ON CLASSICAL OCCUPANCY PROBLEMS II
(Sequential Occupancy)

by

ANDRÁS BÉKÉSSY

1. Introduction. We consider, as in Part I [1], a random distribution of balls in n cells, assuming that the balls are randomly and independently dropped into the cells with the same probability $1/n$. In Part I the number of the balls was taken to be fixed, and the "state" of the set of cells was regarded a random variable. Suppose now that certain parameters, characterising a "state" of cells, are fixed, while the random variable is the number of balls necessary to reach that given state. Let be k the number of cells, which contain less than $m+1$ $(m = 0, 1, 2, \ldots)$ balls, and let be $\nu(n, m, k)$ the number of independent throws.

Most results concern the random variable $\nu(n, 0, k)$ i.e. the number of balls needed to obtain at least one ball in each, except k cells $(k = 0, 1, 2, \ldots)$. Probability distributions, moments and limiting distributions related to $\nu(n, 0, k)$ have been determined [2], [3]. D. J. NEWMANN and L. SHEPP and later on P. ERDŐS and A. RÉNYI have dealt with the expectation and with the limiting distribution of $\nu(n, m, 0)$ [4], [5].[1]

In the present paper two theorems on the limiting distribution of $\nu(n, m, k)$ are given.

2. Probability generating function. The first statement is that for $k < n$ the generating function

$$\sum_{N=0}^{\infty} \mathbf{P}\{\nu(n, m, k) = N\} \cdot x^N$$

can be expressed in the integral form

$$n\binom{n-1}{k}\int_0^{\infty} \exp\left\{nu\left(1-\frac{1}{x}\right)\right\}(1-K_m(u))^{n-k-1}(K_m(u))^k H_m(u)\,du, \tag{1}$$

where

$$H_m(x) = \frac{x^m e^{-x}}{m!}, \quad K_m(u) = \sum_{\mu=0}^{m} H_\mu(u) = \frac{1}{m!}\int_u^{\infty} t^m e^{-t}\,dt.$$

[1] Let p_j $(j = 1, 2, \ldots, n)$ denote the probability of a ball falling into the j-th cell. Papers discussing $\nu(n, m, k)$ under the more general assumption that the p_j's may be different from each other, will be referred later.

Proof. The probability of the event that after the N-th throw there remain $k+1$ cells occupied by not more than m balls, while the $N+1$-th ball is falling into a cell, which contains m balls already, depends only on the number of cells occupied by m and that occupied by less than m balls. Let $p(n, N, m, l_1, l_2)$ be the probability that after N throws there remain l_1 cells having m and l_2 cells having less than m balls. This probability can be expressed by the G-function (13, Part I); it is easy to see that the G-function of $p(n, N, m, l_1, l_2)$ defined as

$$\sum_{N=0}^{\infty} \frac{(nz)^N}{N!} \cdot \sum_{l_1=0}^{\infty} \sum_{l_2=0}^{\infty} p(n, N, m, l_1, l_2)\, x_1^{l_1} x_2^{l_2}$$

is equal to

$$e^{nz}\,[1 + (x_1 - 1)\, H_m(z) + (x_2 - 1)\, K_{m-1}(z)]^n ,$$

hence

$$\sum_{N=0}^{\infty} p(n, N, m, l_1, l_2) \frac{(nz)^N}{N!} =$$

(2)

$$= \frac{n!}{l_1!\, l_2!\, (n - l_1 - l_2)!}\, e^{nz}(1 - K_m(z))^{n-l_1-l_2} (K_{m-1}(z))^{l_2} (H_m(z))^{l_1} .$$

Since the probability $q(n, N+1, m, k)$ of the event

$$\nu\,(n, m, k) = N + 1$$

is equal to

$$\sum_{l_1+l_2=k+1} p(n, N, m, l_1, l_2) \cdot \frac{l_1}{n} ,$$

it follows from (2) that

$$\sum_{N=0}^{\infty} q(n, N+1, m, k) \frac{(nz)^N}{N!} = \binom{n-1}{k} e^{nz}(1 - K_m(z))^{n-k-1} (K_m(z))^k H_m(z) . \tag{3}$$

The function on the left-hand side of (3) being the Borel-transform of the probability generating function, the latter can be written as

$$\binom{n-1}{k} \int_0^{\infty} e^{nzt}(1 - K_m(zt))^{n-k-1} (K_m(zt))^k H_m(zt)\, e^{-t}\, dt \tag{4}$$

so that the final result (1) now immediately follows.

