The limits of x under Σ being infinite, $x+1$ can be replaced by x, consequently

$$w_z(y) = \frac{n-z}{n} w_z(y-1).$$

This difference-equation, in which y is the variable, may easily be integrated. As we have, further,

$$w_z(0) = (-1)^z \beta_a(z),$$

we get

$$w_z(y) = (-1)^z \cdot \beta_a(z) \cdot \left(\frac{n-z}{n}\right)^y.$$

By Oppermann's inverse transformation we find now:

$$u_x(y) = (-1)^x \Sigma \beta_z(x) \cdot (-1)^z \cdot \beta_a(z) \cdot \left(\frac{n-z}{n}\right)^y,$$

Σ taken from $x = -\infty$ to $x = +\infty$. This expression

$$u_x(y) = \beta_a(x) \Sigma (-1)^{z+x} \cdot \beta_{a-x}(z-x) \cdot \left(\frac{n-z}{n}\right)^y$$

has the above mentioned practical short-comings, which are sensible particularly if n, $a-x$, or y are large numbers; in these cases an artifice like that used by Laplace (problem 17) becomes necessary. But our exact solution has a simple interpretation. The sum that multiplies $\beta_a(x)$ in $u_x(y)$, is the $(a-x)^{\text{th}}$ difference of the function $\left(\frac{n-z}{n}\right)^y$, and is found by a table of the values $\left(\frac{n-a}{n}\right)^y$, $\left(\frac{n-a+1}{n}\right)^y$, $\ldots$ $\left(\frac{n-x-1}{n}\right)^y$, $\left(\frac{n-x}{n}\right)^y$, as the final difference formed by all these consecutive values. We learn from this interpretation that it is possible, if not easy, to solve this problem without the integration of any difference-equation, in a way analogous to that used in § 67, example 4.

If we make use of $w_z(y)$ to give us the half-invariants μ_1, μ_2 for the same law of errors as is expressed by $u_x(y)$, then we find for the mean value of x after y drawings

$$\lambda_1(y) = a\left(\frac{n-1}{n}\right)^y$$

and for the square of the mean error

$$\lambda_2(y) = a\left(\left(\frac{n-1}{n}\right)^y - \left(\frac{n-2}{n}\right)^y\right) + a^2\left(\left(\frac{n-2}{n}\right)^y - \left(\frac{n-1}{n}\right)^{2y}\right).$$

XVI. THE DETERMINATION OF PROBABILITIES A PRIORI AND A POSTERIORI.

§ 70. The computations of probabilities with which we have been dealing in the foregoing chapters have this point in common that we always assume one or several probabilities to be given, and then deduce from them the required ones. If now we ask, how we obtain those "given" probabilities, it is evident that other means are necessary than those which we have hitherto been able to mention, and provisionally it must be clear that both theory and experience must cooperate in these original determinations of probabilities. Without experience it is impossible to insure agreement with reality, and without theory in these as well as in other determinations we cannot get any firmness or exactness. In determining probabilities, however, there is special reason to distinguish between two methods, one of which, the *a priori* method seems at first sight to be purely theoretical, while the other, the *a posteriori* method, is as purely empirical.

§ 71. The *a priori* determination of probabilities is based on estimate of equality, inequality, or ratio of the probabilities of the several events, and in this process we always assume the operative causes, or at any rate their mode of operation, to be more or less known.

On the one hand we have the typical cases in which we know nothing else with respect to the events but that each of them is possible, and in the absence of any reason for preferring any one of them to any other, we estimate them to be equally probable — though certainly with the utmost uncertainty. For instance: What is the probability of seeing, in the course of time, the back of the moon? Shall we say $\frac{1}{2}$ or $\frac{2}{3}$?

On the other hand we have the cases — equally typical, but far more important — in which, by virtue of a good theory, we know so much of the causes or combinations of causes at work that, for each of those which will produce one event, we can point out another (or n others) which will produce the opposite event, and which according to the theory must occur as frequently. In this case we must estimate the probability of the result at $\frac{1}{2}$ and $\frac{1}{n+1}$ respectively, and if the conditions stated be strictly fulfilled, such a determination of probability will be exact.

