Notes on the DeMoivre-Laplace Theorem

October 22, 2018

Abstract

These are some notes on the DeMoivre-Laplace Theorem.

0.1 The DeMoivre-Laplace Theorem

Let $\{T_j\}_{j\in\mathbb{N}}$ be an infinite sequence of independent Bernoulli random variables, each with success parameter p. Let $q = 1 - p$. Then for each j,

$$\mathrm{E}(T_j) = p \quad \text{and} \quad \mathrm{Var}(T_j) = pq \ . \tag{0.1}$$

Think of an infinite sequence of coin tosses with a coin that is weighted so that the probability of heads is p. We assume $0 < p < 1$ to avoid trivialities.

Define S_n to be the cumulative number of successes at the n stage:

$$S_n = \sum_{j=1}^{n} T_j \ .$$

It then follows from (0.1) and the independence of the Bernoulli variables, that

$$\mathrm{E}(S_n) = np \quad \text{and} \quad \mathrm{Var}(S_n) = npq \ . \tag{0.2}$$

For each $n \in \mathbb{N}$, define the random variable

$$X_n = \frac{S_n - np}{\sqrt{npq}} \ .$$

It follows that $\mathrm{E}(X_n) = \dfrac{1}{\sqrt{npq}}\mathrm{E}(S_n - np) = \dfrac{1}{\sqrt{npq}}(np - np) = 0$, and then, using (0.2),

$$\mathrm{Var}(X_n) = \frac{1}{npq}\mathrm{Var}(S_n - np) = \frac{1}{npq}\mathrm{Var}(S_n) = 1 \ .$$

The random variables X_n depend on n in a non-trivial way: The maximum value of X_n is

$$\frac{n - np}{\sqrt{npq}} = \sqrt{\frac{nq}{p}} \ ,$$

which certainly depends on n. However, the expected value of X_n and the variance of X_n are independent of n: We have $\mathrm{E}(X_n) = 0$ and $\mathrm{Var}(X_n) = 1$ for all n. It turns out that other important probabilistic characteristics of X_n are essentially independent of n for large values of n – and by large we mean a couple hundred or less, depending on p and q. We are *not* talking about astronomically large numbers.

Specifically, the cumulative distribution function of X_n is nearly independent of n for large n, and in fact it is nearly equal to the cumulative distribution function of the standard normal distribution:

0.1 DEFINITION. The function $\Phi(a)$ given by

$$\Phi(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \mathrm{d}x \ , \tag{0.3}$$

is the cumulative distribution function of the standard normal density $f(x) = \sqrt{2\pi} e^{-x^2/2}$. We always reserve the symbol Φ for this function.

The function Φ cannot be computed in closed form, but its values are tabulated in many references; e.g. the table on page 196 in the text.

It follows from the definition that if X is a standard normal random variable, $P(X > b) = 1 - \Phi(b)$. In particular, from the table we see that

$$P(X > 1) = 0.1587 \ , \quad P(X > 2) = 0.0228 \ , \quad P(X > 3) = 0.0013 \ .$$

Thus, if X is a standard normal variable, it can take on arbitrarily large values, *however*, the probability that it takes on large values is very small. Even the probability that $X \geq 3$ is barely better than 1 in a thousand. Larger values are far less likely: $P(X \geq 10) \leq 7.9199 \times 10^{-24}$. The standard normal distribution is fundamentally important on account of the following theorem:

0.2 THEOREM (DeMoivre-Laplace Theorem)**.** *Let X_n defined as above in terms of an infinite sequence of independent identically distributed Bernoulli variables. Then for all $a < b \in \mathbb{R}$,*

$$\lim_{n\to\infty} P(a < X_n \leq b) = \Phi(b) - \Phi(a) \ . \tag{0.4}$$

Since Φ is continuous, for any a, and any $\epsilon > 0$, we can find an δ so that $\Phi(a+\delta) - \Phi(a-\delta) < \epsilon$. It follows that for all sufficiently large n, $P(X_n = a) \leq P(a - \delta < X_n \leq a + \delta) < \epsilon$, and therefore, $\lim_{n\to\infty} P(X_n = a) = 0$ for all a. Therefore, we can also write (0.4) as

$$\lim_{n\to\infty} P(a \leq X_n \leq b) = \Phi(b) - \Phi(a) \ . \tag{0.5}$$

0.2 Applying the DeMoivre-Laplace Theorem

As our first example, we apply the DeMoivre-Laplace Theorem to coin tossing. Suppose a fair coin is tossed $2m$ times. What is the probability that the result is exactly m heads?

