Characteristic functions

1/13/12

Literature

Rick Durrett, *Probability: Theory and Examples*, Duxbury 2010.

Olav Kallenberg, *Foundations of Modern Probability*, Springer 2001,

Eugene Lukacs, *Characteristic Functions*, Hafner 1970.

Robert G. Bartle, *The Elements of Real Analysis*, Wiley 1976.

Michel Loève, *Probability Theory*, Van Nostrand 1963.

Contents

1 Definition and basic properties

Let μ be a probability distribution of a random variable X. The **characteristic function**, a.k.a. **Fourier transform**, is the complex valued one-parameter function

$$\hat{\mu}(t) = \varphi(t) = \int_{\mathbb{R}} e^{\imath tx}\,\mu(dx) = \mathsf{E}\,e^{itX}.$$

Similarly we define the ch.f. of a probability distribution $\mu = \mathcal{L}(X)$ on $\mathbb{R}^d$ or in a Hilbert space where $tx = \left\langle t, x \right\rangle$ is the inner product. The definition applies also to finite measures, even to signed measures of bounded variation. The term "characteristic function" is restricted to probability measures.

Proposition 1.1 *Every ch.f.* $\varphi(t) = \widehat{\mu}(t) = \mathsf{E}\,e^{itX}$ *has the properties:*

1. $\varphi(0) = 1$;
2. $|\varphi| \le 1$;
3. φ *is uniformly continuous on* $\mathbb{R}$.
4. φ *is semi-positive definite, i.e.,*

$$\sum_j \sum_k \varphi(t_j - t_k)\, z_j\, \overline{z_k} \ge 0, \quad \textit{for every finite sets} \{\, t_j \,\} \subset \mathbb{R}, \quad \{\, z_j \,\} \subset \mathbb{C}$$

Proof.
(3): $|\varphi(s) - \varphi(t)| \le E\left|e^{isX} - e^{itX}\right| \le E\left|1 - e^{i(s-t)X}\right| \le E\,1 \wedge |s-t|X$.

(4): $0 \le \mathsf{E}\left|\sum_j z_j \mathsf{E}\,e^{it_jX}\right|^2 = \sum_j \sum_k \varphi(t_j - t_k) z_j \overline{z_k}$. ∎

A probabilist should know ch.fs. of basic probability distributions by heart and how they behave under simple transformations. To wit:

Proposition 1.2 *1.* $\varphi_{aX}(t) = \varphi_X(at)$, *hence* $\varphi_{-X} = \overline{\varphi}$.

2. *A convex combination of ch.fs. is a ch.f.*

3. *Hence, given a ch.f.* φ, $\Re\varphi = (\varphi + \overline{\varphi})/2$ *is a ch.f.*

4. *The finite product of ch.fs. is a ch.f. Namely,*

$$\varphi_{X_1} \cdots \varphi_{X_n} = \varphi_{X_1' + \cdots + X_n'},$$

where X_k*'s are independent copies of* X_k*. In other words,*

$$\hat{\mu_1} \cdots \hat{\mu_n} = (\mu_1 \otimes \cdots \otimes \mu_n)\hat{}.$$

5. *Hence, given a ch.f.* φ, $|\varphi|$ *and the natural powers* φ^n *and* $|\varphi|^n$ *are ch.fs.*

6. *a ch.f. is real if and only if*[1] X *is symmetric, i.e.* $X \stackrel{\mathcal{D}}{=} -X$.

■

Notice that We will present examples as needed.

Example 1.3 A "duality".
The triangular density $(1-|x|)_+$ has the ch.f. $\dfrac{2(1-\cos t)}{t^2}$. The Polya density $\dfrac{1-\cos x}{\pi x^2}$ has the ch.f. $(1-|t|)_+$.

The symmetrized exponential distribution with the density $e^{-|x|}/2$ has the ch.f. $\dfrac{1}{1+t^2}$. The ch.f. of the Cauchy density $\dfrac{1}{\pi(1+x^2)}$ equals $e^{-|t|}$.

Using the idea from the proof of (3) of Proposition 1.1, for a family $\mu_\alpha = \mathcal{L}(X_\alpha)$ we obtain the upper estimate that involves the standard L^0-metric:

$$\sup_\alpha |\varphi_\alpha(s) - \varphi_\alpha(t)| \le \sup_\alpha \|(s-t)X_\alpha\|_0$$

Corollary 1.4 *If the family* $\{\mu_\alpha\}$ *is tight (i.e.,* $\{X_\alpha\}$ *is bounded in* L^0*), then* $\{\varphi_\alpha\}$ *is uniformly equi-continuous.* ■

The opposite implication is also true.

[1] only the "if" part is obvious now

Lemma 1.5 *Consider $\mu = \mathcal{L}(X)$, $\varphi = \hat{\mu}$. Then, for $r > 0$,*

$$(1) \quad \mathsf{P}(|X| \geq r) = \mu[-r,r]^c \leq \frac{r}{2}\int_{-2/r}^{2/r} (1-\varphi(t))\, dt,$$

$$(2) \quad \mathsf{P}(|X| \leq r) = \mu[-r,r] \leq 2r\int_{-1/r}^{1/r} |\varphi(t)|\, dt.$$

Proof.
W.l.o.g. we may and do assume that $r = 1$ (just consider X/r and change the variable in the right hand side integrals).

(1): By Fubini's theorem the right hand side equals

$$\mathsf{E}\,\frac{1}{2}\int_{-2}^{2}\left(1-e^{itX}\right)\, dt = \mathsf{E}\,2\left(1-\frac{\sin 2X}{2X}\right) \geq \mathsf{E}\,1\!\!1_{\{|X|\geq 1\}} = \mathsf{P}(|X| \geq 1).$$

(2): In virtue of Fubini's theorem the left hand side is estimated as follows, using the formula for the ch.f. of the triangular density:

$$\mathsf{E}\,\frac{1}{2}1\!\!1_{\{|X|\leq 1\}} \leq \mathsf{E}\,\frac{2(1-\cos X)}{X^2} = \mathsf{E}\int_{\mathbb{R}} (1-|t|)_+ e^{itX}\, dt = \int_{\mathbb{R}} (1-|t|)_+ \varphi(t)\, dt \leq \int_{\mathbb{R}} |\phi(t)|\, dt.$$

Corollary 1.6 *If a family $\{\varphi_\alpha\}$ of ch.fs. is equicontinuous at 0, then μ_α is tight.*

Proof. Let $\epsilon > 0$ and $\delta > 0$ be such that $\sup_\alpha |1-\varphi_\alpha(t)| < \epsilon/2$ whenever $|t| < \delta$. Let $r_0 = 2/\delta$. Then (1) in the Lemma entails

$$\sup_\alpha \mathsf{P}(|X_\alpha| \geq r) \leq \epsilon \quad \text{for } r > r_0$$

■

2 Continuity

Theorem 2.1 (Lévy Continuity Theorem)
For ch.fs. $\varphi_n = \widehat{\mu_n}$ and $\varphi_0 = \widehat{\mu_0}$ the following are equivalent:

1. *$\varphi_n \to \varphi_0$ pointwise;*
2. *$\mu_n \stackrel{w}{\to} \mu_0$;*
3. *$\varphi_n \to \varphi_0$ uniformly on every interval.*

Proof. (2) $\Rightarrow$ (1) follows by the definition of weak convergence and (3) $\Rightarrow$ (1) is obvious.

The remaining nontrivial implications (1) $\Rightarrow$ (2) and (2) $\Rightarrow$ (3) would be much easier to prove if the measures would have the common bounded support, i.e., underlying random variables were bounded. However, each of the assumptions implies that the family $\{\mu_n\}$ is tight, i.e. they are almost supported by a compact set.

(1) $\Rightarrow$ (2): Assume the point convergence of ch.fs., which means that $\mu_n e_t \to \mu_0 e_t$ for special functions $e_t(x) = e^{itx}$, and thus for their finite linear combinations, forming an algebra $\mathcal{A}$. We must show that $\mu_n f \to \mu_0 f$ for every continuous bounded function on $\mathbb{R}$.

We infer that $\{\mu_n\}$ is tight. Indeed, let $\epsilon > 0$ and choose $r > 0$ such that $|1-\varphi(t)| < \epsilon/4$ for every $t \in [-r, r]$. By Lemma 1.5.1 and the Dominated Convergence Theorem

$$\limsup_n \mu_n[-r,r]^c \le \limsup_n \frac{r}{2}\int_{-2/r}^{2/r} (1-\varphi_n(t))\,dt = \frac{r}{2}\int_{-2/r}^{2/r} (1-\varphi(t))\,dt < \epsilon/2$$

Then there is n_0 such that

$$\sup_{n>n_0} \mu_n[-r,r]^c < \epsilon.$$

At the same time there is $r' > 0$ such that

$$\sup_{n\le n_0} \mu_n[-r',r']^c < \epsilon$$

Taking $R = r \vee r'$,

$$\sup_n \mu_n[-R,R]^c < \epsilon.$$

For any continuous bounded complex function h on $\mathbb{R}$

$$\left| \int_{[-R,R]^c} h\,\mu_n - \int_{[-R,R]^c} h\,\mu \right| \le 2||h||_\infty\,\epsilon. \tag{2.1}$$

The Stone-Weierstrass Theorem (cf., e.g., Bartle, Theorem 26.2) says:

If $K \subset \mathbb{R}^d$ is compact, and $\mathcal{A}$ is a separating algebra with unit that consists of complex functions on K, then every continuous function on K can be uniformly approximated by members of $\mathcal{A}$.

Take $K = [-R, R]$ and the algebra $\mathcal{A}$, defined above. In virtue of the Stone-Weierstrass Theorem there is $g \in \mathcal{A}$ such that $||f-g||_K < \epsilon$. Hence, using (2.1) for $h = f$ and for $h = g$,

$$\begin{aligned} \Big|(\mu_n-\mu_0)f\Big| &\le \Big|(\mu_n-\mu_0)f\,\mathbb{1}_{K^c}\Big| + \le \Big|(\mu_n-\mu_0)f\,\mathbb{1}_K\Big| \\ &\le 2||f||_\infty\epsilon + \Big|(\mu_n-\mu_0)(f-g)\mathbb{1}_K\Big| + \Big|(\mu_n-\mu_0)g\,\mathbb{1}_{K^c}\Big| + \Big|(\mu_n-\mu_0)g\Big| \\ &\le \Big(2||f||_\infty + 2 + 2||g||_\infty\Big)\epsilon + \Big|(\mu_n-\mu_0)g\Big| \end{aligned}$$

Let $n \to \infty$ and then $\epsilon \to 0$.

(2) $\Rightarrow$ (3): Assume the weak convergence, consider an interval $[-T, T]$, and let $\epsilon > 0$. So, $\{\mu_n : n \ge 0\}$ is tight, hence there is $r > 0$ such that

$$\sup_{n\ge 0}\left|\int_{[-r,r]^c} e^{itx}\,\mu_n(dx)\right| \le \sup_n \mu_n[-r,r]^c \le \frac{\epsilon}{4}. \tag{2.2}$$

Then

$$|\varphi_n(t) - \varphi(t)| \le \left| \int_{[-r,r]^c} e^{itx}\, \mu_n(dx) \right| + \left| \int_{[-r,r]^c} e^{itx}\, \mu_0(dx) \right| + \left| \int_{-r}^{r} e^{itx}\, (\mu_n - \mu_0)(dx) \right| = A + B + C$$

By (2.2), the sum $A + B$ of the two first terms is bounded by $\epsilon/2$.

To estimate the third term, consider a partition (x_k) of $[-r, r]$ of mesh $< \epsilon/(8T)$, chosen from the continuity set of μ_0. In particular, we may enlarge the interval $[-r, r]$, so $-r$, the first point of the partition, and r, the last point of the partition, are also continuity points. In short,

$$\int_{-r}^{r} = \sum_k \int_{x_{k-1}}^{x_k}$$

Adding and subtracting the term e^{itx_k} on each interval $(x_{k-1}, x_k]$, C is bounded from above by the following expression:

$$\left| \sum_k \int_{x_{k-1}}^{x_k} \left(e^{itx} - e^{itx_k}\right) \mu_n(dx) \right| + \left| \sum_k \int_{x_{k-1}}^{x_k} \left(e^{itx} - e^{itx_k}\right) \mu_0(dx) \right| + \sum_k \Big| \mu_n(x_{k-1}, x_k] - \mu_0(x_{k-1}, x_k] \Big|$$

Since $\left| e^{itx} - e^{itx_k} \right| \le t|x - x_k|$, hence

$$\sum_k \left| \int_{x_{k-1}}^{x_k} \left(e^{itx} - e^{itx_k}\right) \mu_n(dx) \right| \le \frac{\epsilon}{8} \sum_k \left| \int_{x_{k-1}}^{x_k} \mu_n(dx) \right| \le \frac{\epsilon}{8},$$

so the sum of the first two terms in the latter estimate is less than $\epsilon/4$. For the third term, choose n_0 such that, for every $n \ge n_0$,

$$\sum_k \Big| \mu_n(x_{k-1}, x_k] - \mu(x_{k-1}, x_k] \Big| < \frac{\epsilon}{4}.$$

So, for every $\epsilon > 0$ and every interval $[-T, T]$ there is n_0 such that for every $t \in [-T, T]$ there holds $|\varphi_n(t) - \varphi(t)| < \epsilon$, which completes the proof of (2) $\Rightarrow$ (3).

∎

Corollaries and remarks

1. The Lévy Continuity Theorem easily extends to $\mathbb{R}^d$. In the language of random vectors

$$X_n \overset{\mathcal{D}}{\to} X \iff a \cdot X_n \overset{\mathcal{D}}{\to} a \cdot X, \quad a \in \mathbb{R}^d$$

2. **Lévy Uniqueness Theorem** $\varphi = \varphi_0 \iff \mu = \mu_0$.

3. The assumption that the limit φ_0 is a ch.f. can be relaxed.

   It suffices to assume that φ_0 is continuous at 0, then it will be a ch.f. of some measure μ_0.

This follows by Prokhorov's theorem. The continuity at 0 is necessary. For example, the ch.f. of the uniform distribution on $[-n,n]$ is $\frac{\sin nt}{nt}$, which converges to $\mathbb{1}_{\{0\}}(t)$. This sequence of measures is not tight. It generates a continuous functional on $C_b(\mathbb{R})$, a sort of generalized integral akin to Césaro sum:

$$\Lambda f = \lim_n \frac{1}{2n} \int_{-n}^{n} f(x)\, dx$$

which does not correspond to a measure.

4. The Lévy Continuity Theorem extends to bounded measures, also to signed measures with bounded variation, after the weak convergence is augmented by the condition $\lim_n \mu_n \mathbb{R} = \mu \mathbb{R}$.

