اختبارات الأرقام العشوائية:

لغرض التحقق من الخصائص التي يفترض ان تتصف بها الأرقام العشوائية (الارقام تتوزع وفق التوزيع المنتظم فضلا عن الاستقلالية)، فانه يمكن إجراء عدد من الاختبارات للتحقق من هذه الخصائص.

1. اختبار التوزيع المنتظم: يمكن استخدام اختبار كولموكروف سميرنوف او اختبار مربع كاي لمقارنة توزيع مجموعة من الارقام المولدة فيما اذا كانت هذ الارقام تتوزع وفق التوزيع المنتظم.

ان الفرضية التي يجري اختبارها لغرض التحقق من ان الارقام تتوزع وفق التوزيع المنتظم المستمر هي:

$$H_0: R_i \sim U[0,1]$$

$$H_1: R_i \nsim U[0,1]$$

اي اختبار الفرضية الصفرية القائلة بان الارقام تتوزع وفق التوزيع المنتظم بالفترة $[0,1]$ مقابل الفرضية البديلة القائلة بان الارقام لاتتوزع وفق التوزيع المنتظم بالفترة $[0,1]$.

ملاحظة: إن القيم الصحيحة التي يتم توليدها ستكون جميعها بين صفر و m وذلك بسبب قيمة المعامل m، وأن هذه الأعداد الصحيحة العشوائية يجب ان تتبع التوزيع المنتظم. كما ويمكن وضع أرقام عشوائية بين الصفر و الـ 1 بواسطة

$$R_i = \frac{X_i}{m},\ \ \mathrm{i} = 1, \dots$$

عليه، سيتم استخدام الارقام العشوائية بين الصفر والواحد في اختبارات التوزيع المنتظم والاستقلالية.

اختبارات التوزيع المنتظم: ان الاختبار الأساسي الذي يجب القيام به لغرض التحقق من صحة الارقام التي يتم توليدها هو اختبار التوزيع المنتظم .يوجد طريقتان مختلفتان للاختبار وهما اختبار كولموغوروف-سميرنوف واختبار مربع كاي، ويستند الاختبارين على فرضية العدم القائلة بعدم وجود فرق معنوي بين توزيع العينة والتوزيع النظري المتمثل بالتوزيع المنتظم.

الاتي شرح لكيفية تطبيق الاختبارين:

1. ختبار كولموجوروف - سميرنوف . في هذا الاختبار يتم مقارنة الدالة التوزيعية التجميعية (التراكمية) المستمرة $F(x)$، للتوزيع المنتظم مع الدالة التوزيعية التجميعية التجريبية، $S_N(x)$ لعينة مكونة من N من المشاهدات.

$$F(x) = x, \quad 0 \le x \le 1$$

اذا كانت عينة الارقام العشوائية التي تم توليدها هي: $R_1, \dots, R_N$، فان الدالة التوزيعية التجميعية التجريبية $S_N(x)$، تعرف بالصيغة الاتية:

$$S_N(x) = \frac{no. of\ R_1, \dots, R_N \le x}{N}$$

خطوات الاختبار:

1. ترتيب الارقام العشوائية ترتيبا تصاعديا، اي ان:
$$R_{(1)} \le R_{(2)} \le \cdots \le R_{(N)}$$

2. حساب القيمة D^+، D^-، وكالاتي:.

$$D^+ = \max_{1 \le i \le N} \left\{ \frac{i}{N} - R_{(i)} \right\}$$

$$D^- = \max_{1 \le i \le N} \left\{ R_{(i)} - \frac{i-1}{N} \right\}$$

3. ايجاد D وكالاتي:
$$D = \max(D^+, D^-)$$
4. استخرج قيمة D_α الجدولية من جدول كولموكروف-سميرنوف عند مستوى المعنوية α وبحجم العينة N.
5. قارن بين قيمة D المحسوبة مع قيمة D_α الجدولية، فاذا كانت قيمة D المحسوبة اكبر من قيمة D_α الجدولية، عندئذ ترفض فرضية العدم، بمعنى ان بيانات العينة لا تتوزع وفق التوزيع المنتظم. اما اذا كانت قيمة D المحسوبة اقل من قيمة D_α الجدولية، عندئذ تقبل فرضية العدم، بمعنى انه لايوجد فرق بين توزيع بيانات العينة وبين التوزيع المنتظم.