Putting $K_m(u) = v$, integral (1) takes the form

$$n \binom{n-1}{k} \int_0^{1} v^k(1 - v)^{n-k-1} \exp\left\{n u_m(v) \left(1 - \frac{1}{x}\right)\right\} dv , \tag{5}$$

where $u_m(v)$ is the inverse of $v = K_m(u)$.

From (5) the expectations $\mathbf{E}\{\nu(n, m, k)\}$ and $\mathbf{E}\{\nu^2(n, m, k)\}$ are

$$n^2 \binom{n-1}{k} \int_0^1 u_m(v)\, v^k (1-v)^{n-k-1}\, dv \tag{6a}$$

and

$$n^3 \binom{n-1}{k} \int_0^1 (u_m(v))^2\, v^k (1-v)^{n-k-1}\, dv - \mathbf{E}\{\nu(n, m, k)\} \tag{6b}$$

respectively.

3. The first limit theorem. *If* $n \to \infty$, $m = \text{const.}$, $k = \text{const.}$, *then*

$$\mathbf{P}\left\{\frac{\nu(n,m,k)}{n} - \log n - m \log\log n - \log\frac{1}{m!} < x\right\} \to$$

$$\to \exp\{-e^{-x}\} \cdot \sum_{\mu=0}^{k} \frac{e^{-\mu x}}{\mu!}$$

holds.

Proof. Introducing the moment generating function by putting $x = e^s$ in (5), and taking for brevity $\vartheta = n(1 - e^{-s/n})$ the relation

$$\begin{aligned} \Psi\left(\frac{s}{n}\right) &= n \binom{n-1}{k} \int_0^\infty v^k (1-v)^{n-k-1} e^{\vartheta u_m(v)}\, dv \sim \\ &\sim \frac{\Gamma(k+1-s)}{\Gamma(k+1)} \exp\left\{s\left(\log n + \text{m} \log\log n + \log\frac{1}{m!}\right)\right\} \end{aligned} \tag{7}$$

has to be proved for $-1/2 \leq s \leq +1/2$ according to Curtiss' theorem used already in Part I.

The asymptotic behaviour of the integral (7) is determined by the values taken up by the integrand in the close vicinity of $v = +0$. Let us divide the integral (7) in two parts:

$$\int_0^\infty = \int_0^\delta + \int_\delta^\infty = I_1 + I_2, \tag{8}$$

where $\delta > 0$ is some constant. In I_2 the variable v cannot be smaller than δ and ϑ being bounded from above as $n \to \infty$, it follows

$$I_2(n) < C(\delta)\,(1-\delta)^n, \tag{9}$$

where $C(\delta)$ may depend on δ but not on n.

In order to determine I_1, the behaviour of the function $u_m(v)$ for $v \to +0$ has to be found. From the definition it is easy to get

$$u_m(v) = \log\frac{1}{v} + m \log\log\frac{1}{v} - \log m! + \eta(v) \tag{10}$$

with $\eta(v) = o(1)$ for $v \to +0$. Hence putting $w = nv$ we have from (7) and (8)

$$I_1 = \frac{(\log n)^{m\vartheta}}{(m!)^{\vartheta} n^{k+1-\vartheta}} \int_0^{\delta n} w^{k-\vartheta} \left(1 - \frac{w}{n}\right)^{n-k-1} dw +$$

$$+ \frac{(\log n)^{m\vartheta}}{(m!)^{\vartheta} n^{k+1-\vartheta}} \int_0^{\delta n} w^{k-\vartheta} \left(1 - \frac{w}{n}\right)^{n-k-1} \left[\left(1 - \frac{\log w}{\log n}\right)^{m\vartheta} e^{\vartheta\eta(w/n)} - 1\right] dw = I_{11} + I_{12}.$$

