Chapter 2

Weak Convergence

2.1 Characteristic Functions

If α is a probability distribution on the line, its characteristic function is defined by

$$\phi(t) = \int \exp[\, i\, t\, x\,]\, d\alpha. \tag{2.1}$$

The above definition makes sense. We write the integrand e^{itx} as $\cos tx + i \sin tx$ and integrate each part to see that

$$|\phi(t)| \le 1$$

for all real t.

Exercise 2.1. Calculate the characteristic functions for the following distributions:

1. α is the degenerate distribution δ_a with probability one at the point a.

2. α is the binomial distribution with probabilities

   $$p_k = Prob[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$$

   for $0 \le k \le n$.

Theorem 2.1. *The characteristic function $\phi(t)$ of any probability distribution is a uniformly continuous function of t that is positive definite, i.e. for any n complex numbers $\xi_1, \cdots, \xi_n$ and real numbers $t_1, \cdots, t_n$*

$$\sum_{i,j=1}^{n} \phi(t_i - t_j)\, \xi_i\, \bar{\xi}_j \geq 0.$$

Proof. Let us note that

$$\begin{aligned} \sum_{i,j=1}^{n} \phi(t_i - t_j)\, \xi_i\, \bar{\xi}_j \; &= \textstyle\sum_{i,j=1}^{n} \xi_i\, \bar{\xi}_j \int \exp[\, i(t_i - t_j)x]\, d\alpha \\ &= \textstyle\int \big| \sum_{j=1}^{n} \xi_j \exp[i\, t_j\, x] \big|^2\, d\alpha \geq 0. \end{aligned}$$

To prove uniform continuity we see that

$$|\phi(t) - \phi(s)| \leq \int |\exp[\, i(t - s)\, x\,] - 1|\, dP$$

which tends to 0 by the bounded convergence theorem if $|t - s| \to 0$. ∎

The characteristic function of course carries some information about the distribution α. In particular if $\int |x|\, d\alpha < \infty$, then $\phi(\cdot)$ is continuously differentiable and $\phi'(0) = i \int x\, d\alpha$.

Exercise 2.2. Prove it!

Warning: The converse need not be true. $\phi(\cdot)$ can be continuously differentiable but $\int |x|\, dP$ could be ∞.

Exercise 2.3. Construct a counterexample along the following lines. Take a discrete distribution, symmetric around 0 with

$$\alpha\{n\} = \alpha\{-n\} = p(n) \simeq \frac{1}{n^2 \log n}\ .$$

Then show that $\sum_n \frac{(1 - \cos nt)}{n^2 \log n}$ is a continuously differentiable function of t.

Exercise 2.4. The story with higher moments $m_r = \int x^r\, d\alpha$ is similar. If any of them, say m_r exists, then $\phi(\cdot)$ is r times continuously differentiable and $\phi^{(r)}(0) = i^r m_r$. The converse is false for odd r, but true for even r by an application of Fatou's lemma.

The next question is how to recover the distribution function $F(x)$ from $\phi(t)$. If we go back to the Fourier inversion formula, see for instance [2], we can 'guess', using the fundamental theorem of calculus and Fubini's theorem, that

$$F'(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp[-i\,t\,x\,]\,\phi(t)\,dt$$

and therefore

$$\begin{aligned} F(b) - F(a) &= \tfrac{1}{2\pi}\int_a^b dx \int_{-\infty}^{\infty} \exp[-i\,t\,x\,]\,\phi(t)\,dt \\ &= \tfrac{1}{2\pi}\int_{-\infty}^{\infty} \phi(t)\,dt \int_a^b \exp[-i\,t\,x\,]\,dx \\ &= \tfrac{1}{2\pi}\int_{-\infty}^{\infty} \phi(t)\tfrac{\exp[-\,i\,t\,b\,]-\exp[-\,i\,t\,a\,]}{-\,i\,t}\,dt \\ &= \lim_{T\to\infty} \tfrac{1}{2\pi}\int_{-T}^{T} \phi(t)\tfrac{\exp[-\,i\,t\,b\,]-\exp[-\,i\,t\,a\,]}{-\,i\,t}\,dt. \end{aligned}$$

We will in fact prove the final relation, which is a principal value integral, provided a and b are points of continuity of F. We compute the right hand side as

$$\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{\exp[-\,i\,t\,b\,]-\exp[-\,i\,t\,a\,]}{-\,i\,t}\,dt\int \exp[\,i\,t\,x\,]\,d\alpha$$

$$\begin{aligned} &= \lim_{T\to\infty}\tfrac{1}{2\pi}\int d\alpha\int_{-T}^{T}\tfrac{\exp[i\,t\,(x-b)\,]-\exp[i\,t\,(x-a)\,]}{-\,i\,t}\,dt \\ &= \lim_{T\to\infty}\tfrac{1}{2\pi}\int d\alpha\int_{-T}^{T}\tfrac{\sin t\,(x-a)-\sin t\,(x-b)}{t}\,dt \\ &= \tfrac{1}{2}\int[\mathrm{sign}(x-a)-\mathrm{sign}(x-b)]\,d\alpha \\ &= F(b)-F(a) \end{aligned}$$

provided a and b are continuity points. We have applied Fubini's theorem and the bounded convergence theorem to take the limit as $T \to \infty$. Note that the Dirichlet integral

$$u(t,z) = \int_0^T \frac{\sin tz}{t}\,dt$$

satisfies $\sup_{T,z}|u(T,z)| \le C$ and

$$\lim_{T\to\infty} u(T,z) = \begin{cases} 1 & \text{if } z>0 \\ -1 & \text{if } z<0 \\ 0 & \text{if } z=0. \end{cases}$$

As a consequence we conclude that the distribution function and hence α is determined uniquely by the characteristic function.

