Test universel pour l'évaluation du caractère aléatoire

Pour que le caractère aléatoire d'une série de nombres puisse être établi valablement à l'aide du "Test Universel", la condition, nécessaire et suffisante, sera l'obtention d'un résultat inférieur à une valeur déterminée en unité σ (sigma) suivant une échelle normalisée. Il est normal d'affirmer qu'une cryptanalyse linéaire ou différentielle sera d'autant plus difficile, voire même impossible, à réaliser, si les caractéristiques définissant le caractère aléatoire des nombres chiffrés, sont mathématiquement les plus proches de l'aléatoire parfait. D'où l'importance de pouvoir établir une évaluation chiffrée et mathématiquement correcte du caractère aléatoire d'une suite de nombres. Le test universel, qui répond à cette demande, peut s'appliquer à toutes séries de nombres aléatoires, de longueurs et de grandeurs quelconques, et permet de comparer les performances de certaines caractéristiques des algorithmes de chiffrement ou encore des générateurs pseudo-aléatoires de types différents. Il ne serait donc plus question de devoir procéder à tous les tests tels que: run, gap, spectral... (Voir Donald Knuth: Vol 2 The Art of Computer Programming)

La plus importante des caractéristiques d'un générateur de nombres pseudo-aléatoires est que les résultats de tous les tirages soient équidistribués. On suppose les résultats de T tirages, des nombres qui sont compris entre 1 et N. Pour des tranches appliquées aux résultats se situant entre $N1$ et $N2$ où ($N1$-$N2$) = N / k, le nombre de résultats dans chaque tranche doit être proche T / k indépendamment des valeurs $N1$ et $N2$. Dans le test universel, la tranche $N1$-$N2$ est réduite à un seul nombre (k = N) et dans ce cas il y a lieu de se référer aux probabilités données par la loi de Poisson pour ce nombre.

A titre d'exemple, soit à vérifier le caractère aléatoire des 500.000 chiffres décimaux après la virgule de PI, lesquels ont été classés en 100.000 nombres de cinq chiffres, chacun d'eux de valeurs de 0 à 99.999. En recherchant pour chacun des 100.000 nombres, combien de fois la présence de chaque nombre de 0 à 99.999 sera présent dans la série et en classant les résultats par le nombre de fois qu'ils y apparaissent, on obtient (i+1) ensembles dont les montants doivent être proches des montants théoriques donnés par la loi de Poisson (http://fr.wikipedia.org/wiki/Loi_de_Poisson). Dans ce cas, le paramètre lambda (λ) est égal à 1. En effet, le choix de la grandeur des nombres concernés de 0 à 99.999 est égal au nombre 100.000 de tirages ou d'essais effectués.

Les résultats théoriques de la loi de Poisson avec la caractéristique λ=1 sont donnés pour i de 0 à imax par la suite 36.787,96.., 36.787,96.., 18.393,98.., 6.131,32.., 1.532,83.., 306,56.., 51,09.., 7,21.., 0,91.., et 0,10.. (où 36.787,96 = 100.000/e (e=2,71828...)). Les résultats expérimentaux sont donnés par la liste des dix nombres entiers suivants: 36.824, 36.808, 18.393, 6.271, 1.493, 315, 56, 4, 2 et 0.

Pour chiffrer l'écart entre le nombre expérimental 36.824 et le nombre théorique 36.787,96.., on se réfère à l'écart quadratique théorique qui est égal à la racine carrée de 36.787,96.., soit 191,80... . La valeur absolue de la différence théorique/expérimentale est de 36,04..,(= 36.824-36.787,96..), ce qui correspond à 0,187.. fois l'écart quadratique théorique (0.187.. = 36,04.. /191,80..) Ce dernier résultat est à multiplier par 0,3678.. (= 36.787.96.. /100.000), c'est à dire la proportion qui intervient dans le calcul de l'écart expérimental σ. On obtient ainsi le résultat partiel de 0,0687 (= 0.3678.. * 0.187..) qui correspond à la part due aux 36.824 essais où un nombre déterminé n'a pas été retrouvé dans la liste des 100.000 premiers nombres de 5 chiffres après la virgule de PI. Le même calcul est à

effectuer pour les neuf autres montants de 1 à "i", ce qui donne finalement le résultat de σ = 0,461, l'écart quadratique pondéré en unité sigma.

