16

Limiting Distributions

In this chapter we consider the limiting distribution of various quantities, the simplest being that of a sequence of real numbers considered modulo 1. We find that the limiting distribution modulo 1 of a sequence $\{u_n\}$ can be described in terms of the asymptotic size of the associated exponential sums $\sum_{n=1}^{N} e(ku_n)$. Here k runs over integral values, and $e(\theta) = e^{2\pi i\theta}$ is the complex exponential with period 1. This motivates us, in the next chapter, to develop methods for estimating exponential sums.

16.1 Uniform distribution (mod 1)

Let $u_1, u_2, \ldots$ be a sequence of real numbers, and for $0 \le \alpha \le 1$ let $Z(N, \alpha)$ denote the number of n, $1 \le n \le N$, such that $0 \le u_n \le \alpha \pmod 1$. We say that the sequence $\{u_n\}$ is *uniformly distributed* (mod 1) if

$$\lim_{N\to\infty} \frac{1}{N} Z(N, \alpha) = \alpha \tag{16.1}$$

for all $\alpha \in [0, 1]$. To characterize uniformly distributed sequences we have

Theorem 16.1 (Weyl's criterion) *The following are equivalent*:
(a) *The sequence* $\{u_n\}$ *is uniformly distributed*;
(b) *For every integer* $k \ne 0$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e(ku_n) = 0;$$

(c) *For each function* f *with period* 1 *that is properly Riemann-integrable on* $[0, 1]$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(u_n) = \int_0^1 f(x)\, dx. \tag{16.2}$$

Proof We note that (a) is equivalent to the assertion that (16.2) holds whenever f is the characteristic function $\chi_{\mathcal{I}}$ of the interval $\mathcal{I} = [0, \alpha] \pmod 1$ where $0 \le \alpha \le 1$. Similarly, (b) is equivalent to the assertion that (16.2) holds when $f(x) = e(kx)$ all integers k (including

$k = 0$, since (16.2) holds trivially when $f \equiv 1$). Moreover, the relation (16.2) is linear, so that if it holds for several functions, then it holds for any linear combination of (finitely many of) them. Hence (a) asserts that (16.2) holds for step functions with period 1, and (b) asserts that (16.2) holds for trigonometric polynomials with period 1. We complete the proof by showing that (c) $\Rightarrow$ (b) $\Rightarrow$ (a) $\Rightarrow$ (c).

The implication (c) $\Rightarrow$ (b) is trivial, since trigonometric polynomials are Riemann-integrable.

To show that (b) implies (a) we appeal to a theorem of Weierstrass that asserts that if f is continuous with period 1 and $\varepsilon > 0$ is given, then there exists a trigonometric polynomial

$$T(x) = \sum_{k=-K}^{K} a_k e(kx)$$

such that $|f(x) - T(x)| < \varepsilon$ for all x. Of course this does not apply directly to the characteristic function $\chi_{\mathcal{I}}(x)$ of $\mathcal{I}$, since $\chi_{\mathcal{I}}$ has jump discontinuities. However, we can approximate to $\chi_{\mathcal{I}}$ by continuous functions. For example, let $L_+(x)$ be the piecewise linear function with period 1 whose graph has the vertices $(0, 1+\varepsilon), (\alpha, 1+\varepsilon), (\alpha+\varepsilon, \varepsilon), (1-\varepsilon, \varepsilon), (1, 1+\varepsilon)$, and similarly let $L_-(x)$ be the piecewise linear function with period 1 whose graph has the vertices $(0, -\varepsilon), (\varepsilon, 1-\varepsilon), (\alpha-\varepsilon, 1-\varepsilon), (\alpha, -\varepsilon), (1, -\varepsilon)$. (We may suppose that $0 < \alpha < 1$ and that ε is so small that $2\varepsilon \le \alpha \le 1 - 2\varepsilon$.) Then the $L_\pm$ are continuous, $L_-(x) + \varepsilon \le \chi_{\mathcal{I}}(x) \le L_+(x) - \varepsilon$ for all x, and the $L_\pm$ are good approximations to $\chi_{\mathcal{I}}$ in the L^1-norm, since $\int_0^1 L_\pm(x)\,dx = \alpha \pm 2\varepsilon$. By Weierstrass's theorem there exist trigonometric polynomials $T_\pm(x)$ such that $|L_\pm(x) - T_\pm(x)| < \varepsilon$ for all x. Hence $T_-(x) \le \chi_{\mathcal{I}}(x) \le T_+(x)$ for all x, $\int_0^1 T_-(x)\,dx \ge \alpha - 3\varepsilon$, and $\int_0^1 T_+(x)\,dx \le \alpha + 3\varepsilon$. But then

$$Z(N, \alpha) = \sum_{n=1}^{N} \chi_{\mathcal{I}}(u_n) \le \sum_{n=1}^{N} T_+(u_n),$$

and by the hypothesis (b) we know that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T_+(u_n) = \int_0^1 T_+(x)\,dx \le \alpha + 3\varepsilon,$$

so it follows that $\limsup_{N\to\infty} Z(N, \alpha)/N \le \alpha + 3\varepsilon$. By arguing similarly with $T_-(x)$, we see that $\liminf_{N\to\infty} Z(N, \alpha)/N \ge \alpha - 3\varepsilon$. Since ε may be taken arbitrarily small, we have (a).

Finally we show that (a) implies (c); our method is the same as the one just completed. If $f(x)$ is properly Riemann-integrable on $[0, 1]$, then for any $\varepsilon > 0$ there exist step functions $S_\pm(x)$ such that $S_-(x) \le f(x) \le S_+(x)$, $\int_0^1 f(x) - S_-(x)\,dx < \varepsilon$, and $\int_0^1 S_+(x) - f(x)\,dx < \varepsilon$. By proceeding as above, but with $\chi_{\mathcal{I}}$ replaced by f and $T_\pm$ replaced by $S_\pm$, we see that

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(u_n) \le \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} S_+(u_n) = \int_0^1 S_+(x)\,dx < \int_0^1 f(x)\,dx + \varepsilon,$$

and similarly for the lim inf. Hence we see that (a) implies (c), and the proof is complete. □

When we consider a real number x (mod 1), or equivalently the fractional part $\{x\}$ of x, we are treating x as a representative of a member of the circle group $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. Similarly, a function with period 1 may be thought of as having domain $\mathbb{T}$. Thus Weyl's criterion can be considered to be a statement concerning the distribution of points $u_1, u_2, \ldots$ in $\mathbb{T}$.

We find it fruitful to cast Weyl's criterion in the language of measure theory. We call a measure μ on $\mathbb{T}$ a *probability measure* if both $\mu(\mathcal{S}) \ge 0$ for all measureable sets $\mathcal{S} \subseteq \mathbb{T}$ and also $\mu(\mathbb{T}) = 1$. Let δ, the *Dirac delta*, denote the probability measure that assigns unit mass to the point 0. Thus $\delta(\mathcal{S}) = 1$ or 0 according as $0 \in \mathcal{S}$ or not. The measure δ and also Lebesgue measure λ are examples of probability measures on $\mathbb{T}$. If $u_1, u_2, \ldots$ is a sequence of points in $\mathbb{T}$, then for each N put

$$\mu_N(x) = \frac{1}{N}\sum_{n=1}^{N} \delta(x - u_n). \tag{16.3}$$

Thus μ_N is a probability measure that places mass $1/N$ at each of the points $u_1, u_2, \ldots, u_N$, and hence

$$\int_{\mathbb{T}} f(x)\, d\mu_N = \frac{1}{N}\sum_{n=1}^{N} f(u_n).$$

In general, if μ is a measure on $\mathbb{T}$, then for integers k we define its *Fourier coefficient* $\widehat{\mu}(k)$ to be

$$\widehat{\mu}(k) = \int_{\mathbb{T}} e(-kx)\, d\mu. \tag{16.4}$$

Thus for the special measures μ_N we see that $\widehat{\mu}_N(k) = \frac{1}{N}\sum_{n=1}^{N} e(-ku_n)$. Hence Weyl's criterion asserts that the following assertions are equivalent:

(a) $\mu_N([0,\alpha]) \to \alpha$ as $N \to \infty$ for all $\alpha \in [0,1]$ (i.e., $\mu_N \to \lambda$ weakly);
(b) For each integer $k \ne 0$, $\widehat{\mu}_N(k) \to 0$ as $N \to \infty$;
(c) If f is properly Riemann-integrable on $\mathbb{T}$, then $\int_{\mathbb{T}} f\, d\mu_N \to \int_{\mathbb{T}} f\, d\lambda$ as $N \to \infty$.

Here the restriction to measures of the special shape (16.3) may be dropped, since it is easy to see that the proof of Theorem 16.1 applies equally to any sequence of probability measures $\mu_1, \mu_2, \ldots$. In §5 of this chapter we consider an analogous (but more difficult) theory of convergence of measures on the real line.

As a first application of Weyl's criterion we have

Theorem 16.2 *If θ is irrational, then the numbers $n\theta$ are uniformly distributed* (mod 1).

We note that the converse of the above is obvious.

Proof We first show that for any real α,

$$\Big|\sum_{n=1}^{N} e(n\alpha)\Big| \le \min\Big(N, \frac{1}{2\|\alpha\|}\Big). \tag{16.5}$$

The bound N is obvious by the triangle inequality. In proving the second bound we may assume that α is not an integer. From the formula for the sum of a geometric progression we see that the left hand side above is

$$= \left|\frac{e((N+1)\alpha) - e(\alpha)}{e(\alpha) - 1}\right| \le \frac{2}{|e(\alpha) - 1|} = \frac{1}{|\sin \pi\alpha|} \le \frac{1}{2\|\alpha\|}.$$

On taking $\alpha = k\theta$ in (16.5), we see that

$$\frac{1}{N}\sum_{n=1}^{N} e(kn\theta) \ll \frac{1}{\|k\theta\| N}.$$

Since θ is irrational, it follows that $k\theta$ is not an integer, so that the above is $\ll 1/N$ with an implicit constant that depends on k and on θ. Thus we have part (b) of Theorem 16.1, and hence the sequence $n\theta$ is uniformly distributed (mod 1). □

16.1.1 Exercises

1. Suppose that $\{u_n\}$ is uniformly distributed (mod 1), and let c be a real number. Put $v_n = u_n + c$. Show that $\{v_n\}$ is uniformly distributed.

2. (a) Suppose that $f \in L^1(\mathbb{T})$. Show that for every $\varepsilon > 0$ there is a trigonometric polynomial $T(x)$ such that $\int_{\mathbb{T}} |f(x) - T(x)|\,dx < \varepsilon$.
   (b) Suppose that f is real valued, has period 1, and that for every $\varepsilon > 0$ there exist trigonometric polynomials T_+ and T_- such that $T_-(x) \le f(x) \le T_+(x)$ for all x and $\int_{\mathbb{T}} T_+(x) - T_-(x)\,dx < \varepsilon$. Show that f is Riemann-integrable on $[0,1]$.