Exercise 2.5. Prove that if two distribution functions agree on the set of points at which they are both continuous, they agree everywhere.

Besides those in Exercise 2.1, some additional examples of probability distributions and the corresponding characteristic functuions are given below.

1. The Poisson distribution of 'rare events', with rate λ, has probabilities $P[X = r] = e^{-\lambda}\frac{\lambda^r}{r!}$ for $r \geq 0$. Its characteristic function is

$$\phi(t) = \exp[\lambda(e^{it} - 1)].$$

2. The geometric distribution, the distribution of the number of unsuccessful attempts preceeding a success has $P[X = r] = pq^r$ for $r \geq 0$.Its characteristic function is

$$\phi(t) = p(1 - qe^{it})^{-1}.$$

3. The negative binomial distribution, the probabilty distribution of the number of accumulated failures before k successes, with $P[X = r] = \binom{k+r-1}{r}p^k q^r$ has the characteristic function is

$$\phi(t) = p^k(1 - qe^{it})^{-k}.$$

We now turn to some common continuous distributions, in fact given by 'densities' $f(x)$ i.e the distribution functions are given by $F(x) = \int_{-\infty}^{x} f(y)\, dy$

1. The 'uniform ' distribution with density $f(x) = \frac{1}{b-a}$, $a \leq x \leq b$ has characteristic function

$$\phi(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}.$$

In particular for the case of a symmetric interval $[-a, a]$,

$$\phi(t) = \frac{\sin at}{at}.$$

2. The gamma distribution with density $f(x) = \frac{c^p}{\Gamma(p)} e^{-cx} x^{p-1}$, $x \geq 0$ has the characteristic function

$$\phi(t) = (1 - \frac{it}{c})^{-p}.$$

where $c > 0$ is any constant. A special case of the gamma distribution is the exponential distribution, that corresponds to $c = p = 1$ with density $f(x) = e^{-x}$ for $x \geq 0$. Its characteristic function is given by

$$\phi(t) = [1 - it]^{-1}.$$

3. The two sided exponential with density $f(x) = \frac{1}{2} e^{-|x|}$ has characteristic function

$$\phi(t) = \frac{1}{1 + t^2}.$$

4. The Cauchy distribution with density $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ has the characteristic function

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

5. The normal or Gaussian distribution with mean μ and variance σ^2, which has a density of $\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ has the characteristic function given by

$$\phi(t) = e^{it\mu - \frac{\sigma^2 t^2}{2}}.$$

In general if X is a random variable which has distribution α and a characteristic function $\phi(t)$, the distribution β of $aX + b$, can be written as $\beta(A) = \alpha\,[x : ax + b \in A]$ and its characteristic function $\psi(t)$ can be expressed as $\psi(t) = e^{ita}\phi(bt)$. In particular the characteristic function of $-X$ is $\phi(-t) = \overline{\phi(t)}$. Therefore the distribution of X is symmetric around $x = 0$ if and only if $\phi(t)$ is real for all t.

2.2 Moment Generating Functions

If α is a probability distribution on $\mathbf{R}$, for any integer $k \geq 1$, the moment m_k of α is defined as

$$m_k = \int x^k d\alpha. \tag{2.2}$$

Or equivalently the k-th moment of a random variable X is

$$m_k = E[X^k] \tag{2.3}$$

By convention one takes $m_0 = 1$ even if $P[X = 0] > 0$. We should note that if k is odd, in order for m_k to be defined we must have $E[|X|^k] = \int |x|^k d\alpha < \infty$. Given a distribution α, either all the moments exist, or they exist only for $0 \leq k \leq k_0$ for some k_0. It could happen that $k_0 = 0$ as is the case with the Cauchy distribution. If we know all the moments of a distribution α, we know the expectations $\int p(x)d\alpha$ for every polynomial $p(\cdot)$. Since polynomials $p(\cdot)$ can be used to approximate (by Stone-Weierstrass theorem) any continuous function, one might hope that, from the moments, one can recover the distribution α. This is not as staright forward as one would hope. If we take a bounded continuous function, like $\sin x$ we can find a sequence of polynomials $p_n(x)$ that converges to $\sin x$. But to conclude that

$$\int \sin x d\alpha = \lim_{n\to\infty} \int p_n(x)d\alpha$$

we need to control the contribution to the integral from large values of x, which is the role of the dominated convergence theorem. If we define $p^*(x) = \sup_n |p_n(x)|$ it would be a big help if $\int p^*(x)d\alpha$ were finite. But the degrees of the polynomials p_n have to increase indefinitely with n because $\sin x$ is a transcendental function. Therefore $p^*(\cdot)$ must grow faster than a polynomial at ∞ and the condition $\int p^*(x)d\alpha < \infty$ may not hold.

