AMSI VACATION RESEARCH SCHOLARSHIPS 2018-2019

Convergence in the Central Limit Theorem

Hugh Entwistle

Supervised by Dr Georgy Sofronov

Macquarie University

Vacation Research Scholarships are funded jointly by the Department of Education and Training and the Australian Mathematical Sciences Institute.

AMSI Australian Mathematical Sciences Institute

Australian Government Department of Education and Training

1 Abstract

In this report we lay out the tools required to prove various Central Limit Theorems with emphasis on the use of characteristic functions and a necessary inequality for a later proof. We then proceed with two different variations of the Central Limit Theorem providing proofs, interpretations of conditions and highlight limitations and points to consider for each theorem. We then move on to a proof of the Berry-Esseen Theorem and an exploration of a lower bound of the constant stated in the inequality before finally ending with some applications by considering the effectiveness of a normal approximation to the Poisson and Binomial Distribution.

2 Introduction

The central limit theorem is an extremely useful theorem in statistics as it describes the 'stabilisation' of samples where, if we take more and more samples from some population, the graph of the data will become normally distributed — to form a bell curve. Properties of this distribution are well understood by mathematicians so this result allows us to understand so much more about the world. There is not one unique central limit theorem, but rather many different versions that all have different requirements and conditions in order for this 'stabilisation' of random variables to occur. It is important to understand and respect the different conditions needed, so this paper presents proofs of two different central limit theorems as well as highlights some examples where the conditions are not completely met to emphasise this aspect. Another question that naturally arises is "How many observations will be required for this effect to manifest itself enough for us to start using the normal approximation to approximate the distributions that we are working with?" which can have ethical, economical and practical consequences depending on the nature of one's research - making this question a very important one to properly answer. Many textbooks and websites propose that you need around 30 or more observations. This answer is a very simplistic one and, while it may be true for some distributions, is not true for all of them. To address this we study and prove the Berry Esseen Theorem which gives an upper bound on the maximum difference between the normal distribution and the distribution, which is dependent on the distribution itself. This upper bound is then studied in more detail.

The majority of this research required the use of characteristic functions and thus a section has been devoted to stating and proving some of their properties. This project involved a close study of proofs supplied in *Probability: A Graduate Course* [**3**] with many of the details expanded and understood independently. This text is highly recommended for the inspired reader.

3 Characteristic Functions

The majority of the proofs presented in this paper involve the use of characteristic functions. In Probability Theory, characteristic functions are particularly useful since they are always defined meaning that they can be often used in more general proofs (unlike, for example, moment generating functions). In this section, we define the characteristic function and derive the characteristic function for five different distributions as we will use these results in later discussions. We present the uniqueness and inversion theorem for characteristic functions without proof and we then prove an important inequality for characteristic functions that will be used in the proof of the Berry-Esseen Theorem.

Definition 1 *The characteristic function of the random variable X is:*

$$\phi_X(t) = \mathbb{E}(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} f_X(x)dx$$

where f_X is the probability density function of X.

3.1 Examples of Characteristic Functions

We will provide five examples of characteristic functions here, which will be referred to in later proofs in this paper.

3.1.1 Bernoulli Distribution

The probability mass function of the Bernoulli Distribution is:

$$P(X = x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \end{cases}$$

So the characteristic function is calculated by:

$$\phi_X(t) = pe^{it} + (1 - p) = 1 + pe^{it} - p.$$

3.1.2 Binomial Distribution

The probability mass function of the Binomial Distribution is

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad x = 0, 1, \ldots, n.$$

So the characteristic function is calculated by:

$$\phi_X(t) = \sum_{x=0}^{n} \binom{n}{x} e^{itx} p^x (1-p)^{n-x} = (1 + pe^{it} - p)^n.$$

We notice the similarities in the two characteristic functions above to be reflective of the fact that the binomial variable can be expressed as the sum of bernoulli variables.

3.1.3 Standard Normal Distribution

By recalling that the probability density function of the standard normal is

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \text{ on } -\infty < x < \infty.$$

We compute the characteristic function as follows:

$$\begin{aligned}\phi_X(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{itx} e^{-\frac{x^2}{2}} dx \\ &= \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{x^2}{2}} \cos(tx) dx.\end{aligned}$$

Differentiating under the integral sign and integrating by parts proves the following differential equation:

$$\frac{\phi_X'(t)}{\phi_X(t)} = -t$$

providing

$$\phi_X(t) = e^{-\frac{t^2}{2}} \text{ noting that } \phi_X(0) = 1.$$

3.1.4 Cauchy Distribution

The probability density function of the standard cauchy distribution is

$$f_X(x) = \frac{1}{\pi(1+x^2)} \text{ on } -\infty < x < \infty.$$

So the characteristic function is calculated by:

$$\phi_X(t) = \int_{-\infty}^{\infty} \frac{1}{\pi(1+x^2)} dx.$$

Considering the contour integral of $\frac{e^{itz}}{1+z^2}$ on the contour that goes along the real axis from $x = -r$ to $x = r$ and then counter-clockwise along the semi-circle with radius $r > 1$ centred at $z = 0$ and applying Cauchy's residue theorem we obtain:

$$\phi_X(t) = e^{-|t|}.$$

3.1.5 Poisson Distribution

The probability mass function of the Poisson distribution is

$$P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!}\text{ for } x=0,1,2,3,\dots \text{ and } \lambda>0.$$

So the characteristic function is calculated by:

$$\begin{aligned}\phi_X(t)&=\sum_{x=0}^{\infty}e^{itx}\frac{e^{-\lambda}\lambda^x}{x!}\\&=e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^{it})^x}{x!}\\&=e^{-\lambda}e^{\lambda e^{it}}\\&=e^{\lambda(e^{it}-1)}.\end{aligned}$$

3.2 Useful Properties of Characteristic Functions

Theorem 1 *Inversion Formula*
If $\int_{-\infty}^{\infty}|\phi_X(t)|dt<\infty$ *then* X *has an absolutely continuous distribution with a bounded, continuous density* f *given by*

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\cdot\phi(t)dt$$

.

We may consult [**3**] for the proof.

Theorem 2 *Let* X *and* Y *be random variables. If* $\phi_X=\phi_Y$*, then* $X\overset{\mathrm{d}}{=}Y$ *and the converse is also true.*

We may consult [**3**] for the proof.

Theorem 3 *Let* X *be a random variable with characteristic function* ϕ*. If* $\mathbb{E}(|X|^n)<\infty$ *for some* $n\in\mathbb{N}$ *then*

$$\left|\phi_X(t)-\sum_{k=0}^{n}\frac{(it)^k}{k!}\mathbb{E}(X^k)\right|\le\mathbb{E}\left(\min\left\{2\frac{|t|^n|X|^n}{n!},\frac{|t|^{n+1}|X|^{n+1}}{(n+1)!}\right\}\right).$$

We provide this proof in the appendix.

Remark: The latter result can be used to establish the property, if the above conditions are satisfied, that

$$\phi_X(t)=1+\sum_{k=1}^{n}\frac{(it)^k}{k!}\mathbb{E}(X^k)+o(|t|^n)\text{ as } t\to 0$$

which plays a useful role in the proof of the Central Limit Theorem.

4 The Central Limit Theorem

The Central Limit Theorem plays an extremely important role in mathematics. It tells us that, upon certain conditions, the distribution of a sum of random variables will converge to the normal distribution. Its utility cannot be understated, since a lot of theorems in statistics and tools in data analysis rely on this property.

