FORMALIZED MATHEMATICS
Vol. 17, No. 2, Pages 173–178, 2009
DOI: 10.2478/v10037-009-0020-z

Probability on Finite and Discrete Set and Uniform Distribution

Hiroyuki Okazaki
Shinshu University
Nagano, Japan

Summary. A pseudorandom number generator plays an important role in practice in computer science. For example: computer simulations, cryptology, and so on. A pseudorandom number generator is an algorithm to generate a sequence of numbers that is indistinguishable from the true random number sequence. In this article, we shall formalize the "Uniform Distribution" that is the idealized set of true random number sequences. The basic idea of our formalization is due to [15].

MML identifier: DIST_1, version: 7.11.02 4.125.1059

The notation and terminology used in this paper are introduced in the following papers: [16], [10], [1], [3], [17], [6], [18], [8], [4], [5], [7], [2], [11], [13], [12], [9], [14], and [19].

1. Probability on Finite and Discrete Set

Let S be a non empty finite set and let s be a finite sequence of elements of S. We introduce the certain event of s as a synonym of $\operatorname{dom} s$.

Let S be a non empty finite set and let s be a non empty finite sequence of elements of S. Then the certain event of s is a non empty finite set.

Next we state the proposition

(1) Let S be a non empty finite set and s be a finite sequence of elements of S. Then the certain event of $s = s^{-1}(S)$.

Let S be a non empty finite set, let s be a finite sequence of elements of S, and let x be a set. We introduce $\mathcal{E}_i(s(i) = x)$ as a synonym of $s^{-1}(x)$.

ISSN 1426–2630(p), 1898-9934(e)

Let S be a non empty finite set, let s be a finite sequence of elements of S, and let x be a set. Then $\mathcal{E}_i(s(i) = x)$ is an event of the certain event of s.

Let S be a non empty finite set, let s be a finite sequence of elements of S, and let x be a set. The functor frequency(x, s) yielding a natural number is defined as follows:

(Def. 1) frequency$(x, s) = \overline{\overline{\mathcal{E}_i(s(i) = x)}}$.

One can prove the following propositions:

(2) Let S be a non empty finite set, s be a finite sequence of elements of S, and e be a set. Suppose $e \in$ the certain event of s. Then there exists an element x of S such that $e \in \mathcal{E}_i(s(i) = x)$.

(3) Let S be a non empty finite set and s be a finite sequence of elements of S. Then $\overline{\overline{\text{the certain event of } s}} = \operatorname{len} s$.

Let S be a non empty finite set, let s be a finite sequence of elements of S, and let x be a set. The functor $\operatorname{Prob}_{\mathrm{D}}(x, s)$ yielding a real number is defined as follows:

(Def. 2) $\operatorname{Prob}_{\mathrm{D}}(x, s) = \frac{\text{frequency}(x,s)}{\operatorname{len} s}$.

Next we state the proposition

(4) For every non empty finite set S and for every finite sequence s of elements of S and for every set x holds frequency$(x, s) = \operatorname{len} s \cdot \operatorname{Prob}_{\mathrm{D}}(x, s)$.

Let S be a non empty finite set and let s be a finite sequence of elements of S. The functor FDprobSEQ s yielding a finite sequence of elements of $\mathbb{R}$ is defined by:

(Def. 3) $\operatorname{dom}\text{FDprobSEQ}\, s = \operatorname{Seg}\overline{\overline{S}}$ and for every natural number n such that $n \in \operatorname{dom}\text{FDprobSEQ}\, s$ holds $(\text{FDprobSEQ}\, s)(n) = \operatorname{Prob}_{\mathrm{D}}((\text{CFS}(S))(n), s)$.

The following proposition is true

(5) Let S be a non empty finite set, s be a non empty finite sequence of elements of S, and x be a set. Then $\operatorname{Prob}_{\mathrm{D}}(x, s) = \mathrm{P}(\mathcal{E}_i(s(i) = x))$.

