Normal Distribution : Basic Properties and Testing for Normality

Yinglan Deng ID : 201800090100

Abstract – It is well-known that the normal distribution is one of the most important distributions for continuous variables. This Project, about the normal distribution, consists of two parts. In Part 1, I will introduce some basic, but important and interesting properties of the normal distribution, most of which we have learned from our Probability course. As the core theorem of the applied statistics, the central limit theorem is specifically discussed. In Part 2 the Histograms is discussed in detail, which is one of the methods of testing whether a data set comes from a normal distribution.

Keywords: *Normal Distribution, Test for Normality*

1. INTRODUCTION

The normal distribution, also called Gaussian distribution, is central in Probability, Statistics and their applications. People are willing to make the a priori assumption that a specific set of data is normally distributed because it can characterize the distribution of large numbers of variables. Another reason for the importance of the normal distribution is that scientists have established a simple and computationally feasible sample theory for a series of vital statistics under the normal distribution. Also for a specific system, as long as the various objects are independent or weakly related, their total average performance, according to the central limit theorem (CLT), is Gaussian or nearly Gaussian.

However, sometimes there is no definite conclusion on the distribution of a large amount of observations, or of samples of big sizes. There exists cases when the data follow a distribution that exhibits a large skewness or kurtosis, then there is a high possibility that the distribution is heavy-tailed, which is far from the normal.

1.1. BASIC PROPERTIES OF THE NORMAL DISTRIBUTION

Density Function and Distribution Function

We say that the random variable (r.v.) X follows a **normal distribution with parameters** μ **and** σ^2, we write $X \sim \mathcal{N}(\mu, \sigma^2)$, if X takes values in the real line $\mathbb{R} = (-\infty, \infty)$ and its density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, \qquad x \in \mathbb{R}, \quad \sigma > 0, \quad \mu \in \mathbb{R}.$$

The corresponding distribution function (d.f.) of X is

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} \mathrm{e}^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2} \mathrm{d}y, \qquad x \in \mathbb{R}.$$

Of great importance is the case when $\mu = 0$, $\sigma = 1$. In this case we use the notation $Z \sim \mathcal{N}(0, 1)$, and say that Z is a **standard normal r.v.**, whose d.f. is called a **standard normal distribution function**. We use Φ and φ as the notations of the corresponding standard normal d.f. and the standard normal density function:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{u^2}{2}} \mathrm{d}u, \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

The normal density function is sometimes informally called a bell curve. This is because the graph of the function has a shape of a bell, which is symmetric about $x = \mu$, and takes the maximum value $f_{max} = f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$.

Property 1. If X is a normal r.v. and $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma}$ is a normal standard r.v.. This can be written as:

$$X \sim \mathcal{N}(\mu, \sigma^2) \quad \Leftrightarrow \quad Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1).$$

By changing the variables when integrating, we can easily derive the d.f. of the standard normal distribution.

Here is an equivalent narrative of Property 1:
If $Z \sim \mathcal{N}(0,1)$, $\mu \in \mathbb{R}$ and $\sigma > 0$, then $X = \sigma Z + \mu \sim \mathcal{N}(\mu, \sigma^2)$. Conversely, if $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$.

We also have the relations: $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$, $\quad f(x) = \varphi\left(\frac{x-\mu}{\sigma}\right)$, $\quad x \in \mathbb{R}$.

Moments of a standard normal r.v. Z

There are many sources containing methods for calculating the moments, e.g., the paper [1]. Our knowledge in Mathematical Analysis should be enough to do this.

For a r.v. $X \sim F$ with arbitrary d.f. F, the moments $\mathsf{E}[X^k]$, $\quad k = 1, 2, \ldots$ exists, if $\mathsf{E}\left[|X|^k\right] < \infty$ for all k: $\int |x|^k dF(x) < \infty$. (If there is a density $f = F'$, then $\int |x|^k f(x) dx < \infty$.)

