

REVUE DE STATISTIQUE APPLIQUÉE

J. J. CLAUSTRIAUX

Génération et contrôle de validité de nombres pseudo-aléatoires sur ordinateur à mots de 16 bits

Revue de statistique appliquée, tome 24, n° 2 (1976), p. 75-88

http://www.numdam.org/item?id=RSA_1976__24_2_75_0

© Société française de statistique, 1976, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

*Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques*
<http://www.numdam.org/>

GÉNÉRATION ET CONTROLE DE VALIDITÉ DE NOMBRES PSEUDO-ALÉATOIRES SUR ORDINATEUR A MOTS DE 16 BITS (1)

J. J. CLAUSTRIAUX

Chaire de Statistique et d'Informatique

Faculté des Sciences Agronomiques de l'Etat à Gembloux

PLAN

	Pages
1. Introduction et résumé.	75
2. Génération de nombres pseudo-aléatoires.	76
2.1. Nombres aléatoires uniformes.	76
2.2. Nombres aléatoires normaux.	78
3. Contrôle de validité.	79
3.1. Tests des générateurs de nombres uniformes.	79
3.2. Tests de normalité à une dimension.	81
3.3. Tests de normalité à trois dimensions.	82
4. Conclusions.	86
Remerciements.	86
Bibliographie.	87

1 – INTRODUCTION ET RESUME.

Les nombres pseudo-aléatoires sont d'une utilisation fréquente dans les domaines scientifiques les plus divers, en vue par exemple de planifier des expériences (FISHER et YATES, 1963) ou de simuler des processus aléatoires pour la détermination des paramètres statistiques (MEYER, 1956; KLEIJNEN, 1973).

En agronomie, notamment, de nombreuses variables étudiées ont une fonction de densité de probabilité répondant à la loi normale ou tout au moins asymptotiquement normale (DAGNELIE, 1970). Aussi, est-il parfois nécessaire d'engendrer des réalisations indépendantes de variables aléatoires

(1) Article remis en Août 1975, révisé en Janvier 1976

de lois normales dans le but de simuler des rendements, des hauteurs d'arbres, des poids de feuilles, etc.

Ce problème sera envisagé pour des ordinateurs dont l'unité en mémoire centrale est un mot de 16 bits et sera étudié à l'aide d'un ordinateur I.B.M. 1130 (CARLETTI et *al.*, 1973). A ce propos, il faut remarquer que deux tiers des miniordinateurs du parc mondial possèdent une unité de 16 bits (CHABANAS, 1974 ; LEGRAND, 1974).

La génération de nombres pseudo-aléatoires uniformes et normaux sera examinée au paragraphe 2. On s'attachera notamment au choix de la constante multiplicative relative à l'algorithme générant les nombres pseudo-aléatoires uniformes ; on présentera un algorithme de grande capacité et on introduira la notion de période observée.

La troisième partie se rapportera au contrôle de validité des nombres pseudo-aléatoires uniformes et normaux. Entre autres, des tests de normalité à trois dimensions seront proposés afin de vérifier la normalité des distributions générées et de montrer la présence de corrélations entre les nombres pseudo-aléatoires uniformes. Ces tests seront ensuite appliqués à des modèles de génération de nombres pseudo-aléatoires normaux.

Enfin, des conclusions en vue de recherches ultérieures seront exposées au paragraphe 4.

2 – GENERATION DE NOMBRES PSEUDO-ALEATOIRES.

2.1. Nombres aléatoires uniformes.

Pour certaines études, il est nécessaire de disposer d'une grande "quantité" de nombres aléatoires. De plus, pour un travail sur ordinateur, on souhaite un générateur rapide et de volume réduit. C'est pourquoi, on a recours à des processus arithmétiques.

La méthode π (3, 14 159...) est relative à la puissance par exemple cinquième, de l'addition à π d'un nombre uniforme (0,1). Le nombre aléatoire résultant est formé des décimales du résultat obtenu.

La méthode du centre des carrés (von NEUMANN, 1951) est basée sur l'élévation au carré d'un nombre de $2n$ chiffres et la prise en considération des $2n$ chiffres centraux.

Vu la période de retour d'un nombre imprévisible et la présence de nombres absorbants (NEWMAN et ODELL, 1971), ces méthodes sont abandonnées au profit des méthodes à récurrence linéaire ou méthodes basées sur les résidus de puissances ou méthodes congruentielles (COVEYOU, 1960 ; LEHMER, 1951).

Ces dernières ont la qualité de reproductibilité des nombres (SMITH, 1968). Les nombres qui sont ainsi tirés à partir d'algorithmes ou processus déterministes sont appelés nombres "pseudo-aléatoires". Dans la suite de l'étude, nous qualifierons les nombres uniquement d'aléatoires.