3 Inversion

Define the signum function as

$$\sigma = \text{sign} = \mathbb{1}_{(0,\infty)} - \mathbb{1}_{(-\infty,0)}.$$

The following formula involves an improper integral of a function that is not integrable:

$$\int_{-\infty}^{\infty} \frac{\sin ux}{x}\, dx = \pi\sigma(u).$$

Its value follows from the Cauchy theorem that states that an analytic function $f(z)$ on an open simply connected domain in the complex plane entails the curve integral vanishing over a rectifiable simple closed curve. So, choosing $u \neq 0$ and then $u = 1$, $f(z) = \frac{e^{iz}}{z}$ is analytic on the complement of any closed disk centered at the origin. Denote by $S(r)$ the semidisk $|z| \le r$, $\Im z \ge 0$, and let $C = C(\epsilon, r)$ be the boundary of $S(r) \setminus S(\epsilon)$, oriented counterclockwise. Then, using the standard parametrization of four fragments - two segments and two semicircles, we have

$$0 = \oint_C f(z)\, dz = \int_\epsilon^r \frac{e^{ix}}{x}\, dx + i \int_0^\pi e^{ire^{i\theta}}\, d\theta + \int_{-r}^{-\epsilon} \frac{e^{ix}}{x}\, dx + i \int_\pi^0 e^{i\epsilon e^{i\theta}}\, d\theta \to i \int_{-\infty}^{\infty} \frac{\sin x}{x} - i\pi$$

as $\epsilon \to 0$, $r \to \infty$. This approach allows to define the integral

$$\int_{-\infty}^{\infty} \frac{e^{iut}}{it}\, dt \stackrel{\text{def}}{=} \lim_{\epsilon \to 0} \int_{|t|>\epsilon} \frac{e^{iut}}{it}\, dt = \pi\sigma(u)$$

Alternatively, we may consider a more standard option

$$\int_{-\infty}^{\infty} \frac{e^{iut} - e^{ivt}}{it}\, dt = \pi\left(\sigma(u) - \sigma(v)\right)$$

which entails

$$\int_{-\infty}^{\infty} \frac{e^{i(x-a)t} - e^{i(x-b)t}}{it}\, dt = 2\pi \mathbb{1}_{(a,b)}(x) + \pi \mathbb{1}_{\{a,b\}}(x) \tag{3.1}$$

As noted previously, the ch.f. of the uniform distribution on $[-T, T]$

$$\upsilon_T(t) = \frac{\sin Tt}{Tt} \to 1\!\!1_{\{0\}}(t) \text{ as } T \to \infty.$$

Hence, for a bounded discrete measure $\mu = \sum_n a_n \delta_{x_n}$, i.e. for which $a_n \geq 0$ and $\sum_n a_n < \infty$,

$$\lim_{T\to\infty} \frac{1}{2T} \int_T^T \hat{\mu}(t) e^{-itx_k} = \lim_{T\to\infty} \sum_n a_n \frac{\sin T(x_n - x_k)}{T(x_n - x_k)} = a_k.$$

On the other hand,

$$\frac{1}{\pi} \int_\infty^\infty \hat{\mu}(t) \frac{e^{-itv}}{it} \, dt = \sum_n a_n \sigma(x_n - v),$$

whence, for $u < v$,

$$\frac{1}{2\pi} \int_\infty^\infty \hat{\mu}(t) \frac{e^{-itu} - e^{-itv}}{it} \, dt = \sum_n a_n \frac{\sigma(x_n - u) - \sigma(x_n - v)}{2} = \mu(u, v) + \frac{\mu\{u, v\}}{2}. \qquad (3.2)$$

Since $\mu = \mu_d + \mu_c$, with the pure discrete and pure continuous part, we may consider only the latter.

Theorem 3.1 (Inversion Theorem) *Let μ be an atomless bounded measure. Then, for every $a < b$,*

$$\mu(a, b) = \frac{1}{2\pi} \int_{-\infty}^\infty \hat{\mu}(t) \frac{e^{-iat} - e^{-itb}}{it} \, dt$$

Proof. By Fubini's theorem for improper integrals (Exercise 2) we rewrite the right hand side and use (3.1):

$$\int_{-\infty}^\infty \frac{1}{2\pi} \int_{-\infty}^\infty \frac{e^{it(x-a)} - e^{it(x-b)}}{it} \, dt = \int_{-\infty}^\infty 1\!\!1_{(a,b)}(x) \, \mu(dx) = \mu(a, b).$$

■

Corollary 3.2 *If $\varphi = \hat{\mu}$ is integrable, then μ is absolutely continuous and its density $f(x)$ is continuous:*

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^\infty \varphi(t) \, e^{-itx} \, dt.$$

Proof. *Choose $a = x, b = x + h$ in the theorem, divide by h, and let $h \to 0$. The proof of continuity is left as an exercise.* ■

4 CLT

4.1 The basic CLT

A ch.f. can be perceived as a path in the unit disk of the complex plane. The ch.f. e^{ita} of a point mass δ_a is a periodic circular path, visiting the point 1 infinitely often. So does the ch.f. of a discrete measure with finitely many co-rational atoms x_n (that is, for some t, tx_n is an integer for every n). Periodicity will result from a lattice distribution of atoms (i.e., when they form an arithmetic sequence). Otherwise, only $\limsup_{t\to\infty} |\varphi(t)| = 1$ is certain (Lukacs, 2.2).

On the other hand, $\lim_{t\to\infty} |\varphi(t)| = 0$ indicates an absolutely continuous (atomless) measure (ibid.)

It follows immediately that the existence of the k-th moment of μ entails the existence of the k-th derivative of $\hat{\mu}$. The inverse implication is not quite simple and we will not go there (but see Exercise 3b).

For a complex valued function $g \in C^{n+1}(\mathbb{R})$, Taylor's theorem states that

$$g(x) = \sum_{k=0}^{n} \frac{g^{(k)}(0)}{k!}\, x^k + R_n(x).$$

Among various versions of the remainder, we choose the integral form:

$$R_n(x) = \frac{1}{n!} \int_0^x g^{(n+1)}(t)\,(x-t)^n\,dt$$

The formula is true under weaker assumptions. For example, it suffices to assume that the n-th derivative is absolutely continuous. Then the $(n+1)$-the derivative exists in the Radon-Nikodym sense and is integrable on every bounded interval. However, we are interested only in the smooth function $g(x) = e^{ix}$, with a simple remainder that can be further refined by integrating by parts:

$$R_n(x) = \frac{i^{n+1}}{n!} \int_0^x e^{it}\,(x-t)^n\,dt = i^n \left(-\frac{x^n}{n!} + \frac{1}{(n-1)!} \int_0^x e^{it}\,(x-t)^{n-1}\,dt \right).$$

Hence we obtain two upper estimates that we merge to one:

$$|R_n(x)| \le \frac{|x|^{n+1}}{(n+1)!} \wedge \frac{2|x|^n}{n!}$$

which is bounded by the second term for $|x| > 2(n+1)$.

Corollary 4.1 *Let* $\mathsf{E}\,|X|^n < \infty$, φ *denote the ch.f. of* X, *and* $m_k = \mathsf{E}\,X^k$, $k = 0, ..., n$. *Then*

$$\varphi(t) = \sum_{k=1}^{n} \frac{i^k m_k}{k!} t^k + R_k(t), \quad \textit{where } |R_k(t)| \le \mathsf{E}\, \frac{|tX|^{n+1}}{(n+1)!} \wedge \frac{2|tX|^n}{n!}$$

E.g., for $n = 2$,

$$\varphi(t) = 1 + it\,\mathsf{E}\,X - \frac{t^2}{2}\,\mathsf{E}\,X^2 + R_2(t)$$

where

$$|R_2(t)| \le \mathsf{E}\,\frac{|tX|^3}{6} \wedge |tX|^2.$$

■

We hardly ever need this "precision" with $\frac{1}{6}$. Let's make it cruder and simpler:

$$|R_2(t)| \le \mathsf{E}\,|tX|^2\,(|tX| \wedge 1) \le c_T\,\mathsf{E}\,X^2(|X| \wedge 1), \tag{4.1}$$

where $|t| \le T$ and $c_T = T^2(T \vee 1)$.

Note the immediate application that involves the standard Gaussian distribution $N(0,1)$:

$$\text{with the density } \frac{1}{\sqrt{2\pi}}\,e^{-x^2/2} \qquad \text{and the ch.f. } e^{-t^2/2}. \tag{4.2}$$

The Central Limit Theorem

Let $X_n \in L^2$ be i.i.d. with ch.f. φ. W.l.o.g. we may and do assume that $\mathsf{E}\,X_n = 0$ and $\mathsf{E}\,X_n^2 = 1$. Then

$$Y_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}} \stackrel{\mathcal{D}}{\to} N(0,1)$$

Proof. The ch.f. of Y_n can be estimated as follows:

$$p_n = \varphi(t/\sqrt{n})^n = \left(1 - \frac{t^2}{2n} + R_2(t/\sqrt{n}\right)^n,$$

That is, we can rewrite it a

$$p_n = \left(1 - \frac{a_n}{n}\right)^n, \text{ where } a_n \to a = t^2/2$$

Given $\epsilon \in (0,a)$, we find n_0 such that $|a_n - a| < \epsilon$ for every $n > n_0$, so

$$\left(1 - \frac{a+\epsilon}{n}\right)^n \le p_n \le \left(1 - \frac{a-\epsilon}{n}\right)^n$$

That is,

$$e^{-a-\epsilon} \le \liminf_n p_n \le \limsup_n p_n \le e^{-a+\epsilon}$$

Hence $\lim_n p_n = e^{-t^2/2}$. ■

4.2 Lindeberg-Feller Condition

The CLT is one of the basic examples of a limit theorem that establishes a limit distribution of a sequence of random variables (S_n), subject to affine transformations:

$$\frac{S_n - b_n}{a_n},$$

where a_n, b_n are scalar sequences, and S_n may depend on observed data, expressed as random variables, and be their function (a.k.a. a **statistic**), e.g., the sum, maximum, minimum, etc. For example, under moment assumptions, the centering scalar b_n could be the mean while the scaling scalar a_n could be the standard deviation of the transformed variable. The independence assumption may be relaxed, the moment assumptions may be dropped, so the centering and scaling constants might not be related to moments at all.

Presently we consider a random array $[\xi_{nk} : n \in \mathbb{N}, k \leq n]$ and denote

$$S_n = \xi_{n1} + \cdots + \xi_{nn}.$$

We assume that

$$\begin{aligned} &(1) \quad \text{For every } n,\ \xi_{n1}, \ldots, \xi_{nn} \text{ are independent;} \\ &(2) \quad \text{For every } n \text{ and } k \leq n,\ \mathsf{E}\,\xi_{nk} = 0,\ \sigma_{nk}^2 = \mathsf{E}\,|\xi_{nk}|^2 < \infty; \\ &(3) \quad s_n^2 = \mathsf{Var}(S_n) = \mathsf{E}\,|S_n|^2 = \sum_{k=1}^{n} \sigma_{nk}^2 = 1. \end{aligned} \tag{4.3}$$

Introduce also the **Lindeberg-Feller condition**

$$\forall\, \epsilon > 0 \quad \lim_n \ell_n(\epsilon) = 0, \ \text{where } \ell_n(\epsilon) = \sum_{j=1}^{n} \mathsf{E}\,\left[\,\xi_{nj}^2; |\xi_{nj}| > \epsilon\,\right] \tag{4.4}$$

Theorem 4.2 *Let a triangular random matrix* $[\xi_{nk}]$ *satisfy* (4.3)*. Then,*

$$\begin{aligned} &(a) \quad \max_k \sigma_{nk}^2 \to 0, \\ &(b) \quad \mathcal{L}(S_n) \to N(0,1). \end{aligned}$$

if and only if the Linderberg-Feller condition (4.4) *is satisfied.*

Proof. Assume (4.4). Then (a) follows since

$$\mathsf{E}\,\xi_{nk}^2 \leq \epsilon + \ell_n(\epsilon).$$

Consider a Gaussian random matrix $[\zeta_{nk}]$ with all characteristics (4.3), and denote $Z_n = \sum_k \zeta_{nk}$. Clearly, $Z_n \sim N(0,1)$. Let $|t| \leq T$.

$$\left|\mathsf{E}\,e^{itS_n} - \mathsf{E}\,e^{itZ_n}\right| = \left|\prod_k \mathsf{E}\,e^{it\xi_{nk}} - \prod_k \mathsf{E}\,e^{it\zeta_{nk}}\right| \leq \sum_k \left|\mathsf{E}\,e^{it\xi_{nk}} - \mathsf{E}\,e^{it\zeta_{nk}}\right|,$$

where the last inequality for products of complex numbers from the unit disk follows by induction. Continuing, the latter term is bounded by

$$\sum_k \left| \mathsf{E}\, e^{it\xi_{nk}} - 1 + \frac{t^2\sigma_{nk}^2}{2} \right| + \sum_k \left| \mathsf{E}\, e^{it\zeta_{nk}} - 1 + \frac{t^2\sigma_{nk}^2}{2} \right|.$$

Using the error estimate (4.1), the first of the above terms is bounded by

$$c_T \sum_k \mathsf{E}\, |\xi_{nk}|^2 (1 \wedge |\xi_{nk}|) \leq \epsilon\, c_T\, s_n^2 + c_T \ell_n(\epsilon)$$

Denoting the p'th absolute moment of a $N(0,1)$ Gaussian r.v. by m_p, the p-th moment of $N(0,\sigma^2)$ Gaussian r.v. ζ equals

$$\mathsf{E}\, |\zeta|^p = \sigma^p \mathsf{E}\, |\zeta/\sigma|^p = m_p\, \sigma^p$$

Hence, the second term is bounded by

$$c_T \sum_k \mathsf{E}\, |\zeta_{nk}|^2 (1 \wedge |\zeta_{nk}|) \leq c_T \sum_k \mathsf{E}\, |\zeta_{nk}|^3 = c_T\, m_3 \sum_k \mathsf{E}\, \sigma_{nk}^{3/2} \leq c_T m_3 \max_k \sigma_{nk}^{1/2}$$

Now, let $n \to \infty$, and then $\epsilon \to 0$.

