

AN UNDERGRADUATE LECTURE ON THE CENTRAL LIMIT THEOREM

N.V. KRYLOV

In the first section we explain why the central limit theorem for the binomial $1/2$ distributions is natural. The second section contains the proof and the third section gives a refinement of our basic estimate.

1. SETTING AND MOTIVATION

Let S_n be the number of successes in n independent Bernoulli trials with success probability $p = 1/2$.

First, one realizes that

$$P(S_n = k) = 2^{-n} \binom{n}{k} = 2^{-n} \frac{n!}{k!(n-k)!}. \quad (1)$$

Then comes the observation that $ES_n = n/2$, $\text{Var } S_n = n/4$. This shows that the random variables

$$(S_n - n/2)/\sqrt{n}$$

are centered and have constant variance. One may hope that their distributions converge to something as $n \rightarrow \infty$. In other words, one may hope that for any $a < b$

$$P(a < (S_n - n/2)/\sqrt{n} \leq b) \quad (2)$$

has a limit as $n \rightarrow \infty$.

Let us try to guess what the limit could be. Set

$$y_{kn} = n^{-1/2}(k - n/2), \quad k = 0, 1, \dots, n.$$

Observe that

$$y_{k+1,n} = y_{kn} + n^{-1/2}, \quad k = y_{kn}\sqrt{n} + n/2.$$

$$P(S_n = k) = P(n^{-1/2}(S_n - n/2) = y_{kn})$$

and, for each n , introduce a function

$$f_n(y_{kn}) = P(n^{-1/2}(S_n - n/2) = y_{kn}),$$

so that

$$f_n(y_{kn}) = P(S_n = k).$$

For $k \leq n - 1$ simple manipulations using (1) show that

$$P(S_n = k) = 2^{-n} \frac{n!}{(k+1)!(n-k-1)!} \frac{k+1}{n-k},$$

implying that

$$\begin{aligned} f_n(y_{kn}) &= f_n(y_{kn} + n^{-1/2}) \frac{y_{kn}\sqrt{n} + n/2 + 1}{n/2 - y_{kn}\sqrt{n}}, \\ f_n(y_{kn} + n^{-1/2}) &= f_n(y_{kn}) \frac{n/2 - y_{kn}\sqrt{n}}{y_{kn}\sqrt{n} + n/2 + 1}, \end{aligned} \quad (3)$$

$$\begin{aligned} f_n(y_{kn} + n^{-1/2}) - f_n(y_{kn}) &= f_n(y_{kn}) \left(\frac{n/2 - y_{kn}\sqrt{n}}{y_{kn}\sqrt{n} + n/2 + 1} - 1 \right) \\ &= -f_n(y_{kn}) \frac{2y_{kn}\sqrt{n} + 1}{y_{kn}\sqrt{n} + n/2 + 1}. \end{aligned} \quad (4)$$

Next, the probability in (2) is

$$\sum_{k: a < y_{kn} \leq b} f_n(y_{kn})$$

and this sum contains roughly speaking $(b - a)\sqrt{n}$ terms. Therefore, each term should be of order \sqrt{n} . One way or another, let us believe that, for a function ϕ ,

$$\phi_n(y_{kn}) := \sqrt{n} f_n(y_{kn}) \rightarrow \phi(y)$$

as $n \rightarrow \infty$ and $y_{kn} \rightarrow y$.

For ϕ_n equation (4) becomes

$$\phi_n(y_{kn} + n^{-1/2}) - \phi_n(y_{kn}) = -\phi_n(y_{kn}) \frac{2y_{kn}\sqrt{n} + 1}{y_{kn}\sqrt{n} + n/2 + 1}. \quad (5)$$

Let us pretend that (5) holds for ϕ rather than ϕ_n and for all y in place of y_{kn} rather than for y_{kn} of a special form, that is

$$\phi(y + n^{-1/2}) - \phi(y) = -\phi(y) \frac{2y\sqrt{n} + 1}{y\sqrt{n} + n/2 + 1}.$$

Devide both parts by $n^{-1/2}$ and let $n \rightarrow \infty$ to see that, naturally, $\phi_n(y)$ should be close to a ϕ that is a solution of

$$\phi'(y) = -4y\phi(y).$$

All solutions of the latter equation are known to be ce^{-2y^2} , and this is a way to explain why the normal distribution arises in this particular instance of the central limit theorem.