But even if such a theory is not absolutely unimpeachable, we can often in this way obtain probabilities, which are so nearly exact and have such infinitely small mean errors, that we may very well make use of them, and compute from them values which may be used as our theoretically given probabilities. We are not more strict in other kinds of computations. In astronomical adjustment, for instance, it is almost an established practice to consider all times of observation as theoretically given. Their real errors, however, will often give occasion to sensible bonds between the observed co-ordinates; but the fact is that it would require great labour to avoid the drawback.

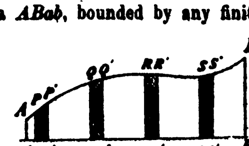
Such an a priori determination of probabilities is particularly applicable in games. For it is essential to the idea of a game that the rules must be laid down in such a way that, on the one hand they exclude all computation beforehand of the result in a particular case, while on the other hand they make a pretty exact computation of the probabilities possible. The procedure employed in a game, e. g. throwing of dice or shuffling of cards, ought therefore to exclude all circumstances that might permit the players to set causes in train, which could bring about or further a certain event (corriger la fortune). But also those circumstances ought to be eliminated, which not only by their incalculability make a judgment of the probabilities very insecure, but, above all, make it depend on the theoretical insight of the parties. Otherwise the game will cease to be a fair game and will become a struggle. The so-called stock-jobbing is rather a war than a game.

When the estimate of the probabilities depends essentially upon personal knowledge, we speak of a *subjective probability*. This too plays a great part, especially in daily life. The fear which ignorant people have of all that is new and unknown, proves that they understand that there is a great uncertainty in the estimate, and that it is greater for those who know but a little, than for those who know more and are therefore better able to judge.

Roulette may be taken as an example of the *objective probability* which arises in a well arranged game. A pointer turns on an almost frictionless pivot and points to the scale of a circle whose center is in the pivot. The pointer is made to revolve quickly, and the result of the game depends on where it stops. If the pointer stops opposite a space — suppose a red one — previously selected as favourable, the game is won.

There we have as essential circumstances: 1) the length of the arc which is traversed, this being determined by the initial velocity and the friction, 2) the initial position, and 3) the manner in which the circle is divided.

The length of the arc is unknown, especially when we take care to exclude very small velocities, and when the friction, as already mentioned, is very slight. So much only may be regarded as given, that the frequency of a given length of the arc must, as function of this length, be expressed by a functional law of errors of a nearly typical form. For the frequency must go down, asymptotically, as far as 0, both below and above limits of the arc which will be separated by many full revolutions of the pointer, and with at least one maximum between these limits. If now, for instance, it depended on, whether the arc traversed was greater or smaller than a certain value, the apparatus would be inexpedient, it would not allow any tolerably trustworthy a priori estimate. But if the winning space (or spaces) is small in proportion to the total circumference and, moreover, repeated regularly for each of the numerous revolutions, then the a priori determination of the probabilities will be even very exact. For an area $ABab$, bounded by any finite, continuous curve whatever (in the present case the curve of errors of the different possible events), by the axis of abscissae, and two ordinates, can always as a first approximation be expressed as the sum of numerous equidistant small areas pP', qQ' .... with a constant base, multiplied by the



interval $pq = qr = \ldots$ and divided by the base $pp' = qq' = \ldots$ And if we speak of the total area of a *curve of errors*, then the series of which the first term is this approximation, is even very convergent, in such a degree as $\theta(x) = 1 + x + x^4 + x^9 + x^{16} + \ldots$ for small x, and the said approximation is sufficient for all practical purposes.

That the initial position of the roulette is unknown, does not essentially change the result of the foregoing, viz. that the probability of winning is $\frac{pp'}{pq}$. This uncertainty can only cause an improvement of the accuracy of this approximation. If we may assume that the pointer will as probably start from any point in the circle as from any other, this determination $\frac{pp'}{pq}$ will even be exact, without any regard to the special kind of the unknown function of frequency.

The ratio of the winning space on the circle pp' to the whole circumference pq, the third essential circumstance, cannot be determined wholly a priori, but demands a measurement or a counting whose mean error it is essential to know.

The a priori determination of probability can thus, according to circumstances, give results of the most different values, from the very poorest through gradual transition up to such exact probabilities as agree with the suppositions in § 65 seqq., and permit the probability to replace the whole law of errors for our predictions. But what the a priori method cannot give, is a quantitative statement of the uncertainty which affects the numerical value of the probability itself. Only when it is evident, as in the example of the roulette, that this uncertainty is infinitely small, can we make use of a priori probabilities in computations that are to be relied on. If in the work and struggles of our life, we cannot entirely avoid building on altogeth r uncertain and subjective a priori estimates, great caution is necessary, and in order not to overdo this caution for want of a proper measure, we must try, by tact or experience, without any real method, to get an estimate of the uncertainty.

