

# STA 711: Probability & Measure Theory

Robert L. Wolpert

## 9 Sums of Independent Random Variables

We continue our study of sums of independent random variables,  $S_n = X_1 + \dots + X_n$ . If each  $X_i$  is square-integrable, with mean  $\mu_i = \mathbb{E}X_i$  and variance  $\sigma_i^2 = \mathbb{E}[(X_i - \mu_i)^2]$ , then  $S_n$  is square integrable too with mean  $\mathbb{E}S_n = \mu_{\leq n} := \sum_{i \leq n} \mu_i$  and variance  $\mathbb{V}S_n = \sigma_{\leq n}^2 := \sum_{i \leq n} \sigma_i^2$ . But what about the actual probability distribution? If the  $X_i$  have density functions  $f_i(x_i)$  then  $S_n$  has a density function too; for example, with  $n = 2$ ,  $S_2 = X_1 + X_2$  has CDF  $F(s)$  and pdf  $f(s) = F'(s)$  given by

$$\begin{aligned} \mathbb{P}[S_2 \leq s] = F(s) &= \iint_{x_1 + x_2 \leq s} f_1(x_1)f_2(x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{s-x_2} f_1(x_1)f_2(x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} F_1(s-x_2)f_2(x_2) dx_2 = \int_{-\infty}^{\infty} F_2(s-x_1)f_1(x_1) dx_1 \\ f(s) = F'(s) &= \int_{-\infty}^{\infty} f_1(s-x_2)f_2(x_2) dx_2 = \int_{-\infty}^{\infty} f_1(x_1)f_2(s-x_1) dx_1, \end{aligned}$$

the *convolution*  $f = f_1 \star f_2$  of  $f_1(x_1)$  and  $f_2(x_2)$ . Even if the distributions aren't absolutely continuous, so no pdfs exist,  $S_2$  has a distribution measure  $\mu$  given by  $\mu(ds) = \int_{\mathbb{R}} \mu_1(dx_1)\mu_2(ds-x_1)$ . There is an analogous formula for  $n = 3$ , but it is quite messy; things get worse and worse as  $n$  increases, so this is not a promising approach for studying the distribution of sums  $S_n$  for large  $n$ .

If CDFs and pdfs of sums of independent RVs are not simple, is there some other feature of the distributions that is? The answer is *Yes*. What is simple about independent random variables is calculating expectations of products of the  $X_i$ , or products of any functions of the  $X_i$ ; the exponential function will let us turn the partial *sums*  $S_n$  into *products*  $e^{S_n} = \prod e^{X_i}$  or, more generally,  $e^{zS_n} = \prod e^{zX_i}$  for any real or complex number  $z$ . Thus for independent RVs  $X_i$  and any number  $z$  we can use independence to compute the expectation

$$\mathbb{E}e^{zS_n} = \prod_{i=1}^n \mathbb{E}e^{zX_i},$$

often called the “moment generating function” and denoted  $M_X(z) = \mathbb{E}e^{zX}$  for any random variable  $X$ .

For real  $z$  the function  $e^{zX}$  becomes huge if  $X$  becomes very large (for positive  $z$ ) or very negative (if  $z < 0$ ), so that even for integrable or square-integrable random variables  $X$  the

expectation  $M(z) = \mathbf{E}e^{zX}$  may be infinite. Here are a few examples of  $\mathbf{E}e^{zX}$  for some familiar distributions (try to verify some of them):

Binomial:	$\text{Bi}(n, p)$	$[1 + p(e^z - 1)]^n$	$z \in \mathbb{C}$
Neg Bin:	$\text{NB}(\alpha, p)$	$[1 - (p/q)(e^z - 1)]^{-\alpha}$	$z \in \mathbb{C}$
Poisson	$\text{Po}(\lambda)$	$e^{\lambda(e^z - 1)}$	$z \in \mathbb{C}$
Normal:	$\text{No}(\mu, \sigma^2)$	$e^{z\mu + z^2\sigma^2/2}$	$z \in \mathbb{C}$
Gamma:	$\text{Ga}(\alpha, \lambda)$	$(1 - z/\lambda)^{-\alpha}$	$\Re(z) < \lambda$
Cauchy:	$\frac{a}{\pi(a^2 + (x-b)^2)}$	$e^{zb - a z }$	$\Re(z) = 0$
Uniform:	$\text{Un}(a, b)$	$\frac{1}{z(b-a)}[e^{zb} - e^{za}]$	$z \in \mathbb{C}$

Aside from the problem that  $M(z) = \mathbf{E}e^{zX}$  may be infinite or fail to exist for some  $z \in \mathbb{C}$ , the approach is promising: we *can* identify the probability distribution from  $M(z)$ , and we can even find important features about the distribution directly from  $M$ . For example, if we can justify interchanging the limits implicit in differentiation and integration, then  $M'(z) = \mathbf{E}[Xe^{zX}]$  and  $M''(z) = \mathbf{E}[X^2e^{zX}]$ , so (upon taking  $z = 0$ )  $M'(0) = \mathbf{E}X = \mu$  and  $M''(0) = \mathbf{E}X^2 = \sigma^2 + \mu^2$ . This lets us calculate the mean and variance (and other moments  $\mathbf{E}X^k = M^{(k)}(0)$ ) from derivatives of  $M(z)$  at zero. We have two problems to overcome: discovering how to infer the distribution of  $X$  from  $M_X(z) = \mathbf{E}e^{zX}$ , and what to do about distributions for which  $M(z)$  is infinite or doesn't exist.

## 9.1 Characteristic Functions

For complex numbers  $z = x + iy$ , Leonard Euler showed that the exponential  $e^z$  can be given in terms of familiar real-valued transcendental functions as  $e^{x+iy} = e^x \cos(y) + ie^x \sin(y)$ . Since both  $\sin(y)$  and  $\cos(y)$  are bounded by one, for *any* real-valued random variable  $X$  and real number  $\omega$  the real and imaginary parts of the complex-valued random variable  $e^{i\omega X}$  are bounded and hence integrable; thus it *always* makes sense to define the *characteristic function*

$$\phi_X(\omega) = \mathbf{E}e^{i\omega X} = \int_{\mathbb{R}} e^{i\omega x} \mu_X(dx), \quad \omega \in \mathbb{R}$$

with finite absolute value  $|\phi_X(\omega)| \leq 1$ . Of course this is just  $\phi_X(\omega) = M_X(i\omega)$  when  $M_X$  exists, but  $\phi_X(\omega)$  exists even when  $M_X$  does not; on the chart above you'll notice that only the *real* part of  $z$  posed problems, and  $\Re(z) = 0$  was always OK, even for the Cauchy. For real-valued  $\omega \in \mathbb{R}$ , some familiar distributions' ch.f.s are:

Binomial:	$\text{Bi}(n, p)$	$\phi(\omega) = [1 + p(e^{i\omega} - 1)]^n$
Neg Bin:	$\text{NB}(\alpha, p)$	$\phi(\omega) = [1 - (p/q)(e^{i\omega} - 1)]^{-\alpha}$
Poisson	$\text{Po}(\lambda)$	$\phi(\omega) = e^{\lambda(e^{i\omega} - 1)}$
Normal:	$\text{No}(\mu, \sigma^2)$	$\phi(\omega) = e^{i\omega\mu - \omega^2\sigma^2/2}$
Gamma:	$\text{Ga}(\alpha, \lambda)$	$\phi(\omega) = (1 - i\omega/\lambda)^{-\alpha}$
Cauchy:	$\frac{a/\pi}{a^2 + (x-b)^2}$	$\phi(\omega) = e^{i\omega b - a \omega }$
Uniform:	$\text{Un}(a, b)$	$\phi(\omega) = \frac{1}{i\omega(b-a)}[e^{i\omega b} - e^{i\omega a}]$

## Another Perspective

Fourier transforms arise frequently in engineering and signal analysis. A recent STA711 student found a youtube video presenting an interesting way of visualizing them; take a look at this URL <https://youtu.be/spUNpyF58BY> if you're interested.

