BULLETIN OF THE
INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
Volume 15, Number 4, December 1987

A UNIFIED APPROACH TO LIMIT THEOREMS ON RANDOMIZED SEQUENTIAL OCCUPANCY PROBLEMS

BY

HSIAW-CHAN YEH (葉小蓁)

Abstract. Consider n cells into which balls are being dropped independently in such a way that the cells are equiprobable and each ball has probability p of falling through. The experiment is stopped until k (unspecified) cells contain at least b balls each. Let Y_i be the number of balls in the ith cell, $1 \leq i \leq n$, and Y_{n+1} be the number of balls not hitting any cell when stopping. For two given arbitrary functions f and h, the characteristic function of the random variable $Z = \sum_{j=1}^{n} f(Y_j) + h(Y_{n+1})$ is derived. The usefulness of this representation is illustrated by two corollaries.

1. Introduction. In many statistical applications the underlying probabilistic problem can be described in terms of the following urn model (Johnson & Kotz (1977)): Consider n cells into which balls are thrown independently of each other such that the probability of hitting and staying in each cell is $(1-p)/n$. Each ball has probability p of falling through or leaking (hence not being available to fill a cell), the balls are thrown until k unspecified cells contain at least b balls each, $b \geq 1$. Let, for $j = 1, \cdots, n$, Y_j be the content, i.e. the number of balls in cell j when this happens and Y_{n+1} is the number of balls not hitting any cell. The random variable of interest in this paper is $Z = \sum_{j=1}^{n} f(Y_j) + h(Y_{n+1})$, where f and h are any given nonnegative functions. Owing to the arbitrary functions f and h, the definition of Z will give us the more general types of the waiting time problems.

Received by the editors October 3, 1986 and in revised form June 8, 1987.
1986 A. M. S. Subject Classification: 60C05
Key words: Urn model; waiting time; randomized sequential occupancy; truncated poisson distribution; poisson process; order statistics; extreme-valued distribution.

The special case $p=0$ and $b=1$ is the classical sequential occupancy problem. See, for example, Holst (1981). The case $0<p<1$ is called the extended occupancy or the randomized occupancy model. There are several papers concerned about the randomized occupancy problems for the case of $b=1$, see for example, Samuel-Cahn (1974), also Anderson, Sobel and Uppuluri (1980). The most recent one is Prasad and Menon (1985). There are no general results on the situation above seem to have been given so far for the case $0<p<1$ and $b\geq 1$. In this paper, we concern $0<p<1$, $b\geq 1$ and find a unified approach to solve many different types of the waiting time problems on the randomized sequential occupancy model.

Consider the conceptual "experiment" of throwing balls randomly into n cells including the balls of falling through. We do not fix in advance the number of balls but let the balls be placed one by one as necessary for a prescribed situation to arise.

If the functions $f(\cdot)$ and $h(\cdot)$ are taken as $f(x)=h(x)\equiv x$, then the random variable Z represents the total number of balls thrown until k cells contain at least b balls each, i.e., Z is the discrete waiting time for this experiment. This case is called the birthday problem, the coupon-collectors problem, or the randomized sequential occupancy problem.

In section 2, a representation of the characteristic function of Z is obtained. This characteristic function can be used to derive exact and asymptotic distributions of Z. To illustrate its applications, two corollaries are given in section 3.

2. **The characteristic function of Z.** Concerning with our urn model, let $T_i=\sum_{j=1}^{b} T_{i,j}$, for $i=1,2,\cdots,n$, where $T_{i,j}$ is the interarrival time between the $(j-1)$ th ball and the jth ball in cell i. Then T_i is the waiting time until cell i gets its bth ball. The requirement for the experiment to terminate is that when k out of the n cells contain at least b balls each. Let Y_i be the number of balls in the ith cell, $1\leq i\leq n$, and Y_{n+1} be the number of balls not entering in any of the n cells when stopping.