Let $T_j = 1$ if the jth toss is heads, and $T_j = 0$ otherwise. Then S_{2m} is the total number of heads, and by the DeMoivre-Laplace Theorem, the cumulative distribution function of

$$X_{2m} = \frac{S_{2m} - m}{\sqrt{m/2}}$$

is approximately equal to Φ. The possible values of X_{2m} are the numbers

$$x_{n,k} = \frac{k - m}{\sqrt{m/2}} \ .$$

Note that $S_{2m} = k$ if and only if $X_{2m} = x_{n,k}$. Hence

$$k - \tfrac{1}{2} \le S_{2m} \le k + \tfrac{1}{2}$$

if and only if

$$x_{n,k} - \frac{1}{\sqrt{2m}} \le X_{2m} \le x_{n,k} + \frac{1}{\sqrt{2m}} \ .$$

Now take $k = m$. Identifying the event $S_{2m} = m$ with the event $m - \frac{1}{2} \le S_{2m} \le m + \frac{1}{2}$, which is known as the *continuum correction*, we have

$$P(S_{2m} = m) = P\left(-\frac{1}{\sqrt{2m}} \le X_{2m} \le \frac{1}{\sqrt{2m}}\right) \ .$$

The DeMoivre-Laplace Theorem says that for *fixed* $a < b$,

$$\lim_{m\to\infty} P(a \le X_{2m} \le b) = \Phi(b) - \Phi(a) \ .$$

In our case, a and b depend on m, but let us make the approximation suggested by the DeMoivre-Laplace Theorem anyway:

$$P(S_{2m} = m) = P\left(-\frac{1}{\sqrt{2m}} \le X_{2m} \le \frac{1}{\sqrt{2m}}\right) \approx \Phi\left(\frac{1}{\sqrt{2m}}\right) - \Phi\left(-\frac{1}{\sqrt{2m}}\right) \ . \tag{0.6}$$

Let's put in some numbers and see how we did. Taking $m = 20$,

$$\Phi\left(\frac{1}{\sqrt{40}}\right) - \Phi\left(-\frac{1}{\sqrt{40}}\right) = 0.12563...$$

The exact value is

$$P(S_{40} = 20) = \binom{40}{20} 2^{-40} = 0.12537...$$

This is really quite good and certainly m is not that large.

Now let's consider a problem in which the exact answer is not so easy to numerically evaluate. Suppose we are given what is claimed is a fair coin, and we toss it !0,000 times and obtain $5,050$ heads. How likely is it that if the coin is fair, $10,000$ tosses result in at least $5,050$ heads?

The exact answer is

$$2^{-10{,}000}\sum_{k=5{,}050}^{10{,}000}\binom{10,000}{k}\ ,$$

but just try to evaluate this on a calculator. Identifying the even $S_{10{,}000} \geq 5,050$ with the event

$$X_{10{,}000} \geq \frac{5,050 - 5,000}{\sqrt{2,500}} = 1\ ,$$

The DeMoivre-Laplace Theorem says that

$$P(X_{10{,}000} \geq 1) \approx \Phi(1) = 0.1587...\ .$$

It is not really very likely that the coin is fair.

0.3 Proof of the DeMoivre-Laplace Theorem and more

We will now prove the DeMoivre-Laplace Theorem and somewhat more. In doing so, we will see why it gives good results even for values of n that are not so large, and why it was justified to make the approximation that we made in (0.6), in which the endpoints on the interval also depend on n.

The basis of the proof is Stirling's formula, which is essentially due to DeMoivre. Stirling's formula says that

$$n! \approx \sqrt{2\pi}n^{n+1/2}e^{-n}\ .$$

Stirlings's contribution was to identify the constant $\sqrt{2\pi}$. DeMoivre had the same formula apart from the constant, for which he had only a good decimal approximation, which was good enough for his purposes. The fact that this constant is exactly $\sqrt{2\pi}$ could even be deduced from the DeMoivre-Laplace Theorem, as we shall see.

