

Chapitre 1 : Génération de nombres aléatoires uniformes

Université Hassiba Benbouali de Chlef

Plan du Cours

Ce chapitre traite de la simulation de nombres (pseudo) aléatoires distribués uniformément sur l'intervalle $[0, 1]$.

- ▶ Générateur congruentiel linéaire

Références

- Premiers pas en simulation. Springer.
- ▶ Dodge, Y., & Melfi, G. (2008).



Exemplaires

Code-barres	Cote	Support	Localisation	Section	Disponibilité
FS2013/0101-2	515.2 DOD	Ouvrage	Bibliothèque de la Faculté des Sciences	Mathématique	Disponible

Objectifs

La **simulation stochastique** est une technique de plus en plus utilisée en statistique comme dans la plupart des sciences appliquées :

- ▶ étude d'estimateurs ou de tests statistiques,
- ▶ méthodes de Monte-Carlo pour le calcul d'intégrales,
- ▶ la simulation de files d'attente,
- ▶ simulation de phénomènes physiques
- ▶ simulation de fonctionnements de réseaux ou de systèmes complexes,
- ▶ cryptographie,
- ▶ imagerie, ...

Avantages

- ▶ la simulation est non destructrice et peu coûteuse ;
- ▶ le système considéré n'a pas besoin d'exister ;
- ▶ la simulation est facile à répéter ;
- ▶ l'évolution dans la simulation peut être plus rapide que dans la réalité ;
- ▶ la simulation permet de considérer des modèles très complexes impossibles à traiter analytiquement.

Générateurs de nombres aléatoires

Théoriquement, la génération de nombres aléatoires suivant une loi donnée se ramène à la génération de suites de variables aléatoires indépendantes de loi uniforme sur $[0, 1]$.

Définition 2.1

On appelle suite de nombres pseudo-aléatoires une suite déterministe qui possède des propriétés statistiques identiques à celles que posséderait une suite réellement aléatoire.

Générateurs de nombres aléatoires (suite)

Caractéristiques

- ▶ Distribution uniforme
- ▶ Non corrélation
- ▶ Reproductibilité

Générateurs physiques

- ▶ Le jeu de "pile" ou "face" permet de générer une variable aléatoire de Bernoulli qui prend les valeurs 0 ou 1 avec des probabilités identiques, égale à $\frac{1}{2}$.



- ▶ Le jet d'un dé parfait permet de générer une variable aléatoire à valeurs dans $\{1, 2, 3, 4, 5, 6\}$ avec des probabilités identiques égales à $\frac{1}{6}$.



Tables de nombres au hasard

- ▶ Ces tables sont telles que la suite des nombres qui y figurent est assimilable à la réalisation de tirages avec remise dans une urne à dix catégories de boules figurant à proportions égales.

Tables de nombres au hasard (suite)

