Funciones características y el teorema central del límite

Pablo L. De Nápoli

Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Probabilidades y Estadística para Matemática
Segundo cuatrimestre de 2021

Parte I

Más sobre funciones características y la transformada de Fourier

Funciones características

Definición

Si X es una variable aleatoria tal que $E(|X|)$ es finita, su función característica se define por

$$\varphi_X(t) = E\left[e^{itX}\right] = \int_{-\infty}^{\infty} e^{itx}\, dF_X(x) \quad t \in \mathbb{R}$$

φ_X es siempre una función uniformemente continua, $|\varphi_X(t)| \leq 1$ y $\varphi_X(0) = 1$.

Si X es una variable aleatoria absolutamente continua con densidad de probabilidad $f(x)$, entonces la función característica viene dada por la transformada de Fourier de f

$$\widehat{f}(t) = \mathcal{F}(f)(t) = \int_{-\infty}^{\infty} f(x) e^{itx}\, dx$$

La transformada de Fourier está definida para cualquier $f : \mathbb{R} \to \mathbb{C}$ tal que

$$\int_{-\infty}^{\infty} |f(x)|\, dx < \infty \quad \textbf{Notación: } f \in L^1(\mathbb{R})$$

Fórmulas de inversión

Teorema (Fórmula de inversión)

Sea X una variable aleatoria con $E(|X|) < +\infty$, función de distribución F_X y función característica φ_X. Entonces

$$F_X(x_0) = \frac{1}{2\pi} \lim_{\sigma \to 0} \int_{-\infty}^{x_0} \left[\int_{-\infty}^{\infty} \varphi_X(y)\, e^{-iz\cdot y} \cdot e^{-(\sigma y)^2/2}\, dy \right] dz$$

en cada punto de continuidad x_0 de F_X.

Teorema (Fórmula clásica de inversión de Fourier)

Sea $f \in L^1(\mathbb{R})$ continua tal que $\widehat{f} \in L^1(\mathbb{R})$, entonces podemos reconstruir f a partir de su transformada mediante la fórmula de inversión

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f}(y)\, e^{-ix\cdot y}\, dy$$

Un ejemplo

Una variable aleatoria X tiene la distribución de Laplace o distribución exponencial doble con parámetros $\mu \in \mathbb{R}$ y $b > 0$ si tiene la densidad de probabilidad

$$f(x) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$

Calculemos su función caracerística. Usando las propiedades que vimos la clase pasada, basta saber hacerlo con $\mu = 0$ y $b = 1$. En ese caso,

$$\begin{aligned}
\varphi_X(t) = \widehat{f}(t) &= \int_{-\infty}^{\infty} e^{ixt} \frac{1}{2} e^{-|x|} \, dx \\
&= \int_{-\infty}^{0} e^{ixt} \frac{1}{2} e^{-|x|} \, dx + \int_{0}^{\infty} e^{ixt} \frac{1}{2} e^{-|x|} \, dx \\
&= \int_{-\infty}^{0} e^{ixt} \frac{1}{2} e^{x} \, dx + \int_{0}^{\infty} e^{ixt} \frac{1}{2} e^{-x} \, dx \\
&= \frac{1}{2} \int_{-\infty}^{0} e^{x(it+1)} \, dx + \frac{1}{2} \int_{0}^{\infty} e^{x(it-1)} \, dx
\end{aligned}$$

Un ejemplo (2)

Siguiendo, las integrales que nos quedaron se pueden calcular con la definición de integral impropia y la regla de Barrow. Nos queda:

$$\begin{aligned}\varphi_X(t) = \widehat{f}(t) &= \frac{1}{2}\left[\frac{1}{it+1} - \frac{1}{it-1}\right] \\ &= \frac{1}{1+t^2}\end{aligned}$$

En general, si μ y b son cualesquiera, la función característica de una variable aleatoria con distribución de Laplace va a ser

$$\varphi_X(t) = \frac{e^{it\mu}}{1+b^2 t}$$

Otro ejemplo (aparece en la práctica: ejercicios 11 y 12)