Since any r.v. $X \sim \mathcal{N}(\mu, \sigma^2)$ can be written as $X = \sigma Z + \mu$, we have

$$\mathsf{E}\left[X^k\right] = \mathsf{E}\left[(\sigma Z + \mu)^k\right] = \sum_{j=0}^{k} \binom{k}{j} \sigma^j \mu^{k-j} \mathsf{E}\left[Z^j\right].$$

Here we used the Newton's binomial formula and changed the order of expectation and summation.

We will only focus on the moments of standard normal r.v. $Z \sim \mathcal{N}(0,1)$, since any moment $\mathsf{E}\left[X^k\right]$ can be found from the above formula.

In this part we are going to derive explicit expressions for the moments $\mathsf{E}[Z^{2k-1}]$ and $\mathsf{E}[Z^{2k}]$ for $k \in \mathbb{N} = \{1, 2, ...\}$, and $\mathsf{E}\left[|Z|^p\right]$ for any real $p > 0$.

For the moment of order n, we use the notation

$$m_n = \mathsf{E}[Z^n], \qquad n = 1, 2, ...; \qquad m_0 = 1.$$

By definition we have

$$m_n = \mathsf{E}[Z^n] = \int_{-\infty}^{\infty} x^n \varphi(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^n \mathrm{e}^{-\frac{x^2}{2}} \mathrm{d}x.$$

For any $k \in \mathbb{N}$, we find that the moments of odd order vanish, i.e.,

$$\mathsf{E}[Z^{2k-1}] = 0,$$

because the integrand is an odd function on $(-\infty, \infty)$.

Let us find the moment of order $n = 2k$, i.e. the even order moment:

$$\begin{aligned} m_{2k} = \mathsf{E}\left[Z^{2k}\right] = & \int_{-\infty}^{\infty} x^{2k} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x^2} \mathrm{d}x \\ = & \sqrt{\frac{2}{\pi}} \int_0^{\infty} (2k-1) x^{2(k-1)} \mathrm{e}^{-\frac{1}{2}x^2} \mathrm{d}x \\ = & \sqrt{\frac{2}{\pi}} (2k-1)!! \int_0^{\infty} \mathrm{e}^{-\frac{1}{2}x^2} \mathrm{d}x \\ = & (2k-1)!! = (2k-1)(2k-3) \ldots 3 \cdot 1. \end{aligned} \tag{1}$$

Starting with $Z \sim \mathcal{N}(0,1)$, we define $|Z|$, which is called a **half-normal r.v.**. The values of $|Z|$ are positive and the density function is

$$f_{|Z|}(x) = 2\varphi(x), \qquad x > 0.$$

Let us find the moments of $|Z|^p$, for any real $p > 0$:

For this part, I refer to the calculation in [5]:

$$\begin{aligned}\widetilde{m}_p = \mathsf{E}\left[|Z|^p\right] = &\int_0^\infty 2x^p \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}x^2} \mathrm{d}x \\ = &\frac{2}{\sqrt{\pi}} \int_0^\infty (\sqrt{2}x)^p \mathrm{e}^{-x^2} \mathrm{d}x \\ = &\frac{2^{p/2}}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right).\end{aligned} \tag{2}$$

Here Γ is the Eluer gamma function, $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\mathrm{e}^{-\mathrm{x}}\mathrm{d}x, \quad \alpha > 0.$

Now, as is said before, we can use our findings for Z to get the moments for any r.v. $X \sim \mathcal{N}(\mu, \sigma^2)$ by using Property 1.

Characteristic Function, Moment Generating Function and Cumulant Generating Function

The **moment generating function** (m.g.f.) of a real r.v. X is the expected value of e^{tX}. For a r.v. $X \sim \mathcal{N}(\mu, \sigma^2)$, the m.g.f. can be easily derived by a standard integration and change of variables:

$$M(t) = \mathsf{E}[\mathrm{e}^{tX}] = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right), \qquad t \in \mathbb{R}.$$

Similarly we can find the **characteristic function** (ch.f.) of X:

$$\Psi(t) = \mathsf{E}\left[\mathrm{e}^{itX}\right] = \exp\left(i\mu t - \frac{1}{2}t^2\sigma^2\right), \quad t \in \mathbb{R}.$$

Notice: $M(t)$ is real-valued and, in general, $\Psi(t)$ is complex-valued, here $t \in \mathbb{R}$. If the r.v. X is symmetric, its mean value is zero, and $\Psi(t)$ is real-valued.

