Short communication

Estimation of probability of incorrect result at application of Kolmogorov – Smirnov test: uniform distribution

L.V. Nedorezov

Research Center for Interdisciplinary Environmental Cooperation RAS, Saint-Petersburg, Russia
E-mail l.v.nedorezov@gmail.com

Keywords: Kolmogorov – Smirnov test, uniform distribution, Monte-Carlo methods

Let's assume that we have two samples of sizes n and m, and we know that these samples were generated by one and the same distribution. Below we'll consider the case with uniform distribution on interval $[0,1]$. Basic question is following: what is the probability of incorrect result if we apply Kolmogorov – Smirnov test (Bolshev, Smirnov, 1983) for analysis of these samples? A'priori we know that both samples were generated by one and the same law; respectively, incorrect result we can get in situation when Kolmogorov – Smirnov test shows that analyzed samples are different.

It is obvious that probabilities of considering events are positive and depend on significance level we use for Kolmogorov – Smirnov test. These probabilities can be considered as *built-in errors* of Kolmogorov – Smirnov test when we apply it for samples generated by uniform distribution. Thus we can use these estimations of probabilities when we compare various methods of homogeneity of two samples. It is better to use methods with smaller value of built-in errors.

Below several tables with obtained results (for 5%, 1%, and 0.1% significance levels of Kolmogorov – Smirnov test) are presented. For all cells with numbers (m,n) $2\cdot10^8$ independent samples (with respective sample sizes) were generated. It is important to note that in various cases built-in errors are very close to significance level. In tables there are tails of numbers only: in table 1 all numbers must start with 0.0 and after that we have to add a tail which is presented in respective cell. In table 2 all numbers must start with 0.00; in table 3 all numbers must start with 0.000.

Table 1 (part 1)
Results for 5% significance level*

	10	11	12	13	14	15
10	1233603					
11	3386457	2070252				
12	2762311	3140848	3150876			
13	2433732	3108253	2040812	4424927		
14	2632404	3006088	3228021	2945959	187314	
15	2956237	2847231	296251	2641095	385578	2618976
16	3121477	3357047	3370655	2691392	2999864	2694484
17	2821582	2699454	31709	2808289	32869	3304706
18	2620634	3321993	2615072	291779	3252303	3329461
19	3532356	3224865	2925379	2817678	3071327	3337027
20	2901367	2906815	3313639	3216718	3041948	3460711
21	2542566	277183	2891202	3124995	3369865	3009077
22	3299383	2049916	2917739	319408	3174864	3408974
23	2865395	3378981	3462194	3238698	3163326	3254038
24	2886002	2876592	3090183	3229899	3248284	3344238
25	2959546	3003621	2588956	2830648	3148971	3067337
26	3154165	2954708	3388846	4322757	3284001	3288334
27	3225627	2974286	3416795	3396064	3562434	3104512
28	2987837	2913738	3083746	310787	3056293	3281327
29	2653841	3231826	3253179	3099548	3094978	2860357
30	3738161	3136103	3526901	3143522	328745	4096141

*In Table tails of numbers are only presented. All numbers has a form 0.0 plus tail. Thus, in cell (10,10) we have number 0.01233603

Table 1 (part 2)
Results for 5% significance level*

	16	17	18	19	20
16	3521333				
17	2374949	4485835			
18	3055475	31078	2067678		
19	3162814	3179859	3863191	2669625	
20	3135293	3011034	308887	3349205	3359546
21	304343	3153422	3396117	3322729	2577166
22	3262488	3172317	3245533	3505289	332743
23	3227486	3135006	3275614	3408172	3193306
24	2689844	3376378	3272641	3192333	3258047
25	3388299	3164604	3336014	3524311	3519521
26	3351796	3057501	329118	3385216	3256655
27	3189612	3327463	311101	3418423	3568751
28	3413735	3572544	3331741	3168122	3459049
29	3383219	3433567	3310282	3413866	3380164
30	3417472	3464374	332418	3532638	3440812

