§ 24 PROBABILITY REDUCES DIFFERENCES TO DEVIATIONS

BINOMIAL CHANCE DETERMINES HOW OFTEN YOU CAN EXPECT A YES OR NO

With the throw of a coin, cross ⊕ or coin ○ both have the same (binomial) chance ('probability') p = ½ (50%).

After throwing twice (k=2) there are 4 possibilities (N=4): ⊕⊕, ⊕○, ○⊕ or ○○ with each ¼ (25%) chance p.

After three times (k=3) you have made one of N=8 combinations: ⊕⊕⊕, ⊕⊕○, ⊕○⊕, ○⊕⊕, ⊕○○, ○⊕○, ○○⊕ or ○○○.

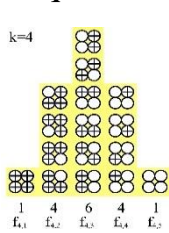With four throws (k=4) there are $N=16=2^4$ possibilities. In short, there are $N=2^k$ possibilities for k throws.

In *Fig. 85* you see, throwing 4 times (or 4 coins at the same time) you rarely throw 4×⊕ or 0×⊕ and most often 6×⊕.

There is only 1 possibility to throw 4×⊕, 4 to throw 3×⊕, and so on. In short, the 'frequency' f = {1, 4, 6, 4, 1}.

Their 'probability' p is divided by the number of possibilities N, respectively p = f/N = {1/16, 4/16, 6/16, 4/16 and 1/16}. The biggest probability p is therefore f/N=6/16=38% to throw two times ⊕ and two times ○.

---

a https://nl.wikipedia.org/wiki/Tensor; http://www.mate.tue.nl/~peters/4K400/VectTensColMat.pdf; http://www.ita.uni-heidelberg.de/~dullemond/lectures/tensor/tensor.pdf; https://www.ese.wustl.edu/~nehorai/Porat_A_Gentle_Introduction_to_Tensors_2014.pdf; https://web2.ph.utexas.edu/~jcfeng/notes/Tensors_Poor_Man.pdf

With more throws or coins, the writing out of all combinations N and sorting on frequencies f becomes laborious.
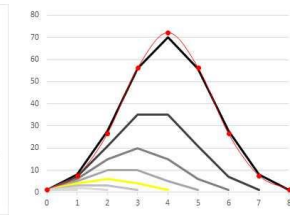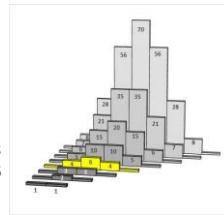

| | Fig. 85 frequencies | Fig. 86 Triangle of Pascal | Fig. 87 The frequencies … | Fig. 88 … approaching a clock shape |

Pascal devised a triangle of numbers to predict the frequencies of each subsequent throw k, by adding two higher numbers each time (**Fig. 86**: for example $10 + 10 = 20$). The sum of each row ($\sum f_k$) is $N = 2^k$.

In row f_8 (k = 8 throws) is it already N = 256 possibilities in 9 frequencies. **Fig. 87** shows all frequencies in a bar graph.

Fig. 88 also suggests intermediate values, but the 'binomial distribution' itself is 'discrete', not 'continuous'.

Galton came up with a plate with rows of nails that resembles the Pascal triangle (**Fig. 89**). Bullets that fall in the middle of the top nail can go in two directions. They fall in the next row on a nail where they also can go two ways and so on. After k = 14 rows, a bullet has covered one of the $N = 2^{14} = 16384$ possible paths.

It is then captured in one of the k+1=15 frequency columns (including column 0: f_0, f_1, f_2 ... f_{14}).

There is only one path to end up in one of the two outer ones. The probability ending up there is 1 in 16384 (0.006%). The other opportunities are calculated with Excel in **Fig. 90**. Newton came up with the formula (**Fig. 91**).

The symbol '!' it is called 'faculty': 0! and 1! are both 1, but $2!=2*1$, $3!=3*2*1$, $4!=4*3*2*1$, and so on.