Write down the ch.f. and the m.g.f. of a standard normal r.v. $Z \sim \mathcal{N}(0,1)$:

$$\Psi_Z(t) = \mathrm{e}^{-\frac{t^2}{2}}, \quad \mathrm{M}_Z(t) = \mathrm{e}^{\frac{t^2}{2}}, \quad t \in \mathbb{R}.$$

The relation between the m.g.f. and the ch.f. of any r.v.X is as follows:

$$\Psi(t) = M(it), \quad M(t) = \Psi(-it).$$

Both M and Ψ can be used to determine uniquely the d.f. of X. This can be done by using the so-called inversion formula from our course in Fourier analysis.

As an illustration, we use m.g.f. to find the moments conveniently. We can calculate the first two moments m_1 and m_2 as follows:

$$M^{'}(t) = (\mathsf{E}[\mathrm{e}^{tX}])^{'}_{\mathrm{t}} = \mathsf{E}[X\mathrm{e}^{tX}] \Rightarrow M^{'}(0) = \mathsf{E}[X] = m_1,$$

$$M^{''}(t) = ((\mathsf{E}[\mathrm{e}^{tX}])^{'}_{\mathrm{t}})^{'}_{\mathrm{t}} = \mathsf{E}[X^2\mathrm{e}^{tX}] \Rightarrow M^{''}(0) = \mathsf{E}[X^2] = m_2.$$

In general, we have

$$m_k = \mathsf{E}[X^k] = M^{(k)}(0), \quad k = 1, 2, \ldots$$

Finally, we recall the definition of the **cumulant generating function** (cum.g.f.), say g. It is the logarithm of the m.g.f. Hence for $X \sim \mathcal{N}(\mu, \sigma^2)$ we have

$$g(t) = \ln M(t) = \mu t + \frac{1}{2}\sigma^2 t^2, \qquad t \in (-\infty, \infty).$$

In parallel with the moments, we can consider other numbers called **cumulants**, or **semi-invariants**. The symbol κ_n is used for the cumulant of order n, $\quad n = 1, 2, ...$ They are the coefficients in the Taylor expansion of $g(t)$. For $X \sim \mathcal{N}(\mu, \sigma^2)$, we have:

$$\kappa_1 = \mu, \quad \kappa_2 = \sigma^2, \quad \kappa_n = 0 \quad for \quad n = 3, 4, ...$$

Since the cum.g.f. of $X \sim \mathcal{N}(\mu, \sigma^2)$ is a quadratic polynomial in t, only the first two cumulants, κ_1 and κ_2, are nonzero, and they are equal to the mean μ and the variance σ^2, respectively. All other cumulants, starting from κ_3, are equal to zero.

Bivariate Normal Distribution

Now we turn to r.v.s of a higher dimension. We start with dimension 2, it is also accepted to use the term 'bivariate', so we are going to discuss the properties of bivariate r.v.s and bivariate distributions.

We say that (X, Y) is a **bivariate normal r.v.** (2-dimensional normal random vector) and write $(X, Y) \sim \mathcal{N}_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, if its density is

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right].$$

Here $\mu_1, \mu_2 \in (-\infty, \infty), \quad \sigma_1, \sigma_2 \in (0, \infty), \quad \rho \in (-1, 1)$.

The values of (X, Y) are in $\mathbb{R}^2$, so we just say that $x, y \in \mathbb{R}$.