We are assuming that the arrival processes $\{T_{i,j}\}_{j=1}^{\infty}$ of the balls in any of the n cells are independent Poisson processes with intensity one and the arrival process for those balls leaking from the n cells is also a Poisson process which is independent of the previous n processes with intensity $np/(1-p)$. It is because that the ratio of the intensities for the two independent Poisson processes is equal to the ratio of probabilities of their occurrence, then in our case, the intensity $np/(1-p)$ is got from the ratio of p to $(1-p)/n$. Under the equiprobable assumption, all the n cells have the same chance $(1-p)/n$ of being hit or with probability p of not hitting any cell on each throw. All the throws are independent Poisson processes. So that T_i, $1 \leq i \leq n$ are i.i.d. $\Gamma(b, 1)$ random variables. Moreover, we find an interesting and important fact that the waiting time until k unspecified cells contain at least b balls each is just the order statistics $T_{k:n}$. Now we derive the characteristic function of Z in the following theorem.

THEOREM. *Given the time interval* $[0, t]$, $t \geq 0$, *let the random variables* ψ_t *be Poisson* (t) *and* ζ_t *be Poisson* $(t(np/(1-p)))$, ψ_t *and* ζ_t *are independent. Then the characteristic function for*

$$Z = \sum_{i=1}^{n} f(Y_i) + h(Y_{n+1})$$

is

$$\begin{aligned} E(e^{i\theta Z}) &= n\binom{n-1}{k-1} e^{i\theta f(b)} \\ &\cdot \int_0^\infty \Big(\sum_{j=b}^{\infty} e^{i\theta f(j)} P(\psi_t = j)\Big)^{k-1} \\ &\cdot \Big(\sum_{j=0}^{b-1} e^{i\theta f(j)} P(\psi_t = j)\Big)^{n-k} \\ &\cdot P(\psi_t = b-1)\Big(\sum_{j=0}^{\infty} e^{i\theta h(j)} P(\zeta_t = j)\Big)\, dt. \end{aligned} \tag{2.1}$$

Proof. From the previous discussion, we have known that $T_{k:n}$ is the (continuous) waiting-time of the experiment and since $Z = \sum_{j=1}^{n} f(Y_j) + h(Y_{n+1})$ is a function of $\sum_{i=1}^{n+1} Y_i$ which is the (discrete) waiting-time of the experiment, so we consider using the conditional expectation

$$E(e^{i\theta Z}) = E(E(e^{i\theta Z} | T_{k:n}))$$

to derive the characteristic function of Z.

Let U_t and V_t have the truncated Poisson distributions

$$P(U_t = j) = P(\psi_t = j | \psi_t \leq b-1), \qquad j = 0, 1, \cdots, b-1. \tag{2.2}$$

and

$$P(V_t = j) = P(\psi_t = j | \psi_t \geq b), \qquad j = b, b+1, \cdots. \tag{2.3}$$

For given fixed time $t > 0$, the event $T_{k:n} = t$ says the following three cases:

(i) By the time t, since

$$T_{1:n} \leq T_{2:n} \leq \cdots \leq T_{k-1:n} \leq T_{k:n} = t$$

and $T_{l:n}$, $1 \leq l \leq k-1$ must be one of T_i, $1 \leq i \leq n$. Thus there are exactly $k-1$ cells have got at least b balls each by time t. Also the corresponding contents of $k-1$ cells are i. i. d. truncated (from below by b) Poisson random variables, say $V_t(1), \cdots, V_t(k-1)$, where V_t is defined as in equation (2.3).

(ii) At time t, consider

$$\text{the given event } T_{k:n} = t,$$

where $T_{k:n}$ is one of the T_i, $1 \leq i \leq n$ which is different from the previous $T_{l:n}$, $1 \leq l \leq k-1$. Hence there is exactly one cell gets its bth ball at time t and thus there is exactly one random variable among $\{Y_1, Y_2, \cdots, Y_n\}$ degenerates at b at time t.

(iii) After time t, since

$$t = T_{k:n} \leq T_{k+1:n} \leq \cdots \leq T_{n:n}.$$

There will be $(n-k)$ remaining cells get their bth ball each later than t. Thus at time t the contents of these $n-k$ cells are i. i. d. truncted (from above by $b-1$) Poisson random variables, say $U_t(1) \cdots$, $U_t,(n-k)$, where U_t is defined as in equation (2.2).