NOMBRES AU HASARD

13407	62899	78937	90525	25033	56358	78902	47008	72488	57949
50230	63237	94083	93634	71652	02656	57532	60307	91619	48916
84980	62458	09703	78397	66179	46982	67619	39254	90763	74056
22116	33646	17545	31321	65772	86506	09811	82848	92211	51178
68645	15068	56898	87021	40115	27524	42221	88293	67592	06430
26518	39122	96561	56004	50260	68648	85596	83979	09041	62350
36493	41666	27871	71329	69212	57932	65281	57233	07732	58439
77402	12994	59892	85581	70823	53338	34405	67080	16568	00854
83679	97154	40341	84741	08967	73287	94952	59008	95774	44927
71802	39356	02981	89107	79788	51330	37129	31898	34011	43304
57494	72484	22676	44311	15356	05348	03582	66183	68392	86844
73364	38416	93128	10297	11419	82937	84389	88273	96010	09843
14499	83965	75403	18002	45068	54257	18085	92626	60911	39137
40747	03084	07734	88940	88722	85717	73810	79866	84853	68647
42237	59122	92855	62097	81276	06318	81607	00565	56626	77422
32934	60227	58707	44858	36081	70981	01291	68707	45427	82145
05764	14284	73069	80830	17231	42936	48472	18782	51646	37564
32706	94879	93188	66049	25988	46656	35365	13800	83745	40141
22190	27559	95668	53261	21676	98943	43618	42110	93402	93997
81616	15641	94921	95970	63506	22007	29966	38144	62556	07864
26099	65801	69870	84446	58248	21282	56938	54729	67757	68412
71874	61692	80001	21430	02305	59741	34262	15157	27545	14522
08774	29689	42245	51903	69179	96682	91819	60812	47631	50609
37294	92028	56850	83380	05912	29830	37612	15593	73198	99287
33912	37996	78967	57201	66916	73998	54289	07147	84313	51939
63610	61475	26980	23804	54972	72068	19403	53756	04021	98022
01570	41702	30282	54647	06077	29354	95704	75928	21811	88274
24159	77787	38973	82178	68802	90245	01805	06550	06785	52147
92834	52941	88301	22127	23459	40229	74678	21859	98645	72388
16178	60063	59284	16729	48003	44634	08623	32752	40472	05470
81808	32980	80660	98391	62243	19678	39551	18398	36918	43543
88628	82612	02809	86689	68337	86689	11120	28638	72850	03650
62249	65757	12273	91261	96983	15082	83851	77682	81728	48365
84541	99891	01585	96711	29712	02877	70955	59693	26838	96011
89052	39061	99811	69831	47234	93263	47386	17462	18874	74210
33645	89282	00384	10888	21789	14093	06268	46460	97660	23490
61618	19275	40744	22482	12424	19860	53166	41836	28265	41836
68136	06546	40029	47946	19526	27122	42515	55048	23912	81105
74005	34558	93779	96120	01695	47720	88646	73520	40050	90553
54437	88825	07943	87943	31709	04629	04626	64838	92133	44221
01990	94762	89926	84764	19159	95355	98213	17704	47400	30837
02404	42408	67981	43684	55467	47030	42545	43920	11199	36521
59253	71535	26149	35629	87127	45581	00185	01041	46662	98897
20471	13914	92330	79938	69649	57964	97149	41628	78664	80727
65946	60766	74084	22484	49514	89820	41310	19722	07045	28808
00939	47818	75949	44707	49105	06777	31998	79942	98351	10265
49952	29123	45950	67578	13524	03023	18046	75287	74989	58152
17328	70732	46319	26950	19037	02831	36558	82712	05590	64941
19420	70215	90476	76400	51553	12158	14658	15656	37895	94559
19121	41190	49145	05373	00755	17817	22757	76116	76977	94570

Tables de nombres au hasard (suite)

- ▶ On choisit arbitrairement un nombre dans la table, il servira de point de départ.
- ▶ On prélève alors les chiffres au fur et à mesure qu'on les trouve dans la table dans une direction et un sens fixés à l'avance.

Exemple

Pour choisir des nombres au hasard dans $[0, 1]$, avec une précision de 10^{-2} , on procède comme suit :

- ▶ On choisit arbitrairement la ligne 5 et la colonne 12, on lit de gauche à droite,
- ▶ On prélève les nombres 68, 98, 87, 02, 14, ... et on obtient donc les nombres aléatoires
 $u_1 = 0.68, u_2 = 0.98, u_3 = 0.87, u_4 = 0.02, u_5 = 0.14, \dots$

Méthode du carré médian

Cette méthode a été introduite par John Von Neumann (1946).

Principe

- ▶ On génère une suite de nombres ayant chacun m chiffres, où m est un nombre pair.
- ▶ Le successeur d'un nombre de cette suite est obtenue en élevant ce nombre au carré puis en retenant les m chiffres du milieu.

Méthode du carré médian (suite)

Exemple

Soit le nombre «2020».

1 $2020^2 = 4080400$

2 on récupère les chiffres du milieu : 8040. C'est la sortie du générateur.

3 $8040^2 = 64641600$

4 on récupère les chiffres du milieu : 6416.