2. FIRST RIGOROUS RESULT

As above we assume that $p = 1/2$ and add the assumption that n is an even number.

Theorem 1. *Let $\beta \in [1/2, 3/4]$ and n be even. Then for $|k - n/2| \leq n^\beta$ (that is $|y_{kn}| \leq n^{\beta-1/2}$)*

$$c_n^{-1} e^{2y_{kn}^2} f_n(y_{kn}) = e^{O(n^{4\beta-3})} = 1 + O(n^{4\beta-3}), \quad (6)$$

where

$$c_n = P(S_n = n/2), \quad y_{kn} = (k - n/2)n^{-1/2}.$$

Furthermore,

$$\lim_{n \rightarrow \infty} c_n \sqrt{n} = \sqrt{2/\pi}, \quad (7)$$

and if $\beta \in [1/2, 3/4)$, then

$$\lim_{n \rightarrow \infty} \max_{|k - n/2| \leq n^\beta} |(n\pi/2)^{1/2} e^{2y_{kn}^2} P(S_n = k) - 1| = 0. \quad (8)$$

Finally, for any $a, b \in (-\infty, \infty)$ such that $a < b$,

$$\lim_{n \rightarrow \infty} P(a\sqrt{n} \leq S_n - n/2 \leq b\sqrt{n}) = \sqrt{2/\pi} \int_a^b e^{-2y^2} dy. \quad (9)$$

Proof. To prove (6) it suffices to concentrate on $k \geq n/2$ when $y_{kn} \geq 0$. Set

$$\begin{aligned} g_n(y_{kn}) &= \ln f_n(y_{kn}), \\ x_{kn} &= 2y_{kn}n^{-1/2} = 2(k - n/2)n^{-1} = 2k/n - 1 \end{aligned}$$

and use (3) to get

$$\begin{aligned} g_n(y_{k+1,n}) - g_n(y_{kn}) &= \ln(1 - x_{kn}) - \ln(1 + x_{k+1,n}) \\ &= -x_{kn} - x_{k+1,n} + (1/2)[x_{k+1,n}^2 - x_{kn}^2] - (1/3)[\bar{x}_{kn}^3 + \bar{x}_{k+1,n}^3], \end{aligned} \quad (10)$$

where $\bar{x}_{kn} \leq x_{kn}$, $\bar{x}_{k+1,n} \leq x_{k+1,n}$. For $n/2 \leq k \leq n/2 + n^\beta$ we have

$$0 \leq y_{kn} \leq n^{\beta-1/2}, \quad 0 \leq x_{kn} \leq 2n^{\beta-1} \leq 2, \quad \bar{x}_{kn}^3 + \bar{x}_{k+1,n}^3 = O(n^{3\beta-3}),$$

$$\sum_{i=n/2}^k [x_{i+1,n}^2 - x_{in}^2] = x_{k+1,n}^2 = O(n^{2\beta-2}) = O(n^{4\beta-3}),$$

where we used that $\beta \geq 1/2$. We also observe that

$$\begin{aligned} \sum_{i=n/2}^k (x_{in} + x_{i+1,n}) &= 4n^{-1} \sum_{i=n/2}^k (i - n/2) + 2n^{-1} \sum_{i=n/2}^k 1 \\ &= 2n^{-1}(k - n/2)(k + 1 - n/2) + 2n^{-1}(k + 1 - n/2) \\ &= 2n^{-1}(k + 1 - n/2)^2 = 2y_{k+1,n}^2. \end{aligned}$$

Then we find that for $n/2 \leq k \leq n/2 + n^\beta$

$$\begin{aligned} g_n(y_{k+1,n}) - g_n(0) &= -2y_{k+1,n}^2 + O(n^{4\beta-3}), \\ c_n^{-1} e^{2y_{kn}^2} f_n(y_{kn}) &= e^{O(n^{4\beta-3})} = 1 + O(n^{4\beta-3}), \end{aligned} \quad (11)$$

which is (6). After this we focus on determining the behavior of c_n as $n \rightarrow \infty$.

