ANNALES DE L'I. H. P., SECTION B

M. MIGNOTTE

J. L. NICOLAS

Statistiques sur $\mathbb{F}_q[X]$

Annales de l'I. H. P., section B, tome 19, nº 2 (1983), p. 113-121

<http://www.numdam.org/item?id=AIHPB_1983__19_2_113_0>

Ann. Inst. Henri Poincaré, Section B :
Vol. XIX, n° 2, 1983, p. 113-121. *Calcul des Probabilités et Statistique.*

Statistiques sur $\mathbb{F}_q[X]$

par

M. MIGNOTTE
Université Louis-Pasteur, Centre de Calcul de l'Esplanade,
7, rue René-Descartes, 67084 Strasbourg, France

et

J. L. NICOLAS
U. E. R. des Sciences, 123, Avenue Albert-Thomas
87060 Limoges, France

RÉSUMÉ. — Soit P un polynôme de degré n à coefficients dans un corps fini K. On étudie le degré de la plus petite extension de K dans laquelle P se factorise complètement. On montre que le logarithme de ce degré est « presque toujours » voisin de $\frac{1}{2}\,\mathrm{Log}^2 n$.

ABSTRACT. — Let P be a polynomial of degree n with coefficients in a finite field K. We study the degree of the splitting field of P over K. We show that the logarithm of this degree is « almost always » close to $\frac{1}{2}\,\mathrm{Log}^2 n$.

I. INTRODUCTION

Ce travail a pour origine l'étude de la factorisation des polynômes en une variable et à coefficients entiers. Les algorithmes actuels de factorisation comportent, comme étape intermédiaire, la factorisation d'un polynôme à coefficients dans un corps fini. Le principe est le suivant :

a) on réduit le polynôme initial G modulo un certain nombre premier p, *b*) le polynôme F de $\mathbb{F}_p[X]$ ainsi obtenu est factorisé dans cet anneau,

c) pour toute décomposition $F = F_1F_2$ du polynôme F dans $\mathbb{F}_p[X]$, on teste si F_1 est l'image modulo p d'un certain facteur de G, ceci en utilisant une variante un lemme de Hensel.

Les étapes *a*) et *b*) ont un coût raisonnable (polynomial en fonction du degré et de la taille de G). Mais l'étape *c*) peut être très coûteuse (pour certains algorithmes, son coût peut être exponentiel en fonction du degré de G). Le coût de l'étape *c*) dépend du nombre de facteurs irréductibles de F (si F possède k facteurs irréductibles il y a 2^k décompositions possibles de la forme $F = F_1F_2$). Il était donc important d'obtenir des renseignements statistiques sur la factorisation de F pour connaître quel est « en général » le coût de l'étape *c*). Dans un algorithme de Mc. Eliece le degré, noté ici $r(F)$, de la plus petite extension de $\mathbb{F}_p$ dans laquelle F se décompose complètement joue un rôle important. Nous étudierons tout particulièrement cette fonction.

La quantité $r(F)$ est égale au p. p. c. m. des degrés des facteurs irréductibles de F. Il est apparu de nombreuses analogies entre le comportement de cette fonction et celui de l'ordre d'une permutation, qui est égal au p. p. c. m. des ordres des cycles de cette permutation. Cette deuxième fonction a été étudiée par Erdös et Turan [Erd$_1$] et la méthode qu'ils ont utilisée a pu être adaptée ici pour l'étude statistique de la fonction r (cf. aussi [Best] et [Nic$_2$]).

Soit $\mathbb{F}_q$ le corps à q éléments, et soit $\mathbb{F}_q[X]$ l'anneau des polynômes à une indéterminée sur $\mathbb{F}_q$. On désigne par E_n l'ensemble des polynômes unitaires de degré n de $\mathbb{F}_q[X]$. On a donc : card $E_n = q^n$. Soit ϑ_m l'ensemble des polynômes irréductibles de degré m. On pose $I_m = \operatorname{card} \vartheta_m$, et l'on a (cf. : [Ber], [Car])

$$I_m = \frac{1}{m} \sum_{d \mid m} \mu(d) q^{m/d}$$

où μ est la fonction de Möbius. Il en résulte l'encadrement :

$$\frac{q^m}{m} - \frac{2}{m} q^{m/2} \leq I_m \leq \frac{q^m}{m}.$$

Pour $A \in E_n$, on écrira la décomposition en facteurs irréductibles :

