An Analytical Introduction to Probability Theory

10. Characteristic Functions

DASN, NTU https://personal.ntu.edu.sg/wptay/

10.1 Characteristic Functions

Given a random variable $X : \Omega \mapsto \mathbb{R}$, the characteristic function of X is $\varphi : \mathbb{R} \mapsto \mathbb{C}$ given by

$$\begin{aligned}\varphi(t) &= \mathbb{E}\big[e^{itX}\big],\\ &= \mathbb{E}[\cos(tX)] + \mathrm{i}\mathbb{E}[\sin(tX)],\end{aligned}$$

where $\mathrm{i} = \sqrt{-1}$. This characteristic function is always well defined as cos and sin are bounded and integrable, even if $\mathbb{E}X$ does not exist. If X has pdf $f \in L^1(\mathbb{R})$, then $\varphi(t) = \int e^{itx} f(x)\,\mathrm{d}x$, which is essentially the Fourier transform of f.

Some simple properties:

$$\begin{aligned}\varphi(0) &= 1,\\ \varphi(-t) &= \overline{\varphi(t)},\\ |\varphi(t)| &\le 1,\\ |\varphi(t+h) - \varphi(t)| &\le \mathbb{E}|e^{ihX} - 1|.\end{aligned}$$

Since $e^{ihX} - 1$ does not depend on t, φ is uniformly continuous in $\mathbb{R}$.

Example 10.1. *The characteristic function of the Gaussian distribution $N(\mu, \sigma^2)$ is*

$$\varphi(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}.$$

In particular, when $\mu = 0, \sigma^2 = 1$,

$$\varphi(t) = e^{-\frac{t^2}{2}}.$$

Lemma 10.1. *If $\mathbb{E}|X|^r < \infty$ for integer $r \ge 1$, then $\varphi(t) \in C^r(\mathbb{R})$ and $\varphi^{(j)}(t) = \mathbb{E}\big[(\mathrm{i}X)^j e^{\mathrm{i}tX}\big]$ for $j \le r$.*

Proof. We start with a simple fact. For $a \le b$, we have

$$\begin{aligned}|e^{\mathrm{i}a} - e^{\mathrm{i}b}| &= \left|\int_a^b \mathrm{i}e^{\mathrm{i}t}\,\mathrm{d}t\right|\\ &\le \int_a^b |\mathrm{i}e^{\mathrm{i}t}|\,\mathrm{d}t\\ &= |b-a|.\end{aligned}$$

The proof of Lemma 10.1 is by induction on r. For $r = 1$, we have $\left|\frac{e^{itX} - e^{isX}}{t-s}\right| \leq |X|$. Since $\mathbb{E}|X| < \infty$, from the DCT (Theorem 4.3), we obtain

$$\varphi'(t) = \lim_{s\to t} \mathbb{E}\left[\frac{e^{\mathrm{i}tX} - e^{\mathrm{i}sX}}{t-s}\right] = \mathbb{E}\big[\mathrm{i}Xe^{\mathrm{i}tX}\big] \in C(\mathbb{R}),$$

where the set inclusion follows from a further application of the DCT. Therefore, $\varphi \in C^1(\mathbb{R})$. Assume that the lemma holds for r. Then for $r+1$, by assumption, we have

$$\begin{aligned} &\varphi^{(r)}(t) = \mathbb{E}\big[(\mathrm{i}X)^r e^{\mathrm{i}tX}\big] \in C^r(\mathbb{R}), \\ &\left|\frac{(\mathrm{i}X)^r e^{\mathrm{i}tX} - (\mathrm{i}X)^r e^{\mathrm{i}sX}}{t-s}\right| \leq |X|^{r+1}, \end{aligned}$$

and from the DCT, we have

$$\varphi^{(r+1)}(t) = \lim_{s\to t} \mathbb{E}\left[\frac{(\mathrm{i}X)^r e^{\mathrm{i}tX} - (\mathrm{i}X)^r e^{\mathrm{i}sX}}{t-s}\right] = \mathbb{E}\big[(\mathrm{i}X)^{(r+1)} e^{\mathrm{i}tX}\big] \in C(\mathbb{R}),$$

and the proof is complete. $\square$

Suppose $X \perp\!\!\!\perp Y$. We then have $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$. We define

