Z-style notation for Probabilities

Maarten M. Fokkinga
DB group, dept INF, fac EWI, University of Twente

Version of April 13, 2006, 20:17

Abstract. A notation for probabilities is proposed that differs from the traditional, conventional notation by making explicit the domains and bound variables involved. The notation borrows from the Z notation, and lends itself well to calculational manipulations, with a smooth transition back and forth to set and predicate notation.

1 Introduction. The notation commonly used in applied probability theory suffers from two drawbacks: the domain of discourse is left implicit, and consequently in predicates the argument is left implicit. To say it in a crude way, the formulas have no meaning without a little verbal story along with them. As a consequence, it is hard to do machine assisted formal calculations (as striven for in, for example, transformational programming [1, 11, 2, 3, 4, 8, 7, 12, 10, 11]); it is simply too hard to feed the machine with the little verbal stories that define the semantics of various sub-expressions. This note presents a possible improvement.

The proposal is not meant to replace existing notion; the current notation has proved its functioning over the years. Rather, the new notation may be beneficial in an educational setting, and every now and then it may help to express one's ideas in a clear and precise way as a stepping-stone to achieve a convenient conventional formulation.

2 The notation. The proposal is fully in the style of the Z notation [15], a notation designed for large scale formal specifications, supported by a range of tools [5] (syntax checker, type checker, pretty printer, proof checker, theorem prover). Below we give a list of some notations available in Z — with in the last line the proposed notation for probabilities. The list alone already clearly shows the systematic approach in the choice of the notation; in the subsequent paragraphs we shall show the advantage of this systematic notation in formal manipulations. In the list and the sequel, we shall use letter D for arbitrary *declarations*, letters P, Q for arbitrary *predicates*, and letter E for arbitrary *expressions*. We do not elaborate the syntax of these categories, but instead leave them to the imagination of the reader. The symbols $|$ and $\bullet$ are *no* operator symbols, and have *no* meaning by themselves; they merely separate the three parts (namely D, P, E, and D, P, Q respectively).

Here is the list, followed by a discussion of each line:

notation	concrete example	=	value
$\{D \mid P \bullet E\}$	$\{x, y : \mathbb{N} \mid 2y = x < 5 \bullet x{+}10\}$	=	$\{10, 12, 14\}$
$(\lambda\, D \mid P \bullet E)$	$(\lambda\, x, y : \mathbb{N} \mid 2y = x < 5 \bullet x{+}10)$	=	$\{(0{,}0){\mapsto}10,\ (2{,}1){\mapsto}12,\ (4{,}2){\mapsto}14\}$
$(\mu\, D \mid P \bullet E)$	$(\mu\, x, y : \mathbb{N} \mid 0 < 2y = x < 4 \bullet x{+}10)$	=	12
$(\forall\, D \mid P \bullet Q)$	$(\forall\, x, y : \mathbb{N} \mid 2y = x < 5 \bullet x{+}10 < 13)$	=	*false*
$(\exists\, D \mid P \bullet Q)$	$(\exists\, x, y : \mathbb{N} \mid 2y = x < 5 \bullet x{+}10 < 13)$	=	*true*
$(\mathcal{P} D \mid P \bullet Q)$	$(\mathcal{P} x, y : \mathbb{N} \mid 2y = x < 5 \bullet x{+}10 < 13)$	=	0.66666...

In the sequel we shall not use (and therefore not explain) the λ-form and the μ-form; they are given here only to demonstrate the variety of forms involving a '$D \mid P \bullet$ ...'-part.

The form $\{D \mid P \bullet E\}$ denotes a set; the values of expression E constitute the members of the set, where the variables in E range over their domains as specified in declaration D — but only as far as predicate P is true. A more traditional notation for the example set is "$\{x{+}10 \mid x, y{\in}\mathbb{N} \wedge 2y{=}x{<}5\}$", in which, alas, the *declaration* '$x, y{\in}\mathbb{N}$' is syntactically indistinguishable from the constraining *predicate* '$2y{=}x{<}5$'.