In general, it is not true that moments determine the distribution. If we look at it through characteristic functions, it is the problem of trying to recover the function $\phi(t)$ from a knowledge of all of its derivatives at $t = 0$. The Taylor series at $t = 0$ may not yield the function. Of course we have more information in our hands, like positive definiteness etc. But still it is likely that moments do not in general determine α. In fact here is how to construct an example.

We need nonnegative numbers $\{a_n\}, \{b_n\} : n \geq 0$, such that

$$\sum_n a_n e^{k\,n} = \sum_n b_n e^{k\,n} = m_k$$

for every $k \geq 0$. We can then replace them by $\{\frac{a_n}{m_0}\}, \{\frac{b_n}{m_0}\} : n \geq 0$ so that $\sum_k a_k = \sum_k b_k = 1$ and the two probability distributions

$$P[X = e^n] = a_n, \qquad P[X = e^n] = b_n$$

will have all their moments equal. Once we can find $\{c_n\}$ such that

$$\sum_n c_n e^{n\,z} = 0 \qquad \textit{for } z = 0, 1, \cdots$$

we can take $a_n = \max(c_n, 0)$ and $b_n = \max(-c_n, 0)$ and we will have our example. The goal then is to construct $\{c_n\}$ such that $\sum_n c_k z^n = 0$ for $z = 1, e, e^2, \cdots$. Borrowing from ideas in the theory of a complex variable, (see Weierstrass factorization theorem, [1]) we define

$$C(z) = \Pi_{n=1}^{\infty}(1 - \frac{z}{e^n})$$

and expand $C(z) = \sum c_n z^n$. Since $C(z)$ is an entire function, the coefficients c_n satisfy $\sum_n |c_n| e^{k\,n} < \infty$ for every k.

There is in fact a positive result as well. If α is such that the moments $m_k = \int x^k d\alpha$ do not grow too fast, then α is determined by m_k.

Theorem 2.2. *Let m_k be such that $\sum_k m_{2k} \frac{a^{2k}}{(2k)!} < \infty$ for some $a > 0$. Then there is atmost one distribution α such that $\int x^k d\alpha = m_k$.*

Proof. We want to determine the characteristic function $\phi(t)$ of α. First we note that if α has moments m_k satisfying our assumption, then

$$\int \cosh(ax) d\alpha = \sum_k \frac{a^{2k}}{(2k)!} m_{2k} < \infty$$

by the monotone convergence theorem. In particular

$$\psi(u + it) = \int e^{(u+it)x} d\alpha$$

is well define as an analytic function of $z = u + it$ in the strip $|u| < a$. From the theory of functions of a complex variable we know that the function $\psi(\cdot)$ is uniquely determined in the strip by its derivatives at 0, i.e. $\{m_k\}$. In particular $\phi(t) = \psi(0 + it)$ is determined as well $\square$

2.3 Weak Convergence

One of the basic ideas in establishing Limit Theorems is the notion of **weak convergence** of a sequence of probability distributions on the line **R**. Since the role of a probability measure is to assign probabilities to sets, we should expect that if two probability measures are to be close, then they should assign for a given set, probabilities that are nearly equal. This suggets the definition

$$d(P_1, P_2) = \sup_{A \in \mathcal{B}} |P_1(A) - P_2(A)|$$

as the distance between two probability measures P_1 and P_2 on a measurable space $(\Omega, \mathcal{B})$. This is too strong. If we take P_1 and P_2 to be degenerate distributions with probability 1 concentrated at two points x_1 and x_2 on the line one can see that, as soon as $x_1 \neq x_2$, $d(P_1, P_2) = 1$, and the above metric is not sensitive to how close the two points x_1 and x_2 are. It only cares that they are unequal. The problem is not because of the supremum. We can take A to be an interval $[a, b]$ that includes x_1 but omits x_2 and $|P_1(A) - P_2(A)| = 1$. On the other hand if the end points of the interval are kept away from x_1 or x_2 the situation is not that bad. This leads to the following definition.

Definition 2.1. *A sequence α_n of probability distributions on* **R** *is said to converge weakly to a probability distribution α if,*

$$\lim_{n \to \infty} \alpha_n[I] = \alpha[I]$$

for any interval $I = [a, b]$ such that the single point sets a and b have probability 0 under α.

One can state this equivalently in terms of the distribution functions $F_n(x)$ and $F(x)$ corrresponding to the measures α_n and α respectively.

Definition 2.2. *A sequence α_n of probability measures on the real line* **R** *with distribution functions $F_n(x)$ is said to converge weakly to a limiting probability measure α with distribution function $F(x)$ (in symbols $\alpha_n \Rightarrow \alpha$ or $F_n \Rightarrow F$) if*

$$\lim_{n \to \infty} F_n(x) = F(x)$$

for every x that is a continuity point of F.

Exercise 2.6. prove the equivalence of the two definitions.

Remark 2.1. One says that a sequence X_n of random variables **converges in law** or in distribution to X if the distributions α_n of X_n converges weakly to the distribution α of X.