$$\begin{aligned} \mathbb{P}_X * \mathbb{P}_Y(A) &\triangleq \mathbb{P}_{X+Y}(A) \\ &= \mathbb{E}\big[\mathbf{1}_{\{X+Y\in A\}}(\omega)\big] \\ &\overset{\text{Fubini}}{=} \int\int \mathbf{1}_{\{x+y\in A\}} \,\mathrm{d}\mathbb{P}_X(x)\,\mathrm{d}\mathbb{P}_Y(y) \\ &= \int\int \mathbf{1}_{\{x\in A-y\}} \,\mathrm{d}\mathbb{P}_X(x)\,\mathrm{d}\mathbb{P}_Y(y) \\ &= \int \mathbb{P}_X(A-y)\,\mathrm{d}\mathbb{P}_Y(y), \end{aligned}$$

where $A - y = \{a - y \,:\, a \in A\}$. If X has pdf p_X, then

$$\begin{aligned} \mathbb{P}_X * \mathbb{P}_Y(A) &= \int\int \mathbf{1}_{\{x+y\in A\}} p_X(x)\,\mathrm{d}x\,\mathrm{d}\mathbb{P}_Y(y) \\ &= \int\int \mathbf{1}_{\{z\in A\}} p_X(z-y)\,\mathrm{d}z\,\mathrm{d}\mathbb{P}_Y(y) \\ &= \int\int_A p_X(z-y)\,\mathrm{d}z\,\mathrm{d}\mathbb{P}_Y(y) \\ &\overset{\text{Fubini}}{=} \int_A\int p_X(z-y)\,\mathrm{d}\mathbb{P}_Y(y)\,\mathrm{d}z, \end{aligned}$$

hence the pdf of $\mathbb{P}_X * \mathbb{P}_Y$ is $\int p_X(z-y)\,\mathrm{d}\mathbb{P}_Y(y)$, i.e., $\mathbb{P}_X * \mathbb{P}_Y \ll$ Lebesgue measure.

If X and Y have pdfs p_X and p_Y respectively, then the pdf of $\mathbb{P}_X * \mathbb{P}_Y$ is given by

$$p_{X+Y}(z) = \int p_X(z-y) p_Y(y)\,\mathrm{d}y,$$

which is the convolution of the pdfs p_X and p_Y.

A special case is

$$\mathbb{P}^\sigma = \mathbb{P}\mathcal{N}(0,\sigma^2) \ll \text{Lebesgue measure},$$

therefore, its pdf exists and it can be shown to be given by

$$p^\sigma(x) = \frac{1}{2\pi}\int \varphi(t) e^{-\mathrm{i}tx - \frac{\sigma^2}{2}t^2}\,\mathrm{d}t, \tag{1}$$

where φ is the characteristic function of $\mathbb{P}$.

Lemma 10.2 (Uniqueness)**.** *If $\varphi_X(t) = \varphi_Y(t)$, then $\mathbb{P}_X = \mathbb{P}_Y$.*

Proof. Assume $\varphi_X(t) = \varphi_Y(t)$, then from (1), we have $\mathbb{P}^\sigma_X = \mathbb{P}^\sigma_Y$. We have

$$X(\omega) + \sigma Z(\omega) \xrightarrow{\sigma\to 0} X(\omega) \ \text{ a. s.},$$
$$Y(\omega) + \sigma Z(\omega) \xrightarrow{\sigma\to 0} Y(\omega) \ \text{ a. s.},$$

and Lemma 9.5 yields

$$\mathbb{P}^\sigma_X \xrightarrow{\mathrm{d}} \mathbb{P}_X,\ \mathbb{P}^\sigma_Y \xrightarrow{\mathrm{d}} \mathbb{P}_Y.$$

Therefore, $\mathbb{P}_X = \mathbb{P}_Y$. □

Lemma 10.3 (Fourier Inversion)**.** *Suppose that $\varphi(t)$ is the characteristic function of $\mathbb{P}$. If $\int |\varphi(t)|\,\mathrm{d}t < \infty$, then $\mathbb{P}$ has pdf*

$$p(x) = \frac{1}{2\pi}\int \varphi(t) e^{-\mathrm{i}tx}\,\mathrm{d}t.$$

Proof. We note that

$$\varphi(t) e^{-\mathrm{i}tx - \frac{\sigma^2}{2}t^2} \xrightarrow{\sigma\to 0} \varphi(t) e^{-\mathrm{i}tx}, \text{ and}$$
$$\left|\varphi(t) e^{-\mathrm{i}tx - \frac{\sigma^2}{2}t^2}\right| \le |\varphi(t)|.$$

By the DCT (Theorem 4.3), we then have

$$p^\sigma(x) \xrightarrow{\sigma\to 0} p(x).$$

Since $\mathbb{P}^\sigma \xrightarrow{\mathrm{d}} \mathbb{P}$ and p^σ is the pdf of $\mathbb{P}^\sigma$, we have

$$\int g(x)p^\sigma(x)\,\mathrm{d}x \xrightarrow{\sigma\to 0} \int g(x)\,\mathrm{d}\mathbb{P}(x),\ \forall\, g \in C_b(\mathbb{R}). \tag{2}$$