The random variable X_n takes on the values

$$x_{n,k} := \frac{k - np}{\sqrt{npq}} \qquad \text{for} \qquad o \leq k \leq n\ . \tag{0.7}$$

These values are evenly spaced with the distance between them being

$$\Delta_n x := \frac{1}{\sqrt{npq}}\ . \tag{0.8}$$

Since $P(S_n = k) = \binom{n}{k}p^k q^{n-k}$,

$$P(Z_n = x_{n,k}) = \binom{n}{k}p^k q^{n-k}\ . \tag{0.9}$$

We are going to prove:

0.3 THEOREM. *For any $R > 0$, and any $\epsilon > 0$, there is an N_0 depending only of R, ϵ and p such that for all $n \geq N_0$ and all k such that $|x_{n,k}| \leq R$,*

$$(1-\epsilon)f(x_{n,k})\Delta_n x \ \leq \ P(X_n = x_{n,k}) \ \leq \ (1+\epsilon)f(x_{n,k})\Delta_n x \ . \tag{0.10}$$

where

$$f(x) := \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \ .$$

Theorem 0.3, known as the *local DeMoivre-Laplace Theorem* implies the DeMoivre-Laplace Theorem as stated in Theorem 0.2. Before turning to the proof of Theorem 0.3, we first give another useful corollary that we will use to deduce the Theorem 0.2 from Theorem 0.3, among other things.

0.4 COROLLARY. *With X_n defined as above, for all piecewise continuous functions g on $\mathbb{R}$ such that for some finite C, $|g(x)| \leq C$ for all x,*

$$\lim_{n\to\infty} \mathrm{E}g(X_n) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} g(x)e^{-x^2/2}\mathrm{d}x \ . \tag{0.11}$$

To see that the corollary implies implies the DeMoivre-Laplace Theorem, take

$$g(x) = \begin{cases} 1 & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \ , \tag{0.12}$$

which is bounded and piecewise continuous, The corollary says that

$$\lim_{n\to\infty} P(a \leq X_n \leq b) = \lim_{n\to\infty} \mathrm{E}g(X_n) = \int_{\mathbb{R}} g(x)f(x)\mathrm{d}x = \int_a^b f(x)\mathrm{d}x = \Phi(b) - \Phi(a) \ ,$$

which is (0.5). We first prove Corollary 0.4, and then Theorem 0.3.

Proof of Corollary 0.4. Since for each n, $\mathrm{E}X_n^2 = 1$,

$$P(|X_n| > R) \leq \mathrm{E}(R^{-2}X_n^2) = R^{-2} \ , \tag{0.13}$$

where we used the fact that the random variable $R^{-2}X_n^2$ is at least 1 on the event $\{|X_n| > R\}$, as in the proof of the law of large numbers. Pick $\epsilon > 0$, and then pick any R large enough that $CR^{-2} < \epsilon$.

By the definition of the expectation,

$$\begin{aligned} \mathrm{E}g(X_n) &= \sum_{k=0}^{n} g(x_{n,k})P(X_n = x_{n,k}) \\ &= \sum_{k\ :\ |x_{n,k}|\leq R} g(x_{n,k})P(X_n = x_{n,k}) + \sum_{k\ :\ |x_{n,k}|>R} g(x_{n,k})P(X_n = x_{n,k}) \ . \end{aligned}$$

By what we have noted above,

$$\begin{aligned}\left|\sum_{k\,:\,|x_{n,k}|>R} g(x_{n,k})P(X_n = x_{n,k})\right| &\leq C \sum_{k\,:\,|x_{n,k}|>R} P(X_n = x_{n,k}) \\ &= CP(|X_n| > R) \leq CR^{-2} \leq \epsilon \ .\end{aligned}$$

Next, by Theorem 0.3, there is an N_0 such that for all $n \geq N_0$ and all k such that $|x_{n,k}| \leq R$,

$$\begin{aligned}|g(x_{n,k})P(X_n = x_{n,k}) - g(x_{n,k})f(x_{n,k})\Delta_n x| &= |g(x_{n,k})||P(X_n = x_{n,k}) - f(x_{n,k})\Delta_n x| \\ &\leq C\epsilon f(x_{n,k})\Delta_n x \ .\end{aligned}$$