Property 2. Both X and Y are normal: $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

Consider the following two decomposition formulas for the bivariate density $f(x, y)$:

$$\begin{aligned} f(x, y) &= \frac{1}{\sqrt{2\pi}\sigma_1} \mathrm{e}^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} \mathrm{e}^{-\frac{\left[y-(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1))\right]^2}{2\sigma_2^2(1-\rho^2)}}, \\ f(x, y) &= \frac{1}{\sqrt{2\pi}\sigma_2} \mathrm{e}^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}} \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \mathrm{e}^{-\frac{\left[x-(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2))\right]^2}{2\sigma_1^2(1-\rho^2)}}. \end{aligned} \tag{3}$$

The key for the first decomposition formula in (3) is to write $\frac{(x-\mu_1)^2}{\sigma_1^2}$ in the exponent part of the density function, as $(1-\rho^2)\frac{(x-\mu_1)^2}{\sigma_1^2} + \rho^2\frac{(x-\mu_1)^2}{\sigma_1^2}$. Similarly, we obtain the second formula in (3).

Now let us find the marginal distribution function F_X and F_Y. Since in general $f_X = F_X'$ and $f_Y = F_Y'$, we first find the marginal densities f_X and f_Y. We have

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy \qquad (by \quad (3)) \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} e^{-\frac{\left[y-(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1))\right]^2}{2\sigma_2^2(1-\rho^2)}} dy. \end{aligned} \tag{4}$$

Setting $t = \frac{1}{\sqrt{1-\rho^2}}\left(\frac{y-\mu_2}{\sigma_2} - \rho\frac{x-\mu_1}{\sigma_1}\right)$, it is clear that the integral in (4) equals the total mass of a density, so it is 1. Hence

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \quad \Rightarrow \quad X \sim \mathcal{N}(\mu_1, \sigma_1^2).$$

Similarly, we get

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}} \quad \Rightarrow \quad Y \sim \mathcal{N}(\mu_2, \sigma_2^2).$$

Knowing the marginal densities f_X and f_Y we easily express the marginal d.f.s F_X and F_Y in an integral form.

Conclusion. If $(X,Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, then $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

Therefore, we know the meaning of the first four parameters:

μ_1 and σ_1^2 are the corresponding mean and variance of X,

μ_2 and σ_2^2 are the corresponding mean and variance of Y.

However, one **Question** still remains: What is ρ?

Answer: The parameter ρ in the family $\mathcal{N}_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ is **the correlation coefficient** between X and Y. The notation is $\rho = corr(X,Y)$.

Recall that, in general for the two r.v.s X and Y with finite variances,

$$\rho = corr(X,Y) = \frac{\mathsf{E}\left[(X - \mathsf{E}X)(Y - \mathsf{E}Y)\right]}{\sqrt{Var(X) \cdot Var(Y)}}.$$

Since we know the density $f(x,y)$ of the random vector (X,Y) and also the means and the variances of X and Y, after a careful calculation we obtain that indeed $\rho = corr(X,Y)$.

If $\rho = 0$, we say that X and Y are **uncorrelated**, which as a property is weaker than independence, If X and Y are independent, then they are uncorrelated. The converse implication is not true. There are **counterexamples**.

Example 1. Consider a r.v. A that is continuous and uniform, with its density function to be $\frac{1}{2}$ on $(-1,1)$, and 0 outside. Define a r.v. B, $B = A^2$. We find that $\mathsf{E}[A] = 0$, $\mathsf{E}[A^2] = \frac{1}{3} = \mathsf{E}[B]$. Also $A \cdot B = A^3$, and $\mathsf{E}[A^3] = 0$. It follows easily from these calculations, that $\rho(A,B) = 0$, i.e., A and B are uncorrelated. However it is obvious that A and B are not independent. They are functionally dependent, hence dependent in probabilistic sense.

Notice that, in this example, there is no normal distribution. So let us go to the next example involving the normal distribution.

Example 2. Start with a r.v. $Z \sim \mathcal{N}(0,1)$ and define $X = Z$ and $Y = Z^2$. What can we say about the bivariate (X,Y)?

