The Characteristic Function

Joseph Johnston

Spring 2018

1 Introduction

In statistics, moments offer critical information about a distribution. For random variable X, the nth moment is defined as $E(X^n)$ [4]. The zero moment is the total area under the curve, which is always 1. The first moment about the origin is the mean of the distribution. The second moment about the mean is the variance. The third moment about the mean is the skewness. Finally, the fourth moment about the mean is tail-thickness. Moments beyond the fourth can reveal more about the distribution's shape. The content in this paper will focus on continuous distributions, but discrete distributions have analogous results.

Moments are typically obtained via the moment-generating function,

$$g_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx \tag{1}$$

for density function f. Differentiating the moment-generating function n times and then evaluating at the origin yields the nth moment about the origin, from which one can derive the more commonly used moments about the mean. Another aspect of this method is the simplicity of working with a linear combination of independent random variables. If $Y = \sum_{j=1}^{n} a_j X_j$ for independent random variables X_j and real constants a_j,

$$\begin{aligned} g_Y(t) &= E(\exp\big(\sum_{j=1}^{n} a_j t X_j\big)) = E\big(\prod_{j=1}^{n} e^{a_j t X_j}\big) \\ &= \prod_{j=1}^{n} E(e^{a_j t X_j}) = \prod_{j=1}^{n} g_{X_j}(a_j t) \end{aligned}$$

.

But one problem often encountered by statisticians that prevents them from taking advantage of this convenience, is that the integral in (1) does not exist due to the unboundedness of the real exponential. A solution to this problem is to replace the real exponential e^{tX} with a complex exponential e^{itX} which *is* bounded as a result of Euler's formula. The price payed for this solution is the shift from analysis on the real line to the complex plane.

2 The Characteristic Function

Evidently, when i is inserted in the exponent of the exponential, the resulting function is equivalent to the moment-generating function with variable t replaced by it. That is, for resulting function ϕ,

$$\phi_X(t) = g_X(it) \tag{2}$$

The function ϕ_X is known as the characteristic function of random variable X, because similar to the moment-generating function it fully characterizes the random variable's distribution. When the moment-generating function exists, the characteristic function can be obtained by simply replacing every t with it in light of (2). When the moment-generating function does not exist, however, ϕ must be computed directly using the following definition.

Definition 2.1. *Characteristic Function.* For random variable X, the complex valued function $\phi : \mathbb{R} \to \mathbb{C}$ is called the characteristic function of X and is defined by

$$\phi_X(t) = E(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} dF(x) = \int_{-\infty}^{\infty} e^{itx} f(x) dx \tag{3}$$

where F and f are the cumulative and density distribution functions of X respectively [1].

The definition of ϕ may appear similar to that of the Fourier transform of f, if only the sign of the exponent was negative. It is indeed the case that ϕ is closely related to the Fourier transform, and in fact it has the special name Fourier-Stieltjes transform, which is another convention of the ordinary Fourier transform.

The characteristic function has several properties worth noting. As mentioned previously, one of the most important is that the characteristic function always exists in contrast to the moment-generating function which often fails to exist.

Theorem 2.2. *The characteristic function ϕ_X exists regardless of the distribution of X.*

Proof. Keeping in mind that density function $f(x) \geq 0$,

$$|\phi_X(t)| \leq \int_{-\infty}^{\infty} |e^{itx} f(x) dx| = \int_{-\infty}^{\infty} |f(x)||dx| = \int_{-\infty}^{\infty} f(x) dx = 1$$

□

Other evident properties are listed below.

Theorem 2.3. $\phi_X(0) = 1$

Proof.

$$\phi_X(0) = E(e^{i(0)X}) = E(1) = 1$$

□

Theorem 2.4. *The characteristic function maps to the boundary and interior of the unit circle.*

Proof. Proof of Theorem (2.2) suffices. □

Theorem 2.5. *If ϕ_X is the characteristic function of X, then $\overline{\phi_X}$ is the characteristic function of $-X$.*

Proof.

