Répartition des fonctions booléennes et des suites aléatoires

Florian Caullery et François Rodier

Darkmatter LLC, Abu Dhabi, UAE, florian.caullery@darkmatter.ae
Institut de Mathématiques de Marseille, francois.rodier@univ-amu.fr

Outline

Fonctions booléennes

- Soit m un entier positif.
- Une **fonction booléenne** avec m variables est une application de l'espace $\mathbb{F}_2^m$ dans $\mathbb{F}_2$.
- Les fonctions booléennes sont utilisées en cryptographie pour chiffrer les messages:
  - par exemple comme **générateurs pseudo-aléatoires dans le chiffrement à flot**,
  - ou pour le **chiffrement par blocs,**
  - ou encore pour **hacher des messages.**
- On peut voir également une séquence binaire de longueur 2^n comme la table de vérité d'une fonction booléenne, pour **vérifier le caractère aléatoire** d'une telle suite.

Mesure de la résistance offerte par une fonction booléenne contre une cryptanalyse spécifique

Il existe plusieurs façons de mesurer cette résistance.
Parmi ceux-ci, il convient de mentionner

- la non linéarité, qui mesure la résistance d'une fonction booléenne contre des **attaques d'approximation** à faible degré.
- l' indicateur absolu et
- la somme des carrés
  - Les deux dernières méthodes sont habituellement regroupés dans le terme de critère global d'avalanche (GAC).
  - Elles mesurent la capacité d'une fonction booléenne d'assurer la propriété de **propagation d'un cryptosystème.**

La non-linéarité

- La non-linéarité est le **nombre minimal de bits** qui doivent changer dans la **table de vérité** d'une fonction booléenne pour atteindre la fonction affine la plus proche.
- On s'intéresse à **l'énumération des fonctions booléennes** en fonction de leur non-linéarité.
- Ceci a été accompli avec succès par Berlekamp et Welch pour jusqu'à **cinq** variables et par Maiorana pour **six** variables.
- Cependant, l'énumération exacte pour un plus grand nombre de variables semble être insoluble.

  Par conséquent, les **estimations asymptotiques de la distribution de la non-linéarité** deviennent pertinentes.

Non-linéarité

- Inégalités sur la non-linéarité

$$\begin{array}{ccccccc} 2^{m/2-1} & & \leq & & 2^{m-1} - NL(f) & \leq & 2^{m-1} \\ \uparrow & & \uparrow & & & \uparrow & \uparrow \\ \textbf{fonctions courbes} & & \text{Parseval} & & & \text{clear} & \textbf{fonctions affines} \end{array}$$

- Distribution de la non-linéarité des fonctions booléennes

  Presque sûrement

$$\frac{2^{m-1} - NL(f)}{\sqrt{2^{m-1} m \log 2}} \longrightarrow 1$$

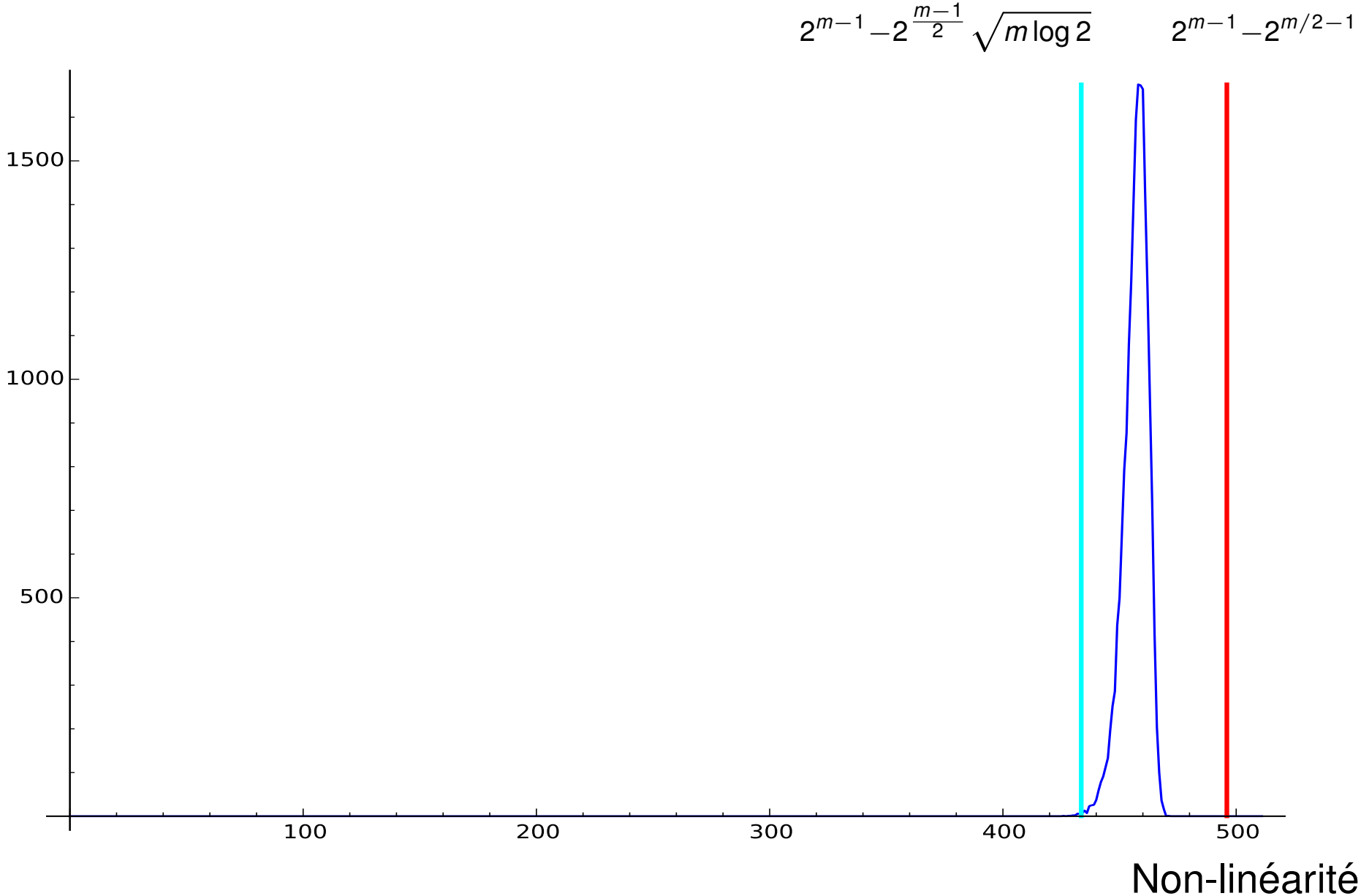
- Combien d'opérations?

