

# Normal Distribution

listenzcc

January 24, 2021

## Abstract

The hand book of Normal distribution. How it is formulated, what it can predict and why we use it.

## Contents

<b>1</b>	<b>Normal Distribution in Classic View</b>	<b>1</b>
1.1	Binominal Distribution . . . . .	1
1.2	Poisson Distribution . . . . .	2
1.3	Normal Distribution . . . . .	3
1.3.1	The Pdf of Normal Distribution . . . . .	3
<b>2</b>	<b>Family Members</b>	<b>4</b>
2.1	Chi-squared Distribution . . . . .	4
2.1.1	Mean and Variance . . . . .	4
2.1.2	The Pdf of Chi-squared distribution . . . . .	4
2.2	Student's T Distribution . . . . .	5
2.2.1	Mean and Variance . . . . .	5
2.2.2	Relationship with Normal Distribution . . . . .	6
2.2.3	The pdf of Student's T Distribution . . . . .	6
<b>3</b>	<b>Examples</b>	<b>7</b>
3.1	Paired and Un-paired T-test . . . . .	8
3.1.1	Paired T-test . . . . .	8
3.1.2	Un-paired T-test . . . . .	8
3.2	Required Sample Numbers for Solid Output . . . . .	8
3.3	Analysis of Variance . . . . .	8

## 1 Normal Distribution in Classic View

### 1.1 Binominal Distribution

Perform experiment for  $n$  times, we assume the trials are independent and follow the same distribution. The output of the experiment is noted as 1 or 0, with no vague. The probability of observing the 1 output is noted as  $p$ . Then, we have

$$P(n, m) = (n, m) \cdot p^m \cdot (1 - p)^{n-m} \quad (1)$$

where  $m$  refers the fact that we observe 1 output for  $m$  times.

It is easy to see that the  $P(n, m)$  produces a distribution since

$$\sum_{i=0}^n P(n, i) = 1, i \in \mathcal{N}$$

Use the computation of **expectation and Variation** of the random variable, we have

$$\begin{aligned}\mathcal{E}(m) &= n \cdot p \\ \mathcal{V}(m) &= n \cdot p \cdot (1 - p)\end{aligned}$$

## 1.2 Poisson Distribution

The poisson distribution is the infinity binominal distribution (see (1)), when  $p$  is **small** and  $n$  is **large**. It is defined as

$$P(k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \lambda = np \quad (2)$$

*Proof.* Use the equation of  $\lambda = np$ , we can rewrite (1) as

$$P(n, k) = (n, k) \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

Since  $p$  is small and  $n$  is large, we have  $k$  is relatively small compared to  $n$ . As a result, in infinity case,

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{(n, k)}{n^k} &= \frac{1}{k!} \\ \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{n-k} &= e^{-\lambda}\end{aligned}$$

Hence proved.  $\square$

In practice, we require  $\lambda < 1$  to produce a valid approximation.

To compute the expectation and variation of the poisson distribution, we use the taylor series of Exp function

$$e^{\lambda t} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}, t_0 = 0$$

it naturally guarantees the property of PDF that

$$\sum_{k=0}^{\infty} P(k) = 1$$

Use the definition of the expectation, we have

$$\begin{aligned}\mathcal{E}(k) &= \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \cdot e^{-\lambda} \\ \mathcal{E}(k) &= \lambda\end{aligned}$$

Use the definition of the variation, we have

$$\begin{aligned}\mathcal{E}(k^2) &= \sum_{k=1}^{\infty} \frac{k \lambda^k}{(k-1)!} \cdot e^{-\lambda} \\ \mathcal{E}(k^2) &= \lambda + \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} \cdot e^{-\lambda} \\ \mathcal{E}(k^2) &= \lambda + \lambda^2\end{aligned}$$

where we used the idea of  $k = 1 + (k-1)$ . Thus, the variation is

$$\mathcal{V}(k) = \lambda$$

### 1.3 Normal Distribution

When  $n$  is large and  $p$  is not so small, the poisson distribution fails on approximate the binominal distribution. The Normal distribution is used as a more general replacement.

Basically, when  $n$ ,  $np$  and  $nq$  are large, the binominal distribution is well approximated by the Normal distribution

$$p(x) = (n, x)p^x q^{n-x} \approx \frac{1}{\sqrt{2\pi npq}} e^{-(x-np)^2/2npq}$$

where  $p + q = 1$ . See the website <sup>1</sup> for detail.