3. Suppose that f has period 1 and that $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(u_n)$ exists whenever $\{u_n\}$ is uniformly distributed. Show that f is properly Riemann-integrable.

4. (a) Show that

$$\limsup_{N\to\infty} \frac{1}{N}\,\mathrm{card}\{n : 1 \le n \le N, \{\log n\} \in [0,1/2]\} = \frac{e - e^{1/2}}{e - 1}.$$

   (b) Show that

$$\liminf_{N\to\infty} \frac{1}{N}\,\mathrm{card}\{n : 1 \le n \le N, \{\log n\} \in [0,1/2]\} = \frac{e^{1/2} - 1}{e - 1}.$$

   (c) Show that

$$\frac{1}{N}\sum_{n=1}^{N} e(k\log n) = \frac{N^{2\pi i k}}{2\pi i k + 1} + O\Big((|k| + 1)\frac{\log N}{N}\Big).$$

(d) Show that the sequence $\{\log n\}$ is not uniformly distributed (mod 1).

5. Suppose that $\{u_n\}$ is a sequence such that $\lim_{n\to\infty} u_{n+1} - u_n = \alpha$. Show that if α is irrational, then $\{u_n\}$ is uniformly distributed (mod 1).

6. Let $\mathfrak{I}$ and $\mathfrak{J}$ be arcs of $\mathbb{T}$, and suppose that α is an irrational number. Put $u_n = n\alpha$. Show that for each non-negative integer m the limit

$$d_m = \lim_{N\to\infty} \frac{1}{N} \operatorname{card}\{n \in [1, N] : u_n \in \mathfrak{I},\ u_{n-m} \in \mathfrak{J}\}$$

exists. Prove that

$$\lim_{M\to\infty} \frac{1}{M} \sum_{m=1}^{M} d_m = |\mathfrak{I}||\mathfrak{J}|.$$

16.2 Quantitative estimates

Suppose that a sequence $\{u_n\}$ is given, and let $Z(N, \alpha)$ be defined as in the preceding section. For $0 \le \alpha \le 1$ put

$$D(N, \alpha) = Z(N, \alpha) - N\alpha.$$

The *discrepancy* of the sequence $\{u_n\}$ is the quantity

$$D^*(N) = \sup_{0\le\alpha\le 1} |D(N, \alpha)|.$$

As a companion to (a)–(c) of Theorem 16.1, consider the assertion

(d) $D(N) = o(N)$ as $N \to \infty$.

On one hand, this is equivalent to asserting that (16.1) holds uniformly in α. Hence (d) implies part (a) of Theorem 16.1. To establish the converse, we first observe that $D(N, \alpha)$ has a sort of one-sided Lipschitz property: If $0 \le \alpha \le \beta \le 1$, then

$$D(N, \beta) - D(N; \alpha) \ge -N(\beta - \alpha)$$

because $Z(N, \alpha)$ is an increasing function of α. Hence if $|D^*(N, m/M)| < \varepsilon N$ for $m = 0, 1, 2, \ldots, M$, then $|D(N, \alpha)| < (\varepsilon + 1/M)N$ for all $\alpha \in [0, 1]$. Thus we see that if (16.1) holds everywhere pointwise, then it holds uniformly, and hence the assertion (d) above is equivalent to the assertions of Theorem 16.1.

Since the discrepancy of a sequence provides a measure of the rate at which the limit (16.1) is attained, it is reasonable to ask for quantitative connections between the size of the discrepancy and of the exponential sums considered in part (b) of Theorem 16.1. Such links can be established in both directions, but since we shall shortly be developing methods for estimating exponential sums, the most useful tool is a bound for the discrepancy in terms of exponential sums.

Theorem 16.3 (The Erdős–Turán inequality) *Let $u_1, u_2, \dots, u_N$ be N numbers in $\mathbb{T}$, let $\mathcal{I} = [\alpha, \beta]$ be an arbitrary arc of $\mathbb{T}$ of length $\beta - \alpha \le 1$, and let K be an arbitrary positive integer. Then*

$$|\operatorname{card}\{n \in [1, N] : u_n \in \mathcal{I}\} - (\beta - \alpha)N| \le \frac{N}{K+1} + 3\sum_{k=1}^{K} \frac{1}{k}\Big|\sum_{n=1}^{N} e(ku_n)\Big|.$$

By taking $\alpha = 0$ and allowing β to vary, we may make the left hand side above as close as we like to $D^*(N)$, and hence this inequality provides the desired bound for the discrepancy.

Proof We proceed in the same manner as in the proof of Theorem 16.1, but now we employ quantitative one-sided trigonometric approximations to $\chi_{\mathcal{I}}$ that are sharp in the L^1-norm. Specifically, suppose that $T_-(x)$ is chosen as in Theorem E.3. Then

$$\begin{aligned}
\operatorname{card}\{n \in [1, N] : u_n \in \mathcal{I}\} &= \sum_{n=1}^{N} \chi_{\mathcal{I}}(u_n) \\
&\ge \sum_{n=1}^{N} T_-(u_n) \\
&= \sum_{k=-K}^{K} \widehat{T}_-(k) \sum_{n=1}^{N} e(ku_n).
\end{aligned}$$

By Theorem E.3(b) we know that $\widehat{T}_-(0) = \beta - \alpha - 1/(K+1)$. To estimate the Fourier coefficients $\widehat{T}_-(k)$ for $k \ne 0$, we recall that if f is in $L^1(\mathbb{T})$, then

$$|\widehat{f}(k)| = \Big|\int_{\mathbb{T}} f(x)e(-kx)\,dx\Big| \le \int_{\mathbb{T}} |f(x)|\,dx = \|f\|_{L^1}.$$

Since

$$\widehat{\chi}_{\mathcal{I}}(k) = e(-k(\alpha+\beta)/2)\frac{\sin \pi k(\beta - \alpha)}{\pi k} \qquad (k \ne 0),$$

by taking $f = \chi_{\mathcal{I}} - T_-$ we see by Theorem E.3(b) that

$$\Big|e(-k(\alpha+\beta)/2)\frac{\sin \pi k(\beta - \alpha)}{\pi k} - \widehat{T}_-(k)\Big| \le \frac{1}{K+1} \qquad (k \ne 0),$$

and hence that

$$|\widehat{T}_-(k)| \le \frac{1}{K+1} + \Big|\frac{\sin \pi k(\beta - \alpha)}{\pi k}\Big| \qquad (k \ne 0).$$

Thus

$$\begin{aligned}
\operatorname{card}\{n \in [1, N] : u_n \in \mathcal{I}\} &\ge (\beta - \alpha)N - \frac{N}{K+1} \\
&\quad - 2\sum_{k=1}^{K}\Big(\frac{1}{K+1} + \Big|\frac{\sin \pi k(\beta - \alpha)}{\pi k}\Big|\Big)\Big|\sum_{n=1}^{N} e(ku_n)\Big|.
\end{aligned} \tag{16.6}$$

But $|\sin u| \le 1$ and $2(\frac{1}{K+1}+\frac{1}{\pi k}) \le 2(\frac{1}{k}+\frac{1}{3k}) < 3/k$, so this gives the desired lower bound. The corresponding upper bound is proved similarly, using the $T_+(x)$ from Theorem E.3. $\square$

The Erdős–Turán inequality provides a good estimate for the discrepancy in terms of exponential sums, but for short intervals we can do better.

Theorem 16.4 *Let $u_1, u_2, \dots, u_N$ be given, and suppose that K is an integer such that*

$$\sum_{k=1}^{K}\Big|\sum_{n=1}^{N} e(ku_n)\Big| < N/8.$$

Then any arc $\mathcal{I}=[\alpha,\beta]$ of $\mathbb{T}$ of length $\beta-\alpha \ge 3/(K+1)$ contains at least $\frac{1}{3}(\beta-\alpha)N$ of the points u_n.

Proof Since $|\sin u| \le |u|$, the lower bound in (16.6) is

$$\ge (\beta-\alpha)N - \frac{1}{3}(\beta-\alpha)N - \frac{8}{3}(\beta-\alpha)\sum_{k=1}^{K}\Big|\sum_{n=1}^{N} e(ku_n)\Big|,$$

which gives the result. $\square$

16.2.1 Exercises

1. Let $\mathcal{A}$ be a dense subset of $[0,1]$, and suppose that a sequence $\{u_n\}$ is given. Show that if the relation (16.1) holds for all $\alpha \in \mathcal{A}$, then (16.1) holds for all α, and hence $\{u_n\}$ is uniformly distributed (mod 1).

2. Let p be an odd prime, and put $u_n = n^2/p$ for $1 \le n \le p$.
   (a) Show that

$$\Big|\sum_{n=1}^{p} e(ku_n)\Big| = \begin{cases} p & \text{if } k \equiv 0 \pmod{p}, \\ \sqrt{p} & \text{if } k \not\equiv 0 \pmod{p}. \end{cases}$$

   (b) Show that $D^*(p) \ll p^{1/2}\log p$.
   (c) This is a special case of what familiar inequality?

3. (a) Show that $D^*(N) \ge 1/2$ for any sequence $\{u_n\}$ and any $N \ge 1$.
   (b) Show that if N points $u_1, u_2, \dots, u_N$ are equally spaced (mod 1), then $D^*(N) \le 1$.