Combining all the discussions from above and by the fact that $T_i \overset{\text{i. i. d.}}{\longleftrightarrow} \Gamma(b, 1)$ and $\psi_t \sim \text{Poisson } (t)$, we get the two events $\{T_i \leq t\}$ and $\{\psi_t \geq b\}$ are equivalent and the density of $T_{k:n}$ is

$$g_{T_{k:n}}(t) = n\binom{n-1}{k-1}(P(\psi_t \geq b))^{k-1} \cdot (P(\psi_t \leq b-1))^{n-k} P(\psi_t = b-1) \tag{2.4}$$

and by the time $t_{k:n} = t$, the requirement for the experiment to terminate is satisfied, also consider conditioning on $T_{k:n} = t$, since all the T_i, $1 \leq i \leq n$, are i. i. d. $\Gamma(b, 1)$ random variables, hence given the condition $T_{k:n} = t$, the order statistics $T_{l:n}$, $1 \leq l \leq k-1$ and $T_{j:n}$, $k+1 \leq j \leq n$, and among themselves are independent. Therefore, we have the random variable

$$\{Y_1, Y_2, \cdots, Y_n \mid T_{k:n} = t\} \stackrel{d}{=} \{V_t(1), V_t(2), \cdots, V_t(k-1), U_t(1), U_t(2), \cdots, U_t(k), \text{ and one integer r. v. degenerates at } b\}$$

where $V_t(i)$, $1 \leq i \leq k-1$, are i. i. d. truncated (from below by b) Poisson r. v.

$U_t(j)$, $1 \leq j \leq n-k$, are i. i. d. truncated (from above by $b-1$) Poisson r. v.

moreover, $V_t(i)$, $U_t(j)$ are independent. As for the leaking balls

$$Y_{n+1} \stackrel{d}{=} \zeta_t \sim \text{Poisson}\left(\text{intensity} = \frac{np}{1-p}\right)$$

where ζ_t is independent of $V_t(i)$, $U_t(j)$ for $1 \leq i \leq k-1$ and $1 \leq j \leq n-k$. Thus for $Z = \sum_{i=1}^n f(Y_i) + h(Y_{n+1})$

$$E(e^{i\theta Z} \mid T_{k:n} = t) = E\left\{\exp\left[i\theta\left(\sum_{i=1}^n f(Y_i) + h(Y_{n+1})\right) \middle| \; T_{k:n} = t\right]\right\}$$
$$= e^{i\theta f(b)}\{E(e^{i\theta f(V_t)})\}^{k-1}\{E(e^{i\theta f(U_t)})\}^{n-k} E(e^{i\theta h(\zeta_t)}).$$

By equation (2.4), the characteristic equation of Z is

$$E(e^{i\theta Z}) = \int_0^\infty E(e^{i\theta Z} \mid T_{k:n} = t)\, g_{T_{k:n}}(t)\, dt$$
$$= n\binom{n-1}{k-1} e^{i\theta f(b)} \int_0^\infty \left(\sum_{j=b}^\infty e^{i\theta f(j)} P(\psi_t = j)\right)^{k-1}$$
$$\cdot \left(\sum_{j=0}^{b-1} e^{i\theta f(j)} P(\psi_t = j)\right)^{n-k}$$
$$\cdot P(\psi_t = b-1)\left(\sum_{j=0}^\infty e^{i\theta h(j)} P(\zeta_t = j)\right) dt.$$

Therefore, the theorem is proved.

Note. Once the characteristic equation of Z is derived as in equation (2.1), then the moment generating function of Z is analogously obtained as

$$M_Z(t) = E(e^{sZ})$$
$$= n\binom{n-1}{k-1} e^{sf(b)} \int_0^\infty \left(\sum_{j=b}^{\infty} e^{sf(j)} P(\psi_t = j)\right)^{k-1}$$
$$\cdot \left(\sum_{j=0}^{b-1} e^{sf(j)} P(\psi_t = j)\right)^{n-k}$$
$$\cdot P(\psi_t = b-1)\left(\sum_{j=0}^{\infty} e^{sh(j)} P(\zeta_t = j)\right) dt$$

for $s \in (-s_0, s_0)$ with $s_0 > 0$ and is fixed.