Since ϑ tends to s uniformly in $-{}^1\!/_2 \leq s \leq +{}^1\!/_2$ for $n \to \infty$, Laplace's method applied to I_{11} gives

$$(11) \quad I_{11} \sim \frac{(\log n)^{ms}}{(m!)^s n^{k+1-s}} \int_0^{\infty} w^{k-s} e^{-w} dw = \frac{\Gamma(k+1-s)}{n^{k+1}} e^{s(\log n + m \log\log n - \log m!)},$$

while the modulus of the term I_{12} is not larger than

$$\frac{(\log n)^{m\vartheta}}{(m!)^{\vartheta} n^{k+1-\vartheta}} \zeta(\delta)$$

with $\lim_{\delta \to 0} \zeta(\delta) = 0$, by the obvious inequality

$$(12) \quad \left|\left(1 - \frac{\log w}{\log n}\right)^{m\vartheta} e^{\eta\vartheta} - 1\right| \leq \left|m\,\vartheta \frac{\log w}{\log n} e^{\eta\vartheta}\right| e^{|m\vartheta \log w/\log n|} + |e^{\eta\vartheta} - 1|.$$

Now, although the value of δ must be kept constant while $n \to \infty$, it may be taken arbitrarily small such that from (11) and (12) we obtain

$$(13) \quad I_1 \sim I_{11} \quad (n \to \infty)$$

and the summarizing of the partial results (9), (11) and (13) leads to (7) we wanted to prove.

4. The second limit theorem. *If* $n \to \infty$, $m = \text{const.}$, $\beta = \text{const.}$, $(0 < \beta < 1)$, $k = n\beta$ *then*

$$(14) \quad \mathbf{P}\left\{\frac{\nu(n, m, k) - E}{D} < x\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

with

$$(15) \quad \begin{cases} E = n\,u_m(\beta), \\ D^2 = n\big(\beta(1-\beta)(u_m'(\beta))^2 - u_m(\beta)\big), \end{cases}$$

where the function $u_m(x)$ *is the inverse of* $x = K_m(u)$ *and* $u_m'(x)$ *is the derivative.*

The quantity D^2 is positive, because it is the asymptotic value of the variance $\mathbf{D}^2\{\nu(n, m, k)\}$ for $n \to \infty$ under the assumptions on k and m mentioned above. This can be seen by deducing asymptotic expressions for

$\mathbf{E}\{\nu(n, m, k)\}$ and $\mathbf{E}\{\nu^2(n, m, k)\}$ (6a, 6b). Laplace's method gives in the present case (see [6])

$$\mathbf{E}\{\nu(n, m, k)\} = nu_m(\beta) + (1-\beta)\, u_m'(\beta) + \frac{1}{2}\beta(1-\beta)\, u_m''(\beta) + O(n^{-1})$$

and

$$\mathbf{E}\{\nu^2(n, m, k)\} = n^2 u_m^2(\beta) + 2\,n(1-\beta)\, u_m(\beta)\, u_m'(\beta) + \\ + n\,\beta(1-\beta)\, u_m(\beta)\, u_m''(\beta) - nu_m(\beta) + O(1)$$

hence

$$\mathbf{D}^2\{\nu(n, m, k)\} = \mathbf{E}\{\nu^2\} - (\mathbf{E}\{\nu\})^2 = n\,\beta(1-\beta)\,(u_m'(\beta))^2 - nu_m(\beta) + O(1)$$

follows.

Introducing the moment generating function, the statement to be proved will be

$$\Psi\left(\frac{s}{D}\right) = n\binom{n-1}{k}\int_0^\infty e^{nSu}(1-K_m(u))^{n(1-\beta)-1}(K_m(u))^{n\beta} H_m(u)\, du \sim \tag{16}$$
$$\sim \exp\left\{\frac{Es}{D} + \frac{s^2}{2}\right\}.$$

where $S = 1 - e^{-s/D}$. By the definition (15) of D we have

$$S = \frac{s}{D} - \frac{s^2}{2\,D^2} + O(n^{-3/2})\,. \tag{17}$$

Putting

$$f(u) = \beta \log K_m(u) + (1-\beta)\log(1-K_m(u))\,,$$

the moment generating function (16) can be rewritten as

$$\Psi\left(\frac{s}{D}\right) = n\binom{n-1}{k}\int_0^\infty e^{nSu+nf(u)}\,\frac{H_m(u)}{1-K_m(u)}\,du\,. \tag{18}$$