Rappel : Congruence et modulo

Définition 3.1


Deux nombres a et b sont dits équivalents, ou **congruents, modulo** m si la différence entre a et b est un entier divisible par m .

$$a \equiv b \pmod{m} \quad \Leftrightarrow \quad \frac{a - b}{m} = k, \quad k \in \mathbb{Z}.$$

Exemple

On a $5 \equiv 14 \pmod{3}$ car $\frac{14 - 5}{3} = 3$.

Remarque

La fonction modulo est définie dans  par %.

Méthode du carré médian (suite)

Le raisonnement de Von Neumann donne lieu à l'algorithme suivant pour générer une suite de nombres en base 10 :

Algorithme

- Initialisation : choisir x_0 à m chiffres ;

$$u_0 = \frac{x_0}{10^m}$$

- Itération :

$$x_{i+1} = \left\lfloor \frac{x_i^2}{10^{m/2}} \right\rfloor \bmod 10^m,$$

$$u_{i+1} = \frac{x_{i+1}}{10^m}.$$

Méthode du carré médian (suite)

Remarque

Ces nombres ne sont pas aléatoires, ils dépendent du premier choisi, et ont, malheureusement, une faible période et donc ils sont peu employés en pratique.

Méthode de Fibonacci

Cette méthode est basée sur la suite de Fibonacci, modulo la valeur maximale désirée :

$$x_n = (x_{n-1} + x_{n-2}) \bmod m$$

avec x_0 et x_1 en entrée.

Générateurs congruentiels linéaires

Définition 4.1

Un générateur de congruence linéaire (GCL) est défini par la fonction de transition suivante :

$$x_i = (ax_{i-1} + c) \bmod m,$$

où $0 \leq x_i < m$ et

- ▶ a est appelé le **multiplicateur** ;
- ▶ c est appelé l'**incrément** ;
- ▶ m est appelé le **module** ;
- ▶ x_0 (un nombre quelconque) est l'**amorce** (**seed**).

Générateurs congruentiels linéaires (suite)

Remarque

- Pour obtenir des nombre uniformes sur $[0, 1]$, il suffit de définir

$$u_i = \frac{x_i}{m}.$$

- Pour $c = 0$, on obtient le **générateur multiplicatif** et pour $c \neq 0$, on obtient le **générateur mixte**.

Générateurs congruentiels linéaires (suite)

Exemple 4.2

Cet exemple illustrant la méthode emploie de petits nombres afin de faciliter les calculs.

Soit

$$x_i = (10x_{i-1} + 5) \bmod 12.$$

On note que les valeurs possibles sont $\{0, 1, 2, \dots, 11\}$. Si on fixe $x_0 = 5$, on trouve :

$$x_1 = 7, x_2 = 3, x_3 = 11, x_4 = 7, x_5 = 3, x_6 = 11, x_7 = 7, \dots$$

Générateurs congruentiels linéaires (suite)

Propriétés

- ▶ Le nombre de valeurs possibles pouvant être fournies par un générateur de congruence linéaire est au plus égal à m .
- ▶ D'autre part, si un nombre apparaît une deuxième fois, tous les nombres qui le suivent apparaissent aussi une deuxième fois et selon le même ordre. Ainsi, un générateur de congruence linéaire est nécessairement périodique. Sa période maximale vaut m .
- ▶ la période maximale n'est pas toujours atteinte (voir exemple 4.2 ci-dessus)

Générateurs congruentiels linéaires (suite)

Théorème de Hull et Dobel (1962)

Le paramètre m étant donné, pour atteindre la période maximale soit m , il faut et il suffit que :

- 1 c et m sont premiers entre eux
- 2 $(a - 1)$ est un multiple de chaque nombre premier qui divise m
- 3 si m est un multiple de 4 alors $(a - 1)$ l'est aussi.

Générateurs congruentiels linéaires (suite)

Théorème de Greenberger 1961

Dans le cas où $m = 2^\ell$ avec $\ell \geq 2$, la période du GCL est maximale (vaut m) si et seulement si

- 1 c est impair
- 2 $a \equiv 1 \pmod{4}$.

