Nims, L. F. 1934 *J. Amer. Chem. Soc.* **56**, 1110.
Nims, L. F. 1936 *J. Amer. Chem. Soc.* **58**, 987.
Owen, B. B. 1934 *J. Amer. Chem. Soc.* **56**, 1695.
Pedersen, K. J. 1933 *J. Phys. Chem.* **37**, 751.
Shilov, E. A. & Gladtchikova, J. N. 1938 *J. Amer. Chem. Soc.* **60**, 490.
Skrabal, A. & Mrazek, D. 1918 *S.B. Akad. Wiss. Wien*, **127**, 301.
Tong, L. K. J. & Olson, A. R. 1941 *J. Amer. Chem. Soc.* **63**, 3406.
West, R. W. 1924 *J. Chem. Soc.* p. 1277.
Wright, D. D. 1934 *J. Amer. Chem. Soc.* **56**, 314.

An invariant form for the prior probability in estimation problems

By Harold Jeffreys, F.R.S.

(*Received* 23 *November* 1945)

It is shown that a certain differential form depending on the values of the parameters in a law of chance is invariant for all transformations of the parameters when the law is differentiable with regard to all parameters. For laws containing a location and a scale parameter a form with a somewhat restricted type of invariance is found even when the law is not everywhere differentiable with regard to the parameters. This form has the properties required to give a general rule for stating the prior probability in a large class of estimation problems.

1. *The consistency of induction.* The consistency of pure mathematics has never been in serious doubt, but work mostly done in the present century has shown that a definite proof that pure mathematics is consistent is a matter of considerable difficulty. Nevertheless, both the creators and the users of pure mathematics have proceeded on the hypothesis that consistency at least cannot be disproved. Suitable axioms for the theory of real numbers have been stated, and if those of any extension can be formulated in terms of those of real numbers that has been held to be sufficient justification.

The usual attitude towards generalization from experience is quite different. It is taken for granted that it is possible, but when it comes to actual application there is seldom any attempt to state formal rules, and different workers proceed by methods that lead to different conclusions from the same data. Each is sure that there is only one right method, but they disagree about the fundamental principles.

The object of my *Theory of Probability* (1939) is to construct a system whose most important properties should be that (1) on the same evidence the probabilities of a given set of propositions shall have a unique order, (2) the axioms of the system shall not deny any law capable of being stated as a prediction concerning observations, (3) with sufficient observational evidence any such law can be established with a probability approaching certainty. Property (1) is the assertion that a consistent scheme is possible, but merely making this assertion does not guarantee that a given scheme is consistent. It does make it possible to derive theorems by equating probabilities found in different ways; and if, in spite of all efforts, probabilities found in different ways are different it makes it impossible to accept the situation as satisfactory. A theorem that has been established suffices to establish consistency over a large part of the subject, namely, that if q_1 and q_2 are two alternative hypotheses, p_1 and p_2 two sets of observational data, then the posterior probabilities of q_1 and q_2, given the same prior probabilities, are in the same ratio whether p_1 is taken into account and then p_2, the order reversed, or both allowed for at once. The proof depends on supposing that the product rule for probabilities holds for the likelihoods, but this has never been questioned. It follows at once from the definition if the probability of p_2, given $p_1 q_r H$, is the same as the probability of p_2, given $q_r H$ alone; that is, if the likelihoods are chances.* It is not obviously true in general, but it does follow that, if the product rule for probabilities is consistent for likelihoods, no inconsistency can arise through the principle of inverse probability.

The question then arises, can any general rule be laid down for assessing the prior probability itself? In existing theory there is a principle that specifies it for any quantity such as a standard error, where it might be equally reasonable to regard some power of the parameter as fundamental; the only rule that is invariant (except for an irrelevant constant) for such changes is the $d\sigma/\sigma$ rule. In other cases uniform distribution is used in estimation problems, and various more or less satisfactory rules are adopted in significance tests, but no general rule is available, and the absence of one makes further extensions difficult. For errors satisfying a Pearson law of Type VII, for instance, it is natural to use the $d\sigma/\sigma$ rule for the scale parameter and uniform distribution for the location parameter, but there is no definite rule to say what should be taken for the index—or, to put it in another way, what function of the index should be taken to have uniform prior probability distribution.

