

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

L. AMY

Recherches statistiques sur l'identification des encres au campêche-chrome

Journal de la société statistique de Paris, tome 91 (1950), p. 188-202

http://www.numdam.org/item?id=JSFS_1950__91__188_0

© Société de statistique de Paris, 1950, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

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

Recherches statistiques sur l'identification des encres au campêche-chrome.

Dans un premier exposé, effectué en juin 1949 devant la Société de Statistique de Paris, nous avons montré comment le calcul des probabilités permettait de mieux préciser la notion d'identification du père dans la recherche de la paternité par les groupes sanguins (1). Ultérieurement, dans un deuxième mémoire publié également dans ce même Journal, nous avons encore appliqué le calcul des probabilités à l'identification des individus par leurs dessins papil-

(1) L. AMY, « Probabilité, groupes sanguins et paternité ». *Journal de la Société de Statistique de Paris*, 86, p. 102-112 (1945).

lares digitaux (1). Exposer les résultats obtenus par les méthodes statistiques dans l'identification des documents manuscrits à base d'encre au campêche-chrome, tel est le but de ce troisième mémoire.

Les encres au campêche-chrome sont des encres noires essentiellement formées par une solution acide d'hématoxyline. Ce dernier pigment est obtenu lui-même par oxydation chromique de l'extrait de bois de campêche. Immédiatement après leur fabrication, les encres au campêche-chrome diffèrent légèrement les unes des autres par des différences de nuances variant du brun foncé au bleu foncé, suivant leur teneur plus ou moins grande en acide. Au fur et à mesure de leur emploi, les différences s'accroissent dans l'encrier en raison de cette même acidité : les encres attaquent les plumes de fer et dissolvent plus ou moins ce métal, ce qui diminue leur acidité. Celle-ci varie également plus ou moins rapidement par évaporation et suivant leur nature. Enfin, une partie du pigment précipite et l'acidité joue encore un rôle important dans ce phénomène car les encres sont d'autant plus instables qu'elles sont plus voisines de la neutralité.

Une fois l'encre déposée sur le papier, le pigment subit une lente transformation : lorsque le pigment est récent, il se redissout presque entièrement dans les solutions acides en donnant une coloration rouge cerise. Au fur et à mesure du vieillissement, l'encre change lentement de couleur et prend une nuance de plus en plus jaunâtre; en même temps le virage au rouge sous l'action des acides se fait de moins en moins franc et il reste sur le papier une fraction insoluble de plus en plus importante. Il est probable que les modalités de cette évolution sont encore en grande partie fonction de l'acidité primitive de l'encre.

En collaboration avec M. le professeur C. Sannie (2), nous avons mis au point une méthode permettant de différencier sur un même document des encres au campêche-chrome d'origine différente. Cette technique consiste essentiellement à évaluer l'importance de la fraction insoluble du pigment dans l'acide chlorhydrique ainsi que leur teneur en fer. Nous avons, en effet, constaté que dans un document écrit en une seule fois, ces deux quantités : fraction insoluble et teneur en fer étaient à peu près constantes et variaient au contraire dans de larges limites pour des encres différentes ou même pour la même encre déposées sur le papier à des époques différentes.

En raison des quantités extrêmement faibles de pigment dont on dispose (l'analyse devant s'effectuer sur des fragments de textes manuscrits, quelques lettres le plus souvent, parfois de simples jambages) les méthodes microchimiques ordinaires gravimétriques, volumétriques, colorimétriques même, seraient encore trop grossières. La méthode que nous avons mise au point consiste à transformer, sur place, les substances à doser et à mesurer photographiquement les variations de densité optiques qui se produisent. On mesure d'abord l'opacité optique du trait d'encre, puis on le traite par une solution chlorhydrique de

(1) L. AMY, « Valeur de la preuve en dactyloscopie ». *Journal de la Société de Statistique de Paris* (1^{re} partie), 87, p. 80 à 87 (1947).
2^e partie, 87, p. 183 à 195 (1947).

(2) « Recherche sur l'analyse de l'encre des documents manuscrits. I. Méthode générale d'analyse quantitative. Applications aux encres au campêche-chrome, Ch. SANNIE et L. AMY, *Annales de Médecine légale*, 18, p. 401 à 419, 1938.

ferrocyanure de potassium : une partie du pigment se dissout, la fraction insoluble vire au rouge, tandis que le fer se transforme en bleu de prusse.