### 9.1.1 Uniqueness

Suppose that two probability distributions  $\mu_1(A) = \mathbb{P}[X_1 \in A]$  and  $\mu_2(A) = \mathbb{P}[X_2 \in A]$  have the same Fourier transform  $\hat{\mu}_1 := \hat{\mu}_2$ , where:

$$\hat{\mu}_j(\omega) = \mathbb{E}[e^{i\omega X_j}] = \int_{\mathbb{R}} e^{i\omega x} \mu_j(dx);$$

does it follow that  $X_1$  and  $X_2$  have the same probability distributions, *i.e.*, that  $\mu_1 = \mu_2$ ? The answer is *yes*; in fact, one can recover the measure  $\mu$  explicitly from the function  $\hat{\mu}(\omega)$ . Thus we regard uniqueness as a corollary of the much stronger result, the Fourier Inversion Theorem.

[Resnick \(1999\)](#) has lots of interesting results about characteristic functions in Chapter 9, [Grimmett and Stirzaker \(2001\)](#) discuss related results in their Chapter 5, and [Billingsley \(1995\)](#) proves several versions of this theorem in his Section 26. I'm going to take a different approach, and stress the two special cases in which  $\mu$  is discrete or has a density function, trying to make some connections with other encounters you might have had with Fourier transforms.

### 9.1.2 Positive Definiteness

Which functions  $\phi(\omega)$  can be characteristic functions? We know that  $|\phi(\omega)| \leq 1$  for every  $\omega \in \mathbb{R}$ , and that  $\phi(0) = 1$ . In a homework exercise you showed that  $\phi(\omega)$  must be uniformly continuous, too— is that enough?

The answer is *no*. Each ch.f. has the interesting property that it is “positive definite,” in the following sense:

**Definition 1** A function  $\phi : \mathbb{R} \rightarrow \mathbb{C}$  is **positive definite** if for every  $n \in \mathbb{N}$ ,  $z \in \mathbb{C}^n$ , and  $\omega \in \mathbb{R}^n$ ,

$$\sum_{j,k=1}^n z_j \phi(\omega_j - \omega_k) \bar{z}_k \geq 0$$

or, equivalently, that each  $n \times n$  matrix  $A_{jk} := \phi(\omega_j - \omega_k)$  is positive-definite.

Here's a proof that  $\phi(\omega) := \int_{\mathbb{R}} e^{i\omega x} \mu(dx)$  is positive definite, for every distribution  $\mu$  on

$(\mathbb{R}, \mathcal{B})$ , with the interchange of summation & integration justified by Fubini's Theorem:

$$\begin{aligned} \sum_{j,k=1}^n z_j \phi(\omega_j - \omega_k) \bar{z}_k &= \sum_{j,k=1}^n \int_{\mathbb{R}} z_j e^{ix(\omega_j - \omega_k)} \mu(dx) \bar{z}_k \\ &= \int_{\mathbb{R}} \left\{ \sum_{j=1}^n z_j e^{ix\omega_j} \right\} \overline{\left\{ \sum_{k=1}^n z_k e^{ix\omega_k} \right\}} \mu(dx) \\ &= \int_{\mathbb{R}} \left| \sum_{j=1}^n z_j e^{ix\omega_j} \right|^2 \mu(dx) \\ &\geq 0. \end{aligned}$$

Interestingly, this condition is also sufficient:

**Theorem 1 (Bochner)** *If  $\phi : \mathbb{R} \rightarrow \mathbb{C}$  is continuous at zero, satisfies  $\phi(0) = 1$ , and is positive definite, then there exists a Borel probability measure  $\mu$  on  $(\mathbb{R}, \mathcal{B})$  such that  $\phi(\omega) = \int_{\mathbb{R}} e^{i\omega x} \mu(dx)$  for each  $\omega \in \mathbb{R}$ .*

Here's a proof sketch for the special (but common) case where  $\phi \in L_1(\mathbb{R}, d\omega)$ . By positive definiteness, for any  $\{\omega_j\} \subset \mathbb{R}$  and  $\{z_j\} \subset \mathbb{C}$ ,

$$0 \leq \sum z_j \phi(\omega_j - \omega_k) \bar{z}_k$$

and in particular, for  $x \in \mathbb{R}$ ,  $\epsilon > 0$ , and  $z_j := \exp(-ix\omega_j - \epsilon\omega_j^2/2)$ ,

$$0 \leq \sum e^{-ix(\omega_j - \omega_k) - \epsilon(\omega_j^2 + \omega_k^2)/2} \phi(\omega_j - \omega_k).$$

Taking  $\omega_j := j/n$  for  $-n^2 \leq j \leq n^2$  and then taking the limit as  $n \rightarrow \infty$ ,

$$0 \leq \iint_{\mathbb{R}^2} e^{-ix(u-v) - \epsilon(u^2 + v^2)/2} \phi(u-v) du dv$$

Now change variables from  $v$  to  $\omega := (u - v)$ :

$$\begin{aligned} &= \iint_{\mathbb{R}^2} e^{-ix\omega - \epsilon[u^2 + (u^2 - 2u\omega + \omega^2)]/2} \phi(\omega) du d\omega \\ &= \int_{\mathbb{R}} e^{-ix\omega - \epsilon\omega^2/2} \left\{ \int_{\mathbb{R}} e^{-\epsilon(u-\omega/2)^2 + \omega^2/4} du \right\} \phi(\omega) d\omega \\ &= \sqrt{\pi/\epsilon} \int_{\mathbb{R}} e^{-ix\omega - \epsilon\omega^2/4} \phi(\omega) d\omega \end{aligned}$$

Re-scaling and then taking  $\epsilon \rightarrow 0$ , we find that  $f(x) := \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ix\omega} \phi(\omega) d\omega \geq 0$  for every  $x \in \mathbb{R}$  and can verify that  $\phi(\omega) = \int_{\mathbb{R}} e^{i\omega x} \mu(dx)$  for the absolutely-continuous distribution given by  $\mu(dx) = f(x) dx$ .

### 9.1.3 Inversion: Integer-valued Discrete Case

Notice that the integer-valued discrete distributions always satisfy  $\phi(\omega + 2\pi) = \phi(\omega)$  (and in particular are *not* integrable over  $\mathbb{R}$ ), while the continuous ones satisfy  $|\phi(\omega)| \rightarrow 0$  as  $\omega \rightarrow \pm\infty$ . For integer-valued random variables  $X$  we can recover the probability mass function  $p_k := \mathbb{P}[X = k]$  by inverting the Fourier series:

$$\begin{aligned}\phi(\omega) &= \mathbb{E}[e^{i\omega X}] = \sum p_k e^{ik\omega}, \text{ so (by Fubini's thm)} \\ p_k &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ik\omega} \phi(\omega) d\omega.\end{aligned}$$

### 9.1.4 Inversion: Continuous Random Variables

Now let's turn to the case of a distribution with a density function; first two preliminaries. For any real or complex numbers  $a, b, c$  it is easy to compute (by completing the square) that

$$\int_{-\infty}^{\infty} e^{-a-bx-cx^2} dx = \sqrt{\frac{\pi}{c}} e^{-a+b^2/4c} \quad (1)$$

if  $c$  has positive real part, and otherwise the integral is infinite. In particular, for any  $\epsilon > 0$  the function  $\gamma_\epsilon(x) := \frac{1}{\sqrt{2\pi\epsilon}} e^{-x^2/2\epsilon}$  satisfies  $\int \gamma_\epsilon(x) dx = 1$  (it's just the normal pdf with mean 0 and variance  $\epsilon$ ).