Générateurs congruentiels linéaires (suite)

Exemples

- ▶ Glibc(GCC) et ANSI C

$$m = 2^{32}, a = 1103515245 \text{ et } c = 12345$$

- ▶ Numerical Recipes

$$m = 2^{32}, a = 1664525 \text{ et } c = 1013904223$$

- ▶ MS Visual/Quick C/C++

$$m = 2^{32}, a = 214013 \text{ et } c = 2531011$$

Exemple

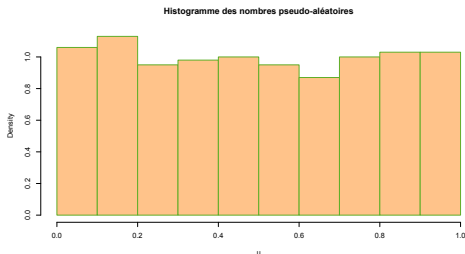


FIGURE 1 – Histogramme des nombres pseudo-aléatoires issus d'un GLC avec $m = 2048$, $a = 1229$ et $c = 1$.

Les paires $(u_1, u_2), (u_2, u_3), \dots, (u_i, u_{i+1}), \dots$ devraient être uniformément réparties dans un carré. Hélas, comme la Figure 2 le montre, cela n'est pas toujours le cas, et ce constat constitue le point faible principal de la méthode de congruence simple.

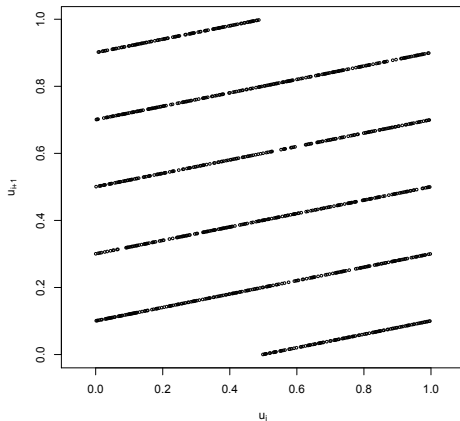


FIGURE 2 – Graphe des couples (u_i, u_{i+1}) issus du même GLC.

Exemple du générateur Randu avec $m = 2^{31}$, $a = 65539$ et $c = 0$

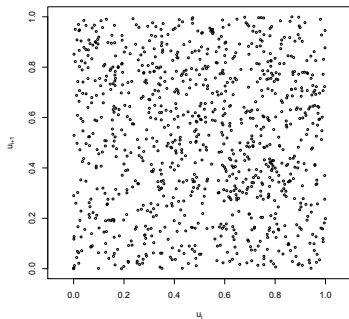


FIGURE 3 – Graphe des couples (u_i, u_{i+1}) .

Exemple du générateur Randu avec $m = 2^{31}$, $a = 65539$ et $c = 0$

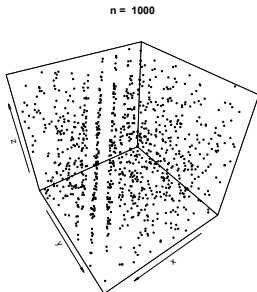


FIGURE 4 – Graphe des triplets (u_i, u_{i+1}, u_{i+2}) .

Exemple générateurs ANSI C

Autres générateurs

- ▶ La méthode de congruence avec retard
- ▶ La méthode de congruence avec mélange
- ▶ La méthode du registre à décalage avec rétroaction linéaire

Test de Kolmogorov-Smirnov

Soit (X_1, \dots, X_n) un échantillon de loi \mathbb{F} et \mathbb{F}_n la fonction de répartition empirique associée.

- ▶ On désire effectuer un test d'hypothèses concernant \mathbb{F} . Les hypothèses à tester sont :

$$\begin{cases} \mathcal{H}_0 & : \mathbb{F} = \mathbb{F}_0 \\ \mathcal{H}_1 & : \mathbb{F} \neq \mathbb{F}_0 \end{cases}$$

avec \mathbb{F}_0 une fonction de répartition continue spécifiée.

