Bull. Nov. Comp. Center, Num. Anal., 14 (2009), 41–56

Numerical characteristics of randomness for binary samples

Sergei M. Prigarin, Angela Kempe, Gerhard Winkler

Abstract. In this paper, several numerical characteristics of sample randomness are discussed. These characteristics are based on the concepts of waiting time and information level of samples. The considered numerical characteristics can be used for developing the novel methods for model selection in statistical analysis.

Introduction

A serious effort was made by mathematicians to define a concept of randomness. These attempts promoted the appearance of new research directions and theories. A primitive formulation of the problem is the following: why a sample '1111111111110000000' for the Bernoulli distribution seems to be less random than '0111010001101111101'? The probability of the samples is the same. So, we need some additional characteristics to measure 'randomness'. One can apply different algorithmic or statistical approaches to define randomness, see, for example, [1–6]. In [7], an approach was proposed which is based on the average time of the first appearance of a sample in a sequence of independent identically distributed random variables. Below, we compare this approach to an alternative based on a family of posterior distributions.

To be more precise, we do not want to measure 'randomness', but we want to measure 'consistency' of a sample with respect to a prior model. Measures of such a consistency can be used for model selection in statistics: among several alternative models we choose a model, for which an observed sample is most consistent. In this paper, we consider several characteristics of **sample consistency** and present the results of consistency computation on an example of the Bernoulli distribution.

1. A sample information level

Assume that $X = (x_1, \ldots, x_N)$ is a sample of independent identically distributed **discrete** random variables $\xi_1, \ldots, \xi_N$, i.e. X is a realization of the random vector $(\xi_1, \ldots, \xi_N)$. By Y_M we denote a set of all M-size subsamples of X:

$$Y_M = \{(x_i, \ldots, x_{i+M}),\ i = 1, \ldots, N - M + 1\}, \quad 1 \le M \le N.$$

There are $N-M+1$ subsamples of the size M in X, but some of them can coincide. Assuming that every subsample $(x_i, \dots, x_{i+M})$, $i = 1, \dots, N-M+1$, can be chosen with the same probability $1/(N-M+1)$, we obtain *a posterior* distribution Q_M^X for samples of the size M concentrated on the set Y_M. As *a prior* distribution it is natural to use a distribution of $(\xi_1, \dots, \xi_M)$, denoted by P_M. As a 'distance' between P_M and Q_M^X we consider the Kullback–Leibler divergence

$$d_{\mathrm{KL}}(P_M, Q_M^X) = I(Q_M^X \mid P_M) = \sum_{Y \in Y_M} Q_M^X(Y) \ln \frac{Q_M^X(Y)}{P_M(Y)}. \tag{1}$$

It can be interpreted as 'the loss of information if we believe in Q_M^X, whereas P_M is the right distribution'.

The value

$$R_{\mathrm{KL}}(X) = \sum_{M=1}^{N} d_{\mathrm{KL}}(P_M, Q_M^X)$$

will be called the **information level** of a sample X and will be used as a characteristic of sample consistency: for smaller $R_{\mathrm{KL}}(X)$ a sample X is more consistent with the prior model.

2. A combined waiting time of a sample

In Tables 1–7, we present the information level of samples for the Bernoulli distribution with probability p to hit 1. In the same tables, one can find values of the **combined waiting time** $\boldsymbol{V}(X)$ of a sample X,

$$\boldsymbol{V}(X) = \sum_{Y \le X} \boldsymbol{W}(Y), \tag{2}$$

where the summation is performed over all subsamples $Y = (x_k, \dots, x_m)$, $1 \le k \le m \le N$, of the sample X and $\boldsymbol{W}(Y)$ is the **waiting time** of a sample Y, i.e., an average time of the first appearance of the sample Y in a sequence of independent identically distributed random variables $\xi_0, \xi_1, \dots, \xi_n, \dots$:

$$\boldsymbol{W}(X) = \mathsf{E}\tau_1(X),$$
$$\{\tau_1(X) = 0\} = \{\xi_0^N = X\},$$
$$\{\tau_1(X) = k\} = \{\xi_k^N = X,\ \xi_{k-1}^N \ne X,\ \dots,\ \xi_0^N \ne X\}$$
$$\xi_k^N = (\xi_k, \xi_{k+1}, \dots, \xi_{k+N-1}).$$

The concept of a combined waiting time was proposed in [7] to characterize the sample consistency.

