Spanos, A. (1999): *Probability Theory and Statistical Inference. Econometric Modeling with Observational Data.* Cambridge University Press.

Spanos, A. (1986): *Statistical Foundations of Econometric Modelling.* Cambridge University Press.

Davidson, J. (2000): *Econometric Theory.* Oxford: Blackwell.

James Bernoulli in 1713 proved the first Law of Large Numbers. This thread was taken up by de Moivre who proved the first Central Limit Theorem in 1718. Laplace in 1812, drew together and extended the previous two limit theorems. Apart from some marginal weakening of the conditions underlying the LLN by Poisson in the 1830s, the next important milestone was the founding of the Russian school of probability by Chebyshev in the 1870s. Chebyshev was the first to recognize the generality of these limit theorems and provided the foundation upon which the other members of this school of thought, his students Markov and Lyapunov, extended the limit theorems to their modern form. The last member of that illustrious school, Kolmogorov, not only improved upon the work of his predecessors but, in 1933, provided probability theory with its modern mathematical foundations. Let us see how the story of limit theorems unfolds in some more detail.

The first limit theorem was proved by **James Bernoulli** (1654–1705) in his book *Ars Conjectandi*, published posthumously in 1713. Bernoulli revealed his views about the importance of the theorem by calling it *the golden theorem*; today it is known as the *Law of Large Numbers* (LLN), a term first introduced by Poisson in 1837. According to Bernoulli's LLN:

if we toss a fair coin n times and it falls k times heads (H's), then, by increasing the number of tosses the probability of the event $\left\{\left|\left(\frac{k}{n}\right)-\frac{1}{2}\right|<\varepsilon\right\}$ goes to one.

As a prelude to the discussion that follows we make it clear at the outset that when we refer to a sequence of random variables $\{X_n\}_{n=1}^{\infty}:=\{X_1,X_2,\ldots,X_n,\ldots\}$ we are in effect talking about a stochastic process as defined in chapter 8. The reader is strongly advised to refer back to chapter 8 for a number of concepts used in this chapter.

The WLLN, in its general form, can be stated crudely as saying that *under certain restrictions* on the sequence of random variables $\{X_n\}_{n=1}^{\infty}$:

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\tfrac{1}{n}\sum_{k=1}^{n}X_k-\tfrac{1}{n}\sum_{k=1}^{n}E(X_k)\right|<\varepsilon\right)=1,\ \text{for any } \varepsilon>0. \tag{9.9}$$

9.3.1 Bernoulli's WLLN

In an attempt to bring out the gradual weakening of the conditions giving rise to the WLLN let us begin with the general form of Bernoulli's WLLN; see Bernoulli (1713).

Bernoulli's WLLN Let $\{X_n\}_{n=1}^{\infty}:=\{X_1,X_2,\ldots,X_n,\ldots\}$ be a sequence of random variables which satisfy the following conditions:

(D)	Bernoulli:	$f(x_k;\theta_k)=\theta_k^{x_k}(1-\theta_k)^{1-x_k},\ x_k=0,1,\ k=1,2,\ldots$
(M)	Independence:	$f(x_1,x_2,\ldots,x_n;\boldsymbol{\varphi})=\prod_{k=1}^{n}f(x_k;\theta_k),$
(H)	Identical Distribution:	$\theta_k=\theta$, for all $k=1,2,\ldots,$

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\tfrac{1}{n}\sum_{k=1}^{n}X_k-\theta\right|<\varepsilon\right)=1,\ \text{for any } \varepsilon>0, \tag{9.10}$$

and denoted by:

$$\tfrac{1}{n}\sum_{k=1}^{n}X_k\overset{\mathbb{P}}{\to}\theta.$$

The first condition to be weakened was that of Identical Distribution (*complete homogeneity*) when **Simeon Denis Poisson** (1781–1840) proved in 1837 that (iii) could be relaxed without affecting the result.