$$A = P_1^{\alpha_1} P_2^{\alpha_2} \dots P_k^{\alpha_k}$$

et l'on pose pour $1 \leq i \leq k$: $\deg P_i = n_i$. On a donc :

$$\sum_{i=1}^{k} \alpha_i n_i = n.$$

On définit $\omega(A) = k = \sum_{P|A,\, P\,\mathrm{irr}} 1$, et aussi :

$$r(A) = \text{p. p. c. m.}\,(n_1, n_2, \ldots, n_k).$$

La fonction $g(n) = \max_{A \in E_n} r(A)$ vérifie $\log g(n) \sim \sqrt{n} \log n$, et a été étudiée dans [Nic$_1$].

Nous nous proposons de démontrer le théorème suivant :

THÉORÈME. — A l'exception d'au plus $O\left(\dfrac{q^n}{(\log n)^{7/32}}\right)$, d'entre eux, les polynômes A de E_n vérifient

$$\log r(A) = \frac{1}{2} \log^2 n + O\,(\log n)^{7/4}.$$

II. ÉTUDE DE LA FONCTION ω_T

LEMME 1. — Soit T un nombre entier, $1 \leq T \leq n$. On pose

$$\omega_T(A) = \sum_{\substack{P|A \\ d^\circ P \leq T}} 1.$$

On a :

$$\sum_{A \in E_n} \omega_T(A) = q^n (\log T + O(1)).$$

Démonstration. — On a

$$\sum_{A \in E_n} \omega_T(A) = \sum_{A \in E_n} \sum_{\substack{P|A \\ d^\circ P \leq T}} 1 = \sum_{i=1}^{T} \sum_{\substack{P\,\mathrm{irr} \\ d^\circ P = i}} \sum_{\substack{A \in E_n \\ P|A}} 1$$

$$= \sum_{i=1}^{T} I_i q^{n-i}.$$

Or $I_i = \frac{q^i}{i} + R_i q^{i/2}$ avec $|R_i| \leq \frac{2}{i}$. On a donc :

$$\sum_{A \in E_n} \omega_T(A) = q^n \left(\sum_{i=1}^{T} \frac{1}{i} + \sum_{i=1}^{T} R_i q^{-i/2} \right) = q^n (\log T + O(1)).$$

Notons que la constante sous entendue par le O(1) est absolue, le plus mauvais cas étant $q = 2$. Pour une estimation plus précise voir [Kno], p. 85.

LEMME 2. — Avec les notations du lemme 1, il existe une constante absolue C telle que, pour tout T, $2 \leq T \leq n$, on ait :

$$\sum_{A \in E_n} (\omega_T(A) - \log T)^2 \leq C q^n \log T.$$

Démonstration. — Suivons la démonstration de [Har] pour la fonction ω des nombres entiers ; $\omega_T(A)(\omega_T(A) - 1)$ est le nombre de paires P, Q avec $P \neq Q$ telles que $P \mid A$ et $Q \mid A$, en comptant $(P, Q) \neq (Q, P)$. On a donc en posant $\tilde{\vartheta}_m = \bigcup_{i=1}^{m} \vartheta_i$:

$$\omega_T(A)(\omega_T(A) - 1) = \sum_{\substack{P,Q \in \tilde{\vartheta}_T \times \tilde{\vartheta}_T \\ P \neq Q \\ PQ \mid A}} 1.$$

D'où il vient :

$$\sum_{A \in E_n} \omega_T(A)(\omega_T(A) - 1) \leq \sum_{\substack{P \in \tilde{\vartheta}_T \\ d^0Q + d^0P \leq n}} \sum_{\substack{A \\ PQ \mid A}} 1$$

$$\leq \sum_{i=1}^{T} I_i \sum_{j=1}^{T} I_j q^{n-i-j} \leq q^n \left(\sum_{i=1}^{T} \frac{1}{i} \right)^2 \leq q^n (1 + \log T)^2.$$

La démonstration du lemme s'achève en utilisant la formule

$$(\omega_T(A) - \log T)^2 = \omega_T(A)(\omega_T(A) - 1) + (1 - 2 \log T)\omega_T(A) + (\log T)^2$$

ainsi que le lemme 1.

LEMME 3 (Inégalité de Chebychev pour la fonction ω). —

$$\operatorname{Card} \{ A \in E_n ; \ | \omega(A) - \log n | \geq \lambda \sqrt{\log n} \} \leq C \lambda^{-2} q^n.$$

Démonstration. — On applique le lemme 2 avec $T = n$.

