9 Sums of Independent Random Variables

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

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

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 μ 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

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

often called the "moment generating function" and denoted $M_X(z) = \mathsf{E}e^{zX}$ 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 $M(z) = \mathsf{E}e^{zX}$ may be infinite. Here are a few examples of $\mathsf{E}e^{zX}$ for some familiar distributions:

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

Aside from the problem that $M(z) = \mathsf{E}e^{zX}$ may 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: if we can justify interchanging the limits implicit in differentiation and integration, then $M'(z) = \mathsf{E}[Xe^{zX}]$ and $M''(z) = \mathsf{E}[X^2e^{zX}]$, so (upon taking $z = 0$) $M'(0) = \mathsf{E}X = \mu$ and $M''(0) = \mathsf{E}X^2 = \sigma^2 + \mu^2$, so we can calculate the mean and variance (and other moments $\mathsf{E}X^k = M^{(k)}(0)$) from derivatives of $M(z)$ at zero. We have two problems to overcome: discovering how to infer the distribution of X from $M_X(z) = \mathsf{E}e^{zX}$, and what to do about distributions for which $M(z)$ doesn't exist.

9.1 Characteristic Functions

For complex numbers $z = x + iy$ 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 ω 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*

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

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.

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

9.1.1 Uniqueness

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

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

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 μ 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 μ 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 $|\phi(\omega)| \le 1$ 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 $\phi : \mathbb{R} \to \mathbb{C}$ is* ***positive definite*** *if for every $n \in \mathbb{N}$, $z \in \mathbb{C}^n$, and $\omega \in \mathbb{R}^n$,*

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

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 μ on

$(\mathbb{R},\mathcal{B})$:

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

Interestingly, this condition is also sufficient:

Theorem 1 (Bochner) *If $\phi:\mathbb{R}\to\mathbb{C}$ is continuous at zero, satisfies $\phi(0)=1$, and is positive definite, then there exists a Borel probability measure μ 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 $\{z_j\}\subset\mathbb{C}$,

$$0\le\sum z_j\phi(\omega_j-\omega_k)\bar z_k$$

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

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

Taking $\omega_j:=(j-n^2)/n$ for $0\le j\le 2n^2$ and then taking the limit as $n\to\infty$,

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

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 $\epsilon\to 0$, we find that $f(x):=\frac{1}{2\pi}\int_{\mathbb{R}} e^{-ix\omega}\phi(\omega)\,d\omega\ge 0$ 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 $|\phi(\omega)| \to 0$ as $\omega \to \pm\infty$. For integer-valued random variables X we can recover $p_k = \mathsf{P}[X = k]$ by inverting the Fourier series:

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

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

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

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 ϵ).

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 $|\phi(\omega)| \le 1$ 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. ω over $\mathbb{R}$. We can compute

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

(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 $\epsilon \to 0$ 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 $\mathsf{P}[X=y]>0$.

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

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

There are several interesting connections between the density function $f(x)$ and characteristic function $\phi(\omega)$. If $\phi(\omega)$ "wiggles" with rate approximately ξ, *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 ξ; 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 $\mathsf{E}[|X|^p]<\infty$ for large p. If $|\phi(\omega)|$ falls off quickly as $\omega\to\pm\infty$ then $\phi(\omega)$ doesn't have large *low*-frequency components and $f(x)$ must be rather tame, without any spikes. Thus ϕ 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 $\mu_\epsilon:=\mu\star\gamma_\epsilon$ converges* a.s *as $\epsilon\to 0$ to an L_1 function $f(x)$, $\hat\mu_\epsilon(\omega)$ converges uniformly to $\hat f(\omega)$, 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 μ and real numbers $a<b$,*

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

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

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

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*

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

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

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

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

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 $z = 0$: $\phi(0) = 1$, $\phi'(0) = i\mathsf{E}[X]$, $\phi''(0) = -\mathsf{E}[X^2]$, so

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

for some $c < \infty$, so by Taylor's theorem we have:[1]

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

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 μ on a measurable space $(\mathcal{X}, \mathcal{E})$ (written $\mu_n \Rightarrow \mu$):

$$\big(\forall\phi \in \mathcal{C}_b(\mathcal{X})\big)\ \lim_{n\to\infty}\int_{\mathcal{X}} \phi(x)\,\mu_n(dx) = \int_{\mathcal{X}} \phi(x)\,\mu(dx) \qquad (4)$$

In fact requiring this convergence for *all* bounded continuous functions ϕ 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 $\phi_\omega(x) = \exp(i\omega\cdot x)$ 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*

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

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

How would you prove this?