To show that Lindeberg's condition is necessary, assume (a) and (b), and fix $\epsilon > 0$ and $t > 0$. Assuming (b), the weak convergence of laws S_n to the symmetric normal law, we infer that

$$\sum_k \ln \mathsf{E}\, \cos t\xi_{nk} = \ln \Re \mathsf{E}\, e^{itS_n} \to -\frac{t^2}{2}.$$

We claim that

$$\sum_k \mathsf{E}\left(1 - \cos t\xi_{nk}\right) \to \frac{t^2}{2}. \tag{4.5}$$

Indeed, let's write

$$b_n = \sum_k \ln \mathsf{E}\, \cos t\xi_{nk} + \sum_k \mathsf{E}\left(1 - \cos t\xi_{nk}\right).$$

Applying the inequality $|\ln z + 1 - z| \leq |1 - z|^2$ to $z = \mathsf{E}\, \cos t\xi_{nk}$, and next using the estimate $1 - \cos u \leq u^2/2$ (with $u = t\xi_{nk}$), we infer that

$$|b_n| \leq \sum_k \left| \mathsf{E}\, (1 - \cos t\xi_{nk}) \right|^2 \leq \sum_k \mathsf{E}\, \left(\frac{t^2 \xi_{nk}^2}{2} \right)^2 \leq \frac{t^4 \max_k \sigma_{nk}^2}{4} \sum_k \sigma_{nk}^2 \to 0,$$

which proves (4.5), by virtue of the assumed condition (a). On the left hand side of (4.5), consider a single term with $\xi = \xi_{nk}$. Then, since $1 - \cos u \leq u^2/2$ and $1 - \cos u \leq 2$,

$$\begin{aligned} \mathsf{E}\, (1 - \cos t\xi) &= \mathsf{E}\, [1 - \cos t\xi; |\xi| \leq \epsilon] + \mathsf{E}\, [1 - \cos t\xi; |\xi| > \epsilon] \\ &\leq \frac{t^2}{2} \mathsf{E}\, [\xi^2; |\xi| \leq \epsilon] + 2\mathsf{P}(|\xi| > \epsilon) \\ &\leq \frac{t^2}{2} \mathsf{E}\, [\xi^2; |\xi| \leq \epsilon] + \frac{2}{\epsilon^2} \mathsf{E}\, \xi^2, \end{aligned}$$

where the second term was estimated with the help of Chebyshev's inequality. In other words,

$$\begin{aligned} \mathsf{E}\,[\xi^2; |\xi| \le \epsilon] &\ge \frac{2}{t^2}\mathsf{E}\,(1-\cos t\xi) - \frac{4}{t^2\epsilon^2}\mathsf{E}\,\xi^2 \\ &= 1 + \frac{2}{t^2}\Big(\mathsf{E}\,(1-\cos t\xi) - \frac{t^2}{2}\Big) - \frac{4}{t^2\epsilon^2}\mathsf{E}\,\xi^2 \end{aligned}$$

Return to ξ_{nk}, sum up along j, and let $n \to \infty$, Then, (4.5) and the normalizing condition $s_n^2 \to 1$ imply

$$\liminf_n \sum_k \mathsf{E}\,[\xi_{nk}^2; |\xi_{nk}| \le \epsilon] \ge 1 - \frac{4}{t^2\epsilon^2}.$$

Although t was fixed, it is arbitrary. Now, let $t \to \infty$, ■

Corollary 4.3 (Lyapunov) *Let ξ_{nk} fulfill assumptions of Lindeberg's theorem, and let $\delta > 0$. Then*

$$\Big[\lambda_n(\delta) = \sum_k \mathsf{E}\,|\xi_{nk}|^{2+\delta} \to 0\Big] \Rightarrow \Big[S_n \stackrel{d}{\to} \zeta\Big]$$

Proof. Indeed, $\ell_n \le \lambda_n(\delta)/\epsilon^\delta$, so Lyapunov's condition implies Lindeberg's. ■

We return to the sequence of independent random variables (ξ_k). As before, assume $\mathsf{E}\,\xi = 0$, $\sigma_k^2 = \mathsf{E}\,\xi^2 < \infty$, $S_n = \xi_1 + \cdots + \xi_n$, $s_n = \sigma_1^2 + \cdots + \sigma_n^2$. Let $F_n = \mathsf{P}(S_n/s_n \le x)$ and $\Phi(x) = \mathsf{P}(\zeta \le x)$. If $F_n \stackrel{w}{\to} \Phi$, it is desirable to know how fast the convergence occur. That is, estimates of

$$\text{dist}\,(F_n, \Phi)$$

are of great practical and theoretical importance, where "dist" - preferably a metric - measures the convergence. Although the Lévy-Prokhorov metric seems to be the most natural choice since it metrizes the weak convergence of measures, its specific definition makes it difficult to examine. The uniform metric is stronger but more appropriate for applications. Of course, to obtain a stronger mode of convergence, a stronger assumption is needed. Let $F(x) = \mathsf{P}(\xi \le x)$, $G(x) = \mathsf{P}(\eta \le x)$ be probability distribution functions. Consider

$$\text{dist}(F, G) = \|F - G\|_\infty = \sup_x |F(x) - G(x)|.$$

Theorem 4.4 (Berry (1941), Esseen (1945), Van Beek 1972) *Assume $\mathsf{E}\,\xi = 0$, $\mathsf{E}\,\xi^2 = 1$, and $\mathsf{E}\,|\xi|^{2+\delta} < \infty$, for some $\delta > 0$. Let (ξ_k) be independent copies of ξ. Then*

$$\|F_n - \Phi\|_\infty \le c\,\frac{\mathsf{E}\,|\xi|^{2+\delta}}{\sqrt{n}},$$

where c is some universal constant, independent of n and of distribution of ξ, although it may depend on δ.

No proof will be presented here (see Durrett).

4.3 Poisson convergence

Let $[\xi_{nk}]$ be a triangular array of random variables:

$$\begin{aligned} &(1) \quad \text{values are whole numbers } 0, 1, 2, \ldots; \\ &(2) \quad \text{for every } n,\ \xi_{nk} \text{ are independent;} \\ &(3) \quad \max_k \|\xi_{nk}\|_0 \to 0 \text{ as } n \to \infty, \text{ where } \|\cdot\|_0 \text{ is any } L^0\text{-metric.} \end{aligned} \tag{4.6}$$

We are free to choose $\|X\|_0 = \mathsf{E}\,(1 \wedge |X|)$, or $\|X\|_0 = \mathsf{E}\,(1 - e^{-tX})$ for a fixed $t > 0$.

As before, denote $S_n = \sum_k \xi_{nk}$. We will discuss the weak convergence of its distribution to the Poisson distribution on $\mathbb{Z}_+$:

$$\mu\{n\} = \mathsf{P}(\xi = n) = e^{-\lambda}\,\frac{\lambda^n}{n!}, \quad \hat{\mu}(t) = e^{\lambda(e^{it}-1)}, \quad \tilde{\mu}(t) = e^{\lambda(1-e^{-t})}.$$

Lemma 4.5 *Assume* (4.6)*. Then the following conditions are equivalent:*

1. $S_n \stackrel{\mathcal{D}}{\to} \xi$
2. $-\sum_k \ln \psi_{nk}(t) \to \lambda(1 - e^{-t}),\ t > 0.$
3. $C_n = \sum_k \Big(1 - \psi_{nk}(t)\Big) \to \lambda(1 - e^{-t}),\ t > 0.$

Proof. The equivalence of the first two conditions follows by the Lévy Continuity Theorem for Laplace transforms $\psi(t) = \mathsf{E}\,e^{-tX}$, after applying the logarithm.

For $u \in [0, 1/2]$ we have the identity

$$-\ln(1-u) = u + r(u), \quad \text{where } 0 \le r(u) \le u^2. \tag{4.7}$$

To show the equivalence of the second and third condition, we apply it with $u = u_{nk} = 1 - \psi_{nk}$, which is arbitrarily small by the third assumption in (4.6). Then we sum up along k. That is, either we assume (3), so $C_n \le C$, or we assume (2) which yields

$$C_n = \sum_k (1 - \psi_{nk}) \le -\sum_k \ln \psi_{nk} \le C.$$

Hence the remainder is bounded by

$$\sum_k r_{nk} \le \sum_k (1 - \psi_{nk})^2 \le C\,\max_k \|\xi_{nk}\|_0 \to 0,$$

where we choose $\|X\|_0 = \mathsf{E}\,(1 - e^{-tX})$. ∎

Theorem 4.6 *Let a random triangular array* $[\xi_{nk}]$ *satisfy* (4.6) *and* ξ *be Poisson*(λ)*. Then*

$$\left.\begin{array}{ll}\text{(a)} & \sum_k \mathsf{P}(\xi_{nk} > 1) \to 0 \\ \text{(b)} & \sum_k \mathsf{P}(\xi_{nk} = 1) \to \lambda\end{array}\right\} \iff S_n = \sum_k \xi_{nk} \xrightarrow{\mathcal{D}} \xi$$

Proof. The necessity. Suppose that $S_n \xrightarrow{\mathcal{D}} \xi \sim \text{Poisson}(\lambda)$, and look at the third condition in Lemma 4.5. For simplicity, denote $s = e^{-t}$. So, $C_n \to \lambda(1-s)$. For a single random variable X with values in $\mathbb{Z}_+$ we have

$$\mathsf{E}\,(1-s^X) = \mathsf{E}\,[1-s^X; X>0] = \mathsf{E}\,[1-s+s-s^X; X>0] = (1-s)\mathsf{P}(X>0) + R(s)$$

where

$$R(s) = \mathsf{E}\,[s-s^X; X>0] = \mathsf{E}\,[s-s^X : X>1], \quad \text{hence}(s-s^2)\mathsf{P}(X>1) \le R(s) \le s\mathsf{P}(X>1).$$

(because $X > 1 \iff X \ge 2$). Since

$$C_n = (1-s)\sum_k \mathsf{P}(X_{nk}>0) + \sum_k R_{nk}(s), \quad \text{where } R_{nk}(0) = 0,$$

thus

$$\sum_k \mathsf{P}(X_{nk}>0) \to \lambda \quad \text{and} \quad \sum_k R_{nk}(s) \to 0, \quad \text{hence} \quad \sum_k P(\xi_{nk}>1) \to 0$$

so (b) and (a) hold true.

The sufficiency. First, we reduce the range of r.vs. to the mere $\{0,1\}$. Then (a) is trivially true. Suppose that [*] "*(b) is sufficient for the Poisson convergence for 0-1 r.vs.*" Denote $\xi'_{nk} = \mathbb{I}\{\xi_{nk} = 1\}$.

Assume (b), which is the same for both ξ_{nk} and ξ'_{nk}. So, by [*] $S'_n = \sum_k \xi'_{nk} \xrightarrow{\mathcal{D}} \xi$. But

$$S_n = S'_n + R_n, \quad \text{where } R_n = \sum_k \xi_{nk} \mathbb{I}_{\{\xi_{nk}>1\}}.$$

By (a), using the subadditivity $1 \wedge \sum_k c_k \le \sum_k (1 \wedge c_k)$,

$$\mathsf{E}\,1 \wedge R_n \le \mathsf{E}\sum_k 1 \wedge \xi_{nk}\mathbb{I}_{\{\xi_{nk}>1\}} = \sum_k \mathsf{P}(\xi_{nk}>1) \to 0,$$

i.e., $R_n \xrightarrow{\mathsf{P}} 0$, so $S_n \xrightarrow{\mathcal{D}} \xi$.

Now, assume (b) for an 0-1 array, and compute the logarithm of the Laplace transform $\mathsf{E}\,\exp\{-tS_n\}$, with the notation $u_{nk} = 1 - \mathsf{E}\,\exp\{-t\xi_{nk}\} = (1-e^{-t})\mathsf{P}(\xi_{nk}=1)$, using (4.7):

$$-\sum_k \ln(1-u_{nk}) = (1-e^{-t})\sum_k \mathsf{P}(\xi_{nk}=1) + R_n, \qquad \text{where } R_n \to 0$$

since

$$R_n \leq \sum_k u_{nk}^2 \leq C \max_k \|\xi_{nk}\|_0.$$

Thus, we obtain the Laplace transform of Poisson(λ) in the limit. ■

Remark 4.7 In elementary probability courses the special case of i.i.d. Bernoulli ξ_{nk}'s is known as **the Poisson approximation of the Binomial**. Indeed, in this case S_n is binomial, where it is also assumed that $p_n = \mathsf{P}(X_{n1} = 1) \to 0$, and then (b) means that $np_n \to \lambda$.

Exercise. What condition imposed on p_n does ensure (or, is necessary and sufficient for) the Lindeberg-Feller condition. i.e. Gaussian rather than Poisson convergence? Note that the each entry ξ_{nk} needs to be standardized to fulfill the standing assumptions for the CLT for random arrays.

4.4 Exercises

1. Verify the relations "density vs. ch.f." in Example 1.3 and formula (4.2).

2. Let $(S, \mathcal{S}, \mu)$ be a bounded measure space and $f(t,s)$ a measurable real or complex function on $\mathbb{R} \times S$. Assume that

   - $f(\cdot, s)$ is locally (i.e., on every interval) integrable functions on $\mathbb{R}$ for almost every $s \in S$;
   - $f(t, \cdot)$ is μ-integrable for almost every $t \in \mathbb{R}$;
   - The improper integral $g(s) = \int_{\mathbb{R}} f(t,s)\,dt \overset{\mathsf{def}}{=} \lim_{T\to\infty} \int_{-T}^{T} f(t,s)\,dt$ exists for almost every $s \in S$;
   - $\sup_T \left| \int_{-T}^{T} f(t,s) dt \right|$ is μ-integrable;

   Then
   $$\int_{\mathbb{R}} \left(\int_S f(t,s)\,\mu(ds) \right) = \int_S g(s)\,\mu(ds).$$

3. A discrete version of the above theorem [**Abel's convergence criterion for infinite series**]. Prove:

   If $a_n \searrow 0$ *and* $\sup_N \left| \sum_{n=1}^{N} b_n \right| < \infty$, *then* $\sum_n a_n b_n$ *converges.*

   Hint: write $d_n = a_{n-1} - a_n$ and $B_n = b_1 + \cdots + b_n$, and split the sum (discovering and proving Abel's "summation by parts" formula):
   $$\sum_{n=0}^{N} a_n\, b_n = a_N\, B_N - \sum_{n=0}^{N-1} d_n B_n.$$

   (a) Let $p > 0$. Show that $\sum_n \frac{\sin an}{n^p}$ converges for every real a, and $\sum_n \frac{\cos an}{n^p}$ converges for $a \notin 2\pi\mathbb{Z}$.

   (b) Let $\mu = \sum_n \frac{C}{n^2 \ln n} \delta_n$, where C makes probabilities of the sequence $(n^2 \ln n)^{-1}$. The first moment does not exist but $\hat{\mu}$ is differentiable for $t \neq 0$. Prove also the same statement when atoms $\{ n \}$ oscillate, i.e., replace δ_n by $\delta_{(-1)^n n}$.

4. Show that the density in Corollary 3.2 is continuous.

5. On the unit disk of the complex plane, $|\prod_k z_k - \prod_k w_k| \le \sum_k |z_k - w_k|$.

6. Find the arbitrary n-th absolute moment $\mathsf{E}\,|\zeta|^n$ of a standard $N(0,1)$ r.v. ζ. Hint: in the first semester we evaluated even moments (while studying Marcinkiewicz-Zygmund-Paley inequalities).