The probability of ending up in the middle column (f_7) is thus the frequency 7 out of 14 (= 3432) from 16384 (21%).

$$\binom{k}{k_x} = \frac{k!}{k_x!\,(k - k_x)!}$$
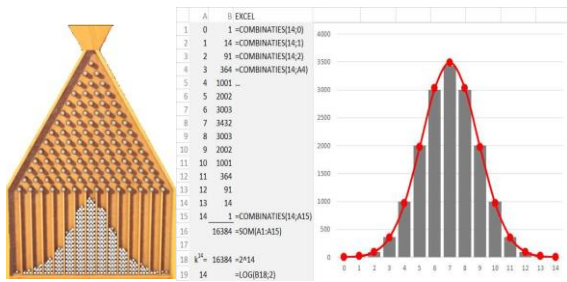
1 over 4   2 over 4   3 over 4

$$\frac{4*3*2*1}{1*\,3*2*1} \qquad \frac{4*3*2*1}{2*1*\,2*1} \qquad \frac{4*3*2*1}{3*2*1*\,1}$$
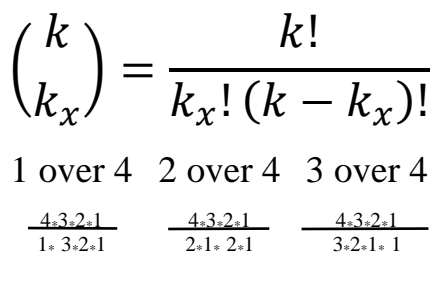
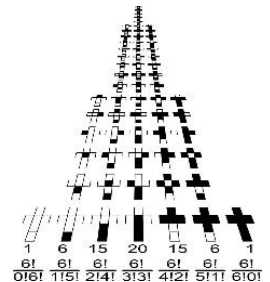| *Fig. 89*	*Fig. 90 2^{14} = 16384*	*Fig. 91 Binomium of*	*Fig. 92*
Galton-	*combinations*	*Newton: k_x over k*	*2^6 combinations*
board[a]	*divided over 15 columns*	*combinations*	

A site with 6 lots, of which 0, 1, 2, 3, 4, 5 or 6 can be built, has in that order f = {1, 6, 15, 20, 15, 6 or 1} (total N = 64) building possibilities (*Fig. 92*). There are, however, surprisingly many cases where a combination must be chosen from an unbeatable number of possibilities ('combinatorics').

Compiling a team from a football club with 25 members can go to '11 out of 25 '= 4 457 400 ways ('combinatoric explosion'). That coach therefore has a considerable choice problem. If he left it to chance, it would amount to throwing 11 dice with 25 sides each (possible outcomes). There is still a chance that some dice will point the same member, so that he has to throw them more often.

A NORMAL DISTRIBUTION DETERMINES HOW OFTEN YOU CAN EXPECT ANYTHING IN BETWEEN

If the number of possible outcomes N is not exactly a power of 2, or with a dice of 6, but of an infinite number, such as the lengths of 17mln people, then you need a continuous probability distribution for all real numbers. Such a distribution has an infinite number of possible outcomes N for every k observations x and their chances p.

Such a 'normal distribution' must condense around an average μ but dilute infinitely to the rare cases.

Gauss came up with a formula[b] that 'continuously' fills the gaps between whole numbers (red in *Fig. 88*, *Fig. 90* and *Fig. 94*). Newton's 'discrete' numbers at k = 8 (*Fig. 88*) differ little from that normal distribution, good for k = ∞.

The average of all deviations x-μ determines how 'flat' Gauss (*Fig. 94*) is. If the - and + deviations on both sides of μ disappear symmetrically, their sum is zero, so you can not calculate that average (0 / k).

---

a https://yourstory.com/2016/08/02373fb268-power-of-thoughts-truth-or-bullshit/

b That formula $\frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ gives the chance of outcome x. In Excel: NORMDIST(x;μ;σ;FALSE). The 'cumulative' probability of a result between two chances is the area, the 'integral' of this formula. It is so complicated that textbooks do provide a table with different exceedance probabilities for values of x, but Excel calculates them all with 1-NORMDIST(x;μ;σ;TRUE).

