Normal Distribution

Haiyue Liu

Dr. Hubert Bray, Duke University

Mathematics of the Universe

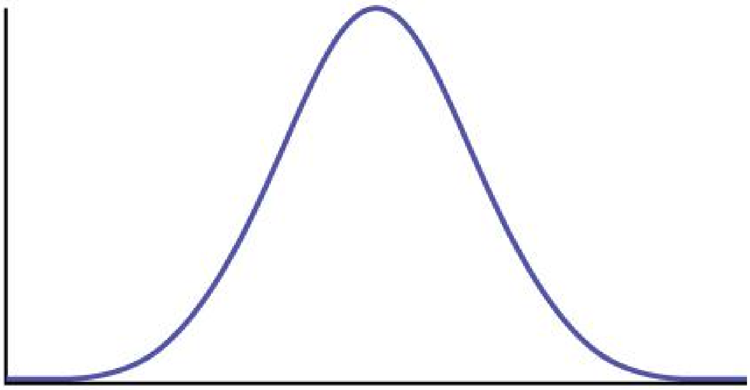
July 31, 2017

Abstract:

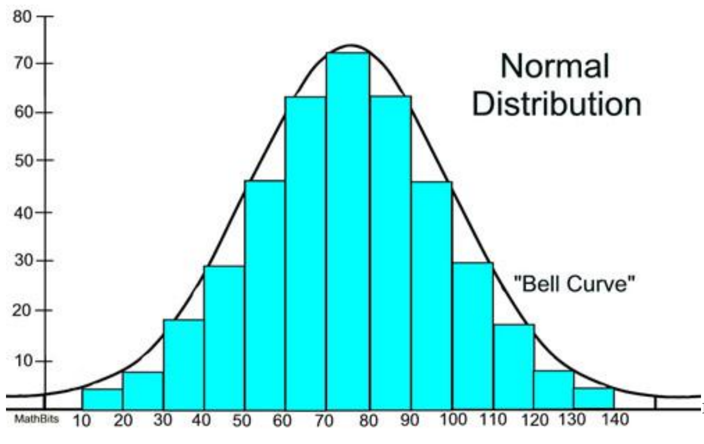
In this paper, we will explore the splendid world of normal distribution in statistics, including its applications in various areas and theories behind it.

Introduction:

First, let’s look at the line below, on which is bestowed perfect symmetry and smooth curve.



This is the image of **normal distribution**, also known as **Gaussian distribution** or **bell curve**. It is possibly the most fundamental distribution in statistics, on the basis of which we can make further inferences. Moreover, a tremendous amount of models in the **natural and social sciences** coincidentally agrees with it. Therefore, in these areas, normal distribution works very well to represent real-valued random variables whose distributions are unknown.



[1]