مثال:

استخدم اختبار كولموغروف- سميرنوف بدرجة حرية $\alpha = 0.05$ لبيان فيما اذا كانت الارقام العشوائية الاتية تتوزع وفق التوزيع المنتظم القياسي.

$$0.44, 0.81, 0.14, 0.05, 0.93$$

$R_{(i)}$	0.05	0.14	0.44	0.81	0.93
i	1	2	3	4	5
$\frac{i}{N}$	0.20	0.40	0.60	0.80	1.00
$\frac{i}{N} - R_{(i)}$	0.15	0.26	0.16	-	0.07
$\frac{i-1}{N}$	0	0.20	0.40	0.60	0.80
$R_{(i)} - \frac{i-1}{N}$	0.05	-	0.04	0.21	0.13

اذن $D^{+} = 0.26$ وان $D^{-} = 0.21$، عليه فان

$$D = \max(D^{+}, D^{-}) = \max(0.26, 0.21) = 0.26$$

تقارن قيمة $D = 0.26$ مع قيمة D_{α} الجدولية عند مستوى المعنوية 0.05 و بعدد مشاهدات 5 ، اذ ان $D_{0.05} = 0.565$. بما ان D المستخرجة اقل من $D_{0.05}$ الجدولية. عليه تقبل فرضية العدم القائلة بان البيانات تتوزع وفق التوزيع المنتظم.

واجب: استخدم اختبار كولموغروف- سميرنوف بدرجة حرية $\alpha = 0.05$ لبيان فيما اذا كانت الارقام العشوائية الاتية تتوزع وفق التوزيع المنتظم $[0, 1]$، علما ان $m = 64$.

$$1, 13, 41, 21, 17, 29, 57, 37, 33, 45, 9, 53, 49, 61, 25, 5$$

	ترتيب البيانات تصاعديا	i	$R_i = \frac{X_i}{m}$	$\frac{i}{N}$	D^{+} $\max_{1 \le i \le N} \left\{ \frac{i}{N} - R_{(i)} \right\}$	D^{-} $\max_{1 \le i \le N} \left\{ R_{(i)} - \frac{i-1}{N} \right\}$
1	1	1	0.0156	0.0625	0.0469	0.0156
13	5	2	0.078	0.1250	0.0470	0.0155
41	9	3	0.141	0.1875	0.0465	0.0160
21	13	4	0.203	0.2500	0.0470	0.1155
17	17	5	0.265	0.3125	0.0475	0.0150
29	21	6	0.328	0.3750	0.0470	0.0155
57	25	7	0.391	0.4375	0.0465	0.0160
37	29	8	0.453	0.5000	0.0475	0.0155
33	33	9	0.516	0.5625	0.0465	0.0160
45	37	10	0.578	0.6250	0.0470	0.0155
9	41	11	0.641	0.6875	0.0465	0.0160
53	45	12	0.703	0.7500	0.0470	0.0155
49	49	13	0.765	0.8125	0.0475	0.0150
61	53	14	0.828	0.8750	0.0470	0.0155
25	57	15	0.890	0.9375	0.0475	0.0150
5	61	16	0.953	1.000	0.0470	0.0155
					$D^{+} = 0.0475$	$D^{-} = 0.0160$

عليه فان

$$D = \max(D^{+}, D^{-}) = \max(0.0475, 0.0160) = 0.0475$$

تقارن قيمة $D = 0.0475$ المستخرجة مع قيمة $D_{0.05} = 0.328$ الجدولية. بما ان D المستخرجة اقل من $D_{0.05}$ الجدولية عند مستوى المعنوية 0.05. عليه تقبل فرضية العدم وترفض الفرضية البديلة، اي ان البيانات تتوزع وفق التوزيع المنتظم.

