A two-dimensional continuum of a priori probability distributions on constituents

Kuipers, Theo A.F.

THEO A. F. KUIPERS

A TWO-DIMENSIONAL CONTINUUM OF A PRIORI PROBABILITY DISTRIBUTIONS ON CONSTITUENTS

I

Hintikka has defined a *one-dimensional* continuum of a priori probability distributions on constituents and has built on it a two-dimensional continuum of inductive methods with the aid of Carnap's λ-continuum ([1]), which plays also a fundamental role in his continuum of a priori distributions, and the formula of Bayes ([2]). Here a *two-dimensional* continuum of a priori probability distributions will be introduced. On its base a three-dimensional continuum of inductive methods can be constructed in the same way as Hintikka has done. The importance of the new continuum of a priori distributions, which is also based on Carnap's λ-continuum but in a completely different way, is that it leaves room for almost all kinds of a priori considerations, whereas Hintikka's continuum admits only considerations that lead to increasing probability for the constituents by increasing size.

In this article we will show furthermore (in section 10) the mathematical equivalence between a *generalized carnapian continuum* and the corresponding (*generalized*) so-called *Polya-distribution*.

II

There is given a set of K ($\geqslant 2$) so-called *Q-predicates*. They are supposed to be mutually exclusive and together exhaustive with respect to the universe considered. The *constituent-structure* S_w tells us that exactly w Q-predicates are exemplified in our universe (this concept was introduced by Carnap in [3]). A *constituent of size* w, abbreviated by C_w, tells us not only that there are w Q-predicates exemplified (i.e. that S_w holds) but it specifies also which ones. In the context of P (probability)--values for constituents we do not make any difference between constituents of the same size. Therefore C_w refers to an arbitrary constituent of size w and we have therefore: $P(S_w) = \binom{K}{w}\ P(C_w)$.

III

Hintikka defines $P(C_w)$ proportional to the probability, according to Carnap's λ-continuum that α individuals turn out to be compatible with C_w, hence $P(C_w)$ is equal to

$$\pi(\alpha, w\cdot\lambda/K)\sum_{v=0}^{K}\binom{K}{v}\pi(\alpha, v\cdot\lambda/K),\ \alpha \geqslant 0,\ \lambda > 0,$$

where $\pi(n, x) =_{\mathrm{df}} x\cdot(x+1)\ \ldots\ (x+n-1)$ if $n = 1, 2, 3, \ldots$ and $\pi(0, x) =_{\mathrm{df}} 1$.

It is easily seen that this distribution is directed to disorder, i.e. it satisfies the inequality $P(C_w) < P(C_v)$ if $0 \leqslant w < v \leqslant K$, as soon as $\alpha > 0$. If $\alpha = 0$ all constituents get the same probability, 2^{-K}.

IV

According to Hintikka a priori considerations of symmetry suggest that P has to satisfy the inequality just mentioned ([2], p. 117). It is not clear whether Hintikka would come to the same conclusion if he did not have assumed that the Q-predicates are defined in terms of a number of primitive, logically independent predicates. How this may be, we only assume that the Q-predicates are mutually exclusive and together exhaustive (on logical, linguistic or empirical grounds) with respect to our universe and we do not even assume that the universe is non--empty. To our opinion we have therefore to leave room for as many as possible different conclusions based on different a priori logical, statistical or metaphysical considerations.

V

In order to obtain such a rich continuum of distributions we start to look more closely to the relation between the universe of individuals, U, and our set of Q-predicates. The latter is in fact a second order universe, U_Q, with K elements and we have that for all $u \in U$ there is exactly one $Q \in U_Q$ such that Qu holds. Now we define two *metapredicates* E and N as follows: if $Q \in U_Q$ then

$$E(Q) \quad \textit{if and only if} \quad \exists_{u\in U}\, Qu$$

$$N(Q) \quad \textit{if and only if} \quad \sim \exists_{u\in U}\, Qu.$$

It is easily seen that these two predicates are mutually exclusive and together exhaustive with respect to U_Q, given U.