And they are linked based on Sterling's formula,

$$n! = n^n e^{-n} \sqrt{2\pi n} [1 + \mathcal{O}(1/n)] \quad (3)$$

See the website <sup>2</sup> for detail.

Formally, the Normal distribution is expressed as

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4)$$

where expectation  $\mu = \lim_{n \rightarrow \infty} np$  and variation  $\sigma^2 = \lim_{n \rightarrow \infty} npq$ .

#### 1.3.1 The Pdf of Normal Distribution

The Normal distribution is usually expressed as  $p(x) \sim \mathcal{N}(\mu, \sigma^2)$ .

*Proof.* The Pdf of Normal distribution is a pdf.

$$\int_{-\infty}^{\infty} p(x) = 1, p(x) > 0$$

$$p(x) \sim \mathcal{N}(\mu, \sigma^2)$$

Firstly, the  $p(x) > 0$  is obvious. Secondly, using the equation of  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ , the integral is 1.

Hence proved.  $\square$

The expectation and variance are

$$\mathcal{E}(x) = \mu$$

$$\mathcal{V}(x) = \sigma^2$$

*Proof.* The expectation and variance of Normal distribution are  $\mu$  and  $\sigma^2$ .

Use the definition of expectation, and variable change with  $y = x - \mu$ , one gets

$$\mathcal{E}(x) = \int_{-\infty}^{\infty} (y + \mu) p(y + \mu) dy$$

Since integrand function of  $y \cdot p(y + \mu)$  is odd function, and use the PDF property of Normal distribution, one gets  $\mathcal{E}(x) = \mu$ .

Use the value of  $\Gamma(\frac{3}{2}) = \frac{\sqrt{\pi}}{2}$ , and variable change with  $y = \frac{x-\mu}{\sqrt{2\sigma^2}}$ , one gets

$$\mathcal{V}(x) = \int_{-\infty}^{\infty} (x - \mu)^2 p(x) dx = \sigma^2$$

Hence proved.  $\square$

<sup>1</sup><http://scipp.ucsc.edu/~haber/ph116C/NormalApprox.pdf>

<sup>2</sup>[https://www.researchgate.net/publication/237571154\\_A\\_Very\\_Short\\_Proof\\_of\\_Stirling's\\_Formula](https://www.researchgate.net/publication/237571154_A_Very_Short_Proof_of_Stirling's_Formula)

## 2 Family Members

There are several common distributions in the family of Normal distribution.

### 2.1 Chi-squared Distribution

If  $Y_i \sim \mathcal{N}(0, 1)$ , and  $Y_i$ s are independent with each other, then

$$\chi^2 \equiv \sum_{i=1}^r Y_i^2 \quad (5)$$

is distributed as Chi-squared  $\chi^2$  distribution with  $r$  degrees of freedom. The symbolic notion is  $p_r(x) \sim \chi^2(r)$ .

The pdf of Chi-squared distribution is

$$p_r(x) = \frac{x^{r/2-1} e^{-x/2}}{\Gamma(r/2) 2^{r/2}}, 0 < x < \infty \quad (6)$$

#### 2.1.1 Mean and Variance

The mean and variance of the chi-squared distribution is

$$\begin{aligned} \text{Mean} &\triangleq E(x) = r \\ \text{Variance} &\triangleq E(x^2) - E^2(x) = 2r \end{aligned}$$

#### 2.1.2 The Pdf of Chi-squared distribution

**Lemma 2.1.** *To get the pdf of a Chi-squared distribution, we have to prove that*

$$p_n(x) \propto x^{n/2-1} \cdot e^{-x/2}$$

in which,  $x = \sum_{i=1}^n y_i^2$  and  $y_i \sim \mathcal{N}(0, 1)$ . Each  $y_i$  are independent.

*Proof.* The joint probability of  $\{y_1, y_2, \dots, y_n\}$  is

$$p_{joint} = \exp\left(\sum_{i=1}^n -y_i^2/2\right)$$

Thus, the cumulative sum of  $p_n(x)$  can be computed using surface integral

$$\begin{aligned} P_n(r < \sqrt{x}) &\propto \int_S p_{joint} ds \\ P_n(r < \sqrt{x}) &\propto \int_S e^{-r^2/2} ds \end{aligned}$$

in which,  $S$  refers the volume of a sphere with radius of  $x$ .