2. اختبار مربع كاي:

ان صيغة اختبار مربع كاي هي الاتي:

$$\chi_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

اذ ان O_i هي الارقام المشاهدة في الفئة i ، وان E_i هي الارقام المتوقعة في الفئة i وتعطى بالصيغة الاتية

$$E_i = \frac{N}{n}$$

اذ ان N هي العدد الكلي للمشاهدات وان n هي عدد الفئات.

يمكن بيان فيما اذا كان توزيع المعاينة χ_0^2 يتوزع تقريبا وفق توزيع مربع كاي بدرجة حرية $n - 1$، وذلك من خلال المقارنة بين قيمة مربع كاي المحسوبة مع قيمة مربع كاي الجدولية، فاذا كانت χ_0^2 المحسوبة اكبر من χ^2 الجدولية، عندئذ ترفض فرضية العدم، بمعنى ان بيانات العينة لا تتوزع وفق التوزيع المنتظم. اما اذا كانت χ_0^2 المحسوبة اقل من χ^2 الجدولية، عندئذ تقبل فرضية العدم، بمعنى انه لايوجد فرق بين توزيع بيانات العينة وبين التوزيع المنتظم.

عند تطبيق اختبار مربع كاي على مجموعة من البيانات بحجم $n = 100$، فانه يمكن استخدام فترات طول متساوية قدرها من خمسة الى عشرة. وبشكل عام فانه يمكن استخدام $E_i \geq 5$.

ملاحظة: يمكن استخدام كل من اختبار كواموكروف-سميرنوف واختبار مربع كاي لاختبار فيما اذا كانت البيانات تتوزع وفق التوزيع المنتظم، مع ملاحظة ان اختبار كولموكروف يعد الافضل في حالة صغر حجم العينة، بينما مربع كاي يكون اكثر ملائمة عندما يكون حجم العينة كبير، اي ان $N \geq 50$.

مثال: استخدم اختبار مربع كاي بدرجة حرية $\alpha = 0.05$ لبيان فيما اذا كانت الارقام العشوائية الاتية تتوزع وفق التوزيع المنتظم القياسي.

0.34, 0.90, 0.25, 0.89, 0.87, 0.44, 0.12, 0.21, 0.46 0.67
0.83 , 0.76, 0.79, 0.64, 0.70, 0.81, 0.94, 0.74, 0.22 0.74
0.96, 0.99, 0.77, 0.67, 0.56, 0.41, 0.52, 0.73, 0.99 0.02
0.47, 0.30, 0.17, 0.82, 0.56, 0.05, 0.45, 0.31, 0.78 0.05
0.79, 0.71, 0.23, 0.19, 0.82, 0.93, 0.65, 0.37, 0.39 0.42

في حالة اختيار عدد الفترات $n = 10$. فان طول الفترة هو $0.1 = \frac{1}{10}$.

اذن الفترات هي $[0, 0.1), [0.1, 0.2), [0.2, 0.3), [0.3, 0.4), \ldots [0.9, 1)$

Intervals	O_i	$E_i = \frac{N}{n}$	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$

$[0, 0.1)$	3	5	-2	4	0.8
$[0.1, 0.2)$	3	5	-2	4	0.8
$[0.2, 0.3)$	5	5	0	0	0.0
$[0.3, 0.4)$	4	5	-1	1	0.2
$[0.4, 0.5)$	6	5	1	1	0.2
$[0.5, 0.6)$	3	5	-2	4	0.8
$[0.6, 0.7)$	5	5	0	0	0.0
$[0.7, 0.8)$	9	5	4	16	3.2
$[0.8, 0.9)$	7	5	2	4	0.8
$[0.9, 1.0)$	5	5	0	0	0.0
	50	50			$\sum = 6.8$

تقارن قيمة $\chi_0^2 = 6.8$ مع قيمة χ^2 الجدولية عند مستوى معنوية 0.05 وبدرجة حرية $n - 1$. بما ان قيمة χ^2 الجدولية هي 16.9. عليه تقبل فرضية العدم القائلة بان البيانات تتبع التوزيع المنتظم القياسي.

