Section 7: Foundations of probability and statistical inference

Jürgen HUMBURG, Schleißheimerstr. 173, D-8000 München 40, Germany West.

A NOVEL AXIOM OF INDUCTIVE LOGIC WHICH IMPLIES A RESTRICTION OF THE CARNAP PARAMETER λ .

НОВАЯ АХИОМА ИНДУКТИВНОЙ ЛОГИКИ, ИЗ КОТОРОЙ СЛЕДУЕТ ОГРАНИЧЕНИЕ КАРНАПСКОГО ПАРАМЕТРА λ .

Резюме: Мой доклад основывается на следующем результате из моей книги [6, 7]: Если мы приймём КАРНАПСКУЮ функцию c_λ для интерпретированной языковой системы L, то мы также должны принять, что для этой языковой системы L существует объективная вероятность p. – Тогда мы можем сравнить объктивную уверенность P конфидентного теста с соответствующей логической вероятностью в КАРНАПСКОЙ системе. Это сравнение даёт нижнюю и верхнюю границу для λ. – Это – значительный результат, поскольку он показывает, что цель Индуктивной Логики по крайней мере для простой функции c_λ достигаема. Индуктиная Логика имеет по-видимому большое значение для Искусственной Интеллигенции и Учащихся Машин (Learning Machines); в этом смысле имеет по-видимому даже простая функция c_λ практическое значение. В частности мы увидем, что правило ЛАПЛАСА, которое соответствует КАРНАПСКОЙ функции c^*, не действительно, если у нас более четырёх основных предикатов.

(1) **Summary:** My paper is based on the following result of my book [6, 7]: If we accept the CARNAP function c_λ for an interpreted language system L we accept also that for this system L there exists an *objective* probability p. – Now it is possible to compare the *objective* certainty P of the confidence test with the corresponding *logical* probability in the CARNAP system. This comparison gives us a lower and an upper bound of λ. – This result is important in so far as it shows that the aim of Inductive Logic is attainable at least for the simple function c_λ. Inductive Logic seems to be important for Artificial Intelligence or *Learning Machines*; already the simple function c_λ seems to have practical importance in this sense; in particular we will see that the LAPLACE rule which corresponds to CARNAP's function c^* is *not acceptable* if we have *more* than *four* logically disjoint fundamental predicates.

(2) Our considerations concern the system L_∞ of CARNAP; it is simple to relate this system to a set-theoretic field and then consider the σ-field F generated by this field. By the usual measure theory the functions m_λ , c_λ are uniquely extendible from L_∞ to F . – First I will briefly explain the result of my book [6, 7] why the adequacy of the functions m_λ , c_λ for a system L_∞ will imply that there exists an *objective probability* p for the related σ-field F:

Let $h_n(M)$ be the relative frequency of a molecular predicate M in a sequence of observations of the individuals a_1, a_2, a_3, ... Then we have for m-functions which are symmetric relative to the individuals the following theorem:

Theorem 1: There exists a function P(M) with the property

$$m(h_n(M) \to P(M)) = 1 .$$

We have logical probability one that the relative frequencies converge to a limit. The next theorem justifies why we are entitled to call this limit P *objective probability:*

Theorem 2: We define, based on P(M), a polynomial distribution **p** on F with the property $p(P_j a_i) := P(P_j)$. Let $m_{\lambda n}$ be the m_λ conditional on n observations, i.e.

$$m_{\lambda n}(A) = c_\lambda(A, P_{i_1}a_1 \ldots P_{i_n}a_n) ,$$

if the individuals are denumerated in the order of their observation; then we have: $m_\lambda(m_{\lambda n} \to \mathbf{p}) = 1$.

We see that the objective probability distribution **p** is the limit of the conditional probabilities after infinitely many observations. Of course we do not know **p**, but we know its inner structure – that it is a polynomial distribution. If we are only interested in the events described by *one* molecular predicate M and its negation, **p** is the *binomial distribution*. For more details see my book [6, 7].

(3) We may now compare the objective certainty of the confidence test with the corresponding logical probability and we will see that this gives a restriction of λ.

We first state the BERNOULLI inequality:

$$\mathbf{p}(|P(M) - \mathbf{h}_n(M)| \geq \varepsilon) \leq 1/4N\varepsilon^2 .$$

This inequality permits the confidence estimation: If we assume the hypothesis $|P(M) - h_n(M)| \leq \varepsilon$, the probability of an error is at most $1/4N\varepsilon^2$.