  Le calcul se fait en $O(m2^m)$ grâce à la formule

$$NL(f) = 2^{m-1} - \frac{1}{2} \sup_{u \in \mathbb{F}_2^m} |\widehat{f}(u)|$$

  qui permet d'utiliser la transformée de Fourier $\widehat{f}$, que l'on calcule avec la transformation de Fourier rapide.

Distribution de la nonlinéarité pour $m = 10$



Le critère global d'avalanche

- Une fonction f est courbe ssi pour tout $u \neq 0$ dans $\mathbb{F}_2^m$, les vecteurs $f(x) + f(x+a)$ sont équilibrés.
- Une fonction f a un **bon critère global d'avalanche** si pour tout $u \neq 0$ dans $\mathbb{F}_2^m$, les vecteurs $f(x) + f(x+a)$ sont presque équilibrés.
- D'où les définitions
  - L'indicateur absolu (ou l'autocorrélation).

$$\Delta(f) = \sup_{u \neq 0} |\sum_x f(x) f(x+u)|$$

  - La somme des carrés

$$\sigma_f = \sum_{u \in \mathbb{F}_2^m} \left(\sum_x f(x) f(x+u) \right)^2 .$$

L'indicateur absolu (ou l'autocorrélation).

- L'indicateur absolu (ou l'autocorrélation).

$$0 \quad \leq \quad \Delta(f) = \sup_{u \neq 0} \left| \sum_x f(x) f(x+u) \right| \quad \leq \quad 2^m$$