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

Examples

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

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

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

Po: Let X_n have Binomial $\mathsf{Bi}(n,p_n)$ distributions with success probabilities $p_n \asymp 1/n$, so that $n\,p_n \to \lambda$ for some $\lambda > 0$ as $n \to \infty$. Then the ch.f.s satisfy

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

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

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

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

the ch.f. of the $\mathsf{No}(0,1)$ 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 μ and variance σ^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 α, β, γ,

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

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

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

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

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

with logarithm

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

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

Note: We assumed X_i were iid with finite third moment $\gamma := \mathsf{E}|X_i|^3 < \infty$. Under those conditions one can prove the uniform "Berry-Esséen" bound $\sup_x |F_n(x) - \Phi(x)| \le \gamma/\big[2\sigma^3\sqrt{n}\big]$ for the CDF F_n of Z_n. Another version of the CLT for iid $\{X_i\}$ asserts weak convergence of Z_n to $\mathsf{No}(0,1)$ assuming only $\mathsf{E}[X_i^2] < \infty$ (*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

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

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

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

as $n \to \infty$ for each $t > 0$. This follows immediately for iid $\{X_j\} \subset L_2$ (where it becomes $\frac{1}{\sigma^2}\mathsf{E}\left[|X_1|^2\,\mathbf{1}_{\{|X_1|^2 > nt^2\sigma^2\}}\right]$ which tends to zero as $n \to \infty$ 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

$$\max_{j \le n} \frac{\sigma_j^2}{s_n^2} \to 0 \qquad\qquad \max_{j \le n} \mathsf{P}\left\{|X_j|/s_n > \epsilon\right\} \to 0$$

as $n \to \infty$, 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,

$$s_n^{-2-\delta} \sum_{j=1}^{n} \mathsf{E}|X_j|^{2+\delta} \to 0$$

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 $\mathsf{E}[X^2] = \infty$; 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 have approximate normal distributions, and may not even be concentrated around a central value.

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

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

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

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

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

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, α-Stable, and Pareto arise when modeling income distribution, weather extremes, volcanic flows, and many other phenomena.

10 Distributionl 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 ν_j and let $u_j \in \mathbb{R}$; then the ch.f. for $Y := \sum u_j X_j$ is

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

for the discrete measure $\nu(du) = \sum \nu_j \delta_{u_j}(du)$ that assigns mass ν_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

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

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

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

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,

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

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

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

where now the Gaussian component $\frac{-\sigma^2\omega^2}{2}$ arises from a point mass for $\kappa(du)$ of size σ^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 ν 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 ν concentrated on the positive half-line and no Gaussian component ($\sigma^2 = 0$), so has the representation

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

for some $a \geq 0$ and some measure ν 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 ν_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 $S_n = X_1 + \ldots + X_n$ be the partial sum of iid random variables. IF the random variables are all square integrable, THEN the Central Limit Theorem applies and necessarily $\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \Longrightarrow \mathsf{No}(0,1)$. 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 $F(x) = \mathsf{P}[X_n \leq x]$ the common CDF of the $\{X_n\}$.

Theorem 6 (Stable Limit Law) *Let $S_n = \sum_{j\le n} X_j$ 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*

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

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 $G = \mathsf{No}(\mu, \sigma^2)$ is the Normal distribution) or there are constants $0 < \alpha < 2$, $M^- \ge 0$, and $M^+ \ge 0$, with $M^- + M^+ > 0$, such that as $x \to \infty$ the following limits hold for every $\xi > 0$:

1. $\dfrac{F(-x)}{1-F(x)} \to \dfrac{M^-}{M^+}$;
2. $M^+ > 0 \Rightarrow \dfrac{1-F(x\xi)}{1-F(x)} \to \xi^{-\alpha}$ $\qquad$ $M^- > 0 \Rightarrow \dfrac{F(-x\xi)}{F(-x)} \to \xi^{-\alpha}$.

