Math/Stat 523, Spring 2020

Jon A. Wellner

Lecture 5

Friday April 10?)

Outline

- 0: Questions: Convergence of sums of independent rv's.
- 1: Classical convergence in distribution;
  Helly's selection theorem
- 2: Convergence of types
- 3: Characteristic functions: basic results
- 4: Uniqueness and inversion; the continuity theorem
- 5: Higher dimensions

0. Questions: convergence in distribution of independent rv's

Problem 1:
Suppose that $X_1, X_2, \ldots$ are i.i.d. F with $E(X_1) = 0$ and $E(X_1^2) = \sigma^2 < \infty$.
(a) Then

$$\frac{X_1 + \cdots + X_n}{\sigma\sqrt{n}} = \frac{\sqrt{n}\overline{X}_n}{\sigma} \to_d Z \sim N(0,1).$$

(b) By the convergence of types theorem, for any $\{a_n\}$ with $a_n > 0$, if $S_n/a_n \to_d W$, then W is normal or degenerate.
(c) How fast does $\to_d$ occur?

Problem 2. Suppose that $X_{n1}, X_{n2}, \ldots, X_{nn}$ are independent and identically distributed for each n, and let $S_n = \sum_{i=1}^n X_{ni}$.

(a) If $S_n \to_d S$, what are all possible limiting distributions for S?

Answer: S is infinitely divisible: i.e. for every n there exist $X_1^{(n)}, \ldots, X_n^{(n)}$ i.i.d. such that $S \stackrel{d}{=} \sum_{j=1}^n X_j^{(n)}$.

Problem 3. Suppose $X_{n1}, \ldots, X_{nn}$ are independent, and let $S_n \equiv \sum_{j=1}^n X_{nj}$. If $S_n \to_d S$ and $\sup_{1 \le k \le n} P(|X_{nk}| > \epsilon) \to 0$ for every $\epsilon \to 0$, what are the possible limiting distributions of S?

Answer: S is infinitely divisible!

Problem 4. Suppose that $X_1, \ldots, X_n$ are i.i.d. and let

$$S_n \equiv \frac{\sum_{j=1}^n X_j}{a_n} - b_n \quad \text{with} \quad a_n > 0.$$

What is the class of all limit laws $\mathcal{L}(Y)$ of $S_n \to_d Y$? (Note that $E(X_1) = 0$ and $E(X_1^2) < \infty$ are not assumed!)
Answer: $\mathcal{L}(Y) \in$ Stable laws.

Definition: Y has a stable law or distribution if for every integer $k > 0$ and $X_1, \ldots, X_k$ independent with $X_j \stackrel{d}{=} Y$ there are constants $a_k > 0$ and $b_k \in \mathbb{R}$ such that

$$X_1 + \cdots + X_k \stackrel{d}{=} a_k Y + b_k. \tag{1}$$

Example 1: $Y \sim N(0,1)$. Then $X_1 + \cdots + X_k \sim N(0,k) \stackrel{d}{=} \sqrt{k}Y$, so (1) holds with $a_k = \sqrt{k}$ and $b_k = 0$.

Example 2: $Y \sim$ Cauchy$(0,1)$ with density

$$f(x) = \pi^{-1}(1 + x^2)^{-1}.$$

Then $X_1 + \cdots + X_k \sim k$ Cauchy$(0,1) \stackrel{d}{=} kY$, so (1) holds with $a_k = k$ and $b_k = 0$.

1. Classical convergence in distribution; Helly's selection theorem

Definition.

- (a) H is a sub-distribution function if H is right-continuous, non-decreasing with $0 \le H(x) \le 1$ for all x. Thus for an extended-valued random variable X, $P(X = -\infty) = H(-\infty)$, $P(X = +\infty) = 1 - H(\infty)$ and $H(x) = P(-\infty \le X \le x)$ for $-\infty < x < \infty$.

- (b) If $F_n(x) \to H(x)$ for each $x \in C_H$, then we say that X_n (or F_n) converges in sub-distribution function to X (or H) and we write $X_n \to_{sd} X$ (or $F_n \to_{sd} H$). (This allows escape of mass to $\pm\infty$.) [This is also called *vague convergence* by some authors.]

- (c) F_n or F will denote bona-fide distribution functions of proper (finite-valued) random variables.

Example 1. If $X_n \stackrel{d}{=} \sigma_n Z \sim N(0, \sigma_n^2)$ with $\sigma_n^2 \to \infty$ and $Z \sim N(0,1)$, then $P(X_n \le x) = P(Z \le x/\sigma_n) = \Phi(x/\sigma_n) \to \Phi(0) = 1/2 \equiv H(x)$ for $x \in \mathbb{R}$, so $X_n \to_{sd} X$ where $P(X = \pm\infty) = 1/2$.

Example 2: If $X_n \sim \text{Uniform}[0, a_n] \stackrel{d}{=} a_n U$ with $a_n \to \infty$ and $U \sim \text{Uniform}[0,1]$, then $F_n(x) = P(X_n \le x) = P(U \le x/a_n) = (x/a_n)1_{[x \le a_n]} + 1_{(a_n, \infty)}(x) \to 0$ for all $x \in \mathbb{R}$. In this case $X_n \to_{sd} X$ where $P(X = \infty) = 1$.