Le bleu de prusse étant à peu près transparent pour la radiation violette 435 μ du mercure, une mesure de densité optique pour cette radiation donne sensiblement la fraction insoluble tandis que le fer est donné par une mesure de densité pour les radiations rouges, radiations pour lesquelles le bleu de prusse présente le maximum, et la fraction insoluble le minimum d'opacité. Finalement, on caractérise les traits d'encre par les rapports :

$$\frac{\text{densité après virage}}{\text{densité avant virage}} \quad \text{en lumière rouge} \quad \text{et} \quad \frac{\text{densité après virage}}{\text{densité avant virage}} \quad \text{en lumière violette}$$

Si ces rapports étaient absolument constants pour la totalité d'un même manuscrit, l'interprétation des résultats ne présenterait aucune difficulté et permettrait dans chaque cas l'identification ou la différenciation des encres. Pour des raisons mal élucidées, ces rapports subissent des fluctuations quelquefois assez importantes. On peut diminuer cet inconvénient en multipliant les mesures et en effectuant des moyennes mais de toute façon on n'obtient jamais des valeurs identiques et d'ailleurs on est souvent limité dans la multiplication des mesures par la brièveté du texte ou la nécessité de n'altérer qu'au minimum la pièce à conviction.

Nous nous proposons de montrer dans ce travail comment les méthodes statiques permettent d'interpréter au mieux les résultats expérimentaux. Certaines des conclusions d'ordre général auxquelles nous avons abouti auraient peut-être pu être obtenues plus directement en multipliant les mesures expérimentales. Nous n'aurions pas manqué de le faire ne serait-ce qu'à titre de contrôle de nos calculs si nous avions encore disposé des instruments physiques et en particulier des appareils photographiques et du microphotomètre indispensables. Nos nouvelles fonctions administratives ne nous permettent plus de telles expériences.

Un cas particulier, particulièrement délicat, est à l'origine de cette étude; nous nous appuyons sur cet exemple pour illustrer nos calculs théoriques. Nous pensons donc utile de donner tout d'abord quelques indications sommaires sur ce cas concret :

Au cours d'un procès l'une des parties présente dans un copie-lettres la reproduction d'une lettre adressée à son adversaire. Celui-ci reconnaît avoir reçu la lettre en question mais prétend que deux mots ajoutés à la fin n'étaient pas dans le document original et que ce dernier présentant de ce fait un intérêt plus limité n'a pas été conservé. La question posée à l'expert est donc la suivante : les deux mots litigieux ont-ils été écrits en même temps que le contexte ou ont-ils été ajoutés ultérieurement?

Pour répondre à la question, un certain nombre de prélèvements ont été effectués sur la page litigieuse (folio n° 441) au niveau des mots suspects et au niveau du contexte. D'autres prélèvements ont été effectués sur onze pages réparties dans l'ensemble du copie-lettres et en particulier sur le folio 442, copie d'une lettre datant du lendemain du document contesté. On a choisi,

dans ces pages, des fragments de texte de densité optique approximativement voisine de celle des mots suspects; la très faible densité des mots suspects a beaucoup limité le nombre des prélèvements. On verra d'ailleurs, par la suite, de cet exposé, qu'il aurait été préférable de les compléter par un certain nombre de mesures effectuées sur des textes de densité optique plus élevée. 273 mesures correspondant à 91 points ont été effectuées en tout sur les 13 documents.

TABLEAU I
Densités optiques en centièmes.