Let  $\mu(dx) = f(x)dx$  be any probability distribution with density function  $f(x)$  and ch.f.  $\phi(\omega) = \hat{\mu}(\omega) = \int e^{i\omega x} f(x) dx$ . Then  $|\phi(\omega)| \leq 1$  so for any  $\epsilon > 0$  the function  $|e^{-iy\omega - \epsilon\omega^2/2}\phi(\omega)|$  is bounded above by  $e^{-\epsilon\omega^2/2}$  and so is integrable w.r.t.  $\omega$  over  $\mathbb{R}$ . We can compute

$$\begin{aligned}\frac{1}{2\pi} \int_{\mathbb{R}} e^{-iy\omega - \epsilon\omega^2/2} \phi(\omega) d\omega &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iy\omega - \epsilon\omega^2/2} \left[ \int_{\mathbb{R}} e^{ix\omega} f(x) dx \right] d\omega \\ &= \frac{1}{2\pi} \int_{\mathbb{R}^2} e^{i(x-y)\omega - \epsilon\omega^2/2} f(x) dx d\omega \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} e^{i(x-y)\omega - \epsilon\omega^2/2} d\omega \right] f(x) dx \quad (2)\end{aligned}$$

$$\begin{aligned}&= \frac{1}{2\pi} \int_{\mathbb{R}} \left[ \sqrt{\frac{2\pi}{\epsilon}} e^{-(x-y)^2/2\epsilon} \right] f(x) dx \quad (3) \\ &= \frac{1}{\sqrt{2\pi\epsilon}} \int_{\mathbb{R}} e^{-(x-y)^2/2\epsilon} f(x) dx \\ &= [\gamma_\epsilon \star f](y) = [\gamma_\epsilon \star \mu](y)\end{aligned}$$

(where the interchange of orders of integration in (2) is justified by Fubini's theorem and the calculation in (3) by equation (1)), the convolution of the normal kernel  $\gamma_\epsilon(\cdot)$  with  $f(y)$ . As  $\epsilon \rightarrow 0$  this converges

- uniformly (and in  $L_1$ ) to  $f(y)$  if  $f(\cdot)$  is bounded and continuous, the most common case;
- pointwise to  $\frac{f(y^-)+f(y^+)}{2}$  if  $f(x)$  has left and right limits at  $x = y$ ; and
- to infinity if  $\mu(\{y\}) > 0$ , *i.e.*, if  $\mathbf{P}[X = y] > 0$ .

This is the Fourier Inversion Formula for  $f(x)$ : we can recover the density  $f(x)$  from its Fourier transform  $\phi(\omega) = \hat{\mu}(\omega)$  by  $f(x) = \frac{1}{2\pi} \int e^{-i\omega x} \phi(\omega) d\omega$ , if that integral exists, or otherwise as the limit

$$f(x) = \lim_{\epsilon \rightarrow 0} \frac{1}{2\pi} \int e^{-i\omega x - \epsilon \omega^2 / 2} \phi(\omega) d\omega.$$

There are several interesting connections between the density function  $f(x)$  and characteristic function  $\phi(\omega)$ . If  $\phi(\omega)$  “wiggles” with rate approximately  $\xi$ , *i.e.*, if  $\phi(\omega) \approx a \cos(\omega\xi) + b \sin(\omega\xi) + c$ , then  $f(x)$  will have a spike at  $x = \xi$  and  $X$  will have a high probability of being close to  $\xi$ ; if  $\phi(\omega)$  is very smooth (*i.e.*, has well-behaved continuous derivatives of high order) then it does *not* have high-frequency wiggles and  $f(x)$  falls off quickly for large  $|x|$ , so  $\mathbf{E}[|X|^p] < \infty$  for large  $p$ . If  $|\phi(\omega)|$  falls off quickly as  $\omega \rightarrow \pm\infty$  then  $\phi(\omega)$  doesn’t have large *low*-frequency components and  $f(x)$  must be rather tame, without any spikes. Thus  $\phi$  and  $f$  both capture information about the distribution, but from different perspectives. This is often useful, for the vague descriptions of this paragraph can be made precise:

**Theorem 2** If  $\int_{\mathbb{R}} |\hat{\mu}(\omega)| d\omega < \infty$  then  $\mu_\epsilon := \mu * \gamma_\epsilon$  converges a.s as  $\epsilon \rightarrow 0$  to an  $L_1$  function  $f(x)$ ,  $\hat{\mu}_\epsilon(\omega) := \int e^{i\omega x} \mu_\epsilon(dx)$  converges uniformly to  $\hat{f}(\omega) := \int e^{i\omega x} f(x) dx$ , and  $\mu(A) = \int_A f(x) dx$  for each Borel  $A \subset \mathbb{R}$ . Also  $f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\omega x} \hat{\mu}(\omega) d\omega$  for almost-every  $x$ .

**Theorem 3** For any distribution  $\mu$  and real numbers  $a < b$ ,

$$\mu(a, b) + \frac{1}{2} \mu(\{a, b\}) = \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-i\omega a} - e^{-i\omega b}}{2\pi i \omega} \hat{\mu}(\omega) d\omega.$$

**Theorem 4** If  $\int_{\mathbb{R}} |x|^k \mu(dx) < \infty$  for an integer  $k \geq 0$  then  $\hat{\mu}(\omega)$  has continuous derivatives of order  $k$  given by

$$\hat{\mu}^{(k)}(\omega) = \int_{\mathbb{R}} (ix)^k e^{i\omega x} \mu(dx). \quad (1)$$

Conversely, if  $\hat{\mu}(\omega)$  has a derivative of finite **even** order  $k$  at  $\omega = 0$ , then  $\int_{\mathbb{R}} |x|^k \mu(dx) < \infty$  and

$$\mathbf{E}X^k = \int_{\mathbb{R}} x^k \mu(dx) = (-1)^{k/2} \hat{\mu}^{(k)}(0). \quad (2)$$

To prove (1) first note it's true by definition for  $k = 0$ , then apply induction:

$$\begin{aligned}\hat{\mu}^{(k+1)}(\omega) &= \lim_{\epsilon \rightarrow 0} \int_{\mathbb{R}} (ix)^k \left( \frac{e^{i\epsilon x} - 1}{\epsilon} \right) e^{i\omega x} \mu(dx) \\ &= \int_{\mathbb{R}} (ix)^{k+1} e^{i\omega x} \mu(dx)\end{aligned}$$

by LDCT since  $|e^{i\epsilon x} - 1| \leq \epsilon|x|$ .

