MATHÉMATIQUES ET SCIENCES HUMAINES

M. PETRUSZEWYCZ

A. A. Markov, ses probabilités en chaîne et les statistiques linguistiques

Mathématiques et sciences humaines, tome 66 (1979), p. 5-42

<http://www.numdam.org/item?id=MSH_1979__66__5_0>

Math. Sci. hum. (*17*^e année, n°66, *1979*, p.5-42)

A.A. MARKOV, SES PROBABILITES EN CHAINE ET LES STATISTIQUES LINGUISTIQUES

M. PETRUSZEWYCZ [1]

Nous nous proposons de rendre compte pour le lecteur français d'un texte considéré comme un classique de la linguistique mathématique : "Un exemple de recherche statistique sur le texte d'*Evgenii Onegin*[2] illustrant la liaison des épreuves en chaîne"[1][3], publié en russe par le grand mathématicien Andrej Andreevič Markov, (voir Annexe I), en 1913. Cet article n'est généralement connu des linguistes que par le résumé qu'en a présenté Herdan [15] et les références bibliographiques données par Guiraud [14]. Les psychologues le connaissent mieux grâce à la présentation qu'en ont fait G. Miller et N. Chomsky dans leur chapitre "Finitary models of language Users" [27].

Constituer la notion de chaîne et en faire un objet propre de la théorie mathématique a été l'une des contributions essentielles de Markov à la théorie des probabilités - en tout cas la plus féconde - et le distingue de chercheurs à peu près contemporains dont les travaux peuvent être jugés, avec le recul du temps, très proches des siens. La lecture des articles théoriques [2] et [3] et de son livre [4] est à la fois passionnante et difficile. Difficile parce que la pensée de l'auteur se situe à l'apogée de la pensée probabiliste de l'Ecole dite de Čebyšev ou plus fréquemment de Saint-Pétersbourg et en constitue un tournant. On trouve dans ces travaux la genèse de notions très raffinées telles les notions de dépendance - et d'indépendance - de variables aléatoires et de probabilité *a priori*. Mais lecture passionnante quand on

---

(1) Centre de Mathématique Sociale de l'Ecole des Hautes Etudes en Sciences Sociales

(2) Nous translitérons l'alphabet cyrillique selon la norme internationale Ref. ISO/R 9 1968(F) recommandée par l'AFNOR et de plus en plus répandue au risque de troubler ceux qui sont accoutumés à la tradition Tchebychev que nous écrivons ici Čebyšev.

(3) Les références bibliographiques se trouvent sous le titre BIBLIOGRAPHIE p.91-97 dans ce numéro.

aborde [1] et [4] parce que si Markov est, semble-t-il, arrivé à la notion de chaîne à partir de raisonnements abstraits internes au Calcul des Probabilités, il en a donné cette remarquable application : la succession des "voyelles" et des "consonnes" dans deux textes littéraires russes.

Dans une première partie nous essayons de situer la pensée de Markov dans le contexte où elle s'est développée avant d'esquisser les notions nécessaires pour aborder le résumé des articles théoriques [2] et [3] que nous présentons parce que l'auteur les cite et en utilise les résultats dans [1]. De ce dernier article nous ferons, dans un deuxième paragraphe, une paraphrase aussi fidèle que possible à l'original ; nous espérons être assez explicite pour qu'il n'y ait pas de confusion possible entre le texte de l'auteur et les explications que nous avons cru nécessaire parfois d'y ajouter. Ainsi conçue cette partie de notre étude est plus proche de l'original que le texte de Herdan, seul accessible jusqu'ici au lecteur dont la compétence est essentiellement linguistique. Herdan présente un résumé théorique et les manipulations concrètes de Markov en les appliquant aux dix mille premières lettres d'une traduction anglaise d'*Evgenii Onegin*. Le lecteur plus accoutumé au discours mathématique pourra aborder trois références fondamentales en français : Markov [6], Fréchet [11] et Hostinsky [16].

A peu près à l'époque où Markov publie, de 1898 à 1922, des mathématiciens ou des chercheurs travaillant dans d'autres domaines essaient de donner un statut mathématique aux mêmes notions que celles qu'il va utiliser. Ainsi à la fin du XIXe siècle les Ehrenfest présentent leur modèle de diffusion ; Poincaré étudiant le problème du battage des cartes démontre l'ergodicité ; von Smoluchowski et Einstein étudient les mouvements browniens. Ceci nous savons le voir maintenant que la théorie des processus markoviens a été généralisée par Bernstein, Fréchet, Romanovsky,... et complètement axiomatisée par Kolmogorov en 1930. Mais Markov, bien que nourri aux sources de la pensée probabiliste française - il faut lire son cours de Calcul des Probabilités [4] pour voir l'importance historique et philosophique accordée aux auteurs français du XVIIIè siecle - ne saura pas voir l'importance des travaux de Poincaré par exemple. Ceci peut-être en raison d'une "certaine étroitesse d'abord" et d'une "attitude sceptique" à l'égard des résultats des mathématiciens occidentaux que Maistrov [25] attribue aux membres de l'Ecole de Saint-Pétersbourg dans son très remarquable livre sur l'histoire de la théorie des probabilités.

Il nous semble que pour comprendre l'apport de Markov il faut partir des travaux de Čebyšev dans le domaine de la théorie des probabilités qui selon celui-ci "est une partie des mathématiques qui s'occupe de déterminer certaines probabilités à partir de probabilités connues" dont il ne dit pas comment elles sont obtenues. L'intention qui sert de base à ses recherches est de donner des bornes précises aux erreurs commises dans l'application des théorèmes limites c'est-à-dire la loi des grands nombres telle qu'elle avait été formulée par Jacques I Bernoulli et le théorème de Moivre-Laplace. Si le problème de la loi des grands nombres reçut une démonstration simple et définitive dans l'article de Čebyšev "Sur les valeurs moyennes"(1), le théorème central limite exigera à la fois plus de temps et les travaux conjugués de Čebyšev, Markov et Ljapunov. Dans ses travaux Markov se montrera un véritable disciple de Čebyšev en ce qu'il s'appliquera à n'employer que les instruments mathématiques que celui-ci avait adaptés au Calcul des Probabilités : la méthode des fractions dites continues et surtout la méthode des moments. Celle-ci consiste à étudier l'espérance mathématique des puissances successives de la variable aléatoire, mais il peut arriver que certains moments, en particulier les moments d'ordre élevé, n'existent pas. Markov améliorera la méthode par l'introduction de la notion de variable aléatoire tronquée, x_k', dont la définition permet d'affirmer que tous ses moments existent. Pour un choix convenable, c'est-à-dire assez grand de N, la somme des valeurs de la variable tronquée s'écarte assez peu de la somme des valeurs de la variable initiale et les deux sommes ont une même distribution limite. On définit la variable aléatoire tronquée à partir de la variable x_k :

$$x_k' = \begin{cases} x_k & \text{si } x_k < N \\ 0 & \text{si } x_k \geq N \end{cases}$$

On recourt encore maintenant dans l'étude de certaines distributions aléatoires à cette approche que Markov mit longtemps à trouver. Ljapunov, lui,

---

(1) Nous indiquons en note, et non en bibliographie les références des articles ou livres que nous n'avons pas lus : Čebyšev P.L.,"Des valeurs moyennes", *J. Math. pures appl.*, 12, (1867), 177-184. Repris dans "On mean values", *Mat. Sb.*, II, (1867), 1-9.

faisant montre d'une plus grande indépendance vis-à-vis de Čebyšev donnera la première démonstration rigoureuse du théorème central limite : "Nouvelle forme du théorème sur la limite de probabilité", (1901)[1]. Il utilisera à cette occasion un instrument mathématique très puissant, la fonction caractéristique qui définit complètement la distribution de la variable, que les moments de celle-ci existent ou non :

$$\varphi_{(x)}\ t \text{ est définie par } E\ (e^{itx})$$

où x est la variable aléatoire, E étant l'opérateur espérance mathématique.

Cette mésaventure n'est pas la seule qu'on puisse citer en exemple dans la vie de Markov. Précédemment, et toujours pour obtenir une démonstration rigoureuse de résultats énoncés par Čebyšev, il s'est affronté avec Stieltjes qui acceptera de sortir du cadre de l'analyse classique et de ce fait pourra généraliser l'intégrale classique en ce que nous appelons encore intégrale de Stieltjes.

Les recherches de Markov ont en général pour point de départ des résultats de Čebyšev soit qu'il les reformule de façon rigoureuse, soit qu'il en présente une extension, ceci jusqu'au moment où ce faisant il mettra en place la notion de variables aléatoires dépendantes et où il établira ses modèles successifs de "*chaînes de Markov*". A partir des années 1907-1908 il concentrera ses efforts sur cette invention dont la théorie constitue maintenant une partie très importante du Calcul des Probabilités, après les travaux de divers mathématiciens et grâce à l'axiomatisation générale établie par Kolmogorov. Ceci constitue d'ailleurs un juste retour des choses quand on sait que l'un des aspects nouveaux de la mathématique européenne que Markov dédaigna quelque peu fut précisément l'apparition des préoccupations axiomatiques.

I.1. RESUME DES ARTICLES THEORIQUES FONDAMENTAUX.

I.1.1. <u>1907 : "Recherches sur un cas remarquable d'épreuves dépendantes"</u> [2].

Markov est parti du résultat établi par Čebyšev pour des grandeurs indépendantes ("Sur deux théorèmes concernant les probabilités", 1887) et en cherchant à le généraliser il a d'abord trouvé un cas d'application à des "grandeurs dépendant l'une de l'autre", ceci en utilisant l'examen de l'espérance

---

(1) Ljapunov, A.M., "Nouvelle forme du théorème sur la limite de probabilité", *Notes Acad. Sci. Phys.-Math.* Sect. Ser.8, 12, (1901), 1-24.

mathématique du carré de leur somme. Markov a publié une première extension à ce cas particulier dans le Bulletin de la Société de Physique-Mathématique de l'Université de Kazan, en 1898. Dans cet article-ci Markov va montrer, et ceci pour la première fois, que dans le cas d'un grand nombre d'épreuves *dépendantes*, la loi des grands nombres telle qu'elle a été formulée par Čebyšev s'applique et que par conséquent la probabilité de leur somme a pour limite l'intégrale de Laplace et reste comprise dans des limites fixées. Autrement dit, il démontre un théorème limite pour des variables dépendantes en employant la méthode des moments.

Markov définit la *chaîne simple homogène* à un nombre fini d'états, précisément deux états E et F. Il ne lui donne d'ailleurs pas ce nom ; le mot "цеп = cep = *chaîne*" n'apparaîtra qu'en 1908 dans un autre article de Markov dont nous ne parlerons pas : "Extension des théorèmes limites du calcul des probabilités à la somme de grandeurs liées en *chaîne*".

Voici les conditions posées :

1) La probabilité p de l'apparition de l'événement E au cours de n épreuves successives liées entre elles est constante au cours de l'expérience.

2) A chaque épreuve l'un des deux événements complémentaires E ou F se réalise.

Cela permet de définir ce que nous appelons aujourd'hui des probabilités conditionnelles :

a) si l'épreuve précédente a amené la réalisation de E, p_1 désigne la probabilité d'apparition de E ,

b) si l'épreuve précédente a amené la réalisation de F, p_0 désigne la probabilité d'apparition de E[1].