Therefore,

$$\sum_{k\,:\,|x_{n,k}|\leq R} g(x_{n,k})P(X_n = x_{n,k}) = \sum_{k\,:\,|x_{n,k}|\leq R} g(x_{n,k})f(x_{n,k})\Delta_n x \pm C\epsilon(1-\epsilon)^{-1}$$

where we have used the fact that

$$\sum_{k\,:\,|x_{n,k}|\leq R} f(x_{n,k})\Delta_n x \leq (1-\epsilon)^{-1} \sum_{k\,:\,|x_{n,k}|\leq R} P(X_n = x_{n,k}) \leq 1 \ ,$$

which is again a consequence of Theorem 0.3. Altogether,

$$\mathrm{E}g(X_n) = \sum_{k\,:\,|x_{n,k}|\leq R} g(x_{n,k})f(x_{n,k})\Delta_n x \pm \epsilon(C(1-\epsilon)^{-1} + 1) \ .$$

Then since

$$\lim_{n\to\infty} \sum_{k\,:\,|x_{n,k}|\leq R} g(x_{n,k})f(x_{n,k})\Delta_n x = \int_{-R}^{R} g(x)f(x)e^{-x^2/2}\mathrm{d}x \ , \tag{0.14}$$

which is true since g is bounded and piecewise continuous,

$$\lim_{n\to\infty} \mathrm{E}g(X_n) = \int_{-R}^{R} g(x)e^{-x^2/2}\mathrm{d}x \pm \epsilon(C(1-\epsilon)^{-1} + 1) \ .$$

Finally,

$$\begin{aligned}\left|\int_{-R}^{R} g(x)e^{-x^2/2}\mathrm{d}x - \int_{\mathbb{R}} g(x)e^{-x^2/2}\mathrm{d}x\right| &\leq \sqrt{\frac{2}{\pi}}C\int_{R}^{\infty} e^{-x^2/2}\mathrm{d}x \\ &\leq C\int_{R}^{\infty} \frac{x}{R}e^{-x^2/2}\mathrm{d}x = C\frac{e^{-R^2/2}}{R} \ .\end{aligned}$$

It is readily checked that $Re^{-R^2/2} \leq 1$ for all R, and so the final bound is no more than $CR^{-2} \leq \epsilon$. Then we have

$$\lim_{n\to\infty} \mathrm{E}g(X_n) = \int_{\mathbb{R}} g(x)e^{-x^2/2}\mathrm{d}x \pm \epsilon(C(1-\epsilon)^{-1} + 2) \ .$$

Since ϵ is arbitrary this proves the claim. □

0.5 Remark. The estimate

$$\int_R^\infty e^{-x^2/2}\mathrm{d}x \le \frac{1}{R}\int_R^\infty xe^{-x^2/2}\mathrm{d}x = \frac{e^{-R^2/2}}{R}$$

is often useful.

0.6 Remark. Looking back over the error terms in the proof of Corollary 0.4, it can be seen that the rate of convergence in (0.11) depends on g only through C and the rate of convergence in (0.14). In particular, if g is Lipschizt with Lipschitz constant L, then the rate of convergence in (0.14) can be controlled using L and p. (Note that $\Delta_n x$ depends on n and p.) Hence the rate of convergence in (0.11) is uniform over the class of all bounded, Lipschitz continuous g with bound C and Lipschitz constant L. Taking Lipschitz approximations to the functions g defined in (0.14), one can then easily prove that the rate of convergence in (0.5) is uniform in a and b.

Proof of Theorem 0.3. A more precise version of Stirling's formula is:

$$\sqrt{2\pi}n^{n+1/2}e^{-n}e^{\frac{1}{12n+1}} \le n! \le \sqrt{2\pi}n^{n+1/2}e^{-n}e^{\frac{1}{12n}} \ .$$

Taking logarithms, it follows that

$$\left|\log n! - \frac{1}{2}\log(2\pi n) - n\log n + n\right| \le \frac{1}{12n} \ .$$

For n a natural number and $0 < k < n$ an integer, we compute

$$\begin{aligned}
\log\binom{n}{k} &= \log n! - \log k! - \log(n-k)! \\
&\approx \frac{1}{2}\log(2\pi) + \left(n+\frac{1}{2}\right)\log n - \left(k+\frac{1}{2}\right)\log k - \left(n-k+\frac{1}{2}\right)\log(n-k) \\
&= \frac{1}{2}\log\left(\frac{2\pi}{n}\right) - \left(k+\frac{1}{2}\right)\log\left(\frac{k}{n}\right) - \left(n-k+\frac{1}{2}\right)\log\left(\frac{n-k}{n}\right)
\end{aligned}$$

where we have used

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

to obtain the last line. Therefore

$$\log\left(\binom{n}{k}p^k q^{n-k}\right) \approx$$

$$\frac{1}{2}\log\left(\frac{2\pi}{npq}\right) - \left(k+\frac{1}{2}\right)\log\left(\frac{k}{np}\right) - \left(n-k+\frac{1}{2}\right)\log\left(\frac{n-k}{nq}\right) \ . \tag{0.15}$$

Note that the error made in (0.15) is no greater that

$$\frac{1}{12}\left(\frac{1}{n}+\frac{1}{k}+\frac{1}{n-k}\right) \tag{0.16}$$

in magnitude.