7. What condition imposed on p_n in the triangular matrix of Bernoulli r.vs. (i.i.d. in each row) will ensure (or be implied by) the Lindeberg-Feller condition. i.e. Gaussian rather than Poisson convergence?

5 Poisson random measure

In science and beyond the most typical activity is counting. Scientists and beyondists count everything, stars in sky sectors, pollutant particles in water or air, bird nests per area of Alabama (or Alaska), coins in collections, gold nuggets in mines, Burmese pythons in Everglades, Occupieds per city, customers in burger joints, votes in the GOP primary per Florida county, etc.

Typically, the count involves the number of items per region, may vary from 0 through all natural numbers, and there is no reason to assume that there is a definite upper bound. All the listed - and unlisted - examples involve measurable regions - linear, planar, spatial, etc. It stands to reason to suppose that the count depend more on the measurement (length, area, volume) than on other aspects like geometry or topology. Also, the counts in separate regions should be independent.

Of course, both assumptions are ideal but so are all human made models.

The randomness is entailed by the random distribution of items Y_n - wether arrival moments on the temporal line, or scatter points in the plane or surface, or in the space or 3D-manifold, or just in an abstract set S. So, the count is

$$N(A) = \sum_n \mathbb{1}_A(Y_n). \tag{5.1}$$

The formula can be viewed from the measure-theoretic point of view. Denoting by δ_a the atomic measure at a point a, we may write

$$N = \sum_n \delta_{Y_n}, \tag{5.2}$$

and then for $f \geq 0$,

$$Nf = \int_S f\, dN = \sum_n f(Y_n).$$

Let $(S, \mathcal{S}, \lambda)$ be a σ-finite continuous measure space and $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space, both entailing the L^0-spaces of measurable function. While $L^0(\Omega)$ with convergence in P is metrizable by the traditional metric $\mathsf{E}\,(1 \wedge |X|)$, the analogous metric $\int_S (1 \wedge |f|)\, d\lambda$ yields a topology essentially stronger than the convergence in measure although weaker than L^1. On $L^\infty(S)$ it is L^1, though.

A sequence of random elements Y_n in S (i.e., measurable mappings $\Omega \mapsto S$) entails a random counting measure on $\mathcal{S}$, a so called **point process**. It is not immediately clear whether the converse is true, that is, a random counting measure requires random points to be counted.

The counting random measure is just one example with the concept of a **random measure**, i.e. a mapping $X : \mathcal{S} \to L^0(\Omega, \mathcal{F}, \mathsf{P})$ such that

$$X \bigcup_n A_n = \sum_n X A_n, \quad A_n \in \mathcal{S} \text{ are disjoint.} \tag{5.3}$$

The series on the right should converge in probability and, a fortiori, the convergence must be unconditional, i.e., independent of permutations of the indices. The range might be a narrower subspace of L^0 such as L^1 or L^2. Thus, a random measure is factually a **vector measure** which extends the classical concept of a nonnegative countably additive set function. For example, a signed measure is an $\mathbb{R}$-valued vector measure.

A deterministic **control measure** is a very convenient tool:

$$XA_n \to 0 \iff \lambda A_n \to 0.$$

Then it would suffice to introduce the random measure on a generator $\mathcal{S}_0$. For example, when $S = \mathbb{R}^d$, and $\mathcal{S}$ consists of Borel sets, with a control measure it suffices to define X on simple figures such as intervals.

5.1 Poisson measure and integral

The function $x \mapsto 1 \wedge x$ on the positive half-line can be replaced by another more convenient function. Below we shall use $\psi(x) = 1 - e^{-x}$ for reasons that will soon become clear. So, for random variables, the L^0-metric is

$$\|X\|_0 = \mathsf{E}\,\psi(|X|).$$

The mapping $\xi : \mathcal{S} \to L^0(\Omega, \mathcal{F}, \mathsf{P})$ is called a **Poisson random measure** (PRM) if

1. ξA is Poisson(λA), for every $A \in S$ of finite measure λ;
2. ξA and ξB are independent if $A \cap B = \emptyset$.

We call λ the **control measure** of ξ. At this moment the issue of existence is not yet resolved but properties can be easily derived.

Proposition 5.1 *Let ξ be a PRM with a control measure λ. Let $A_1, \ldots, A_n$ be disjoint measurable sets of finite measure. Then $\xi A_1, \cdots, \xi A_n$ are independent, and their joint Laplace transform is*

$$E\,\exp\left\{ -\sum_k t_k\, \xi A_k \right\} \;=\; \exp\left\{ -\sum_k \left(1 - e^{-it_k}\right) \lambda A_k \right\}\;.$$

Proof. The Laplace transform formula follows by induction and utilizes the property of Poisson distribution: the sum of two independent Poisson random variables is again Poisson, and the parameters add up. ■

First, we note that ξ is factually a countably additive function (in the sense to be explained) on the δ-ring of $\mathcal{S}_0$ subsets of $\mathcal{S}$ of finite measure. Let $A \in \mathcal{S}_0$ and $A = \bigcup_k A_k$, where $A_k \in \mathcal{S}_0$ are disjoint. Then

$$\mathsf{E}\,\xi A = \lambda A = \sum_k \lambda A_k = \sum_k \mathsf{E}\,\xi A_k = \mathsf{E}\,\sum_k \xi A_k, \quad \text{or } \mathsf{E}\left(\xi(A) - \sum_k \xi A_k\right) = 0$$

(the r.v. in parentheses is nonnegative[2]). That is, ξ is countably additive as a mapping with values in $L^1(\Omega, \mathcal{F}, \mathcal{P})$.

Corollary 5.2 From the measure-theoretic point of view the Laplace transform formula appears as the assignments

$$f = \sum_k t_k\, \mathbb{1}_{A_k} \quad \mapsto \quad \xi f \stackrel{\mathsf{def}}{=} \sum_k t_k\, \xi(A_k) = \int_S f(s)\, \xi(ds).$$

$$\mathsf{E}\;\exp\{-\xi f\} \;=\; \exp\left\{ -\int_S \left(1 - e^{-f(s)}\right) \lambda(ds) \right\}$$

[2] why?

The quantity entails a complete metric vector subspace of measurable functions on S. Recall $\psi(x) = 1 - e^{-x}$, $x \geq 0$.

$$\begin{aligned} \|f\|_0 &= \int_S \psi(|f|)\, d\lambda \\ \mathbb{L} &= \{ f \in L^0(S) : \|f\|_0 < \infty \}, \quad d(f,g) = \|f-g\|_0, \end{aligned}$$

where simple functions form a dense subset. Write $\lambda F = \int_S F d\lambda$. Then the last formula in the Corollary can be rewritten as

$$\mathsf{E}\, \psi(\xi f) = \psi(\lambda \psi(f)).$$

In other words,

$$\|\xi f\|_0 = \psi(\|f\|_0) \tag{5.4}$$

which establishes a homeomorphism between the space of simple f's and their Poisson integrals.

Proposition 5.3 *For a Poisson random measure ξ consider the positive cone $\mathbb{L}_+ = \{ f \in \mathbb{L} : f \geq 0 \}$. Then the mapping ξf, defined originally for simple functions, extends to a continuous positive-linear mapping from $\mathbb{L}_+$ into $L^0_+(\Omega, \mathcal{F}, \mathsf{P})$, and* (5.4) *continues to hold*

Then ξ extends to a continuous linear mapping on $\mathbb{L}$, defined as $\xi f = \xi f_+ - \xi f_-$.

Proof. Let $f \in \mathbb{L}_+$ and $f_n \geq 0$ be increasing simple measurable functions such that $f = \lim_n f_n = \sup_n f_n$ and $\|f - f_n\|_0 \to 0$. Clearly, the well defined ξf_n increase a.s. and ξf_n is a Cauchy sequence in L^0. Indeed, for $n \geq m$, by (5.4)

$$\|\xi f_n - \xi f_m\|_0 = \psi(\|f_n - f_m\|_0) \to 0$$

So $\xi f = \lim_n \xi f_n$ exists in probability and hence a.s. (since the sequence increases). Further, (5.4) is preserved in the limit, which ensures the other listed properties,

■

5.2 About stochastic processes

The definition of PRM contains a family of finite dimensional distributions

$$\{ \mu_{A_1,\ldots,A_n} : \text{ disjoint } A_k \in \mathcal{S} \}.$$

Although it is easy to create a random vector $(\xi_1, \ldots, \xi_k)$ with independent Poisson (λA_k) components, the existence of a robust mapping $\xi : \mathcal{S} \to L^0$ is not immediately obvious. It is a special case of a more general problem.

Let T be a nonempty set and $X = (X_t : t \in T)$ be a family of real random variables, a.k.a. **stochastic process**. By **finite dimensional distributions** (FDD) of X we understand the Borel probability measures

$$\mu_{t_1,\dots,t_n} = \mathcal{L}\Big(X_{t_1},\dots,X_{t_n}\Big) \text{ on } \mathcal{B}(\mathbb{R}^n), \quad n \in \mathbb{N}, \quad t_1,\dots,t_n \in T,$$

μ_τ in short, where $\tau = \{\, t_1,\dots,t_n \,\}$. So, more precisely, μ_τ is a Borel measure on $\mathbb{R}^\tau$. We notice the obvious relation for $m > n$

$$\mathsf{P}\Big(X_{t_1} \in A_1,\dots,X_{t_n} \in A_n, X_{t_{n+1}} \in \mathbb{R},\dots,X_{t_m} \in \mathbb{R}\Big) = \mathsf{P}\Big(X_{t_1} \in A_1,\dots,X_{t_n} \in A_n\Big), \qquad A_k \in \mathcal{B}(\mathbb{R})$$

In terms of probability measures, we say that their family is **consistent**. That is, for finite $\tau' \supset \tau$

$$\mu_{\tau'}\Big(A \times \mathbb{R}^{\tau'\setminus\tau}\Big) = \mu_\tau(A), \quad A \in \mathcal{B}(\mathbb{R}^\tau).$$

So, the passage from the family of random variables (X_t) to the consistent family of multidimensional probability distribution (μ_τ) is immediate but the inverse implication is highly nontrivial, and is known as **Kolmogorov Extension Theorem** (cf., e.g. Theorem 6.16 in Kallenberg, or the special case in Appendix A7 in Durrett).

Even the existence of an infinite sequence of independent random variables belongs to this category. However, we introduced a countable product measure in the first semester. That is, if $(\Omega_k, \mathcal{F}_k, P_k)$ is an infinite sequence of probability spaces, then there is a product measure $P = P_1 \otimes \mathsf{P}_2 \otimes \cdots$ on

$$(\otimes, \mathcal{F}) = (\prod_k \Omega_k, \mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \cdots).$$

In particular, if μ_k are Borel probability measures on $\mathbb{R}$, then the well defined product measure

$$\mu = \mu_1 \otimes \mu_2 \otimes \cdots \text{ on } (\mathbb{R}^\mathbb{N}, \mathcal{B}(\mathbb{R}^\mathbb{N}))$$

entails independent random variables

$$X_n(\omega) = \omega_n, \quad \omega = (\omega_n) \in \mathbb{R}^\mathbb{N}.$$

Therefore, any constructive and intuitive approach should be appreciated.

5.3 Classical Poisson Process

There is one-to-one correspondence between increasing sequences y_n on $[0,\infty)$ and nondecreasing piecewise constant right continuous functions $n(t)$ with unit jumps (CF - for "counting functions") on $[0,\infty)$:

$$\text{given } y_n \nearrow \qquad \text{put} \qquad n(t) = \sum_n 1\!\!1_{[0,t]}(y_n);$$

$$\text{given a CF } n(t) \qquad \text{put} \qquad y_n = n^{\text{th}} \text{ jump of } n(t);$$

In other words, for every $t \geq 0$ and $n = 0, 1, 2, ...$

$$n(t) \geq n \iff y_n \leq t, \quad \text{or} \quad n(t) = \sup\{\, k : y_k \leq t \,\}. \tag{5.5}$$

For any nonnegative Borel function f on $[0,\infty)$, the Lebesgue-Stieltjes integral is well defined although it could be infinite

$$nf = \int_0^\infty f(t)\, dn(t) = \sum_n f(y_n).$$

Hence any increasing random sequence Y_n entails the CF N_t and a random counting measure NA (5.1), and then the integral of a nonnegative function. Conversely, a counting random measure defines the random CF N_t, and its discontinuities define Y_n's. We shall call them **signals**.

If (V_n) are i.i.d. and $Y_n = V_1 + \ldots + V_n$, then the CF N_t defined by (5.5) is called a **renewal process**. The most important case involves the exponential distribution of the summands V_k. Denote the parameter, also called the **intensity**, by λ. We will show that N_t induces a Poisson random measure with the scaled Lebesgue measure as a counting measure.

Proposition *The r.v. N_t has the Poisson(λt) distribution.*

Proof: Y_n has the Gamma distribution with the density

$$f_n(x) = \frac{\lambda^n}{(n-1)!}\, x^{n-1}\, e^{-\lambda x}$$

Hence, by conditioning and since $Y_{n+1} = Y_n + V_{n+1}$

$$\begin{aligned}
\mathsf{P}(N_t = n) = \mathsf{P}(N_t \geq n, N_t < n+1) &= \mathsf{P}(Y_n \leq t, Y_n + V_{n+1} > t) \\
&= \int_0^t \mathsf{P}(V_{n+1} > t - x) f_n(x)\, dx \\
&= \frac{\lambda^n}{(n-1)!} \int_0^t e^{-\lambda(t-x)} x^{n-1}\, e^{-\lambda x}\, dx = \frac{(\lambda t)^n}{n!} e^{-\lambda t}
\end{aligned}$$

■

We call N_t the **Poisson process**. Note that the name does not and should not apply to the sequence Y_n, although it determines N_t. For that reason the terminology is often abused, and the sequence is improperly called "the Poisson process".