In statistical practice, one does not use the BERNOULLI inequality, because it is a very crude estimation. But to formulate a preliminary postulate, the bound for the confidence test given by the BERNOULLI inequality is useful.

The statement $|P(M) - h_n(M)| \leq \varepsilon$ describes an event of our σ-field F, because P(M) is defined as the limit of $h_n(M)$. Therefore the following logical probability is determind: $c(|P(M) - r/n| \leq \varepsilon , h_n(M) = r/n)$.

It seems plausible to demand that only such c-functions are admitted for which this logical probability is at least as great as the bound given by the BERNOULLI inequality for the above confidence test; hence we have the following preliminary postulate:

Postulate: Only c-functions are admitted for which, for all M, r, n, ε, the following inequality is valid:

$$c(|P(M) - r/n| \leq \varepsilon , h_n(M) = r/n) \geq 1 - 1/4n\varepsilon^2 .$$

(4) The distribution $m(P(M) \leq x)$ is given by the DE FINETTI representation of the function m; to evaluate the above postulate we need therefore the DE FINETTI representation of the functions m_λ. – Using the DE FINETTI representation the above postulate for the functions c_λ is equivalent to the following integral inequality:

$$\int_{r/n-\varepsilon}^{r/n+\varepsilon} x^{\alpha+r}(1-x)^{\beta+n-r}dx \Big/ \int_0^1 x^{\alpha+r}(1-x)^{\beta+n-r}dx \geq 1 - 1/4n\varepsilon^2 \cdot$$

$\alpha := (\lambda w/k) - 1$, $\beta := (\lambda(k-w)/k) - 1$, w:=logical width of M, k:=number of logically disjoint fundamental predicates.

The left side of the above inequality is the incomplete Beta-function which cannot be integrated analytically.

The evaluation of the inequality by computer gives the result:
$\lambda \leq 8$ for k=2, $\lambda \leq 4$ for k=3, $\lambda \leq 3$ for $4 \leq k \leq 6$, $\lambda \leq 2$ for $7 \leq k \leq 177$, $\lambda \leq 1.5$ for $k \geq 178$.

The violation of the above inequality for larger values of λ occurs in the domain $n \leq 20$. I have only tested integral values of λ, and once the value 1.5 . That the inequality is fulfilled for the given values has been tested up to n=1000. I am still seeking an analytical proof that the inequality is also fulfilled for the given values of λ for all large n; till yet I do not have such a proof.

(5) **The Axiom:** Now I will briefly develop the announced axiom, which is based on an *exact* calculation of the objective certainty of the confidence test. The exact confidence test is characterized as follows: Depending on the

occurring relative frequency r/n at n observations and the chosen objective certainty P we determine the numbers pu(r,n,P) , po(r,n,P) . The confidence test consits in the proposition that we have for the objective probability P(M): pu(r,n,P) ≤ P(M) ≤ po(r,n,P) . The bounds pu, po are determined as solutions of the following equations:

$$\sum_{i=r}^{n} \binom{n}{i} pu^i (1-pu)^{n-i} = E \;,\quad \sum_{i=0}^{r} \binom{n}{i} po^i (1-po)^{n-i} = E \;,\quad E:=(1-P)/2 \;.$$

These solutions we determine by computer. — This confidence test has the objective certainty P; that means: Independently of the value of P(M), we always have an objective probability at least as great as P that we will have: pu(r,n,P) ≤ P(M) ≤ po(r,n,P) ; formally:

$$p(\{r:\ pu(r,n,P) \le P(M) \le po(r,n,P)\}) \ge P \;.$$

For more details see [4].

We can now calculate the logical probability of this confidence test, i.e

$$c_\lambda(pu(r,n,P) \le P(M) \le po(r,n,P) \;,\; h_n(M)=r/n) \;,$$

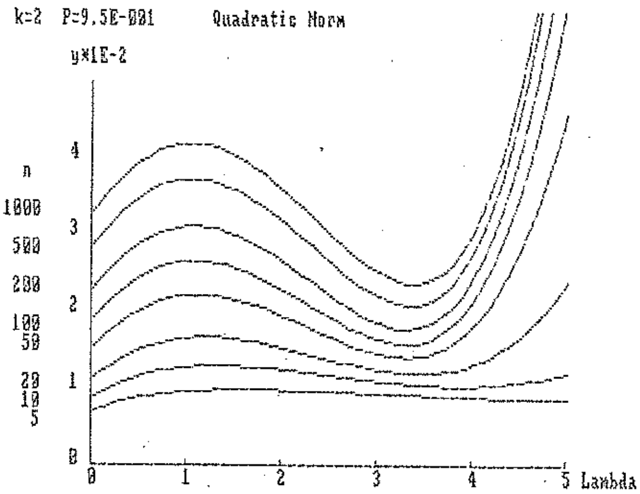
which is the integral of section (4) taken from pu to po instead of from $r/n-\varepsilon$ to $r/n+\varepsilon$.