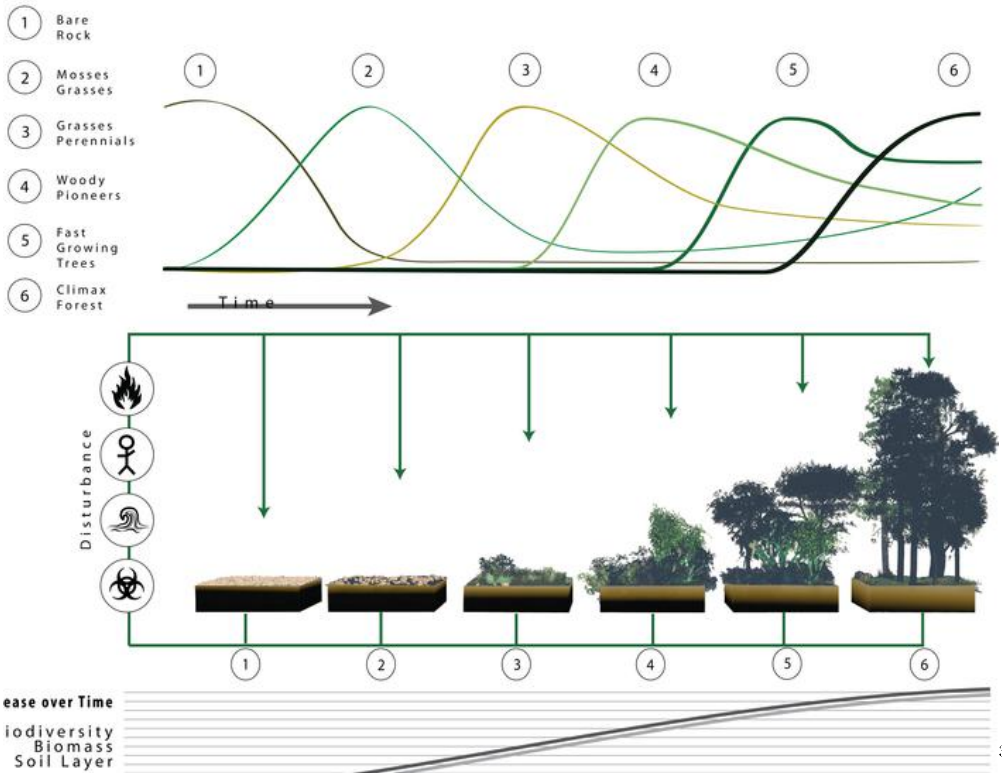
---

[1] https://mathbitsnotebook.com/Algebra2/Statistics/normalturqa.jpg

But, of course, normal distribution does not work simply because others resemble it. The broad applicability is credited to the **Central Limit Theorem (CLT)**, which I will introduce in this paper. Before unveiling the essence of normal distribution, we are supposed to learn its applications in almost every walk of life.

Normal Distribution in Ecological Succession(in forest):

Ecological succession is the natural replacement of plant or animal species, or species associations, in an area over time. When we discuss forest succession, we are usually talking about replacement of tree species or tree associations, the whole process of which is usually divided into six stages as below.[2]



[3]

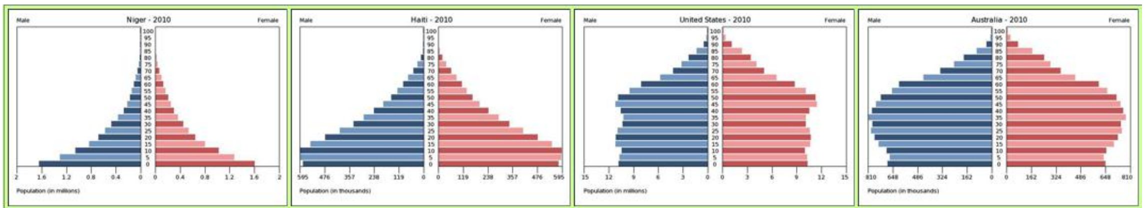
2 Forest Succession, Jeff Martin and Tom Gower, November, 1996

3 https://upload.wikimedia.org/wikipedia/commons/thumb/4/41/Forest_succession_depicted_over_time.png/800px-Forest_succession_depicted_over_time.png

As time goes by, each curve except the sixth one reaches its vertex and then decreases; to put it in another way, as the evolution takes place, these species dominate the ecosystem in turn. The six curves, especially the second one, resemble the normal distribution a lot. With the model, we can simulate the succession in forests and calculate the possible biomass in different succession stages, showing the environment of previous ecosystem.

Normal Distribution in Demographics (population pyramid):

A **population pyramid**, also called an **age pyramid** or **age picture** is a graphical illustration that shows the distribution of various age groups in a population.[4] It takes role of an indication of the reproductive capabilities and likelihood of the continuation of a species. There are three types of population pyramids: expansive, constrictive and stationary. The picture below will shed light on them.

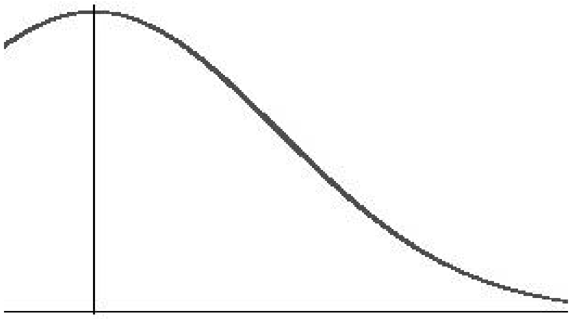


Stage 1	Stage 2	Stage 3	Stage 4
Expansive.	Expansive.	Stationary.	Contractive.
Concave sides.	Straight sides.	Convex sides.	Convex sides.
High birth rate.	Still high birth rate.	Declining birth rate.	Very low birth rate.
High death rate.	Falling death rate.	Low death rate.	Low death rate.
Short life expectancy.	Slightly longer life expectancy.	Long life expectancy.	Longer life expectancy.
Rapid fall in each upward age group due to high DR.	Fall in DR so more people living into middle age.	An increasing proportion of the population is in the 65+ age group.	Higher dependency ratio.
Niger	Haiti	Morocco	Australia