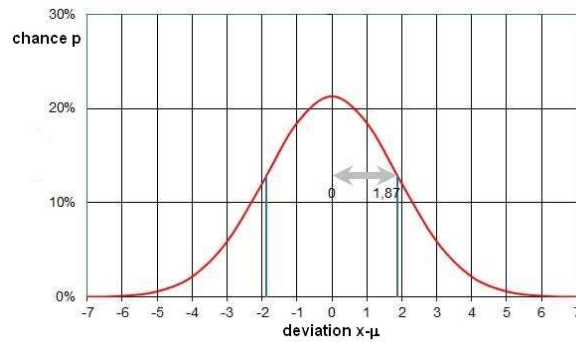| k	f	u=k-μ	(k-μ)²	f(k-μ)²	Gauss	/N	deviation
0	1	-7	49	49	3,2	0%	0,0%
1	14	-6	36	504	20,4	0%	0,0%
2	91	-5	25	2275	98,2	1%	0,0%
3	364	-4	16	5824	355,3	2%	-0,1%
4	1001	-3	9	9009	965,9	6%	-0,2%
5	2002	-2	4	8008	1973,0	12%	-0,2%
6	3003	-1	1	3003	3028,7	18%	0,2%
7	3432	0	0	0	3493,8	21% μ	0,4%
8	3003	1	1	3003	3028,7	18%	0,2%
9	2002	2	4	8008	1973,0	12%	-0,2%
10	1001	3	9	9009	965,9	6%	-0,2%
11	364	4	16	5824	355,3	2%	-0,1%
12	91	5	25	2275	98,2	1%	0,0%
13	14	6	36	504	20,4	0%	0,0%
14	1	7	49	49	3,2	0%	0,0%
N=	16384	Σ(x-μ)²	280	57344 =Σf(k-μ)²			
				16384 =N			
			variance	3,5 =Σf(k-μ)²/n			
μ = 7			σ=	1,870829 =√Σf(k-μ)²/n			

Fig. 93 Calculation of σ from the frequencies of Fig. 90 and…

Fig. 94, …its normal Gauss-distribution with μ=0 and σ=1,87

Squaring makes everything positive. The square sum of all deviations $\Sigma(x-\mu)^2$ is therefore a positive measure for the total of all deviations of μ. That is on average $\Sigma(x-\mu)^2/N$ ('variance'). If you measure the different lengths of people in cm, then you get that variance in cm^2. That is a surface, and the deviation is not a surface, but a length measure.

The root $\sqrt{\Sigma(x-\mu)^2/n}$ is then again in cm. That root of the variance is the 'standard deviation' σ.

The standard deviation σ has some special characteristics. Its deviation is the inflection point where densification around μ must pass into dilution to infinity (the graph becomes from convex into hollow).

Within the area bounded by both -σ and +σ on both sides, you will find 68.2% of the results.

At 1,96σ this is already 95%. The rest is then called 5% 'exceedance chance'.

For Gauss (note dddddddd p118) you only need to know μ and σ in order to predict the probability of occurrence of each outcome.

If you do not know all the outcomes x, but their frequencies f in the neighbourhood of each k (their 'class'), take k instead of x (*Fig. 93*) and put another f in the formula: $\sigma=\sqrt{\Sigma f(k-\mu)^2/n}$.

If you do know all N results x, then k is x and f=1, so that you can leave out that f again.

119

SAMPLES ALLOW CONCLUSIONS ABOUT THE WHOLE BY PROBABILITY TESTS

With the results of a sample of n from a much larger mass N you can already estimate μ and σ quite nicely.

You then write m, s, z instead of μ, σ, x. Gauss (m, s) predicts all outcomes within set 'confidence limits'.

By comparing your outcomes with that prediction, you can find deviations that require further explanation.
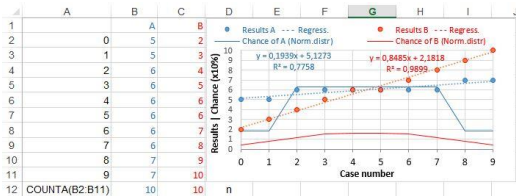

Fig. 95 Two series of outcomes A and B, their chances

Fig. 96 Comparative values and tests in Excel

You can now also compare two different series of outcomes A and B.

For example, **Fig. 95** shows two data sets. **Fig. 96** shows the differences and 'tests' some relations between A and B.

For example, the 'χ^2-test' ('chi-square test') shows that A and B still have a 5.5% chance of actually being one normal distribution ('not independent'). The r-test shows a considerable correlation. The t-test shows at least 100% chance of an equal average, but the F-test proves a considerable difference in variance and therefore in standard deviation.

Of course we already knew that from **Fig. 96**, but you can use these tests if you do not have all the data yet.

SOME CASES FROM COUNTLESS CASES STILL HAVE A CHANCE

A year counts $365*24*60*60 = 31\ 536\ 000$ seconds. If you stay awake you will see about 16 frames per second.

