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

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

Then comes the observation that $ES_n = n/2$, $\operatorname{Var} S_n = n/4$. 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 $n \to \infty$. In other words, one may hope that for any $a < b$

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

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

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

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

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

$$f_n(y_{kn})=f_n(y_{kn}+n^{-1/2})\frac{y_{kn}\sqrt{n}+n/2+1}{n/2-y_{kn}\sqrt{n}},$$

$$f_n(y_{kn}+n^{-1/2})=f_n(y_{kn})\frac{n/2-y_{kn}\sqrt{n}}{y_{kn}\sqrt{n}+n/2+1}, \tag{3}$$

$$f_n(y_{kn}+n^{-1/2})-f_n(y_{kn})=f_n(y_{kn})\big(\frac{n/2-y_{kn}\sqrt{n}}{y_{kn}\sqrt{n}+n/2+1}-1\big)$$

$$=-f_n(y_{kn})\frac{2y_{kn}\sqrt{n}+1}{y_{kn}\sqrt{n}+n/2+1}. \tag{4}$$

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 $1/\sqrt{n}$. One way or another, let us believe that, for a function ϕ,

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

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

For ϕ_n equation (4) becomes

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

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 $n\to\infty$ 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 resulat

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 $|y_{kn}| \leq n^{\beta-1/2}$)*

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

where

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

Furthermore,

$$\lim_{n\to\infty} c_n \sqrt{n} = \sqrt{2/\pi}, \tag{7}$$

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

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

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

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

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

$$g_n(y_{kn}) = \ln f_n(y_{kn}),$$

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

and use (3) to get

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

where $0 \leq \bar{x}_{kn} \leq x_{kn}, 0 \leq \bar{x}_{k+1,n} \leq x_{k+1,n}$. Since $x_{kn} \leq 2n^{\beta-1}$ and $2n^{\beta-1} \leq 1/2$ if n is large, the last bracket in (16) is less than or equal to $2 \cdot 8 \cdot 2^3 n^{3\beta-3}$. The sum of these brackets over $k = n/2, ..., n/2 + n^\beta$ is therefore less than or equal to $128n^{4\beta-3}$. Also for $n/2 \leq k \leq n/2 + n^\beta$ we have

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

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

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

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

$$g_n(y_{k+1,n}) - g_n(0) = -2y_{k+1,n}^2 + z_{kn}n^{2\beta-2},$$

where $|z_{kn}| \leq 132$. This yields

$$c_n^{-1}e^{2y_{kn}^2}f_n(y_{kn}) = e^{z_{kn}n^{4\beta-3}} = 1 + \bar{z}_{kn}n^{4\beta-3}, \tag{11}$$

where $|\bar{z}_{kn}|$ are bounded by a constant (e^{132}) independent of k, n. This proves (6). After this we focus on determining the behavior of c_n as $n \to \infty$.

Rewrite (11) as

$$n^{-1/2}c_n^{-1}P(S_n = k) = n^{-1/2}e^{-2y_{kn}^2} + n^{-1/2}e^{-2y_{kn}^2}\bar{z}_{kn}n^{4\beta-3} \tag{12}$$

and using that $|\bar{z}_{kn}|$ are bounded and $n^{4\beta-3} \to 0$ conclude that

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

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

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

which along with (13) yields

$$\lim_{n\to\infty} n^{-1/2}c_n^{-1} = \lim_{n\to\infty} J_n. \tag{14}$$

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

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

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

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

we infer that

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

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

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

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 $a < b$

$$\sqrt{\pi/2}\lim_{n\to\infty} P(a\sqrt{n} \le S_n - n/2 \le b\sqrt{n})$$

$$= \lim_{n\to\infty} n^{-1/2} \sum_{k:a\le y_{kn}\le b} e^{-2y_{kn}^2} = \int_a^b e^{-2y^2}\,dy.$$

This proves (9).

Finally, we can rewrite (11) as

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

where $d_n = c_n(n\pi/2)^{1/2} \to 1$. 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 $n \to \infty$. The following theorem implies, in particular, that

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

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

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

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

where $\alpha > 0$ is a constant.

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

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

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

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

Proof. It follows from (3) that

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

where $0\leq\bar{x}_{kn}\leq x_{kn}, 0\leq\bar{x}_{k+1,n}\leq x_{k+1,n}$. For $n/2\leq k\leq n/2+n^{\beta}$ we have

$$0\leq y_{kn}\leq n^{\beta-1/2},\quad 0\leq x_{kn}\leq 2n^{\beta-1},$$
$$0\leq y_{k+1,n}\leq n^{\beta-1/2}+n^{-1/2}\leq 2n^{\beta-1/2},\quad 0\leq x_{k+1,n}\leq 4n^{\beta-1},$$

and if n is large enough

$$0\leq(1-\bar{x}_{kn})^{-5}x_{kn}^5+(1-\bar{x}_{k+1,n})^{-5}x_{k+1,n}^5\leq 2\cdot 2^5\cdot 4^5n^{5\beta-5}=:Nn^{5\beta-5}$$

and the sum of those terms as k runs between $n/2$ and $n/2+n^{\beta}$ is dominated by $Nn^{6\beta-5}$. Next,

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

We also observe that

$$\sum_{i=n/2}^{k}[x_{i+1,n}^4-x_{in}^4]=x_{k+1,n}^4\leq 2^8n^{4\beta-4}\leq 2^8n^{2\beta-2},$$

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

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 $n/2 \leq k \leq n/2 + n^{\beta}$

$$g_n(y_{k+1,n}) - g_n(0) \leq -2y_{k+1,n}^2 - (4/3)n^{-1}y_{k+1,n}^4 + 72n^{2(\beta-1)},$$

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

where we used that $\beta \geq 1/2$ and $\alpha = N/5 + 16/3$. The above inequalities yield (15), that easily implies all other assertions of the theorem. The theorem is proved.