Example 3: If $X \sim \text{Bernoulli}(p)$, then $1/X$ has sub-distribution function H given by $H(x) = 0$, $-\infty < x < 1$, $H(x) = p$, $1 \le x < \infty$, and $H(\infty) = p$, so that $P(1/X = \infty) = P(X = 0) = 1 - H(+\infty) = 1 - p$.

Definition:

- (a) A family of distributions $\mathcal{P}$ on $\mathbb{R}$ (or a family of distribution functions F on $\mathbb{R}$ is called tight if for each $\epsilon > 0$ there is a compact set K_ϵ (a closed and bounded set) with $P(K_\epsilon) = P(X \in K_\epsilon) \geq 1 - \epsilon$ for all $P \in \mathcal{P}$. (Equivalently, $K_\epsilon = [a, b]$ such that $P(X \in [a, b]) = F(b) - F(a-) > 1 - \epsilon$ for all $F \in \mathcal{F}$.

- (b) A sequence of random variables $\{X_n\}$ is called asymptotically tight if

$$\lim_{M\to\infty} \overline{\lim}_n P(|X_n| > M) = 0.$$

The following two theorems are familiar to us from Math/Stat 521 and 522:

Theorem (Helly - Bray) If $F_n \to F$ and g is bounded and continuous a.s. F, then

$$Eg(X_n) = \int g dF_n \to \int g dF = Eg(X).$$

Theorem (Mann-Wald, Continuous Mapping) Suppose that $X_n \to_d X$ and that g is continuous a.s. $F = P^X$. Then $Y_n \equiv g(X_n) \to_d g(X) \equiv Y$.

Proof: Let $\psi \in C_b(\mathbb{R})$. Then $\psi \circ g$ is bounded and continuous a.s. F, so

$$\begin{aligned} E\psi(Y_n) = E\psi(g(X_n)) &\to E\psi(g(X)) \quad \text{by Helly - Bray} \\ &= E\psi(Y). \end{aligned}$$

Hence $Y_n = g(X_n) \to_d g(X) = Y$ by the Helly-Bray converse. $\square$

A theorem that clarifies the role of tightness in conjunction with $\to_{sd}$ is Helly's selection theorem.

Theorem (Helly's selection theorem) Let $F_1, F_2, \ldots$ be any sequence of distribution functions. Then there necessarily exists a subsequence $\{F_{n'}\}$ and a sub-distribution function H for which $F_{n'} \to_{sd} H$. If the subsequence of d.f.'s is tight, then the limit H is necessarily a bona-fide (proper) distribution function.

Corollary Let $F_1, F_2, \ldots$ be any sequence of distribution functions. Let H be a fixed sub - distribution function. Suppose every $sd-$convergent subsequence $\{F_{n'}\}$ satisfies $F_{n'} \to_{sd}$ to this same H. Then the whole sequence satisfies $F_n \to_{sd} H$.

Proof: Let $r_1, r_2, \ldots$ be a sequence dense in $\mathbb{R}$ (e.g. an enumeration of the rationals). Using the Bolzano-Weierstrass theorem, choose a subsequence n_{1j} such that

$$F_{n_{1,j}}(r_1) \to \text{ some } a_1;$$

a further subsequence also satisfies $F_{n_{2,j}}(r_2) \to$ some a_2. Continue in this fashion. The diagonal subsequence $n_{j,j}$ has

$$F_{n_{j,j}}(r_i) \to a_i \quad \text{for all} \quad i \geq 1.$$

Define H_0 on the r_i's via $H_0(r_i) = a_i$, and then define H on $\mathbb{R}$ by

$$H(x) \equiv \inf\{H_0(r_i) : \; r_i > x\}.$$

This H is clearly non-decreasing and takes values in $[0,1]$. We now verify that H is right-continuous and that $F_{n_{j,j}} \to_{sd} H$. Let $n_{j,j} \equiv n'$. The monotonicity of H_0 trivially gives $\inf_{y \searrow x} H(y) \geq H(x)$. Furthermore

$$
\begin{aligned}
H(x) &> H_0(r_{k_\epsilon}) - \epsilon \quad \text{for some } x < r_{k_\epsilon} \text{ close to } x \\
&\geq H(y) - \epsilon \quad \text{for any } x < y < r_{k_\epsilon},
\end{aligned}
$$

and this yields $\inf_{y \searrow x} H(y) \leq H(x)$. Hence $\inf_{y \searrow x} H(y) = H(x)$, and H is right-continuous.

We next show that $F_{n'}(x) = F_{n_{jj}}(x) \to H(x)$ for any $x \in C_H$. First note that

$$
F_{n'}(r_k) \leq F_{n'}(x) \leq F_{n'}(r_l) \quad \text{for all } r_k < x < r_l.
$$

Passing to the limit gives

$$
H_0(r_k) \leq \underline{\lim} F_{n'}(x) \leq \overline{F}_{n'}(x) \leq H_0(r_l)
$$

for all $r_k < x < r_l$. Now let $r_k \nearrow x$ and $r_l \searrow x$ to get

$$
\lim_{n'} F_{n'}(x) = H(x) \quad \text{for all } \ x \in C_H;
$$

i.e. $F_{n'} \to_{sd} H$. $\square$

Proof of the Corollary:

Fact 1: Any bounded sequence of real numbers $\{a_n\}$ contains a convergent subsequence.

