Z. Wahrscheinlichkeitstheorie verw. Gebiete
41, 1–11 (1977)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete
© by Springer-Verlag 1977

A Central Limit Theorem for Empirical Histograms

D.A. Freedman*
Statistics Department, University of California, Berkeley, California 94720, USA

1. Introduction

The central limit theorem is often used heuristically to justify the approximation of histograms for data by the normal curve. This argument can be made precise through the following model. There is some basic random variable X. Take the sum of n independent copies of X, and then take k independent copies of this sum: the data is considered as the sequence of observed values of these sums. Provided X is well-behaved, n is large, and k is large in relation to n, this model rigorously justifies the use of the curve to approximate the histogram for the data. The object is to explore this idea in detail.

Let $X_1, X_2, \ldots$ be independent, identically distributed random variables. Suppose the X_i are integer-valued, and have period 1:

$$\text{g.c.d. } \{m\colon P(X_i=m)>0\}=1.$$

Suppose also $E|X_1^3|<\infty$. Let $S_n=X_1+\cdots+X_n$. Take k independent copies of S_n, and let $H_{n,k}$ be their empirical histogram, but rescaled by the mean and standard deviation of S_n. If k increases rapidly enough, $H_{n,k}$ should tend to the normal curve. What is the critical rate of increase for k? To begin with, S_n has of order $n^{\frac{1}{2}}$ values in the center of its distribution, and there must be a large number of observations at each of these values, so $k/n^{\frac{1}{2}}\to\infty$ is a natural guess for the critical rate. However, this turns out to be a bit too slow: the right condition is that $k/n^{\frac{1}{2}} \log n\to\infty$.

(1) **Theorem.** *If k and n approach infinity in such a way that $k/(n^{\frac{1}{2}}\log n)\to\infty$, then $H_{n,k}$ converges uniformly to the normal curve, in probability.*

(2) **Proposition.** *If $k/n^{\frac{1}{2}}\to\infty$ but $k=O(n^{\frac{1}{2}}\log n)$, then $H_{n,k}$ converges to the normal curve pointwise but not uniformly, in probability.*

(3) **Proposition.** *If $k=O(n^{\frac{1}{2}})$, then $H_{n,k}$ does not even converge pointwise to the normal curve, in probability.*

* Prepared with the partial support of National Science Foundation Grant MCS75-09459

To state a more precise estimate, let

$$\mu = E(X_1) \quad \text{and} \quad \sigma^2 = \operatorname{Var} X_1.$$

Let N_j be the number of copies of S_n which equal j. So N_j depends also on n and k. And $H_{n,k}(x) = \sigma n^{\frac{1}{2}} N_j/k$ for x in the interval

(4) $(j - \frac{1}{2} - n\mu)/\sigma n^{\frac{1}{2}} < x < (j + \frac{1}{2} - n\mu)/\sigma n^{\frac{1}{2}}$.

Let H_n be the probability histogram for S_n, rescaled to have mean 0 and variance 1. Thus, $H_n(x) = \sigma n^{\frac{1}{2}} P(S_n = j)$ for x in the interval (4).

Theorem (1) follows from the sharper result.

(5) **Theorem.** *If* $k/(n^{\frac{1}{2}} \log n) \to \infty$, *then*

$$\max(H_{n,k} - H_n)/(n^{\frac{1}{2}} \log n/k)^{\frac{1}{2}} \to \sigma^{\frac{1}{2}}(2\pi)^{-\frac{1}{4}},$$
$$\min(H_{n,k} - H_n)/(n^{\frac{1}{2}} \log n/k)^{\frac{1}{2}} \to -\sigma^{\frac{1}{2}}(2\pi)^{-\frac{1}{4}}.$$

The convergence is in probability. The max and min are taken over the whole line $(-\infty, \infty)$.

Theorem (5) will be proved in Section 3, after some preliminaries are dealt with in Section 2. Proposition (2) can be sharpened in a similar way, and this will be done in Section 4, where Proposition (3) will also be established.

