History of the Normal Distribution

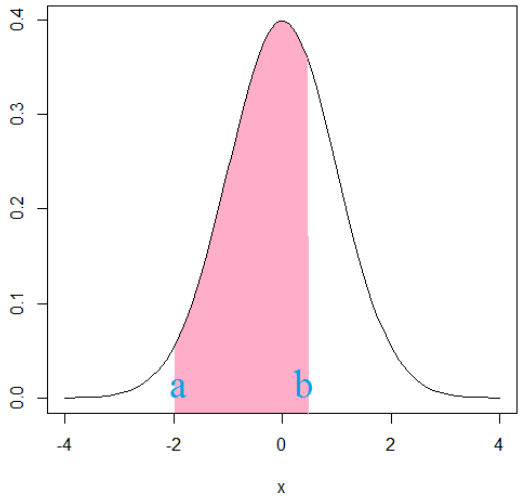
Jenny Kenkel

November 8th, 2016

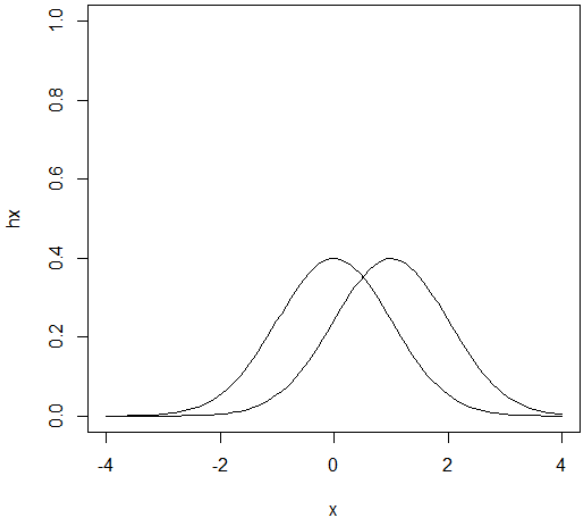
Probabiltiy Distributions

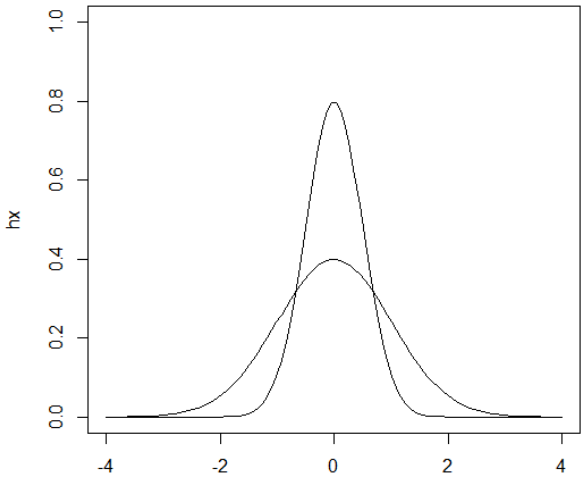
A **probability distribution** is a function $f(x)$ so that

$$P(a < X < b) = \int_a^b f(x)dx$$



The Normal Distribution





The normal distribution is a *family* of distributions, given by

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

The **Standard Normal** has $\mu = 0$ and $\sigma = 1$, i.e.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}}$$

Changing μ changes the location of the curve, and changing σ changes the spread of the curve

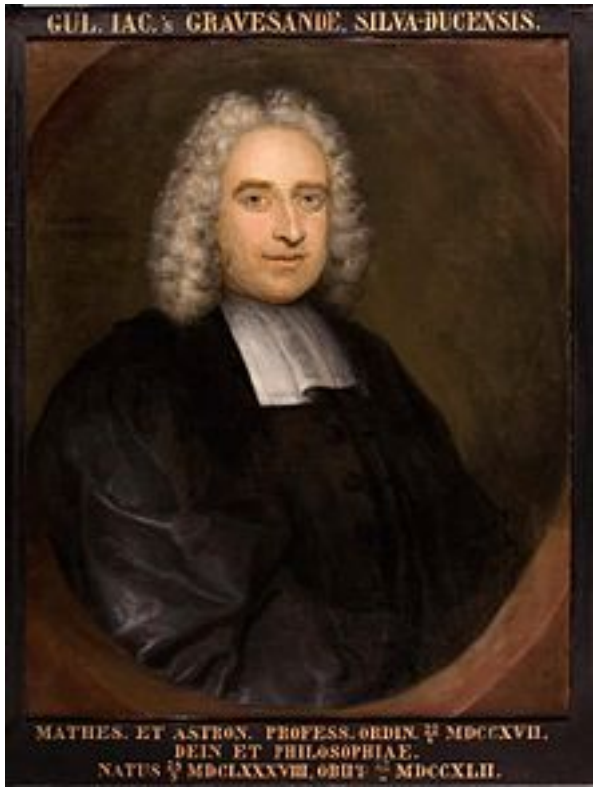
Arbuthnot: "An argument for Divine Providence, taken from the constant Regularity observ'd in the Births of both Sexes"

- Argued the probability that more males than females were born every year from 1629 to 1710, 82 years, was $\left(\frac{1}{2}\right)^{82}$

Arbuthnot: "An argument for Divine Providence, taken from the constant Regularity observ'd in the Births of both Sexes"

- Argued the probability that more males than females were born every year from 1629 to 1710, 82 years, was $\left(\frac{1}{2}\right)^{82}$
- But the fact that not *many* more males than females were born every year was proof polygamy is contrary to the laws of nature

‘sGravesande



- Felt evidence for divine providence was *even stronger* than Arbuthnot had argued

Bernoulli Trials

- A *Bernoulli Trial* is an experiment with only two possible outcomes. A *Binomial Experiment* is a series of repeated Bernoulli trials.