N° DU FOLIO	x	y	z	δx	δy	δz	${}_1x$	${}_1y$	δ_{1x}	δ_{1y}	
441 C	28	17	17	+ 1,9	+ 1,1	- 1,4	112	142	+ 19	- 0	
	35	20	59	+ 2,5	+ 0,2	- 1,4	157	60	+ 41	+ 7	
	21	7	36	+ 1,9	- 4,6	+ 0,5	67	9	- 4	- 31	
	16	2	30	+ 0,9	- 7,2	+ 1,9	37	2	- 22	- 25	
	38	31	54	+ 5,4	+11,2	+ 1,9	176	132	+ 70	+ 54	
	26	18	25	+ 8,7	+ 7,5	- 7,1	99	48	+ 50	+ 26	
	22	19	55	- 6,2	+ 1,9	+ 2,7	73	54	- 35	+ 5	
	24	18	54	- 4,1	+ 0,9	- 6,6	86	48	- 20	0	
	17	15	42	- 4,6	+ 1,9	+ 1,9	42	32	- 41	- 5	
	20	24	78	-15,3	+ 2,5	+ 7,4	60	86	- 84	+ 21	
	32	15	31	+11,5	+ 2,5	- 7,0	138	32	+ 77	+ 4	
	12	5	29	- 2,3	- 3,7	+ 2,5	21	6	- 36	- 20	
	12	7	23	- 0,2	- 0,4	+ 0,3	21	9	- 24	- 12	
	19	6	31	+ 2,3	- 4,1	+ 0,1	54	8	- 7	- 20	
	25	11	41	+ 2,6	- 2,6	- 0,6	92	18	+ 11	- 19	
	36	18	81	- 5,0	- 6,9	+ 4,9	164	48	+ 5	- 24	
	441 Q	31	16	38	+ 7,1	+ 3,7	- 5,0	132	37	+ 47	+ 9
		30	15	42	+ 4,9	+ 2,1	- 3,3	125	32	+ 31	+ 1
		33	26	74	- 6,9	+ 5,5	+ 2,2	144	99	- 22	+ 45
		33	19	62	- 1,3	+ 1,4	+ 0,3	144	54	+ 5	+ 9
21		7	39	0,0	- 3,8	+ 1,1	67	9	- 21	- 19	
26		11	50	- 1,0	- 2,9	+ 1,4	99	18	- 13	- 18	
32		12	66	- 2,8	- 5,9	+ 3,3	138	21	- 10	- 24	
1		27	30	40	- 1,6	+ 4,3	- 1,5	106	125	+ 5	+ 43
	29	22	59	- 5,3	- 8,9	+ 0,2	118	73	- 31	- 48	
	27	17	39	+ 1,8	- 5,7	+ 2,3	106	42	+ 8	- 38	
	27	17	23	+ 7,7	- 0,4	- 5,1	106	42	+ 43	- 5	
	26	19	12	+10,6	+ 5,1	-10,4	99	54	+ 69	+ 29	
	20	18	31	- 0,7	- 0,6	+ 0,9	60	48	- 18	- 16	
	14	30	43	-12,4	+ 6,2	+ 4,7	28	125	- 80	+ 37	
52	47	41	68	+ 1,2	+ 2,0	- 1,9	234	196	+ 49	+ 54	
	28	21	41	+ 1,2	- 1,8	+ 0,2	112	67	+ 1	- 19	
	26	20	40	+ 0,3	- 1,8	+ 0,8	99	60	- 10	- 24	
	19	20	34	- 2,7	+ 1,6	+ 0,9	54	60	- 39	- 11	
124	39	17	74	+ 1,8	- 6,7	+ 1,3	133	42	+ 8	- 45	
	39	24	93	- 6,1	- 4,8	+ 4,7	133	86	- 37	- 23	
	35	31	54	+ 4,3	+11,5	- 6,0	157	132	+ 29	+ 68	
163	45	20	59	+ 6,7	- 5,3	- 2,0	221	60	+ 73	- 6	
	20	22	25	+ 1,3	+ 9,7	- 4,8	60	73	- 1	+ 45	
	15	14	48	-10,3	- 2,7	+ 7,6	32	28	- 85	- 26	
	26	14	37	+ 2,3	- 1,6	- 0,8	99	28	+ 9	- 13	
240	12	21	53	- 0,3	+ 1,6	- 0,5	21	67	- 5	+ 12	
	21	19	50	+ 9,1	+ 0,1	- 2,1	67	54	+ 42	- 2	
	11	20	65	- 3,5	- 2,9	+ 1,8	18	60	- 14	- 8	
	9	20	43	- 1,1	+ 4,0	- 1,2	13	60	- 8	+ 15	
	7	15	51	- 4,2	- 2,8	+ 2,0	9	32	- 16	- 21	
305	41	28	89	- 1,0	+ 0,7	+ 0,3	196	112	+ 3	+ 14	
	39	18	66	+ 6,7	- 3,0	- 2,3	133	48	+ 40	- 25	
	28	16	58	+ 0,7	- 1,8	+ 0,2	112	37	- 14	- 27	
	38	25	84	- 1,4	- 0,6	+ 0,8	176	92	- 6	- 1	
	41	27	88	- 0,5	0,0	+ 0,2	196	106	+ 5	+ 9	
	33	29	80	- 4,5	+ 4,6	+ 0,7	144	118	- 29	+ 30	
367	26	19	46	+ 0,2	- 1,3	+ 0,5	99	54	- 14	- 23	
	29	26	20	+11,1	+11,9	-11,6	118	99	+ 69	+ 65	
	43	32	107	-11,3	-10,7	+11,1	208	138	- 55	- 42	
417	28	16	8	+11,7	+ 6,6	-12,4	112	37	+ 82	+ 23	
	34	12	47	- 0,5	- 7,8	+ 4,0	151	21	+ 22	- 34	
	28	16	46	- 4,7	- 2,9	+ 5,1	112	37	- 34	- 17	
	17	9	18	+ 1,6	+ 0,1	- 1,3	42	13	- 16	- 8	
	34	21	42	- 0,1	+ 1,4	- 0,6	151	67	+ 23	+ 13	
	24	21	45	- 7,9	+ 2,6	+ 5,1	86	67	- 57	+ 14	