3. **Limit theorems for two special cases of Z.** By the unique existence of the characteristic function for every random variable then the representation of the characteristic function of Z is very useful both for obtaining exact and asymptotic distributions for the random variable Z. For different choices of f and h, we get different types of the waiting time problems.

In this section we give two corollaries to illstrate the usefulnees of the Theorem in section 2.

Under the same probabilistic structure as in the Introduction, i.e., the experiment is continued as soon as k unspecified cells containing at least b balls each, we will consider two types of Z in this section:

(1) $Z = \sum_{i=1}^{n+1} Y_i$, i.e. Z is the total number of balls in the n cells with those are leaking. As we have mentioned in the Introduction. Z is the discrete waiting time for the experiment.

(2) $Z = \sum_{i=1}^{n} I(Y_i = b)$, then Z represents the number of cells with exactly b balls when the experiment is stopped. This case is called the randomized occupancy problem.

We will study the limiting distribution of Z for each case respectively in Corollary 1 and 2.

COROLLARY 1. *If $Z = \sum_{i=1}^{n+1} Y_i$, then $((1-p)/n) Z \xrightarrow{d} T_{k:n}$ as n is very large. Moreover, if*

(i) *let $n, k \to \infty$ with $n - k = m$ as a constant, then after suitably normalized, Z is extreme-valued distributed in the limit, or*

(ii) *if $n \to \infty$ with $k/n \to \lambda$, $0 < \lambda < 1$, then Z is asymptotically normally distributed.*

Proof. Referring to the Theorem, the corresponding functions f and h should be taken as the identity function on the case $Z = \sum_{i=1}^{n+1} Y_i$. For convenience of derivation, we use probability-generating function instead of the characteristic function of Z in the Theorem, then we get

$$\begin{aligned}
E(s^Z) &= s^b \int_0^\infty (E(s^{V_t}))^{k-1}(E(s^{U_t}))^{n-k}(E(s^{\zeta_t}))\, g_{T_{k:n}}(t)\, dt \\
&= n\binom{n-1}{k-1} s^b \int_0^\infty \Big(\sum_{j=b}^{\infty} \frac{s^j e^{-t} t^j}{j!}\Big)^{k-1} \\
&\quad \cdot \Big(\sum_{j=0}^{b-1} \frac{s^j e^{-t} t^j}{j!}\Big)^{n-k} \frac{e^{-t} t^{b-1}}{(b-1)!} e^{(np/(1-p))t(s-1)}\, dt \\
&= \int_0^\infty e^{n(-t+ts+(tp/(1-p))(s-1))}\, g_{T_{k:n}}(ts)\, s dt \\
&= E\{e^{(n/(1-p))(1-s^{-1})T_{k:n}}\}.
\end{aligned}$$

Hence the moment-generating function of Z is

$$E(e^{\theta Z}) = E\{e^{(n/(1-p))(1-e^{-\theta})T_{k:n}}\}, \tag{3.1}$$

where $T_{k:n}$ is the kth order statistics of $\{T_1, T_2, \cdots, T_n\}$, T_i's are i.i.d. $\Gamma(b, 1)$ r.v. We consider the general case $b \geq 1$. The asymptotic distribution of Z can be obtaind from the relation of equation (3.1).

From equation (3.1), we get

$$\frac{Z}{\sqrt{n}} \stackrel{d}{=} \frac{n}{\theta(1-p)}(1 - e^{-\theta/\sqrt{n}})\, T_{k:n},$$

for any $0 \neq |\theta| \leq 1$ and fixed large n. In particular, let $\theta \to 0$ and by Taylor's expansion, we have

$$\frac{1-p}{n} Z \xrightarrow{d} T_{k:n}, \tag{3.2}$$

as n is very large. Thus the first part of the corollary follows. This tells us that if the number of cells, n, is large but fixed, then when the experiment is stopped, the numbers of balls in any of the n cells are identically distributed as the waiting time of the experiment, $T_{k:n}$.

The asymyptotic distribution of Z can be obtained by Stirling's formula and Taylor's expansion. We consider two special cases:

Case (i). Let n, $k \to \infty$ with $n - k = m$ as a constant and keep b as a constant. By using the fact equation (3.2), we get that

$$\frac{Z}{n} \xrightarrow{d} \frac{1}{1-p} T_{k:n}$$

as n is very large.