4.1 Central Limit Theorem (i.i.d. case)

Theorem 4 *Central Limit Theorem*
Let $X_1, X_2, X_3, \dots$ be independent, identically distributed random variables with finite expectation μ and positive, finite variance σ^2. Let $S_n = \sum_{k=1}^{n} X_k \ n \geq 1$. Then

$$F_{\frac{S_n - n\mu}{\sigma\sqrt{n}}} = \Phi(x) \text{ as } n \to \infty$$

where $F_X(x) = Pr(X \leq x)$ is the cumulative density function (cdf) and $\Phi(x)$ is the cdf of the standard normal distribution.

Proof: It suffices to show, by exploiting the uniqueness of characteristic functions, that:

$$\phi_{\frac{S_n - n\mu}{\sigma\sqrt{n}}}(t) = e^{-\frac{x^2}{2}} \text{ as } n \to \infty.$$

Consider

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{X_k - \mu}{\sigma}$$

where we may assume without loss of generality that $\mu = 0$ and $\sigma = 1$. Thus we obtain:

$$\begin{aligned}
\phi_{\frac{S_n - n\mu}{\sigma\sqrt{n}}}(t) &= \phi_{\frac{S_n}{\sqrt{n}}}(t) \\
&= \phi_{S_n}(\frac{t}{\sqrt{n}}) \\
&= \left(\phi_X(\frac{t}{\sqrt{n}})\right)^n \text{ due to the linearity of } \mathbb{E}(\cdot) \\
&= \left(1 + \sum_{k=1}^{n} \frac{(\frac{it}{\sqrt{n}})^k}{k!}\mathbb{E}(X^k) + o(|t|^n)\right)^n \\
&= \left(1 + \frac{it}{\sqrt{n}}\mathbb{E}(X) + \frac{1}{2!}(\frac{it}{\sqrt{n}})^2\mathbb{E}(X^2) + o(|\frac{t}{\sqrt{n}}|^2)\right)^n \\
&= \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \\
&\to e^{-\frac{t^2}{2}} \text{ as } n \to \infty.
\end{aligned}$$

Remark: It must be noted here that the conditions are quite strict and that there exist sequences of independent, identically distributed random variables that do not satisfy this Theorem.
For example if we consider the Cauchy Distribution, we see that the Theorem does not hold by inspecting the characteristic function:

$$\phi_{\frac{S_n}{n}}(t) = \mathbb{E}\left(\prod_{k=1}^{n} e^{\frac{itX_k}{n}}\right) = \left(\phi_X\left(\frac{t}{n}\right)\right)^n = e^{-|t|}.$$

Which is noticed to be the characteristic function of a Cauchy variable. It follows that the sample mean of n independent, identically distributed Cauchy variables will always follow the Cauchy distribution and thus will never converge to the normal distribution. We notice here that

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} \frac{1}{\pi(1+x^2)} dx = \frac{1}{\pi} \underbrace{\int_{-\infty}^{\alpha} \frac{1}{1+x^2} dx}_{I_1} + \frac{1}{\pi} \underbrace{\int_{\alpha}^{\infty} \frac{1}{1+x^2} dx}_{I_2}.$$

$I_1 = -\infty$ and $I_2 = +\infty$ regardless of what $a \in \mathbb{R}$ is chosen. So the expected value of a Cauchy variable is undefined. In a similar fashion, we can also show that the variance is undefined - and so the conditions in the Central Limit Theorem stated are not met.

4.2 The Lindeberg-Levy-Feller Theorem

Let $X_1, X_2, \ldots$ be independent random variables with finite mean and variances. Set $\mathbb{E}(X_k) = \mu_k, Var(X_k) = \sigma_k^2$ and $S_n = \sum_{k=1}^n X_k$ and $s_n^2 = \sum_{k=1}^n \sigma_k^2 \neq 0$.

Feller Condition:

$$L_1(n) = \max_{1 \leq k \leq n} \frac{\sigma_k^2}{s_n^2} \to 0 \text{ as } n \to \infty.$$

Lindeberg Condition:

$$L_2(n) = \frac{1}{s_n^2} \sum_{k=1}^{n} \mathbb{E}\left((X_k - \mu_k)^2 I\{|X_k - \mu_k| \geq \epsilon s_n\}\right) \to 0 \text{ as } n \to \infty$$

where $I(\cdot)$ is the indicator function.

Theorem 5 *Lindeberg-Levy-Feller*
If the Lindeberg Condition is satisfied then so is the Feller condition. Moreover

$$\frac{1}{s_n} \sum_{k=1}^{n} (X_k - \mu_k) \xrightarrow{d} N(0,1) \text{ as } n \to \infty$$

where $\xrightarrow{d}$ is convergence in distribution. We prove this theorem in the appendix.

4.2.1 Lyapounov's Condition

Theorem 6 *Assume that $\mathbb{E}(|X_k|^r) < \infty$ for all k and some $r > 2$. If*

$$\beta(n,r) = \frac{1}{s_n^r}\sum_{k=1}^{n}\mathbb{E}(|X_k - \mu_k|^r) \to 0 \text{ as } n \to 0,$$

then the Lindeberg condition is satisfied.

We prove this theorem in the appendix.

Remark: Lindeberg's condition is quite hard to verify so it is justified to provide an expression that may be easier at times to evaluate that implies that the Lindeberg condition is true. We must realise that the converse will not always be true and that there will exist sequences of random variables that do not satisfy Lyaponouv's Condition but still satisfy the Lindeberg.

We can consider the independent identically distributed sequences of random variables that follow the distribution:

$$f_X(x) = \frac{c}{|x|^3(\log(|x|))^2} \text{ for } |x| > 2.$$

Then

$$\sigma^2 = 2c\int_2^\infty \frac{dx}{x(\log(x))^2} = \frac{2c}{\log(2)} < \infty.$$

But for $r > 2$

$$\mathbb{E}(|X_k|^r) = 2c\int_2^\infty \frac{x^r}{x^3(\log(x))^2}dx = \infty.$$

Lyapounov's condition is not satisfied. However by symmetry

$$\mu = c\int_{|x|>2} \frac{x}{|x|^3(\log(|x|))^2}dx = 0.$$

So

$$\begin{aligned} L_1(n) &= \frac{1}{s_n^2}\sum_{k=1}^{\infty}\mathbb{E}(|X_k - \mu_k|^2 I\{|X_k - \mu_k| > \epsilon s_n\}) \\ &= \frac{2c}{\sigma^2}\int_{\epsilon\sigma\sqrt{n}}^\infty \frac{dx}{x(\log(x))^2} \text{ , and using integration by parts to evaluate,} \\ &= \frac{2c}{\sigma^2\log(\epsilon\sigma\sqrt{n})} \\ &\to 0 \text{ as } n \to \infty. \end{aligned}$$

Thus the Lindeberg condition is satisfied, but the Lyapounov condition was not.

It should be noted that it was obvious that the Central Limit Theorem would work here as the (i.i.d.) case still applies.

4.2.1.1 Interpretations of the Conditions

We can observe that

$$\begin{aligned} Pr\left(\max_{1\leq k\leq n}|X_k-\mu_k|>\epsilon s_n\right) &\leq \sum_{k=1}^{n} Pr\left(|X_k-\mu_k|>\epsilon s_n\right) \\ &\leq \sum_{k=1}^{n}\frac{\mathbb{E}(|X_k-\mu_k|I\{|X_k-\mu_k|>\epsilon s_n\})}{(\epsilon s_n)^2} \\ &= \frac{1}{\epsilon^2}L_2(n). \end{aligned}$$

So the Lindeberg condition can be interpreted as guaranteeing that no individual summand will 'dominate' any of the others.