LEMME 4. — Soit $e^{10000} \leq T' \leq n$. Le nombre de polynômes $A \in E_n$ et vérifiant pour tout T, $T' \leq T \leq n$,

$$|\omega_T(A) - \log T| > 6(\log T)^{3/4}$$

est au plus égal à $\dfrac{C}{\sqrt[4]{\log T'} - 2} \cdot q^n$.

Démonstration (Cf. [Erd$_2$]). — D'après le lemme 2, le nombre de $A \in E_n$ tels que

$$|\omega_T(A) - \log T| > Z\sqrt{\log T}$$

est majoré par Cq^n/Z^2.

On applique ce résultat pour $T = T_i = \exp(i^4)$, $Z = Z_i = i$: le nombre de polynômes $A \in E_n$ et vérifiant pour un certain $i > i_0$

$$|\omega_{T_i}(A) - \log T_i| > (\log T_i)^{3/4}$$

est au plus $Cq^n \displaystyle\sum_{i > i_0} \frac{1}{Z_i^2} \leq Cq^n/i_0$.

On choisit i_0 de façon que $T_{i_0+1} \leq T'$: on prend $i_0 = [\sqrt[4]{\log T'} - 1]$, donc $i_0 \geq 9$. Excepté pour au plus $\dfrac{Cq^n}{\sqrt[4]{\log T'} - 2}$ polynômes de E_n, on a, pour $A \in E_n$ et pour tout $i > i_0$,

$$|\omega_{T_i}(A) - \log T_i| \leq (\log T_i)^{3/4}.$$

Soit maintenant T vérifiant $T' \leq T \leq n$. Il existe $i \geq i_0 + 1$ tel $T_i \leq T < T_{i+1}$. On a alors, grâce à la croissance en T de $\omega_T(A)$:

$$\begin{aligned} \omega_T(A) - \log T &\leq \omega_{T_{i+1}} - \log T_{i+1} + \log \frac{T_{i+1}}{T_i} \\ &\leq (\log T_{i+1})^{3/4} + (i+1)^4 - i^4 = 5i^3 + 9i^2 + 7i + 2, \end{aligned}$$

et comme $i \geq 10$,

$$\omega_T(A) - \log T \leq 6i^3 = 6(\log T_i)^{3/4} \leq 6(\log T)^{3/4}.$$

De même :

$$\begin{aligned} \omega_T(A) - \log T \geq \omega_{T_i}(A) - \log T_i + \log \frac{T_i}{T_{i+1}} \\ \geq -(\log T_i)^{3/4} + i^4 - (i+1)^4 = -5i^3 - 6i^2 - 4i - 1 \\ \geq -6i^3 = -6(\log T_i)^{3/4} \geq -6(\log T)^{3/4}. \end{aligned}$$

Ce qui achève la démonstration.

III. COMPARAISON DU P. P. C. M. ET DU PRODUIT

LEMME 5. — Soit $n > e^e$, et soit m un nombre entier $> \log n$. Le nombre de polynômes $A \in E_n$ qui ont deux diviseurs irréductibles dont le degré est un multiple de m est majoré par $q^n \frac{(\log n)^2}{m^2}$.

Démonstration. — Comment fabriquer un tel polynôme A ? On choisit un polynôme irréductible de degré am, $a \geq 1$, et un polynôme irréductible de degré bm, $b \geq 1$, avec $(a + b)m \leq n$. Enfin on multiplie par un polynôme quelconque de degré $n - (a + b)m$. Le nombre de tels polynômes est donc inférieur ou égal à :

$$\sum_{a=1}^{n/m} I_{am} \sum_{b=1}^{\frac{n}{m}-a} I_{bm} q^{n-(a+b)m} \leq \frac{q^n}{m^2}\left(\sum_{a=1}^{n/m} \frac{1}{a}\right)^2 \leq \frac{q^n}{m^2}(1 + \log n - \log m)^2 \leq \frac{q^n}{m^2}(\log n)^2.$$

LEMME 6. — On définit, pour $m \in \mathbb{N}^*$, $\rho(m) = \max_{p|m, p \text{ premier}} p$. Avec les notations de l'introduction, le nombre de polynômes $A \in E_n$ pour lesquels il existe i et j, $1 \leq i < j \leq k$ tels que $\rho((n_j, n_i)) \leq (\log n)^3$ est au plus $\frac{q^n}{\log n}$ pour $n \geq e^e$.