Note that $T_{k:n}$ is the kth order statistics of n's i.i.d. $T_i \sim \Gamma(b, 1)$ random variables. We apply the well known theorem on the asymptotic distribution of $T_{k:n}$ (Leadbetter, Lindgren & Rootzen (1983), p. 33, Theorem 2.2.2.). We denote this reference as L-L-R later.

The *cdf* of T_i is

$$F(t) = \frac{1}{\Gamma(b)} \int_0^t s^{b-1} e^{-s} ds = \sum_{j=b}^{\infty} \frac{e^{-t} t^j}{j!}.$$

First, we have to find the asymptotic distribution of $T_{n:n} = \max\{T_1, T_2, \cdots, T_n\}$, say $G(x)$, and normalizing constants $\{a_n, b_n\}$

$$P\{a_n(T_{n:n} - b_n) \leq x\} \xrightarrow{d} G(x)$$

i.e.

$$P\left\{T_{n:n} \leq \frac{x}{a_n} + b_n\right\} \xrightarrow{d} G(x).$$

Let

$$U_n = \frac{x}{a_n} + b_n.$$

By Theorem 1.5.1. of L-L-R (1983), we may choose U_n as large as possible so that there exists τ, $0 \leq \tau < \infty$, such that

$$n(1 - F(U_n)) \longrightarrow \tau \tag{3.3}$$

where τ is fixed and finite, and

$$\begin{aligned} 1 - F(U_n) &= \sum_{j=0}^{b-1} \frac{e^{-U_n} U_n^j}{j!} \\ &= e^{-U_n}\left(1 + U_n + \frac{U_n}{2!} + \cdots + \frac{U_n^{b-1}}{(b-1)!}\right) \\ &= e^{-U_n} \frac{U_n^{b-1}}{(b-1)!} \{1 + o(U_n^{b-1})\}, \end{aligned}$$

taking logarithm on relation (3.3), we get

$$l_n n + l_n(1 - F(U_n)) \longrightarrow l_n \tau,$$

i. e.

$$(3.4) \qquad l_n n - U_n + (b-1)\, l_n(U_n) - l_n(b-1)! + l_n\{1 + o(U_n^{b-1})\} \longrightarrow l_n \tau.$$

Dividing U_n on both sides and let $U_n \to \infty$, then

$$(3.5) \qquad \frac{l_n n}{U_n} - 1 + (b-1)\frac{l_n(U_n)}{U_n} - \frac{l_n(b-1)!}{U_n} \longrightarrow \frac{l_n \tau}{U_n}.$$

Since τ is finite and U_n approximates to infinite, so the above relation (3.5) is reduced to

$$\frac{l_n n}{U_n} - 1 \longrightarrow 0, \text{ and then } \frac{l_n n}{U_n} \to 1.$$

Thus

$$U_n \overset{\circ}{\sim} l_n n,$$

i. e. U_n is asymptotically equivalent to $l_n n$.

By relation (3.4), we get

$$U_n \overset{\circ}{\sim} - l_n \tau + l_n n + (b-1)\, l_n(l_n n) - l_n(b-1)!$$

Comparing this with $U_n = (x/a_n) + b_n$, so take $a_n = 1$, $x = - l_n \tau$,

$$b_n = l_n n + (b-1)\, l_n(l_n n) - l_n(b-1)!.$$

Then

$$\tau = e^{-x}, \text{ so } G(x) = e^{-e^{-x}}$$

i. e.

$$(3.6) \quad P\{T_{n:n} - (l_n n + (b-1)\, l_n(l_n n) - l_n(b-1)!) \leq x\} \xrightarrow{d} e^{-e^{-x}}.$$

So by Theorem 2.1.2 of L-L-R, we get the asymptotic distribution of $T_{k:n}$ as

$$(3.7) \quad P\{T_{k:n} - (l_n n + (b-1)\, l_n(l_n n) - l_n(b-1)!) \leq x\} \longrightarrow e^{-e^{-x}} \sum_{j=0}^{m} \frac{(e^{-x})^j}{j!}.$$