By Theorem 4 the first few moments of the distribution, if they exist, can be determined from derivatives of the characteristic function *or its logarithm*  $\log \phi(z)$  at  $z = 0$ :  $\phi(0) = 1$ ,  $\phi'(0) = i\mathbb{E}[X]$ ,  $\phi''(0) = -\mathbb{E}[X^2]$ , so

$$\begin{array}{llll} \text{Mean:} & [\log \phi](0) & = \log \phi(0) & = \log 1 & = 0 \\ & [\log \phi]'(0) & = \phi'(0)/\phi(0) & = i\mathbb{E}[X] & = i\mu \\ \text{Variance:} & [\log \phi]''(0) & = \frac{\phi''(0)\phi(0) - (\phi'(0))^2}{\phi(0)^2} & = \mathbb{E}[X]^2 - \mathbb{E}[X^2] & = -\sigma^2 \\ \text{Etc.:} & [\log \phi]'''(0) & = -i\mathbb{E}[X^3] - 3\sigma^2\mu - \mu^3 & & \leq c\mathbb{E}|X|^3 \end{array}$$

for some  $c < \infty$ , so by Taylor's theorem we have:<sup>1</sup>

$$\begin{aligned}\log \phi(\omega) &= 0 + i\mu\omega - \sigma^2\omega^2/2 + \mathcal{O}(\omega^3) \\ \phi(\omega) &\approx e^{i\mu\omega - \sigma^2\omega^2/2 + \mathcal{O}(\omega^3)}\end{aligned}\tag{3}$$

### 9.1.5 Convergence in Distribution

In the Week 6 Notes we defined *convergence in distribution* of a sequence of distributions  $\{\mu_n\}$  to a limiting distribution  $\mu$  on a measurable space  $(\mathcal{X}, \mathcal{E})$  (written  $\mu_n \Rightarrow \mu$ ):

$$(\forall \phi \in \mathcal{C}_b(\mathcal{X})) \lim_{n \rightarrow \infty} \int_{\mathcal{X}} \phi(x) \mu_n(dx) = \int_{\mathcal{X}} \phi(x) \mu(dx)\tag{4}$$

In fact requiring this convergence for *all* bounded continuous functions  $\phi$  is more than what is necessary. For  $\mathcal{X} = \mathbb{R}^d$ , for example, it is enough to verify (4) for infinitely-differentiable  $\mathcal{C}_b^\infty$ , or even just for complex exponentials  $\phi_\omega(x) = \exp(i\omega'x)$  for  $\omega \in \mathbb{R}^d$ , *i.e.*,

**Theorem 5** *Let  $\{\mu_n(dx)\}$  and  $\mu(dx)$  be distributions on Euclidean space  $(\mathbb{R}^d, \mathcal{B})$ . Then  $\mu_n \Rightarrow \mu$  if and only if the characteristic functions converge pointwise, *i.e.*, if*

$$\phi_n(\omega) := \int_{\mathbb{R}^d} e^{i\omega'x} \mu_n(dx) \rightarrow \phi(\omega) := \int_{\mathbb{R}^d} e^{i\omega'x} \mu(dx)\tag{5}$$

for all  $\omega \in \mathbb{R}^d$ .

How would you prove this?

---

<sup>1</sup>The “big oh” notation “ $f = \mathcal{O}(g)$  at  $a$ ” means that for some  $M < \infty$  and  $\epsilon > 0$ ,  $|f(x)| \leq Mg(x)$  whenever  $|x - a| < \epsilon$ —roughly, that  $\limsup_{x \rightarrow a} |f(x)/g(x)| < \infty$ . Here (implicitly)  $a = 0$ .

## Examples

Un: Let  $X_n$  have the discrete uniform distribution on the points  $j/n$ , for  $1 \leq j \leq n$ . Then its ch.f. is

$$\begin{aligned}\phi_n(\omega) &= \frac{1}{n} \sum_{j=1}^n e^{i\omega j/n} \\ &= \frac{e^{i\omega/n} - e^{i(n+1)\omega/n}}{n(1 - e^{i\omega/n})} \\ &= \frac{1 - e^{i\omega}}{n(e^{-i\omega/n} - 1)} \\ &\rightarrow \frac{1 - e^{i\omega}}{-i\omega} = \frac{e^{i\omega} - 1}{i\omega} = \int_0^1 e^{i\omega x} dx,\end{aligned}$$

the ch.f. of the  $\text{Un}(0, 1)$  distribution.

Po: Let  $X_n$  have Binomial  $\text{Bi}(n, p_n)$  distributions with success probabilities  $p_n \asymp \lambda/n$ , so that  $np_n \rightarrow \lambda$  for some  $\lambda > 0$  as  $n \rightarrow \infty$ . Then the ch.f.s satisfy

$$\begin{aligned}\phi_n(\omega) &= \sum_{k=0}^n \binom{n}{k} e^{i\omega k} p_n^k (1 - p_n)^{n-k} \\ &= [1 + p_n(e^{i\omega} - 1)]^n \rightarrow e^{(e^{i\omega} - 1)\lambda},\end{aligned}$$

the ch.f. of the  $\text{Po}(\lambda)$  distribution. This is an example of a “law of small numbers”.

No: Let  $X_n$  have Binomial  $\text{Bi}(n, p_n)$  distributions with success probabilities  $p_n$  such that  $\sigma_n^2 := np_n(1 - p_n) \rightarrow \infty$  as  $n \rightarrow \infty$ , and set  $\mu_n := np_n$ . Then the ch.f.s of  $Z_n := (X_n - \mu_n)/\sigma_n$  satisfy

$$\begin{aligned}\phi_n(\omega) &= [1 + p_n(e^{i\omega/\sigma_n} - 1)]^n e^{-i\omega\mu_n/\sigma_n} \\ &\approx \exp \left\{ \mu_n(e^{i\omega/\sigma_n} - 1) - p_n\mu_n(e^{i\omega/\sigma_n} - 1)^2/2 - i\omega\mu_n/\sigma_n \right\} \\ &\approx \exp \left\{ i\mu_n\omega/\sigma_n - \mu_n\omega^2/2\sigma_n^2 - p_n\mu_n(-\omega^2/\sigma_n^2)/2 - i\omega\mu_n/\sigma_n \right\} \rightarrow e^{-\omega^2/2},\end{aligned}$$

the ch.f. of the  $\text{No}(0, 1)$  distribution. This result is called the “DeMoivre-Laplace” theorem, a pre-cursor (and special case) of the Central Limit Theorem.

## 9.2 Limits of Partial Sums and the Central Limit Theorem

Let  $\{X_i\}$  be iid and  $L_2$ , with common mean  $\mu$  and variance  $\sigma^2$ , and set  $S_n := \sum_{i=1}^n X_i$  for  $n \in \mathbb{N}$ . We’ll need to center and scale the distribution of  $S_n$  before we can hope to make sense of  $S_n$ ’s distribution for large  $n$ , so we’ll need some facts about characteristic functions



of linear combinations of independent RVs. For independent  $X$  and  $Y$ , and real numbers  $\alpha$ ,  $\beta$ ,  $\gamma$ ,

$$\phi_{\alpha+\beta X+\gamma Y}(\omega) = \mathbf{E}e^{i\omega(\alpha+\beta X+\gamma Y)} = \mathbf{E}e^{i\omega\alpha}\mathbf{E}e^{i\omega\beta X}\mathbf{E}e^{i\omega\gamma Y} = e^{i\omega\alpha}\phi_X(\omega\beta)\phi_Y(\omega\gamma).$$

In particular, for iid  $L_3$  random variables  $\{X_i\}$  with characteristic function  $\phi(t)$ , the normalized sum  $[S_n - n\mu]/\sqrt{n\sigma^2}$  has characteristic function

$$\phi_n(\omega) = \prod_{j=1}^n \left[ \phi(\omega/\sqrt{n\sigma^2})e^{-i\omega\mu/\sqrt{n\sigma^2}} \right]$$

Setting  $s := \omega/\sqrt{n\sigma^2}$ , this is

$$= [\phi(s)e^{-is\mu}]^n = e^{n[\log \phi(s) - is\mu]}$$

with logarithm

$$\begin{aligned} \log \phi_n(\omega) &= n[\log \phi(s) - is\mu] \\ &= n[0 + i\mu s - \sigma^2 s^2/2 + \mathcal{O}(s^3)] - nis\mu \quad (\text{by (3)}) \\ &= -n[\sigma^2(\omega^2/n\sigma^2)/2 + \mathcal{O}(n^{-3/2})] \quad (\text{since } s^2 = \omega^2/n\sigma^2) \\ &= -\omega^2/2 + \mathcal{O}(n^{-1/2}), \end{aligned}$$

so  $\phi_n(\omega) \rightarrow e^{-\omega^2/2}$  for all  $\omega \in \mathbb{R}$  and hence  $Z_n := [S_n - n\mu]/\sqrt{n\sigma^2} \Rightarrow \text{No}(0, 1)$ , the Central Limit Theorem.

