THREE NEW PROOFS AND A GENERALIZATION OF A THEOREM OF IRVING WEISS

by

A. RÉNYI

Dedicated to Prof. P. Erdős, at his 50th birthday.

Introduction

Let us suppose that N balls are distributed at random among n boxes, so that each ball may fall into any box with the same probability $\frac{1}{n}$, independently of what happens to the other balls. Let $\zeta_{n,N}$ denote the number of boxes which remain empty. I. WEISS [1] has proved that if $n \to +\infty$ and $N = N(n)$ is a function of n such that $\lim_{n\to+\infty} \frac{N}{n} = \alpha > 0$ then $\zeta_{n,N}$ is in the limit normally distributed, its expectation and variance being asymptotically equal to $ne^{-\frac{N}{n}}$ and $ne^{-\frac{N}{n}}\left[1-\left(1+\frac{N}{n}\right)e^{-\frac{N}{n}}\right]$ respectively; by other words we have (denoting by $\mathbf{P}(\ldots)$ the probability of the event in the brackets) for any real x

$$\lim_{\substack{n\to+\infty \\ \frac{N}{n}\to\alpha}} \mathbf{P}\left(\frac{\zeta_{n,N}-ne^{-\frac{N}{n}}}{\sqrt{ne^{-\frac{N}{n}}\left[1-\left(1+\frac{N}{n}\right)e^{-\frac{N}{n}}\right]}} < x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}\,du\,. \tag{1}$$

WEISS used the method of moments, which in this case leads to rather cumbersome calculations. In view of the simplicity of the question (which is in striking contrast with the difficulties of the proof) and the importance of its various possible practical applications, we considered it worth while trying to find a simpler proof.

We succeeded in finding three new proofs for the result of Weiss all using the method of characteristic functions. All these proofs are definitely simpler than that of I. WEISS, though a certain amount of subtle asymptotic analysis is indispensable. Besides their simplicity, these proofs are of a certain methodological interest, as they may serve as patterns for the solution of other similar problems.

These proofs yiels also a more general result, namely that (1) remains valid also if $\frac{N}{n} \to 0$ or $\frac{N}{n} \to +\infty$ provided that

$$ne^{-\frac{N}{n}}\left(1-\left(1+\frac{N}{n}\right)e^{-\frac{N}{n}}\right) \to +\infty$$

$\left(\text{this is equivalent in case } \frac{N}{n} \to 0 \text{ with } \frac{N^2}{n} \to +\infty \text{ and in case } \frac{N}{n} \to +\infty\right.$ with $\left.\frac{N}{n} - \log n \to -\infty\right)$. We prove the generalization of the theorem of WEISS to the case $\frac{N}{n} \to +\infty$, $\frac{N}{n} - \log n \to -\infty$ by the second method, and the generalization to the case $\frac{N}{n} \to 0$, $\frac{N^2}{n} \to +\infty$ by the third method.

The problem in question is essentially that of proving the central limit theorem for certain not independent but weakly (and symmetrically) dependent random variables.) The situation is similar to that of sampling from a finite population (see [2]). The variables in question are the indicators of the emptiness of the different boxes; as a matter of fact $\zeta_{n,N}$ can be written in the form

$$\zeta_{n,N} = \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_n$$

where ε_k is 1 or 0 according to whether the k-th box is empty or not $(k = 1, 2, \ldots, n)$. The dependence is in this case not negligible in the limit, as is seen from the formula for the variance.

As a matter of fact, the mean value of ε_k being $\left(1 - \frac{1}{n}\right)^N \sim e^{-\frac{N}{n}}$ if the dependence would be in the limit negligible, the variance would be asymptotically equal to $ne^{-\frac{N}{n}}\left(1 - e^{-\frac{N}{n}}\right)$; while as mentioned above, the variance is definitely smaller $\left(\text{by the factor } 1 - \frac{N}{n(e^{N/n} - 1)} < 1\right)$ which shows that there is a not negligible negative correlation among the variables ε_k.