Since

$$\frac{Z}{n} \stackrel{d}{=} \frac{1}{1-p} T_{k:n}.$$

So the asymptotic distribution of Z is

$$\text{(3.8)} \quad \frac{Z}{n} - \frac{1}{1-p} (l_n n + (b-1) l_n(l_n n) - l_n(b-1)!) \simeq \frac{1}{1-p} W$$

where W has the extreme value (type I) distribution (Galambos (1978) with distribution function,

$$F_W(x) = e^{-e^{-x}} \sum_{j=0}^{m} \frac{(e^{-x})^j}{j!}.$$

Case (ii). If $n \to \infty$ with $k/n \to \lambda$, $0 < \lambda < 1$. Then by equation (3.3), and by the well-known result about the asymptotic distribution of order statistics (Wilks (1962)), we get the asymptotic distribution of $T_{k:n}$ is

$$T_{k:n} \xrightarrow{d} N\left(T_{(\lambda)}, \frac{\lambda(1-\lambda)}{n\{(1/(b-1)!)\, T_{(\lambda)}^{b-1} e^{-T_{(\lambda)}}\}^2}\right),$$

where $T_{(\lambda)}$ is the λth quantile of T_i. Thus,

$$\lim_{n\to\infty} \Pr\left[\frac{Z}{\sqrt{n}} - \frac{\sqrt{n}}{1-p} T_{(\lambda)} < x\right] = \Phi^*(x)$$

with $k/n \to \lambda$, where $\Phi^*(x)$ is a normal distribution with mean $= 0$, variance $= \lambda(1-\lambda)\{(b-1)!/((1-p)\, T_{(\lambda)}^{b-1} e^{-T_{(\lambda)}})\}^2$. Thus, the waiting time Z suitably normalized is in the limit, extrme-valued distributed in case (i) and normally distributed in case (ii).

COROLLARY 2. *If $Z = \sum_{i=1}^{n} I(Y_i = b)$, and we denote it by N, let $n, k \to \infty$ with $n - k = m$ as a constant, then after suitably normalized, N is asymptotically $\Gamma(m+1, 1/b)$ distributed, i.e.*

$$\frac{N}{\{l_n n + l_n(b-1)!\}^b} = \frac{\#\{\text{cells with exactly } b \text{ balls}\}}{\{l_n n + l_n(b-1)!\}^b} \xrightarrow{d} \Gamma\left(m+1, \frac{1}{b}\right).$$

Proof. Referring to the Theorem in section 2, we take function f as the indicator function

$$f(x)=I(x=b)=\begin{cases}1 & \text{if } x=b\\ 0 & \text{o. w.}\end{cases}$$

and function $h\equiv 0$. Then according to eq. (2.1),

$$E(e^{i\theta N})=e^{i\theta}\int_0^\infty g_{T_{k:n}}(t)\{E(e^{i\theta f(V_t)})\}^{k-1}\{E(e^{i\theta f(U_t)})\}^{n-k}\,dt, \tag{3.9}$$

wher $T_{k:n}=$ the kth order statistic of $\{T_1, T_2,\cdots, T_n\}$, T_i are i. i. d. $\Gamma(b, 1)$ random variables.

U_t, V_t are defined in equation (2.2), (2.3) respectively. Then in equation (3.9)

$$\begin{aligned}E(e^{i\theta f(V_t)})&=\sum_{j=b}^{\infty}e^{i\theta f(j)}P(V_t=j)\\&=1+\frac{(e^{-t}t^b/b!)(e^{i\theta}-1)}{1-\sum_{j=0}^{b-1}e^{-t}t^j/j!}\end{aligned} \tag{3.10}$$

and

$$E(e^{i\theta f(U_t)})=\sum_{j=0}^{b-1}e^{i\theta f(j)}P(U_t=j)=1, \tag{3.11}$$

by equation (2.4), the density of $T_{k:n}$ is

$$\begin{aligned}g_{T_{k:n}}(t)=&\frac{n!}{(k-1)!(n-k)!}\left\{\sum_{j=b}^{\infty}\frac{e^{-t}t^j}{j!}\right\}^{k-1}\\&\cdot\left\{\sum_{j=0}^{b-1}\frac{e^{-t}t^j}{j!}\right\}^{n-k}\frac{1}{\Gamma(b)}t^{b-1}e^{-t},\end{aligned}$$