These are therefore N = 504 576 000 observations. If you see a lightning (1 to 0.0001 sec) on average 4 times a year in 1/16 of a second, then you would calculate a negligible average probability of 4 out of more than 500 million, practically zero. The complementary chance of no lightning strike is almost 100%

This is a binomial case (cross or coin) of whether or not a lightning occurs, but the deviations, variance or standard deviation, are unpredictable. How can you now know how much chance you run in an entire year, to experience no or 10 lightning strikes? Such a probability problem with innumerable observations N in time occurs in insurances, waiting times and the like.

120

Poisson came up with a formula[a] for a given μ (eg 4) with a standard deviation σ=√μ (eg 2) for each period (eg a year).

You only need to know μ, but with Poisson that is called λ. That formula gives the chance of k occurrences (**Fig. 97**).

If λ is far enough from 0, then the distribution seems to be quite similar to that of Gauss.

However, if the average λ is close to 0 (eg 4, *Fig. 98*), the Poisson distribution becomes skewed.

With an average of 4 lightning strikes per year (20% chance) you have 10% chance that it will be 4 + 2 = 6, but on the other hand 15% chance of 4-2 = 2. That division "collides with the zero limit".

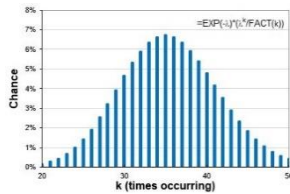A lightning strike can indeed occur 0 times, but not -1 times.

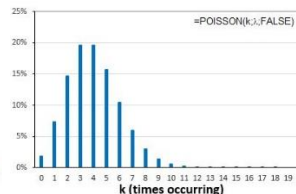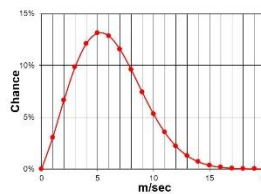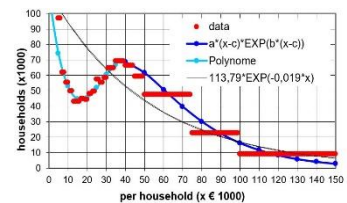**Fig. 97 Poisson at λ=35,5** **Fig. 98 Poisson at λ=4** **Fig. 99 Weibull[b]** **Fig. 100 An income distribution[c]**

NON-NORMAL DISTRIBUTIONS SUPPOSE DEVIATING DEVIATIONS

There are other skewed probability distributions that start at 0, such as the continuous 'Weibull distribution'.

For example, it is used for wind speeds (*Fig. 99*). The wind also can not blow less than with a speed 0.

The distribution of incomes is also distorted, but by a bureau for statistics data on incomes (the red dots in *Fig. 100*) are often given per 'class' (from .. to). At the higher incomes, the frequency classes in *Fig. 100* are horizontal red lines.

The lower incomes are given in smaller classes. As a result, around 15 thousand you can see a dent that would disappear from one large class of 0 to 40,000. The class-middle would then reasonably be on an 'exponential' downward trend.

Exponential functions are usually written as powers of the number e (= 2.718 ...), because they integrate easily (adding 'cumulative' opportunities as a surface). In Excel you write e^x as 'EXP (x)'. If you click on an exponential 'trend line' in Excel for the data of *Fig. 100*, it draws the continuously falling curve $113.79 * e^{-0.019x}$.[d]

---

a The formula $\dfrac{e^{-\lambda}.\lambda^k}{k!}$ is calculated in Excel with POISSON(k;λ;FALSE). With TRUE the cumulative version is calculated.

b $P(v, C, a): a * C * v^{C-1} * e^{\{(-a*v)^C\}}$

c p_1(x,a,f)= a+bx+cx^2+dx^3+ex^4+fx^5 and p_2(x,a,b,c)= a(x-c)e$^{b(x-c)}$

d This formula is printed with the 'correlation' curve (eg R^2 = 81%) if you have checked these options.

This 'regression line' therefore deviates considerably from the more detailed data with this moderate correlation.

In the formula $a*e^{b*x}$, a and b (the 'parameters') in **Fig. 100** get the values a = 113.79 and b = -0.019.
You should set these parameters differently each year if the data changes. That formula does not *predict* anything yet, but it is useful to be able to estimate intermediate values ('interpolate'). If, after a number of years, you find a trend in the course of the parameters yourself, you can perhaps risk a prediction ('extrapolate').