Even by the best a priori determinations of probability caution is not superfluous; the dice may be false, the pivot of the roulette may be worn out or bent, and so on.

§ 72. By the *a posteriori determination of probability* we build on the law of the large numbers, inferring from a law of actual errors in the form of frequency to the law of presumptive errors in that of the probability. We repeat the trial or the observation, and count the numbers m for the favourable and n for the unfavourable events.

Owing to the signification of a probability as mean value, the single values being 0 for every unfavourable event, 1 for every favourable event, the probability p for the favourable event must be transferred unchanged from the law of actual errors to that of presumptive errors; consequently

$$p = \frac{m}{m+n}. \tag{130}$$

Since, according to the same consideration, the square of the mean deviation for a single trial is $\frac{s_2 s_0 - s_1^2}{s_0^2} = \frac{mn}{(m+n)^2}$, and the number s_0 of the repetitions is $= m+n$, the square of the mean errors must, according to (47), be

$$\lambda_2 = \frac{mn}{(m+n)(m+n-1)}, \tag{131}$$

which is, therefore, the square of the mean error for a single trial, whether this is one of those which we have made, or is a repetition which we are still to make, and for which we are to compute the uncertainty.

If we then ask for the mean error of the probability $p = \frac{m}{m+n}$, got from the $m+n$ repetitions, we have

$$\lambda_2(p) = \frac{mn}{(m+n)^2(m+n-1)} = \frac{p(1-p)}{m+n-1} \tag{132}$$

as the square of this mean error.

The identity

$$\frac{mn}{(m+n)^2} + \frac{mn}{(m+n)^2(m+n-1)} = \frac{mn}{(m+n)(m+n-1)}$$

or

$$pq + \lambda_2(p) = \lambda_2 \tag{133}$$

shows that the mean error at a single trial, when the probability p is determined a posteriori by $m+n$ repetitions, can be computed by (34), as originating in two mutually free sources of errors, one of which is the normal uncertainty belonging to the probability, for which $\lambda_2 = pq$ (123), while the other is the inaccuracy of the a posteriori determination, for which $\lambda_2(p)$ is the square of the mean error.

The a posteriori determination therefore never gives an exact result, but only an approximation to the probability. Only when the number of repetitions we employ is so large that their reduction by a unit may be regarded as insignificant, we can immediately employ the probabilities found by means of them as complete expressions for the law of errors. But even by the very smallest number of repetitions of the trial, we not only obtain some knowledge of the probability, but also a determination of the mean error, which may be useful in predictions, and may serve as a measure of the caution that is necessary. It must be admitted that it is not such a simple thing to employ these mean errors as those in the ideal theory of probability, but it is not at all difficult.

As above mentioned, the a posteriori determination of probability seems to be purely empiric; theory, however, takes part in it, but is concealed in the demand, that all the trials we make use of must be repetitions, in the same way as the future trials whose results and uncertainty are predicted by the a posteriori probabilities. Transgressions of this rule, which reveal themselves by unsuccessful predictions, are by no means rare, and compel statistics and the other sciences which work with probabilities, to many alterations of their theories and hypotheses, and to the division of the materials obtained by trial into more and more homogeneous subdivisions.

Example. A die is inaccurate and suspected of being false. On trial, however, we have on throwing it 126 times got "six" exactly 21 times, and so far, all is right. The probability of "six" is found, consequently, to be $p = \frac{21}{126} = \frac{1}{6}$; the square of the mean error is $\lambda_2(p) = \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{1}{125} = \frac{1}{900}$; the limits indicated by the mean errors are consequently $\frac{1}{6} \pm \frac{1}{30}$, or $\frac{2}{15}$ and $\frac{1}{5}$.