- ▶ La statistique $D_n(\mathbb{F}_0)$ de **Kolmogorov-Smirnov** est définie par

$$D_n(\mathbb{F}) = \sup_{x \in \mathbb{R}} |\mathbb{F}_n(x) - \mathbb{F}(x)|,$$

qui est la distance de la norme uniforme entre fonctions de répartitions.

Test de Kolmogorov-Smirnov (suite)

- ▶ Intuitivement, $D_n(\mathbb{F}_0)$ doit être petite si \mathcal{H}_0 est vraie. On peut montrer que sous \mathcal{H}_0 pour $n \rightarrow \infty$, $D_n(\mathbb{F}_0) \rightarrow 0$. La région de rejet du test est définie par $D_n(\mathbb{F}_0) > C$.
- ▶ Sous l'hypothèse \mathcal{H}_0 , on a, pour tout $t \geq 0$:

$$\lim_{n \rightarrow \infty} \mathbb{P}(\sqrt{n}D_n(\mathbb{F}) \leq t) = 1 - 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2 t^2}.$$

- ▶ Pour $n \leq 100$ et quelques valeurs de α les valeurs d_n telles que $\mathbb{P}(D_n(\mathbb{F}) \leq d_n) = \alpha$, sont tabulées.

TABLE 1 – Table du test de Kolmogorov-Smirnov

$n \backslash \alpha$	0.2	0.15	0.1	0.05	0.01
1	0.900	0.925	0.950	0.975	0.995
2	0.684	0.726	0.776	0.842	0.929
3	0.565	0.597	0.642	0.708	0.828
4	0.494	0.525	0.564	0.624	0.733
5	0.446	0.474	0.510	0.565	0.669
6	0.410	0.436	0.470	0.521	0.618
7	0.381	0.405	0.438	0.486	0.577
8	0.358	0.381	0.411	0.457	0.543
9	0.339	0.360	0.388	0.432	0.514
10	0.322	0.342	0.368	0.410	0.490
11	0.307	0.326	0.352	0.391	0.468
12	0.295	0.313	0.338	0.375	0.450
13	0.284	0.302	0.325	0.361	0.433
14	0.274	0.292	0.314	0.349	0.418
15	0.266	0.283	0.304	0.338	0.404
16	0.258	0.274	0.295	0.328	0.392
17	0.250	0.266	0.286	0.318	0.381
18	0.244	0.259	0.278	0.309	0.371
19	0.237	0.252	0.272	0.301	0.363
20	0.231	0.246	0.264	0.294	0.356
25	0.210	0.220	0.240	0.270	0.320
30	0.190	0.200	0.220	0.240	0.290
35	0.180	0.190	0.210	0.230	0.270
> 35	$\frac{1.07}{\sqrt{n}}$	$\frac{1.14}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

- ▶ De manière équivalente on utilise la statistique de test D_n définie comme

$$D_n(\mathbb{F}) = \max \left\{ D_n^+(\mathbb{F}), D_n^-(\mathbb{F}) \right\}$$

avec

$$D_n^+(\mathbb{F}) = \sup_{x \in \mathbb{R}} [\mathbb{F}_n(x) - \mathbb{F}(x)] \quad \text{et} \quad D_n^-(\mathbb{F}) = \sup_{x \in \mathbb{R}} [\mathbb{F}(x) - \mathbb{F}_n(x)] .$$

- ▶ Soit $X_{(1)} < \dots < X_{(n)}$ les statistiques d'ordre de l'échantillon. La fonction de répartition empirique est définie par :

$$\mathbb{F}_n(x) = \begin{cases} 0 & \text{pour } x < X_{(1)}, \\ \frac{i}{n} & \text{pour } X_{(i)} \leq x < X_{(i+1)}, \\ 1 & \text{pour } x > X_{(n)}. \end{cases}$$

► Alors

$$\begin{aligned} D_n^+(\mathbb{F}) &= \max_{0 \leq i \leq n} \sup_{X_{(i)} \leq x < X_{(i+1)}} \left[\frac{i}{n} - \mathbb{F}(x) \right] \\ &= \max_{0 \leq i \leq n} \left[\frac{i}{n} - \inf_{X_{(i)} \leq x < X_{(i+1)}} \mathbb{F}(x) \right] \\ &= \max_{0 \leq i \leq n} \left[\frac{i}{n} - \mathbb{F}(X_{(i)}) \right]. \end{aligned}$$