On a donc les probabilités p et $q = 1-p$, p_1 et $q_1 = 1-p_1$, p_0 et $q_0 = 1-p_0$. Les probabilités p, p_1 et p_0 sont liées par la relation $p = p\,p_1 + q\,p_0$. Markov pose aussi une autre relation fondamentale de la chaîne : $\delta = p_1 - p_0$. A partir de là sont définies les relations fondamentales :

$$p_1 = p + \delta q \qquad q_1 = q - \delta q$$

$$p_0 = p - \delta p \qquad q_0 = q + \delta p$$

(1) Markov a changé de notation au cours de la rédaction des articles étudiés ; nous avons choisi d'homogénéiser les indices avec ceux de la notation de [1] .

Markov a introduit un nouvel instrument mathématique, la fonction génératrice qui détermine la probabilité pour que l'événement E, au cours de n épreuves considérées apparaisse un nombre déterminé de fois. Soit la probabilité de m succès au cours de k premières épreuves : $P_{(m\,,\,k)}$, on a après introduction d'un nombre arbitraire ξ :

$$\omega_k = \sum_{(m)} P_{m,k}\ \xi^m \qquad 0 \leq m \leq k$$

et on peut poser après introduction d'un nouveau nombre arbitraire t :

$$\Omega(\xi,t) = \omega_0 + \omega_1 t + \omega_2 t^2 + \dots$$

Markov donne l'expression suivante :

(13)[1] $$\Omega(\xi,t) = \frac{1 - \delta(q\xi + p)t}{1 - \{p\xi + q + \delta(q\xi + p)t + \delta\xi t^2\}}$$

qui va lui permettre d'étudier les moments factoriels. Mais en fait après avoir obtenu la dérivée, effectué une différentiation, il est ramené à ne considérer que certaines fractions constituant l'expression trouvée et à développer ces fractions en séries (convergentes) ordonnées par degrés croissants de δ et indépendantes de n. Après certaines considérations sur les exposants respectifs des quantités p, q et n, il arrive à exprimer l'espérance mathématique de $(m - pn)^k$ sous la forme du polynôme

(23)[1] $$R_k^{(k)} n^k + R_{(k-1)}^{(k)} n^{(k-1)} + \dots + R_i^k n^i + \dots + R_o^k$$

dont les coefficients sont des fonctions entières des quantités p, q et δ. Puis par l'étude de la parité de k au moyen de l'introduction des nombres pair et impair 2ℓ et $2\ell - 1$, en effectuant un passage à la limite il peut écrire $R_\ell^{2\ell}$ sous forme d'une somme :

(28)[1] $$R_\ell^{2\ell} = (a_0 + a_1 + a_2 + \dots + a_\ell)\, p^\ell q^\ell$$

invariable quand on permute p et q. Après avoir montré que $a_0, a_1, \dots$ sont indépendants de p et q et que leur somme a une limite quand n augmente

---

(1) Les numéros des équations sont ceux de l'article original.

indéfiniment, il énonce son théorème limite.

THEOREME LIMITE.- La probabilité des inégalités

$$n\,p + t_1\sqrt{2p\,q\,\frac{1+\delta}{1-\delta}\,n} < m < n\,p + t_2\sqrt{2p\,q\,\frac{1+\delta}{1-\delta}\,n}\ ,$$

où n *est le nombre des épreuves,* t_1 *et* t_2 *sont deux nombres arbitraires,* m *le nombre d'apparitions de l'événement* E*, converge à la limite vers*

$$\frac{1}{\sqrt{\pi}}\int_{t_1}^{t_2} e^{-t^2}\,dt$$

quand le nombre n *croît sans limite,* p*,* q*,* t_1*,* t_2 *et* δ *restant constants.*

Markov a donné, en 1910, dans *Acta Mathematica* [6] une rédaction de cet article, en français, légèrement différente. Il utilise les mêmes méthodes mathématiques mais développe d'une façon plus homogène au reste du texte sa réponse à une critique faite par Ljapunov.

I.1.2. <u>1911 : "Sur un cas d'épreuves liées en chaîne multiple".</u>

Le deuxième texte auquel Markov fera référence est son article paru en 1911 : "Sur un cas d'épreuves liées en chaîne multiple" [3]. Il rappelle l'expression limite de "l'écart probable" qu'il a donnée dans l'article de 1907 :

$\sqrt{2\,p\,q\,\frac{1+\delta}{1-\delta}\,n}$, expression différente de celle qu'on utiliserait pour approximer une distribution binomiale engendrée dans le cas de chaîne simple par une distribution normale : $\sqrt{2\,p\,q\,n}$ (1). Dans cet article il va montrer que sous certaines conditions propres à la chaîne multiple ces deux expressions sont égales.

---

(1) Le coefficient 2 qui apparaît ici est un archaïsme dont l'origine est l'expression analytique utilisée au XIXe siècle ; on écrivait alors

$\frac{1}{\sqrt{\pi}}\int e^{-x^2}dx$ alors que nous écrivons $\frac{1}{\sqrt{2\pi}}\int e^{-1/2\,u^2}du$ $\quad (x = \frac{u}{\sqrt{2}})$.

Il rappelle d'abord les équations ci-dessus citées qui caractérisent la chaîne simple dont on peut donner le schéma suivant :

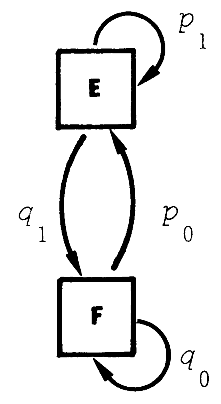


Figure 1. Schéma markovien de la chaîne d'ordre 1

Puis il introduit les notations définissant une *chaîne multiple* d'ordre 2 ou à un nombre fini d'états, ici précisément deux états E et F. La probabilité d'apparition de E dépend des résultats des épreuves précédentes. Si les résultats des deux épreuves précédentes ont été E E alors on désigne la probabilité d'apparition de l'événement E par p_{11}, puis respectivement si on a eu E F : p_{10} , F E : p_{01} ; et enfin F F : p_{00}. Les probabilités de l'événement F seront dans les mêmes conditions respectivement q_{11} , q_{10} , q_{01} et q_{00} , complémentaires des probabilités d'apparition de l'événement E.

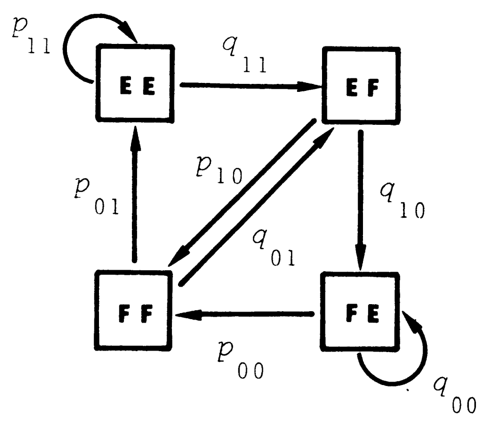


Figure 2. Schéma markovien de la chaîne d'ordre 2

On peut représenter la chaîne multiple soit par le schéma de la Fig.2, çi-dessus, soit par la matrice des probabilités de transition. Markov ne donne pas cette matrice ni dans le texte ni dans la rédaction qu'il en refera pour l'introduire en Annexe dans son cours de Calcul des Probabilités [4] dont nous parlerons plus loin. Les termes de matrice stochastique qu'utilisent les éditeurs soviétiques des Oeuvres choisies [5] dans leur évocation des travaux probabilistes de Markov n'appartiennent pas, pour autant que nous ayons pu en juger au vocabulaire de cet auteur, non plus que les termes de processus ni états initial et final.

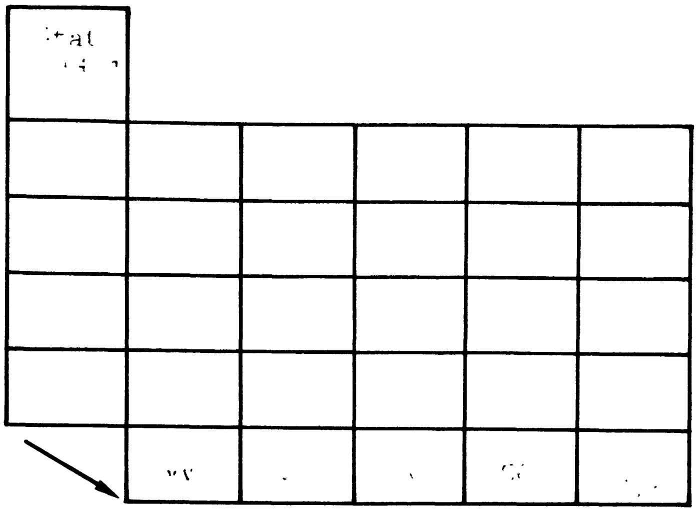


Figure 3. Matrice des probabilités de passage [1]

Revenons au texte de Markov. Toutes ces probabilités sont liées comme nous l'avons déjà vu dans le cas de la chaîne simple. Rappelons les équations de base de celle-ci :

$$p = p\,p_1 + q\,p_o \qquad \delta = p_1 - p_o \qquad p_1 = p + \delta\,q \qquad q_1 = q - \delta\,q$$

$$p_o = p - \delta\,p \qquad q_o = q + \delta\,p$$

Les équations liant les huit probabilités de transition sont :

$$p_1 = p_1\,p_{11} + q_1\,p_{o1} \qquad \text{ou} \qquad p_1\,q_{11} = q_1\,p_{o1}$$

$$p_o = p_o\,p_{1o} + q_o\,p_{oo} \qquad\qquad p_o\,q_{1o} = q_o\,p_{oo}$$

Il introduit alors deux paramètres ε et η et en tire les expressions :

$$q_{11} = q_1(1-\varepsilon) \qquad p_{o1} = p_1(1-\varepsilon) \qquad q_{1o} = q_o(1-\eta) \qquad p_{oo} = p_o(1-\eta)$$

$$p_{11} = p_1 + \varepsilon\,q_1 \qquad q_{o1} = q_1 + \varepsilon\,p_1 \qquad p_{1o} = p_o + \eta\,q_o \qquad q_{oo} = q_o + \eta\,p_o$$

déterminées par les nombres p , q , δ , ε , η .

---

(1) Nous avons remplacé ici les événements désignés dans le texte de Markov par les lettres E et F par les initiales des événements du phénomène étudié : apparition dans un texte d'un graphème Vocalique et d'un graphème Consonantique.

On aura remarqué que les équations de départ sont évidemment complémentaires. Il faut envisager les divers cas possibles concernant les paramètres. Dans ces divers cas nous exprimerons les probabilités de transition en fonction des seuls nombres p, q, δ, ε et η, ce que ne fait pas Markov. Nous ne détaillerons pas les substitutions.

