Phys 443
Computational Physics I

Probability Distributions

Probability and Statistics

Probability:

Know parameters of the theory →Predict distributions of possible experiment outcomes

Statistics

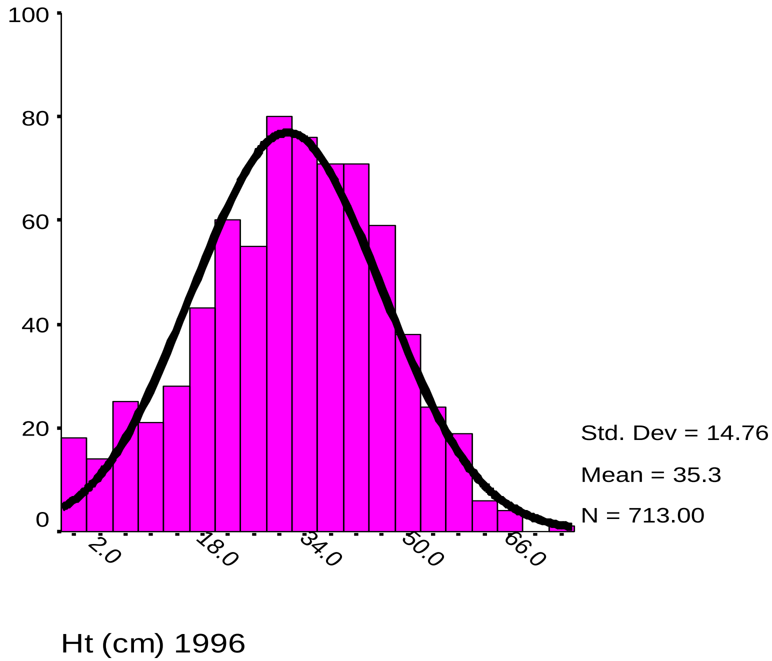
Know the outcome of an experiment → Extract information about the parameters and/or the theory

Probability distributions

These distributions represent good approximations to "real life " .

- We use probability distributions because they work –they fit lots of data in real world

Sarı kantaron



Height (cm) of *Hypericum cumulicola* at Archbold Biological Station

Distributions

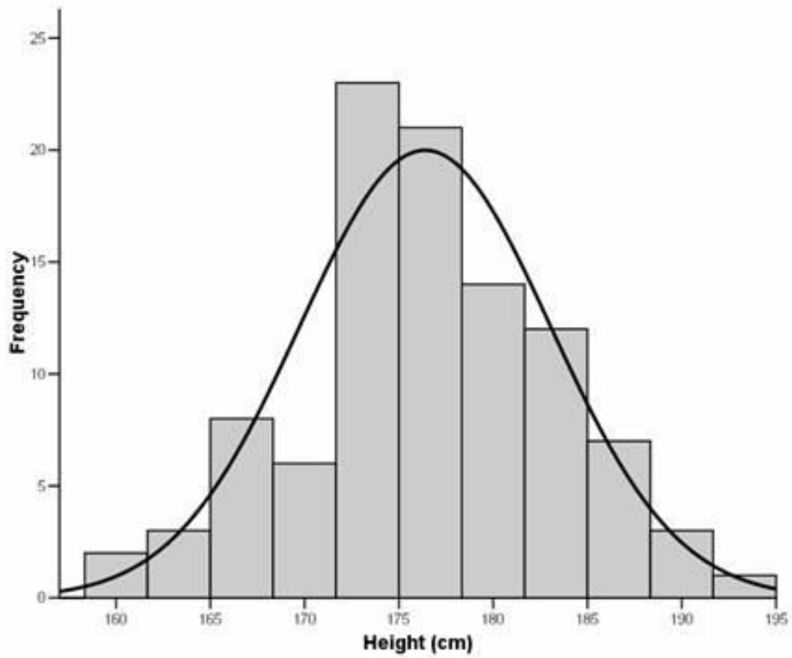
Distributions

In general, the result of repeating the same measurement many times does not lead to the same result.

Experiment:

- Measure the length of one side of your table 10 times and display the results in a histogram.
- What happens if you repeat the measurement 50 times?

Distributions



Pdf/cdf

If p(x) is a density function for some characteristic of a population, then

$$\int_a^b p(x)\,dx = \begin{pmatrix} \text{fraction of the} \\ \text{population for} \\ \text{which } a \le x \le b \end{pmatrix}$$

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

Pdf

We also interpret density functions as probabilities If p(x) is a probability density function (pdf), then

$$\int_a^b p(x)\,dx = \begin{pmatrix} \text{probability} \\ \text{that} \\ a \le x \le b \end{pmatrix}$$

Pdf

Suppose p(x) is a density function for a quantity. The cumulative distribution function (cdf) for the quantity is defined as

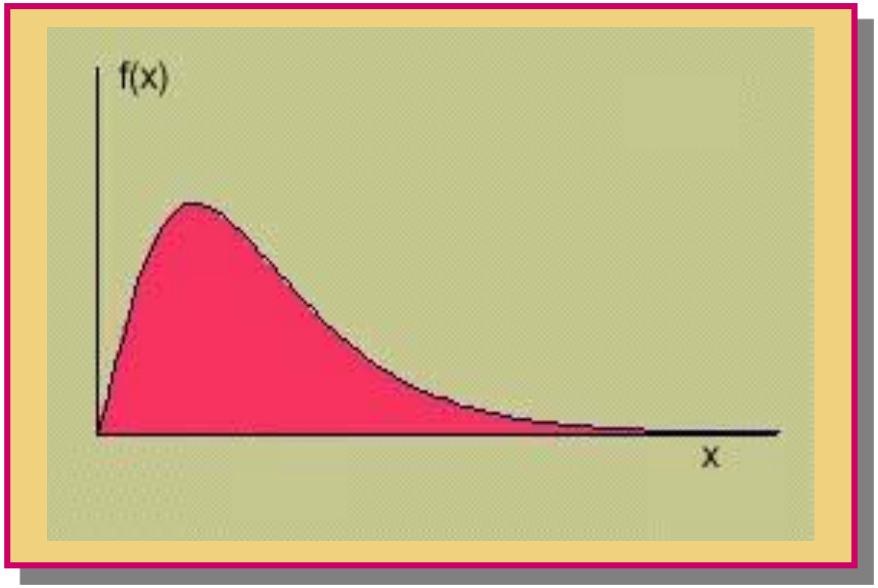
$$P(x) = \int_{-\infty}^{x} p(t)\, dt$$

Gives:

- The proportion of population with value less than x
- The probability of having a value less than x.

Probability Distribution for a Continuous Random Variable

Probability distribution describes how the probabilities are distributed over all possible values. A **probability distribution** for a **continuous random variable x** is specified by a mathematical function denoted by f(x) which is called the **density function**. The graph of a density function is a smooth curve.



Distributions

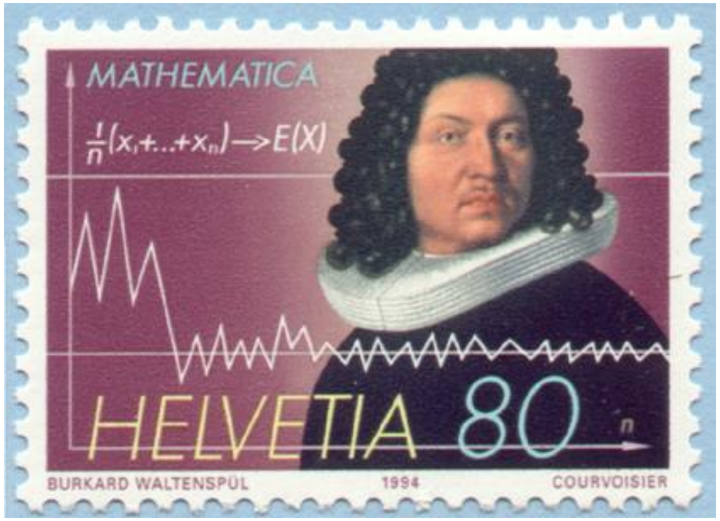
There are two categories of random variables.
These are:

- Discrete random variables
- Continuous random variables

The Binomial Distributions

Many outcomes are binary---yes/no, heads/tails, etc.

- Imagine a simple trial with only two possible outcomes
  - Success (*S*)
  - Failure (*F*)



Jacob Bernoulli (1654-1705)

- Examples
  - Toss of a coin (heads or tails)
  - Sex of a newborn (male or female)
  - Survival of an organism in a region (live or die)

Because the binomial distribution was originally discovered by Jacob Bernoulli (1654-1705), it is sometimes called the Bernouilli distribution.

The Binomial Distributions

- The procedure must have fixed number of trials.
- The trials must be independent.
- Each trial must have all outcomes classified into two categories.
- The probability of success remains the same in all trials.

The Binomial Distributions

- Suppose that the probability of success is p
- What is the probability of failure?
  - $q = 1 - p$
- Examples
  - Toss of a coin (S = head): $p = 0.5 \Rightarrow q = 0.5$
  - Roll of a die (S = 1): $p = 0.1667 \Rightarrow q = 0.8333$
  - Fertility of a chicken egg (S = fertile): $p = 0.8 \Rightarrow q = 0.2$

The Binomial Distributions

- Imagine that a trial is repeated n times
- Examples
  - A coin is tossed 5 times
  - A die is rolled 25 times
- Assume p remains constant from trial to trial and that the trials are statistically independent of each other

The Binomial Distributions

- What is the probability of obtaining x successes in n trials?

- Example
  - What is the probability of obtaining 2 heads from a coin that was tossed 5 times?

$$P(HHTTT) = (1/2)^5 = 1/32$$

The Binomial Distributions

- But there are more possibilities:

HHTTT	HTHTT	HTTHT	HTTTH
	THHTT	THTHT	THTTH
		TTHHT	TTHTH
			TTTHH

$P(2 \text{ heads}) = 10 \times 1/32 = 10/32$

The Binomial Distributions

- In general, if trials result in a series of success and failures,

$$FFSFFFFSFSFSSFFFFFSF\ldots$$

Then the probability of r successes <u>in that order</u> is

$$P(r) = q \cdot q \cdot p \cdot q \cdot \ldots$$
$$= p^r \cdot q^{n-r}$$

Binomial coefficients: number of ways of taking N things m at time

$$C_{N,m} = \binom{N}{m} = \frac{N!}{m!(N-m)!}$$

The Binomial Distributions

- However, if order is not important, then

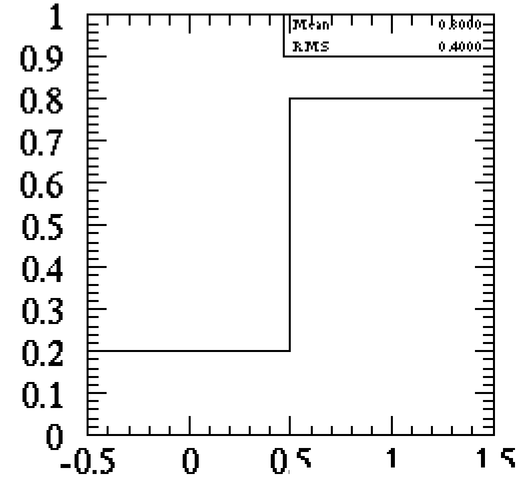
$$P(r) = \frac{n!}{r!(n-r)!} \; p^r \cdot q^{n-r}$$

where $\dfrac{n!}{r!(n-r)!}$ is the number of ways to obtain r successes

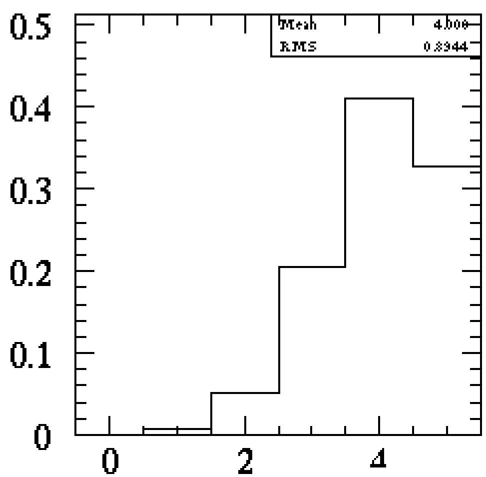
in n trials, and $i! = i \cdot (i-1) \cdot (i-2) \cdot \ldots \cdot 2 \cdot 1$

First term: number of different ways to pick *r* different coins from a collection of *n* total be to heads.
Second term: probability of r coins all getting heads
Third term: probability of n-r coins all getting tails

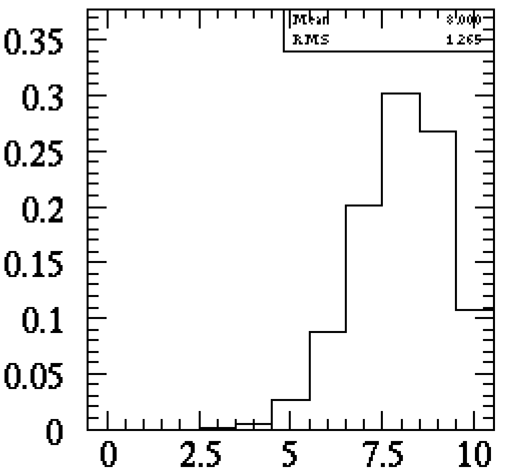
The Binomial Distributions



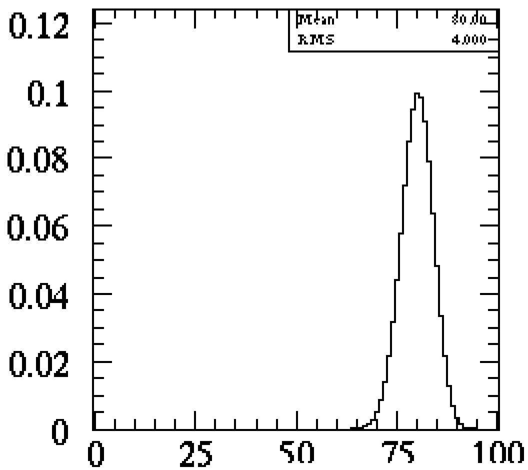
p=0.8,n=1



p=0.8, n=5



p=0.8,n=10



p=0.8, n=100

Mean = np
Variance = np(1-p)

Notice that the mean and variance both scale linearly with n. This is understandable ---flipping n coins is the sum of n independent binomial variables.

The Binomial Distribution

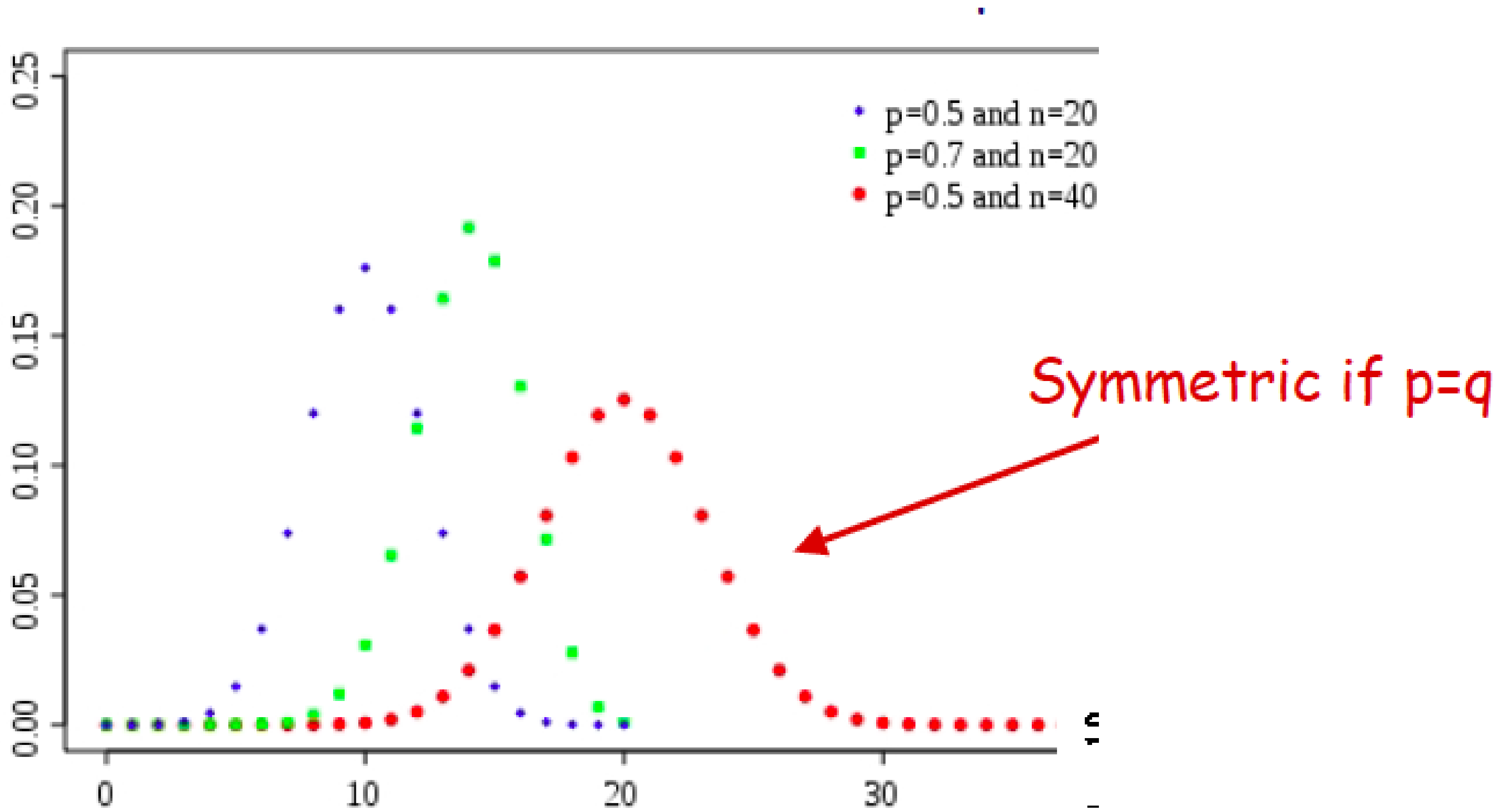
Mean:

$$\mu = \sum_{r=0}^{N} \left[r \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r} \right] = Np$$

Variance:

$$\sigma^2 = \sum_{r=0}^{N} \left[(r-\mu)^2 \frac{N!}{r!(N-r)!} p^r (1-p)^{N-r} \right] = Np(1-p) = Npq$$

The Binomial Distribution



The Binomial Distribution

Use a binomial distribution to model most processes with two outcomes:

- Detection efficiency (either we detect or we don't)
- Cut rejection
- Win-loss records

Example: Suppose you observed *m* special events (success) in a sample of *N* events
The measured probability (“efficiency”) for a special event to occur is:

$$\varepsilon = \frac{m}{N}$$

What is the error (standard deviation) on the probability ("error on the efficiency"):

$$\sigma_\varepsilon = \frac{\sigma_m}{N} = \frac{\sqrt{Npq}}{N} = \frac{\sqrt{N\varepsilon(1-\varepsilon)}}{N} = \sqrt{\frac{\varepsilon(1-\varepsilon)}{N}}$$

The Binomial Distribution

Example: What is the probability of rolling a 1 on a six sided die exactly 10 times when the die is rolled for a total of 24 times.

$r = 10$, $n = 24$, $p = 1/6$, $P_{binom}(r=10) = 0.0025 \sim 1$ in 400

Example: What is the probability that a clinical trial will include 100 smokers in a random cohort of 10,000 when the probability a person is a smoker is 1%.

$r = 100$, $n = 10{,}000$, $p = 1\%$

The Binomial Distribution

What is the probability of obtaining 4 heads out of 7 tosses of an unbiased coin?

