13 Characteristic functions

Definition and basic properties

A main tool in our proof of the central limit theorem will be that of characteristic functions. The basic idea will be to show that

\[
\lim_{n \to \infty} \mathbb{E} \left[ g \left( \frac{S_n - n\mu}{\sqrt{n}\sigma} \right) \right] = \mathbb{E} g(N(0,1))
\]

for a sufficiently large family of functions \( g \). It turns out that the family of functions of the form

\[
g_t(x) = e^{itx}, \quad (t \in \mathbb{R})
\]

is ideally suited for this purpose. (Here and throughout, \( i = \sqrt{-1} \)).

Definition

The characteristic function of a r.v. \( X \), denoted \( \varphi_X \), is defined by

\[
\varphi_X(t) = \mathbb{E}[e^{itX}] = \mathbb{E} \cos(tX) + i \mathbb{E} \sin(tX), \quad (t \in \mathbb{R}).
\]

Note that we are taking the expectation of a complex-valued random variable (which is a kind of two-dimensional random vector, really). However, the main properties of the expectation operator (linearity, the triangle inequality etc.) that hold for real-valued random variables also hold for complex-valued ones, so this will not pose too much of a problem.
Here are some simple properties of characteristic functions. For simplicity we denote \( \varphi = \varphi_X \) where there is no risk of confusion.

1. \( \varphi(0) = \expec e^{i0X} = 1 \).
2. \( \varphi(-t) = \expec e^{-itX} = \overline{\varphi(t)} \) (where \( \overline{z} \) denotes the complex conjugate of a complex number \( z \)).
3. \( |\varphi(t)| \leq 1 \) by the triangle inequality.
4. \( |\varphi(t) - \varphi(s)| \leq \expec |e^{itX} - e^{isX}| = \expec |e^{isX}(e^{i(t-s)X} - 1)| \). Note also that \( \expec |e^{iuX} - 1| \to 0 \) as \( u \downarrow 0 \) by the bounded convergence theorem. It follows that \( \varphi \) is a uniformly continuous function on \( \mathbb{R} \).
5. \( \varphi_{aX}(t) = \expec e^{iatX} = \varphi_X(at) \), \( a \in \mathbb{R} \).
6. \( \varphi_{X+b}(t) = \expec e^{it(X+b)} = e^{ibt} \varphi_X(t) \), \( b \in \mathbb{R} \).
7. Important: If \( X,Y \) are independent then \[ \varphi_{X+Y}(t) = \expec e^{it(X+Y)} = \expec \varphi_X(t) \varphi_Y(t) \] Note that this is the main reason why characteristic functions are such a useful tool for studying the distribution of a sum of independent random variables.

A note on terminology: If \( f \) has a density function \( f(x) \), then the characteristic function can be computed as \[ \varphi_X(t) = \int_{-\infty}^{\infty} f_X(x) e^{itx} \, dx. \]

In all other branches of mathematics, this would be called the Fourier transform of \( f \). Well, more or less -- it is really the inverse Fourier transform; but it will be the Fourier transform if we replace \( f(t) \) by \( f(-t) \), so that is almost the same thing. So the concept of a characteristic function generalizes the Fourier transform. If \( \mu(t) \) is the distribution measure of \( X \), some authors write \[ \varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} d\mu(x) \]

(which is an example of a Lebesgue-Stieltjes integral) and call this the Fourier-Stieltjes transform (or just the Fourier transform) of the measure \( \mu(t) \).

Examples

No study of characteristic functions is complete without `dirtying your hands" a little to compute the characteristic function for some important cases. The following exercise is highly recommended

Exercise

Compute the characteristic functions for the following distributions.

1. **Coin flips:** Compute \( \varphi_{X_1} \) when \( \prob(X=-1) = \prob(X=1) = 1/2 \) (this comes out slightly more symmetrical than the usual Bernoulli r.v. for which \( \prob(X=0) = \prob(X=1) = 1/2 \)).
2. **Symmetric random walk:** Compute \(\varphi_{S_n}\) where \(S_n = \sum_{k=1}^n X_k\) is the sum of \(n\) i.i.d. copies of the coin flip distribution above.

