Moments and Characteristic Functions
Instructor's Notes

Fu Tianwen Yao Chaorui Zhao Feng

April 6, 2019

Contents

1 Moments

1.1 Definition of Moments

Generally, in math, the n-th moment of a real-valued continuous function about center c is: [1]

$$\mu_n = \int_{-\infty}^{\infty} (x-c)^n f(x) dx$$

In particular, for probability density functions f (or cumulative density function F), the moments are given by

$$\mu_n' = E[X^n] = \int_{-\infty}^{\infty} x^n dF(x) = \int_{-\infty}^{\infty} x^n f(x) dx$$

Also we have the definition of the central moment [2]:

$$\mu_n = E[(X - E[X])^n] = \int_{-\infty}^{\infty} (x-\mu)^n f(x) dx$$

Generally central moments are more useful. Not to be confused with mean μ.

1.2 Description of Moments

The first moment is the mean of a random variable, i.e.

$$\mu = E[X]$$

The second moment is related to the variance of a random variable:

$$\mathrm{Var}[X] = E[X^2] - E[X]^2$$

In fact the variance is just the second central moment:

$$\mathrm{Var}[X] = \mu_2 = E[(X - E[X])^2]$$

As for the third central moment, a related concept is skewness. Below shows two random variables with the same mean variance however different in skewness[3]:

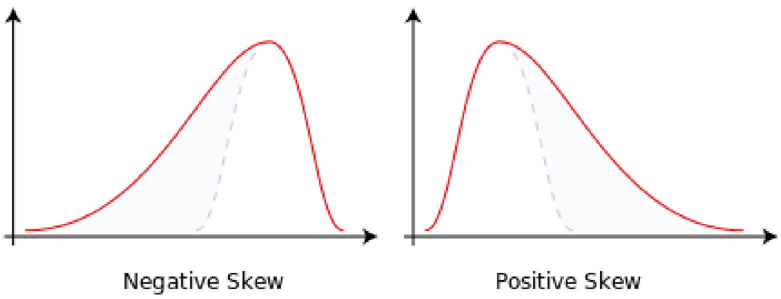

Figure 1: Negative and Positive Skew Diagrams

With all moments up to the order of infinity we can describe the **characteristics** of a probability distribution.

2 Characteristic Functions

2.1 Moment Generating Functions

Definition 2.1. Let X be a random variable with probability density function $f(x)$. If there is a positive number h such that

$$\int_{-\infty}^{\infty} e^{tx} f(x) dx$$

exists and is finite for $h < t < h$, then the function defined by

$$M(t) = E[e^{tX}]$$

is called the moment-generating function of X (or of the distribution of X). [4]

The r-th moment about the origin can be achieved from the moment generating function by evaluating the r-th derivative[5]:

$$M^{(r)}(0) = E[X^r]$$

Also notice the relation between the Taylor Expansion and the moments.

2.2 Characteristic Function

Notice that e^{tx} is not a "good" function in the sense that it is not bounded and may not converge under some circumstances. Before going to characteristic functions, we first get acquainted with knowledge of complex numbers:

2.2.1 Basic information about complex numbers

Let $z = a + bi$, where $a, b \in \mathbb{R}$, and $i = \sqrt{-1}$ is the imaginary unit. z is then called a complex number and a, b are called the real and imaginary parts of z, denoted by $a = \mathrm{Re}(z), b = \mathrm{Im}(z)$ respectively. (Consider i as rotation by $\frac{\pi}{2}$ counterclockwise in the complex plane)
The conjugate of a complex number $z = a + bi, a, b \in \mathbb{R}$ is $\hat{z} = a - bi$, we also define the modulus (or length) of z to be $|z| = z\hat{z}$. Notice that $|z|$ is a non-negative real number.
Euler's formula:

$$e^{i\theta} = \cos\theta + i\sin\theta$$

The formula comes from Taylor's Series. It also gives rise to the polar representation of a complex number, i.e. $z = re^{i\theta}$, where r is the modulus and θ is the phase.
From this we also have that $|e^{i\theta}| = 1$ for any θ.

2.2.2 Definition of Characteristic Functions

Definition 2.2. Let X be a random variable and denote by F the cumulative distribution function of X. The characteristic function $\varphi = \varphi_X$ of X (or of F, in which case we also write φ_F) is defined by [6]

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

2.2.3 Basic Properties

Theorem 2.1 (Uniqueness Theorem [7]). *Let X be a real random variable with distribution function F and characteristic function φ. Similarly, let Y have distribution function G and characteristic function ψ. If $\varphi(t) = \psi(t)$ for all $t \in \mathbb{R}$ then $F(x) = G(x)$ for all $x \in \mathbb{R}$.*

Properties from here on come from Bisgaard and Zoltan's book [6].