1. $\varepsilon > \eta$

$$p_{11} \begin{cases} = p_1 + \varepsilon q_1 \\ = p + q\,(\delta + \varepsilon\,(1 - \delta)) \end{cases} \qquad q_{11} \begin{cases} = q_1\,(1 - \varepsilon) \\ = q\,(1 - \delta - \varepsilon + \varepsilon\,\delta) \end{cases}$$

$$p_{o1} \begin{cases} = p_1\,(1 - \varepsilon) \\ = p\,(1 - \varepsilon) + q\,(\delta - \varepsilon\,\delta) \end{cases} \qquad q_{o1} \begin{cases} = q_1 + \varepsilon q_1 \\ = p\,\varepsilon + q\,(1 - \delta + \varepsilon\,\delta) \end{cases}$$

$$p_{1o} \begin{cases} = p_o + \eta\,q_o \\ = p\,(1 - \delta + \eta\,\delta) + \eta\,q \end{cases} \qquad q_{1o} \begin{cases} = q_o\,(1 - \eta) \\ = p\,(\delta - \eta\,\delta) + q\,(1 - \eta) \end{cases}$$

$$p_{oo} \begin{cases} = p_o\,(1 - \eta) \\ = p\,(1 - \delta - \eta + \eta\,\delta) \end{cases} \qquad q_{oo} \begin{cases} = q_o + \eta q_o \\ = q + p\,(\delta + \eta - \eta\,\delta) \end{cases}$$

2. $\varepsilon = \eta$. Les probabilités p_{11} et q_{11}, p_{o1} et q_{o1} demeurent inchangées.

$$p_{1o} = p_1\,(1 - \delta + \varepsilon\,\delta) + \varepsilon q \qquad q_{1o} = p\,(\delta - \varepsilon\,\delta) + q\,(1 - \varepsilon)$$

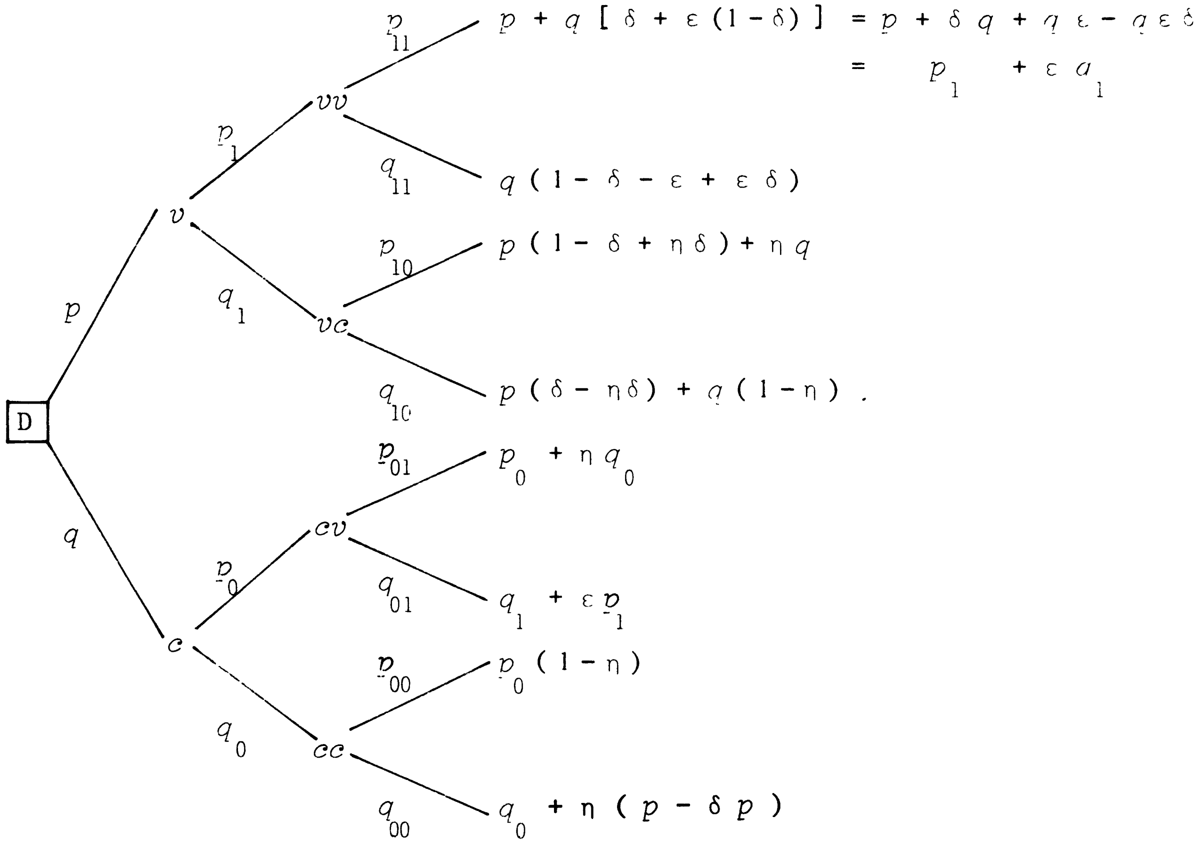
$$p_{oo} = p\,(1 - \delta - \varepsilon - \varepsilon\,\delta) \qquad q_{oo} = p\,(\delta + \varepsilon - \varepsilon\,\delta) + q$$

3. Enfin, dans le cas où l'on a $\varepsilon = -\delta$:

$$p_{11} = p + \delta^2\,q \qquad q_{11} = q - \delta^2\,q$$

$$p_{o1} \begin{cases} = p_{11} + \delta \\ = p + \delta\,(1 + \delta q) \end{cases} \qquad q_{o1} \begin{cases} = q_{11} - \delta \\ = q\,(1 - \delta^2) - \delta \end{cases}$$

On peut aussi représenter la chaîne sous forme d'arbre (voir Fig.4). Pour l'établissement des théorèmes limite que démontre Markov dans les différents cas possibles, il procède à l'étude des systèmes d'équations aux différences finies à l'aide des déterminants. En particulier par l'examen de la fonction entière $F(\xi, t)$ en faisant $\xi = 1$, en développant en un polynôme du 1er degré, et en se donnant deux nombres arbitraires u_1 et u_2, il obtient les bornes de l'expression limite de la probabilité de m apparitions



p_1 : une voyelle apparaît derrière une voyelle (nous symboliserons par $v\underline{v}$).

p_0 : une voyelle apparaît derrière une consonne (nous symboliserons par $c\underline{v}$).

p_{11} : une voyelle apparaît après une voyelle qui suit elle-même une voyelle ($vv\underline{v}$).

p_{10} : une voyelle apparaît après une consonne qui suit elle-même une voyelle ($vc\underline{v}$).

p_{01} : une voyelle apparaît après une voyelle qui suit elle-même une consonne ($cv\underline{v}$).

p_{00} : une voyelle apparaît après deux consonnes ($cc\underline{v}$).

Pour la variable aléatoire : "nombre de voyelles dans une suite de n lettres" :

$$Var = n\,p\,q \left[\frac{1+\delta}{1-\delta} \left\{ \frac{1}{2} \left[\frac{1+\varepsilon}{1-\varepsilon} + \frac{1+\eta}{1-\eta} \right] \right\} + \frac{(q-p)(\eta-\varepsilon)}{(1-\varepsilon)(1-\eta)} \right]$$

$$E = n\,p$$

Figure 4 . L'arbre des états successifs de la chaîne complexe.

de E au cours de n épreuves : $n\,a + u_1\sqrt{2bn}$ et $n\,a + u_2\sqrt{2bn}$ où a et b sont exprimés en fonction des dérivées première et seconde de la fonction F. A l'aide de l'introduction d'une nouvelle variable z telle que $\xi\, t = z$ Markov étudie, toujours par la méthode des déterminants, la fonction transformée $\Phi(z,t)$ et obtient a et b en fonction des valeurs connues :

$$a = \frac{-p(1-\delta)(1-\varepsilon)(1-\eta)}{-(1-\delta)(1-\varepsilon)(1-\eta)} = p \qquad \text{d'où} \quad 1-a = q$$

$$b = p\,q\,\frac{\left\{q(1-3\varepsilon)(1-\eta) + p(1-3\eta)(1-\varepsilon) - 2(1-\varepsilon)(1-\eta)\right\}(1-\delta) + 2(1-\varepsilon\eta)}{(1-\delta)(1-\varepsilon)(1-\eta)}$$

Dans le cas où $\varepsilon = \eta$, on obtient pour b l'expression $b = p\,q\,\dfrac{(1+\delta)(1+\varepsilon)}{(1-\delta)(1-\varepsilon)}$;

il énonce alors un deuxième théorème limite.

THEOREME LIMITE.- La valeur de la probabilité de m est bornée par les expressions :

$$n\,p + u_1\sqrt{2npq\,\frac{(1+\delta)(1+\varepsilon)}{(1-\delta)(1-\varepsilon)}} < m < n\,p + u_2\sqrt{2npq\,\frac{(1+\delta)(1+\varepsilon)}{(1-\delta)(1-\varepsilon)}}$$

Dans le cas où $\varepsilon = -\delta$ *, la probabilité de l'événement* $\left\{np + u_1\sqrt{2npq} < m < np + u_2\sqrt{2npq}\right\}$*, avec les symboles déjà définis,* u_1 *et* u_2 *étant des nombres arbitraires tels que* $u_2 > u_1$ *, a pour limite*

$$\frac{1}{\sqrt{\pi}}\int_{u_1}^{u_2} e^{-x^2}\,dx$$

lorsque le nombre n croît sans limite.

Markov termine son article par l'étude théorique de la chaîne multiliée et montre que le théorème limite ci-dessus reste valable.

I.2. PRESENTATION DETAILLEE DE L'APPLICATION DU MODELE DE CHAÎNE AUX TEXTES LITTERAIRES RUSSES.

I.2.1. Application au roman en vers de Puškin *Evgenii Onegin*.

"Un exemple d'étude statistique sur le texte d'*Evgenii Onegin* illustrant la liaison des épreuves en chaîne" a paru dans le *Bulletin Impérial des Sciences de Saint-Pétersbourg* en 1913. Markov l'avait présenté à la session de la Section de Physique-Mathématique le 23 janvier [1]. Il le reprendra pratiquement textuellement, à une ou deux inversions près de phrases dans [4] et y ajoutera un autre exemple dont nous parlerons plus loin. C'est le texte de l'article que nous suivons, sinon phrase à phrase, du moins de très près.

L'étude concerne la succession de 20.000 lettres russes, sans compter ъ (signe dur) et ь (signe mou) (voir Annexe II), dans le roman en vers d'A.S. Puškin : *Evgenii Onegin*. Ce fragment occupe tout le premier chapitre et seize strophes du second. Cette suite nous fournit 20.000 épreuves liées, chacune d'elles donnant une voyelle ou une consonne. Ici introduisons un commentaire : en fait l'auteur appelle consonne ce que tout linguiste reconnaîtrait comme tel mais il a appelé voyelle toutes les autres lettres de l'alphabet russe (voir Annexe II). Pour cette raison, et aussi parce qu'on peut reprocher à la procédure de Markov de négliger les espaces entre mots, strophes et livres, Herdan a pu affirmer que l'article "donne prise à certaines critiques du point de vue linguistique" [15]. Nous nous sommes posé la question de l'origine de cette application "linguistique" de la théorie des chaînes née, pensons-nous, à l'intérieur de la théorie des probabilités. Aucune réponse précise à cette question n'est apparue au cours des lectures que nous avons faites des diverses biographies de Markov. Une possibilité serait les préoccupations des poètes futuristes énoncées à travers ou peut-être en dépit des scandales et qui provoquèrent une très grande effervescence artistique. Une autre serait l'influence de Jan Baudouin de Courtenay soit par l'intermédiaire de A.V. Vassil'ev de Kazan, soit dans le cadre de l'Académie à Saint-Pétersbourg. Malgré le "psychologisme" de Baudouin de Courtenay des écoles de pensée différentes lui reconnaissent un grand rôle dans le développement pré-troubetzkien de la phonologie russe et une grande notoriété dans le monde savant de son époque [29] [39]. Nous n'avons pas trouvé de preuve d'un lien direct entre les deux savants. Cependant Markov semble avoir porté aux études linguistiques un intérêt suffisant pour critiquer en 1916 [7] un article de "linguistique quantitative" publié en 1915 par A.N. Morozov [28] (voir Annexe II). L'article [7] est fréquemment cité dans les travaux soviétiques contemporains [13] et [46].