Transfer the integral into sphere coordinates, we have

$$P_n(r < \sqrt{x}) \propto \int_{r=0}^{\sqrt{x}} e^{-r^2/2} r^{(n-1)} dr$$

Derivate to  $x$ , we have

$$\begin{aligned} \frac{\partial}{\partial x} P_n(r < \sqrt{x}) &\propto e^{-r^2/2} r^{(n-1)} x^{-1/2} \\ \frac{\partial}{\partial x} P_n(r < \sqrt{x}) &\propto x^{n/2-1} \cdot e^{-x/2} \end{aligned}$$

The first step is because of the Newton's integral rule, the second step is based on the replacement of  $r = \sqrt{x}$ .

Hence proved.  $\square$

**Lemma 2.2.** *Next, we have to prove that the integral of  $p_n(x)$  with  $p_n(x) \sim \chi^2(n)$  is*

$$\int_0^\infty p_n(x) dx = \Gamma(n/2) \cdot 2^{n/2}$$

*Proof.* Use the definition of  $\Gamma$  function

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx$$

Use variable replacement of  $z = 2x$ , we have

$$\Gamma(n) = 2^{-n} \int_0^\infty z^{n-1} e^{-z/2} dz$$

Then, use substitution of  $n = n/2$ , we have

$$\Gamma(n/2) \cdot 2^{n/2} = \int_0^\infty z^{n/2-1} e^{-z/2} dz$$

Hence proved.  $\square$

## 2.2 Student's T Distribution

The student's t distribution describes a random variable  $T$  of the form

$$T = \frac{\bar{x} - m}{s/\sqrt{N}} \quad (7)$$

where  $\bar{x}$  is the sample mean value of all  $N$  samples,  $m$  is the population mean value and  $s$  is the population standard deviation.

Or, in a more formal one

$$T = \frac{X}{\sqrt{Y/r}} \quad (8)$$

where  $X \sim \mathcal{N}(0, 1)$  and  $Y \sim \chi_r^2$ .

The pdf of Student's t-distribution is

$$t_r(x) = \frac{\Gamma(\frac{r+1}{2})}{\Gamma(\frac{r}{2})\sqrt{r\pi}} \left(1 + \frac{x^2}{r}\right)^{-\frac{r+1}{2}}, -\infty < x < \infty \quad (9)$$

### 2.2.1 Mean and Variance

The mean and variance of the Student's t-distribution is

$$Mean \triangleq E(x) = 0$$

$$Variance \triangleq E(x^2) - E^2(x) = \frac{r}{r-2}$$

### 2.2.2 Relationship with Normal Distribution

It is easy to see that  $\lim_{r \rightarrow \infty} t_r(x) \sim \mathcal{N}(0, 1)$ . It demonstrates that when  $r$  is large enough, the Student's t-distribution is equalize to Normal Distribution. It should to be noted that the calculation of  $\frac{\Gamma(\frac{r+1}{2})}{\Gamma(\frac{r}{2})\sqrt{r\pi}}$  is somehow difficult, however, and fortunately, the value is constant with  $x$  when  $r \rightarrow \infty$ . Since the other factor can be formulated as the form of  $e^{\frac{x^2}{2}}$ , the constant can be calculated using the property of Normal distribution. Thus, the equation is also an useful approximation to the constant.

### 2.2.3 The pdf of Student's T Distribution

Here, we provide a simple computation of the pdf of the Student's t-distribution.

$$T = \frac{X}{\sqrt{Y/r}}$$

in which  $X \sim \mathcal{N}(0, 1)$  and  $Y \sim \chi^2(r)$ , and they are independent. Thus, we have

$$\begin{aligned} p(x) &\propto e^{-x^2/2} \\ p(y) &\propto y^{r/2-1} \cdot e^{-y/2} \end{aligned}$$

The random variable  $t$  follows the equation  $t = \frac{x}{\sqrt{y/r}}$ .