Theorem 2.2. *If X and Y are independent random variables then the characteristic function of their sum is*

$$\varphi_{X+Y}(t) = \varphi_X \cdot \varphi_Y.$$

Corollary 2.2.1. *The product of two characteristic functions is a characteristic function.*

Remark. If X and Y are random variables such that $\varphi_{X+Y} = \varphi_X \cdot \varphi_Y$, then in general we do not conclude X and Y are independent. (See page 13 in [6])

Theorem 2.3. *For any $a, b \in \mathbb{R}$,*

$$\varphi_{aX+b}(t) = e^{ibt}\varphi_X(at).$$

Theorem 2.4. *Every characteristic function φ has the following properties:*

(i) $\varphi(0) = 1$,

(ii) $|\varphi(t)| <= 1$,

(iii) $\varphi(-t) = \overline{\varphi(t)}$

(iv) *φ is continuous on $\mathbb{R}$*

Theorem 2.5 (Inversion Formula [8]). *If $\int_{\mathbb{R}} |\varphi(t)|dt < \infty$ then X has bounded continuous density*

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

2.3 Common Distributions and Their Characteristic Functions

Table 1: Characteristic Functions for Common Distributions[9]

Distribution	PMF/PDF	Characteristic Function
Constant $X \equiv a$	-	$\varphi_X(t) = e^{iat}$.
Binomial $X \sim Binomial(m, p)$	$p_X(n) = \binom{n}{m}p^n(1-p)^{m-n}$	$\varphi_X(t) = (pe^{it} + (1-p))^m$
Poisson $X \sim Poisson(\lambda)$	$p_X(n) = \frac{\lambda^n}{n!}e^{-\lambda}$	$\varphi_X(t) = e^{\lambda(e^{it}-1)}$
Exponential $X \sim Exponential(\lambda)$	$p_X(n) = \lambda e^{-\lambda x}$	$\varphi_X(t) = \frac{\lambda}{\lambda - it}$
Normal $X \sim N(0,1)$	$f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	$\varphi_X(t) = e^{-\frac{t^2}{2}}$
Normal $Y \sim N(\mu, \sigma^2)$	$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y-\mu)^2}{2\sigma^2}}$	$\varphi_Y(t) = e^{it\mu - \frac{\sigma^2 t^2}{2}}$

3 Examples and Applications of Characteristic Functions

Example 3.1. Rain falls on your head at λ drops per second on average. What is the distribution of rain drops on your head in two seconds?

Solution. Our intuition tells us that it should be $Poisson(2\lambda)$.
Let X, Y be two independent $Poisson(\lambda)$ random variables. Let $Z = X + Y$. Notice that characteristic functions for X (and respectively Y) is $\varphi_X(t) = e^{\lambda(e^{it}-1)}$. Therefore we have $\varphi_Z(t) = \varphi_{X+Y}(t) = (\varphi_X(t))^2 = e^{2\lambda(e^{it}-1)}$. By uniqueness of characteristic functions we know that $Z \sim Poisson(2\lambda)$.

Remark. By similar ideas, one can show that the sum of two independent poisson random variables has a possion distribution with an expectation of the sum of both expectations.

Example 3.2. $X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2)$. X_1, X_2 are independent. Distribution of $Y = X_1 + X_2$?

Solution. Similarly to last example

$$\varphi_Y(t) = \varphi_{X_1+X_2}(t) = e^{it\mu_1 - \frac{\sigma_1^2 t^2}{2}} \cdot e^{it\mu_2 - \frac{\sigma_2^2 t^2}{2}}$$

Therefore we have

$$\varphi_Y(t) = e^{it(\mu_1+\mu_2) - \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}}$$

which implies that $Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Example 3.3 (Central Limit Theorem[10])**.** If X_i are independent identically distributed random variables with $E[X_i] = \mu, Var[X_i] = \sigma^2$, then $S_n^* = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n}(X_i - \mu)$ converges weakly to $N(0, 1)$.

Warning: The proof is not rigorous and serves only for intuitively demonstrating the usage of Characteristic Functions. See the references[10, 11] for detailed proofs.

Lemma 3.1. *For any random variable X with $E[X] = 0, Var[X] = 1$, we have $\varphi_X(t) = 1 - \frac{1}{2}t^2 + o(t^2)$.*