Table 1. Characteristics of sample consistency for the Bernoulli distribution, $p = 0.5$ $N = 2$

X	$\boldsymbol{V}(X)$	$R_{\mathrm{KL}}(X)$	$R_{\chi}(X)$	$R_H(X)$
00	6	2.079	4	1.59
11	6	2.079	4	1.59
01	4	1.386	3	1
10	4	1.386	3	1

Table 2. Characteristics of sample consistency for the Bernoulli distribution, $p = 0.5$, $N = 3$

X	$\boldsymbol{V}(X)$	$R_{\mathrm{KL}}(X)$	$R_{\chi}(X)$	$R_H(X)$
000	22	4.159	11	2.88
111	22	4.159	11	2.88
010	14	2.829	8.11	1.91
101	14	2.829	8.11	1.91
001	14	2.829	8.11	1.91
100	14	2.829	8.11	1.91
011	14	2.829	8.11	1.91
110	14	2.829	8.11	1.91

Table 3. Characteristics of sample consistency for the Bernoulli distribution, $p = 0.6$, $N = 3$

X	$\boldsymbol{V}(X)$	$R_{\mathrm{KL}}(X)$	$R_{\chi}(X)$	$R_H(X)$	$\mathsf{P}(X)$
000	39.38	5.50	21.38	3.43	0.064
001	20.00	3.42	11.32	2.19	0.096
100	20.00	3.42	11.32	2.19	0.096
010	17.92	3.22	10.80	2.07	0.096
111	12.96	3.06	6.07	2.32	0.216
101	12.78	2.68	7.05	1.86	0.144
011	11.39	2.48	6.67	1.70	0.144
110	11.39	2.48	6.67	1.70	0.144

Table 4. Characteristics of sample consistency for the Bernoulli distribution, $p = 0.5$, $N = 4$

X	$\boldsymbol{V}(X)$	$R_{\mathrm{KL}}(X)$	$R_{\chi}(X)$	$R_H(X)$
0000	64	6.93	26	4.38
1111	64	6.93	26	4.38
0001	42	5.04	19.47	3.17
0111	42	5.04	19.47	3.17
0101	40	4.91	19.22	3.11
1010	40	4.91	19.22	3.11
0100	38	4.58	18.58	2.84
0010	38	4.58	18.58	2.84
0110	36	4.45	18.33	2.77
0011	36	4.45	18.33	2.77

Table 5. Characteristics of sample consistency for the Bernoulli distribution, $p = 0.5$, $N = 5$.

X	$\boldsymbol{V}(X)$	$R_{\mathrm{KL}}(X)$	$R_{\chi}(X)$	$R_H(X)$
00000	163	10.40	57	6.03
11111	163	10.40	57	6.03
01111	111	8.00	43.30	4.69
00001	111	8.00	43.30	4.69
01010	103	7.70	42.48	4.55
10101	103	7.70	42.48	4.55
00010	95	7.07	40.53	4.11
01000	95	7.07	40.53	4.11
00100	95	7.07	40.53	4.11
11011	95	7.07	40.53	4.11
01110	91	6.90	40.21	4.02
01001	91	6.90	40.21	4.02
10010	91	6.90	40.21	4.02
00011	91	6.90	40.21	4.02
00111	91	6.90	40.21	4.02
00101	91	6.90	40.21	4.02
01011	91	6.90	40.21	4.02
00110	87	6.55	39.71	3.72
01100	87	6.55	39.71	3.72

Table 6. Characteristics of sample consistency for the Bernoulli distribution, $p = 0.5$, $N = 6$

