2.5. THE CENTRAL-LIMIT AND ITERATED-LOGARITHM THEOREMS

The DeMoivre-Laplace limit theorem of Exercise 2.5 can be generalized very broadly, to yield one of the most important, indeed "central", results in the Theory of Probability. We shall state and discuss this generalization in the present section, but defer its proof to section 3.3, after we have studied questions of Convergence of Probability Measures (section 3.2) as well as the properties of Fourier Transforms (section 3.1).

In section 9 of the present chapter we shall present a different, and very elementary and succinct, approach to the Central Limit Theorem, based on A.V. Skorohod's idea of 'embedding' a Random Walk with zero mean and unit variance in the Brownian Motion process; see subsection 9.C on this *Skorohod Embedding*.

5.1 THEOREM : CENTRAL LIMIT THEOREM. *Let* $X_1, X_2, \cdots$ *be I.I.D. random variables with* $\mathbf{E}(|X_1|^2) < \infty$. *If* $m := \mathbf{E}(X_1)$, $\sigma := \sqrt{Var(X_1)} > 0$ *are the expectation and standard deviation of the underlying commom distribution* F, *and* $S_n = \sum_{k=1}^n X_k$, *then in the notation of (2.7) we have as* $n \to \infty$:

$$\mathbf{P}\left[a \le \frac{S_n - nm}{\sigma\sqrt{n}} \le b\right] \longrightarrow \Phi(b) - \Phi(a) = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx \tag{5.1}$$

for $-\infty \le a < b \le \infty$. *In other words, the sequence* $\frac{1}{\sigma\sqrt{n}} \sum_{k=1}^n \big(X_k - m\big)$, $n \in \mathbf{N}$ *converges vaguely to a standard normal (Gaussian) random variable.*

• *Heuristic Discussion: Small Deviations in the Law of Large Numbers.* From the Strong Law of Large Numbers we know that, for n large, we have $\overline{X}_n := (S_n / n) \sim m = \mathbf{E}(X_1)$, and we are led to ask the question: *What is the order of "small fluctuations" (deviations) of the "sample average"* $\overline{X}_n$ *around the "ensemble average"* m*?* In other words, can we find a positive sequence $f(n) \downarrow 0$ (as $n \to \infty$) such that, for all n enough, we have $|\overline{X}_n - m| = O(1) \cdot f(n)$ almost everywhere, or equivalently

$$\mathbf{P}\Big(|\overline{X}_n - m| \le K f(n)\Big) = 1, \qquad \text{for some } K \in (0,\infty)\,?$$

Requiring that this probability be equal to one, turns out to be too much to ask for; nevertheless, according to the Central Limit Theorem

$$\mathbf{P}\left(\frac{\overline{X}_n - m}{f(n)} \in A\right) \sim \int_A \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx\,, \qquad A \in \mathcal{B}(\mathbf{R})\,, \tag{5.2}$$

we can manage to *make this probability very high,* provided that we choose the order of "small deviations" to be

$$f(n) \,=\, \sigma \,/\, \sqrt{n}\,. \tag{5.3}$$

Indeed, with $A = [-K, K]$, the probability of (5.2) is approximately $2\,\Phi(K) - 1$ for n large; for instance, this quantity equals 98.76% for $K = 2.5$, and 99.74% for $K = 3$. These numbers are readily available from the ubiquitous tables of the standard normal distribution.

It is rather remarkable that in all of this, only two characteristics from the underlying common distribution of $X_1, X_2, \cdots$ matter at all: the expectation $m = \mathbf{E}(X_1)$ and the variance $\sigma^2 = \mathrm{Var}(X_1)$. This feature makes the Central Limit Theorem such an important tool in applications.

The rate of convergence in the relation (5.1) of the Central Limit Theorem is quite slow, namely of the order $(1/\sqrt{n}\,)$. This can be seen most clearly by considering the *symmetric Bernoulli distribution* $\mathbf{P}[X_1 = \pm 1] = 1/2$ in Theorem 5.1; in this case it is straightforward to see that

$$\big|\mathbf{P}(S_{2n} < 0) - \Phi(0)\big| \,=\, \frac{1}{2}\,\mathbf{P}(S_{2n} = 0) \,=\, \frac{(2n)!}{2\,(n!)^2}\left(\frac{1}{2}\right)^{2n} \sim\, \frac{1}{\sqrt{2\pi\,(2n)}}\ , \tag{5.2}$$

using the Stirling formula of Exercise 2.5. The following result shows that this order of magnitude is typical.