Define the **age time** from the given moment to the last signal that precedes it, and the **excess time** from the moment t to the next signal:

$$A_t = t - Y_{N_t}, \quad W_t = Y_{N_t+1} - t$$

In what follows the crucial role is played by the "**lack of memory**" of an exponential distribution:

$$\mathsf{P}(U_1 > t + s | U_1 > t) = \mathsf{P}(U_1 > s).$$

Proposition. *The excess time W_t is independent of N_t and $\mathcal{L}(W_t) = \mathcal{L}(U_1)$.*
Proof: Compute

$$\begin{aligned} \mathsf{P}(Y_{n+1} \geq t+s,\, Y_n \leq t) &= \int_0^t \mathsf{P}(U_{n+1} \geq t+s-x)\, f_n(x)\, dx \\ &= \frac{\lambda^n}{(n-1)!} \int_0^t e^{-\lambda(t+s-x)}\, x^n\, e^{-\lambda x}\, dx = \frac{(\lambda t)^n}{n!}\, e^{-\lambda(t+s)} \end{aligned}$$

Since $\{ N_t = n \} = \{ Y_n \leq t,\, Y_{n+1} \geq t \}$, this yields

$$\mathsf{P}(W_t \geq s | N_t = n) = \mathsf{P}(Y_{N_t+1} \geq t + s | N_t = n) = \frac{\mathsf{P}(Y_{n+1} \geq t+s,\, Y_n \leq t)}{\mathsf{P}(N_t = n)} = e^{-\lambda s}$$

and so

$$\mathsf{P}(W_t \geq s) = \sum_n \mathsf{P}(W_t \geq s | N_t = n) \mathsf{P}(N_t = n) = e^{-\lambda s},$$

which also entails the independence: $\mathsf{P}(W_t \geq s | N_t = n) = \mathsf{P}(W_t \geq s)$. ■

Corollary 5.4

1. *The Poisson process starts afresh and independently after any time t.*

   More precisely, given $t > 0$, define i.i.d. $\exp(\lambda)$ r.vs:

   $$U_1' = W_t = Y_{N_t+1} - t, \quad U_2' = Y_{N_t+2} - Y_{N_t+1}, ... U_k' = Y_{N_t+k} - Y_{N_t+k-1}, ...$$

   and also

   $$Y_n' = U_1' + \cdots + U_n'.$$

   Then, by Proposition, U_1' is independent of N_t, so it is independent of $(U_k', k \geq 2)$. Then, for $m \geq 2$,

   $$\begin{aligned} \mathsf{P}(U_k' \leq u_k,\, k = 2, \ldots, m) &= \sum_n \mathsf{P}(Y_{n+k} - Y_{n+k-1} \leq u_k,\, k = 2, \ldots, m | N_t = n) \mathsf{P}(N_t = n) \\ &= \mathsf{P}(U_k \leq u_k,\, k = 2, \ldots, m) = \mathsf{P}(U_2 \leq u_2) \cdots \mathsf{P}(U_m \leq u_m) \end{aligned}$$

2. *$N(t, t+s] = N_{t+s} - N_t$ is independent of N_t and is distributed as N_s:*

$$N(t, t+s] = \sum_n \mathbb{1}_{(t,t+s]}(Y_n) = \sum_n \mathbb{1}_{(t,t+s]}(Y'_n).$$

In other words, the distribution of increments depends only on their durations not on their locations, and we often say that the process is **stationary**.

By induction, the independence holds true for any finite number of disjoint of increments.

3. *N_t entails a Poisson measure that starts with $N(a, b] = N_b - N_a$.*

4. *The age time A_t and excess time W_t have the same exp(λ) distribution.*

Hence we encounter a paradox: if at time $t > 0$ the previous and the next signals are observed, then the epoch - the time distance between them - has the expectation twice as long than the average epoch between two arbitrary signals:

$$\mathsf{E}\, U_{N_t+1} = \mathsf{E}\left(Y_{N_t+1} - Y_{N_t}\right) = \mathsf{E}\left(A_t + W_t\right) = 2\,\mathsf{E}\, U_1$$

Is it really a paradox?

5.4 Transformations of Poisson process

5.4.1 Nonhomogeneous Poisson process

Let $\phi : (0, \infty) \to (0, \infty)$ be a strictly monotonic function with the inverse $\Lambda = \phi^{-1}$. We assume the strict monotonicity for the sake of clarity of presentation. Otherwise, for function that may be piecewise constant we would have to use the generalized inverse.

Given a Poisson process N_t with unit intensity, transform its Gamma-distributed signals Y_n into $Z_n = \phi(Y_n)$, and denote the new counting process by M_t or the counting measure by MA. That is

$$MA = \sum_n \mathbb{1}_A(Z_n) = \sum_n \mathbb{1}_A(\phi(Y_n)) = \sum_n \mathbb{1}_{\Lambda(A)}(Y_n) = N\Lambda(A).$$

Hence $MA_1, \ldots, MA_n$ are independent when $A_1, \ldots, A_n$ are disjoint, and $M(A)$ is Poisson with parameter $\Lambda(A)$. Notice that ΛA is a measure, e.g.

$$\Lambda(a, b] = |\Lambda(b) - \Lambda(b)|.$$

Thus M is a Poisson random measure on the range $\phi(0, \infty)$. If the measure Λ is absolutely continuous with respect to the Lebesgue measure, then denoting its density by $\lambda(t)$, also called the **intensity function** we obtain

$$\Lambda A = \int_A \lambda(t)\, dt.$$

Examples.

1. Poisson process often serves as a model of customer service. However, its original setup would require the 24/7 servicing, in contrast to the usual piecewise service periods as in banking hours 9-5 for example. So, we can use two-valued $\{0, \lambda\}$ intensity function, with hours as time units:
$$\lambda(t) = \lambda \sum_{n=0}^{\infty} \mathbb{1}_{(9+24n, 17+24n]}(t).$$
The above "square wave" is just one example of a periodic intensity function.

2. Say, $\phi(t) = t^2, t > 0$. Then signals are $Y_1^2, Y_2^2, \ldots$. Then $\Lambda(t) = \sqrt{t}$ and its intensity is $\lambda(t) = \Lambda'(t) = \dfrac{1}{2\sqrt{t}}$.

   For a general power $\phi(t) = t^p$, $\lambda(t) = t^{1/p-1}/|p|$. E.g., the transformation $\phi(t) = 1/t$ entails the intensity $\lambda(t) = 1/t^2$, so with probability 1 the number of signals Z_n in every half-line $[a, \infty)$ with $a < 0$ is finite.

5.4.2 Reward or compound Poisson process

Write the Poisson process, Poisson measure or Poisson integral again:
$$N_t = \sum_n \mathbb{1}_{[0,t]}(Y_n), \quad NA = \sum_n \mathbb{1}_A(Y_n), \quad Nf = \sum_n f(Y_n).$$
Let R_n be i.i.d. r.vs. ("rewards"), independent of N that replace unit size jumps by R_n's. Define ad rewrite
$$M_t = \sum_n R_n \, \mathbb{1}_{[0,t]}(Y_n) = \sum_n R_n \, \mathbb{1}_{\{Y_n \le t\}} = \sum_n R_n \, \mathbb{1}_{\{N_t \ge n\}} = \sum_{n=1}^{N_t} R_n \tag{5.6}$$
with the convention $\sum_{n=1}^0 = 0$. We may also write
$$Mf = \sum_n R_n \, f(Y_n).$$
Let us compute the Laplace transform using Fubini's Theorem (subscripts at the expectations indicate the suitable integrals) and abbreviating $R_1 = R$:
$$\mathsf{E}\, e^{-Mf} = \mathsf{E}_N \, \mathsf{E}_R \exp\left\{ -\sum_n R_n f(Y_n) \right\} = \mathsf{E}_N \prod_n \mathsf{E}_R \exp\{ -R f(Y_n) \}.$$
Introducing the function
$$g(x) = -\ln \mathsf{E}_R \exp\{ -R f(x) \}$$
we obtain the formula (no need to use the subscript anymore)
$$\mathsf{E}\, e^{-Mf} = \mathsf{E} \prod_n \exp\{ -g(Y_n) \} = \mathsf{E} \exp\left\{ -\sum_n g(Y_n) \right\} = \mathsf{E}\, e^{-Ng} = \exp\left\{ -\int_0^\infty \left(1 - e^{-g(x)}\right) dx \right\}$$

Now, removing the function g we arrive at the identity

$$\mathsf{E}\, e^{-Mf} = \exp\left\{ -\int_0^\infty \mathsf{E}\left(1 - e^{-Rf(x)}\right)\, dx \right\}. \tag{5.7}$$

One more time, denote by μ the probability distribution of R, supported by $[0,\infty)$, and let $S = [0,\infty)^2$ with Borel sets and the product measure $\lambda = \mu \otimes \mathrm{Leb}$ ("Leb" of course denotes the Lebesgue measure). Also, define the positively linear operator

$$[0,\infty)^2 \ni (u,x) = s \quad \mapsto \quad Lf(s) = u\, f(x).$$

Thus, finally we see that the "reward Poisson process" is factually identical (in regard to its FDD) with a Poisson random measure ξ on the product space, $Mf \stackrel{\mathcal{D}}{=} \xi Tf$:

$$\mathsf{E}\, e^{-Mf} = \exp\left\{ -\int_S \left(1 - e^{-Lf(s)}\right) \lambda(ds) \right\} = \mathsf{E}\, e^{-\xi\, Lf}.$$

Example 5.5 Let us examine one more time formula (5.6)

$$M_t = \sum_{n=1}^{N_t} R_n$$

An alternative name for M_t is a "**compound Poisson process**".

1. Let R_n be i.i.d. Bernoulli with $\mathsf{P}(R_n = 1) = p$. That is, with probability p a signal is recorded (or taken, or colored) while with probability $1-p$ the signal is neglected (or left out, or whitened out). Then M_t is a Poisson process with intensity $p\lambda$, a "thinned" Poisson process.

   In other words, if X is a binomial r.v. with parameters n and p, $bin(n,p)$ and then n is "randomized" by a Poisson random variable N independent of X, then $b(N,p)$ is Poisson.

2. The remaining process with "rewards" $1-R_n$ is also Poisson with intensity $(1-p)\lambda$. Further, both processes are independent.

   This property can be generalized to a finite decomposition of the unit (as in $1 = R_n + (1-R_n)$). To wit, let $R = \sum_{j=1}^d R_j$, where $R_j R_k = 0$ for $j \neq k$, and R_j is Bernoulli with parameter p_j. We may think of a wheel-of-fortune like spinner, with slices marked by numbers or colors $j = 1, ..., d$. Let (R_{nj}) be independent copies of (R_j). When a signal Y_n of a Poisson process is recorded, the spinner is spun and the signal is marked by the outcome shown, one between 1 and d. We claim that the resulting process

   $$M_j f = \sum_n R_{nj}\, f(Y_n)$$

   are independent Poisson with parameters $p_j \lambda$.

5.5 A few constructions of Poisson random measure

5.5.1 adding new atoms

Using the setting of formula (5.2), we look at the reward Poisson process in Example 5.5 as an extension of an already defined Poisson random measure on $(S, \mathcal{S}, \lambda)$ to a product space $S \times T$, where $(T, \mathcal{T}, \mu)$ is a probability space, and τ_n are i.i.d. random elements in T with probability distribution μ:

$$M = \sum_n \delta_{(Y_n, \tau_n)}.$$

We may think of τ_n's as "marks", that are not necessarily numbers. That is why this Poisson random measure (as we will see) is often called a **marked Poisson process**.

In the integral form, for a function $F(t, y) = \alpha(t) g(y)$ with separable variables

$$MF = \sum_n \alpha(\tau_n)\, f(Y_n).$$

So, the reward Poisson measure is just the special case of the marked Poisson measure, $R_n = \alpha(\tau_n)$. For a general F,

$$MF = \sum_n F(\tau_n, Y_n).$$

It remains to verify that M is a Poisson random measure.

$$\mathsf{E}\, e^{-MF} = \mathsf{E}_N \prod_n \mathsf{E}_\tau e^{-F(\tau, Y_n)}.$$

Denote

$$g(y) = -\ln \mathsf{E}\, F(\tau, y).$$

So

$$\mathsf{E}\, e^{-MF} = \mathsf{E} \exp\left\{ -\sum_n g(Y_n) \right\} = \mathsf{E}\, e^{-Ng} = \exp\left\{ -\int_S \left(1 - e^{-g(y)}\right) \lambda(dy) \right\}$$
$$= \exp\left\{ -\int_S \int_T \left(1 - e^{-F(t,y)}\right) \mu(dt)\, \lambda(dy) \right\}.$$

Hence M is a Poisson random measure on $\mathcal{S} \times \mathcal{T}$ with intensity $\lambda \otimes \mu$.

Example: A Poisson random measure in $\mathbb{R}^d$. We shall use spherical coordinates (when $d = 2$ they are called polar coordinates) $(r, \boldsymbol{t})$ where $r \geq 0$ and $\boldsymbol{t}$ is a point from the $(d-1)$-sphere $T = S_{d-1}$ (e.g., the unit circle when $d = 2$, the two-dimensional unit sphere when $d = 3$, etc.). Let Y_n be signals of a unit intensity Poisson process on $[0, \infty)$ and let independent $\boldsymbol{\tau}_n$, also independent of N, be uniformly distributed on S_{d-1}. For a cone C described by $r \leq a$, $\boldsymbol{t} \in B$, where B is a Borel subset of S_{d-1},

$$\mathbb{1}_C(r, \boldsymbol{t}) = \mathbb{1}_{[0,a]}(r)\, \mathbb{1}_B(\boldsymbol{t}),$$

the Poisson random variable MC has the expectation $a \cdot |B| = Leb_d(C)$ (where $|B|$ denotes the normalized Lebesgue measure on the sphere). So, the intensity of M is the Lebesgue measure in $\mathbb{R}^d$.

5.5.2 gluing the pieces

The last case of Example 5.5 can be generalized (and simplified at the same time) as follows. Let $(S, \mathcal{S}, \mu)$ be a probability space and let $X : \Omega \to S$ be a random element with distribution μ. That is, $\mathsf{P}(X \in A) = \mu A$ for $A \in \mathcal{S}$. Let X_n be its independent copies and let N be a unit intensity Poisson process with signals (Y_n), independent of (X_n). Define

$$MA = \sum_{n=1}^{N_1} 1\!\!1_A(X_n) \quad \text{or} \quad Mf = \sum_{n=1}^{N_1} f(X_n) = \sum_{n=1}^{\infty} f(X_n) 1\!\!1_{\{N_1 \geq n\}} = \sum_{n=1}^{\infty} f(X_n) 1\!\!1_{[0,1]}(Y_n),$$

where $A \in \mathcal{S}$ or $f \geq 0$ is a Borel measurable function on S. For simplicity, denote $I = 1\!\!1_{[0,1]}$. Then

$$\mathsf{E}\, e^{-Mf} = \mathsf{E}_N \prod_n \mathsf{E}\, e^{-f(X_n) I(Y_n)}.$$

Wit the help of the function

$$g(y) = -\ln \mathsf{E}\, e^{-f(X) I(y)}.$$

the latter formula reads

$$\mathsf{E}\, e^{-Mf} = \mathsf{E} \prod_n e^{-g(Y_n)} = \mathsf{E}\, e^{-\sum_n g(Y_n)} = \mathsf{E}\, e^{-Ng} = \exp\left\{ -\int_0^\infty \left(1 - e^{-g(y)}\right) dy \right\}.$$

Removing g and bringing up $I = 1\!\!1_{[0,1]}$, the last expression

$$= \exp\left\{ -\int_0^\infty \left(1 - \mathsf{E}\, e^{-f(X) I(y)}\right) dy \right\} = \exp\left\{ -\int_0^1 \left(1 - \mathsf{E}\, e^{-f(X)}\right) dy \right\}$$
$$= \exp\left\{ -\int_S \left(1 - e^{-f(s)}\right) \mu(ds) \right\}.$$

In other words, M is a Poison measure on $(S, \mathcal{S}, \mu)$.

