Fourier Transforms of Measures

Sums of Independent Random Variables

We begin 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 $\mathsf{V}S_n = \sigma^2_{\le n} = \sum_{i\le n}\sigma_i^2$. 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

$$\begin{aligned}
\mathsf{P}[S_2 \le s] = F(s) &= \iint_{x_1+x_2\le s} f_1(x_1)f_2(x_2)\,dx_1dx_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{s-x_2} f_1(x_1)f_2(x_2)\,dx_1dx_2 \\
f(s) = F'(s) &= \int_{-\infty}^{\infty} f_1(s-x_2)f_2(x_2)\,dx_2 \\
&= \int_{-\infty}^{\infty} f_1(x_1)f_2(s-x_1)\,dx_1,
\end{aligned}$$

the *convolution* of $f_1(x_1)$ and $f_2(x_2)$. Even if the distributions aren't absolutely continuous, so no pdf's 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 CDF's and pdf's of sums of independent RV's 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 RV's 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:	$\mathsf{Ga}(\alpha,\beta)$	$(1 - z/\beta)^{-\alpha}$	$\Re(z) < \beta$
Cauchy:	$\dfrac{a}{\pi(a^2 + (x-b)^2)}$	$e^{zb-a\lvert z\rvert}$	$\Re(z) = 0$

Aside from the problem that $M(z) = \mathsf{E}e^z X$ 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 $M''(z) = \mathsf{E}[X^2 e^{zX}]$, 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.

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*

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

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:	$\mathsf{Bi}(n, p)$	$[1 + p(e^{i\omega} - 1)]^N$
Neg Bin:	$\mathsf{NB}(\alpha, p)$	$[1 - (p/q)(e^{i\omega} - 1)]^{-\alpha}$
Poisson	$\mathsf{Po}(\lambda)$	$e^{\lambda(e^{i\omega} - 1)}$
Normal:	$\mathsf{No}(\mu, \sigma^2)$	$e^{i\omega\mu - \omega^2\sigma^2/2}$
Gamma:	$\mathsf{Ga}(\alpha, \beta)$	$(1 - i\omega/\beta)^{-\alpha}$
Cauchy:	$\dfrac{a/\pi}{a^2 + (x - b)^2}$	$e^{i\omega b - a\lvert\omega\rvert}$

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

Grimmett and Stirzaker mention related results in Chapter 5, and Billingsley proves several versions of this theorem in 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.

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), while the continuous ones satisfy $|\phi(\omega)| \to 0$ as $\omega \to \pm\infty$. For integer-valued random variables X we can recover $p_k = \Pr[X = k]$ by inverting the Fourier series:

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

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 $\gamma_\epsilon(x) \equiv \frac{1}{\sqrt{2\pi\epsilon}} e^{-x^2/2\epsilon}$ satisfies $\int \gamma_\epsilon(x) \, dx = 1$ (it's just the normal probability density 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 integrable and 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 $f(y)$. This converges

- uniformly to $f(y)$ as $\epsilon \to 0$ if $f(\cdot)$ is bounded and continuous (the most common case),
- pointwise to $\frac{f(y-)+f(y+)}{2}$ if $f(x)$ has a jump discontinuity at $x = y$, and
- to infinity if $\mu(\{y\}) > 0$, *i.e.*, if $\Pr[X = y] > 0$.

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 1. *If $\int_{\mathbb{R}} |\hat\mu(\omega)|\,d\omega < \infty$ then $\mu_\epsilon \equiv \mu \star \gamma_\epsilon$ converges a.s. to an L^1 function $f(x)$, $\hat\mu_\epsilon(\omega)$ converges uniformly to $\hat f(\omega)$, and $\mu(A) = \int_A f(x)\,dx$. Also $f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-i\omega x}\hat\mu(\omega)\,d\omega$ for almost-every x.*

Theorem 2. *For any μ and any $a < b$, $\mu\big((a,b)\big) + {}^1\!/\!_2\,\mu\big(\{a,b\}\big) = \lim_{T\to\infty}\int_{-T}^{T} \frac{e^{-i\omega a} - e^{-i\omega b}}{2\pi i\omega}\hat\mu(\omega)\,d\omega$.*

Theorem 3. *If $\int_{\mathbb{R}} |x|^k\,\mu(dx) < \infty$ for an integer $k > 0$ then $\hat\mu(\omega)$ has continuous derivatives of order k given by*