2. اختبار الارتباط الذاتي: لاختبار ارتباط الارقام مع بعضها البعض، بمعنى مقارنة ارتباط بيانات العينة مع الارتباط المتوقع لهذه البيانات والذي يفترض ان لا يكون هناك من ارتباط بين هذه الارقام.

اما اختبار الاستقلالية، فان الفرضية هي الاتي:

$$H_0: R_i \sim independetly$$

$$H_1: R_i \nsim independetly$$

اي اختبار الفرضية الصفرية القائلة بان الارقام التي تم توليدها هي ارقام مستقلة، مقابل الفرضية البديلة القائلة بان الارقام غير مستقلة.

ان احصاءة اختبار البيانات هل مرتبطة ام لا من خلال الصيغة التالية

$$Z_0 = \frac{\hat{\rho}_{i\ell}}{\sigma_{\hat{\rho}_{i\ell}}}$$

والتي يتم توزيعها طبيعيا بمتوسط صفر وتباين 1، تحت افتراض الاستقلالية، و M الكبيرة.

ويتم حساب كل من $\hat{\rho}_{i\ell}$ و $\sigma_{\hat{\rho}_{i\ell}}$ والذي يمثلان الارتباط والانحراف المعياري للمقدر من خلال الصيغ التالية

$$\hat{\rho}_{im} = \frac{1}{M+1}\left[\sum_{k=0}^{M} R_{i+km}R_{i+(k+1)m}\right] - 0.25$$

$$\sigma_{\hat{\rho}_{im}} = \frac{\sqrt{13M+7}}{12(M+1)}$$

يمثل i بداية الرقم الذي يتم اختيارة من حجم العينة ويمثل

Ri, Ri+, Ri+2, . . . , Ri+(M+1)m تسلسل لارقام العشوائية التي يتم اختيارها و m تمثل مقدار الزيادة و M تساوي m-1

$$i + (M+1)m \leq N$$

مثال

1. اختبر عشوائية الارقام الاتية باستخدام اختبار الارتباط الذاتي وبحسب المعطيات الاتية:
$i = 3$، $m = 5$. $Z_{0.025} = 1.96$.

0.12	0.01	0.23	0.28	0.89	0.31	0.64	0.28	0.83	0.93
0.99	0.15	0.33	0.35	0.91	0.41	0.60	0.27	0.75	0.88
0.68	0.49	0.05	0.43	0.95	0.58	0.19	0.36	0.69	0.87

الحل

M تحدد من المشاهدات وتكون اقل من 30 من خلال المعادلة التالية

$$i + (M+1)m \leq N$$

عددها في المثال يساوي 4

$3 + (M + 1)5 \leq 30$). Then,

ويمثل تسلسل الرقم العشوائي الثالث 0.23 $R_{i+km} = R_3$

ويمثل تسلسل الرقم العشوائي الثامن 0.28 $R_{i+(k+1)m} = R_8$

يتم اولا حساب الارتباط من خلال المعادلة التالية

$$\hat{\rho}_{im} = \frac{1}{M+1}\left[\sum_{k=0}^{M} R_{i+km}R_{i+(k+1)m}\right] - 0.25$$

$$\begin{aligned}\hat{\rho}_{35} &= \frac{1}{4+1}[(0.23)(0.28) + (0.28)(0.33) + (0.33)(0.27) + (0.27)(0.05) \\ &\quad +(0.05)(0.36)] - 0.25 \\ &= -0.1945\end{aligned}$$