Now let $(S, \mathcal{S}, \lambda)$ be an infinite but σ-finite measure space. Assume that is continuous (atomless). Let $S = \bigcup S_k$, where $S_k \in \mathcal{S}$ are probability spaces. Create independent Poisson meaures M_k on $(S_k, \mathcal{S}_k, \lambda_k)$, where $\mathcal{S}_k = \mathcal{S} \cap S_k$ and $\lambda_k = \lambda|_{S_k}$ according to the previous construction. Finally, there comes the Poisson random measure with intensity λ:

$$MA = \sum_k M_k(A \cap S_k).$$

5.5.3 using a density of a random element

Let $(S, \mathcal{S}, \lambda)$ be an atomless infinite σ-finite measure space and τ be a random element in S whose distribution is absolutely continuous with respect to λ and its density $p(s)$ is strictly positive. Let τ_n be independent copies of τ. Let N_t be a unit intensity Poisson process with signals Y_n, independent of (τ_n). Finally, let A be a Borel set on $[0, \infty)$ with Lebesgue measure 1. Put $\alpha = 1\!\!1_A$ and define the integral process for $f \in L^0_+(S)$ by the formula

$$\xi f = \sum_n \alpha\left(Y_n\, p(\tau_n)\right)\, f(\tau_n).$$

Theorem 5.6 *ξ is a Poisson measure on $\mathcal{S}$ with intensity λ.*

Proof. Let us compute the Laplace transform

$$\mathsf{E}\, e^{-\xi f} = \mathsf{E}_N \prod_n \mathsf{E}_\tau\, e^{-\alpha(Y_n\, p(\tau) f(\tau)}.$$

With the help of the function

$$g(y) = -\ln \mathsf{E}\, e^{-\alpha(y\, p(\tau))\, f(\tau)}$$

and the identity $1 - e^{-\alpha c} = (1 - e^{-c})\alpha$, where $\alpha \in \{0, 1\}$, we rewrite the latter expression as

$$\begin{aligned}
\mathsf{E} \prod_n e^{-g(Y_n)} = \mathsf{E}\, e^{-Ng} &= \exp\left\{ - \int_0^\infty \left(1 - e^{-g(y)}\right) dy \right\} \\
&= \exp\left\{ - \int_0^\infty \mathsf{E}\left(1 - e^{-\alpha(y\, p(\tau)) f(\tau)}\right) dy \right\} \\
&= \exp\left\{ - \int_0^\infty \int_S \left(1 - e^{-f(s)}\right) \alpha(y\, p(s))\, p(s)\, \lambda(ds)\, dy \right\}
\end{aligned}$$

Using Fubini's Theorem, in the "dy-integral" we substitute $x = y\, p(s)$, so $dx = p(s)\, dy$, and since $\int_0^\infty \alpha(x)\, dx = |A| = 1$, the latter quantity becomes

$$\exp\left\{ - \int_S \left(1 - e^{-f(s)}\right) \lambda(ds) \right\}.$$

That is, ξ is a Poisson measure with intensity λ. ■

Example. Let us construct a planar Poisson measure, for which we need a strictly positive densitiy. E.g., we may pick the Gaussian density, for $u = (u_1, u_2)$,

$$p(u) = \frac{1}{2\pi}\, e^{-(u_1^2 + u_2^2)/2}, \quad q(u) = \frac{1}{p(u)} = 2\pi\, e^{||u||^2/2}$$

so $\tau_n = (\gamma_{n1}, \gamma_{n2})$, where γ_{nk} are independent N(0,1) random variables. Also, we choose $A = [0,1]$. Let (Y_n) form a unit intensity Poisson process N_t, independent of (τ_n). We observe that $V_n = \|\tau_n\|^2/2$ are exponential r.vs. with unit intensity. So we obtain

$$\xi f = \sum_n \mathbb{I}_{\{Y_n \le q(\tau_n)\}} \, f(\tau_n).$$

5.6 Exercises

1. A Poisson random measure ξ is countably additive in every L^p, $0 < p < \infty$

2. Show that the only solution of the functional Cauchy equation,

   $$f(x+y) = f(x) + f(y), \quad x, y \in \mathbb{R},$$

   in the class of real continuous functions on $\mathbb{R}$ is the linear function $f(x) = ax$. Equivalently, within this class, the only solution of the functional equation

   $$g(s+t) = g(s)g(t), \quad s, t \ge 0$$

   is the exponential $g(s) = e^{as}$. Hence the only continuous distribution that enjoys the lack of memory property is the exponential distribution.

3. Why do the age time A_t and the excess time W_t have the same distribution? Is this a property of Poisson process or any renewal process?

   Find the probability distribution of $U_{N_t+1} = A_t + W_t$ for the Poisson process.

4. A Poisson process N_t on the positive half-line entails immediately a finite additive set function $N(a,b] = N_b - N_a$ on the field spanned by the intervals $(a,b]$. Since its control measure is the Lebesgue measure times λ, show in few lines how this additive set function extends to a true random measure on Bore sets. Note: it is easier to construct the Poisson integral Nf first! Then the random measure is simply $N\mathbb{I}_A$. Clean details need to be written down.

5. In Corollary 5.4.4 a "paradox" is shown. Say, Auburn Transit buses arrive at a bus stop according to a Poisson distribution, say, with the average interarrival time 20 min. You come to the bus stop, there is no bus yet so you wait. How long, in average? 10 minutes, 15, 20? Yes, 20 is the answer. Also, the time between the moment of departure of the last bus before your arrival and the moment of your forthcoming ride would be... yes, 40 minutes, in average.

   It's a paradox, isn't it? Or, perhaps not...

   Similarly, if there are two lines to a service, say to a cash register or a ticket booth at a rock concert, and you choose one line, the other will move faster. So you'll change the line. But then the line that you just left will be mowing faster. That's the fact and it has a logical explanation (the same phenomenon as in waiting for a bus).

Explain!

6. Show that the split Poisson processes M_j in Example 5.5.2 are independent Poisson with parameters $p_j\lambda$. Hint: for a fixed f show that, for $M_j = M_j f$,

$$\mathsf{E}\, e^{-\sum_j c_j M_j} = \prod_j \mathsf{E}\, e^{-c_j M_j}$$

Then, for finitely many f_k with disjoint supports (so Nf_k are independent):

$$\mathsf{E}\, e^{-\sum_j \sum_k c_j M_j f_k} = \prod_j \prod_k \mathsf{E}\, e^{-c_j M_j f_k}.$$

Argue that these relations prove the statement.

7. Let $(S, \mathcal{S}, \lambda)$ be an atomless (continuous) space. Let $a \le \lambda(S)$. Then there exists $A \in \mathcal{S}$ such that $\lambda A = c$. In particular, an infinite σ-finite measure space enjoys a partition into the union of probability spaces.

8. Let $(S, \mathcal{S}, \mu)$ be a probability space. Consider the standard probability space $(\Omega, \mathcal{F}, \mathsf{P})$ as the unit interval with Borel sets and the Lebesgue measure. Argue that there exists a measurable mapping $X : \Omega \to S$ such that $\mathsf{P}(X \in A) = \mu A$ for every $A \in \mathcal{S}$.

5.7 Non-positive awards

Let (ζ_n) be i.i.d. and copies of a ζ with distribution μ, independent of a Poisson process N_t with signals (Y_n) and intensity λ. The integral

$$Xf = \sum_n \zeta_n \, f(Y_n)$$

is well defined, e.g., when f has a bounded support, e.g., for $f = 1\!\!1_{(a,b]}$ and linear combinations of such functions, say,

$$f = a_0 1\!\!1_{\{0\}} + \sum_{k=1}^n a_k 1\!\!1_{[t_{k-1},t_k]} = \sum_k af_k, \quad 0 = t_0 < t_1 \cdots < t_n = t$$

Then its ch.f. equals

$$\mathsf{E}\, e^{iXf} = \exp\left\{ -\lambda \int_0^\infty \mathsf{E}\left(1 - e^{i\zeta f(t)}\right) dt \right\} = \exp\left\{ -\lambda \int_0^\infty \int_{\mathbb{R}} \left(1 - e^{ixf(t)}\right) \mu(dx)\, dt \right\}.$$

For the specific simple function listed above, it equals

$$\exp\left\{ -\lambda \sum_k (t_k - t_{k-1}) \int_{\mathbb{R}} \left(1 - e^{ia_k f(t)}\right) \mu(dx) \right\} = \prod_k \mathsf{E}\, e^{ia_k f_k},$$

which shows that X is an independently scattered random measures with stationary increments. Therefore, its FDD are fully described by one dimensional distributions, for $f = 1\!\!1_{[0,t]}$

$$\mathsf{E}\, e^{iaX_t} = \exp\left\{ -\lambda t \int_{\mathbb{R}} \left(1 - e^{iax}\right) \mu(dx) \right\}$$

By the "gluing technique", the introduced concept of a random measure can be extended even to infinite but σ-finite measure μ on $\mathbb{R}$, restricted by the existence of the integral that appears in the characteristic function. Clearly, a sufficient condition is

$$\int_{|x|\le 1} x^2 \, \mu(dx) + \mu([-1,1]^c) = \int_{\mathbb{R}} 1 \wedge x^2 \, \mu(dx) < \infty.$$

The finiteness of the first term on the left is obviously necessary. It can be shown that the second term must be finite necessarily but it requires some tedious reasoning, and we will not show it.

However, we will show details in the symmetric case.

Let's begin with the simplest case of symmetric ± 1-valued rewards. Let ξ be a Poisson random measure counting random points Y_n in $(S, \mathcal{S}, \lambda)$, and let ε_n be a Rademacher sequence independent of ξ (and of (Y_n)). Define

$$\tilde{\xi} f = \sum_n \varepsilon_n \, f(Y_n).$$

By Fubini' theorem and properties of Rademacher series:

$$\sum_n \varepsilon_n a_n \text{ converges} \iff \sum_n a_n^2 < \infty,$$

the series converges in probability, or, equivalently a.s., if and only if

$$\xi f^2 = \sum_n |f(Y_n)|^2 < \infty$$

and this happens if and only if

$$\int_S \left(1 - e^{-f^2(s)}\right) \lambda(ds) < \infty.$$

Observe that we do not need to restrict ourselves to nonnegative functions (or differences of such). Instead of Laplace transforms we rather use the characteristic functions. Because of symmetry, the ch.f. is real and equals

$$\mathsf{E}\, e^{i\tilde{\xi} f} = \exp\left\{ -\int_S \left(1 - \cos f(s)\right) \lambda(ds) \right\}.$$

We shall call $\tilde{\xi}$ a **symmetrized Poisson random measure** (SPRM) with intensity λ. The above existence condition can be replaced by a more elegant condition:

$$\int_S 1 \wedge f^2(s)\, \lambda(ds) < \infty. \tag{5.8}$$

Now, we will examine some of previously discussed variants in this new context.

Symmetric rewards. Let R_n be independent copies of a symmetric r.v. R, i.e. $R \stackrel{\mathcal{D}}{=} -R$. Therefore, $R \stackrel{\mathcal{D}}{=} \varepsilon\, R$, where ε and R are independent. Assume also that (R_n) is independent of the Poisson measure ξ. As before, put

$$Mf = \sum_n R_n\, f(Y_n),$$

where the series converges if and only if

$$\int_S \mathsf{E}\, 1 \wedge R^2 f^2(s)\, ds < \infty.$$

The ch.f. $\mathsf{E}\, e^{iMf}$ equals

$$\exp\left\{ -\int_S \mathsf{E}\left(1 - \cos R\, f(s)\right) \lambda(ds) \right\} = \exp\left\{ -\int_S \int_{\mathbb{R}} \left(1 - \cos x\, f(s)\right) \mu(dx)\, \lambda(ds) \right\} \tag{5.9}$$

So M is a SPRM on $S \times \mathbb{R}$ with intensity $\lambda \otimes \mu$, where $\mu = \mathcal{L}(R)$. In fact

$$Mf \stackrel{\mathcal{D}}{=} \xi L f, \quad \text{where } L : S \times \mathbb{R} \to \mathbb{R}, \quad L(s,x) = x\, f(s). \tag{5.10}$$

We observe that the restriction to a probability or even finite measure μ is not necessary. A potential extension is controlled by condition (5.8). Consider a standard Poisson process on $S = \mathbb{R}_+$ with unit intensity. Let μ be a measure on $\mathbb{R}$ whose properties need to be found and let ξ be a PRM on $\mathbb{R} \times [0, \infty)$ with intensity $\mu \otimes Leb$.

Lemma 5.7 *The inner integral in* (5.9) *is finite over the class of functions f that contains indicators iff*

$$\int_{\mathbb{R}} 1 \wedge x^2 \, \mu(dx) < \infty \tag{5.11}$$

Proof. The statement follows from (5.8) and the inequalities

$$(a^2 \wedge 1)\,(1 \wedge x^2) \le 1 \wedge (ax)^2 \le (a^2 \vee 1)\,(1 \wedge x^2)$$

■

Definition. A Borel measure μ on $\mathbb{R}$ is called a **Lévy measure** if (5.11) holds.

Thus, a PRM ξ on $\mathbb{R} \times \mathbb{R}_+$ with intensity $d\mu \otimes dt$, where μ is a Lévy measure, entails a process Mf by (5.10) with the ch.f.

$$\mathsf{E}\, e^{iMf} = \exp\left\{ - \int_0^\infty \int_{\mathbb{R}} \Big(1 - \cos x\, f(s)\Big)\, \mu(dx)\, ds \right\}$$

In particular, for functions $f_1, \ldots, f_n$ with disjoint supports, $Mf_1, \ldots, Mf_n$ are independent. Hence, if $f_k = \mathbb{1}_{(a+t_{k-1}, a+t_k]}$, $k = 1, \ldots, n$, where $a \ge 0$ and $t_0 = 0 < t_1 \cdots < t_n$,

$$\mathsf{E}\, \exp\left\{ -i \sum_k c_k M f_k \right\} = \prod_k \exp\left\{ -(t_k - t_{k-1}) \int_{\mathbb{R}} (1 - \cos x)\, \mu(dx) \right\}.$$

In other words, the stochastic process $M_t = M\mathbb{1}_{[0,t]}$ has independent and stationary increments.