Fact 2: A sequence of real numbers $\{a_n\}$ converges if and only if all subsequential limits are the same. [Or, if every subsequence $\{a_{n'}\}$ contains a further subsequence $\{a_{n''}\}$ that converges to one fixed number a_0, then $a_n \to a_0$.]

The proof of Helly's theorem shows that every subsequence $\{F_{n'}\}$ contains a further subsequence $\{F_{n''}\}$ for which $F_{n''}(x) \to$ the same $H(x)$ for each $x \in C_H$. Thus the whole sequence has $F_n(x) \to H(x)$ for $x \in C_H$. So $F_n \to_{sd} H$. $\square$

2. Convergence of types

Definition: (Type) When $Y \stackrel{d}{=} (X - b)/a$ for some $a \neq 0$, we say that X and Y are of the same *type*.

Theorem 9.1.5 (convergence of types) Suppose that $(X_n - b_n)/a_n \to_d X \sim F$, and $(X_n - \beta_n)/\alpha_n \to Y \sim G$ where $a_n > 0$, $\alpha_n > 0$.and both X and Y are non-degenerate. Then there exists $a > 0$ and $b \in \mathbb{R}$ such that

$$\frac{a_n}{\alpha_n} \to \text{some } a > 0, \quad \text{and} \quad \frac{\beta_n - b_n}{a_n} \to \text{some real } b$$

and $Y \stackrel{d}{=} (X - b)/a$ (or, equivalently, $G(x) = F(ax + b)$ for all $x \in \mathbb{R}$).

Proof: For a detailed proof of Theorem 9.1.5 see Billingsley (1986), *Probability and Measure*, Theorem 14.2, page 195. $\square$

3. Characteristic functions: basic results

Elementary Facts

Definition 9.3.1 (Characteristic Function) Let X be an arbitrary rv, and let F denote its d.f. The characteristic function of X, or chf, is defined for all $t \in \mathbb{R}$ by

$$\begin{aligned} \phi(t) &\equiv \phi_X(t) \equiv Ee^{itX} \equiv \int_{-\infty}^{\infty} e^{itx} d\mu_F(x) = \int_{-\infty}^{\infty} e^{itx} dF(x) \\ &\equiv \int_{-\infty}^{\infty} \cos(tx) dF(x) + i \int_{-\infty}^{\infty} \sin(tx) dF(x). \end{aligned}$$

With dF replaced by $hd\mu$, we call this the *Fourier transform* of the signed measure $hd\mu$. (We note that the chf $\phi_X(t)$ exists for $-\infty < t < \infty$ for *all* rv's X, since $|e^{itX(\omega)}| \leq 1$ for all t and all ω.)

Proposition 9.3.1. (Elementary properties) Let ϕ denote an arbitrary chf.

(a) $\phi(0) = 1$ and $|\phi(t)| \le 1$ for all $t \in \mathbb{R}$.

(b) $\phi_{aX+b}(t) = e^{itb}\phi_X(at)$ for all $t \in \mathbb{R}$.

(c) $\phi_{\sum_1^n X_i}(t) = \prod_{i=1}^n \phi_{X_i}(t)$ when $X_1, \ldots, X_n$ are independent.

(d) $\overline{\phi}_X(t) = \phi_{-X}(t) = E\cos(tX) - iE\sin(tX)$ for all $t \in \mathbb{R}$.

(e) ϕ is real-valued if and only if $X \stackrel{d}{=} -X$.

(f) $|\phi(\cdot)|^2$ is a chf [of the rv $X^s \equiv X - X'$, with X and X' i.i.d. with chf ϕ].

(g) ϕ is uniformly continuous on $\mathbb{R}$.

Proof: (a), (b), (c), and (d) are trivial. (e) If $X \stackrel{d}{=} -X$, then $\phi_X = \overline{\phi}_X$; so ϕ_X is real.

If ϕ_X is real, then $\phi_X = \overline{\phi}_X = \phi_{-X}$; so $X \stackrel{d}{=} -X$ by the uniqueness theorem below.