Poisson's WLLN Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables which satisfy the following conditions:

(D)	Bernoulli:	$f(x_k;\theta_k)=\theta_k^{x_k}(1-\theta_k)^{1-x_k},\ x_k=0,1,\ k=1,2,\ldots$
(M)	Independence:	$f(x_1,x_2,\ldots,x_n;\boldsymbol{\varphi})=\prod_{k=1}^{n}f(x_k;\theta_k),$
(H)	Heterogeneity:	$\theta_i\neq\theta_j,\ i,j=1,2,\ldots$

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\tfrac{1}{n}\sum_{k=1}^{n}X_k-\tfrac{1}{n}\sum_{i=1}^{n}\theta_k\right|<\varepsilon\right)=1,\ \text{for any } \varepsilon>0. \tag{9.12}$$

The first general (in its modern form) WLLN was proved by Chebyshev (1821–1884), the founder of the Russian school of thought which included Markov (1856–1922), Lyapunov (1857–1918), and Kolmogorov (1903–1989). This school of thought had a profound effect on probability theory.

In addition to the complete homogeneity relaxed by Poisson, Chebyshev noticed that when using the inequality bearing his name to prove the WLLN:

(a) the Bernoulli distributed assumption seemed totally unnecessary; it is only role in the above proof was in deriving the mean and variance of $\frac{1}{n}\sum_{k=1}^{n}X_k$,
(b) the Independence assumption was unnecessarily restrictive; its only role is in ensuring that the variance of the sum is equal to the sum of the individual variances. In the case of dependence:

$$Var\left(\tfrac{1}{n}\sum_{i=1}^{n}X_i\right)=\tfrac{1}{n^2}\left[\sum_{i=1}^{n}Var(X_i)+\sum\sum_{i\neq j}Cov(X_i,X_j)\right]. \tag{9.13}$$

For the last term to be zero, however, one does not need to assume complete independence; non-correlation will suffice. Chebyshev in 1867 went on to impose the somewhat stronger dependence restriction of *pairwise independence*, because the difference between the latter condition and non-correlation was not very clear at the time.

Chebyshev's WLLN Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables which satisfy the following conditions:

(D) Bounded moments: $E(X_k)<\infty,\ Var(X_k)<c<\infty,\ k=1,2,\ldots$
(M) Pairwise independence: $f(x_i,x_j;\boldsymbol{\varphi})=f_i(x_i;\boldsymbol{\theta}_i)\cdot f_j(x_j;\boldsymbol{\theta}_j),\ i\neq j,\ i,j=1,2,\ldots$
(H) Heterogeneity: $E(X_k)=\mu_k,\ Var(X_k)=\sigma_k^2,\ k=1,2,\ldots$

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\tfrac{1}{n}\sum_{k=1}^{n}X_k-\tfrac{1}{n}\sum_{k=1}^{n}E(X_k)\right|<\varepsilon\right)=1,\ \text{for any } \varepsilon>0. \tag{9.14}$$

Andrei Markov, a student of Chebyshev, was the first to exploit in full the opportunities offered by the proof of the WLLN using Chebyshev's inequality in order the relax the assumptions giving rise to the result. He saw that even the non-correlation was too restrictive.

Markov's LLN Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables which satisfy the following conditions:

(D)	Bounded moments:	$E(X_k)<\infty,\ Var(X_k)<c<\infty,\ k=1,2,\ldots$
(M)	Asymptotic non-correlation:	$\left(\frac{1}{n^2}\right)Var\left(\sum_{k=1}^{n}X_k\right)\underset{n\to\infty}{\to}0,$
(H)	Heterogeneity:	$E(X_k)=\mu_k,\ Var(X_k)=\sigma_k^2,\ k=1,2,\ldots$

Then (9.14) holds. Condition (M) is called asymptotic non-correlation because in view of (9.13), it holds only if:

$$\tfrac{1}{n^2}\left[\sum\sum_{i\neq j}Cov(X_i,X_j)\right]\underset{n\to\infty}{\to}0.$$

The discerning reader would have noticed that, in addition to the gradual weakening of the initial conditions used by James Bernoulli, the above theorems also show a trade off between the restrictiveness of the three types of conditions. For instance Poisson, by retaining the Bernoulli assumption, was able to relax the complete homogeneity condition to asymptotic homogeneity. This trade off is made in Khintchine's WLLN, proved in 1928 by retaining the IID assumptions, we can relax the boundedness of the variance; we do not need to assume a finite variance.