We see that $X \sim \mathcal{N}(0,1)$, its density function is $f_X = \varphi$; $Y \sim \chi^2_1$, its density function is $f_Y(y) = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} \mathrm{e}^{-\frac{x}{2}}$, $x > 0$. Since $X \cdot Y = Z^3$, we easily check that the density function of Z^3 is symmetric w.r.t. 0, hence all odd order moments of Z^3 are equal to zero, e.g., $\mathsf{E}[Z^3] = 0$. This implies that $\rho(X,Y) = 0$, i.e., X and Y are uncorrelated. However it is obvious that X and Y are not independent. They are functionally dependent, hence dependent in probabilistic sense.

In the case of normal distributions, we have the following remarkable fact:

Corollary 1. $\rho = 0 \iff X$ and Y are independent.

In this case, the density of (X, Y) is the product of the densities of X and Y:

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2} exp\left[-\frac{1}{2}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right], \quad x \in \mathbb{R}, y \in \mathbb{R}.$$

Here comes an interesting phenomenon that, in our real life, many things fit the normal distribution well. For example, when we take a look at the English score at many schools, we can see most students' score to be around the mean score, but we could hardly see someone to have super high or super low grades.

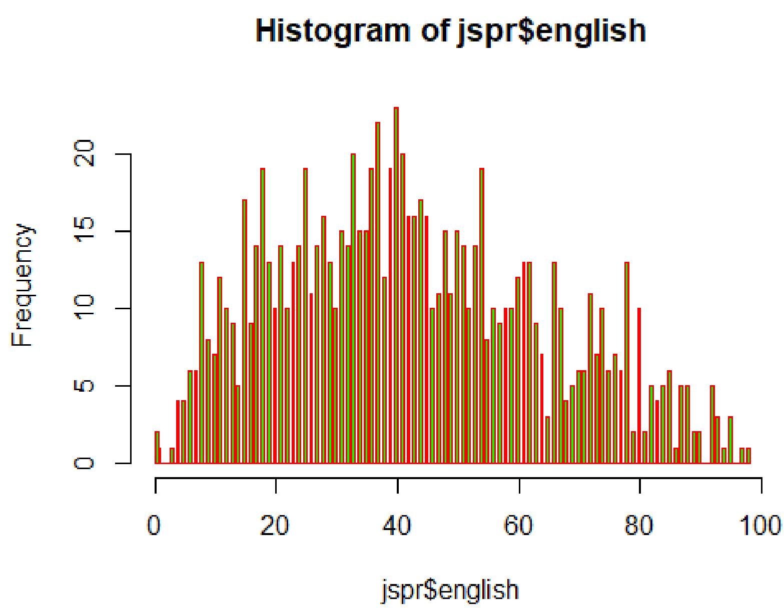

Figure 1. Histogram of scores.

The main tool to make this plot is the **Histogram**, which will be discussed later.

Now we are going to take a look at the main reason behind those phenomenon, that is, the Central Limit Theorem.

Central Limit Theorem (CLT)

The **central limit theorem (CLT)** states that, under certain conditions, the sum of a large number of r.v.s, properly centered and normalized, follows the normal distribution, no matter what their original distributions are.

Formally, we observe a sequence of independent r.v.s, say $X_1 \sim F_1$, $X_2 \sim F_2$,..., $X_n \sim F_n$. We assume that the mean values and the variances are finite, and we use the following notations:

$$S_n = X_1 + \cdots + X_n, \quad a_j = \mathsf{E}[X_j], \quad A_n = \mathsf{E}[S_n] = a_1 + \cdots + a_n,$$

$$b_j^2 = Var[X_j], \quad B_n^2 = Var[S_n] = b_1^2 + \cdots + b_n^2.$$

We are mainly interested in the following random quantity Z_n and its d.f. F_n:

$$Z_n = \frac{S_n - A_n}{B_n}, \quad F_n(x) = \mathsf{P}[Z_n \le x], \quad x \in \mathbb{R}.$$

Notice, this means, that we deal with the **standardized** r.v. Z_n based on the sequence $\{X_j\}$, since we have $\mathsf{E}[Z_n] = 0$ and $Var[Z_n] = 1$ for any n.