Bernoulli Trials

- A *Bernoulli Trial* is an experiment with only two possible outcomes. A *Binomial Experiment* is a series of repeated Bernoulli trials.
- The probability, p, of *success* stays constant as more trials are performed

Bernoulli Trials

- A *Bernoulli Trial* is an experiment with only two possible outcomes. A *Binomial Experiment* is a series of repeated Bernoulli trials.
- The probability, p, of *success* stays constant as more trials are performed
- The probability of k successes in n trials is

$$\binom{n}{k} p^k (1-p)^{n-k}$$

‘sGravesande was very patient

- ‘sGravesande let $n = 11,429, p = 0.5$

‘sGravesande was very patient

- ‘sGravesande let $n = 11,429, p = 0.5$
- The lowest ratio, which occurred in 1703, was .5027, which corresponded to about 5745 out of 11429
- The highest ratio, which occurred in 1661, was .5362, which corresponded to about 6128 out of 11429

‘sGravesande was very patient

- ‘sGravesande let $n = 11,429, p = 0.5$
- The lowest ratio, which occurred in 1703, was .5027, which corresponded to about 5745 out of 11429
- The highest ratio, which occurred in 1661, was .5362, which corresponded to about 6128 out of 11429

‘sGravesande was very patient

- ‘sGravesande let $n = 11,429, p = 0.5$
- The lowest ratio, which occurred in 1703, was .5027, which corresponded to about 5745 out of 11429
- The highest ratio, which occurred in 1661, was .5362, which corresponded to about 6128 out of 11429
- ‘sGravesande calculated

$$P[5745 \leq x \leq 6128] = \sum_{5745}^{6128} \binom{11429}{x} \left(\frac{1}{2}\right)^{11429} \approx 0.292$$

‘sGravesande was very patient

- ‘sGravesande let $n = 11,429, p = 0.5$
- The lowest ratio, which occurred in 1703, was .5027, which corresponded to about 5745 out of 11429
- The highest ratio, which occurred in 1661, was .5362, which corresponded to about 6128 out of 11429
- ‘sGravesande calculated

$$P[5745 \leq x \leq 6128] = \sum_{5745}^{6128} \binom{11429}{x} \left(\frac{1}{2}\right)^{11429} \approx 0.292$$

- Thus the probability of this difference in birth rates recurring 82 years in a row is $.292^{82}$

DeMoivre and Stirling's Approximation

In 1733, in a second supplement that only appeared with certain editions of his book[3], De Moivre noted that:

- $n! \sim Bn^{n+\frac{1}{2}}e^{-n}$

DeMoivre and Stirling's Approximation

In 1733, in a second supplement that only appeared with certain editions of his book[3], De Moivre noted that:

- $n! \sim Bn^{n+\frac{1}{2}}e^{-n}$
- James Stirling determined that $B = \sqrt{2\pi}$

DeMoivre and Stirling's Approximation

In 1733, in a second supplement that only appeared with certain editions of his book[3], De Moivre noted that:

- $n! \sim Bn^{n+\frac{1}{2}}e^{-n}$
- James Stirling determined that $B = \sqrt{2\pi}$
- The probability of k successes in n trials is

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}}$$

DeMoivre and Stirling's Approximation

In 1733, in a second supplement that only appeared with certain editions of his book[3], De Moivre noted that:

- $n! \sim Bn^{n+\frac{1}{2}}e^{-n}$
- James Stirling determined that $B = \sqrt{2\pi}$
- The probability of k successes in n trials is

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}}$$

- Let $\sigma^2 = np(1-p)$ and $\mu = np$, and you have

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(k-\mu)^2}{\sigma^2}}$$

From Binomials to Errors

In 1733, DeMoivre first used the Normal distribution as an approximation for probabilities of binomial experiments where n is very large.
About 70 years later, it would be used as the probability distribution of **random errors**.

The beginning of the average

Astronomy called for accurate measurements

- In the 1600s, Tycho Brahe advocated incorporating repeated measurements but did not specify how to *use* repeated measurements

The beginning of the average

Astronomy called for accurate measurements

- In the 1600s, Tycho Brahe advocated incorporating repeated measurements but did not specify how to *use* repeated measurements
- Sometimes astronomers used median observations, sometimes mean observations, some times ????

