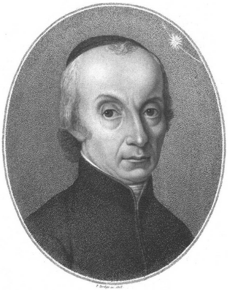
The Normal Distribution

Ceres

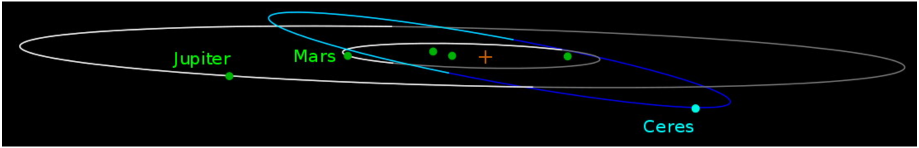


Portrait of Giuseppe Piazzi (1746-1826). Bordiga, F. 1808. Smithsonian institute Library. File:Giuseppe Piazzi.jpg. (2013, November 2). Wikimedia Commons, the free media repository.

On January 1st, 1801, the comet Ceres (later: planet, asteroid, dwarf planet) was discovered by the Italian priest Giuseppe Piazzi. He observed it 24 times until February 11th. When his observations were finally published in September 1801, Ceres could not be observed any more due to the sun's glare. So Piazzi's discovery could not be confirmed.

To find Ceres once it would become visible again at the end of the year, its position would need to be calculated.

But Piazzi had observed only around 1% of Ceres's orbit.



Orbit of Ceres File:Ceres Orbit.svg. (2016, January 12). Wikimedia Commons, the free media repository.

Carl Friedrich Gauß

The young mathematician Carl Friedrich Gauß heard about the Ceres problem and set to work. Within three months, he derived a prediction for the expected position of Ceres and published it in early December 1801. On December 31st, Ceres was found very close to the predicted position. This achievement of the 24-year-old Gauß established his reputation.

C. F. Gauß at 50 (1777-1855). Bendixson, S. D. 1828. Lithograph. Smithsonian institute Library. File:Bendixen - Carl Friedrich Gau, 1828.jpg. (2020, March 5). Wikimedia Commons, the free media repository.

Gauß developed several new mathematical tools (such as the least-squares method).

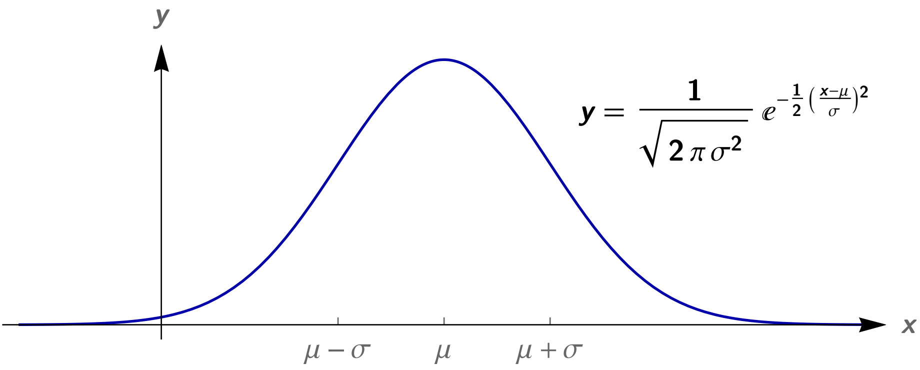
But the most important idea was that a prediction would be impossible without understanding the mathematical function which described the errors and uncertainties in the Piazzi's observations. Starting from the premise that such a function exists in the first place, he derived the distribution which became known as the ***Gaußian*** or ***normal distribution***.

Normal (Gauß) Distribution

8.1. Definition. Let $\mu \in \mathbb{R}$, $\sigma > 0$. A continuous random variable (X, f_X) with density

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

is said to follow a normal distribution with parameters μ and σ.



Normal Distribution

It is easily verified that $\int_{\mathbb{R}} f_X(x)\,dx = 1$ by using polar coordinates.

We write

$$X \sim N(\mu, \sigma)$$

whenever a random variable X follows a normal distribution with mean μ and variance σ^2.