1. L'ensemble des méthodes à récurrence linéaire peut se résumer par la formule :

$$x_{i+1} = (A \cdot x_i + M) \text{ (modulo B)}, \quad (1)$$

où x_i et x_{i+1} sont deux réalisations successives de l'algorithme, A et B des constantes entières et M une réalisation ou une constante ; x_{i+1} est le reste obtenu après multiplication de x_i par A, addition de M et division par B. Le terme M étant présent et le facteur A égalé à un, l'expression (1) est dite additive. L'absence ou non de M définit une suite multiplicative ou mixte.

Diverses publications, notamment celles de DAVIES (1971), de DELTOUR (1967), d'I.B.M. (1959), de SAUTEREAU (1972), nous ont conduit à choisir un algorithme multiplicatif. L'expression (1) s'écrit donc :

$$x_{i+1} = A \cdot x_i \text{ (modulo B)}. \quad (2)$$

2. Les valeurs des constantes A et B doivent conduire à l'uniformité de la distribution et aux caractères aléatoire et simple des nombres. La constante B est la plus grande valeur entière que peut contenir l'ordinateur ; la constante A est égale à $(8t \pm 3)^p$ pour une période théorique de B/4 ou à $(8t \pm 1)^p$ pour une période théorique de B/6, avec l'exposant p impair (HAMMING, 1962).

Le travail de DELTOUR (1967) a montré que les qualités citées plus haut étaient obtenues pour une valeur de A égale à la plus grande puissance impaire de 5 d'une unité de mémoire.

Pour une unité de 16 bits, la valeur de B, en enlevant un bit pour le signe, est théoriquement égale à 32.768 et la période théorique maximale est de 8.192. La constante A est donc du type $(8t \pm 3)^p$ ou 3.125 en prenant la plus grande puissance impaire de 5. Le constructeur prend la valeur 899 qui correspond à $(8.112 + 3)^1$ (IBM, 1969).

Deux sous-routines de nombres aléatoires uniformes, l'une avec la valeur 3.125 et l'autre avec le nombre 899, sont écrites en langage FORTRAN (RAND 1 et RAND 2).

3. Elles se distinguent de celle du constructeur par un test sur la longueur de la période théorique. En effet, pour la génération des 8.192 premiers nombres, il est nécessaire de choisir une première valeur impaire donc le premier x_i , par exemple à l'aide d'une table de nombres aléatoires. Un indice teste le nombre de valeurs obtenues ; cet indicateur varie de 1 à 8.192. Lorsque, le compteur atteint le nombre 8.192, l'ensemble de la période a été envisagé. Si, on souhaite obtenir plus de 8.192 nombres aléatoires, les générations suivantes débutent après avoir égalé la valeur de x_i à la valeur du nombre utilisé au début de la séquence précédente augmentée d'un nombre ou pas. Afin d'obtenir le plus grand nombre de valeurs différentes, le pas vaut deux. Au cours d'une séquence, si un des 8.192 nombres est égal à la valeur de départ de la séquence plus deux unités, le générateur débute automatiquement la génération de la série suivante. Dès lors, le 134.217.729^e nombre n'est pas nécessairement la valeur choisie au départ.

4. Puisque d'une part, les valeurs sont prélevées dans un ensemble nul ou positif variant de 0 à 32.767 et d'autre part l'algorithme a la qualité de reproductibilité des nombres, il est certain que des corrélations existent entre les séries dès qu'une valeur est commune aux séries. Afin de connaître

cette dépendance entre séries, un programme écrit en langage FORTRAN recherche les valeurs identiques à des séries successives.

Ainsi, à partir d'une valeur (x_1) fixée à priori ou déterminée dans une table de nombres aléatoires, on génère 8.192 nombres. Au fur et à mesure de la génération, on recherche dans la série la présence d'un nombre (x_k) égal au premier (x_1) plus un pas de deux unités :

$$x_k = x_1 + 2. \quad (3)$$

Lorsque les 8 192 nombres sont engendrés, une nouvelle série est produite sur base de la valeur x_k . Une valeur x_1 est alors recherchée dans la nouvelle séquence :

$$x_1 = x_k + 2. \quad (4)$$

1.719 séries successives de 8.192 valeurs sont produites soit 14.082.048 nombres uniformes. Jamais, on ne trouve dans une séquence le nombre de départ de la séquence suivante mais bien celui de la seconde voisine.

D'ailleurs, si le pas est dans (3) et (4) de quatre unités, la valeur recherchée peut survenir dans la série après la génération de 79 valeurs comme après avoir engendré 8.005 nombres.

En conséquence, en adoptant les formules (2) et (3), si la période théorique est de 8.192, il existe une période de plus grande longueur, que nous appelons période observée et qui s'élève au moins à 16.384.