The beginning of the average

Astronomy called for accurate measurements

- In the 1600s, Tycho Brahe advocated incorporating repeated measurements but did not specify how to *use* repeated measurements
- Sometimes astronomers used median observations, sometimes mean observations, some times ????
-

On 1600 January 13/23 at $11^h\ 50^m$ the right ascension of Mars was:

	°	′	″
using the bright foot of Gemini	134	23	39
using Cor Leonis	134	27	37
using Pollux	134	23	18
at $12^h\ 17^m$, using the third in the wing of Virgo	134	29	48
The mean, treating the observations impartially:	134	24	33

Kepler's choice of data representative is baffling. Note that

Average: 134° 26′ 5.5″

Median: 134° 25′ 38″

Early reasoning about the probability distributions of errors

In 1632, Galileo reasoned that

1. There is a true distance, which is only one number

Early reasoning about the probability distributions of errors

In 1632, Galileo reasoned that

1. There is a true distance, which is only one number
2. All observations have errors

Early reasoning about the probability distributions of errors

In 1632, Galileo reasoned that

1. There is a true distance, which is only one number
2. All observations have errors
3. The errors are symmetric around the true value

Early reasoning about the probability distributions of errors

In 1632, Galileo reasoned that

1. There is a true distance, which is only one number
2. All observations have errors
3. The errors are symmetric around the true value
4. Small errors are more common than large ones

The average didn't take on immediately

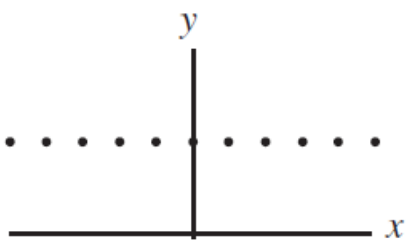
- In the 1660s, Robert Boyle argued that, rather than averaging at all, astronomers should just focus on one very careful experiment

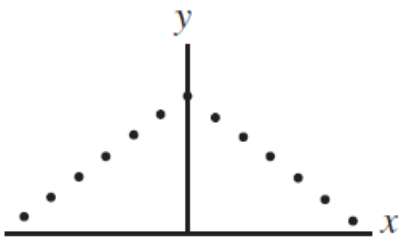
The average didn't take on immediately

- In the 1660s, Robert Boyle argued that, rather than averaging at all, astronomers should just focus on one very careful experiment
- In 1756, Simpson wrote that: *"some persons, of considerable note, have been of opinion, and even publickly maintained, that one single observation, taken with due care, was as much to be relied on as the Mean of a great number "*

Simpson's probability distributions of errors!

- Also in Simpson's 1756 paper was the notion of *probability distributions of errors* or *error curve*





Simpson's probability distributions of errors!

- Also in Simpson's 1756 paper was the notion of *probability distributions of errors* or *error curve*
- Simpson computed the probabilities when the errors could take on values

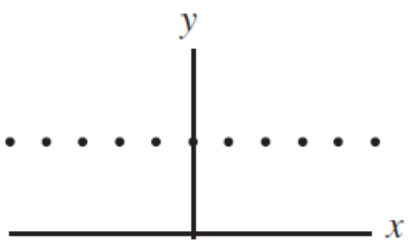
$$-v, \ldots, -2, -1, 0, 1, 2, \ldots, v$$

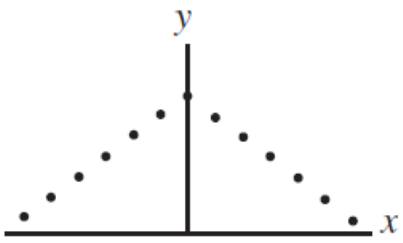
to have probabilities proportional to either

$$r^{-v}, \ldots, r^{-2}, r^{-1}, r^0, r^1, r^2, \ldots, r^4$$

or

$$r^{-v}, 2r^{1-v}, 3r^{2-v} \ldots, (v+1)r^0, \ldots, 2r^{v-1}, r^4$$





Laplace's Error Curves

In 1774, Laplace proposed the first of his probability distributions for errors .

- Probability distribution, ϕ, must be symmetric about 0

Laplace's Error Curves

In 1774, Laplace proposed the first of his probability distributions for errors .

- Probability distribution, ϕ, must be symmetric about 0
- Probability distribution, ϕ, must be monotone decreasing with x

Laplace's Error Curves

In 1774, Laplace proposed the first of his probability distributions for errors .

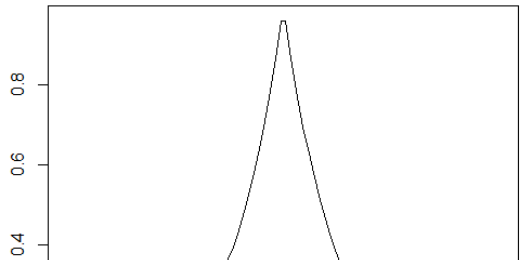
- Probability distribution, ϕ, must be symmetric about 0
- Probability distribution, ϕ, must be monotone decreasing with x
- "as we have no reason to suppose a different law for the ordinates than for their differences...", assumed $\phi(x) \sim \frac{d\phi(x)}{dx}$,

Laplace's Error Curves

In 1774, Laplace proposed the first of his probability distributions for errors .