An attempt to give a general rule will require a new set of conditions. As long as each type of problem is treated separately, no problem of consistency arises; the prior probability in one type is irrelevant to that in another, and there is no more to be said. But it may turn out (1) that no general rule can be used consistently, (2) that there may be a choice of several, or that (3) a general rule may be self-consistent but contrary to the fundamental principles that we are trying to express in our formalism, or (4) conflict with other less fundamental principles that we are not prepared to abandon lightly. A satisfactory rule must, in particular, provide for the distinction between problems of estimation and significance.

2. *Invariants expressing the difference between two distributions of chance.* For simplicity take the laws as referring to one variable x; let the respective chances that x is not greater than an assigned value be P, P'. Then consider the integrals

$$I_1 = \int (\sqrt{dP'} - \sqrt{dP})^2, \quad I_2 = \int \log \frac{dP'}{dP}\, d(P' - P), \tag{1}$$

* This is used in the proof of Theorem 9 (1939, p. 24) but has escaped explicit statement.

defined in the Stieltjes manner, by taking δP, $\delta P'$ for the same interval δx, forming the approximating sums, and letting the intervals of x tend to zero. It is obvious from the form of these integrals, since they depend only on the distribution of P, P', that they are invariant for all transformations of the parameters in the laws and also for all non-singular transformations of x, the range of integration being over the range permitted to x. Further, they are both positive definite. They therefore provide measures of the discrepancy between the two laws.

The greatest possible discrepancies between the laws will be when $dP' = 0$ in all ranges where P varies, and conversely. In that case $I_1 = 2$ and I_2 is infinite. The same extreme values occur if P is continuous as a function of x but P' varies only at isolated points.

These are not the only invariants with similar properties; any integral of the form

$$\int |(dP')^m - (dP)^m|^{1/m} \tag{2}$$

would behave similarly; but I_1 and I_2 are apparently the only ones that are ordinarily of the second order in the differences of the parameters in the laws when these differences are small.

If the range of x is subdivided and $p_r = \delta P_r$, $p_r = \delta P'_r$ for the rth interval, and if p'_r is differentiable with regard to all parameters in the law, then to the first order

$$\sqrt{(\delta P'_r)} - \sqrt{(\delta P_r)} = \sqrt{p'_r} - \sqrt{p_r} = \frac{1}{2\sqrt{p_r}}(p'_r - p_r), \tag{3}$$

$$\log \frac{\delta P'_r}{\delta P_r} = \frac{1}{p_r}(p'_r - p_r), \tag{4}$$

and to the second order

$$I_1 = \lim \Sigma \frac{1}{4}\frac{(p'_r - p_r)^2}{p_r}, \quad I_2 = \lim \Sigma \frac{(p'_r - p_r)^2}{p_r}. \tag{5}$$

For laws referring to more than one variable only formal changes are needed. The limit taken corresponds to $(\delta x_r)_{\max} \to 0$. Now if the P_r law depends on parameters α_i ($i = 1$ to m), and P'_r on $\alpha_i + \Delta\alpha_i$,

$$\begin{aligned} I_2 &= \lim \Sigma \frac{1}{p_r}\left(\frac{\partial p'_r}{\partial \alpha_i}\Delta\alpha_i\right)\left(\frac{\partial p'_r}{\partial \alpha_k}\Delta\alpha_k\right) \\ &= g_{ik}\Delta\alpha_i\Delta\alpha_k, \end{aligned} \tag{6}$$

where

$$g_{ik} = \lim_{\delta x_r \to 0} \Sigma \frac{1}{p_r}\frac{\partial p'_r}{\partial \alpha_i}\frac{\partial p'_r}{\partial \alpha_k}. \tag{7}$$

Hence for small variations of the parameters I_2, and therefore I_1, have the form of the square of an element of distance in curvilinear co-ordinates. If we transform to any other parameters α'_j, I_2 is unaltered, and

$$I_2 = g'_{jl}\Delta\alpha'_j\Delta\alpha'_l, \tag{8}$$

$$g'_{jl} = g_{ik}\frac{\partial\alpha_i}{\partial\alpha'_j}\frac{\partial\alpha_k}{\partial\alpha'_l}, \tag{9}$$

$$\|g'_{jl}\| = \|g_{ik}\|\left\|\frac{\partial\alpha_i}{\partial\alpha'_j}\right\|^2. \tag{10}$$

But
$$d\alpha_1 d\alpha_2 \ldots d\alpha_m = \left\|\frac{\partial\alpha_i}{\partial\alpha'_j}\right\| d\alpha'_1 \ldots d\alpha'_m$$
$$= \left\{\frac{\|g'_{jl}\|}{\|g_{ik}\|}\right\}^{\frac{1}{2}} d\alpha'_1 \ldots d\alpha'_m. \tag{11}$$

Hence
$$\|g_{ik}\|^{\frac{1}{2}} d\alpha_1 d\alpha_2 \ldots d\alpha_m \text{ is invariant,} \tag{12}$$

and if the prior probability density of the parameters is taken as proportional to $\|g_{ik}\|^{\frac{1}{2}}$, the prior probability over any region will be invariant for all ways of choosing the parameters.