Let S be a non empty finite set and let s, t be finite sequences of elements of S. We say that s and t are probability equivalent if and only if:

(Def. 4) For every set x holds $\operatorname{Prob}_{\mathrm{D}}(x, s) = \operatorname{Prob}_{\mathrm{D}}(x, t)$.

Let us notice that the predicate s and t are probability equivalent is reflexive and symmetric.

The following proposition is true

(6) Let S be a non empty finite set and s, t, u be finite sequences of elements of S. Suppose s and t are probability equivalent and t and u are probability equivalent. Then s and u are probability equivalent.

Let S be a non empty finite set and let s be a finite sequence of elements of S. The equivalence class of s yielding a non empty subset of S^* is defined by

the condition (Def. 5).

(Def. 5) The equivalence class of $s = \{t; t$ ranges over finite sequences of elements of S: s and t are probability equivalent$\}$.

Next we state three propositions:

(7) Let S be a non empty finite set and s, t be finite sequences of elements of S. Then s and t are probability equivalent if and only if $t \in$ the equivalence class of s.

(8) Let S be a non empty finite set and s be a finite sequence of elements of S. Then $s \in$ the equivalence class of s.

(9) Let S be a non empty finite set and s, t be finite sequences of elements of S. Then s and t are probability equivalent if and only if the equivalence class of $s =$ the equivalence class of t.

Let S be a non empty finite set. The distribution family of S yielding a non empty family of subsets of S^* is defined by the condition (Def. 6).

(Def. 6) Let A be a subset of S^*. Then $A \in$ the distribution family of S if and only if there exists a finite sequence s of elements of S such that $A =$ the equivalence class of s.

Next we state two propositions:

(10) Let S be a non empty finite set and s, t be finite sequences of elements of S. Then s and t are probability equivalent if and only if FDprobSEQ $s =$ FDprobSEQ t.

(11) Let S be a non empty finite set and s, t be finite sequences of elements of S. If $t \in$ the equivalence class of s, then FDprobSEQ $s =$ FDprobSEQ t.

Let S be a non empty finite set. The functor GenProbSEQ S yields a function from the distribution family of S into $\mathbb{R}^*$ and is defined by the condition (Def. 7).

(Def. 7) Let x be an element of the distribution family of S. Then there exists a finite sequence s of elements of S such that $s \in x$ and (GenProbSEQ S)$(x) =$ FDprobSEQ s.

One can prove the following proposition

(12) Let S be a non empty finite set and s be a finite sequence of elements of S. Then (GenProbSEQ S)(the equivalence class of s) = FDprobSEQ s.

Let S be a non empty finite set. Observe that GenProbSEQ S is one-to-one.

Let S be a non empty finite set and let p be a finite probability distribution finite sequence of elements of $\mathbb{R}$. Let us assume that $\operatorname{len} p = \overline{\overline{S}}$ and there exists a finite sequence s of elements of S such that FDprobSEQ $s = p$. The functor distribution(p, S) yielding an element of the distribution family of S is defined by:

(Def. 8) (GenProbSEQ S)(distribution$(p, S)) = p$.

Let S be a non empty finite set and let s be a finite sequence of elements of S. The functor freqSEQ s yields a finite sequence of elements of $\mathbb{N}$ and is defined by:

(Def. 9) $\operatorname{dom} \operatorname{freqSEQ} s = \operatorname{Seg} \overline{\overline{S}}$ and for every natural number n such that $n \in \operatorname{dom} \operatorname{freqSEQ} s$ holds $(\operatorname{freqSEQ} s)(n) = \operatorname{len} s \cdot (\operatorname{FDprobSEQ} s)(n)$.

One can prove the following propositions:

(13) Let S be a non empty finite set, s be a non empty finite sequence of elements of S, and n be a natural number. If $n \in \operatorname{Seg} \overline{\overline{S}}$, then there exists an element x of S such that $(\operatorname{freqSEQ} s)(n) = \operatorname{frequency}(x, s)$ and $x = (\operatorname{CFS}(S))(n)$.

