О критериях проверки равномерности, использующих оценки энтропии

П.Ю. Блинов, Б.Ю. Лемешко[1]
Новосибирский государственный технический университет

Рассматриваются критерии проверки гипотезы о принадлежности выборки равномерному закону распределения, базирующиеся на непараметрических оценках энтропии. Исследуются распределения статистик критериев, мощность критериев относительно различных конкурирующих гипотез. Опираясь на результаты исследований, даны рекомендации по применению критериев. Рассмотрена проблема выбора размера окна.

Ключевые слова: равномерное распределение, оценка энтропии, критерий, статистика критерия, мощность, размер окна.

1. Введение

Для проверки гипотезы о равномерности наблюдаемых случайных величин предложено множество специальных критериев. Это обусловлено интересом, который проявляется к использованию модели равномерного закона в различных приложениях. Частота применения модели равномерного закона в задачах статистического анализа не в последнюю очередь определяется тем, что использование такой простой модели во многих ситуациях позволяет найти решение задачи с опорой только на аналитические методы. Если применение модели равномерного закона обосновано, то многие статистические выводы оказываются проще.

Пусть случайные величины $X_1, X_2, ..., X_n$ принадлежат некоторому закону с функцией распределения вероятностей $F(x)$, тогда случайные величины $Y_i = F(X_i)$, $i = \overline{1,n}$ распределены равномерно на интервале $[0,1]$. В связи с этим во многих случаях задачу проверки гипотезы о принадлежности выборки $X_1, X_2, ..., X_n$ некоторому непрерывному закону распределения можно заменить задачей проверки гипотезы о принадлежности выборки $Y_1, Y_2, ..., Y_n$ равномерному закону.

При проверке гипотезы о принадлежности наблюдаемой случайной величины равномерному закону **простая** проверяемая гипотеза имеет вид H_0: $X \in Rav(0,1)$ или H_0: $X \in Rav(a,b)$, где a и b известны. Эту же гипотезу можно записать как H_0: $F(x) = x$, $x \in [0,1]$ или H_0: $F(x) = \frac{x-a}{b-a}$, $x \in [a,b]$. Проверяемая гипотеза будет **сложной**, если по данной выборке находится и область определения равномерной случайной величины.

Со статистической проверкой гипотез связаны два вида ошибок. Ошибка 1-го рода, вероятность которой α (уровень значимости), как правило задаётся, заключается в отклонении справедливой проверяемой гипотезы H_0. Если выдвигается и некоторая конкурирующая гипотеза H_1, то с ней связывают ошибку 2-го рода (и её вероятность β), которая заключается в том, что при справедливости H_1 не отклоняется проверяемая гипотеза H_0. Естественно желание, чтобы при проверке гипотезы вероятности α и β были минимальны, но при выдви-

[1] Работа выполнена при поддержке Министерства образования и науки РФ в рамках проектной части государственного задания (№ 2.541.2014/К).

жении конкретной гипотезы H_1 задание одной из вероятностей определяет и другую (для данного критерия и объёма выборки n).

С вероятностью β ошибки 2-го рода связана мощность критерия $1-\beta$. Понятно, что при выборе критерия, если есть такая возможность, предпочтение следует отдать критерию с наибольшей мощностью. Естественно, что наибольший интерес вызывает способность критериев различать близкие альтернативы, то есть, отличать от H_0 близкие конкурирующие гипотезы.

В данной работе в качестве конкурирующих рассматриваются три гипотезы H_1, H_2 и H_3, соответствующие бета-распределению 1-го рода $B_I(\theta_0,\theta_1,\theta_2,\theta_3)$ с функцией плотности

$$f(x)=\frac{1}{\theta_2 \mathrm{B}(\theta_0,\theta_1)}\left(\frac{x-\theta_3}{\theta_2}\right)^{\theta_0-1}\left(1-\frac{x-\theta_3}{\theta_2}\right)^{\theta_1-1}$$

при различных значениях параметров:

$$H_1 : F(x)=B_I(1.5, 1.5, 1, 0),\ x\in[0,1];$$

$$H_2 : F(x)=B_I(0.8, 1, 1, 0),\ x\in[0,1];$$

$$H_3 : F(x)=B_I(1.1, 0.9, 1, 0),\ x\in[0,1].$$

Функции распределения вероятностей, соответствующие рассматриваемым гипотезам, достаточно близки. При этом конкурирующей гипотезе H_1 соответствует закон, функция распределения которого пересекается с функцией распределения равномерного закона, а при H_2 и H_3 функции распределения законов лежат выше и ниже функции равномерного. В то же время плотности законов существенно отличаются от равномерного (см. рис. 1).

