NOTES

PROBABLE INFERENCE, THE LAW OF SUCCESSION, AND STATISTICAL INFERENCE

BY EDWIN B. WILSON, *Harvard School of Public Health*

Probable Inference (*Usual*). If there be observed a certain frequency or rate p_0 in a population of n and if the corresponding standard deviation $(p_0q_0/n)^{1/2}=\sigma_0$ be computed, the common statement of probable inference is to say that: The probability that the true value of the rate p lies outside its limits $p_0-\lambda\sigma_0$ and $p_0+\lambda\sigma_0$ is less than or equal to P_λ. It is assumed that P_λ decreases with an increase of λ. If the criterion of Tchebycheff is used, P_λ is itself less than $1/\lambda^2$; but if the probability table is used, P_λ is the area under the probability curve beyond the ordinates $\pm\lambda\sigma_0$. The rule of Tchebycheff is exceedingly conservative in its estimate of P_λ, whereas the probability table gives a radical estimate.

Strictly speaking, the usual statement of probable inference as given above is elliptical. Really the chance that the true probability p lies outside a specified range is either 0 or 1; for p actually lies within that range or does not. It is the observed rate p_0 which has a greater or less chance of lying within a certain interval of the true rate p. If the observer has had the hard luck to have observed a relatively rare event and to have based his inference thereon, he may be fairly wide of the mark.

Probable Inference (*Improved*). A better way to proceed is to reason as follows: There is some rate p. Its standard deviation is $(pq/n)^{1/2}=\sigma$. The probability that an observation as bad as p_0 will occur, where p_0 lies outside the limits $p-\lambda\sigma$ and $p+\lambda\sigma$, is less than or equal to P_λ. This form of statement throws the emphasis upon the fallibility of a particular observation in respect to being typical of a general situation.

It is still possible to state the criterion in terms of the observed rate p_0 for the equation $(p_0-p)^2=\lambda^2pq/n$, where $q=1-p$, is quadratic in p and may be solved to find p. If $\lambda^2/n=t$, the solution is

$$p=\frac{p_0+t/2}{1+t}\pm\frac{\sqrt{p_0q_0t+t^2/4}}{1+t}.$$

The rule then may be stated as: If the true value of the probability p lies outside the range

$$\frac{p_0+t/2}{1+t}-\frac{\sqrt{p_0q_0t+t^2/4}}{1+t} \quad\text{and}\quad \frac{p_0+t/2}{1+t}+\frac{\sqrt{p_0q_0t+t^2/4}}{1+t},$$

the chance of having such hard luck as to have made an observation so bad as p_0 is less or equal to P_λ. And this form of statement is not elliptical. It is the proper form of probable inference.

Concerning the range indicated, it may be remarked that it is not centered at the value p_0 but at the value $(p_0+t/2)/(1+t)$ which differs from p_0 by being displaced toward the value 1/2 by the amount

$$\frac{p_0+t/2}{1+t}-p_0=\frac{t(1/2-p_0)}{1+t}=\frac{(q_0-p_0)t/2}{1+t}.$$

Moreover, the interval on either side of the mean is

$$R=\sqrt{p_0q_0/n+\lambda^2/4n^2}/(1+\lambda^2/n),$$

which is not identical with $\lambda\sigma_0$ computed from p_0 nor with that value $\lambda\sigma_c$ which might be computed from the central value $(p_0+t/2)/(1+t)$ of the range indicated. In fact $R<\lambda\sigma_c$ and $\lambda\sigma_0<\lambda\sigma_c$, but R may be either less than or greater than $\lambda\sigma_0$—less if p_0 lies between .067 and .933, greater if p_0 lies outside those limits unless $t=\lambda^2/n$ be considerable compared with 2. The precise lines of division are $p_0=\frac{1}{2}\pm\frac{1}{2}\sqrt{1-(2+t)^{-2}}$.