X	$\boldsymbol{V}(X)$	$R_{\mathrm{KL}}(X)$	$R_{\chi}(X)$	$R_H(X)$
000000	382	14.56	120	7.78
111111	382	14.56	120	7.78
000001	268	11.71	92.05	6.38
011111	268	11.71	92.05	6.38
010101	244	11.17	89.97	6.15
101010	244	11.17	89.97	6.15
000010	224	10.32	85.54	5.64
010000	224	10.32	85.54	5.64
011110	218	10.14	85.20	5.54
000011	218	10.14	85.20	5.54
001111	218	10.14	85.20	5.54
010010	214	10.03	84.88	5.49
101101	214	10.03	84.88	5.49
001001	214	10.03	84.88	5.49
011011	214	10.03	84.88	5.49
001010	214	10.03	84.88	5.49
010100	214	10.03	84.88	5.49
000100	216	9.98	84.54	5.43
001000	216	9.98	84.54	5.43
010001	206	9.69	83.88	5.29
011101	206	9.69	83.88	5.29
000101	206	9.69	83.88	5.29
010111	206	9.69	83.88	5.29
010110	204	9.63	83.77	5.26
011010	204	9.63	83.77	5.26
000111	204	9.63	83.77	5.26
001100	202	9.41	83.56	5.02
000110	202	9.41	83.56	5.02
011000	202	9.41	83.56	5.02
001101	200	9.35	83.45	5.00
001110	200	9.35	83.45	5.00
011100	200	9.35	83.45	5.00
011001	200	9.35	83.45	5.00
001011	200	9.35	83.45	5.00
010011	200	9.35	83.45	5.00

Table 7. Characteristics of sample consistency for the Bernoulli distribution, $p = 0.5$, $N = 9$

X		$\boldsymbol{V}(X)$	$R_{\mathrm{KL}}(X)$	$R_{\chi}(X)$	$R_H(X)$	Group size
000101110 001011100 010001110 011100010	000111010 001110100 010111000 011101000	20.593	701.889	9.046	1909	16
010001101 010110001 011000101 011100101	010011101 010111001 011010001 011101001	20.659	702.014	9.080	1913	16
000101100 001011000	000110100 001101000	20.709	702.112	9.105	1917	8
001001110 010011100	001110010 011100100	20.791	702.215	9.202	1917	8
000010111 000100111 000101111 000110111 000111011 001000111 001011001 001011110 001101001 001110101 010000111 010001111 010011011 010100011 010110011 011000111 011010011 011100011 011110010	000011101 000101101 000110101 000111001 000111101 001001101 001011101 001100101 001110001 001111010 010001011 010010111 010011110 010100111 010111100 011001011 011100001 011110001 011110100	20.857	702.340	9.236	1921	76
000010110 000100110 001000110 010000110 010011000 011000010 011001000	000011010 000110010 001100010 010001100 010110000 011000100 011010000	20.907	702.439	9.262	1925	28
000111100 010010110 010100110 011001010	001111000 010011010 010110010 011010010	20.922	702.465	9.270	1925	16

Table 7. Continuation

X		$\boldsymbol{V}(X)$	$R_{\mathrm{KL}}(X)$	$R_{\chi}(X)$	$R_H(X)$	Group size
000011100 001110000	000111000	21.015	702.689	9.314	1933	6
000011011 001010011 001101110 001111001 011001110 011100110	001001111 001101011 001110110 011000011 011011100 011101100	21.055	702.667	9.393	1929	24
001001100	001100100	21.105	702.765	9.418	1933	4
000011110 010110100	001011010 011110000	21.120	702.792	9.427	1933	8
000010011 001000011 001101111 011001111	000011001 001100001 001111011 011110011	21.213	703.015	9.470	1941	16
000110110 011000110	001101100 011011000	21.220	703.430	9.478	1941	8
000101011 001001011 001011011 001100111 010011001 011001101	000110011 001010111 001100011 001110011 011001001 011011001	21.286	703.555	9.512	1945	24
010111010		21.293	702.917	9.634	1937	2
000101001 010000101 010111101 011101011	001010001 010100001 011010111 011110101	21.344	703.015	9.660	1941	16
001010110 010101100	001101010 011010100	21.351	703.680	9.546	1949	8
010001010	010100010	21.419	703.342	9.685	1949	4
000100011 001110111	000110001	21.444	703.904	9.590	1957	8
000111110 011111000	001111100	21.468	704.132	9.590	1961	6
000010100 001010000	000101000	21.493	703.338	9.735	1953	6