4. (a) Suppose that the M points $u_1, u_2, \dots, u_M$ have discrepancy $D^*(M;\boldsymbol{u})$ and the the N points $v_1, v_2, \dots, v_N$ have discrepancy $D^*(N;\boldsymbol{v})$. Let $w_1, w_2, \dots, w_{M+N}$ be the concatenation of these two sequences (i.e., $w_m = u_m$ for $1 \le m \le M$, $w_{M+n} = v_n$ for $1 \le n \le N$), and let $D^*(M+N;\boldsymbol{w})$ be its discrepancy. Show that $D^*(M+N;\boldsymbol{w}) \le D^*(M;\boldsymbol{u}) + D^*(N;\boldsymbol{v})$.
   (b) Show that if $\|u_n - v_n\| \le \delta$ for all n, then $|D^*(N;\boldsymbol{u}) - D^*(N;\boldsymbol{v})| \le \delta N$ for all

$N \ge 1$.
(c) Suppose that $u_n = n\theta + \beta$. Show that if $|\theta - a/q| \le A/q^2$ and $(a, q) = 1$, then $D^*(q) \le A + 1$.
(d) In the remaining parts of this exercise, let θ denote the 'golden ratio', $\theta = (1+\sqrt{5})/2$, and let F_h denote the h^{th} Fibonacci number. Show that $F_h\theta = F_{h+1} + (-1)^{h+1}\theta^{-h}$. Deduce that $|\theta - F_{h+1}/F_h| \le F_h^{-2}$.
(e) Show also that $(F_h, F_{h+1}) = 1$.
(f) Deduce that if $N = F_h$ for some h and if the numbers $u_1, u_2, \ldots, u_N$ are any N consecutive members of the sequence $\{n\theta\}$, then $D^*(N) \le 2$.
(g) Show that any positive integer N may be written in the form $N = F_{h_1} + F_{h_2} + \cdots + F_{h_R}$ where $h_1 > h_2 > \cdots > h_R$.
(h) Show that if $u_n = n\theta$ where θ is the golden ratio, then $D^*(N) \ll \log N$.
(i) Show that

$$\sum_{k=1}^{F_h} \frac{1}{\|k\theta\|} \asymp F_h \log F_h.$$

(j) Deduce that

$$\sum_{k=1}^{K} \frac{1}{k\|k\theta\|} \asymp (\log K)^2.$$

(k) Conclude that the Erdős–Turán inequality gives a bound weaker than in (h), namely $D^*(N) \ll (\log N)^2$.

5. Suppose that a sequence $\{u_n\}$ is given, and let

$$D(N) = \sup_{0 \le \alpha \le \beta \le 1} |\operatorname{card}\{n \in [1, N] : \alpha \le u_n \le \beta \pmod 1\} - (\beta - \alpha)N|.$$

(a) Show that $D(N) = \sup_{0\le\alpha\le1} D(N, \alpha) - \inf_{0\le\alpha\le1} D(N, \alpha)$.
(b) Show that $D^*(N) \le D(N) \le 2D^*(N)$.
(c) Suppose that $v_n = u_n + c$ for all n. Show that $D(N; \boldsymbol{u}) = D(N; \boldsymbol{v})$.

6. Let μ denote a probability measure on $\mathbb{T}$. Show that if $\mathcal{I} = [\alpha, \beta]$ is an arc of $\mathbb{T}$, $0 \le \beta - \alpha \le 1$, then for any positive integer K

$$|\mu([\alpha, \beta]) - (\beta - \alpha)| \le \frac{1}{K} + 3\sum_{k=1}^{K} \frac{1}{k}|\widehat{\mu}(k)|$$

where $\widehat{\mu}(k)$ is defined as in (16.4).

In the next exercise we establish a quantitative version of the implication (d) $\Rightarrow$ (c) for a restricted—but important–class of functions.

7. Suppose that $\{u_n\}$ is a given sequence, and that f has bounded variation on $\mathbb{T}$. Show that

$$\left|\sum_{n=1}^{N} f(u_n) - \int_0^1 f(x)\,dx\right| \le D^*(N)\operatorname{Var}_{\mathbb{T}}(f).$$

(Careful! The Riemann-Stieltjes integral $\int f(\alpha)\,dD(N,\alpha)$ does not exist if f is has a jump discontinuity at any of the points u_n.)

Next we establish a quantitative version of the implication (d) $\Rightarrow$ (b).

8. Let $\{u_n\}$ be a given sequence.
(a) Show that if $k \ne 0$, then
$$\sum_{n=1}^{N} e(ku_n) = -2\pi ik\int_0^1 D(N,\alpha)e(k\alpha)\,d\alpha.$$
(b) Show that if $k \ne 0$, then
$$\Big|\sum_{n=1}^{N} e(ku_n)\Big| \le 2\pi|k|D^*(N).$$
(c) Write $\sum_{n=1}^{N} e(ku_n) = \rho e(\theta)$ in polar coordinates, so that $\rho = |\sum_{n=1}^{N} e(ku_n)| = \sum_{n=1}^{N} \cos 2\pi(ku_n - \theta)$. Show that
$$\Big|\sum_{n=1}^{N} e(ku_n)\Big| = 2\pi k\int_0^1 D(N,\alpha)\sin 2\pi(k\alpha-\theta)\,d\alpha \le 4|k|D^*(N).$$
(d) Construct examples to show that the inequality above would be false if the constant 4 were replaced by a number < 4.

9. (a) Suppose that the points u_n are distinct from 0 (mod 1) and from α (mod 1). Show that
$$D(N,\alpha) = \sum_{n=1}^{N} s(u_n-\alpha) - s(u_n)$$
where $s(x)$ is the 'saw-tooth function' as in Lemma D.1.
(b) Show that
$$\int_0^1 D(N,\alpha)\,d\alpha = \sum_{n=1}^{N} \frac{1}{2} - \{u_n\}.$$
(c) By using Lemma D.1, or otherwise, show that if α is distinct (mod 1) from the points u_n, then
$$D(N,\alpha) = \Big(\sum_{n=1}^{N}\frac{1}{2} - \{u_n\}\Big) + \frac{1}{2\pi i}\lim_{K\to\infty}\sum_{0<|k|\le K}\frac{1}{k}\Big(\sum_{n=1}^{N} e(-ku_n)\Big)e(k\alpha).$$
(d) Deduce that
$$\int_0^1 D(N,\alpha)^2\,d\alpha = \Big(\sum_{n=1}^{N}\frac{1}{2} - \{u_n\}\Big)^2 + \frac{1}{2\pi^2}\sum_{k=1}^{\infty}\frac{1}{k^2}\Big|\sum_{n=1}^{N} e(ku_n)\Big|^2.$$

10. (a) Suppose that δ is given. Show that if α and $\alpha+\delta$ are both distinct from the u_n (mod 1), then

$$D(N,\alpha+\delta)-D(N,\alpha)=\frac{1}{2\pi i}\lim_{K\to\infty}\sum_{0<|k|\le K}\frac{1}{k}\Big(\sum_{n=1}^{N}e(-ku_n)\Big)(e(k\delta)-1)e(k\alpha).$$

(b) Deduce that

$$\int_0^1 (D(N,\alpha+\delta)-D(N,\alpha))^2\,d\alpha=\sum_{k\ne 0}\Big(\frac{\sin\pi\delta k}{\pi k}\Big)^2\Big|\sum_{n=1}^{N}e(ku_n)\Big|^2.$$

(c) Show that

$$\sum_{k=1}^{K}\Big|\sum_{n=1}^{N}e(ku_n)\Big|^2\le 2\pi^2K^2D^*(N)^2.$$

11. (a) Show that

$$\sum_{|k|<K}\Big(1-\frac{|k|}{K}\Big)\Big|\sum_{n=1}^{N}e(ku_n)\Big|^2=\sum_{m=1}^{N}\sum_{n=1}^{N}\Delta_K(u_m-u_n)\ge NK$$

where $\Delta_K(\alpha)$ is Fejér's kernel,

$$\Delta_K(\alpha)=\sum_{|k|<K}\Big(1-\frac{|k|}{K}\Big)e(k\alpha)=\frac{1}{K}\Big(\frac{\sin\pi K\alpha}{\sin\pi\alpha}\Big)^2.$$

(b) Show that

$$\max_{1\le k\le 2N}\Big|\sum_{n=1}^{N}e(ku_n)\Big|\ge (N/2)^{1/2}.$$

12. Suppose that $0<u_1\le u_2\le\ldots\le u_N=1$, and put $\delta_n=u_n-n/N$.
(a) Show that

$$\max_{0\le\alpha\le1}D(N,\alpha)=-N\min_{1\le n\le N}\delta_n.$$

(b) Show that

$$\inf_{0\le\alpha\le1}D(N,\alpha)=-1-\max_{1\le n\le N}\delta_n.$$

(c) Show that

$$\int_0^1 D(N,\alpha)^2\,d\alpha=\sum_{n=1}^{N}\int_{u_{n-1}}^{u_n}(n-1-N\alpha)^2\,d\alpha$$

where $u_0 = 0$.
(b) Deduce that

$$\sum_{n=1}^{N} \delta_n^2 = \frac{1}{N}\int_0^1 D(N,\alpha)^2\,d\alpha + \frac{1}{N}\int_0^1 D(N,\alpha)\,d\alpha + \frac{1}{6N}.$$

(c) Show that if $N > 1$, then

$$\sum_{n=1}^{N} e(u_n) = \sum_{n=1}^{N} (e(\delta_n) - 1)e(n/N).$$

(d) Deduce that

$$\Big|\sum_{n=1}^{N} e(u_n)\Big| \le 2\pi \sum_{n=1}^{N} |\delta_n|.$$

13. Take the u_n to be the Farey fractions $a/q \in [0,1)$ of order Q. Thus $(a,q) = 1$, $q \le Q$, and

$$N = N(Q) = \sum_{q=1}^{Q} \varphi(q) \sim \frac{3}{\pi^2}Q^2.$$

(a) By considering the contribution of the interval $[1 - 1/Q, 1)$, show that

$$\int_0^1 D(N,\alpha)^2\,d\alpha \ge \frac{N^2}{3Q^3} \asymp Q.$$

(b) Use properties of Ramanujan's sum $c_q(k)$ to show that

$$\sum_{q=1}^{Q} \sum_{\substack{a=1 \\ (a,q)=1}}^{q} e(ak/q) = \sum_{d|k} dM(Q/d)$$

where $M(x) = \sum_{n\le x} \mu(n)$ is the summatory function of the Möbius function.
(c) Show that $D^*(N) = o(N)$ as $N \to \infty$.
(d) Show that for every $Q \ge 1$,

$$\int_0^1 D(N,\alpha)^2\,d\alpha \ge \frac{M(Q)^2}{2\pi^2}.$$

(e) Show that if $Q \ge 1$, then

$$\int_0^1 D(N,\alpha)^2\,d\alpha = \frac{1}{4} + \frac{1}{12}\sum_{r\le Q}\Big(\prod_{p|r}(1-p^{-2})\Big)\Big(\sum_{s\le Q/r}\frac{1}{s}M(Q/(rs))\Big)^2.$$

(f) (Franel 1924) Show that the Riemann Hypothesis is equivalent to the assertion that

$$\int_0^1 D(N,\alpha)^2 \, d\alpha \ll_\varepsilon Q^{1+\varepsilon}$$

for every $\varepsilon > 0$.
(g) (Franel 1924) Let the numbers δ_n be defined as in Exercise 12. Show that the Riemann Hypothesis is equivalent to the estimate

$$\sum_{n=1}^{N} \delta_n^2 \ll_\varepsilon Q^{-1+\varepsilon}.$$

(h) (Landau 1924) Show that the Riemann Hypothesis is equivalent to the estimate

$$\sum_{n=1}^{N} |\delta_n| \ll_\varepsilon Q^{1/2+\varepsilon}.$$

14. Let b be an integer > 1, and suppose that the representation of x in base b is $x = 0.a_1a_2a_3\ldots$ where $0 \le a_n < b$ for all n. Suppose that $c_1, c_2, \ldots, c_K$ are integers such that $0 \le c_k < b$ for all k. We say that x is *normal base* b if

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{card}\{n \in [1,N] : a_{n+k} = c_k \ (1 \le k \le K)\} = \frac{1}{b^K}$$

for each $K \ge 1$ and each of the b^K admissible choices of the c_k.
(a) Show that x is normal base b if and only if the sequence $\{xb^n\}$ is uniformly distributed (mod 1).
(b) Show that the numbers normal to base b form a set of first Baire category (i.e., the set can be expressed as a countable union of nowhere dense sets).
(c) Show that

$$\int_0^1 \Big| \sum_{n=1}^{N} e(xb^n) \Big|^2 \, dx = N.$$

(d) Let $D^*(N; \boldsymbol{u}_x)$ denote the discrepancy of the sequence $\{xb^n\}_{n=1}^N$. Show that

$$\int_0^1 D^*(N; \boldsymbol{u}_x) \, dx \ll N^{1/2} \log N.$$

(e) Show that almost all real numbers x are normal base b.