Tenant compte de ces résultats, un second modèle de génération de nombres aléatoires uniformes est proposé.

Au cours d'une première étape, un fichier est créé sur disque magnétique et y sont enregistrés 16.384 nombres uniformes produits par RAND 1 ou RAND 2. Le choix de l'un ou l'autre de ces deux sous-programmes dépend des résultats exposés au paragraphe 3.1.

Ensuite, un autre sous-programme est réalisé (RAND 3) et prélève des nombres uniformes dans le fichier ou table selon l'équation (2).

Enfin, lorsque le nombre de valeurs prélevées s'élève à 16.384, la table est transformée. Le nombre ayant l'indice 1 est placé en 16.384^e position, le deuxième nombre devient le premier, le 16.384^e nombre devient le 16.383^e et ainsi de suite. Après la permutation du fichier, la nouvelle série de prélèvements (RAND 3) débute avec le nombre de départ de la série antérieure. Les valeurs sont permutées après chaque génération de 16.384 nombres. Il y a donc 268.435.456 prélèvements possibles.

2.2. Nombres aléatoires normaux.

La génération de variables aléatoires de fonction de densité de probabilité distribuée normalement se réalise par l'intermédiaire des variables aléatoires uniformes et résulte de l'application des propriétés limites ou exactes des distributions théoriques.

1. En vertu du théorème de Lindeberg-Levy, la somme de n variables aléatoires uniformes indépendantes est une variable asymptotiquement normale.

Si la sommation porte sur 12 variables (X_i) uniformes (0,1), la relation engendrant une variable normale quelconque (N_j) s'écrit :

$$N_j = (m - 6 s) + s \sum_{i=1}^{12} X_i ; \quad (5)$$

cette égalité fait intervenir la moyenne (m) et l'écart-type (s) de la distribution normale souhaitée.

NEWMAN et ODELL (1971) estime le nombre 12 trop petit pour des approximations raisonnables ; les mêmes auteurs signalent une représentation non conforme des valeurs extrêmes de la distribution normale pour une sommation de 50 variables uniformes. Il faut toutefois remarquer que l'augmentation du nombre de valeurs uniformes diminue la périodicité du générateur des variables normales.

2. En se basant sur les propriétés exactes des distributions théoriques et sur les propriétés trigonométriques, on obtient les équations de BOX et MULLER (1958) :

$$U_1 = \sqrt{-2 \log_e X_1} \cdot \cos (2 \pi X_2) , \quad (6)$$

$$U_2 = \sqrt{-2 \log_e X_1} \cdot \sin (2 \pi X_2) . \quad (7)$$

Si X_1 et X_2 sont des variables aléatoires indépendantes toutes deux distribuées uniformément dans l'intervalle (0,1), U_1 et U_2 sont des variables aléatoires normales réduites indépendantes.

3. Des sous-routines faisant appel aux sous-programmes cités au point 2.1 sont écrites en vue de générer des nombres aléatoires normaux ; GAUS 1, GAUS 2, GAUS 3 se basent respectivement sur les relations (5), (6), (7) et font appel à RAND 1 ou RAND 2. Les résultats des tests des paragraphes 3.1 et 3.2 permettront d'effectuer un choix entre RAND 1 ou RAND 2 et entre GAUS 1, GAUS 2 ou GAUS 3.

Pour un ordinateur I.B.M. 1130, le temps nécessaire à la génération d'un nombre aléatoire normal s'élève à 0,13 seconde avec le sous-programme GAUS 1 et 0,05 seconde avec les sous-routines GAUS 2 et GAUS 3.

Enfin, il faut remarquer que la précision obtenue avec GAUS 2 et GAUS 3 dépend aussi des fonctions logarithmique, trigonométriques et racine carrée.

3 – CONTROLE DE VALIDITE.

3.1. Tests des générateurs de nombres uniformes.

L'uniformité de la distribution, les caractères aléatoire et simple des nombres peuvent être testés de diverses façons (CHASSE et *al.*, 1974 ; DEVILLERS et *al.*, 1973 ; IBM, 1959 ; LINDER, 1970 ; NEWMAN et ODELL, 1971 ; SAUTEREAU, 1972). Nous avons repris certains tests d'ajustement moins classiques cités dans les travaux de DELTOUR (1967) et LINDER (1970).

3.1.1. Méthodes.

1. L'homogénéité de la distribution (test 1) se base sur un échantillon de 20.000 nombres et un intervalle de classe d'un vingtième.

2. L'homogénéité de la distribution des chiffres dans les nombres est envisagée comme suit : tout d'abord, on détermine la fréquence de chaque chiffre en chaque position du nombre considéré dans l'intervalle 0-9.999 et indépendamment de sa position (test 2) ; ensuite, le test de Poker appliqué à des nombres de 4 chiffres vérifie l'indépendance des chiffres composant un même nombre (test 3).