There are equivalent formulations in terms of expectations and characteristic functions.

Theorem 2.3. (Lévy-Cramér continuity theorem) *The following are equivalent.*

1. $\alpha_n \Rightarrow \alpha$ *or* $F_n \Rightarrow F$

2. *For every bounded continuous function* $f(x)$ *on* $\mathbf{R}$

$$\lim_{n\to\infty} \int_{\mathbf{R}} f(x)\, d\alpha_n = \int_{\mathbf{R}} f(x)\, d\alpha$$

3. *If* $\phi_n(t)$ *and* $\phi(t)$ *are respectively the characteristic functions of* α_n *and* α*, for every real* t*,*

$$\lim_{n\to\infty} \phi_n(t) = \phi(t)$$

Proof. We first prove $(a \Rightarrow b)$. Let $\epsilon > 0$ be arbitrary. Find continuity points a and b of F such that $a < b$, $F(a) \leq \epsilon$ and $1 - F(b) \leq \epsilon$. Since $F_n(a)$ and $F_n(b)$ converge to $F(a)$ and $F(b)$, for n large enough, $F_n(a) \leq 2\epsilon$ and $1 - F_n(b) \leq 2\epsilon$. Divide the interval $[a, b]$ into a finite number $N = N_\delta$ of small subintervals $I_j = (a_j, a_{j+1}], 1 \leq j \leq N$ with $a = a_1 < a_2 < \cdots < a_{N+1} = b$ such that all the end points $\{a_j\}$ are points of continuity of F and the oscillation of the continuous function f in each I_j is less than a preassigned number δ. Since any continuous function f is uniformly continuous in the closed bounded (compact) interval $[a, b]$, this is always possible for any given $\delta > 0$. Let $h(x) = \sum_{j=1}^N \chi_{I_j} f(a_j)$ be the simple function equal to $f(a_j)$ on I_j and 0 outside $\cup_j I_j = (a, b]$. We have $|f(x) - h(x)| \leq \delta$ on $(a, b]$. If $f(x)$ is bounded by M, then

$$\left| \int f(x)\, d\alpha_n - \sum_{j=1}^{N} f(a_j)[F_n(a_{j+1}) - F_n(a_j)] \right| \leq \delta + 4M\epsilon \qquad (2.4)$$

and

$$\left| \int f(x)\, d\alpha - \sum_{j=1}^{N} f(a_j)[F(a_{j+1}) - F(a_j)] \right| \le \delta + 2M\epsilon. \tag{2.5}$$

Since $\lim_{n\to\infty} F_n(a_j) = F(a_j)$ for every $1 \le j \le N$, we conclude from equations (2.4), (2.5) and the triangle inequality that

$$\limsup_{n\to\infty} \left| \int f(x)\, d\alpha_n - \int f(x)\, d\alpha \right| \le 2\delta + 6M\epsilon.$$

Since ϵ and δ are arbitrary small numbers we are done.

Because we can make the choice of $f(x) = \exp[i\,t\,x\,] = \cos tx + i \sin tx$, which for every t is a bounded and continuous function $(b \Rightarrow c)$ is trivial.

$(c \Rightarrow a)$ is the hardest. It is carried out in several steps. Actually we will prove a stronger version as a separate theorem. ☐

Theorem 2.4. *For each $n \ge 1$, let $\phi_n(t)$ be the characteristic function of a probability distribution α_n. Assume that $\lim_{n\to\infty} \phi_n(t) = \phi(t)$ exists for each t and $\phi(t)$ is continuous at $t = 0$. Then $\phi(t)$ is the characteristic function of some probability distribution α and $\alpha_n \Rightarrow \alpha$.*

Proof.

Step 1. Let $r_1, r_2, \cdots$ be an enumeration of the rational numbers. For each j consider the sequence $\{F_n(r_j) : n \ge 1\}$ where F_n is the distribution function corresponding to $\phi_n(\cdot)$. It is a sequence bounded by 1 and we can extract a subsequence that converges. By the diagonalization process we can choose a subseqence $G_k = F_{n_k}$ such that

$$\lim_{k\to\infty} G_k(r) = b_r$$

exists for every rational number r. From the monotonicity of F_n in x we conclude that if $r_1 < r_2$, then $b_{r_1} \le b_{r_2}$.

Step 2. From the skeleton b_r we reconstruct a right continuous monotone function $G(x)$. We define

$$G(x) = \inf_{r > x} b_r.$$

Clearly if $x_1 < x_2$, then $G(x_1) \le G(x_2)$ and therefore G is nondecreasing. If $x_n \downarrow x$, any $r > x$ satisfies $r > x_n$ for sufficiently large n. This allows us to conclude that $G(x) = \inf_n G(x_n)$ for any sequence $x_n \downarrow x$, proving that $G(x)$ is right continuous.