$\uparrow$ fonction courbe

$\uparrow$ fonction ayant une structure linéaire

$D_e f$ est constant

- La limite

  On a (Caullery, Rodier) presque sûrement

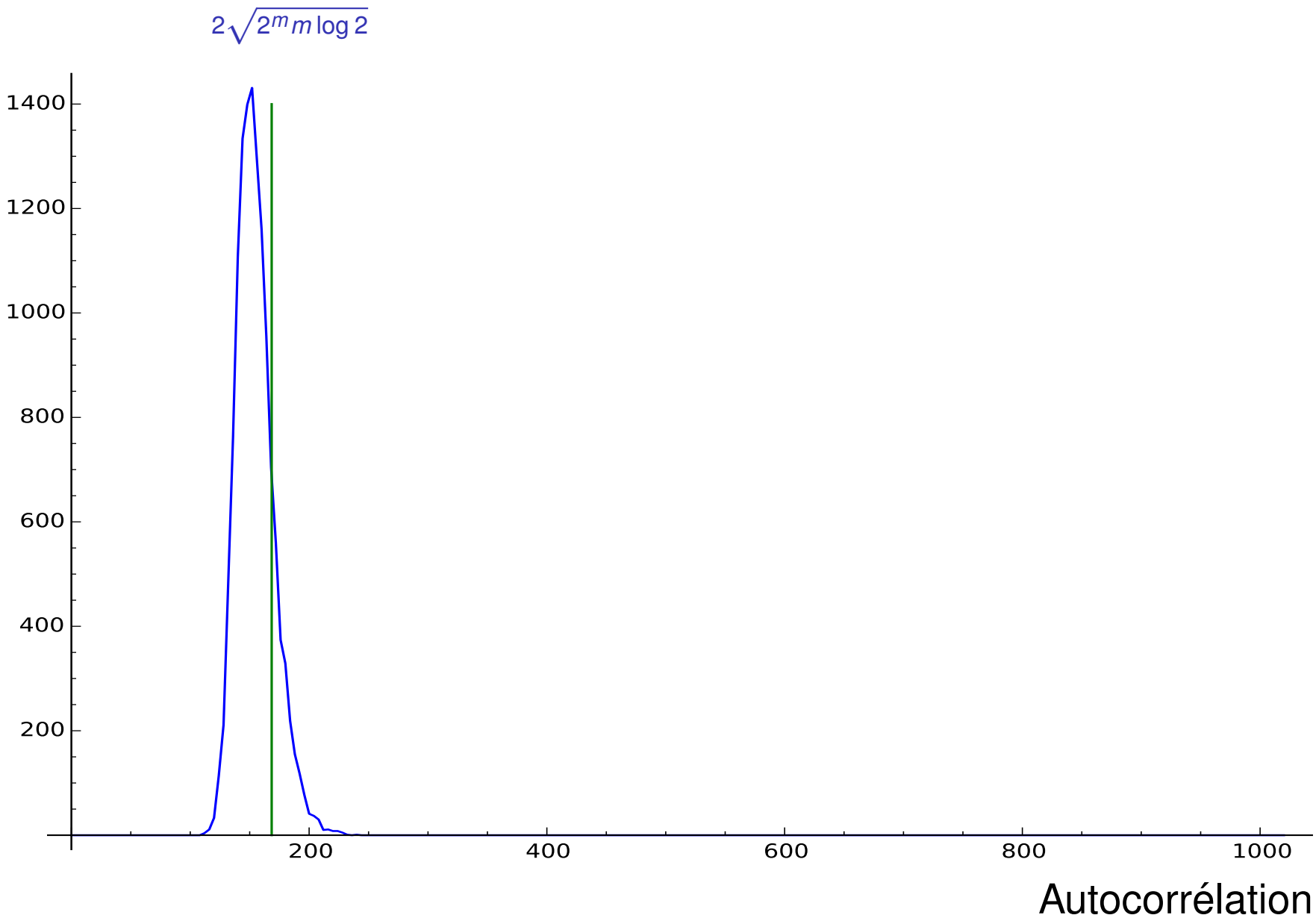
$$\frac{\Delta(f)}{\sqrt{2^m m \log 2}} \longrightarrow 2$$

- Combien d'opérations?

$$\Delta_f(u) = f \otimes f \quad \text{et} \quad \widehat{f \otimes f} = (\widehat{f})^2$$

  d'où $O(2^m m)$ opérations.

L’indicateur absolu (ou l’autocorrélation).



La somme des carrés

- La somme des carrés

$$2^{2m} \quad \leq \quad \sigma_f = \sum_{u \in \mathbb{F}_q} \left(\sum_x f(x) f(x+u) \right)^2 \quad \leq \quad 2^{3m}$$