Следует отметить, что для многих критериев проверки равномерности, в том числе и для непараметрических критериев согласия, камнем преткновения оказывается различение H_0 и H_1 при малых объёмах выборок n.

В данной работе рассматриваемые критерии проверки равномерности, как и в предшествующих работах [1–3], исследовались методами статистического моделирования. При исследовании распределений статистик критериев количество экспериментов, осуществляемых при статистическом моделировании принималось равным 1 660 000. Такое количество экспериментов позволяет, с одной стороны, проследить качественную картину, отражающую изменение распределений статистик в зависимости от различных факторов, с другой – обеспечить приемлемую точность получаемых оценок мощности и искомых вероятностей.

2. Энтропийные критерии проверки равномерности

Пусть $X_1, X_2, ..., X_n$ – выборка независимых наблюдений случайной величины X с функцией плотности $f(x)$, тогда энтропия $H(f)$ этой величины, предложенная Шенноном [4], имеет вид:

$$H(f)=-\int_{-\infty}^{\infty}\ln\left(f(x)\right)f(x)dx. \tag{1}$$

Впервые оценку энтропии предложил Васичек [5], используя тот факт, что (1) можно привести к виду:

$$H(f)=\int_0^1 \ln\left\{\frac{d}{dp}F^{-1}(p)\right\}dp.$$

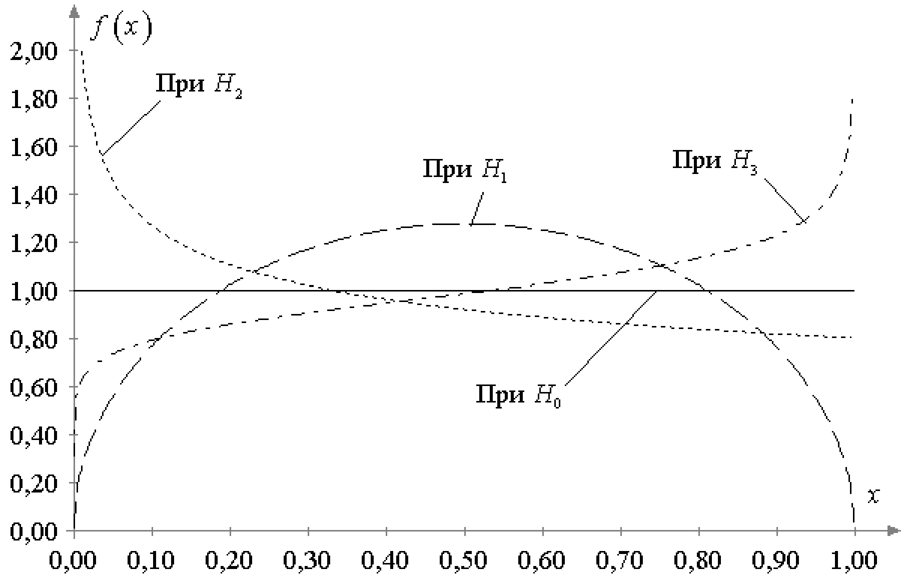

Рис. 1. Плотности распределения законов, соответствующие конкурирующим гипотезам

В конечном виде предложенная им оценка описывается формулой:

$$H = \frac{1}{n}\sum_{i=1}^{n}\ln\left\{\frac{n}{2m}\left(U_{i+m}-U_{i-m}\right)\right\}, \tag{2}$$

где $1 \le m < \frac{n}{2}$ целочисленный параметр, называемый размером окна; $U_i = X_{(i)} -$ i-й элемент вариационного ряда, построенного по выборке $X_1, X_2, ..., X_n$; $U_{i-m} = U_1$, если $i-m<1$, и $U_{i+m} = U_n$, если $i+m>n$. В [5] было доказано, что если $n,m \to \infty$ и $\frac{m}{n} \to 0$ то $H \to H(f)$.