Démonstration. — On doit compter les polynômes de E_n pour lesquels il existe p premier $> (\log n)^3$ tels que $p \mid n_i$ et $p \mid n_j$. D'après le lemme 5, il y en a au plus :

$$\sum_{\substack{p > (\log n)^3 \\ p \text{ premier}}} \frac{q^n (\log n)^2}{p^2} \leq q^n (\log n)^2 \sum_{\substack{p > (\log n)^3 \\ p \text{ impair}}} \frac{1}{p^2} \leq \frac{q^n}{\log n}.$$

LEMME 7. — Soit $R = \{ m \in \mathbb{N}^* \,;\, p \mid n \Rightarrow p \leq (\log n)^3 \}$. Pour $n \geq n_0$, on a :

$$\sum_{\substack{e^{(\log \log n)^4} \leq m \leq n \\ m \in R}} \frac{1}{m} \leq \frac{1}{\log^3 n}.$$

Démonstration. — On peut suivre la démonstration du lemme V de [Erd₁], qui permet d'expliciter n_0. On peut aussi utiliser les résultats de De Bruijn sur la fonction

$$\psi(x, y) = \sum_{\substack{n \leq x \\ p|n \Rightarrow p \leq y}} 1$$

et écrire la somme à majorer au moyen de l'intégrale de Stieltjes

$$\int_{e^{(\log\log n)^4}}^{n} \frac{d\,[\psi(x, (\log n)^3]}{x}.$$

On obtient une meilleure majoration, mais les constantes ne sont pas explicitées dans les estimations de De Bruijn (cf. [Bru]).

LEMME 8. — Soit $n_1, \ldots, n_k$ les degrés des polynômes irréductibles divisant $A \in E_n$. On écrit

$$n_i = a(n_i)b(n_i)$$

avec :

$$a(n_i) = \prod_{\substack{p^\alpha || n_i \\ p \leq (\log n)^3}} p^\alpha \quad \text{et} \quad b(n_i) = \prod_{\substack{p^\alpha || n_i \\ p > (\log n)^3}} p^\alpha.$$

Alors le nombre de polynômes $A \in E_n$ pour lesquels il existe i tel que $a(n_i) > \exp((\log\log n)^4)$, $1 \leq i \leq k$, est au plus $\dfrac{q^n}{\log n}$ dès que n est assez grand.

Démonstration. — Pour un tel polynôme A, $\exists m \in R$, $m > \exp((\log\log n)^4)$ tel que A ait un facteur irréductible dont le degré est multiple de m. Par un raisonnement déjà fait dans la démonstration du lemme 5, pour un m fixe, il y a au plus $q^n\left(\dfrac{1+\log n}{m}\right)$ polynômes A ayant un facteur irréductible de degré multiple de m. Le nombre total de polynômes A est donc majoré par

$$q^n(1+\log n) \sum_{\substack{m \in R \\ e^{(\log\log n)^4} \leq m \leq n}} \frac{1}{m} \leq \frac{q^n}{\log n} \quad \text{(d'après le lemme 7)}.$$

LEMME 9. — Pour n assez grand, à l'exception d'au plus $0\left(\dfrac{q^n}{\log n}\right)$ polynômes, les éléments de E_n vérifient

$$\exp(-2\log n(\log\log n)^4)(n_1 n_2 \ldots n_k) \leq \text{p. p. c. } n(n_1, \ldots, n_k)$$
$$\leq n_1 n_2 \ldots n_k.$$

Démonstration. — On enlève d'abord de E_n les polynômes tels que $k > 2\log n$. D'après le lemme 3, avec $\lambda = \sqrt{\log n}$, il y en a au plus $\dfrac{Cq^n}{\log n}$. On enlève ensuite les polynômes tels que pour un couple i, j, $(i \neq j)$ on ait $\rho((n_i, n_j)) > (\log n)^3$. D'après le lemme 6, il y en a au plus $\dfrac{q^n}{\log n}$. On enlève enfin les polynômes pour lesquels il existe i, tel que $a(n_i) > \exp((\log\log n)^4)$. D'après le lemme 8, il y en a au plus $\dfrac{q^n}{\log n}$. Pour les polynômes restants, les $b(n_i)$ sont premiers entre eux, et l'on a :

$$\text{p. p. c. m.}\,(n_1, \ldots, n_k) \geq \text{p. p. c. m.}\,(b(n_1), \ldots, b(n_k)) = \prod_{i=1}^{k} b(n_i)$$

$$= \frac{n_1 n_2 \ldots n_k}{\prod\limits_{1\leq i\leq k} a(n_i)} \geq (n_1 n_2 \ldots n_k) \exp(-2\log n\,(\log\log n)^4).$$