As a matter of fact, the variance $\mathbf{D}^2(\zeta_{n,N})$ of $\zeta_{n,N}$ is

$$\mathbf{D}^2(\zeta_{n,N}) = n^2\left[\left(1 - \frac{2}{n}\right)^N - \left(1 - \frac{1}{n}\right)^{2N}\right] + n\left[\left(1 - \frac{1}{n}\right)^N - \left(1 - \frac{2}{n}\right)^N\right]$$

and therefore

$$\mathbf{D}^2(\zeta_{n,N}) = ne^{-\frac{N}{n}}\left[1 - \left(1 + \frac{N}{n}\right)e^{-\frac{N}{n}}\right]\cdot\left(1 + O\left(\frac{N}{n^2}\right)\right).$$

In § 1 we give the first proof which is worked out only for the case when $\frac{N}{n} \to \alpha > 0$. In § 2 we present the second proof. In § 3 we give the third proof and in § 4 add some remarks.

This paper was written during the stay of the author, as a visiting professor, at the Department of Statistics of the Michigan State University during the summer term of the year 1961. The author is indebted to Prof. H. RUBIN for calling his attention to the paper of I. WEISS.

§ 1. The first proof

Let $P_{n,N}(k)$ denote the probability that there are exactly k empty boxes if n denotes the total number of boxes and N the number of balls. By other words we put

$$P_{n,N}(k) = \mathbf{P}(\zeta_{n,N} = k) \qquad (k = 0, 1, \ldots, n).$$

The following two recursion formulae hold:

$$P_{n,N+1}(k) = P_{n,N}(k)\left(1 - \frac{k}{n}\right) + P_{n,N}(k+1)\frac{k+1}{n} \tag{2}$$

and

$$P_{n,N}(k) = P_{n-1,N}(k-1)\cdot\left(1 - \frac{1}{n}\right)^N + \sum_{l=1}^{N}\binom{N}{l}\frac{1}{n^l}\left(1 - \frac{1}{n}\right)^{N-l}\cdot P_{n-1,N-l}(k). \tag{3}$$

We obtain (2) by considering what happens if after throwing N balls into the boxes, one ball more is added: this may fall in a box which already contains a ball, or into an empty box. The two terms on the right correspond to these two possibilities.

As regards the less evident recursion formula (3) this can be obtained by the following argument: Let us label the boxes by the numbers 1, 2,. . ., n and let us denote by l the number of balls which fall into the first box. We distinguish between the cases $l = 0$ and $l \geqq 1$. If $l = 0$ (the probability of which is $\left(1 - \frac{1}{n}\right)^N$) then the N balls are distributed among the remaining $n - 1$ boxes so that among these $k - 1$ boxes remain empty. This gives the first term on the right of (3). If however l is some positive integer, $1 \leqq l \leqq n$ (the probability of which is $\binom{N}{l}\frac{1}{n^l}\left(1 - \frac{1}{n}\right)^{N-l}$) then the remaining $N - l$ balls are distributed among $n - 1$ boxes so that k boxes remain empty. Thus we obtain the sum on the right of (3).

Now let us introduce the characteristic funtion $\varphi_{n,N}(t)$ of $\zeta_{n,N}$, i.e. we put

$$\varphi_{n,N}(t) = \sum_{k=0}^{n} P_{n,N}(k)\, e^{ikt}. \tag{4}$$

We obtain from (3) the recursion formula

$$\varphi_{n,N}(t) = \left(1 - \frac{1}{n}\right)^N \cdot e^{it}\cdot\varphi_{n-1,N}(t) + \sum_{l=1}^{N}\binom{N}{l}\frac{1}{n^l}\left(1 - \frac{1}{n}\right)^{N-l}\cdot\varphi_{n-1,N-l}(t). \tag{5}$$

Let us now introduce the generating function

$$G_n(t,z) = \sum_{N=0}^{\infty}\frac{\varphi_{n,N}(t)\cdot(nz)^N}{N!}. \tag{6}$$