Now consider a $g \in C_c(\mathbb{R})$.[1] Since $|p^\sigma(x)| \le \frac{1}{2\pi}\int |\varphi(t)|\,\mathrm{d}t < \infty$, from the DCT (Theorem 4.3), we obtain

$$\int g(x)p^\sigma(x)\,\mathrm{d}x \xrightarrow{\sigma\to 0} \int g(x)p(x)\,\mathrm{d}x.$$

Together with (2), we have $\int g(x)\,\mathrm{d}\mathbb{P}(x) = \int g(x)p(x)\,\mathrm{d}x$ and

$$\int_A \mathrm{d}\mathbb{P}(x) = \int_A p(x)\,\mathrm{d}x,$$

for all compact sets A. From Ulam's Theorem (Theorem 3.3), $p(x)$ is the pdf of $\mathbb{P}$. $\square$

We want to find out under what condition convergence of a sequence of characteristic functions $\varphi_n(t)$ implies weak convergence of the corresponding probability distributions $\mathbb{P}_n$. The following example shows that this implication is not always true.

Example 10.2. *The distribution $\mathcal{N}(0,n)$ has characteristic function $\varphi_n(t) = e^{-\frac{nt^2}{2}} \xrightarrow{n\to\infty} 0,\ \forall\, t \neq 0$ and $\varphi_n(0) = 1,\ \forall\, n$. Therefore $\varphi_n(t)$ converges. But for all $x \in \mathbb{R}$,*

$$\mathbb{P}_n((-\infty,x)) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x}{\sqrt{2n}}\right)\right) \to \frac{1}{2},$$

when $n \to \infty$, which implies that this sequence of distributions does not converge weakly.

On the other hand, we have the following converse.

Lemma 10.4. *If $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}$ then $\varphi_n(t) \to \varphi(t),\ \forall\, t \in \mathbb{R}$.*

Proof. Obvious because $\varphi_n(t) = \mathbb{E}\left[e^{\mathrm{i}tX_n}\right] = \mathbb{E}[\cos tX_n] + \mathrm{i}\mathbb{E}[\sin tX_n]$, $\cos(t\cdot)$ and $\sin(t\cdot)$ are bounded continuous functions, and $X_n \xrightarrow{\mathrm{d}} X$. $\square$

Lemma 10.5. *Suppose that $(\mathbb{P}_n)_{n\ge1}$ is uniformly tight on $\mathbb{R}$ and $\varphi_n(t) = \int e^{\mathrm{i}tx}\,\mathrm{d}\mathbb{P}_n(x) \xrightarrow{n\to\infty} \varphi(t)$. Then $\varphi(t) = \int e^{\mathrm{i}tx}\,\mathrm{d}\mathbb{P}(x)$ for some probability distribution $\mathbb{P}$ and $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}$.*

Proof. From Helly's selection theorem (Theorem 9.2), for every subsequence $N = (n(k))_{k\ge1}$, there exists a subsequence $(n(k(r)))_{r\ge1}$ such that $\mathbb{P}_{n(k(r))} \xrightarrow{\mathrm{d}} \mathbb{P}_N$ for some $\mathbb{P}_N$ as $r \to \infty$. Therefore, since $e^{\mathrm{i}tx}$ is bounded and continuous, we have from Definition 9.1 that

$$\int e^{\mathrm{i}tx}\,\mathrm{d}\mathbb{P}_{n(k(r))} \to \int e^{\mathrm{i}tx}\,\mathrm{d}\mathbb{P}_N.$$

From the lemma assumption, the L.H.S. converges to $\varphi(t)$, therefore $\int e^{\mathrm{i}tx}\,\mathrm{d}\mathbb{P}_N = \varphi(t)$. By Lemma 10.2, $\mathbb{P}_N = \mathbb{P}$ is the same for all choices of $N = (n(k))_{k\ge1}$. The result then follows from Lemma 9.4. $\square$

[1] $C_c(\mathbb{R})$ is the set of continuous functions with compact support.