It remains to be seen whether such a choice agrees with those that have been used hitherto. Take first the normal law of error. Here

$$p_r \doteqdot \frac{1}{\sqrt{(2\pi)}\,\sigma}\exp\left\{-\frac{(x_r-\alpha)^2}{2\sigma^2}\right\}\delta x_r \tag{13}$$

and, exactly,

$$I_1 = \int_{-\infty}^{\infty}\frac{1}{\sqrt{(2\pi)}}\left[\frac{1}{\sqrt{\sigma'}}\exp\left\{-\frac{(x-\alpha')^2}{4\sigma'^2}\right\}-\frac{1}{\sqrt{\sigma}}\exp\left\{-\frac{(x-\alpha)^2}{4\sigma^2}\right\}\right]^2 dx$$
$$= 2-\int_{-\infty}^{\infty}\frac{2}{\sqrt{(2\pi)}\sqrt{(\sigma\sigma')}}\exp\left\{-\frac{(x-\alpha')^2}{4\sigma'^2}-\frac{(x-\alpha)^2}{4\sigma^2}\right\}dx$$
$$= 2\left[1-\frac{\sqrt{2}}{\sqrt{(\sigma/\sigma'+\sigma'/\sigma)}}\exp\left\{-\frac{(\alpha-\alpha')^2}{4(\sigma^2+\sigma'^2)}\right\}\right]. \tag{14}$$

Writing $\sigma = \sigma_0 e^{-\zeta}$, $\sigma' = \sigma_0 e^{\zeta}$, we have

$$I_1 = 2\left[1-\operatorname{sech}^{\frac{1}{2}} 2\zeta\exp\left(-\frac{(\alpha-\alpha')^2}{8\sigma_0^2\cosh 2\zeta}\right)\right], \quad \doteqdot 2\zeta^2+\frac{(\alpha-\alpha')^2}{4\sigma_0^2}, \tag{15}$$

when ζ is small.

Again,

$$I_2 = \int_{-\infty}^{\infty}\frac{1}{\sqrt{(2\pi)}}\left\{\log\frac{\sigma}{\sigma'}+\frac{(x-\alpha)^2}{2\sigma^2}-\frac{(x-\alpha')^2}{2\sigma'^2}\right\}\left[\frac{1}{\sigma'}\exp\left\{-\frac{(x-\alpha')^2}{2\sigma'^2}\right\}-\frac{1}{\sigma}\exp\left\{-\frac{(x-\alpha)^2}{2\sigma^2}\right\}\right]dx$$
$$= -1+\frac{\sigma'^2}{2\sigma^2}+\frac{\sigma^2}{2\sigma'^2}+\frac{(\alpha-\alpha')^2(\sigma^2+\sigma'^2)}{2\sigma^2\sigma'^2} \tag{16}$$
$$\doteqdot 8\zeta^2+\frac{(\alpha-\alpha')^2}{\sigma_0^2} \tag{17}$$
$$\doteqdot 2\left(\frac{\delta\sigma}{\sigma}\right)^2+\frac{(\delta\alpha)^2}{\sigma^2}.$$

Three cases arise. If α is fixed, $g_{\sigma\sigma} = 2/\sigma^2$, and the probability density for σ is proportional to $1/\sigma$, in agreement with the previous rule. If σ is fixed, $g_{\alpha\alpha}$ is constant and the probability density for α is uniform, again in accordance with the previous rule. But if α and σ are both to be treated as unknown, the determinant is $2/\sigma^4$, and then

$$P(d\sigma\, d\alpha \mid H) \propto d\sigma\, d\alpha/\sigma^2, \tag{18}$$

instead of the previous form $d\sigma\, d\alpha/\sigma$. The effect on the t distribution would be substantial, the index being increased by $\frac{1}{2}$. Further, if the same method was applied to a joint distribution for several variables about independent true values, an extra factor $1/\sigma$ would appear for each, and the index in the t distribution for n observations would always be $\frac{1}{2}(n+1)$ whatever the number of unknowns estimated. These consequences are unacceptable. To see what (18) really says, let us suppose first that α is known to lie between fixed limits independent of σ, which corresponds fairly well to the usual practical case. Then by integration the probability density for σ is proportional to $1/\sigma^2$, and there is probability 1 that σ is less than any assigned positive value. To avoid this we should have to suppose that the range permitted to α is proportional to σ. There would be nothing self-contradictory in this, but it is not the normal state of affairs in an estimation problem.