We obtain from (5) easily

$$G_n(t,z) = G_{n-1}(t,z)\,(e^{it} + e^z - 1). \qquad (n = 2, 3, \ldots). \tag{7}$$

As evidently

$$\varphi_{1,N}(t) = \begin{cases} e^{it} & \text{for } N = 0 \\ 1 & \text{for } N \geq 1 \end{cases} \tag{8}$$

and thus

$$G_1(t, z) = e^{it} + e^z - 1 \tag{9}$$

we obtain by (7) the surprisingly simple formula

$$G_n(t, z) = (e^{it} + e^z - 1)^n . \tag{10}$$

One can deduce (10) also from (2); as a matter of fact from (2) we get for G_n the partial differential equation $\frac{\partial G_n}{\partial z} = nG_n + \frac{e^{-it} - 1}{i} \frac{\partial G_n}{\partial t}$; taking into account that $G_n(t, 0) = \varphi_{n,0}(t) = e^{int}$, we get (10).

Now to prove (1) we apply the method of Laplace to obtain from (10) an asymptotic formula for $\varphi_{n,N}(t)$. We have by Cauchy's formula

$$\varphi_{n,N}(t) = \frac{N!}{n^N \cdot 2\pi i} \oint \frac{(e^{it} + e^z - 1)^n}{z^{N+1}} dz \tag{11}$$

where the integration may be carried out on any circle around the point $z = 0$.

We shall prove that putting

$$\alpha_n = \frac{N}{n} \tag{12}$$

and

$$D_n^2 = e^{-\frac{N}{n}} \left(1 - \left(1 + \frac{N}{n} \right) \cdot e^{-\frac{N}{n}} \right) \tag{13}$$

we have

$$\lim_{\substack{n \to +\infty \\ \frac{N}{n} \to \alpha}} \varphi_{n,N} \left(\frac{t}{D_n \sqrt{n}} \right) \cdot e^{-\frac{itn \cdot e^{-\alpha_n}}{D_n \sqrt{n}}} = e^{-\frac{t^2}{2}} . \tag{14}$$

((14) is clearly equivalent to (1)).

Let us choose as the path of integration in (11) the circle $|z| = \alpha_n$, by putting $z = \alpha_n \zeta$ with $\zeta = e^{iu}$ $(-\pi \leq u \leq +\pi)$. Then we obtain, by Stirling's formula, putting $u = \frac{v}{\sqrt{n \alpha_n}}$

$$\varphi_{n,N} \left(\frac{t}{D_n \sqrt{n}} \right) \sim \frac{1}{\sqrt{2\pi}} \int_{-\pi\sqrt{n\alpha_n}}^{+\pi\sqrt{n\alpha_n}} E_n^n \cdot e^{-iv\sqrt{n\alpha_n}} dv \tag{15}$$

where

$$E_n = \left(e^{\frac{it}{D_n \sqrt{n}}} - 1 \right) \cdot e^{-\alpha_n} + e^{\alpha_n \left(e^{\frac{iv}{\sqrt{n\alpha_n}}} - 1 \right)} . \tag{16}$$

Clearly for $\left|\frac{v}{\sqrt{n\alpha_n}}\right| \geqq \delta > 0$ we have for $n \geqq n_1$ (D_n and α_n being bounded from below by a positive constant)

$$E_n \leqq q < 1 \tag{17}$$

where q depends only on δ and α. It follows that

$$\lim_{n\to+\infty} \int\limits_{\delta\sqrt{\alpha_n n} \leq |v| \leq \pi\sqrt{\alpha_n \cdot n}} |E_n|^n\, dv = 0\,. \tag{18}$$