2. Some Lemmas

First, some results on the binomial distribution. Let N be binomial with parameters k and p, where $0 < p < 1$. Both are varying.

(6) **Lemma.** *Let* $u > 0$. *Then*

$$P\{N > kp(1+u)\} \leqq \exp\{-g(u)\, kp\},$$

where

$$g(u) = (1+u)\log(1+u) - u.$$

This function of u is convex and strictly increasing. For all positive u,

$$\tfrac{1}{2}u^2/(1+u) < g(u) < \tfrac{1}{2}u^2.$$

In particular,

$$P\{N > kp + m\} \leqq \exp\{-\tfrac{1}{2}m^2/(kp+m)\}.$$

The main inequality is well known. For a proof in a general martingale context, see [2]. The behavior of g is easily checked.

(7) **Lemma.** *Fix* $\varepsilon > 0$. *There is a positive* δ *such that for all positive integers* a, b, k *satisfying*

$$a > 1/\delta, \quad b > 1/\delta, \quad k > 1/\delta, \quad a + b < \delta k, \quad b < \delta(k - a - b)^{\frac{1}{2}}$$

the ratio $P(N=a+b)/P(N=a)$ is bounded between $(1\pm\varepsilon)$ times

$$\left[\frac{a}{a+b}\right]^{\frac{1}{2}}\exp\left[-g\left(\frac{b}{a}\right)a\right]\left[\frac{(k-a)\,p}{a(1-p)}\right]^{b}.$$

Proof. By Stirling's formula, the ratio of the two probabilities is asymptotic to

$$\left(\frac{k-a}{k-a-b}\right)^{\frac{1}{2}}\cdot\left(\frac{a}{a+b}\right)^{\frac{1}{2}}\cdot\frac{a^a}{(a+b)^{a+b}}\cdot\frac{(k-a)^{k-a}}{(k-a-b)^{k-a-b}}\cdot\left(\frac{p}{1-p}\right)^{b}.$$

Since $b/(k-a)$ is small, the first factor is essentially one. Next

$$\frac{(k-a)^{k-a}}{(k-a-b)^{k-a-b}}=(k-a)^b\left(1+\frac{b}{k-a-b}\right)^{k-a-b}\approx(k-a)^b e^b,$$

because $b^2/(k-a-b)$ is small. Then

$$\frac{a^a}{(a+b)^{a+b}}\,e^b=\exp\left[-g\left(\frac{b}{a}\right)a\right]\frac{1}{a^b}.$$

The lemma follows by collecting factors.

(8) **Corollary.** *Fix* $\varepsilon>0$. *There is* $\delta>0$ *such that*

$$p<\delta,\qquad kp>1/\delta,\qquad m<\delta k^{\frac{1}{2}},\qquad m<\delta kp$$

imply

$$P(N>kp+m)>(1-\varepsilon)(kp/2\pi)^{\frac{1}{2}}m^{-1}\exp\{-\tfrac{1}{2}m^2/kp\}.$$

Proof. Use Lemma (7), with a the least integer above kp, and b running through the integers between m and $2m$. Clearly, $m=o(kp)$ makes $a/(a+b)$ and $\left[\frac{(k-a)\,p}{a(1-p)}\right]^b$ both tend to 1, uniformly for b in the range m to $2m$. And $g(b/a)\leqq\frac{1}{2}b^2/a^2$ by (6), so eventually $-g(b/a)\,a\geqq-\frac{1}{2}b^2/kp$.

As usual, $P(N=a)\approx(2\pi kp)^{-\frac{1}{2}}$, which eventually exceeds $(1-\frac{1}{2}\varepsilon)\,(2\pi kp)^{-\frac{1}{2}}$. From then on, $P(N=a+b)$ exceeds

$$(1-\tfrac{1}{2}\varepsilon)(2\pi kp)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}b^2/kp\}.$$

The sum of this expression for integer b's between m and $2m$ can be bounded below by

$$(1-\tfrac{1}{2}\varepsilon)(2\pi kp)^{-\frac{1}{2}}\int_{m+2}^{2m}\exp\{-\tfrac{1}{2}u^2/kp\}\,du.$$

By the usual manipulations, this can in turn be bounded below by

$$(1-\varepsilon)(kp/2\pi)^{\frac{1}{2}}m^{-1}\exp\{-\tfrac{1}{2}m^2/kp\}.$$

Here is another corollary of (7).