Theorem 10.1 (Levy's Continuity Theorem)**.** *Suppose $\mathbb{P}_n$ has characteristic function $\varphi_n(t) \to \varphi(t)$, and $\varphi(t)$ is continuous at $t = 0$. Then there exists $\mathbb{P}$ s.t. $\varphi(t)$ is the characteristic function of $\mathbb{P}$ and $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}$.*

Proof. From Lemma 10.5, it suffices to show that $(\mathbb{P}_n)$ is uniformly tight. Since $\varphi(0) = \lim_{n\to\infty} \varphi_n(0) = 1$, $\forall\, \epsilon > 0,\ \exists\, \delta > 0$ *s.t.* $|\varphi(t) - 1| < \epsilon$ if $|t| \le \delta$. Letting $\Re(\cdot)$ denote the real part, we then have

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{\delta}\int_0^\delta (1 - \Re(\varphi_n(t)))\,\mathrm{d}t \\
&\overset{\text{DCT}}{=} \frac{1}{\delta}\int_0^\delta (1 - \Re(\varphi(t)))\,\mathrm{d}t \\
&\le \frac{1}{\delta}\int_0^\delta |1 - \varphi(t)|\,\mathrm{d}t \\
&< \epsilon.
\end{aligned} \tag{3}$$

For each $n \ge 1$, we have

$$\begin{aligned}
\frac{1}{\delta}\int_0^\delta (1 - \Re(\varphi_n(t)))\,\mathrm{d}t &= \frac{1}{\delta}\int_0^\delta \int_{\mathbb{R}} (1 - \cos(tx))\,\mathrm{d}\mathbb{P}_n(x)\,\mathrm{d}t \\
&\overset{\text{Fubini}}{=} \frac{1}{\delta}\int_{\mathbb{R}}\int_0^\delta (1 - \cos(tx))\,\mathrm{d}t\,\mathrm{d}\mathbb{P}_n(x) \\
&= \int_{\mathbb{R}} \left(1 - \frac{\sin(\delta x)}{\delta x}\right)\mathrm{d}\mathbb{P}_n(x) \\
&\ge \int_{|\delta x| \ge \pi} \left(1 - \frac{\sin(\delta x)}{\delta x}\right)\mathrm{d}\mathbb{P}_n(x) \\
&\ge \frac{1}{2}\mathbb{P}(-\,|\,NoValue - n)(\left\{x\,:\,|x| \ge \frac{\pi}{\delta}\right\}),
\end{aligned}$$

since $\operatorname{sinc} y = \frac{\sin \pi y}{\pi y} \le 1$ for $|y| \le 1$ and $\le 1/2$ for $|y| \ge 1$. From (3), for all n sufficiently large, we have

$$\mathbb{P}(-\,|\,NoValue - n)(\left\{x\,:\,|x| \ge \frac{\pi}{\delta}\right\}) \le 4\epsilon.$$

Therefore $(\mathbb{P}_n)$ is uniformly tight. □

An application of Theorem 10.1 is to show that if $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P},\ \mathbb{Q}_n \xrightarrow{\mathrm{d}} \mathbb{Q}$, then $\mathbb{P}_n \times \mathbb{Q}_n \xrightarrow{\mathrm{d}}$

$\mathbb{P} \times \mathbb{Q}$. For $t = (t_1, t_2)$ and $x = (x_1, x_2)$, we have

$$\begin{aligned}
&\int e^{\mathrm{i}\langle t,x\rangle}\, \mathrm{d}\mathbb{P}_n \times \mathbb{Q}_n(x) \\
&\overset{\text{Fubini}}{=} \int e^{\mathrm{i}t_1x_1}\, \mathrm{d}\mathbb{P}_n(x_1) \int e^{\mathrm{i}t_2x_2}\, \mathrm{d}\mathbb{Q}_n(x_2) \\
&\to \int e^{\mathrm{i}t_1x_1}\, \mathrm{d}\mathbb{P}(x_1) \int e^{\mathrm{i}t_2x_2}\, \mathrm{d}\mathbb{Q}(x_2) \text{ since } \mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P},\ \mathbb{Q}_n \xrightarrow{\mathrm{d}} \mathbb{Q} \\
&\overset{\text{Fubini}}{=} \int e^{\mathrm{i}\langle t,x\rangle}\, \mathrm{d}\mathbb{P} \times \mathbb{Q}(x),
\end{aligned}$$

which is the characteristic function of $\mathbb{P}\times\mathbb{Q}$ and is thus continuous at 0. From Theorem 10.1, we then have

$$\mathbb{P}_n \times \mathbb{Q}_n \xrightarrow{\mathrm{d}} \mathbb{P} \times \mathbb{Q}.$$

Exercise 10.1. *Show that if $\mathbb{P}_n \xrightarrow{\mathrm{d}} \mathbb{P}$, $\mathbb{Q}_n \xrightarrow{\mathrm{d}} \mathbb{Q}$, then $\mathbb{P}_n * \mathbb{Q}_n \xrightarrow{\mathrm{d}} \mathbb{P}\mathbb{Q}$. (Hint: Let $g(x, y) = x + y$ and consider $\mathbb{P}_n \times \mathbb{Q}_n \circ g^{-1}$.)*