On compte la fréquence des nombres formés soit de 4 chiffres différents, soit de 2 chiffres identiques parmi les 4 (une paire), soit de 3 chiffres identiques (un brelan), soit de deux paires, soit de 4 chiffres semblables. Ces tests portent sur 20.000 nombres.

3. L'absence de corrélation entre les nombres (test 4) peut se réaliser en déterminant le nombre de valeurs successives qui se situent au-dessus ou au-dessous d'une certaine constante fixée a priori (DELTOUR, 1967). Un échantillon de 10.000 nombres est utilisé pour ce test.

Enfin, les coefficients de corrélation d'ordre 1 à 20 sont calculés sur des échantillons de 10.000 nombres.

3.1.2. Résultats.

Le tableau 1 reprend pour chaque test le nombre d'ajustements effectués par constante (RAND 1 : 3.125 ; RAND 2 : 899), pour chaque constante le nombre de rejets de l'hypothèse nulle au niveau de probabilité 0,05 et le nombre de valeurs χ^2 de Pearson supérieures à celles de l'autre constante.

Tableau 1
Tests d'ajustements

numéro du test	nombre d'ajustements par constante	nombre de rejets		nombre de valeurs sup.	
		3.125	899	3.125	899
1	1	0	0	0	1
2	5	0	0	2	3
3	1	0	0	0	1
4	10	1	1	5	5

En se basant sur la signification des différents tests, il n'est pas possible d'orienter le choix de la constante vu l'acceptation de l'hypothèse d'ajustement dans tous les cas sauf un aussi bien pour la valeur 899 que 3.125. Les valeurs des coefficients de corrélation sont aussi tous inférieurs au seuil de 0.02, valeur critique correspondant au niveau de signification de 0.05.

Néanmoins, constatant que la valeur de la variable χ^2 relative à la constante 3.125 est inférieure à celle dépendant de la constante 899 dans 10 tests sur 17, la valeur 3.125 (RAND 1) est utilisée comme constante A pour la suite du travail.

3.2. Tests de normalité à une dimension.

3.2.1. Méthodes.

Les propos de ce paragraphe ne sont ni d'inventer un nouveau test de normalité, ni de comparer les tests existants pour en choisir le "meilleur", mais simplement de soumettre pour chaque méthode (GAUS 1, GAUS 2, GAUS 3) 30 échantillons de 1.000 nombres aléatoires dits normaux à quelques tests en vue d'étudier la normalité des distributions.

On trouvera dans la littérature de nombreuses méthodes pour tester la normalité (D'AGOSTINO et TIETJEN, 1971 ; D'AGOSTINO et PEARSON, 1973 ; DURBIN, 1961 ; LINDER et CZEGLÉDY, 1973 ; SHAPIRO et FRANZIA, 1972 ; SHAPIRO et al., 1968).

Vu l'effectif élevé de nos échantillons, les objections formulées envers le test χ^2 d'ajustement peuvent être repoussées. Aussi, le test χ^2 de Pearson (DAGNELIE, 1970) (test 1) est appliqué à une distribution complètement définie comportant 24 classes de fréquences attendues supérieures à 5.

Le second test (test 2) est basé sur la comparaison de la fonction cumulative de fréquence de l'échantillon à la fonction de répartition de la population ou test de Kolmogorov et Smirnov (DAGNELIE, 1968).

Les tests 3 et 4 se rapportent au coefficient g_1 de Fisher et au coefficient b_2 de Pearson (PEARSON et HARTLEY, 1965).

Comme prolongement au test de Shapiro (SHAPIRO et al., 1968) pour des effectifs supérieurs à 50, le test d'Agostino (D'AGOSTINO, 1971) est utilisé comme cinquième test. Il nécessite la mise par ordre croissant des n observations et tient compte du rang de chaque valeur et de la dispersion des observations.

Les tests 6 et 7 sont aussi associés à l'estimation de la dispersion. Le test de Geary (GEARY, 1947) rapporte l'écart moyen absolu à l'écart-type et le test de David (DAVID, et al., 1954) calcule le rapport entre l'amplitude et l'écart-type estimé.

Un programme (NORMA) écrit en langage FORTRAN a permis la réalisation des calculs.

3.2.2. Résultats.

Le tableau 2 rapporte pour chaque sous-routine et pour chaque test, le nombre de rejets significatifs de l'hypothèse de normalité sur les 30 cas étudiés.

Pour le niveau de signification fixé à 0.10, il apparaît pour la sous-routine basée sur la méthode asymptotique (GAUS 1), un nombre de rejets inférieur aux valeurs relatives aux autres méthodes. En particulier, la sous-routine GAUS 3 présente pour trois tests, un pourcentage de rejet supérieur ou égal au niveau fixé.