The form $(\forall\, D \mid P \bullet Q)$ is the familiar universal quantification; it denotes the claim that for all conglomerates of variables described by D and satisfying constraint P, predicate Q holds true. The form $(\exists\, D \mid P \bullet Q)$ is its dual: the existential quantification.

The form $(\mathcal{P} D \mid P \bullet Q)$ is the proposed notation for probabilities; it denotes the probability that an arbitrary conglomerate of variables drawn from D satisfies predicate Q, given that the variables already satisfy P. We shall elaborate upon this notation later.

In all forms, when P is *true*, it may be omitted — together with the preceding symbol $|$. Also, in the set notation, when E is exactly the conglomerate of variables declared by D, it may be omitted together with the preceding symbol $\bullet$. Thus $\{x, y : \mathbb{N} \mid 2y{=}x{<}5\}$ stands for $\{x, y : \mathbb{N} \mid 2y{=}x{<}5 \bullet (x, y)\}$, which equals $\{(1, 0), (2, 1), (4, 2)\}$. This abbreviation could also be done for the λ-form and μ-form, but it is not customary to do so. Similarly, it is not customary to omit '$\bullet\ Q$' when Q is *true*, but there is no formal objection to it. It is customary to omit outer parentheses when no confusion can result.

3 Formal manipulations. Several laws hold for forms involving '$D \mid P \bullet$ ...'. These laws facilitate formal manipulations. Not only is it easy for humans to apply those rules, but also a machine can easily support them since all ingredients are available in the notation — there is no need for an informal verbal explanations along with the formulas. By way of illustration we give only a few of these laws. Each line is discussed below:

$$
\begin{array}{lcll}
(\exists\, D \mid P \bullet Q) & = & (\exists\, D \bullet P \wedge Q) & \\
(\forall\, D \mid P \bullet Q) & = & (\forall\, D \bullet P \Rightarrow Q) & \\
\neg\, (\exists\, D \bullet P) & = & (\forall\, D \bullet \neg\, P) & \\
\neg\, (\forall\, D \bullet P) & = & (\exists\, D \bullet \neg\, P) & \\
\neg\, (\exists\, D \mid P \bullet Q) & = & (\forall\, D \mid P \bullet \neg\, Q) & \\
\neg\, (\forall\, D \mid P \bullet Q) & = & (\exists\, D \mid P \bullet \neg\, Q) & \\
\{D \mid (\exists\, D' \mid P' \bullet Q') \wedge P \bullet E\} & = & \{D; D' \mid P' \wedge Q' \wedge P \bullet E\} & \\
\quad D \mid (\exists\, D' \mid P' \bullet Q') \wedge \ldots & = & \quad D; D' \mid P' \wedge Q' \wedge \ldots & \text{(Shunting)}
\end{array}
$$

The first two lines show how to eliminate the "constraint" P, and obtain a more traditional form having no constraint part. Note that the elimination gives rise to a *conjunction* in case of existential quantification, and to an *implication* in case of a universal quantification.

> Remark. Newcomers to the field of predicate logic often erroneously write '$\forall D \bullet P \wedge Q$' (similarly to their correct use of the form $\exists D \bullet P \wedge Q$) when they actually mean $\forall D \bullet P \Rightarrow Q$. The reason is that they view P as a constraint upon the the domain of discourse D, and therefore treat P the same way in both quantifications. However, that is not possible with the two-part traditional notation '$D \bullet$ ...' in contrast to the three-part Z notation '$D \mid P \bullet$ ...'. Thus the three-part notation '$D \mid P \bullet$ ...' is practically appealing.

The next two lines show, as a refresher, the familiar duality between universal and existential quantification. No surprise here.