$$\begin{aligned}
\phi_X &= \int_{-\infty}^{\infty} e^{itx} f(x) dx = \int_{-\infty}^{\infty} cos(tx) f(x) dx + i \int_{-\infty}^{\infty} sin(tx) f(x) dx = u(t) + iv(t) \\
\phi_{-X} &= \int_{-\infty}^{\infty} e^{-itx} f(x) dx = \int_{-\infty}^{\infty} cos(tx) f(x) dx - i \int_{-\infty}^{\infty} sin(tx) f(x) dx = u(t) - iv(t) \\
&\implies \phi_{-X} = \overline{\phi_X}
\end{aligned}$$

A similar result can be more quickly obtained by noting the conjugation property of Fourier transforms. □

Theorem 2.6. *ϕ_X is real if and only if the distribution of X is symmetric about the origin.*

Proof. The statement that ϕ_X is real valued is equivalent to the statement that $\phi_X = \overline{\phi_X}$. From Theorem (2.4) above, this occurs strictly when the distribution of X is identical to that of $-X$. In this case, if $f(x)$ is the density function of X, then $f(x) = f(-x)$. This is the definition of a symmetric distribution. □

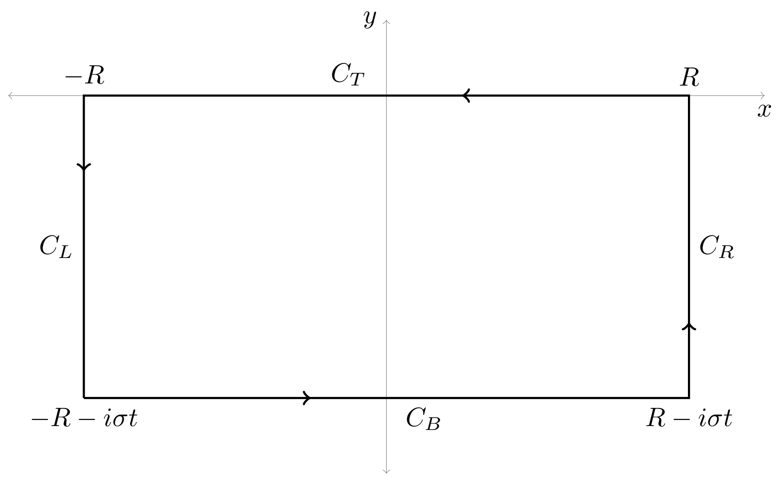

Figure 1: The closed contour around which we integrate for Example 2.7

Example 2.7. Normal Distribution Characteristic function.
This example will find the characteristic function for the normal distribution.

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

The exponent of the exponential in this integrand can be transformed as follows by completing the square

$$\begin{aligned} itx - \frac{(x-\mu)^2}{2\sigma^2} &= \frac{-1}{2\sigma^2}(x^2 - 2(\mu + it\sigma^2)x + \mu^2) \\ &= \frac{-1}{2\sigma^2}((x - (\mu + i\sigma^2 t))^2 + (\sigma^4 t^2 - i2\mu\sigma^2 t)) \\ &= \frac{-1}{2}\left(\left(\left(\frac{x-\mu}{\sigma}\right) - i\sigma t\right)^2 + \left(\frac{-t^2\sigma^2}{2} + i\mu t\right)\right) \end{aligned}$$

Let $z = (x-\mu)/\sigma - i\sigma t$. Then

$$\begin{aligned} &dz = dx/\sigma, \\ &x = \infty \implies z = \infty - i\sigma t \\ &x = -\infty \implies z = -\infty - i\sigma t \end{aligned}$$

Then the integral becomes

$$\frac{e^{-t^2\sigma^2/2 + i\mu t}}{\sqrt{2\pi}\sigma} \int_{-\infty - i\sigma t}^{\infty - i\sigma t} e^{-z^2/2} \sigma dz$$

and the task is to solve the complex integral

$$\lim_{R\to\infty} \int_{-R - i\sigma t}^{R - i\sigma t} e^{-z^2/2} dz$$

This is done by integrating along each side of C, the simple, closed contour in Figure 1. Because the integrand is analytic in the finite plane,

$$\int_C e^{-z^2/2} dz = \left(\int_{C_B} + \int_{C_R} + \int_{C_T} + \int_{C_L}\right) e^{-z^2/2} dz = 0$$