3. **Poisson distribution:** \(X \sim \text{Poisson}(\lambda)\).

4. **Uniform distribution:** \(X \sim U[a,b]\), and in particular \(X \sim [-1,1]\) which is especially symmetric and useful in applications.

5. **Exponential distribution:** \(X \sim \text{Exp}(\lambda)\).

6. **Symmetrized exponential:** A r.v. \(Z\) with density function \(f_Z(x) = e^{-|x|}\). Note that this is the distribution of the exponential distribution after being "symmetrized" in either of two ways: (i) We showed that if \(X,Y \sim \text{Exp}(1)\) are independent then \(X-Y\) has density \(e^{-|x|}\); (ii) alternatively, it is the distribution of an "exponential variable with random sign", namely \((\varepsilon \cdot X)\) where \(X \sim \text{Exp}(1)\) and \(\varepsilon\) is a random sign (same as the coin flip distribution mentioned above) that is independent of \(X\).

The normal distribution has the nice property that its characteristic function is equal, up to a constant, to its density function.

**Lemma**

If \(Z \sim \text{N}(0,1)\) then

\[
\varphi_Z(t) = e^{-t^2/2}.
\]

**Proof**

\[
\begin{eqnarray*}
\varphi_Z(t) &=& \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-x^2/2} \, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-it)^2/2} \, dx \\
&=& e^{-t^2/2} \left( \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{(x-it)^2/2} \, dx \right).
\end{eqnarray*}
\]

As Durrett suggests in his "physics proof" (p. 92 in [Dur2010], 91 in [Dur2004]), the expression in parentheses is \(1\), since it is the integral of a normal density with mean \(it\) and variance \(1\). This is a nonsensical argument, of course (\(it\) being an imaginary number), but the claim is true, easy, and is proved in any complex analysis course using contour integration.

Alternatively, let \(S_n = \sum_{k=1}^n X_k\) where \(X_1, X_2, \ldots\) are i.i.d. coin flips with \(\prob(X_k) = 1/2\). We know from the de Moivre-Laplace theorem (Theorem~\ref{thm-demoivre-laplace}) that

\[
\varphi_{S_n} = \varphi_{\sqrt{n}} \implies \text{N}(0,1),
\]
so that
\[
\varphi_{\{S_n/\sqrt{n}\}}(t) = \exp\left( i t S_n/\sqrt{n} \right) \xrightarrow{n\to\infty} \varphi_Z(t), \quad (t \in \mathbb{R}),
\]
since the function \(x \to e^{i t x}\) is bounded and continuous. On the other hand, from the exercise above it is easy to compute that \(\varphi_{\{S_n\}}(t) = \cos^n(t)\), which implies that
\[
\varphi_{\{S_n/\sqrt{n}\}}(t) = \cos^n\left(\frac{t}{\sqrt{n}}\right) = \left(1 - \frac{t^2}{2n} + O\left(\frac{t^4}{n^2}\right)\right)^n \xrightarrow{n\to\infty} e^{-t^2/2}.
\]
As a consequence, let \(X \sim N(0,\sigma_1^2)\) and \(Y \sim N(0,\sigma_2^2)\) be independent, and let \(Z = X + Y\). Then
\[
\varphi_X(t) = e^{-\sigma_1^2 t^2/2}, \quad \varphi_Y(t) = e^{-\sigma_2^2 t^2/2},
\]
so \(\varphi_Z(t) = e^{-\left(\sigma_1^2 + \sigma_2^2\right)/2}\). This is the same as \(\varphi_W(t)\), where \(W \sim N(0,\sigma_1^2 + \sigma_2^2)\). It would be nice if we could deduce from this that \(Z \sim N(0,\sigma_1^2 + \sigma_2^2)\) (we already proved this fact in a homework exercise, but it's always nice to have several proofs of a result, especially an important one like this one). This naturally leads us to an important question about characteristic functions, which we consider in the next section.