And the Feller condition states that the contribution of any individual random variable to the variance s_n^2 is arbitrarily small.

5 The Berry Esseen Theorem

Remark: (Where are all of my thumbs?)

Most textbooks and undergraduate courses that teach the Central Limit Theorem do not properly discuss how many variables are required before the Theorem can be used accurately. Instead, many rules of thumb are offered without much insight or explanation. This became a source of frustration for me as an undergraduate student - and the following theorem is presented to give an insight into how many variables are needed for various distributions. We turn to the celebrated Berry-Esseen Theorem, discovered independently by Andrew C. Berry and Carl-Gustav Esseen. We will first state the general version of the Theorem and provide a proof - before moving on to state the independent, identically distributed case and provide an insight into how a lower bound for the constant stated in the theorem can be found and improved on.

Theorem 7 *Berry-Esseen Theorem for Non-Identically Distributed Variables*
Let $X_1, X_2, \ldots$ be independent random variables with zero mean and partial sums $\{S_n; n \geq 1\}$. Suppose $\gamma_k^3 = \mathbb{E}(|X_k|^3) < \infty$ for all k and set $\sigma_k^2 = Var(X_k), s_n^2 = \sum_{k=1}^n \sigma_k^2, \beta_n^3 = \sum_{k=1}^n \gamma_k^3$. Then

$$\sup_x \left|F_{\frac{S_n}{s_n}}(x) - \Phi(x)\right| \leq C \cdot \frac{\beta_n^3}{s_n^3}$$

where C is a universal constant.

Remarkably, the upper bound only depends on the variance and absolute third raw moment - making this theorem applicable to many distributions.

The proof is based on estimating the difference of cumulative distribution functions by the nearness of characteristic functions. We thus state and prove the following result:

Theorem 8 *Let $F_U(x)$ and $F_V(x)$ be cumulative distribution functions for the random variables U and V respectively. Suppose $F_V(x)$ has a bounded derivative so that*

$$\sup_{x\in\mathbb{R}} F_V'(x) \leq A \text{ where } A \in \mathbb{R}.$$

Then

$$\sup_{x\in\mathbb{R}} |F_U(x) - F_V(x)| \leq \frac{1}{\pi}\int_{-T}^{T} \left|\frac{\phi_U(t) - \phi_V(t)}{t}\right| dt + \frac{24A}{\pi T}.$$

We refer to the appendix for this proof.

We now set $U = \frac{S_n}{s_n}$ and V to follow the standard normal distribution in Theorem 8.

We progress towards the Berry-Esseen inequality by first showing:

Lemma 1

$$\left|\phi_{\frac{S_n}{s_n}}(t) - e^{-\frac{t^2}{2}}\right| \leq 16\frac{\beta_n^3}{s_n^3}|t|^3 e^{-\frac{t^2}{3}} \text{ for } |t| \leq \frac{s_n^3}{4\beta_n^3}.$$

Proof:

The proof is separated into two components:

1. $|\phi_{\frac{S_n}{s_n}}(t)| \leq e^{-\frac{t^2}{3}}$ for $|t| \leq \frac{s_n^3}{4\beta_n^3}$.
2. The lemma holds for $\frac{s_n}{2\beta_n} \leq |t| \leq \frac{s_n^3}{4\beta_n^3}$.

For 1, we consider the symmetrized sequence of random variables $\{X_k^s; k \geq 1\}$ which are independent and identically distributed such that

$\mathbb{E}(X_k^s) = 0, Var(X_k^s) = 2\sigma_k^2, \mathbb{E}(|X_k^s|^3) \leq 8\gamma_k^3$

where the symmetrized random variable X_k^s is defined to be $X_k - X'$ where X' follows the same distribution as X_k. From this definition it is immediately clear that $\mathbb{E}(X_k^s) = 0$ and $Var(X_k^s) = 2\sigma_k^2$. The last result follows from the c_r-inequality since:

$$\mathbb{E}(|X_k - X'|^3) \leq \mathbb{E}((|X_k| + |X'|)^3) \leq 2^{3-1}(\mathbb{E}(|X_k|^3) + \mathbb{E}(|X'|^3)) = 8\gamma_k^3.$$

Now $\phi_{X_k^s}(t) = \phi_{X_k - X'}(t) = |\phi_{X_k}(t)|^2$ is real valued. And so, once again applying the inequality result for characteristic functions:

$$\left|\phi_{X_k^s}(t) - \sum_{j=0}^{2}\frac{(it)^j}{j!}\mathbb{E}((X_k^s)^j)\right| = \left|\phi_{X_k^s}(t) - (1 - t^2\sigma_k^2)\right|$$

$$\leq \mathbb{E}\left(\min\left\{t^2(X_k^s)^2, \frac{|t|^3|X_k^s|^3}{6}\right\}\right)$$
$$= \mathbb{E}\left(\min\left\{t^2(X_k^s)^2, \frac{|t|^3 8\gamma_k^3}{6}\right\}\right)$$
$$\leq \frac{|t|^3 4\gamma_k^3}{3}.$$

From this we obtain that

$$\left|\phi_{\frac{X_k^s}{s_n}}(t) - (1 - \frac{t^2\sigma_k^2}{s_n^2})\right| \leq \frac{|t|^3 4\gamma_k^3}{3s_n^3}$$

and so

$$\left|\phi_{\frac{X_k^s}{s_n}}(t)\right| \leq 1 - \frac{t^2\sigma_k^2}{s_n^2} + \frac{4|t|^3\gamma_k^3}{3s_n^3} \leq e^{-\frac{t^2\sigma_k^2}{s_n^2} + \frac{4|t|^3\gamma_k^3}{3s_n^3}}.$$

Thus

$$\begin{aligned}
\left|\phi_{\frac{S_n}{s_n}}(t)\right|^2 &= \phi_{\frac{S_n^s}{s_n}}(t) \\
&= \prod_{k=1}^{n} \phi_{\frac{X_k^s}{s_n}}(t) \\
&\leq \prod_{k=1}^{n} e^{-\frac{t^2\sigma_k^2}{s_n^2} + \frac{4|t|^3\gamma_k^3}{3s_n^3}} \\
&= e^{\sum_{k=1}^{n}\left(-\frac{t^2\sigma_k^2}{s_n^2} + \frac{4|t|^3\gamma_k^3}{3s_n^3}\right)} \\
&= e^{-t^2 + \frac{t^2}{3}\cdot|t|\cdot\frac{4\beta_n^3}{s_n^3}} \\
&\leq e^{-\frac{2t^2}{3}} \text{ for } |t| \leq \frac{s_n^3}{4\beta_n^3}.
\end{aligned}$$

So

$$\left|\phi_{\frac{S_n}{s_n}}(t)\right| \leq \sqrt{e^{-\frac{2t^2}{3}}} = e^{-\frac{t^2}{3}} \text{ as required.}$$

For 2:

If $\frac{s_n}{2\beta_n} \leq |t|$ then $1 \leq \frac{2|t|\beta_n}{s_n} \leq \frac{8|t|^3\beta_n^3}{s_n^3}$ and so

$$\begin{aligned}
\left|\phi_{\frac{S_n}{s_n}}(t) - e^{-\frac{t^2}{2}}\right| &\leq \left|\phi_{\frac{S_n}{s_n}}(t)\right| + e^{-\frac{t^2}{2}} \text{(Triangular Inequality)} \\
&\leq e^{-\frac{t^2}{3}} + e^{-\frac{t^2}{2}} \text{(From (1))} \\
&\leq 2e^{-\frac{t^2}{3}} \\
&\leq \frac{8|t|^3\beta_n^3}{s_n^3}\cdot 2e^{-\frac{t^2}{3}}
\end{aligned}$$

$$= \frac{16|t|^3\beta_n^3}{s_n^3}e^{-\frac{t^2}{3}} \text{ as required.}$$

We now have enough to prove the Berry-Esseen Theorem:

5.1 Proof of Berry-Esseen Theorem

Recalling that $U = \frac{S_n}{s_n}$ and $V \sim N(0,1)$, we note that $\sup_{x\in\mathbb{R}} F_V'(x) \leq \frac{1}{\sqrt{2\pi}} \Rightarrow A \leq \frac{1}{\sqrt{2\pi}}$ and we let $T = \frac{s_n^3}{4\beta_n^3}$ obtaining:

$$\begin{aligned}
\sup_{x\in\mathbb{R}} \left|F_{\frac{S_n}{s_n}}(x) - \Phi(x)\right| &\leq \frac{1}{\pi}\int_{-T}^{T}\left|\frac{\phi_{\frac{S_n}{s_n}}(t) - e^{-\frac{t^2}{2}}}{t}\right| dt + \frac{24A}{\pi T} \\
&\leq \frac{1}{\pi}\int_{-T}^{T}\frac{1}{|t|}16\frac{\beta_n^3}{s_n^3}|t|^3 e^{-\frac{t^2}{3}}dt + \frac{24}{\pi}\cdot\frac{1}{\sqrt{2\pi}}\cdot\frac{4\beta_n^3}{s_n^3} \text{ (using the previous lemma.)} \\
&= \frac{16}{\pi}\cdot\frac{\beta_n^3}{s_n^3}\int_{-T}^{T} t^2 e^{-\frac{t^2}{3}}dt + \frac{96\beta_n^3}{\pi\sqrt{2\pi}s_n^3} \\
&\leq \frac{16}{\pi}\cdot\frac{\beta_n^3}{s_n^3}\int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{3}}dt + \frac{96\beta_n^3}{\pi\sqrt{2\pi}s_n^3} \\
&= \frac{\beta_n^3}{s_n^3}\frac{24}{\pi}\cdot\sqrt{\frac{3}{2}}\cdot\sqrt{2\pi} \qquad \underbrace{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\cdot\sqrt{\frac{3}{2}}}e^{-\frac{1}{2}\left(\frac{t-0}{\sqrt{\frac{3}{2}}}\right)^2}dt}_{\text{This is the integral of a normal density function}} \qquad + \frac{96\beta_n^3}{\pi\sqrt{2\pi}s_n^3} \\
&= \frac{16}{\pi}\cdot\frac{\beta_n^3}{s_n^3}\times\left(\frac{3}{2}\sqrt{3\pi}\right) + \frac{96\beta_n^3}{\pi\sqrt{2\pi}s_n^3} \\
&= \frac{\beta_n^3}{s_n^3}\left(\frac{24\sqrt{3}}{\sqrt{\pi}} + \frac{96}{\pi\sqrt{2\pi}}\right) \\
&\approx \frac{\beta_n^3}{s_n^3}\times 35.643698\ldots \\
&\leq 36\frac{\beta_n^3}{s_n^3}.
\end{aligned}$$

From this it follows that $C \leq 36$ but of course this is only a proof of existence and C has been drastically improved upon through different results:

- $C \leq 36$.
- (Feller, 1979) $C \leq 3$.
- (Beek, 1972) $C \leq 0.7975$.
- (Shiganov, 1986) $C \leq 0.7915$.

- (Shevtsova, 2012) $C \leq 0.5600$.

Before illustrating a method to obtain a lower bound for C, we first state the Berry-Esseen Theorem for identically distributed variables.

Theorem 9 *Berry-Esseen Theorem for Identically Distributed Random Variables Let $X_1, X_2, \dots$ be independent, identically distributed random variables with partial sums $\{S_n : n \geq 1\}$. Set $\mu = \mathbb{E}(X), \sigma^2 = Var(X)$ and $\gamma^3 = \mathbb{E}(|X-\mu|^3) < \infty$. Then:*

$$\sup_{x \in \mathbb{R}} |F_{\frac{S_n - n\mu}{\sigma\sqrt{n}}}(x) - \Phi(x)| \leq C \cdot \frac{\gamma^3}{\sigma^3 \sqrt{n}}$$

where C is a universal constant.

The proof of this theorem is obtained from the previous proof.
Bounds for C are different from the non-identically distributed variation of the theorem.

- (Esseen, 1942) $C \leq 7.59$.
- (Beek, 1972) $C \leq 0.7882$.
- (Shiganov, 1986) $C \leq 0.7655$.
- (Shevtsova, 2012) $C \leq 0.4784$.

5.2 Obtaining a Lower Bound for C

By considering the discrete random variable X with masses placed at $x = \pm\frac{1}{2}$ with equal probability and taking random variables $X_1, X_2, \dots$ from this distribution. Then $\frac{2X_k+1}{2}$ will follow the $Bern(\frac{1}{2})$ distribution. With the insight that the sum of n independent, identically distributed Bernoulli random variables will be binomially distributed, which can be easily verified using the uniqueness of characteristic functions since (letting $Y_k \sim Bern(1/2)$)

$$\begin{aligned} \phi_{Y_1+Y_2+\cdots+Y_n}(t) &= (\phi_{Y_1}(t))^n \\ &= \left(e^{it}+1\right)^n \left(\frac{1}{2}\right)^n \\ &= \sum_{k=0}^{n} \binom{n}{k} e^{itk} \left(\frac{1}{2}\right)^n \\ &= \phi_{Bin(n,\frac{1}{2})}(t). \end{aligned}$$

So we can compute a lower bound for observing that:

$$\begin{aligned}
Pr\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} = 0\right) &= Pr\left(\frac{2}{\sqrt{n}}\sum_{k=0}^{n} X_k = 0\right) \\
&= Pr\left(\sum_{k=0}^{n}\frac{2X_k+1}{2} = \frac{n}{2}\right) \\
&= \binom{n}{n/2}\left(\frac{1}{2}\right)^n \text{ assuming } n \text{ is even.} \\
&\sim \frac{\sqrt{2\pi n}}{\pi n}\left(\frac{1}{2}\right)^n\left(\frac{n}{e}\right)^n\left(\frac{2e}{n}\right)^n \\
&= \frac{2}{\sqrt{2\pi n}}.
\end{aligned}$$

We observe that that the largest 'jump' in the difference of cumulative distribution functions stated in the inequality will occur at $x = 0$. Thus we evaluate the bound as we approach $x = 0$ from the right.

$$\begin{aligned}
\lim_{n\to\infty}\sqrt{n}\sup_{x\in\mathbb{R}}\left|F_{\frac{S_n-n\mu}{\sigma\sqrt{n}}}(x) - \Phi(x)\right| &= \lim_{n\to\infty}\sqrt{n}\sup_{x\in\mathbb{R}}\left|F_{\frac{2S_n}{\sqrt{n}}}(x) - \Phi(x)\right| \\
&\geq \lim_{n\to\infty}\sqrt{n}\lim_{x\to 0^+}\left|F_{\frac{2S_n}{\sqrt{n}}}(x) - \Phi(x)\right| \\
&= \frac{1}{\sqrt{2\pi}}.
\end{aligned}$$

Noticing that $\sigma = \gamma = \frac{1}{2}$ we obtain that $C \geq \frac{1}{\sqrt{2\pi}} \approx 0.3989$.
Esseen further observed that we can use another two-point distribution to yield $C \geq \frac{\sqrt{10}+3}{6\sqrt{2\pi}} \approx 0.4097$ which, as of January 2019, is the tightest known lower bound for C in both variations of the Theorem stated in this paper.