The conclusion to be drawn is that it would be a self-consistent and generally applicable procedure to take the prior probability density proportional to $\| g_{ik} \|^{\frac{1}{2}}$ subject to the chances being differentiable functions of the parameters, and therefore that under this condition a self-consistent general theory of induction is possible. It does not, however, cover the whole ground because where it has been seen to succeed it gives the estimation prior probability, not one suitable for significance tests; and if a scale parameter σ is taken as an unknown, irrelevant to the values of the other parameters, a correcting factor is needed.

Next consider simple sampling. Denote the chance of a success by $p = \sin^2\alpha$, that of a failure by $\cos^2\alpha$. Then for variations of α

$$I_1 = (\sin\alpha - \sin\alpha')^2 + (\cos\alpha - \cos\alpha')^2 = 4\sin^2\tfrac{1}{2}(\alpha' - \alpha), \tag{19}$$

$$I_2 = (\sin^2\alpha' - \sin^2\alpha)\log\frac{\tan^2\alpha'}{\tan^2\alpha} \doteqdot 4(\alpha' - \alpha)^2. \tag{20}$$

Then the rule gives, since $0 \leqslant \alpha \leqslant \frac{1}{2}\pi$,

$$P(d\alpha \mid H) = \frac{2}{\pi}\, d\alpha = \frac{dp}{\pi\sqrt{\{p(1-p)\}}}. \tag{21}$$

According to the Bayes-Laplace rule the prior probability of p should be taken uniform. Haldane has suggested

$$P(dp \mid H) \propto \frac{dp}{p(1-p)}, \tag{22}$$

which has the property that the expectation of p, given the sampling ratio, is equal to the sampling ratio. The present rule is intermediate. It is easy to construct models to make any of these correspond to states called ignorance in ordinary language, but the present more precise language makes it necessary to distinguish between them.

For the comparison of two Poisson distributions of the form

$$P(m \mid rH) = e^{-r}\frac{r^m}{m!},$$

then
$$I_1 = 2-2e^{-\frac{1}{2}(\sqrt{r'}-\sqrt{r})^2}$$

leading to
$$P(dr \mid H) \propto dr/\sqrt{r}.$$

This conflicts with the rule dr/r previously adopted for the Poisson parameter, and would say that r is practically certain to exceed any suggested value. Alternatively, in most cases of the Poisson law it would be equally reasonable to take r or $1/r$ as the fundamental parameter; and the dr/r law is the only one that has the same form for both.

3. *Limitations of the rule.* The requirement that the chances shall be differentiable with respect to the parameters is not always satisfied. One case is the rectangular distribution. If the laws are that the chance of a measure in a range dx is dx/α for $0<x<\alpha$, or dx/α' for $0<x<\alpha'$, take $\alpha'>\alpha$. Then according to the first law $dP=0$ for any interval included in $\alpha<x<\alpha'$, and I_2 is infinite. I_1 is still finite; it is, in fact,

$$I_1 = \int_0^\alpha \left(\frac{1}{\sqrt{\alpha'}}-\frac{1}{\sqrt{\alpha}}\right)^2 dx + \int_\alpha^{\alpha'} \frac{dx}{\alpha'} = 2\left(1-\sqrt{\frac{\alpha}{\alpha'}}\right).$$

For α' slightly greater than α, I_1 is of the first order in $\alpha'-\alpha$ instead of the second.

Another exceptional case is where the unknown itself is capable of taking only discrete values, as in the sampling of a finite population or in the 'tramcar' problem (Jeffreys 1939, p. 186). Here we cannot even try to differentiate. If the suggested numbers of the class are n and n', then, for a given n, I_1 is finite but takes only a discrete set of values; I_2 is infinite except for $n'=n$.