It turns out that, under general conditions, we have

$$Z_n \quad \xrightarrow{d} \quad Z \quad as \quad n \to \infty, \quad where \quad Z \sim \mathcal{N}(0,1).$$

The symbol $\xrightarrow{d}$ means convergence of Z_n to Z in distribution:

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P[Z_n \le x] = \Phi(x), \quad x \in \mathbb{R}.$$

Recall that Φ is the standard normal d.f., $Z \sim \Phi$.

The above statement is called the **central limit theorem** (CLT), and its universality is obvious, it is true for arbitrary d.f.s F_j. There is a rich literature, books and papers on this topic, and the exact conditions and the proofs can be learned during special advanced university courses.

It is useful to formulate the CLT for a sequence of independent and identically distributed (i.i.d.) r.v.s. Now start with a r.v. $X \sim F$, the mean value $a = \mathsf{E}[X]$ and variance $b^2 = Var[X]$. As is said before, here F is arbitrary. Now consider a sequence of i.i.d. r.v.s $X_1, \ldots, X_n$ that follows the same distribution as X. We have

$$S_n = X_1 + \cdots + X_n, \quad \mathsf{E}[S_n] = na, \quad Var[S_n] = nb^2.$$

Moreover,

$$Z_n = \frac{S_n - na}{b\sqrt{n}} \quad \xrightarrow{d} \quad Z, \quad as \quad n \to \infty.$$

There are several limit theorems showing convergence of the d.f. F_n of Z_n to Φ (Chebyshev, Lyapunov, Lindeberg, Feller). Some of the proofs are using m.g.f.s. Let us briefly illustrate this in the case of i.i.d. r.v.s.

Indeed, if we write the m.g.f. of the r.v. X, $M(t) = \mathsf{E}[\mathrm{e}^{tX}]$, which exists for $t \in (-t_0, t_0)$, fixed $t_0 > 0$, we expand the exponent and obtain

$$M(t) = 1 + \frac{a}{1!}t + \frac{b^2}{2!}t^2 + o(t^2) \quad \Rightarrow \quad \widetilde{M}(t) = 1 + \frac{1}{2}t^2 + o(t^2),$$

where $\widetilde{M}(t)$ is the m.g.f. of $\frac{X-a}{b}$.

The next step is to calculate the m.g.f. of Z_n:

$$M_n(t) = \mathsf{E}\left[\mathrm{e}^{tZ_n}\right]$$

By using the independence of X_i and the linearity of the expectation, we obtain

$$M_n(t) = \left(\widetilde{M}\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 + \frac{1}{2n}t^2 + o(t^2)\right)^n \to \mathrm{e}^{\frac{t^2}{2}}, \quad \text{as} \quad n \to \infty.$$

Since $e^{\frac{t^2}{2}}$ is the m.g.f. of $Z \sim \mathcal{N}(0,1)$, we conclude the convergence in distribution of Z_n to Z, i.e. the validity of the CLT.

Historically, the first result of this kind is the De Moivre-Laplace theorem, establised in 1733. It concerns the sum

$$S_n = X_1 + \cdots + X_n$$

of n independent Bernoulli r.v.s, each X_j takes value 1 with probability p and value 0 with probability $1-p$. Thus S_n has a binomial distribution, $S_n \sim Bin(n,p)$, and it was shown that

$$\mathsf{P}\left[\frac{S_n - np}{\sqrt{np(1-p)}} \leq x\right] \to \Phi(x), \quad as \quad n \to \infty, \quad x \in \mathbb{R}.$$

We have met the similar results for other distributions such as the Poisson, Exponential and the χ^2 distributions in our Probability course. We do not provide details here. If anyone is interested, a lot of details can be seen in [4].