Solution:

$$P(4)=\frac{7!}{4!\,3!}\left(\frac{1}{2}\right)^7=\frac{35}{128}.$$

Find the mean, variance and standard deviation

$$\mu = Np = 7 \times \frac{1}{2} = \frac{7}{2} = 3.5$$

$$\sigma^2 = Npq = 7 \times \frac{1}{2} \times \frac{1}{2} = \frac{7}{4} = 1.75 \text{ and}$$

$$\sigma = \sqrt{1.75} \approx 1.323$$

The Binomial Distribution

Example:
Parents each have one brown (B) and one blue (b) gene.
Brown is dominant: Bb → brown eyes (p=0.75).
Parents have 4 children.
X: number of children with brown eyes
Solution:

$$P(X=3)=\binom{4}{3}(0.75)^3(0.25)^1=4\cdot\frac{3^3}{4^4}\approx 0.422$$

The Binomial Distribution: Example

Easiest approach may be simply to list the possibilities:

i. Win in 4 straight games. Probability = p^4

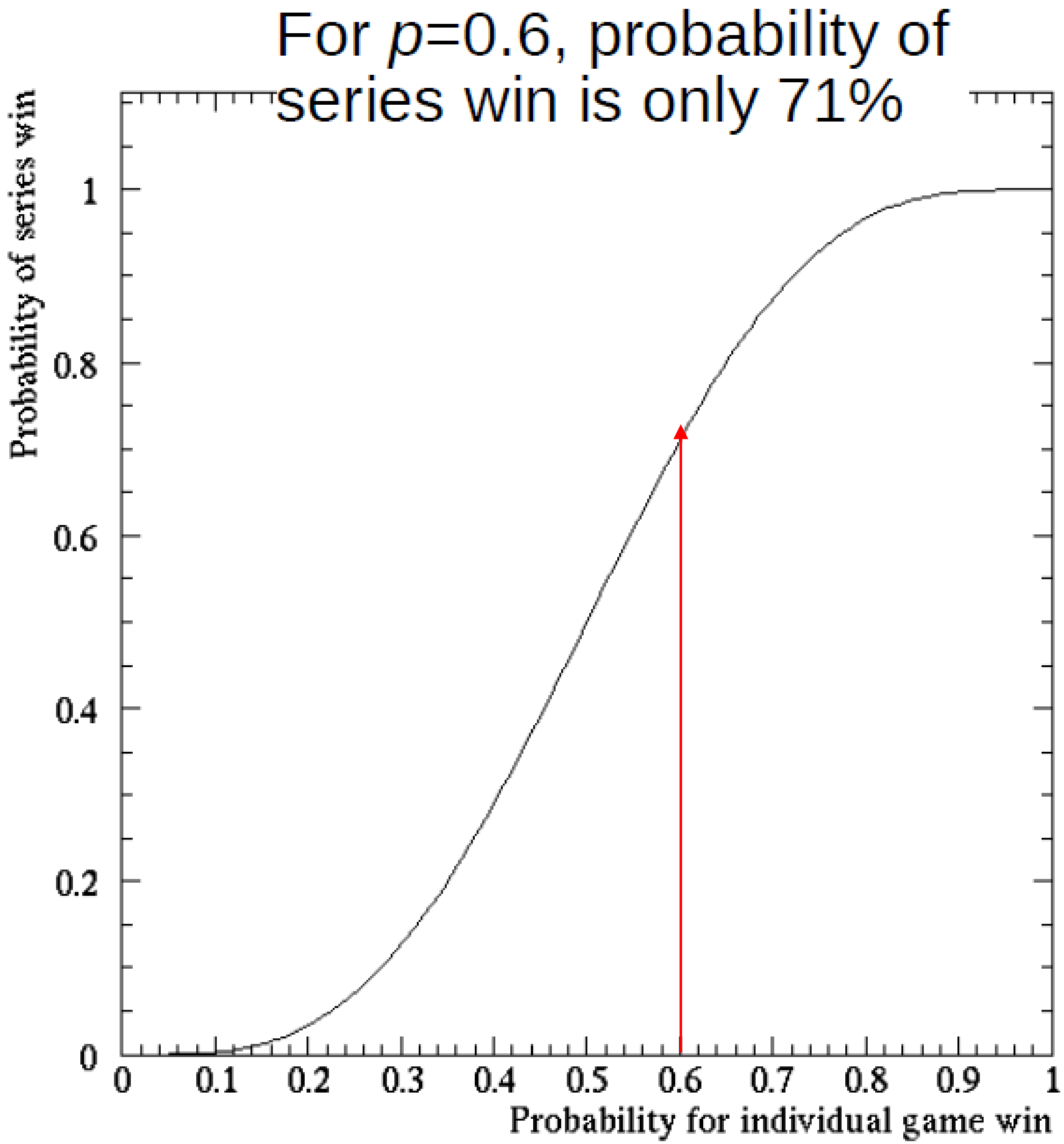
ii. Win in 5 games. Four choices for which game the team gets to lose. Probability = $4p^4(1-p)$

iii. Win in 6 games. Choose 2 of the previous five games to lose. Probability = $C(5,2)p^4(1-p)^2 = 10p^4(1-p)^2$

iv. Win in 7 games. Choose 3 of the previous six games to lose. Probability = $C(6,3)p4(1-p)^3 = 20p^4(1-p)^3$

$$Prob(p) = p^4\left(1+4(1-p)+10(1-p)^2+20(1-p)^3\right)$$

The Binomial Distribution: Example



Negative Binomial distribution

In the negative binomial distribution, you decide how many heads you want to get, then calculate the probability that you have to flip the coin N times before getting that many heads. This gives you a probability distribution for N:

$$P(N|k,p)=\binom{N-1}{k-1}p^k(1-p)^{N-k}$$

Multinomial distribution

We can generalize a binomial distribution to the case where there are more than two possible outcomes. Suppose there are k possible outcomes, and we do N trials. Let n_i be the number of times that the i^{th} outcome comes up, and let p_i be the probability of getting outcome i in one trial. The probability of getting a certain distribution of n_i is then:

$$P(n_1, n_2, \ldots, n_k | p_1 \ldots p_k) = \frac{N!}{n_1! n_2! \ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}$$

Note that there are important constraints on the parameters:

$$\sum_i^k p_i = 1 \qquad \sum_i^k n_i = N$$

Multinomial distribution

Any problem in which there are several discrete outcomes (binomial distribution is a special case).

Note that unlike the binomial distribution, which basically predicts one quantity (the number of heads---you get the number of tails for free), the multinomial distribution is a joint probability distribution for several variables (the various n_i, of which all but one are independent).

The Poisson Distribution

- The Poisson distribution is a widely used discrete probability distribution.
- With binomial conditions it sometimes happens that the rate p of "sucesses" is very small. In a long series of N trials the total numberof successes *Np* may, however, st ill be considerable. It is therefore appealing to examine mathematically the limiting case of the binomial distri bution , when $p \rightarrow 0$, $N \rightarrow \propto$ in such a way that the product Np remains constant and equal to μ, say.

p is very small and approaches 0

- example: a 100 sided dice instead of a 6 sided dice, $p = 1/100$ instead of 1/6
- example: a 1000 sided dice, $p = 1/1000$

N is very large and approaches ∞

- example: throwing 100 or 1000 dice instead of 2 dice

The product *Np* is finite

The Poisson Distribution

- Other phenomena that often follow a Poisson distribution are death of infants, the number of misprints in a book, the number of customers arriving, and the number of activations of a Geiger counter.

- The distribution was derived by the French mathematician Siméon Poisson in 1837, and the first application was the description of the number of deaths by horse kicking in the Prussian army.

- radioactive decay
- number of Prussian soldiers kicked to death by horses per year !
- quality control, failure rate predictions

The Poisson Distribution

Suppose that some event happens at random times with a constant rate R (probability per unit time). (For example, supernova explosions.)

If we wait a time interval dt, then the probability of the event occurring is R dt. If dt is very small, then there is negligible probability of the event occurring twice in any given time interval.

We can therefore divide any time interval of length T into N=T/dt subintervals. In each subinterval an event either occurs or doesn't occur. The total number of events occurring therefore follows a binomial distribution:

$$P(k|p=R\,dt,N)=\frac{N!}{k!(N-k)!}p^k(1-p)^{N-k}$$

The Poisson Distribution

The ***Poisson distribution*** is based on the Poisson process.

- The occurrences of the events are independent in an interval.
- An infinite number of occurrences of the event are possible in the interval.
- The probability of a single event in the interval is proportional to the length of the interval.
- In an infinitely small portion of the interval, the probability of more than one occurrence of the event is negligible.

The Poisson Distribution

Let dt=T/N → 0, so that N goes to infinity. Then

$$P(k|p=R\,dt,N)=\lim_{N\to\infty}\frac{N!}{k!(N-k)!}(RT/N)^k(1-RT/N)^{N-k}$$

$$P(k|p=R\,dt,N)=\lim_{N\to\infty}\frac{N^k}{k!}\left(\frac{RT}{N}\right)^k(1-RT/N)^N(1-RT/N)^{-k}$$

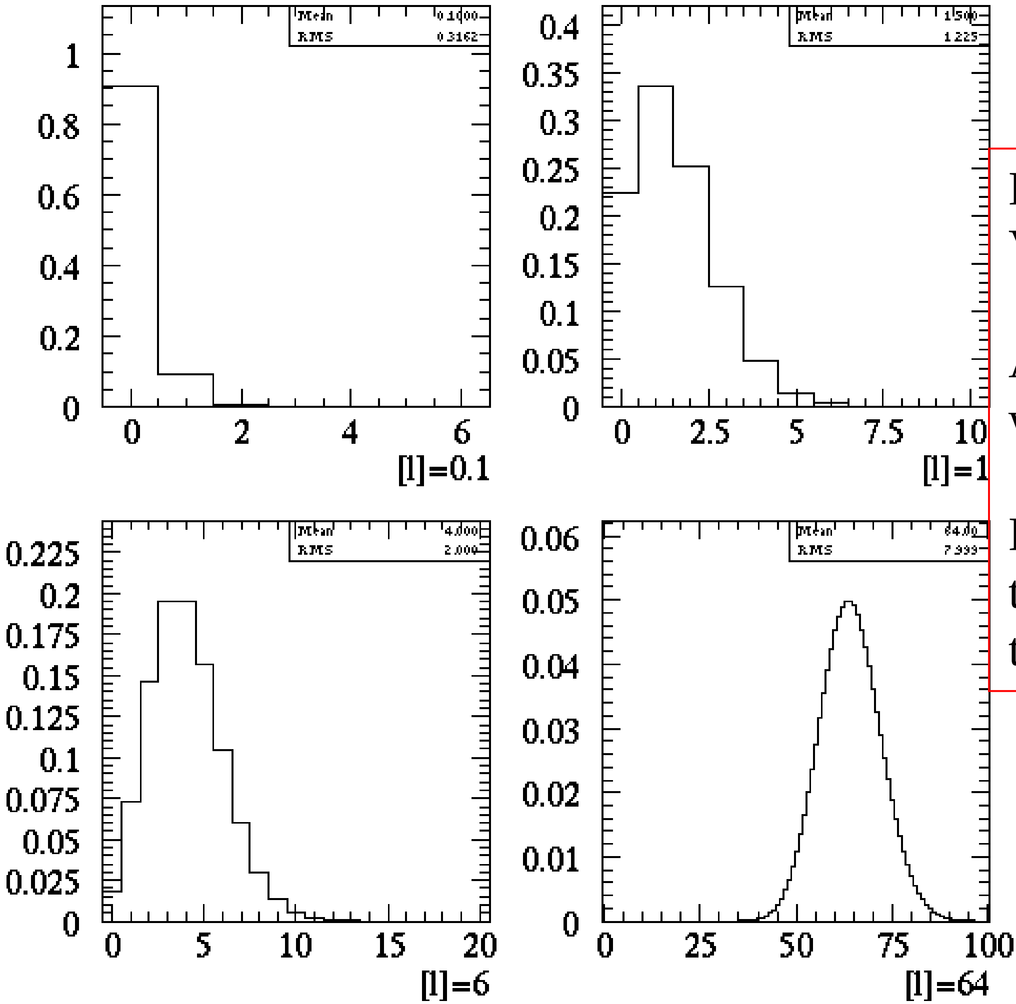
$$=(RT)^k\frac{e^{-RT}}{k!}\equiv\frac{e^{-\lambda}\lambda^k}{k!}$$

P(k|λ) is called the Poisson distribution. It is the probability of seeing k events that happen randomly at constant rate R within a time interval of length T.

From the derivation, it's clear that the binomial distribution approaches a Poisson distribution when p is very small.

λ is the mean number of events expected in interval T.

The Poisson Distribution



Mean = λ
Variance = λ

Approaches Gaussian distribution when λ gets large.

Note that in this case, the standard deviation is in fact equal to sqrt(N).

The Poisson Distribution

- For a Poisson distribution:
  - The **expected value** ($E(X)$) is: $E(X) = \lambda$
  - The **variance** ($Var(X)$) is: $Var(X) = \sigma^2 = \lambda$
  - The **standard deviation** ($SD(X)$) is: $SD(X) = \sigma = \sqrt{\lambda}$

The Poisson Distribution

- Rutherford, Geiger, and Bateman (1910) counted the number of α-particles emitted by a film of polonium in 2608 successive intervals of one-eighth of a minute
  - What is n?
  - What is p?

- Do their data follow a Poisson distribution?

The Poisson Distribution

Emission of α-particles

- Calculation of μ:

$$\mu = \text{No. of particles per interval}$$
$$= 10097/2608$$
$$= 3.87$$

- Expected values:

$$2680 \times P(x) = \; 2608 \times \frac{e^{-3.87}(3.87)^x}{x!}$$

No. α-particles	Observed
0	57
1	203
2	383
3	525
4	532
5	408
6	273
7	139
8	45
9	27
10	10
11	4
12	0
13	1
14	1
Over 14	0
Total	2608

The Poisson Distribution

No. α-particles	Observed	Expected
0	57	54
1	203	210
2	383	407
3	525	525
4	532	508
5	408	394
6	273	254
7	139	140
8	45	68
9	27	29
10	10	11
11	4	4
12	0	1
13	1	1
14	1	1
Over 14	0	0
Total	2608	2680

The Poisson Distribution

Consider you are interested in the radioactive decay of a sample

- The number of nuclei in one milligram of radioactive material is of the order of 10^{29}
- To determine the decay rate, the number of disintegrating nuclei per unit time has to be measured
- It is many order of magnitude smaller than the number of nuclei
- **The binomial distribution correctly describes the probability of observing r events, provided each one has probability p of occurring**

The Poisson Distribution

However,

- The large number of possible events **N** makes exact evaluation impossible
- In an experiment like the one measuring the decay rate of a radiactive material neither the number of possible events **N** nor probability **p** for each one is known

In turn,

- We know the average number of observed events per unit time λ or its estimate r
- $\lambda \ll N$ and $p \ll 1$

The Poisson Distribution

- Poisson Probability Function

Patients arrive at the emergency room of Ankara Hospital at the average rate of 6 per hour on weekend evenings. What is the probability of 4 arrivals in 30 minutes on a weekend evening?

Example

- Births in a hospital occur randomly at an average rate of 1.8 births per hour.

What is the probability of observing 4 births in a given hour at the hospital?

Let X = No. of births in a given hour

(i) Events occur randomly
(ii) Mean rate λ= 1.8

We can now use the formula to calculate the probability of observing exactly 4 births in **a given hour**

$$P(X=4) = e^{-1.8}\frac{1.8^4}{4!} = 0.0723$$

Example

- What about the probability of observing more than or equal to 2 births in a given hour at the Hospital?

$$P(X \geq 2) = P(X = 2) + P(X = 3) + \ldots$$

Example

$$
\begin{aligned}
P(X \geq 2) &= P(X=2) + P(X=3) + \ldots \\
&= 1 - P(X < 2) \\
&= 1 - (P(X=0) + P(X=1)) \\
&= 1 - (\mathrm{e}^{-1.8}\frac{1.8^0}{0!} + \mathrm{e}^{-1.8}\frac{1.8^1}{1!}) \\
&= 1 - (0.16529 + 0.29753) \\
&= 0.537
\end{aligned}
$$

Example

- A radioactive source is found to have a count rate of 5 counts/second. What is probability of

    a) observing no counts in a period of 2 seconds?

    b) five counts in 2 seconds?

– Mean count rate: 5 cnts/sec. --> 10 cnts/2 sec.

– mean = λ, observed cnts = r:

$$P_{\text{Poisson}}(r=0) = \frac{(\lambda=10)^{r=0}\exp(-(\lambda=10))}{(r=0)!} = 4.54 * 10^{-5}$$

$$P_{\text{Poisson}}(r=5) = \frac{(10)^{5}\exp(-10)}{(5)!} = 0.038$$

Example

The average number of traffic accidents on a certain section of highway is two per week. Find the probability of exactly one accident during a one-week period.

$$P(x=1)=\frac{\lambda^{k}e^{-\lambda}}{k!}=\frac{2^{1}e^{-2}}{1!}=2e^{-2}=.2707$$

Neutrinos from supernovae

VOLUME 58, NUMBER 14 PHYSICAL REVIEW LETTERS 6 APRIL 1987

Observation of a Neutrino Burst in Coincidence with Supernova 1987A in the Large Magellanic Cloud

R. M. Bionta,[12] G. Blewitt,[4] C. B. Bratton,[5] D. Casper,[2,14] A. Ciocio,[14] R. Claus,[14] B. Cortez,[16] M. Crouch,[9] S. T. Dye,[6] S. Errede,[10] G. W. Foster,[15] W. Gajewski,[1] K. S. Ganezer,[1] M. Goldhaber,[3] T. J. Haines,[1] T. W. Jones,[7] D. Kielczewska,[1,8] W. R. Kropp,[1] J. G. Learned,[6] J. M. LoSecco,[13] J. Matthews,[2] R. Miller,[1] M. S. Mudan,[7] H. S. Park,[11] L. R. Price,[1] F. Reines,[1] J. Schultz,[1] S. Seidel,[2,14] E. Shumard,[16] D. Sinclair,[2] H. W. Sobel,[1] J. L. Stone,[14] L. R. Sulak,[14] R. Svoboda,[1] G. Thornton,[2] J. C. van der Velde,[2] and C. Wuest[12]

[1]*The University of California, Irvine, Irvine, California 92717*
[2]*The University of Michigan, Ann Arbor, Michigan 48109*
[3]*Brookhaven National Laboratory, Upton, New York 11973*
[4]*California Institute of Technology, Jet Propulsion Laboratory, Pasadena, California 91109*
[5]*Cleveland State University, Cleveland, Ohio 44115*
[6]*The University of Hawaii, Honolulu, Hawaii 96822*
[7]*University College, London WC1E 6BT, United Kingdom*
[8]*Warsaw University, Warsaw, Poland*
[9]*Case Western Reserve University, Cleveland, Ohio 44106*
[10]*The University of Illinois, Urbana, Illinois 61801*
[11]*The University of California, Berkeley, California 94720*
[12]*Lawrence Livermore National Laboratory, Livermore, California 94550*
[13]*The University of Notre Dame, Notre Dame, Indiana 46556*
[14]*Boston University, Boston, Massachusetts 02215*
[15]*Fermi National Accelerator Laboratory, Batavia, Illinois 60510*
[16]*AT&T Bell Laboratories, Summit, New Jersey 07910*

(Received 13 March 1987)

Neutrinos from supernovae

TABLE II. The frequency distribution of events in 10-s intervals of the 6.4-h period containing the neutrino burst.

No. of events	No. of 10-s intervals
0	1043
1	860
2	307
3	78
4	15
5	3
6	0
7	0
8	0
9	1
≥ 10	0

Neutrinos from supernovae

the number of neutrinos detected in 10-second intervals by the IMB detector on 23 February 1987 was:

No. events	0	1	2	3	4	5	6	7	8	9
No. intervals	1042	860	307	78	15	3	0	0	0	1

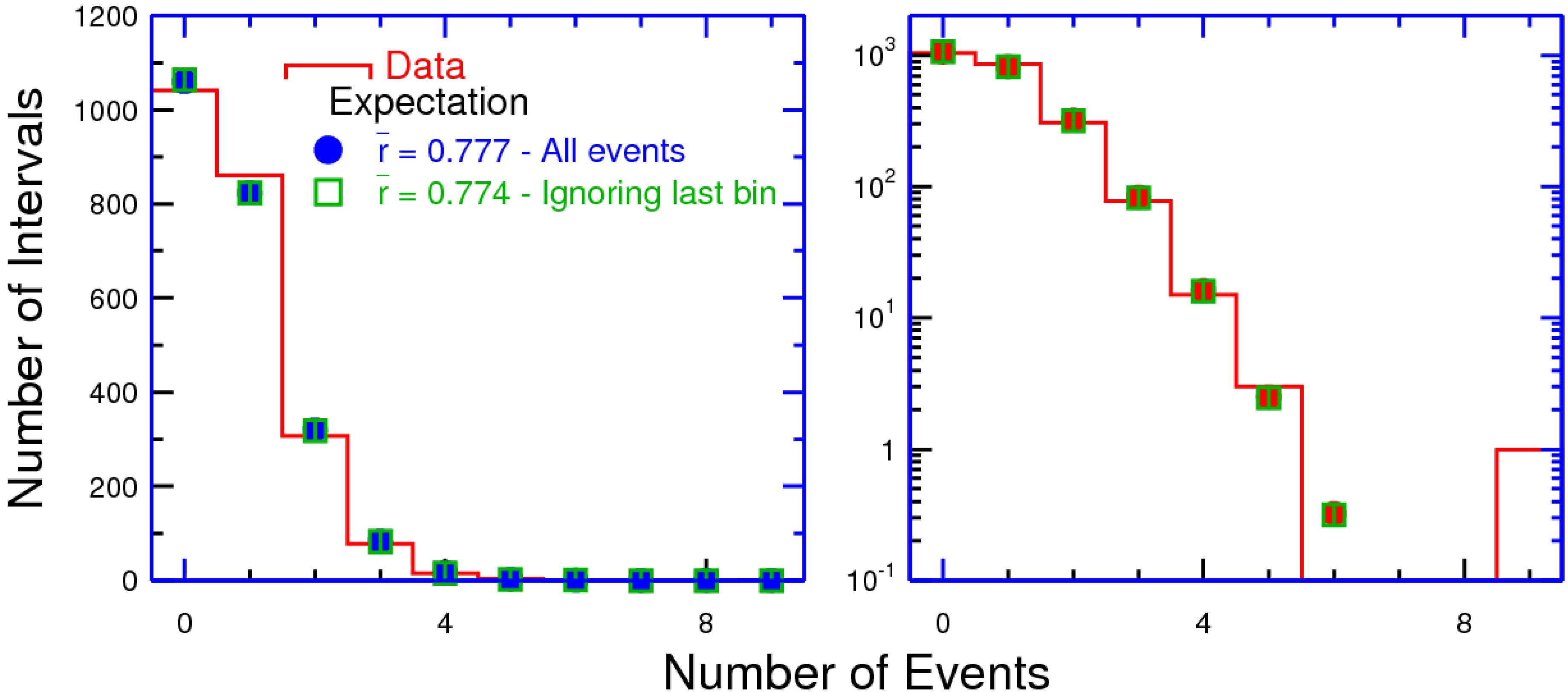
The prediction comes from a Poisson distribution with λ obtained by calculating the weighted average

$$\bar{m} = \hat{\lambda} = \sum_{i=0}^{8} w_i c_i \Big/ \sum_{i=0}^{8} w_i = \frac{0 \cdot 1042 + 1 \cdot 860 + \ldots}{1042 + 860 + \ldots} = 0.77$$

In Class Exercise

Write a python code that

- reads supernova data,
- calculate prediction using poisson distribution
- fit data with poisson distribution



Sum of two poisson variables

Here we will consider the sum of two independent Poisson variables X and Y. If the mean number of expected events of each type are A and B, we naturally would expect that the sum will be a Poisson with mean A+B.

