Randomness and Computation
Joint Workshop "New Horizons in Computing" and "Statistical Mechanical Approach to Probabilistic Information Processing" (18-21 July, 2005, Sendai, Japan)

On a definition of random sequences with respect to parametric models

Hayato Takahashi [1]
Tokyo Institute of Technology, Department of Mathematical and Computing Sciences, 2-12-1-W8-25, O-okayama, Meguro-ku, Tokyo 152-8552, Japan.

1 Introduction

In [2] Martin-Löf succeeded to define random sequences with respect to two important class of probability distributions: computable distributions and Bernoulli model for all parameters. There are many subsequent studies on random sequence with respect to computable distributions, for example see surveys [3]. However from probabilistic or statistical point of view the assumption of computability of distributions is inadequate. In parametric statistical inference we consider a family of distributions $P_\theta,\ \theta \in \Theta$. Here P_θ is a probability distribution for each fixed θ, and P_θ is called parametric model. Since usually the parameter space Θ consists of uncountable elements, e.g., $[0,1], R$, etc., assumption on computability of parametric model is inappropriate. For example let us consider Bernoulli model (in the following we consider parametric distributions $P(\ \cdot\ ;\theta)$ $(= P_\theta)$ on infinite binary sequences Ω)

$$P(x^n;\theta) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}, \tag{1}$$

where $x^n = x_1x_2\cdots x_n \in \{0,1\}^n$ and $\theta \in [0,1]$. Then we see that θ is computable iff P_θ is computable, see pp. 300 problem 4.5.3 in [1]. Here we say that a real number θ is computable if there is a computable function that approximates θ with any given precision, and probability measure P is computable if there is a computable function $A(x)$ that approximates $P(x)$ with any given precision for all finite binary string x. Therefore we cannot apply Martin-Löf theory for computable distributions to Bernoulli model if the parameter is not computable.

In order to define randomness of Bernoulli model for all parameter, in [2] Martin-Löf studied conditional probability by sufficient statistic, which is uniform (and of course computable) distribution, then by applying that distribution to his theory for computable distributions he succeeded to define randomness with respect to Bernoulli model for all parameters.

In this paper, we give a definition of random sequences with respect to parametric distributions, which is slightly different from [2]. We consider the parameter to be not only unknown variable but also random variable; thus our definition is based on Bayesian statistics. We introduce a universal Bayes test, which is a Bayesian version of Martin-Löf universal test. Then we define random sequences with respect to parametric model by our universal Bayes test. In [4], a similar problem is studied in a different way, however they considered only computable parameters.

2 Universal Bayes test

We introduce a Bayesian version of Martin-Löf theory in a natural way. For simplicity, let the prameter space $\Theta = [0,1)$. Let S be the set of finite binary strings. For $s \in S$, let $\Delta(s) = [\sum_{k=1}^{l(s)} s_k 2^{-k}, \sum_{k=1}^{l(s)} s_k 2^{-k} + 2^{-l(s)}]$. Let m be a prior, i.e., probability measure on Θ. We say that m is *computable* if there is a computable function $A(r,\epsilon)$ such that $\forall \epsilon \in Q^+ \forall r \in S^k,\ |A(r,\epsilon) - m(\Delta(r))| < \epsilon$. Let $\{P(\cdot;\theta)\}_{\theta\in\Theta}$ be a parametric model, and let

$$P_m(x,s) = \int_{\Delta(s)} P(x;\theta)dm(\theta), \tag{2}$$

for $x, s \in S$. P_m uniquely defines a probability measure on Borel σ-algebra of $\Omega \times \Theta$. Computability of P_m is defined in a similar way. We assume that (a) m is computable, (b) P_m is computable, and (c) θ is computable iff P_θ is computable.

[1]E-mail: Hayato.Takahashi@is.titech.ac.jp

In the following for sets $U_n \subseteq S \times S$, we set $\tilde{U_n} = \cup_{(x,s)\in U_n} \Delta(x) \times \Delta(s)$. Let $U = \{(n,x,s) | (x,s) \in U_n, n \in N\}$. If U is recursively enumerable and $P_m(\tilde{U}_n) < 2^{-n}$ holds for all n, then we call U *Bayesian test*. If U is maximal among the Bayesian test, U is called *universal*. The following is a direct consequence of the theorem of Martin-Löf.