$\uparrow$ fonction courbe — $\uparrow$ fonction affine

- La limite

  Si f est une fonction aléatoire, on a presque sûrement

$$\lim_m \frac{\sigma_f}{2^{2m}} = 3.$$

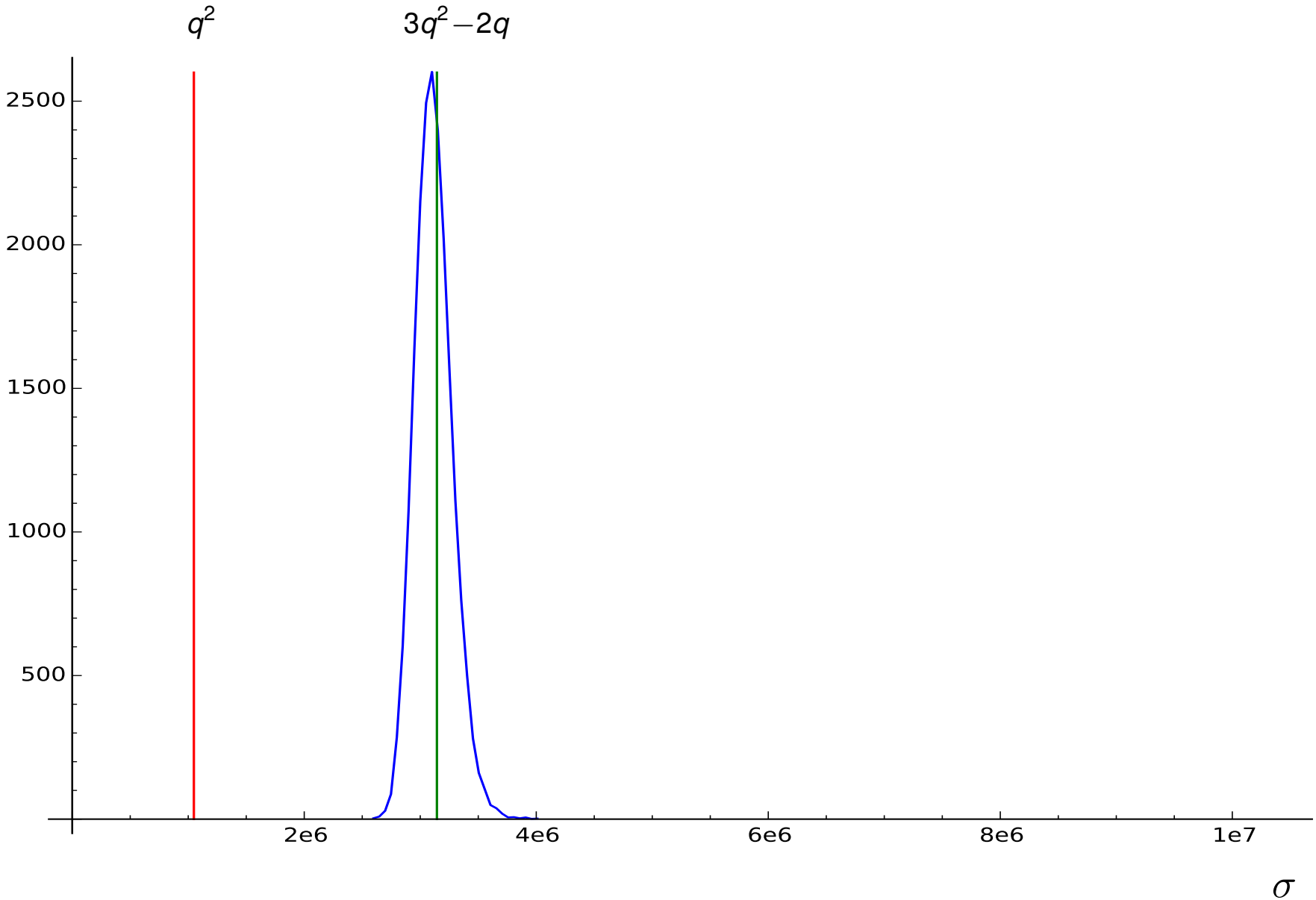
- Les bornes absolues

  Si f est une fonction booléenne avec m variables, et t un réel positif, et $q = 2^m$,

$$\underline{\mathrm{P}}\left(\left| \frac{\sigma_f}{q^2} - 3 + \frac{2}{q} \right| \geq t \right) \leq \frac{40}{t^2 q}.$$

- Combien d'opérations? Il y en a aussi $O(2^m m)$

La somme des carrés



Autocorrelation apériodique d'une suite

- **Autocorrelation aperiodique** d'une suite $f : [1,\ n] \to \pm 1$, donnée par

$$C_u(f) = \sum_{x=1}^{n-u} f(x)f(x+u)$$

et

$$M(f) = \sup_{0<u<n} |\sum_{x} f(x)f(x+u)|$$

- Les séquences binaires avec une petite autocorrélation pour les décalages non nuls ont une large gamme d'applications dans les communications numériques, y compris la synchronisation et le radar
- On sait que le **minimum de** $M(A)$ sur toutes les séquences binaires A de longueur n est ≥ 1 et que il est égal à 1 pour $n \in \{2, 3, 4, 5, 7, 11, 13\}$ (Séquences de Barker).

  Le problème de savoir si ce minimum est > 1 pour tout $n > 13$ est encore ouvert.