N° DU FOLIO	x	y	z	δx	δy	δz	${}_1x$	${}_1y$	δ_1x	δ_1y	
442	35	24	76	- 1,4	+ 8,0	- 1,1	157	86	- 1	+ 37	
	42	25	97	- 3,5	+ 5,7	+ 0,5	202	92	0	+ 30	
	35	25	87	- 5,6	+ 7,8	+ 1,1	157	92	- 24	+ 36	
	34	20	84	- 4,9	+ 3,5	+ 1,6	151	60	- 24	+ 6	
	38	16	78	- 1,3	+ 1,4	+ 0,3	144	37	- 8	- 10	
	43	23	110	- 7,4	+ 1,6	+ 3,2	208	80	- 21	+ 9	
	27	7	57	+ 0,4	- 4,3	+ 0,7	106	9	- 18	- 28	
	27	1	48	+ 4,2	- 8,7	- 0,3	106	1	+ 6	- 30	
	28	2	34	+ 10,2	- 5,6	- 3,7	112	2	+ 41	- 20	
	33	0	48	+ 9,2	- 10,1	- 2,3	144	0	+ 44	- 31	
	443	17	6	68	- 0,7	- 0,5	+ 0,2	42	8	- 7	0
		16	7	65	- 0,9	+ 0,7	+ 0,2	37	9	- 10	+ 1
17		7	70	- 1,2	+ 0,3	+ 0,3	42	9	- 8	0	
23		7	77	+ 2,8	- 0,5	- 0,7	80	9	+ 25	- 1	
451	4	10	58	- 3,4	- 2,1	+ 0,9	5	15	- 11	- 12	
	10	17	76	+ 0,1	+ 0,9	- 0,2	15	42	- 5	+ 6	
	8	17	78	- 2,1	+ 0,6	+ 0,2	11	42	- 10	+ 5	
	4	15	69	- 4,9	+ 0,5	+ 0,5	5	32	- 14	- 1	
20	16	74	+ 10,2	+ 0,1	- 1,3	60	37	+ 40	+ 2		
455	18	8	59	- 1,2	- 4,6	+ 1,4	48	11	- 26	- 24	
	21	13	66	- 0,8	- 1,3	+ 0,6	67	24	- 15	- 16	
	41	24	106	+ 5,2	+ 0,4	- 1,8	196	86	+ 64	+ 22	
	36	20	95	+ 4	- 1,1	- 1,1	164	60	+ 46	+ 3	
	26	17	83	- 1,4	- 1,0	+ 0,7	99	42	- 4	- 8	
	25	15	84	- 2,4	- 3,1	+ 1,5	92	32	- 13	- 18	
	24	23	49	+ 6,2	+ 11,3	- 4,5	84	80	+ 23	+ 51	
	25	23	58	+ 4,5	+ 9,5	- 3,6	92	80	+ 20	+ 45	
	26	18	131	- 15,2	- 9,2	+ 7,1	99	48	- 64	- 31	
	21	12	57	+ 1,9	- 0,8	- 0,5	67	21	- 4	- 18	
	10	7	33	- 0,9	- 0,2	+ 0,3	15	9	- 26	- 11	

Nous appellerons x , y et z les valeurs expérimentales des densités optiques avant virage, après virage en lumière bleue et après virage en lumière rouge.

Le tableau I résume les différentes valeurs pour x , y et z dans les colonnes 2, 3, 4.

Reportons sur trois diagrammes rectangulaires les valeurs de x et y , x et z , y et z , correspondant à un même texte, par exemple x et z pour le contexte du document suspect n° 441 (fig. 1). On constate immédiatement que les points sont très dispersés autour de la droite moyenne.

Les écarts observés proviennent-ils des variations réelles dans les densités optiques ou proviennent-ils de l'insuffisance de la méthode de mesure? Avant d'essayer de discuter cette question, nous avons tout d'abord recherché si les écarts étaient répartis symétriquement autour des moyennes ou au contraire s'il y avait des variations systématiques.

Les mesures sont trop peu nombreuses dans chaque folio pour qu'un écart systématique apparaisse sur les diagrammes précédents. Mais on peut grouper les résultats de tous les feuillets sur un même diagramme en multipliant pour chaque feuillet l'une des mesures par un coefficient convenable de manière à faire coïncider toutes les droites moyennes. En multipliant les valeurs de z par un coefficient correctif convenable, c'est avec le diagramme x, z que les résultats sont les plus nets : les points du diagramme ne sont pas groupés symétriquement autour de la moyenne générale. Bien qu'un peu moins net, le résultat est le même pour les diagrammes x, y et y, z . Ceci montre déjà qu'il n'y a proportionnalité stricte entre aucune des trois variables. Il n'est guère facile d'aller plus loin en utilisant simultanément les trois diagrammes précédents parce que l'on obtient deux valeurs différentes de l'écart pour chaque mesure. Il faut

drait supprimer l'un des diagrammes, ce qui revient à dire qu'il faudrait choisir une des variables comme indépendante et étudier les deux autres en fonction de la première. Avant d'effectuer ce choix, nous avons essayé une représentation plus symétrique en étudiant la position de chaque point dans un espace à trois dimensions x, y, z .