let $n, k\to\infty$ with $n-k=m$ as a constant, then the asymptotic distribution of $T_{k:n}$ is equation (3.7), hence the *pdf* of the normalized variable

$$\begin{aligned}u&=T_{k:n}-l_n n-(b-1)\,l_n(l_n n)-l_n(b-1)!\\&=T_{k:n}-l_n n\left[1+(b-1)\frac{l_n(l_n n)}{l_n n}\right]-l_n(b-1)!\\&=T_{k:n}-l_n n[1+(b-1)(l_n n)]-l_n(b-1)!\\&\simeq T_{k:n}-l_n n-l_n(b-1)!\end{aligned}$$

becomes asymptotically as

(3.12) $$g_{T_{k:n}-l_n n-l_n(b-1)!}(u) \longrightarrow e^{-u}\, e^{-e^{-u}}\, \frac{(e^{-u})^m}{m!}.$$

Insert equation (3.10), (3.11) into equation (3.9), we have

(3.13) $$E(e^{i\theta N}) = e^{i\theta}\int_0^{\infty} g_{T_{k:n}}(t)\left\{1 + \frac{(e^{-t}\,t^b/b!)(e^{i\theta}-1)}{1-\sum_{j=0}^{b-1} e^{-t}\,t^j/j!}\right\}^{k-1} dt,$$

using the transformation $t = u + l_n n + l_n(b-1)!\ (= u_n)$ and normalizing N by $N/[l_n n + l_n(b-1)!]^b$. Then equation (3.13) becomes

$$E(e^{i(\theta/[l_n n+l_n(b-1)!]^b)N})$$

$$= e^{i(\theta/[l_n n+l_n(b-1)!]^b)} \int_{-\infty}^{\infty} g_{T_{k:n}-l_n n-l_n(b-1)!}(u)$$

(3.14) $$\left\{1 + \frac{(e^{-u-l_n n-l_n(b-1)!}(u+l_n n+l_n(b-1)!)^b)/b!\ (e^{i(\theta/[l_n n+l_n(b-1)!]^b)}-1)}{1-\sum_{j=0}^{b-1} e^{-u_n}\,u_n^j\, j!}\right\}^{k-1} dt.$$

In equation (3.14), the denominator

$$1-\sum_{j=0}^{b-1}\frac{e^{-u_n}u_n^j}{j!} = F(u_n)$$

where $F(\cdot)$ is the *cdf* of a $\Gamma(b, 1)$ random variable, then by relation (3.3)

$$1-F(U_n) \simeq \frac{\tau}{n} = \frac{e^{-u}}{n}.$$

Hence $1-\sum_{j=0}^{b-1} e^{-u_n}u_n^j\, j! \simeq 1-(e^{-u}/n)$.

In the numerator, the terms

$$(u+l_n n+l_n(b-1)!)^b\,(e^{i(\theta/[l_n n+l_n(b-1)!]^b)}-1)$$

can be approximated by binomial and Taylor's expansions to the constant, $i\theta$.

Also the term in the numerator

$$e^{-u-l_n n-l_n(b-1)!} = \frac{e^{-(u+l_n(b-1)!)}}{n}.$$

Insert all these approximation into the bracket of equation (3.14) and replace k by $n-m$, note that m is finite. Then the contents in bracket become

$$\left\{1+\frac{(e^{-(u+l_n(b-1)!)}/b!\ n)}{1-e^{-u}/n}\,i\theta\right\}^{n-m-1}$$

$$=\left\{\frac{1+e^{-u}/n((e^{-l_n(b-1)!}/b!\ i\theta-1)}{1-e^{-u}/n}\right\}^{n}$$

$$\cdot\left\{1+\frac{(e^{-(u+l_n(b-1)!)}/b!\ n)}{1-e^{-u}/n}\,i\theta\right\}^{-m-1}$$

$$\text{(3.15)}\qquad \longrightarrow \frac{e^{e^{-u}(e^{-l_n(b-1)!}/b!\,i\theta-1)}}{e^{-e^{-u}}}=e^{(e^{-(u+l_n(b-1)!)}/b!)i\theta}.$$