(9) **Corollary.** *Fix $u>0$. For any $\varepsilon>0$, there is a $\delta>0$ such that*

$$p<\delta, \quad kp>1/\delta, \quad k^{\frac{1}{2}}p<\delta$$

imply

$$P\{N>kp(1+u)\}>\exp\{-(1+\varepsilon)\,g(u)\,kp\}.$$

Proof. Since $k^{\frac{1}{2}}p$ is small, Lemma (7) can be used, with a the least integer exceeding kp, and b the least integer exceeding kpu. This time, $a/(a+b)\to 1/(1+u)$.

Let $a=kp+\theta$, so

$$\left[\frac{(k-a)\,p}{a(1-p)}\right]^b=\left(\frac{1-\theta/k\bar{p}}{1+\theta/kp}\right)^b.$$

This has $\liminf \geqq e^{-\theta}\geqq 1/e$. Finally

$$g\left(\frac{b}{a}\right)a\approx g(u)\,kp.$$

The constants and the factor $(kp)^{-\frac{1}{2}}$ can be absorbed by $\exp\{-\varepsilon\,g(u)\,kp\}$.

Now another topic.

(10) **Lemma.** *Let $X_1, X_2, \ldots$ be independent and identically distributed, taking the values $\{1,2,3\}$. Fix k. Let N_i be the number of $\{X_1,\ldots,X_k\}$ equal to i. Then*

$$P\{N_1>n_1 \text{ and } N_2>n_2\}\leqq P\{N_1>n_1\}\cdot P\{N_2>n_2\}.$$

Proof. Let $p(i)=P(X_1=i)$. Given $N_2=m$, N_1 is distributed like a binomial random variable with parameters $k-m$ and $p(1)/[p(1)+p(3)]$. So $f(m)=P(N_1>n_1|N_2=m)$ decreases as m increases. Let $q_m=P(N_2=m)$ and $q=P(N_2>n_2)$. Clearly,

$$\frac{1}{1-q}\sum_0^{n_2} f(m)\,q_m\geqq\frac{1}{q}\sum_{n_2+1}^{k} f(m)\,q_m.$$

Rearranging,

$$\sum_0^{n_2} f(m)\,q_m\geqq\sum_{n_2+1}^{k} f(m)\,q_m\left(\frac{1}{q}-1\right)$$

or

$$\sum_0^{k} f(m)\,q_m\geqq\sum_{n_2+1}^{k} f(m)\,q_m/q.$$

That is,

$$P(N_1>n_1)\geqq P\{N_1>n_1|N_2>n_2\}.$$

The next fact is the local central limit theorem recorded here for ease of reference. Recall that $X_1, X_2, \ldots$ are independent, identically distributed, integer-valued, aperiodic, $\mu=E(X_1)$, $\sigma^2=\operatorname{Var} X_1$, and $E|X_1^3|<\infty$. The local central limit theorem states:

(11) $n^{\frac{1}{2}} P(S_n=j)=\sigma^{-1}(2\pi)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\frac{(j-n\mu)^2}{\sigma^2 n}\right]+O(n^{-\frac{1}{2}})$, uniformly in j.

In particular

(12) If $\varepsilon>0$, then $n^{\frac{1}{2}} \max_j P(S_n=j)<(1+\varepsilon)/\sigma(2\pi)^{\frac{1}{2}}$ eventually.

(13) Let $C_n(\theta)=\{j\colon |j-n\mu|<\theta\sigma n^{\frac{1}{2}}\}$.