5.3 Brief Applications of the Berry-Esseen Theorem

We will now show how the Berry-Esseen theorem can be used to quantify the rate of convergence in using the normal approximation to the binomial and poisson distribution. This can be then used to formulate our own 'rules of thumb' based on our own requirements of how much error is acceptable in using an approximation. We provide a sample table of values for each of these approximations in the appendix.

5.3.1 Normal Application to the Poisson Distribution

Letting X be a Poisson random variable with parameter λ and $X_1, X_2, \ldots, X_n$ be independent Poisson random variables with parameter $\frac{\lambda}{n}$, we notice that:

$$\phi_{X_1+X_2+\cdots+X_n}(t) = (\phi_{X_1}(t))^n = (e^{\frac{\lambda}{n}(e^{it}-1)})^n = e^{\lambda(e^{it}-1)} = \phi_X(t)$$

which alerts us of the fact that X can be expressed as the sum of n independent, identically distributed Poisson random variables and so the Central Limit Theorem can be used here.
So in the (iid) version of the Berry-Esseen Theorem we have $\mu = \frac{\lambda}{n}, \sigma^2 = \frac{\lambda}{n}$ and

$$\gamma^3 = \mathbb{E}(|X-\mu|^3) < \mathbb{E}((X+\mu)^3) = \frac{\lambda}{n} + 6\left(\frac{\lambda}{n}\right)^2 + 8\left(\frac{\lambda}{n}\right)^3$$

where $\mathbb{E}(X^k)$ can be easily computed as they are Touchard polynomials in λ.
So we obtain:

$$\frac{C\gamma^3}{\sigma^3\sqrt{n}} < \frac{C\left(\frac{\lambda}{n} + 6\left(\frac{\lambda}{n}\right)^2 + 8\left(\frac{\lambda}{n}\right)^3\right)}{\frac{\lambda}{n}\sqrt{\frac{\lambda}{n}}\sqrt{n}} = C\left(\frac{8\lambda^{\frac{3}{2}}}{n^2} + \frac{6\sqrt{\lambda}}{n} + \frac{1}{\sqrt{\lambda}}\right) \to \frac{C}{\sqrt{\lambda}} \text{ as } n \to \infty.$$

Using $C \leq 0.4784$, we can produce a table of values to compare and look at the largest differences for each λ. A sample table is provided in the appendix. It should be noted that In reality, the normal gives a better approximation - and one can further improve the accuracy with a continuity correction.

5.3.2 Normal Application to the Binomial Distribution

As the Binomial variable is defined as the sum of Bernoulli variables, we may similarly apply the Central Limit Theorem and the Berry Esseen Inequality.
We have $\mu = p, \sigma^2 = pq$ and $\gamma^3 = pq(p^2+q^2)$. So

$$\frac{C\gamma^3}{\sigma^3\sqrt{n}} = \frac{C(p^2+q^2)}{\sqrt{npq}}$$

where ideally we require $p \approx \frac{1}{2}$ for the best result and n as large as possible. Rules of thumb aren't as straightforward here as acceptable values for n vary on the different values for p. A table of values for various values of n and p is provided in the appendix.

6 Discussion and Conclusion

The Central Limit Theorem is an extremely important theorem in mathematics but we must be careful when applying it. There are cases where the Central Limit Theorem will not work which thus requires a better understanding of the conditions stated in each theorem and an awareness of the distribution that we are using. Even when the Central Limit Theorem can be applied - we must take extra care in how many variables are 'good enough' to justify using the theorem. Historically, undergraduate students are simply taught various rules of thumb for different distributions - but, by applying the Berry-Esseen Theorem, we must appreciate that it is possible to produce a quantitative result for the maximum error that we are to expect when using a normal approximation. This result will be different for each distribution as the error is dependent on the variance and third absolute moment which varies for each distribution. The universal constant C stated in the Berry-Esseen Theorem remains to be unknown with upper and lower bounds being continuously developed by mathematicians makes this an extremely relevant and exciting area of research.

7 Appendix

7.1 Proof of Theorem 3

Let $y > 0$ and $k \geq 0$

$$\int_0^y e^{ix}(y-x)^k dx = \frac{y^{k+1}}{k+1} + \frac{i}{k+1}\int_0^y e^{ix}(y-x)^{k+1}dx$$

$$\int_0^y e^{ix}dx = \frac{e^{iy}-1}{i}.$$

Thus

$$\int_0^y e^{ix}dx = \frac{e^{iy}-1}{i} = y + i\int_0^y e^{ix}(y-x)dx.$$

Solving for e^{iy} gives:

$$e^{iy} = 1 + iy + i^2\int_0^y e^{ix}(y-x)dx$$

and by induction, we obtain:

$$e^{iy} = \sum_{k=0}^{n}\frac{(iy)^k}{k!} + \frac{i^{n+1}}{n!}\int_0^y e^{ix}(y-x)^n dx.$$

So

$$\begin{aligned}\left|e^{iy} - \sum_{k=0}^{n}\frac{(iy)^k}{k!}\right| &= \left|\frac{i^{n+1}}{n!}\int_0^y e^{ix}(y-x)^n dx\right| \\ &\leq \left|\frac{1}{n!}\int_0^y e^{ix}(y-x)^n dx\right| \\ &\leq \left|\frac{1}{n!}\int_0^y (y-x)^n dx\right| \\ &= \frac{y^{n+1}}{(n+1)!}.\end{aligned}$$

Or, similarly

$$\begin{aligned}e^{iy} &= \sum_{k=0}^{n-1}\frac{(iy)^k}{k!} + \frac{i^n}{(n-1)!}\int_0^y e^{ix}(y-x)^{n-1}dx \\ &= \sum_{k=0}^{n}\frac{(iy)^k}{k!} - \frac{(iy)^n}{n!} + \frac{i^n}{(n-1)!}\int_0^y e^{ix}(y-x)^{n-1}dx \\ &= \sum_{k=0}^{n}\frac{(iy)^k}{k!} + \frac{i^n}{(n-1)!}\left(\int_0^y e^{ix}(y-x)^{n-1}dx - \frac{y^n}{n}\right) \\ &= \sum_{k=0}^{n}\frac{(iy)^k}{k!} + \frac{i^n}{(n-1)!}\int_0^y e^{ix}(y-x)^{n-1} - (y-x)^{n-1}dx \\ &= \sum_{k=0}^{n}\frac{(iy)^k}{k!} + \frac{i^n}{(n-1)!}\int_0^y (e^{ix}-1)(y-x)^{n-1}dx.\end{aligned}$$

Thus

$$\left|e^{iy}-\sum_{k=0}^{n}\frac{(iy)^k}{k!}\right| \leq \frac{1}{(n-1)!}\int_0^y |e^{ix}-1||y-x|^{n-1}dx$$
$$\leq \frac{2}{(n-1)!}\int_0^y |y-x|^{n-1}dx$$
$$=\frac{2|y|^n}{n}.$$

Where, by the Triangular Inequality we have that

$$|e^{ix}-1| \leq |e^{ix}|+1 = 2.$$

Hence, from the two inequalities

$$\left|e^{iy}-\sum_{k=0}^{n}\frac{(iy)^k}{k!}\right| \leq \min\left\{2\frac{|y|^n}{n}, \frac{|y|^{n+1}}{(n+1)!}\right\}.$$

Replacing y with tX and applying $\mathbb{E}(\cdot)$ obtains the required result.