The inversion formula

A fundamental question about characteristic functions is whether they contain all the information about a distribution, or in other words whether knowing the characteristic function determines the distribution uniquely. This question is answered (affirmatively) by the following theorem, which is a close cousin of the standard inversion formula from analysis for the Fourier transform.

**Theorem: The Inversion Formula**

If \(X\) is a r.v. with distribution \(\mu_X\), then for any \((a,b)\) we have
\[
\begin{align*}
\lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-iat} - e^{-ibt}}{it} \varphi_X(t) \, dt &= \mu_X((a,b)) + \frac{1}{2} \mu_X(\{a\}) + \frac{1}{2} \mu_X(\{b\})
\end{align*}
\]

**Corollary: Uniqueness of Character Function**
If \((X,Y)\) are r.v.s such that \(\varphi_X(t) \equiv \varphi_Y(t)\) for all \(t \in \mathbb{R}\) then \(X \eqdist Y\).

**Example**

Explain why Corollary \ref{cor:charfun-uniqueness} follows from the inversion formula.

**Proof: Inversion Theorem**

Throughout the proof, denote \(\varphi(t) = \varphi_X(t)\) and \(\mu = \mu_X\). For convenience, we use the notation of Lebesgue-Stieltjes integration with respect to the measure \(\mu\), remembering that this really means taking the expectation of some function of the r.v. \(X\). Denote

\[
I_T = \int_{-T}^T \frac{e^{-i a t} - e^{-i b t}}{i t} \varphi(t) \, dt
= \int_{-T}^T \int_{-\infty}^{\infty} \frac{e^{-i a t} - e^{-i b t}}{i t} e^{i t x} \, d\mu(x) \, dt.
\]

Since \(\frac{e^{-i a t} - e^{-i b t}}{i t} = \int_a^b e^{-i t y} \, dy\) is a bounded function of \(t\) (it is bounded in absolute value by \(|b-a|\)), it follows by **Fubini's theorem** that we can change the order of integration, so

\[
\begin{align*}
I_T &= \int_{-\infty}^{\infty} \left[ \int_{-T}^T \frac{\sin(t(x-a))}{t} \, dt - \int_{-T}^T \frac{\sin(t(x-b))}{t} \, dt \right] \, d\mu(x) \\
&= \int_{-\infty}^{\infty} \left( R(x-a,T) - R(x-b,T) \right) \, d\mu(x),
\end{align*}
\]

where we denote \(R(\theta,T) = \int_{-T}^{T} \frac{\sin(x)}{x} \, dx\). Note that in the notation of expectations this can be written as \(\expec \left( R(X-a,T) - R(X-b,T) \right)\).

This can be simplified somewhat; in fact, observe also that

\[
\int (R(\theta, T) = \int_{-T}^{T} \frac{\sin(t(x-a))}{t} \, dt - \int_{-T}^{T} \frac{\sin(t(x-b))}{t} \, dt).
\]

Note that the notation of expectations this can be written as \(\expec \left( R(X-a,T) - R(X-b,T) \right)\).

This can be simplified somewhat; in fact, observe also that

\[
\int (\theta, T = 2 \text{ textrm} \{\text{sgn}\} (\theta) \text{ textt{int}0^{T}} \{\theta} \text{ text{frac}\{\sin x\} \{x\}} \, dx = 2 \text{ textt{exr{trm}m} \{\text{sgn}\} (\theta) S(|\theta| T), \}
\]