On the other hand for $\sqrt[7]{n\alpha_n} \leqq |v| \leqq \delta\sqrt{n\alpha_n}$, in view of

$$|E_n|^2 = e^{2\alpha_n\left(\cos\frac{v}{\sqrt{n\alpha_n}} - 1\right)} + 2\,e^{-2\alpha_n}\left(1 - \cos\frac{t}{D_n\sqrt{n}}\right) - \tag{19}$$

$$- 4\,e^{\alpha\left(\cos\frac{v}{\sqrt{n\alpha_n}} - 2\right)} \cdot \sin\frac{t}{2\sqrt{n}}\sin\left(\frac{t}{2\sqrt{n}} - \frac{v}{\sqrt{n\alpha_n}}\right),$$

we get

$$|E_n|^2 \leqq 1 - C_\delta \frac{v^2}{n} + O\left(\frac{t^2 + |tv|}{n}\right) \tag{20}$$

where $C_\delta > 0$ depends only on δ, which implies

$$\lim_{n\to+\infty} \int\limits_{\sqrt[7]{n\alpha_n}}^{\delta\sqrt{n\alpha_n}} |E_n|^n\, dv = 0\,. \tag{21}$$

Thus we obtain from (13), (18) and (21)

$$\varphi_{n,N}\left(\frac{t}{D_n \cdot \sqrt{n}}\right) e^{-\frac{itne^{-\alpha_n}}{D_n\sqrt{n}}} \sim \frac{1}{\sqrt{2\pi}} \int\limits_{-\sqrt[7]{n\alpha_n}}^{+\sqrt[7]{n\alpha_n}} E_n^n \cdot e^{-iv\sqrt{n\alpha_n} - \frac{itne^{-\alpha_n}}{D_n\sqrt{n}}}\, dv\,. \tag{22}$$

Now we get by an elementary calculation

$$E_n^n \cdot e^{-iv\sqrt{n\alpha_n} - \frac{itne^{-\alpha_n}}{D_n\sqrt{n}}} = e^{-\frac{t^2}{2} - \frac{\left(v - \frac{t\sqrt{\alpha_n}}{D_n} e^{-\alpha_n}\right)^2}{2}} + O\left(\frac{v^3}{\sqrt{n}}\right). \tag{23}$$

Thus it follows from (22) that (14) holds. As mentioned already, this proves (1).

§ 2. The second proof

In this § we give a second proof for (1) which leads at the same time to the following more general

Theorem 1. *If we distribute at random N balls among n boxes where $N = N(n)$ is a function of n such that $\frac{N(n)}{n}$ is bounded from below and $\lim_{n\to+\infty} \frac{N(n)}{n} - \log n \to -\infty$, then we have, putting $\alpha_n = \frac{N(n)}{n}$ for any real x*

$$\lim_{n\to+\infty} \mathbf{P}\left(\frac{\zeta_{n,N(n)} - ne^{-\alpha_n}}{\sqrt{ne^{-\alpha_n}[1-(1+\alpha_n)e^{-\alpha_n}]}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du\,.$$

Remark. In case $\alpha_n \to +\infty$ of course the result can be written also in the simpler form

$$\lim_{n\to+\infty} \mathbf{P}\left(\frac{\zeta_{n,N(n)} - ne^{-\alpha_n}}{\sqrt{ne^{-\alpha_n}}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du\,.$$

Proof. The idea of the proof which will be given in this § is one already applied in the paper [3]. Let us put

$$S_r = \mathbf{M}\Big(\sum_{1\le k_1<k_2<\ldots<k_r\le n} \varepsilon_{k_1}\varepsilon_{k_2}\ldots\varepsilon_{k_r}\Big) \tag{24}$$

where $\mathbf{M}(\ldots)$ denotes the expectation of the random variable in the brackets. Then evidently

$$S_r = \sum_{j=r}^{n} P_{n,N}(j)\binom{j}{r} \tag{25}$$

and thus

$$\sum_{r=0}^{n} S_r(x-1)^r = \sum_{j=0}^{n} P_{n,N}(j)\,x^j\,. \tag{26}$$

Now we can give a simple formula for S_r; in fact we have

$$S_r = \binom{n}{r}\left(1-\frac{r}{n}\right)^N. \tag{27}$$

Thus we get for the characteristic function $\varphi_{n,N}(t)$ defined by (4) the explicit formula

$$\varphi_{n,N}(t) = \sum_{r=0}^{n}\binom{n}{r}\left(1-\frac{r}{n}\right)^N (e^{it}-1)^r\,. \tag{28}$$

Of course (28) can be deduced also directly from (11).