8.2. Theorem. Let (X, f_X) be a normally distributed random variable with parameters μ and σ.

(i) The moment-generating function of X is given by

$$m_X\colon \mathbb{R} \to \mathbb{R}, \qquad\qquad m_X(t) = e^{\mu t + \sigma^2 t^2/2}.$$

(ii) $\mathrm{E}[X] = \mu$.

(iii) $\mathrm{Var}[X] = \sigma^2$.

Normal Distribution

Proof.

We will verify the moment-generating function only.

$$\begin{aligned} m_X(t) = \mathrm{E}[e^{tX}] &= \int_{-\infty}^{\infty} \frac{e^{tx}}{\sqrt{2\pi}\sigma} e^{-((x-\mu)/\sigma)^2/2}\, dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{tx-((x-\mu)/\sigma)^2/2}\, dx \end{aligned}$$

We complete the square in the exponent to gain

$$tx - \frac{(x-\mu)^2}{2\sigma^2} = -\frac{(x-(\mu+\sigma^2 t))^2}{2\sigma^2} + \mu t + \sigma^2 t^2/2$$

Normal Distribution

Proof (continued).

Substituting into the integral,

$$m_X(t) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{(x-(\mu+\sigma^2 t))^2}{2\sigma^2}+\mu t+\sigma^2 t^2/2}\,dx$$

$$= e^{\mu t+\sigma^2 t^2/2}\underbrace{\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{(x-(\mu+\sigma^2 t))^2}{2\sigma^2}}\,dx}_{=1}$$

□

Standard Normal Distribution

8.3. Definition. A normally distributed random variable with parameters $\mu = 0$ and $\sigma = 1$ is called a ***standard normal*** random variable and denoted by Z.

The standard normal distribution is particularly important because any normally distributed random variable can be transformed into a standard-normally distributed one.

8.4. Theorem. Let X be a normally distributed random variable with mean μ and standard deviation σ. Then

$$Z := \frac{X - \mu}{\sigma}$$

has standard normal distribution.

Transformation of Random Variables

It is easily seen that $Z = \frac{X-\mu}{\sigma}$ has mean $\mathrm{E}[Z] = 0$ and variance $\operatorname{Var} Z = 1$, but it is not clear that Z is normally distributed. To see this, we need to find the density of Z.

Hence it is worth studying the density of transformed variables in general.

8.5. Theorem. Let X be a continuous random variable with density f_X. Let $Y = \varphi \circ X$, where $\varphi\colon \mathbb{R} \to \mathbb{R}$ is strictly monotonic and differentiable. The density for Y is then given by

$$f_Y(y) = f_X(\varphi^{-1}(y)) \cdot \left| \frac{d\varphi^{-1}(y)}{dy} \right| \qquad \text{for } y \in \operatorname{ran} \varphi$$

and

$$f_Y(y) = 0 \qquad \text{for } y \notin \operatorname{ran} \varphi.$$

Transformation of Random Variables

Proof.

We assume without loss of generality that φ is strictly decreasing. (The case where φ is strictly increasing is analogous.)

The cumulative distribution function for Y is given by

$$F_Y(y) = P[Y \le y] = P[\varphi(X) \le y].$$

Since φ is strictly decreasing, φ^{-1} exists and is also decreasing. Suppose that $y \in \operatorname{ran}\varphi$. Then

$$\begin{aligned} F_Y(y) &= P[\varphi(X) \le y] \\ &= P[\varphi^{-1}(\varphi(X)) \ge \varphi^{-1}(y)] \\ &= P[X \ge \varphi^{-1}(y)] \\ &= 1 - P[X \le \varphi^{-1}(y)] \\ &= 1 - F_X(\varphi^{-1}(y)). \end{aligned}$$

Transformation of Random Variables

Proof (continued).

Since φ is strictly continuous, the range of φ is an interval in $\mathbb{R}$. If $y \notin \operatorname{ran}\varphi$, then either $y > z$ for all $z \in \operatorname{ran}\varphi$ or $y < z$ for all $z \in \operatorname{ran}\varphi$.