De même :

$$D_n^-(\mathbb{F}) = \max_{0 \leq i \leq n} \left[\mathbb{F}(X_{(i)}) - \frac{i-1}{n} \right]$$

Procédure

- 1 classer les valeurs observées par ordre croissant ;
- 2 calculer $\mathbb{F}_n(X_{(i)})$ et $\mathbb{F}(X_{(i)})$;
- 3 calculer les valeurs absolues des écarts $|\mathbb{F}_n(X_{(i)}) - \mathbb{F}(X_{(i)})|$ entre \mathbb{F} et la distribution empirique ;
- 4 la distance de Kolmogorov-Smirnov est le plus grand de tous ces écarts ;
- 5 on conclut le test en acceptant l'hypothèse \mathcal{H}_0 si la distance calculée est inférieure à la valeur critique donnée dans la table et en la rejetant sinon.

Exemple

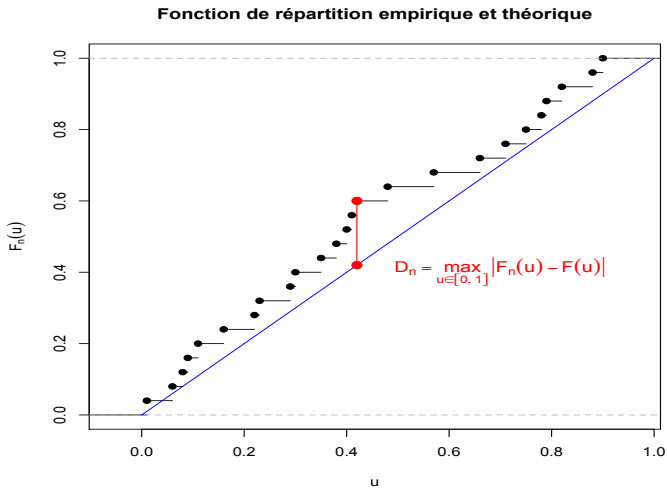
Exemple

Utilisez le test de Kolmogorov-Smirnov pour tester l'hypothèse selon laquelle les 25 valeurs du tableau 2 forment un échantillon aléatoire à partir de la distribution uniforme sur l'intervalle $[0, 1]$.

TABLE 2 – Une suite de 25 nombres pseudo-aléatoires

0.42	0.06	0.88	0.40	0.90
0.38	0.78	0.71	0.57	0.66
0.48	0.35	0.16	0.22	0.08
0.11	0.29	0.79	0.75	0.82
0.30	0.23	0.01	0.41	0.09


u_i	$u_{(i)}$	$\mathbb{F}_n(u_{(i)})$	$\mathbb{F}(u_{(i)})$	$ \mathbb{F}_n(u_{(i)}) - \mathbb{F}(u_{(i)}) $
0.42	0.01	0.04	0.01	0.03
0.06	0.06	0.08	0.06	0.02
0.88	0.08	0.12	0.08	0.04
0.40	0.09	0.16	0.09	0.07
0.90	0.11	0.20	0.11	0.09
0.38	0.16	0.24	0.16	0.08
0.78	0.22	0.28	0.22	0.06
0.71	0.23	0.32	0.23	0.09
0.57	0.29	0.36	0.29	0.07
0.66	0.30	0.40	0.30	0.10
0.48	0.35	0.44	0.35	0.09
0.35	0.38	0.48	0.38	0.10
0.16	0.40	0.52	0.40	0.12
0.22	0.41	0.56	0.41	0.15
0.08	0.42	0.60	0.42	0.18
0.11	0.48	0.64	0.48	0.16
0.29	0.57	0.68	0.57	0.11
0.79	0.66	0.72	0.66	0.06
0.75	0.71	0.76	0.71	0.05
0.82	0.75	0.80	0.75	0.05
0.30	0.78	0.84	0.78	0.06
0.23	0.79	0.88	0.79	0.09
0.01	0.82	0.92	0.82	0.10
0.41	0.88	0.96	0.88	0.08
0.09	0.90	1.00	0.90	0.10



- ▶ La distance la plus élevée s'établit à 0.18.
- ▶ La table de Kolmogorov-Smirnov pour $n = 25$ et $\alpha = 5\%$ donne la valeur critique 0.26931. Puisque $0.18 < 0.26931$, on **accepte** l'hypothèse \mathcal{H}_0 .