In order to prove theorem 1 we consider the function

$$\psi_{n,N}(t) = \varphi_{n,N}\left(\frac{t}{D_n\sqrt{n}}\right) e^{-\frac{it\sqrt{n}\,e^{-\alpha_n}}{D_n}\left(1-\frac{N-n\alpha_n}{n}\right)} \tag{29}$$

where N is now not identical with $N(n)$, but is an independent nonnegative integral-valued variable. Then we have by (28), putting $y = e^{\frac{it}{D_n\sqrt{n}}}$,

$$\psi_{n,N}(t) = y^{-ne^{-\alpha_n}(1+\alpha_n)} \sum_{r=0}^{n} \binom{n}{r} \left(1 - \frac{r}{n}\right)^N (y-1)^n y^{Ne^{-\alpha_n}}. \tag{30}$$

It follows that putting

$$K_n(t) = \sum_{N=0}^{\infty} \psi_{n,N}(t) \frac{(\alpha_n \cdot n)^N e^{-\alpha_n \cdot n}}{N!} \tag{31}$$

we have

$$K_n(t) = e^{\alpha_n \cdot n(y^{e^{-\alpha_n}}-1)} y^{-ne^{-c_n}(1+\alpha_n)} \left(1 + (y-1) e^{-\alpha_n y^{e^{-\alpha_n}}}\right)^n. \tag{32}$$

Now if α_n is bounded from below and $\alpha_n - \log n \to -\infty$ the right hand side of (32) can be evaluated as follows

$$K_n(t) \sim e^{-\frac{t^2}{2} + O\left(\frac{1}{\sqrt{n}}\right)}. \tag{33}$$

Thus we obtain

$$\lim_{n\to+\infty} K_n(t) = e^{-\frac{t^2}{2}}. \tag{34}$$

Now evidently $|\psi_{n,N}(t)| \leq 1$ and thus by the central limit theorem, if $\omega_n = \log(n\alpha_n)$ then we have

$$\lim_{n\to+\infty} \left| \sum_{|N-\alpha_n \cdot n| > \omega_n \sqrt{n\alpha_n}} \psi_{n,N}(t) \frac{(n\alpha_n)^N e^{-n\alpha_n}}{N!} \right| = 0 \tag{35}$$

and

$$\lim_{n\to+\infty} \sum_{|N-n\alpha_n| \leq \omega_n \sqrt{n\alpha_n}} \frac{(n\alpha_n)^N e^{-n\alpha_n}}{N!} = 1. \tag{36}$$

We shall show now that for

$$|N - n\alpha_n| \leq \omega_n \sqrt{n\alpha_n} \tag{37}$$

we have

$$|\psi_{n,N}(t) - \psi_{n,N(n)}(t)| = O\left(\frac{\log^3 n}{\sqrt{n}}\right). \tag{38}$$

This implies by virtue of (34), (35) and (36) that

$$\lim_{n\to+\infty} \psi_{n,N(n)}(t) = e^{-t^2/2}. \tag{39}$$

As we have

$$\psi_{n,N(n)}(t) = \varphi_{n,N(n)}\left(\frac{t}{D_n\sqrt{n}}\right) e^{-\frac{it\sqrt{n}e^{-\alpha_n}}{D_n}} \tag{40}$$

it will follow from (39) that (14), and thus Theorem 1 holds.