We then see that

$$F_Y(y) = P[Y \le y] = P[\varphi(X) \le y] = \begin{cases} 0 & \text{if } y < z \text{ for all } z \in \operatorname{ran}\varphi \\ 1 & \text{if } y > z \text{ for all } z \in \operatorname{ran}\varphi \end{cases}$$

To obtain the density f_Y, we differentiate F_Y. For $y \in \operatorname{ran}\varphi$ we have

$$\begin{aligned} f_Y(y) = F_Y'(y) &= -f_X(\varphi^{-1}(y))\frac{d\varphi^{-1}(y)}{dy} \\ &= f_X(\varphi^{-1}(y)) \cdot \left|\frac{d\varphi^{-1}(y)}{dy}\right|. \end{aligned}$$

If $y \notin \operatorname{ran}\varphi$, F_Y is constant and hence $f_Y = F_Y' = 0$. $\square$

Standard Normal Distribution

We can now prove Theorem 8.4. We have $Z = \varphi \circ X$, where $\varphi(x) = \frac{x-\mu}{\sigma}$ is strictly increasing and differentiable with $\operatorname{ran} \varphi = \mathbb{R}$. Note that

$$\varphi^{-1}(z) = \sigma z + \mu, \qquad \frac{d\varphi^{-1}(z)}{dz} = \sigma > 0.$$

Using

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

we have

$$f_Z(z) = f_X(\varphi^{-1}(z)) \cdot \left| \frac{d\varphi^{-1}(z)}{dz} \right| = \frac{1}{\sqrt{2\pi}\sigma} e^{-(z)^2/2} \cdot \sigma = \frac{1}{\sqrt{2\pi}} e^{-z^2/2},$$

which is the density of the standard normal distribution. Hence the variable $Z = \frac{X-\mu}{\sigma}$ is standard normal.

Transforming Variables

We can verify Theorem 8.4 with Mathematica:

```
TransformedDistribution[(X - μ)/σ, X ≈ NormalDistribution[μ, σ]]
```

```
NormalDistribution[0, 1]
```

Question. If X is standard normal, what is the density of X^2?

```
PDF[TransformedDistribution[X^2, X ≈ NormalDistribution[0, 1]], x]
```

$$\begin{cases} \dfrac{e^{-x/2}}{\sqrt{2\pi}\,\sqrt{x}} & x > 0 \\ 0 & \text{True} \end{cases}$$

Note that the function $f(x) = x^2$ is not monotonic, so Theorem 8.5 can not be applied and a formal calculation needs to be done by hand!

Standard Normal Distribution

The cumulative distribution function of the standard normal distribution is often denoted by Φ,

$$\Phi(z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt.$$

The values of Φ are given in Table V of Appendix A. In Mathematica, the cumulative distribution function is expressed through the error function, defined as

$$\mathbf{erf(z)} := \frac{2}{\sqrt{\pi}} \int_{0}^{z} e^{-t^2}\, dt, \qquad \mathbf{erfc(z)} := 1 - \mathbf{erf(z)}.$$

Hence,

```
CDF[NormalDistribution[0, 1], x]
```

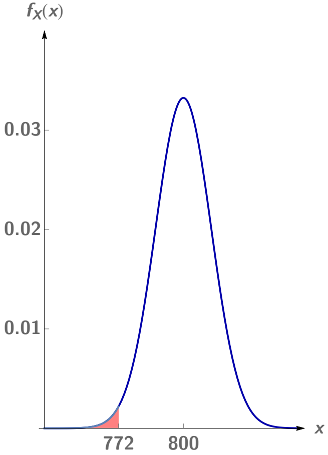
$$\frac{1}{2}\,\mathrm{Erfc}\left[-\frac{x}{\sqrt{2}}\right]$$

Standard Normal Distribution

8.6. Example. The breaking strength of a synthetic fabric is denoted X, and it is normally distributed with mean $\mu = 800\,\text{N}$ and standard deviation $\sigma = 12\,\text{N}$.

A purchaser of the fabric requires the fabric to have a strength of at least $772\,\text{N}$. A fabric sample is randomly selected and tested. To find $P[X \geq 772]$, we calculate

$$\begin{aligned} P[X < 772] &= P\left[\frac{X-\mu}{\sigma} < \frac{772-800}{12}\right] \\ &= P[Z < -2.33] \\ &= \Phi(-2.33) = 0.01. \end{aligned}$$



Hence the sample is 99% likely to pass inspection.

Standard Normal Distribution

8.7. Example. Let X denote the amount of radiation that can be absorbed by an individual before death ensues. Assume that X is normal with a mean dosage of 500 roentgens and a standard deviation of 150 roentgens. Above what dosage level will only 5% of those exposed survive?