Васичек использовал оценку (2) в качестве статистики критерия для проверки нормальности. Использовать данную оценку в качестве статистики критерия проверки равномерности предложили Дудевич и ван дер Мюлен [6]. Они отметили, что при любых выборках, распределенных на интервале $[0,1]$, всегда выполняется неравенство $H \le 0$, и предложили для проверки равномерности правосторонний критерий со статистикой:

$$H(m,n) = -\frac{1}{n}\sum_{i=1}^{n}\ln\left\{\frac{n}{2m}\left(U_{i+m}-U_{i-m}\right)\right\}. \tag{3}$$

Во многих источниках (3) называется энтропийным критерием Дудевича–ван дер Мюлена.

В таблице 1 приведены процентные точки критерия со статистикой (3), расширенные и уточнённые нами в ходе исследований.

Ибрагими [7] на основе (2) предложил собственную оценку энтропии:

$$H_c = \frac{1}{n}\sum_{i=1}^{n}\ln\left\{\frac{n}{cm}\left(U_{i+m}-U_{i-m}\right)\right\}, \tag{4}$$

где

$$c = \begin{cases} 1+(i-1)/m, & i \le m, \\ 2, & m+1 \le i \le n-m, \\ 1+(n-i)/m, & i \ge n-m+1. \end{cases}$$

Таблица 1. Критические значения критерия со статистикой $H(m,n)$

n	m	$1-\alpha$				
		0.8	0.85	0.9	0.95	0.99
10	1	0.700	0.758	0.835	0.960	1.225
	2	0.547	0.592	0.653	0.754	0.971
	3	0.542	0.584	0.642	0.735	0.942
	4	0.575	0.617	0.673	0.765	0.969
20	1	0.505	0.538	0.582	0.651	0.795
	2	0.350	0.375	0.409	0.463	0.579
	3	0.322	0.345	0.376	0.426	0.532
	4	0.325	0.346	0.376	0.423	0.525
	5	0.339	0.360	0.388	0.435	0.535
	6	0.359	0.380	0.408	0.454	0.554
	7	0.383	0.404	0.432	0.477	0.576
	8	0.410	0.431	0.459	0.504	0.603
	9	0.439	0.459	0.487	0.532	0.631
30	1	0.437	0.462	0.494	0.544	0.648
	2	0.283	0.301	0.326	0.364	0.445
	3	0.248	0.264	0.286	0.321	0.395
	4	0.242	0.257	0.277	0.310	0.381
	5	0.246	0.261	0.281	0.313	0.381
	6	0.256	0.271	0.290	0.321	0.389
	7	0.269	0.283	0.303	0.333	0.401
	8	0.285	0.299	0.317	0.348	0.415
	9	0.301	0.315	0.334	0.364	0.431
	10	0.319	0.333	0.351	0.382	0.448
40	1	0.403	0.423	0.449	0.490	0.573
	2	0.249	0.264	0.283	0.314	0.378
	3	0.211	0.224	0.241	0.268	0.327
	4	0.200	0.212	0.228	0.254	0.309
	5	0.200	0.211	0.226	0.251	0.305
	6	0.205	0.216	0.231	0.255	0.307
	7	0.213	0.224	0.238	0.262	0.314
	8	0.223	0.234	0.248	0.272	0.322
	9	0.234	0.245	0.259	0.282	0.333
	10	0.246	0.257	0.271	0.294	0.345
	15	0.316	0.326	0.341	0.364	0.413
50	1	0.382	0.399	0.421	0.456	0.526
	2	0.229	0.241	0.258	0.283	0.337
	3	0.188	0.199	0.214	0.236	0.284
	4	0.175	0.185	0.198	0.219	0.265
	5	0.172	0.181	0.194	0.214	0.258
	6	0.174	0.183	0.195	0.215	0.258
	7	0.179	0.188	0.200	0.219	0.261
	8	0.186	0.195	0.206	0.226	0.267
	9	0.194	0.203	0.214	0.234	0.274