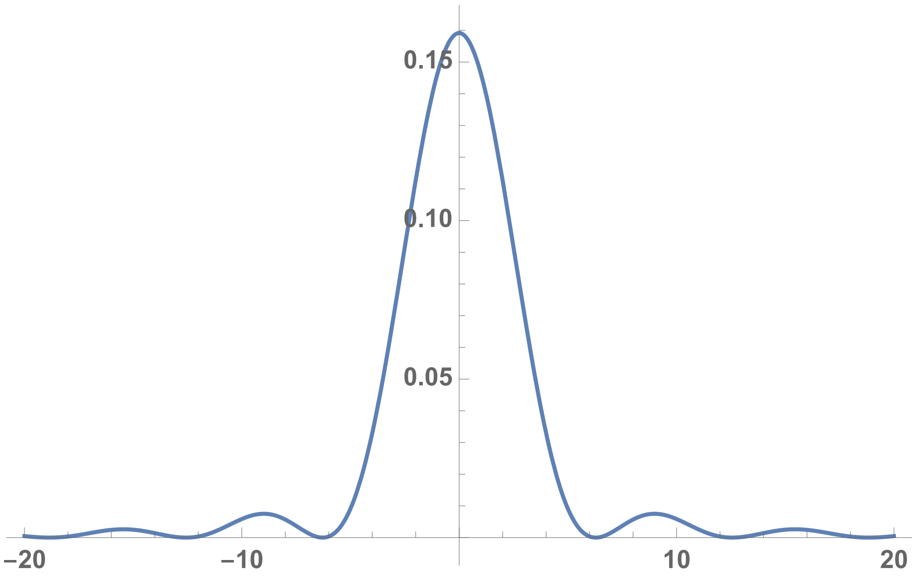
(f) If X and X' are independent with characteristic function ϕ, then $\phi_{X-X'} = \phi_X \phi_{-X} = \phi\overline{\phi} = |\phi|^2$.

(g) Note that for all t,

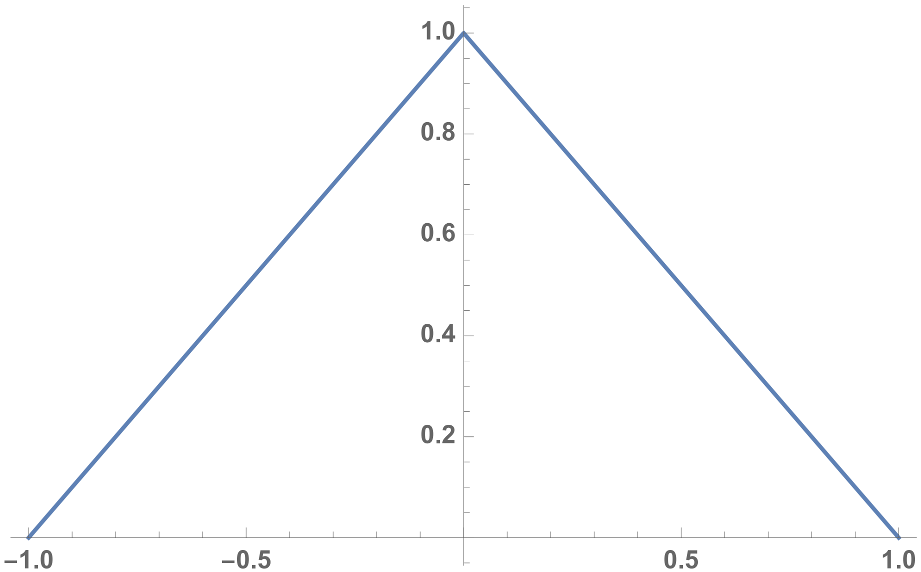
$$\begin{aligned} |\phi(t+h) - \phi(t)| &= \left| \int [\exp(i(t+h)x) - e^{itx}] dF(x) \right| \\ &\leq \int |e^{itx}||e^{ihx} - 1| dF(x) \leq \int |e^{ihx} - 1| dF(x) \to 0 \end{aligned}$$

as $h \to 0$ by the dominated convergence theorem with dominating function 2.

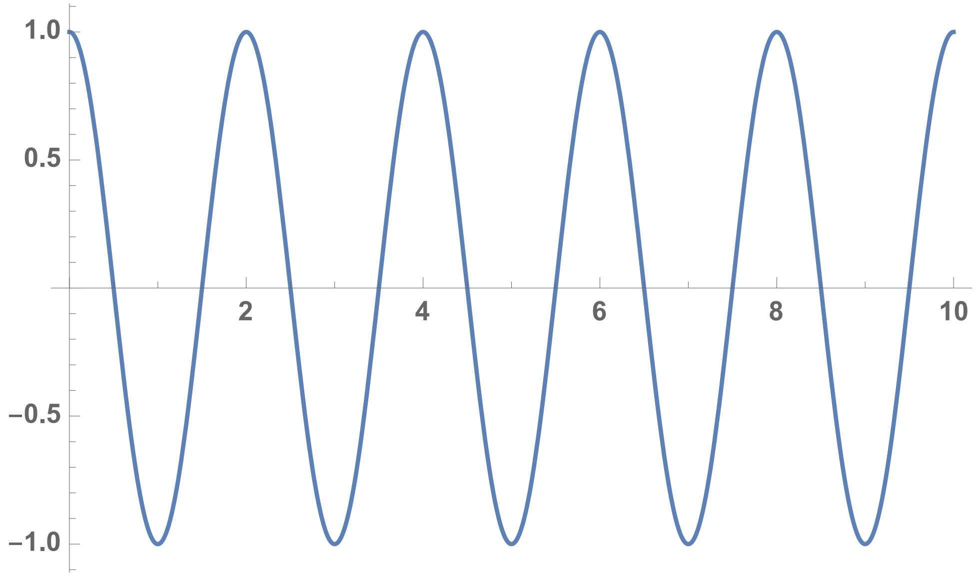
The converse of (c) is false. Let $X_1 \equiv X_2$ and X_3 by two i.i.d Cauchy$(0,1)$ rv's. We will see below that $\phi_{Cauchy}(t) = \exp(-|t|)$, so $\phi_{2X_1}(t) = \phi_{X_1+X_2}(t) = \phi_{X_1+X_3}(t)$ for all t. $\square$