Here we want to find x_0 such that $P[X \geq x_0] = 0.05$. Standardizing,

$$P[X \geq x_0] = P\left[\frac{X - 500}{150} \geq \frac{x_0 - 500}{150}\right] = P\left[Z \geq \frac{x_0 - 500}{150}\right] \overset{!}{=} 0.05$$

From Table V, $P[Z \geq 1.64] = 0.0505$ and $P[Z \geq 1.65] = 0.0495$.

Interpolating, we take $P[Z \geq 1.645] \approx 0.0500$, so we have

$$\frac{x_0 - 500\,\text{roentgen}}{150\,\text{roentgen}} = 1.645 \qquad \Leftrightarrow \qquad x_0 = 746.75\,\text{roentgen}\,.$$

Estimates on Variability

In general, the following estimates are often useful:

8.8. Theorem. Let X be normally distributed with parameters μ and σ. Then

$$\begin{aligned} P[-\sigma < X - \mu < \sigma] &= 0.68 \\ P[-2\sigma < X - \mu < 2\sigma] &= 0.95 \\ P[-3\sigma < X - \mu < 3\sigma] &= 0.997 \end{aligned}$$

Hence 68% of the values of a normal random variable lie within one standard deviation of the mean, 95% lie within two standard deviations, and 99.7% lie within three standard deviations. This rule of thumb will be especially important in statistics, where the number of "extraordinary" events needs to be judged.

Estimates on Variability

8.9. Example. Table of mean weights and heights of 12-month-old babies from a Chinese infant care book.

男童的标准

项　目	–2SD	中位数	+2SD
体重(kg)	8.1	10.2	12.4
身高(cm)	70.7	76.1	81.5

女童的标准

项　目	–2SD	中位数	+2SD
体重(kg)	7.4	9.5	11.6
身高(cm)	68.6	74.3	80.0

Jiping Liu, ***0-12 Months infant health and parenting guide***, Sichuan Children's Publishing House, 2008, ISBN 978-7-5365-4392-8
[Chinese; original title 0-12个月婴儿健康养育指南 by 刘纪平.]

The Chebyshev Inequality

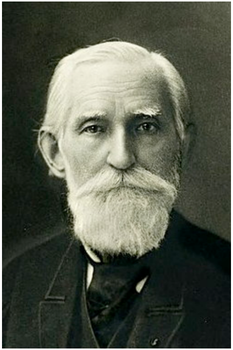
Let $c > 0$ be any real number. Then

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx \geq \int_{|x|\geq c} x^2 f_X(x)\,dx$$
$$\geq c^2 \int_{|x|\geq c} f_X(x)\,dx$$
$$= c^2 \cdot P[|X| \geq c]$$

More generally, for $k \in \mathbb{N} \setminus \{0\}$,

$$P[|X| \geq c] \leq \frac{\mathrm{E}[|X|^k]}{c^k}. \tag{8.1}$$

This is one version of ***Chebyshev's inequality***. The inequality also holds for discrete random variables, with an analogous proof.



Pafnuty Lvovich Chebyshev (1821-1894). File:Pafnuty Lvovich Chebyshev.jpg. (2017, December 2). Wikimedia Commons, the free media repository.

Variability Estimate from Chebyshev's Inequality

If we replace X with $X - \mu$ in (8.1) and set $k = 2$ and $c = m \cdot \sigma$, $m > 0$, we obtain the estimate

$$P\big[|X - \mu| \geq m\sigma\big] \leq \frac{1}{m^2}. \tag{8.2}$$

or, equivalently,

$$P[-m\sigma < X - \mu < m\sigma] \geq 1 - \frac{1}{m^2} \tag{8.3}$$

Comparing (8.2) with Theorem 8.8, we see that the estimates in the theorem are tighter.

This is not surprising, as Chebyshev's rule is valid for any random variable with finite second moment, while the previous theorem uses the specific properties of the normal distribution.

(Im-)Practical Application of Chebyshev's Inequality

8.10. Example. From an analysis of company records, a materials control manager estimates that the mean and standard deviation of the "lead time" required in ordering a small valve are 8 days and 1.5 days, respectively. She does not know the distribution of the lead time, but she is willing to assume the estimates of the mean and standard deviation to be absolutely correct.