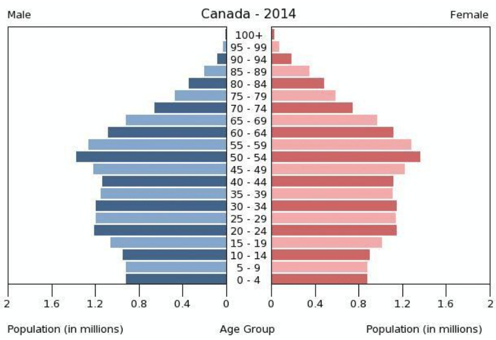
[5]

[4] Weeks, John (2001). *Population An introduction to concepts and issues.*
[5] https://jemimacooper.files.wordpress.com/2012/03/picture-6.png

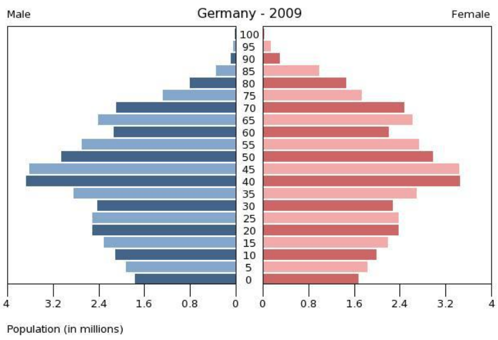
As you can see, the “stationary” type indicates that the peak of age appears a bit below middle; if we draw a line graph instead of the bar graph, we can find the curve approximately a normal distribution, though starts a bit higher.



The stationary type, resembling normal distribution, showcases the stability of population in certain areas. And this is an ideal model for most developed countries.



[6]



[7]

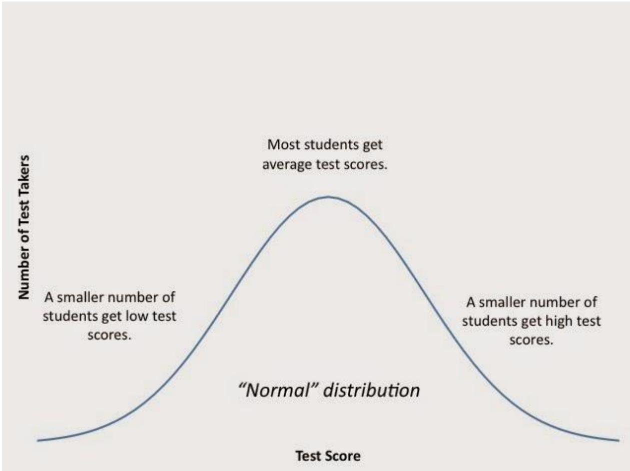
P.S. In fact, normal distribution is not accurate enough to analyze an event. However, due to its symmetry, it can be conveniently calculated and analyzed, on the basis of which we can narrow down the choices of more complicated models and establish the correct one faster. This is why statisticians like normal distributions and try their best to unearth the potential ones in every corner of life.

[6]https://userscontent2.emaze.com/images/37b94807-c54d-48da-b01a-1e61a531b6c3/947bc3d8a1b09a3cf2e9922aec3c33d6.gif

[7] https://wasatchecon.files.wordpress.com/2010/08/germanypyr.jpeg

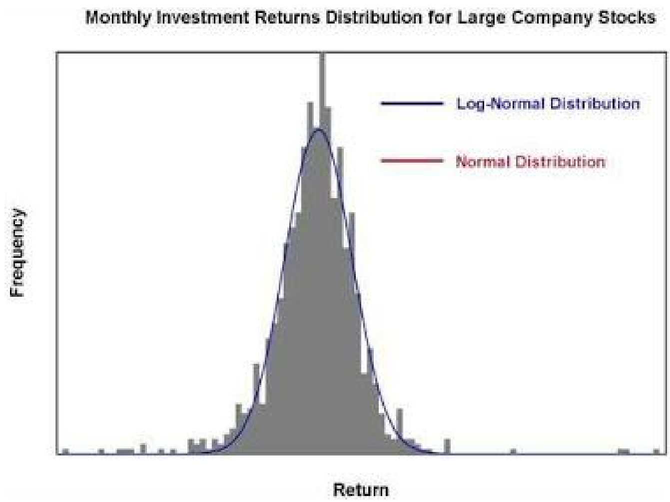
Normal Distribution in Other Cases:

It also works in tests:

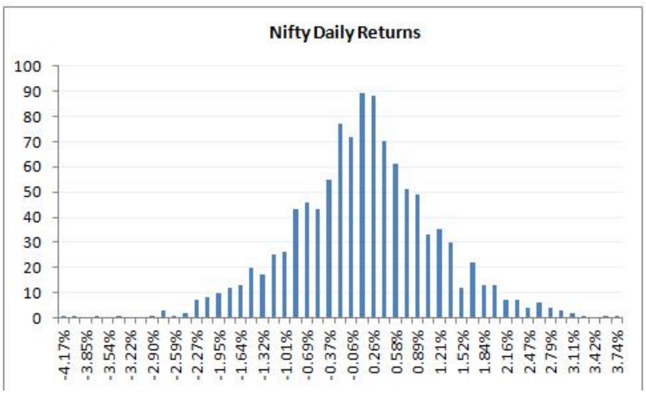


[8]

in investments:



[9]



[10]

---

8 http://4.bp.blogspot.com/-sV049r3-jMA/VWMuqsAG5dI/AAAAAAAADdY/IBjgxxvmRc0/s1600/Slide2.jpg
9 http://4.bp.blogspot.com/_Dl6zNDJDfVg/SEl8lU-bDAI/AAAAAAAAAJg/zU4WGP5HBDQ/s400/PMPT3-Graph+1.JPG
10 https://pbs.twimg.com/media/CLx_DAeUsAAWhxw.png

Central Limit Theorem(CLT)[11]:

Normal distribution never emerges out of thin air; the CLT firmly supports it.

Intuitively, the CLT says that a large sum of i.i.d.(independent and identically distributed) random variables, properly normalized, will always have approximately a **normal distribution**. This shows that the normal distribution is extremely fundamental in probability and statistics.

First, I will show the CLT and **then explain it**.

Suppose $\{X_1, \ldots, X_n\}$ is an i.i.d. sequence of random variables each having finite mean μ and finite variance σ^2. The sample sum is given by: $S_n = X_1 + \cdots + X_n$

And the sample mean is: $M_n \doteq S_n / n$

The central limit theorem is concerned with the distribution of the random variable:

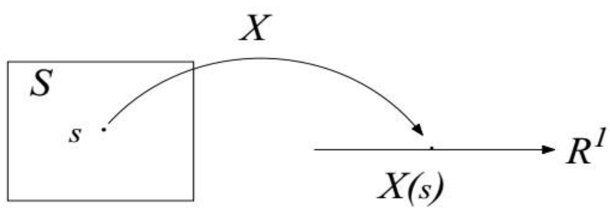
$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{M_n - \mu}{\sigma / \sqrt{n}} = \sqrt{n}\left(\frac{M_n - \mu}{\sigma}\right)$$

where σ = sqr(σ^2). We know E(M_n) = μ, and Var(M_n) = σ^2/n, which implies that E(Z_n) = 0 and Var(Z_n) = 1. The variable Z_n is thus obtained from the sample mean by subtracting its mean and dividing by its standard deviation, and it shares two characteristics with the $N(0,1)$ distribution; namely, it has mean 0 and variance 1. So here comes the CLT:

Theorem 4.4.3 (*The central limit theorem*) Let $X_1, X_2, \ldots$ be i.i.d. with finite mean μ and finite variance σ^2. Let $Z \sim N(0, 1)$. Then as $n \to \infty$, the sequence $\{Z_n\}$ converges in distribution to Z, i.e., $Z_n \xrightarrow{D} Z$.

Mind blown off? Here is explanations from the basic terms to brief proof:

1) random variable



11 Most equations and theorems from Probability and Statistics The Science of Uncertainty 2nd Edition

According to the definition, random variable is a function from sample space S to the set R^1 of all real numbers. This is a basic concept in statistics, every theory stands on its basis. You may take it as a real number.

2) density function

> **Definition 2.4.2** Let $f : R^1 \to R^1$ be a function. Then f is a *density function* if $f(x) \geq 0$ for all $x \in R^1$, and $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

In fact, you can take the density function just as the absolute value of **likelihood function** introduced in the second paper.

3) normal distribution

$N(\mu, \sigma^2)$ distribution has been introduced in the previous paper,

$$f(x) = \frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

where μ is the expectation of the distribution, σ is the standard deviation, and σ^2 is the variance.

So here I put X ~ N(0,1) normal distribution:

$$f_X(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

where the expectation equals 0 and variance equals 1. Now you know why Z_n shares two characteristics with N(0,1) distribution.