Table 7. Continuation

X		$\boldsymbol{V}(X)$	$R_{\mathrm{KL}}(X)$	$R_{\chi}(X)$	$R_H(X)$	Group size
000001011 000001110 001111101 010011111 011100000	000001101 001011111 010000011 011000001 011111001	21.519	704.231	9.616	1965	20
010101110	011101010	21.524	703.805	9.754	1953	4
000011000	000110000	21.534	703.914	9.616	1969	4
000100101 010110111	001000101 010111011	21.575	703.904	9.779	1957	8
011011110	011110110	21.584	703.467	9.834	1953	4
010111110	011111010	21.692	704.356	9.823	1969	4
011010110		21.722	704.132	9.910	1961	2
000001100	001100000	21.765	704.803	9.736	1985	4
000010101 010010001 011011101	010001001 010101111 011101101	21.773	704.231	9.936	1965	12
001001010 010010100	001010010 010100100	21.848	704.557	9.962	1973	8
000001111 010010101 010101101	000011111 010100101 010110101	21.905	704.709	9.988	1977	12
000010010 010000100	001000010 010010000	21.922	704.553	10.011	1977	8
001100110	011001100	21.927	707.206	9.847	1997	4
001101101	010010011	21.992	707.331	9.881	2001	4
000001010 010100000	010000010	21.997	704.880	10.037	1985	6
000001001 000100001 011011111 011110111	000010001 001000001 011101111 011111011	22.136	705.255	10.100	1997	16
010110110	011011010	22.231	707.581	10.122	2009	4
011101110		22.324	707.805	10.166	2017	2
000101010 010101000	001010100	22.356	708.006	10.174	2021	6
010101001	011010101	22.413	708.157	10.200	2025	4

Table 7. Ending

X		$\boldsymbol{V}(X)$	$R_{\mathrm{KL}}(X)$	$R_{\chi}(X)$	$R_H(X)$	Group size
000100010 001000100 010001000	000100100 001001000	22.431	708.002	10.223	2025	10
001111110	011111100	22.575	709.923	10.106	2065	4
000001000 000100000	000010000	22.605	706.704	10.319	2041	6
000000110	011000000	22.682	710.121	10.163	2073	4
000000111	000111111	22.748	710.048	10.313	2069	4
000000101 010111111	010000001 011111101	22.855	710.246	10.370	2077	8
000000100	001000000	23.201	711.042	10.551	2105	4
001001001 011011011	010010010	23.409	719.782	10.616	2141	6
001010101	010101011	23.628	721.148	10.687	2169	4
000000011 011111110	001111111	24.481	727.838	10.976	2301	6
000000010	010000000	24.662	728.135	11.085	2313	4
010101010		25.740	765.845	11.522	2493	2
000000001	011111111	27.213	786.197	11.996	2793	4
000000000		31.192	1013.000	13.385	3797	2

There is a nice mathematical result (see, e.g., [8,9]) that gives us a simple rule to calculate the waiting time $\boldsymbol{W}(X)$ for a sample $X = (x_1, x_2, \ldots, x_N)$ from the discrete distribution:

$$\boldsymbol{W}(X) = t(1) + t(2) + \ldots + t(N) - N, \tag{3}$$

where

$$t(1) = \frac{1}{\mathsf{P}(x_1, \ldots, x_N)},$$

$$t(2) = \begin{cases} 1/\mathsf{P}(x_2, \ldots, x_N) & \text{if } x_2 = x_1,\ x_3 = x_2,\ \ldots,\ x_N = x_{N-1}, \\ 0, & \text{otherwise,} \end{cases}$$

$$t(3) = \begin{cases} 1/\mathsf{P}(x_3, \ldots, x_N) & \text{if } x_3 = x_1,\ x_4 = x_2,\ \ldots,\ x_N = x_{N-2}, \\ 0, & \text{otherwise,} \end{cases}$$

$$\ldots$$

$$t(N) = \begin{cases} 1/\mathsf{P}(x_N) & \text{if } x_N = x_1, \\ 0, & \text{otherwise.} \end{cases}$$

One can make the following trivial observations:

- Two samples with the same probability can have different waiting time.
- If one sample has a larger probability (as compared to another one), this does not mean that it has a smaller waiting time.