IV. DÉMONSTRATION DU THÉORÈME

On choisit, dans le lemme 4, $\log T' = (\log n)^{7/8}$. En enlevant $O\left(\dfrac{q^n}{(\log n)^{7/32}}\right)$ polynômes de E_n, on aura pour ceux qui restent :

$$\omega_T(A) = \log T + O(\log T)^{3/4} \qquad \text{pour} \qquad T' \leq T \leq n.$$

On a par ailleurs :

$$\omega_T(A) \leq \omega_{T'}(A) \leq \log T' + O(\log T')^{3/4} \qquad \text{pour} \qquad 1 \leq T \leq T',$$

ce qui entraîne :

$$\omega_T(A) = \log T + O(\log T') \qquad \text{pour} \qquad 1 \leq T \leq T'.$$

Il vient alors :

$$\sum_{i=1}^{k} \log n_i = \int_1^n \log T d[\omega_T(A)] = [\omega_T(A) 1\oplus T]_1^n - \int_1^n \frac{\omega_T(A)}{T} dT$$

$$= \log^2 n + O(\log n)^{7/4} - \int_1^n \frac{\log T}{T} dT - E$$

$$= \frac{1}{2}\log^2 n + O(\log n)^{7/4} - E,$$

avec

$$E = \int_1^n \frac{\omega_T(A) - \log T}{T} dT = \int_1^{T'} \frac{O(\log T')}{T} dT + \int_{T'}^n \frac{O(\log T)^{3/4}}{T} dT$$

$$= O(\log n)^{7/4}.$$

On applique ensuite le lemme 9 : en enlevant $O(q^n/\log n)$ polynômes, on aura pour les polynômes A restants :

$$\log r(A) = \sum_{i=1}^{k} \log n_i + O(\log n (\log \log n)^4))$$
$$= \frac{1}{2}(\log n)^2 + O(\log n)^{7/4}.$$

RÉFÉRENCES

[Ber] E. R. BERLEKAMP, *Algebraic coding theory*, Mc Graw Hill, New York, 1968.

[Best] M. R. BEST, The distribution of some variables on symmetric groups, Nederl. Akademie Wetens. *Proc.*, *ser. A*, t. **73**, n° 5, 1970, p. 385-402.

[Bru] N. G. DE BRUIJN, On the number of positive integers $\leqslant x$ and free of prime factors $> y$. I, Nederl. Akad. Wetensch. *Proc. ser. A*, t. **54**, 1951, p. 50-60.

[Car] L. CARLITZ, Some topics in the arithmetic of polynomials, *Bull. Amer. Math. Soc.*, t. **48**, 1942, p. 679-691.

[Erd_1] P. ERDÖS and P. TURAN, On some problems of a statistical group-theory I. *Zeit. f. Wahrscheinlichkeitstheorie und verw. Gebiete*, t. **4**, 1965, p. 175-186.

[Erd_2] P. ERDÖS, Some unconventional problems in number theory, *Astérisque*, t. **61**, 1979, p. 73-82.

[Har] G. H. HARDY and E. M. WRIGHT, *An introduction to the theory of numbers*, 4th ed. Oxford, at the Clarendon Press, 1960.

[Kno] J. KNOPFMACHER, Analytic arithmetic of algebraic function fields, Marcel Kekker inc., 1979, *Lecture Notes in pure and applied mathematics*, vol. 50.

[Mc. E] R. J. Mc ELIECE, Factorization of polynomials over finite fields, *Math. of Comp.*, t. **23**, n° 108, 1969, p. 861-868.

[Nic 1] J. L. NICOLAS, Ordre maximal d'un élément du groupe S*n* des permutations et « highly composite numbers », *Bull. Soc. Math. France*, t. 97, 1969, p. 129-191.

[Nic 2] J. L. NICOLAS, *Statistiques sur le groupe symétrique*, Séminaire Delange-Pisot-Poitou (théorie des nombres), 13ᵉ année, 1971-1972, n° G2, 6 p.

(Manuscrit reçu le 30 octobre 1981)