15. *Grössencharaktere* for $\mathbb{Q}(\sqrt{-1})$, continued from Exercise 11.3.1.14. Show that the number of pairs (a,b) of integers such that $a^2 + b^2 \le x$, $a^2 + b^2$ is prime, and $0 \le \arg(a + ib) \le \theta$ is

$$\frac{2\theta}{\pi} \operatorname{li}(x) + O\big(x \exp(-c\sqrt{\log x})\big)$$

uniformly for $0 \le \theta \le 2\pi$.

16.3 Kronecker's theorem

We now generalize Theorem 16.2 to R dimensions: We describe the distribution of the points $(\{n\theta_1\}, \{n\theta_2\}, \dots, \{n\theta_R\})$ in $[0,1)^R$. Since it is a nuisance to have to take the fractional part of real numbers, we simply consider $\boldsymbol{p}_n = (n\theta_1, n\theta_2, \dots, n\theta_R)$ modulo $\mathbb{Z}^R$, or, equivalently, we consider $\boldsymbol{p}_n$ to represent a member of the R-dimensional circle group $\mathbb{T}^R = (\mathbb{R}/\mathbb{Z})^R$.

For $1 \le r \le R$ let $\mathcal{I}_r = [\alpha_r, \beta_r]$ be an arc of $\mathbb{T}$ with $0 \le \beta_r - \alpha_r \le 1$, so that $\mathcal{B} = \mathcal{I}_1 \times \mathcal{I}_2 \times \cdots \times \mathcal{I}_R$ is a box in $\mathbb{T}^R$. For a given sequence $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots$ of points in $\mathbb{T}^R$, let

$$Z(N, \mathcal{B}) = \operatorname{card}\{n \in [1, N] : \boldsymbol{u}_n \in \mathcal{B}\}$$

and set

$$D(N, \mathcal{B}) = Z(N, \mathcal{B}) - N \prod_{r=1}^{R} (\beta_r - \alpha_r).$$

We say that the points $\boldsymbol{u}_n$ are *uniformly distributed* in $\mathbb{T}^R$ if $D(N, \mathcal{B}) = o(N)$ as $N \to \infty$ for every such box $\mathcal{B} \subseteq \mathbb{T}^R$. This is all in parallel with our treatment of the case $R = 1$, and we can also define a discrepancy function,

$$D(N) = \sup_{\mathcal{B} \subseteq \mathbb{T}^R} |D(N, \mathcal{B})|$$

where the supremum is over all boxes as described above. Weyl's criterion extends to this situation in an obvious manner.

Theorem 16.5 *Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots$ be a given sequence of points in $\mathbb{T}^R$. Then the following assertions are equivalent:*

(a) *The sequence $\{\boldsymbol{u}_n\}$ is uniformly distributed in $\mathbb{T}^R$;*

(b) *If $\boldsymbol{k}$ is a non-zero lattice point (i.e., $\boldsymbol{k} \in \mathbb{Z}^R$, $\boldsymbol{k} \ne \boldsymbol{0}$), then*

$$\sum_{n=1}^{N} e(\boldsymbol{k} \cdot \boldsymbol{u}_n) = o(N) \qquad (N \to \infty);$$

(c) *If f is properly Riemann-integrable on $\mathbb{T}^R$, then*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f(\boldsymbol{u}_n) = \int_{\mathbb{T}^R} f(\boldsymbol{x})\, d\boldsymbol{x};$$

(d) *$D(N) = o(N)$ as $N \to \infty$.*

Proof The arguments of §16.1 carry over to the present context without change, except for the issue of constructing trigonometric majorants and minorants in several dimensions. We consider the majorants first. For $1 \le r \le R$ suppose that $T_r(x)$ is a trigonometric polynomial such that $\chi_{\mathcal{I}_r}(x) \le T_r(x)$ for all x, and that $\int_0^1 T_r(x)\, dx \le \beta_r - \alpha_r + \varepsilon$. If we set

$$T_+(\boldsymbol{x}) = \prod_{r=1}^{R} T_r(x_r),$$

then $\chi_{\mathcal{B}}(\boldsymbol{x}) \le T_+(\boldsymbol{x})$ and

$$\int_{\mathbb{T}^R} T_+(\boldsymbol{x})\,d\boldsymbol{x} \le \prod_{r=1}^{R}(\beta_r - \alpha_r + \varepsilon) \le \mathrm{vol}(\mathcal{B}) + \big((1+\varepsilon)^R - 1\big).$$

This suffices as a majorant. As for minorants, we observe that $\mathcal{I}_r$ and its complement $\mathcal{I}_r^c$ partition $\mathbb{T}$ into two subsets. Hence the cartesian products of the $\mathcal{I}_r$ and their complements partition $\mathbb{T}^R$ into 2^R boxes, say $\mathcal{B} = \mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_{2^R}$. That is,

$$\sum_{m=1}^{2^R} \chi_{\mathcal{B}_m}(\boldsymbol{x}) \equiv 1.$$

Hence if $\chi_{\mathcal{B}_m}(x) \le T_m(\boldsymbol{x})$ for all $\boldsymbol{x}$ and $\int_{\mathbb{T}^R} T_m(\boldsymbol{x})\,d\boldsymbol{x} \le \mathrm{vol}(\mathcal{B}_m) + \varepsilon$, then

$$\chi_{\mathcal{B}}(\boldsymbol{x}) = 1 - \sum_{m=2}^{2^R} \chi_{\mathcal{B}_m}(\boldsymbol{x}) \ge 1 - \sum_{m=2}^{2^R} T_m(\boldsymbol{x}) = T_-(\boldsymbol{x}),$$

say, and $\int_{\mathbb{T}^R} T_-(\boldsymbol{x})\,d\boldsymbol{x} \ge \mathrm{vol}(\mathcal{B}) - (2^R - 1)\varepsilon$. This suffices to construct the required trigonometric minorant. □

We consider several forms of Kronecker's theorem, the simplest being a natural extension of Theorem 16.2.

Theorem 16.6 *Let $\theta_1, \theta_2, \dots, \theta_R$ be real numbers. If the points $\boldsymbol{p}_n = (n\theta_1, n\theta_2, \dots, n\theta_R)$ are dense in $\mathbb{T}^R$, then $1, \theta_1, \theta_2, \dots, \theta_R$ are linearly independent over $\mathbb{Q}$. Conversely, if $1, \theta_1, \theta_2, \dots, \theta_R$ are linearly independent over $\mathbb{Q}$, then the points $\boldsymbol{p}_n$ are not only dense in $\mathbb{T}^R$ but are uniformly distributed in $\mathbb{T}^R$.*

Proof Suppose that

$$c_0 + c_1\theta_1 + c_2\theta_2 + \cdots + c_R\theta_R = 0$$

where the c_r are integers, not all 0. Clearly at least one of $c_1, c_2, \dots, c_R$ is non-zero; without loss of generality, we may suppose that $c_R \ne 0$. If $\|n\theta_r\| \le \varepsilon$ for $1 \le r < R$, then

$$\|c_R n\theta_R\| = \Big\|\sum_{r=1}^{R-1} c_r n\theta_r\Big\| \le \sum_{r=1}^{R-1}\|c_r n\theta_r\| \le \sum_{r=1}^{R-1}|c_r|\|n\theta_r\| \le \varepsilon\sum_{r=1}^{R-1}|c_r|.$$

Suppose that ε is so small that this last quantity above is $\le |5c_R|^{-1}$. Then the box

$$\|x_1\| \le \varepsilon, \|x_2\| \le \varepsilon, \dots, \|x_{R-1}\| \le \varepsilon, \Big\|x_R - \frac{1}{2c_R}\Big\| \le \frac{1}{5c_R}$$

contains no point $\boldsymbol{p}_n$, so the $\boldsymbol{p}_n$ are not dense.

Suppose now that $1, \theta_1, \theta_2, \ldots, \theta_R$ are linearly independent over $\mathbb{Q}$. Hence if $\boldsymbol{k} \in \mathbb{Z}^R$, $\boldsymbol{k} \neq \nvdash$, then $\boldsymbol{k} \cdot \boldsymbol{\theta}$ is not an integer. Consequently by (16.5) it follows that

$$\left| \sum_{n=1}^{N} e(\boldsymbol{k} \cdot \boldsymbol{p}_n) \right| = \left| \sum_{n=1}^{N} e(n\boldsymbol{k} \cdot \boldsymbol{\theta}) \right| \le \frac{1}{2\|\boldsymbol{k} \cdot \boldsymbol{\theta}\|} = o(N),$$

and so condition (b) of Theorem 16.5 is satisfied, and hence the $\boldsymbol{p}_n$ are uniformly distributed in $\mathbb{T}^R$. $\square$

Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots$ be a sequence of points in $\mathbb{T}^R$. We may define a probability measure μ_N by placing a mass $1/N$ at each of the points $\boldsymbol{u}_n$ for $1 \le n \le N$. Then Theorem 16.5 could be formulated in terms of the μ_N, and indeed both the theorem and its proof apply equally to any sequence of probability measures. That is, the following assertions are equivalent:

(a) If $\mathcal{B} = \prod_{r=1}^{R} \mathcal{I}_r$ is a box in $\mathbb{T}^R$, then $\lim_{N\to\infty} \mu_N(\mathcal{B}) = \operatorname{vol} \mathcal{B}$;
(b) If $\boldsymbol{k} \in \mathbb{Z}^R$, $\boldsymbol{k} \neq \boldsymbol{0}$, then $\lim_{N\to\infty} \widehat{\mu}_N(\boldsymbol{k}) = 0$ where $\widehat{\mu}(\boldsymbol{k}) = \int_{\mathbb{T}^R} e(-\boldsymbol{k} \cdot \boldsymbol{x})\, d\mu(\boldsymbol{x})$;
(c) If f is properly Riemann-integrable on $\mathbb{T}^R$, then

$$\lim_{N\to\infty} \int_{\mathbb{T}^R} f(\boldsymbol{x})\, d\mu_N(\boldsymbol{x}) = \int_{\mathbb{T}^R} f(\boldsymbol{x})\, d\boldsymbol{x};$$

(d) $\lim_{N\to\infty} \sup_{\mathcal{B}} |\mu_N(\mathcal{B}) - \operatorname{vol} \mathcal{B}| = 0$ where the supremum is over all boxes $\mathcal{B} = \prod_{r=1}^{R} \mathcal{I}_r \subseteq T^R$.