Proof. Directly from the definition and Taylor Series we have

$$\varphi_X(t) = E[e^{itX}] = E[1 + itX + \frac{1}{2}i^2t^2X^2 + o((tX)^2)]$$

By linearity of expectation,

$$\varphi_X(t) = E[1] + itE[X] - \frac{1}{2}t^2E[X^2] + o(t^2)$$

Also

$$E[X^2] = Var[X] - (E[X])^2 = 1$$

Therefore

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

□

Now with Lemma 3.1 we denote the characteristic function of $\frac{X_i-\mu}{\sigma}$ by $\varphi(t)$, then $\varphi(t) = 1 - \frac{1}{2}t^2 + o(t^2)$. Therefore the characteristic function of S_n^* is

$$\varphi^n(t/\sqrt{n}) = [1 - \frac{t^2}{2n} + o(t^2/n)]^n$$

Take $n \to \infty$ and we get the characteristic function of $\lim_{n\to\infty} S_n^*$

$$\Phi(t) = \lim_{n\to\infty} \varphi^n(t/\sqrt{n}) = \lim_{n\to\infty} [1 - \frac{t^2}{2n}]^n = e^{-\frac{t^2}{2}}$$

Therefore $\lim_{n\to\infty} S_n^*$ converges to $N(0,1)$.

Example 3.4. Show that $Poisson(\lambda)$ is the same as $\lim_{n\to\infty} Binomial(n, \frac{\lambda}{n})$.

Solution. Take the limit of the characteristic function of the binomial distribution.

$$\begin{aligned}\lim_{n\to\infty} \varphi_B(t) &= \lim_{n\to\infty} (\frac{\lambda}{n}e^{it} + (1 - \frac{\lambda}{n}))^n \\ &= \lim_{n\to\infty} (1 + (e^{it} - 1)\frac{\lambda}{n})^n \\ &= e^{\lambda(e^{it}-1)}\end{aligned}$$

Remark. This is not a standard way of defining Poisson distribution; however, this gives us a nice intuition of Poisson distribution and can be seen as a proof for the formula of the characteristic function of it.

Lemma 3.2 (Parseval identity[9]). *F(A) and G(A) are probability measures on a real line; $\phi(t) = \int_{-\infty}^{\infty} e^{itx}F(dx)$ and $\psi(t) = \int_{-\infty}^{\infty} e^{itx}G(dx)$ are their characteristic functions. Then the Parseval Identity is as follows:*

$$\int_{-\infty}^{\infty} e^{-ity}\phi(t)G(dt) = \int_{-\infty}^{\infty} \psi(x-y)F(dx).$$

Proof. By the definition of characteristic functions, we have:

$$e^{-ity}\phi(t) = \int_{-\infty}^{\infty} e^{it(x-y)}F(dx)$$

Integrating it w.r.t. G(A), we get:

$$\int_{-\infty}^{\infty} e^{-ity}\phi(t)G(dt) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{it(x-y)}F(dx)G(dt)$$

Using Fubini's Theorem, we get:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{it(x-y)}F(dx)G(dt) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{it(x-y)}G(dt)F(dx) = \int_{-\infty}^{\infty} \psi(x-y)F(dx)$$

□

References

[1] Wikipedia contributors, "Moment (mathematics) — Wikipedia, the free encyclopedia." https://en.wikipedia.org/w/index.php?title=Moment_(mathematics)&oldid=886051824, 2019. [Online; accessed 8-March-2019].

[2] Wikipedia contributors, "Central moment — Wikipedia, the free encyclopedia." https://en.wikipedia.org/w/index.php?title=Central_moment&oldid=866322488, 2018. [Online; accessed 8-March-2019].

[3] Wikipedia contributors, "Skewness — Wikipedia, the free encyclopedia." https://en.wikipedia.org/w/index.php?title=Skewness&oldid=886052950, 2019. [Online; accessed 8-March-2019].

[4] R. V. Hogg, *Probability and statistical inference*. 9th ed., 2015.

[5] Penn State University, "Stat 414 / 415 probability theory and mathematical statistics — finding moments." https://newonlinecourses.science.psu.edu/stat414/node/73/.

[6] T. M. Bisgaard and S. Zoltán, *Characteristic functions and moment sequences : positive definiteness in probability*. Huntington, N.Y.: Nova Science Publishers, 2000.

[7] M. J. Wichura, " Characteristic functions." https://galton.uchicago.edu/~wichura/Stat304/Handouts/L12.cf2.pdf. [Online; accessed 4-April-2019].

[8] R. Durrett, "Probability: Theory and examples." https://services.math.duke.edu/~rtd/PTE/PTE5_011119.pdf, 2019. [Online; accessed 5-April-2019].

[9] Aktuella kurssidor vid Matematiska institutionen Stockholms universitet , "Lecture 10: Characteristic Functions." https://kurser.math.su.se/pluginfile.php/5149/mod_resource/content/1/lecture-10b.pdf. [Online; accessed 4-April-2019].

[10] D. L. McLeish, "Characteristic functions and the central limit theorem." http://sas.uwaterloo.ca/~dlmcleis/s901/chapt6.pdf. [Online; accessed 5-April-2019].

[11] R. Su, "Lecture 7." http://www.math.nus.edu.sg/~matsr/ProbI/Lecture7.pdf. [Online; accessed 5-April-2019].