5.8 $S S\alpha$ - symmetric α-stable processes

For $\alpha > 0$ define the symmetric[3] measure μ_α by the formula

$$\mu_\alpha[x, \infty) = \frac{1}{x^\alpha}, \quad x > 0$$

Equivalently, μ_α has the density

$$g_\alpha(x) = \frac{\alpha}{|x|^{\alpha+1}}, \quad x \neq 0.$$

Lemma 5.8 *μ_α is a Lévy measure if and only if $\alpha < 2$.* ■

Put $S = \mathbb{R} \setminus \{0\}$. Consider the Poisson measure M on $S \times \mathbb{R}_+$. That is,

$$\mathsf{E}\, e^{iMf} = \exp\left\{ - \int_0^\infty \int_S \Big(1 - \cos x f(t)\Big) \frac{\alpha}{|x|^{\alpha+1}}\, dx\, dt \right\}$$

By symmetry we may consider the integral for $x > 0$, and then change[4] the variable $x|f(t)| \mapsto x$, so the ch.f. equals

$$\exp\left\{ -c_\alpha \int_0^\infty |f(t)|^\alpha\, dt \right\}, \quad \text{where } c_\alpha = 2\alpha \int_0^\infty (1 - \cos x) \frac{dx}{x^{\alpha+1}}.$$

In particular, taking $f = c_\alpha^{-1/\alpha} \mathbb{I}_{(a,b]}$,

$$\mathsf{E}\, e^{it(M_b - M_a)} = e^{-(b-a)|t|^\alpha}.$$

Definition A random variable X with the ch.f.

$$\mathsf{E}\, e^{itX} = e^{-a|t|^\alpha}$$

is called **symmetric $\boldsymbol{\alpha}$-stable**, or $S S\alpha$ in short. Mf, MA, M_t are called then $S S\alpha$ integral, measure, process - respectively.

Also, taking $f_k = c_\alpha^{-1/\alpha} \mathbb{I}_{A_k}$, where A_k are disjoint of unit Lebesgue measure

$$X_k = Mf_k,\ k \in \mathbb{N} \ \Rightarrow\ f_k \text{ are i.i.d. and } \sum_k a_k f_k \stackrel{\mathcal{D}}{=} \left(\sum_k |a_k|^\alpha \right)^{1/\alpha} X_1.$$

Example 5.9 (Le Page representation) Let Y_n be Poisson points with unit intensity, τ_n be i.i.d. uniform on $[0, 1]$, ε_n be Rademacher r.vs., and the three sequences be independent. Let $\alpha \in (0, 2)$. Then

$$Mf = \sum_n \varepsilon_n \frac{f(\tau_n)}{Y_n^{1/\alpha}}$$

---

[3] $\mu(A) = \mu(-A)$

[4] the cosine is an even function

is a $S S\alpha$ process/integral/measure.

Indeed, even in a more general case

$$Mf = \sum_n \varepsilon_n \, f(\tau_n) \, \phi(Y_n) \ \text{ converges iff } \ \sum_n f^2(\tau_n) \, \phi^2(Y_n) < \infty \text{ a.s.}$$

and the necessary and sufficient condition is

$$\int_0^\infty \int_0^1 1 \wedge f^2(t) \, \phi^2(y) \, dt \, dy < \infty.$$

The ch.f. equals

$$\mathsf{E} \, e^{iMf} = \exp\left\{ - \int_0^\infty \int_0^1 \Big(1 - \cos f(t) \, \phi(y)\Big) \, dt \, dy \right\}.$$

Returning to the original function $\phi(y) = y^{-1/\alpha}$, using Fubini's theorem and the substitution $y = |f(t)|^\alpha x^{-\alpha}$, we obtain

$$\mathsf{E} \, e^{iMf} = \exp\left\{ - \int_0^1 \int_0^\infty \Big(1 - \cos f(t) \, y^{-1/\alpha}\Big) \, dt \, dy \right\} = \exp\left\{ -c_\alpha \int_0^1 |f(t)|^\alpha \, dt \right\}$$

with

$$c_\alpha = \alpha \int_0^\infty \frac{1 - \cos x}{x^{1+\alpha}} \, dx.$$

5.9 Exercises

1. Let N_t be a standard Poisson process with arrivals Y_n and A be a bounded Borel set. Show that $\mathsf{P}(Y_n \in A \text{ for infinitely many } n) = 0$. Deduce then that $f(Y_n) = 0$ eventually with probability 1 when f has a bounded support.

2. Verify that the symmetric measure μ on $\mathbb{R} \setminus 0$ with the tail $\mu(x, \infty) = x^{-\alpha}$ is a Lévy measure iff $\alpha \in (0, 2)$.

3. Show that the p the moment $\mathsf{E}\,|X|^p$ of a $SS\alpha$ r.v. is finite iff $p < \alpha$.

4. Let X_k be i.i.d. $SS\alpha$. Let $p \in [0, \alpha)$. Show that

$$\sum_k a_k X_k \text{ converges in } L^0 \text{ and a.s.} \iff \sum_k |a_k|^\alpha < \infty.$$

   In particular, for $p \in (0, \alpha)$

$$\left\| \sum_k a_k X_k \right\|_p = \|a\|_\alpha,$$

   i.e., every F-space (for $\alpha < 1$) or Banach space (for $\alpha \in (1, 2)$) contains a subspace isometric with ℓ^α. That it is true also for $\alpha = 2$ was proved previously (in lieu of stable we can use Rademacher or Gaussian i.i.d. r.vs.)

5. (added here although it belongs to the previous topic). Consider the paraboloid of revolution given by the equation $z = x^2 + y^2$. Project the disk of radius r that lies on the xy-plane to the paraboloid's surface, obtaining a set A. Let Y_n be Poisson points on the paraboloid, controlled by the surface area. Find the probability that A has no Poisson points.

   More difficult: Construct Poisson points on the paraboloid.

   More difficult: Let S be a smooth connected unbounded surface, say, given by a parametric equation $\boldsymbol{r} = \boldsymbol{r}(u, v)$, where $(u, v) \in D$, where D is an open domain in $\mathbb{R}^2$, and $\boldsymbol{r} \in C^1$. Construct Poisson points on S.

   (Hint: Show that w.l.o.g $D = \mathbb{R}^2$. Construct Poisson points on the plane. Carry them by some mapping into S. The Jacobian will be involved.).

6 Infinitely divisible distributions

6.1 Preliminaria

Recall that the ch.funs. $\varphi_1, \ldots, \varphi_n$ of independent r.vs $X_1, \ldots, X_n$ satisfy the formula

$$\mathsf{E}\,\exp\left\{ i \sum_k t_k X_k \right\} = \varphi_1(t_1)\cdots\varphi(t_n)$$

(which is also sufficient for independence). For just two independent r.vs. X_1, X_2 that make the sum $X = X_1 + X_2$ we have, in terms of their probability laws μ_1, μ_2, and μ:

$$\mu A = \int_{\mathbb{R}} \mu_1(A - x)\,\mu_2(dx) = \int_{\mathbb{R}} \mu_2(A - x)\,\mu_1(dx),$$

which can be equivalently stated (**Exercise: Prove it**) in terms of the integrals

$$\mu F = \int_{\mathbb{R}} F(x)\,\mu(dx) = \mathsf{E}\,F(X) = \int_{\mathbb{R}} \mu_1 F(\cdot + y)\,\mu_2(dy) = \int_{\mathbb{R}} \mu_2 F(\cdot + x)\,\mu_1(dx).$$

The "dot" inside indicates the integration along a hidden variable, e.g.:

$$\mu_1 F(\cdot + y) = \int_{\mathbb{R}} F(x + y)\,\mu_1(dx).$$

When both partial measures are absolutely continuous, and f_1, f_2 denote their densities, then the density of the sum equals

$$f_{X_1+X_2}(z) = \int_{\mathbb{R}} f_1(z - y)\,f_2(y)\,dy = \int_{\mathbb{R}} f_2(z - x)\,f_1(x)\,dx.$$

The operation produces a new measure or a new density which is called the **convolution** of measures or densities, and denoted by $\mu_1 * \mu_2$ or $f_1 * f_2$. The extension to any finite number of terms follows immediately. If $\mu_1 = \cdots = \mu_n = \mu$ we may write $\mu^{*n} = \mu_1 * \cdots * \mu_n$. In the language of random variables the convolution n-th power is the probability law of the sum $X_1 + \cdots + X_n$ of i.i.d. r.vs. with $\mathcal{L}(X_1) = \mu$.

Now, let us look at this pattern from an opposite point of view. Let $\mathcal{L}(X) = \nu$ and $\hat{\nu} = \psi$. Is it possible to write

$$X = X_1 + \cdots + X_n, \quad \text{where } X_k \text{ are i.i.d rvs.?}$$

Equivalently, does there exist a ch.f. ϕ such that $\psi = \phi^n$? In other words, is $\psi^{1/n}$ a ch.f.? [5]. If the measure is supported by the positive halfline, we may use the Laplace transform in lie of the ch.f. So, if L is a Laplace transform of a probability measure, is $L^{1/n}$ such, too?

First, we will find a counterexample. Suppose that ψ is ID. Then so is $\overline{\psi}$ and consequently, $|\psi|^2$ is ID. The limit of ch.f.

$$\phi = \lim_n |\psi|^{2/n}$$

[5] It doesn't matter which root of n complex roots we consider; for simplicity we choose the principal root

takes only two values, 0 and 1. Since $|\psi| > 0$ on a neighborhood of 0, hence $\phi = 1$ on that neighborhood, and as a continuous function must equal 1 everywhere. Hence $\psi \neq 0$ everywhere. Let us repeat:

An ID ch.f. never vanishes.

Thus, as an example of a non-ID distribution it suffices to take one with a ch.f. vanishing at some point. For example, consider the uniform r.v. on $[0, 1]$. Its ch.f.

$$\frac{e^{it} - 1}{it}$$

vanishes for $t = 2n\pi$, $n \in \mathbb{Z}$.

Also the "tent function" that is a ch.f. by Polya criterion is not ID.

6.2 A few theorems

We note that the class of ID distributions is closed under convolution (Cf. an exercise).

Theorem 6.1 *The class of ID probablity distributions is closed under the weak limit.*

Proof. Let φ_k be ID and $\varphi_k \to \varphi_0$. Let $n \in \mathbb{N}$. Then $|\varphi_k|^2$ are real ch.funs. for $k = 0, 1, 2, ..$ and ID for $k = 1, 2, ...$ So, in the latter case $|\varphi_k|^{2/n}$ are ch.funs. But

$$|\varphi_0|^{2/n} = \lim_k |\varphi_k|^{2/n},$$

and is continuous at 0, so by the continuity theorem $|\varphi_0|^{2/n}$ is a ch.fun. for every $n \in \mathbb{N}$. That is, $|\varphi_0|^2$ is ID. As such, it has no zeros, but then φ_0 has no zeros, and we can thus define its n-th root as well as the n-th roots of φ_k's:

$$\varphi_0^{1/n}(t) = \exp\left\{\frac{1}{n}\ln\varphi_0(t)\right\} = \lim_k \exp\left\{\frac{1}{n}\ln\varphi_k(t)\right\} = \lim_k \varphi_k^{1/n}, \quad n \in \mathbb{N}.$$

That is, $\varphi_0^{1/n}$ is continuous at 0 and is the limit of ch.funs., so itself it is a ch.fun. ■

Corollary 6.2 *Let ϕ be a ID ch.f. and $\alpha > 0$. Then ϕ^α is ID.*

Proof. For a rational α the statement is obvious. Then we pass to any real limit. ■

If α is irrational then the latter property is hard, if not impossible, to express in the language of real variables or probability distribution

6.3 A side trip: decomposable distributions

A probability distribution μ is called **decomposable** if there are nontrivial probability distributions μ_1, μ_2 such that $\mu = \mu_1 * \mu_2$. In the language of random variables, X is decomposable, if there exist independent X_1 and X_2, none degenerate, such that

$$X = X_1 + X_2$$

We exclude degenerate r.vs. from the class of decomposable ones to avoid the triviality:

$$X = (X - a) + a.$$

Note that the decomposability may have a finite depth, that is, some of the summands may be non-decomposable. Even if the law can be split into an arbitrary finite number of parts, these might not be identical.

Call a r.v. X **self-similar** or **c-decomposable**, if

$$X \stackrel{\mathcal{D}}{=} cX + R,$$

where X and R are independent, and R is non-degenerate. It follows by iteration, that X can be decomposed into a sum of any length that consists of independent summands:

$$X \stackrel{\mathcal{D}}{=} cX + R \stackrel{\mathcal{D}}{=} c(cX + R') + R = c^2 X + cR' + R \stackrel{\mathcal{D}}{=} \cdots \stackrel{\mathcal{D}}{=} c^n X + \sum_{k=1}^{n} c^{k-1} R_k,$$

where all r.vs on the right side are independent and R_k's are copies of each other. In particular, if X is c-decomposable, then it is c^n-decomposable for every $n \in \mathbb{N}$. We observe that this an attribute of the probability distribution or its transform rather than of the random variable itself. That is, the property reads, for $\mu = \mathcal{L}(X)$, with $\mu_c = \mathcal{L}(cX)$ and $\varphi = \hat{\mu}$:

$$\exists\, \nu \quad \mu = \mu_c * \nu \quad or \quad \frac{\varphi(t)}{\varphi(ct)} \text{ is a ch.fun.}$$

We note that $S S\alpha$ and Gaussian distributions are c-decomposable for every $c \in (-1, 1)$.

The uniform random variable V on $[-1, 1]$ is $1/k$-decomposable, for every natural number k. Indeed, since its Fourier transform is $\sin t/t$, then

$$\frac{\sin kt}{k \sin t} = \Big(1 - \frac{1}{k}\Big) \frac{\sin(k-1)t}{(k-1)\sin t}\, \cos t + \frac{1}{k} \cos(k-1)t,$$

i.e.,

$$V \stackrel{\mathcal{D}}{=} \frac{1}{k} V + R_k, \qquad R_k \stackrel{\mathcal{D}}{=} D_{1-1/k}\Big(R_{k-1} + \varepsilon_{k-1}\Big) + \Big(1 - D_{1/k}\Big)(k-1)\varepsilon_k,$$

where $D_{1/k}$ are $(1/k)$-Bernoulli, ε_k are Rademacher variables (i.e., $(1+\varepsilon_k)/2$ are $1/2$-Bernoulli), and all sequences are independent. Therefore, all uniform random variables are $1/k$-decomposable, being affine transformations of each other, $U_{[a,b]} = \frac{b-a}{2} U_{[-1,1]} + \frac{b+a}{2}$. In general, if a variable with possible negative values (or bounded away from 0) has the property $Y \stackrel{\mathcal{D}}{=} cY + R$, where the residue R has an atom at the minimum, $\mathsf{P}(R = m(R)) > 0$, then $Y - m(Y)$ is c-decomposable for some c. In the uniform case, there is a simpler direct argument.