Suppose now that we were going to investigate the elements of U_Q, i.e. observe for each Q in U_Q, whether $E(Q)$ or $N(Q)$, then we would certainly use a generalized carnapian continuum for a family with two predicates in order to calculate a priori and a posteriori probabilities. By a *generalized carnapian continuum* we mean a λ-continuum with not necessarily equal logical weights for the members of the family of predicates and with room for 'negative induction', i.e. negative values for λ are permitted as far as they do not lead to negative 'probabilities'.

The foregoing includes, among others, an assignment of a priori probabilities to constituents and therefore to constituent-structures. However, in fact we are not going to investigate U_Q, at least not in a direct way, but U. But we need a priori probabilities for constituents and our proposal is to take the same values as those which we would take if we were going to investigate U_Q. Because there is no logical difference before we start our investigations this proposal seems to be reasonable.

VI

The assignment proposed in section V is the following. Assign weights p and q to E ('to be exemplified') and N ('not to be exemplified') resp. such that $p = 1-q$ and $0 < p < 1$, and therefore $0 < q < 1$. If exactly m_E of the Q-predicates $Q_1, Q_2, \ldots, Q_m$ $(m < K)$ turn out to be exemplified in U then the probability that Q_{m+1} will also be exemplified is, according to Carnap's λ-continuum (with $\lambda > 0$), equal to $(m_E+p\lambda)/(m+ +\lambda)$ and therefore we get, by applying the product rule, as the a priori probability of a constituent C_w:

$$\frac{p\lambda\cdot(p\lambda+1)\ldots(p\lambda+w-1)\cdot q\lambda(q\lambda+1)\ldots(q\lambda+K-w-1)}{\lambda(\lambda+1)(\lambda+2)\ldots(\lambda+K-1)}$$

which is equal to:

(1) $\qquad \pi(w, p\lambda)\cdot\pi(K-w, q\lambda)/\pi(K, \lambda),\ \lambda > 0.$

We have assumed for a moment that λ is positive, but (1) is defined for all real values of λ except for $\lambda = 0, -1, -2, \ldots, -K+1$. However, not all these permitted negative values for λ lead to non-negative values for (1) and this is necessary if we want to have a genuine probability distribution. Now if λ is negative, say $-\lambda = \gamma > 0$, we get positive values by (1) as soon as $p\gamma > K-1$ and $q\lambda > K-1$ (and therefore $\gamma > 2K-2$), which is easily seen by rewriting (1) in terms of γ,

$$\frac{p\gamma(p\gamma-1)\ldots(p\gamma-w+1)\cdot q\gamma(q\gamma-1)\ldots(q\gamma-K+w+1)}{\gamma(\gamma-1)\ldots(\gamma-K+1)},$$

or in terms of $\varrho(n, x) =_{\mathrm{df}} x\cdot(x-1)\ldots(x-n+1)$, if $n = 1, 2, 3, \ldots$ and $\varrho(0, x) =_{\mathrm{df}} 1$.

(2) $\quad \varrho(w, p\gamma)\cdot\varrho(K-w, q\gamma)/\varrho(K, \gamma), \quad -\lambda = \gamma > 0.$

As we will see at the end of section 8 there are included in (2) non-negative assignments for which $p\gamma$ and/or $q\gamma$ do not exceed $K-1$.

VII

It is possible to reformulate (1) and (2) in terms of the *gamma-function* and in terms of the *beta-function*. If $\lambda > 0$ then (1) is equivalent to

$$(1^*) \qquad \frac{\Gamma(p\lambda+w)\cdot\Gamma(q\lambda+K-w)}{\Gamma(\lambda+K)} \Big/ \frac{\Gamma(p\lambda)\cdot\Gamma(q\lambda)}{\Gamma(\lambda)}$$

which is also equivalent to

$$(1^{**}) \qquad B(p\lambda+w, q\lambda+K-w)/B(p\lambda, q\lambda).$$

If $0 > \lambda = -\gamma$ and $p\gamma > K-1$ and $q\gamma > K-1$ then (2) is equivalent to