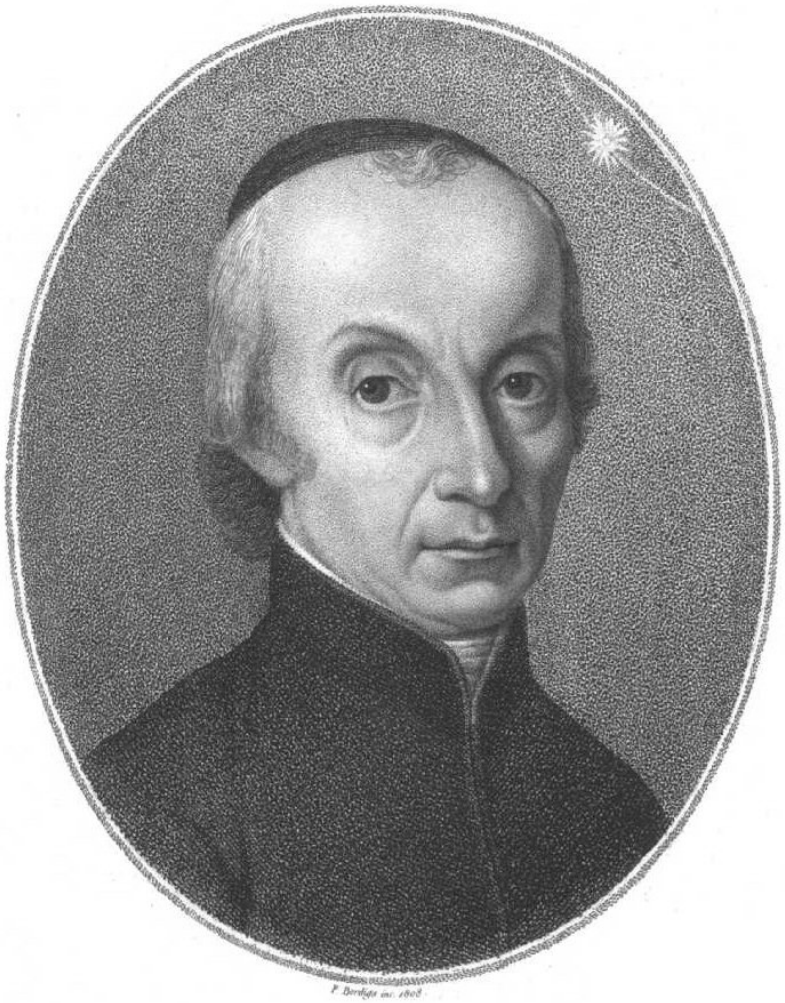
- Probability distribution, ϕ, must be symmetric about 0
- Probability distribution, ϕ, must be monotone decreasing with x
- "as we have no reason to suppose a different law for the ordinates than for their differences...", assumed $\phi(x) \sim \frac{d\phi(x)}{dx}$,
- Thus $\phi(x) = \frac{m}{2} e^{-m|x|}$
  Recall:

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



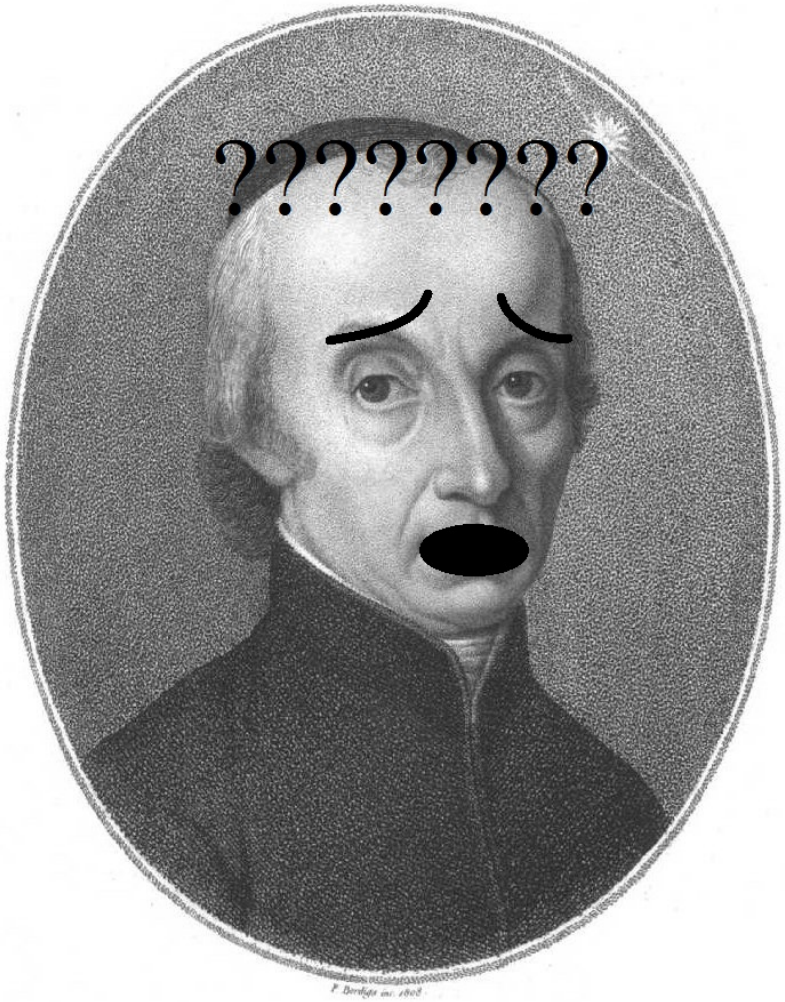
Gauss gets involved

On January 1, 1801, Giuseppe Piazzi spotted something he believed to be a new planet and named Ceres.



Gauss gets involved

Six weeks later, he lost it!



Gauss gets involved

Gauss suggested looking in a different area of the sky than most other astronomers... and he was right!