where we denote \(|S(x) = \text{ int}_0^{x} \text{ frac}\{\sin(u)\} \{u\} \, du\) and \(|\text{ textt{exr{trm}m} \{\text{sgn}\} (\theta)\}) is \(|1\) if \(|\theta| > 0\), \(-1\) if \(|\theta| < 0\) and \(|0\) if \(|\theta| = 0\). By a standard convergence test for integrals, the improper integral \(|\text{ int}_0^{\infty} \text{ frac}\{\sin u\} \{u\} \, du = \lim_{x \to \infty} S(x)\) converges; denote its value by \(|C/4\). Thus, we have shown that

\[
\int R(x-a,T) - R(x-b,T) \xrightarrow[T \to \infty]{} \begin{cases}
C & a<x<b, \\
C/2 & x=a \text{ or } x=b,
\end{cases}
\]

\[\]
0 & x<\text{or} & x>b.
\end{cases}

Furthermore, the function \( (R(x-a,T)-R(x-b,T)) \) is bounded in absolute value by \( 2\sup_{x\geq 0} S(x) \). It follows that we can apply the bounded convergence theorem in \eqref{eq:inversionproof} to get that

\[
\int I_T \rightarrow_{T\to\infty} C \mathbb{E}(1_{a<X<b}) + (C/2) \mathbb{E}(\mathbb{1}_{\{X=a\}} + \mathbb{1}_{\{X=b\}}) = C \mu((a,b)) + (C/2)\mu\{\{a,b\}\}.
\]

This is just what we claimed, minus the fact that \( (C=2\pi) \). This fact is a well-known integral evaluation from complex analysis. We can also deduce it in a self-contained manner, by applying what we proved to a specific measure \( (\mu) \) and specific values of \( (a) \) and \( (b) \) for which we can evaluate the limit in \eqref{eq:inversionproof} directly. This is not entirely easy to do, but one possibility, involving an additional limiting argument, is outlined in the next exercise; see also Exercise 1.7.5 on p. 35 in [Dur2010], (Exercise 6.6, p. 470 in Appendix A.6 of [Dur2004]) for a different approach to finding the value of \( (C) \).

**Exercise: Recommended for aspiring analysts**

For each \( (\sigma>0) \), let \( (X_\sigma) \) be a r.v. with distribution \( (N(0,\sigma^2)) \) and therefore with density

\( f_X(x)=(\sqrt{2\pi}\sigma)^{-1} e^{-x^2/2\sigma^2} \) and characteristic function \( \varphi_X(t) = e^{-\sigma^2 t^2/2} \). For fixed \( (\sigma) \), apply Theorem \ref{thm-inversion} in its weak form given by \eqref{eq:almostinversion} (that is, without the knowledge of the value of \( (C) \)), with parameters \( (X=X_\sigma) \), \( (a=-1) \) and \( (b=1) \), to deduce the identity

\[
\frac{C}{\sqrt{2\pi}\sigma} \int_{-1}^1 e^{-x^2/2\sigma^2} \, dx = \int_{-\infty}^{\infty} \frac{2\sin t}{t} e^{-\sigma^2 t^2/2} \, dt.
\]

Now multiply both sides by \( (\sigma) \) and take the limit as \( (\sigma\to\infty) \). For the left-hand side this should give in the limit (why?) the value \( (2C)\sqrt{2\pi} \). For the right-hand side this should give \( (2\sqrt{2\pi}) \). Justify these claims and compare the two numbers to deduce that \( (C=2\pi) \).

The following theorem shows that the inversion formula can be written as a simpler connection between the characteristic function and the density function of a random variable, in the case when the characteristic function is integrable.

**Theorem**

If \( \int_{-\infty}^{\infty} \varphi_X(t) \, dt < \infty \), then \( (X) \) has a bounded and continuous density function \( (f_X) \), and the density and characteristic function are related by
\begin{eqnarray*}
\varphi_X(t) &=& \int_{-\infty}^{\infty} f_X(x) \, e^{itx} \, dx, \\
f_X(x) &=& \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi_X(t) \, e^{-itx} \, dt.
\end{eqnarray*}

In the lingo of Fourier analysis, this is known as the \textit{inversion formula for Fourier transforms.}

**Proof**

This is a straightforward corollary of Theorem \ref{thm-inversion}. See p.~95 in either [Dur2010] or [Dur2004].