La distribución de Cauchy $\mathcal{C}(\mu, \lambda)$ tiene densidad de probabilidad dada por:

$$f(x) = \frac{1}{\pi} \frac{\lambda}{\lambda^2 + (x-\mu)^2}$$

De vuelta, no bastaría calcular su función característica con $\mu = 0$ y $\lambda = 1$. Sería

$$\varphi_X(t) = \int_{-\infty}^{\infty} \frac{1}{\pi} \frac{1}{x^2+1} e^{ixt} \, dx$$

No es fácil calcular esta integral directamente. Pero si observamos que la densidad $f(x) = \frac{1}{1+x^2}$ es el resultado del ejemplo anterior, podemos calcularla usando la fórmula de inversión de Fourier (como f es par, no cambia la integral si reemplazamos x por $-x$). Obtenemos

$$\varphi_X(t) = e^{-|t|}$$

En general, si μ y λ son cualesquiera,

$$\varphi_X(t) = e^{i\mu t - \lambda |t|}$$

Transformada de Fourier de una derivada

Proposición

Si $f : \mathbb{R} \to \mathbb{C}$ *es una función en* L^1 *que es de clase* C^1 *y* $f(x) \to 0$ *cuando* $|x| \to +\infty$,

$$\mathcal{F}(f')(t) = (-it)\mathcal{F}f(t)$$

Demostración:

$$\begin{aligned}
\mathcal{F}(f')(t) &= \int_{-\infty}^{\infty} f'(x)e^{ixt}\,dx \\
&= \lim_{R\to+\infty} \int_{-R}^{R} f'(x)e^{ixt}\,dx \\
&= \lim_{R\to+\infty} \left\{ f(x)e^{ixt}\Big|_{-R}^{R} - \int_{-R}^{R} f(x)ite^{ixt}\,dx \right\} \\
&= -\int_{-\infty}^{\infty} f(x)ite^{ixt}\,dx \\
&= (-it)\mathcal{F}f(t)
\end{aligned}$$

Derivada de la transformada de Fourier

Proposición

Si $f : \mathbb{R} \to \mathbb{C}$ *es una función en* L^1 *tal que* $x \cdot f(x) \in L^1$ *entonces* $\widehat{f}$ *es derivable y*

$$\frac{d}{dt}\mathcal{F}f(t) = \mathcal{F}(ixf)(t)$$

Demostración:

$$\begin{aligned}\frac{d}{dt}\mathcal{F}f(t) &= \int_{-\infty}^{\infty} f(x)\,\frac{d}{dt}[e^{ixt}]\,dx \\ &= \int_{-\infty}^{\infty} f(x) ixe^{ixt}\,dx \\ &= \mathcal{F}(ixf)(t)\end{aligned}$$

Para justificar la derivación bajo el signo de integral, se usa un teorema de análisis real (corolario del teorema de convergencia mayorada que mencioné en la clase anterior).

El espacio de Schwartz

Definimos el espacio de Schwartz $\mathcal{S}(\mathbb{R})$ como el conjunto de funciones $f : \mathbb{R} \to \mathbb{C}$ de clase C^∞ tales que para todo par de índices j y k en $\mathbb{N}_0$ existe una constante $M_{j,k}$ tal que

$$|x^j f^{(k)}(x)| \leq M_{j,k} \text{ para todo } x \in \mathbb{R}$$

La idea es que si una función está en $\mathcal{S}(\mathbb{R})$, ella y todas sus derivadas decaen en infinito más rápido que x^{-k} para todo k. Es un espacio muy chico, pero las funciones C^∞ de soporte compacto están en él, así como las funciones gaussianas

$$f(x) = e^{-ax^2} \quad \text{con } a > 0$$

En particular, si $f \in \mathcal{S}(\mathbb{R})$, $f^{(k)}$ y $x^k f$ estarán en $\mathcal{S}(\mathbb{R})$ para todo $k \in \mathbb{N}$.