**Note:** We assumed  $X_i$  were iid with finite third moment  $\gamma := \mathbf{E}|X_i|^3 < \infty$ . Under those conditions one can prove the uniform ‘‘Berry-Esséen’’ bound  $\sup_x |F_n(x) - \Phi(x)| \leq \gamma/[2\sigma^3\sqrt{n}]$  for the CDF  $F_n$  of  $Z_n$ . Another version of the CLT for iid  $\{X_i\}$  asserts weak convergence of  $Z_n$  to  $\text{No}(0, 1)$  assuming only  $\mathbf{E}[X_i^2] < \infty$  (*i.e.*, no  $L_3$  requirement), but this version gives no bound on the difference of the CDFs. Another famous version, due to Lindeberg and Feller, asserts that

$$\frac{S_n}{s_n} \Longrightarrow \text{No}(0, 1)$$

for partial sums  $S_n = X_1 + \dots + X_n$  of independent mean-zero  $L_2$  random variables  $X_j$  that need not be identically distributed, but whose variances  $\sigma_j^2 = \mathbf{V}[X_j]$  aren’t too extreme. The specific condition, for  $s_n^2 := \sigma_1^2 + \dots + \sigma_n^2$ , is

$$\frac{1}{s_n^2} \sum_{j=1}^n \mathbf{E} \{ X_j^2 \mathbf{1}_{\{|X_j| > ts_n\}} \} \rightarrow 0$$

as  $n \rightarrow \infty$  for each  $t > 0$ . This follows immediately for iid  $\{X_j\} \subset L_2$  (where it becomes  $\frac{1}{\sigma^2} \mathbf{E} [ |X_1|^2 \mathbf{1}_{\{|X_1|^2 > nt^2\sigma^2\}} ]$  which tends to zero as  $n \rightarrow \infty$  by the DCT), but for applications it's important to know that independent but non-iid summands still lead to a CLT.

This “Lindeberg Condition” implies both of

$$\max_{j \leq n} \frac{\sigma_j^2}{s_n^2} \rightarrow 0 \qquad \max_{j \leq n} \mathbf{P} \{ |X_j|/s_n > \epsilon \} \rightarrow 0$$

as  $n \rightarrow \infty$ , for any  $\epsilon > 0$ ; roughly, no single  $X_j$  is allowed to dominate the sum  $S_n$ . This condition follows from the easier-to-verify Liapunov Condition,  $(\exists \delta > 0)$  s.t.:

$$s_n^{-2-\delta} \sum_{j=1}^n \mathbf{E} |X_j|^{2+\delta} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Other versions of the CLT apply to non-identically distributed or nonindependent  $\{X_j\}$ , but  $S_n$  *cannot* converge to a normally-distributed limit if  $\mathbf{E}[X^2] = \infty$ ; ask for details (or read [Gnedenko and Kolmogorov \(1968\)](#)) if you're interested.

More recently an interesting new approach to proving the Central Limit Theorem and related estimates with error bounds was developed by Charles Stein ([Stein, 1972, 1986](#); [Barbour and Chen, 2005](#)), described later in these notes.

### 9.3 Failure of Central Limit Theorem

The CLT only applies to square-integrable random variables  $\{X_j\} \subset L_2$ . Some contemporary statistical work, both theoretical and applied, entails heavier-tailed distributions that do not have a finite variance (or, often, even a finite mean). In these cases, sums and averages of independent random variables do *not* have approximate normal distributions, and may not even be concentrated around a central value.

For example, if  $\{X_i\} \stackrel{\text{iid}}{\sim} \text{Ca}(m, s)$  are IID Cauchy random variables with pdf and ch.f.

$$f(x) = \frac{s/\pi}{s^2 + (x - m)^2} \qquad \chi(\omega) = \exp(im\omega - s|\omega|)$$

then the sample mean  $\bar{X}_n \sim \text{Ca}(m, s)$  *also* has the same Cauchy distribution— and, in particular, no weak or strong LLN applies and no CLT applies.

Worse— if  $\{\zeta_i\} \stackrel{\text{iid}}{\sim} \text{No}(0, 1)$  are IID standard Normals, then the random variables  $X_i := 1/\zeta_i |\zeta_i|$  have the symmetric  $\alpha$ -stable  $\text{St}(\frac{1}{2}, 0, 1, 0)$  distribution with ch.f.  $\chi_1(\omega) := \exp(-|\omega|^{1/2})$ , so the sample average  $\bar{X}_n$  has ch.f.

$$\chi_n(\omega) = \left[ \chi_1(\omega/n) \right]^n = \exp(-|n\omega|^{1/2}),$$

the same distribution as  $nX_1$ . The average of several independent replicates has a much *wider* distribution than the individual terms.

Heavy-tailed distributions like the Fréchet,  $\alpha$ -Stable, and Pareto arise when modeling income distribution, weather extremes, volcanic flows, and many other phenomena.

## 10 Distributional Limits of Heavy-Tailed Sums

Although one cannot expect  $\bar{X}_n$  to have an approximate normal distribution for  $\{X_i\} \not\subset L_2$ , other distributional limit theorems may still apply. Before we introduce those, we introduce some tools useful for studying heavy-tailed distributions.

### 10.1 Compound Poisson Distributions

Let  $X_j$  have independent Poisson distributions with means  $\nu_j$  and let  $u_j \in \mathbb{R}$ ; then the ch.f. for  $Y := \sum u_j X_j$  is

$$\begin{aligned} \phi_Y(\omega) &= \prod \exp [\nu_j (e^{i\omega u_j} - 1)] \\ &= \exp \left[ \sum (e^{i\omega u_j} - 1) \nu_j \right] \\ &= \exp \left[ \int_{\mathbb{R}} (e^{i\omega u} - 1) \nu(du) \right] \end{aligned}$$

for the discrete measure  $\nu(du) = \sum \nu_j \delta_{u_j}(du)$  that assigns mass  $\nu_j$  to each point  $u_j$ . Evidently we could take a limit using a sequence of discrete measures that converges to a continuous measure  $\nu(du)$  so long as the integral makes sense, *i.e.*,  $\int_{\mathbb{R}} |e^{i\omega u} - 1| \nu(du) < \infty$ ; this will follow from the requirement that  $\int_{\mathbb{R}} (1 \wedge |u|) \nu(du) < \infty$ . Such a distribution is called *Compound Poisson*, at least when  $\nu_+ := \nu(\mathbb{R}) < \infty$ ; in that case we can also write represent it in the form

$$Y = \sum_{i=1}^N X_i, \quad N \sim \text{Po}(\nu_+), \quad X_i \stackrel{\text{iid}}{\sim} \nu(dx)/\nu_+.$$

We'll now see that it includes an astonishingly large set of distributions, each with ch.f. of the form  $\exp \left\{ \int (e^{i\omega u} - 1) \nu(du) \right\}$  with ‘‘Lévy measure’’  $\nu(du)$  as given:

	Distribution	Log Ch Function $\phi(\omega)$	Lévy Measure $\nu(du)$
	Poisson $\text{Po}(\lambda)$	$\lambda(e^{i\omega} - 1)$	$\lambda \delta_1(du)$
	Gamma: $\text{Ga}(\alpha, \lambda)$	$-\alpha \log(1 - i\omega/\lambda)$	$\alpha e^{-\lambda u} u^{-1} \mathbf{1}_{\{u>0\}} du$
	Normal: $\text{No}(0, \sigma^2)$	$-\omega^2 \sigma^2/2$	$-\frac{1}{2} \sigma^2 \delta_0''(du)$
	Neg Bin: $\text{NB}(\alpha, p)$	$-\alpha \log[1 - \frac{p}{q}(e^{i\omega} - 1)]$	$\sum_{k=1}^{\infty} \frac{\alpha p^k}{k} \delta_k(du)$
	Cauchy: $\text{Ca}(\gamma, 0)$	$-\gamma \omega $	$\frac{\gamma}{\pi} u^{-2} du$
	Stable: $\text{St}_\alpha(\alpha, \beta, \gamma)$	$-\gamma \omega ^\alpha [1 - i\beta \tan \frac{\pi\alpha}{2} \text{sgn}(\omega)]$	$\gamma c_\alpha [1 + \beta \text{sgn} u] \alpha  u ^{-1-\alpha} du,$

where  $c_\alpha := \frac{1}{\pi} \Gamma(\alpha) \sin \frac{\pi\alpha}{2}$ . Try to verify the measures  $\nu(du)$  for the Negative Binomial and Cauchy distributions. All these distributions share the property called *infinite divisibility* (‘‘ID’’ for short), that for every integer  $n \in \mathbb{N}$  each can be written as a sum of  $n$  independent

identically distributed terms. In 1936 the French probabilist Paul Lévy and Russian probabilist Alexander Ya. Khinchine discovered that every distribution with this property must have a ch.f. of a very slightly more general form than that given above,

$$\log \phi(\omega) = ia\omega - \frac{\sigma^2}{2}\omega^2 + \int_{\mathbb{R}} [e^{i\omega u} - 1 - i\omega h(u)]\nu(du),$$

where  $a \in \mathbb{R}$  and  $\sigma \in \mathbb{R}_+$  are constants and where  $h(u)$  is any bounded Borel function that acts like  $u$  for  $u$  close to zero (for example,  $h(u) = \arctan(u)$  or  $h(u) = \sin(u)$  or  $h(u) = u/(1+u^2)$ ). The measure  $\nu(du)$  need not quite be finite, but we must have  $u^2$  integrable near zero and 1 integrable away from zero... one way to write this is to require that  $\int(1 \wedge u^2)\nu(du) < \infty$ , another is to require  $\int \frac{u^2}{1+u^2}\nu(du) < \infty$ . Some authors consider the finite measure  $\kappa(du) = \frac{u^2}{1+u^2}\nu(du)$  and write

$$\log \phi(\omega) = ia\omega + \int_{\mathbb{R}} [e^{i\omega u} - 1 - i\omega h(u)] \frac{1+u^2}{u^2} \kappa(du),$$

where now the Gaussian component  $-\frac{\sigma^2\omega^2}{2}$  arises from a point mass for  $\kappa(du)$  of size  $\sigma^2$  at  $u = 0$ .

If  $u$  is locally integrable, *i.e.*, if  $\int_{-\epsilon}^{\epsilon} |u| \nu(du) < \infty$  for some (and hence every)  $\epsilon > 0$ , then the term “ $-i\omega h(u)$ ” is unnecessary (it can be absorbed into  $ia\omega$ ). This *always* happens if  $\nu(\mathbb{R}_-) = 0$ , *i.e.*, if  $\nu$  is concentrated on the positive half-line. Every increasing stationary independent-increment stochastic process  $X_t$  (or *subordinator*) has increments which are infinitely divisible with  $\nu$  concentrated on the positive half-line and no Gaussian component ( $\sigma^2 = 0$ ), so has the representation

$$\log \phi(\omega) = ia\omega + \int_0^{\infty} [e^{i\omega u} - 1] \nu(du)$$

for some  $a \geq 0$  and some measure  $\nu$  on  $\mathbb{R}_+$  with  $\int_0^{\infty} (1 \wedge u) \nu(du) < \infty$ . In the compound Poisson example,  $\nu(du) = \sum \nu_j \delta_{u_j}(du)$  was the sum of point masses of size  $\nu_j$  at the possible jump magnitudes  $u_j$ . This interpretation extends to help us understand all ID distributions: every ID random variable  $X$  may be viewed as the sum of a constant, a Gaussian random variable, and a compound Poisson random variable, the sum of independent Poisson jumps of sizes  $u \in E \subset \mathbb{R}$  with rates  $\nu(E)$ .

## 10.2 Stable Limit Laws

Let  $S_n = X_1 + \dots + X_n$  be the partial sum of iid random variables. IF the random variables are all square integrable, THEN the Central Limit Theorem applies and necessarily  $\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \implies \text{No}(0, 1)$ . But what if each  $X_n$  is *not* square integrable? We have already seen that the CLT fails for Cauchy variables  $X_j$ . Denote by  $F(x) = \mathbb{P}[X_n \leq x]$  the common CDF of the  $\{X_n\}$ .

**Theorem 6 (Stable Limit Law)** Let  $S_n = \sum_{j \leq n} X_j$  be the sum of iid random variables. There exist constants  $A_n > 0$  and  $B_n \in \mathbb{R}$  and a non-trivial distribution  $G$  for which the scaled and centered partial sums converge in distribution

$$\frac{S_n - B_n}{A_n} \implies G$$

to a non-trivial limit  $G$  if and only if  $\{X_j\} \subset L_2$  (in which case  $A_n \asymp \sqrt{n}$ ,  $B_n = n\mu + O(\sqrt{n})$ , and  $G = \text{No}(\mu, \sigma^2)$  is the Normal distribution) or there are constants  $0 < \alpha < 2$ ,  $M^- \geq 0$ , and  $M^+ \geq 0$ , with  $M^- + M^+ > 0$ , such that as  $x \rightarrow \infty$  the following limits hold for every  $\xi > 0$ :

$$\begin{aligned} M_+ &= \lim_{x \rightarrow \infty} x^\alpha [1 - F(x)] \quad \text{and, if } M_+ > 0, \quad \frac{1 - F(x\xi)}{1 - F(x)} \rightarrow \xi^{-\alpha} \\ M_- &= \lim_{x \rightarrow \infty} x^\alpha [F(-x)] \quad \text{and, if } M_- > 0, \quad \frac{F(-x\xi)}{F(-x)} \rightarrow \xi^{-\alpha} \end{aligned} \tag{6}$$

In this case the limit is the  $\alpha$ -**Stable Distribution**, with index  $\alpha$ , with characteristic function

$$\mathbb{E}[e^{i\omega Y}] = \exp \left\{ i\delta\omega - \gamma|\omega|^\alpha \left[ 1 - i\beta \tan \frac{\pi\alpha}{2} \text{sgn}(\omega) \right] \right\} \tag{7a}$$

in the  $\text{St}_\alpha$  parametrization, where  $\beta = \frac{M^+ - M^-}{M^- + M^+}$ ,  $\gamma = (M^- + M^+)$ , and  $\delta \in \mathbb{R}$  is arbitrary. The sequences  $A_n, B_n$  must be (see Section (10.5))

$$A_n = (\mathcal{L}_n^A) n^{1/\alpha} \quad B_n = \begin{cases} (\mathcal{L}_n^B) (\beta\gamma \tan \frac{\pi\alpha}{2} - \delta) (n^{1/\alpha} - n) & \alpha \neq 1 \\ (\mathcal{L}_n^B) \frac{2\beta\gamma}{\pi} n \log n & \alpha = 1 \end{cases} \tag{8}$$

for constants or sequences  $\mathcal{L}_n^A, \mathcal{L}_n^B$  that, like  $(\log n)^p$  for any power  $p \in \mathbb{R}$  but not like  $n^p$  for any  $p \neq 0$ , are “slowly varying” in the sense that