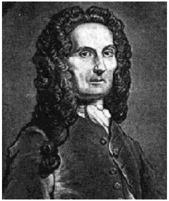
The manager would like to determine a time interval such that the probability is at least 8/9 that the order will be received during that time. That is,

$$1 - \frac{1}{k^2} = \frac{8}{9},$$

so that $k = 3$ and $\mu \pm k\sigma = (8 \pm 4.5)$ days.

This interval may well be too large to be of any value to the manager, in which case she may elect to learn more about the distribution of lead times.

Approximating the Binomial Distribution



Abraham De Moivre (1667-1754). Portrait. Faber. 1736. File:Abraham de moivre.jpg. (2015, October 30). Wikimedia Commons, the free media repository.

Pierre-Simon de Laplace (1749-1827). Engraving. File:Pierre-Simon-Laplace (1749-1827).jpg. (2017, June 6). Wikimedia Commons, the free media repository.

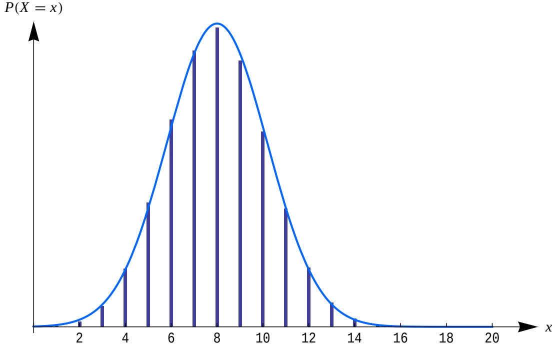
Long before Gauß discovered the normal distribution in 1801, it had been published 60 years earlier, in 1738. De Moivre had wanted to approximate the shape of the binomial distribution, considering the behavior of 3600 coin tosses. In 1810, Laplace proved the general result for $0 < p < 1$.

8.11. Theorem of De Moivre-Laplace. Denote by S_n the number of successes in a sequence of n i.i.d. Bernoulli trials with probability of success $0 < p < 1$. Then

$$\lim_{n\to\infty} P\left[a < \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx.$$

Approximating the Binomial Distribution

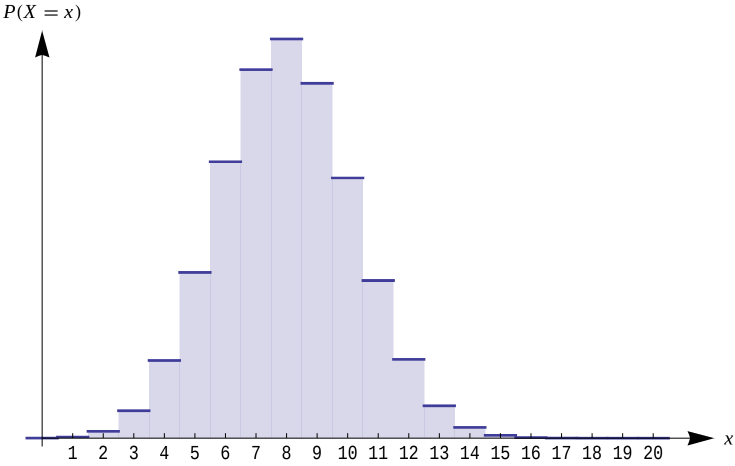
Intuitively, for large n, the binomial distribution with parameters n and p behaves as a normal distribution with mean $\mu = np$ and variance $\sigma^2 = npq$. This is illustrated below for $n = 20$ and $p = 0.4$:



The height of the vertical bars represents the values of $P[X = x]$ according to the binomial distribution, while the density curve of the corresponding normal distribution has been superimposed.