2. TESTS FOR NORMALITY

A fundamental problem in Applied Statistics is the following. We have n observations, that is $X_1, \ldots, X_n$, of a r.v.X whose d.f. F is unknown, and our goal is to find F. The method from Kolmogrov-Smirov, which can be seen in [2], may be one that can be applied widely. It introduces the empirical d.f. $\widehat{F}_n$ based on the observations, and the fact that $\widehat{F}_n(x) \to F(x)$, as $n \to \infty$, at any $x \in \mathbb{R}$ is a continuity point of F. We do not give details here and move to a similar question assuming that the unknown distribution is the normal distribution. Among the many methods, we have chosen to consider the graphical method in detail.

2.1. GRAPHICAL METHOD

Assume Y is a continuous r.v., i.e. $Y \sim F$, where the d.f. F has a derivative, $F' = f$, the density of Y. Assume that we have in our disposal n independent observations, $Y_1, ..., Y_n$; we also say that $Y_1, ..., Y_n$ is a sample of size n.

We are going to use **histograms**, representing the available data graphically. First we show the arguments for this. For a fixed positive integer m, we take points $t_0 < ... < t_m$ and distribute the values of Y into m intervals. Notice that t_0 should be smaller than the minimal observation, and t_m should be larger than the maximum observation. The value range of the observed values of Y is covered by the interval

$$(t_0, t_m] = \bigcup_{k=1}^{m} (t_{k-1}, t_k].$$

Let $p_k = \mathsf{P}[t_{k-1} < Y \leq t_k] = \int_{t_{k-1}}^{t_k} f(y)dy$. According to the mean value theorem from the Mathematical Analysis course, we have

$$p_k = f(\xi_k)(t_k - t_{k-1}), \quad where \quad \xi_k \in (t_{k-1}, t_k)$$

The number of observations of Y falling in the interval $(t_{k-1}, t_k]$ is denoted by U_k, then the ratio $\frac{U_k}{n}$ indicates the probability that Y belongs to $(t_{k-1}, t_k]$.

There is a general result, in fact the LLN, telling that if we choose the points $\{t_k\}_{k=1}^{m}$ s.t. $\Delta = \max_k\{|t_k - t_{k-1}|\} \to 0$, then

$$\frac{U_k}{n} \to p_k, \quad as \quad n \to \infty, \quad for \quad k = 1, 2, ..., m.$$

This implies that, as $n \to \infty$, we have

$$\left| \frac{U_k}{n(t_k - t_{k-1})} - f(y) \right| \to 0 \quad a.s., \quad y \in (t_{k-1}, t_k].$$

Suppose that the density f is unknown, and we only get the samples $Y_1, ..., Y_n$.

So we define the empirical density function f_n as follows:

$$f_n(y) = \begin{cases} \sum_{k=1}^{m} \frac{U_k}{n(t_k - t_{k-1})} \mathbf{1}_{(t_{k-1}, t_k]}(y), & \text{if} \quad t_0 < y \leq t_m, \\ 0, & \text{if} \quad y \leq t_0 \quad or \quad y > t_m. \end{cases}$$

From the above, we have the important relation that

$$f_n(y) \to f(y), \quad as \quad n \to \infty.$$

The interpretation of this result is that the empirical density f_n is a "good" approximation for the unknown density f; larger sample size n, better approximation.

Let us use the numbers $U_1, U_2, \ldots, U_m$ and the intervals $(t_0, t_1], (t_1, t_2], \ldots, (t_{m-1}, t_m]$. We choose a scale in the direction Ox and a scale in the direction Oy in the plane, and draw m rectangles. Rectangle number j has a base $(t_{j-1}, t_j]$ and a height U_j, where $j = 1, 2, \ldots, m$.

As a result we obtain a picture called a **histogram**. Hence, the histogram is an easy way to represent graphically samples coming from a r.v. $Y \sim f$. For a specific density f, the histogram has the same shape, and for large $n-$samples, the histogram will almost coincide with f. This means that from the shape of the histogram, we can identify the density f.

Now based on the available data, we can draw the histogram to present the distribution of the data. And if the histogram has the shape of a "Bell Curve", then one can preliminary claim that $f = \varphi$, the data follow the normal distribution.