Theorem 2.1 (Martin-Löf) *There exists a universal Bayes test.*

If U is an universal Bayes test, then the set $\cap_{n=1}^{\infty} \tilde{U}_n$ is called *nonrandom pair with respect to prior* m, and its complement $\{\cap_{n=1}^{\infty} \tilde{U}_n\}^c$ is called *random pair with respect to prior* m; let

$$\mathcal{R}^{<P,\ m>} = \{\cap_{n=1}^{\infty} \tilde{U}_n\}^c,$$

and

$$\mathcal{R}_{\theta}^{<P,\ m>} = \{x^{\infty} \mid (x^{\infty}, \theta) \in \{\cap_{n=1}^{\infty} \tilde{U}_n\}^c\}.$$

Let $\mathcal{R}^m$ be the set of random reals with respect to m. We have the following:

Lemma 2.1 *$\theta \notin \mathcal{R}^m$ iff $\mathcal{R}_{\theta}^{<P,\ m>} = \emptyset$. Let $\tilde{\mathcal{R}}^m = \{\theta\ | P(\mathcal{R}_{\theta}^{<P,\ m>}; \theta) = 1\}$ then $\tilde{\mathcal{R}}^m \subseteq \mathcal{R}^m$ and $m(\tilde{\mathcal{R}}^m) = 1$.*

Thus we may say that for $\theta \in \tilde{\mathcal{R}}^m$, $\mathcal{R}_{\theta}^{<P,\ m>}$ is the set of random sequences with respect to $P(\cdot;\theta)$ and m. The author do not know whether $\mathcal{R}^m = \tilde{\mathcal{R}}^m$ or not, however the author conjecture that $\mathcal{R}^m = \tilde{\mathcal{R}}^m$. We call θ *atom* or we say that m has point mass at θ if $m(\{\theta\}) > 0$.

Theorem 2.2 *If θ is computable and an atom of m then $\theta \in \mathcal{R}^m$ and $\mathcal{R}_{\theta}^{<P,\ m>} = \mathcal{R}^{P_{\theta}}$.*

Next we study relations between random samples and random parameters. For $x \in S$, let

$$P_{mix}(x) = P_m(x, \lambda) = \int_{\Theta} P(x;\theta) dm(\theta). \tag{3}$$

P_{mix} is called mixture distribution. Since by assumption P_m is computable, we see P_{mix} is computable. Let $\mathcal{R}^{P_{mix}}$ be the set of random sequences with respect to P_{mix}.

Lemma 2.2 *$\mathcal{R}^{<P,\ m>} \subseteq \mathcal{R}^{P_{mix}} \times \mathcal{R}^m$.*

We see that if (x^{∞}, θ) is a random pair then x^{∞} is random with respect to the mixture distribution, and θ is random with respect to the prior. Let Km be the monotone complexity, for example see [3].

Theorem 2.3 *If $(x^{\infty}, \theta) \in \mathcal{R}^{<P,m>}$ then*

$$\begin{aligned} Km(x^n, \theta^m) &= Km(x^n) + Km(\theta^m | x^n) + O(1) \\ &= Km(\theta^m) + Km(x^n | \theta^m) + O(1), \end{aligned}$$

where x^n is the initial segment of x^{∞} of length n, and θ^m is the initial segment of binary expansion of θ of length m.

Let $I(x,y) = Km(x) + Km(y) - Km(x,y)$. By Theorem 2.3 we have the following symmetry:

Colorally 2.1 *If $(x^{\infty}, \theta) \in \mathcal{R}^{<P,m>}$ then*

$$\begin{aligned} I(x^n, \theta^m) &= Km(x^n) - Km(x^n | \theta^m) + O(1) \\ &= Km(\theta^m) - Km(\theta^m | x^n) + O(1). \end{aligned}$$

References

[1] M. Li and P. M. B. Vitanyi. *An Introduction to Kolmogorov Complexity and its Applications*. Springer, New York, second edition, 1997.

[2] P. Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–619, 1966.

[3] V. A. Uspenskii, A. L. Semenov, and A. Kh. Shen. Can an individual sequence of zeros and ones be random? *Russian Math. Surveys*, 45(1):121–189, 1990.

[4] V. G. Vovk and V. V. V'yugin. On the empirical validity of the Bayesian method. *J. R. Stat. Soc. B*, 55(1):253–266, 1993.