We now rewrite (0.15) in terms of $x_{n,k}$. From the definition (0.7),

$$k = np + x_{n,k}\sqrt{npq} \qquad \text{and} \qquad n - k = nq - x_{n,k}\sqrt{npq}\ , \tag{0.17}$$

and then

$$\frac{k}{np} = 1 + x_{n,k}\sqrt{\frac{q}{np}} \qquad \text{and} \qquad \frac{n-k}{nq} = 1 - x_{n,k}\sqrt{\frac{p}{nq}}\ . \tag{0.18}$$

Let N_0 be such that

$$\max\left\{R\sqrt{\frac{q}{N_0 p}}\ ,\ R\sqrt{\frac{p}{N_0 q}}\right\} \le \frac{1}{2}\ . \tag{0.19}$$

Then for $n \ge N_0$, $k \ge np/2$ and $n - k \ge nq/2$, and so the quantity in (0.16) is no more than

$$\frac{1}{12}\left(\frac{1}{n} + \frac{2}{np} + \frac{2}{nq}\right)\ ,$$

and this goes to zero as n increases to infinity. This bounds the size of the error made in (0.15) uniformly, as required.

Having dealt with estimating the error, we return to the right side of (0.15) and apply the Taylor expansion for the natural logarithm, which is

$$\log(1+t) = t - \tfrac{1}{2}t^2 + \tfrac{1}{3}t^3 + \dots\ ,$$

which converges for $|t| < 1$, and is an alternating decreasing series, meaning that

$$\left|\log(1+t) - t + \tfrac{1}{2}t^2\right| \le \tfrac{1}{3}|t|^3\ . \tag{0.20}$$

which allows up to control the errors in the approximation $\log(1+t) \approx t - \frac{1}{2}t^2$. Making this approximation, and using the first formula in (0.17),

$$\begin{aligned}\left(k+\frac{1}{2}\right)\log\left(\frac{k}{np}\right) &\approx \left(np + x_{n,k}\sqrt{npq} + \frac{1}{2}\right)\left(x_{n,k}\sqrt{\frac{q}{np}} - \frac{1}{2}x_{n,k}^2\frac{q}{np}\right)\\ &= x_{n,k}\sqrt{npq} + \frac{1}{2}qx_{n,k}^2 + \left(\frac{1}{2}x_{n,k} - x_{n,k}^3 q\right)\sqrt{\frac{q}{np}} - \frac{1}{4}x_{n,k}^2\frac{q}{np}\end{aligned}$$

By (0.20), the error we have made here is bounded in magnitude by

$$\frac{1}{3}R^3\left(\sqrt{\frac{p}{nq}}\right)^3$$

since we apply (0.20) with $t = x_{n,k}\sqrt{\frac{q}{np}}$, and for $n \ge N_0$, $|t| < \frac{1}{2}$ by (0.19).

Likewise, using the second formula in (0.17),

$$\begin{aligned}\left(n-k+\frac{1}{2}\right)\log\left(\frac{n-k}{nq}\right) &\approx \left(nq - x_{n,k}\sqrt{npq} + \frac{1}{2}\right)\left(-x_{n,k}\sqrt{\frac{p}{nq}} - \frac{1}{2}x_{n,k}^2\frac{p}{nq}\right)\\ &= -x_{n,k}\sqrt{npq} + \frac{1}{2}px_{n,k}^2 + \left(\frac{1}{2}x_{n,k} - x_{n,k}^3 q\right)\sqrt{\frac{p}{nq}} - \frac{1}{4}x_{n,k}^2\frac{p}{nq}\end{aligned}$$

and the analysis of the error is the same as we saw just above, only with the roles of p and q interchanged, so we do not repeat it.