Khintchine's WLLN
Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables which satisfy the following conditions:

(D)	Bounded mean:	$E(X_k)=\mu<\infty,\ k=1,2,\ldots$
(M)	Independence:	$f(x_1,x_2,\ldots,x_n;\boldsymbol{\varphi})=\prod_{k=1}^{n}f_k(x_k;\boldsymbol{\theta}_k),(x_1,x_2,\ldots,x_n)\in\mathbb{R}^n,$
(H)	Identical Distribution:	$f_k(x_k;\boldsymbol{\theta}_k)=f(x_k;\boldsymbol{\theta})$, for all $k=1,2,\ldots$

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\tfrac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|<\varepsilon\right)=1,\ \text{for any } \varepsilon>0. \tag{9.17}$$

9.3 The central limit theorem

As with the WLLN and SLLN, it was realised that LT2 was not contributing in any essential way to the De Moivre–Laplace theorem and the literature considered sequences of r.v.'s with restrictions on the first few moments. Let $\{X_n, n\geqslant 1\}$ be a sequence of r.v.'s and $S_n=\sum_{i=1}^{n}X_i$, the CLT considers the limiting behaviour of

$$Y_n=\frac{S_n-E(S_n)}{\sqrt{[\mathrm{Var}(S_n)]}}, \tag{9.27}$$

which is a normalised version of $S_n-E(S_n)$, the subject matter of the WLLN and SLLN.

Lindeberg–Levy theorem

Let $\{X_n,\ n\geqslant 1\}$ *be a sequence of IID r.v.'s such that* $E(X_i)=\mu$, $\mathrm{Var}(X_i)=\sigma^2<\infty$ *for all i. Then for* $F_n(y)$ *the DF of* Y_n,

$$\lim_{n\to\infty}F_n(y)=\lim_{n\to\infty}P(Y_n\leqslant y)=\int_{-\infty}^{y}\frac{1}{\sqrt{(2\pi)}}\exp\{-\tfrac{1}{2}u^2\}\,du. \tag{9.28}$$

Liapunov's theorem

Let $\{X_n,\ n\geqslant 1\}$ *be a sequence of independent r.v.'s with*

$$E(X_i)=\mu_i,\quad \mathrm{Var}(X_i)=\sigma_i^2<\infty,\quad E(|X_i|^{2+\delta})<\infty,\quad \delta>0.$$

9.6.3 Lindeberg–Feller's CLT

The most well known Central Limit Theorem is known as the Lindeberg–Feller theorem. This theorem assumes the existence of the second moment and provides both necessary (proposed by Feller in 1935) as well as sufficient conditions (Lindeberg (1922)).

Let $\{X_n, n\geqslant 1\}$ *be a sequence of independent r.v.'s with distribution functions* $\{F_n(x), n\geqslant 1\}$ *such that*

$$\left.\begin{array}{l}(i)\ \ E(X_i)=\mu_i\\ (ii)\ \ \mathrm{Var}(X_i)=\sigma_i^2<\infty,\quad i=1,2,\ldots\end{array}\right\}. \tag{9.31}$$

This chapter introduces some fundamental theory for the treatment of dynamic regression models. Although some of the material is quite general, we structure the analysis around the simplest possible dynamic model, the first-order autoregression. Consider

$$x_t=\lambda x_{t-1}+u_t \tag{6.1.1}$$

where x_t is a scalar random variable. The intercept is omitted here solely for simplicity, but the omission implies that the process has a mean of zero if u_t does.

By an elementary substitution, write

$$\sqrt{n}(\hat{\lambda}-\lambda)=\frac{n^{-1/2}\sum_{t=2}^{n}x_{t-1}u_t}{n^{-1}\sum_{t=2}^{n}x_{t-1}^2}. \tag{6.1.6}$$

Focusing attention on the numerator on the right-hand side, notice that $E(x_{t-1}u_t)=0$ by construction, and this term also has a finite, albeit unknown, variance.

Consider (6.1.6) once again. Since the probability limit of $n^{-1}\sum_{t=2}^{n}x_{t-1}^2$ has been shown to be a positive constant, asymptotic normality of $\sqrt{n}(\hat{\lambda}-\lambda)$ follows from asymptotic normality of $n^{-1/2}\sum_{t=2}^{n}x_{t-1}u_t$ by Cramér's Theorem (Theorem 3.3.5)

The terms $x_{t-1}u_t$ are uncorrelated but not independent, since x_{t-1} depends on u_{t-j} for all $j>0$. Appeal must therefore be made to a central limit theorem for dependent processes. The original result was proved in a famous paper by Mann and Wald (1943b) who showed that the Liapunov central limit theorem could be applied in an adapted form, given the asymptotic independence of the terms $x_{t-1}u_t$. However, the martingale approach is neater and simpler.