The following hypothesis was put forward in [7].

Hypothesis. Let P_1 and P_2 be two discrete distributions, and $X = (x_1, ..., x_N)$ be a sample of length N. If $\mathsf{P}_1(X) \geq \mathsf{P}_2(X)$, then $\boldsymbol{W}_{P_1}(X) \leq \boldsymbol{W}_{P_2}(X)$, where $\mathsf{P}_i(X) = P_i(x_1) \cdots P_i(x_N)$, $i = 1, 2$, and $\boldsymbol{W}_{P_1}$, $\boldsymbol{W}_{P_2}$ are waiting times with respect to the corresponding distributions. In other words, if a sample with respect to the first model has a larger probability than with respect to the second one, then this sample has a smaller waiting time with respect to the first model.

The statement below can easily be proved using formula (3).

Statement. A) In the case of the Bernoulli distribution with equiprobable events ($p = 0.5$), the values of the waiting time $\boldsymbol{W}(X)$ are integers for all samples X of any length N.

B) Let $\boldsymbol{W}^N_{\min}$ and $\boldsymbol{W}^N_{\max}$ denote a minimum and a maximum of waiting times for all samples X of length N. Then, in case of the Bernoulli distribution with equiprobable events, the following relations are fulfilled:

$$\boldsymbol{W}^{N+1}_{\min} = \boldsymbol{W}^N_{\max} + 1, \qquad \boldsymbol{W}^N_{\max} = 2\boldsymbol{W}^N_{\min} + N - 2.$$

It is easy to check that samples like ‘00000000’ or ‘01010101’ have the largest values of the waiting time for the Bernoulli distribution with equiprobable events. On the other hand, samples of the type ‘00000001’ have the smallest waiting time (follows from (3)). That is why the the concept of ‘combined waiting time’ was proposed to characterize the sample consistency.

Remark. In addition to formula (2), another version of the combined waiting time was considered in [7]. It can be defined by the following recursive procedure

$$\boldsymbol{V}_2(x_1, \ldots, x_N) = \boldsymbol{W}(x_1, \ldots, x_N) + \boldsymbol{V}_2(x_1, \ldots, x_{N-1}) + \boldsymbol{V}_2(x_2, \ldots, x_N),$$
$$\boldsymbol{V}_2(x_1) = \boldsymbol{W}(x_1).$$

It is obvious that $\boldsymbol{V}_2(X) \geq \boldsymbol{V}(X)$. The value $\boldsymbol{V}_2(X)$ presents the sum of waiting times for all subsamples just as $\boldsymbol{V}(X)$, but for many subsamples the waiting time is presented in $\boldsymbol{V}_2(X)$ more than once (if a subsample is closer to the center of a sample X, then more often it makes a contribution).

3. The general approach based on posterior distributions families

To compute the information level of a sample in Section 1, we used the following general approach. On the basis of a sample X of size N we constructed a family of posterior distributions Q_M^X on the spaces of samples of various sample size M. Then, we calculated a 'distance' between the family of posterior distributions Q_M^X and the corresponding family of natural prior distributions P_M.