El espacio de Schwartz (2)

Teorema

Sea $\mathcal{S} = \mathcal{S}(\mathbb{R})$ el espacio de Schwartz. La transformada de Fourier pensada como una transformación lineal $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ es biyectiva. Su inversa viene dada por la fórmula de inversión clásica

$$\mathcal{F}^{-1}(g)(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(y)\, e^{-ix\cdot y} f(y)\, dy$$

Para todo índice k tenemos:

$$\mathcal{F}(f^{(k)})(t) = (-it)^k \mathcal{F} f(t)$$

$$\frac{d^k}{dt^k} \mathcal{F} f(t) = \mathcal{F}((ix)^k f)(t)$$

El Teorema de Continuidad de Paul Levy

Teorema

Sea $(X_n)_{n\in\mathbb{N}}$ *una sucesión de variables aleatorias finitas en casi todo punto, y* X *otra variable aleatoria finita en casi todo punto. Entonces*

$$X_n \xrightarrow{D} X \Leftrightarrow \varphi_{X_n}(t) \to \varphi(t) \; \forall \; t \in \mathbb{R}$$

Nota: En realidad vamos a ver que si

$$\varphi_{X_n}(t) \to \varphi(t) \text{ para casi todo } t$$

entonces

$$X_n \xrightarrow{D} X$$

Un teorema de análisis real

Para la demostración vamos a usar otro resultado que enunciamos en la clase pasada,

Teorema (de convergencia mayorada de Lebesgue)

Sea (f_n) *una sucesión de funciones en* $L^1(\mathbb{R}^n)$. *Si existe una función* g *en* $L^1(\mathbb{R})$ *tal que*

$$|f_n(x)| \leq g(x) \text{ para casi todo } x \in \mathbb{R}$$

y

$$\lim_{n\to+\infty} f_n(x) = f(x) \text{ para casi todo } x \in \mathbb{R}$$

entonces

$$\lim_{n\to+\infty} \int_{-\infty}^{\infty} f_n(x)\, dx = \int_{-\infty}^{\infty} f(x)\, dx$$

En este teorema, "para casi todo x´´ quiere decir "salvo quizás para los x en un conjunto cuya medida de Lebesgue es cero´´. En este teorema en realidad es muy importante en realidad usar la integral de Lebesgue, porque el límite puntual de funciones integrables en el sentido de Riemann podría no serlo.

Demostración del teorema de continuidad de Levy (1)

Supongamos primero que $X_n \xrightarrow{D} X$. Para ver que $\varphi_{X_n}(t) \to \varphi(t)$ basta aplicar el teorema de Helly-Bray a la función $\varphi(t) = e^{itx}$ que es continua y acotada (Este teorema se extiende a funciones con valores complejos).

Ahora queremos probar el recíproco. Supongamos que $\varphi_{X_n}(t) \to \varphi(t)$ para todo t. Queremos probar que $X_n \xrightarrow{D} X$. Usando el recíproco fuerte del teorema de Helly-Bray, esto es equivalente a probar que

$$E[\psi(X_n)] \to E[\psi(X)]$$

para toda ψ de clase C^∞ con soporte compacto. Como obervamos antes, ψ está en el espacio de Swartz, así que podemos escribir $\psi = \mathcal{F}(g)$ donde

$$g(x) = \mathcal{F}^{-1}(\psi)(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi(y)\, e^{-ix\cdot y}\, dy$$

será otra función en el espacio de Swartz.

Demostración del teorema de continuidad de Levy(2)

Entonces escibimos

$$E[\psi(X_n)] = \int_{-\infty}^{\infty} \psi(x)\, dF_{X_n}(x) \qquad = \int_{-\infty}^{\infty} \widehat{g}(x)\, dF_{X_n}(x)$$
$$= \int_{-\infty}^{\infty} g(x)\, \varphi_{X_n}(x)\, dx$$

por la indentidad de Plancherel, que vimos en la clase anterior. Como

$$|g(x)\, \varphi_{X_n}(x)| \leq |g(x)|$$

y g está en L^1, podemos pasar al límite cuando $n \to +\infty$ usando el teorema de convergencia mayorada (que enuncié en la clase anterior). Y se obtiene:

$$E[\psi(X_n)] \to \int_{-\infty}^{\infty} g(x)\, \varphi_X(x)\, dx = E[\psi(X)]$$

haciendo la misma cuenta que antes, con X_n en lugar de X. Como vale para toda ψ con soporte compacto, deducimos que

$$X_n \xrightarrow{D} X$$

Un ejemplo

Ejercicio

Supongaamos que las X_n son variables de Rademacher con probabilidad de éxito 1/2, o sea

$$P\{X_n = -1\} = P\{X_n = 1\} = 1/2$$

y son independientes. Probar que

$$Y_n = \sum_{k=1}^{n} \frac{X_k}{2^k} \xrightarrow{D} \mathcal{U}(-1,1) \text{ cuando } n \to +\infty$$

Ayuda: usar la siguiente identidad trigonométrica para calcular φ_{Y_n}.

$$\text{sen}\, t = 2 \sin\left(\frac{t}{2}\right) \cdot \cos\left(\frac{t}{2}\right)$$

Notamos que

$$\varphi_{X_n}(t) = E[e^{itX_n}] = \frac{1}{2} \cdot e^{it} + \frac{1}{2} e^{-it} = \cos(t)$$

Solución

Calculemos la función característica de Y_n. Como las X_n son independientes:

$$\varphi_{Y_n}(t) = \prod_{k=1}^{n} \varphi_{X_k}\left(\frac{t}{2^k}\right) = \prod_{k=1}^{n} \cos\left(\frac{t}{2^k}\right)$$

¿Cómo calcular este producto ? La identidad trigonométrica

$$\operatorname{sen} t = 2 \operatorname{sen}\left(\frac{t}{2}\right) \cdot \cos\left(\frac{t}{2}\right)$$

permite probar por inducción que

$$\operatorname{sen} t = 2^n \operatorname{sen}\left(\frac{t}{2^n}\right) \left[\prod_{k=1}^{n} \cos\left(\frac{t}{2^k}\right)\right]$$

Estas fórmulas las saqué del artículo de Wikipedia sobre la fórmula de Viète para π.

Solución (2)

Entonces despejando vemos que si $t \notin \pi\mathbb{Z}$,

$$\varphi_{Y_n}(t) = \frac{\sin(t)}{2^n \operatorname{sen}\left(\frac{t}{2^n}\right)} = \frac{\sin(t)}{t} \cdot \frac{\frac{t}{2^n}}{\operatorname{sen}\left(\frac{t}{2^n}\right)}$$

Como

$$\lim_{x \to 0} \frac{\operatorname{sen}\ x}{x} = 1$$

vemos que

$$\varphi_{Y_n}(t) \to \frac{\operatorname{sen} t}{t} \quad \forall\ t \notin \pi\mathbb{Z}$$

Deducimos que

$$\varphi_{Y_n}(t) \to \varphi_Y(t) \text{ para todo } t \notin \pi\mathbb{Z}$$

donde $Y \sim \mathcal{U}(-1,1)$. Por el teorema de continuidad de Paul Levy,

$$Y_n \xrightarrow{D} Y$$

Parte II

El teorema del Límite Central

El teorema del límite central

Teorema (Teorema del Límite Central, versión sencilla)

Sea $(X_k)_{k\in\mathbb{N}} : \Omega \to \mathbb{R}$ *una sucesión de variables aleatorias independientes e idénticamente distribuidas con* $0 < \sigma^2 = Var(X_k) < +\infty$. *Sea* $\mu = E[X_k]$ *(como suponemos que las* X_k *tienen todas la misma distribución, tendrán todas la misma esperanza y varianza). Notemos:*

$$S_n = X_1 + X_2 + \ldots + X_n$$

$$S_n^* = \frac{S_n - E[S_n]}{\sqrt{Var(S_n)}} = \frac{S_n - n\mu}{\sqrt{n}\,\sigma}$$

Entonces

$$S_n^* \xrightarrow{D} N(0,1)$$

Nota: Si las X_n son variables de Bernoulli, S_n representa el número de éxitos en n ensayos de Bernoulli, y se obtiene el teorema de De Moivre-Laplace que enuciamos en la clase 7.