Rewrite (6) as

$$n^{-1/2} c_n^{-1} P(S_n = k) = n^{-1/2} e^{-2y_{kn}^2} + O(n^{4\beta-3.5}) \quad (12)$$

and conclude that

$$\lim_{n \rightarrow \infty} n^{-1/2} c_n^{-1} P(|S_n - n/2| \leq n^{3/5}) = \lim_{n \rightarrow \infty} J_n, \quad (13)$$

where

$$J_n := n^{-1/2} \sum_{k: |y_{kn}| \leq n^{1/10}} e^{-2y_{kn}^2}$$

Here by Chebyshev's inequality

$$1 \geq P(|S_n - n/2| \leq n^{3/5}) = 1 - P(|S_n - n/2| \geq n^{3/5}) \geq 1 - n^{-1/5},$$

implying that

$$\lim_{n \rightarrow \infty} P(|S_n - n/2| \leq n^{3/5}) = 1,$$

which along with (13) yields

$$\lim_{n \rightarrow \infty} n^{-1/2} c_n^{-1} = \lim_{n \rightarrow \infty} J_n. \quad (14)$$

To find $\lim_{n \rightarrow \infty} J_n$ observe that for $k \geq n/2$ we have

$$e^{-2y_{kn}^2} \leq e^{-2y^2} \quad \text{for } 0 \leq y \leq y_{kn}, \quad e^{-2y_{kn}^2} \geq e^{-2y^2} \quad \text{for } y_{k,n} \leq y.$$

Therefore, for $k > n/2$

$$\int_{y_{k-1,n}}^{y_{kn}} e^{-2y^2} dy \geq n^{-1/2} e^{-2y_{kn}^2} \geq \int_{y_{kn}}^{y_{k+1,n}} e^{-2y^2} dy.$$

Since

$$J_n = n^{-1/2} + 2 \sum_{k > n/2: y_{kn} \leq n^{1/10}} e^{-2y_{kn}^2},$$

we infer that

$$\int_0^{\bar{y}_n} e^{-2y^2} dy \geq (J_n - n^{-1/2})/2 \geq \int_0^{\bar{y}_n} e^{-2y^2} dy - \int_0^{n^{-1/2}} e^{-2y^2} dy,$$

where \bar{y}_n is the largest y_{kn} such that $y_{kn} \leq n^{1/10}$. Obviously, $\bar{y}_n \rightarrow \infty$ as $n \rightarrow \infty$. This yields

$$\lim_{n \rightarrow \infty} J_n = 2 \int_0^\infty e^{-2y^2} dy = \sqrt{\pi/2}.$$

Upon combining this with (14) we come to (7).

Now we can replace $n^{-1/2}c_n^{-1}$ in (13) with $\sqrt{\pi/2}$ and by the definition of integral obtain that for any $a, b \in (-\infty, \infty)$ such that $a < b$

$$\begin{aligned} & \sqrt{\pi/2} \lim_{n \rightarrow \infty} P(a\sqrt{n} \leq S_n - n/2 \leq b\sqrt{n}) \\ &= \lim_{n \rightarrow \infty} n^{-1/2} \sum_{k: a \leq y_{kn} \leq b} e^{-2y_{kn}^2} = \int_a^b e^{-2y^2} dy. \end{aligned}$$

This proves (9).

Finally, we can rewrite (6) as

$$(n\pi/2)^{1/2} e^{2y_{kn}^2} f_n(y_{kn}) = e^{O(n^{4\beta-3})} = d_n + O(n^{4\beta-3}),$$

where $d_n = c_n(n\pi/2)^{1/2} \rightarrow 1$. Then (8) follows. The theorem is proved.

3. A REFINEMENT OF THEOREM 1

One can improve (6) and (8). For $\beta = 3/4$ and $y_{kn} \sim n^{1/4}$ equation (6) only says that

$$c_n^{-1} e^{2y_{kn}^2} f_n(y_{kn})$$

is bounded without asserting that it converges to 1 as $n \rightarrow \infty$. The following theorem implies, in particular, that

$$c_n^{-1} e^{2y_{kn}^2} f_n(y_{kn}) \rightarrow e^{-4c^4/3}$$

if $n \rightarrow \infty$, $k = k_n \rightarrow \infty$, and $y_{kn}n^{-1/4} \rightarrow c$. The refinement comes from taking two more terms in Taylor's expansions in (16).