Soit alors dans cet espace un point M de coordonnées x, y et z (ces trois quan-

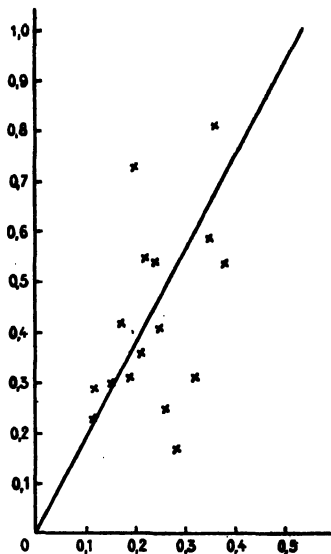


Fig. 1.

tités étant les valeurs expérimentales). Projetons M en N sur la droite moyenne du folio correspondant et soit X, Y, Z, les coordonnées du point N. Les valeurs les plus probables des écarts sont évidemment :

$$\delta x = x - X \qquad \delta y = y - Y \qquad \delta z = z - Z$$

Les différentes valeurs des écarts figurent au tableau I, colonnes 5, 6 et 7.

Les écarts ainsi calculés ne sont pas indépendants. En effet, les vecteurs x, y, z et $\delta x, \delta y, \delta z$, étant orthogonaux, on a :

$$X. \delta x + Y. \delta y + Z. \delta z = 0$$

ce qui a pour conséquence que deux des écarts sont nécessairement de même signe et le troisième de signe contraire.

Si les écarts expérimentaux avaient pour origine des écarts de mesures des trois grandeurs x, y et z , que la dispersion ait été la même pour chaque grandeur et enfin qu'il n'y ait pas eu de corrélation, la répartition des écarts de même signe aurait été symétrique et par conséquent deux quelconques d'entre elles seraient une fois sur trois de même signe et deux fois sur trois de signe contraire et les coefficients de corrélation seraient tous trois négatifs et égaux. Or, δx et δy ont le même signe dans 43 cas et ont des signes contraires dans 46. Le coefficient de corrélation moyen est $+ 0,187$; δx et δz ont le même signe dans 17 cas et ont des signes contraires dans 73. Le coefficient de corrélation moyen est de $- 0,791$. Enfin δy et δz ont le même signe dans 30 cas et ont des signes contraires dans 60 cas. Le coefficient de corrélation moyen est de $- 0,476$.

D'autre part, si l'on considère les sommes $\Sigma (X. \delta x)$, $\Sigma (Y. \delta y)$ et $\Sigma (Z. \delta z)$ leur signe donnera le sens de l'écart systématique moyen. Nous avons trouvé respectivement : — 0,2078, — 0,0456, + 0,2534.

Ainsi on voit que z se présente nettement différent de x et de y tant au point de vue écart systématique qu'au point de vue fluctuation. Il était donc naturel de prendre z comme variable indépendante pour l'étude systématique des écarts. Ce choix gardait peut-être un certain arbitraire, mais on verra plus loin que les résultats obtenus l'ont confirmé à *posteriori*.

Écarts systématiques.

Nous avons d'abord supposé que x et y étaient des fonctions linéaires de z avec ordonnées à l'origine constante autrement dit que l'on avait :

$$x = c_i z + a \tag{1}$$

$$y = d_i z + b \tag{2}$$

les coefficients c_i et d_i variant avec chaque folio tandis que a et b sont les mêmes pour tous les folios. Pour obtenir les valeurs les plus probables de a et des c_i il suffit d'écrire que l'expression :

$$U = \Sigma \Sigma_i (x - c_i z - a)^2 \tag{3}$$

est minimum, la sommation d'indice i correspondant à la somme de tous les termes d'un même folio i tandis que celle sans indice correspond à l'ensemble de tous les folios. Dérivant par rapport à a puis par rapport à chaque c_i et éliminant ces derniers, on trouve facilement en appelant n_i le nombre de points correspondant au folio i :

$$a = \frac{\Sigma \Sigma_i \left(1 - \frac{z_i \Sigma_i z}{\Sigma_i z^2} \right)^2}{\Sigma \left(n_i - \frac{\Sigma_i^2 z}{\Sigma_i z^2} \right)} \tag{4}$$

On calculera ensuite chaque valeur de c_i et en se reportant dans l'équation (1) on obtiendra finalement la valeur de U .

Les x étant des valeurs aléatoires, a se présente comme une somme de valeurs aléatoires. Supposons que l'écart quadratique sur x soit une constante égale à ϵ (on trouvera plus loin la justification de cette hypothèse qui n'est du reste qu'approchée), on peut facilement calculer une valeur de l'écart quadratique de l'erreur α que l'on risque de commettre en calculant a par la formule (4). On trouvera facilement :

$$\alpha^2 = \frac{\epsilon^2}{\Sigma \left(n_i - \frac{\Sigma_i^2 z}{\Sigma_i z^2} \right)^2}$$

D'autre part, on peut avoir une estimation de ϵ à partir de U . En effet, on peut poser :

$$\epsilon^2 = \frac{U}{76}$$

car on dispose de 91 prélèvements mais 15 valeurs des constantes sont fixées par les équations 1 : la constante a et 14 valeur de c_i il ne reste donc que 76 degrés de liberté.