With a formula $a*(x-c)*e^{b*(x-c)}$ you approach the data after x=40 much better, but before x=40 it is totally wrong (in **Fig. 100** that part is omitted). There (x≤40) a 'polynomial' function fits much better $(p(x) = 2,3x^0-81x + 67x^2 + 0,37x^3-0,021x^4-0,0004x^5)$, but that on its turn does not work after x=40. The two functions that are connected at x=40 together describe the data very nice, but if you do not find trends in all those parameters, then you still do not have a prediction.[a]

REGRESSION REDUCES DEVIATIONS
In **Fig. 100** is a trend line in a point cloud drawn as a global relationship between income and the chance of it.
The method of finding a mathematical relation between two real variables between single data is called 'regression'.
Suppose a shopkeeper notes the average price of his products and his turnover for a few months to find out whether a price reduction is worthwhile (**Fig. 101**)[b]. The simplest is, of course, if that relationship shows a straight line y=ax+b.

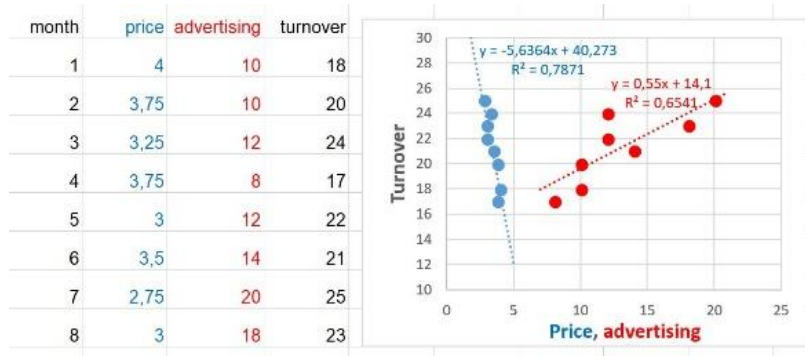| month	price	advertising	turnover
1	4	10	18
2	3,75	10	20
3	3,25	12	24
4	3,75	8	17
5	3	12	22
6	3,5	14	21
7	2,75	20	25
8	3	18	23

Fig. 101 Pricing, advertising and sales of a shopkeeper

Excel now gives a trend formula that promises that turnover doubles if he halves the price. That is of course not true, because if you see the price line in **Fig. 101** goes to zero (extrapolate to the top left) he would get a huge turnover according to that formula if he gives away everything for free. That straight ('linear') line is only reasonable (R2 = 79%) for those 8 observations, so that you can interpolate there, but extrapolating can lead to absurdities.

a Mathematicians do not like this kind of mess.
b The figures from this list come from Buijs(2003)Statistiek om mee te werken(Groningen)Stenfert Kroese p384

An advertising expert, to whom the retailer has spent a lot of money each month (column 'advertising' in *Fig. 101*), makes Excel claiming that sales will increase from 20 to 25 if you double advertising spending from 10 to 20.
The shopkeeper feels that something is not right. Who guarantees that the increase in turnover is not caused by a simultaneous price reduction? The 'pairwise' relationship between price and advertising per month is missing.

There are now three variables in the game. Statisticians then switch to 'multivariate analysis', and there are all sorts of mathematical methods for this.[a] I limit myself to 'multiple regression analysis'. Excel has an add-in for data analysis that provides a formula for this case: turnover = - 4.33*price+0.17*advertising+33.63. This shows that a price reduction for the turnover of this retailer has weighed much heavier (-4.33) than advertising costs (0.17).

Perhaps there are other variables (purchasing costs, competition, seasonal influences, fashion) that pull turnover up or down. Moreover, the question remains whether those regression lines are straightforward. Perhaps you can approach some better parabolic, exponential or with a polynomial. In any case, a wrong choice of two variables (turnover and advertising) or the omission of other (pricing) for the reality can lead to incorrect conclusions.

More coherent variables are difficult to understand, but we have to deal with them on a daily basis.
They can not be represented in one graph, but often people in a situation in which they have to take into account many things at the same time can still make the right decision.
How do our brains deal with so many variables from different senses and memories?

The increasing understanding of our neural system led computer experts to simulate the functioning of nerve cells (neurons) in a network (see § 26 p138). This has produced remarkably useful results in this century after a period of trial and error. It is an iterative process in which the weighting of different signals is constantly being adjusted ('learning').