Theorem 2. For $\beta \in [1/2, 1)$ and $|k - n/2| \leq n^\beta$ we have

$$e^{-\alpha n^{6\beta-5}} \leq c_n^{-1} e^{2y_{kn}^2(1+2y_{kn}^2/(3n))} P(S_n = k) \leq e^{\alpha n^{2\beta-2}}, \quad (15)$$

where

$$\alpha = 16/3 + (2^{10} + 2^5)/5.$$

In particular, if $\beta \in [1/2, 5/6)$, then

$$\lim_{n \rightarrow \infty} \max_{|k-n/2| \leq n^\beta} |c_n^{-1} e^{2y_{kn}^2(1+2y_{kn}^2/(3n))} P(S_n = k) - 1| = 0.$$

Furthermore, if $\beta \in [3/4, 5/7)$, then

$$\lim_{n \rightarrow \infty} \sum_{|k-n/2| \leq n^\beta} |c_n^{-1} e^{2y_{kn}^2(1+2y_{kn}^2/(3n))} P(S_n = k) - 1| = 0.$$

Proof. It follows from (3) that

$$\begin{aligned}
g_n(y_{k+1,n}) - g_n(y_{kn}) &= \ln(1 - x_{kn}) - \ln(1 + x_{k+1,n}) \\
&= -x_{kn} - x_{k+1,n} + (1/2)[x_{k+1,n}^2 - x_{kn}^2] - (1/3)[x_{kn}^3 + x_{k+1,n}^3] \\
&\quad + (1/4)[x_{k+1,n}^4 - x_{kn}^4] - (1/5)[\bar{x}_{kn}^5 + \bar{x}_{k+1,n}^5], \tag{16}
\end{aligned}$$

where $\bar{x}_{kn} \leq x_{kn}$, $\bar{x}_{k+1,n} \leq x_{k+1,n}$. For $n/2 \leq k \leq n/2 + n^\beta$ we have

$$\begin{aligned}
0 &\leq y_{kn} \leq n^{\beta-1/2}, \quad 0 \leq x_{kn} \leq 2n^{\beta-1}, \\
0 &\leq y_{k+1,n} \leq n^{\beta-1/2} + n^{-1/2} \leq 2n^{\beta-1/2}, \quad 0 \leq x_{k+1,n} \leq 4n^{\beta-1}, \\
0 &\leq \bar{x}_{kn}^5 + \bar{x}_{k+1,n}^5 \leq (2^{10} + 2^5)n^{5\beta-5}, \\
&\sum_{i=n/2}^k [x_{i+1,n}^2 - x_{in}^2] = x_{k+1,n}^2 \leq 16n^{2\beta-2}.
\end{aligned}$$

We also observe that

$$\begin{aligned}
&\sum_{i=n/2}^k [x_{i+1,n}^4 - x_{in}^4] = x_{k+1,n}^4 \leq 2^8 n^{4\beta-4} \leq 2^8 n^{2\beta-2}, \\
&\sum_{i=n/2}^k [x_{kn}^3 + x_{k+1,n}^3] = 8n^{-3} \sum_{i=0}^{k-n/2} i^3 + 8n^{-3} \sum_{i=1}^{k+1-n/2} i^3 \\
&= 2n^{-3}(k - n/2)^2(k + 1 - n/2)^2 + 2n^{-3}(k + 1 - n/2)^2(k + 2 - n/2)^2 \\
&= 4n^{-1}y_{k+1,n}^2[y_{k+1,n}^2 + n^{-1}] = 4n^{-1}y_{k+1,n}^4 + 4n^{-2}y_{k+1,n}^2.
\end{aligned}$$

As a consequence

$$4n^{-1}y_{k+1,n}^4 \leq \sum_{i=n/2}^k [x_{kn}^3 + x_{k+1,n}^3] \leq 4n^{-1}y_{k+1,n}^4 + 16n^{2\beta-3}.$$

Now upon summing up (16) with respect to k , we find that for $n/2 \leq k \leq n/2 + n^\beta$

$$\begin{aligned}
g_n(y_{k+1,n}) - g_n(0) &\leq -2y_{k+1,n}^2 - (4/3)n^{-1}y_{k+1,n}^4 + 72n^{2(\beta-1)}, \\
g_n(y_{k+1,n}) - g_n(0) &\geq -2y_{k+1,n}^2 - (4/3)n^{-1}y_{k+1,n}^4 \\
&\quad - n^{2\beta-3}16/3 - n^{6\beta-5}(2^{10} + 2^5)/5 \\
&\geq -2y_{k+1,n}^2 - (4/3)n^{-1}y_{k+1,n}^4 - \alpha n^{6\beta-5},
\end{aligned}$$

where we used that $\beta \geq 1/2$. The above inequalities yield (15), that easily implies all other assertions of the theorem. The theorem is proved.