Gauss' Proof: Assumptions and Notation

Gauss concluded that the probability density for the error is:

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

- Let p be the true but unknown error
- Let $M_1, \ldots, M_n$ be estimates of p
- Let $\phi(x)$ be the probability density function of the random error
  (recall: $P(a < x < b) < \int_a^b \phi(x)$)
- $\phi(x)$ has a maximum at $x = 0$
- $\phi(-x) = \phi(x)$
- The average, $\bar{M} = \frac{1}{n} \sum_{i=1}^{n} M_i$, is the most likely value of p

Gauss' Proof: Part 1 of 4

- Let $f(x) = \frac{\phi'(x)}{\phi(x)}$

Gauss' Proof: Part 1 of 4

- Let $f(x) = \frac{\phi'(x)}{\phi(x)}$
- Then $f(-x) = -f(x)$

Gauss' Proof: Part 1 of 4

- Let $f(x) = \frac{\phi'(x)}{\phi(x)}$
- Then $f(-x) = -f(x)$
- If p is the true value and M_i is the ith estimate, then $M_i - p$ is the ith error

Gauss' Proof: Part 1 of 4

- Let $f(x) = \frac{\phi'(x)}{\phi(x)}$
- Then $f(-x) = -f(x)$
- If p is the true value and M_i is the ith estimate, then $M_i - p$ is the ith error
- If the errors are *independent*, then their joint probability distribution is

$$\Omega = \phi(M_1 - p)\phi(M_2 - p)\cdots\phi(M_n - p)$$

Gauss' Proof: Part 1 of 4

- Let $f(x) = \frac{\phi'(x)}{\phi(x)}$
- Then $f(-x) = -f(x)$
- If p is the true value and M_i is the ith estimate, then $M_i - p$ is the ith error
- If the errors are *independent*, then their joint probability distribution is

$$\Omega = \phi(M_1 - p)\phi(M_2 - p)\cdots\phi(M_n - p)$$

- If $\overline{M} = \frac{1}{n}\sum_{i=1}^{n} M_i$ is the most likely value of p, then $\overline{M}$ should maximize Ω

Gauss' Proof: Part 1 of 4

- Let $f(x) = \frac{\phi'(x)}{\phi(x)}$
- Then $f(-x) = -f(x)$
- If p is the true value and M_i is the ith estimate, then $M_i - p$ is the ith error
- If the errors are *independent*, then their joint probability distribution is

$$\Omega = \phi(M_1 - p)\phi(M_2 - p)\cdots\phi(M_n - p)$$

- If $\overline{M} = \frac{1}{n}\sum_{i=1}^{n} M_i$ is the most likely value of p, then $\overline{M}$ should maximize Ω
- That is, $\frac{d\Omega}{dp}|_{\overline{M}} = 0$

Gauss' Proof: Part 2 of 4

$$\Omega = \phi(M_1 - p)\phi(M_2 - p)\cdots\phi(M_n - p)$$

so $0 = \frac{d\Omega}{dp}\Big|_{p=\overline{M}} =$

Gauss' Proof: Part 2 of 4

$$\Omega = \phi(M_1 - p)\phi(M_2 - p)\cdots\phi(M_n - p)$$

so $0 = \frac{d\Omega}{dp}\Big|_{p=\overline{M}} =$

$$\begin{aligned}
&= -\phi'(M_1 - \overline{M})\phi(M_2 - \overline{M})\cdots\phi(M_n - \overline{M}) \\
&- \phi(M_1 - \overline{M})\phi'(M_2 - p)\cdots\phi(M_n - \overline{M}) \\
&- \ldots \\
&- \phi(M_1 - \overline{M})\phi(M_2 - \overline{M})\cdots\phi'(M_n - \overline{M})
\end{aligned}$$

Gauss' Proof: Part 2 of 4

$$\Omega = \phi(M_1 - p)\phi(M_2 - p)\cdots\phi(M_n - p)$$

so $0 = \frac{d\Omega}{dp}\Big|_{p=\overline{M}} =$

$$\begin{aligned}
&= -\phi'(M_1 - \overline{M})\phi(M_2 - \overline{M})\cdots\phi(M_n - \overline{M}) \\
&- \phi(M_1 - \overline{M})\phi'(M_2 - p)\cdots\phi(M_n - \overline{M}) \\
&- \ldots \\
&- \phi(M_1 - \overline{M})\phi(M_2 - \overline{M})\cdots\phi'(M_n - \overline{M})
\end{aligned}$$

$$= \left(\frac{\phi'(M_1 - \overline{M})}{\phi(M_1 - \overline{M}}) + \frac{\phi'(M_2 - \overline{M})}{\phi(M_2 - \overline{M}}) + \ldots + \frac{\phi'(M_n - \overline{M})}{\phi(M_n - \overline{M}}) \right)\Omega$$

Gauss' Proof: Part 3 of 4

$$0 = \left(\frac{\phi'(M_1 - \overline{M})}{\phi(M_1 - \overline{M}}) + \frac{\phi'(M_2 - \overline{M})}{\phi(M_2 - \overline{M}}) + \ldots + \frac{\phi'(M_n - \overline{M})}{\phi(M_n - \overline{M}}) \right) \Omega$$