Par ailleurs nous présentons dans ce numéro (p.43) un compte-rendu d'un article de V.Ja. Bunjakovskij qui a peut-être joué un rôle dans le choix de cette application "linguistique".

Revenons au texte de Markov.

On admet l'existence d'une probabilité inconnue constante p qu'une lettre soit une voyelle et on cherche la valeur approximative de p (nous dirions plutôt aujourd'hui une estimation de p) en comptant le nombre d'apparitions des voyelles et des consonnes. On trouve également, d'après les observations, la valeur approximative des nombres p_1 et p_o, p_{11}, p_{1o}, p_{o1} et p_{oo} représentant les probabilités des événements du système (probabilités conditionnelles, dans la terminologie contemporaine), dont nous avons explicité les significations sous la fig.4 que ne donne pas Markov.

p_1 : une voyelle apparaît derrière une voyelle (nous symboliserons par $v\underline{v}$)

p_o : une voyelle apparaît derrière une consonne (nous symboliserons par $c\underline{v}$)

p_{11} : une voyelle apparaît après une voyelle qui suit elle-même une voyelle ($vv\underline{v}$)

p_{1o} : une voyelle apparaît après une consonne qui suit elle-même une voyelle ($vc\underline{v}$)

p_{o1} : une voyelle apparaît après une voyelle qui suit elle-même une consonne ($cv\underline{v}$)

p_{oo} : une voyelle apparaît après deux consonnes ($cc\underline{v}$)

L'auteur fait alors référence à [2] et [3] ; puis précise que les notations similaires désigneront par la lettre q les probabilités complémentaires. Pour estimer le nombre p nous en calculons 200 valeurs approximatives à partir desquelles nous déduisons la moyenne arithmétique. En effet on divise la succession des 20000 lettres en 200 suites de 100 lettres et on compte le nombre de voyelles dans chaque centaine. On obtient 200 nombres qui divisés par 100 donnent 200 valeurs approximatives de p. Pour se garder la possibilité de constituer d'autres assemblages de 100 lettres (commentons en anticipant : Markov va mettre en place des manipulations qui détruiront l'effet de chaîne qui existe dans le texte), chaque centaine de lettres est disposée en carré de dix lignes et dix colonnes selon l'ordre du schéma suivant :

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
....									
91	92	93	94	95	96	97	98	99	100

Ensuite on compte combien il y a de voyelles dans chaque colonne séparément et on ajoute les effectifs par deux de la façon suivante :

colonnes : 1 et 6 , 2 et 7 , 3 et 8 , 4 et 9 , enfin 5 et 10.

On obtient ainsi pour chaque centaine de lettres cinq chiffres désignés par les symboles : (1,6) (2,7) (3,8) (4,9) (5,10), dont la somme est égale au nombre de voyelles de cette centaine. Markov donne en exemple les 200 premières lettres d'*Evgenii Onegin*. Les manipulations qu'il a fait subir au texte sont représentées par le Tableau I.

On peut constituer des groupes de cinq centaines de lettres : le premier à partir des premières et sixièmes colonnes ; le deuxième à partir des deuxièmes et septièmes colonnes et ainsi de suite. Le nombre de voyelles de ces nouvelles centaines est déterminé évidemment par les sommes $\sum(1,6)$, $\sum(2,7)$, $\sum(3,8)$, $\sum(4,9)$ et $\sum(5,10)$ se composant des cinq termes correspondants. Les résultats sont présentés en quarante tables dont chacune contient dans la première ligne - cinq nombres (1,6) et leur somme, dans la deuxième ligne - cinq nombres (2,7) et leur somme, et ainsi de suite. Enfin dans la dernière ligne - le nombre de voyelles dans la première centaine, dans la deuxième centaine et ainsi de suite, et pour finir le nombre de voyelles dans les cinq centaines ensemble. Sur la reproduction des 40 tableaux de Markov (Tableau II) on verra que la somme globale de chacun des tableaux a été diminuée systématiquement de 200 pour des raisons typographiques. Le lecteur retrouvera les résultats du Tableau I tout au début du Tableau II, ce qui lui facilitera la lecture en éclairant la construction de ce dernier.

TABLEAU I. Les deux premiers tableaux élémentaires du texte d'Evgenii Onegin.

м	о	й	д	я	ц	я	с	а	м	5
ы	х	ч	е	с	т	н	ы	х	п	3
р	а	в	и	л	к	о	г	д	а	4
н	е	в	ш	у	т	к	у	з	а	4
н	е	м	о	г	о	н	у	в	а	5
ж	а	т	с	е	б	я	з	а	с	4
т	а	в	и	л	и	л	у	ч	ш	4
е	в	ы	д	у	м	а	т	н	е	5
м	о	г	е	г	о	п	р	и	м	4
е	р	д	р	у	г	и	м	н	а	4
3	7	2	5	5	3	5	4	3	5	42

(1,6) : 3 + 3 = 6
(2,7) : 7 + 5 = 12
(3,8) : 2 + 4 = 6
(4,9) : 5 + 3 = 8
(5,10) : 5 + 5 = 10
42

у	к	а	н	о	б	о	ж	е	м	5
о	й	к	а	к	а	я	с	к	у	6
к	а	с	б	о	л	н	ы	м	с	3
и	д	е	т	и	д	е	н	и	н	5
о	ч	н	е	о	т	х	о	д	я	5
н	и	ш	а	г	у	п	р	о	ч	4
к	а	к	о	е	н	и	з	к	о	5
е	к	о	в	а	р	с	т	в	о	4
п	о	л	у	ж	и	в	о	г	о	5
з	а	б	а	в	л	я	т	е	м	4
5	6	3	6	6	3	5	3	4	5	46

(1,6) : 5 + 3 = 8
(2,7) : 6 + 5 = 11
(3,8) : 3 + 3 = 6
(4,9) : 6 + 4 = 10
(5,10) : 6 + 5 = 11
46

Les lettres soulignées sont des "voyelles".

TABLEAU II. Les quarante tableaux de Markov constitués sur les 20.000 premières lettres du texte. (1)

6	8	11	11	13	49	16	11	9	8	7	51	14	12	7	3	6	42	5	11	10	6	10	42
12	11	7	7	5	42	4	8	9	11	10	42	5	5	11	9	11	41	12	8	8	11	7	46
6	6	6	7	13	38	9	9	9	7	10	44	8	10	6	10	7	41	7	7	12	10	9	45
8	10	11	9	4	42	12	9	6	10	7	44	11	11	8	3	10	43	8	12	7	9	9	45
10	11	5	10	8	44	3	8	10	8	9	38	4	4	11	14	8	41	12	8	10	9	8	47
42	46	40	44	43	15	44	45	43	44	43	19	42	42	43	39	42	8	44	46	47	45	43	25

10	6	6	6	7	35	8	7	8	7	10	40	11	11	8	7	7	44	11	10	10	12	6	49
9	12	15	6	9	51	10	9	9	8	8	44	9	6	10	11	11	47	4	4	9	7	9	33
9	3	6	10	9	37	8	9	8	8	8	41	12	9	9	5	6	41	11	13	6	9	10	49
9	11	8	5	6	39	10	6	13	6	12	47	10	8	6	11	11	46	6	7	11	8	6	38
9	10	10	10	9	48	8	12	5	13	6	44	7	6	8	9	8	38	8	6	10	7	12	43
46	42	45	37	40	10	44	43	43	42	44	16	49	40	41	43	43	16	40	40	46	43	43	12

12	9	8	10	10	49	8	9	9	5	8	39	7	7	7	7	9	37	12	7	7	6	8	40
3	10	12	9	10	44	7	9	9	11	7	43	9	13	6	8	4	40	6	8	7	10	8	39
11	11	6	11	10	49	10	6	6	9	9	40	9	7	11	12	14	53	9	10	10	8	7	44
10	8	11	6	7	42	7	8	15	6	9	45	7	11	8	9	7	42	9	5	6	7	7	34
6	8	7	9	6	36	11	7	6	11	10	45	8	10	10	11	9	48	7	11	9	13	7	47
42	46	44	45	43	20	43	39	45	42	43	12	40	48	42	47	43	20	43	41	39	44	37	4

7	4	11	5	7	34	5	5	7	5	9	31	8	6	5	14	11	44	10	9	13	6	12	50
11	14	9	11	9	54	12	6	10	10	8	46	8	12	10	7	4	41	8	8	8	9	5	38
7	6	9	8	9	39	8	14	11	11	10	54	8	10	9	8	14	49	10	10	8	9	10	47
10	9	8	10	5	42	4	3	9	5	9	30	9	5	9	9	6	38	7	9	10	7	10	43
11	10	8	9	11	49	13	14	9	11	7	54	8	13	11	5	10	47	9	8	3	11	7	38
46	43	45	43	41	18	42	42	46	42	43	15	41	46	44	43	45	19	44	44	42	42	44	16

4	11	10	12	5	42	5	11	10	6	5	37	4	4	10	11	5	34	13	11	13	10	10	57
14	9	8	7	14	52	8	9	8	10	10	45	6	12	9	8	10	45	7	10	9	6	2	34
4	8	9	8	4	33	8	8	6	9	9	40	13	4	10	8	6	41	8	8	7	8	12	43
8	14	11	12	6	51	10	6	9	7	6	38	7	10	7	12	11	47	9	11	9	10	6	45
11	6	7	4	14	42	11	9	8	10	12	50	9	13	8	1	8	39	6	3	7	9	9	34
41	48	45	43	43	20	42	43	41	42	42	10	39	43	44	40	40	6	43	43	45	43	39	13

(à suivre)

(1) Pour des raisons de lisibilité ce sont les p. 568 et 569 de [4] qui sont reproduites mais les tableaux sont absolument identiques aux tableaux de [1].

TABLEAU II. suite.