$$(2^*) \qquad \frac{(\gamma+1)\Gamma(p\gamma+1)\cdot\Gamma(q\gamma+1)}{\Gamma(\gamma+2)} \Big/ \frac{(\lambda-K+1)\Gamma(p\gamma-w+1)\cdot\Gamma(q\gamma-K+w+1)}{\Gamma(\gamma-K+2)}$$

which is also equivalent to

$$(2^{**}) \qquad (\gamma+1)\cdot B(p\gamma+1, q\gamma+1)/(\gamma-K+1)B(p\gamma-w+1, q\gamma-K+w+1).$$

In order to obtain these alternative formulations one only needs the following theorem

$$\Gamma(x+1) = x\cdot\Gamma(x), x > 0$$

and the definition of the beta-function:

$$B(x, y) =_{df} \Gamma(x)\Gamma(y)/\Gamma(x+y), \ x > 0, \ y > 0.$$

VIII

Now we will consider some particular symmetric distributions ($p = q$) that result if λ assumes a particular value. Note first that $p = q$ implies $P(C_w) = P(C_{K-w})$ ($w = 0, 1, 2, \ldots, K$).

$\lambda \to \pm\infty$ $\quad P(C_w) = 2^{-K}$, $w = 0, 1, 2, \ldots, K$.
here all constituents get the same probability and therefore we have that the constituent-structure of size w gets a probability exactly proportional to the number of constituents of size w; this assignment corresponds to *Hintikka's assignment* (see section 3) with $\alpha = 0$.

$\lambda = 2$: $\quad P(S_w) = (K+1)^{-1}$, $w = 0, 1, 2, \ldots, K$.
here all constituent-structures get the same probability; this assignment has been suggested by Carnap in [3].

$-\lambda = \gamma = 2K$: $\quad P(C_w) = \binom{K}{w} \Big/ \binom{2K}{K}$, $w = 0, 1, \ldots, K$.
here the probability of a constituent of size w is exactly proportional to the number of constituents with size w.

The last assignment ($\gamma = 2K$) is an extreme case of the following two generalizations for which a satisfactory interpretation has still to be found:

$-\lambda = \gamma = 2rK$:
$r = 1, 2, 3, \ldots$ $\quad P(S_w) = \binom{rK}{w}\binom{rK}{K-w} \Big/ \binom{2rK}{K}$, $w = 0, 1, 2, \ldots, K$.

$-\lambda = \gamma = 2K - 2\dot{Z}$: $\quad P(C_{\dot{Z}-i}) = 0 = P(C_{K-(\dot{Z}-i)})$, $i = 1, 2, \ldots, \dot{Z}$

$K = 2L$ *or* $2L+1$, $\quad$ *and*

$\dot{Z} = 0, 1, 2, \ldots, L.$ $\quad P(C_{\dot{Z}+i}) = P(C_{K-(\dot{Z}+i)}) =$

$$= \binom{K-2\dot{Z}}{i} \Big/ \binom{2K-2\dot{Z}}{K-\dot{Z}},$$

$$i = 0, 1, 2, \ldots, L-\dot{Z}.$$

IX

In the previous section we have seen that the proposed continuum of distributions includes a great variety of symmetric distributions. As to asymmetric distributions we will restrict our attention to the case $\lambda > 0$. Our aim is to indicate that under certain restrictions on the possible values of the parameters p and λ the a priori distribution satisfies certain relations of monotony which may be seen as reflecting specific a priori considerations.

We will only consider $p > \frac{1}{2}$ because all relations to be given change in the opposite one if $p < \frac{1}{2}$ and if p and q are interchanged in the conditions.

I *for all* $w < K-w$: $P(C_w) < P(C_{K-w})$ *and therefore* $P(S_w) < P(S_{K-w})$.