Gauss' Proof: Part 3 of 4

$$0 = \left(\frac{\phi'(M_1 - \overline{M})}{\phi(M_1 - \overline{M}}) + \frac{\phi'(M_2 - \overline{M})}{\phi(M_2 - \overline{M}}) + \ldots + \frac{\phi'(M_n - \overline{M})}{\phi(M_n - \overline{M}}) \right) \Omega$$

Recall, we defined $f(x) = \frac{\phi'(x)}{\phi(x)}$, so

$$0 = f(M_1 - \overline{M}) + \ldots + f(M_n - \overline{M})$$

Gauss' Proof: Part 3 of 4

$$0 = \left(\frac{\phi'(M_1 - \overline{M})}{\phi(M_1 - \overline{M}}) + \frac{\phi'(M_2 - \overline{M})}{\phi(M_2 - \overline{M}}) + \ldots + \frac{\phi'(M_n - \overline{M})}{\phi(M_n - \overline{M}}) \right) \Omega$$

Recall, we defined $f(x) = \frac{\phi'(x)}{\phi(x)}$, so

$$0 = f(M_1 - \overline{M}) + \ldots + f(M_n - \overline{M})$$

For arbitrary real numbers M and N, let $M_1 = M$, and $M_2 = M_3 = \ldots = M_n = M - nN$.

Gauss' Proof: Part 3 of 4

$$0 = \left(\frac{\phi'(M_1 - \overline{M})}{\phi(M_1 - \overline{M}}) + \frac{\phi'(M_2 - \overline{M})}{\phi(M_2 - \overline{M}}) + \ldots + \frac{\phi'(M_n - \overline{M})}{\phi(M_n - \overline{M}}) \right) \Omega$$

Recall, we defined $f(x) = \frac{\phi'(x)}{\phi(x)}$, so

$$0 = f(M_1 - \overline{M}) + \ldots + f(M_n - \overline{M})$$

For arbitrary real numbers M and N, let $M_1 = M$, and $M_2 = M_3 = \ldots = M_n = M - nN$.
Then $\overline{M} = M - (n-1)N$

$$0 = f((n-1)N) + (n-1)f(-N) \text{ or } f((n-1)N) = (n-1)f(N)$$

Gauss' Proof: Part 4

$$f((n-1)N) = (n-1)f(N)$$

Gauss' Proof: Part 4

$$f((n-1)N) = (n-1)f(N)$$

If f is continuous, then $f(x) = kx$

Gauss' Proof: Part 4

$$f((n-1)N) = (n-1)f(N)$$

If f is continuous, then $f(x) = kx$

Thus, $\frac{\phi'(x)}{\phi(x)} = kx$

Gauss' Proof: Part 4

$$f((n-1)N) = (n-1)f(N)$$

If f is continuous, then $f(x) = kx$

Thus, $\frac{\phi'(x)}{\phi(x)} = kx$

Inegration with respect to x gives us

$$\ln(\phi(x)) = \frac{k}{2}x^2 + c$$

Gauss' Proof: Part 4

$$f((n-1)N) = (n-1)f(N)$$

If f is continuous, then $f(x) = kx$

Thus, $\frac{\phi'(x)}{\phi(x)} = kx$

Inegration with respect to x gives us

$$\ln(\phi(x)) = \frac{k}{2}x^2 + c$$

$$\phi(x) = Ae^{\frac{kx^2}{2}}$$

Gauss' Proof: Part 4

$$f((n-1)N) = (n-1)f(N)$$

If f is continuous, then $f(x) = kx$
Thus, $\frac{\phi'(x)}{\phi(x)} = kx$
Inegration with respect to x gives us

$$\ln(\phi(x)) = \frac{k}{2}x^2 + c$$

$$\phi(x) = Ae^{\frac{kx^2}{2}}$$

In order for $\phi(x)$ to have a maximum at $x = 0$, we must have that k is negative, so set $k = -h^2$.

Gauss' Proof: Part 4

$$f((n-1)N) = (n-1)f(N)$$

If f is continuous, then $f(x) = kx$
Thus, $\frac{\phi'(x)}{\phi(x)} = kx$
Inegration with respect to x gives us

$$\ln(\phi(x)) = \frac{k}{2}x^2 + c$$

$$\phi(x) = Ae^{\frac{kx^2}{2}}$$

In order for $\phi(x)$ to have a maximum at $x = 0$, we must have that k is negative, so set $k = -h^2$.
We require that $\int_{-\infty}^{\infty} Ae^{-h^2x^2}dx = 1$, so $A = \frac{h}{\sqrt{\pi}}$

Gauss' Proof: Part 4

$$f((n-1)N) = (n-1)f(N)$$

If f is continuous, then $f(x) = kx$
Thus, $\frac{\phi'(x)}{\phi(x)} = kx$
Inegration with respect to x gives us