(14) For every non empty finite set S and for every finite sequence s of elements of S holds $\operatorname{freqSEQ} s = \operatorname{len} s \cdot \operatorname{FDprobSEQ} s$.

(15) For every non empty finite set S and for every finite sequence s of elements of S holds $\sum \operatorname{freqSEQ} s = \operatorname{len} s \cdot \sum \operatorname{FDprobSEQ} s$.

(16) Let S be a non empty finite set, s be a non empty finite sequence of elements of S, and n be a natural number. Suppose $n \in \operatorname{dom} s$. Then there exists a natural number m such that $(\operatorname{freqSEQ} s)(m) = \operatorname{frequency}(s(n), s)$ and $s(n) = (\operatorname{CFS}(S))(m)$.

(17) Let n be a natural number, S be a function, and L be a finite sequence of elements of $\mathbb{N}$. Suppose that

(i) S is disjoint valued,

(ii) $\operatorname{dom} S = \operatorname{dom} L$,

(iii) $n = \operatorname{len} L$, and

(iv) for every natural number i such that $i \in \operatorname{dom} S$ holds $S(i)$ is finite and $L(i) = \operatorname{Card} S(i)$.

Then $\bigcup \operatorname{rng} S$ is finite and $\operatorname{Card} \bigcup \operatorname{rng} S = \sum L$.

(18) Let S be a function and L be a finite sequence of elements of $\mathbb{N}$. Suppose S is disjoint valued and $\operatorname{dom} S = \operatorname{dom} L$ and for every natural number i such that $i \in \operatorname{dom} S$ holds $S(i)$ is finite and $L(i) = \operatorname{Card} S(i)$. Then $\bigcup \operatorname{rng} S$ is finite and $\operatorname{Card} \bigcup \operatorname{rng} S = \sum L$.

(19) For every non empty finite set S and for every non empty finite sequence s of elements of S holds $\sum \operatorname{freqSEQ} s = \operatorname{len} s$.

(20) For every non empty finite set S and for every non empty finite sequence s of elements of S holds $\sum \operatorname{FDprobSEQ} s = 1$.

(21) Let S be a non empty finite set and s be a non empty finite sequence of elements of S. Then $\operatorname{FDprobSEQ} s$ is finite probability distribution.

Let S be a non empty finite set. A finite probability distribution finite sequence of elements of $\mathbb{R}$ is said to be a probability distribution finite sequence on S if:

(Def. 10) $\operatorname{len} \mathrm{it} = \overline{\overline{S}}$ and there exists a finite sequence s of elements of S such that $\mathrm{FDprobSEQ}\, s = \mathrm{it}$.

The following proposition is true

(22) Let S be a non empty finite set and p be a probability distribution finite sequence on S. Then

(i) p is a finite probability distribution finite sequence of elements of $\mathbb{R}$,

(ii) $\operatorname{len} p = \overline{\overline{S}}$,

(iii) there exists a finite sequence s of elements of S such that $\mathrm{FDprobSEQ}\, s = p$,

(iv) $\mathrm{distribution}(p, S)$ is an element of the distribution family of S, and

(v) $(\mathrm{GenProbSEQ}\, S)(\mathrm{distribution}(p, S)) = p$.

2. Uniform Distribution

Let S be a non empty finite set and let s be a finite sequence of elements of S. We say that s is uniformly distributed if and only if:

(Def. 11) For every natural number n such that $n \in \operatorname{dom} \mathrm{FDprobSEQ}\, s$ holds $(\mathrm{FDprobSEQ}\, s)(n) = \frac{1}{\overline{\overline{S}}}$.

We now state four propositions:

(23) Let S be a non empty finite set and s be a finite sequence of elements of S. If s is uniformly distributed, then $\mathrm{FDprobSEQ}\, s$ is constant.