If now we seek the probability that we shall not get "six" in 6 throws, the probability is still as by an accurate die $(1-p)^6 = \frac{15625}{46656} = \frac{1}{3} + \ldots$, but what is now the mean error? Ideally, its square should be $(1-p)^6(1-(1-p)^6) = \frac{2}{9} + \ldots$ But if p can have a small error dp, the consequent error in $(1-p)^6$ will be $-6(1-p)^5 dp$; if then the square of the mean error of p is $= \frac{1}{900} = p(1-p)\frac{1}{s_0 - 1}$, the total square of the mean error of the probability of not getting "six" in 6 throws will be

$$\lambda_2 = (1-p)^6(1-(1-p)^6) + 36(1-p)^{10} \cdot p(1-p)\frac{1}{s_0 - 1}$$
$$= \frac{2}{9} + \ldots + \frac{2}{311} + \ldots = \frac{8}{35} + \ldots$$

In every single game of this sort the mean error is therefore only slightly larger than with an accurate die, but its actual value is so large (nearly $\frac{1}{2}$) as to call for so much caution on the part both of the player and of his opponent, that there is not much chance of their laying a wager. This may be remedied by stipulating for a large number of repetitions of the game. Let us examine the conditions if we are to play this game of making 6 throws without "six" 72 times. With the above approximate fractions there will be expectation of winning $72 \cdot \frac{1}{3} = 24$ games. In the computation of the square of the mean error of this result, the first term in the above λ_2 must be multiplied by 72, but the second by 72^2; hence

$$\lambda_2 = \frac{2}{9} \cdot 72 + \frac{2}{311} \cdot 5184$$
$$= 16 + 33 = 49.$$

The mean error will be about 7, while it would only have been 4, if the die had been quite trustworthy.

§ 73. We have mentioned already, in § 66, the skewness of the laws of errors which is peculiar to all probability. It does not disappear, of course, in passing from the law of actual errors to that of presumptive errors, and in the a posteriori determination of probability it produces what we may call the *paradox of unanimity*: if all the repetitions we have made agree in giving the same event, the probability deduced from this, a posteriori, must not only be 1 or 0, but the square of the mean error $\lambda_2(p)$ of these determinations (as well as the higher half-invariants) becomes $= 0$. Must we infer then, respectively, to certainty or to impossibility, only because a smaller or greater number of repetitions mutually agree? must we consider a unanimous agreement as a proof of the absolute correctness of that which is thus agreed upon? Of course not; nor can this inference be maintained, if we look more closely at the law of errors $\mu_2 = 0$, $\mu_3 = 0$, ... $\mu_\infty = 0$. Such a law of errors, to be sure, *may* signify certainty, but not when, as here, the ratio $\mu_n : \mu_2^{\frac{n}{2}} = \infty$. A law of errors which is skew in an infinitely high degree, must indicate something peculiar, even though the mean error be ever so small. Add to this that it is not a strict consequence in practical calculations that, because the square of a number, here that of the mean error, is $= 0$, the number itself must be $= 0$, but only that it must be so small that it may be treated as a differential, which otherwise is indeterminate. The paradox being thus explained, it follows that no objections against the use of a posteriori probabilities in general can be based on it. But it must warn us to be cautious in computations with such probabilities as observed values, where the computation, as the method of the least squares, presupposes typical laws of errors. For this reason, we must for such computations reject all unanimously or nearly unanimously determined probabilities as unsuitable material of observation. Another thing is that we must also reject the hypothesis or theory of the computation, if it does not explain the unanimity. As an example we may take an examination of the probability of marriage at different ages. The a posteriori statistics before the c. 20[th] year and after the c. 60[th] must not be used in the computation of the sought constants of the formula, but the formula can be employed only when it has the quality of a functional law of errors so that it approaches asymptotically towards 0, both for low and high ages.

The paradox of unanimity has played rather a considerable part in the history of the theory of probabilities. It has even been thought that we ought to compute a posteriori probabilities by another formula

$$p = \frac{m+1}{m+n+2} \text{ (Bayes's Rule)} \tag{134}$$

and not, as above, by the formula of the mean number

$$p = \frac{m}{m+n}.$$

The proofs some authors have tried to give of Bayes's rule are open to serious objections. In the "Tidsskrift for Mathematik" (Copenhagen, 1879), Mr. *Bing* has given a crushing criticism of these proofs and their traditional basis, to which I shall refer those of my readers who take an interest in the attempts that have been made to deduce the theory of probabilities mathematically from certain definitions.