10.2 Central Limit Theorem for I.I.D. Sequences

Suppose that $f : \mathbb{R} \mapsto \mathbb{R}$ is k-differentiable at $a \in \mathbb{R}$. Then the Taylor series of f evaluated at a is given by

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k + o(|x-a|^k).$$

Furthermore, if $f^{(k+1)}$ exists and is continuous in an open interval I containing a, then

$$o(|x-a|^k) \le \sup_I \left|f^{(k+1)}\right| \frac{|x-a|^{k+1}}{(k+1)!}.$$

Theorem 10.2 (CLT for i.i.d. sequences)**.** *Consider i.i.d. random variables $X_1, X_2, \ldots$, with $\mathbb{E}X_i = \mu$ and $\operatorname{var} X_i = \sigma^2 < \infty$. Let $S_n = \sum_{i=1}^n X_i$. Then*

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1). \tag{4}$$

Proof. Without loss of generality, we assume that $\mu = 0, \sigma = 1$. Let

$$\begin{aligned}
\varphi_n(t) &= \mathbb{E}\left[\exp\left(\mathrm{i}t\frac{S_n}{\sqrt{n}}\right)\right] \\
&= \prod_{j=1}^n \mathbb{E}\left[\exp\left(\mathrm{i}t\frac{X_j}{\sqrt{n}}\right)\right] \\
&= \left(\mathbb{E}e^{\mathrm{i}t\frac{X_1}{\sqrt{n}}}\right)^n \\
&= \varphi_1\left(\frac{t}{\sqrt{n}}\right)^n.
\end{aligned}$$

Since $\mathbb{E}X_1^2 < \infty$ we have $\varphi_1(t) \in C^2(\mathbb{R})$ from Lemma 10.1, and its Taylor series is

$$\varphi_1(t) = \varphi_1(0) + \varphi_1'(0)t + \frac{\varphi_1''(0)}{2!} + o(t^2).$$

Since $\varphi_1(0) = 1$, $\varphi_1'(0) = \mathbb{E}\big[\mathrm{i}X_1 e^{i0X_1}\big] = \mathrm{i}\mathbb{E}X_1 = 0$, $\varphi_1''(0) = \mathbb{E}[(\mathrm{i}X_1)^2] = -1$, we obtain

$$\varphi_1(t) = 1 - \frac{t^2}{2} + o(t^2).$$

Therefore,

$$\begin{aligned}\varphi_n(t) &= \varphi_1\left(\frac{t}{\sqrt{n}}\right)^n \\ &= \left(1 - \frac{t^2}{2n} + o(t^2/n)\right)^n \xrightarrow{n\to\infty} e^{-\frac{t^2}{2}},\end{aligned}$$

the characteristic function of $\mathcal{N}(0,1)$. From Theorem 10.1, we then have

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

□

10.3 Stable Distributions

The above proof of the CLT does not give us much intuition of why the amazing result that is the CLT holds. Some intuition can be had from the following observation: If $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent, then $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. By induction, if $X_i \sim \mathcal{N}(0,1)$ for $i = 1, \ldots, n$, we have $\frac{S_n}{\sqrt{n}} \sim \mathcal{N}(0,1)$. Therefore, the normal distribution has a "stability" property.

In general, suppose $X \sim \mathbb{P}$ and $X_1, X_2, \ldots$ are i.i.d. copies of X. If

$$X \overset{d}{=} \frac{\sum_{j=1}^k X_j - a_k}{b_k}$$

for some sequences of constants (a_k) and (b_k), then we say that $\mathbb{P}$ is a stable distribution or stable law.

It turns out that all stable laws have characteristic functions given by

$$\varphi(t) = \exp(\mathrm{i}tc - b|t|^\alpha(1 + \mathrm{i}\kappa \operatorname{sgn}(t)\omega_\alpha(t))),$$

where $-1 \le \kappa \le 1$, $0 < \alpha \le 2$ and

$$\omega_\alpha(t) = \begin{cases} \tan(\pi\alpha/2), & \text{if } \alpha \ne 1, \\ \frac{2}{\pi}\log|t|, & \text{if } \alpha = 1. \end{cases}$$

In particular, if $\alpha = 2$, $\varphi(t) = \exp(\mathrm{i}tc - bt^2) \sim \mathcal{N}(c, 2b)$, and this is the only stable law with finite variance. Therefore, for (4) to converge in distribution, it **must** converge to the normal distribution. Of course, this does not explain why (4) must converge in distribution in the first place.