Let Z=X+Y. Consider P(X,Y):

$$P(X,Y)=P(X)P(Y)=\frac{e^{-A}A^X}{X!}\frac{e^{-B}B^Y}{Y!}=\frac{e^{-(A+B)}A^X B^Y}{X!Y!}$$

To find P(Z), sum P(X,Y) over all (X,Y) satisfying X+Y=Z

$$P(Z)=\sum_{X=0}^{Z}\frac{e^{-(A+B)}A^X B^{(Z-X)}}{X!(Z-X)!}=\frac{e^{-(A+B)}}{Z!}\sum_{X=0}^{Z}\frac{Z!A^X B^{(Z-X)}}{X!(Z-X)!}$$

$$P(Z)=\frac{e^{-(A+B)}}{Z!}(A+B)^Z$$

Sum of two poisson variables

Now suppose we know that in hospital A births occur randomly at an average rate of 2.3 births per hour and in hospital B births occur randomly at an average rate of 3.1 births per hour.

What is the probability that we observe 7 births in total from the two hospitals?

If $X \sim Po(\lambda_1)$ on 1 unit interval,
and $Y \sim Po(\lambda_2)$ on 1 unit interval,
then $X + Y \sim Po(\lambda_1 + \lambda_2)$ on 1 unit interval.

Sum of two poisson variables

So if we let X = No. of births in a given hour at hospital A
and Y = No. of births in a given hour at hospital B

Then X ~ Po(2.3), Y ~Po(3.1) and X + Y ~Po(5.4)

$$P(X+Y=7) = e^{-5.4}\frac{5.4^7}{7!} = 0.11999$$

Using the Poisson to approximate the Binomial

The Binomial and Poisson distributions are both discrete probability distributions. In some circumstances the distributions are very similar.

In general,

If n is large (say > 50) and p is small (say < 0.1) then a Bin(n, p) can be approximated with a Po(λ) where λ= np

In many situations it is extremely difficult to use the exact distribution and so approximations are very useful.

Example
Given that 5% of a population are left-handed, use the Poisson distribution to estimate the probability that a random sample of 100 people contains 2 or more left-handed people.

Using the Poisson to approximate the Binomial

X = No. of left handed people in a sample of 100
X ~Bin(100, 0.05)

Poisson approximation) X ~Po(λ) with λ= 100 * 0.05 = 5
We want P(X ≥2)?

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) \\ &= 1 - \Big(P(X=0) + P(X=1)\Big) \\ &\approx 1 - \left(\mathrm{e}^{-5}\frac{5^0}{0!} + \mathrm{e}^{-5}\frac{5^1}{1!}\right) \\ &\approx 1 - 0.040428 \\ &\approx 0.959572 \end{aligned}$$

Table 1

The table gives values of $B(r;n,p) = \binom{n}{r} p^r (1-p)^{n-r}$ for specified values of n,p and r, where $0 \le r \le n$.

The table only has entries for $p \le 0.50$, but it can be used to find values of B(r;n,p) for p > 0.50 by means of the relation

$$B(r;n,p) = B(n-r;n,1-p).$$

n	r	p = .01	.02	.03	.05	.10	.15	.20	.25	.30	.40	.50
1	0	.9900	.9800	.9700	.9500	.9000	.8500	.8000	.7500	.7000	.6000	.5000
	1	.0100	.0200	.0300	.0500	.1000	.1500	.2000	.2500	.3000	.4000	.5000
2	0	.9801	.9604	.9409	.9025	.8100	.7225	.6400	.5625	.4900	.3600	.2500
	1	.0198	.0392	.0582	.0950	.1800	.2550	.3200	.3750	.4200	.4800	.5000
	2	.0001	.0004	.0009	.0025	.0100	.0225	.0400	.0625	.0900	.1600	.2500
3	0	.9703	.9412	.9127	.8574	.7290	.6141	.5120	.4219	.3430	.2160	.1250
	1	.0294	.0576	.0847	.1354	.2430	.3251	.3840	.4219	.4410	.4320	.3750
	2	.0003	.0012	.0026	.0071	.0270	.0574	.0960	.1406	.1890	.2880	.3750
	3	.0000	.0000	.0000	.0001	.0010	.0034	.0080	.0156	.0270	.0640	.1250
4	0	.9606	.9224	.8853	.8145	.6561	.5220	.4096	.3164	.2401	.1296	.0625
	1	.0388	.0753	.1095	.1715	.2916	.3685	.4096	.4219	.4116	.3456	.2500
	2	.0006	.0023	.0051	.0135	.0486	.0975	.1536	.2109	.2646	.3456	.3750
	3	.0000	.0000	.0001	.0005	.0036	.0115	.0256	.0469	.0756	.1536	.2500
	4	.0000	.0000	.0000	.0000	.0001	.0005	.0016	.0039	.0081	.0256	.0625
5	0	.9510	.9039	.8587	.7738	.5905	.4437	.3277	.2373	.1681	.0778	.0313
	1	.0480	.0922	.1328	.2036	.3281	.3915	.4096	.3955	.3601	.2592	.1563
	2	.0010	.0038	.0082	.0214	.0729	.1382	.2048	.2637	.3087	.3456	.3125
	3	.0000	.0001	.0003	.0011	.0081	.0244	.0512	.0879	.1323	.2304	.3125
	4	.0000	.0000	.0000	.0000	.0005	.0022	.0064	.0146	.0284	.0768	.1563
	5	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010	.0024	.0102	.0313
6	0	.9415	.8858	.8330	.7351	.5314	.3771	.2621	.1780	.1176	.0467	.0156
	1	.0571	.1085	.1546	.2321	.3543	.3993	.3932	.3560	.3025	.1866	.0937
	2	.0014	.0055	.0120	.0305	.0984	.1762	.2458	.2966	.3241	.3110	.2344
	3	.0000	.0002	.0005	.0021	.0146	.0415	.0819	.1318	.1852	.2764	.3125
	4	.0000	.0000	.0000	.0001	.0012	.0055	.0154	.0330	.0595	.1382	.2344
	5	.0000	.0000	.0000	.0000	.0001	.0004	.0015	.0044	.0102	.0369	.0937
	6	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0007	.0041	.0156
7	0	.9321	.8681	.8080	.6983	.4783	.3206	.2097	.1335	.0824	.0280	.0078
	1	.0659	.1240	.1749	.2573	.3720	.3960	.3670	.3115	.2471	.1306	.0547
	2	.0020	.0076	.0162	.0406	.1240	.2097	.2753	.3115	.3177	.2613	.1641
	3	.0000	.0003	.0008	.0036	.0230	.0617	.1147	.1730	.2269	.2903	.2734
	4	.0000	.0000	.0000	.0002	.0026	.0109	.0287	.0577	.0972	.1935	.2734
	5	.0000	.0000	.0000	.0000	.0002	.0012	.0043	.0115	.0250	.0774	.1641
	6	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0013	.0036	.0172	.0547
	7	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0016	.0078
8	0	.9227	.8508	.7837	.6634	.4305	.2725	.1678	.1001	.0576	.0168	.0039
	1	.0746	.1389	.1939	.2793	.3826	.3847	.3355	.2670	.1977	.0896	.0313
	2	.0026	.0099	.0210	.0515	.1488	.2376	.2936	.3115	.2965	.2090	.1094
	3	.0001	.0004	.0013	.0054	.0331	.0839	.1468	.2076	.2541	.2787	.2188
	4	.0000	.0000	.0001	.0004	.0046	.0185	.0459	.0865	.1361	.2322	.2734
	5	.0000	.0000	.0000	.0000	.0004	.0026	.0092	.0231	.0467	.1239	.2188
	6	.0000	.0000	.0000	.0000	.0000	.0002	.0011	.0038	.0100	.0413	.1094
	7	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0012	.0079	.0313
	8	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0007	.0039
9	0	.9135	.8337	.7602	.6302	.3874	.2316	.1342	.0751	.0404	.0101	.0020
	1	.0830	.1531	.2116	.2985	.3874	.3679	.3020	.2253	.1556	.0605	.0176
	2	.0034	.0125	.0262	.0629	.1722	.2597	.3020	.3003	.2668	.1612	.0703
	3	.0001	.0006	.0019	.0077	.0446	.1069	.1762	.2336	.2668	.2508	.1641
	4	.0000	.0000	.0001	.0006	.0074	.0283	.0661	.1168	.1715	.2508	.2461
	5	.0000	.0000	.0000	.0000	.0008	.0050	.0165	.0389	.0735	.1672	.2461
	6	.0000	.0000	.0000	.0000	.0001	.0006	.0028	.0087	.0210	.0743	.1641
	7	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0012	.0039	.0212	.0703
	8	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0035	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0020

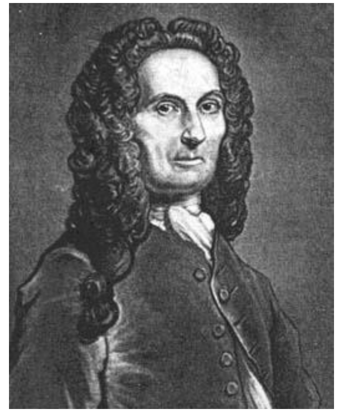
The Gaussian Distribution

- Discovered in 1733 by de Moivre as an approximation to the binomial distribution when the number of trails is large

- Derived in 1809 by Gauss

- Importance lies in the Central Limit Theorem, which states that the sum of a large number of independent random variables (binomial, Poisson, etc.) will approximate a normal distribution

  – Example: Human height is determined by a large number of factors, both genetic and environmental, which are additive in their effects. Thus, it follows a normal distribution.



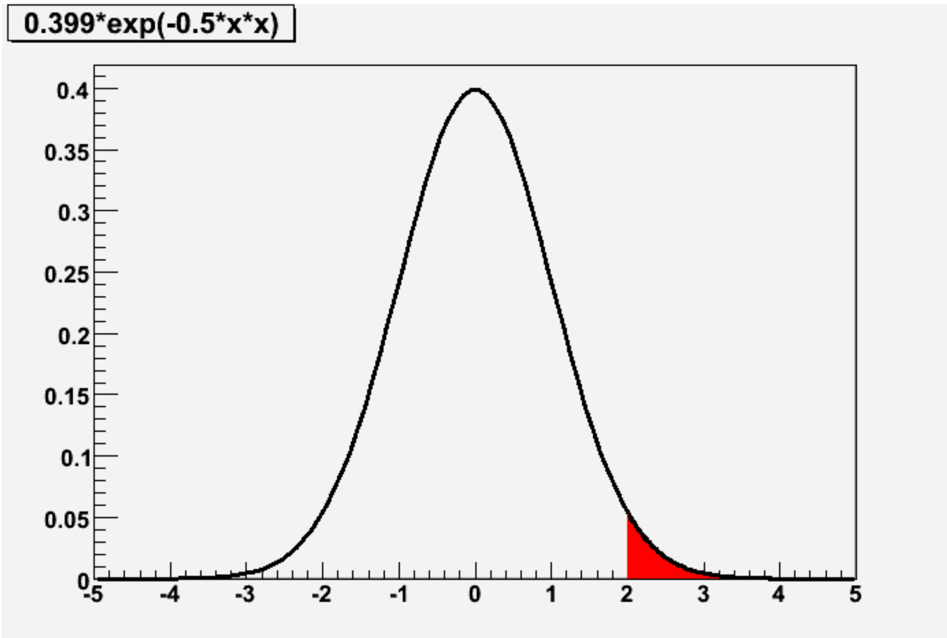
Abraham de Moivre
(1667-1754)

Karl F. Gauss
(1777-1855)

The Gaussian Distribution

By far the most useful distribution is the Gaussian (normal) distribution:

$$P(x|\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



68.27% of area within $\pm 1\sigma$

95.45% of area within $\pm 2\sigma$

99.73% of area within $\pm 3\sigma$

Mean = μ, Variance=σ^2

Note that width scales with σ.

Area out on tails is important---use lookup tables or cumulative distribution function.

In plot to left, red area ($>2\sigma$) is 2.3%.

90% of area within $\pm 1.645\sigma$

95% of area within $\pm 1.960\sigma$

99% of area within $\pm 2.576\sigma$

The Gaussian Distribution

Gaussian distribution

General form:

Continuous variable

Mean $\mu = \int_{-\infty}^{+\infty} xP(x)dx$

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

Variance $\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 P(x)dx$

Normalization

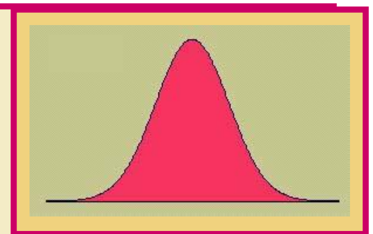
$$\int_{-\infty}^{+\infty} dx\ P(x) = 1$$

Limiting case of the binomial distribution when n → $\propto$, np >>1

The Normal Distribution

The formula that generates the normal probability distribution is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for } -\infty < x < \infty$$

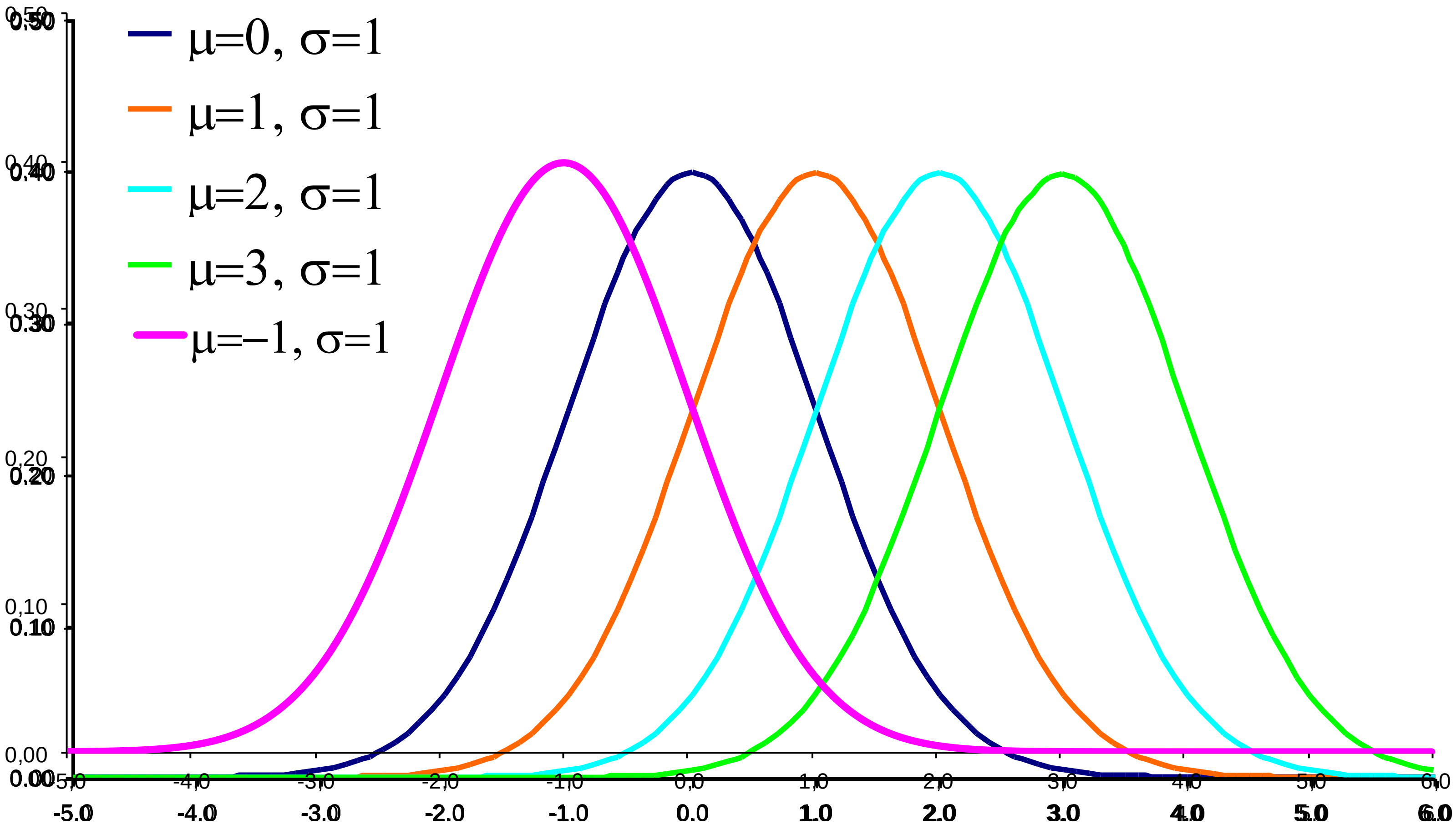


$$e = 2.7183 \qquad \pi = 3.1416$$

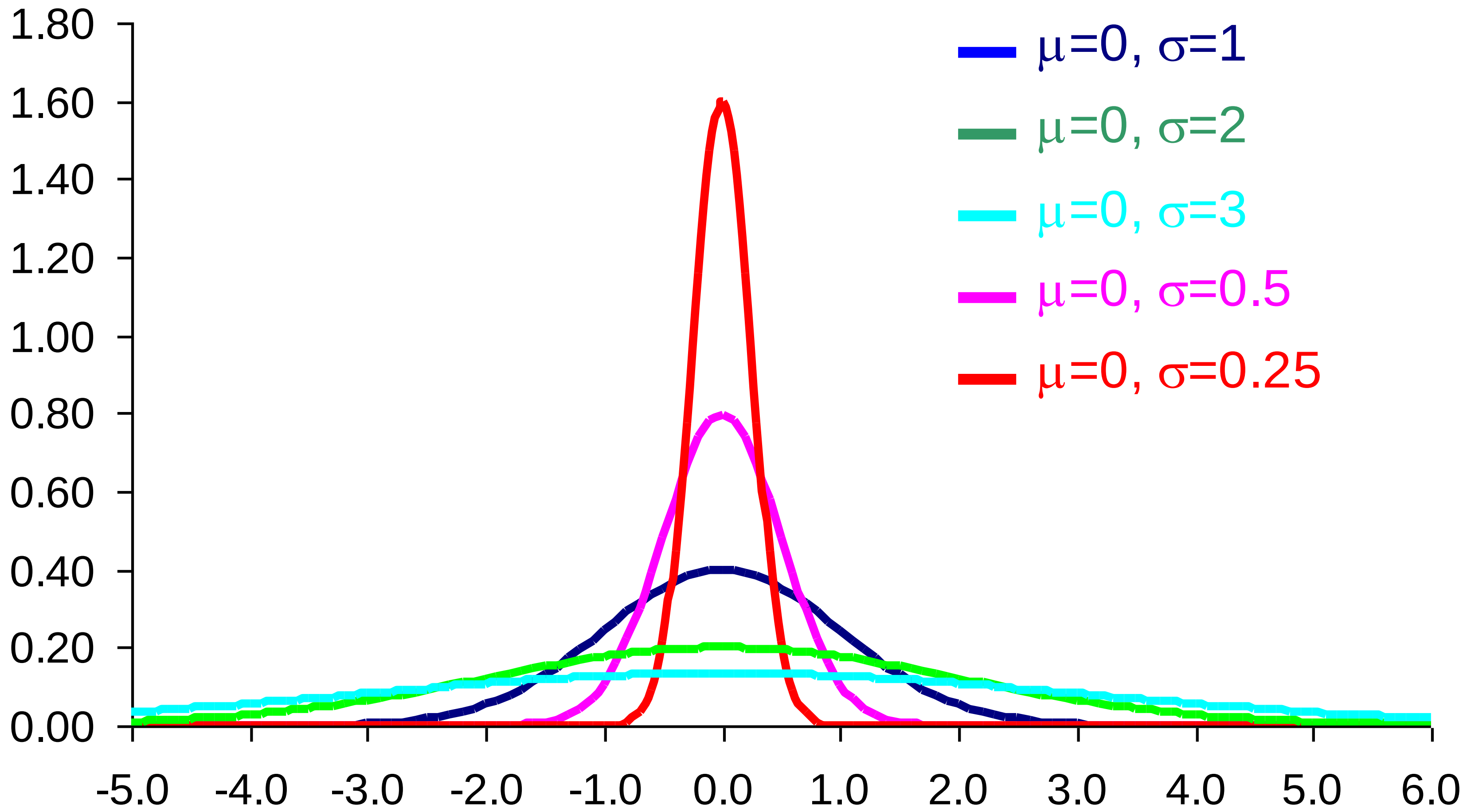
μ and σ are the population mean and standard deviation.