Step 3. Next we show that at any continuity point x of G

$$\lim_{n\to\infty} G_n(x) = G(x).$$

Let $r > x$ be a rational number. Then $G_n(x) \le G_n(r)$ and $G_n(r) \to b_r$ as $n \to \infty$. Hence

$$\limsup_{n\to\infty} G_n(x) \le b_r.$$

This is true for every rational $r > x$, and therefore taking the infimum over $r > x$

$$\limsup_{n\to\infty} G_n(x) \le G(x).$$

Suppose now that we have $y < x$. Find a rational r such that $y < r < x$.

$$\liminf_{n\to\infty} G_n(x) \ge \liminf_{n\to\infty} G_n(r) = b_r \ge G(y).$$

As this is true for every $y < x$,

$$\liminf_{n\to\infty} G_n(x) \ge \sup_{y<x} G(y) = G(x-0) = G(x)$$

the last step being a consequence of the assumption that x is a point of continuity of G i.e. $G(x-0) = G(x)$.

Warning. This does not mean that G is necessarily a distribution function. Consider $F_n(x) = 0$ for $x < n$ and 1 for $x \ge n$, which corresponds to the distribution with the entire probability concentrated at n. In this case $\lim_{n\to\infty} F_n(x) = G(x)$ exists and $G(x) \equiv 0$, which is not a distribution function.

Step 4. We will use the continuity at $t = 0$, of $\phi(t)$, to show that G is indeed a distribution function. If $\phi(t)$ is the characteristic function of α

$$\begin{aligned}
\tfrac{1}{2T}\int_{-T}^{T}\phi(t)\,dt &= \int\Big[\tfrac{1}{2T}\int_{-T}^{T}\exp[i\,t\,x\,]\,dt\Big]\,d\alpha \\
&= \int \tfrac{\sin Tx}{Tx}\,d\alpha \\
&\leq \int \big|\tfrac{\sin Tx}{Tx}\big|\,d\alpha \\
&= \int_{|x|<\ell}\big|\tfrac{\sin Tx}{Tx}\big|\,d\alpha + \int_{|x|\geq\ell}\big|\tfrac{\sin Tx}{Tx}\big|\,d\alpha \\
&\leq \alpha[|x|<\ell] + \tfrac{1}{T\ell}\,\alpha[|x|\geq\ell].
\end{aligned}$$

We have used Fubini's theorem in the first line and the bounds $|\sin x| \leq |x|$ and $|\sin x| \leq 1$ in the last line. We can rewrite this as

$$\begin{aligned}
1-\tfrac{1}{2T}\int_{-T}^{T}\phi(t)\,dt &\geq 1-\alpha[|x|<\ell] - \tfrac{1}{T\ell}\,\alpha[|x|\geq\ell] \\
&= \alpha[|x|\geq\ell] - \tfrac{1}{T\ell}\,\alpha[|x|\geq\ell] \\
&= \big(1-\tfrac{1}{T\ell}\big)\alpha[|x|\geq\ell] \\
&\geq \big(1-\tfrac{1}{T\ell}\big)[1-F(\ell)+F(-\ell)]
\end{aligned}$$

Finally, if we pick $\ell = \frac{2}{T}$,

$$[1-F(\frac{2}{T})+F(-\frac{2}{T})] \leq 2\Big[1-\frac{1}{2T}\int_{-T}^{T}\phi(t)\,dt\Big].$$

Since this inequality is valid for any distribution function F and its characteristic function ϕ, we conclude that, for every $k \geq 1$,

$$[1-F_{n_k}(\frac{2}{T})+F_{n_k}(-\frac{2}{T})] \leq 2\Big[1-\frac{1}{2T}\int_{-T}^{T}\phi_{n_k}(t)\,dt\Big]. \tag{2.6}$$

We can pick T such that $\pm\frac{2}{T}$ are continuity points of G. If we now pass to the limit and use the bounded convergence theorem on the right hand side of equation (2.6), we obtain

$$[1-G(\frac{2}{T})+G(-\frac{2}{T})] \leq 2\Big[1-\frac{1}{2T}\int_{-T}^{T}\phi(t)\,dt\Big].$$

Since $\phi(0) = 1$ and ϕ is continuous at $t = 0$, by letting $T \to 0$ in such a way that $\pm\frac{2}{T}$ are continuity points of G, we conclude that

$$1-G(\infty)+G(-\infty) = 0$$

or G is indeed a distribution function

Step 5. We now complete the rest of the proof, i.e. show that $\alpha_n \Rightarrow \alpha$. We have $G_k = F_{n_k} \Rightarrow G$ as well as $\psi_k = \phi_{n_k} \to \phi$. Therefore G must equal F which has ϕ for its characteristic function. Since the argument works for any subsequence of F_n, every subsequence of F_n will have a further subsequence that converges weakly to the same limit F uniquely determined as the distribution function whose characteristic function is $\phi(\cdot)$. Consequently $F_n \Rightarrow F$ or $\alpha_n \Rightarrow \alpha$. ☐

Exercise 2.7. How do you actually prove that if every subsequence of a sequence $\{F_n\}$ has a further subsequence that converges to a common F then $F_n \Rightarrow F$?