In this case the limit is the α-**Stable Distribution**, with index α, with characteristic function

$$\mathsf{E}\big[e^{i\omega Y}\big] = e^{i\delta\omega - \gamma|\omega|^\alpha\left[1 - i\beta\tan\frac{\pi\alpha}{2}\,\mathrm{sgn}(\omega)\right]}, \tag{6}$$

where $\beta = \frac{M^+ - M^-}{M^- + M^+}$ and $\gamma = (M^- + M^+)$. The sequence A_n must be essentially $A_n \propto n^{1/\alpha}$ (more precisely, the sequence $C_n := n^{-1/\alpha} A_n$ is *slowly varying* in the sense that

$$1 = \lim_{n\to\infty} \frac{C_{cn}}{C_n}$$

for every $c > 0$). For $\alpha > 1$ the sample means converge at rate $n^{(1-\alpha)/\alpha}$, more slowly (*much* more slowly, if α is close to one) than in the L_2 case where the central limit theorem applies. The limits in “2.” above are equivalent to the requirement that $F(x) = |x|^{-\alpha}\mathcal{L}_-(x)$ as $x \to -\infty$ and $F(x) = 1 - x^{-\alpha}\mathcal{L}_+(x)$ as $x \to +\infty$ for slowly varying functions $\mathcal{L}_\pm$— roughly, that $F(-x)$ and $1 - F(x)$ are both $\propto x^{-\alpha}$ (or zero) as $x \to \infty$.

The simplest case is the *symmetric* α-stable (SαS). For $0 < \alpha < 2$ and $0 < \gamma < \infty$, the $\mathsf{St}(\alpha, 0, \gamma, 0)$ has ch.f. of the form

$$e^{-\gamma|\omega|^\alpha}$$

This includes the centered Cauchy ($\alpha = 1$, $\gamma = s$) and the centered Normal ($\alpha = 2$, $\gamma = \sigma^2/2$). The SαS 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 where the 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 $\mathsf{P}[X > u]$ does not die off quickly with increasing u. The symmetric (SαS) ones all have bell-shaped pdfs.

Moments are easy enough to compute but, for $\alpha < 2$, moments $\mathsf{E}|X|^p$ 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 does not have a well-defined mean.

Condition 2. says that each tail must be fall off like a power (sometimes called *Pareto tails*), and the powers must be identical; Condition 1. gives the tail ratio. A common special case is $M^- = 0$, 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 $\mathsf{P}[X_n > t] = (k/t)^\alpha$ for $t \ge k$ will have a stable limit for their partial sums if $\alpha < 2$, and (by CLT) a normal limit if $\alpha \ge 2$. 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 (6) for the α-stable ch.f. behaves badly as $\alpha \to 1$ if $\beta \ne 0$, because the tangent function has a pole at $\pi/2$. For $\alpha \approx 1$ the complex part of the log ch.f. is:

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

where the last term is bounded as $\alpha \to 1$, so (following V. M. Zolotarev, 1986) the α-stable is often parametrized for $\alpha \ne 1$ as

$$\log \mathsf{E}[e^{i\omega Y}] = -\gamma|\omega|^\alpha + i\delta^*\omega - i\beta\gamma\tan\tfrac{\pi\alpha}{2}\,\omega\left(1 - |\omega|^{\alpha-1}\right)$$

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 Key Idea of the Stable Limit Laws

The stable limit law 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 sum $S_n := \sum_{j\le n} X_j$ of iid random variables $\{X_j\}$ satisfies

$$\frac{S_n - B_n}{A_n} \Longrightarrow G \tag{7}$$

then G must be either the normal distribution or an α-stable distribution for some $0 < \alpha < 2$. The key idea behind the theorem is that *if* a distribution μ with cdf G satisfies (7) *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

$$\begin{aligned}\chi(\omega) &= \mathsf{E} \exp\{i\omega(c_n S_n + d_n)\}\\ &= \exp(i\omega d_n)\ \chi(\omega c_n)^n\\ \psi(\omega) &= i\omega d_n + n\psi(c_n\omega) \qquad (8)\end{aligned}$$

whose only solutions are the normal and α-stable distributions. Here's a sketch of the proof for the symmetric (SαS) case, where $\psi(-\omega) = \psi(\omega)$ and so $d_n = 0$. Set $\gamma := -\psi(1)$ and note that (8) with $\omega = c_n^k$ for $k = 0, 1, \ldots$ implies successively:

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

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 α 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 S$\alpha$S for $0 < \alpha < 2$ and for $\mathsf{No}(0, 2\gamma)$ for $\alpha = 2$.

References

Barbour, A. D. and Chen, L. H. Y. (2005), *An Introduction to Stein's Method*, *Lecture Note Series, Institute for Mathematical Sciences*, volume 4, New Jersey, USA: National University of Singapore.

Billingsley, P. (1995), *Probability and Measure*, New York, NY: Wiley, 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.