5.2 Theorem : Berry-Esseen. *With the same assumptions and notation as in Theorem 5.1, and with the additional condition* $\mathbf{E}(|X_1|^3) < \infty$, *we have*

$$\sup_{x \in \mathbf{R}} \left| \mathbf{P}\left[\frac{S_n - nm}{\sigma\,\sqrt{n}} \le x\right] - \Phi(x)\right| \,\le\, \frac{C \cdot \mathbf{E}(|X_1|^3)}{\sigma^3\,\sqrt{n}}\,, \quad \forall\, n \in \mathbf{N} \tag{5.3}$$

for some universal (that is, not depending on the distribution of X_1*) real constant* $C > 0$.

The example of (5.2) makes clear that $C \ge (1\,/\,\sqrt{2\pi}\,)$, and it can be shown that $C < 0.8$. For a proof of this result, see for instance Bolthausen (1984).

We have also the following multi-dimensional version of the Central Limit Theorem; recall Example 2.3 and Exercise 4.7.

5.3 Theorem : Multivariate Central Limit Theorem. *Let* $X^{(1)} = \big(X_1^{(1)}, \cdots, X_d^{(1)}\big)'$, $X^{(2)} = \big(X_1^{(2)}, \cdots, X_d^{(2)}\big)', \cdots$ *be I.I.D. random vectors with* $\mathbf{E}(||X^{(1)}||^2) < \infty$. *Denote by* $(m_1, \cdots, m_d)'$ *the vector of expectations* $m_i := \mathbf{E}\big(X^{(1)_i}\big)$, $i = 1, \cdots, d$ *and* *by* $\Sigma = \{\Sigma_{ij}\}_{1\le i,j\le d}$ *the (symmetric, non-singular) variance/covariance matrix with* $\Sigma_{ij} := \mathbf{E}\big[\big(X_i^{(1)} - m_i\big)\,\big(X_j^{(1)} - m_j\big)\big]$ *. Then the distribution of the random vector*

$$\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \left(X_i^{(k)} - m_i \right) , \cdots , \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \left(X_i^{(k)} - m_i \right) \right)'$$

converges vaguely as $n \to \infty$ *to a multivariate normal (Gaussian) distribution with mean vector zero, and variance/covariance matrix* Σ *.*

A: LAW OF THE ITERATED LOGARITHM.

Another celebrated result of classical Probability Theory is the following *Law of the Iterated Logarithm* – the last in a long series of efforts culminating with Hartman & Wintner (1941). This is a deep result that provides precise information about the growth of the zero-mean, finite-variance Random Walk $S_n = \sum_{k=1}^n X_k$ on the real line. We shall not prove this result here; but in Section 4.4 we shall use results of martingale theory to provide a rather elementary argument for the simple case of X_k's with Gaussian distribution.

This same proof establishes the Law of the Iterated Logarithm for the Brownian Motion process – the continuous analogue of the random walk. Then a rather simple argument, based once again on the Skorohod embedding of Random Walk into the Brownian Motion process, provides the proof of Theorem 5.4 as well; see subsection 9.C of the present chapter for the detail.

5.4 THEOREM : LAW OF THE ITERATED LOGARITHM. *Let* $X_1, X_2, \cdots$ *be I.I.D. random variables with* $\sigma = \sqrt{\mathbf{E}(|X_1|^2)} < \infty$ *and* $\mathbf{E}(X_1) = 0$*, and consider the Random Walk* $S_n = \sum_{k=1}^n X_k$ *for* $n \in \mathbf{N}$ *with* $S_0 = 0$ *. Then we have*

$$\limsup_{n\to\infty} \left(\frac{S_n}{\sigma\sqrt{2\,n\,\log\log n}} \right) = 1\,, \qquad \liminf_{n\to\infty} \left(\frac{S_n}{\sigma\sqrt{2\,n\,\log\log n}} \right) = -1\,. \tag{5.4}$$