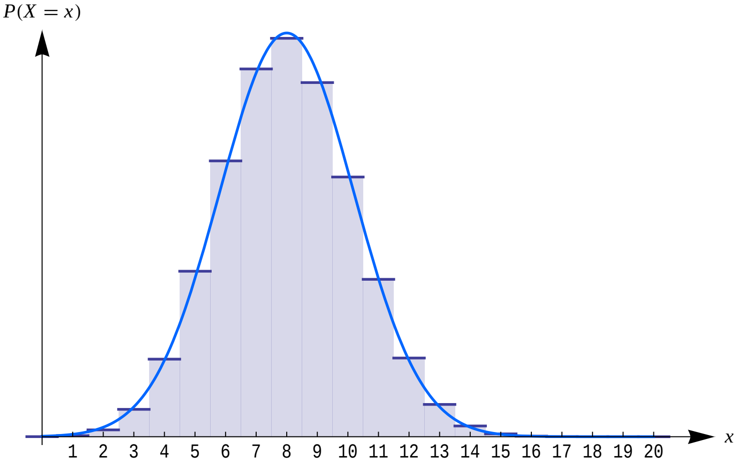
Approximating the Binomial Distribution

We would like to use the normal distribution to approximate the cumulative distribution function of the binomial distribution.



Approximating the Binomial Distribution

It is clear that for each $y = 0, \dots, 20$ the sum over all $x \le y$ corresponds to the area of the bars to the left of y. Superimposing the normal distribution, we see that we can approximate this sum by integrating to $y + 1/2$:



Approximation and the Half-Unit Correction

Hence, for $y = 0, \dots, n$,

$$P[X \le y] = \sum_{x=0}^{y} \binom{n}{x} p^x (1-p)^{n-x} \approx \Phi\left(\frac{y + 1/2 - np}{\sqrt{np(1-p)}}\right).$$

This additional term $1/2$ is known as the ***half-unit correction*** for the normal approximation to the cumulative binomial distribution function. It is necessary because in practice we do not have the limit $n \to \infty$ but rather a finite value of n, which may not even be especially large.

This approximation is good if p is close to $1/2$ and $n > 10$. Otherwise, we require that

$$np > 5 \qquad \text{if } p \le 1/2 \qquad \text{or} \qquad n(1-p) > 5 \qquad \text{if } p > 1/2.$$

Approximating the Binomial Distribution

8.12. Example. In sampling from a production process that produces items of which 20% are defective, a random sample of 100 items is selected each hour of each production shift. The number of defectives in a sample is denoted by X.

To find, say, $P[X \leq 15]$ we might use the normal approximation as follows:

$$\begin{aligned} P[X \leq 15] &\approx P\left[Z \leq \frac{15 - 100 \cdot 0.2}{\sqrt{100 \cdot 0.2 \cdot 0.8}}\right] = P[Z \leq -1.25] \\ &= \Phi(-1.25) = 0.1056 \end{aligned}$$

The half-unit correction would instead give

$$P[X \leq 15] \approx P\left[Z \leq \frac{15.5 - 20}{4}\right] = 0.130$$

The correct result is $P[X \leq 15] = \sum_{k=0}^{15} \binom{100}{k} 0.2^k 0.8^{100-k} = 0.1285.$

Lyapunov's Central Limit Theorem

Aleksandr Mikhailovich Lyapunov (1857-1918). File:Alexander Ljapunow jung.jpg. (2017, January 29). Wikimedia Commons, the free media repository.

Today there exist various "Central Limit Theorems" that generalize the Theorem of De Moivre - Laplace. The following is due to Lyapnuov and was established a century after Laplace's work.

8.13. Central Limit Theorem. Let (X_i) be a sequence of independent, but not necessarily identical random variables whose third moments exist and satisfy a certain technical condition.

Let

$$Y_n = X_1 + \cdots + X_n.$$

Then for any $z \in \mathbb{R}$,

$$P\left[\frac{Y_n - \mathrm{E}[Y_n]}{\sqrt{\mathrm{Var}[Y_n]}} \le z\right] \xrightarrow{n\to\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2}\, dx.$$

Experimental Error

Lyapunov's Central Limit theorem is at the core of the belief by experimentalists that "random error" may be described by the normal distribution. The idea is that any "random disturbance" of a measurement is the sum of many inscrutable and random effects which individually cannot be tracked. However, their sum will be well-described by the normal distribution.

The French physicist Gabriel Lippman wrote to Henri Poincare:

> *Tout le monde y croit cependent, car les experimenteurs s'imaginent que c'est un theorem de mathematiques, et les mathematiciens que c'est un fait experimental.*

"Everybody believes in the exponential law of errors: the experimenters, because they think it can be proved by mathematics; and the mathematicians, because they believe it has been established by observation."