Tableau 2
Tests de normalité

numéro du test	nombre de rejets		
	GAUS 1	GAUS 2	GAUS 2
1	2	0	0
2	0	1	0
3	0	1	2
4	0	1	4
5	0	1	3
6	1	2	2
7	0	2	4

3.3. Tests de normalité à trois dimensions.

Les tests du caractère normal des distributions effectués jusqu'à présent, portent au maximum sur des échantillons de 12.000 nombres uniformes.

Sur base des points 3 et 4 du paragraphe 2.1 et des résultats du paragraphe 3.2, il est envisagé de poursuivre l'étude sur un plus grand nombre de valeurs aléatoires en vue d'une part de comparer les sous-routines GAUS 1, GAUS 2, GAUS 4 et GAUS 5 et d'autre part de détecter ou non une concentration anormale, non directement évidente, entre les nombres aléatoires uniformes utilisés pour générer les variables normales.

Les sous-programmes GAUS 4 et GAUS 5 emploient la sous-routine RAND 3 (paragraphe 2.1, point 4) et génèrent des nombres aléatoires normaux sur base des équations 5 et 6.

3.3.1. Méthode.

COLDWELL (1974) envisage l'étude de l'uniformité d'une distribution en générant, dans un espace à trois dimensions des nombres aléatoires uniformes.

Un principe identique est adopté. Sur chaque axe, un nombre normal, en particulier de distribution normale réduite, est généré.

On détermine ainsi, les coordonnées d'un point dans l'espace comme le montre la figure 1.

La distance du point A a l'origine se calcule par la relation :

$$\delta = \sqrt{x^2 + y^2 + z^2} \quad (8)$$

où x, y, et z sont les grandeurs OD, OB, et OE.

Dès lors, si OD, OB et OE sont trois valeurs des variables aléatoires X, Y, Z, normales réduites indépendantes, la quantité :

$$\delta^2 = X^2 + Y^2 + Z^2 \quad (9)$$

est une variable aléatoire χ^2 de Pearson à trois degrés de liberté.

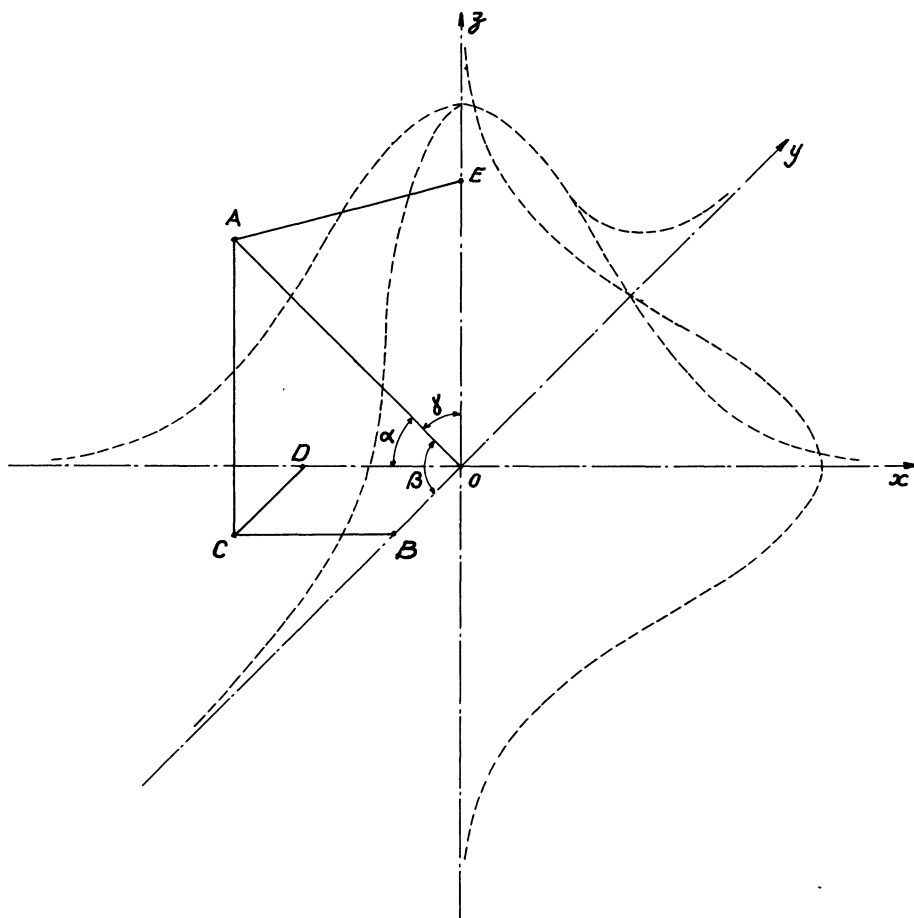


Figure 1 – Distributions normales et coordonnées.