---

The continuity theorem

**Theorem: The Continuity Theorem**

Let \((X_n)_{n=1}^{\infty}\) be r.v.'s. Then:

1. If \((X_n) \implies X) for some r.v. \((X), then \(\varphi_{X_n}(t) \to \varphi_X(t)\) for all \(t \in \mathbb{R}\).
2. If the limit \(\varphi(t) = \lim_{n \to \infty} \varphi_{X_n}(t)\) exists for all \(t \in \mathbb{R}\), and \(\varphi\) is continuous at 0, then \(\varphi\equiv \varphi_X\) for some r.v. \((X), and \(X_n) \implies X)\).

**Proof**

Part (i) follows immediately from the fact that convergence in distribution implies that \(\expec g(X_n) \to \expec g(X)\) for any bounded continuous function. It remains to prove the less trivial claim in part (ii). Assume that \(\varphi_{X_n}(t) \to \varphi(t)\) for all \(t \in \mathbb{R}\) and that \(\varphi\) is continuous at 0. First, we show that the sequence \((X_n)_{n=1}^{\infty}\) is tight. Fixing an \(M>0\), we can bound the probability \(\prob(|X_n|>M)\), as follows:

\begin{eqnarray*}
\prob(|X_n|>M) &=& \expec \left( \ind_{\{|X_n|>M\}} \right) \\
&\le& \expec \left[ 2 \left( 1-\frac{M}{2|X_n|} \right) \ind_{\{|X_n|>M\}} \right] \\
&\le& \expec \left[ 2 \left( 1-\frac{\sin(2X_n/M)}{2X_n/M} \right) \ind_{\{|X_n|>M\}} \right].
\end{eqnarray*}

But this last expression can be related to the behavior of the characteristic function near \(0\). Denote \(\delta=2/M\). Reverting again to the Lebesgue-Stieltjes integral notation, we have:

\[
\left[ \expec \left( 1-\frac{\sin(2X_n/M)}{2X_n/M} \right) \ind_{\{|X_n|>M\}} \right]
\]
Now use Fubini's theorem to get that this bound can be written as

\[
\frac{1}{\delta} \int_{-\delta}^{\delta} \int_{-\infty}^{\infty} (1-e^{itx}) \, d\mu_{X_n}(x) \, dt
\]

(the convergence follows from the bounded convergence theorem).

So we have shown that

\[
\limsup_{n\to\infty} \prob(|X_n|>M) \le \frac{1}{\delta} \int_{-\delta}^{\delta} (1-\varphi(t)) \, dt.
\]

But, because of the assumption that $\varphi(t) \to \varphi(0)=1$ as $t \to 0$, it follows that if $\delta$ is sufficiently small then $\lim_{\delta \to 0} \int_{-\delta}^{\delta} (1-\varphi(t)) \, dt \le \epsilon$, where $\epsilon > 0$ is arbitrary; so this establishes the tightness claim.

Finally, to finish the proof, let $(n_k)_{k=1}^\infty$ be a subsequence (guaranteed to exist by tightness) such that $(X_{n_k})_{k=1}^\infty$ implies $Y$) for some r.v. $Y$. Then $\varphi_{X_{n_k}}(t) \to \varphi_Y(t)$ as $k \to \infty$ for all $t \in \mathbb{R}$, so $\varphi_Y(t) \equiv \varphi_Y$. This determines the distribution of $Y$, which means that the limit in distribution is the same no matter what convergent in distribution subsequence of the sequence $(X_n)_{n=1}^\infty$ we take. But this implies that $(X_n)_{n=1}^\infty$ implies $Y$) (why? The reader is invited to verify this last claim; it is best to use the definition of convergence in distribution in terms of expectations of bounded continuous functions).