II *if* $\lambda(p-q) > K-1$ *then for all* $w < v$: $P(C_w) < P(C_v)$ (as we have seen in section 3 this relation holds generally in Hintikka's distribution as soon as $\alpha > 0$)

III *if* $q\lambda - 1 < 0 < p\lambda - 1$ *or* $0 \leqslant q\lambda - 1 < (p\lambda - 1)/K$ *or* $0 \leqslant 1 - p\lambda < (1 - q\lambda)/K$ *then for all* $w < v$: $P(S_w) < P(S_v)$

IV *if* $q\,K < \lambda q(p-q) + q < p$ *or* $K - p\lambda \leqslant Kq\lambda - p\lambda < K-1 < \lambda(p-q)$ *then for all* $w < v$: $P(C_w) < P(C_v)$ *and* $P(S_w) < P(S_v)$ (the first disjunctive condition combines the (sole) condition of II with the first of III, the second combines that of II with the second of III; the condition of II is incompatible with the third of III)

The proofs for all these theorems are elementary: the simplest way to start is to write $P(C_w) < P(C_{w+1})$ or $P(S_w) < P(S_{w+1})$ in terms of the gamma-function (see section 7).

From these theorems we may conclude that the parameters p (if $\neq q$) and λ determine in a complex way the character of an asymmetric assignment. In section 11 we will see that $p \cdot K$ is the a priori expectation of the size of the true constituent whereas λ determines the a priori correlation coefficient between the possibilities of being exemplified of two different Q-predicates.

X

In section V we have defined a generalized carnapian continuum as a λ-continuum with suitable real values for λ and a logical weight for each Q-predicate. In order to give some a priori and limit properties of our particular continuum of distributions we prove in this section the mathematical equivalence between such a generalized carnapian continuum and the (generalized) so-called *Polya-distribution* ([4], [5], [6]).

Suppose we have an urn containing N balls such that there are N_i with property (colour) $Q_i(i = 1, 2, \ldots, K^*)$, $\sum_{i=1}^{K^*} N_i = N$. Let p_i be equal to N_i/N, $\sum_{i=1}^{K^*} p_i = 1$.

Our experiments consist out of succesive random selections of one ball with replacement of $\Delta+1$ balls with the same colour. Here Δ is some integer; if Δ is negative then there are some obvious restrictions to the total number of experiments. Note that if $\Delta = 0$ we have drawings with replacement and that if $\Delta = -1$ we have drawings without replacement. Let N/Δ be equal to λ.

The result of the first m trials is indicated by e_m. Then e_m specifies for each i how many trials have resulted in Q_i, say m_i. Let h_i indicate the hypothesis that the next (the $m+1-$th) trial will result in Q_i. Now it is easy to prove that the objective conditional probability of h_i, given e_m, is equal to

$$\frac{m_i \cdot \Delta + N_i}{m \cdot \Delta + N} = \frac{m_i + N_i/\Delta}{m + N/\Delta} = \frac{m_i + p_i \lambda}{m + \lambda},$$

which is equivalent to the corresponding 'special value' in a generalized carnapian continuum.

By applying the product rule we obtain as objective unconditional probability of e_m: $\prod_{i=1}^{K^*} \pi(m_i, p_i \lambda)/\pi(m, \lambda)$
and this is equivalent to the a priori probability of the state-description corresponding to e_m in a generalized carnapian continuum.

XI

Friedman has given in [6], a number of a priori properties of the (actual) Polya distribution in which $K^* = 2$ and this corresponds therefore to our continuum defined by (1). Let x_m be a random variable that equals 1 if Q_m tourns out to be exemplified and 0 if not. Let the random variable X be defined as $x_1+x_2+ \dots +x_K$. In the following E indicates the (mathematical) expectation operator.