Using these two expressions in (0.15), we see that for all $n \geq N_0$ and k such that $|x_{n,k}| \leq R$,

$$\log\left(\binom{n}{k}p^k q^{n-k}\right) = \frac{1}{2}\log\left(\frac{2\pi}{npq}\right) - \frac{1}{2}x_{n,k}^2 \pm \frac{C}{\sqrt{n}} ,$$

where the constant C depends only on R and p.

Now exponentiating,

$$\left[\frac{1}{\sqrt{2\pi}}e^{-x_{n,k}^2/2}\Delta_n x\right] e^{-C/\sqrt{n}} \leq \binom{n}{k}p^k q^{n-k} \leq \left[\frac{1}{\sqrt{2\pi}}e^{-x_{n,k}^2/2}\Delta_n x\right] e^{C/\sqrt{n}} \tag{0.21}$$

Now for any $\epsilon > 0$ choose n_0 such that $e^{C/\sqrt{n_0}} \leq 1+\epsilon$. Then $e^{-C/\sqrt{n_0}} \geq 1-\epsilon$. Replacing N_0 with the maximum of n_0 and the previously chosen N_0, we have completed the proof. □

0.7 Remark. The constant $\sqrt{2\pi}$ in Stirling's formula shows up as the constant $\sqrt{2\pi}$ in the standard normal density function. If we only knew that

$$\left|\log n! - \frac{1}{2}\log(Cn) - n\log n + n\right| \leq \frac{K}{n}$$

for some constants C and K, the exact same reasoning we have just gone through would show that for all $R >$,

$$\lim_{n\to\infty} P(-R \leq X_n \leq R) = \frac{1}{C}\int_{-R}^{R} e^{-x^2/2}\mathrm{d}x .$$

Since $\mathrm{E}(X_n^2) = 1$ for each n, $P(|X_n| > R) \leq= \mathbb{R}^{-2}$, as in the proof of Corollary0.4. It follows that for all n,

$$1 - R^{-2} \leq P(-R \leq X_n \leq R) \leq 1 .$$

It then follows that for all R,

$$1 - R^{-2} \leq \frac{1}{C}\int_{-R}^{R} e^{-x^2/2}\mathrm{d}x \leq 1 .$$

Taking $R \to \infty$ yields $1 \leq \frac{1}{C}\sqrt{2\pi} \leq 1$, and this shows that $C = \sqrt{2\pi}$, Stirling's contribution to Stirling's formula.

0.4 Universality

A remarkable and important aspect of Corollary 0.4 may be hidden by our compact notation. Recall that if $\{T_j\}$ is a sequence of independent Bernoulli random variables with success parameter p, $0 < p < 1$, and $q = 1-p$,

$$X_n = \sum_{j=1}^{n}\frac{T_j - p}{\sqrt{npq}} = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{T_j - p}{\sqrt{pq}} ,$$

and so the corollary says that, for appropriate g,

$$\lim_{n\to\infty} \mathrm{E}g\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{T_j - p}{\sqrt{pq}}\right) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} g(x)e^{-x^2/2}\mathrm{d}x\ .$$

The success parameter appears on the left, but not on the right. The random variables

$$\widetilde{T}_j := \frac{T_j - p}{\sqrt{pq}} \tag{0.22}$$

have distributions that depend on p, though they have the same mean, 0, and the same variance, 1. The limiting behavior of the distribution of the normalized sums

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{T_j - p}{\sqrt{pq}} = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\widetilde{T}_j$$

is *universal*; it does not matter what the success parameter is.

In fact, this is only a small hint of the scope of this universality: It really does not matter at all that the underlying random variables T_j are Bernoulli variables. All that matters is the mean and the variance. However, to keep the proof of the next lemma from getting too technical, we will make some mild additional assumptions on g and on the random variables $\widetilde{T}_j$. Note that we are *not* assuming that the random variables $\widetilde{T}_j$ in the lemma are constructed out of Bernoulli variables as in (0.22).