(14) Fix $\varepsilon>0$. For sufficiently small θ, for all large n,

$$n^{\frac{1}{2}} \min_{j\in C_n(\theta)} P(S_n=j)>(1-\varepsilon)/\sigma(2\pi)^{\frac{1}{2}}.$$

3. The Proof of the Main Theorem

For a moment, keep n fixed. There are k independent copies of S_n. Recall that N_j is the number of these copies equal to j. After rescaling, the discrepancy $H_{n,k}(x)-H_n(x)$ between the empirical histogram and the probability histogram can be expressed as

$$\sigma n^{\frac{1}{2}}(N_j-k p_j)/k, \; p_j=p(S_n=j)$$

for x in the interval (4). This discrepancy exceeds $\sigma y[(n^{\frac{1}{2}}\log n/k]^{\frac{1}{2}}$ when

$$N_j>k p_j+y[(k\log n)/n^{\frac{1}{2}}]^{\frac{1}{2}}.$$

Let $A_j=A_j(n,y)$ be the event that this inequality is satisfied. Here, $y>0$.

(15) **Lemma.** *Let T_n be the set of j with $|j-n\mu|>\sigma(n\log n)^{\frac{1}{2}}$. Let $B_n=\bigcup_j\{A_j\colon j\in T_n\}$. Then $P(B_n)\to 0$, for each y.*

Proof. Clearly, $E(N_j)=kp_j$ and $\operatorname{Var} N_j\leqq k p_j$. By Čebyšev's inequality,

$$P(A_j)\leqq\frac{1}{y^2}\frac{n^{\frac{1}{2}}}{\log n}p_j.$$

So

$$P(B_n)\leqq\frac{1}{y^2}\frac{n^{\frac{1}{2}}}{\log n}p(S_n\in T_n).$$

But the Berry-Eseen bound [1, p. 542] shows

$$P(S_n\in T_n)\approx n^{-\frac{1}{2}}.$$

So

$$P(B_n)=O(1/\log n)\to 0.$$

(16) **Lemma.** *Suppose $k/(n^{\frac{1}{2}}\log n)\to\infty$. Let $y=(1+\varepsilon)\sigma^{-\frac{1}{2}}(2\pi)^{-\frac{1}{4}}$ for some $\varepsilon>0$. Define T_n as in* (15). *Let $D_n=\bigcup_j\{A_j\colon j\notin T_n\}$. Then $P(D_n)\to 0$.*

Proof. Lemma (6) implies

$$P(A_j) \leqq \exp\left[-\tfrac{1}{2} y^2 (\log n)/(n^{\frac{1}{2}} p_j + \theta_n)\right]$$

where

$$\theta_n = y[n^{\frac{1}{2}}(\log n)/k]^{\frac{1}{2}} \to 0.$$

Lemma (12) implies

$$n^{\frac{1}{2}} p_j \leqq \sigma^{-1}(2\pi)^{-\frac{1}{2}} + o(1) \qquad \text{for all } j.$$

For large enough n,

$$\theta_n + n^{\frac{1}{2}} p_j \leqq (1+\varepsilon)\,\sigma^{-1}(2\pi)^{-\frac{1}{2}} \qquad \text{for all } j$$

and then

$$P(A_j) \leqq \exp[-\tfrac{1}{2}(1+\varepsilon)\log n] = n^{-\frac{1}{2}(1+\varepsilon)}.$$

But the number of j's not in T_n is only $2\sigma(n \log n)^{\frac{1}{2}}$, so

$$P(B_n) = O[(\log n)^{\frac{1}{2}}/n^{\varepsilon/2}] \to 0.$$

By combining (15) and (16):

(17) **Corollary.** *Suppose* $k/n^{\frac{1}{2}} \log n \to \infty$. *Let* $\varepsilon > 0$. *Then*

$$P\{\max(H_{n,k} - H_n) > (1+\varepsilon)\,\sigma^{\frac{1}{2}}(2\pi)^{-\frac{1}{4}}[n^{\frac{1}{2}}(\log n)/k]\} \to 0.$$

In a similar way, under the same conditions one proves

(18) $P\{\min(H_{n,k} - H_n) < -(1+\varepsilon)\,\sigma^{\frac{1}{2}}(2\pi)^{-\frac{1}{4}}[n^{\frac{1}{2}}(\log n)/k]\} \to 0.$

This proves half of theorem (5). The proof of the other half begins with

(19) **Lemma.** *Suppose* $k/n^{\frac{1}{2}} \log n \to \infty$. *Let* $\theta > 0$ *and let* $C_n(\theta) = \{j\colon |j - n\mu| < \theta n^{\frac{1}{2}}\}$, *as in* (13).