Two parameters, mean and standard deviation, completely determine the Normal distribution. The shape and location of the normal curve changes as the mean and standard deviation change.

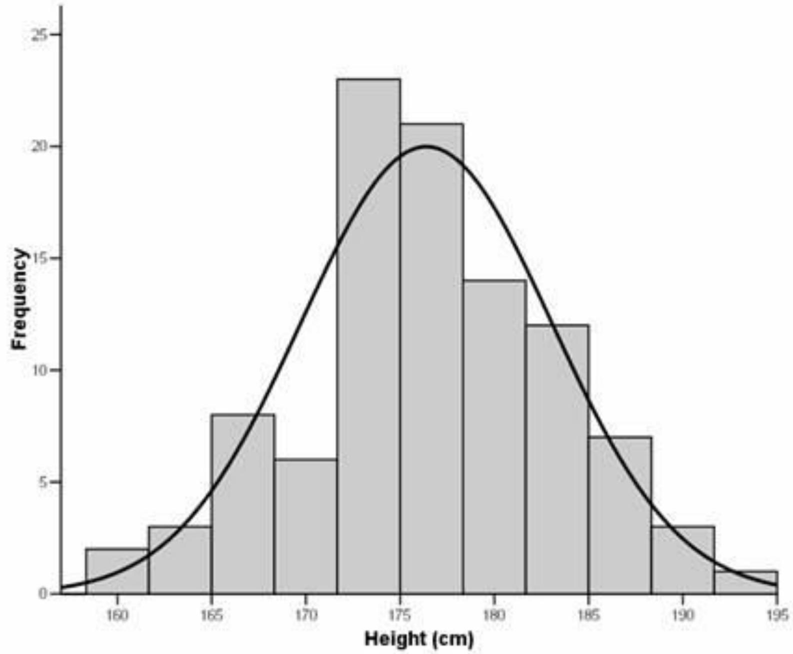
Normal Distributions: σ=1



The Normal Distribution



The Gaussian Distribution



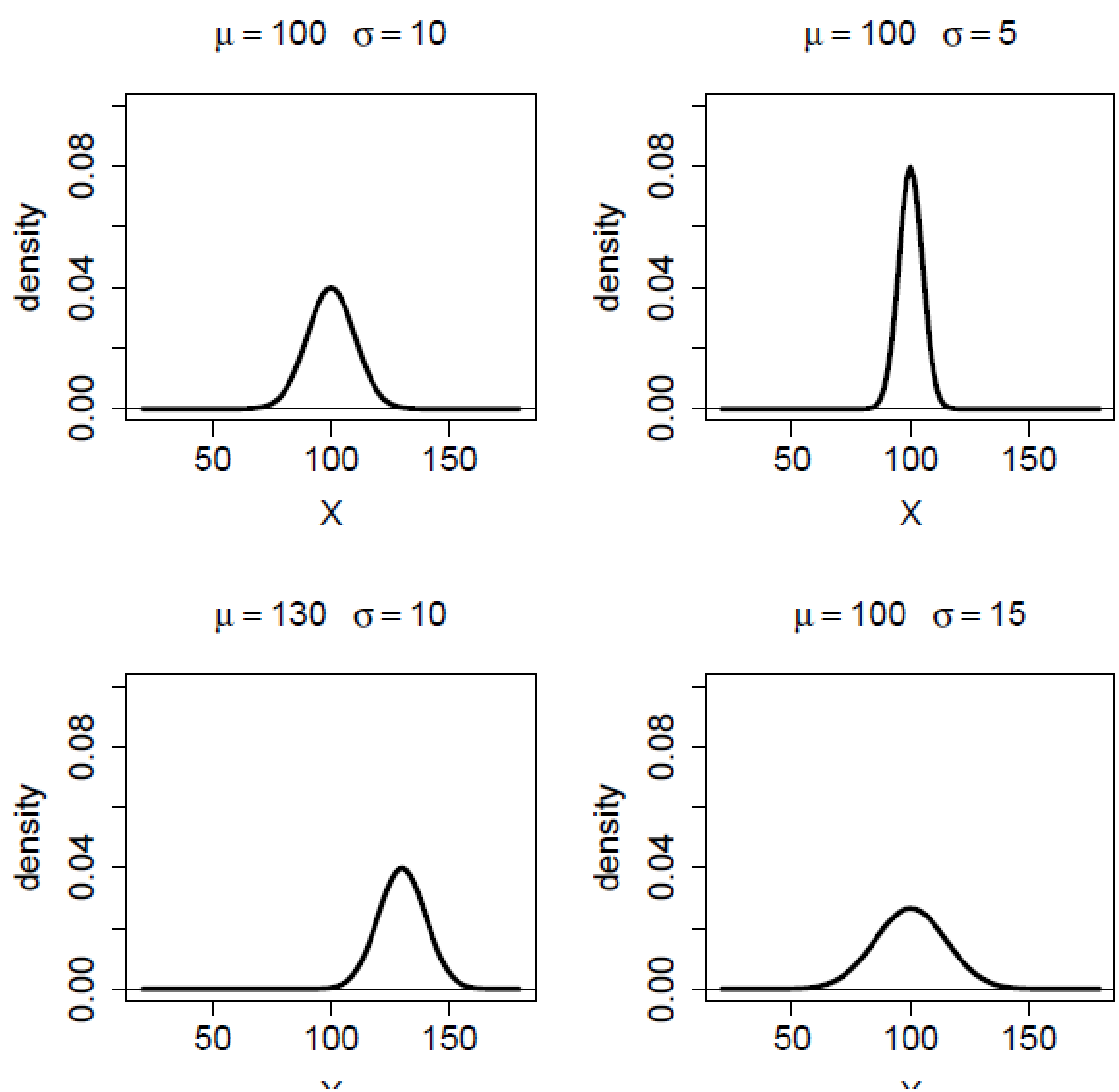
The Gaussian Distribution

Where the Gaussian distribution apply?

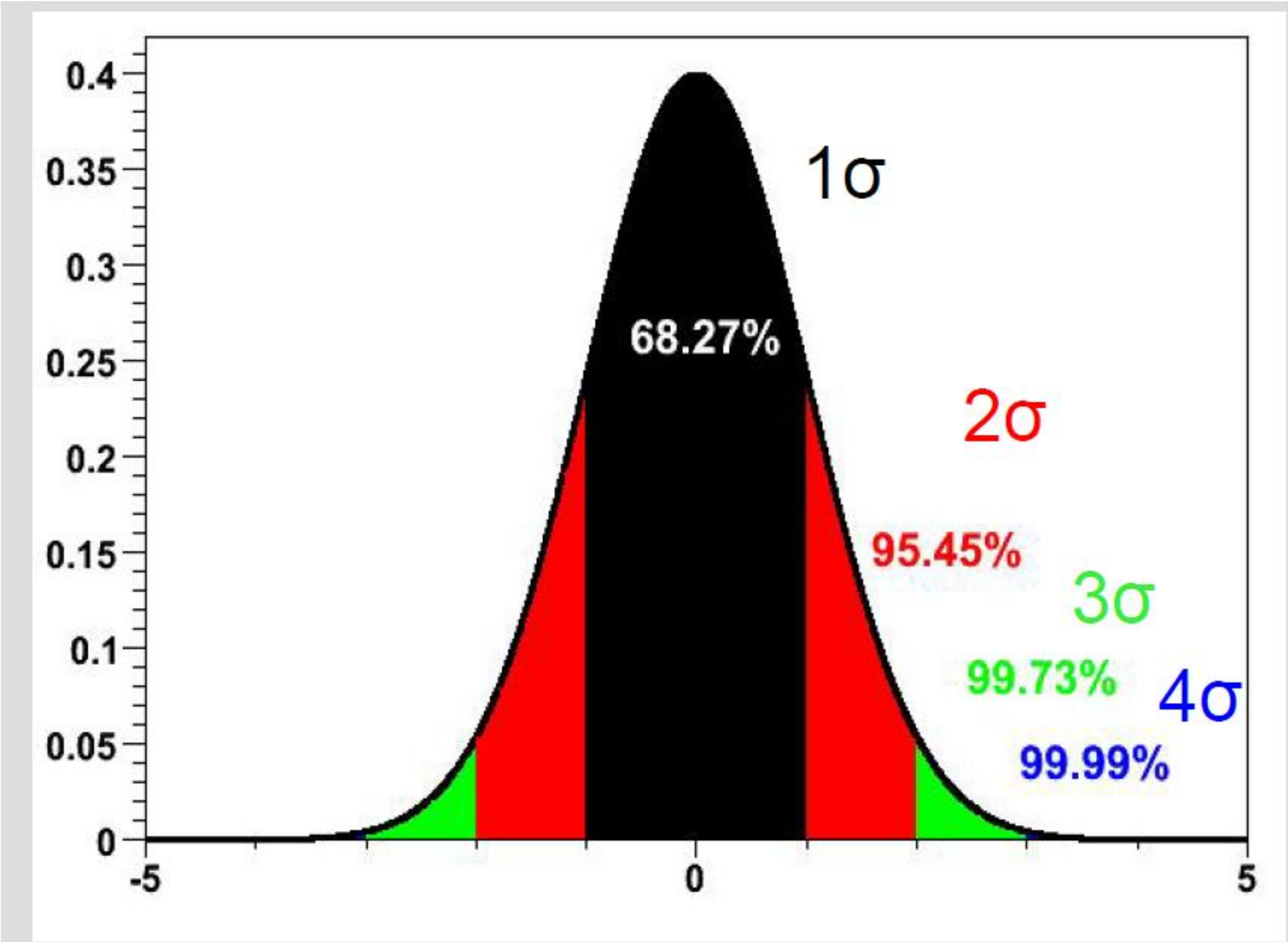
The reult of repeating an experiment many times produces a spread of answers whose distribution is approsimately Gaussian

If the individual errors contributing to the final answer is small, the approximation to a Gaussian is especially good

The Gaussian Distribution



The Gaussian Distribution

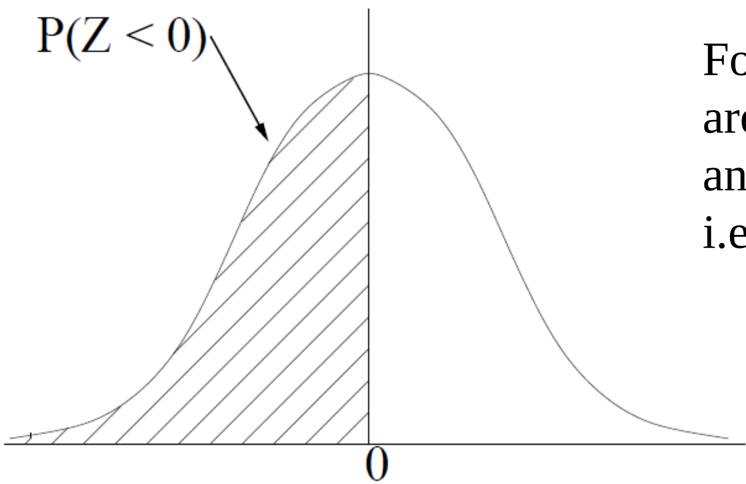


Calculating probabilities from Gaussian

For a discrete probability distribution we calculate the probability of being less than some value x, i.e. $P(X < x)$, by simply summing up the probabilities of the values less than x.

For a continuous probability distribution we calculate the probability of being less than some value x, i.e. $P(X < x)$, by calculating the area under the curve to the left of x.

For example, suppose $X \sim N(0, 1)$ and we want to calculate $P(X < 0)$?

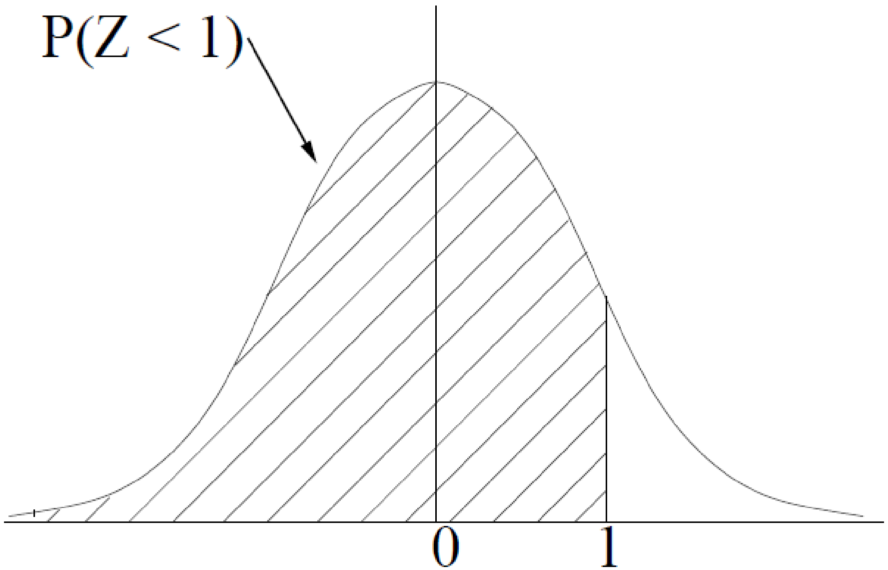


For this example we can calculate the required area as we know the distribution is symmetric and the total area under the curve is equal to 1, i.e. $P(X < 0) = 0:5$.

Calculating probabilities from Gaussian

What about P(X <1.0) ?

Calculating this area is not easy and so we use probability tables. Probability tables are tables of probabilities that have been calculated on a computer. All we have to do is identify the right probability in the table and copy it down! Obviously it is impossible to tabulate all possible probabilities for all possible Normal distributions so only one special Normal distribution, N(0, 1), has been tabulated.



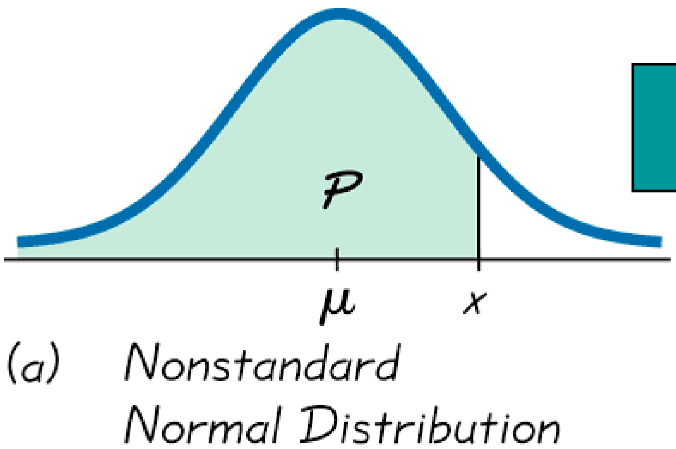
The N(0, 1) distribution is called the standard Normal distribution

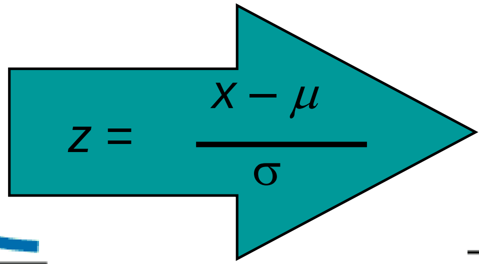
The Gaussian Distribution

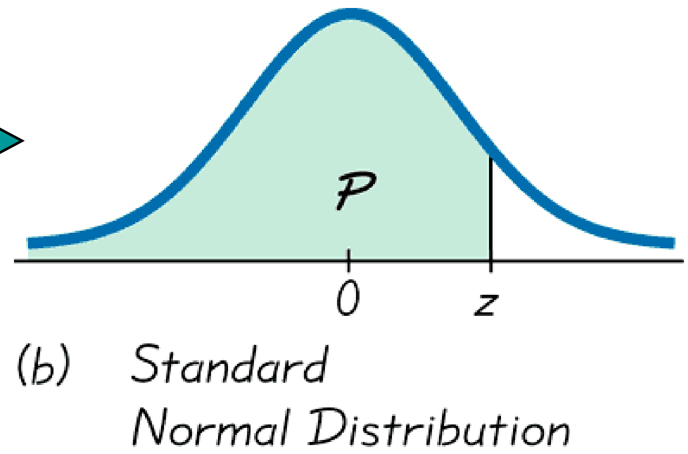
- To find P(a < *x* < b), we need to find the area under the appropriate normal curve.
- To simplify the tabulation of these areas, we **standardize** each value of *x* by expressing it as a *z*-score, the number of standard deviations σ it lies from the mean μ.