D'autre part, soit le triangle rectangle ABO ; en vertu des propriétés des triangles rectangles, on peut écrire :

$$\operatorname{tg} \beta = AB / OB. \quad (10)$$

De plus, à partir du triangle ACB, on obtient une autre égalité :

$$AB^2 = AC^2 + BC^2. \quad (11)$$

Remarquons que les grandeurs AC et BC sont respectivement égales à OE et OD. L'équation 10 devient alors :

$$\operatorname{tg} \beta = \frac{\sqrt{AC^2 + BC^2}}{OB} = \frac{\sqrt{Z^2 + X^2}}{Y} \quad (12)$$

c'est-à-dire le rapport de la racine carrée d'une variable χ^2 à deux degrés de liberté et d'une variable normale réduite, cette dernière indépendante de la première variable.

Partant de l'équation 12, on peut écrire :

$$\sqrt{2} / \operatorname{tg} \beta = Y \cdot \sqrt{2} / \sqrt{\chi_2^2}. \quad (13)$$

Cette quantité est une distribution t de Student à deux degrés de liberté.

Un raisonnement identique peut être mené pour les angles α et γ .

En conséquence, après la génération d'un certain nombre de points dans un espace à trois dimensions, si les variables X, Y, Z sont des variables normales réduites indépendantes, les ajustements des carrés des distances des points à l'origine et des inverses des tangentes des angles multipliés par $\sqrt{2}$ doivent conduire à l'acceptation des hypothèses de distributions χ^2 et t.

Les sous-routines (GAUS 1, GAUS 2, GAUS 4, GAUS 5) produisent les coordonnées des points dans l'espace. L'ajustement des distributions proposées se réalise par le test χ^2 de Pearson. Toutes les classes définies ont des valeurs attendues supérieures ou égales à 10.

3.3.2. Résultats.

Le tableau 3 reprend, pour chaque modèle proposé, outre le nombre de valeurs normales générées, les nombres de degrés de liberté (d.l.) et les valeurs χ^2 observées pour les quatre distributions étudiées.

Selon que l'hypothèse est rejetée au niveau 0,05, 0,01 ou 0,001, une, deux ou trois astérisques indiquent que l'écart observé est significatif, hautement significatif ou très hautement significatif.

Les nombres de valeurs uniformes qui correspondent aux nombres de valeurs normales générées et les estimations des temps nécessaires à la génération d'une valeur aléatoire normale sont d'autres éléments de comparaison.

C'est pourquoi, ces données sont citées dans le tableau 4.

Sur base des résultats réunis dans les tableaux 3 et 4, il apparaît pour le modèle GAUS 5, un rejet d'une des hypothèses de départ dès la génération de 1.362 nombres uniformes.

Remarquons que pour la classe 681, les modèles GAUS 1 et GAUS 2 sont semblables respectivement à GAUS 4 et GAUS 5. Avec le sous-programme GAUS 1, une conclusion identique est formulée pour la génération de 154.836 nombres uniformes. Par contre, c'est seulement aux environs de 56 utilisations de la période théorique relative aux nombres uniformes qu'un rejet significatif apparaît avec GAUS 4.

Ces résultats sont à mettre en rapport avec le temps moyen nécessaire à la génération d'une valeur aléatoire normale (GAUS 4 : 5 secondes). Dans nos conditions de travail, le rapport des temps moyens entre GAUS 4 et GAUS 1 vaut approximativement 38. Certainement, ce nombre est élevé ; mais, il faut savoir que le temps d'accès moyen à un fichier se trouvant sur disque magnétique s'élève à 0,775 seconde alors que le temps d'accès moyen en mémoire centrale est de $3,6 \cdot 10^{-6}$ seconde.

En supposant pour un miniordinateur, un temps d'accès moyen sur disque magnétique de l'ordre de 0,050 à 0,060 seconde, le rapport des temps de génération sera environ 10 fois moins élevé et le temps pour la génération d'une valeur aléatoire normale avec GAUS 4 sera vraisemblablement de l'ordre de 0,4 seconde.