4. On inspection it is seen that most of these anomalies concern laws with a common feature; they include one or more parameters with an infinite range of possible values, and both for the normal law (with any number of parameters), the rectangular law, and the Poisson law the previous solution was quite satisfactory. These parameters are dimensional (though the Poisson parameter is itself a number, it arises as the product of two dimensional quantities, one of which is the arbitrary unit taken as a 'trial'). Suppose then that a law depends on a location parameter λ, a scale parameter σ of the same dimensions as λ, and on a number of numerical parameters α_i. Examples will be the indices in various Pearson laws and the negative binomial. It is just the occurrence of these parameters that lead in practice to a doubt about what should be taken as the best way of stating the law; for instance, whether any asymmetrical law should be stated with the first moment or the mode as location parameter. But different ways of stating the law that would ever be considered fall into the type represented by the transformation

$$\lambda' = \lambda+\sigma f(\alpha_i), \quad \sigma' = \sigma g(\alpha_i), \quad \alpha'_j = \alpha'_j(\alpha_i). \tag{23}$$

In practice, again, there is usually some more or less vague information indicating that λ and σ are practically certain to lie within wide but finite limits. With a transformation of this form the ranges permitted to λ' and $\log\sigma'/a$ (where a is any fixed quantity of the same dimensions as σ), for given α_i, will be practically the same as for λ and $\log\sigma/a$. Now

$$\frac{\partial(\lambda', \sigma', \alpha'_1 \dots \alpha'_m)}{\partial(\lambda, \sigma, \alpha_1 \dots \alpha_m)} = \frac{\partial(\lambda', \sigma', \alpha'_1 \dots \alpha'_m)}{\partial(\lambda, \sigma, \alpha'_1 \dots \alpha'_m)} \frac{\partial(\lambda, \sigma, \alpha'_1 \dots \alpha'_m)}{\partial(\lambda, \sigma, \alpha_1 \dots \alpha_m)}$$

$$= g(\alpha_i) \left\| \frac{\partial \alpha'_j}{\partial \alpha_i} \right\|. \tag{24}$$

Hence if I_1 or I_2 is formed for changes of the α_i only, then

$$d\lambda \frac{d\sigma}{\sigma} \| g_{ik} \|^{\frac{1}{2}} \Pi d\alpha_i = d\lambda' \frac{d\sigma'}{\sigma'} \| g'_{jl} \|^{\frac{1}{2}} \Pi d\alpha'_j, \tag{25}$$

in the sense that the integrals of both quantities over corresponding regions in the two cases are equal, and the same constant factor will be needed to make the total probability of all possible values equal to unity. Thus the condition of invariance for this type of transformation gives again the same type of law for the numerical parameters, but permits us to retain the previous form $d\alpha d\sigma/\sigma$ for location and scale parameters.

Most of the difficulty for laws that are not everywhere differentiable with regard to some parameter also disappears. It is well known that with some values of the indices Pearson Type I laws behave like $(x-c)^m$ near a terminus, where m may be less than 1, and that $m = 1$ is a critical value in the fitting of the law to observations. If $m > 1$, the law is differentiable at $x = c$, and the uncertainty of an estimate of c from the observations is nearly proportional to $n^{-\frac{1}{2}}$, where n is the number of observations. If $m < 1$, the law is not differentiable at $x = c$, and the uncertainty decreases faster than $n^{-\frac{1}{2}}$ (like n^{-1} for the rectangular law). The present device covers both cases, if we agree to take the termini in these laws always as $\lambda \pm \sigma$; for with fixed termini I_1 is a quadratic for small variations of the indices over the whole range of intelligibility of the laws.

Consequently the form (25) for the prior probability can be applied consistently over most estimation problems where both a scale and a location parameter have to be determined.

The outstanding exception is the case where a parameter can take only discrete values. This difficulty does not appear insuperable. Consider, for instance, the sampling of a finite population of number n, containing r members with property ϕ. This is often itself a sample of n derived from a chance x. Then

$$P(dx \mid nH) = \frac{dx}{\pi\sqrt{\{x(1-x)\}}},$$

$$P(r \mid n, x, H) = \frac{n!}{r!(n-r)!} x^r (1-x)^{n-r},$$

$$P(r\,dx \mid nH) = \frac{n!}{\pi\, r!(n-r)!} x^{r-\frac{1}{2}} (1-x)^{n-r-\frac{1}{2}}\, dx,$$

and by integration $$P(r \mid nH) = \frac{(r-\frac{1}{2})!\,(n-r-\frac{1}{2})!}{\pi\, r!(n-r)!},$$

which is finite both for $r = 0$ and $r = n$. If used when the population is not derived from a chance x, it will still be satisfactory.