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

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

By Theorem 3 the first few moments of the distribution, if they exist, can be determined from derivatives of the characteristic function *or its logarithm* at zero: $\phi(0) = 1$, $\phi'(0) = i\mathsf{E}[X]$, $\phi''(0) = -\mathsf{E}[X^2]$, so

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

so by Taylor's theorem, we have

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

Limits of Partial Sums

We'll need to re-center and re-scale the distribution of $S_n = \sum_{i=1}^n X_i$ 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 RV's: 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 *i.i.d.* L^2 random variables X_i with characteristic function $\phi(t)$, the normalized sum $[S_n - n\mu]/\sqrt{n\sigma^2}$ has characteristic function

$$\begin{aligned} \phi_S(\omega) &= \prod_{j=1}^{n} \left[\phi(\omega/\sqrt{n\sigma^2})e^{-i\omega\mu/\sqrt{n\sigma^2}}\right] \\ &= \left[\phi(s)e^{-is\mu}\right]^n, \text{ where } s = \omega/\sqrt{n\sigma^2} \\ &= e^{n[\log\phi(s) - is\mu]} \end{aligned}$$

with logarithm

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

so $\phi_S(\omega) \to e^{-\omega^2/2}$ for all $\omega \in \mathbb{R}$ and hence $[S_n - n\mu]/\sqrt{n\sigma^2} \Rightarrow N(0,1)$, the Central Limit Theorem.

Note: We assumed X_i were *i.i.d.* with finite third moment; the method of proof really only requires $\mathsf{E}[X^2] < \infty$, and can be extended to the non-identically-distributed case (and even independence can be weakened), but S_n *cannot* converge to a normally-distributed limit if $\mathsf{E}[X^2] = \infty$; ask for details (or read *Glivenko & Kolmogorov*) if you're interested.

Generalized 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 $Y \equiv \sum u_j X_j$ is

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

for the discrete measure $\nu(du)$ 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 *Generalized Poisson*; we'll now see that it includes an astonishingly large set of distributions.

Distribution			**Log Characteristic Function**
Poisson	$\mathsf{Po}(\lambda)$	$\lambda(e^{i\omega} - 1)$	$= \int_0^\infty (e^{i\omega u} - 1)\, \lambda\delta_1(du)$
Gamma:	$\mathsf{Ga}(\alpha, \beta)$	$-\alpha \log(1 - i\omega/\beta)$	$= \int_0^\infty (e^{i\omega u} - 1)\, \alpha e^{-u\beta} u^{-1}(du)$
Normal:	$\mathsf{No}(0, \sigma^2)$	$-\omega^2\sigma^2/2$	$= \lim_{\epsilon\to 0} \int (e^{i\omega u} - 1)\, \frac{\sigma^2}{\epsilon^2}\delta_\epsilon(du) - \frac{i\omega\sigma^2}{\epsilon}$
Neg Bin:	$\mathsf{NB}(\alpha, p)$	$-\alpha \log[1 - \frac{p}{q}(e^{i\omega} - 1)]$	$= \int (e^{i\omega u} - 1)\nu(du), \quad \nu(du) = \sum_{k=1}^\infty \frac{\alpha p^k}{k}\delta_k(du)$
Cauchy:	$\mathsf{Ca}(a, 0)$	$-a\|\omega\|$	$= \int (e^{i\omega u} - 1)\nu(du), \quad \nu(du) = a\, u^{-2}\, du$
Stable:	$\mathsf{St}(\alpha, \beta)$	$-\|\omega\|^\alpha[1 - i\beta\tan(\frac{\pi\alpha}{2})\,\mathrm{sgn}(\omega)]$	$= \int (e^{i\omega u} - 1)\nu(du), \quad \nu(du) = \frac{\alpha\Gamma(\alpha)\sin(\frac{\pi\alpha}{2})[1 + \beta\,\mathrm{sgn}(u)]}{\pi\, u^{1+\alpha}}$

Try to verify the measures $\nu(du)$ for the Negative Binomial and Cauchy distributions. All these distributions share the property called *infinite divisibility*, that each can be written as a sum of

n independent identically distributed pieces for every number n; independently the French probabilist Paul Lévy and Russian probabilist A. 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,

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

where $h(u)$ is any bounded continuous 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

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

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

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

for some $a \geq 0$ and some measure ν on $\mathbb{R}_+$ satisfying $\int_0^\infty (1 \wedge u)\, \nu(du) < \infty$. In the generalized 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 generalized Poisson random variable, the sum of independent Poisson jumps of sizes $u \in E \subset \mathbb{R}$ with rates $\nu(E)$.