Un lema de análisis complejo

Para la prueba vamos a necesitar el siguiente lema,

Lema

Si (c_n) es una sucesión de números complejos tal que $c_n \to c$, entonces

$$\left(1 + \frac{c_n}{n}\right)^n \to e^c$$

Recordamos que si $z \in \mathbb{C}$, $z = x + iy$ con $x, y \in \mathbb{R}$

$$e^z = e^x(\cos y + i \operatorname{sen} y)$$

Este lema generaliza una propiedad análoga que vale para la exponencial de variable real, y lo vamos a aceptar sin demostración. En variable real, se lo puede demostrar usando el dearrollo de Taylor del logaritmo. En variable compleja funciona la misma cuenta, pero hay que conocer el logaritmo complejo.

Demostración del teorema (1)

Sin pérdida de generalidad, podemos suponer que $\mu = 0$, cambiando sino las X_k por las variables centradas

$$\widetilde{X}_k = X_k - \mu$$

Calculemos la función característica de S_n^*. Como las (X_k) son idependientes, y tienen todas la misma distribución será

$$\varphi_{S_n^*}(t) = \varphi\left(\frac{t}{\sigma\sqrt{n}}\right)^n$$

donde $\varphi(t) = \varphi_{X_k}(t)$ para todo k.

Demostración del teorema (2)

Hagamos el desarrollo de Taylor de $\varphi(t)$ a segundo orden. Como vimos en la clase anterior, las derivadas de φ_X en $t = 0$ están relacionadas con los momentos de X

$$\begin{aligned} \varphi_{X_k}(t) &= 1 + \varphi'(0)t + \frac{1}{2}\varphi''(0)t^2 + t^2 e_2(t) \\ &= 1 - \frac{\sigma^2}{2}t^2 + t^2 e_2(t) \\ &= 1 + \left[-\frac{\sigma^2}{2} + e_2(t)\right] t^2 \end{aligned}$$

donde

$$\lim_{t \to +\infty} e_2(t) = 0 \qquad (1)$$

por la propiedad que tiene el resto de Taylor.

Entonces:

$$\varphi_{S_n^*}(t) = \left\{ 1 + \left[-\frac{\sigma^2}{2} + e_2\left(\frac{t}{\sigma\sqrt{n}}\right)\right] \left(\frac{t}{\sigma\sqrt{n}}\right)^2 \right\}^n$$
$$= \left\{ 1 + \left[-\frac{1}{2} + \frac{1}{\sigma^2}\, e_2\left(\frac{t}{\sigma\sqrt{n}}\right)\right] \frac{t^2}{n} \right\}^n$$

Fijado un t, si llamamos

$$c_n = \left[-\frac{1}{2} + \frac{1}{\sigma^2}\, e_2\left(\frac{t}{\sigma\sqrt{n}}\right)\right] t^2$$

vemos que

$$c_n \to c = -\frac{t^2}{2}$$

Demostración del teorema (4)

Usando el lema, deducimos que

$$\varphi_{S_n^*}(t) \to e^c = e^{-t^2/2} \text{ cuando } n \to +\infty$$

pero esta función es justamente la función característica de la distribución normal estándar $N(0,1)$.

Por el teorema de continuidad de Paul Levy), se deduce que S_n^* converge en distribución a la normal estándar, como afirma el teorema.