This general approach can be realized in different ways. For example, instead of the Kullback–Leibler divergence (1), we can consider other distances. Together with the information level R_{KL} in Tables 1–7, we present the values

$$R_\chi(X) = \sum_{M=1}^{N} d_\chi(P_M, Q_M^X), \qquad R_H(X) = \sum_{M=1}^{N} d_H(P_M, Q_M^X)$$

corresponding to the Chi-square statistic

$$d_\chi(P_M, Q_M^X) = \sum \frac{\Big[Q_M^X(Y) - P_M(Y)\Big]^2}{P_M(Y)} \tag{4}$$

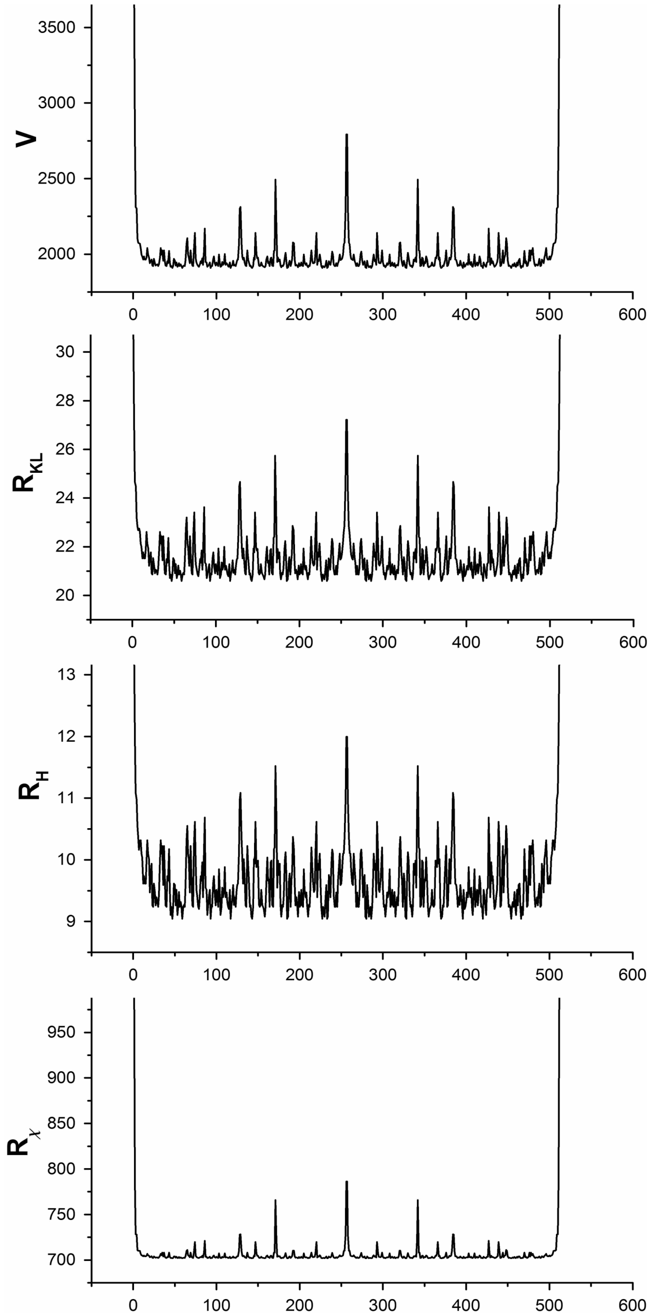
and the Hellinger distance

$$\begin{aligned} d_H(P_M, Q_M^X) &= \sum \Big[\sqrt{Q_M^X(Y)} - \sqrt{P_M(Y)}\Big]^2 = \\ &= 2 - 2\sum \sqrt{Q_M^X(Y) P_M(Y)}. \end{aligned} \tag{5}$$

The sums in (4) and (5) are taken over all samples Y of size M. The characteristics R_{KL}, R_χ, R_H, and V of the sample consistency are presented in Tables 1–7 and in the figure.

Moreover, characteristics of consistency can be introduced by different ways of constructing the posterior distributions. For instance, assume that a sample $Y = (y_1, \dots, y_M)$ of size M is chosen from a random vector $(\xi_1, \dots, \xi_{N+2M-2})$ in such a way that $(y_1, \dots, y_M) = (\xi_{k+1}, \dots, \xi_{k+M})$ and k is uniformly distributed in the set $\{0, 1, \dots, N+M-2\}$, while the observed sample $X = (x_1, \dots, x_N)$ is fixed just in the middle of the vector $(\xi_1, \dots, \xi_{N+2M-2})$:

$$(x_1, \dots, x_N) = (\xi_M, \dots, \xi_{M+N-1}).$$



Characteristics $V(X)$, $R_{\mathrm{KL}}(X)$, $R_H(X)$, $R_\chi(X)$ of sample consistency for the Bernoulli distribution, $N = 9$, $p = 0.5$ (binary samples X correspond to numbers $0, \dots, 511$ on the horizontal axis)