11	6	8	9	5	39	10	10	4	7	9	40	10	8	7	8	8	41	10	3	11	13	5	42
6	10	6	8	13	43	11	10	13	13	9	56	6	9	9	8	7	39	7	11	9	7	10	44
10	5	11	11	6	43	10	7	5	9	6	37	15	9	11	13	9	57	10	10	4	7	7	38
9	12	6	8	10	45	10	5	8	10	10	43	5	10	5	4	7	31	7	7	14	13	7	48
7	11	9	10	10	47	6	13	10	5	6	40	8	9	10	12	9	48	11	9	9	6	15	50
43	44	40	46	44	17	47	45	40	44	40	16	44	45	42	45	40	16	45	40	47	46	44	22
8	8	13	5	8	42	12	7	12	5	12	48	10	14	7	6	6	43	9	6	7	10	5	37
9	10	7	14	9	49	10	8	5	13	4	40	4	6	8	10	14	42	11	10	7	8	9	45
9	11	6	8	7	41	10	13	8	7	9	47	13	6	12	8	5	44	10	10	9	9	10	48
7	9	12	6	9	43	9	4	12	6	9	40	7	13	5	8	10	43	8	6	12	10	10	46
10	9	9	12	9	49	4	12	9	9	8	42	8	5	15	10	9	47	9	11	8	5	11	44
43	47	47	45	42	24	45	44	46	40	42	17	42	44	47	42	44	19	47	43	43	42	45	20
12	13	5	9	11	50	5	11	8	12	10	46	9	11	10	6	13	49	5	9	7	10	6	37
7	7	10	5	8	37	12	8	9	8	6	43	9	8	6	8	6	37	10	9	11	7	7	44
7	7	9	14	7	44	8	11	9	8	7	43	7	7	12	10	9	45	11	11	11	10	8	51
12	13	7	8	10	50	8	5	7	11	8	39	12	12	6	8	8	46	7	7	5	10	10	39
4	4	12	11	9	40	11	11	10	6	8	46	5	7	9	11	4	36	13	8	9	8	10	48
42	44	43	47	45	21	44	46	43	45	39	17	42	45	43	43	40	13	46	44	43	45	41	19
8	6	8	7	14	43	7	9	8	6	7	37	9	11	11	8	8	47	5	7	4	3	7	26
8	14	13	8	4	47	9	8	6	10	11	44	10	8	5	9	10	42	14	10	13	9	5	51
12	4	6	9	11	42	10	9	10	8	10	47	6	8	16	12	11	53	7	8	6	8	9	38
6	8	9	10	8	41	8	7	4	9	4	32	12	11	5	7	8	43	7	10	9	5	9	40
6	8	11	8	6	39	11	8	10	8	9	46	6	5	9	10	8	38	9	10	11	16	7	53
40	40	47	42	43	12	45	41	38	41	41	6	43	43	46	46	45	23	42	45	43	41	37	8
4	7	9	11	10	41	10	8	7	8	7	40	12	10	11	4	5	42	12	13	6	6	10	47
10	7	9	4	9	39	10	8	11	10	7	46	5	9	10	11	11	46	6	3	10	10	4	33
8	13	9	12	10	52	6	11	11	10	10	48	10	8	10	7	13	48	11	11	9	7	14	52
7	5	7	7	12	38	12	8	7	6	5	38	11	8	8	11	5	43	5	8	8	9	9	39
13	10	10	9	5	47	5	9	11	12	11	48	4	8	8	9	11	40	11	6	11	12	7	47
42	42	44	43	46	17	43	44	47	46	40	20	42	43	47	42	45	19	45	41	44	44	44	18

Markov dresse ensuite la statistique des totaux en colonnes ou marges horizontales.

TABLEAU III. Statistique des sommes en bas des carrés.

Nombre des voyelles par centaine	37	38	39	40	41	42	43	44	45	46	47	48	49
Effectif des centaines	3	1	6	18	12	31	43	29	25	17	12	2	1

(voir aussi le graphique Fig.5).

Il effectue ensuite un changement de variable et calcule la moyenne arithmétique pondérée par les distances au mode 43 (qui se trouve avoir accidentellement pour effectif le même nombre 43) et il écrit :

$$43 + \frac{29 + 25 \times 2 + 17 \times 3 + 12 \times 4 + 2 \times 5 + 6 - 31 - 12 \times 2 - 18 \times 3 - 6 \times 4 - 5 - 3 \times 6}{200} = 43{,}19$$

d'où il tire $p \simeq 0{,}4319 \simeq 0{,}432$.

Il calcule ensuite la somme des carrés des écarts des 200 nombres à 43,2 et obtient 1022,8. Il effectue ensuite $\frac{1022{,}8}{200} = 5{,}114$; c'est la variance des 200 nombres observés. Revenons au texte : divisant ce nombre à son tour par 200 on a $\frac{5{,}114}{200} \simeq 0{,}02557$ qui est l'espérance mathématique du carré de l'erreur commise quand on écrit $100\,p \simeq 43{,}2$, estimation de la moyenne arithmétique.

Markov en conclut que ces résultats s'accordent avec l'hypothèse habituelle de la méthode des moindres carrés, savoir que l'on a affaire à des grandeurs indépendantes, hypothèse qui ne se justifie pas plus mal que dans de nombreux autres cas car le lien entre les nombres, en raison de la méthode employée pour les obtenir, est très faible. Il en donne une autre preuve en utilisant ce qu'on appelle "*l'écart probable* ω" [41]. Commentaire : l'écart ou fluctuation probable est défini à partir d'une fonction de répartition connue $F(t)$ par le fait que la variable x ait une probabilité $\frac{1}{2}$

de se trouver entre $F(Z+\omega)$ et $F(Z-\omega)$. L'écart probable ω (ou fluctuation probable est peu connu aujourd'hui dans l'enseignement français (encore qu'il soit cité dans [41] par exemple).
Cette notion construite par les statisticiens "continentaux" et très utilisée à l'époque des travaux de Markov a pris un caractère vétuste au fur et à mesure que l'école britannique a privilégié la notion à peu près correspondante d'écart-type σ. C'est une notion claire et dont l'interprétation est immédiate, alors qu'on ne peut en dire autant de σ. Les avantages pédagogiques qu'elle présente amènent quelques statisticiens français à l'introduire à nouveau dans leur enseignement. La notion d'écart probable figure par contre dans certains manuels de Calcul de probabilité soviétiques. Ceci en raison d'une part sans doute de la tradition : Markov en parle [4], et d'autre part en raison de son importance dans la théorie du tir [40].
L'auteur de ce dernier manuel très classique signale cependant qu'il tend à disparaître au profit de l'écart-type.

Pour une distribution normale $\omega = 0{,}6745\ \sigma$. Dans son texte Markov constate en effet que : $0{,}67 \times \sqrt{5{,}11} \simeq 1{,}5$ et qu'entre les valeurs $43{,}2 - 1{,}5 = 41{,}7$ et $43{,}2 + 1{,}5 = 44{,}7$, se trouvent environ la moitié de ses observations : 31 fois 42 + 43 fois 43 + 29 fois 44 = 103 valeurs sur 200. Il en conclut donc que sa procédure de lecture a fait du nombre de voyelles une variable dont on peut rendre compte par une loi normale.

L'auteur reprend : A l'indépendance de nos grandeurs correspond le fait qu'en les associant par deux, par quatre et par cinq (il veut dire par là qu'il regroupe les valeurs marginales par deux, quatre ou cinq - ceci revenant à ne considérer que le dernier chiffre à droite en bas des quarante tableaux) et en calculant pour ces 100, 50 et 40 combinaisons les sommes des carrés de leurs écarts à

86,4 172,8 216 (moyennes des groupements respectifs)

nous trouvons les nombres

827,6 975,2 1004

Et, dit-il, ces trois nombres ne se distinguent pas beaucoup du nombre trouvé précédemment

$$1022{,}8$$

Si maintenant on revient aux épreuves élémentaires, on remarque que

$$\frac{5{,}114}{100} = 0{,}05114$$

est très différent de la variance théorique : $p \times q$, (de la loi binomiale) :

$$0{,}432 \times 0{,}568 = 0{,}245376 \ .$$

Le rapport entre variance observée et variance théorique est égal à

$$\frac{5{,}114}{24{,}537} \simeq 0{,}208$$

c'est-à-dire n'est que du $\frac{1}{5}$ environ, ce qui s'explique par le caractère lié des épreuves.

Pour expliciter ce lien on peut se servir du calcul approximatif de p_1 et p_o. Examinant le texte composé des 20.000 lettres nous y dénombrons les successions

voyelle , voyelle ;

nous en trouvons 1104 qui divisés par le nombre total de voyelles donne pour valeur de p_1 :

$$\frac{1104}{8638} \simeq 0{,}128 \ .$$

De la même façon comptant le nombre des successions

consonne , consonne

et le divisant par $20000 - 8638 = 11362$, on aurait pu trouver la valeur approximative (estimation) de q_o et par la suite $p_o = 1 - q_o$. Mais on peut remplacer le décompte direct de la façon suivante : en soustrayant 1104 de 8638 on obtient le nombre de consonnes suivant une voyelle : 7534. Comme la première exceptée (il y a évidemment un couple de moins que de lettres), toutes les consonnes doivent suivre soit une consonne, soit une voyelle, le nombre des successions consonne - consonne s'obtient par la différence

$$11361 - 7534 = 3872 \ .$$

De là on tire pour p_o la valeur approximative

$$\frac{7534}{11361} \simeq \frac{7534}{11362} \simeq 0,663 \ .$$

On constate que la probabilité qu'une lettre soit une voyelle varie considérablement selon qu'elle est précédée d'une voyelle ou d'une consonne. La différence, que nous appelons δ (delta grec) est égale à :

$$0,128 - 0,663 = -0,535 \ .$$

Si on admet que la succession des 20000 lettres constitue une chaîne simple on peut prendre, en raison des résultats de [3] pour "coefficient de dispersion : коэффициент дисперсии " [(1)] (coefficient dispersii) :

$$\frac{1+\delta}{1-\delta} = \frac{465}{1535} \simeq 0,3 \ .$$

Il est vrai que ce nombre n'est pas égal à 0,208 mais il en est plus proche que du nombre un qui lui correspondrait si les épreuves étaient indépendantes.

Mais si on adopte le *modèle de chaîne multiple*, on va voir qu'on obtient une meilleure adéquation.

Markov a effectué le décompte des successions VVV et CCC, dont les effectifs sont respectivement de 115 et 505. D'où

$$p_{11} = \frac{115}{1104} \simeq 0,104 \qquad\qquad q_{oo} = \frac{505}{3827} \simeq 0,132$$

[(1)] La mesure de la dispersion est à l'époque un objet d'étude pour de nombreux savants (voir Annexe III).

Nous avons donc :

$$p \simeq 0{,}432 \qquad p_1 \simeq 0{,}128 \qquad p_o \simeq 0{,}663 \qquad p_{11} \simeq 0{,}104$$

$$q \simeq 0{,}568 \qquad q_1 \simeq 0{,}872 \qquad q_o \simeq 0{,}337 \qquad q_{oo} \simeq 0{,}132$$

De là on tire :

$$\delta = -0{,}535 \qquad \varepsilon = \frac{-24}{872} \simeq -0{,}027 \qquad \eta = -\frac{205}{663} \simeq -0{,}309$$

On peut en effet exprimer ε et η en fonction des probabilités de passage, ce que ne fait pas Markov :

$$\varepsilon = 1 - \frac{(1-p_{11})}{1-p_1} \qquad \eta = \frac{1}{p_o}(q_{oo} - q_o)$$