**Lemma 2.3.** *Since then we want to prove that*

$$p(t) \propto \left(1 + \frac{t^2}{r}\right)^{-\frac{r+1}{2}} \quad (10)$$

*Proof.* The joint probability of  $p(x, y)$  matches

$$p(x, y) \propto e^{-x^2/2} \cdot y^{r/2-1} \cdot e^{-y/2}$$

And the divergence of  $p(x, y)$  is  $p(x, y)dx dy$ . We can use the variable replacement of

$$\begin{aligned} y &= \frac{x^2}{t^2} \cdot r \\ \frac{dy}{dt} &\propto \frac{x^2}{t^3} \end{aligned}$$

Thus we have the joint probability of  $p(x, t)$  matches

$$p(x, t) \propto e^{-x^2/2} \cdot \left(\frac{x^2}{t^2}\right)^{r/2-1} \cdot e^{-\frac{x^2}{2t^2}r} \cdot \frac{x^2}{t^3}$$

The probability of  $p(t)$  can be expressed as

$$p(t) \propto \int_x p(x, t) dx$$

Analysis the expression, we have

$$\begin{aligned} p(t) &\propto t^{-r-1} \int_x x^r \cdot e^{-\frac{1}{2}(1+\frac{r}{t^2})x^2} dx \\ p(t) &\propto t^{-r-1} \cdot \left(1 + \frac{r}{t^2}\right)^{-\frac{r-1}{2}} \int_z z^r \cdot e^{-z^2} dz \\ p(t) &\propto (t^2 + r)^{-\frac{r+1}{2}} \\ p(t) &\propto \left(1 + \frac{t^2}{r}\right)^{-\frac{r+1}{2}} \end{aligned}$$

The process uses the integral of  $\Gamma$  function is constant, and  $r$  is constant.  $\square$

After that, combining with the following, we should finally have the pdf function.

**Lemma 2.4.** *The values of  $t_r(x)$  is positive and the integral is 1.*

$$\int_{-\infty}^{\infty} t_r(x) dx = 1$$

*Proof.* Consider the variable part of Student's t-distribution

$$f(x) = (1 + \frac{x^2}{r})^{-\frac{r+1}{2}}, -\infty < x < \infty$$

use a replacement as following

$$x^2 = \frac{y}{1-y}$$

it is easy to see that  $\lim_{y \rightarrow 0} x = 0$  and  $\lim_{y \rightarrow 1} x = \infty$ . Additionally, the  $x^2$  is even function. Thus we can write the integral of  $f(x)$

$$\int_{-\infty}^{\infty} f(x) dx = 2\sqrt{r} \int_0^1 (\frac{1}{1-y})^{-\frac{r+1}{2}} d(\frac{y}{1-y})^{\frac{1}{2}}$$

it is not hard to find out that the integral may end up with

$$\sqrt{r} \int_0^1 (1-y)^{\frac{r}{2}-1} y^{\frac{1}{2}-1} dy = \sqrt{r} B(\frac{r}{2}, \frac{1}{2})$$

Finally the Normalization factor has to be

$$\frac{\Gamma(\frac{r+1}{2})}{\sqrt{r}\Gamma(\frac{r}{2})\Gamma(\frac{1}{2})}$$

which makes the integral of  $t_r(x)$  is 1.  $\square$

### 3 Examples

One important usage of normal distribution is to valuate the output of the experiment. Since a large number of observations can be fitted into a normal distribution.

In an experiment, the results can be variance on different experiment trails. In the standard analysis pipeline, the *random variables* which can be fitted into the normal distribution, are formulated based on the obtained outputs.

The statistical analysis is commonly used to obtain the underlying ground truth value of the outputs. Particularly, we are interested in the *expectation* and *variance* of the random variables. The expectation is used to estimate the ideal value, often refers noise-free situation, of one trial. The variance is used to estimate the reasonable range of the output of one trial. Additionally, the distribution of the selected random variable is also used to *valuate* or *compare* the output of the experiment.

### 3.1 Paired and Un-paired T-test

The t-test method is the useful method to compare the effects of *two* factors. A common situation is to compare the difference of two methods doing the same job. We assume the outputs can be evaluated as the random variable fitting in *student's t distribution*. Then the sum of the variables is also fitting in *student's t distribution*, but the only question is the parameters being left as unknown.

To perform valid comparison of the variables of two methods, the *generation* of the random variables and the *estimation* of the parameters should be investigated. Back to the question of binary comparison. If for each experiment trial, we can parallel apply the two methods to it, the *paired t-test* method will be used. If we can only apply the method on a trial once, the *un-paired t-test* method will be used.

#### 3.1.1 Paired T-test

#### 3.1.2 Un-paired T-test

### 3.2 Required Sample Numbers for Solid Output

### 3.3 Analysis of Variance