Окончание таблицы 1

n	m	$1-\alpha$				
		0.8	0.85	0.9	0.95	0.99
50	10	0.203	0.212	0.223	0.242	0.283
	15	0.256	0.264	0.276	0.294	0.335
	20	0.315	0.324	0.335	0.354	0.394
100	1	0.337	0.348	0.362	0.384	0.427
	2	0.186	0.194	0.204	0.220	0.251
	3	0.142	0.149	0.157	0.171	0.199
	4	0.123	0.129	0.137	0.150	0.175
	5	0.115	0.120	0.128	0.139	0.164
	6	0.111	0.116	0.123	0.135	0.158
	7	0.110	0.115	0.122	0.134	0.156
	8	0.111	0.116	0.123	0.133	0.156
	9	0.113	0.118	0.125	0.135	0.157
	10	0.116	0.121	0.127	0.138	0.159
	15	0.138	0.142	0.148	0.158	0.179
	20	0.164	0.169	0.175	0.184	0.205
	30	0.222	0.227	0.233	0.242	0.263
	40	0.283	0.288	0.295	0.305	0.325

В [7] показано, что оценка (4) быстрее чем (2) сходится к (1), и доказано равенство:

$$H_c = H + \frac{2}{n}\left[m\ln 2m + \ln\frac{(m-1)!}{(2m-1)!} \right]. \tag{5}$$

Так как существует прямая зависимость вида (5) между оценками, замена 2 на коэффициент c в статистике (3) не повлияет на мощность критерия.

Совсем недавно было предложено два критерия, использующих другие оценки энтропии:

$$HY_1 = -\frac{1}{n}\sum_{i=1}^{n}\ln\left(\frac{U_{i+m}-U_{i-m}}{\hat{F}(U_{i+m})-\hat{F}(U_{i-m})}\right), \tag{6}$$

$$HY_2 = -\sum_{i=1}^{n}\ln\left(\frac{U_{i+m}-U_{i-m}}{\hat{F}(U_{i+m})-\hat{F}(U_{i-m})}\right)\left(\frac{\hat{F}(U_{i+m})-\hat{F}(U_{i-m})}{\sum_{j=1}^{n}\left(\hat{F}(U_{j+m})-\hat{F}(U_{j-m})\right)}\right), \tag{7}$$

где

$$\hat{F}(U_i) = \frac{n-1}{n(n+1)}\left(i + \frac{1}{n-1} + \frac{U_i - U_{i-1}}{U_{i+1}-U_{i-1}}\right), i = 2,\dots,n-1,$$

$$\hat{F}(U_1) = 1 - \hat{F}(U_n) = \frac{1}{(n+1)}.$$

Несколько ранее в [8] была представлена оценка (7), а позже в [9] было предложено использовать оценки (6) и (7) в качестве статистик критериев проверки равномерности.

Некоторая неопределенность при использовании энтропийных критериев со статистиками (3), (6) и (7) связана с выбором размера окна m, так как от этого зависят не только распределения статистик при справедливости проверяемой гипотезы H_0, но и мощность критериев.

В [9] был рекомендован выбор m (см. табл. 2) в зависимости от объёмов выборки n. Именно рекомендованные значения m^* были использованы при исследовании мощности критериев.

Таблица 2. Оптимальные значения m^* [9]

n	m^*	n	m^*
$n \le 5$	1	$19 \le n \le 29$	4
$6 \le n \le 8$	2	$30 \le n \le 39$	5
$9 \le n \le 18$	3	$40 \le n \le 100$	6

3. Исследование мощности критериев

В ходе работы был проведен сравнительный анализ мощности трёх критериев. Оценки мощности критериев со статистиками (3) и (7), полученные в ходе моделирования, можно увидеть в таблицах 3-5. Мощности критерия со статистикой (6) в таблицы не включены, так как они отличаются от мощности критерия со статистикой (3) на величину не более погрешности моделирования. Оценки мощности получены при $m = m^*$ (см. табл. 2).

Рассмотренные критерии показывают очень высокую мощность относительно конкурирующей гипотезы H_1. Причем бòльшую мощность показывает критерий со статистикой (7).

В то же время относительно конкурирующих гипотез H_2 и H_3 критерий со статистикой (7) демонстрирует меньшую мощность по сравнению с критериями со статистиками (3) и (6).

Как упоминалось выше, многие из критериев равномерности страдают следующим недостатком: при малых n и малых α они оказываются смещёнными относительно конкурирующей гипотезы H_1. Этого недостатка не отмечается у критериев (3), (6), (7), использующих оценки энтропии.