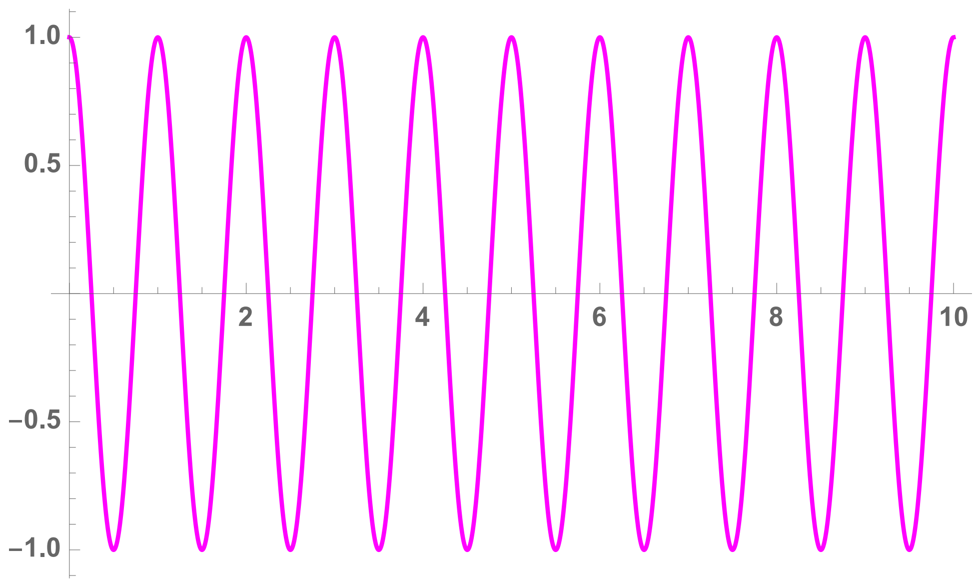
$f(x) = (1 - \cos(x))/(\pi x^2)$



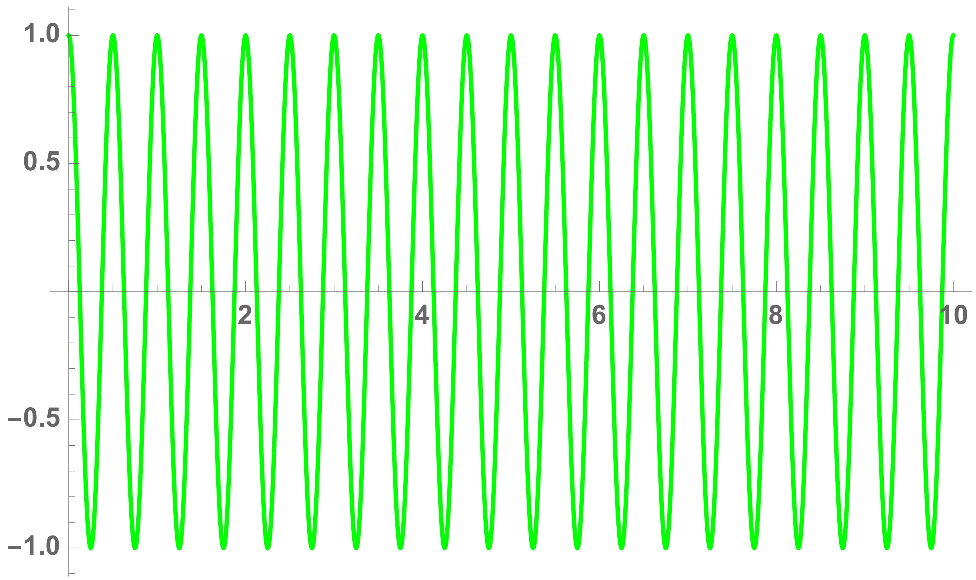
$\phi(t) = (1 - |t|)1_{[-1,1]}(t)$



$\cos(x)$



$\cos(2x)$



$\cos(4x)$

4. Characteristic functions: uniqueness and inversion

Before proving the main theorems we need the notion of complementary pairs of densities and characteristic functions.

Complementary pairs

Suppose U and W are rv's as follows:

$$\begin{aligned} U &\sim F_U \text{ with density } f_U, \\ W &\sim F_W \text{ with density } f_W, \\ f_U(t) &= c\phi_W(-t) \text{ for all real } t. \end{aligned}$$

Then U and W are called a complementary pair. Here are some examples:

$$\begin{array}{ccc} U \sim F_U & W \sim F_W & c \\ U = Z \sim N(0,1) & W = Z \sim N(0,1) & c = 1/\sqrt{2\pi} \\ U = T \sim triangular & W = D & c = 1 \\ U = D & W = T & c = 1/(2\pi) \end{array}$$

where

$$f_T(x) = (1 - |x|)1_{[-1,1]}(x) \quad \text{with} \quad \phi_T(t) = 2(1 - \cos(t))/t^2,$$
$$f_D(x) = (1 - \cos(x))/(\pi x^2), x \in \mathbb{R}, \quad \text{with} \quad \phi_D(t) = 1 - |t|.$$

Theorem 9.4.1: (Uniqueness theorem) Every distribution function on $\mathbb{R}$ has a unique characteristic function.