According to [6] we have the following a priori properties of our continuum of distributions defined by (1):

$$\mathrm{E}(x_m) = p$$

i.e. *the a priori expectation of* x_m, *or the a priori probability that* Q_m *appears to be exemplified*; note the independence (at least explicitely) *of* K, λ and m.

$$\mathrm{E}(X) = Kp = \sum_{w=0}^{K} wP(S_w)$$

i.e. *the a priori expectation of* X, *or the a priori expectation of the size of the actual constituent*: E (*size*); note the independence of λ.

$$\mathrm{E}(x_m - \mathrm{E}(x_m))^2 = p \cdot q$$

i.e. *the a priori variance of* x_m;
note the independence of K, λ and m.

$$\mathrm{E}\{X-\mathrm{E}(X)\}^2 = Kpq(\lambda+K)/(\lambda+1) =$$

$$= \sum_{w=0}^{K} P(S_w)\left\{w - \sum_{w=0}^{K} wP(S_w)\right\}^2$$

i.e. *the a priori variance of* X: VAR (*size*);
note the independence on K, λ *and* p.

$$\mathrm{E}\{(x_m - \mathrm{E}(x_m))\,(x_n - \mathrm{E}(x_n))\} = p \cdot q/(\lambda+1)$$

i.e. *the a priori covariance of* x_m *and* x_n $(m \neq n)$.

$$\frac{\mathrm{E}\{(x_m - \mathrm{E}(x_m))\,(x_n - \mathrm{E}(x_n))\}}{\sqrt{\mathrm{E}(x_m - \mathrm{E}(x_m))^2 \cdot \mathrm{E}(x_n - \mathrm{E}(x_n))^2}} = 1/(\lambda+1) =_{\mathrm{df}} c.c.$$

i.e. *the a priori correlation coefficient between* x_m *and* x_n $(m \neq n)$.

$$\mathrm{E}(X - \mathrm{E}(X))^2 \Big/ \sum_{m=1}^{K} \mathrm{E}(x_m - \mathrm{E}(x_m))^2 = (\lambda+K)/(\lambda+1) =_{\mathrm{df}} Q^2$$

i.e. *the a priori dispersion coefficient of* X *in relation to the* x_m*'s*.

XII

In [5] Polya has presented the limit-distributions of the Polyadistribution. Here we are only interested in discrete limit-distributions.

K *finite*, $\lambda \to \infty, 0 < p < 1$

$$P(S_w) = \binom{K}{w} p^w q^{K-w} \qquad \textit{(binominal distribution)}$$

$\mathrm{E}(\textit{size}) = Kp$; VAR $(\textit{size}) = Kpq$; $Q^2 = 1$; $c.c. = 0$

$K \to \infty, \lambda \to \infty, p \to 0$, such that $Kp \to h > 0$ and $K/\lambda \to 0$

$$P(S_w) = e^{-h} h^{w/w}! \qquad \textit{(Poisson distribution)}$$

$\mathrm{E}\,(\textit{size}) = h$; VAR $(\textit{size}) = h$; $Q^2 = 1$; $c.c. = 0$

$K \to \infty, \lambda \to \infty, p \to 0$, such that $Kp \to h > 0$ and $K/\lambda \to \to d > 0$

$$P(S_w) = \binom{h/d + w - 1}{w} \left(\frac{1}{1+d}\right)^{h/d} \left(\frac{d}{1+d}\right)^w$$

(a special case of the negative binominal distribution)

$\mathrm{E}\,(\textit{size}) = h$; VAR $(\textit{size}) = h(1+d)$; $Q^2 = 1+d$; $c.c. = 0$

From the last two results we may conclude that our proposal leaves room for a denumerably infinite number of Q-predicates. In this case we get zero probability for all constituents, but positive probabilities for constituent-structures.

XIII

In this section we will present the a posteriori distribution on constituent-structures that results if we substitute our a priori continuum of distributions in *Hintikka's frame-work* ([2]) based on Bayes' formula. Suppose we are investigating the individuals of U one by one, with or without replacement. Let e_n be a state-description of n investigations, according to which there are n_i exemplifications of Q_i and c is the number of Q_i' s such that $n_i > 0$. In the following we assume that e_n is compatible with C_w in the expression $P(e_n/C_w)$.