It is easy to show that the function $f'(u)$ is decreasing in $0 \leq u < \infty$, whereas $\lim\limits_{u\to+0} f'(u) = +\infty$, $\lim\limits_{u\to\infty} f'(u) = -\beta$ such that for sufficiently large n there exists a "saddle point" b defined by

$$f'(b) = -S\,. \tag{19}$$

Equation (19) leads to

$$S = H_m(b)\,\frac{\beta - K_m(b)}{K_m(b)\,(1-K_m(b))} \tag{20}$$

or to

$$K_m(b) = \beta - \frac{K_m(b)\,(1-K_m(b))}{H_m(b)}\cdot s = \beta + O(n^{-1/2})\,. \tag{21}$$

— since $S = O(n^{-1/2})$, and since the factor $K_m(b)\big(1 - K_m(b)\big)/H_m(b)$ is bounded from above. This means that the point b lies in the close vicinity of $u_m(\beta)$.

Let us now write $\Psi(s/D)$ in the form

$$\Psi(s/D) = F.J, \tag{22}$$

where the two factors are

$$F = \sqrt{2\pi n}\binom{n-1}{k}\exp\{nSb + nf(b)\} \tag{23}$$

and

$$J = \sqrt{\frac{n}{2\pi}}\int_0^\infty \exp\{nS(u-b) + nf(u) - nf(b)\}\,\frac{H_m(u)}{1 - K_m(u)}\,du\,. \tag{24}$$

The calculations on F, leading to

$$F \sim \beta^{-1/2}(1-\beta)^{1/2}\exp\left\{\frac{Es}{D} + \frac{s^2}{2}\right\} \qquad (n \to \infty) \tag{25}$$

are rather elementary but somewhat cumbersome, so that it does not seem to be superfluous to give some details. First, using Stirling's formula to the factor $\binom{n-1}{k}$ and taking logarithms in (23) we obtain

$$\log F = \log\big(\beta^{-1/2}(1-\beta)^{1/2}\big) + nbS - n\beta\log\left(1 + S\,\frac{1 - K_m(b)}{H_m(b)}\right) -$$
$$- n(1-\beta)\log\left(1 - S\,\frac{K_m(b)}{H_m(b)}\right) + o(1)$$

and then, expanding the logarithms in powers of S up to S^2, we have

$$\begin{aligned}\log F &= \log\big(\beta^{-1/2}(1-\beta)^{1/2}\big) + nbS - nS\,\frac{\beta - K_m(b)}{H_m(b)} + \\ &+ \frac{nS^2}{2}\left(\beta\,\frac{(1-K_m(b))^2}{(H_m(b))^2} + (1-\beta)\,\frac{(K_m(b))^2}{(H_m(b))^2}\right) + o(1)\,.\end{aligned} \tag{26}$$

With regard to (20), the term $\big(\beta - K_m(b)\big)/H_m(b)$ in (26) may be replaced by $SK_m(b)\big(1 - K_m(b)\big)/\big(H_m(b)\big)^2$, and, since by (19) the quantity $K_m(b)$ differs from β only by an $O(n^{-1/2})$ term, the latter can be substituted to the former in (26). Thus the expression (26) for F becomes

$$\log F = \log\big(\beta^{-1/2}(1-\beta)^{1/2}\big) + nb\,S - \frac{nS^2}{2}\,\frac{\beta(1-\beta)}{H_m(b)^2} + o(1)\,. \tag{27}$$

Let us now consider the variable b. From (20) b is equal to

$$u_m\left(\beta - S\,\frac{K_m(b)\big(1 - K_m(b)\big)}{H_m(b)}\right),$$

which can be expanded to give

$$b = u_m(\beta) + \frac{\beta(1-\beta)}{(H_m(b))^2} S + O\left(\frac{1}{n}\right), \tag{28}$$

having substituted $\beta + O(n^{-1/2})$ for $K_m(b)$ and $-(H_m(b))^{-1} + O(n^{-1/2})$ for $u'_m(\beta)$. Replacing b in (27) by the right-hand expression of (28) and then S by $s/D - s^2/2D^2 + O(n^{-3/2})$, the asymptotic relation (25) now easily follows.