Thus it remains only to prove (38). In view of (2) we obtain,

$$(41)\quad \psi_{n,N+1}(t)-\psi_{n,N}(t)=\left(e^{\frac{ite^{-\alpha_n}}{D_n\sqrt{n}}}-1\right)\psi_{n,N}(t)+\frac{\left(e^{-\frac{it}{D_n\sqrt{n}}}-1\right)}{n}\sum_{k=1}^{\infty}P_{n,N}(k)\,k\cdot e_k$$

where

$$e_k=e^{\frac{it}{D_n\sqrt{n}}\left(k-ne^{-\alpha_n}(1+\alpha_n)+(N+1)e^{-\alpha_n}\right)}.$$

Thus we obtain

$$(42)\quad \begin{aligned}\psi_{n,N+1}(t)-\psi_{n,N}(t)&=\left[e^{\frac{ite^{-\alpha_n}}{D_n\sqrt{n}}}-1-\left(e^{+\frac{it}{D_n\sqrt{n}}}-1\right)\left(1-\frac{1}{n}\right)^N\right]\psi_{n,N}(t)+\\&+\frac{\left(e^{-\frac{it}{D_n\sqrt{n}}}-1\right)}{n}\sum_{k=0}^{\infty}P_{n,N}(k)\left(k-n\left(1-\frac{1}{n}\right)^N\right)\cdot e_k\,.\end{aligned}$$

Thus we get for $|N-n\,\alpha_n|=O(\omega_n\sqrt{n\,\alpha_n})$

$$(43)\quad \begin{aligned}|\psi_{n,N+1}(t)-\psi_{n,N}(t)|&=O\left(\frac{(\log n)^{3/2}}{n}\right)+\\&+O\left(\frac{1}{D_n\cdot n^{3/2}}\sum_{k=0}^{\infty}P_{n,N}(k)\cdot\left|k-n\left(1-\frac{1}{n}\right)^N\right|\right).\end{aligned}$$

As clearly

$$(44)\quad \sum P_{n,N}(k)\left|k-n\left(1-\frac{1}{n}\right)^N\right|\leq \mathbf{D}(\zeta_{n,N})$$

where $\mathbf{D}^2(\zeta_{n,N})$ denotes the variance of $\zeta_{n,N}$ and we have

$$\mathbf{D}^2(\zeta_{n,N})\sim n\,D_n^2$$

it follows

$$|\psi_{n,N+1}(t)-\psi_{n,N}(t)|=O\left(\frac{(\log n)^{3/2}}{n}\right).$$

Thus if $|N-n\,\alpha_n|\leq\omega_n\sqrt{n\,\alpha_n}$ we have

$$(45)\quad |\psi_{n,N}(t)-\psi_{n\,N(n)}(t)|=O\left(\frac{\log^3 n}{\sqrt{n}}\right).$$

This proves (39), and thus Theorem 1.

§ 3. The third proof

In this § we shall prove the following

Theorem 2. *If we distribute N balls at random among n boxes, where $N=N(n)$ is a function of n such that for $n\to+\infty$ $\frac{N(n)}{n}$ is bounded from above and*

$$(46)\quad \lim_{n\to+\infty}\frac{N^2(n)}{n}=+\infty,$$

and $\zeta_{n,N(n)}$ *denotes the number of empty boxes, then we have, putting* $\alpha_n = \frac{N(n)}{n}$ *for any real* x

$$\lim_{n\to+\infty} \mathbf{P}\left(\frac{\zeta_{n,N(n)} - ne^{-\alpha_n}}{\sqrt{ne^{-\alpha_n}[1-(1+\alpha_n)e^{-\alpha_n}]}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du . \tag{47}$$

Remark. In case $\alpha_n \to 0$, the result can be written in the simpler form

$$\lim_{n\to+\infty} \mathbf{P}\left(\frac{\zeta_{n,N(n)} - ne^{-\alpha_n}}{\alpha_n \sqrt{\frac{n}{2}}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du \tag{47a}$$

or, putting $\zeta^*_{n,N(n)} = \zeta_{n,N(n)} - n(1-\alpha_n)$

$$\lim_{n\to+\infty} \mathbf{P}\left(\frac{\zeta^*_{n,N(n)} - \frac{n\alpha_n^2}{2}}{\alpha_n \sqrt{\frac{n}{2}}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du . \tag{47b}$$