$$\lim_{n \rightarrow \infty} \frac{\mathcal{L}_{cn}}{\mathcal{L}_n} = 1$$

for every  $c > 0$ . Don’t get mesmerized by this— think of the  $\mathcal{L}_n$ s as proportionality *constants*. For  $\alpha \in (1, 2)$  the sample means converge at rate  $n^{(1-\alpha)/\alpha}$  to  $\mathbb{E}Y$ , more slowly (*much* more slowly, if  $\alpha$  is close to one) than in the  $L_2$  case where the central limit theorem applies and  $\bar{X}_n \rightarrow \mu$  at rate  $n^{-1/2}$ . No means exist for  $\alpha \leq 1$ . The limits in (6) above are equivalent to the requirement that  $F(x) \propto |x|^{-\alpha} \mathcal{L}_-(x)$  as  $x \rightarrow -\infty$  and  $(1 - F(x)) \propto x^{-\alpha} \mathcal{L}_+(x)$  as  $x \rightarrow +\infty$  for slowly varying functions  $\mathcal{L}_\pm$ — roughly, that  $F(-x)$  and  $1 - F(x)$  are both  $\propto x^{-\alpha}$  (or zero) as  $x \rightarrow \infty$ , *i.e.*, that the pdf falls off like  $|x|^{-\alpha-1}$  in both directions.

The simplest case is the *symmetric*  $\alpha$ -stable (SaS). For  $0 < \alpha \leq 2$  and  $0 < \gamma < \infty$ , the  $\text{St}(\alpha, 0, \gamma, 0)$  has ch.f.

$$\phi(\omega) = e^{-\gamma|\omega|^\alpha}$$

This includes the centered Cauchy  $\text{Ca}(0, s)$  (with  $\alpha = 1$ ,  $\gamma = s$ ) and the centered Normal  $\text{No}(0, \sigma^2)$  (with  $\alpha = 2$ ,  $\gamma = \sigma^2/2$ ). The SaS family interpolates between these (for  $1 < \alpha < 2$ ) and extends them (for  $0 < \alpha < 1$ ) to distributions with even heavier tails.

Although each Stable distribution has an absolutely continuous distribution with continuous unimodal probability density function  $f(y)$ , these two cases and the “inverse Gaussian” or “Lévy” distribution with  $\alpha = 1/2$  and  $\beta = \pm 1$  are the only ones whose pdf is available in closed form. Perhaps that’s the reason these are less studied than normal distributions; still, they are very useful for problems with “heavy tails”, *i.e.*, where  $\text{P}[X > u]$  does not die off quickly with increasing  $u$ . The symmetric (SaS) ones all have bell-shaped pdfs.

Moments are easy enough to compute but, for  $\alpha < 2$ , moments  $\text{E}|X|^p$  are only finite for  $p < \alpha$ . In particular, means only exist for  $\alpha > 1$  and *none* of them has a finite variance. The Cauchy has finite moments of order  $p < 1$ , but (despite its symmetry) does not have a well-defined mean.

Equation (6) says that each tail must be fall off like a power (sometimes called *Pareto tails*), and the powers must be identical; it also gives the tail ratio. A common special case is  $M^- = 0$  (or equivalently  $\beta = 1$ ), the “one-sided” or “fully skewed” Stable. For  $0 < \alpha < 1$  these take only values in  $[\delta, \infty)$  ( $\mathbb{R}_+$  if  $\delta = 0$ ). For example, random variables  $X_n$  with the Pareto distribution (often used to model income) given by  $\text{P}[X_n > t] = (1 + t/k)^{-\alpha}$  for  $t \in \mathbb{R}_+$  and some  $\alpha > 0$ ,  $k > 0$  will have a stable limit for their partial sums if  $\alpha < 2$ , and (by CLT) a normal limit if  $\alpha \geq 2$ . There are close connections between the theory of Stable random variables and the more general theory of statistical extremes. Ask me for references if you’d like to learn more about this exciting area.

Expression (7a) for the  $\alpha$ -stable ch.f. (called the “A-parametrization”  $\text{St}_A(\alpha, \beta, \gamma, \delta)$  by [Zolotarev \(1986\)](#)) behaves badly as  $\alpha \rightarrow 1$  if  $\beta \neq 0$ , because the tangent function has a pole at  $\pi/2$ — so, for  $\beta > 0$ , the imaginary part of  $\log \text{E}[e^{i\omega Y}]$  converges to  $+\infty$  as  $\alpha \nearrow 1$  and to  $-\infty$  as  $\alpha \searrow 1$  (the signs reverse for  $\beta < 0$ ). For  $\alpha \approx 1$  the complex part of the log ch.f. is:

$$\begin{aligned} \Im \{ \log \text{E}[e^{i\omega Y}] \} &= i\delta\omega + i\beta\gamma \tan \frac{\pi\alpha}{2} |\omega|^\alpha \text{sgn}(\omega) \\ &= i\delta\omega + i\beta\gamma \tan \frac{\pi\alpha}{2} |\omega|^{\alpha-1} \omega \\ &= i\omega \left[ \delta + \beta\gamma \tan \frac{\pi\alpha}{2} \right] - i\beta\gamma \tan \frac{\pi\alpha}{2} \omega (1 - |\omega|^{\alpha-1}) \end{aligned}$$

where the last term is bounded as  $\alpha \rightarrow 1$ , so (following V. M. [Zolotarev, 1986](#)) the  $\alpha$ -stable is often parametrized differently in the “M-parametrization”  $\text{St}_M(\alpha, \beta, \gamma, \delta)$  as

$$\text{E}[e^{i\omega Y}] = \exp \left\{ -\gamma|\omega|^\alpha + i\delta^*\omega - i\beta\gamma \tan \frac{\pi\alpha}{2} \omega (1 - |\omega|^{\alpha-1}) \right\} \quad (9)$$

for  $\alpha \neq 1$  with shifted “drift” term  $\delta^* = \delta + \beta\gamma \tan(\pi\alpha/2)$ . You can find out more details by asking me or reading [Breiman \(1968, Chapter 9\)](#).

### 10.3 Examples

Let  $\{U_n\} \stackrel{\text{iid}}{\sim} \text{Un}(0, 1)$ , and set  $X_n := 1/U_n$ . How would you describe the distribution of  $S_n := \sum_{j \leq n} X_j$ , or of  $\bar{X}_n := \frac{1}{n} S_n$ ? Evidently  $\mathbf{E}X_n = \int_0^1 \frac{1}{u} du = \infty$ , so  $\frac{1}{n} S_n \rightarrow \infty$  *a.s.*; to get any interesting limit we will have to rescale or recenter or both.