Theorem 9.4.2: (Inversion theorem) If X has d.f. F_X and characteristic function ϕ_X, we can always write

$$F_X(r_2) - F_X(r_1) = \lim_{a \searrow 0} \int_{r_1}^{r_2} f_a(t)dt \quad \text{for all} \quad r_1 < r_2 \in C_{F_X}.$$

where f_a is the density of $X + aU \equiv X_a \to_d X$ as $a \searrow 0$, and where

$$f_a(y) = \int_{-\infty}^{\infty} e^{-iyv} \phi_X(v) c f_W(av) dv \quad \text{for all} \quad y \in \mathbb{R}.$$

Theorem 9.4.3: (Inversion formula for densities) If a random variable X has a chf ϕ_X that satisfies

$$\int_{-\infty}^{\infty} |\phi_X(t)| dt < \infty, \tag{2}$$

then X has a uniformly continuous density f_X given by

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_X(t) dt. \tag{3}$$

Remark: The uniqueness theorem can be rephrased as follows: the set of complex exponentials $\mathcal{G} = \{g_t(x) \equiv e^{itx} : \ t \in \mathbb{R}\}$ is a determining class. That is, knowing all values of $\phi_X(t) = Ee^{itX} = Eg_t(X)$ allows the d.f. F to be determined via the inversion formula.

Proof: By the convolution formula and $X_a = X + aU$,

$$
\begin{aligned}
f_a(y) &= \int_{-\infty}^{\infty} \frac{1}{a} f_U\left(\frac{y-x}{a}\right) dF_X(x) \\
&= \int_{-\infty}^{\infty} \frac{c}{a} \phi_W\left(\frac{x-y}{a}\right) dF_X(x) \quad \text{since } f_U(t) = c\phi_W(-t) \\
&= \frac{c}{a} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i(x-y)w/a} f_W(w) dw dF_X(x) \\
&= \frac{c}{a} \int_{-\infty}^{\infty} e^{-iyw/a} f_W(w) \int_{-\infty}^{\infty} e^{ixw/a} dF_X(x) dw \\
&= \frac{c}{a} \int_{-\infty}^{\infty} e^{-iyw/a} f_W(w) \int_{-\infty}^{\infty} e^{ixw/a} dF_X(x) dw \text{ by Fubini} \\
&= \frac{c}{a} \int_{-\infty}^{\infty} e^{-iyw/a} \phi_X(w/a) f_W(w) dw \\
&= c \int_{-\infty}^{\infty} e^{-iyv} \phi_X(v) f_W(av) dv. \qquad (4)
\end{aligned}
$$

Since $X_a \to X$, at continuity points $r_1 < r_2$ of F we have (with $X_a \sim F_a$)

$$F_X(r_2) - F_X(r_1) = \lim_{a \to 0} \{F_a(r_2) - F_a(r_1)\}$$
$$= \lim_{a \to 0} \int_{r_1}^{r_2} f_a(y)\,dy. \quad (5)$$

This completes the proofs of Theorems 9.4.1 and 9.4.2. Now suppose that the integrability hypothesis (2) on $|\phi_X|$ holds. Then by the dominated convergence theorem (with dominating function $\|f_W\|_\infty \cdot |\phi_X(t)|$) we conclude that

$$f_a(y) \to f(y) \equiv [c f_W(0)] \int_{-\infty}^{\infty} e^{-iyv} \phi_X(v)\,dv$$

since f_W is bounded and continuous. For the uniform continuity of f, note that

$$
\begin{aligned}
|f(y+h)-f(y)| &= cf_W(0)\left|\int_{-\infty}^{\infty}(e^{-i(y+h)v}-e^{-iyv})\phi_X(v)dv\right| \\
&\le cf_W(0)\int_{-\infty}^{\infty}|e^{-ivh}-1|\cdot|\phi_X(v)|dv \to 0 \\
&\quad \text{as } h\to 0
\end{aligned}
$$

by the DCT with dominating function $2\|f_W\|_\infty|\phi_X(v)|$. Uniform convergence of f_a to f on any bounded interval just involves a term $|f_W(0)-f_W(av)|$ under the integral. Applying this uniform convergence to (5) yields

$$
F_X(r_2)-F_X(r_1) = \int_{r_1}^{r_2} f(y)dy.
$$

The conclusion (3) holds since we may take $U=W=Z\sim N(0,1)$, so that $cf_W(0)=1/(2\pi)$ (which always holds for conjugate pairs). $\square$