Moments

The final step in our lengthy preparation for the proof of the central limit theorem will be to tie the behavior of the characteristic function $\varphi_X(t)$ near $t=0$ to the moments of $X$. Note that, computing formally without regards to rigor, we can write

\[
\varphi_X(t) = \expec(e^{itX}) = \expec\left[\sum_{n=0}^{\infty} \frac{i^n t^n X^n}{n!}\right] = \sum_{n=0}^{\infty} \frac{i^n t^n \expec(X^n)}{n!}.
\]
So it appears that the moments of \(\langle X \rangle\) appear as (roughly) the coefficients in the Taylor expansion of \(\langle \varphi_X \rangle\) around \(t=0\). However, for CLT we don't want to assume anything beyond the existence of the second moment, so a (slightly) more delicate estimate is required.

**Lemma: Taylor Estimate**

\[
\left|e^{ix}-\sum_{m=0}^n \frac{(ix)^m}{m!}\right| \leq \min\left(\frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!}\right). 
\]

**Proof**

Start with the identity

\[
R_n(x) := e^{ix}-\sum_{m=0}^n \frac{(ix)^m}{m!} = \frac{i^{n+1}}{n!}\int_0^x (x-s)^n e^{is}\,ds,
\]

which follows from Lemma~\ref{lem-prevlemma} that we used in the proof of Stirling’s formula. Taking the absolute value and using the fact that \(|e^{is}|=1\) gives

\[
|R_n(x)| \leq \frac{1}{n!} \left|\int_0^x |x-s|^n\,ds\right| = \frac{|x|^{n+1}}{n!}.
\]

To get a bound that is better-behaved for large \(|x|\), note that

\[
\begin{align*}
R_n(x) &= R_{n-1}(x) - \frac{(ix)^n}{n!} = R_{n-1}(x) - \frac{i^n (x-s)^{n-1}}{(n-1)!}\int_0^x e^{is}-1\,ds.
\end{align*}
\]

So, since \(|e^{is}-1|\leq 2\), we get that

\[
|R_n(x)| \leq \frac{2}{(n-1)!} \left|\int_0^x |x-s|^{n-1}\,ds\right| = \frac{2|x|^n}{(n-1)!}.
\]

Combining \eqref{eq:firstbound} and \eqref{eq:secondbound} gives the claim.

Now let \(\langle X \rangle\) be a r.v. with \(\langle \expec|X|\rangle < \infty\). Letting \(\langle X=tX \rangle\) in Lemma~\ref{lem-taylor-estimate}, taking expectations and using the triangle inequality, we get that

\[
\begin{align*}
\left|\varphi_X(t)-\sum_{m=0}^n \frac{i^m \expec X^m}{m!}t^m\right| &\leq \expec\left[\min\left(\frac{|t|^{n+1} |X|^{n+1}}{(n+1)!}, \frac{2 |t|^n |X|^n}{n!}\right)\right].
\end{align*}
\]
Note that in this minimum of two terms, when \( \var(t) \) is very small the first term gives a better bound, but when taking expectations we need the second term to ensure that the expectation is finite if \( \var(X) \) is only assumed to have a finite \( (n) \)-th moment.

**Theorem: Second Moments of Character Functions**

If \( X \) is a r.v. with mean \( \mu = \expec X \) and \( \var(X) < \infty \) then

\[
\varphi_X(t) = 1 + i \mu t - \frac{\expec X^2}{2} t^2 + o(t^2) \quad \text{as } t \to 0.
\]

**Proof**

By \eqref{eq:taylor-expec} above, we have

\[
\frac{1}{t^2} \left| \varphi_X(t) - \left(1 + i \mu t - \frac{\expec X^2}{2} t^2 \right) \right| \le \expec \left[ \min \left( |t| \cdot |X|^3/6, X^2 \right) \right].
\]

As \( t \to 0 \), the right-hand side converges to 0 by the dominated convergence theorem.

**Contributors**

- Dan Romik (UC Davis)