Historical Discussion: Suppose that we have a sequence $X_1, X_2, \cdots$ of *I.I.D.* random variables with as in Theorem 5.1 with $m = \mathbf{E}(X_1) = 0$. Then we know from the Strong Law of Large Numbers that the Random Walk $S_n = \sum_{j=1}^n X_j$, $n \in \mathbf{N}$ "grows slower than n", since $\lim_{n\to\infty}(S_n\,/\,n) = 0$ holds a.e. On the other hand, at least in the case of the simple, symmetric Random Walk with $\mathbf{P}(X_1 = \pm 1) = 1/2$, it is not hard to show that

$$\overline{\lim}_{n\to\infty} \frac{S_n}{\sqrt{n}} = \infty\,, \qquad \underline{\lim}_{n\to\infty} \frac{S_n}{\sqrt{n}} = -\infty \qquad \text{and} \qquad \lim_{n\to\infty} \left(\frac{S_n}{\sqrt{n}\,\log n} \right) = 0\,, \quad \text{a.e.}$$

(see, for instance, Theorem 4.2.8 and its corollaries). In the same vein, Haussdorff (1910) proved that

$$\overline{\lim}_{n\to\infty}\left(\frac{|S_n|}{\sqrt{n^{1+\varepsilon}}}\right) \qquad \text{is a.e. finite, for any } \varepsilon > 0\,,$$

provided that X_1 has moments of all orders; and Hardy & Littlewood (1915) showed that

$$\overline{\lim}_{n\to\infty}\left(\frac{|S_n|}{\sqrt{n\,\log n}}\right) \qquad \text{is a.e. finite }\,,$$

when $||X_1||_\infty < \infty$. The *exact* growth behavior of the order $\sqrt{2\,n\,\log\log n}$ as in (5.4) was first obtained by Ȟinčin (1924) when $||X_1||_\infty < \infty$. Under the general conditions of Theorem 5.4, this growth behavior was establihed by Hartman & Wintner (1941).

5.1 Exercise : Show $\lim_{n\to\infty} e^{-n}\sum_{k=1}^n n^k\,/\,k! \,=\, 1/2$.

5.2 Exercise : Gamma Distribution. If $X_1, X_2, \cdots$ are *I.I.D.* random variables with common

(a) Poisson distribution $\mathbf{P}[X_1 \,=\, k] \,=\, e^{-\lambda}\left(\lambda^k\,/\,k!\right)$, $k \,=\, 0, 1, \cdots$ with parameter $\lambda > 0$, then $S_n = \sum_{j=1}^n X_j$ has also Poisson distribution, with parameter $n\lambda$;

(b) exponential distribution $\lambda\, e^{-\lambda\, x}\,\chi_{(0,\infty)}(x)\,dx$, then $S_n = \sum_{j=1}^n X_j$ has the *Gamma distribution* $(\lambda^n/\Gamma(n))\,x^{n-1}\,e^{-\lambda\,x}\,\chi_{(0,\infty)}(x)\,dx$.

(*Hint:* Use Exercise 2.3 and induction.)

5.3 Exercise : Maxwell Distribution. Let $Y = \sqrt{X_1^2 + X_2^2 + X_3^2}$ denote the magnitude of the velocity vector $(X_1, X_2, X_3)'$ of a gas-molecule with mass M at temperature T. According to the kinetic theory of gases, Y obeys the so-called *Maxwell distribution*, whose probability density function is

$$f(y) \;=\; \frac{4y^2}{a^3\sqrt{\pi}}\cdot \exp\left\{-\frac{y^2}{a^2}\right\}\,, \quad 0 < y < \infty\,.$$

Here $a = \sqrt{2kT/M}$ and k is the Boltzmann constant.

(i) Show that the kinetic energy $Z = \frac{1}{2}MY^2$ has a Gamma distribution (Exercise 5.1(b)) with $\lambda = 1/(kT)$, $n = 3/2$.

(ii) If the components of the velocity vector (X_1, X_2, X_3) are independent normal random variables, with zero expectation and the same variance, show that Y has indeed the Maxwell distribution.