Fix $\varepsilon > 0$. *Let* $y = (1-\varepsilon)\,\sigma^{-\frac{1}{2}}(2\pi)^{-\frac{1}{4}}$. *Then*

$$\sum\{P(A_j)\colon\ j \in C_n(\theta)\} \to \infty.$$

Proof. Since j is confined to $C_n(\theta)$, Lemma (8) can be used with $m = y(k \log n/n^{\frac{1}{2}})^{\frac{1}{2}}$. (The relationship (11) can be used to verify this, from the condition $k/n^{\frac{1}{2}} \log n \to \infty$.) For large enough n,

$$P(A_j) \geqq 0.9(2\pi)^{-\frac{1}{2}} y^{-1}(n^{\frac{1}{2}} p_j/\log n)^{\frac{1}{2}} \exp[-\tfrac{1}{2} y^2 \log n/(n^{\frac{1}{2}} p_j)].$$

Now use (14). For sufficiently small θ, for all large n,

$$n^{\frac{1}{2}} p_j > (1-\varepsilon)\,\sigma^{-1}(2\pi)^{-\frac{1}{2}} \qquad \text{for all } j.$$

Then

$$-y^2/(n^{\frac{1}{2}} p_j) > (1-\varepsilon)$$

and

$$P(A_j) \geqq K(\log n)^{-\frac{1}{2}} n^{-\frac{1}{2}(1-\varepsilon)}$$

for all $j \in C_n(\theta)$. Here, K is a positive constant, and θ must be small. Still, $C_n(\theta)$ contains $2\theta n^{\frac{1}{2}}$ indices j, so

$$\sum_j \{P(A_j)\colon\ j \in C_n(\theta)\} \geqq 2\theta K(\log n)^{-\frac{1}{2}} n^{\varepsilon/2} \to \infty.$$

Plainly, this last holds for any θ.

(20) **Proposition.** *Suppose* $k/n^{\frac{1}{2}} \log n \to \infty$. *Fix* $\varepsilon > 0$. *Then*

$$P\{\max(H_{n,k} - H_n) > (1-\varepsilon)\,\sigma^{\frac{1}{2}}(2\pi)^{-\frac{1}{4}}[n^{\frac{1}{2}}(\log n)/k]\} \to 1.$$

Proof. Use the notation of (19). It is enough to prove that

$$\bigcup \{A_j\colon\ j \in C_n(\theta)\}$$

has probability near 1. Here, $y = (1-\varepsilon)\,\sigma^{-\frac{1}{2}}(2\pi)^{-\frac{1}{4}}$, and θ is small. Confine j to $C_n(\theta)$. Let

$$X = \sum_j 1_{A_j}.$$

What (19) says that $E(X) \to \infty$. The problem is to conclude that $P(X > 0) \to 1$. Now

$$X^2 = \sum_j 1_{A_j} + \sum_{j \neq j'} 1_{A_j} 1_{A'_j}$$

so

$$E(X^2) = E(X) + \sum_{j \neq j'} P(A_j \cap A_{j'}).$$

Lemma (10) shows

$$P(A_j \cap A_{j'}) \leqq P(A_j) \cdot P(A_{j'}).$$

So

$$\begin{aligned} E(X^2) &\leqq E(X) + \sum_{j \neq j'} P(A_j) \cdot P(A_{j'}) \\ &\leqq E(X) + [\sum_j P(A_j)]^2 \\ &= E(X) + [E(X)]^2. \end{aligned}$$