According to Bayes' formula the a posteriori probability of S_w ($w \geqslant c$), given e_n:

$$P(S_w/e_n) = \binom{K-c}{w-c} P(C_w)P(e_n/C_w) \Bigg| \Bigg/ \sum_{v=c}^{K} \binom{K-c}{v-c} P(C_v)P(e_n/C_v).$$

In the same way as Hintikka has done we take $P(e_n/C_w)$ equal to the probability of e_n in a carnapian continuum with w Q-predicates. Because the parameter λ occurs already in our a priori distribution we replace here however $\lambda(\lambda(w))$ by $\eta(\eta(w))$. Then we get

$$P(e_n/C_w) = \prod_{i-1}^{K} \pi(n_i, \eta(w)/w)/\pi(n, \eta(w))$$

For convenience we strict our attention to the case that $\eta(w) = w$. Then $P(e_n/C_w)$ is equal to

$$\prod_{i=1}^{K} n_i!/\pi(n, w) \tag{4}$$

Replacing $P(C_w)$ by (1) (section 6) and $P(e_n/C_w)$ by (4) in (3) leads to:

$$P(S_w/e_n) = \frac{\binom{K-c}{w-c} \pi(w, p\lambda)\cdot\pi(K-w, q\lambda)/\pi(n, w)}{\sum_{v=c}^{K} \binom{K-c}{v-c} \pi(v, p\lambda)\cdot\pi(K-v, q\lambda)/\pi(n, v)}$$

In case of symmetric a priori distributions $\left(p = q = \frac{1}{2}\right)$ we get e.g.

$$\textit{if } \lambda \to \pm\infty: \quad P(S_w/e_n) = \binom{K}{w}\binom{w}{c}\Big/\pi(n, w)\Big/ \left\{\sum_{v=c}^{K}\binom{K}{v}\binom{v}{c}/\pi(n, v)\right\}$$

$$\textit{if } \lambda = 2: \quad P(S_w/e_n) = \binom{w}{c}\Big/\pi(n, w)\Big/\left\{\sum_{v=c}^{K}\binom{v}{c}\Big/\pi(n, v)\right\},$$

$$\textit{note that } P(S_w/e_1) = 1/K(w = 1, 2, \ldots, K)$$

$$\textit{if } \lambda = -2K: \quad P(S_w/e_n) = \binom{K-c}{w-c}\binom{K}{w}\Big/\pi(n, w)\Big/ \left\{\sum_{w=c}^{K}\binom{K-c}{v-c}\binom{K}{v}\Big/\pi(n, v)\right\}$$

XIV

In summaiy we hope to have shown that the proposed two-dimensional continuum of a priori distributions leaves indeed room for many kinds of considerations that we may have in a particular inductive situation. However, the problem of finding an acceptable a priori distribution is not a simple question of deductive nature, for any distribution includes e.g. an a priori expectation of the size of the true constituent and it is clear that such an expectation cannot be deductively derived from what we actually know about the universe of individuals and the set of Q-predicates if we are confronted with an inductive situation.

BIBLIOGRAPHY

[1] Carnap, R., *The Continuum of Inductive Methods*, The University of Chicago Press, Chicago 1952.

[2] Hintikka, J., 'A two-dimensional continuum of inductive methods', in *Aspects of Inductive Logic* (ed. by J. Hintikka and P. Suppes), North-Holland Publ. Comp. Amsterdam 1966, pp. 113–132.

[3] Carnap, R., 'The concept of constituent-structure', in *The Problem of Inductive Logic* (ed. by I. Lakatos), North-Holland Publ. Comp., Amsterdam 1968, pp. 218–220.

[4] Eggenberger, F. and Polya, G., 'Ueber die Statistik verketterter Vorgänge'; *Zeitschrift für Angewandte Mathematik und Mechanik*, vol. **3**, pp. 279–289 (1923).

[5] Polya, G., 'Sur quelques points de la théorie des probabilités' *Annales de l'Institut Henri Poincaré*, vol. **1**, pp. 117–161 (1931).

[6] Friedman, B., 'A simple urn model'; *Communications on Pure and Applied Mathematics*, vol. **2** pp. 59–70 (1949).