(24) Let S be a non empty finite set and s, t be finite sequences of elements of S. Suppose s is uniformly distributed and s and t are probability equivalent. Then t is uniformly distributed.

(25) Let S be a non empty finite set and s, t be finite sequences of elements of S. Suppose s is uniformly distributed and t is uniformly distributed. Then s and t are probability equivalent.

(26) For every non empty finite set S holds $\mathrm{CFS}(S)$ is uniformly distributed.

Let S be a non empty finite set. The uniform distribution S yielding an element of the distribution family of S is defined by the condition (Def. 12).

(Def. 12) Let s be a finite sequence of elements of S. Then $s \in$ the uniform distribution S if and only if s is uniformly distributed.

Let S be a non empty finite set. One can check that there exists a probability distribution finite sequence on S which is constant.

Let S be a non empty finite set. The functor $\mathrm{UniformFDprobSEQ}\, S$ yielding a constant probability distribution finite sequence on S is defined as follows:

(Def. 13) $\mathrm{UniformFDprobSEQ}\, S = \mathrm{FDprobSEQ}\, \mathrm{CFS}(S)$.

We now state the proposition

(27) For every non empty finite set S holds the uniform distribution S = distribution(UniformFDprobSEQ S, S).

References

[1] Grzegorz Bancerek. Cardinal numbers. *Formalized Mathematics*, 1(**2**):377–382, 1990.
[2] Grzegorz Bancerek. The fundamental properties of natural numbers. *Formalized Mathematics*, 1(**1**):41–46, 1990.
[3] Grzegorz Bancerek. The ordinal numbers. *Formalized Mathematics*, 1(**1**):91–96, 1990.
[4] Grzegorz Bancerek and Krzysztof Hryniewiecki. Segments of natural numbers and finite sequences. *Formalized Mathematics*, 1(**1**):107–114, 1990.
[5] Czesław Byliński. Finite sequences and tuples of elements of a non-empty sets. *Formalized Mathematics*, 1(**3**):529–536, 1990.
[6] Czesław Byliński. Functions and their basic properties. *Formalized Mathematics*, 1(**1**):55–65, 1990.
[7] Czesław Byliński. Functions from a set to a set. *Formalized Mathematics*, 1(**1**):153–164, 1990.
[8] Czesław Byliński. Partial functions. *Formalized Mathematics*, 1(**2**):357–367, 1990.
[9] Czesław Byliński. The sum and product of finite sequences of real numbers. *Formalized Mathematics*, 1(**4**):661–668, 1990.
[10] Agata Darmochwał. Finite sets. *Formalized Mathematics*, 1(**1**):165–167, 1990.
[11] Jarosław Kotowicz. Real sequences and basic operations on them. *Formalized Mathematics*, 1(**2**):269–272, 1990.
[12] Andrzej Nędzusiak. Probability. *Formalized Mathematics*, 1(**4**):745–749, 1990.
[13] Jan Popiołek. Introduction to probability. *Formalized Mathematics*, 1(**4**):755–760, 1990.
[14] Piotr Rudnicki. Little Bezout theorem (factor theorem). *Formalized Mathematics*, 12(**1**):49–58, 2004.
[15] Victor Shoup. A computational introduction to number theory and algebra. *Cambridge University Press*, 2008.
[16] Zinaida Trybulec. Properties of subsets. *Formalized Mathematics*, 1(**1**):67–71, 1990.
[17] Edmund Woronowicz. Relations and their basic properties. *Formalized Mathematics*, 1(**1**):73–83, 1990.
[18] Edmund Woronowicz. Relations defined on sets. *Formalized Mathematics*, 1(**1**):181–186, 1990.
[19] Bo Zhang and Yatsuka Nakamura. The definition of finite sequences and matrices of probability, and addition of matrices of real elements. *Formalized Mathematics*, 14(**3**):101–108, 2006, doi:10.2478/v10037-006-0012-1.

Received May 5, 2009