The fourth and fifth line show the beauty of the '$D \mid P \bullet$ ...' notation: the duality between universal and existential quantification holds even in the presence of a constraint P. In view of the elimination given in the first two lines, this might come as a surprise! In practice, it is often the case that in '$(\exists D \bullet P \wedge Q)$' and '$(\forall D \bullet P \Rightarrow Q)$' the parts P play the role of an additional constraint on the domain of interest D. By making that role explicit, and writing $(\exists D \mid P \bullet Q)$ and $(\forall D \mid P \bullet Q)$, respectively, we see that the duality respects those roles!

The one-but-last line shows one example of the interactions between various forms that involve a '$D \mid P \bullet$ ...'; it assumes that the variables declared in D' do not occur free in P and ... . Thanks to the consistency of the Z-notation the declaration part D' of the existential quantification can be taken over into the declaration part of a set notation — without any change. Apart from this syntactic convenience, the line is also semantically interesting; we urge the reader to check (and understand) the equation. In fact, the rewriting is valid not only in a set context, but also in an arbitrary context:

$$D \mid (\exists D' \mid P' \bullet Q') \wedge \ldots \quad = \quad D; D' \mid P' \wedge Q' \wedge \ldots \qquad \text{(Shunting)}$$

We will apply the rule in a $\mathcal{P}$ context (page 8), thus showing the smooth transition between probability and set/predicate notation. There are some more such laws about the interaction between the D-part and P-part; e.g.:

$$x : S \mid x \in T \wedge \ldots \quad = \quad x : S \cap T \mid \ldots$$

Much more can be said about the Z notation, but this is not the place to do so. The interested reader may consult the Z literature [5] and Dijkstra [6] who has been a co-initiator of the three-part '$D \mid P \bullet$ ...' notation.

4 Example: calculations. Recall our proposed notation for probabilities:

$$(\mathcal{P} D \mid P \bullet Q)$$

It denotes the probability that Q holds for a random draw from D that satisfies P. The traditional notation is $\mathcal{P}(Q \mid P)$, thus leaving the domain implicit and making it impossible to refer in P and Q to the variables declared in D. Unfortunately, the places of P and Q are reversed between the new and the traditional notation.

In the current paragraph we do some calculations with this notation without referring to concrete examples; in the following paragraphs we'll actually use the notation for concrete examples.

When all sets involved are finite, and the probability distributions are uniform, we may take a frequentist view of probability, and put:

$$(\mathcal{P}\ D \mid P \bullet Q) \quad = \quad \frac{\#\{D \mid P \wedge Q\}}{\#\{D \mid P\}} \qquad \text{(Freq)}$$

Having done so, we are able to derive several theorems that are commonly taken as axioms about $\mathcal{P}$. We are the first to admit that these theorems are worthless when the above equation is false (because the probability distribution is not uniform) or doesn't make sense (because the sets are infinite); in that case we can still take the formulas below as axioms. However, we are mainly interested in the proofs, because *these show how our proposed notation nicely interacts with the set and predicate notation.* Here is a theorem that we shall use later on (page 6):

Let P_i, Q_i be predicates that do not use variables from D_j, for $j \neq i$. Then:

$$(\mathcal{P} D_1;\, D_2 \mid P_1 \wedge P_2 \bullet Q_1 \wedge Q_2) = (\mathcal{P} D_1 \mid P_1 \bullet Q_1) \times (\mathcal{P} D_2 \mid P_2 \bullet Q_2) \qquad \text{(Independence)}$$

Proof (using $\times$ for both the number product and cross product):