La distribution

- La limite

  Soit A_n une suite binaire de longueur n. Alors (K-U Schmidt),

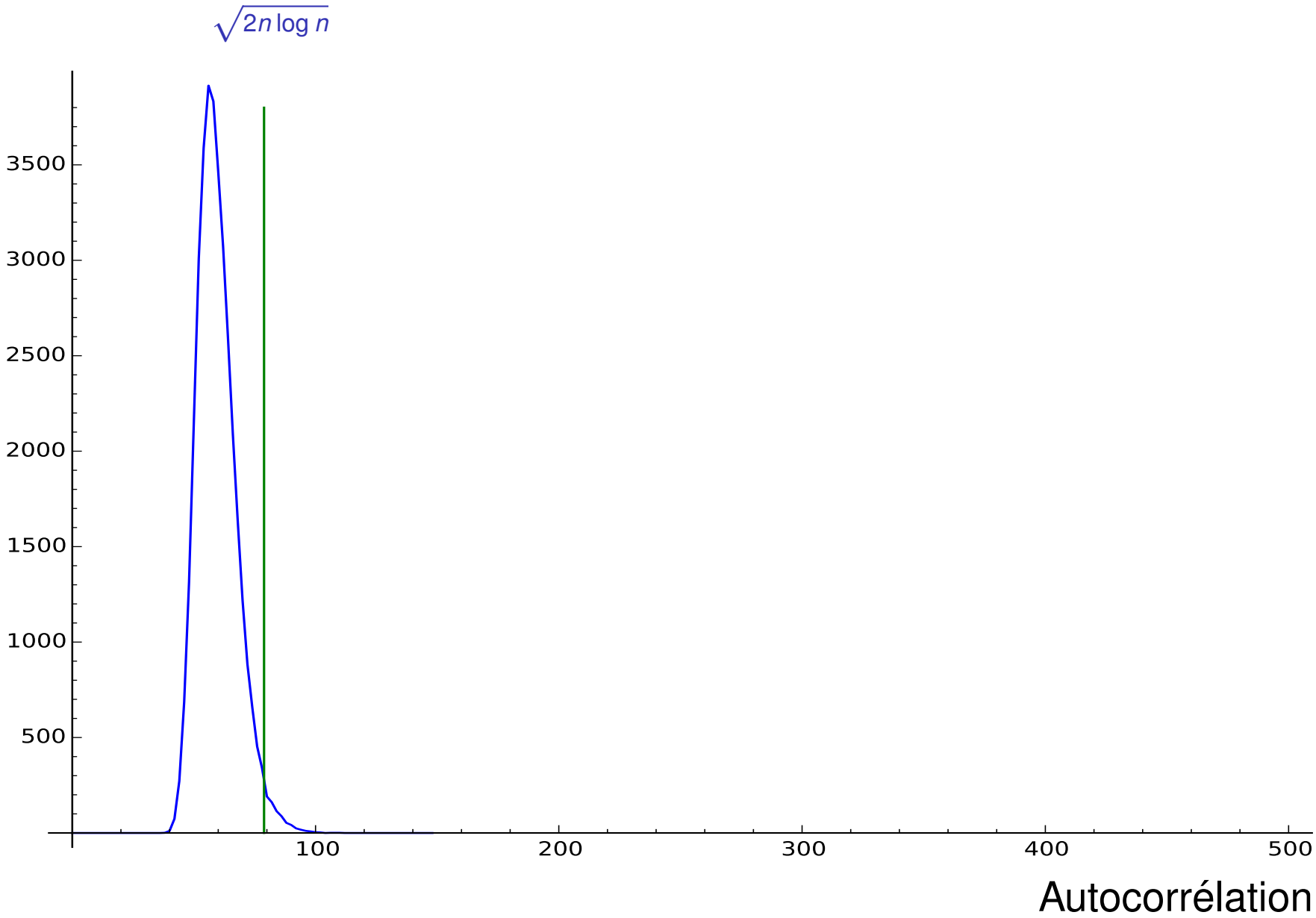
$$P\left(\left|\frac{M(A_n)}{\sqrt{2n\log n}} - 1\right| > \epsilon\right) \longrightarrow 0$$

  pour tout ϵ quand $n \to \infty$.

- Combien d'opérations?

  On a en général $O(n^2)$ opérations.

L'autocorrélation apériodique.



Autocorrelation périodique d'une suite

- **Autocorrelation periodique** d'une suite binaire, donnée par

$$\widehat{C}_u(f) = \sum_{x=1}^{n-u} f(x)f(x+u) + \sum_{x=n-u+1}^{n} f(x)f(x+u-n)$$