7.2 Proof of Theorem 5 (Lindeberg Levy Feller)

We first prove that the Lindeberg condition implies the Feller condition. We may, without loss of generality, assume that $\mu_k = 0$. For any $\epsilon > 0$ we have that

$$\begin{aligned}
L_1(n) &= \max_{1\leq k\leq n}\frac{\sigma_k^2}{s_n^2} \\
&= \frac{1}{s_n^2}\max_{1\leq k<n}\int_{\mathbb{R}} x^2 f_{X_k}(x)dx \\
&= \frac{1}{s_n^2}\max_{1\leq k<n}\int_{|x|<\epsilon s_n} x^2 f_{X_k}(x)dx + \frac{1}{s_n^2}\max_{1\leq k\leq n}\int_{|x|\geq\epsilon s_n} x^2 f_{X_k}(x)dx \\
&\leq \frac{1}{s_n^2}\max_{1\leq k<n}\int_{|x|<\epsilon s_n} x^2 f_{X_k}(x)dx + \frac{1}{s_n^2}\sum_{k=1}^{n}\int_{|x|\geq\epsilon s_n} x^2 f_{X_k}(x)dx \\
&\leq \frac{1}{s_n^2}(\epsilon s_n)^2 + L_2(n) \\
&= L_2(n) + \epsilon^2.
\end{aligned}$$

So $L_2(n) \to 0$ as $n \to \infty \Rightarrow L_1(n) \to 0$ as $n \to \infty$ as required.

Now using characteristic functions, we prove that the normalised sum converges to the standard normal distribution:

$$\phi_{\frac{1}{s_n}\sum_{k=1}^{n}(X_k-\mu_k)}(t) = \phi_{\frac{S_n}{s_n}}(t)$$

$$
\begin{aligned}
&= \prod_{k=1}^{n} \phi_{X_k}(\frac{t}{s_n}) \\
&= e^{\sum_{k=1}^{n} \log(\phi_{X_k}(\frac{t}{s_n}))} \\
&\to e^{-\frac{t^2}{2s_n^2}\sum_{k=1}^{n}\sigma_k^2} \text{ as } n \to \infty \qquad (1)\\
&= e^{-\frac{t^2}{2}}.
\end{aligned}
$$

For the details for (1) we consider the use of the following results:

$$\left|\sum_{k=1}^{n}\left(\log\left(\phi_{X_k}\left(\frac{t}{s_n}\right)\right) + \left(1-\phi_{X_k}\left(\frac{t}{s_n}\right)\right)\right)\right| \to 0 \text{ as } n \to \infty. \qquad (2)$$

$$\left|\sum_{k=1}^{n}\left(\phi_{X_k}\left(\frac{t}{s_n}\right) - \left(1-\frac{\sigma_k^2 t^2}{2s_n^2}\right)\right)\right| \to 0 \text{ as } n \to 0. \qquad (3)$$

For (2) we recall the inequality result for characteristic functions, replacing t with $\frac{t}{s_n}$:

$$
\begin{aligned}
\left|\phi_{X_k}\left(\frac{t}{s_n}\right) - 1\right| &\le \mathbb{E}\left(\min\left\{\frac{2|tX_k|}{s_n}, \frac{t^2X_k^2}{2s_n^2}\right\}\right) \\
&\le \mathbb{E}\left(\frac{t^2X_k^2}{2s_n^2}\right) \\
&= \frac{t^2}{2}\left(\frac{\sigma_k^2}{s_n^2}\right) \\
&\le \frac{t^2}{2}L_1(n) \\
&\to 0 \text{ as } n \to \infty \text{ since the Feller condition is satisfied.}
\end{aligned}
$$

So

$$
\begin{aligned}
\left|\sum_{k=1}^{n}\left(\log\left(\phi_{X_k}\left(\frac{t}{s_n}\right)\right) + \left(1-\phi_{X_k}\left(\frac{t}{s_n}\right)\right)\right)\right| &\le \sum_{k=1}^{n}\left|\log\left(\phi_{X_k}\left(\frac{t}{s_n}\right)\right) + \left(1-\phi_{X_k}\left(\frac{t}{s_n}\right)\right)\right| \\
&= \sum_{k=1}^{n}\left|\log\left(1-(1-\phi_{X_k}(\frac{t}{s_n})\right) + (1-\phi_{X_k}(\frac{t}{s_n}))\right| \\
&\le \sum_{k=1}^{n}\left|1-\phi_{X_k}(\frac{t}{s_n})\right|^2 \text{ Since } |\log(1-z)+z| \le |z|^2 \text{ if } |z| \le \frac{1}{2} \\
&\le \max_{1\le k\le n}\left|1-\phi_{X_k}(\frac{t}{s_n})\right| \cdot \sum_{k=1}^{n}\left|1-\phi_{X_k}(\frac{t}{s_n})\right| \\
&\le \underbrace{\frac{t^2}{2}L_1(n)}_{\text{From the starting inequality}} \cdot \sum_{k=1}^{n}\mathbb{E}(\frac{t^2X_k^2}{2s_n^2})
\end{aligned}
$$

$$= \frac{t^4}{4} \cdot L_1(n)$$
$$\to 0 \text{ as } n \to \infty.$$

For (3) we make use of the inequality:

$$\left|\sum_{k=1}^{n}\left(\phi_{X_k}\left(\frac{t}{s_n}\right) - \left(1 - \frac{\sigma_k^2 t^2}{2s_n^2}\right)\right)\right| \leq \sum_{k=1}^{n}\left|\phi_{X_k}\left(\frac{t}{s_n}\right) - \left(1 - \frac{\sigma_k^2 t^2}{2s_n^2}\right)\right|$$
$$\leq \sum_{k=1}^{n}\mathbb{E}\left(\min\left\{\frac{t^2X_k^2}{s_n^2}, \frac{|t|^3|X_k|^3}{6s_n^3}\right\}\right)$$
$$= \sum_{k=1}^{n}\mathbb{E}\left(\min\left\{\frac{t^2X_k^2}{s_n^2}, \frac{|t|^3|X_k|^3}{6s_n^3}\right\} I\{|X_k| > \epsilon s_n\}\right)$$
$$+ \sum_{k=1}^{n}\mathbb{E}\left(\min\left\{\frac{t^2X_k^2}{s_n^2}, \frac{|t|^3|X_k|^3}{6s_n^3}\right\} I\{|X_k| \leq \epsilon s_n\}\right)$$
$$\leq \sum_{k=1}^{n}\mathbb{E}\left(\frac{|t|^3|X_k|^3}{6s_n^3} I\{|X_k| > \epsilon s_n\}\right) + \sum_{k=1}^{n}\mathbb{E}\left(\frac{t^2X_k^2}{s_n^2} I\{|X_k| \leq \epsilon s_n\}\right)$$
$$= \sum_{k=1}^{n}\left(\frac{|t|^3}{6s_n^3}\mathbb{E}\left(|X_k|^3 I\{|X_k| \leq \epsilon s_n\}\right)\right) + t^2 L_2(n)$$
$$\leq \sum_{k=1}^{n}\frac{|t|^3\epsilon s_n}{6s_n^3}\mathbb{E}\left(|X_k|^2 I\{|X_k| \leq \epsilon s_n\}\right) + t^2 L_2(n)$$

since $|X_k|^3 = |X_k||X_k|^2 \leq \epsilon s_n |X_k|^2$ under the indicator function.

$$\leq \sum_{k=1}^{n}\frac{|t|^3\epsilon}{6s_n^2}\mathbb{E}\left(|X_k|^2\right) + t^2 L_2(n)$$
$$\leq \sum_{k=1}^{n}\frac{|t|^3\epsilon}{6} + t^2 L_2(n)$$

where $L_2(n) \to 0$ as $n \to \infty$.