Definition 2.3. *A subset $\mathcal{A}$ of probability distributions on* $\mathbf{R}$ *is said to be totally bounded if, given any sequence α_n from $\mathcal{A}$, there is subsequence that converges weakly to some limiting probability distribution α.*

Theorem 2.5. *In order that a family $\mathcal{A}$ of probability distributions be totally bounded it is necessary and sufficient that either of the following equivalent conditions hold.*

$$\lim_{\ell\to\infty} \sup_{\alpha\in\mathcal{A}} \alpha[x : |x| \geq \ell] = 0 \tag{2.7}$$

$$\lim_{h\to 0} \sup_{\alpha\in\mathcal{A}} \sup_{|t|\leq h} |1 - \phi_\alpha(t)| = 0. \tag{2.8}$$

Here $\phi_\alpha(t)$ is the characteristic function of α.

The condition in equation (2.7) is often called the uniform tightness property.

Proof. The proof is already contained in the details of the proof of the earlier theorem. We can always choose a subsequence such that the distribution functions converge at rationals and try to reconstruct the limiting distribution function from the limits at rationals. The crucial step is to prove that the limit is a distribution function. Either of the two conditions (2.7) or (2.8) will guarantee this. If condition (2.7) is violated it is straight forward to pick a sequence from $\mathcal{A}$ for which the distribution functions have a limit which is not a distribution function. Then $\mathcal{A}$ cannot be totally bounded. Condition (2.7) is therefore necessary. That a)$\Rightarrow$ b), is a consequence of the estimate

$$\begin{aligned} |1-\phi(t)| &\le \int |\exp[\,i\,t\,x\,]-1|\,d\alpha \\ &= \int_{|x|\le\ell} |\exp[\,i\,t\,x\,]-1|\,d\alpha + \int_{|x|>\ell} |\exp[\,i\,t\,x\,]-1|\,d\alpha \\ &\le |t|\ell + 2\alpha[x : |x| > \ell] \end{aligned}$$

It is a well known principle in Fourier analysis that the regularity of $\phi(t)$ at $t = 0$ is related to the decay rate of the tail probabilities. $\square$

Exercise 2.8. Compute $\int |x|^p\,d\alpha$ in terms of the characteristic function $\phi(t)$ for p in the range $0 < p < 2$.

Hint: Look at the formula

$$\int_{-\infty}^{\infty} \frac{1-\cos tx}{|t|^{p+1}} dt = C_q |x|^p$$

and use Fubini's theorem.

We have the following result on the behavior of $\alpha_n(A)$ for certain sets whenever $\alpha_n \Rightarrow \alpha$.

Theorem 2.6. *Let $\alpha_n \Rightarrow \alpha$ on $\mathbf{R}$. If $C \subset \mathbf{R}$ is closed set then*

$$\limsup_{n\to\infty} \alpha_n(C) \le \alpha(C)$$

while for open sets $G \subset \mathbf{R}$

$$\liminf_{n\to\infty} \alpha_n(G) \ge \alpha(G)$$

If $A \subset \mathbf{R}$ is a continuity set of α i.e. $\alpha(\partial A) = \alpha(\bar{A} - A^o) = 0$, then

$$\lim_{n\to\infty} \alpha_n(A) = \alpha(A)$$

Proof. The function $d(x, C) = \inf_{y\in C} |x - y|$ is a continuous and equals 0 precisely on C.

$$f(x) = \frac{1}{1 + d(x, C)}$$

is a continuous function bounded by 1, that is equal to 1 precisely on C and

$$f_k(x) = [f(x)]^k \downarrow \chi_{_C}(x)$$

as $k \to \infty$. For every $k \geq 1$, we have

$$\lim_{n\to\infty} \int f_k(x)\, d\alpha_n = \int f_k(x)\ d\alpha$$

and therefore

$$\limsup_{n\to\infty} \alpha_n(C) \leq \lim_{n\to\infty} \int f_k(x)\, d\alpha_n = \int f_k(x) d\alpha.$$

Letting $k \to \infty$ we get

$$\limsup_{n\to\infty} \alpha_n(C) \leq \alpha(C).$$

Taking complements we conclude that for any open set $G \subset \mathbf{R}$

$$\liminf_{n\to\infty} \alpha_n(G) \geq \alpha(G).$$

Combining the two parts, if $A \subset \mathbf{R}$ is a continuity set of α i.e. $\alpha(\partial A) = \alpha(\bar{A} - A^o) = 0$, then

$$\lim_{n\to\infty} \alpha_n(A) = \alpha(A).$$

$\square$

We are now ready to prove the converse of Theorem 2.1 which is the hard part of a theorem of Bochner that characterizes the characteristic functions of probability distributions as continuous positive definite functions on $\mathbf{R}$ normalized to be 1 at 0.

Theorem 2.7. (Bochner's Theorem). *If $\phi(t)$ is a positive definite function which is continuous at $t = 0$ and is normalized so that $\phi(0) = 1$, then ϕ is the characteristic function of some probability ditribution on $\mathbf{R}$.*

Proof. The proof depends on constructing approximations $\phi_n(t)$ which are in fact characteristic functions and satisfy $\phi_n(t) \to \phi(t)$ as $n \to \infty$. Then we can apply the preceeding theorem and the probability measures corresponding to ϕ_n will have a weak limit which will have ϕ for its characteristic function.

Step 1. Let us establish a few elementary properties of positive definite functions.

1) If $\phi(t)$ is a positive definite function so is $\phi(t)exp[\,i\,t\,a\,]$ for any real a. The proof is elementary and requires just direct verification.