Таблица 3. Мощности критериев относительно гипотезы H_1 при m^*

n	Статистика	α				
		0.15	0.1	0.05	0.025	0.01
10	$H\left(m^*, n\right)$	0.347	0.254	0.145	0.080	0.035
	HY_2	0.362	0.265	0.151	0.083	0.036
20	$H\left(m^*, n\right)$	0.464	0.361	0.228	0.140	0.071
	HY_2	0.515	0.407	0.262	0.163	0.082
30	$H\left(m^*, n\right)$	0.562	0.459	0.311	0.204	0.112
	HY_2	0.636	0.532	0.373	0.250	0.143
40	$H\left(m^*, n\right)$	0.648	0.548	0.395	0.273	0.160
	HY_2	0.733	0.638	0.481	0.346	0.211
50	$H\left(m^*, n\right)$	0.696	0.601	0.449	0.324	0.199
	HY_2	0.789	0.704	0.557	0.421	0.274
100	$H\left(m^*, n\right)$	0.853	0.790	0.669	0.546	0.399
	HY_2	0.925	0.883	0.793	0.687	0.540

В целом критерии показывают неплохую мощность и относительно H_2 и H_3, но уступают многим, как непараметрическим критериям согласия, так и специальным критериям равномерности (например, критериям Хегази–Грина [10], Фросини [11], Неймана–Бартона [12]). Подчеркнём, что относительно конкурирующей гипотезы H_2 для всех 3-х критериев отмечается небольшая смещённость при объемах выборок $n = 10$.

Таблица 4. Мощности критериев относительно гипотезы H_2 при m^*

n	Статистика	α				
		0.15	0.1	0.05	0.025	0.01
10	$H\left(m^*,n\right)$	0.144	0.097	0.050	0.026	0.011
	HY_2	0.140	0.095	0.049	0.025	0.011
20	$H\left(m^*,n\right)$	0.166	0.115	0.062	0.033	0.015
	HY_2	0.154	0.106	0.056	0.030	0.013
30	$H\left(m^*,n\right)$	0.192	0.136	0.076	0.042	0.020
	HY_2	0.169	0.118	0.065	0.036	0.017
40	$H\left(m^*,n\right)$	0.217	0.158	0.092	0.053	0.026
	HY_2	0.184	0.131	0.074	0.042	0.020
50	$H\left(m^*,n\right)$	0.252	0.188	0.113	0.069	0.035
	HY_2	0.210	0.153	0.089	0.052	0.026
100	$H\left(m^*,n\right)$	0.407	0.327	0.224	0.151	0.088
	HY_2	0.340	0.266	0.174	0.113	0.063

Таблица 5. Мощности критериев относительно гипотезы H_3 при m^*

n	Статистика	α				
		0.15	0.1	0.05	0.025	0.01
10	$H\left(m^*,n\right)$	0.169	0.115	0.060	0.031	0.013
	HY_2	0.168	0.114	0.059	0.030	0.013
20	$H\left(m^*,n\right)$	0.192	0.134	0.071	0.038	0.017
	HY_2	0.191	0.133	0.071	0.038	0.016
30	$H\left(m^*,n\right)$	0.215	0.152	0.084	0.046	0.021
	HY_2	0.212	0.150	0.082	0.045	0.020
40	$H\left(m^*,n\right)$	0.237	0.172	0.098	0.055	0.026
	HY_2	0.233	0.167	0.094	0.053	0.024
50	$H\left(m^*,n\right)$	0.260	0.191	0.111	0.064	0.031
	HY_2	0.254	0.185	0.107	0.061	0.029
100	$H\left(m^*,n\right)$	0.355	0.275	0.175	0.110	0.058
	HY_2	0.347	0.267	0.169	0.105	0.055

4. Выбор размера окна m

Выбор оптимального размера окна m для рассмотренных критериев представляет собой актуальную проблему. Большинство авторов предлагают использовать такие m, при которых значения непараметрических оценок энтропии ближе к теоретическому. Но размер окна m влияет также на мощность критериев. Причём оптимальное m зависит и от конкурирующей гипотезы.

В данной работе нами были получены оптимальные значения m, при которых критерий Дудевича–ван дер Мюлена (3) показывает наибольшую мощность относительно конкурирующих гипотез H_1, H_2 и H_3. Полученные результаты относительно гипотез H_2 и H_3 представлены в таблицах 6 и 7. А относительно гипотезы H_1 мощность всегда возрастает при увеличении m, другими словами максимальная мощность в этом случае всегда будет при $m = \left[\frac{n-1}{2}\right]$.