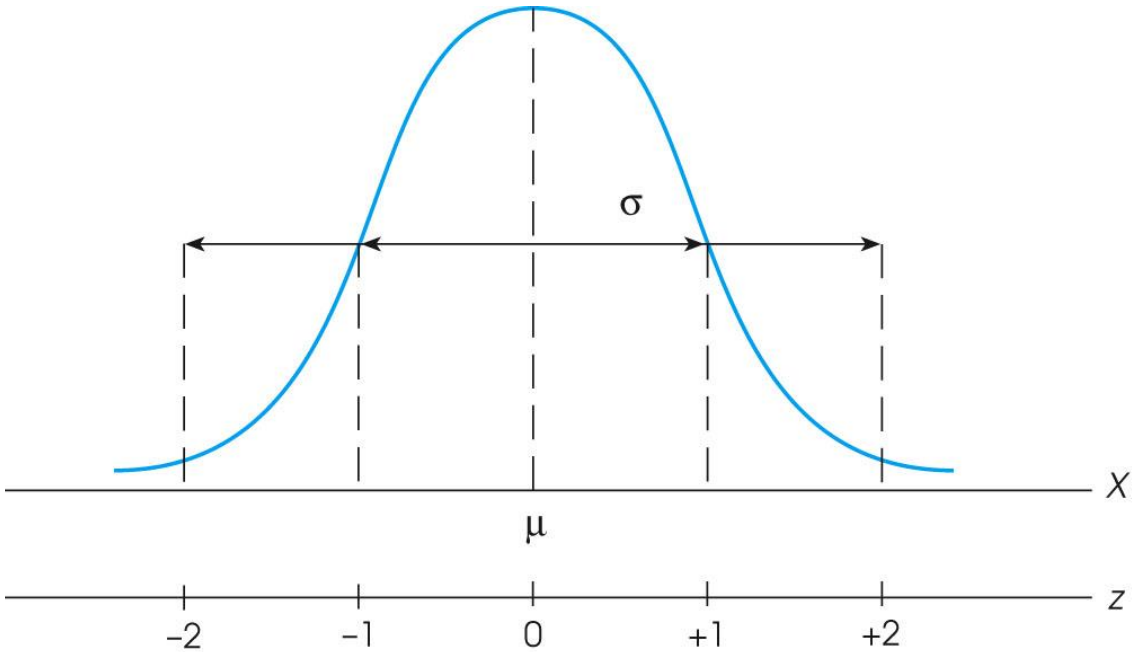
$$\boxed{z = \frac{x - \mu}{\sigma}}$$

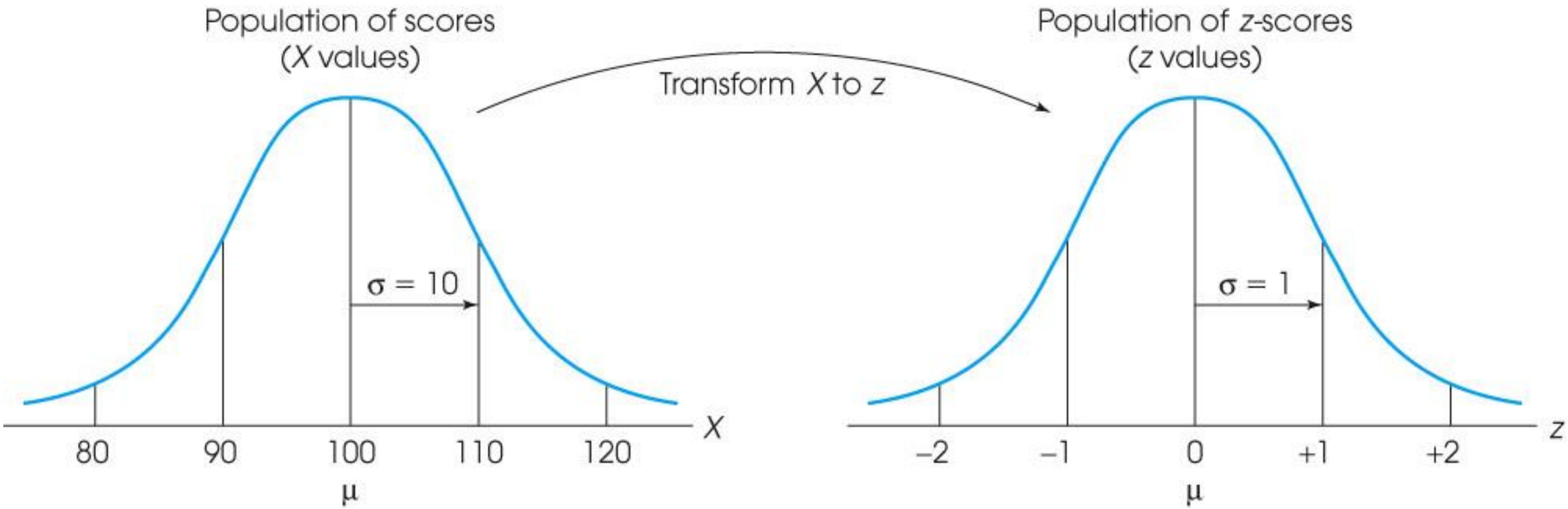
The Gaussian Distribution





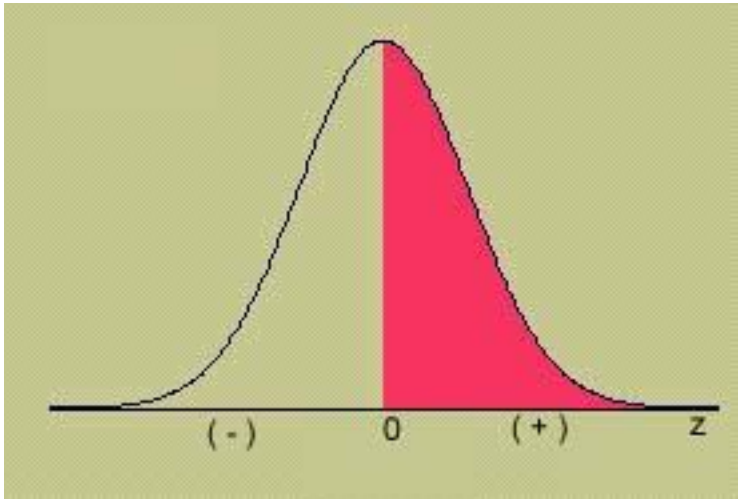



σ
X
μ
z
−2
−1
0
+1
+2


Population of scores
(X values)
Transform X to z
Population of z-scores
(z values)
σ = 10
80
90
100
110
120
μ
X
σ = 1
−2
−1
0
+1
+2
μ
z

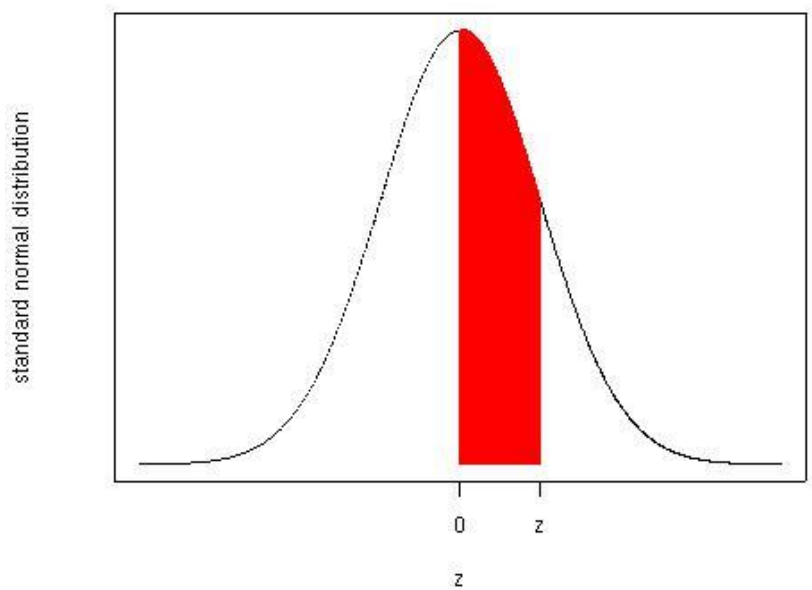
The Standard Gaussian (z) Distribution

- Mean = 0; Standard deviation = 1
- When $x = \mu$, $z = 0$
- Symmetric about $z = 0$
- Values of z to the left of center are negative
- Values of z to the right of center are positive
- Total area under the curve is 1.
- Areas on both sides of center equal .5



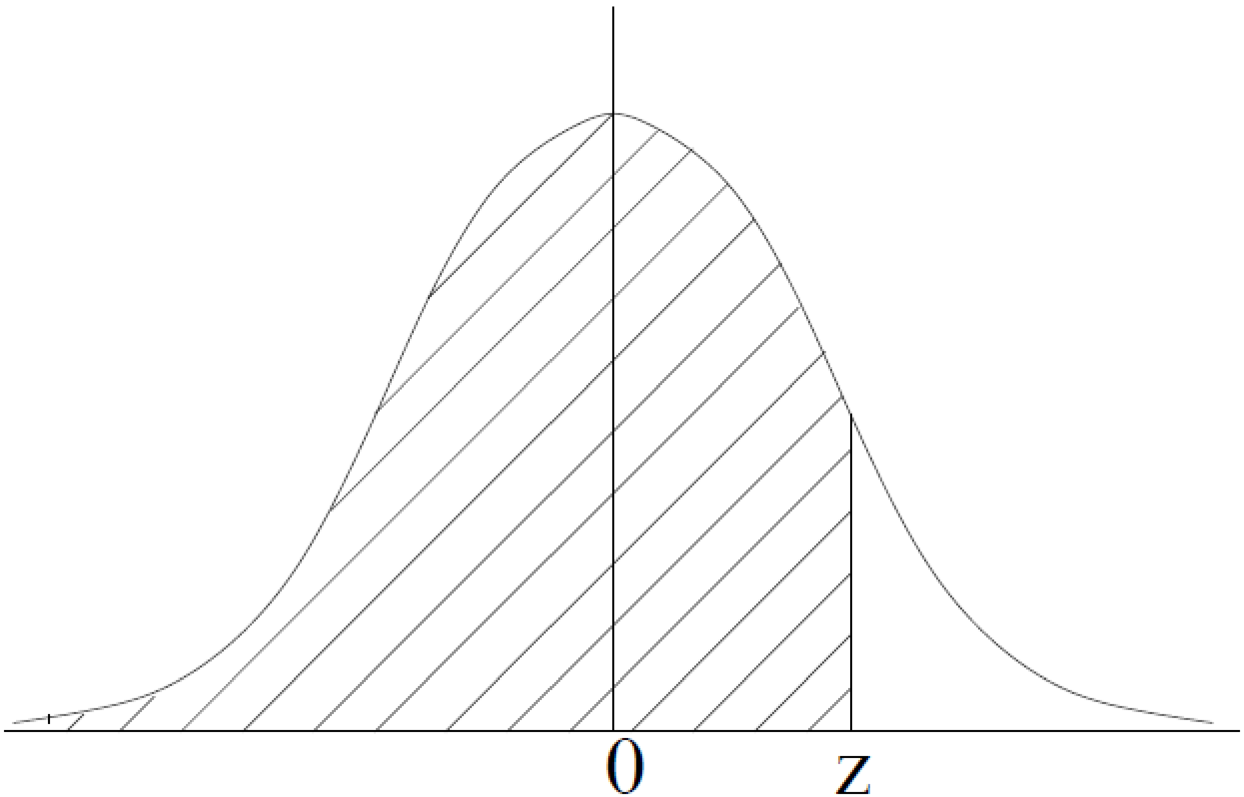
The Standard Gaussian (z) Distribution

The four digit probability in a particular row and column of Table3 gives the area under the standard normal curve between 0 and a positive value z. This is enough because the standard normal curve is symmetric.



The Gaussian Distribution

The tables allow us to read off probabilities of the form P(Z < z). Most of the table in the formula book has been reproduced in a Table . From this table we can identify that **P(Z < 1.0) = 0.8413** (this probability has been highlighted with a box)



The Gaussian Distribution

z	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0.1	0.5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0.2	0.5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0.3	0.6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0.4	0.6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0.5	0.6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
0.6	0.7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0.7	0.7580	7611	7642	7673	7704	7734	7764	7794	7823	7852
0.8	0.7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
0.9	0.8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1.0	$\boxed{0.8413}$	8438	8461	8485	8508	8531	8554	8577	8599	8621
1.1	0.8643	8665	8686	8708	8729	8749	8770	8790	8810	8830

N(0, 1) probability table

Table 3

Second decimal place in z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	* 1.0000									

* For values of $z \geq 3.90$, the areas are 1.0000 to four decimal places

Table 3

Second decimal place in z

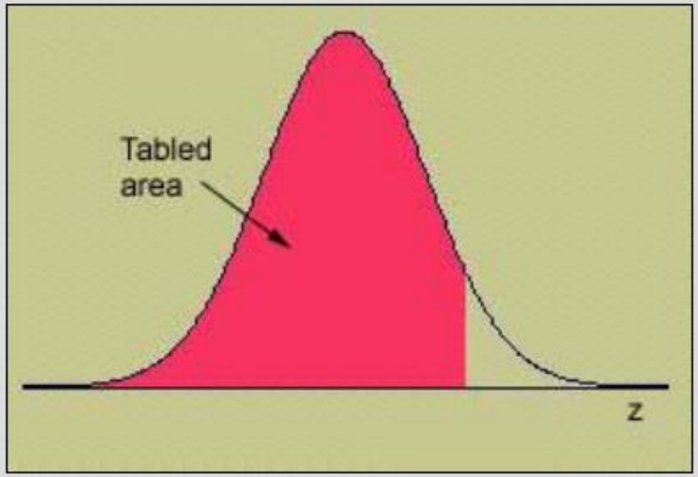
0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00	z
									* 0.0000	-3.9
0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	-3.8
0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	-3.7
0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002	-3.6
0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	-3.5
0.0002	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	-3.4
0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005	0.0005	0.0005	-3.3
0.0005	0.0005	0.0005	0.0006	0.0006	0.0006	0.0006	0.0006	0.0007	0.0007	-3.2
0.0007	0.0007	0.0008	0.0008	0.0008	0.0008	0.0009	0.0009	0.0009	0.0010	-3.1
0.0010	0.0010	0.0011	0.0011	0.0011	0.0012	0.0012	0.0013	0.0013	0.0013	-3.0
0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019	-2.9
0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026	-2.8
0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035	-2.7
0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047	-2.6
0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062	-2.5
0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082	-2.4
0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107	-2.3
0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139	-2.2
0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179	-2.1
0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228	-2.0
0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287	-1.9
0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359	-1.8
0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446	-1.7
0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548	-1.6
0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668	-1.5
0.0681	0.0694	0.0708	0.0721	0.0735	0.0749	0.0764	0.0778	0.0793	0.0808	-1.4
0.0823	0.0838	0.0853	0.0869	0.0885	0.0901	0.0918	0.0934	0.0951	0.0968	-1.3
0.0985	0.1003	0.1020	0.1038	0.1056	0.1075	0.1093	0.1112	0.1131	0.1151	-1.2
0.1170	0.1190	0.1210	0.1230	0.1251	0.1271	0.1292	0.1314	0.1335	0.1357	-1.1
0.1379	0.1401	0.1423	0.1446	0.1469	0.1492	0.1515	0.1539	0.1562	0.1587	-1.0
0.1611	0.1635	0.1660	0.1685	0.1711	0.1736	0.1762	0.1788	0.1814	0.1841	-0.9
0.1867	0.1894	0.1922	0.1949	0.1977	0.2005	0.2033	0.2061	0.2090	0.2119	-0.8
0.2148	0.2177	0.2206	0.2236	0.2266	0.2296	0.2327	0.2358	0.2389	0.2420	-0.7
0.2451	0.2483	0.2514	0.2546	0.2578	0.2611	0.2643	0.2676	0.2709	0.2743	-0.6
0.2776	0.2810	0.2843	0.2877	0.2912	0.2946	0.2981	0.3015	0.3050	0.3085	-0.5
0.3121	0.3156	0.3192	0.3228	0.3264	0.3300	0.3336	0.3372	0.3409	0.3446	-0.4
0.3483	0.3520	0.3557	0.3594	0.3632	0.3669	0.3707	0.3745	0.3783	0.3821	-0.3
0.3859	0.3897	0.3936	0.3974	0.4013	0.4052	0.4090	0.4129	0.4168	0.4207	-0.2
0.4247	0.4286	0.4325	0.4364	0.4404	0.4443	0.4483	0.4522	0.4562	0.4602	-0.1
0.4641	0.4681	0.4721	0.4761	0.4801	0.4840	0.4880	0.4920	0.4960	0.5000	-0.0

* For values of z ≤ -3.90, the areas are 0.0000 to four decimal places

The Standard Gaussian (z) Distribution

- The four digit probability in a particular row and column of Table gives the area under the standard normal curve between 0 and a positive value z. This is enough because the standard normal curve is symmetric.

z	.00	.01	.02	.03	.04	.05	.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051
0.9	0.8159	0.8186	0.8212	0.8328	0.8264	0.8289	0.8315
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278



Area for z = 1.36

Example

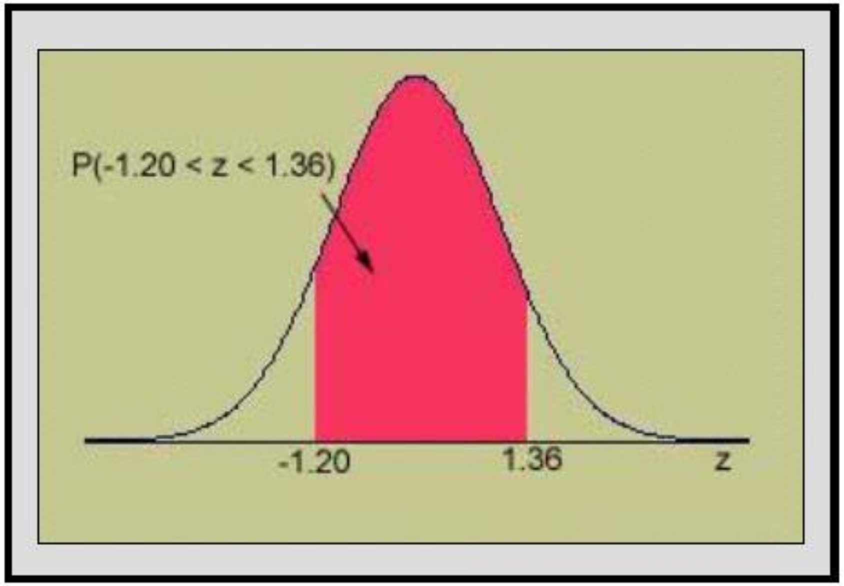
Use Table 3 to calculate these probabilities:

$P(z \leq 1.36) = .9131$

$P(z > 1.36)$

$= 1 - .9131 = .0869$

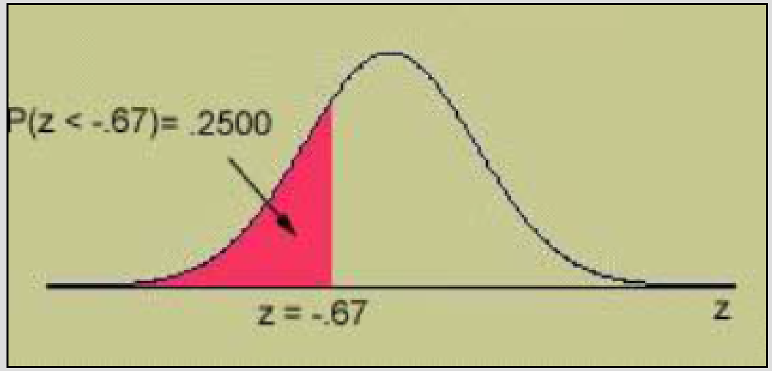
$P(-1.20 \leq z \leq 1.36) =$
$.9131 - .1151 = .7980$



Example

Find the value of *z* that has area .25 to its left.

1. Look for the four digit area closest to .2500 in Table 3.
2. What row and column does this value correspond to?
3. **3.** $z = -.67$



		.03	.04	.05	.06	.07	.08
0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635
0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894
0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177
	.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483
	.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810

4. What percentile does this value represent?

25th percentile, or 1st quartile (Q_1)

Example

$P(z<?) = .75$

$P(z<?)=P(z<0)+P(0<z<?)=.5+P(0<z<?)=.75$

$P(0<z<?)=.25$

$z = .67$

What percentile does this value represent?

75th percentile, or the third quartile.

The Gaussian Distribution

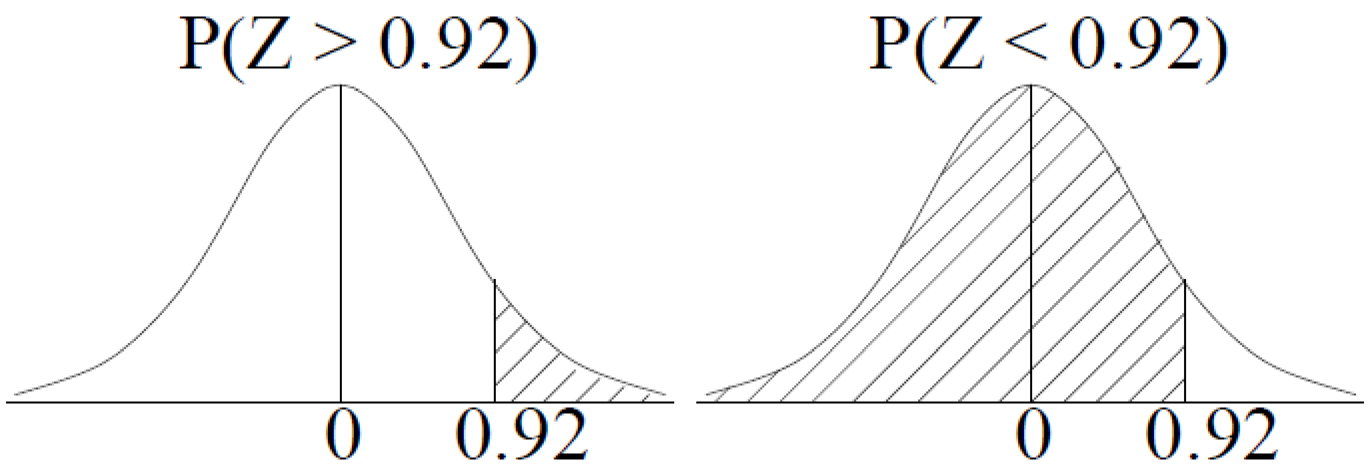
Consider P (X<0.567) ?