et

$$\widehat{C}(f) = \sup_{0<u<n} |\widehat{C}_u(f)|$$

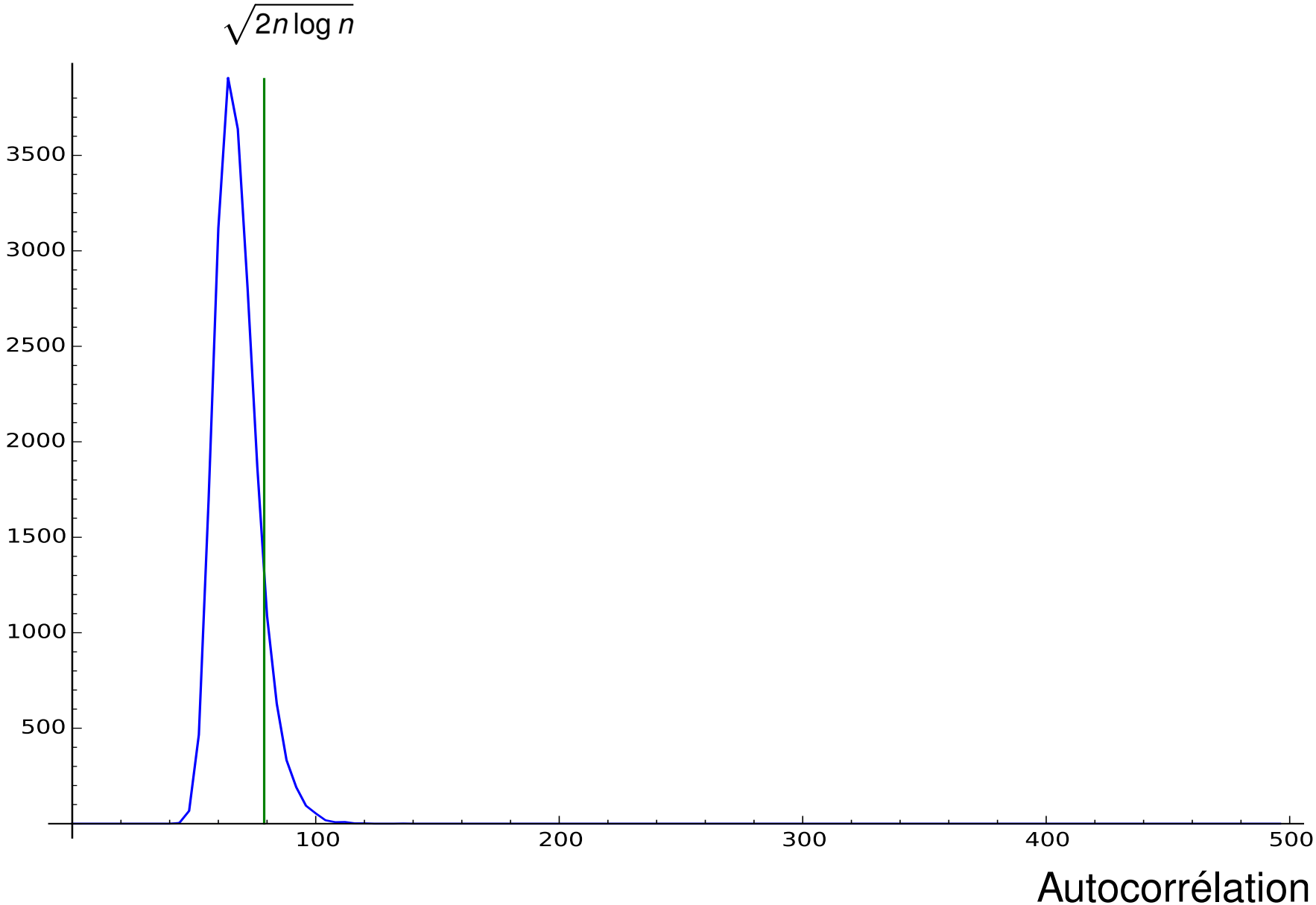
- La limite?

$$P\left(\left|\frac{\widehat{C}(f)}{\sqrt{2n\log n}} - 1\right| > \epsilon\right) \to 0$$

pour tout $\epsilon > 0$ quand $n \to \infty$.

- Combien d'opérations? On a en général $O(n^2)$ opérations.

L’autocorrélation périodique.