4)

> **Definition 4.4.1** Let $X, X_1, X_2, \ldots$ be random variables. Then we say that the sequence $\{X_n\}$ *converges in distribution* to X, if for all $x \in R^1$ such that $P(X = x) = 0$ we have $\lim_{n\to\infty} P(X_n \leq x) = P(X \leq x)$, and we write $X_n \xrightarrow{D} X$.

> **Corollary 4.4.1** If $X_n \to X$ with probability 1, then $X_n \xrightarrow{D} X$

This definition and its corollary sheds light on $Z_n \xrightarrow{D} Z$ in CLT theorem.

5) strong law of large numbers

Theorem 4.3.2 (*Strong law of large numbers*) Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables, each having finite mean μ. Then

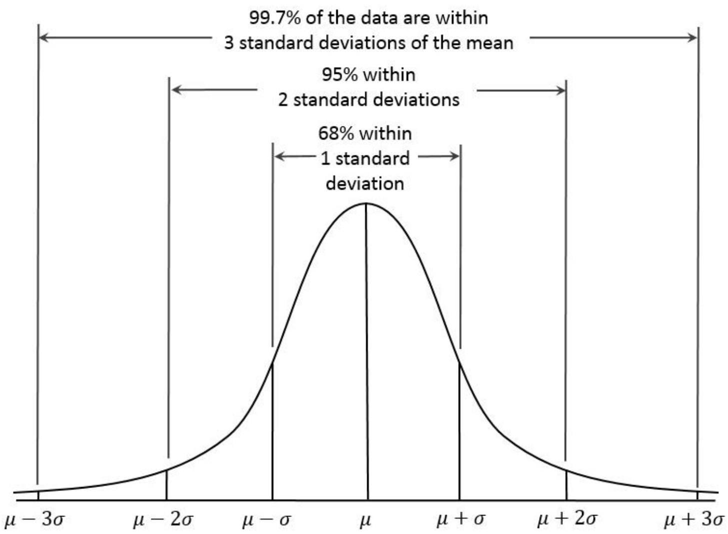
$$P\left(\lim_{n\to\infty} M_n = \mu\right) = 1.$$

That is, the averages converge with probability 1 to the common mean μ or $M_n \overset{a.s.}{\to} \mu$.

This law demonstrates the reason that $E(M_n) = \mu$.

And it’s done now.

6) Verification



Above is the universally acknowledged properties of normal distribution. Let’s see if Z_n in the CLT abides by this rule.

$$\begin{aligned}\Phi(3) - \Phi(-3) &= \lim_{n\to\infty} P\left(-3 < \frac{M_n - \mu}{\sigma/\sqrt{n}} < 3\right) \\ &= \lim_{n\to\infty} P\left(M_n - 3\frac{\sigma}{\sqrt{n}} < \mu < M_n + 3\frac{\sigma}{\sqrt{n}}\right).\end{aligned}$$

$$\Phi(3) - \Phi(-3) = 0.9974.$$

$\Phi(3)$ turns out to be correct. So do $\Phi(2)$ and $\Phi(1)$.

Conclusion:

In this paper, we know the ubiquitous normal distribution and the CLT behind it. Normal distribution, though in a rather complicated form, can analyze the practical models more easily. Besides its functions, normal distribution also maintains elegant curve that fascinates every person who loves symmetry. It is the representative of nature in statistics!

References:

https://mathbitsnotebook.com/Algebra2/Statistics/normalturqa.jpg

Forest Succession, Jeff Martin and Tom Gower, November, 1996

https://upload.wikimedia.org/wikipedia/commons/thumb/4/41/Forest_succession_depicted_over_time.png/800px-Forest_succession_depicted_over_time.png

Weeks, John (2001). *Population An introduction to concepts and issues.*

https://jemimacooper.files.wordpress.com/2012/03/picture-6.png

https://userscontent2.emaze.com/images/37b94807-c54d-48da-b01a-1e61a531b6c3/947bc3d8a1b09a3cf2e9922aec3c33d6.gif

https://wasatchecon.files.wordpress.com/2010/08/germanypyr.jpeg

http://4.bp.blogspot.com/-sV049r3-jMA/VWMuqsAG5dI/AAAAAAAADdY/IBjgxxvmRc0/s1600/Slide2.jpg

http://4.bp.blogspot.com/_Dl6zNDJDfVg/SEl8lU-bDAI/AAAAAAAAAJg/zU4WGP5HBDQ/s400/PMPT3-Graph+1.JPG

https://pbs.twimg.com/media/CLx_DAeUsAAWhxw.png

Probability and Statistics The Science of Uncertainty 2nd Edition