In particular, if $\boldsymbol{p}(t)$ is the position vector of a continuous curve, then we can define a probability measure

$$\mu_T(\mathcal{S}) = \frac{1}{T} \operatorname{meas}\{t \in [0, T] : \boldsymbol{p}(t) \in \mathcal{S}\}.$$

Thus if f is Riemann-integrable over $\mathbb{T}^R$, then

$$\int_{\mathbb{T}^R} f(\boldsymbol{x})\, d\mu_T(\boldsymbol{x}) = \frac{1}{T} \int_0^T f(\boldsymbol{p}(t))\, dt,$$

and hence we see that the curve $\boldsymbol{p}(t)$ is uniformly distributed in $\mathbb{T}^R$ if and only if

$$\int_0^T e(\boldsymbol{k} \cdot \boldsymbol{p}(t))\, dt = o(T) \qquad (T \to \infty)$$

for every non-zero lattice point $\boldsymbol{k} \in \mathbb{Z}^R$. A situation of this kind arises in our second formulation of Kronecker's theorem.

Theorem 16.7 *Let $\theta_1, \theta_2, \ldots, \theta_R$ be real numbers and let $\boldsymbol{p}(t) = (t\theta_1, t\theta_2, \ldots, t\theta_R)$ where t is a real parameter. If the set $\mathcal{P} = \{\boldsymbol{p}(t) : t \in \mathbb{R}\}$ is dense in $\mathbb{T}^R$, then $\theta_1, \theta_2, \ldots, \theta_R$ are linearly independent over $\mathbb{Q}$. Conversely, if $\theta_1, \theta_2, \ldots, \theta_R$ are linearly independent over $\mathbb{Q}$, then $\mathcal{P}$ is not only dense in $\mathbb{R}^R$ but also uniformly distributed in the sense that*

$$\lim_{T\to\infty} \sup_{\mathcal{B}} \left|\frac{1}{T} \operatorname{meas}\{t \in [0, T] : \boldsymbol{p}(t) \in \mathcal{B}\} - \operatorname{vol} \mathcal{B}\right| = 0$$

where the supremum is taken over all boxes $\mathcal{B} = \prod_{r=1}^{R} \mathcal{I}_r \subseteq \mathbb{T}^R$.

It is easy to demonstrate by elementary reasoning that Theorems 16.6 and 16.7 are equivalent (see Exercise 1 below). Thus it is possible to present Theorem 16.7 as a consequence of Theorem 16.6, but we find it more instructive to derive it independently.

Proof Suppose that the θ_r are linearly dependent, say $\boldsymbol{c}\cdot\boldsymbol{\theta} = 0$ where $\boldsymbol{c} \in \mathbb{Z}^R$ and $c_R > 0$. Hence if $\boldsymbol{p} \equiv t\boldsymbol{\theta} \pmod{\mathbb{Z}^R}$ for some t, then $\boldsymbol{c}\cdot\boldsymbol{p} \in \mathbb{Z}$. Let $\mathcal{B}$ denote the box of points $\boldsymbol{x}$ for which $|x_r| \le \varepsilon$ for $1 \le r < R$ and $|x_R - 1/(2c_R)| \le 1/(5c_R)$. If $\boldsymbol{x} \in \mathcal{B}$, then

$$|\boldsymbol{c}\cdot\boldsymbol{x} - \frac{1}{2}| \le |c_R x_R - 1/2| + \sum_{r=1}^{R-1} |c_r x_r| \le \frac{1}{5} + \varepsilon \sum_{r=1}^{R-1} |c_r|.$$

Hence if ε is chosen so small that this last term is $< 1/5$, then $\boldsymbol{c}\cdot\boldsymbol{x}$ is not an integer, $\mathcal{B}$ contains no point of the curve $t\boldsymbol{\theta}$, and so the curve is not dense in $\mathbb{T}^R$.

Now suppose that the θ_r are linearly independent over $\mathbb{Q}$. On defining the measure μ_T as above, we see that if $\boldsymbol{k} \in \mathbb{Z}^R$, $\boldsymbol{k} \ne \boldsymbol{0}$, then

$$\widehat{\mu}_T(\boldsymbol{k}) = \frac{1}{T}\int_0^T e(\boldsymbol{k}\cdot t\boldsymbol{\theta})\,dt = \frac{1}{T}\cdot\frac{e(\boldsymbol{k}\cdot T\boldsymbol{\theta}) - 1}{2\pi i(\boldsymbol{k}\cdot\boldsymbol{\theta})} \to 0$$

as $T \to \infty$. Hence the curve $t\boldsymbol{\theta}$ is uniformly distributed in $\mathbb{T}^R$. □

Theorem 16.7 provides useful information concerning the values taken by exponential polynomials, as follows.

Corollary 16.8 *Suppose that* $f(t) = \sum_{r=1}^{R} a_r e(\lambda_r t)$ *where the* λ_r *are real numbers. Suppose also that* $|a_1| \ge |a_2| \ge \cdots \ge |a_R|$. *If* $|a_1| \le \sum_{r=2}^{R} |a_r|$, *then the values of* $f(t)$ *for* $t \in \mathbb{R}$ *lie in the disk* $|z| \le \sum_{r=1}^{R} |a_r|$. *If* $|a_1| > \sum_{r=2}^{R} |a_r|$, *then the values of* $f(t)$ *lie in the annulus* $|a_1| - \sum_{r=2}^{R} |a_r| \le |z| \le |a_1| + \sum_{r=2}^{R} |a_r|$. *If the* λ_r *are linearly independent over* $\mathbb{Q}$, *then the values of* $f(t)$ *are dense in this disk* (*or annulus*).

Proof It is clear that the disk (or annulus) described is the set of points z that can be written in the form $z = \sum_{r=1}^{R} a_r e(\alpha_r)$. If the λ_r are linearly independent over $\mathbb{Q}$, then for any $\varepsilon > 0$ there exist real numbers t such that $\|\lambda_r t - \alpha_r\| \le \varepsilon$ and hence $|f(t) - z| \le C\varepsilon$ where $C = 2\pi\sum_{r=1}^{R} |a_r|$. □

We have found that if the numbers θ_r are linearly independent over $\mathbb{Q}$, then the curve $t\boldsymbol{\theta}$ passes through any given box $\mathcal{B}$ infinitely many times. But we can actually prove a little more, namely that the gaps between returns to $\mathcal{B}$ are uniformly bounded. This is critical to our discussion of almost periodic functions in the next section.

Corollary 16.9 *Suppose that the real numbers* $\theta_1, \theta_2, \ldots, \theta_R$ *are linearly independent over* $\mathbb{Q}$, *and let* $\varepsilon > 0$ *be given. Then there is a number* $H > 0$, *depending only on* ε *and the numbers* $\theta_1, \theta_2, \ldots, \theta_R$, *such that for any real number* T *and any* $\boldsymbol{\alpha} \in \mathbb{T}^R$ *there is a real number* t, $T \le t \le T + H$, *such that* $\|t\theta_r - \alpha_r\| \le \varepsilon$ *for* $1 \le r \le R$.

In order to clarify the relation of this new result to our earlier ones, we provide two proofs.

First Proof By Theorem 16.7 we know that if H is sufficiently large, then

$$|\operatorname{meas}\{t \in [0, H] : t\boldsymbol{\theta} \in \mathcal{B}\} - \operatorname{vol}\mathcal{B}| < (2\varepsilon)^R$$

for all boxes $\mathcal{B}$. Let $\mathcal{B}_0 = [-\varepsilon, \varepsilon]^R$ be a small box centred at $\mathbf{0}$. Then $\boldsymbol{c} + \mathcal{B}$ is a box of the same size, centred at $\boldsymbol{c}$. By taking $\mathcal{B} = \boldsymbol{c} + \mathcal{B}_0$ in the above, we see that for every $\boldsymbol{c}$ there is a t, $0 \le t \le H$ such that $t\boldsymbol{\theta} \in \boldsymbol{c} + \mathcal{B}_0$, which is to say that $\|t\theta_r - c_r\| \le \varepsilon$ for all r. Now take $c_r = \alpha_r - T\theta_r$. Then $\|(t+T)\theta_r - \alpha_r\| \le \varepsilon$ for all r, and $T \le T + t \le T + H$. □

Second Proof Let $\mathcal{B}_0 = [-\varepsilon, \varepsilon]^R$ be a small box centred at $\mathbf{0}$, and let $T_-(\boldsymbol{x})$ be a trigonometric polynomial in R variables such that

$$T_-(\boldsymbol{x}) \le \chi_{\mathcal{B}_0}(\boldsymbol{x})$$

for all $\boldsymbol{x}$, and

$$\int_{\mathbb{T}^R} T_-(\boldsymbol{x})\, d\boldsymbol{x} > 0.$$

Then

$$\int_T^{T+H} \chi_{\mathcal{B}_0}(t\boldsymbol{\theta} - \boldsymbol{\alpha})\, dt \ge \int_T^{T+H} T_-(t\boldsymbol{\theta} - \boldsymbol{\alpha})\, dt = \sum_{\boldsymbol{k}} \widehat{T}_-(\boldsymbol{k}) e(-\boldsymbol{k}\cdot\boldsymbol{\alpha}) \int_T^{T+H} e(t\boldsymbol{k}\cdot\boldsymbol{\theta})\, dt.$$

Now if $\theta \ne 0$, then

$$\Big|\int_T^{T+H} e(t\theta)\, dt\Big| = \Big|\frac{e((T+H)\theta) - e(T\theta)}{2\pi i\theta}\Big| = \Big|\frac{\sin \pi H\theta}{\pi\theta}\Big| \le \frac{1}{\pi|\theta|}.$$

By hypothesis the θ_r are linearly independent over $\mathbb{Q}$, which is to say that $\boldsymbol{k}\cdot\boldsymbol{\theta} \ne 0$ when $\boldsymbol{k} \in \mathbb{Z}^R$, $\boldsymbol{k} \ne \mathbf{0}$. Hence

$$\int_T^{T+H} \chi_{\mathcal{B}_0}(t\boldsymbol{\theta} - \boldsymbol{\alpha})\, dt \ge \widehat{T}_-(\mathbf{0})H - \sum_{\boldsymbol{k}\ne\mathbf{0}} \frac{|\widehat{T}_-(\boldsymbol{k})|}{\pi|\boldsymbol{k}\cdot\boldsymbol{\theta}|}.$$

Here the sum is finite because T_- is a trigonometric polynomial. Since $\widehat{T}_-(\mathbf{0}) > 0$, we see that if H is large enough, then the right hand side above is positive, and so there is a t, $T \le t \le T + H$, such that $\|t\theta_r - \alpha_r\| \le \varepsilon$ for $1 \le r \le R$. □

In the proof of Theorem 19.5, we constructed our minorant by rather inefficient means, which would be quantitatively inferior in higher dimensions. When quantitative precision is desired, the following construction may be useful.