Review of some complex analysis

A function f is called *analytic* on a *region* (a connected open subset of the complex plane) if it has a derivative at each point of the region; if it does, then it necessarily has derivatives of all orders at each point in the region. If z_0 is an isolated singularity of f and $f(z) = \sum_{n=0}^{\infty}(z-z_0)^n + \sum_{n=1}^{m} b_n(z-z_0)^{-n}$, then $k \equiv$ (the *residue* of f at z_0) $= b_1$. Thus if f has a pole of order m at z_0 (that is, $b_n = 0$ for $n > m$ in the expansion above), then $g(z) \equiv (z-z_0)^m f(z) = b_m + \cdots + b_1(z-z_0)^{m-1} + \sum_{n=0}^{\infty} a_n(z-z_0)^{m+n}$ has $b_1 = g^{(m-1)}(z_0)/(m-1)!$. Thus

$$\begin{aligned} b_1 = k \;&=\; (\text{residue of } f \text{ at } z_0) \\ &\qquad \{= \lim_{z\to z_0}(z-z_0)f(z) \quad \text{for a simple pole at } z_0\}. \end{aligned}$$

A *smooth arc* is described via $x = \phi(t)$ and $y = \psi(t)$ for $a \le t \le b$ and when ϕ' and ψ' are continuous and not simultaneously zero. A *contour* is a continuous chain of a finite number of smooth arcs that do not cross the same point twice. *Closed* means that the starting and ending points are identical.

Lemma 3.1. (Residue theorem) If f is analytic on a region containing a closed contour C except for a finite number of singularities $z_1, \ldots, z_m$ interior to C at which f has residues $k_1, \ldots, k_n$, then (for counterclockwise integration over C),

$$\int_C f(z)dz = 2\pi i \sum_{j=1}^{n} k_j \begin{cases} = 0 & \text{if } f \text{ is analytic,} \\ 2\pi i(z - z_0)f(z_0) & \text{for one simple pole at } z_0. \end{cases}$$

The Continuity Theorem

Theorem 9.5.1 (Continuity theorem for chfs: Cramér - Lévy)
(i) If $\phi_n \to \phi$ (pointwise) where ϕ is continuous at 0, then ϕ is the chf of a bona fide d.f. F and $F_n \to_d F$.
(ii) $F_n \to_d F$ implies $\phi_n \to \phi$ uniformly on any finite interval $|t| \leq T$.

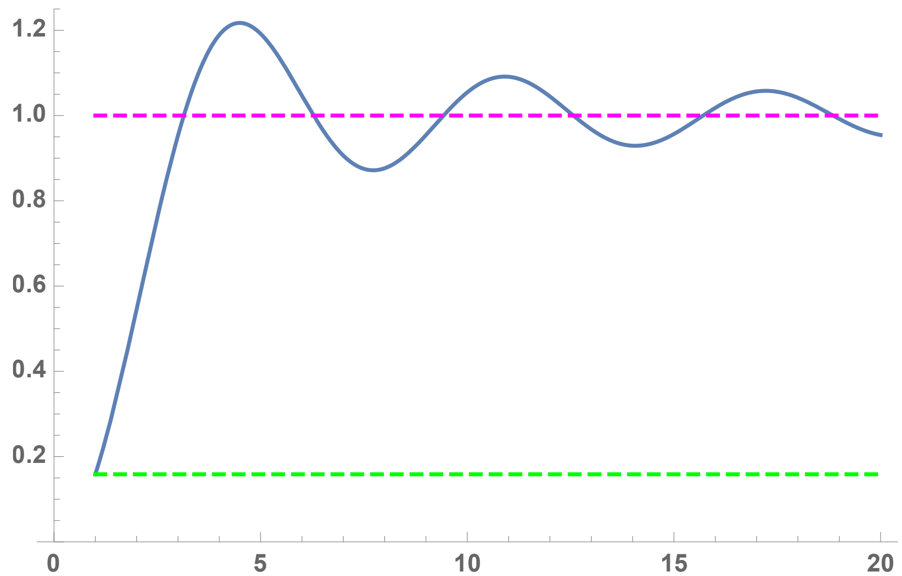
Inequality 9.5.1 (Chf bound on the tails of a d.f.) For any rv X with d.f. F we have

$$P(|X| \geq \lambda) \leq 7\lambda \int_0^{1/\lambda} (1 - Re(\phi(t)))\,dt \quad \text{for all } \lambda > 0. \tag{6}$$

Proof: Note that

$$\begin{aligned}
\lambda \int_0^{1/\lambda} (1 - Re(\phi(t)))\, dt &= \lambda \int_0^{1/\lambda} \int_{-\infty}^{\infty} (1 - \cos(tx)) dF(x) dt \\
&= \int_{-\infty}^{\infty} \lambda \int_0^{1/\lambda} (1 - \cos(tx)) dt dF(x) \\
&= \int_{-\infty}^{\infty} \left\{ \lambda t \left(1 - \frac{\sin(xt)}{xt} \right) \right\} \Big|_0^{1/\lambda} dF(x) \\
&= \int_{-\infty}^{\infty} \left(1 - \frac{\sin(x/\lambda)}{(x/\lambda)} \right) dF(x) \\
&\geq \int_{[|x|/\lambda \geq 1]} \left(1 - \frac{\sin(x/\lambda)}{(x/\lambda)} \right) dF(x) \\
&= \inf_{|y| \geq 1} \left(1 - \frac{\sin(y)}{y} \right) P(|X| \geq \lambda) = [1 - \sin(1)] P(|X| \geq \lambda) \\
&= (.1585\ldots) P(|X| \geq \lambda) \geq P(|X| \geq \lambda)/7.
\end{aligned}$$

Thus (6) holds. $\square$



The function $\left(1 - \frac{\sin(y)}{y}\right)$; green $= 1 - \sin(1)$

Proof of Theorem 9.5.1: (i) By the uniqueness theorem, the collection $\mathcal{G}$ of complex exponentials form a determining class and the expectations of these are hypothesized to converge. Thus by the Helly-Bray theorem it suffices to show that $\{F_n\}$ is tight.

