum_{t=0}^{k-1} \frac{p^t}{t!} \left[\sum_{n=0}^{\infty} \frac{(mn)t}{n+1} q^{mn-t} - \sum_{n=1}^{\infty} \frac{(mn)t}{n} q^{mn-t} \right]$$

$$= \sum_{t=0}^{k-1} \frac{p^t}{t!} \left[\sum_{n=0}^{\infty} \frac{d^t}{dq^t} \left(\frac{q^{mn}}{n+1} \right) - \sum_{n=1}^{\infty} \frac{d^t}{dq^t} \left(\frac{q^{mn}}{n} \right) \right]$$

$$= \sum_{t=0}^{k-1} \frac{p^t}{t!} \frac{d^t}{dq^t} \left[\sum_{n=0}^{\infty} \frac{q^{mn}}{n+1} - \sum_{n=1}^{\infty} \frac{q^{mn}}{n} \right]$$

$$= \sum_{t=0}^{k-1} \frac{p^t}{t!} \frac{d^t}{dq^t} \left[- \frac{\ln(1-q^m)}{q^m} + \ln(1-q^m) \right]$$

$$= \sum_{t=0}^{k-1} \frac{p^t}{t!} \frac{d^t}{dq^t} \left[\frac{(1-q^m)}{q^m} \ln\left(\frac{1}{1-q^m}\right) \right]$$

$$= \sum_{t=0}^{k-1} \frac{p^t}{t!} \frac{d^t}{dq^t} \left[\frac{\ln Q_m}{q^m Q_m} \right].$$

The Distribution of X

$$P(X=x \mid N=n) = \frac{1}{P(N=n)} \sum_{t=0}^{k-1} P(X_1 + \dots + X_{n-1} = t)\, P(X_n = x-t)$$

$$= \frac{1}{P(N=n)} \sum_{t=0}^{k-1} b(t;\, m(n-1),\, p) \cdot b(x-t;\, m,\, p)$$

$$= \frac{1}{P(N=n)} \sum_{t=0}^{k-1} C_t^{m(n-1)} p^t q^{m(n-1)-t} C_{x-t}^{m} p^{x-t} q^{m-x+t}$$

$$= \frac{p^x q^{mn-x}}{P(N=n)} \sum_{t=0}^{k-1} C_t^{m(n-1)} C_{x-t}^{m}$$

$$P(X=x) = p^x \sum_{n=1}^{\infty} q^{mn-x} \sum_{t=0}^{k-1} C_t^{m(n-1)} C_{x-t}^{m}$$

$$= p^x \sum_{t=0}^{k-1} C_{x-t}^{m} \frac{q^{m-x+t}}{t!} \sum_{n=0}^{\infty} (mn)_t q^{mn-t}$$

$$= p^x \sum_{t=0}^{k-1} C_{x-t}^{m} \frac{q^{m-x+t}}{t!} \frac{d^t Q_m}{dq^m}$$

$$E(X) = \sum_{x=k}^{m+k-1} xp^x \sum_{t=0}^{k-1} C^m_{x-t} \frac{q^{m-x+t}}{t!} \frac{d^tQ_m}{dq^t}$$

$$= \sum_{t=0}^{k-1} \frac{p^t}{t!} \frac{d^tQ_m}{dq^t} \sum_{x=k}^{m+k-1} xC^m_{x-t}p^{x-t}q^{m-x+t}$$

$$= mpE(N) + \sum_{t=0}^{k-1} t \frac{p^t}{t!} \frac{d^tQ_m}{dq^t}$$

$$- \sum_{s=0}^{k-1} C^m_{k-s-1}p^{k-s-1}q^{m-k+s+1} \sum_{t=0}^{s}(t+k-s-1) \frac{p^t}{t!} \frac{d^tQ_m}{dq^t}$$

The Distribution of Y

$$P(Y=y|N=n) = C^{nu}_y p^y q^{nu-y}$$

$$E(Y) = E[E(Y|N)] = E(Nup) = upE(N)$$

$$E(Y^2) = E[E(Y^2|N)] = E[(Nup)^2 + Nupq] = u^2p^2E(N^2) + upqE(N)$$

$$\mathrm{var}(Y) = u^2p^2\mathrm{var}(N) + upqE(N)$$

The Joint Distribution of N, X, Y

$$P(X=x,\ Y=y \mid N=n) = P(X=x \mid N=n)\cdot P(Y=y \mid N=n)$$

$$= \frac{1}{P(N=n)} C^{nu}_y p^{x+y} q^{n(m+u)-(x+y)} \sum_{t=0}^{k-1} C_t^{m(n-1)} C^m_{x-t}$$

$$P(X=x,\ Y=y,\ N=n) = C^{nu}_y p^{x+y} q^{n(m+u)-(x+y)} \sum_{t=0}^{k-1} C_t^{m(n-1)} C^m_{x-t}$$

Some Results for the case k=1

$$P(N=n) = \frac{q^{m(n-1)}}{Q_m},\ E(N) = Q_m,\ E\left(\frac{1}{N}\right) = \frac{\ln Q_m}{Q_m q^m},\ \mathrm{var}(N) = q^m Q_m^2$$

$$P(X=x) = C^m_x p^x q^{m-x} Q_m,\ E(X) = mpQ_m,\ \mathrm{var}(X) = mpqQ_m$$

$$E(Y) = upQ_m,\ \mathrm{var}(Y) = upqQ_m + u^2p^2q^mQ_m^2$$

$$P(X=x,\ Y=y,\ N=n) = C^m_x C^{nu}_y p^{x+y} q^{n(m+u)-(x+y)}$$

$$E\left[\frac{m+1}{x+1} - m(N-1)\right] = \frac{1}{p}$$

$$E\left[Y(N-1)\right] = 2E(Y)E(N-1) = 2upq^m Q_m^2$$

$$E\left[X(N-1)\right] = E(X)E(N-1) = mpq^m Q_m^2$$

$$EY\left[\frac{m+1}{x+1} + \frac{m(N-1)}{2}\right] = uQ_m$$