The integral J defined in (24) can be rewritten as

$$J = \sqrt{\frac{n}{2\pi}} \int_0^\infty e^{ng_n(u)} \frac{H_m(u)}{1 - K_m(u)} \, du, \tag{29}$$

hence, Laplace's method gives

$$J \sim \frac{H_m(b)}{1 - K_m(b)} \cdot \frac{1}{\sqrt{-g''_n(b)}} \sim \beta^{1/2}(1-\beta)^{-1/2}, \qquad (n \to \infty), \tag{30}$$

and from (25) and (30) we obtain

$$\Psi\left(\frac{s}{D}\right) = F \cdot J \sim \exp\left\{\frac{Es}{D} + \frac{s^2}{2}\right\}$$

as stated in (16).

When applying Laplace's method to J, care must be taken, because the function $g_n(u) = S(u-b) + f(u) - f(b)$ slightly depends on n through S and b. The fact, however, that $S = O(n^{-1/2})$ and $b = \text{const.} + O(n^{-1/2})$ makes possible to carry out the routine estimating procedures.

5. Remark. The proof of the second theorem is now complete, it may be, however, conjectured that the condition implied upon k is too strong; the standardized variable $\nu(n, m, k)$ tends to be normally distributed for $n \to \infty$, $k \to \infty$ even if only $D \to \infty$. In the simple special case $m = 0$ this can be shown as follows.

For $m = 0$ the corresponding expressions will be (see (1), (5), (15))

$$\begin{cases} H_0(u) = K_0(u) = e^{-u}, \quad u_0(v) = \log\dfrac{1}{v}; \\[2mm] \Psi\left(\dfrac{s}{D}\right) = \dfrac{\Gamma(n+1)}{\Gamma(k+1)} \cdot \dfrac{\Gamma(k+1-nS)}{\Gamma(n+1-nS)}, \quad S = 1 - e^{-s/D}; \\[2mm] E = n\log\dfrac{1}{\beta}, \quad D^2 = n\left(\dfrac{1}{\beta} - 1 - \log\dfrac{1}{\beta}\right), \quad \beta = \dfrac{k}{n}; \end{cases} \tag{31}$$

– where now β is not supposed to be constant.

By Stirling's formula we have

$$\log \Psi\left(\frac{s}{D}\right) = (n + 1/2)\log n - (k + 1/2)\log k + \tag{32}$$
$$+ (k + 1/2 - nS)\log(k - nS) - (n + 1/2 - nS)\log(n - nS) + o(1).$$

If $\limsup_{n\to\infty} \beta < 1$, then $S = O(n^{-1/2})$ and the calculations leading to

$$\log \Psi\left(\frac{s}{D}\right) = \frac{Es}{D} + \frac{s^2}{2} + o(1)$$

are obvious, however, if $\beta \to 1$, — since $nS^3 = o(1)$ does not hold in that case — we have to transform (32) by putting $\beta^* = (1 - \beta)/\beta$ in order to lend it the form

$$\log \Psi\left(\frac{s}{D}\right) = nS \log \frac{n}{k} + k(1 - S - \beta^* S) \log \left(1 - \frac{\beta^* S}{1 - S}\right) -$$

$$- k\beta^* \log (1 - S) + o(1) .$$

Since now the estimates

$$n\beta^{*2} S^3 = O(D^{-1}) = o(1) ,$$

$$k\beta^{*3} S^3 = O(n^{-1/2})$$

are valid, it will be enough to expand the logarithmic term containing $\beta^* S/(1 - S)$ up to $(\beta^* S)^2/(1 - S)^2$. With regard to

$$D = O(n^{1/2}\beta^*) , \quad S = \frac{s}{D} - \frac{s^2}{2D^2} + O(n^{-3/2}\beta^{*-3})$$