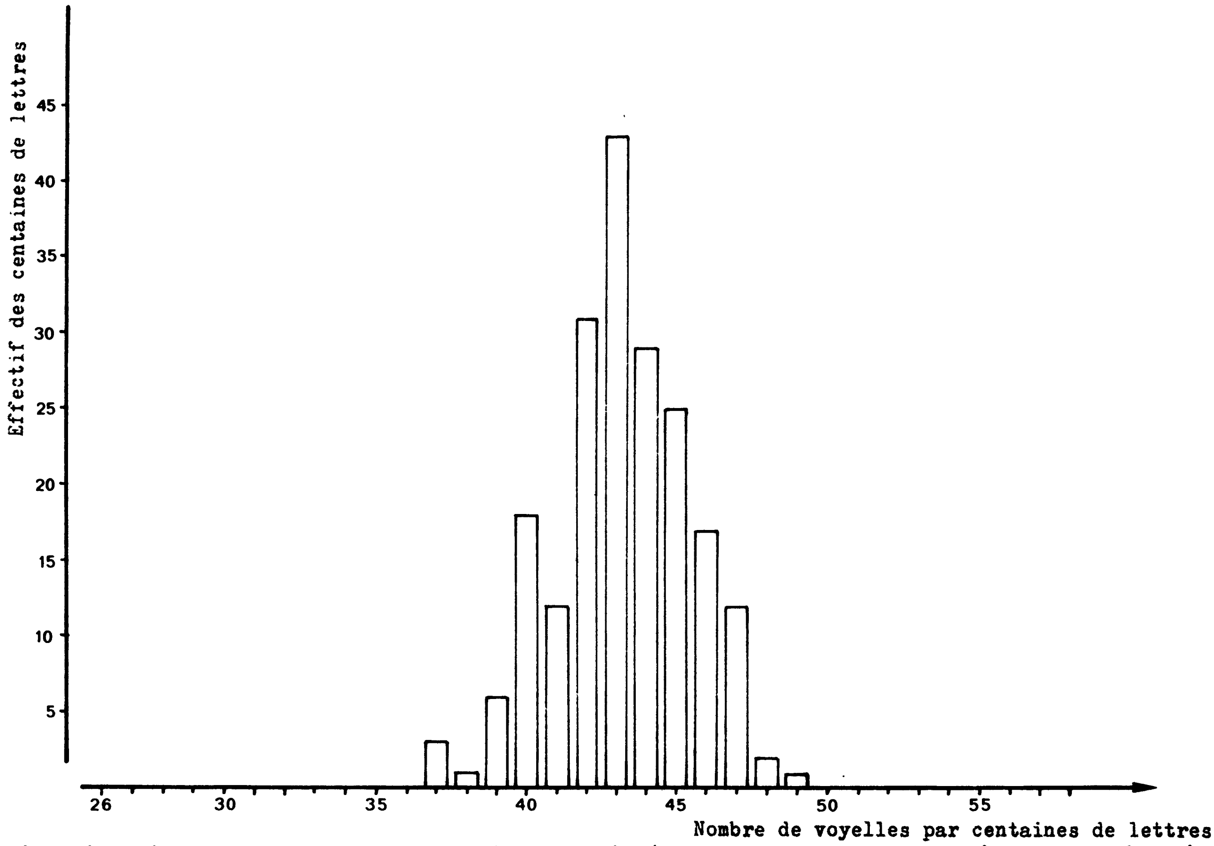
Il faut alors calculer le coefficient de dispersion dans le cas de chaîne multiple d'après la formule donnée dans[3] et que nous rappelons :

$$C = \frac{\{q(1-3\varepsilon)(1-\eta) + p(1-3\eta)(1-\varepsilon) - 2(1-\varepsilon)(1-\eta)\}(1-\delta) + 2(1-\varepsilon\eta)}{(1-\delta)\quad(1-\varepsilon)\quad(1-\eta)}$$

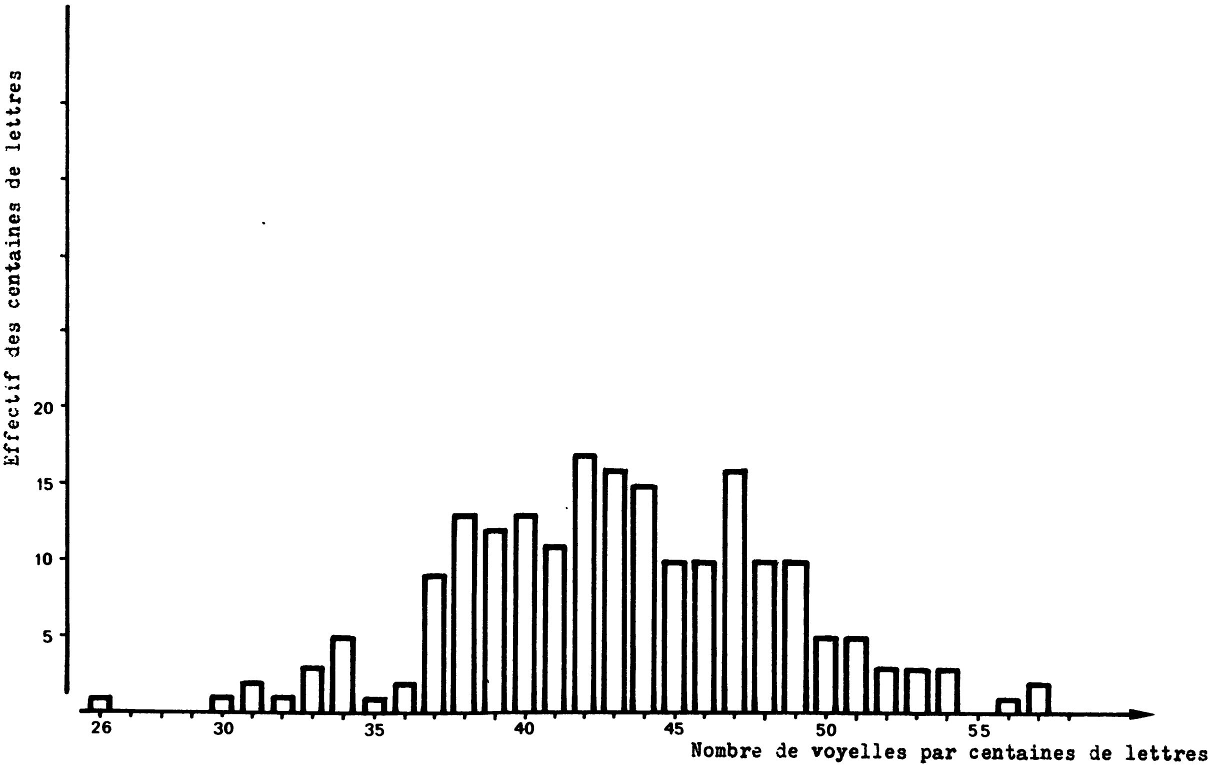
$$= \frac{1+\delta}{1-\delta}\left\{\frac{1+\varepsilon}{2(1-\varepsilon)} + \frac{1+\eta}{2(1-\eta)}\right\} + \frac{(q-p)(\eta-\varepsilon)}{(1-\varepsilon)(1-\eta)}$$

En substituant les valeurs numériques ci-dessus on obtient 0,195. Markov dit qu'il n'est pas possible d'exiger une concordance plus grande avec la valeur 0,208 et que cette concordance ne peut être imputée au hasard mais qu'elle traduit la correspondance des hypothèses qu'il a formulées pour définir la chaîne multiple avec les conditions réalisées par l'exemple de la succession des lettres dans le texte.

On passe ensuite à l'examen des totaux en lignes c'est-à-dire des marges verticales des quarante tableaux reproduits ci-dessus.



Distribution des sommes au bas des carrés (nombre de v par centaines consécutives)



Distribution des sommes à droite des carrés (nombre de v par centaines séparées)

Figure 5. Distribution des sommes partielles des 40 tableaux du Tableau II.
Puskin : *Evgenii Onegin* (20000 lettres décomptes de Markov) .

TABLEAU IV. Statistique des sommes à droite des carrés.

26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	**Nombre de voyelles par centaine**
1	0	0	0	1	2	1	3	5	1	2	9	13	12	13	11	**Effectif des centaines**
42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	**Nombre de voyelles par centaine**
17	16	15	10	10	16	10	10	5	5	3	3	3	0	1	2	**Effectif des centaines**

La moyenne arithmétique de ces 200 nouveaux nombres est bien sûr égale à la précédente : 43,19 , mais la somme des carrés de leurs écarts à 43,2 est beaucoup plus grande que dans le cas précédent : 5788,2. Ici nous passons un paragraphe où l'auteur renvoie à son cours [4] et précisément au chapitre sur la méthode des moindres carrés et la condition d'indépendance des variables mais ce n'est pas, pour nous, le centre du débat.

On verra par ailleurs le graphique Fig.5. C'est une représentation des distributions des marges établie à partir des données des Tableaux III et IV. Mais elle ne figure pas dans l'article de Markov ; nous l'avons insérée ici car elle nous paraît très parlante. Deux distributions de même moyenne, $\bar{x}$ = 43,2 mais de dispersion très différente 5,113 et 28,943 alors que la symétrie -0,18 et 0,06 montre que dans les deux cas les petites valeurs de la variable (nombre de voyelles) l'emportent sur les grandes, ce à quoi on s'attendait.

L'auteur revient à l'examen des marges verticales en faisant remarquer que chacune est le représentant ou plus exactement la somme de grandeurs presqu'indépendantes (en effet rappelons qu'il s'agit de cinq termes (1,6) sur la première ligne, cinq termes (2,7) sur la deuxième ligne, etc...) ; mais en revanche, précise-t-il, *ces termes sont liés par cinq* (c'est nous qui soulignons). En effet, dans chaque groupe de 5000 lettres, les lettres de la première centaine sont contigües à celles de la seconde centaine, les lettres de la deuxième centaine sont contigües à celles de la première et de la troisième centaines et ainsi de suite.

Dans ces conditions le nombre $\frac{5788,8}{200} \simeq 28,944$ peut être considéré comme la valeur approximative de l'espérance mathématique du carré de nos 200 nouveaux nombres

$$49 \ , \ 42 \ , \ 38 \ , \ 42 \ , \ 44 \ , \ \ldots$$

à leur espérance mathématique commune approximativement égale à

$$43,2 \ .$$

Revenant aux épreuves élémentaires nous remarquons que le nombre 0,28944 n'est pas fortement différent de

$$0,432 \times 0,568 = 0,245376 \quad ,$$

(variance théorique de la loi binomiale). Le rapport de la variance observée à la variance théorique $\frac{28944}{24537,6} \simeq 1,18$.

Si on considère la moyenne arithmétique 43,19 , l'espérance mathématique du carré de son erreur ne peut être exprimée par le nombre

$$\frac{28,944}{200} \simeq 0,14472$$

en raison du caractère lié des nombres 49 , 42 , 38 , 42 , 44 , ..., mais conformément à la distribution initiale des lettres en centaines par le nombre

$$\frac{5,114}{200} \simeq 0,02557 \ .$$

On reprend alors l'assemblage des nombres par deux, quatre et surtout cinq centaines dont les moyennes respectives sont :

86,4 172,8 216

(évidemment puisqu'ici on retrouve le même assemblage)

et si on calcule les sommes des carrés de leurs écarts à ces moyennes, au lieu du nombre 5788,8 on obtient

3551,6 3089,2 1004

et on remarque que ce dernier est presque six fois plus petit que 5788,8.

Nous terminerons par un rapide commentaire de la réédition de ce texte dans le cours Calcul des Probabilités [4] mais auparavant nous allons revenir sur le problème de la définition de la probabilité. Dans son cours Markov, en tant que pédagogue, part de la définition de la probabilité d'un événement comme une fraction dont le numérateur est le nombre des cas équipossibles favorables à cet événement et dont le dénominateur est le nombre de tous les cas équipossibles correspondants au problème. Cela constitue une sorte de recul par rapport à la définition de Čebyšev énoncée ci-dessus. Mais il prend ensuite des positions qui traduisent une mise en question de cette définition. D'abord quand il dit : "Dans certains cas ils [les théorèmes fondamentaux] nous donnent les probabilités cherchées directement ; dans d'autres cas ils nous procurent une équation pour la recherche de probabilités dont nous devons supposer préalablement l'existence comme grandeurs connues" [4, p.24]. Ensuite quand à propos du théorème de Bernoulli il dit : "L'apparition d'écarts de $\frac{m}{n}$ par rapport à p pour de grandes valeurs de n sont très peu probables. Mais la possibilité que se présentent de tels écarts n'est pas pour autant complètement écartée et ces cas très peu probables peuvent parfois cependant se réaliser" [25]. Nous écartant du commentaire qu'en fait Maistrov nous préfèrerons dire que nous voyons là un avertissement concernant la soi-disant définition de la probabilité comme "limite" d'une fréquence $\frac{m}{n}$.