Proposition 6.3 *A uniform random variable U on $[0,1]$ belongs to the class $\mathfrak{S}(c)$ if and only if $c = p = m^{-1}$, for some natural number m. In this case*

$$U \stackrel{\mathcal{D}}{=} \frac{1}{m} U + D_m \cdot Z_m$$

where D_m denotes a $(1 - 1/m)$-Bernoulli r.v. and Z_m has the discrete uniform distribution

$$\frac{1}{m-1} \sum_{k=1}^{m-1} \delta_{k/m}$$

on $\{ k/m : k = 1, \ldots, m-1 \}$. The three variables U, D_m, Z_m are independent.

Proof. Consider the binary series representation of U:

$$U = \sum_{n=0}^{\infty} \frac{D_n}{2^{n+1}} \quad \text{i.e.} \quad U \stackrel{\mathcal{D}}{=} \frac{1}{2} U + D \frac{1}{2},$$

where D_n are i.i.d. (1/2)-Bernoulli. Other admissible parameters c come from the equation

$$M(s) = L(s)/L(cs) = a \frac{1 - e^{-s}}{1 - e^{-cs}} = p + (1-p) H(s).$$

Clearly, $c = p$. Hence, the sought-for Laplace transform $H(s)$ would be equal to

$$\frac{p}{1-p} \frac{e^{-ps} - e^{-s}}{1 - e^{-ps}}.$$

Then, denoting the Dirac delta measure at point c by δ_c, we have

$$\frac{e^{-ps}}{1 - e^{-ps}} = \mathcal{L}\left(\sum_{k=1}^{\infty} \delta_{pk} \right), \qquad \frac{e^{-s}}{1 - e^{-ps}} = \mathcal{L}\left(\sum_{k=0}^{\infty} \delta_{pk+1} \right)$$

Thus, the signed measure whose $H(s)$ is the Laplace transform is nonnegative,

$$\mathcal{L}^{-1}(H) = \frac{p}{1-p} \left(\sum_{k=1}^{\infty} \delta_{pk} - \sum_{k=0}^{\infty} \delta_{pk+1} \right),$$

if and only if $p = 1/m$, for some $m = 1, 2, \ldots$. Thus, $H(s)$ is the Laplace transform of the uniform discrete probability on $\{ k/m,\ k = 1, \ldots, m-1 \}$. ■

Example 6.4 The replacement of the parameter $c = 1/2$ by $1/3$ yields a c-decomposable variable with the singular Cantor-Lebesgue distribution:

$$V = \sum_{n=0}^{\infty} \frac{2D_n}{3^{n+1}}, \quad \text{i.e.} \quad V \stackrel{\mathcal{D}}{=} \frac{1}{3} V + D \frac{2}{3}.$$

Whether its decomposability semigroup extends beyond $\{ 1/3^n \}$ is yet to be determined.

6.4 ID of Poisson type

The inspiration: Section 5.4 in Lukacs' book.

Let N_λ be a Poisson r.v. with intensity λ and $a > 0$. Then the ch.fun. of aN is $\exp\left\{ \lambda e^{ita} - 1 \right\}$. A Poisson integral of a simple function is said to be of **Poisson type** in the literature. In other words, a r.v. is of Poisson type, if for a finite choice of parameters $a_k, \lambda_k > 0$, and independent Poisson r.vs. N_{λ_k}

$$X = \sum_k a_k N_{\lambda_k} = Ng = \int_S g dN,$$

where $g = \sum_k a_k \mathbb{1}_{A_k}$ and A_k are disjoint with $\lambda A_k = \lambda_k$. We carry the name to probability distributions and ch.funs. So, a ch.fun. ψ is of Poisson type iff

$$\psi(t) = \prod_k \exp\left\{ a_k \left(e^{ita_k} - 1\right) \right\} \; = \; \exp\left\{ p \sum_k p_k \left(e^{ita_k} - 1\right) \right\} \tag{6.1}$$

where $p = \sum_k a_k$ and $p_k = a_k/p$ make a discrete probability distribution $\mu = \sum_k p_k \delta_{a_k}$. In other words, a Poisson type ch.fun. ψ, obtained from $\varphi = \hat{\mu}$ by the formula

$$\psi = \exp\left\{ p\,(\varphi - 1) \right\} , \tag{6.2}$$

is an ID ch.fun.

Lemma 6.5 *Every* (6.2) *is ID.*

Proof. Let φ be a ch.fun. Let $p > 0$ and $n > p$. Then the power of the convex combination

$$\psi_n = \left(\left(1 - \frac{p}{n}\right) + \frac{p}{n}\,\varphi \right)^n$$

is a ch.fun., so is its limit as $n \to \infty$. Let us repeat (*): (6.2) is a ch.fun. for every $p > 0$.

However, the limit is equal to the right hand side of (6.2). We must see that $\psi^{1/m}$ is a ch.fun. for every $m \in \mathbb{N}$. But

$$\psi^{1/m} = \lim_{k\to\infty} \psi_{km}^{1/m} = \lim_{k\to\infty} \left(\left(1 - \frac{p}{km}\right) + \frac{p}{km}\,\varphi \right)^k = \exp\left\{ (p/m)\,(\varphi - 1) \right\}$$

which is a ch.fun. by (*). ■

Proposition 6.6 (De Finetti's Theorem) *A ch.fun. is ID iff it has the form*

$$\psi(t) = \lim_m \exp\left\{ p_m\,(\varphi_m(t) - 1) \right\} \tag{6.3}$$

Proof. The sufficiency follows by the continuity theorem. To prove the necessity, let ψ be ID. Then ψ^α is ID for every $\alpha > 0$. Hence

$$\exp\left\{\frac{1}{\alpha}(\psi^\alpha - 1)\right\}$$

is an ID ch.fun. by the preceding argument. So is ψ, obtained as the limit for $\alpha \to 0$. Now, choose $\alpha = 1/m$, $p_m = m$, and $\varphi_m = \psi^{1/m}$. That is, ψ can be represented as the desired limit. ■

Now, we will see that t suffices to consider only the Poisson types among the above ϕ_m's.

Theorem 6.7 *A ch.fun. ψ is ID iff ψ in* (6.3) *involves only* (6.1).

Proof. The sufficiency follows from the continuity (or De Finetti's) theorem. Let ψ be an ID ch.fun. and consider its form (6.3), ensured by De Finetti's Theorem, with $\varphi_m = \hat{\mu_m}$. We choose discrete $\mu_{mk} \overset{w}{\to} \mu_m$ as $k \to \infty$. That is, $\varphi_{mk} = \hat{\mu}_{mk} \to \varphi_m$.

Then the statement follows by the diagonal argument. ■

6.5 Lévy-Khinchin formula

This is inspired by the presentation in Loeève's book, Section 22.1. However, the original approach that used analysis, Riemann-Stieltjes integrals, etc., in a great detail and carefulness, has been "translated" into the language of Poisson integrals with the help of our sufficient background in measure theory.

Recall that a Lévy measure M on $\mathbb{R}$ is defined by the condition

$$\int_{\mathbb{R}} 1 \wedge |x|^2 \, M(dx) < \infty.$$

In this condition we may replace the function $1 \wedge |x|^2$ by any bounded monotonic continuous function that behaves like $|x|^2$ near 0. So, as the defining condition we may prefer

$$\int_{\mathbb{R}} \frac{x^2}{1+x^2} \, M(dx) < \infty,$$

or in other words, that the measure $x^2/(1+x^2)\, M(dx)$ is finite, or a probability up to a positive scalar multiplier. This probability μ follows the formula

$$c\,\mu(dx) = \frac{1+x^2}{x^2} \, M(dx), \tag{6.4}$$

for some $c > 0$. For a fixed t let us examine the behavior of the function

$$f(x) = e^{itx} - 1 - \frac{itx}{1+x^2}.$$

We see that $f(x)$ is bounded away from 0, and

$$f(x) \approx itx - \frac{t^2x^2}{2} - \frac{itx}{1+x^2} \approx -\frac{t^2x^2}{2}, \quad |x| \to 0.$$

Hence, for every Lévy measure M

$$\int_{\mathbb{R}} f(x)M(dx)$$

exists, and thus the continuous function

$$\psi(t) = \exp\left\{ iat + \int_{\mathbb{R}} \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) M(dx) \right\} \tag{6.5}$$

is well defined, and $\psi(0) = 1$. Using (6.4), we rewrite it:

$$\psi(t) = \exp\left\{ iat + c\int_{\mathbb{R}} \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) \frac{1+x^2}{x^2}\, \mu(dx) \right\}$$

Consider discrete $\mu_m = \sum_k p_{mk}\delta_{x_{mk}}$ of finite supports without 0 that converge to μ. Then

$$\begin{aligned} \ln \hat{\mu_m}(t) = & \ iat + c\sum_k p_{mk} \left(e^{itx_{mk}} - 1 - \frac{itx_{mk}}{1+x_{mk}^2} \right) \frac{1+x_{mk}^2}{x_{mk}^2} \\ = & \ \sum_k c\, p_{mk} \frac{1+x_{mk}^2}{x_{mk}^2} \left(e^{itx_{mk}} - 1 \right) - itc\sum_k \frac{p_{mk}}{x_{mk}} + iat \\ = & \ \varphi_m(t) - ita_m, \end{aligned}$$

where φ_m are of Poisson type (6.1). Thus μ_m are ID ch.fun. and so is ψ.

Let us denote the ch.f. given by (6.5) by (a, M).

Proposition 6.8 ((a,M)-uniqueness Theorem)
The pair (a, M) is unique, i.e. $(a, M) = (a', M')$ implies that $a = a'$ and $M = M'$.

Proof. Notice that the functions iat and the one defined by the integral are linearly independent. Hence $(a, M) = (a', M')$ implies that $a = a'$. Assume that $a = 0$ and $\ln \psi$ has two integral representations with M and M', or, equivalently with two corresponding probability measures μ and μ'.

We need to show that the function

$$\phi(t) = \int_{\mathbb{R}} \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) M(dx)$$

uniquely determines the probability μ. To this end, scale the variable, integrate, use Fubini's theorem, and exponentiate:

$$\exp\left\{ \frac{1}{2}\int_{-1}^{1} \phi(ut)\, dt \right\} = \exp\left\{ \int_{\mathbb{R}} (\cos ux - 1)\, M(dx) \right\} = \mathsf{E} \exp\left\{ \int_0^u \int_{\mathbb{R}} (\cos xv - 1)\, dv\, M(dx) \right\}.$$

This entails the distribution of a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ with intensity $Leb \times M$. It remains to see that M is unique. ■

Proposition 6.9 ((a,M)-convergence Theorem)
$(a_n, M_n) \to (a, M)$ iff $a_n \to a$ and $M_n \to M$ weakly (which can be expressed in terms of the convergence of the corresponding measures μ_n).

Further, if $(a_n, M_n) \to \psi$ continuous at the origin, then $\psi = (a, M)$ for some real a and some Lévy measure M.

Proof. The sufficiency is obvious. The necessity will follow by the previous approach. We infer that $\phi_n \to \phi$ implies that the distributions of the Poisson measures with mean $Leb \otimes M_n$ converge weakly to the distribution of a Poisson measure with mean $Leb \otimes M$. Hence $M_n \to M$. ■

Proposition 6.10 (Lévy-Khinchin) *Every ID ch.fun. has the unique representation* (6.5).

Proof. As we have seen, (6.5) entails ID. Now, let ψ be an ID ch.fun, so is $\psi^{1/n}$ for every n. Let the latter be a ch.fun. of some probability μ_n. Thus

$$\psi(t) = \lim_n \exp\left\{ n\Big(\psi^{1/n}(t) - 1\Big) \right\} = \lim_n \exp\left\{ \int_{\mathbb{R}} \Big(e^{itx} - 1\Big)\, n\mu_n(dx) \right\}$$

Rewrite the integral in the exponent:

$$it \int_{\mathbb{R}} \frac{nx}{1+x^2}\mu_n(dx) + \int_{\mathbb{R}} \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1+x^2}{x^2} \cdot \frac{x^2}{1+x^2}\, n\mu_n(dx) = \ln(a_n, M_n)$$

By the convergence theorem $a_n \to a$, $M_n \to M$. So, $\psi = (a, M)$. ■

6.6 Exercises

1. Find an example of a r.v. or probability distribution that is not ID although its ch.f. does not vanish ever. Hint: try simple discrete (even two-valued) r.vs.

2. **Examples of ID laws**: Normal, Poisson, stable, exponential, gamma. Prove (or just observe): *If a one-parameter family of ch.funs. is of the form* $\varphi(t) = \varphi_\theta(t) = \exp\{-\theta p(t)\}$, *where* θ *may vary through* $\mathbb{R}$ *or* $[0, \infty)$*, then* φ *is ID.*

3. Convolutions of ID are ID. If φ is ID so is $|\varphi|$.

4. Let V be exponential, so it is ID. Is V^α (so called, **Weibull r.v.**) ID?

5. Show that a discrete distribution is not ID. It might be not even decomposable! Find an example (e.g., the binomial distribution is decomposable but...).

6. Let X be c-decomposable. Prove that necessarily $|c| < 1$. Infer from this that every c-decomposable r.v. can be written as

$$X \stackrel{\mathcal{D}}{=} \sum_{n=0}^{\infty} R_n c^n,$$

where R_n are i.i.d., and so this provides a large class of examples.

7. Let $X = \sum_k a_k N_{\lambda_k}$, where $a_k, \lambda_k > 0$, N_{λ_k} are independent Poisson, and the infinite series converges in distribution. Show that it converges a.s.

8. In the proof of Theorem 6.7 the "diagonal argument" was used. Write precisely all of its details (beginning with "*Let* $\epsilon > 0$*...*") Hint: Let (x_{mk}) be a matrix of elements of a metric space such that $x_m = \lim_k x_{mk}$ exists for every m, and also $x = \lim_m x_m$ exists. Then there is a subsequence k_m such that $\lim_m x_{m\,k_m} = x$. Although the statement in the theorem involves a pointwise convergence of functions, which is not metrizable in general, the metric convergence must be used somehow.

9. Prove that the function defined in (6.5) is continuous.

10. Prove that the functions iat and the one defined by the integral in (6.5) are linearly independent.

11. At the end of the uniqueness and convergence theorems for (a, M) there were three statements.

(a) First, w.l.o.g., we may assume that an atomless Lévy measure M is a probability measure. In fact, $M = \sum_k M_k$, where M_k are probabilities. Examine the details in both statements.

(b) *Let ξ, ξ' be Poisson measures on $\mathbb{R}_+ \times \mathbb{R}$ with intensities Leb $\otimes M$ and Leb $\otimes M'$. If $\xi \overset{\mathcal{D}}{=} \xi'$, then $M = M'$.* Prove it.

(c) *If the distributions of the Poisson measures with mean Leb $\otimes M_n$ converge weakly to the distribution of a Poisson measure with mean Leb $\otimes M$, then $M_n \to M$.* Prove it.

(d) *Let the distributions of the Poisson measures with mean Leb $\otimes M_n$ converge weakly to some distribution. Then it must be of some random Poisson measure with mean Leb$\otimes M$.*