2) If $\phi_j(t)$ are positive definite for each j then so is any linear combination $\phi(t) = \sum_j w_j \phi_j(t)$ with nonnegative weights w_j. If each $\phi_j(t)$ is normalized with $\phi_j(0) = 1$ and $\sum_j w_j = 1$, then of of course $\phi(0) = 1$ as well.

3) If ϕ is positive definite then ϕ satisfies $\phi(0) \geq 0$, $\phi(-t) = \overline{\phi(t)}$ and $|\phi(t)| \leq \phi(0)$ for all t.

We use the fact that the matrix $\{\phi(t_i - t_j) : 1 \leq i, j \leq n\}$ is Hermitian positive definite for any n real numbers $t_1, \cdots, t_n$. The first assertion follows from the the positivity of $\phi(0)|z|^2$, the second is a consequence of the Hermitian property and if we take $n = 2$ with $t_1 = t$ and $t_2 = 0$ as a consequence of the positive definiteness of the 2×2 matrix we get $|\phi(t)|^2 \leq |\phi(0)|^2$

4) For any s, t we have $|\phi(t) - \phi(s)|^2 \leq 4\phi(0)|\phi(0) - \phi(t-s)|$

We use the positive definiteness of the 3×3 matrix

$$\begin{bmatrix} 1 & \phi(t-s) & \phi(t) \\ \overline{\phi(t-s)} & 1 & \phi(s) \\ \overline{\phi(t)} & \overline{\phi(s)} & 1 \end{bmatrix}$$

which is $\{\phi(t_i - t_j)\}$ with $t_1 = t, t_2 = s$ and $t_3 = 0$. In particular the determinant has to be nonnegative.

$$\begin{aligned} 0 &\leq 1 + \phi(s)\phi(t-s)\overline{\phi(t)} + \overline{\phi(s)\phi(t-s)}\phi(t) - |\phi(s)|^2 \\ &\quad -|\phi(t)|^2 - |\phi(t-s)|^2 \\ &= 1 - |\phi(s) - \phi(t)|^2 - |\phi(t-s)|^2 - \phi(t)\overline{\phi(s)}(1 - \phi(t-s)) \\ &\quad -\overline{\phi(t)}\phi(s)(1 - \overline{\phi(t-s)}) \\ &\leq 1 - |\phi(s) - \phi(t)|^2 - |\phi(t-s)|^2 + 2|1 - \phi(t-s)| \end{aligned}$$

Or

$$\begin{aligned} |\phi(s) - \phi(t)|^2 &\leq 1 - |\phi(s-t)|^2 + 2|1 - \phi(t-s)| \\ &\leq 4|1 - \phi(s-t)| \end{aligned}$$

5) It now follows from 4) that if a positive definite function is continuous at $t = 0$, it is continuous everywhere (in fact uniformly continuous).

Step 2. First we show that if $\phi(t)$ is a positive definite function which is continuous on $\mathbf{R}$ and is absolutely integrable, then

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp[-\,i\,t\,x\,]\phi(t)\,dt \geq 0$$

is a continuous function and

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

Moreover the function

$$F(x) = \int_{-\infty}^{x} f(y)\,dy$$

defines a distribution function with characteristic function

$$\phi(t) = \int_{-\infty}^{\infty} \exp[i\,t\,x\,]f(x)\,dx. \tag{2.9}$$

If ϕ is integrable on $(-\infty,\infty)$, then $f(x)$ is clearly bounded and continuous. To see that it is nonnegative we write

$$\begin{aligned}
f(x) &= \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\big(1-\frac{|t|}{T}\big)e^{-\,i\,t\,x}\phi(t)\,dt && (2.10)\\
&= \lim_{T\to\infty}\frac{1}{2\pi T}\int_{0}^{T}\int_{0}^{T} e^{-\,i\,(t-s)\,x}\phi(t-s)\,dt\,ds && (2.11)\\
&= \lim_{T\to\infty}\frac{1}{2\pi T}\int_{0}^{T}\int_{0}^{T} e^{-\,i\,t\,x}\;e^{i\,s\,x}\phi(t-s)\,dt\,ds && (2.12)\\
&\geq 0.
\end{aligned}$$

We can use the dominated convergence theorem to prove equation (2.10), a change of variables to show equation (2.11) and finally a Riemann sum approximation to the integral and the positive definiteness of ϕ to show that the quantity in (2.12) is nonnegative. It remains to show the relation (2.9). Let us define

$$f_\sigma(x) = f(x)\exp[-\frac{\sigma^2x^2}{2}]$$

and calculate for $t \in \mathbf{R}$, using Fubini's theorem

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{i\,t\,x} f_\sigma(x)\,dx &= \int_{-\infty}^{\infty} e^{i\,t\,x} f(x)\exp[\,-\frac{\sigma^2 x^2}{2}]\,dx \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i\,t\,x}\phi(s)e^{-i\,s\,x}\exp[\,-\frac{\sigma^2 x^2}{2}]\,ds\,dx \\
&= \int_{-\infty}^{\infty}\phi(s)\frac{1}{\sqrt{2\pi}\sigma}\exp[\,-\frac{(t-s)^2}{2\sigma^2}]\,ds. \qquad (2.13)
\end{aligned}$$