Таблица 6. Оптимальные значения m относительно гипотезы H_2

n	m	n	m
$n \le 17$	1	$55 \le n \le 71$	4
$17 \le n \le 34$	2	$72 \le n \le 89$	5
$35 \le n \le 54$	3	$89 \le n \le 100$	6

Таблица 7. Оптимальные значения m относительно гипотезы H_3

n	m	n	m
$5 \le n \le 10$	2	$39 \le n \le 50$	6
$11 \le n \le 21$	3	$51 \le n \le 60$	7
$22 \le n \le 30$	4	$61 \le n \le 72$	8
$31 \le n \le 38$	5	$73 \le n \le 100$	9

Аналогичные исследования были выполнены для других критериев. Для критерия со статистикой (6) результаты оказались идентичными полученным для критерия Дудевича–ван дер Мюлена со статистикой (3).

Однако для критерия со статистикой (7) результаты отличаются. Относительно гипотезы H_1 (как и в случае других критериев) мощность всегда возрастает при увеличении m. Но относительно гипотез H_2 и H_3 мощность достигает максимального значения при других размерах окна m.

Некоторые оценки мощности критерия со статистикой (7) относительно конкурирующих гипотез H_2 и H_3 при $\alpha = 0.05$ представлены в таблице 8.

5. Заключение

Критерии, базирующиеся на оценках энтропии, представляют собой достаточно эффективные критерии проверки гипотез о принадлежности наблюдений равномерному закону. В частности они, как правило, существенно превосходят критерии равномерности, в которых используются разности последовательных порядковых статистик [13–15] (например, критерии Шермана [16], Кимбелла [17], Янга [18]).

Относительно конкурирующих гипотез вида H_2 или H_3 эти критерии несколько уступают в мощности критериям типа Хегази–Грина [10], Фросини [11], Неймана–Бартона [12], но у них отсутствует смещённость относительно конкурирующих гипотез вида H_1 и, даже более того, относительно такого рода гипотез они имеют преимущество в мощности перед большинством критериев, включая непараметрические критерии согласия, особенно при больших размерах окна m.

Таблица 8. Оценки мощности критерия со статистикой (7) относительно гипотез H_2 и H_3

n	Размер окна m					
	1	2	3	4	5	6
Относительно H_2						
10	0.057	0.052	0.049	0.046	–	–
20	0.072	0.067	0.062	0.056	0.052	0.049
30	0.085	0.082	0.076	0.070	0.065	0.060
40	0.097	0.0968	0.092	0.086	0.080	0.074
50	0.108	0.111	0.108	0.102	0.095	0.089
100	0.155	0.176	0.1836	0.184	0.180	0.174
Относительно H_3						
10	0.059	0.060	0.059	0.058	–	–
20	0.068	0.071	0.072	0.071	0.069	0.068
30	0.075	0.081	0.0834	0.0835	0.082	0.081
40	0.081	0.090	0.094	0.0953	0.0953	0.094
50	0.087	0.098	0.104	0.1072	0.1075	0.1073
100	0.111	0.133	0.148	0.159	0.165	0.169

Но так как при слишком больших m с использованием этих критериев хуже распознаются другие гипотезы (вида H_2 или H_3), рекомендуется применять критерий при нескольких размерах окна m или использовать критерий совместно с рядом других критериев, хорошо отличающих равномерное распределение от законов, соответствующих H_2 и H_3 (например, наряду с критериями равномерности Неймана–Бартона, Хегази–Грина, Фросини или критерием согласия Андерсона–Дарлинга).

С точки зрения оценивания энтропии лучше использовать оценки (4), (6) и (7) с авторскими рекомендациями по выбору размера окна, поскольку они принимают более близкие значения к величине $H(f)$ чем оценка (2).