Aplicación a las distribuciones χ^2_n

Para dar un ejemplo del teorema del límite central, consideremos las variables

$$Z_n = X_1^2 + X_2^2 + \ldots + X_n^2$$

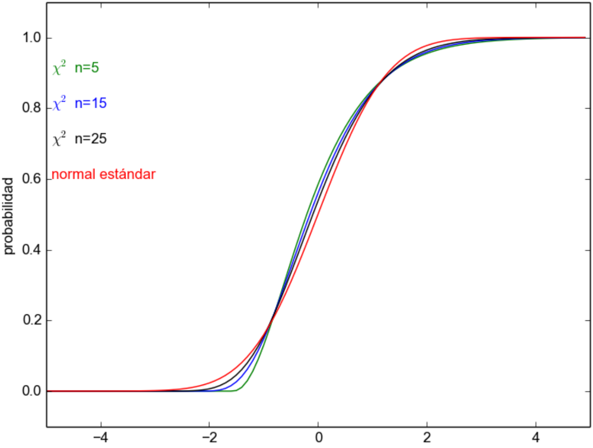
donde las (X_k) son variables con distribución normal estándar independientes. Entonces, por definición Z_n tiene distribución χ^2_n y sabemos que $E[Z_n] = n$ y $\mathrm{Var}(Z_n) = 2n$. Por el teorema del límite central, para n grande, la distribución normal proporciona una buena aproximación de la distribución χ^2_n en el sentido que las variables normalizadas

$$Z_n^* = \frac{Z_n - n}{\sqrt{2n}}$$

convergen en distribución a una normal estándar.

Aplicación a las distribuciones χ^2_n: gráfico

El siguiente gráfico compara las funciones de distribución de Z_n^* con la de la distribución normal, para n grande:



Generalizaciones y comentarios adicionales

El teorema del límite central no está limitado al caso de variables idénticamente distribuidas: se aplica en general a sumas de variables aleatorias independientes con varianza finita, donde la varianza de cada variable contribuye (en algún sentido) a la varianza total.
Voy a enunciar a continuación dos versiones más generales de este teorema (pero no veremos sus demostraciones en el curso).
Una condición muy general para su validez está dada por el siguiente teorema de Lindeberg:

El Teorema del Límite central de Lindeberg

Teorema (Teorema Límite central de Lindeberg)

Sea $(X_k)_{k\in\mathbb{N}}$ *una sucesión de variables aleatorias independientes tales que* $\mu_k = E[X_k]$ *y* $\sigma_k^2 = Var(X_k)$*, donde* σ_k *es finita y al menos algún* $\sigma_{k_0} > 0$*. Sean*

$$S_n = X_1 + X_2 + \ldots + X_k, \quad s_n = \sqrt{Var(S_n)} = \sqrt{\sigma_1^2 + \ldots + \sigma_n^2}$$

y supongamos que se cumple la siguiente condición de Lindeberg:

$$\forall\, \varepsilon > 0 \ \lim_{\varepsilon\to 0} \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x-\mu_k|>\varepsilon s_n} (x-\mu_k)^2 \, dF_{X_k}(x) = 0$$

entonces si definimos

$$S_n^* = \frac{S_n - E[S_n]}{s_n} = \frac{S_n - (\mu_1 + \mu_2 + \ldots + \mu_n)}{s_n}$$

tenemos que

$$S_n^* \xrightarrow{D} N(0,1)$$

Teorema del Límite central de Lyapunov

El teorema de Lindeberg implica el siguiente teorema de Lyapunov que da una condición más fuerte, pero quizás más fácil de entender:

Teorema (Teorema del Límite central de Lyapunov)

Sea $(X_k)_{k\in\mathbb{N}}$ una sucesión de variables aleatorias independientes tales que $\mu_k = E[X_k]$ y $\sigma_k^2 = Var(X_k)$, donde σ_k es finita y al menos algún $\sigma_{k_0} > 0$. Sean

$$S_n = X_1 + X_2 + \ldots + X_k, \quad s_n = \sqrt{Var(S_n)} = \sqrt{\sigma_1^2 + \ldots + \sigma_n^2}$$

y supongamos que existe algún $\delta > 0$ tal que se cumple la siguiente condición de Lyapunov:

$$\forall\, \varepsilon > 0 \ \lim_{\varepsilon \to 0} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} E[|X_k - \mu_k|^{2+\delta}] = 0$$

entonces

$$S_n^* = \frac{S_n - E[S_n]}{s_n} = \frac{S_n - (\mu_1 + \mu_2 + \ldots + \mu_n)}{s_n} \xrightarrow{D} N(0,1)$$