En se référant aux 4.000.000 chiffres après la virgule de Pi, on trouve les 8 résultats suivants: σ = 0,461, 0,197, 0,200, 0,352, 0,362, 0,656, 0,514 et 0,430 correspondant à une moyenne de σ = 0,396.

La même méthode a été utilisée pour évaluer le caractère aléatoire du générateur pseudo-aléatoire congruentiel cité par WIKIPEDIA : X(n+1)=X(n)*16.807 MOD (2^31-1). Pour les besoins de l'utilisation du test universel, le résultat de chaque nombre aléatoire généré est multiplié par (100.000/2^31-1) dont on prend la partie entière en obtenant ainsi un nombre entier composé de cinq chiffres. Onze groupes de 100.000 tirages ont donné les résultats suivants: σ = 0,535, 0,482, 0,576, 0,892, 0,522, 0,416, 0,212, 0,463, 0,299, 0,981 et 0,439, donnant une moyenne de σ = 0,528. On voit tout de suite que le caractère aléatoire de ce dernier générateur est inférieur à celui de PI

Au cours des années 1960, la société IBM propose pour ses gros ordinateurs le générateur pseudo-aléatoire "RANDU" où X(n+1)=X(n)*65.539 MOD 2^31 (voir WIKIPEDIA). Onze groupes de 100.000 tirages ont donné les résultats suivants : σ = 60,424, 60,446, 59,925, 60.482, 60,372, 60,293, 60,317, 60,197, 59,858, 60,201 et 60,127, ce qui donne lieu à une moyenne de σ = 60,249 et un écart quadratique pour ces onze valeurs de σ = 0.193. Ces résultats se passent de commentaires, mais ne sont pas dénués d'enseignement utile. La moyenne 60,249 signifie que ce générateur est excessivement mauvais, par contre les fluctuations autour de la moyenne sont normales pour des tests portant sur 100.000 tirages. La cotation de cet algorithme donnée par WIKIPEDIA est tout à fait correcte : « biaisé et fortement déconseillé ».

En vue d'obtenir un résultat σ très petit, une simulation de l'algorithme de chiffrement SED a été effectuée dans deux corps finis de (2^19-1), les polynômes caractéristiques étant P0+P3+P7+P10+P12+P13+P14+P15 = P19 et P0+P2+P4+P5+P7+P8+P14+P15 = P19. Le test universel a donné le résultat de **σ = 0,137,** en chiffrant tous les nombres de 1 à 524287 comme clef et comme nombre clair. L'algorithme SED est exposé dans l'URL indiquée ci-après.

Des tests portant sur des ensembles de 131071 nombres composés de 17 bits (131.071=2^17-1) extraits des résultats de chiffrements effectués à l'aide de l'algorithme SED de 128 bits donnent les résultats suivants: σ = 0,301, 0,670, 0,805, 0,450, 0,220, 0,315 et 0,508 ce qui correspond à une moyenne de σ = 0,467.

Dans les tests qui précèdent, certains se rapportent à des listes de longueurs finies (2^19-1 et 2^17-1), d'autres à des listes de longueurs infinies (PI et le SED). Dans le premier cas, le résultat sigma est donné par un nombre exact, tandis que dans le second cas, le résultat est approximatif. Pour disposer d'un résultat valable, il a lieu de procéder à un grand nombres de tests pour lesquels on établira une moyenne des résultats sigma accompagnés de l'écart quadratique attaché à la valeur moyenne. On pourra ainsi établir, avec suffisamment de précision, les niveaux des caractères aléatoires relatifs aux algorithmes de chiffrement concurrents.

En conclusion de ce qui précède, vu la structure scientifiquement établie du test universel et les tests expérimentaux afférents à ce test, il est normal d'accepter la conjecture suivante: la condition nécessaire et suffisante pour que le caractère aléatoire d'une suite de nombres puisse être valablement établi par tous les tests existants ou futurs, serait d'être accepté par le test universel. (Pour plus de détails, voir : http://fr.wikipedia.org/wiki/Utilisateur:Emile_Musyck)

Emile Musyck