after some simple calculations we obtain

$$\log \Psi\left(\frac{s}{D}\right) = (1 - e^{-s/D})\left(n \log \frac{1}{\beta} - k\beta^*\right) + \frac{k\beta^* s}{D} + \frac{1}{2}\frac{k\beta^{*2} s^2}{D^2} + o(1) =$$

$$= \frac{Es}{D} - \frac{1}{2}\frac{s^2}{D^2}\left(n \log \frac{1}{\beta} - k\beta^*\right) + \frac{1}{2}\frac{k\beta^{*2} s^2}{D^2} + o(1) =$$

$$= \frac{Es}{D} + \frac{1}{4}\frac{k\beta^{*2} s^2}{D^2} + o(1) ,$$

thus the normal limit distribution law holds, if only D^2 is asymptotically equal to $k\beta^{*2}/2$. In fact, for $\beta \to 1$ the expression (31) of D^2 is asymptotically equal to $k\beta^{*2}/2$.

(Received November 26, 1963)

LITERATURE

[1] BÉKÉSSY, A.: "On Classical Occupancy Problems I." *MTA Mat. Kut. Int. Közleményei* **8** (1963) 59—71.

[2] BARTON, E. E.—DAVID, F. N.: "Sequential Occupancy". *Biometrica* **46** (1959) 218—223.

[3] BARTON, E. E.—DAVID, F. N.: *Combinatorial Chance.* London, 1962 (Griffin).

[4] NEWMAN, D. J.—SHEPP, L.: "The double dixie cup problem." *The American Mathematical Monthly,* **67** (1960) 58—61.

[5] ERDŐS, P.—RÉNYI, A.: "On a classical problem of probability theory." *MTA Mat. Kut. Int. Közleményei* **6** (1961) 215—220.

[6] ФУКС, Б. А. — ЛЕВИН, В. И.: *Функции комплексного переменного и их приложения.* Москва 1951 (ГТТИ).

О КЛАССИЧЕСКИХ ЗАДАЧАХ ЗАПОЛНЕНИЯ ЯЧЕЕК II
(Последовательные проблемы заполнения)

A. BÉKÉSSY

Резюме

Разыгрываем шарики в n ячеек независимо друг от друга и от состояний ячеек. Каждый шарик может попасть в каждую ячейку с вероятностью $1/n$. Пусть будет $\nu(n, m, k)$ случайная величина — число шариков, расположенных в ячейках, в тот момент, когда система ячеек первый раз принимает состояние, в котором хотя бы $n - k$ ячеек содержат хотя бы по $m + 1$ шариков. Рассмотрим предельное распределение $\nu(n, m, k)$, когда n безгранично растёт. Согласно первой предельной теореме, если $n \to \infty$; k; m константы, тогда

$$\mathbf{P}\left\{\frac{\nu(n, m, k)}{n} - \log n - m \log\log n - \log\frac{1}{m!} < x\right\} \to \exp\{-e^{-x}\} \sum_{\mu=0}^{k} \frac{e^{-\mu x}}{\mu!}.$$

Согласно второй предельной теореме, если $n \to \infty$, $k = n\beta$, $\beta =$ константа, $(0 < \beta < 1)$; $m =$ константа $(m = 0, 1, 2, \ldots)$ тогда

$$\mathbf{P}\left\{\frac{\nu(n, m, k) - E}{D} < x\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

где

$$E = n u_m(\beta)$$

$$D^2 = n\big(\beta(1-\beta)\,(u'_m(\beta))^2 - u_m(\beta)\big)$$

$$u_m(\beta) \text{ — обратная функция функции } \beta = \sum_{\mu=0}^{m} \frac{u^\mu}{\mu!} e^{-\mu}$$

$$u'_m(\beta) \text{ — производная функций } u_m(\beta).$$

E — ассимптотическое значение математического ожидания величины $\nu(n, m, k)$ а D^2 — ассимптотическое значение дисперсии этой величины согласно предельному распределению, данному во второй теореме.

Вышеупомянутые теоремы являются обобщениями теорем, известных в литературе [2], [3], [4], [5]. Доказательства исходят из интегралной формы производящей функции моментов величины $\nu(n, m, k)$ доказанной во втором параграфе, и в них основную роль играет метод Лапласа.