$$\begin{aligned}
&(\mathcal{P} D_1;\, D_2 \mid P_1 \wedge P_2 \bullet Q_1 \wedge Q_2) \\
&= \#\{D_1;\, D_2 \mid P_1 \wedge P_2 \wedge Q_1 \wedge Q_2\} \ / \ \#\{D_1;\, D_2 \mid P_1 \wedge P_2\} && \text{[Freq]} \\
&= \#(\{D_1 \mid P_1 \wedge Q_1\} \times \{D_2 \mid P_2 \wedge Q_2\}) \ / \ \#(\{D_1 \mid P_1\} \times \{D_2 \mid P_2\}) && \text{[set calc.; premise]} \\
&= (\#\{D_1 \mid P_1 \wedge Q_1\} \times \#\{D_2 \mid P_2 \wedge Q_2\}) \ / \ (\#\{D_1 \mid P_1\} \times \#\{D_2 \mid P_2\}) && \text{[set calc.]} \\
&= (\#\{D_1 \mid P_1 \wedge Q_1\} \ / \ \#\{D_1 \mid P_1\}) \ \times \ (\#\{D_2 \mid P_2 \wedge Q_2\} \ / \ \#\{D_2 \mid P_2\}) && \text{[nbr calc.]} \\
&= (\mathcal{P} D_1 \mid P_1 \bullet Q_1) \times (\mathcal{P} D_2 \mid P_2 \bullet Q_2) && \text{[Freq]}
\end{aligned}$$

Many more properties can be proved in this *algebraic* way: decomposing the expression and composing it in another way while preserving the semantics — and in this case there is also a smooth switch between probability and set notation.

Since they are used in the sequel, we mention two further laws but leave the simple algebraic proofs to the industrious reader:

$$\begin{aligned}
&(\mathcal{P} D \mid P \bullet Q_1 \vee Q_2) && \text{(Distribution)} \\
&\quad = \quad (\mathcal{P} D \mid P \bullet Q_1) + (\mathcal{P} D \mid P \bullet Q_2) - (\mathcal{P} D \mid P \bullet Q_1 \wedge Q_2) \\
&\quad = \quad (\mathcal{P} D \mid P \bullet Q_1) + (\mathcal{P} D \mid P \bullet Q_2) && \text{if } \forall\, D \mid P \bullet \neg\, (Q_1 \wedge Q_2)
\end{aligned}$$

And the divide and conquer law:

$$\begin{aligned}
&(\forall\, D \bullet P_1 \neq P_2) \ \Rightarrow && \text{(Divide\&Conquer)} \\
&(\mathcal{P} D \bullet Q) \quad = \quad (\mathcal{P} D \bullet P_1)(\mathcal{P} D \mid P_1 \bullet Q) + (\mathcal{P} D \bullet P_2)(\mathcal{P} D \mid P_2 \bullet Q)
\end{aligned}$$

Note that $P_1 \neq P_2$ means that P_1, P_2 are each others negation, that is, it is equivalent to $(P_1 \vee P_2) \wedge \neg\, (P_1 \wedge P_2)$, also known as exclusive-or.

5 Example: information retrieval. We set out to reformulate part of Section 2.3.3 of Hiemstra's PhD thesis[9] in our style and notation. We shall also make a comparison between our and his notation.

The scene is information retrieval. Here is a rough introduction. A set *Doc* of documents is given, and a user is in need for some relevant documents. The user poses a query to the system, and it is the systems task to rank the documents in order of increasing probability of being relevant (and then show the top-ranked documents to the user); the documents that contain the same "query terms" should be ranked the same. We bypass the problem of the way in which the set of relevant documents can be made known to the system.

With this introduction in mind, the following is our formalization. First, a set *Doc* of documents is postulated. Next, in order to avoid defining the internal structure of documents and queries, we postulate for each query q an equivalence relation $\approx_q$ on *Doc*, with the following interpretation:

$$d \approx_q d' \quad = \quad \text{“}d \text{ contains the same query terms of } q \text{ as } d' \text{ does”}$$

Now, the ranking function $rnk_{q,R}$ related to a query q and a set $R \subseteq Doc$ of "relevant" documents, is defined as follows:

$$rnk_{q,R}(d_0) \quad = \quad (\mathcal{P}\, d : Doc \mid d \approx_q d_0 \bullet d \in R)$$

In words: document d_0 is ranked with the probability that an arbitrary document with the same terms as d_0, happens to be relevant. (Much more can be said about the alternatives and variations for *rnk*, but this note is not the place to do so.)