Remarque

- ▶ Les logiciels de calcul statistique expriment souvent le résultat d'un test en fournissant une grandeur appelée **p-valeur** (en anglais p-value).
- ▶ la règle de décision, en termes de p-valeur, stipule qu'on **rejette** l'hypothèse nulle \mathcal{H}_0 dès que **la p-valeur est inférieure au risque α choisi**.

- ▶ Voici la sortie du logiciel  pour le test précédent :
`> ks.test(u,"punif",alternative="two.sided")`

One-sample Kolmogorov-Smirnov test

data: u

$D = 0.18$, p-value = 0.3501

alternative hypothesis: two-sided

- ▶ On retrouve bien la valeur de la distance $D_n = 0.18$. La p-valeur 0.3501 est nettement supérieure à 0.05, donc on accepte effectivement l'hypothèse \mathcal{H}_0 .

Test du χ^2

- ▶ On désire tester si $U_1, \dots, U_n \sim \mathcal{U}[0, 1]$.
- ▶ Diviser l'intervalle unité en k sous-intervalles : $\left[0, \frac{1}{k} \right[$,
 $\left[\frac{1}{k}, \frac{2}{k} \right[$, \dots , $\left[\frac{k-1}{k}, 1 \right]$.
- ▶ Comptez combien des n observations tombent dans les k classes. On dispose alors des effectifs des k classes : n_1, n_2, \dots, n_k .
- ▶ On définit les effectifs théoriques pour la loi uniforme $n_{t,i} := \frac{n}{k}$.
- ▶ On calcule la statistique : $\chi^2 = \sum_{i=1}^k \frac{(n_i - n_{t,i})^2}{n_{t,i}}$.

Test du χ^2 (suite)

- ▶ On rejette l'hypothèse \mathcal{H}_0 si $\chi^2 > \chi_{k-1, 1-\alpha}^2$, où $\chi_{k-1, 1-\alpha}^2$ est le quantile d'ordre $1 - \alpha$ de la loi Khi-deux à $k - 1$ degrés de liberté.

Remarque

- ▶ Le test du χ^2 convient aux données qui présentent un nombre fini de modalités (des variables discrètes finies ou des variables continues qu'on a regroupées en un nombre fini de classes).
- ▶ Pour que ce test soit valide, il faut que $n_i \geq 5$ pour tout i . En particulier, on ne doit pas l'appliquer si une classe a un effectif nul.

Exemple

On veut tester la validité d'un générateur de nombres aléatoires défini par

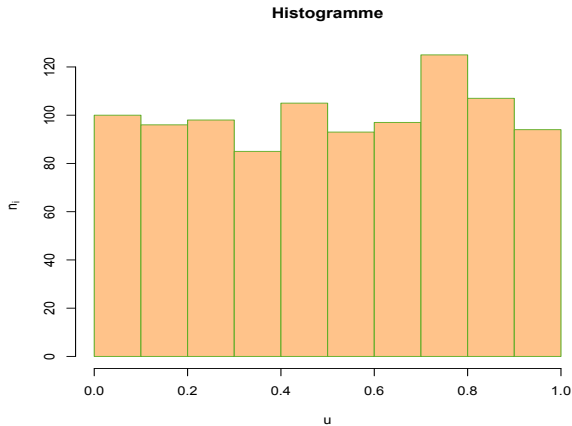
$$x_n = (125x_{n-1} + 1) \bmod 2^{12}.$$

On génère un échantillon de taille $n = 1000$ avec $x_0 = 1$.

- ▶ On regroupe la série simulée en 10 classes, par exemple.