Précédé d'une présentation théorique de la chaîne simple puis multiple reprenant les deux articles résumés ci-dessus, l'exemple statistique est introduit par la phrase suivante : "Cet exemple éclaire le fait que les sommes de nombreuses grandeurs liées peuvent constituer (presque) [(1)] des grandeurs indépendantes.

Notre exercice ne nécessite rien d'autre que n'importe quel livre et c'est pourquoi il peut être reproduit par chacun sur une plus grande ou plus petite échelle". Ensuite vient la réédition de [1].

I.2.2. Application au roman en prose d'Aksakov "*Les jeunes années du petit-fils Bagrov*".

Mais après vient un compte-rendu parfois peu explicite d'un même type d'expérience effectué sur les 10000 puis 100000 premières lettres relevées dans un roman en prose d'un auteur du XIXe siècle, agréable mais mineur : S.T. Aksakov (voir Annexe II).

Nous reproduisons ici la distribution des dizaines de lettres selon leur nombre de voyelles.

TABLEAU V. Distribution des dizaines selon le nombre de voyelles.

Nombre de voyelles dans une dizaine	2	3	4	5	6	7	Probabilité de voyelle	Coefficient dispersion
Effectif des dizaines	17	835	4276	4011	827	34	0,44898	0,25

(1) Les parenthèses sont de l'auteur et n'indiquent pas ici un commentaire de notre part.

Ensuite il a effectué le compte direct[1] des doublets de voyelles : il en trouve 6588 d'où il peut calculer pour le texte de 100000 lettres :

$$p_1 = \frac{6588}{44898} \simeq 0,147 \qquad p_0 = \frac{38310}{55102} \simeq 0,695 \qquad \delta = -0,548 \qquad \frac{1+\delta}{1-\delta} \simeq 0,29$$

Puis il donne un tableau des valeurs observées et théoriques (c'est-à-dire calculées en faisant l'hypothèse d'indépendance, en prenant pour p la valeur 0,44898 donc $q = 0,5512$, $n = 10$ et en multipliant par 10000) des distributions des dizaines de lettres selon leur nombre de voyelles.

TABLEAU VI. Distributions observée et théorique (en bas) des dizaines de lettres selon leur nombre de voyelles.

0	1	2	3	4	5	6	7	8	9	10	Probabilité de voyelle	Coefficient dispersion
26	233	793	1699	2320	2319	1548	740	261	59	2	0,44898	1,05
26	210	771	1675	2389	2335	1586	758	226	41	3		

L'adéquation est de toute évidence très satisfaisante ; notons que dans cette étude Markov a utilisé pour "casser" la dépendance le même procédé que celui utilisé pour le texte de Puškin mais en prenant une "distance" différente : précisément il a considéré comme consécutives les lettres séparées par neuf lettres du texte (toujours avec les conventions énoncées ci-dessus).

---

[1] Pour les décomptes sur le texte d'Aksakov l'auteur exprime des réserves sur leur exactitude parce qu'il a travaillé sur une copie manuscrite du roman alors qu'il garantit explicitement la précision de ceux concernant le texte de Puškin. Certains des tableaux concernant le texte d'Aksakov nous ont paru ininterprétables pour ne pas dire contradictoires. Il faut se rappeler par ailleurs que la 4e édition du Cours est posthume et donc que les épreuves n'ont pas été relues par l'auteur.

Nous dirons ici notre sentiment à ce sujet. A la suite de Markov nous avons pris beaucoup de temps pour vérifier nos décomptes qui ne coïncident pas avec les siens. Dans les deux cas s'il subsiste quelques erreurs nous pensons qu'elles sont assez peu nombreuses pour ne pas influer sur la validité générale des raisonnements qui les utilisent.

Il est donc, nous semble-t-il, indiscutable que ce qu'il cherche à définir ce sont bien les notions de variables aléatoires dépendantes ou indépendantes et à montrer que sous certaines conditions, qu'il a précisées dans ses textes théoriques, la chaîne multiple engendre une distribution qui, à la limite, est bien approximée par la loi normale mais avec une dispersion toute différente de celle qu'on emploie traditionnellement ($\sqrt{n\,p\,q}$).

L'auteur termine son étude par la comparaison des valeurs de p_1 et p_0 dans les corpus de 10000 lettres et 100000 lettres. Il utilise une autre "distance" ; en effet ici il considère comme consécutives les lettres séparées par une lettre. Dans les 10000 premières lettres il y avait 4462 voyelles et les doublets de voyelles séparées dans le texte par une lettre étaient au nombre de 2470. D'où :

$$p_1 = \frac{2470}{4462} \simeq 0,55 \qquad p_0 = \frac{1992}{5538} \simeq 0,36 \qquad \delta \simeq 0,19 \qquad \frac{1+\delta}{1-\delta} \simeq 1,5$$

Les décomptes pour les 100000 lettres permettent de calculer :

$$p_1 = \frac{24773}{44898} \simeq 0,552 \qquad p_0 = \frac{20125}{55102} \simeq 0,365 \qquad \delta \simeq 0,187 \qquad \frac{1+\delta}{1-\delta} \simeq 1,46$$

Markov présente ainsi une vérification expérimentale de la stabilité des p_1 et p_0 qui nous permet de considérer les valeurs prises comme des "caractéristiques" du texte ou peut-être de l'auteur dans ce texte. Mais alors il paraît intéressant d'examiner si ces caractéristiques peuvent permettre de distinguer des corpus : nous avons tenté l'expérience et nous en rendrons compte dans un autre article.

A N N E X E I

Notice biographique sur Andrej Andreevič Markov (1856-1922).

C'est le fils d'un fonctionnaire des Eaux et Forêts qui, à sa retraite, devient régisseur du domaine d'une dame C.A. Val'vat'eva. Il eut un demi-frère Vladimir qui mourut précocement de tuberculose non sans avoir donné une élégante démonstration d'un théorème énoncé par Andrej. Celui-ci se marie en 1883 avec la fille aînée de C.A. Val'Vat'eva, Marie (1860-1942) et a un fils, lui aussi appelé Andrej Andreevič, née en 1903, et qui est le très réputé mathématicien soviétique spécialiste de la théorie des algorithmes principalement et vivant aujourd'hui à Moscou.

Né le 2 juin 1856, à Riazan, il arrive à Saint-Pétersbourg dans les années soixante du siècle ; il y fait toutes ses études d'ailleurs médiocres sauf en mathématiques.

En 1874 il entre à l'Université de Saint-Pétersbourg où il suit les cours et séminaires de Zockin, Zolotariev et Čebyšev dont il sera le véritable disciple.

1884 : il soutient sa thèse de Doctorat : "Sur quelques applications algébriques des fractions continues".

1886 : il est nommé professeur extraordinaire à l'Université et, sur proposition de Čebyšev, stagiaire à l'Académie des Sciences.

1896 : il est nommé académicien titulaire.

En 1883, date du retrait de Čebyšev, il fait pour la première fois le cours de Calcul des Probabilités qu'il assurera sans interruption de 1885-86 à sa mort à Petrograd le 20 juillet 1922 (1).

A.A. Markov fut l'un des plus éminents représentants de l'Ecole de Saint-Pétersbourg et ses travaux tant en théorie des Nombres qu'en Calcul des Probabilités ont contribué de façon décisive aux progrès accomplis dans ces domaines au cours des années où il a publié. La bibliographie de [5] comprend cent vingt six références dont les dernières sont posthumes.

---

(1) Si l'on en croit le Professeur A.A.Markov dont nous pensons qu'il faut prendre pour définitive la biographie qu'il a été chargé de rédiger pour l'édition des Oeuvres Choisies [5]. Nous ne pouvons mentionner toutes les biographies que nous avons lues. Nous signalerons cependant les articles de l'Encyclopédie Soviétique sur les trois Markov, Gnedenko et A.A. Youschkevitch de l'Académie des Sciences de l'U.R.S.S. dans Dictionnary of Scientific Biography, Ch.C.Gillespie, Princeton University, Ed.in chief, Charles Scribner's sons, New-York, 1974.

Trois points nous ont particulièrement frappé bien qu'ils soient signalés de façon plus ou moins insistante ou parfois omis.

Le premier qui est partout signalé est non pas sa dévotion mais sa fidélité aux méthodes mathématiques utilisées par Čebyšev. Il ne s'agit nullement d'un conservatisme aveugle. Les termes dans lesquels Markov exprime les réserves qui s'imposent, quant à la démonstration du théorème limite donné en 1887 par son maître, dans sa correspondance avec A.V. Vassil'ev de Kazan, ne laissent aucun doute sur la lucidité de son jugement. Mais il va s'employer à perfectionner la démonstration esquissée par Čebyšev et y parviendra en 1898. Peu après, 1900-1901, paraissent les démonstrations de Ljapunov. Et l'on reste confondu devant l'acharnement, à la limite même on pourrait dire l'entêtement dont fera preuve Markov puisqu'il lui faudra huit années de travail pour démontrer, par la méthode des moments qu'il perfectionne à cette occasion, la formulation très générale du théorème central limite telle qu'elle avait été exprimée par son rival. Cette démonstration, donnée en 1913, sera publiée en annexe dans la troisième édition de son cours de Calcul des Probabilités (1913).

Un autre trait qui mérite d'être souligné chez Markov est son anticonformisme notoire. Dans sa biographie, son fils cite plusieurs des lettres que le savant adressera aux autorités compétentes pour protester courageusement contre la cassation, sur ordre du Tsar Nicolas II, de l'élection de Gorki (Peskov de son vrai nom) à l'Académie des Sciences.

Lorsque les Romanov célébrèrent en grande pompe leurs trois siècles de règne, il organisera la même année, 1913, la célébration du deux centième anniversaire de la loi des Grands Nombres ; c'est en effet en 1713 que *L'ars Conjectandi* de Jacques I Bernoulli a été publié, à titre posthume.

Enfin un dernier point qui est absent de certaines biographies nous paraît particulièrement important et rejoint d'ailleurs le premier. Markov participera à la mise en place et à la gestion d'une Caisse de retraite ; dans ce genre d'activité, il n'était pas question d'être "transcendantal" pour reprendre le mot même dont Čebysev qualifiait l'attitude d'esprit de Ljapunov, mais au contraire, comme il l'enseignait avec insistance à ses élèves de s'en tenir à des problèmes concrets, d'aboutir à des solutions effectives, de connaître les marges de l'erreur commise quand on devait se contenter d'une approximation. On trouve ici, dans la vie même d'un grand probabiliste, l'une des origines incontestables et les plus anciennes de la science aux progrès de laquelle il a contribué si remarquablement.

A N N E X E   II

Quelques points d'orthographe russe et notices sur les auteurs littéraires et scientifiques cités.

I - L'ORTHOGRAPHE.

L'alphabet russe comprend vingt signes consonantiques que Markov a compté comme consonnes ; dix signes vocaliques qu'il a compté comme voyelles. Pour le yod /j/, Jakobson Cherry et Halle [18] disent qu'il l'a considéré tantôt comme voyelle (ce que Markov avait annoncé dans ses conventions) tantôt pas sans expliquer comment ils s'en sont aperçus ni combien de fois. Enfin il existe deux signes orthographiques (les jers) : ъ (signe dur) et ь (signe mou) qu'il n'a pas comptés. Il peut y avoir à cela deux raisons :

- certains membres des commissions de réforme de l'orthographe qui se sont succédées de 1904 à 1917 envisageaient de les supprimer. Le décret du 5 janvier 1918 établit la nouvelle orthographe ; il ne touche pas au signe mou et s'il supprime tous les signes durs finaux il les maintient à l'intérieur des mots (il y en a six sur les 20 000 lettres décomptés par Markov, presque tous entre préverbe et verbe dans les verbes de mouvement).