Regarding C_R and C_L, $z = \pm R - iy$, $0 < y < \sigma t$ such that

$$\begin{aligned} &Re\{-z^2/2\} = -\frac{1}{2}((\pm R)^2 - y^2) < 0 \text{ for large } R \\ &\implies \lim_{R\to\infty} Re\{-z^2/2\} = -\infty \end{aligned}$$

As a result, the integrals on C_R and C_L vanish. As for C_T, the integral is real and can be computed using the square root of the following integral

$$\begin{aligned}&\iint_{\mathbb{R}^2} e^{-(x^2+y^2)/2}dxdy = \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2} r dr d\theta \\ &= 2\pi \int_0^{\infty} e^{-p} r \frac{dp}{r} \quad \text{where } p = r^2/2,\ dp = rdr \\ &= 2\pi(-0-(-1)) = 2\pi\end{aligned}$$

But the C_T integral is integrated in the negative direction, so the value of the integral in $\lim_{R\to\infty}$ is $-\sqrt{2\pi}$. From this it can be deduced that

$$\lim_{R\to\infty}\int_{-R-i\sigma t}^{R-i\sigma t} e^{-z^2/2}dz = \sqrt{2\pi}$$

Finally, the original integral reduces to

$$\phi_X(t) = \frac{e^{-t^2\sigma^2/2+i\mu t}}{\sqrt{2\pi}\sigma}\sigma\sqrt{2\pi} = e^{-t^2\sigma^2/2+i\mu t} \tag{4}$$

This is the characteristic function of the normal distribution.

3 Linear Combinations of Random Variables

Similar to the convenience of moment-generating functions, characteristic functions are well suited for linear combinations of independent random variables. Using expected value notation, the characteristic function of a linear combination can be shows to be the product of characteristic functions of the terms in the combination.

$$\begin{aligned}&\phi_{a_1X_1+a_2X_2}(t) = E(\exp(it(a_1X_1+a_2X_2))) \\ &= E(\exp(ita_1X_1)\exp(ita_2X_2)) = E(\exp(ita_1X_1))E(\exp(ita_2X_2)) \\ &= \phi_{a_1X_1}(t)\phi_{a_2X_2}(t)\end{aligned}$$

This is also evident by interpretation of the characteristic function as a Fourier transform. From statistics the density function of a linear combination of two independent random variables can be written as the convolution of their density functions.

$$f_{a_1X_1+a_2X_2}(x_1) = \int_{-\infty}^{\infty} f_{a_1X_1}(x_1-x_2)f_{a_2X_2}(x_2)dx_2$$

Then from Fourier analysis it is applicable that the Fourier transform of a convolution of two functions is the product of the Fourier transform of those two functions [3]. That is, $\phi_{a_1X_1+a_2X_2}(t) = \phi_{a_1X_1}(t)\phi_{a_2X_2}(t)$, same as with expected value notation.

Also take note of the fact that a_j need not be paired with X_j as above, but could instead be pared with t, because either way the integral definition of ϕ is the same. That is, $\phi_{a_jX_j}(t) = \phi_{X_j}(a_jt)$.

4 Central Limit Theorem

The central limit theorem is fundamental to the theory of statistics because it offers normal distribution methods for other distributions. It states that the mean of independent and identically distributed random variables tends to a normal distribution. The variables can belong to any distribution under the condition that both mean and variance exist. The characteristic function will be the basis for the following proof of this theorem.

Theorem 4.1. *Central Limit Theorem: Suppose X_i for $i = 1, \ldots, n$ are iid random variables from a distribution with mean μ and variance σ^2. Then their mean tends to a normal distribution with $n \to \infty$ having mean μ and variance σ^2/n [3].*

Proof. The following equivalent statement will be proven instead.

$$\begin{aligned} S &= \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \\ &= \frac{n\frac{1}{n}(\sum_{i=1}^{n} X_i) - n\mu}{n\sigma/\sqrt{n}} \\ &= \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sigma\sqrt{n}} \end{aligned}$$

$$\lim_{n\to\infty} S \sim N(0,1) \tag{5}$$

First expand e^{itX} to second degree by Taylor series with little o notation for the error.