Exemple (suite)



Exemple (suite)

- ▶ On calcule maintenant la statistique

$$\begin{aligned}\chi^2 &= \sum_{i=1}^{10} \frac{(n_i - n_{t,i})^2}{n_{t,i}} \\ &= \frac{(100 - 100)^2}{100} + \frac{(96 - 100)^2}{100} + \dots + \frac{(94 - 100)^2}{100} \\ &= 10.38\end{aligned}$$

- ▶ Comme on a 10 classes, le nombre de degrés de liberté est 9. La table du Khi-deux, donne un quantile égal à 16.92 si on a fixé un seuil $\alpha = 5\%$.
- ▶ Comme la statistique calculée est nettement en dessous de cette valeur, on accepte l'hypothèse \mathcal{H}_0 .

Classes	n_i	$n_{t,i}$	$\frac{(n_i - n_{t,i})^2}{n_{t,i}}$
$[0, 0.1[$	100	100	0.00
$[0.1, 0.2[$	96	100	0.16
$[0.2, 0.3[$	98	100	0.04
$[0.3, 0.4[$	85	100	2.25
$[0.5, 0.6[$	105	100	0.25
$[0.6, 0.7[$	93	100	0.49
$[0.7, 0.8[$	97	100	0.09
$[0.8, 0.9[$	125	100	6.25
$[0.9, 0.1]$	107	100	0.49
$[0, 0.1[$	94	100	0.36
Total	1000	1000	10.38

Test des séquences de Wald-Wolfowitz

- C'est un test non-paramétrique. Il s'agit de tester le caractère aléatoire d'une suite

$$\begin{cases} \mathcal{H}_0 : & \text{La suite est i.i.d (aléatoire)} \\ \mathcal{H}_1 : & \text{La suite est non i.i.d (non aléatoire)} \end{cases}$$

- Une séquence R , ou **run** est une succession de un ou plusieurs éléments identiques qui sont précédés et/ou suivi d'un symbole différent (alternance)
 - $+++ \rightarrow$ 1 séquence sans alternance
 - $-+++-$ \rightarrow 3 séquences et 2 alternances
 - $--++-+$ \rightarrow 4 séquences et 3 alternances

Test des séquences de Wald-Wolfowitz (suite)

- ▶ Soit R le nombre de séquences dans un échantillon de taille n . Théoriquement, la moyenne et la variance sont :

$$\mathbb{E}[R] = \frac{2n - 1}{3} \text{ et } \mathbb{V}[R] = \frac{16n - 29}{90}.$$

- ▶ La série est non-aléatoire si les séquences sont trop longues ou trop courtes ou si le nombre d'alternances est trop petit ou trop grand.
- ▶ On rejette l'hypothèse \mathcal{H}_0 si $Z = \left| \frac{R - \mathbb{E}[R]}{\sqrt{\mathbb{V}[R]}} \right| > z_{1-\frac{\alpha}{2}}$ où $z_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi $\mathcal{N}(0, 1)$.

Test des séquences de Wald-Wolfowitz (suite)

Exemple

On considère la suite de nombres pseudo-aléatoire suivante :
0.08, 0.93, 0.15, 0.96, 0.26, 0.84, 0.28, 0.79, 0.36, 0.57.

- ▶ On marque chaque nombre avec un $+$ s'il est suivi d'un plus grand nombre et avec un $-$ s'il est suivi d'un plus petit nombre :

$+ - + - + - + - +$

- ▶ $R = 9$ séquences, $\mathbb{E}[R] = 6.33$, $\mathbb{V}[R] = 21.45$ et $Z = 2.21$
- ▶ $|Z| > 1.96$, alors on rejette l'hypothèse \mathcal{H}_0 .

Générateurs de R

- ▶ On obtient des nombres uniformes sur un intervalle quelconque avec la fonction `runif` dans R.
- ▶ La fonction `set.seed` permet de spécifier la valeur de l'amorce du générateur aléatoire, ce qui est utile si on veut répéter une simulation absolument à l'identique.
- ▶ Par défaut, R utilise le générateur **Mersenne–Twister**, réputé pour sa qualité.