From tables we know that $P(X < 0.56) = 0.7123$ and P (X<0.57) $= 0.7157$
To calculate $P(X < 0.567)$ we *interpolate* between these two values

$$P(X < 0.567) = 0.3 \times 0.7123 + 0.7 \times 0.7157 = 0.71468$$

The Gaussian Distribution

P(Z > 0.92)

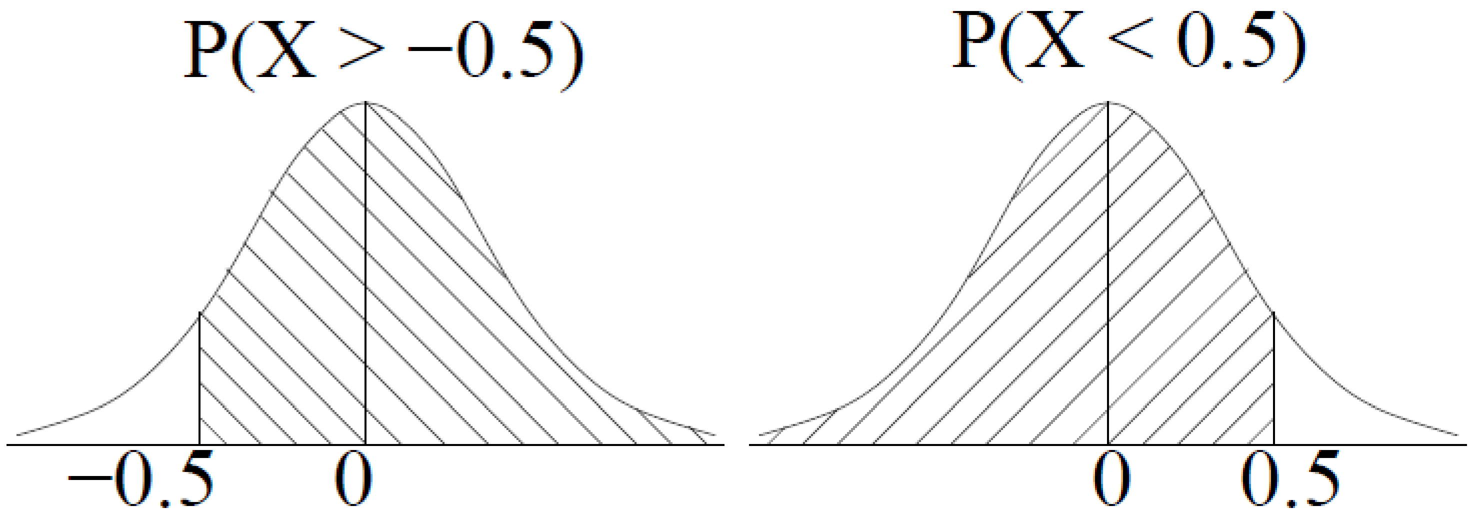


We know that P(Z > 0.92) = 1- P(Z < 0.92) and we can calculate P(Z <0.92) from the tables.

P(Z > 0.92) = 1 -0.8212 = 0.1788

The Gaussian Distribution

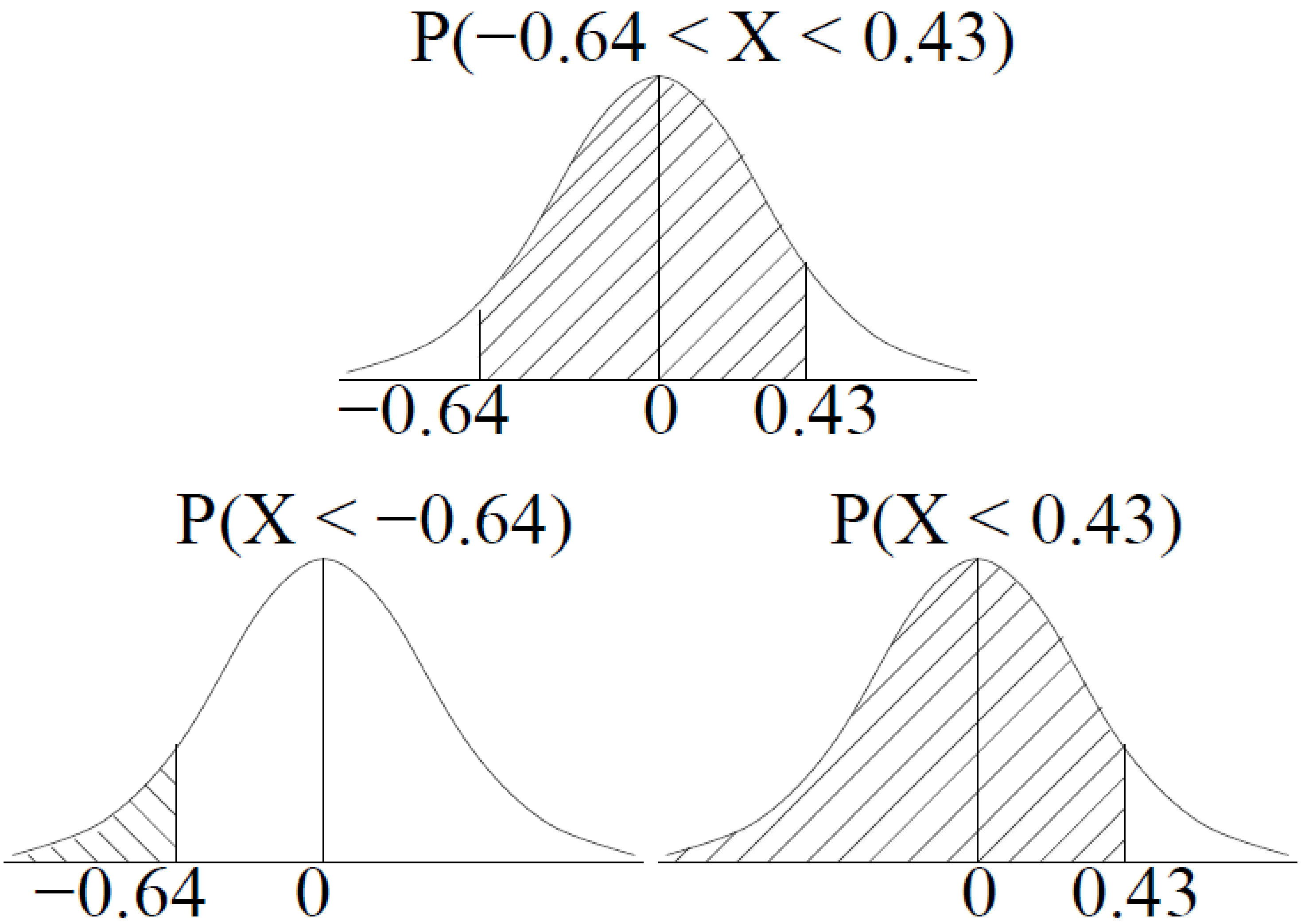
P(Z > -0.5)



The Normal distribution is symmetric so we know that P(Z > -0.5) = P(Z <0.5) = 0.6915

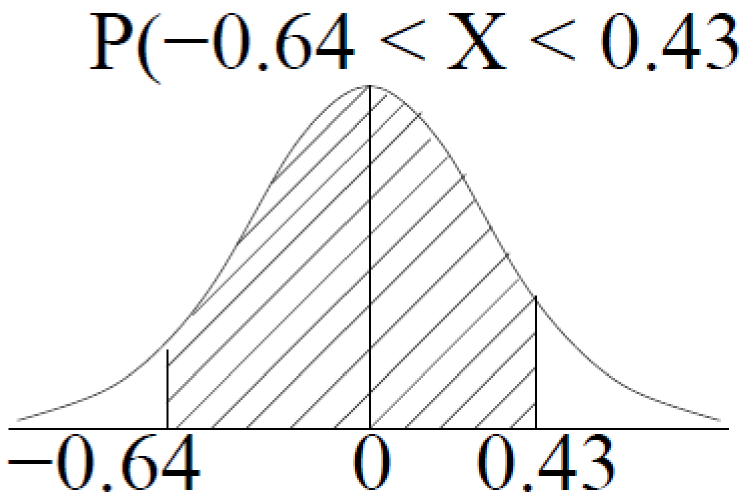
The Gaussian Distribution

P(-0.64<Z < 0.43)



The Gaussian Distribution

P(-0.64<Z < 0.43)

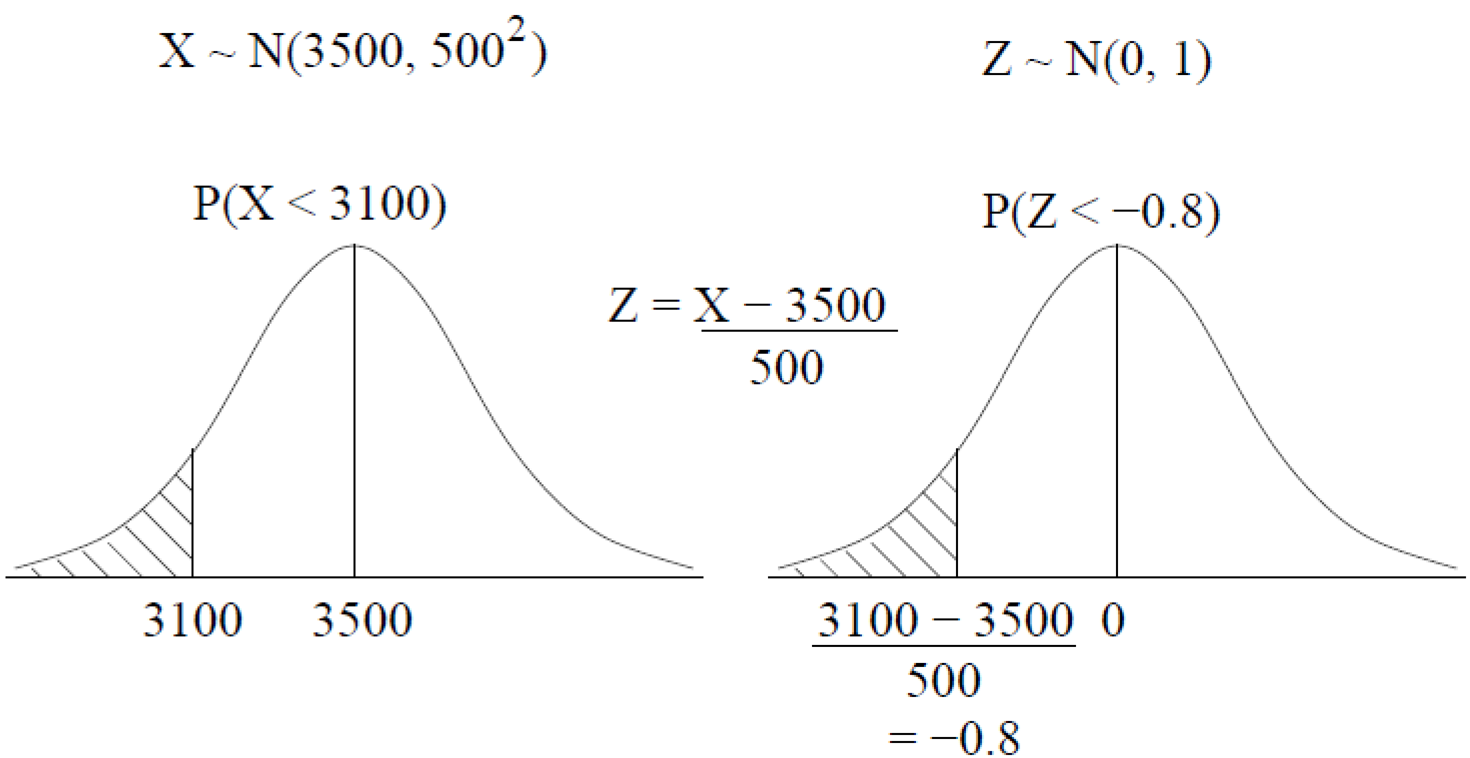


$$
\begin{aligned}
P(-0.64 < X < 0.43) &= P(X < 0.43) - P(X < -0.64) \\
&= 0.6664 - (1 - 0.7389) \\
&= 0.4053
\end{aligned}
$$

Example

Suppose we know that the birth weight of babies is Normally distributed with mean 3500g and standard deviation 500g. What is the probability that a baby is born that weighs less than 3100g?

That is $X \sim N(3500, 500^2)$ and we want to calculate $P(X < 3100)$?

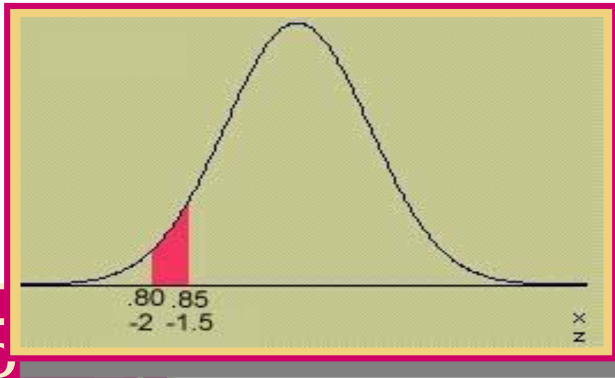


Example

$$
\begin{aligned}
P(X < 3100) = P\left(\frac{X-3500}{500} < \frac{3100-3500}{500}\right) &= P(Z < -0.8) \qquad \text{where } Z \sim \mathbf{N}(0,1) \\
&= 1 - P(Z < 0.8) \\
&= 1 - 0.7881 \\
&= 0.2119
\end{aligned}
$$

Example

- The weights of packages of ground beef are normally distributed with mean 1 kg and standard deviation .10. What is the probability that a randomly selected package weighs between 0.80 and 0.85 kg?



$$P(.80 < x < .85) = P(\frac{.80 - 1}{.1} < z < \frac{.85}{.1})$$

$$= P(-2 < z < -1.5) = P(1.5 < z < 2)$$

$$= .9772 - .9332 = .0440$$

Example

- What is the weight of a package such that only 5% of all packages exceed this weig

$$P(x > ?) = .05$$

$$P(z > \frac{?-1}{.1}) = .05$$

$$P(z < \frac{?-1}{.1}) = 1 - .05 = .95$$

$$\text{From Table 3, } \frac{?-1}{.1} = 1.645$$

$$? = 1.645(.1) + 1 = 1.16$$

Second decimal place in z

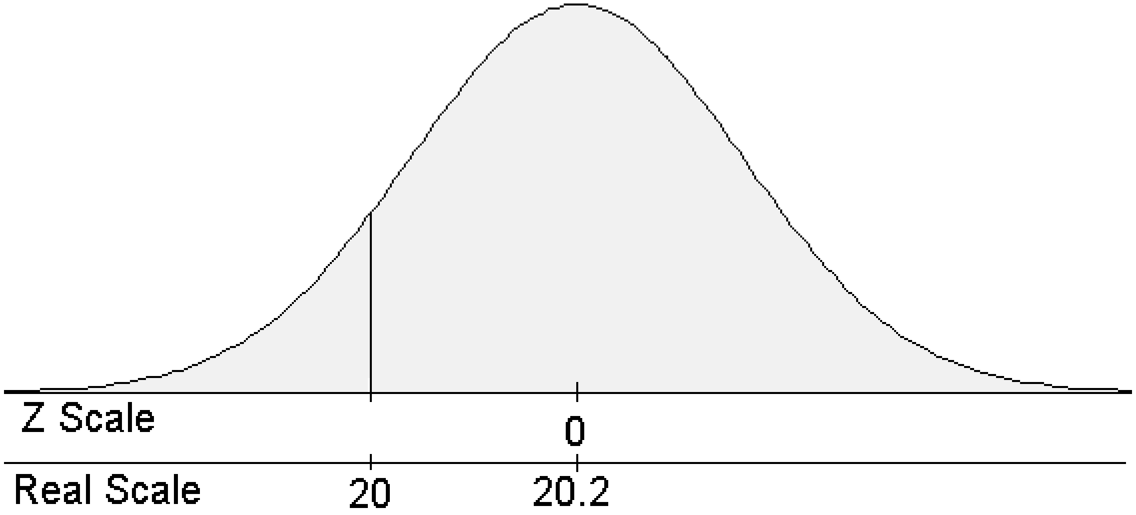
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	* 1.0000									

* For values of $z \geq 3.90$, the areas are 1.0000 to four decimal places

Example

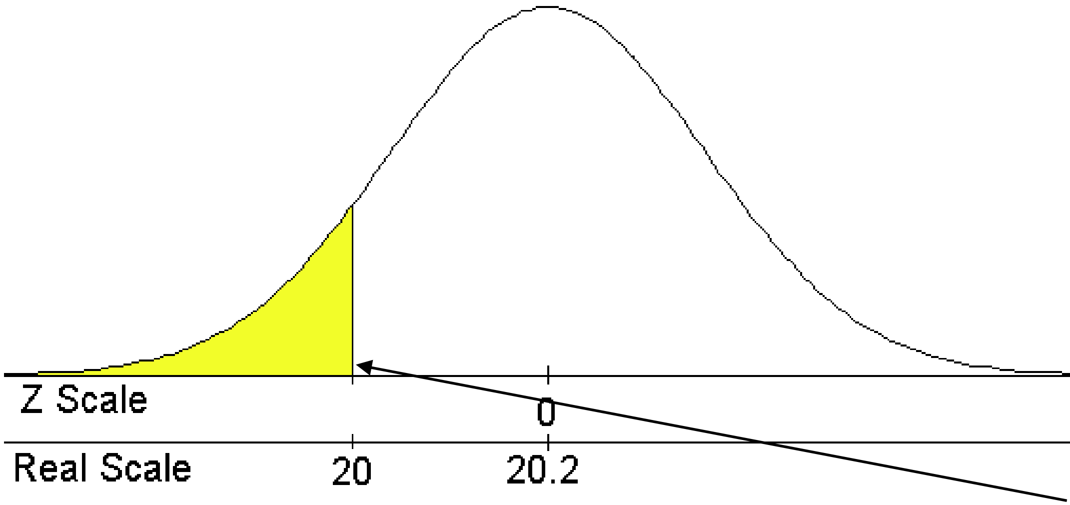
A Company produces "20 gr" chocolate.

Suppose the companies "20 gr" chocolate follow a normally distribution with a mean μ=20.2 gr with a standard deviation σ=0.125 gr .



Example

- What proportion of the chocolate less than 20 gr?

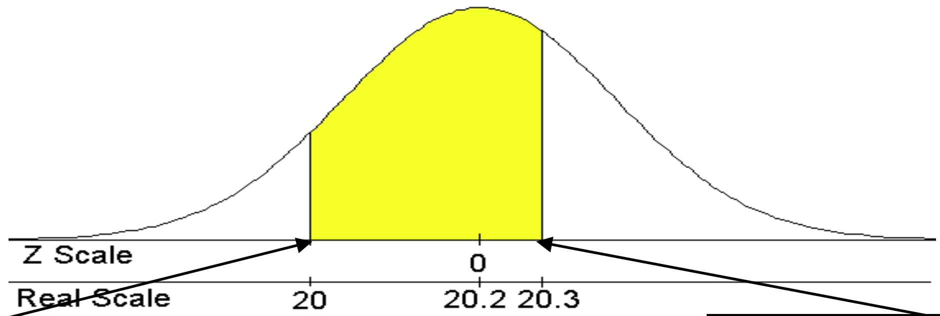


$$z = \frac{x - \mu}{\sigma} = \frac{20 - 20.2}{0.125} = -1.60$$

P(*z*<-1.60) = 1-P(z>1.60) = 1-.9452 = .0548. The proportion of the chocolate less than 20 gr is .0548

Example

- What proportion of the chocolate between 20 and 20.3 gr?



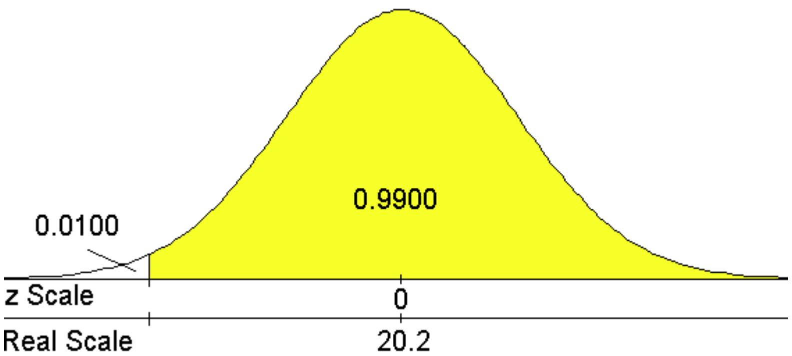
$$Z=\frac{20-20.2}{0.125}=-1.60$$

$$Z=\frac{20.3-20.2}{0.125}=0.80$$

P(-1.60<z<.80) = P(z<.80)-P(z<-1.60)=.7871-.0548 =.7333

Example

99% of chocolate will contain more than what amount chocolate?



$$.99 = P(x > ?) = P(z > \frac{? - 20.2}{.125})$$

$$\text{From Table 3, } \frac{20.2 - ?}{.125} = 2.33$$

$$? = 20.2 - 2.33(.125) = 19.91$$

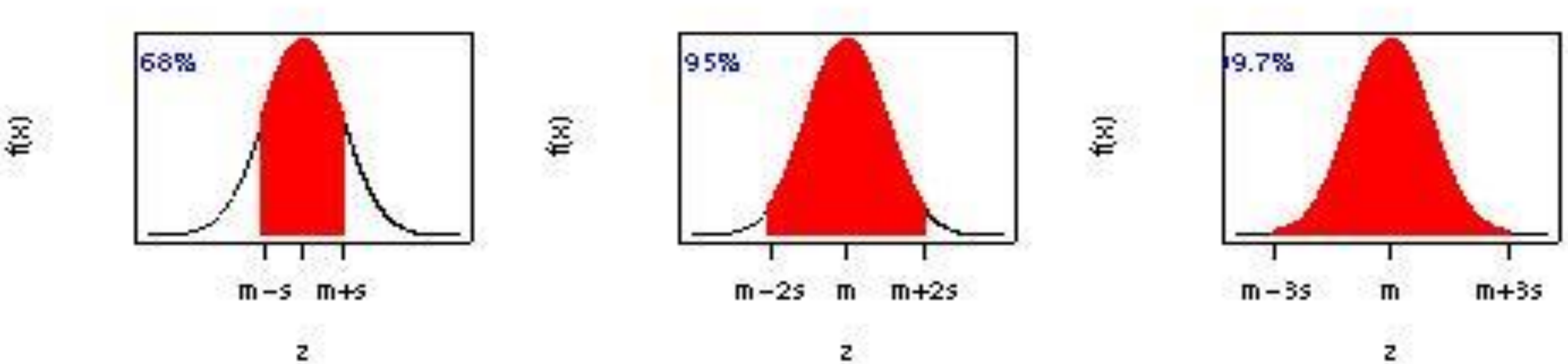
Second decimal place in z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	* 1.0000									

* For values of $z \geq 3.90$, the areas are 1.0000 to four decimal places

How Probabilities Are Distributed

- The interval $\mu \pm \sigma$ contains approximately 68% of the measurements.
- The interval $\mu \pm 2\sigma$ contains approximately 95% of the measurements.
- The interval $\mu \pm 3\sigma$ contains approximately 99.7% of the measurements.