Applications to determine posterior probabilities have not been developed in the present paper, because it is known that, within any limits that anybody would ordinarily suggest, changes of the prior probability make little difference to the posterior probability in practical cases, and the modifications of existing solutions would be easy. The use of the prior probability is to provide a starting-point, and it now appears that a consistent rule for it can be stated so as to be applicable to the majority of estimation problems.

Some significance tests based on the use of I_1 and I_2 when not small have been developed, and will, it is hoped, be the subject of a further paper.

5. *Relation to χ^2 and maximum likelihood.* Suppose the possible values of the variable to be discrete; if the laws are continuous, suppose the values grouped. If there are N observations the expectations of the numbers of observations in the rth group according to the two laws are Np_r, Np'_r. Suppose that the set of parameters α_i are used as a trial hypothesis, and that the set α'_i are in fact correct. Let the observed number in the rth group be n_r. Then

$$\chi^2 = \Sigma \frac{(n_r - Np_r)^2}{Np_r}, \tag{26}$$

and if N is large the expectation of χ^2, where n is the number of groups, is

$$\begin{aligned} E(\chi^2) &= n - 1 + \Sigma \frac{(p'_r - p_r)^2}{p_r} \\ &= n - 1 + NI_2. \end{aligned} \tag{27}$$

Again, let α_i be taken as a set of approximate values, put $\alpha'_i = \alpha_i + \Delta\alpha_i$, and consider the estimation of $\Delta\alpha_i$ by maximum likelihood. The function to be made a maximum is

$$L = \Pi(p'), \tag{28}$$

the product being over the observed values of x. In the ordinary case the terms in $\Delta\alpha_i \Delta\alpha_k$ in $\log L$ will be

$$\tfrac{1}{2}\Delta\alpha_i \Delta\alpha_k \Sigma \frac{\partial^2}{\partial\alpha_i \partial\alpha_k} \log p'. \tag{29}$$

The expectation of the coefficient, if the correct values are α_i, will be

$$\begin{aligned} \tfrac{1}{2}N \int p \frac{\partial^2}{\partial\alpha_i \partial\alpha_k} \log p \, dx &= \tfrac{1}{2}N \int \frac{\partial^2 p}{\partial\alpha_i \partial\alpha_k} dx - \tfrac{1}{2}N \int \frac{1}{p} \frac{\partial p}{\partial\alpha_i} \frac{\partial p}{\partial\alpha_k} dx \\ &= -\tfrac{1}{2}N g_{ik}, \end{aligned} \tag{30}$$

by differentiation and then integration by parts, since $\int p\,dx = 1$. Hence $-\frac{1}{2}Ng_{ik}\Delta\alpha_i\Delta\alpha_k$ is the expectation of the quadratic terms in the logarithm of the likelihood, given that the parameters have the value α'_i.

6. *Joint probabilities for several observations.* Let N observed values be derived independently from each law; their joint chances will be the products of their separate chances, and I_{1N}, I_{2N} can be defined as measures of the difference between the joint chances. Then

$$I_{1N} = \Sigma \ldots \Sigma (\sqrt{(\Pi p'_r)} - \sqrt{(\Pi p'_r)})^2. \tag{31}$$

Π refers to a product over the N values and the summations are over the values of r. Then

$$I_{1N} = 2 - 2\Sigma\Sigma(\sqrt{(\Pi p'_r)}\sqrt{(\Pi p_r)}) = 2 - 2(1 - \tfrac{1}{2}I_1)^N. \tag{32}$$

Hence $\frac{1}{N}\log(1 - \frac{1}{2}I_{1N})$ is independent of N. Also

$$I_{2N} = \Sigma \ldots \Sigma (\Pi p'_r - \Pi p)\log\frac{\Pi p'_r}{\Pi p_r} = \Sigma \ldots \Sigma(\Pi p'_r - \Pi p_r)\,\Sigma\log\frac{p'_r}{p_r}$$

$$= N\,\Sigma p'_r \log\frac{p'_r}{p_r} - N\,\Sigma p_r \log\frac{p'_r}{p_r} = NI_2 \tag{33}$$

and I_{2N}/N is independent of N.

Reference

Jeffreys, H. 1939 *Theory of probability*. Oxford, Clarendon Press.