But by Lemma 9.5.1 we have

$$\begin{aligned}
\overline{\lim}_n P(|X_n| \geq \lambda) &\leq \overline{\lim}_n 7\lambda \int_0^{1/\lambda} [1 - \mathrm{Re}\phi_n(t)]dt \\
&= 7\lambda \int_0^{1/\lambda} [1 - \mathrm{Re}\phi(t)]dt \\
&\qquad \text{by the DCT with dominating function 2} \\
&\to 0 \quad \text{as} \quad \lambda \to \infty
\end{aligned}$$

since ϕ is continuous at 0. (Note that

$$\begin{aligned}
&\lim_{\lambda\to\infty} \lambda \int_0^{1/\lambda} [1 - \mathrm{Re}\phi(t)]dt \\
&= \lim_{\eta\to 0} \eta^{-1} \int_0^{\eta} [1 - \mathrm{Re}\phi(t)] = \lim_{\eta\to 0} [1 - \mathrm{Re}\phi(\eta)] = 0
\end{aligned}$$

by L'Hopital's rule.) Thus $\{F_n\}$ is tight.

(ii) Since $X_n \to_d X$, there exist rv's $Y_n \stackrel{d}{=} X_n$ and $Y \stackrel{d}{=} X$ such that $Y_n \to_{a.s.} Y$, it follows that

$$
\begin{aligned}
|\phi_n(t) - \phi(t)| &\leq E|e^{itY_n} - e^{itY}| \\
&\leq E|e^{it(Y_n - Y)} - 1| \\
&\leq E \sup_{|t| \leq T} |e^{it(Y_n - Y)} - 1| \\
&\to 0 \text{ by the DCT with dominating function } 2
\end{aligned}
$$

since $\sup\{|it(Y_n - Y)| : \ |t| \leq T\} \leq T|Y_n - Y| \to_{a.s.} 0$. $\square$

5. Higher dimensions

If $\underline{X} = (X_1, \ldots, X_d)'$ is a random vector with values in $\mathbb{R}^d$, then the characteristic function $\phi_{\underline{X}}(\underline{t})$ is given by

$$\phi_{\underline{X}}(\underline{t}) = Ee^{i\underline{t}'\underline{X}} = Ee^{i\sum_{j=1}^{d} t_j X_j} \quad \text{for } \underline{t} \in \mathbb{R}^d. \tag{7}$$

The uniqueness theorem ($\{g_t \equiv e^{i\underline{t}'\underline{x}} : \ \underline{t} \in \mathbb{R}^d\}$ is a determining class), and the Cramér - Lévy continuity theorem still holds with minor modifications of the previous proofs.

But we now develop an important tool (or device) for proving convergence in distribution in higher dimension $d > 1$. For $\underline{\lambda} \in \mathbb{R}^d$, the characteristic function of the (one-dimensional !) linear combination $\underline{\lambda}'\underline{X}$ is

$$\phi_{\underline{\lambda}'\underline{X}}(t) = E\text{exp}(it\underline{\lambda}'\underline{X}) \quad \text{for} \quad t \in \mathbb{R}.$$

Comparison of this with (7) shows that knowing the joint chf $\phi_{\underline{X}}(\underline{t})$ for all $\underline{t} \in \mathbb{R}^d$ is equivalent to knowing the one-dimensional chf $\phi_{\underline{\lambda}'\underline{X}}(t)$ for all $t \in \mathbb{R}$ and $\underline{\lambda}$ for which $|\underline{\lambda}| = 1$. This leads to:

Theorem 9.5.3 (Cramér - Wold device) If $\underline{X}_n = (X_{n1}, \ldots, X_{nd})'$ satisfy

$$\phi_{\underline{\lambda}'\underline{X}_n}(t) \to_d \phi_{\underline{\lambda}'\underline{X}}(t) \tag{8}$$

for all $t \in \mathbb{R}$ and for each $\underline{\lambda} \in \mathbb{R}^d$, then $\underline{X}_n \to_d \underline{X}$. It suffices to show (8) for all unit vectors $\underline{\lambda}$. Thus it also suffices to show that $\underline{\lambda}'\underline{X}_n \to_d \underline{\lambda}'\underline{X}$ for all such $\underline{\lambda}$.

G. Walther (1997) has given a proof of the Cramér - Wold device which avoids characteristic functions completely; see Pollard (2002).

Theorem 9.5.3. The random variables $X_1, \ldots, X_d$ are independent if and only if the joint chfs satisfy

$$\phi_{\underline{X}}(\underline{t}) = \prod_{i=1}^{d} \phi_{X_i}(t_i) \quad \text{for all} \quad \underline{t} = (t_1, \ldots, t_d) \in \mathbb{R}^d.$$