That is, $\operatorname{Var} X \leqq E(X)$. By Čebyšev's inequality

$$\begin{aligned} P(X \leqq 0) &\leqq \operatorname{Var} X/[E(X)]^2 \\ &\leqq 1/E(X) \to 0. \end{aligned}$$

In a similar way, under the same conditions one proves

(21) $$P\{\min(H_{n,k} - H_n) < -(1-\varepsilon)\,\sigma^{\frac{1}{2}}(2\pi)^{-\frac{1}{4}}[n^{\frac{1}{2}}(\log n)/k]\} \to 1.$$

Theorem (5) follows by combining (17) and (20) for the first assertion, (18) and (21) for the second.

4. The Case $k \sim n^{\frac{1}{2}} \log n$

(22) **Theorem.** *Suppose $k/n^{\frac{1}{2}} \log n \to \lambda$, a finite positive constant. Then*

$$\max(H_{n,k} - H_n)/(n^{\frac{1}{2}} \log n/k)^{\frac{1}{2}}$$

converges in probability to the number $c(\lambda)$, which exceeds $\sigma^{\frac{1}{2}}(2\pi)^{-\frac{1}{4}}$. The number $c(\lambda)$ depends monotonically on λ; as λ tends to 0 it goes to ∞, as λ goes to infinity it goes to $\sigma^{\frac{1}{2}}(2\pi)^{-\frac{1}{4}}$. In fact, let $u(\lambda)$ be the unique solution to the equation

$$g(u) = \tfrac{1}{2}\sigma(2\pi)^{\frac{1}{2}}/\lambda$$

where $g(u)$ was defined by (6) as

$$(1+u)\log(1+u) - u.$$

Then $c(\lambda) = (\lambda/2\pi)^{\frac{1}{2}} u(\lambda)$.

Note. $\max(H_{n,k} - H_n)$ is therefore tending to $u(\lambda)/(2\pi)^{\frac{1}{2}}$.

There is a similar result for the min, with $-c(\lambda)$ in place of $c(\lambda)$.
First, the analysis.

(23) **Lemma.** *Let $0 < \lambda < \infty$. There is a unique solution $u(\lambda)$ to the equation*

$$g(u) = \tfrac{1}{2}\sigma(2\pi)^{\frac{1}{2}}/\lambda.$$

Let $c(\lambda) = (\lambda/2\pi)^{\frac{1}{2}} u(\lambda)$. Then c is monotone decreasing, $c(0+) = \infty$, $c(\infty) = \sigma^{\frac{1}{2}}(2\pi)^{-\frac{1}{4}}$.

Proof. The function g is continuous and strictly increasing, from 0 at 0 to ∞ at ∞. So $u(\lambda)$ is well defined. Since the right side of the equation is monotone in λ, so is $u(\lambda)$, and $u(0+) = \infty$. Since $g(u) \approx \frac{1}{2}u^2$ for small u, if λ is large then $u(\lambda)$ is asymptotic to the solution of

$$\tfrac{1}{2}u^2 = \tfrac{1}{2}\sigma(2\pi)^{\frac{1}{2}}/\lambda,$$

that is

$$u(\lambda) \approx \sigma^{\frac{1}{2}}(2\pi)^{\frac{1}{4}}/\lambda^{\frac{1}{2}} \quad \text{as } \lambda \to \infty.$$

So

$$c(\lambda) \to \sigma^{\frac{1}{2}}(2\pi)^{-\frac{1}{4}} \quad \text{as } \lambda \to \infty.$$

(24) **Lemma.** *Suppose the conditions and notations of (22). Let $\varepsilon > 0$. Let $u = (1+\varepsilon)u(\lambda)$. As in (15), define T_n as the set of j with $|j - n\mu| > \sigma(n \log n)^{\frac{1}{2}}$. Let*

$$A_j = \{N_j > k p_j + u k \sigma^{-1}(2\pi n)^{-\frac{1}{2}}\}.$$

Let D_n be the union of A_j for $j \notin T_n$. Then $P(D_n) \to 0$.