On trouve finalement :

$$\begin{array}{ll} a = + 15,5 & \alpha = 2,1 \\ b = + 6,7 & \beta = 2,0 \end{array}$$

Or, ce résultat est physiquement inacceptable. Il est évident, en effet, que la densité optique du trait d'encre primitif peut prendre toutes valeurs positives et pour que x soit inférieur à 15,5 il faudrait que z soit négatif, ce qui est absurde.

L'hypothèse qui consiste à supposer des relations linéaires entre x et z d'une part, y et z d'autre part et des constantes uniques à l'origine a et b est donc insuffisante.

On se heurterait à la même difficulté en supposant que toutes les droites se coupent en un même point de l'axe des z car ce point serait aussi négatif.

Nous avons alors supposé que les courbes $x = f_i(z)$ et $y = g_i(z)$ se déduisaient les unes des autres par une simple homothécie de la variable z autrement dit que l'on pouvait trouver pour chaque folio deux coefficients c_i et d_i tels que l'on ait $x = f(c_i z)$ et $y = g(d_i z)$ les fonctions f et g étant identiques pour tous les folios. Calculons alors pour chaque folio la droite moyenne correspondante; on peut admettre en première approximation qu'elles se confondent avec les tangentes au centre de gravité des courbes f et g . En raison de l'incertitude des coefficients c_i et d_i les pentes de ces droites sont sans intérêt, mais leurs ordonnées à l'origine ne dépendent que des fonctions f et g .

Cette dernière remarque permet de comparer entre elles les deux fonctions inconnues f et g . En effet, classons les folios par valeurs croissantes de \bar{x} et éliminons les folios donnant les moyennes les plus élevées et faisons de même pour \bar{y} en éliminant les folios donnant les moyennes les plus basses. Lorsque les moyennes générales de \bar{x} et de \bar{y} sont à peu près égales, calculons comme précédemment une valeur unique a et une valeur b avec les groupes de folios ainsi conservés.

Nous obtenons les valeurs du tableau II.

TABLEAU II

	\bar{x}	\bar{y}
Nombre de mesures	34	32
Moyenne générale	22,2	21,0
Ordonnée à l'origine	10,31	12,98
Écart quadratique	3,9	3,0

La comparaison des valeurs des ordonnées à l'origine ainsi obtenues avec les valeurs précédentes montre que ces quantités varient avec la moyenne de x et de y et en outre que rien ne s'oppose à ce que les fonctions f et g soient identiques.

Nous supposons donc qu'elles le sont effectivement et pour la recherche plus précise de cette fonction, nous confondrons x et y sous le même vocable t .

On doit donc avoir :

$$t = f(c_i, z)$$

ou

$$t = f(d_i, z)$$

Prenons alors suivant les cas :

$$c_i z = u$$

ou

$$d_i z = u$$

On aura finalement la fonction unique :

$$t = f(u)$$

Classons maintenant les folios suivant les valeurs croissantes des moyennes de t puis partageons les en quatre classes; dans chaque classe calculons la moyenne générale et toujours comme précédemment l'ordonnée à l'origine ω de la tangente. A partir des quatre points ainsi obtenus, on peut tracer très approximativement une courbe $\omega = h(t)$.

On a :

$$\frac{dt}{du} = \frac{t - \omega}{u}$$

$$\frac{dt}{t - h(t)} = \frac{du}{u}$$

L'intégration graphique donne ainsi une première relation approximative entre u et t , c'est-à-dire entre t et z à des constantes multiplicatrices près.

Au moyen de la courbe précédente, calculons pour chaque valeur de $t(x$ ou $y)$ la valeur correspondante.

Dans chaque feuillet calculons :

$$c_i = \frac{\bar{u}}{\bar{z}} \quad \text{ou} \quad d_i = \frac{\bar{u}}{\bar{z}}$$

Prenons alors :

$${}_1u = c_i z \quad \text{ou} \quad {}_1u = d_i z$$

Dressons le tableau de correspondance de chaque valeur de t en fonction de ${}_1u$, ces dernières valeurs étant classées par valeur croissante. Découpons 9 classes de 20 valeurs environ dans ce tableau puis calculons dans chaque classe \bar{t} et \bar{u} et traçons la courbe passant au mieux par les 9 points correspondants. On obtient ainsi une deuxième approximation de $t = {}_1f({}_1u)$.

Recommençons une troisième puis une quatrième fois l'opération : nous aurons une troisième puis une quatrième approximation. Ces deux dernières courbes étant à peu près confondues (Voir fig. 2), on peut admettre définitivement les résultats obtenus par ces deux dernières approximations.

Le tableau III donne colonnes 1 et 2 les valeurs des moyennes définitives des 9 classes.