0.8 LEMMA. *Let g be a function with three continuous derivatives such that for some finite constant C,*

$$\max\{g(x), g'(x), g''(x), g'''(x)\} \le C$$

for all x. Let $\{\widetilde{T}_j\}_{j\in\mathbb{N}}$ be any sequence of independent, identically distributed random variables such that $\mathrm{E}(\widetilde{T}_j) = 0$ and $\mathrm{Var}(\widetilde{T}_j) = 1$. Suppose further that for some $K < \infty$, $\mathrm{E}|\widetilde{T}_j|^3 = K$. Then

$$\lim_{n\to\infty} \mathrm{E}g\left(\sum_{j=1}^{n}\frac{\widetilde{T}_j}{\sqrt{n}}\right) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} g(x)e^{-x^2/2}\mathrm{d}x\ .$$

Proof. Let $\{T_j\}_{j\in\mathbb{N}}$ be an infinite independent sequence of such Bernoulli variables. Define $\widetilde{S}_n = \sum_{j=1}^{n}\widetilde{T}_j$, and put $S_n = \sum_{j=1}^{n} T_j$ as before. For each $k = 0, \ldots, n$, define

$$W_{n,k} := \sum_{j=0}^{k} 2(T_j - \tfrac{1}{2}) + \sum_{j=k+1}^{n}\widetilde{T}_j$$

Then $W_{n,0} = \widetilde{S}_n$ and $W_{n,n} = S_n - \frac{1}{2}n$. Therefore, we have the telescoping sum

$$g\left(\frac{2(S_n - \frac{1}{2}n)}{\sqrt{n}}\right) - g\left(\frac{\widetilde{S}_n}{\sqrt{n}}\right) = g\left(\frac{W_{n,n}}{\sqrt{n}}\right) - g\left(\frac{W_{n,0}}{\sqrt{n}}\right) = \sum_{k=0}^{n-1}\left[g\left(\frac{W_{n,k+1}}{\sqrt{n}}\right) - g\left(\frac{W_{n,k}}{\sqrt{n}}\right)\right]\ .$$

By linearity of the expectation,

$$\mathrm{E}g\left(\frac{2(S_n-\frac{1}{2}n)}{\sqrt{n}}\right)-\mathrm{E}g\left(\frac{\widetilde{S}_n}{\sqrt{n}}\right)=\sum_{k=0}^{n-1}\mathrm{E}\left[g\left(\frac{W_{n,k+1}}{\sqrt{n}}\right)-g\left(\frac{W_{n,k}}{\sqrt{n}}\right)\right]\ .$$

The sums defining $W_{n,k+1}$ and $W_{n,k+1}$ differ only in the $(k+1)$st term. Define

$$U_{n,k}=\sum_{j=0}^{k}2(T_j-\tfrac{1}{2})+\sum_{j=k+2}^{n}\widetilde{T}_j$$

Then

$$W_{n,k+1}=U_{n,k}+2(T_{k+1}-\tfrac{1}{2})\quad\text{and}\quad W_{n,k}=U_{n,k}+\widetilde{T}_{k+1}\ .$$

Therefore, by Taylor's Theorem,

$$\begin{aligned}g\left(\frac{W_{n,k+1}}{\sqrt{n}}\right)=g\left(\frac{U_{n,k}}{\sqrt{n}}\right)\ &+\ g'\left(\frac{U_{n,k}}{\sqrt{n}}\right)\frac{2(T_{k+1}-\frac{1}{2})}{\sqrt{n}}\\ &+\ \frac{1}{2}g''\left(\frac{U_{n,k}}{\sqrt{n}}\right)\left(\frac{2(T_{k+1}-\frac{1}{2})}{\sqrt{n}}\right)^2\\ &\pm\ C\left(\frac{2|T_{k+1}-\frac{1}{2}|}{\sqrt{n}}\right)^3\ .\end{aligned}$$

Taking the expectation, since $\mathrm{E}2(T_{k+1}-\frac{1}{2})=0$, $\mathrm{E}(2(T_{k+1}-\frac{1}{2}))^2=1$ and $\mathrm{E}|2(T_{k+1}-\frac{1}{2})|^3=1$, and using the independence of T_{k+1} and $U_{n,k}$, we have

$$\mathrm{E}g\left(\frac{W_{n,k+1}}{\sqrt{n}}\right)=\mathrm{E}g\left(\frac{U_{n,k}}{\sqrt{n}}\right)+\frac{1}{2}\mathrm{E}g''\left(\frac{U_{n,k}}{\sqrt{n}}\right)\frac{1}{n}\pm\frac{C}{n^{3/2}}\ .$$