The Law of Succession. The law of succession of Laplace states that if we have experienced S successes and F failures out of $S+F=n$ trials, the chance of success on the $(n+1)$st trial is $(S+1)/(n+2)$. Thus the law of succession purports to give the probability from experience not as $p_0=S/n$ but as $p=(S+1)/(n+2)$. This chance is, however, not the true chance of success, because the chance of success p on every trial must be the same. The proof of the law depends on inverse probabilities and in particular on the assumption that all probabilities are *a priori* equi-probable. The proof has been much criticized, for it has been held that the experience $p_0=S/n$ does not permit the assumption that all probabilities are equi-probable, but indicates that those in the neighborhood of p_0 must be much more probable than those remote from p_0. The simplest, if crudest, form of the argument of equi-probability is found in interpreting the formula $(S+1)/(n+2)$ as giving two new trials of which one is assumed to be a success and the other a failure.

If we apply the criterion in terms of the standard deviation as above developed we may state that the center of the range for p is $(p_0+t/2)/(1+t)$. If we now replace t by λ^2/n, and $p_0 n$ by S, the center of the range becomes $(S+\lambda^2/2)/(n+\lambda^2)$, and the probable inference is this: If the true probability lies outside the range

$$\frac{S+\lambda^2/2}{n+\lambda^2}-\lambda\frac{\sqrt{SF/n+\lambda^2/4}}{n+\lambda^2} \text{ and } \frac{S+\lambda^2/2}{n+\lambda^2}+\lambda\frac{\sqrt{SF/n+\lambda^2/4}}{n+\lambda^2},$$

the chance of our having the hard luck to realize the observed value $p_0=S/n$ is less than or equal to P_λ. As the distribution of the chances of an observation is asymmetric, it is perhaps unfair to take the central value of the range as the best estimate of the true probability; but this is what is actually done in practice.

In terms, therefore, of the practical criterion the forecasted value of the true probability is

$$\text{not } \frac{S+1}{n+2}, \text{ nor } \frac{S}{n}, \text{ but } \frac{S+\lambda^2/2}{n+\lambda^2};$$

and the value that should be assigned depends on the value of λ, *i. e.*, on our readiness to gamble on the typicalness of our realized experience. From this viewpoint, only those who believe that their experience is absolutely typical will set $\lambda=0$ and use as a forecast the realized frequency S/n. Those who use the law of succession, set $\lambda^2=2$ and allow a total variation in their experience of 2.8σ, *i. e.*, they wish to assert that they have not had an experience so rare that it or one less probable would arise, on the basis of the probability table as an estimate of P_λ, less than 16 times in 100. Those who make the usual allowance of 2σ for drawing an inference would use $(S+2)/(n+4)$ as a law of succession.

A particularly interesting and instructive case is that in which there has been total failure, $p_0=0$, $\sigma_0=0$. Here clearly the first form of the inference, namely, that the true value of p must lie between $p_0-\lambda\sigma_0=0$ and $p_0+\lambda\sigma_0=0$ is out of the question. The true form states that the experience is not so unusual as P_λ if p is less than $\lambda^2/(n+\lambda^2)n$ or if the expected number of instances is less than $\lambda^2/(n+\lambda^2)$, which for n large is practically λ^2/n. If this were applied to the classic case of determining the chance that the sun should fail to rise, one would take λ very small compared to 1 because general considerations of astronomy make it highly probable that our past experience is very nearly typical. If the application were to the fact that there were no deaths from leprosy in Massachusetts ($n=4{,}000{,}000$) in 1924, λ would also be taken small because leprosy is so rare, perhaps $\lambda=2/3$, meaning that we would take an even chance. But in the case of paratyphoid fever, we might prefer to use the ordinary criterion with $\lambda = 2$.

Statistical Inference. This brings us to statistical inference which had best be differentiated from probable inference by requiring that something over and above the value of p_0 be known, something that will motivate a choice among values for λ in drawing the inference. It is well known that some phenomena show less and some show more variation than that due to chance as determined by the Bernoulli expansion $(p+q)^n$. The value L of the Lexian ratio is precisely the ratio of the observed dispersion to the value of $(npq)^{1/2}$ or $(pq/n)^{1/2}$ as the case may be. If we have general information which leads us to believe that the variation of a particular phenomenon be supernormal ($L > 1$), we naturally shall allow for some value of L in drawing the inference. Thus if the Lexian ratio is presumed from previous analysis of similar phenomena to be in the neighborhood of 5, we may use $\lambda = 10$ as properly as we should use $\lambda = 2$ if the phenomenon were believed to be normal (Bernoullian).