Tableau 3
 Comparaison des 4 modèles (valeurs χ^2).

modèles			nombres de valeurs normales générales			
			681	12.903	18.798	38.403
GAUS 1	distance	d.1.	19	35	35	35
		χ^2	10.7	57.4*	87.1***	126.7***
	angles	d.1.	18	20	20	20
		χ^2	9.9	15.5	23.5	27.4
		χ^2	18.6	16.1	17.1	27.2
χ^2	16.4	26.7	31.1	46.3***		
GAUS 4	distance	d.1.	19	35	35	35
		χ^2	21.4	37.0	36.2	41.2
	angles	d.1.	18	20	20	20
		χ^2	16.8	18.9	24.1	19.6
		χ^2	24.1	13.0	19.7	37.3*
		χ^2	17.6	12.7	9.5	26.2
GAUS 2	distance	d.1.	19	35	35	35
		χ^2	25.3	36.0	41.8	75.0***
	angles	d.1.	18	20	20	20
		χ^2	27.1	16.1	17.4	21.3
		χ^2	19.0	21.4	33.6*	63.9***
		χ^2	14.9	23.5	32.6*	33.5*
GAUS 5	distance	d.1.	19	35	35	35
		χ^2	29.5	31.1	34.8	46.4
	angles	d.1.	18	20	20	20
		χ^2	26.9	17.0	28.1	34.0*
		χ^2	33.3*	13.4	23.0	23.3
χ^2	23.6	25.9	34.1*	28.4		

Tableau 4
 Comparaison des 4 modèles (nombres de valeurs uniformes pour générer les nombres normaux et temps moyens de la génération d'une valeur aléatoire normale).

modèles	nombres de valeurs uniformes par classe de valeurs normales				temps moyens (seconde)
	681	12.903	18.798	38.403	
GAUS 1	8.172	154.836	225.576	460.836	0.13
GAUS 4	8.172	154.836	225.576	460.836	5.00
GAUS 2	1.362	25.806	37.596	76.806	0.05
GAUS 5	1.362	25.806	37.596	76.806	0.45

Enfin, en comparant GAUS 1 à GAUS 4 et GAUS 2 à GAUS 5, il semble qu'une méthode de génération des nombres aléatoires normaux basée sur les considérations du paragraphe 2.1. (RAND 3) par rapport au modèle sans permutation (RAND 1) permet de générer un plus grand nombre de valeurs normales.

4 – CONCLUSIONS (1).

Au cours de cette étude, abordée dans le cas d'un ordinateur possédant des mots de 16 bits (la majorité des miniordinateurs), on a tout d'abord montré que la période théorique de retour d'un nombre aléatoire uniforme (0,1) généré par une méthode congruentielle est supérieure à 8.192 si, après la génération de 8.192 nombres, on prend comme nombre de départ d'une nouvelle série de 8.192 valeurs celui de la série antérieure augmenté de deux unités.

Sur base de ces résultats, un algorithme multiplicatif fut présenté et quelques tests ont été effectués, d'une part, pour contrôler le caractère uniforme de la distribution et les qualités aléatoire et simple des nombres et, d'autre part, comparer deux constantes multiplicatives (899 et 3.125).

Après avoir choisi la valeur 3.125 (plus grande puissance impaire de 5 que l'ordinateur peut contenir), nous avons présenté et comparé par des tests de normalité à une dimension deux méthodes de génération de nombres aléatoires normaux (méthode approchée et méthode exacte).

Ensuite, des tests de normalité à trois dimensions ont été proposés en vue d'affiner le contrôle d'ajustement à une distribution normale des valeurs aléatoires dites normales et de détecter des corrélations éventuelles entre des nombres aléatoires uniformes.

Ces tests ont été appliqués à deux catégories de modèles de génération de nombres aléatoires normaux. Ils ont montré la supériorité statistique d'une catégorie de modèles sur l'autre. Mais des estimations de temps ont fait apparaître non seulement qu'un générateur était plus rapide qu'un autre mais aussi qu'un statistiquement moins bon générateur était supérieur quant à sa rentabilité sur ordinateur.

Enfin, il fut constaté que la simulation de variables, dont la fonction de densité de probabilité répond à la loi normale, pouvait se réaliser par la sommation de 12 variables uniformes, les caractères de normalité et d'indépendance des réalisations étant, dans certaines limites considérés comme assurés.

REMERCIEMENTS.

Nous tenons à exprimer notre gratitude à Messieurs P. DAGNELIE et J. DELTOUR, Professeurs à la Faculté des Sciences agronomiques de Gembloux, pour leurs suggestions et critiques.

(1) Les "listings" de tous les programmes employés sont disponibles à la Chaire de Statistique de la Faculté des Sciences agronomiques de l'Etat à Gembloux (BELGIQUE).

BIBLIOGRAPHIE.