Application à la vérification des générateurs de suites aléatoire

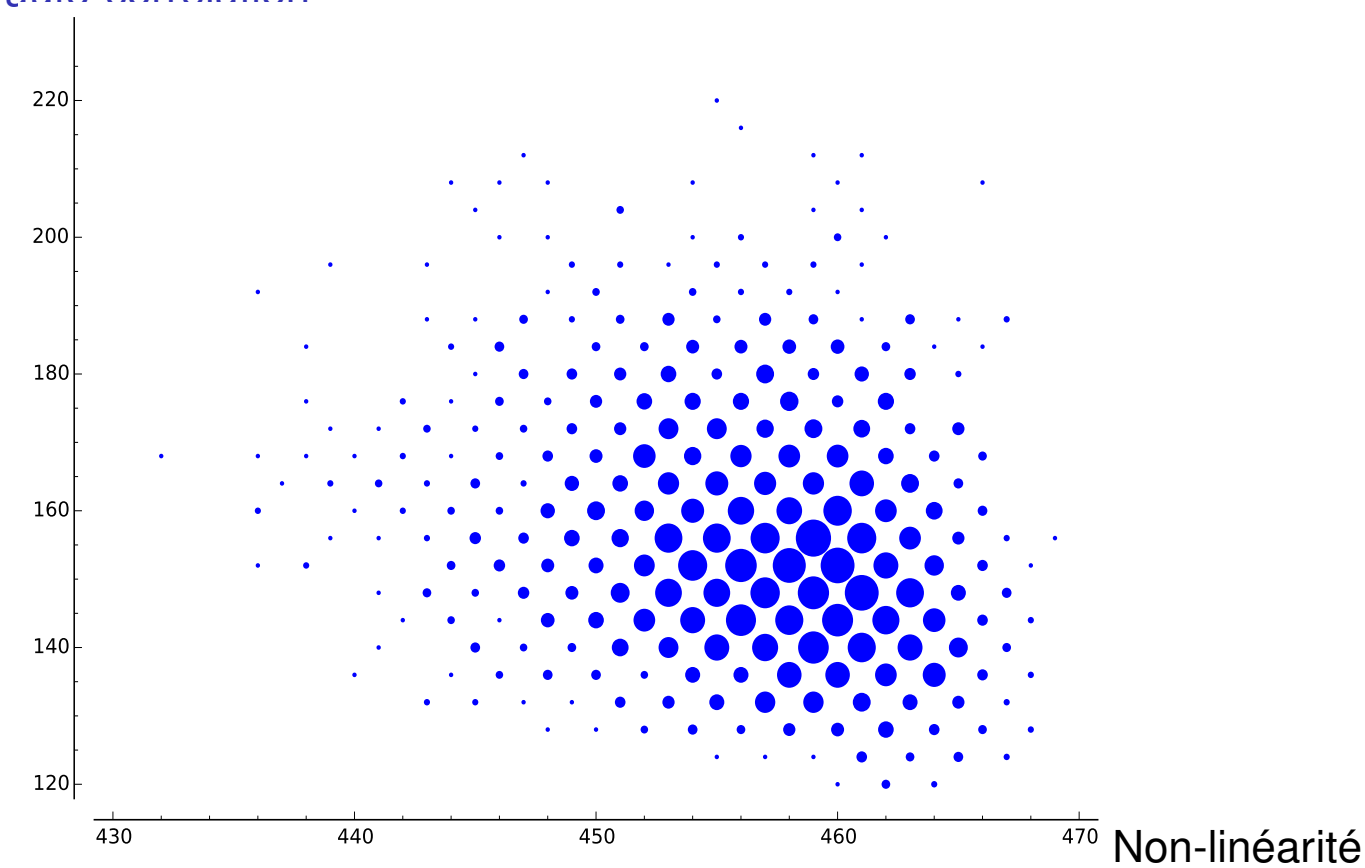
Dans le cas d'un vrai générateur de suites aléatoire (TRNG):

- Problème: **déterminer si une courte séquence binaire semble aléatoire** dans des **conditions de faible puissance de calcul et de mémoire**.
- Les mesures classiques du caractére aléatoire (Knuth, Mauduit-Sakozy) ont une trop grande complexité de calcul.
- Une autre utilisation possible des mesures précédentes a été proposée par Florian Caullery lors d'un atelier réunissant des représentants industriels et des universitaires (une SEME).

Non-linéarité et auto-corrélation

- Ils ont proposé de voir une séquence binaire de longueur 2^n comme la **table de vérité** d'une fonction booléenne, calculer sa **non-linéarité**, son **indicateur absolu** et sa **somme des carrés** et la comparer aux valeurs attendues pour les fonctions booléennes aléatoires.
- L'idée est venue du fait que la non-linéarité des fonctions booléennes aléatoires est concentrée autour de sa valeur espérée.
- J'ai fait le tableau suivant (non-linéarité en abscisses, autocorrélation en ordonnées) pour vérifier s'il y avait un relation entre non-linéarité et autocorrélation pour des fonctions booléennes.

Non-linéarité et auto-corrélation



- La non-linéarité va de 430 à 470 et l'autocorrélation de 120 à 220 (par pas de 4).
- Pour ces valeurs il semble qu'il n'y ait aucune relation entre ces deux critères, donc il faut regarder et la **non-linéarité** et l'**autocorrélation.**

Non-linéarité et auto-corrélation pour les fonctions répétitives

- J'ai pris la **table de vérité** d'une fonction booléenne aléatoire f.

  0 0 0 1 0 1 1 0 1 1 0 0 1 1 1 0 1 0 0 0 ...