We thus obtain that

$$\left|\sum_{k=1}^{n}\left(\phi_{X_k}\left(\frac{t}{s_n}\right) - \left(1 - \frac{\sigma_k^2 t^2}{2s_n^2}\right)\right)\right| \leq \frac{|t|^3\epsilon}{6} \to 0$$

As $n \to \infty$ and ϵ is arbitrarily small.

So now (1) is justified through

$$e^{\sum_{k=1}^{n}\log(\phi_{X_k}(\frac{t}{s_n}))} \to e^{-\sum_{k=1}^{n}(1-\phi_{X_k}(\frac{t}{s_n}))} \text{ by (2)}$$
$$\to e^{-\sum_{k=1}^{n}(1-(1-\frac{\sigma_k^2 t^2}{2s_n^2}))} \text{ by (3)}$$
$$= e^{-\frac{t^2}{2s_n^2}\sum_{k=1}^{n}\sigma_k^2}$$

as required.

7.2.1 Proof of Lyapounov's Condition

Let $\epsilon > 0$ and without loss of generality assume that $\mu_k = 0$. Then

$$\begin{aligned}
\frac{1}{s_n^2}\sum_{k=1}^{n}\mathbb{E}\left(|X_k|^2 I\{|X_k| > \epsilon s_n\}\right) &= \frac{1}{s_n^2}\sum_{k=1}^{n}\mathbb{E}\left(\frac{|X_k|^r}{|X_k|^{r-2}} I\{|X_k| < \epsilon s_n\}\right) \\
&\leq \frac{1}{s_n^2}\sum_{k=1}^{n}\frac{1}{(\epsilon s_n)^{r-2}}\mathbb{E}\left(|X_k|^r I\{|X_k| > \epsilon s_n\}\right) \\
&\leq \frac{1}{\epsilon^{r-2}s_n^r}\sum_{k=1}^{n}\mathbb{E}(|X_k|^r) \\
&= \frac{1}{\epsilon^{r-2}}\beta(n,r) \\
&\to 0 \text{ as } n \to 0 \text{ for any } \epsilon.
\end{aligned}$$

7.3 Proof of Theorem 8

We consider the triangular distribution $Tri(-1,1)$ which has the following probability density function:

$$f_X(x) = \begin{cases} 1-|x| & |x| \leq 1 \\ 0 & |x| > 1 \end{cases}$$

and characteristic function:

$$\begin{aligned}
\phi_X(t) &= \int_{-1}^{1} e^{itx}(1-|x|)dx \\
&= \left(\frac{\sin(t/2)}{t/2}\right)^2.
\end{aligned}$$

And so it follows that the characteristic function for a $Tri(-T,T)$ variable is

$$\left(\frac{\sin(t/2)}{t/2}\right)^2$$

with probability density function

$$f_X(x) = \begin{cases} 1-\frac{|x|}{T} & |x| \leq T \\ 0 & |x| > T \end{cases}.$$

So by using the inversion formula for characteristic functions, stated previously:

$$\frac{1}{T}\left(1-\frac{|x|}{T}\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\cdot\frac{4\sin^2(tT/2)}{(tT)^2}dt$$

and so

$$1 - \frac{|t|}{T} = \int_{-\infty}^{\infty} e^{ixt} \cdot \frac{1 - \cos(xT)}{\pi T x^2} dx$$

And we now consider Z_T to be a random variable with density

$$\frac{1 - \cos(xT)}{\pi T x^2}$$

and thus

$$\phi_{Z_T}(t) = 1 - \frac{|t|}{T}; |t| \leq T.$$

Now consider

$$\int_{-T}^{T} \left| \frac{\phi_U(t) - \phi_V(t)}{t} \right| dt$$

and set:

$$\begin{aligned}
\Delta(x) &= F_U(x) - F_V(x) \\
\Delta_T(x) &= F_{U+Z_T}(x) - F_{V+Z_T}(x) \\
&= \int_{-\infty}^{\infty} \Delta(x-y) f_{Z_T}(y) dy \\
\Delta^* &= \sup_{x \in \mathbb{R}} |\Delta(x)| \\
\Delta_T^* &= \sup_{x \in \mathbb{R}} |\Delta_T(x)|.
\end{aligned}$$

Since $U + Z_T$ and $V + Z_T$ are sums of independent random variables and so the 'characteristic function' corresponding to $\Delta_T(x)$ is

$$(\phi_U(t) - \phi_V(t))\phi_{Z_T}(t).$$

Now $U + Z_T$ and $V + Z_T$ are both continuous $\Rightarrow \Delta_T(x)$ is also continuous.
Hence $\Delta_T(x) \to \Delta(x)$ as $T \to \infty$.
We now estimate Δ_T^* and use this to estimate Δ^*. Once again by the inversion theorem:

$$f_{U+Z_T}(x) - f_{V+Z_T}(x) = \frac{1}{2\pi} \int_{-T}^{T} e^{-itx} (\phi_U(t) - \phi_V(t)) \phi_{Z_T}(t) dt$$

and integrating with respect to x:

$$\Delta_T(x) = \frac{1}{2\pi} \int_{-T}^{T} e^{-itx} \cdot \frac{\phi_U(t) - \phi_V(t)}{-it} \cdot \phi_{Z_T}(t) dt.$$

So

$$
\begin{aligned}
\Delta_T^* &= \frac{1}{2\pi} \sup_{x\in\mathbb{R}} \left| \int_{-T}^{T} e^{-itx} \cdot \frac{\phi_U(t)-\phi_V(t)}{-it} \cdot \phi_{Z_T}(t)dt \right| \\
&\leq \frac{1}{2\pi} \int_{-T}^{T} \left| \frac{\phi_U(t)-\phi_V(t)}{t} \right| \left(1 - \frac{|t|}{T}\right) dt \\
&\leq \frac{1}{2\pi} \int_{-T}^{T} \left| \frac{\phi_U(t)-\phi_V(t)}{t} \right| dt.
\end{aligned}
$$

At this point, it remains to show that

$$\Delta^* \leq 2\Delta_T^* + \frac{24A}{\pi T}.$$

Since

$$
\begin{aligned}
\Delta(-\infty) &= F_U(-\infty) - F_V(-\infty) = 0 \\
\text{and } \Delta(\infty) &= F_U(\infty) - F_V(\infty) = 0
\end{aligned}
$$

and $\Delta(x)$ is the difference of two distribution functions, then it will be right-continuous with left-hand limits. Thus Δ^* will either be $|\Delta(x)|$ evaluated at some $x = x_0$ or at $x \to x_0^-$. This will bring about four possibilities:

1. $\Delta^* = \Delta(x_0)$
2. $\Delta^* = -\Delta(x_0)$
3. $\Delta^* = -\lim_{x\to x_0^-} \Delta(x)$
4. $\Delta^* = \lim_{x\to x_0^-} \Delta(x)$.