Proof. Let us note first that putting

$$D_n^2 = e^{-\alpha_n}\left(1-(1+\alpha_n)e^{-\alpha_n}\right) \tag{48}$$

our conditions on $N(n)$ imply that

$$\lim_{n\to+\infty} nD_n^2 = +\infty . \tag{49}$$

Let us introduce the random variables ν_k ($k = 1, 2, \ldots$) defined as follows: ν_k is the least number such that after distributing ν_k balls among the n boxes, exactly k boxes will be occupied. Let us put further $\delta_1 = \nu_1$, $\delta_k = \nu_k - \nu_{k-1}$ ($k = 2, 3, \ldots$). The random variables δ_k are clearly independent, and we have $\delta_1 \equiv 1$ and

$$\mathbf{P}(\delta_k = j) = \left(\frac{k-1}{n}\right)^{j-1}\left(1-\frac{k-1}{n}\right) \qquad (k = 2, 3, \ldots) . \tag{50}$$

It follows that the characteristic function of δ_k is

$$f_k(t) = \mathbf{M}(e^{it\delta_k}) = \frac{e^{it}\left(1-\frac{k-1}{n}\right)}{1-\left(\frac{k-1}{n}\right)e^{it}} \tag{51}$$

and thus the characteristic function of ν_k is

$$g_k(t) = \prod_{j=1}^{k} f_j(t) = \prod_{j=0}^{k-1} \frac{e^{it}\left(1-\frac{j}{n}\right)}{1-\frac{je^{it}}{n}} . \tag{52}$$

Let us calculate now the expectation and the variance of ν_k. We have evidently

$$\mathbf{M}(\delta_k) = \frac{n}{n-k+1} \tag{53}$$

and thus

$$\mathbf{M}(\nu_k) = \sum_{j=0}^{k-1} \frac{n}{n-j} = n \log \frac{n}{n-k} + O(1) \tag{54}$$

further

$$\mathbf{D}^2(\delta_k) = \frac{1}{\left(1 - \frac{k-1}{n}\right)^2} - \frac{1}{\left(1 - \frac{k-1}{n}\right)} \tag{55}$$

and thus

$$\mathbf{D}^2(\nu_k) = \frac{k}{1 - \frac{k}{n}} - n \log \frac{n}{n-k} + O(1)\,. \tag{56}$$

Now we calculate the third absolute moment of δ_k. We have

$$\mathbf{M}\left(\left|\delta_k - \frac{n}{n-k+1}\right|^3\right) = O\left(\frac{k}{n}\right). \tag{57}$$

Thus Liapounoff's theorem can be applied to $\nu_k = \delta_1 + \delta_2 + \ldots + \delta_k$ if $\frac{k}{\sqrt{n}} \to +\infty$.

It follows that if $k = k(n)$ is a function of n such that $\frac{k(n)}{\sqrt{n}} \to +\infty$ then for any real x

$$\lim_{n \to +\infty} \mathbf{P}\left(\frac{\nu_k - n \log \frac{1}{1-\frac{k}{n}}}{\sqrt{\frac{k}{1-\frac{k}{n}} - n \log \frac{1}{1-\frac{k}{n}}}} < x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\, du\,. \tag{58}$$

Now let us take into account that

$$\mathbf{P}(\nu_k \leq N) = \mathbf{P}(\zeta_{n,N} \leq n - k)\,. \tag{59}$$

Let us suppose that

$$N = n \log \frac{1}{1-\frac{k}{n}} + x \sqrt{\frac{k}{1-\frac{k}{n}} - n \log \frac{1}{1-\frac{k}{n}}} + O(1) \tag{60}$$

and solve this approximate equation approximately for k.