Theorem 16.10 *For $1 \le i \le n$, let I_i be intervals of $\mathbb{R}$ or arcs of $\mathbb{T}$, and let $\mathcal{B}$ be their cardesian product. Suppose that $S_i^-(x) \le \chi_{I_i}(x) \le S_i^+(x)$ are respectively minorants and majorants of the characteristic function of I_i. Then*

$$S^-(\boldsymbol{x}) \;=\; \prod_{i=1}^n S_i^+(x_i) \;-\; \sum_{i=1}^n \big(S_i^+(x_i) - S_i^-(x_i)\big) \prod_{\substack{1\le j\le n\\ j\ne i}} S_j^+(x_j)$$

is a minorant of the characteristic function of $\mathcal{B}$.

This construction is also useful when we want to combine several sieve methods.

Proof Write $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$. Suppose first that there is a k such that $S_k^-(x_k) \le 0$. Then

$$\begin{aligned} S^-(\boldsymbol{x}) &\le \prod_{i=1}^n S_i^+(x_i) - \big(S_k^+(x_k) - S_k^-(x_k)\big) \prod_{\substack{1\le j\le n \\ j\ne k}} S_j^+(x_j) \\ &= S_k^-(x_k) \prod_{\substack{1\le j\le n \\ j\ne k}} S_j^+(x_j) \\ &\le 0 \end{aligned}$$

since the first factor is ≤ 0 and all factors in the product are ≥ 0.

Suppose now that $S_i^-(x_i) \ge 0$ for all i. By induction on n we show that in this case,

$$S^-(\boldsymbol{x}) \le \prod_{i=1}^n S_i^-(x_i)\,. \tag{16.7}$$

This is obvious when $n = 1$. We observe that

$$S^-(\boldsymbol{x}) = S_1^-(x_1)\prod_{i=2}^n S_i^+(x_i) - \sum_{i=2}^n \big(S_i^+(x_i) - S_i^-(x_i)\big) \prod_{\substack{1\le j\le n \\ j\ne i}} S_j^+(x_j)\,.$$

Since $i \ge 2$ in the second term, the factor $S_1^+(x_1)$ always occurs in the second product. If we replace $S_1^+(x_1)$ by $S_1^-(x_1)$, then the product is made smaller, and the overall contribution larger. Hence the above is

$$\le S_1^-(x_1)\Big(\prod_{i=2}^n S_i^+(x_i) - \sum_{i=2}^n \big(S_i^+(x_i) - S_i^-(x_i)\big) \prod_{\substack{2\le j\le n \\ j\ne i}} S_j^+(x_j)\Big)\,.$$

By the inductive hypothesis, the quantity inside the large parentheses is

$$\le \prod_{i=2}^n S_i^-(x_i),$$

so we have (16.7), and the proof is complete. $\square$

16.3.1 Exercises

1. (a) Apply Theorem 7 to the $R+1$ numbers $1, \theta_1, \dots, \theta_R$. Deduce that there exist real numbers t such that $\|t - c_0\| < \varepsilon$, $\|t\theta_r - c_r\| < \varepsilon$ for $1 \le r \le R$. Take $c_0 = 0$, and hence deduce that t is near some integer, say n. Show that the numbers $\|n\theta_r - c_r\|$ are small, and hence deduce Theorem 16.6.
   (b) Suppose that $\theta_1, \dots, \theta_R$ are given linearly independent numbers, and choose α to be linearly independent of them. Thus $1, \theta_1/\alpha, \theta_2/\alpha, \dots, \theta_R/\alpha$ are linearly independent over $\mathbb{Q}$. Apply Theorem 16.6 to obtain Theorem 16.7 with $t = n/\alpha$.

2. Let $f(t) = \sum_{r=1}^{R} a_r \cos(\lambda_r t + \theta_r)$ where the a_r, the λ_r and the θ_r are real numbers. Show that if the λ_r are linearly independent over $\mathbb{Q}$, then $(-A, A) \subseteq \operatorname{range} f \subseteq [-A, A]$ where $A = \sum_{r=1}^{R} |a_r|$.

3. In this exercise we develop an alternative to the construction of Theorem 16.10. We suppose that $0 \le A_i(x) \le 1$ for all x, that $P_i(x) \ge 0$ for all x, and that $A_i(x) - P_i(x) \le 0$ when $x \notin I_i$. (We think of A_i as being an approximation to the characteristic function of I_i, and of $P_i(x)$ as a peak function that compensates for the error in this approximation.) Put
$$S^-(\boldsymbol{x}) \;=\; \prod_{i=1}^{n} A_i(x_i) \;-\; sum_{i=1}^{n} P_i(x_i) \prod_{\substack{1 \le j \le n \\ j \ne i}} A_j(x_j)\,.$$
   (a) Show that $S^-(\boldsymbol{x}) \le 1$ for all $\boldsymbol{x}$.
   (b) Suppose that $x_k \notin I_k$. Show that
$$S^-(\boldsymbol{x}) \le \prod_{i=1}^{n} A_i(x_i) - P_k(x_k) \prod_{\substack{1 \le j \le n \\ j \ne k}} A_j(x_j) = \big(A_k(x_k) - P_k(x_k)\big) \prod_{\substack{1 \le j \le n \\ j \ne k}} A_j(x_j) \le 0.$$
   (c) Conclude that $S^-(\boldsymbol{x})$ minorizes the characteristic function of the box $\mathcal{B} = \times_{i=1}^{n} I_i$.

16.4 Almost periodicity

The definition of almost periodicity is governed by our desire to characterize those functions $f(x)$ of a real variable that can be uniformly approximated by exponential polynomials, i.e., by finite sums of the form

$$P(x) = \sum_{m=1}^{M} a_m e(\lambda_m x). \tag{16.8}$$

To this end we call t an ε *almost-period* if $|f(x+t) - f(x)| < \varepsilon$ for all real x. The appropriate definition of almost periodicity is a little elusive because the mere existence of large almost-periods does not ensure that $f(x)$ can be uniformly approximated by exponential polynomials. The little bit more that is required is suggested by Corollary 16.9.

Definition *Suppose that $f(x)$ is a continuous function of a real variable. Then $f(x)$ is* almost periodic *if for every $\varepsilon > 0$ there exists a number H (depending on f and ε) such that every interval $[T, T+H]$ contains an ε almost-period of f.*

We use Corollary 16.9 to show that an exponential polynomial of the form (16.8) is almost periodic.

Theorem 16.11 *Let $P(x)$ be defined as in* (16.8). *Then $P(x)$ is an almost periodic function.*

Proof Let $\tau_1, \tau_1, \ldots, \tau_R$ be a maximal linearly independent subset of the λ_m. Thus there is an $M \times R$ matrix A with rational elements such that $\boldsymbol{\lambda} = A\boldsymbol{\tau}$. Let q be the least common denominator of the elements of A. Put $B = qA$, $\boldsymbol{\theta} = \frac{1}{q}\boldsymbol{\tau}$. Then the θ_r are linearly independent over $\mathbb{Q}$, and $\boldsymbol{\lambda} = B\boldsymbol{\theta}$, which is to say that each λ_m is an integral linear combination of the θ_r. That is, there is a trigonometric polynomial

$$T(\boldsymbol{x}) = \sum_{\boldsymbol{k}} c(\boldsymbol{k}) e(\boldsymbol{k} \cdot \boldsymbol{x})$$

in R variables such that $P(x) = T(x\boldsymbol{\theta})$. By Corollary 16.9 we know that for any $\varepsilon > 0$ there is an H such that for any T there is a t, $T \le t \le T + H$, such that $\|t\theta_r\| \le \varepsilon$. Then

$$\begin{aligned}
|P(x+t) - P(x)| &= |T((x+t)\boldsymbol{\theta}) - T(x\boldsymbol{\theta})| \\
&= \Big| \sum_{\boldsymbol{k}} c(\boldsymbol{k}) e(\boldsymbol{k} \cdot x\boldsymbol{\theta})(e(\boldsymbol{k} \cdot t\boldsymbol{\theta}) - 1) \Big| \\
&\le 2\pi \sum_{\boldsymbol{k}} |c(\boldsymbol{k})| \, \|\boldsymbol{k} \cdot t\boldsymbol{\theta}\| \\
&\le 2\pi\varepsilon \sum_{\boldsymbol{k}} |c(\boldsymbol{k})| \sum_{r=1}^{R} |k_r| \\
&= C\varepsilon,
\end{aligned}$$

say. Thus t is a $C\varepsilon$ almost period of P. □

The theory of almost periodic functions is quite extensive. We confine ourselves here to describing only some of the highlights of this theory, and refer the reader to other sources for proofs and more detailed accounts.

If $f(x)$ is almost periodic, then for every real number λ the limit

$$c(\lambda) = \lim_{T \to \infty} \frac{1}{T} \int_0^T f(x) e(-\lambda x) \, dx$$

exists. Let $\mathcal{L}(f)$ denote the set of those λ for which $c(\lambda) \ne 0$. The set $\mathcal{L}(f)$ is at most countable, and indeed there is a sort of Parseval identity:

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T |f(x)|^2 \, dx = \sum_{\lambda \in \mathcal{L}(f)} |c(\lambda)|^2.$$

If $f(x)$ is almost periodic, then for every $\varepsilon > 0$ there is an almost periodic polynomial $T(x)$ of the shape (16.8) such that $|f(x) - T(x)| < \varepsilon$ for all x, and indeed such a $T(x)$ can be constructed so that $\lambda_m \in \mathcal{L}(f)$ for all m.