TABLEAU III

u	i		
	obtenu	calculé	
		formule 5	formule 6
6,7	6,8	4,6	5,5
11,5	8,0	7,3	8,8
15,9	10,2	9,7	10,8
19,7	12,4	11,6	11,7
24,7	13,8	13,2	13,2
32,3	14,0	15,6	15,1
41,3	16,4	18,2	16,8
54,3	18,6	20,8	19,0
79,4	23,8	22,0	23,0

INTERPRÉTATION DES RÉSULTATS

Nous avons pensé que la non proportionnalité entre la quantité d'encre totale ou insoluble dans l'acide chlorhydrique et la densité optique s'expliquait par la répartition non uniforme du pigment dans le papier.

Si l'on examine, en effet, un trait d'encre au campêche au microscope, on

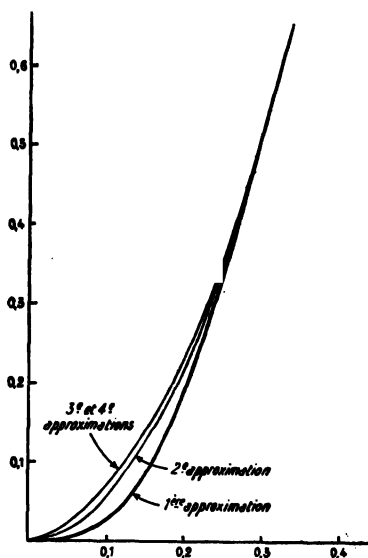


Fig. 2.

constate que le pigment est réparti uniquement sur les fibres végétales du papier tandis que le trait de bleu de prusse est homogène, ce qui est dû à ce que le fer diffuse uniformément dans les fibres et l'encollage.

Supposons d'abord que seules les fibres soient teintes à l'exclusion du reste, soit s la fraction occupée par les fibres en calculant l'opacité du trait d'encre on voit facilement que l'on doit avoir :

$$10^{-t} = s 10^{-u} + 1 - s \tag{5}$$

Cherchons la valeur de s qui rende minima le carré des écarts entre les valeurs calculées par cette formule et les valeurs expérimentales moyennes du tableau III.

On trouve par approximations successives $s = 0,575$, mais les valeurs obtenues (tableau III, 3^e colonne et la figure 4) montrent que manifestement la courbe théorique s'écarte notablement des points expérimentaux.

Dans une deuxième approximation, nous avons supposé que le colorant se fixe à la fois sur les fibres et sur l'encollage, mais inégalement. Soit p le rapport des taux de fixation par unité de surface. La formule 5 devient alors :

$$10^{-t} = s 10^{-\frac{p}{p+1-s} u} + (1-s) 10^{-\frac{1}{p+1-s} u} \quad (6)$$

On trouve cette fois par approximations successives :

$$s = 0,22 \text{ et } p = 26$$

Cette fois-ci, les valeurs ainsi calculées (tableau III, 4^e colonne et la fig. 3) montrent que cette formule exprime très convenablement les résultats expérimentaux. La somme des carrés des écarts est du reste 3,8 fois plus petite que dans le cas précédent. Or, le nombre de degrés de liberté étant réduit de 8 à 7,

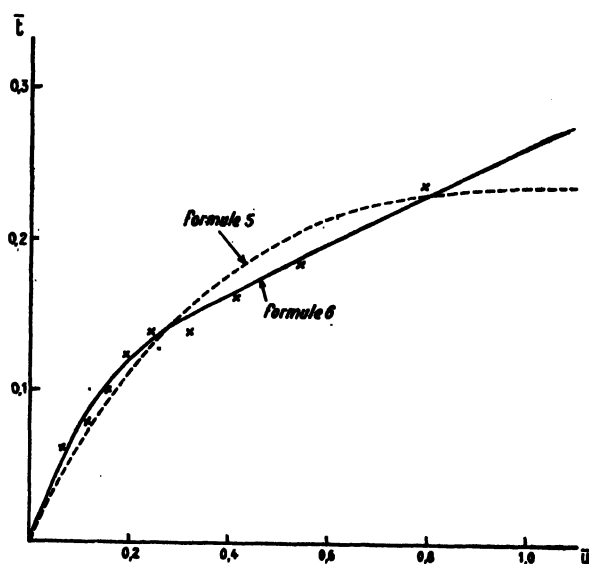


Fig. 3

ce rapport n'aurait été que de 1,14 si les deux formules théoriques avaient été équivalentes.

D'autre part la valeur 0,22 correspond d'une manière très satisfaisante à l'aspect microscopique du trait d'encre. L'explication proposée est donc tout à fait acceptable.

Appliquant alors la formule 6 avec les valeurs numériques 7, nous avons entièrement recalculé les valeurs de x et de y . Soit ${}_1x$ et ${}_1y$ les valeurs ainsi obtenues (8^e et 9^e colonnes du tableau I).