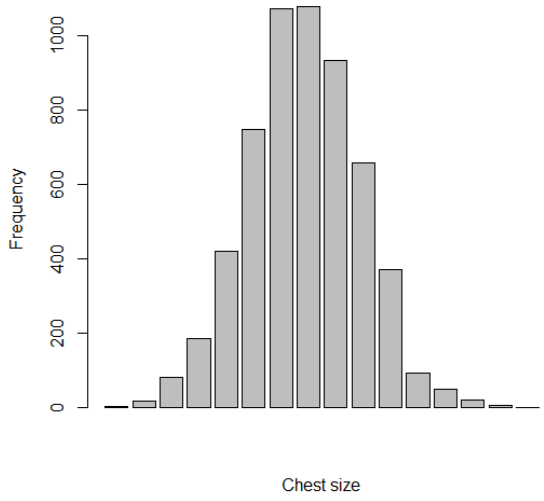
$$\ln(\phi(x)) = \frac{k}{2}x^2 + c$$

$$\phi(x) = Ae^{\frac{kx^2}{2}}$$

In order for $\phi(x)$ to have a maximum at $x = 0$, we must have that k is negative, so set $k = -h^2$.
We require that $\int_{-\infty}^{\infty} Ae^{-h^2x^2}dx = 1$, so $A = \frac{h}{\sqrt{\pi}}$
Thus $\phi(x) = \frac{h}{\sqrt{\pi}}e^{kx^2}$

Not Just For Errors Anymore

- In 1846, Adolphe Quetelet examined chest measurements of Scottish soldiers from a medical journal.



- He realized that measurements of *different* objects of the same type would also often take on the Normal Distribution.
- His idea was that individuals were errors from some true, abstract form: the average man.

The Average Man

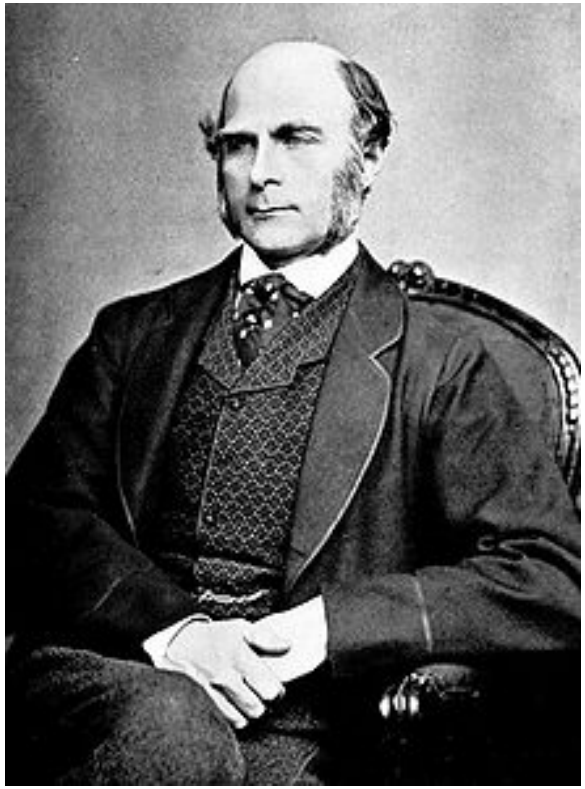
> *"The more knowledge becomes widespread, the more do the deviations from the average diminish, and the more do we tend to approach that which is beautiful and that which is good."*

Quetelet, 1835

Florence Nightingale

- Mentored by Quetelet, translated one of his works to English
- Popularized the pie chart, invented the circular histogram
- Was the first female member of the Royal Statistical Society

Galton



- Quetelet's discovery became very popular in the social sciences
- Galton noticed the normal curve appearing in heights, chest measurements, exam scores, weight of sweet peas
- Galton was one of the first statisticians to use the term "normal" to describe the curve outlined by De Moivre and Gauss

Galton

distinguishable; nevertheless it is well to operate with the eyes shut. The scents chiefly used were peppermint, camphor, carbolic acid, ammonia, and aniseed. I taught myself to associate two whiffs of peppermint with one whiff of camphor; three of peppermint with one of carbolic acid, and so on. Next, I practised at some small sums in addition; at first with the scents themselves, and afterwards altogether with the imagination of them. There was not the slightest difficulty in banishing all visual and auditory images from the mind, leaving nothing in the consciousness besides real or imaginary scents. In this way, without, it is true, becoming very apt at the process, I convinced myself of the possibility of doing sums in simple addition with considerable speed and accuracy solely by means of imaginary scents. Further than this I did not go, so far as addition was concerned. It seemed a serious waste of time to continue the experiments further, because their difficulty and complexity rapidly increased. There were also provoking lapses of memory. For instance, at the present moment, having discontinued the experiments for three months, I find my old lessons almost wholly forgotten. Few persons appreciate the severity of the task imposed on children in making them learn the simple multiplication table, with its 81 pairs of values each associated with a third value. No wonder that they puzzle over it for months, notwithstanding the remarkable receptivity of their fresh brains. I did not attempt multiplication by smell.

Subtraction succeeded as well as addition. I did not go so far as to associate separate scents with the attitudes of mind severally appropriate to subtraction and addition, but determined by my ordinary mental processes which attitude to assume, before isolating myself in the world of scents.

Few experiments were made with taste. Salt, sugar, citric acid, and quinine seemed suitable for the purpose, and there appeared to be little difficulty in carrying on the experiments to a sufficient extent to show that arithmetic by taste was as

Pearson

- In 1893, Pearson called the curve' the *normal* curve (rather than the Gaussian).

Pearson

- In 1893, Pearson called the curve' the *normal* curve (rather than the Gaussian).
- He was trying to give some credit to Laplace, rather than Gauss- not even realizing that DeMoivre had discovered it earlier!