We have to handle these four cases separately.
For $(1) : \Delta^* = \Delta(x_0)$, this implies that

$$\Delta(x_0 + s) \geq \Delta^* - As \text{ for some } s > 0 \text{ by definition of supremum.}$$

In particular, choose $s = \frac{\Delta^*}{2A} + y$ and so

$$
\begin{aligned}
\Delta(x_0 + \frac{\Delta^*}{2A} + y) &\geq \Delta^* - A(\frac{\Delta^*}{2A} + y) \\
&= \frac{\Delta^*}{2} - Ay \text{ when } |y| \leq \frac{\Delta^*}{2A}.
\end{aligned}
$$

In the case where $|y| > \frac{\Delta^*}{2A}$, we are not as interested. We simply use the fact that $-\Delta^*$ is by definition a lower bound for $\Delta(\cdot)$. Thus:

$$\Delta(x_0 + \frac{\Delta^*}{2A} + y) \geq -\Delta^* \text{ when } |y| > \frac{\Delta^*}{2A}.$$

Then we obtain:

$$\begin{aligned}
\Delta_T\left(x_0 + \frac{\Delta^*}{2A}\right) &= \int_{-\infty}^{\infty} \Delta(x_0 + \frac{\Delta^*}{2A} - y) f_{Z_T}(y) dy \\
&= \int_{|y| \leq \frac{\Delta^*}{2A}} \Delta(x_0 + \frac{\Delta^*}{2A} - y) f_{Z_T}(y) dy + \int_{|y| > \frac{\Delta^*}{2A}} \Delta(x_0 + \frac{\Delta^*}{2A} - y) f_{Z_T}(y) dy \\
&= \int_{|y| \leq \frac{\Delta^*}{2A}} \Delta(x_0 + \frac{\Delta^*}{2A} + y) f_{Z_T}(y) dy + \int_{|y| > \frac{\Delta^*}{2A}} \Delta(x_0 + \frac{\Delta^*}{2A} + y) f_{Z_T}(y) dy \\
&\geq \int_{|y| \leq \frac{\Delta^*}{2A}} (\frac{\Delta^*}{2} - Ay) f_{Z_T}(y) dy - \int_{|y| > \frac{\Delta^*}{2A}} \Delta^* f_{Z_T}(y) dy \\
&= \frac{\Delta^*}{2} \int_{|y| \leq \frac{\Delta^*}{2A}} f_{Z_T}(y) dy - A \underbrace{\int_{|y| \leq \frac{\Delta^*}{2A}} y f_{Z_T}(y) dy}_{=0 \text{ by symmetry}} - \Delta^* \int_{|y| > \frac{\Delta^*}{2A}} f_{Z_T}(y) dy \\
&= \frac{\Delta^*}{2} Pr(|Z_T| \leq \frac{\Delta^*}{2A}) - \Delta^* Pr(|Z_T| > \frac{\Delta^*}{2A}) \\
&= \frac{\Delta^*}{2}\left(1 - 3Pr\left(|Z_T| > \frac{\Delta^*}{2A}\right)\right).
\end{aligned}$$

And since

$$\Delta_T^* \geq \Delta_T(x_0 + \frac{\Delta^*}{2A}) \geq \frac{\Delta^*}{2}\left(1 - 3Pr\left(|Z_T| > \frac{\Delta^*}{2A}\right)\right).$$

We obtain

$$\Delta^* \leq 2\Delta_T^* + 3\Delta^* Pr\left(|Z_T| > \frac{\Delta^*}{2A}\right).$$

By recalling that we require to show that

$$\Delta^* \leq 2\Delta_T^* + \frac{24A}{\pi T}.$$

It only remains to show that

$$3\Delta^* Pr\left(|Z_T| > \frac{\Delta^*}{2A}\right) \leq \frac{24A}{\pi T}.$$

Now

$$\begin{aligned}
Pr\left(|Z_T| > \frac{\Delta^*}{2A}\right) &= \int_{|x| > \frac{\Delta^*}{2A}} f_{Z_T}(x) dx \\
&= 2\int_{\frac{\Delta^*}{2A}}^{\infty} \frac{1 - \cos(xT)}{\pi T x^2} dx \\
&= \frac{2}{\pi} \int_{\frac{\Delta^* T}{4A}}^{\infty} \frac{\sin^2 u}{u^2} du
\end{aligned}$$

$$\leq \frac{2}{\pi}\int_{\frac{\Delta^*T}{4A}}^{\infty}\frac{1}{u^2}du$$
$$= \frac{8A}{\pi\Delta^*T}$$

and thus the result has been proven. (The cases for (2), (3) and (4) are similar).

7.4 Tables for the Accuracy of the Normal Approximation to the Binomial and Poisson Distribution

7.4.1 Poisson Distribution

λ	$\frac{C}{\sqrt{\lambda}}$	λ	$\frac{C}{\sqrt{\lambda}}$	λ	$\frac{C}{\sqrt{\lambda}}$
1	0.47840	11	0.14424	25	0.09568
2	0.33828	12	0.13810	30	0.08734
3	0.27620	13	0.13268	35	0.08086
4	0.23920	14	0.12786	40	0.07564
5	0.21395	15	0.12352	45	0.07132
6	0.19531	16	0.11960	50	0.06766
7	0.18082	17	0.11603	55	0.06451
8	0.16914	18	0.11276	60	0.06176
9	0.15947	19	0.10975	65	0.05934
10	0.15128	20	0.10697	100	0.04784

7.4.2 Binomial Distribution

n \ p	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
1	1.30763	0.92463	0.75496	0.58479	0.41351	0.29239	0.23874	0.18493	0.13076
2	0.99814	0.70579	0.57628	0.44638	0.31564	0.22319	0.18223	0.14116	0.09981
3	0.81328	0.57508	0.46955	0.36371	0.25718	0.18185	0.14848	0.11502	0.08133
5	0.69051	0.48826	0.39867	0.30881	0.21836	0.15440	0.12607	0.09765	0.06905
10	0.60549	0.42815	0.34958	0.27078	0.19147	0.13539	0.11055	0.08563	0.06055
20	0.54663	0.38653	0.31560	0.24446	0.17286	0.12223	0.09980	0.07731	0.05466
30	0.50780	0.35907	0.29318	0.22709	0.16058	0.11355	0.09271	0.07181	0.05078
50	0.48562	0.34338	0.28037	0.21718	0.15357	0.10859	0.08866	0.06868	0.04856
100	0.47840	0.33828	0.27620	0.21395	0.15128	0.10697	0.08734	0.06766	0.04784

References

[1] Esseen, CG 1956, ‘A moment inequality with an application to the central limit theorem’, *Scandinavian Actuarial Journal*, no.2, pp.160-170.

[2] Esseen, CG 1942, ‘On the Liapunov limit error in the theory of probability’, *Arkiv fr Matematik, Astronomi och Fysik*, vol.28, pp.1-19.

[3] Gut, A 2013, *Probability: a graduate course*, 2nd edn., Springer Science & Business Media, NY.

[4] Petrov, V 1995, *Limit Theorems of Probability Theory: Sequences of Independent Random Variables.*, Oxford University Press, NY.

[5] Shevtsova, IGE 2007, ‘Sharpening of the upper bound of the absolute constant in the Berry Esseen inequality’, *Theory of Probability & Its Applications*, vol.51, no.3, pp.549-553.

[6] Shiganov, IS 1986, ‘Refinement of the upper bound of a constant in the remainder term of the central limit theorem’, *Journal of Mathematical Sciences*, vol.35, no.3, pp.2545-2550.

[7] Beek, P 1972, ‘An approximation of Fourier methods to the problem of sharpening the Berry-Esseen inequality’, *Probability Theory and Related Fields*, vol.23, no.3, pp.187-196.