Proof. Let $u_j = u\sigma^{-1}(2\pi n)^{-\frac{1}{2}} p_j^{-1}$. So $A_j = \{N_j > k p_j(1+u_j)\}$ and (6) implies

$$P(A_j) \leqq \exp[-g(u_j)\, k p_j].$$

But

$$g(u_j)\, k p_j = \frac{g(u_j)}{u_j}\, k u \sigma^{-1}(2\pi n)^{-\frac{1}{2}}.$$

Now (12) implies that for large enough n,

$$u_j > u/(1+\varepsilon) \qquad \text{for all } j.$$

Since g is convex

$$\begin{aligned} g(u_j)/u_j &> g[u/(1+\varepsilon)]/[u/(1+\varepsilon)] \\ &= g[u(\lambda)]/u(\lambda). \end{aligned}$$

So

$$\begin{aligned} g(u_j)\, k p_j &> (1+\varepsilon)\, g[u(\lambda)]\, k \sigma^{-1}(2\pi n)^{-\frac{1}{2}} \\ &> (1+\tfrac{1}{2}\varepsilon)\, g(u(\lambda))\, \lambda \sigma^{-1}(2\pi)^{-\frac{1}{2}} \log n \\ &= \tfrac{1}{2}(1+\tfrac{1}{2}\varepsilon) \log n. \end{aligned}$$

Thus

$$P(D_n) < \sigma(n \log n)^{\frac{1}{2}}\, n^{-\frac{1}{2}(1+\frac{1}{2}\varepsilon)} \to 0.$$

Since

$$u k \sigma^{-1}(2\pi n)^{-\frac{1}{2}} \sim u \log n$$

and

$$y(k \log n / n^{\frac{1}{2}})^{\frac{1}{2}} \sim y \log n,$$

Lemma (15) shows $P(B_n) \to 0$ for any u, where B_n is the union of A_j for $j \in T_n$.

(25) **Corollary.** *With the conditions and notation of* (22), *for any* $\varepsilon > 0$,

$$\max(H_{n,k} - H_n)/(n^{\frac{1}{2}} \log n/k)^{\frac{1}{2}} < (1+\varepsilon)\, c(\lambda)$$

with probability tending to one.

In the other direction,

(26) **Proposition.** *With the conditions and notation of* (22), *for any* $\varepsilon > 0$,

$$\max(H_{n,k} - H_n)/(n^{\frac{1}{2}} \log n/k)^{\frac{1}{2}} > (1-\varepsilon)\, c(\lambda)$$

with probability tending to one.

Proof. Define $C_n(\theta) = \{j\colon |j - n\mu| < \theta n^{\frac{1}{2}}\}$, as in (13), with θ small and positive. Define A_j as in (24). Let $u = (1-\varepsilon)\, u(\lambda)$. Then

(27) $\sum \{P(A_j)\colon j \in C_n(\theta)\} \to \infty.$

Indeed, corollary (9) may be used for all $j \in C_n(\theta)$: since $p_j \sim n^{-\frac{1}{2}}$, $k p_j$ gets large and $k^{\frac{1}{2}} p_j$ gets small. (This can be made rigorous using (12) and (14).) Fix $\varepsilon' > 0$. Eventually,

$$P(A_j) \geqq \exp[-(1+\varepsilon')\, g(u)\, k\, p_j].$$

For some $\varepsilon^* > 0$, which depends on ε,

$$g(u) = (1-\varepsilon^*)\, \tfrac{1}{2}\, \sigma (2\pi)^{\frac{1}{2}}/\lambda.$$

Eventually,

$$k < (1+\varepsilon')\, \lambda\, n^{\frac{1}{2}} \log n,$$

$$\max_j p_j < (1+\varepsilon')\, n^{-\frac{1}{2}} \sigma^{-1} (2\pi)^{-\frac{1}{2}}.$$

So

$$P(A_j) \geqq \exp[-(1-\varepsilon^*)(1+\varepsilon')^3\, \tfrac{1}{2} \log n].$$

Choose ε' so small that $(1-\varepsilon^*)(1+\varepsilon')^3 < 1 - \frac{1}{2}\varepsilon^*$. Then $P(A_j) \geqq n^{-\frac{1}{2}+\frac{1}{4}\varepsilon^*}$, so the sum in (27) is of order $n^{\varepsilon^*/4}$.