Table 8. Improper characteristics of sample consistency for the Bernoulli distribution, $p = 0.5$, $N = 6$

X	$R_{\mathrm{KL}}(X)$	$R_{\chi}(X)$	$R_H(X)$
000000	1.16057	5.42975	0.59948
101010	0.68376	2.45455	0.33078
000100	0.64865	1.66116	0.38796
001000	0.64865	1.66116	0.38796
000010	0.59847	1.46281	0.34784
010000	0.59847	1.46281	0.34784
010010	0.59298	1.79339	0.30127
000001	0.56143	1.66116	0.26481
100000	0.56143	1.66116	0.26481
001100	0.55563	1.46281	0.32045
001010	0.52147	1.33058	0.29377
010100	0.52147	1.33058	0.29377
001001	0.52104	1.59504	0.26079
011011	0.52104	1.59504	0.26079
001111	0.50502	1.19835	0.29689
000011	0.50502	1.19835	0.29689
000111	0.48720	1.06612	0.30912
000110	0.48034	1.19835	0.26442
011000	0.48034	1.19835	0.26442
000101	0.45569	1.06612	0.26329
101000	0.45569	1.06612	0.26329
101001	0.45358	1.19835	0.23237
011010	0.45358	1.19835	0.23237
100010	0.44314	1.26446	0.21502
010001	0.44314	1.26446	0.21502
011001	0.43700	1.26446	0.20971
011100	0.39938	1.06612	0.20613
001110	0.39938	1.06612	0.20613
001101	0.39801	1.00000	0.22118
101100	0.39801	1.00000	0.22118
001011	0.38594	0.93388	0.22880
100001	0.35432	1.00000	0.17106

In this case, the intersection of X and Y is not empty for any Y, and distribution Q_M^X is not singular even for $M = N$. The latter fact tempts one to consider only the distribution Q_N^X instead of a family of distributions and to use a distance between P_N and Q_N^X as characteristic of consistency. But this idea is not good because samples like '1000001' are among the most consistent according to such characteristic (Table 8), where

$$R^0_{\mathrm{KL}}(X) = d_{\mathrm{KL}}(P_N, Q_N^X), \quad R^0_\chi(X) = d_\chi(P_N, Q_N^X), \quad R^0_H(X) = d_H(P_N, Q_N^X).$$

Remarks. 1. A difference between the maximum likelihood approach and the sample consistency is demonstrated in Table 3, where the most probable sample '111' is not the most consistent.

2. In Tables 4–8, for $p = 0.5$, some of the samples that can be obtained by 0–1 inversion (with the same consistency characteristics) are omitted.

3. For $p = 0.5$ and $N = 2, \ldots, 5$ (Tables 1, 2, 4, 5), all the considered consistency characteristics (R_{KL}, R_χ, R_H, V) reproduce the same ranking of samples. For $N = 6$, only the combined waiting time generates a slightly different ranking. More diverse are samples in Table 7 for $N = 9$ (sample groups are ordered here with respect to the information level R_{KL}).

4. On application to model selection

Assume that we have a priori information about the distribution of X and observe a sample Y after the transformation A of a random sample X, $Y = A(X)$. We want to choose the 'most appropriate' model of A among several given reversible transformations $A_1, \ldots, A_k$. To solve this problem, one can find samples $X_i = A_i^{-1}(Y)$, $i = 1, \ldots, k$, and take the transformation A_i as a solution if the sample X_i is the most consistent. The concept of sample consistency can be used even in situations, when conventional methods of mathematical statistics are inapplicable.