Linear combinations of Normal random variables

Suppose two rats A and B have been trained to navigate a large maze. The time it takes rat A is normally distributed with mean 80 seconds and standard deviation 10 seconds. The time it takes rat B is normally distributed with mean 78 seconds and standard deviation 13 seconds. On any given day what is the probability that rat A runs the maze faster than rat B?

X = Time of run for rat A X $\sim N(80, 10^2)$
Y = Time of run for rat B Y $\sim N(78, 13^2)$

Let D = X - Y be the difference in times of rats A and B

If rat A is faster than rat B then $D < 0$ so we want $P(D < 0)$?

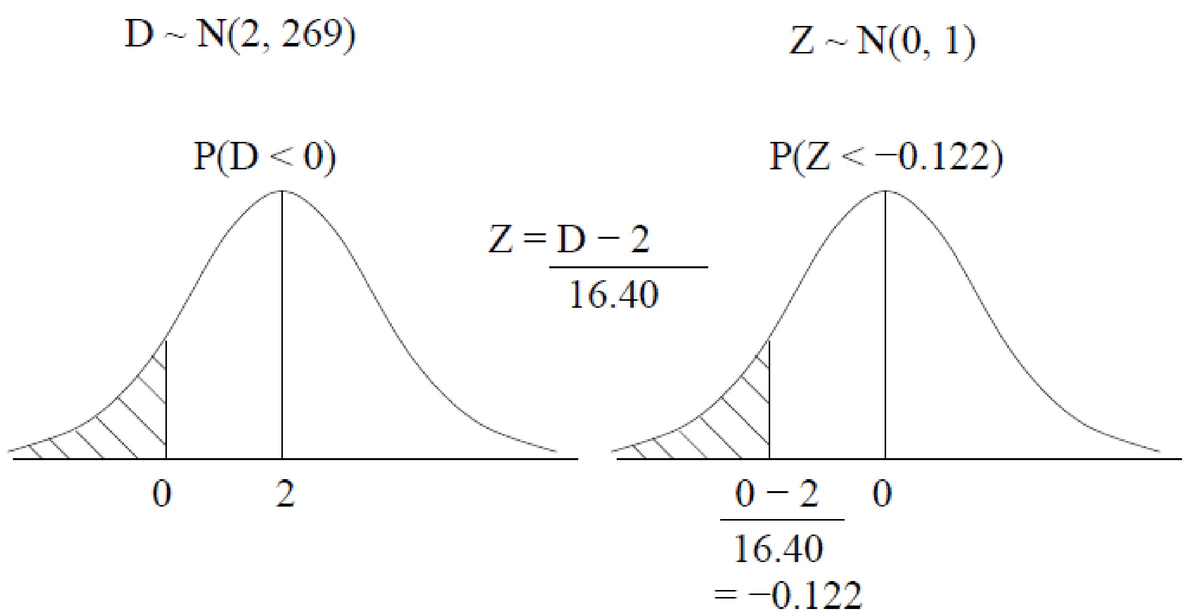
Linear combinations of Normal random variables

If X and Y are two independent normal variable such that

$X \sim N(\mu_1,\sigma_1^2)$ and $X \sim N(\mu_2,\sigma_2^2)$

then $X - Y \sim N(\mu_1-\mu_2,\sigma_1^2+\sigma_2^2)$

In this example
$D = X - Y \sim N(80 - 78, 10^2 + 13^2) = N(2, 269)$



Linear combinations of Normal random variables

$$P(D<0) = P\left(\frac{D-2}{\sqrt{269}} < \frac{0-2}{\sqrt{269}}\right) = P(Z < -0.122) \qquad \text{where } Z \sim \mathbf{N}(0,1)$$

$$= 1 - (0.8 \times 0.5478 + 0.2 \times 0.5517)$$

$$= 0.45142$$

If X and Y are two independent normal variable such that

$$X \sim \mathbf{N}(\mu_1, \sigma_1^2) \text{ and } Y \sim \mathbf{N}(\mu_2, \sigma_2^2)$$

then

$$\begin{aligned} X+Y &\sim \mathbf{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2) \\ aX &\sim \mathbf{N}(a\mu_1, a^2\sigma_1^2) \\ aX+bY &\sim \mathbf{N}(a\mu_1+b\mu_2, a^2\sigma_1^2+b^2\sigma_2^2) \end{aligned}$$

Linear combinations of Normal random variables

Suppose two rats A and B have been trained to navigate a large maze. The time it takes rat A is normally distributed with mean 80 seconds and standard deviation 10 seconds. The time it takes rat B is normally distributed with mean 78 seconds and standard deviation 13 seconds. On any given day what is the probability that the average time the rats take to run the maze is greater than 82 seconds?

X = Time of run for rat A $\quad X \sim N(80, 10^2)$

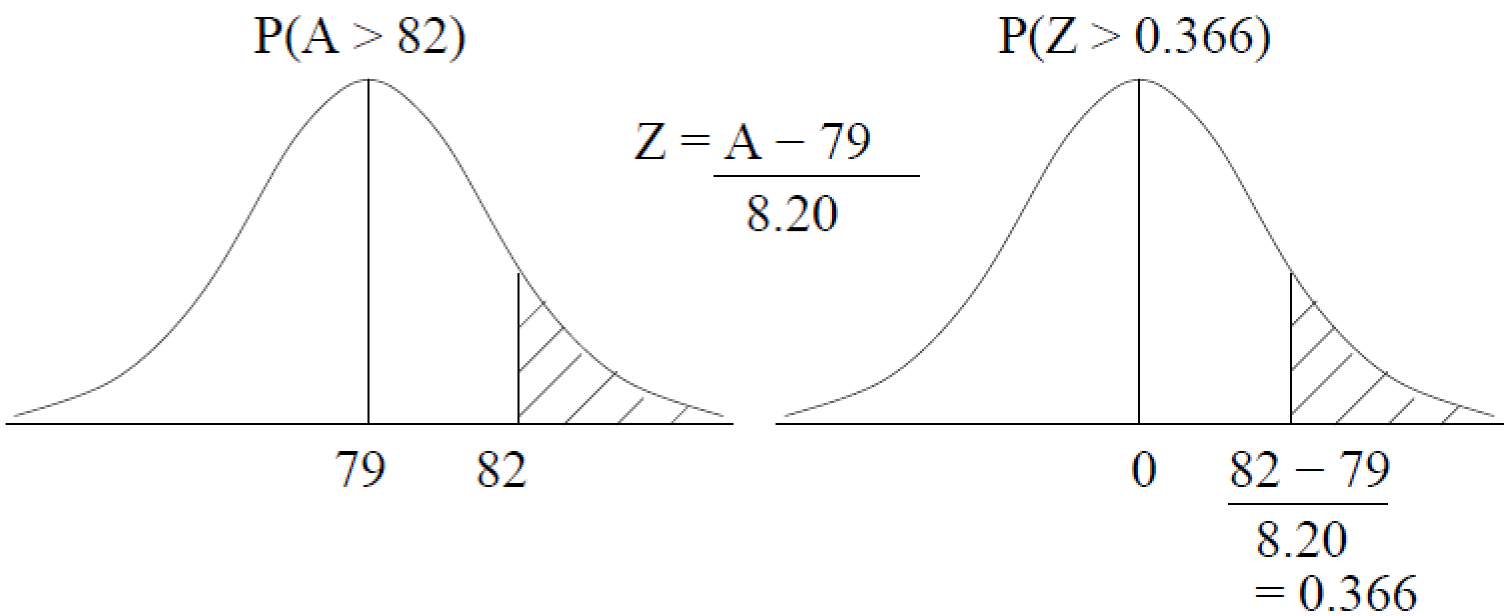
Y = Time of run for rat B $\quad Y \sim N(78, 13^2)$

Let A = (X+Y)/2 = 1/2X + 1/2Y be the average time of rats A and B

Then $A \sim N(½\ 80 + ½\ 78, (1/2)^2 10^2 + (1/2)^2 13^2 = N(79, 67.25)$

We want $P(A > 82)$

Linear combinations of Normal random variables

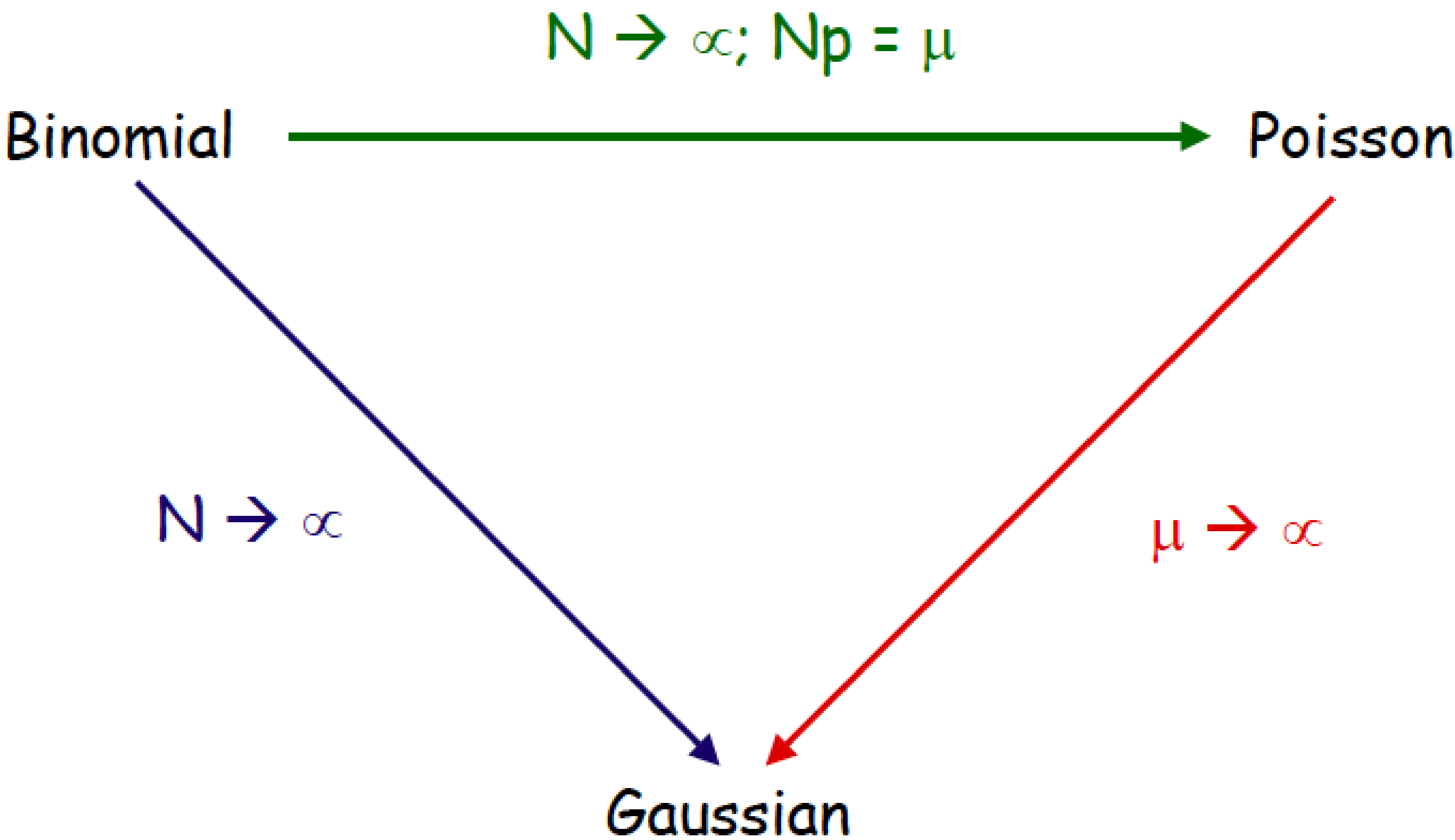


$$
\begin{aligned}
P(A > 82) = P\left(\frac{A-79}{\sqrt{67.25}} < \frac{82-79}{\sqrt{67.25}}\right) &= P(Z > 0.366) \qquad \text{where } Z \sim \mathrm{N}(0,1) \\
&= 1 - (0.4 \times 0.6406 + 0.6 \times 0.6443) \\
&= 0.35718
\end{aligned}
$$

Central Limit Theorem

- It gives conditions under which the mean of a sufficiently large number of independent random variables, each with finite mean and variance will be approximately normally distributed
- It provides a partial explanation for prevalence of the Gaussian distribution in the real world.
- Justifies the approximation of the large – sample statistics to the Gaussian distribution in controlled experiments.

Central Limit Theorem



Central Limit Theorem

1. The distribution of sample $\overline{x}$ will, as the sample size increases, approach a normal distribution.

2. The mean of the sample means is the population mean μ.

3. The standard deviation of all sample means is $\sigma/\sqrt{n}$

Central Limit Theorem

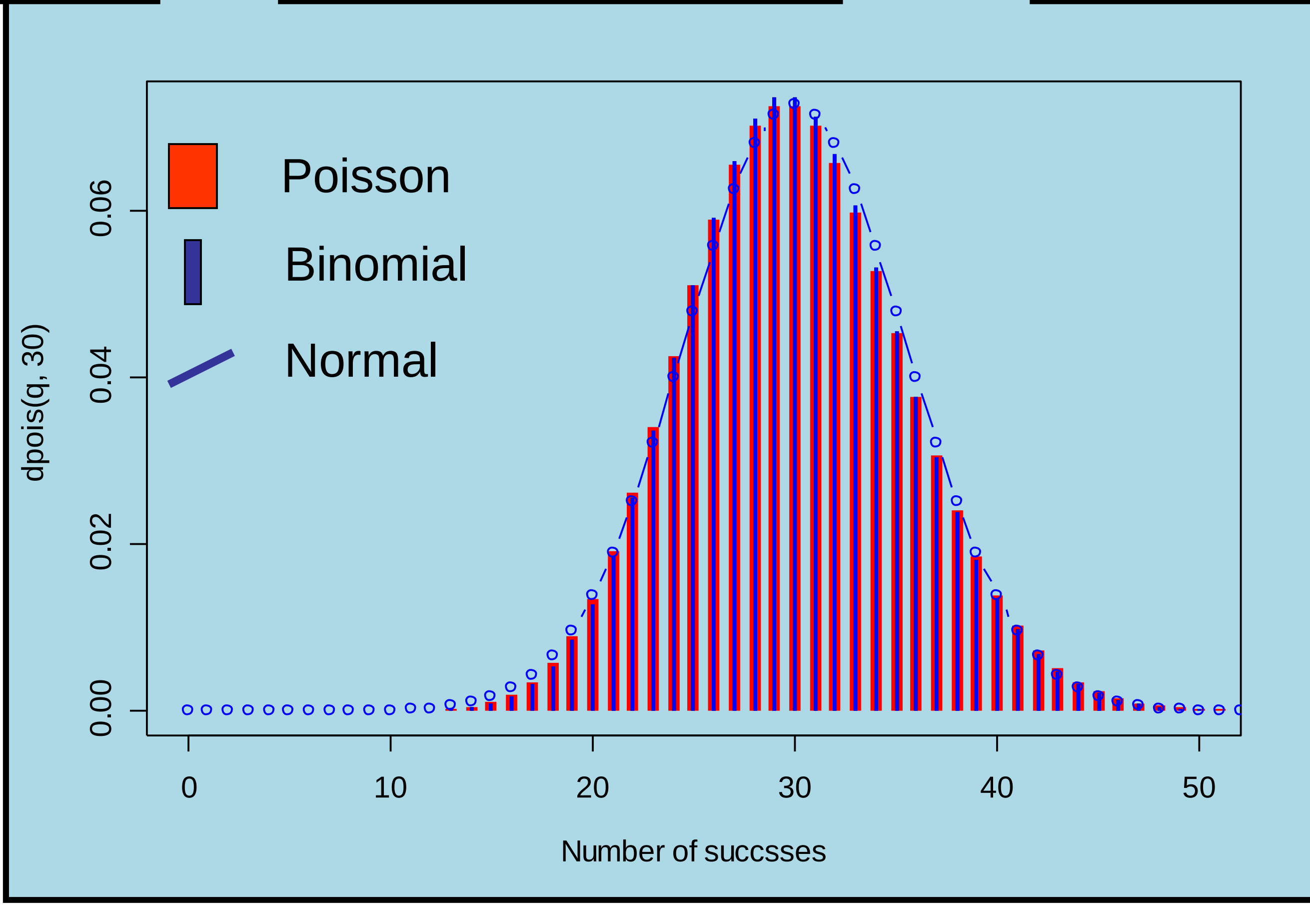
1. For samples of size n larger than 30, the distribution of the sample means can be approximated reasonably well by a normal distribution. The approximation gets better as the sample size n becomes larger.
2. If the original population is itself normally distributed, then the sample means will be normally distributed for any sample size n (not just the values of n larger than 30).

Binomial Poisson and Normal ?

Poisson(30)

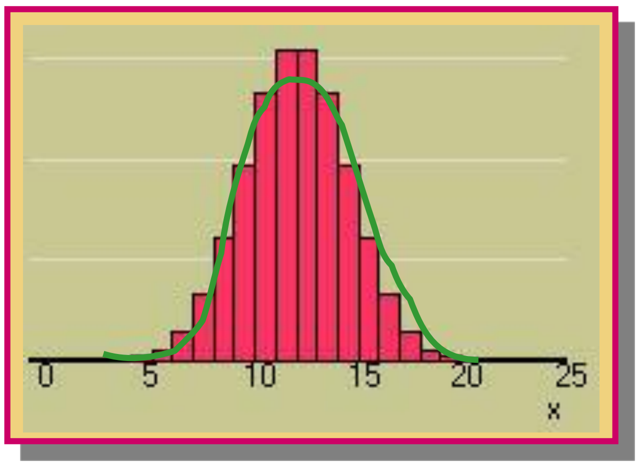
Binomial(1000,.03)

N(30,sqrt(30))



The Normal Approximation to the Binomial

- We can calculate binomial probabilities using
  - The binomial formula
  - The cumulative binomial tables
- When n is large, and p is not too close to zero or one, areas under the normal curve with mean np and variance npq can be used to approximate binomial probabilities.