- Markov voulait donner un exemple d'application simple et donc à deux états ; retenir les signes orthographiques l'obligeait à considérer trois états. Il en est de même pour l'espace,vide ou contenant une ponctuation.

Quand Markov entreprend son étude sur la succession des voyelles et des consonnes sur le texte de Puskin il travaille en orthographe ancienne, nous sommes en 1913 ; mais nous pouvons refaire ce décompte sur un texte en orthographe moderne - ce que nous avons fait, effectivement - cela ne change pas les décomptes, du moins en ce qui concerne le nombre des doublets, triplets, mais parfois ce ne sont pas les mêmes voyelles.
Dans tous les décomptes que nous avons faits nous avons suivi les conventions de Markov, bien qu'elles soient critiquables du point de vue linguistique, et ceci dans tous les cas, que nous ayons travaillé sur un texte en orthographe moderne (*Evgenii Onegin* et Lénine) ou sur un texte en orthographe ancienne (*Kapitanskaja dočka*).

II - BREVES NOTICES SUR LES AUTEURS CITES.

1. *Alexandre Sergeevič Puškin* (1799-1837). D'une famille de boyards appauvris - la première mention historique du nom a trait à l'affaire Boris Godounov -

et descendant par sa mère du célèbre "maure" de Pierre le Grand, Puškin est le plus grand des poètes en langue russe. Son génie très précoce se révèle alors qu'il est lycéen à Carskoe Sielo. Il va montrer à ses contemporains que le russe est une langue littéraire au même titre que le français, l'allemand, le polonais. Bien qu'il s'en défende dans le roman même, *Evgenii Onegin*, l'histoire de ce dandy peu sympathique, est un autoportrait des folles années de l'auteur et une prédiction de sa propre mort en duel. Formellement le roman est composé de huit livres (et du voyage qui ne figure pas dans toutes les éditions). L'ensemble comporte un peu plus de 5000 vers dont environ 99% est rimé. Chaque livre est composé de strophes de quatorze vers de quatre ïambes chacun sur le schéma rimé *ababeeeciddiff*. A une ou deux exceptions près, chaque strophe constitue un poème indépendant.

Traduite, "la poésie de Puškin perd sa qualité d'envoûtement, son éclat sonore. Le miracle né de la concordance parfaite de la pensée, du rythme, des sons disparaît".

La fille du Capitaine, sa dernière oeuvre publiée de son vivant, a été rééditée chez Aubier-Montaigne, en édition bilingue, en ancienne orthographe.

2. *S.T. Aksakov* (1791-1859), père des slavophiles bien connus, se met a écrire à l'âge de soixante-dix ans. Parmi les romans qui le rendront célèbre "*Les années d'enfance du petit-fils Bagrov*".

3. *Nicolas Alexandrovič Morozov* (1854-1946), fils d'un riche propriétaire terrien du gouvernement de Iaroslav et d'une serve, il est membre actif de divers mouvements révolutionnaires puis terroristes des années soixante-dix - quatre-vingt du XIXe siècle : initialement c'est un "narodnik". Il émigre une première fois en 1874 et collabore alors au journal de Bakounine. Arrêté à son retour en Russie en 1875, il est inculpé au "procès des 193" et ne fut libéré qu'en 1878. Il émigre à nouveau en 1880 et rencontre Karl Marx. Arrêté une deuxième fois quand il essaie de rentrer en Russie en 1881, il est inculpé au "procès des 29" et condamné à la prison à vie. Libéré en 1905, il se consacre à divers travaux littéraires et scientifiques et devient, après avoir assumé diverses fonctions officielles membre de l'Académie des Sciences en 1932. Dans l'article que lui consacre l'Encyclopédie soviétique aucune allusion n'est faite à la statistique lexicale comme instrument de caractérisation d'un auteur ; d'après les documents dont nous avons pu disposer, il semble très probable qu'à travers Lutosławski (et donc Baudouin de Courtenay) il soit dans la langue russe

l'initiateur peut-être, en tout cas un pionnier. Pour imparfaits qu'aient été ses essais Morozov mériterait, à notre avis, que mention en soit faite.

4. *Victor Jakovlevič Bunjakovskij* (1804-1889). Eminent mathématicien russe qui enseigne à l'Université de Saint-Pétersbourg à partir de 1846 date où il fait paraître le premier *Cours de Calcul des Probabilités* en langue russe qu'avait précédé l'établissement d'un lexique de mathématique pure et appliquée (1839). Il est Académicien titulaire à partir de 1841. Il se consacre essentiellement à l'extension de l'enseignement des mathématiques à tous les niveaux universitaires. Il fut le maître de Čebyšev.

5. *Victor Vladimirovič dit Vélimir Hlebnikov* (1885-1922). L'un des plus remarquables poètes de l'Ecole Futuriste russe. Luda Schnitzer a présenté une très intéressante édition bilingue [20](1). Sans trop accorder d'importance aux "chatoyants oripeaux du formaliste", une étude statistique ne pouvait rester indifférente à une poétique prétendant altérer, à travers des recherches sur la composition morphématique , les fréquences des voyelles et des consonnes dans la langue russe. R. Jakobson le considère comme "l'un des plus grands poètes russes et peut-être le plus important au monde des poètes modernes" [7]. Majakovskij, disait : "la gloire poétique de Hlebnikov est incommensurablement moins grande que son importance".
La lecture de [23] est une façon rapide et documentée de revivre cette extraordinaire époque de foisonnement artistique dont Jakobson (dans "Retrospect" in Selected Writing, vol.II) dit l'influence déterminante qu'elle eut sur lui.

6. *Ivan Petrovič Caharov* (1807-1863). Un remarquable ethnolinguiste qui a laissé un relevé impressionnant de littérature populaire, de contes, jeux, dictons, chants [48] couvrant toute la Russie d'Europe à partir des origines jusqu'au XIXe siècle.

---

(1) Voir aussi les revues : *Change*, *Action Poétique*, et les *Manifestes futuristes russes* traduits et présentés par L. Robel.

A N N E X E   III

Coefficients de dispersion

La notion de dispersion a été l'objet de nombreuses études qui ont été menées selon des voies différentes, à la fin du XIXe siècle, par les statisticiens britanniques et par ce qu'ils ont appelé "l'école continentale" dont des représentants éminents sont Charlier, statisticien français, Dormoy un actuaire français et la statisticien allemand Lexis. [4] [16].

Il s'agit de rendre compte des variations enregistrées lorsqu'en examinant N échantillons de n événements on observe la proportion des apparitions d'un certain caractère ou événement. On admet que ces observations permettent d'estimer la proportion véritable du caractère dans la population d'où sont tirés les échantillons.

Dormoy (*Théorie mathématique des assurances sur la vie*, Paris, 1878) pour construire son *coefficient de divergence* part de la notion d'<u>écart moyen</u> : soit la suite de fréquences (de l'événement considéré) :

$f_1, f_2, \ldots, f_N$ et $f_o = \dfrac{f_1 + f_2 + \ldots + f_N}{N}$. L'auteur fait le rapport $\dfrac{\sigma_o}{\sigma'_o}$, des écarts moyens observé et théorique correspondant à une loi de Bernoulli, où $\sigma_o = \dfrac{1}{N}\left[|f_1 - f_o| + |f_2 - f_o| + \ldots\right]$ et

$$\sigma'_o = \sqrt{\frac{2}{\pi}}\sqrt{\frac{f_o(1-f_o)}{N}} .$$

Lexis, lui, part de la notion d'écart quadratique moyen ; son *coefficient de dispersion* Q est le rapport $\frac{\mu_o}{\mu'_o}$ où $\mu_o = \sqrt{\frac{(f_1 - f_o)^2 + (f_2 - f_o)^2 + \ldots}{N-1}}$ est l'écart quadratique moyen observé et $\mu'_o = \sqrt{\frac{f_o(1-f_o)}{N}}$ est l'écart quadratique moyen théorique.

On peut aussi le présenter sous une forme faisant intervenir l'écart probable, dont nous avons parlé ci-dessus, considéré alors comme la demi-longueur de l'intervalle délimité par le quartile inférieur ou Q_1 et le quartile supérieur Q_3 pour une loi normale d'écart type unité. Le coefficient de dispersion est alors $\frac{R}{r}$ où $r = 0{,}67449\ [f_o(1-f_o)\ /\ N\ ]^{1/2}$ estime l'erreur probable propre à l'échantillon de la valeur f et $R = 0{,}67449\ [\Sigma(f_i - f_o)^2\ /\ N-1\ ]^{1/2}$ estime l'erreur probable de f. Pour les grands échantillons les deux quantités sont très voisines. La valeur du rapport décèle un cas appelé "normal" de dispersion quand elle est voisine de 1 ; si elle est supérieure à 1 on dit qu'on est en présence d'un cas supra ou hypernormal, si elle est inférieure à 1 il s'agit d'un cas hypo ou subnormal de dispersion. Markov qui cite dans son manuel Lexis (*Zur theorie der Massenerscheinungen in der menschlichen Gesellschaft*, 1877), Bortkiewicz (*Gesetz der Zahlen* et 2 articles dont celui de la *Math. Enzyklopädie*) et Dormoy (op.cité ci-dessus), s'inscrit bien entendu dans la tradition "continentale". Il a consacré à l'étude de la dispersion diverses publications qui ne figurent pas à notre bibliographie car il y aborde la question à son niveau le plus général c'est-à-dire dans le cadre du Calcul des Probabilités ; il donne de la dispersion des expressions complexes faisant intervenir les racines du déterminant de la fonction caractéristique de la variable. Dans le cadre du modèle de chaîne simple appliqué à la succession des voyelles et des consonnes, son coefficient est d'une interprétation immédiate : δ exprimant la liaison entre les probabilités d'apparition d'une voyelle au moment t selon l'état du système au moment $t-1$: s'il n'y avait pas effet de chaîne c'est-à-dire si les événements voyelle et consonne étaient des variables aléatoires indépendantes δ serait négligeable et le coefficient de dispersion serait voisin de l'unité.

Dans la correspondance que Markov échangeait dans les années 1910-1917 avec A.A. Čuprov [21] il étudiait les travaux de Lexis relatifs à la stabilité des séries statistiques. Les deux savants, bien qu'ils ne s'accordent pas sur les raisons profondes de la dispersion ont évidemment vu le rapport de Q avec le χ^2 de Pearson, car c'est là l'autre voie, la voie britannique qui permet d'aborder la mesure de la dispersion. En raison de l'hégémonie que devait exercer l'école britannique c'est ce critère qui fut le plus utilisé par les statisticiens au détriment des coefficients continentaux. Ceci malgré les travaux de Bortkiewicz qui utilisera constamment Q après avoir calculé son espérance et son écart-type, et la théorie extensive et complète qu'en donna Čuprov en 1922. Pour revenir à la correspondance des deux mathématiciens russes, l'étude de la "normalité" de Q leur permit de donner quelques exemples particuliers permettant la généralisation du théorème central limite à des cas d'épreuves faiblement liées.
Sir R.A. Fisher démontra en 1928 l'équivalence formelle de χ^2 et de Q[(1)].

(1) "Moments of sampling distributions", Proceedings of the London Mathematical Society vol. 30 (1928), reproduit in Contributions to Mathematical Statistics, New York, Wiley, 1950.