The notion of almost periodicity that we have described here is known as *uniform almost periodicity* because it is based on the uniform norm. Other kinds of almost periodicity can be defined by using other norms. For example, the function $f(y) = (\psi(e^y) - e^y)/e^{y/2}$ is not uniformly almost-periodic, but it can be shown that it is mean-square almost-periodic if the Riemann Hypothesis is true.

16.4.1 Exercises

1. Show that if $f(x)$ is almost periodic, then $f(x)$ is uniformly bounded.

2. (Bohl 19xx; cf Bohr 1925) Suppose that $f_1(x), f_2(x), \ldots, f_R(x)$ are periodic continuous functions. Show that $f_1(x) + f_2(x) + \cdots + f_R(x)$ is almost-periodic.

3. Let $f(x) = \sum_{m=1}^\infty a_m e(\lambda_m x)$ where $\sum_{m=1}^\infty |a_m| < \infty$ and the λ_m are distinct real numbers.
   (a) Show that $f(x)$ is almost-periodic.
   (b) Show that
$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(x)e(-\lambda x)\,dx = \begin{cases} a_m & \text{if } \lambda = \lambda_m \text{ for some } m,\\ 0 & \text{otherwise.}\end{cases}$$
   (c) Show that
$$\lim_{T\to\infty} \frac{1}{T}\int_0^T |f(x)|^2\,dx = \sum_{m=1}^\infty |a_m|^2.$$

4. Let $\alpha(s) = \sum_{n=1}^\infty a_n n^{-s}$ be a Dirichlet series with abscissa of absolute convergence σ_a. Let σ be fixed, $\sigma > \sigma_a$, and put $f(t) = \alpha(\sigma + it)$.
   (a) Show that the preceding exercise applies to $f(t)$.
   (b) Show that $\mathcal{L}(f) = \{\frac{-1}{2\pi}\log n : a_n \ne 0\}$.

16.5 The weak convergence of measures

We now develop an analogue for the real line $\mathbb{R}$ of uniform distribution in the circle group $\mathbb{T}$. This is useful in discussing the limiting distribution of the error term in the Prime Number Theorem, and (in Chapter 26) the limiting distribution of additive functions.

Let μ be a measure on the real line. We say that μ is a *probability measure* if it is non-negative and if its total mass is 1. That is, $\mu(\mathcal{S}) \ge 0$ for all measurable sets $\mathcal{S}$ and $\lim_{A\to+\infty} \mu([-A, A]) = 1$. The *distribution function* of μ is the function $F(x) = \mu((-\infty, x])$. Clearly μ is a probability measure if and only if F is increasing and

$$\lim_{x\to-\infty} F(x) = 0, \qquad \lim_{x\to+\infty} F(x) = 1. \tag{16.9}$$

Conversely, if F is increasing, right-continuous (i.e., $F(x^+) = F(x)$ for all real x), and satisfies (16.9), then there is a probability measure μ of which F is the distribution function. If μ is a measure, we say that a sequence of probability measures μ_N *tends weakly* to μ if $\lim_{N\to\infty} F_N(x) = F(x)$ whenever x is a point of continuity of F. We note that the distribution function F of a non-negative measure can have at most countably many jump discontinuities. As a first observation about sequences of measures, we have

Theorem 16.12 (Helly's theorem) *Let $\mu_1, \mu_2, \ldots$ be a sequence of probability measures. Then there is a strictly increasing sequence n_k of positive integers such that the subsequence μ_{n_k} is weakly convergent.*

Proof We proceed by a Cantor diagonal process. Let $x_1, x_2, \ldots$ be distinct and dense in $\mathbb{R}$. The numbers $F_N(x_1)$ have a limit point. Thus we may choose indices $N(j,1)$ so that $F_{N(j,1)}(x_1)$ converges. From the indices $N(j,1)$ we choose a subsequence $N(j,2)$ so that $F_{N(j,2)}(x_2)$ converges. We continue in this manner, and find that the sequence $F_{N(j,j)}(x_k)$ converges for all k. Hence there is a unique right-continuous function F such that $F_{N(j,j)}(x)$ tends to $F(x)$ for all points of continuity x of F. $\square$

Consider now the particular case in which μ is the probability measure that attaches weight 1 to the integer N. The sequence μ_N converges to the measure that is identically 0. Thus a limit of probability measures need not be a probability measure. This can be explained by noting that (16.9) need not hold uniformly for all measures μ_N of a sequence. We say, however, that a sequence of measures is *tight* if (16.9) holds uniformly for all measures μ of the family. Our first observation concerning tight families of measures is obvious.

Theorem 16.13 *If $\mu_1, \mu_2, \ldots$ is a sequence of probability measures converging weakly to a measure μ, and if the sequence is tight, then μ is a probability measure.*

If μ is a probability measure and f is bounded and continuous, then the integral $\int_{-\infty}^{+\infty} f(x)\, d\mu(x)$ is well-defined. In particular, for real t we let $\widehat{\mu}(t)$ denote the Fourier transform of the measure,

$$\widehat{\mu}(t) = \int_{-\infty}^{\infty} e(-tx)\, d\mu(x).$$

We note that $\widehat{\mu}(0)$ is the total mass of a measure, so that $\widehat{\mu}(0) = 1$ for a probability measure.

Lemma 16.14 *Let μ be a probability measure. Then $|\mu(t)| \le 1$ for all t, and $\widehat{\mu}(t)$ is continuous. Let $\mathcal{S}$ be a tight family of probability measures. Then $\widehat{\mu}(t)$ is uniformly continuous in t, and uniformly so for $\mu \in \mathcal{S}$.*

Proof Clearly $|\widehat{\mu}(t)| \le \int_{-\infty}^{\infty} 1\, d\mu = 1$. Suppose that $\varepsilon > 0$ is given, and that A is so large that $\mu([-A, A]) > 1 - \varepsilon$. Then for any real t,

$$\left| \int_{-A}^{A} e(-tx)\, d\mu(x) - \widehat{\mu}(t) \right| < \varepsilon.$$

Hence by the triangle inequality,

$$|\widehat{\mu}(t_1) - \widehat{\mu}(t_2)| < 2\varepsilon + \int_{-A}^{A} |e(-t_1x) - e(-t_2x)|\,d\mu(x).$$

Here the integrand is $|\sin\pi(t_1 - t_2)x| \le \pi|t_1 - t_2|A$. Thus if $|t_1 - t_2| < \varepsilon/A$, then the integrand is uniformly $< \pi\varepsilon$. Hence the integral is $< \pi\varepsilon$, so that $|\widehat{\mu}(t_1) - \widehat{\mu}(t_2)| < 6\varepsilon$. Thus $\widehat{\mu}$ is uniformly continuous. If $\mathcal{S}$ is a tight family of measures, then the choice of A depends on ε and $\mathcal{S}$, but is independent of the individual $\mu \in \mathcal{S}$. $\square$

Lemma 16.15 *If $f(t) \in L^1(\mathbb{R})$ and μ is a probability measure, then*

$$\int_{-\infty}^{\infty} \widehat{f}(x)\,d\mu(x) = \int_{-\infty}^{\infty} f(t)\widehat{\mu}(t)\,dt.$$

We note that $\widehat{f}$ is bounded and continuous, so that the integral on the left can be considered to be a Riemann–Stieltjes integral $\int \widehat{f}\,dF$. As $|f\widehat{\mu}| \le |f|$, the integral on the right is a Lebesgue integral for $f \in L^1(\mathbb{R})$.

Proof By the definitions of $\widehat{\mu}$ and $\widehat{f}$, the above asserts that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(t)e(-tx)\,dt\,d\mu(x) = \int_{-\infty}^{\infty} f(t)\int_{-\infty}^{\infty} e(-tx)\,d\mu(x)\,dt.$$

The interchange of integrals is justified by Fubini's theorem, in view of the joint integrability of $f(t)e(-tx)$. $\square$

Corollary 16.16 *If μ_1 and μ_2 are probability measures such that $\widehat{\mu}_1(t) = \widehat{\mu}_2(t)$ for all real t, then $\mu_1 = \mu_2$.*

Proof Let $\mathcal{I} = [a, b]$ be a given interval, and let $f(x)$ be determined so that $\widehat{f}(t)$ is the piecewise linear function whose graph passes through the points $(-\infty, 0)$, $(a - \delta, 0)$, $(a, 1)$, $(b, 1)$, $(b + \delta, 0)$, $(\infty, 0)$. Then $f(x) \ll_\delta x^{-2}$, and hence $f \in L^1(\mathbb{R})$. By Lemma 16.15 we see that

$$\int_{-\infty}^{\infty} \widehat{f}\,d\mu_1 = \int_{-\infty}^{\infty} \widehat{f}\,d\mu_2.$$

On the other hand, as δ tends to 0, the respective sides of this decrease to $\mu_i([a, b])$. Hence $\mu_1(\mathcal{I}) = \mu_2(\mathcal{I})$ for all closed bounded intervals $\mathcal{I}$, and it follows that $\mu_2 = \mu_2$. $\square$

We now characterize weak convergence in a number of useful ways.

Theorem 16.17 *If μ, μ_1, $\mu_2, \ldots$ are probability measures, then the following are equivalent:*

(a) *The μ_N tend weakly to μ;*
(b) *For every bounded continuous function f, $\int_{-\infty}^{\infty} f\,d\mu_N$ tends to $\int_{-\infty}^{\infty} f\,d\mu$;*
(c) *For each real number t, $\widehat{\mu}_N(t)$ tends to $\widehat{\mu}(t)$;*
(d) *For each real number $T > 0$, $\widehat{\mu}_N(t)$ tends to $\widehat{\mu}(t)$ uniformly for $|t| \le T$.*

Proof We first demonstrate that any one of the above conditions implies that the μ_N are tight. Suppose that (a) holds. There is an A such that $\mu([-A, A]) > 1 - \varepsilon$. Hence

$\mu_N([-A,A]) > 1-2\varepsilon$ for all $N > N_0$, and consequently there is a B such that $\mu_N([-B,B]) > 1-2\varepsilon$ for all N. Since both (b) and (d) include (c), we now suppose that (c) holds. Take $f(t) = \chi_{[-\delta,\delta]}(t)/(2\delta)$ in Lemma 16.15. Then

$$\frac{1}{2\delta}\int_{-\delta}^{\delta}\widehat{\mu}_N(t)\,dt = \int_{-\infty}^{\infty}\frac{\sin 2\pi\delta x}{2\pi\delta x}\,d\mu_N(x). \tag{16.10}$$

By the principle of dominated convergence, the left hand side tends to $\frac{1}{2\delta}\int_{-\delta}^{\delta}\widehat{\mu}(t)\,dt$ as N tends to infinity. But $\widehat{\mu}(0) = 1$ and $\widehat{\mu}$ is continuous at 0, so if δ is small, then

$$\left|\frac{1}{2\delta}\int_{-\delta}^{\delta}\widehat{\mu}(t)\,dt - 1\right| < \varepsilon.$$

Hence

$$\int_{-\infty}^{\infty}\frac{\sin 2\pi\delta x}{2\pi\delta x}\,d\mu_N(x) > 1-2\varepsilon$$

for all $N > N_0$. Here the integrand is ≤ 1 for all real x, and is $\le 1/2$ when $|x| \ge B = 1/\delta$, so that this integral is

$$\le \mu_N([-B,B]) + \frac{1}{2}\mu_N\{|x| > B\} = \frac{1}{2} + \frac{1}{2}\mu_N([-B,B]).$$

Hence $\mu_N([-B,B]) > 1-\varepsilon$ for $N > N_0$, and therefore the μ_N are tight.