Bayes's rule has not been employed in practice to any greater extent, particularly not in statistics, though this science works entirely with a posteriori probability. But as it makes the paradox of unanimity disappear in a convenient way, and as, after all, we can neither prove nor disprove the exact validity of a formula for the determination of an a posteriori probability, any more than we can do so for any transition whatever from the law of actual errors to that of presumptive errors, the rule certainly deserves to be tested by its consequences in practice before we give it up altogether. The result of such a test will be that the hypothesis that Bayes's rule will give the true probability, can never deviate more than at most the amount of the mean error from the result of the series of repetition, viz. that m events out of $m+n$ have proved favourable. In order to demonstrate this proposition we shall consider a somewhat more general problem.

If we assume that trials have been previously made which have given μ favourable, ν unfavourable events, and that we have now in continuing the trials found m favourable and n unfavourable events, then the probability, being looked upon as the mean value, is determined by

$$p = \frac{m+\mu}{m+n+\mu+\nu}, \tag{135}$$

of which Bayes's formula is the special case corresponding to $\mu = \nu = 1$. Bayes's rule would therefore agree with the general rule, if we knew before the a posteriori determination so much of the probability of both cases, as a report of one earlier favourable event and one unfavourable event.

In the more general case the square of the mean error at the single trial is now

$$\lambda_2 = \frac{(m+\mu)(n+\nu)}{(m+n+\mu+\nu)(m+n+\mu+\nu-1)},$$

and for the $m+n$ trials is

$$\lambda_2(m+n) = (m+n)\lambda_2.$$

If we now compare with this the square of the deviation between the new observation and its computed value, that is, between m and $(m+n)p$, we find

$$\frac{(m-(m+n)p)^2}{\lambda_2(m+n)} = \frac{(\mu n - \nu m)^2}{(m+\mu)(n+\nu)(m+n)} \cdot \frac{m+n+\mu+\nu-1}{m+n+\mu+\nu}$$
$$= (\mu+\nu)\left(\frac{\mu}{\mu+\nu} - \frac{m}{m+n}\right)\left(\frac{\mu}{m+\mu} - \frac{\nu}{n+\nu}\right)\frac{m+n+\mu+\nu-1}{m+n+\mu+\nu}. \tag{136}$$

It appears at once from the latter formula that the greatest imaginable value of the ratio is the greatest of the two numbers μ and ν. In Bayes's rule $\mu = \nu = 1$. Here, therefore, 1 is the absolute maximum of the ratio of the square of deviation to that of the mean error. With respect to Bayes's rule the postulated proposition is hereby demonstrated. But at the same time it will be seen that we can replace Bayes's rule by a better one, if there is only an a priori determination, however uncertain, of the probability we are seeking. If we take the a priori probabilities ϖ for, and $(1-\varpi)$ against, instead of μ and ν, so that

$$p = \frac{m+\varpi}{m+n+1}, \tag{137}$$

then we are certain to avoid the paradox of unanimity where it might do harm, without deviating so much as the mean error from the observation in the a posteriori determination.

Neither Bayes's rule nor this latter one can be of any great use; but we can always employ them, when the found probabilities can be looked upon as definitive results. On the other hand, the formula of the mean value *may* be used in all cases, if we interpret the paradox of unanimity correctly. Where the found probabilities are to be subjected to adjustment, the latter formula, as I have said, *must* be employed; nor can the other rules be of any help in the cases where observed probabilities have to be rejected on account of the skewness of the law of errors.

XVII. MATHEMATICAL EXPECTATION AND ITS MEAN ERROR.

§ 74. Whether the theory of probability is employed in games, in insurances, or elsewhere, in all cases nearly in which we can speak of a favourable event, the prediction of the practical result is won through a computation of the mathematical expectation. The gain which a favourable event entails, has a value, and the chance of winning it must as a rule be bought by a stake. The question is: How are we to compare the value of the latter with that of which the game gives us expectation? Imagine the game to be repeated, and the number of repetitions N to become indefinitely large, then it is clear, according to the definition of probability, that the sum of the prizes won, if each of them is V, must be pNV, when p indicates the probability. The gain to be expected from every single game is consequently pV, and this product of the probability and the value of the prize is what we call mathematical expectation.

The adjective "mathematical" warns us not to consider pV as the real value which the possible gain has for a single player. This value, certainly, depends, not only objectively on the quantity of good things which form the prize, but also on purely subjective circumstances, among others on how much the player previously possesses and requires of the same sort of good things. An attempt which has been made to determine by means of what is called the "moral expectation", whether a game is advantageous or not, must certainly be regarded as a failure. For it takes into account the probable change in the logarithm of