$$e^{itX} = 1 + itX - \frac{t^2X^2}{2} + o(t^2X^2)$$

Suppose we let

$$Y_i = \frac{X_i - \mu}{\sigma\sqrt{n}} = \left(\frac{1}{\sqrt{n}}\right)\frac{X_i - \mu}{\sigma/\sqrt{n}}$$

in which case

$$\begin{aligned} E(Y_i) &= \frac{1}{\sqrt{n}}E\left(\frac{X_i - \mu}{\sigma/\sqrt{n}}\right) = \frac{1}{\sqrt{n}}(0) \\ \text{Var}(Y_i) &= \frac{1}{(\sqrt{n})^2}\text{Var}\left(\frac{X_i - \mu}{\sigma/\sqrt{n}}\right) = \frac{1}{n}(1) \\ \text{Var}(Y_i) &= E(Y_i^2) - E(Y_i)^2 \implies E(Y_i^2) = \frac{1}{n} \end{aligned}$$

Then plugging Y_i into the exponential and taking the expected value gives

$$\phi_{Y_i}(t) = E(e^{itY_i}) = 1 + itE(Y_i) - \frac{t^2}{2}E(Y_i^2) + o\left(t^2\frac{(X_i-\mu)^2}{\sigma^2 n}\right) = 1 - \frac{t^2}{2n} + o\left(t^2\frac{(X_i-\mu)^2}{\sigma^2 n}\right)$$

By the property of the characteristic function of sums of independent random variables

$$\begin{aligned} \phi_S(t) = \lim_{n\to\infty}\prod_{i=1}^{n}\phi_{Y_i}(t) &= \lim_{n\to\infty}\left(1 - \frac{t^2}{2n} + \prod_{i=1}^{n} o\left(t^2\frac{(X_i-\mu)^2}{\sigma^2 n}\right)\right)^n = \lim_{n\to\infty}\left(1 - \frac{t^2}{2n}\right)^n \\ \text{because } \lim_{n\to\infty} o\left(t^2\frac{(X_i-\mu)^2}{\sigma^2 n}\right) &= 0 \end{aligned}$$

Reaching the desired outcome requires using the following identity which will not be proven here

$$\lim_{n\to\infty}(1 + c/n)^n = e^c, c \in \mathbb{R}$$

[3] A particularly well known case is for $c = 1$. In the case of this proof, $c = -t^2/2$.

$$\phi_S(t) = \lim_{n\to\infty}\left(1 - \frac{t^2}{2n}\right)^n = e^{-t^2/2}$$

The last expression is the characteristic function for the standard normal distribution, confirming equation (4). □

5 A Geometrical Interpretation

A more intuitive understanding of the characteristic function can come by visualizing e^{itX} as a random variable that maps onto the unit circle. More specifically e^{itX} can be thought of as "wrapping" the density function of X multiplied by t around the unit circle [2]. That is, if X

takes the value x, then the random variable e^{itX} maps to the point on the unit circle with angle tx. Different angles are mapped to with different probabilities, depending on the distribution of X. Suppose we wish to determine the probability e^{itX} will map to angle tx_0. This occurs when $tx = tx_0 + j2\pi$ for integer j due to the 2π periodicity of the exponential with imaginary exponent. Solving for x reveals that the probability of mapping to angle tx_0 is the sum of the values of the density function of X at points $x_0 + j2\pi/t$. At $t = 0$ the random variable maps to 1.

The characteristic function is the expected value of the random variable e^{itX} which can be interpreted as the center of mass of the distribution of tX wrapped around the unit circle. As t varies, the distribution of tX changes and so does its center of mass.

References

[1] Kai Lai Chung. *A course in probability theory.* Academic press, 2001.

[2] TW Epps. Characteristic functions and their empirical counterparts: geometrical interpretations and applications to statistical inference. *The American Statistician*, 47(1):33–38, 1993.

[3] William Feller. *An introduction to probability theory and its applications*, volume 1. Wiley, New York, 1968.

[4] Andrey Feuerverger and Philip McDunnough. On some fourier methods for inference. *Journal of the American Statistical Association*, 76(374):379–387, 1981.