Литература

1. *Лемешко Б. Ю.* Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход : монография / Б.Ю. Лемешко, С.Б. Лемешко, С.Н. Постовалов, Е.В. Чимитова. – Новосибирск : Изд-во НГТУ, 2011. – 888 с.
2. *Лемешко Б. Ю.* Непараметрические критерии согласия: Руководство по применению / Б. Ю. Лемешко.– М.: НИЦ ИНФРА-М, 2014. – 163 с.
3. *Лемешко Б. Ю.* Критерии проверки отклонения распределения от нормального закона: Руководство по применению / Б. Ю. Лемешко. – М.: НИЦ ИНФРА-М, 2015. – 160 с. – (Научная мысль). – www.dx.doi.org/10.12737/6086.
4. *Shannon C. E.* A mathematical theory of communications / C. E. Shannon // Bell System Technical Journal. – 1948. – V.27. – P. 379-423, 623-656.
5. *Vasicek O.* A test for normality based on sample entropy. / O. Vasicek // J. R. Statist. Soc. – 1976. – Ser. B. V.38. – P. 54-59.

6. *Dudewics E. J., van der Meulen E. C.* Entropy-based test of uniformity // J. Amer. Statist. Assoc. – 1981. – V.76. No. 376. – P. 967-974.
7. *Ebrahimi N., Pflughoert K., Soofi E.* Two measures of sample entropy // Statistics & Probability Letters. – 1994. – V.20. – P. 225-234.
8. *Yousefsadeh F., Arghami N. R.* Testing exponentiality based on type II censored data and a new CDF estimator. // Communications in Statistics - Simulation and Computation. – 2008. – V.37. – P. 1479-1499.
9. *Zamanzade E.* Testing uniformity based on new entropy estimators // Journal of Statistical Computation and Simulation. – 2014. DOI: 10.1080/00949655.2014.958085.
10. *Hegazy Y. A. S.* Some new goodness-of-fit tests using order statistics / Y. A. S. Hegazy, J. R. Green // Applied Statistics. – 1975. – V.24, №3. – P. 299-308.
11. *Frosini B. V.* On the distribution and power of goodness-of-fit statistic with parametric and nonparametric applications, “Goodness-of-fit” / Ed. by Revesz P., Sarkadi K., Sen P.K. //Amsterdam-Oxford-New York: North-Holland Publ. Comp. – 1987. – P. 133-154.
12. *Neyman J.* “Smooth” tests for goodness-of-fit / J. Neyman // Scandinavisk Aktuarietidskrift. – 1937. – V.20. – P. 149-199.
13. *Блинов П. Ю., Лемешко Б. Ю.* О мощности критериев, используемых для проверки гипотез о принадлежности выборок равномерному закону // Материалы Российской НТК “Обработка информационных сигналов и математическое моделирование”, Новосибирск. 2013. – С.35-38.
14. *Blinov P. Yu., Lemeshko B. Yu.* A review of the properties of tests for uniformity // 2014 12th Iinternational Conference on Actual Problems of Electronics Instrument Engineering (APEIE) 34006 Proceedings. Vol. 1. Novosibirsk, 2014. – P.540-547.
15. *Блинов П. Ю., Лемешко Б. Ю.* Обзор свойств критериев равномерности // Труды XII международной конференции “Актуальные проблемы электронного приборостроения” АПЭП-2014. Т.6, Новосибирск, 2014. – С.29-36.
16. *Sherman B.* A random variable related to the spacing of sample values / B. Sherman // The Annals of Mathematical Statistics. – 1950. – V.21, №3. – P. 339-361.
17. *Kimball B. F.* Some basic theorems for developing tests of fit for the case of the non-parametric probability distribution function./ B. F. Kimball // The Annals of Mathematical Statistics. – 1947. – V.18, №1. – P. 540-548.
18. *Young D. L.* The linear nearest neighbour statistic / D. L. Young // Biometrika. – 1982. – V.69, №2. – P. 477-480.

Лемешко Борис Юрьевич

Д.т.н., профессор, г.н.с. кафедры теоретической и прикладной информатики НГТУ (630073, Новосибирск, просп. Карла Маркса, 20), e-mail: `lemeshko@ami.nstu.ru`

Блинов Павел Юрьевич

Аспирант кафедры теоретической и прикладной информатики НГТУ (630073, Новосибирск, просп. Карла Маркса, 20), e-mail: `blindizer@ya.ru`.

The uniformity tests based on entropy estimators

P. Yu. Blinov, B. Yu. Lemeshko

The uniformity tests based on entropy estimators are considered. Distributions of test statistics, power of tests under different competing hypotheses are studied. We provide recommendations for using tests based on results of studies. Problem of choosing a values of window size is considered. Underlines that this type of uniformity tests has clear advantage in some cases.

Key words: uniform distibution, entropy estimator, test, test statistic, test power, window size