If we take $t = 0$ in equation (2.13), we get

$$\int_{-\infty}^{\infty} f_\sigma(x)\,dx = \int_{-\infty}^{\infty}\phi(s)\frac{1}{\sqrt{2\pi}\sigma}\exp[\,-\frac{s^2}{2\sigma^2}]\,ds \le 1. \qquad (2.14)$$

Now we let $\sigma \to 0$. Since $f_\sigma \ge 0$ and tends to f as $\sigma \to 0$, from Fatous's lemma and equation (2.14), it follows that f is integearable and in fact $\int_{-\infty}^{\infty} f(x)dx \le 1$. Now we let $\sigma \to 0$ in equation (2.13). Since $f_\sigma(x)e^{itx}$ is dominated by the integrable function f, there is no problem with the left hand side. On the other hand the limit as $\sigma \to 0$ is easily calculated on the right hand side of equation (2.13)

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{itx} f(x)dx &= \lim_{\sigma\to 0}\int_{-\infty}^{\infty}\phi(s)\frac{1}{\sqrt{2\pi}\sigma}\exp[\,-\frac{(s-t)^2}{2\sigma^2}]\,ds \\
&= \lim_{\sigma\to 0}\int_{-\infty}^{\infty}\phi(t+\sigma s)\frac{1}{\sqrt{2\pi}}\exp[\,-\frac{s^2}{2}]\,ds \\
&= \phi(t)
\end{aligned}$$

proving equation (2.9).

Step 3. If $\phi(t)$ is a positive definite function which is continuous, so is $\phi(t)\exp[\,i\,t\,y\,]$ for every y and for $\sigma > 0$, as well as the convex combination

$$\begin{aligned}
\phi_\sigma(t) &= \textstyle\int_{-\infty}^{\infty}\phi(t)\exp[\,i\,t\,y\,]\frac{1}{\sqrt{2\pi}\sigma}\exp[\,-\frac{y^2}{2\sigma^2}]\,dy \\
&= \phi(t)\exp[\,-\frac{\sigma^2 t^2}{2}].
\end{aligned}$$

The previous step is applicable to $\phi_\sigma(t)$ which is clearly integrable on $\mathbf{R}$ and by letting $\sigma \to 0$ we conclude by Theorem 2.3. that ϕ is a characteristic function as well. $\square$

Remark 2.2. There is a Fourier Series analog involving distributions on a finite interval, say $S = [0, 2\pi)$. The right end point is omitted on purpose, because the distribution should be thought of as being on $[0, 2\pi]$ with 0 and 2π identified. If α is a distribution on S the characteristic function is defined as

$$\phi(n) = \int e^{i\,n\,x} d\alpha$$

for integral values $n \in Z$. There is a uniqueness theorem, and a Bochner type theorem involving an analogous definition of positive definiteness. The proof is nearly the same.

Exercise 2.9. If $\alpha_n \Rightarrow \alpha$ it is not always true that $\int x\, d\alpha_n \to \int x\, d\alpha$ because while x is a continuous function it is not bounded. Construct a simple counterexample. On the positive side, let $f(x)$ be a continuous function that is not necessarily bounded. Assume that there exists a positive continuous function $g(x)$ satisfying

$$\lim_{|x|\to\infty} \frac{|f(x)|}{g(x)} = 0$$

and

$$\sup_n \int g(x)\, d\alpha_n \le C < \infty.$$

Then show that

$$\lim_{n\to\infty} \int f(x)\, d\alpha_n = \int f(x)\, d\alpha$$

In particular if $\int |x|^k\, d\alpha_n$ remains bounded, then $\int x^j\, d\alpha_n \to \int x^j\, d\alpha$ for $1 \le j \le k-1$.

Exercise 2.10. On the other hand if $\alpha_n \Rightarrow \alpha$ and $g : \mathbf{R} \to \mathbf{R}$ is a continuos function then the distribution β_n of g under α_n defined as

$$\beta_n[A] = \alpha_n[x : g(x) \in A]$$

converges weakly to β the corresponding distribution of g under α.

Exercise 2.11. If $g_n(x)$ is a sequence of continuous functions such that

$$\sup_{n,x} |g_n(x)| \le C < \infty \quad \textit{and} \quad \lim_{n\to\infty} g_n(x) = g(x)$$

uniformly on every bounded interval, then whenever $\alpha_n \Rightarrow \alpha$ it follows that

$$\lim_{n\to\infty} g_n(x) d\alpha_n = \int g(x) d\alpha.$$

Can you onstruct an example to show that even if g_n, g are continuous just the pointwise convergence $\lim_{n\to\infty} g_n(x) = g(x)$ is not enough.

Exercise 2.12. If a sequence $\{f_n(\omega)\}$ of random variables on a measure space are such that $f_n \to f$ in measure, then show that the sequence of distributions α_n of f_n on $\mathbf{R}$ converges weakly to the distribution α of f. Give an example to show that the converse is not true in general. However, if f is equal to a constant c with probability 1, or equivalently α is degenerate at some point c, then $\alpha_n \Rightarrow \alpha = \delta_c$ implies the convrgence in probability of f_n to the constant function c.