Comparison with Hiemstra's formulation.
Hiemstra [9] formulates the ranking value as follows (the main ingredient of equation (2.8) in [9, page 19], see also [9, equation (2.10)]):

$$\mathcal{P}(L{=}1 \mid D_1, \cdots, D_n)$$

This is all the accompanying explanation:

- The domain of discourse is a set of documents; no further formalization is given. A document may be indexed with a query term, meaning that the query term occurs in the document. The query under consideration is supposed to have n query terms.
- Citing from page 19, line 10 from the bottom:
  Let L be the random variable "document is relevant" with a binary sample space {0,1}, 1 indicating relevance and 0 non-relevance.
- Citing again, line 8 from the bottom:
  Let D_k $(1 \leq k \leq n)$ be a random variable indicating "document belongs to the subset indexed with the k-th query term" with a binary sample space $\{0, 1\}$.
- Rephrasing line 6 from the bottom:
  A document satisfying a particular state of $D_1, \ldots, D_n$ is assigned the value given above (with the same state).

Although it is perfectly possible to do *numeric* calculations with such a probability notation, it is hard to take the ingredients of this expression and use them in set or predicate notation: the conventions and semantics of phrases like '$L = 1$' and 'D_k' are just too far away from the conventions and semantics of set and predicate notation. Probability theory and set theory use quite different languages, here, whereas in our opinion that is not at all necessary.

6 Relaxed notation. There is no objection against abbreviations in order to make the notation more compact. For example, when the discussion is about documents from the set *Doc* for pages and pages, then we may convene to omit the indication '$: Doc$' from the declaration part, thus writing:

$$(\mathcal{P}d \mid d \approx_q d_0 \bullet d \in R)$$

Going one step further, we may define $D(d) \Leftrightarrow d{\approx_q}d_0$ and $L(d) \Leftrightarrow d{\in}R$, and then write:

$$(\mathcal{P}d \mid D(d) \bullet L(d))$$

As another convention we might now abbreviate this to:

$$(\mathcal{P} \mid D \bullet L)$$

This comes very close to Hiemstra's notation $\mathcal{P}(L{=}\,1 \mid D_{1,\ldots,n})$. The point is that the semantics is still given by an expression of the form $(\mathcal{P}D \mid P \bullet Q)$, and if the need arises we can fall back to that form. Even with the above abbreviations we do not really leave the conventions of set and predicate notation.

7 Example: tossing. The probability that head *and* tail turn up together with *one* throw of two *fair* coins is 0.5. To formalize the claim, let C be a set consisting of just two distinct symbols H and T, that is, $C = \{H, T\}$; letters C, H, and T are mnemonic for coin, head and tail. The claim then reads:

$$(\mathcal{P}x, y : C \bullet \{x, y\} = \{H, T\}) \quad = \quad 1/2$$

Proof. Fairness means that the probability distribution is uniform:

$$(\mathcal{P}x : C \bullet x{=}H) \quad = \quad (\mathcal{P}x : C \bullet x{=}T) \quad = \quad 1/2 \qquad \text{(Fairness)}$$

Now:

$$\begin{aligned}
&(\mathcal{P}x, y : C \bullet \{x, y\} = \{H, T\}) \\
&= (\mathcal{P}x, y : C \bullet (x{=}H \wedge y{=}T) \vee (x{=}T \wedge y{=}H)) \\
&= (\mathcal{P}x, y : C \bullet x{=}H \wedge y{=}T) + (\mathcal{P}x, y : C \bullet x{=}T \wedge y{=}H) && \text{[distribution]} \\
&= (\mathcal{P}x : C \bullet x{=}H){\times}(\mathcal{P}y : C \bullet y{=}T) \;+\; (\mathcal{P}x : C \bullet x{=}T){\times}(\mathcal{P}y : C \bullet y{=}H) && \text{[indep.]} \\
&= \tfrac{1}{2}{\times}\tfrac{1}{2} \;+\; \tfrac{1}{2}{\times}\tfrac{1}{2} && \text{[fairness]} \\
&= 1/2
\end{aligned}$$

Unfair coins can be dealt with analogously; simply give a different probability distribution.