Approximating the Binomial

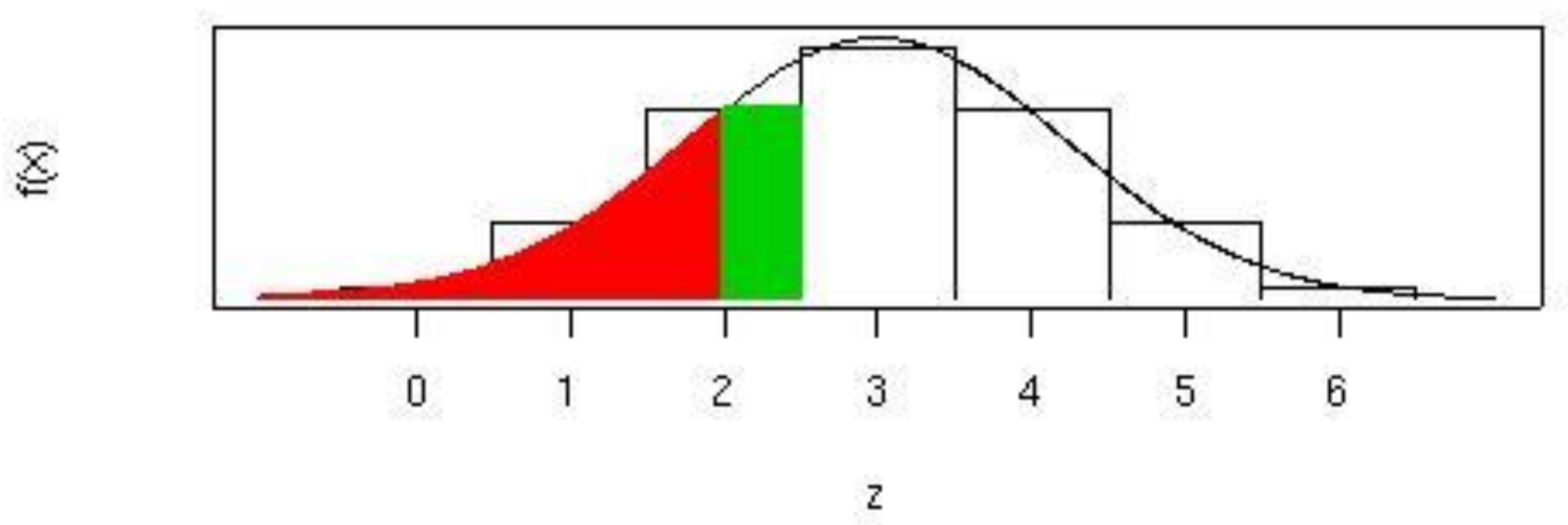
- ✓Make sure to include the entire rectangle for the values of x in the interval of interest. This is called the continuity correction.
- ✓Standardize the values of x using

$$\boxed{z = \frac{x-\mu}{\sigma}, \mu = np, \sigma = \sqrt{npq}}$$

- ✓Make sure that np and nq are both greater than 5 to avoid inaccurate approximations! Or
- ✓n is large and $\mu \pm 2\sigma$ falls between 0 and n

Correction for Continuity

Add or subtract .5 to include the entire rectangle. For illustration, suppose x is a Binomial random variable with n=6, p=.5. We want to compute P(x≤ 2). Using 2 directly will miss the green area. P(x≤ 2)=P(x≤ 2.5) and use 2.5.



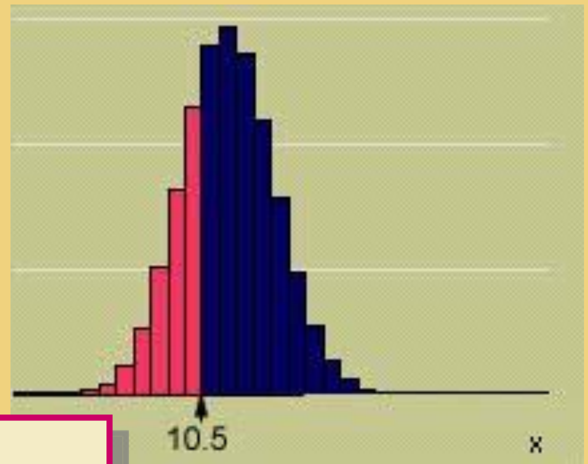
Example

Suppose x is a binomial random variable with $n = 30$ and $p = .4$. Using the normal approximation to find $P(x \le 10)$.

$n = 30$ $p = .4$ $q = .6$

$np = 12$ $nq = 18$

The normal approximation is ok!

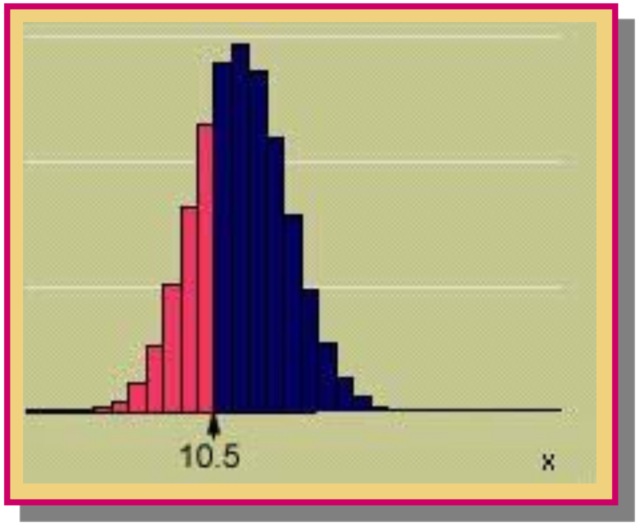


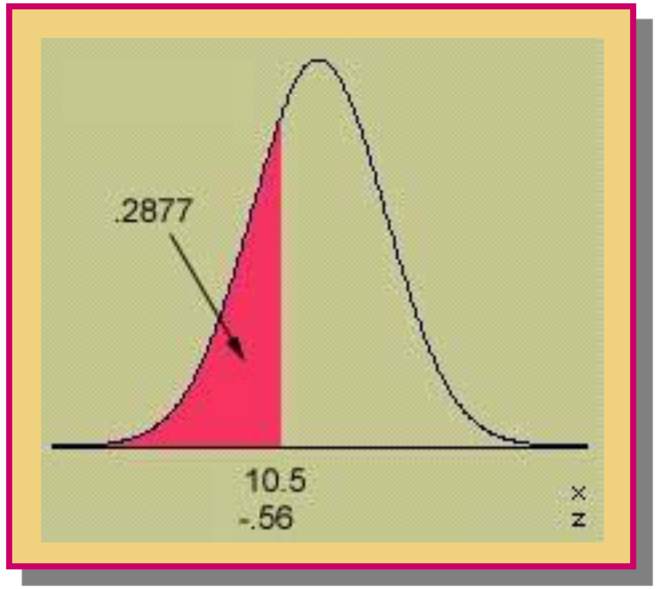
Calculate

$$\mu = np = 30(.4) = 12$$

$$\sigma = \sqrt{npq} = \sqrt{30(.4)(.6)} = 2.683$$

Example





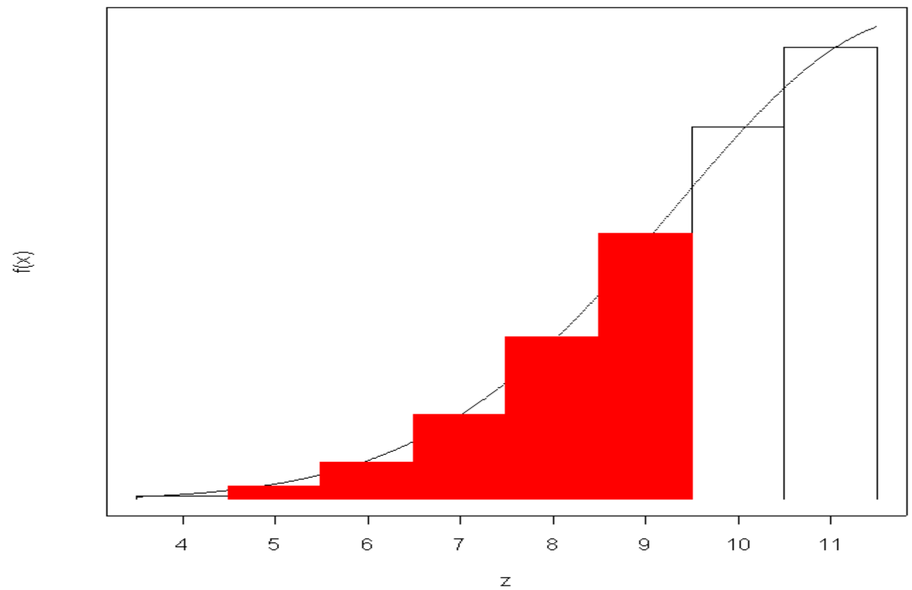
$$P(x \le 10) \approx P(z \le \frac{10.5-12}{2.683})$$

$$= P(z \le -.56) = .2877$$

Example

$$P(x < 10) = P(x < 9.5)$$

$$P(x \geq 5) = P(x \geq 4.5)$$



$$P(x > 5) = P(x > 5.5)$$

$$P(5 < x < 10) = P(5.5 < x < 9.5)$$

$$P(5 \leq x < 10) = P(4.5 < x < 9.5)$$

Example

A production line produces AA batteries with a reliability rate of 95%. A sample of n = 200 batteries is selected. Find the probability that at least 195 of the batteries work.

Success = working battery $n = 200$

$p = .95$ $np = 190$ $nq = 10$

The normal approximation is ok!

$$P(x \geq 195) \approx P(z \geq \frac{194.5 - 190}{\sqrt{200(.95)(.05)}})$$

$$= P(z \geq 1.46) = 1 - .9278 = .0722$$

Useful NumPy Methods for Statistics

Let x be a numpy array of values.....

x.mean() returns the mean of the values contained in array x.

x.var() returns the variance of the values about their mean.

x.std() returns the standard deviation of the values about their mean.

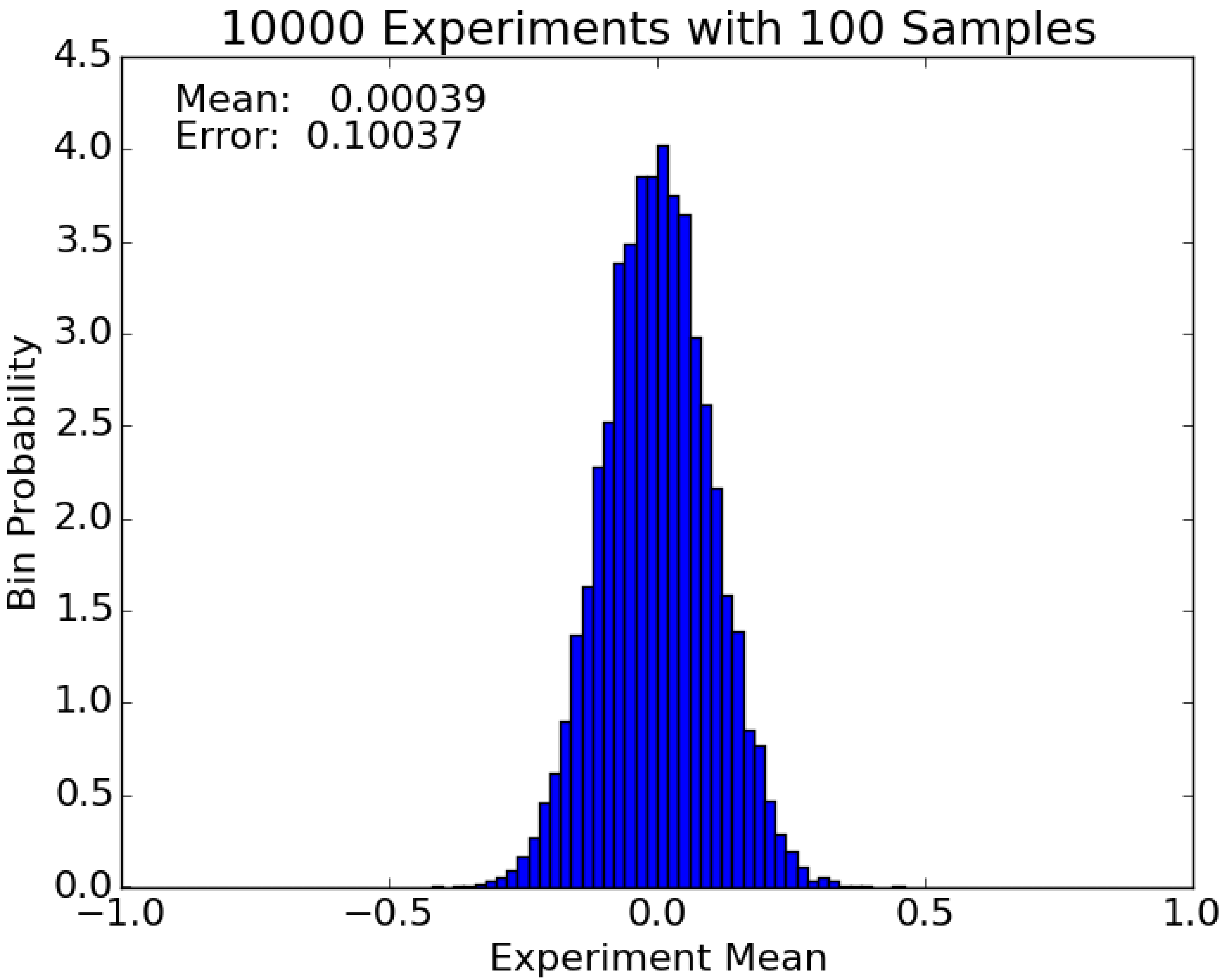
Useful NumPy Methods for Statistics

```
import numpy as np
import math
import matplotlib.pyplot as pl
from numpy.random import RandomState

r = RandomState()
Nexp = 10000 # number of experimentsi
Nsam = 100 # number of samples per experiment

# initialize array to hold experiment results
# experiment_results = np.zeros(nexp)
# now conduct nexp experiments with nsam samples
for experiment in range(nexp):
   x = r.randn(nsam)
   experiment_results[experiment] = x.mean()

fr = experiment_results.mean()
fe = experiment_results.std()
ee = 1./math.sqrt(nsam) # expected error on the mean

print ('Final results')
print ('mean ={0:8.5f} Expected = 0'.format(fr))
print ('error={0:8.5f} Expected ={1:8.5f}'.format(fe,ee))
```

For nexp experiments we take nsam samples from a normal distribution and compute the mean.

Find mean value and standard deviation of the experiment results.

Useful NumPy Methods for Statistics



The χ^2 distribution

Suppose that you generate N random numbers from a normal distribution with μ=0, σ=1: Z_1 ... Z_N.

Let X be the sum of the squared variables:

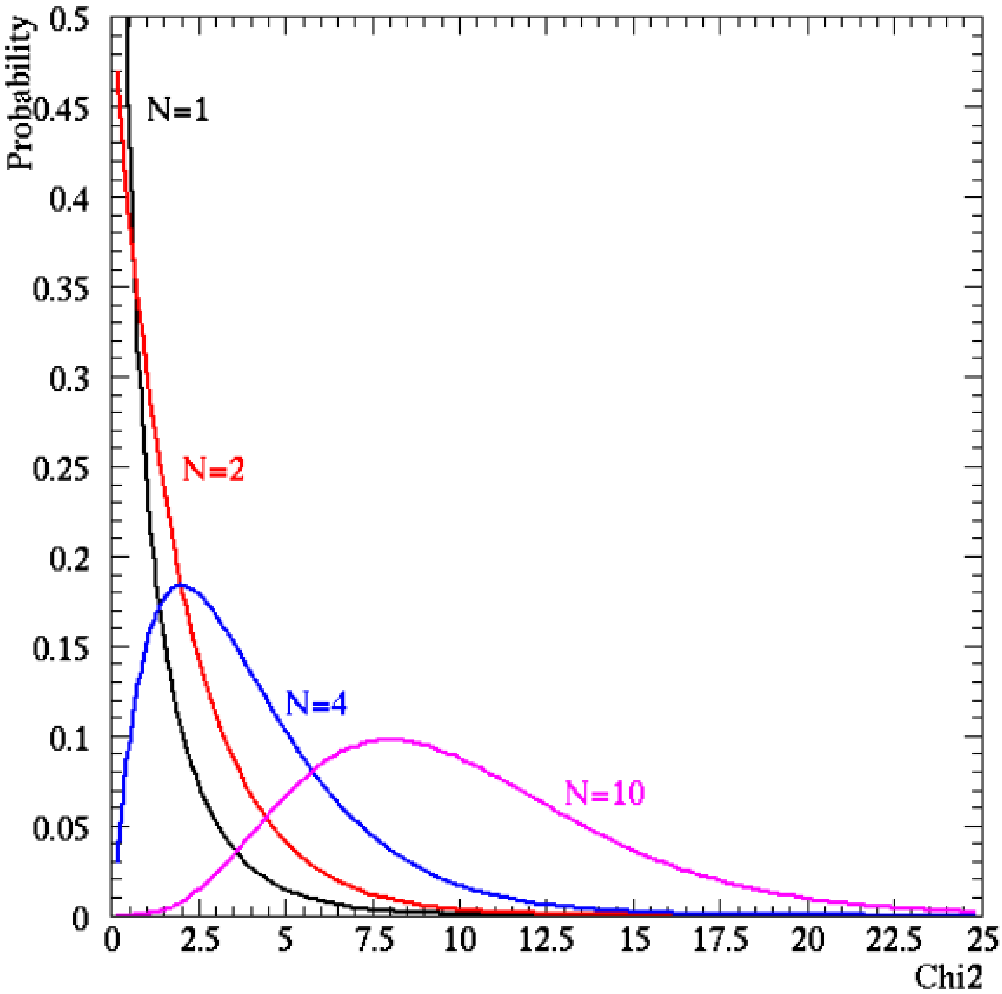
$$X = \sum_{i=1}^{N} Z_i^2$$

The variable X follows a χ^2 distribution with N degrees of freedom:

$$P(\chi^2 | N) = \frac{2^{-N/2}}{\Gamma(N/2)} (\chi^2)^{(N-2)/2} e^{-\chi^2/2}$$

Recall that $\Gamma(N) = (N-1)!$ if N is an integer.

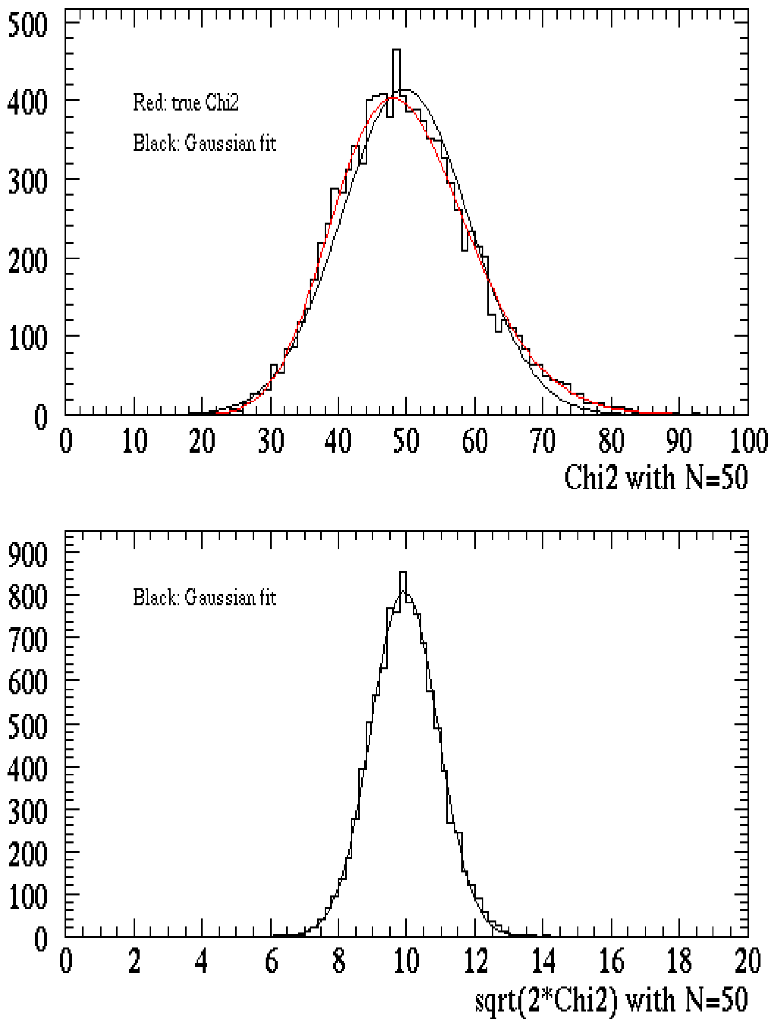
Properties of χ^2 distribution



A χ^2 distribution has mean=N, but variance=2N.

This makes it relatively easy to estimate probabilities on the tail of a χ^2 distribution.

Properties of χ^2 distribution



Since χ^2 is a sum of N independent and identical random variables, it is true that it tends to be Gaussian in the limit of large N (central limit theorem) ...

But the quantity sqrt($2\chi^2$) is actually much more Gaussian, as the plots to the left show! It has mean of sqrt(2N-1) and unit variance.

Calculating a χ^2 tail probability

You're sitting in a talk, and someone shows a dubious looking fit, and claims that the χ^2 for the fit is 70 for 50 degrees of freedom. Can you work out in your head how likely it is to get that large of a χ^2 by chance?

More accurate estimate: $\mathrm{sqrt}(2\chi^2) = \mathrm{sqrt}(140)=11.83$. Mean should be $\mathrm{sqrt}(2N-1)=9.95$. This is really more like a 1.88σ fluctuation.

Uses of the χ^2 distribution

The dominant use of the χ^2 statistics is for least squares fitting.

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - f(x_i|\vec{\alpha})}{\sigma_i} \right)^2$$

The "best fit" values of the parameters α are those that minimize the χ^2.

If there are m free parameters, and the deviation of the measured points from the model follows Gaussian distributions, then this statistic should be a χ^2 with N-m degrees of freedom. More on this later.

χ^2 is also used to test the goodness of the fit

An exponential distribution

Consider for example the distribution of measured lifetimes for a decaying particle:

$$P(t) = \frac{1}{\tau} e^{-t/\tau} \qquad (\text{both } t, \tau > 0)$$

mean: $\langle t \rangle = \tau$ RMS: $\sigma = \tau$