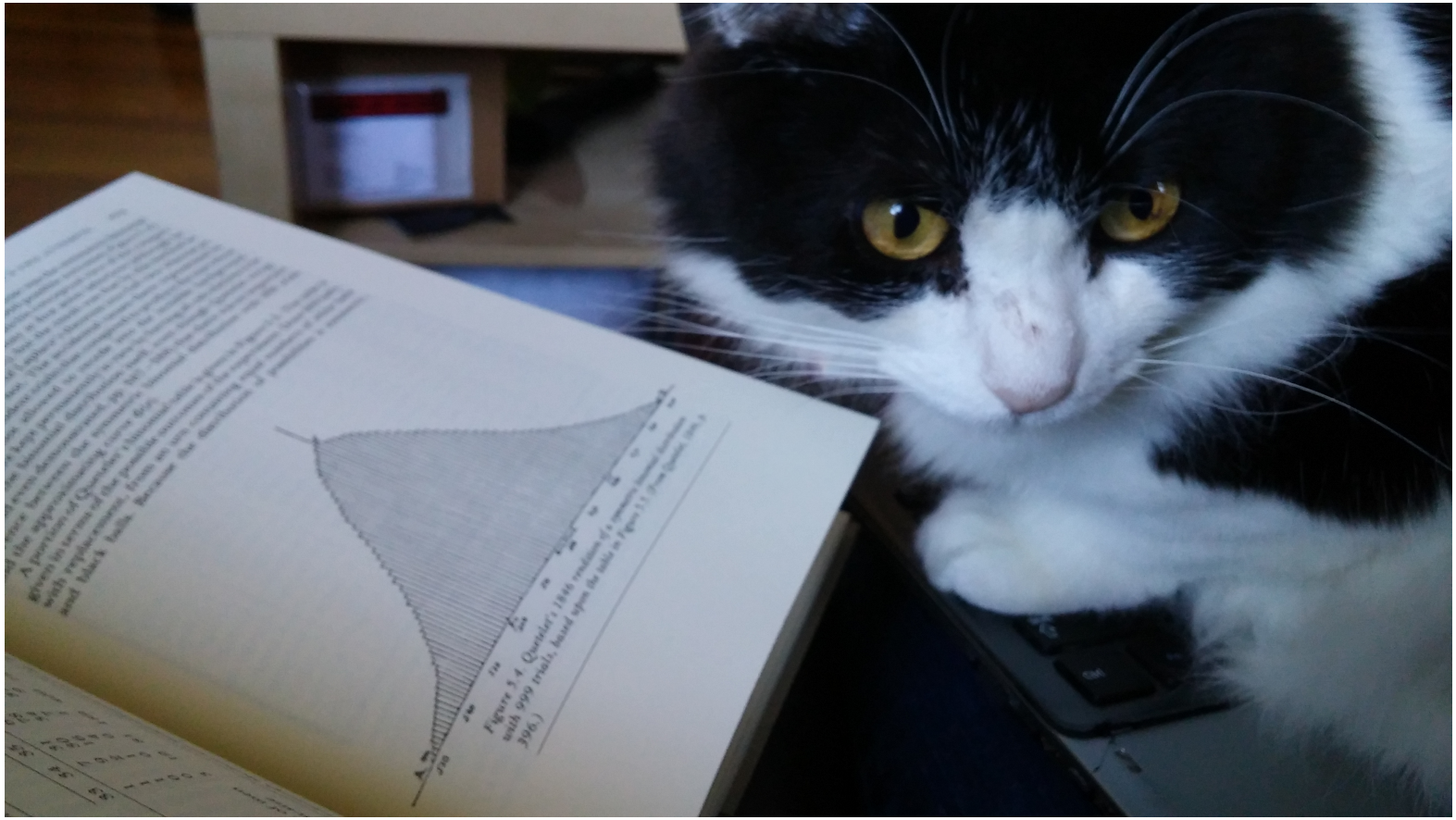
Pearson

- In 1893, Pearson called the curve' the *normal* curve (rather than the Gaussian).
- He was trying to give some credit to Laplace, rather than Gauss- not even realizing that DeMoivre had discovered it earlier!
- In 1900, Pearson developed a method of testing *goodness of fit*

Value at Risk Model

- Value at Risk is a way of quantifying risk of investment into a single number
- For example, if a VaR of 5% is $1 million, that means the probability the portfolio value will decrease more than $ 1 million is 5%
- The original VaR measurement used the normal distribution to model amount a portfolio would gain or lose
- This assumption contributed to the housing crisis.

Everyone Loves The Normal Distribution



Anders Hald. *History of Probability and Statistics and Their Applications before 1750*. John Wiley & Sons, Inc., New Jersey, 2003.

Anders Hald. *A History of Mathematical Statistics From 1750 to 1930*. John Wiley & Sons, Inc., New York, 1998.

Norman L. Johnson and Samual Kotz, editors. *Leading Personalities in Statistical Sciences from teh seventeenth century to the present*. John Wiley & Sons, Inc., New York, 1997.

Stephen M. Stigler. *The History of Statistics: The Measurement of Uncertainty before 1900*. The Belknap Press of Harvard University Press, Massachusetts. 1986.

Saul Stahl. *The Evolution of the Normal Distribution*. Mathematics Magazine, 96-113.