We now derive (b) from (a). Suppose that $|f(x)| \le M$ for all x, that $F(x)$ is the distribution function of μ, and that F_N is the distribution function of μ_N. Since F is increasing, the set of points of discontinuity of F is at most countable. The same is true of each F_N, and hence there a set $\mathcal{D}$ that is at most countably infinite, such that if $x \notin \mathcal{D}$, then F and each of the functions F_N is continuous at x. Suppose that $\varepsilon > 0$ is given. Choose $A \notin \mathcal{C}$ so large that $\mu\{|x| > A\} < \varepsilon$ and $\mu_N\{|x| > A\} < \varepsilon$ for all N. Such an A exists because the family is tight. Hence

$$\left|\int_{-\infty}^{\infty} f(x)\,d\mu_N - \int_{-\infty}^{\infty} f(x)\,d\mu\right| \le \left|\int_{-A}^{A} f(x)\,d\mu_N - \int_{-A}^{A} f(x)\,d\mu\right| + 2M\varepsilon.$$

Since $f(x)$ is uniformly continuous on $[-A,A]$, we may choose numbers x_k, $-A = x_0 < x_1 < \ldots < x_k = A$, so that $|f(x) - f(x')| < \varepsilon$ if $x_{k-1} \le x \le x' \le x_k$ for some k. Moreover, these x_k may be chosen so that none of them is a member of $\mathcal{D}$. Then

$$\left|\int_{x_{k-1}}^{x_k} f(x)\,d\mu - f(x_k)\big(F(x_k) - F(x_{k-1})\big)\right| \le \varepsilon\big(F(x_k) - F(x_{k-1})\big).$$

This also holds with μ replaced by μ_N. On summing over k and using the triangle inequality, we find that the difference in question has absolute value not exceeding

$$M\sum_{k=1}^{K}|\mu_N([x_{k-1},x_k]) - \mu([x_{k-1},x_k])| + 2(M+1)\varepsilon.$$

From the hypothesis (a) it follows that this first term tends to 0 as N tends to infinity. Since ε is arbitrarily small, this gives (b).

We note that (c) is the special case $f(x) = e(-tx)$ of (b), and hence (b) implies (c).

Suppose that (c) holds. Since the μ_N are tight, it follows from Lemma 16.14 that the $\mu_N(t)$ are uniformly continuous both in t and in N. Hence (d) follows from (c).

To complete the proof it remains to show that (d) implies (a). Suppose that $f \in L^1(\mathbb{R})$. Then by the principle of dominated convergence

$$\lim_{N\to\infty} \int_{-\infty}^{\infty} f(t)\widehat{\mu}_N(t)\,dt = \int_{-\infty}^{\infty} f(t)\widehat{\mu}(t)\,dt,$$

and hence by Lemma 16.15,

$$\lim_{N\to\infty} \int_{-\infty}^{\infty} \widehat{f}(x)\,d\mu_N(x) = \int_{-\infty}^{\infty} \widehat{f}(x)\,d\mu(x).$$

Let f be chosen as in the proof of Corollary 16.16, so that $f \in L^1(\mathbb{R})$ and f is a piecewise-linear majorant of $\chi_{[a,b]}$. Then

$$\mu_N([a,b]) \le \int_{-\infty}^{\infty} \widehat{f}(x)\,d\mu_N(x) \longrightarrow \int_{-\infty}^{\infty} \widehat{f}(x)\,d\mu(x) \le \mu([a-\delta, b+\delta])$$

as $N \to \infty$. Here the last member tends to $\mu([a,b])$ as δ tends to 0, and thus we conclude that $\limsup_{N\to\infty} \mu_N([a,b]) \le \mu([a,b])$. Supposing that $a < b$, we may construct $f \in L^1(\mathbb{R})$ so that $\widehat{f}$ forms a piecewise-linear minorant of $\chi_{[a,b]}$, and thus similarly show that $\liminf_{N\to\infty} \mu_N([a,b]) \ge \lim_{\delta\to 0^+} \mu([a+\delta, b-\delta])$. This last expression is equal to $\mu([a,b])$ provided that $\mu(\{a\}) = \mu(\{b\}) = 0$, i.e. if the distribution function F is continuous at a and at b. Thus we have shown that

$$\lim_{N\to\infty} F_N(b) - F_N(a) = F(b) - F(a)$$

whenever a and b are points of continuity of F. Since the family μ_N is tight, the numbers $F_N(a)$ and $F(a)$ are near 0 if a is large and negative. Hence $\lim_{N\to\infty} F_N(b) = F(b)$, which is (a). □

The derivation of (a) from (c) could be constructed by appealing to Theorem 16.12, Theorem 16.13, and to Corollary 16.16, without using the mechanism used in the proof of Corollary 16.16. Conversely, we can derive Corollary 16.16 from Theorem 16.17 by taking $\mu = \mu_1$ and $\mu_N = \mu_2$ for all N.

We now sharpen Theorem 16.17 by showing that the limiting measure need not be given in advance.

Theorem 16.18 *Let $\mu_1, \mu_2, \ldots$ be probability measures, and suppose that for each t the sequence $\widehat{\mu}_1(t), \widehat{\mu}_2(t), \ldots$ converges. Call the limit $r(t)$. If r is continuous at 0, then the μ_N are weakly convergent to a probability measure μ, and $r(t) = \widehat{\mu}(t)$.*

Proof We note that $\widehat{\mu}_N(0) = 1$ for all N, so that $r(0) = 1$. In proving Theorem 16.17 we showed that condition (c) implies that the μ_N are tight. The only properties of $\widehat{\mu}(t)$ used in that argument were that $\widehat{\mu}(0) = 1$ and that $\widehat{\mu}(t)$ is continuous at 0. Hence by the same method we may show that the μ_N are tight. By Theorem 16.12 there is a subsequence $\mu_{N(k)}$ that is weakly convergent to a measure μ. By Theorem 16.13 we see that μ is a probability measure. By Theorem 16.17 applied to the subsequence $\mu_{N(k)}$, we deduce that $\widehat{\mu}_{N(k)}(t) \to \widehat{\mu}(t)$ as $k \to \infty$, for all real t. Hence $\widehat{\mu} = r$, and $\widehat{\mu}_N(t) \to \widehat{\mu}(t)$. Then by Theorem 16.18 again we conclude that the μ_N tend weakly to μ. □

We now characterize continuous measures in terms of the behaviour of $\widehat{\mu}$.

Theorem 16.19 *Let μ be a probability measure. Then μ is continuous if and only if*

$$\int_{-T}^{T} |\widehat{\mu}(t)|^2 \, dt = o(T)$$

as T tends to ∞.

Proof The left hand side is

$$\begin{aligned}\int_{-T}^{T}\int_{-\infty}^{\infty} e(-tx)\,d\mu(x)\int_{-\infty}^{\infty} e(ty)\,d\mu(y)\,dt &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-T}^{T} e(t(y-x))\,dt\,d\mu(x)\,d\mu(y)\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\sin 2\pi T(y-x)}{\pi(y-x)}\,d\mu(x)\,d\mu(y).\end{aligned}$$

If we divide by $2T$, then by the principle of dominated convergence this latter integral tends to

$$\iint_{x=y} d\mu(x)\,d\mu(y) = \sum_i d_i^2$$

as T tends to infinity, where the d_i are the heights of the jump discontinuities of μ. But μ is continuous if and only if $\sum d_i^2 = 0$, so we have the result. □

16.5.1 Exercises

1. Suppose that $\mu_1, \mu_2, \ldots$ are probability measures such that for each convergent subsequence the limit is a probability measure. Show that the sequence is tight.

2. Construct an example to show that the condition (d) of Theorem 16.17 need not hold uniformly for all t, but only in compact ranges.

3. Suppose that μ_N is a sequence of probability measures on $\mathbb{T}$, and that μ is also a probability measure on T. We say that the μ_N tend weakly to μ if $\mu_N([\alpha,\beta]) \to \mu([\alpha,\beta])$ whenever $0 \le \beta - \alpha \le 1$ and $\mu(\{\alpha\}) = \mu(\{\beta\}) = 0$. Show that the following are equivalent:
   (a) The μ_N tend weakly to μ;
   (b) For each integer k, $\widehat{\mu}_N(k)$ tends to $\widehat{\mu}(k)$ as N tends to infinity;
   (c) If f is a continuous function defined on $\mathbb{T}$, then $\int_{\mathbb{T}} f(x)\,d\mu_N(x)$ tends to $\int_{\mathbb{T}} f(x)\,d\mu$ as N tends to infinity.

4. Let μ be the measure on $\mathbb{R}$ defined by $d\mu = (2\pi)^{-1/2}e^{-x^2/2}\,d\lambda$ where λ denotes Lebesgue measure. Show that μ is a probability measure. (This is called *normal distribution* with mean 0 and variance 1.) Show that $\int_{\mathbb{R}} x\,d\mu(x) = 0$, that $\int x^2\,d\mu(x) = 1$, and that $\widehat{\mu}(t) = e^{-2\pi^2 t^2}$.

5. Let p be a fixed real number, $0 < p < 1$. For each positive integer N let μ_N be the discrete measure that has a point mass of weight $\binom{N}{n}p^n(1-p)^{N-n}$ at $(n - Np)/\sigma$ for $n = 0, 1, 2, \ldots, N$ where $\sigma = \sqrt{np(1-p)}$. Show that μ_N is a probability measure. Show that $\widehat{\mu}_N(t) = \big(pe(-t(1-p)/\sigma) + (1-p)e(-tp/\sigma)\big)^N$. Show that if T is given, then $\widehat{\mu}_N(t) = e^{-2\pi^2 t^2} + O(n^{-1/2})$ as N tends to infinity, uniformly for $|t| \le T$. Conclude that the measures μ tend weakly to the measure μ of the preceding exercise. (This is a special case of the Central Limit Theorem of probability theory.)