8 Example: event spaces. The problems and solutions discussed by Robertson [14] form yet another confirmation that our notation works well. Robertson essentially proposes to distinguish the various *event spaces* $\mathcal{E}$ that play a role, and to indicate them somehow in the notation $\mathcal{P}(X \mid Y)$, say as $\mathcal{P}_{\mathcal{E}}(X \mid Y)$. Here we cite the case described by Robertson (the table on the right is ours):

Example: we have stars $\mathcal{S}$, and planets $\mathcal{T}$. Stars either have (X=1) or do not have (X=0) magnetic fields. Planets either have (Y=1) or do not have (Y=0) magnetic fields. We have a (complete) universe consisting of 2 stars and 3 planets. Star s_1 has x_1=1; it has two planets t_{11} and t_{12} with y_{11}=1 and y_{12}=0. Star s_2 has x_2=0; it has one planet t_{21} with y_{21}=0.

The universe:

star	X	planet	Y
s_1	1	t_{11}	1
		t_{12}	0
s_2	0	t_{21}	0

In this universe, the following probabilities may be calculated exactly:

$$\begin{aligned}
\mathcal{P}(X{=}1) &= \tfrac{1}{2} \\
\mathcal{P}(Y{=}1 \mid X{=}1) &= \tfrac{1}{2} \\
\mathcal{P}(Y{=}1 \mid X{=}0) &= 0
\end{aligned}$$

From these we would infer that $\mathcal{P}(Y{=}1) = \frac{1}{4}$; the inference uses the following law [a generalization of our Divide&Conquer, but written in conventional notation]:

$$\mathcal{P}(Y) \;=\; \sum\nolimits_X \mathcal{P}(X)\mathcal{P}(Y \mid X) \qquad \text{(Divide and Conquer)}$$

But we have three planets, one of which has a magnetic field, so actually we have $\mathcal{P}(Y{=}1) = \frac{1}{3}$.

What is the problem here? In short, it is the event space. The laws of probability are written in terms of a single event space with a single probability measure on it; for historical reasons, the standard notation $\mathcal{P}(. \mid .)$ does not provide for the denotation of the event space.

Robertson proposes to denote the probability for a particular event space $\mathcal{E}$ as $\mathcal{P}_{\mathcal{E}}(. \mid .)$. Thus, writing $\mathcal{S}$ and $\mathcal{T}$ for the event space of stars and planets, respectively, he rewrites the calculated probabilities as:

$$\begin{aligned}
\mathcal{P}_{\mathcal{S}}(X{=}1) &= \tfrac{1}{2} \\
\mathcal{P}_{\mathcal{T}}(Y{=}1 \mid X{=}1) &= \tfrac{1}{2} \\
\mathcal{P}_{\mathcal{T}}(Y{=}1 \mid X{=}0) &= 0
\end{aligned}$$

He even distinguishes some more event spaces: $\mathcal{S}^+$, $\mathcal{T}^+$, and $\mathcal{ST}$. Thanks to this notational distinction, it is immediately clear that the Divide and Conquer law cannot be applied to these probabilities; they apply to different event spaces, whereas the equation apparently assumes them to apply to the same event space.

In our notation, the event space is mentioned in the declaration part D. Writing $Z(z)$ for "celestial body z has a magnetic field" (there is no need to invent two names X and Y for the same predicate!) and $star(t)$ for "the star of planet t", we thus have:

$$\begin{aligned}
(\mathcal{P}s : \mathcal{S} \bullet Z(s)) &= \tfrac{1}{2} \\
(\mathcal{P}t : \mathcal{T} \mid Z(star(t)) \bullet Z(t)) &= \tfrac{1}{2} \\
(\mathcal{P}s : \mathcal{S};\, t : \mathcal{T} \mid Z(s) \bullet Z(t)) &= \tfrac{1}{3} \\
(\mathcal{P}t : \mathcal{T} \mid \neg\, Z(star(t)) \bullet Z(t)) &= 0 \\
(\mathcal{P}s : \mathcal{S};\, t : \mathcal{T} \mid \neg\, Z(s) \bullet Z(t)) &= \tfrac{1}{3}
\end{aligned}$$