- BOX G.E.P., MULLER M.E. (1958) – A note on the generation of random normal deviates. *Ann. Math. Statist.*, 29, 610-611.
- CARLETTI G., CLAUSTRIAUX J.J., DAGNELIE P., DEBOUCHE C., IN K., OGER R., ROUSSEAU G. (1973) – Organisation d'une bibliothèque de programmes statistiques pour ordinateur. *Revue Belge de Stat., d'Infor. et de Rech. Op.*, 12,4, 2-16.
- CHABANAS J.M. (1974) – Miniordinateurs, un prince et ses barons. *01 Informatique*, 80, 44-49.
- CHASSE J.C., DEBOUZIE D. (1974) – Utilisation des tests de KIVELIOVITHC et VIALAR dans l'étude de quelques générateurs de nombres pseudo-aléatoires. *Revue Stat. Appl.*, 22, 3, 83-90.
- COLDWELL R.L. (1974) – Correlational defects in the standard IBM 360 random number generator and the classical idea GAS correlation function. *Journal of Comp. Physics*, 14, 2, 223-226.
- COVEYOU R.R. (1960) – Serial correlation in the generation of pseudo-random numbers. *Journal Assoc. Comput. Mach.*, 7, 72-74.
- DAGNELIE P. (1968) – A propos de l'emploi du test de KOLMOGOROV-SMIRNOV comme test de normalité. *Biom. Praxim.*, 9, 3-13.
- DAGNELIE P. (1970) – Théorie et méthodes statistiques (2 vol.) Presses Agron., Gembloux, 378 + 451.
- D'AGOSTINO R.B. (1971) – A omnibus test of normality for moderate and large size samples. *Biometrika*, 58, 341-348.
- D'AGOSTINO R.B., PEARSON E.S. (1973) – Test for departure from normality. Empirical results for the distribution of b_2 and $\sqrt{b_1}$. *Biometrika*, 60, 613-622.
- D'AGOSTINO R.B., TIETJEN G.L. (1971) – Simulation probability points of b_2 for small samples. *Biometrika*, 58, 669-672.
- DAVID H.A., HARTLEY H.O., PEARSON E.S. (1954) – The distribution of the ratio, in a single normal sample, of range to standard deviation. *Biometrika*, 41, 482-493.
- DAVIES R.G. (1971) – Computer programming in quantitative biology. Academic Press, London, 492.
- DELTOUR J. (1967) – Etude d'une distribution de nombres pseudo-aléatoires. *Bull. Rech. Agron. Gembloux*, 3, 450-460.
- DEVILLERS R., DUMONT J.J., LATOUCHE G. (1973) – Tests de generateurs pseudo-aléatoires. *Bull. de la Classe des Sciences, Acad. roy. de Belgique*, 59, 703-724.
- DURBIN J. (1961) – Some methods of constructing exact tests. *Biometrika*, 48, 41-55.
- FISHER R.A., YATES F. (1963) – Statistical tables for use in biological, agricultural, and medical research. Oliver and Boyd, Edinburgh, 138.
- GEARY R.C. (1947) – Testing for normality. *Biometrika*, 34, 209-242.

- HAMMING R.W. (1962) – Numerical methods for scientists and engineers. Mc. Graw-Hill, New-York, 411.
- I.B.M. (1959) – Random number generation and testing. I.B.M., Data processing techniques, 1-12.
- I.B.M. (1969) – 1130 scientific subroutine package. I.B.M. corporation, 191.
- KLEIJNEN J.P.C. (1973) – Statistical techniques in simulation. Marcel Dekker, New-York.
- LEGRAND M. (1974) – Petits, minis, micros... Une grande famille mais des personnalités très marquées. *01 Informatique*, 77, 27-33.
- LEHMER D.H. (1951) – Mathematical methods in large-scale computing units. *Ann. Comp. Lab., Harvard*, 26, 141-146.
- LINDER A. (1970) – Testing a table of random numbers. In : Probability and statistics, University of North Carolina Press, 469-478.
- LINDER A., CZEGLÉDY P. (1973) – Normality test by fractile graphical analysis. *The Indian j. of stat.*, B, 35, 1-14.
- MEYER H.A. (1956) – Symposium on Monte Carlo Methods. John Wiley and sons, New-York.
- NEWMAN T.G., ODELL P.L. (1971) – The generation of random variates. *Griffin's Statistical Monographs and Courses*, 1-88.
- PEARSON E.S., HARTLEY H.O. (1965) – Tables of percentage points of the distributions of $\sqrt{b_1}$ and b_2 . *Biometrika*, 52, 282-285.
- SAUTEREAU C. (1972) – Génération de variables aléatoires et techniques d'échantillonnage des lois de probabilité. *Document C.I.R.O.*, 7, 1-46.
- SHAPIRO S.S., FRANCA R.S. (1972) – An approximate analysis of variance test for normality. *J. Amer. Stat. Assoc.*, 67, 215-216.
- SHAPIRO S.S., WILK M.B., CHEN H.J. (1968) – A comparative study of various tests for normality. *J. Amer. Stat. Assoc.*, 63, 1.343-1.372.
- SMITH J. (1968) – Computer simulation models. Griffin, 1-112.
- von NEUMANN J. (1951) – Various techniques used in connection with random digits. *U.S. Nat. Bur. Stand. Appl. Math. Ser.*, 12, 36-38.