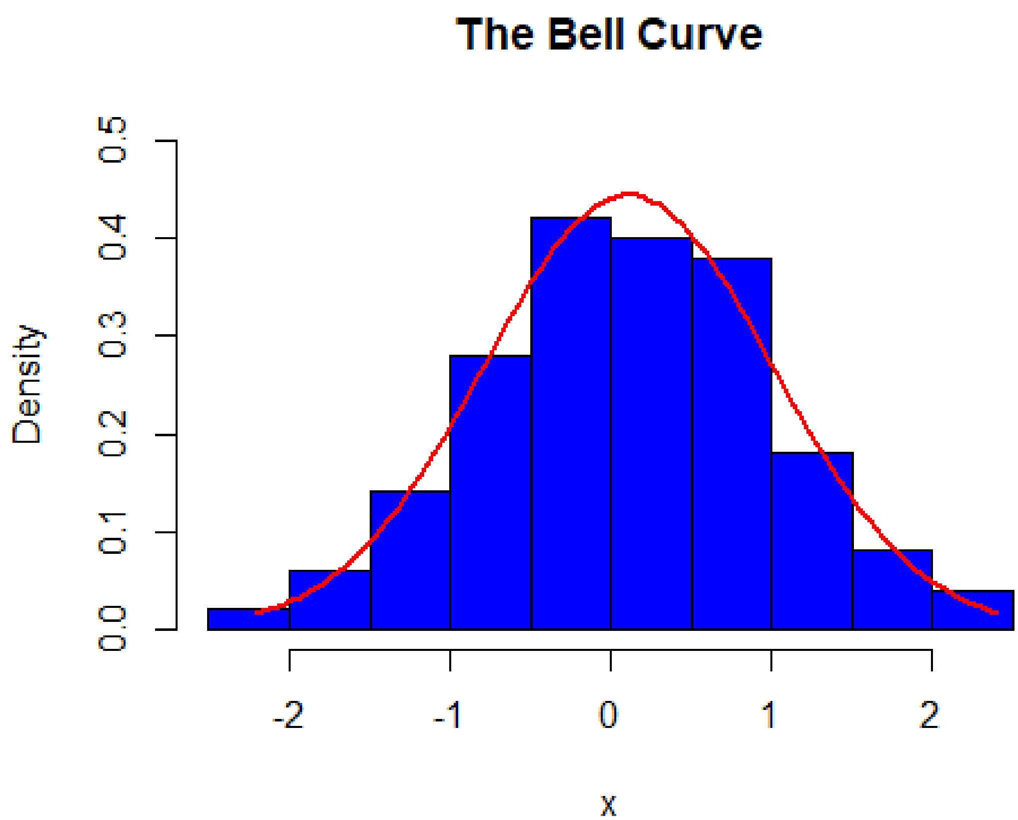

Figure 2. The Bell Curve

Remark. As is said before, there are many and diverse methods to test the normality of a data set. If someone is interested in this topic, there are many resources, e.g., [6] and [3]. And when it comes to a specific real world problem, we will choose a method which we find to be the most appropriate, and our choice depends on our theoretical knowledge and based on our practical experience. Important is to be well familiar with the numerous case studies available in the literature.

Reference

[1] Lefebvre, M. (2010). Nearly Gaussian distributions and application. *Communications in Statistics - Theory and Methods*, 39(5):823–836.

[2] Mohd Razali, N. and Yap, B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics*, Vol.2:21–33.

[3] Thode, H. C. (2002). *Testing For Normality.* CRC Press, 1 edition.

[4] Wikipedia (2021). Article: Normal distribution — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Normal%20distribution&oldid=1023017445. [Online; accessed 15-May-2021].

[5] Winkelbauer, A. (2014). Moments and Absolute Moments of the Normal Distribution. https://arxiv.org/abs/1209.4340v2. [math.ST].

[6] Xinghua, M. and Jinxin, Z. (2014). The comparison among the common normality tests for numerical variables[in Chinese]. *The Journal of Evidence-Based Medicine*, Vol.14,No.2:123–128.