In our notation the calculation of "P(Y)" goes without error:

$$(\mathcal{P}t : \mathcal{T} \bullet Z(t))$$

$$
\begin{aligned}
&= (\mathcal{P}t : \mathcal{T} \mid \mathit{true} \bullet Z(t)) \\
&= (\mathcal{P}t : \mathcal{T} \mid (\exists\, s : \mathcal{S} \bullet Z(s) \vee \neg\, Z(s)) \bullet Z(t)) \\
&= (\mathcal{P}t : \mathcal{T};\, s : \mathcal{S} \mid Z(s) \vee \neg\, Z(s) \bullet Z(t)) && [\text{ (Shunting)}] \\
&= (\mathcal{P}t : \mathcal{T};\, s : \mathcal{S} \mid Z(s) \bullet Z(t))((\mathcal{P}t : \mathcal{T};\, s : \mathcal{S} \bullet Z(s)) + \\
&\quad (\mathcal{P}t : \mathcal{T};\, s : \mathcal{S} \mid \neg\, Z(s) \bullet Z(t))((\mathcal{P}t : \mathcal{T};\, s : \mathcal{S} \bullet \neg\, Z(s)) && [\text{(Divide\&Conquer)}] \\
&= \tfrac{1}{3} \times \tfrac{1}{2} + \tfrac{1}{3} \times \tfrac{1}{2} && [\text{probabilities given above}] \\
&= \tfrac{1}{3}
\end{aligned}
$$

If we want to discuss several probability distributions on the same space, we need to distinguish them in the notation, say by an index. Thus we may talk about $(\mathcal{P}_1\ D \mid P \bullet Q)$ and $(\mathcal{P}_2\ D \mid P \bullet Q)$ at the same time, with the same D, P and Q, while postulating different probability distributions for $\mathcal{P}_1$ and $\mathcal{P}_2$.

9 Conclusion. We have proposed a notation for probabilities that nicely interfaces with set and predicate notation, and other forms. The advantage is ease of understanding (similar aspects are denoted in the same way, in particular the aspect of free and bound variables), and ease of manipulations (transformation to and from set and predicate notation are possible without any change, and laws that already exists in the context of sets and predicates can now be applied as well). An important advantage of the conventional notation is its brevity. However, as we have shown, with suitable abbreviations we achieve the same brevity.

References

[1] R.S. Bird. An introduction to the theory of lists. In M. Broy, editor, *Logic of Programming and Calculi of Discrete Design*, pages 3–42. Springer Verlag, 1987. Also Technical Monograph PRG–56, Oxford University, Computing Laboratory, Programming Research Group, October 1986.

[2] R.S. Bird. Shall we calculate? *The Squiggolist*, 1(1), 1989.

[3] R.S. Bird. Shall we calculate – II? *The Squiggolist*, 1(3), 1990.

[4] R.S. Bird and O. de Moor. *Algebra of Programming.* Prentice-Hall International, 1996.

[5] Jonathan Bowen. The Z notation web site. `http://www.comlab.ox.ac.uk/archive/z.html`.

[6] E.W. Dijkstra and C.S. Scholten. *Predicate Calculus and Program Semantics.* Springer Verlag, 1990.

[7] Maarten M. Fokkinga. An exercise in transformational programming: Backtracking and Branch-and-Bound. *Science of Computer Programming*, 16(1):19–48, July 1991.

[8] M.M. Fokkinga. Calculate categorically! *Formal Aspects of Computing*, 4(4):673–692, 1992. Obtainable by `ftp://ftp.cs.utwente.nl/pub/doc/Parlevink/fokkinga/mmf91j.ps.Z`.