The proof is completed as in (19).

(28) **Lemma.** *Suppose* $k/n^{\frac{1}{2}} \log n \to 0$ *but* $k/n^{\frac{1}{2}} \to \infty$. *Then* $\max(H_{n,k} - H_n) \to \infty$ *in probability.*

Proof. Fix $u > 0$. Arguing as in (26), for any positive ε, eventually

$$P(A_j) \geqq \exp[-(1+\varepsilon^*)\, g(u)\, k\, p_j],$$

$$p_j < \sigma (1+\varepsilon)\, n^{-\frac{1}{2}} \sigma^{-1} (2\pi)^{-\frac{1}{2}},$$

$$k < \lambda\, \varepsilon\, n^{\frac{1}{2}} \log n$$

so that for all $j \in C_n(\theta)$,

$$P(A_j) \geqq \exp[-q(u, \varepsilon) \log n]$$

where

$$q(u, \varepsilon) = \varepsilon (1+\varepsilon)^2\, g(u)\, \sigma^{-1} (2\pi)^{-\frac{1}{2}}.$$

Choose ε so small that $q(u, \varepsilon) < \frac{1}{2}$. Then the sum in (27) becomes infinite, and the argument for (19) shows

$$P\Big(\bigcup_{j \in C_n(\theta)} A_j\Big) \to 1.$$

This shows that

$$\max(H_{n,k} - H_n) > u/(2\pi)^{\frac{1}{2}}$$

with probability near one. Since u can be arbitrarily large, the results follows.

The Proof of (2). Theorem (22) and Lemma (28) show that $\max(H_{n,k} - H_n)$ tends in probability to a positive constant or $+\infty$, preventing $H_{n,k}$ from converging uniformly to the normal curve.

For pointwise convergence, note that

$$n^{\frac{1}{2}}(N_j - k\,p_j)/k$$

has mean 0 and variance $n\,p_j(1-p_j)/k \sim n^{\frac{1}{2}}/k$.

4. The Case $k = O(n^{\frac{1}{2}})$

To begin with, suppose $k/n^{\frac{1}{2}} \to \lambda$, a finite positive number. Fix x, a real number. Then

$$H_{n,k}(x) - H_n(x) = \sigma\, n^{\frac{1}{2}}(N_j - k\,p_j)/k$$

where

$$(j - n\,\mu)/\sigma\, n^{\frac{1}{2}} \to x.$$

Now N_j is binomial with parameters k and p_j, and (11) shows that $k\,p_j$ tends to

$$\theta_x = \lambda\,\sigma^{-1}(2\pi)^{-\frac{1}{2}}\exp(-x^2/2).$$

So N_j converges in distribution to N_x which is Poisson with parameter θ_x, and $H_{n,k}(x) - H_n(x)$ is distributed like

$$(\sigma/\lambda)\,(N_x - \theta_x).$$

(Asymptotically, the N_x's are mutually independent. This is a bit surprising, for c independent variables are involved. As a result, it is easy to see that

$\max(H_{n,k} - H_n) \to \infty$ in probability.)

Now suppose $k/n^{\frac{1}{2}} \to 0$. The same argument shows that $P(N_j = 0) \to 1$, so

$$H_{n,k}(x) - H_n(x) \to -(2\pi)^{-\frac{1}{2}}\exp(-x^2/2)$$

in probability.

Combining these results proves (3).

References

1. Feller, W.: An Introduction to the Theory of Probability and Its Applications, Vol. II, 2nd ed. New York: Wiley
2. Freedman, D.: Another note on the Borel-Cantelli Lemma and Strong Law, with the Poisson approximation as a By Product. Ann. Probability, **6**, 910–925 (1973)

Received September 10, 1976