We could now apply the same postulate as in section (3), namely that the logical probability of this confidence test is at least as great as the objective certainty and we would get by it sharper upper bounds for λ than we have in section (4), for instance the following results: $\lambda \le 2.5$ for k=2, $\lambda \le 1.2$ for $k \ge 100$.

Now it does not seem reasonable to demand that the logical probability of this confidence test be equal to or greater than the objective certainty P, although as mentioned this postulate is mathematically possible. It seems more convincing to demand that the *deviation* between these two probabilities be as small as possible. Therefore we look for that λ, for which the following quantity y is minimal:

$$y(n,P,\lambda) = \sum_{r=0}^{n} (P - c_\lambda(pu(r,n,P) \le P(M) \le po(r,n,P) \;,\; h_n(M)=r/n))^2 \;.$$

Having calculated this quantity y, I have discovered that we have independently of n, P, with 80% ≤ P ≤ 99.9%, nearly always the same minimum of y. The following graph shows for instance $y(n,P,\lambda)$ with y dependent on λ for k=2, P=95% and n=5 up to n=1000:



For k=2 and the different values of P, 80% ≤ P ≤ 99.9%, the above minima of y lie between 3.0 and 3.8. For k=100 and the logical width w=1 the minima lie between 1.5 and 1.9.

Because one may think that the use of the quadratic norm in calculating y as above is a bit arbitrary, I have calculated y also for the linear and the TSCHEBISCHEV norm with the following result: The minima are shifted only a bit (about 0.5) as compared with the minima of the quadratic norm; but the results are not so beautiful as for the quadratic norm; the minima for the different n oscillate a bit; they do not lie so impressively in one vertical line as for the quadratic norm.

Hence we present the following axiom:

Axiom: We presuppose that we have inductive probability functions which are symmetric relative to the individuals and depend on some parameters. Then we demand: Only those parameters are admissible for which the quadratic deviation of the logical probability from the objective certainty of the confidence test has a *minimum*.

(6) The application of the axiom to CARNAP's function c_λ gives, as already mentioned, for k=2 the following result: $3.0 \leq \lambda \leq 3.8$ for k=2.

If k is greater than 2 the situation is more complicated because we can form molecular predicates. The integrand in the integral of section (4) depends on the relative logical width w/k. If we have a molecular predicate M with w/k=1/2, then the same values of λ are optimal as those we have according to the axiom for k=2. Therefore we have a larger spread of admissible λ for k>2. For $k \geq 100$ we get, according to the result mentioned in section (5), the following spread of admissible λ: $1.5 \leq \lambda \leq 3.8$ for $k \geq 100$.

For k between 2 and 100 we have a continuous variation of the two bordering domains, which I will not present in this paper.

(7) It is obvious that the axiom is also applicable to the more complicated functions of CARNAP as well as to the HINTIKKA functions.

(8) For their contribution to this paper I thank Werner v. ZEPPELIN, Dr. Eckehart KÖHLER and my wife Angelika HUMBURG.

References

1. Carnap, R. and Stegmüller, W.: Induktive Logik und Wahrscheinlichkeit. *Wien 1958.*
2. Carnap, R. and Jeffrey, C.R.: Studies in Inductive Logic and Probability. Vol. I. *Berkeley. Los Angeles. London 1971.*
3. Jeffrey, C.R.: Studies in Inductive Logic and Probability. Vol.II. *Berkeley. Los Angeles. London 1980.*
4. Dinges, H. and Rost, H.: Prinzipien der Stochastik. *Stuttgart 1982.*
5. Hintikka, J.: A two-dimensional continuum of inductive methods. *Hintikka and Suppes (eds) 1966, no.7.*
6. Humburg, J.: Grundzüge zu einem neuen Aufbau der Wahrscheinlichkeits-theorie. *Frankfurt a. M., Bern 1981.*
7. Humburg, J.: Foundations of a New System of Probability Theory. *Topoi 5, 1; 1986.*