[9] Djoerd Hiemstra. *Using Language Models for Information Retrieval.* PhD thesis, University of Twente, Netherlands, 2001. CTIT Ph.D. Thesis Series, No. 01-32.

[10] L. Meertens. Algorithmics — towards programming as a mathematical activity. In J.W. de Bakker and J.C. van Vliet, editors, *Proceedings of the CWI Symposium on Mathematics and Computer Science*, pages 289–334. North-Holland, 1986.

[11] Lambert Meertens. Calculate polytypically! In H. Kuchen and S. D. Swiestra, editors, *Proceedings 8th Int. Symp. on Programming Languages: Implementations, Logics, and Programs, PLILP'96, Aachen, Germany, 24–27 Sept 1996*, volume 1140 of *Lecture Notes in Computer Science*, pages 1–16. Springer-Verlag, Berlin, 1996.

[12] E. Meijer, M.M. Fokkinga, and R. Paterson. Functional programming with bananas, lenses, envelopes and barbed wire. In *FPCA91: Functional Programming Languages and Computer Architecture*, volume 523 of *Lect. Notes in Comp. Sc.*, pages 124–144. Springer Verlag, 1991. Obtainable by `http://www.cs.utwente.nl/~fokkinga/mmf91m.ps`.

[13] Jeff Miller. Earliest Uses of Various Mathematical Symbols. `http://members.aol.com/jeff570/mathsym.html`.

[14] Stephen Robertson. On Baysian models and event spaces in information retrieval. Technical note, authors e-maili: `ser@microsoft.com`, May 2002.

[15] J.M. Spivey. *The Z notation: a reference manual (2nd edition).* Prentice Hall International, UK, 1992.

Appendix: History of the notation. The following has been taken literally from `http://members.aol.com/jeff570/stat.html` [13], a web-page about the origin of symbols in mathematics.

Apart from the combinatorial symbols very little of the notation of modern probability dates from before the 20th century.

Probability. Symbols for the probability of an event A on the pattern of $P(A)$ or $Pr(A)$ are a relatively recent development given that probability has been studied for centuries. A.N. Kolmogorov's *Grundbegriffe der Wahrscheinlichkeitsrechnung* (1933) used the symbol $\mathsf{P}(A)$. The use of upper-case letters for events was taken from set theory. H. Cramér's *Random Variables and Probability Distributions* (1937), "the first modern book on probability in English," used $P(A)$. In the same year J.V. Uspensky (*Introduction to Mathematical Probability*) wrote simply (A). W. Feller's influential *An Introduction to Probability Theory and its Applications volume 1* (1950) uses $Pr\{A\}$ and $\mathsf{P}\{A\}$ in later editions. See also the "Earliest Uses of Symbols of Set Theory and Logic" page of this website [13].

Conditional probability. Kolmogorov's (1933) symbol for conditional probability ("die bedingte Wahrscheinlichkeit") was $\mathsf{P}_B(A)$. Cramér (1937) referred to the "relative probability" and wrote $\mathsf{P}_B(A)$. Uspensky (1937) used the term "conditional probability" with the symbol (A, B). The vertical stroke notation $Pr\{A \mid B\}$ was made popular by Feller (1950), though it was used earlier by H. Jeffreys. In his *Scientific Inference* (1931) $P(p \mid q)$ stands for "the probability of the proposition p on the data q." Jeffreys mentions that Keynes and Johnson, earlier Cambridge writers, had used p/q; Jeffreys himself had used $P(p : q)$. The symbols p and q came from Whitehead and Russell's *Principia Mathematica.* See also the "Earliest Uses of Symbols of Set Theory and Logic" page of this website [13].

Random variable. The use of upper and lower case letters to distinguish a random variable from the value it takes, as in $Pr\{X = x_j\}$, became popular around 1950. The convention is used in Feller's *Introduction to Probability Theory.*