ثم يتم حساب الانحراف المعياري من الصيغة التالية

$$\sigma_{\hat{\rho}_U} = \frac{\sqrt{13M+7}}{12(M+1)}$$

$$\sigma_{\hat{\rho}_{35}} = \frac{\sqrt{13(4)+7}}{12(4+1)} = 0.1280$$

ثم يتم حساب اختبار الارتباط الذاتي من الصيغة التالية

Then, the test statistic assumes the value

$$Z_0 = -\frac{0.1945}{0.1280} = -1.516$$

Now, the critical value from Table A.3 is

$$z_{0.025} = 1.96$$

وبالمقارنة مع القيمة الجدولية نلاحظ ان الجدولية اكبر من المحسوبة يفشل في الرفض

Appendix 3

Kolmogorov–Smirnov Tables

Critical values, $d_{alpha};(n)^a$, of the maximum absolute difference between sample $F_n(x)$ and population $F(x)$ cumulative distribution.

Number of trials, n	Level of significance, α			
	0.10	0.05	0.02	0.01
1	0.95000	0.97500	0.99000	0.99500
2	0.77639	0.84189	0.90000	0.92929
3	0.63604	0.70760	0.78456	0.82900
4	0.56522	0.62394	0.68887	0.73424
5	0.50945	0.56328	0.62718	0.66853
6	0.46799	0.51926	0.57741	0.61661
7	0.43607	0.48342	0.53844	0.57581
8	0.40962	0.45427	0.50654	0.54179
9	0.38746	0.43001	0.47960	0.51332
10	0.36866	0.40925	0.45662	0.48893
11	0.35242	0.39122	0.43670	0.46770
12	0.33815	0.37543	0.41918	0.44905
13	0.32549	0.36143	0.40362	0.43247
14	0.31417	0.34890	0.38970	0.41762
15	0.30397	0.33760	0.37713	0.40420
16	0.29472	0.32733	0.36571	0.39201
17	0.28627	0.31796	0.35528	0.38086
18	0.27851	0.30936	0.34569	0.37062
19	0.27136	0.30143	0.33685	0.36117
20	0.26473	0.29408	0.32866	0.35241
21	0.25858	0.28724	0.32104	0.34427
22	0.25283	0.28087	0.31394	0.33666
23	0.24746	0.27490	0.30728	0.32954
24	0.24242	0.26931	0.30104	0.32286

Practical Reliability Engineering, Fifth Edition. Patrick D. T. O'Connor and Andre Kleyner.

Critical values, $d_{alpha};(n)^a$, of the maximum absolute difference between sample $F_n(x)$ and population $F(x)$ cumulative distribution.

Number of trials, n	Level of significance, α			
	0.10	0.05	0.02	0.01
25	0.23768	0.26404	0.29516	0.31657
26	0.23320	0.25907	0.28962	0.31064
27	0.22898	0.25438	0.28438	0.30502
28	0.22497	0.24993	0.27942	0.29971
29	0.22117	0.24571	0.27471	0.29466
30	0.21756	0.24170	0.27023	0.28987
31	0.21412	0.23788	0.26596	0.28530
32	0.21085	0.23424	0.26189	0.28094
33	0.20771	0.23076	0.25801	0.27677
34	0.20472	0.22743	0.25429	0.27279
35	0.20185	0.22425	0.26073	0.26897
36	0.19910	0.22119	0.24732	0.26532
37	0.19646	0.21826	0.24404	0.26180
38	0.19392	0.21544	0.24089	0.25843
39	0.19148	0.21273	0.23786	0.25518
40[b]	0.18913	0.21012	0.23494	0.25205

[a]Values of $d_\alpha(n)$ such that $p(\max)|F^n(x) - F(x)|d^\alpha(n) = \alpha$.

[b]$N > 40 \approx \frac{1.22}{N^{1/2}}, \frac{1.36}{N^{1/2}}, \frac{1.51}{N^{1/2}}$ and $\frac{1.63}{N^{1/2}}$ for the four levels of significance.