5.4 Exercise : Chi-Square Distribution. Suppose that $X_1, X_2, \cdots$ are independent normal random variables $\mathbf{P}[X_j \,\le\, x] \,=\, \Phi\left((x - m_j)\,/\,\sigma_j\right)$, $x \,\in\, \mathbf{R}$ with expectation $m_j \in \mathbf{R}$ and standard deviation $\sigma_j > 0$, for every $j = 1, \cdots, n$. Show that

(a) $(1/\sqrt{n})\,\sum_{j=1}^n (X_j - m_j)\,/\,\sigma_j$ has standard normal distribution.

(b) $\sum_{j=1}^{n}(X_j - m_j)^2 / \sigma_j^2$ has a χ_n^2–distribution (*"chi-square" with n degrees of freedom*), whose probability density function is

$$\frac{2^{-n/2}}{\Gamma(n/2)} \cdot x^{(n/2)-1} e^{-x/2}, \qquad x > 0 .$$

In particular, if $m_j \equiv m$ and $\sigma_j \equiv \sigma$ for all $j = 1, \cdots, n$, then:

(c) the random variable $\sum_{j=1}^{n}((X_j - m) / \sigma\sqrt{n}) = \sqrt{n}(\overline{X}_n - m) / \sigma$ has standard normal distribution; and the statement (5.1) of the Central Limit Theorem holds exactly for each $n \in \mathbf{N}$, not just in the limit as $n \to \infty$.

(d) the random variable $\sum_{j=1}^{n}(X_j - m)^2 / \sigma^2$ has a χ_n^2–distribution; whereas the random variable $\sum_{j=1}^{n}(X_j - \overline{X}_n)^2 / \sigma^2$ has a χ_{n-1}^2–distribution.

5.5 Exercise : Confidence Intervals. With the assumptions and notation of Theorem 5.1:

(a) Show that $\mathcal{S}_n := \sqrt{(1/(n-1)) \sum_{j=1}^{n}(X_j - \overline{X}_n)^2} \longrightarrow \sigma$ as $n \to \infty$, $\mathbf{P}$–a.e.

Suppose now that m is an unknown constant, that we wish to estimate on the basis of the observed random X's.

(b) Assume that σ is known. Show that, for n large, we have

$$\mathbf{P}\left[\overline{X}_n - (0.96)\frac{\sigma}{\sqrt{n}} \leq m \leq \overline{X}_n + (0.96)\frac{\sigma}{\sqrt{n}}\right] \sim 95\%,$$

and that this statement is *exact, for all* $n \in \mathbf{N}$, if F is normal. We say then, that $\left[\overline{X}_n - (0.96)\frac{\sigma}{\sqrt{n}}, \overline{X}_n + (0.96)\frac{\sigma}{\sqrt{n}}\right]$ is an approximate *confidence interval for* m, *in the case of known variance,* with coëfficient 95%.

(c) Assume that σ is unknown. Show that, for n large, we have

$$\mathbf{P}\left[\overline{X}_n - (0.96)\frac{\mathcal{S}_n}{\sqrt{n}} \leq m \leq \overline{X}_n + (0.96)\frac{\mathcal{S}_n}{\sqrt{n}}\right] \sim 95\%.$$

Again, $\left[\overline{X}_n - (0.96)\frac{\mathcal{S}_n}{\sqrt{n}}, \overline{X}_n + (0.96)\frac{\mathcal{S}_n}{\sqrt{n}}\right]$ is called an approximate *confidence interval for* m, *in the case of unknown variance*, with coëfficient 95%.

5.6 Exercise: Let $X_1, X_2, \cdots$ be a sequence of independent random variables, with common distribution $F(\cdot)$ that satisfies $\lim_{x\to\infty} e^x [1 - F(x)] = b > 0$.
With $M_n := \max_{1\leq k\leq n} X_k$, show that

$$\lim_{n\to\infty} \mathbf{P}[M_n - \log(bn) \leq y] = G(y), \quad \forall\, y \in \mathbf{R}$$

holds, for an appropriate distribution function $G(\cdot)$. (*Hint:* Start with the special case of the 'logistic distribution' $F(x) = 1/(1 + e^{-x})$.)