To illustrate how the concept of sample consistency can be applied, let us specify the problem in the following way. We assume $X = (x_1, \ldots, x_9)$ to be the Bernoulli sample with probability 0.5 to hit zero, and we know about the transformation A that it is either identical or it inverts elements of X with even indices. Then for an observed sample

$$Y = (y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9)$$

we have

$$X_1 = (y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9), \quad X_2 = (y_1, \overline{y_2}, y_3, \overline{y_4}, y_5, \overline{y_6}, y_7, \overline{y_8}, y_9),$$

where $\overline{y}$ is inversion of a binary digit y. All binary samples X including X_1 and X_2 have equal probabilities, and it seems that no classical statistical method can help us to choose the most appropriate model. Using the concept of sample consistency, we can choose the 'most appropriate' model if X_1 and X_2 have different consistency characteristics. For example, if Y = '010001101', then X_1 = '010001101', X_2 = '000100111', and we should choose an identical model because the sample X_1 is more consistent, see Table 7:

$$R_{\text{KL}}(X_1) = 20.659, \quad R_\chi(X_1) = 702.014, \quad R_H(X_1) = 9.080, \quad V(X_1) = 1913,$$
$$R_{\text{KL}}(X_2) = 20.857, \quad R_\chi(X_2) = 702.340, \quad R_H(X_2) = 9.236, \quad V(X_2) = 1921.$$

If Y = '010010010', then X_1 = '010010010', X_2 = '000111000', and we should choose a model with inversion:

$$R_{\text{KL}}(X_1) = 23.409, \quad R_\chi(X_1) = 719.782, \quad R_H(X_1) = 10.616, \quad V(X_1) = 2141,$$
$$R_{\text{KL}}(X_2) = 21.015, \quad R_\chi(X_2) = 702.689, \quad R_H(X_2) = 9.314, \quad V(X_2) = 1933.$$

For Y = '000001111', he have X_1 = '000001111', X_2 = '010100101', and the consistency characteristics do not enable us to choose an appropriate model:

$$R_{\text{KL}}(X_1) = R_{\text{KL}}(X_2) = 20.905, \quad R_\chi(X_1) = R_\chi(X_2) = 704.709,$$
$$R_H(X_1) = R_H(X_2) = 9.988, \quad V(X_1) = V(X_2) = 1977.$$

Finally, we would like to mention that the concept of sample consistency generates various problems concerning analysis and comparison of the consistency characteristics, the justification of model selection criteria, the generalization of the concept for continuous distributions and samples with dependent elements, and so on. But, according to our preliminary consideration, this concept can be useful for the development of new model selection methods.

Acknowledgements. We would like to thank Dr. Andreas Martin and Dr. Klaus Hahn for the helpful discussions.

References

[1] Kolmogorov A.N. On tables of random numbers // Theoretical Computer Sciences. — 1998. — Vol. 207. — P. 387–395 (reprinted from Sankhya: The Indian Journal of Statistics. — 1963. — Series A. – Vol. 25. – Part 4).

[2] Kolmogorov A.N. Combinatorial foundations of information theory and the calculus of probabilities // Russian Mathematical Surveys. — 1983. — Vol. 38, No. 4. — P. 29–40.

[3] Shen A. Discussion on Kolmogorov complexity and statistical analysis // The Computer Journal. — 1999. — Vol. 42, No. 4. — P. 340–342.

[4] Gammerman A., Vovk V. Kolmogorov complexity: sources, theory and application // The Computer Journal. — 1999. — Vol. 42, No. 4. — P. 252–255.

[5] Muchnik An.A., Semenov A.L., Uspensky V.A. Mathematical metaphysics of randomness // Theoretical Computer Science. — 1998. — Vol. 207, Iss. 2. — P. 263–317.

[6] Muchnik An.A., Semenov A.L. On the role of the law of large numbers in the theory of randomness // Problems of Information Transmission. — 2003. — Vol. 39, No. 1. — P. 119–147.

[7] Prigarin S.M., Kempe A., Winkler G. An attempt to define new criteria for model selection in mathematical statistics. — 2005. — (Preprint / GSF Neuherberg; 05-18).

[8] Ross S. Stochastic Processes. — New York: Wiley, 1983.

[9] Williams D. Probability with Martingales. — Cambridge University Press, 1991.