Refer back to equation (3.14) and put the relations (3.12), (3.15) in equation (3.14). Then

$$\text{(3.16)}\qquad E(e^{i\theta(N/[l_n n+l_n(b-1)!]^b)})\longrightarrow\int_{-\infty}^{\infty}e^{-u}\,e^{-e^{-u}}\,\frac{(e^{-u})^m}{m!}\,e^{(e^{-(u+l_n(b-1)!)}/b!)i\theta}\,du.$$

Change variable

$$w=\frac{e^{-u-l_n(b-1)!}}{b!}.$$

Then

$$e^{-u}=bw,\qquad du=-\frac{dw}{w}.$$

Thus relation (3.16) becomes

$$\int_0^\infty (bw)e^{-bw}\,\frac{(bw)^m}{m!}\,e^{i\theta w}\,\frac{dw}{w}$$

$$=\int_0^\infty\left(\frac{b^{m+1}}{m!}\,w^m\,e^{-bw}\right)e^{i\theta w}\,dw$$

i. e.

$$E(e^{i\theta(N/[l_n n+l_n(b-1)!]^b)})\longrightarrow\int_0^\infty e^{i\theta w}\cdot f(w)\,dw,$$

where

$$f(w)=\frac{b^{m+1}}{m!}\,w^m\,e^{-bw}$$

is the *pdf* of a $\Gamma(m+1,\ 1/b)$ random variable. By the uniqueness of the characteristic function, we get

$$\frac{N}{[l_n n + l_n(b-1)!]^b} \xrightarrow{d} \Gamma\left(m+1, \frac{1}{b}\right)$$

as $n, k \to \infty$ with $n-k=m$ a constant and $b \geq 1$ is fixed.

It means that when the process is stopped, the number of cells with exactly b balls under the normalization $[l_n n + l_n(b-1)!]^b$ will be asymptotically distributed as a $\Gamma(m+1, 1/b))$ variable.

4. **Conclusion and extension.** From the above corollaries we find the expression in the Theorem is very useful, it can solve many different types of randomized sequential occupancy problems. Actually, on the same problem, we can consider the much more general case that the probabilities of each ball staying in any of the n cells are different or the case that when the experiment is stopped the k unspecified urns contain at least b's balls, the b_i's, $1 \leq i \leq k$, may be different (Yeh (1986)).

REFERENCES

1. K. Anderson, M. Sobel and V.R.R. Uppuluri, *Multivariate Hypergeometric and Multinomial Waiting Times*, Technical Report No. 147 (1980), Department of Statistics, Standford University.

2. E. Cinlar, *Introduction to Stochastic Processes*, Prentice-Hall, Inc. Englewood Cliffs, N.J. (1975).

3. H.A. David, *Order Statistics*, 2nd. edition. Wiley, New York (1980).

4. W. Feller, *An Introduction to Probability Theory and Its Applications*, Wiley, New York (1968).

5. J. Galmbos, *The Asymptotic Theory of Extreme Order Statistics*, Wiley, New York.

6. L. Holst, *On sequential occupancy problems*, J. Appl. Prob. 18 (1981), pp. 435-442.

7. N.L. Johnson and S. Kotz, *Urn Models and Their Application*. Wiley, New York (1977).

8. M.R. Leadbetter, G. Lindgren and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes* (1983).

9. B. Prasad and V.V Menon, *Poisson approximation to the distribution of empty cells in randomized occupancy*, Commun. Statist.-Theor. Meth., 14 (1) (1985), 41-48.

10. E. Samuel-Cahn, *Asymptotic distributions for occupancy and waiting time problems with positive probability of falling through the cells*, Ann. Prob. 2 (1974), 515-521.

11. S. Wilks, *Mathematical Statistics*, Wiley, New York (1962).

12. Hsiaw-Chan, Yeh. *Randomized Occupancy Problems and Related Waiting Time Distribution*, Conference Paper, 1986 International Statistical Symposium, Taipei, R.O.C. 999-1017, Taipei (1986).

Department of Finance
National Taiwan University
Taipei, Taiwan, R.O.C.