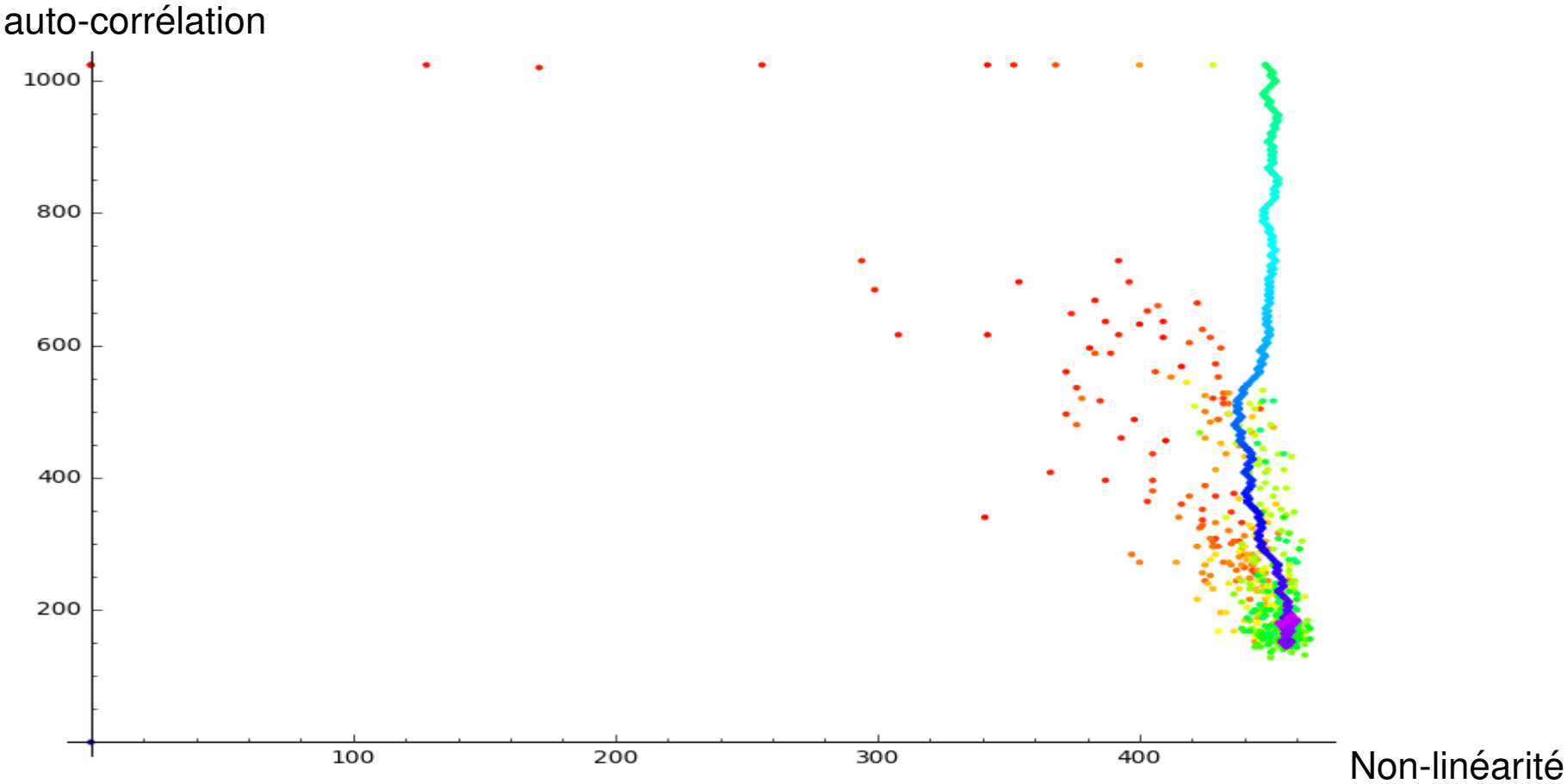
- J'ai construit une **autre fonction booléenne,** en prenant les w premières valeurs de f et en concaténant ces w valeurs autant de fois qu'il fallait.

  $w = 4$ $\quad$ 0 0 0 1 | 0 0 0 1 | 0 0 0 1 | 0 0 0 1 | ...

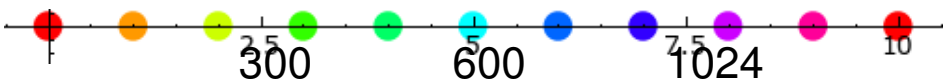
- J'ai fait **varier w de 1 à 1024,**

  $w = 6$ $\quad$ 0 0 0 1 0 1 | 0 0 0 1 0 1 | 0 0 0 1 0 1 | 0 0 0 1 ...

- J'ai construit le même tableau que précédemment (non-linéarité en abscisses, autocorrélation en ordonnées).



- Les points sont colorés suivant la valeur du w correspondant, suivant ce schéma:



- La **ligne de droite** correspond à $512 < w < 1024$. Pour ces valeurs de w, la **non-linéarité varie peu,** tandis que **l'autocorrélation varie** de 1024 à sa valeur moyenne.
- Par conséquent il est utile de regarder les deux critéres.

Conclusion

- Il y a un **point de concentration** pour les fonctions booléennes et pour les suites binaires selon les critères étudiés.
- Ce point de concentration **n'est pas très éloigné du comportement extrêmal** donc la plupart des fonctions booléennes de même que la plupart des suites binaires ont des bons comportements pour les critères étudiés.
- Ces critères sont de plus utiles pour déterminer un **mauvais fonctionnement** d'un générateur de nombres aléatoires.
- L'une des propositions récurrente en crypto appliquée est de trouver de bonnes fonctions booléennes avec une expression simple et de les appliquer en série.

Merci