It is easy to see that we get

$$k = n(1 - e^{-\alpha_n}) - x\sqrt{ne^{-\alpha_n}(1 - (1 + \alpha_n)\, e^{-\alpha_n})} + O(1) \tag{61}$$

further $\frac{N^2}{n} \to +\infty$ implies $\frac{k^2}{n} \to +\infty$. Thus it follows from (58) and (59) that

$$\lim_{n \to +\infty} \mathbf{P}\left(\frac{\zeta_{n,N} - ne^{-\alpha_n}}{\sqrt{ne^{-\alpha_n}(1 - (1 + \alpha_n)\, e^{-\alpha_n})}} < x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\, du\,. \tag{62}$$

This proves Theorem 2.

§ 4. Some remarks

Theorems 1 and 2 settle the question of the asymptotic distribution of the number of empty boxes if there are n boxes at all, and denoting by $N = N(n)$ the number of balls, $\frac{N(n)}{\sqrt{n}} \to +\infty$ and $\frac{N(n)}{n} - \log n \to -\infty$. In the limiting case $\frac{N(n)}{n} - \log n \to \gamma$ with some real γ it can be proved by much more elementary methods that the number of empty boxes is in the limit distributed according to Poisson's law, with mean value $e^{-e^{-\gamma}}$ (see e.g. [4] and also [5].) If $\frac{N(n)}{n} - \log n \to +\infty$ then with probability tending to 1 there will be no empty boxes at all. On the other hand if $N = o(\sqrt{n})$ then it is almost sure that all balls are in different boxes and thus with probability tending to 1 the number of empty boxes will be exactly equal to $n - N$. Finally for $N \sim \delta\sqrt{n}$ with $\delta > 0$ the quantity $\zeta_{n,N} - n + N$ will have in the limit a Poisson distribution with mean value $\frac{\delta^2}{2}$. As these results are quite elementary, we do not go into details. We wanted only to point out that from the results of the present paper a complete picture is obtained concerning the behaviour of the distribution of $\zeta_{n,N}$ in the limit if $N = N(n)$ tends to infinity with $n \to +\infty$ in an arbitrary manner.

(Received April 9, 1962.)

REFERENCES

[1] WEISS, I.: "Limiting distribution in some occupancy problems". *Applied Math. and Stat. Laboratory, Stanford University, Technical Report* No. 28, 1955.

[2] ERDŐS, P.—RÉNYI, A.: "On the central limit theorem for samples from a finite population". *MTA Mat. Kut. Int. Közl.* **4** (1959) 49—61.

[3] ERDŐS, P.—RÉNYI, A.: "On the evolution of random graphs." *MTA Mat. Kut. Int. Közl.* **5** (1960) 17—61.

[4] BERNSTEIN, S. N.: *Теория вероятностей*. Москва, 1946. p. 75—76.

[5] RÉNYI, A.: *Wahrscheinlichkeitsrechnung*. VEB Deutscher Verlag der Wissenschaften, Berlin, 1962.

ТРИ НОВЫХ ДОКАЗАТЕЛЬСТВА И ОБОБЩЕНИЕ ОДНОЙ ТЕОРЕМЫ I. WEISS-А

A. RÉNYI

Резюме

В n ящиков брошено наудачу N дробинок; пусть $\zeta_{n,N}$ означает число пустых ящиков. В работе даны три новых доказательства теоремы, доказанной I. Weiss-ом [1], что если $N = N(n)$ есть такая функция от n, что для $n \to +\infty$ имеем $\frac{N(n)}{n} \to \alpha$ $(0 < \alpha < +\infty)$, тогда имеет место

$$\text{(I)} \qquad \lim_{n \to +\infty} \mathbf{P}\left(\frac{\zeta_{n,N} - ne^{-\frac{N}{n}}}{\sqrt{ne^{-\frac{N}{n}}\left(1 - \left(1 + \frac{N}{n}\right) e^{-\frac{N}{n}}\right)}} < x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du .$$

При этом обобщается результат Weiss-а, так как доказывается, что (1) имеет место также в случае $\frac{N(n)}{n} \to 0$ если $\frac{N^2(n)}{n} \to +\infty$ и в случае $\frac{N(n)}{n} \to +\infty$ если $\frac{N(n)}{n} - \log n \to -\infty$. Доказательства используют метод характеристических функций.