Since we also have $\mathrm{E}(\widetilde{T}_{k+1})=0$, $\mathrm{E}(\widetilde{T}_{k+1})^2=1$ and $\mathrm{E}|\widetilde{T}_{k+1}|^2=K<\infty$, we have by the exact same reasoning that

$$\mathrm{E}g\left(\frac{W_{n,k}}{\sqrt{n}}\right)=\mathrm{E}g\left(\frac{U_{n,k}}{\sqrt{n}}\right)+\frac{1}{2}\mathrm{E}g''\left(\frac{U_{n,k}}{\sqrt{n}}\right)\frac{1}{n}\pm\frac{CK}{n^{3/2}}\ .$$

This shows that

$$\left|\mathrm{E}g\left(\frac{W_{n,k+1}}{\sqrt{n}}\right)-\mathrm{E}g\left(\frac{W_{n,k}}{\sqrt{n}}\right)\right|\le\frac{C(K+1)}{n^{3/2}}\ ,$$

and hence

$$\left|\mathrm{E}g\left(\frac{2(S_n-\frac{1}{2}n)}{\sqrt{n}}\right)-\mathrm{E}g\left(\frac{\widetilde{S}_n}{\sqrt{n}}\right)\right|\le\frac{C(K+1)}{n^{3/2}}\ .$$

Since we know that

$$\lim_{n\to\infty}\mathrm{E}g\left(\frac{2(S_n-\frac{1}{2}n)}{\sqrt{n}}\right)=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}g(x)e^{-x^2/2}\mathrm{d}x\ ,$$

this proves that

$$\lim_{n\to\infty}\mathrm{E}g\left(\frac{\widetilde{S}_n}{\sqrt{n}}\right)=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}g(x)e^{-x^2/2}\mathrm{d}x\ .$$

□

0.9 THEOREM. *Let* $\{Y_j\}_{j\in\mathbb{N}}$ *be a sequence of independent identically distributed random variables such that* $\mathrm{E}Y_j = \mu$, $\mathrm{Var}(Y_j) = \sigma^2$ *and* $\mathrm{E}|Y_j|^3 < \infty$. *Let* g *be a function with three continuous derivatives such that for some finite constant* C,

$$\max\{g(x), g'(x), g''(x), g'''(x)\} \leq C$$

for all x. *Then*

$$\lim_{n\to\infty} \mathrm{E}g\left(\frac{\sum_{j=1}^n Y_j}{\sqrt{n}}\right) = \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} g(y)e^{-(y-\mu)^2/2\sigma^2}\mathrm{d}y \ .$$

Proof. Define $\widetilde{T}_j = (Y_j - \mu)/\sigma$. Then $\mathrm{E}\widetilde{T}_j = 0$, $\mathrm{Var}(\widetilde{T}_j) = 1$ and $\mathrm{E}|\widetilde{T}_j|^3 < \infty$. Define a function h by

$$h(x) = g\left(\frac{x-\mu}{\sigma}\right) \ . \tag{0.23}$$

Then $g\left(\frac{\sum_{j=1}^n Y_j}{\sqrt{n}}\right) = h\left(\frac{\sum_{j=1}^n \widetilde{T}_j}{\sqrt{n}}\right)$, and by the previous lemma and (0.23),

$$\begin{aligned}
\lim_{n\to\infty} \mathrm{E}g\left(\frac{\sum_{j=1}^n Y_j}{\sqrt{n}}\right) &= \lim_{n\to\infty} \mathrm{E}h\left(\frac{\sum_{j=1}^n \widetilde{T}_j}{\sqrt{n}}\right) \\
&= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} h(x)e^{-x^2/2}\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} g\left(\frac{x-\mu}{\sigma}\right)e^{-x^2/2}\mathrm{d}x = \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}} g(y)e^{-(y-\mu)^2/2\sigma^2}\mathrm{d}y \ .
\end{aligned}$$

□

The assumptions in the theorem can be relaxed, using simple results form analysis. The requirement that $\mathrm{E}|Y_j|^3 < \infty$ was used only because we used the simplest form of the remainder in Taylor's Theorem. Working harder with the integral form, one can show that nothing more is needed than $\mathrm{E}|Y_j|^2 < \infty$, which is already required to have a finite variance.

Next, any uniformly continuous function g may be uniformly approximated by other functions with three continuous derivatives that satisfy the hypotheses, It is enough to assume that g is bounded and uniformly continuous. Finally since large values of x are largely irrelevant, and since every continuous function on $\mathbb{R}$ is uniformly continuous on all bounded, closed subsets of $\mathbb{R}$, it is actually enough to assume that g is continuous and does not get too large as x increases.