*In Table tails of numbers are only presented. All numbers has a form 0.0 plus tail. Thus, in cell (16,16) we have number 0.03521333

Table 2 (part 1)
Results for 1% significance level*

	10	11	12	13	14	15
10	199877					
11	437543	433543				
12	474888	259711	786483			
13	41814	388866	485552	287303		
14	429953	488148	384289	489527	490365	
15	549782	391407	504141	483079	305194	762439
16	371965	369291	521413	50733	546743	488518
17	384346	477522	462143	451047	493329	452223
18	414915	451929	569761	450764	465777	469908
19	572855	470214	514458	477029	550437	498656
20	478595	463249	53142	460609	468227	487095
21	35193	4973	485378	509006	531322	558331
22	461965	351717	479927	544003	551597	598612
23	438252	624155	428993	511508	581557	488591
24	459169	50201	637228	548437	57971	515035
25	454226	544621	474535	604615	547748	601253
26	517892	521989	49497	450771	547819	552196
27	519466	453361	552434	551052	429479	56843
28	475678	507333	471953	556645	722756	503742
29	375118	550244	476775	567286	552967	646934
30	592829	522607	491534	518161	513548	510023

*In Table tails of numbers are only presented. All numbers has a form 0.00 plus tail. Thus, in cell (10,10) we have number 0.00199877

Table 2 (part 2)
Results for 1% significance level*

	16	17	18	19	20
16	299456				
17	647764	461074			
18	487708	385399	666387		
19	490918	558537	437149	277828	
20	541352	540519	500403	614101	398797
21	58561	517993	573372	474262	666549
22	519663	507462	536729	596452	563234
23	545142	497285	562883	522629	547578
24	476976	518465	562387	53798	521008
25	544994	516104	583211	580962	61522
26	564711	585335	563965	608911	582053
27	604745	588141	626194	602442	527944
28	515951	597091	608212	634496	603592
29	574352	584734	608844	546949	608984
30	612245	569371	601093	580486	515576

*In Table tails of numbers are only presented. All numbers has a form 0.00 plus tail. Thus, in cell (16,16) we have number 0.00299456

Table 3
Results for 0.1% significance level*

	10	11	12	13	14	15	16	17	18	19	20
10	22263										
11	1141	05767									
12	06772	36844	20505								
13	30209	21547	11447	49395							
14	18318	25934	2961	29109	16968						
15	30743	32382	27417	17572	28118	35756					
16	19748	20385	3052	28357	3755	21023	10295				
17	27317	30543	32948	28896	28966	24792	4124	23211			
18	29043	20423	21067	31929	32804	34638	25774	36296	43873		
19	19623	26362	30008	36047	28592	31468	37407	43777	26338	15043	
20	13387	24054	35479	39795	30321	41468	31935	32076	31732	44293	24698
21	3604	21555	2938	29397	28054	31328	38318	34158	43117	31191	48576
22	25561	40825	24689	28567	37275	3668	38132	3291	36566	38666	44113
23	29498	29658	44871	31978	33885	39198	38375	34626	41452	42119	3494
24	21463	21345	32824	42046	35098	31472	31529	36877	44748	43315	43839
25	2567	34015	22964	34208	35063	40526	39091	38543	37424	38206	41489
26	28457	33501	32558	24371	36203	38346	43025	47396	46164	38159	47211
27	26223	3182	32111	42722	25803	39331	42231	3841	3158	38933	48354
28	3139	35756	33999	38264	50959	34039	39668	36357	38748	371	45776
29	31035	30191	34451	33192	37361	51671	3676	42628	39863	42054	43379
30	2699	27667	42251	32259	32297	37162	48077	39359	4102	42414	50454

*In Table tails of numbers are only presented. All numbers has a form 0.000 plus tail. Thus, in cell (10,10) we have number 0.00022263

References

Bolshev L.N., Smirnov N.V., 1983. Tables of Mathematical Statistics. Moscow: Nauka, 416 p.