By Theorem 6, there exist sequences  $\{A_n\}, \{B_n\}$  such that  $(S_n - B_n)/A_n \Rightarrow G$  for a non-trivial distribution  $G$  if and only if  $X_n \in L_2$  (which fails in this case) or if the CDF

$$F(z) = \mathbf{P}[X_n \leq z] = \mathbf{P}[U_n \geq 1/z] = \begin{cases} 1 - 1/z & z > 1 \\ 0 & z \leq 1 \end{cases}$$

satisfies for some  $0 < \alpha < 2$  the conditions  $M_+ = \lim_{x \rightarrow \infty} x^\alpha [1 - F(x)] \geq 0$ ,  $M_- = \lim_{x \rightarrow \infty} x^\alpha [F(-x)] \geq 0$ ,  $(M_- + M_+) > 0$  and, if  $M_+ > 0$ ,

$$\frac{1 - F(x\xi)}{1 - F(x)} = \frac{1/x\xi}{1/x} \rightarrow \xi^{-\alpha}.$$

Evidently the conditions hold for  $\alpha = 1$ ,  $M_- = 0$ , and  $M_+ = 1$ , so  $\beta = M_+/(M_- + M_+) = 1$  and  $\gamma = (M_- + M_+) = 1$ . By (8), with  $\alpha = \beta = \gamma = 1$ , suitable sequences are given by  $A_n = n$  and  $B_n = \frac{2}{\pi} n \log n$ , so

$$\frac{S_n - \frac{2}{\pi} n \log n}{n} = \frac{S_n}{n} - \frac{2}{\pi} \log n \approx \text{St}_m(\alpha = 1, \beta = 1, \gamma = 1, \delta = 0),$$

or

$$\frac{S_n}{n} \approx \text{St}_m(\alpha = 1, \beta = 1, \gamma = 1, \delta = \frac{2}{\pi} \log n).$$

Sample averages  $\bar{X}_n$  of  $X_j := (1/U_j)$  will grow to infinity at approximate rate  $(2/\pi) \log n$ , and  $\bar{X}_n - (2/\pi) \log n$  will have an asymptotic  $\text{St}_m(1, 1, 1, 0)$  distribution. What do you think happens to  $\bar{X}_n$  for  $X_n := U_n^{-1/\alpha}$  for other values of  $\alpha > 0$ ?

### 10.4 Key Idea of the Stable Limit Laws

The stable limit law of Theorem 6 says that if there exist nonrandom sequences  $A_n > 0$  and  $B_n \in \mathbb{R}$  and a nondegenerate distribution  $G$  such that the partial sums  $S_n := \sum_{j \leq n} X_j$  of iid random variables  $\{X_j\}$  satisfy

$$\frac{S_n - B_n}{A_n} \Longrightarrow G \tag{10}$$

then  $G$  must be either the normal distribution or an  $\alpha$ -stable distribution for some  $0 < \alpha < 2$ . The key idea behind the theorem is that *if* a distribution  $\mu$  with cdf  $G$  satisfies (10) *then* also for any  $n$  the distribution of the sum  $S_n$  of  $n$  independent random variables with cdf  $G$

must *also* (after suitable shift and scale changes) have cdf  $G$ — *i.e.*, that  $c_n S_n + d_n \sim G$  for some constants  $c_n > 0$  and  $d_n \in \mathbb{R}$ , so the characteristic function  $\chi(\omega) := \int e^{i\omega x} G(dx)$  and log ch.f.  $\psi(\omega) := \log \chi(\omega)$  must satisfy

$$\begin{aligned}\chi(\omega) &= \mathbf{E} \exp \{i\omega(c_n S_n + d_n)\} \\ &= \exp(i\omega d_n) \chi(\omega c_n)^n \\ \psi(\omega) &= i\omega d_n + n\psi(c_n \omega)\end{aligned}\tag{11}$$

whose only solutions are the normal and  $\alpha$ -stable distributions. Here's a sketch of the proof for the symmetric (SaS) case, where  $\psi(-\omega) = \psi(\omega)$  and so  $d_n = 0$ . Set  $\gamma := -\psi(1)$  and note that (11) with  $\omega = c_n^k$  for  $k = 0, 1, \dots$  implies successively:

$$\psi(c_n) = \frac{-\gamma}{n} \quad \psi(c_n^2) = \psi(c_n) \frac{1}{n} = \frac{-\gamma}{n^2} \quad \dots \quad \psi(c_n^k) = \frac{-\gamma}{n^k}.$$

Results from complex analysis imply this must hold for *all*  $k \geq 0$ , not just integers. Thus, with  $|w| = c_n^k$  and  $k = \log |w| / \log c_n$ ,

$$\begin{aligned}\psi(w) &= -\gamma n^{-k} \\ &= -\gamma \exp \{-(\log |w|)(\log n) / (\log c_n)\} \\ &= -\gamma |w|^{-(\log n) / (\log c_n)} \\ &= -\gamma |w|^\alpha,\end{aligned}$$

where  $\alpha$  is the constant value of  $\frac{-\log n}{\log c_n}$ . It follows that  $c_n = n^{-1/\alpha}$  (*i.e.*,  $S_n/n^{1/\alpha} \sim G$ ) and that  $\chi(\omega) = e^{-\gamma|\omega|^\alpha}$ , the ch.f. for SaS for  $0 < \alpha < 2$  and for No(0, 2 $\gamma$ ) for  $\alpha = 2$ .

## 10.5 The constants $A_n$ and $B_n$

Here we show where the constants given in (8) come from. Let  $\phi(\omega)$  be the ch.f. of iid random variables  $Y_i \sim \text{St}_M(\alpha, \beta, \gamma, \delta)$  (given in (9)), and let  $\psi(\omega) := \log \phi(\omega)$  be its logarithm. For the partial sums  $S_n := \sum_{j=1}^n Y_j$  to have the property that  $(S_n - B_n)/A_n \sim \text{St}_M(\alpha, \beta, \gamma, \delta)$  has the same distribution as  $Y_j$  for each  $n$ , we need

$$\begin{aligned}\psi(\omega) &= -\gamma|\omega|^\alpha + i\delta^* \omega - i\beta\gamma \tan \frac{\pi\alpha}{2} \omega (1 - |\omega|^{\alpha-1}) \\ &= n\psi(\omega/A_n) - i\omega B_n/A_n \\ &= -\left[\frac{n}{A_n^\alpha}\right] \gamma|\omega|^\alpha + i\omega \frac{n}{A_n} \left\{ \delta^* - \beta\gamma \tan \frac{\pi\alpha}{2} - B_n/n \right\} \\ &\quad + \left[\frac{n}{A_n^\alpha}\right] i\beta\gamma \tan \frac{\pi\alpha}{2} \omega |\omega|^{\alpha-1}.\end{aligned}\tag{12}$$

From the real part of (12) we find  $A_n = n^{1/\alpha}$ , and from the imaginary part we find  $B_n = (\beta\gamma \tan \frac{\pi\alpha}{2} - \delta)(n^{1/\alpha} - n)$  for  $\alpha \neq 1$ . Find the value of  $B_n$  for  $\alpha = 1$  as a limit.



## References

- Barbour, A. D. and Chen, L. H. Y., eds. (2005), *An Introduction to Stein's Method, Lecture Note Series, Institute for Mathematical Sciences*, volume 4, Singapore: National University of Singapore.
- Billingsley, P. J. (1995), *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics, New York, NY: John Wiley & Sons, third edition.
- Breiman, L. (1968), *Probability*, Reading, MA: Addison-Wesley.
- Gnedenko, B. V. and Kolmogorov, A. N. (1968), *Limit Distributions for Sums of Independent Random Variables (revised)*, Reading, MA: Addison-Wesley.
- Grimmett, G. R. and Stirzaker, D. R. (2001), *Probability and Random Processes*, Oxford, UK: Oxford University Press, third edition.
- Khinchine, A. Y. and Lévy, P. (1936), "Sur les lois stables," *Comptes rendus hebdomadaires des seances de l'Académie des sciences. Académie des science (France). Serie A. Paris*, 202, 374–376.
- Resnick, S. I. (1999), *A Probability Path*, New York, NY: Birkhäuser.
- Stein, C. (1972), "A bound for the error in the normal approximation to the distribution of a sum of dependent random variables," in *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*, eds. L. M. Le Cam, J. Neyman, and E. L. Scott, Berkeley, CA: University of California Press, volume 2, pp. 583–602.
- Stein, C. (1986), *Approximate computation of expectations, IMS Lecture Notes-Monograph Series*, volume 7, Hayward, CA: Institute of Mathematical Statistics.
- Zolotarev, V. M. (1986), *One-dimensional Stable Distributions, Translations of Mathematical Monographs*, volume 65, Providence, RI: American Mathematical Society.