Fluctuations. — ${}_1x$ et ${}_1y$ sont proportionnels à z aux fluctuations près. Soit ${}_1c_i$ et ${}_1d_i$ les facteurs moyens de proportionnalités :

$${}_1c_i = \frac{\overline{{}_1x_i}}{z_i} \quad {}_1d_i = \frac{\overline{{}_1y_i}}{z_i}$$

Les fluctuations sont alors :

$$\begin{aligned} \delta_1x &= {}_1x - {}_1c_i z \\ \delta_1y &= {}_1y - {}_1d_i z \end{aligned}$$

(Les valeurs de δ_1x et δ_1y figurent au tableau I, 10^e et 11^e colonnes.)

Nous avons espéré que la théorie précédente permettrait d'expliquer l'importance des écarts. En effet, le petit rectangle éclairé par l'image du filament de la lampe du microphotomètre enregistreur a des dimensions de même ordre de grandeur que les fibres végétales teintes du papier. Il en résulte que, suivant le hasard des circonstances, s peut varier dans de larges limites et l'application de la formule permet de calculer les variations de s en fonction de celles de la densité. Malheureusement les résultats ainsi obtenus sont invraisemblables car on trouve que s pourrait varier entre — 4,78 et 0,93; or, une valeur négative de s n'a évidemment aucun sens.

La variation de s qui existe certainement, ne peut donc expliquer que partiellement les fluctuations des rapports densité encre/densité fer et densité insoluble/densité fer. D'autres causes interviennent également : soit une variation du facteur p , soit une inégale répartition du fer, soit des causes d'erreurs insoupçonnées dans la mesure des densités.

Pour poursuivre l'étude de la comparaison des encres, il importait cependant de connaître les lois relatives à la répartition des fluctuations. Dans ce but, nous avons recherché :

1^o Si les fluctuations observées sur ${}_1x$ et sur ${}_1y$ étaient indépendantes. Il n'en est rien et l'on trouve un coefficient de corrélation positif $\gamma = + 0,38$.

2^o Si les fluctuations étaient variables avec $\overline{{}_1x}$ et $\overline{{}_1y}$ ou au contraire constantes.

Le nombre des mesures effectuées est trop faible pour répondre, d'une manière précise, à cette deuxième question. Il existe une faible corrélation positive entre ${}_1x$ et $|\delta_1x|$ d'une part, ${}_1y$ et $|\delta_1y|$ d'autre part. La différence entre les deux valeurs ainsi obtenues n'est pas significative. Cela revient à dire que l'on ne peut affirmer que la relation entre ${}_1x$ et $|\delta_1x|$ est différente de celle qui relie ${}_1y$ et $|\delta_1y|$. On ne peut d'ailleurs pas connaître cette relation. Si l'on suppose toutefois qu'elle est bien la même et qu'elle est linéaire, on trouve en appelant ${}_1t$ comme précédemment l'ensemble des valeurs ${}_1x$ et ${}_1y$ indifféremment.

$$\overline{|\delta_1t|} = 0,20 + 0,052 {}_1t$$

d'où approximativement :

$$\overline{\delta_1^2 t} = 0,063 + 0,045 {}_1t \quad (7)$$

Nous sommes maintenant en mesure de comparer les moyennes entre les différents éléments constituant le copie-lettres en question.

COMPARAISON DES DIFFÉRENTS TEXTES

Chaque folio peut être caractérisé par les rapports $\frac{1x}{z}$ et $\frac{1y}{z}$ ou mieux par les valeurs moyennes $\frac{\sum_1 x}{\sum z}$ et $\frac{\sum_1 y}{\sum z}$.

Mais cette manière de voir présente un inconvénient sérieux. Il est possible, en effet, qu'un texte ne renferme que très peu de fer; z sera alors très petit devant $1x$, ou $1y$, et les rapports très grands fausseront complètement les moyennes générales. Nous avons alors choisi comme variables pour caractériser un texte les angles ξ et η définis par les équations :

$$\text{tg } \xi = \frac{\sum_1 x}{\sum z} \quad \text{et} \quad \text{tg } \eta = \frac{\sum_1 y}{\sum z}.$$

Erreurs sur ξ et η .

En remarquant que le carré de l'erreur quadratique moyenne sur $\sum_1 x$ est égal à la somme des carrés des erreurs quadratiques sur chaque valeur de $1x$, l'équation 7 donne approximativement :

$$\overline{\delta^2 \sum_1 x} = 0,063 n_i + 0,0045 \sum_1 x$$

de même

$$\overline{\delta^2 \sum_1 y} = 0,063 n_i + 0,045 \sum_1 y$$

Or, si l'on appelle σ_ξ l'erreur quadratique sur ξ , on a :

$$\frac{\sigma_\xi}{\cos^2 \xi} = \frac{\sqrt{\overline{\delta^2 \sum_1 x}}}{\sum z}$$

d'où

$$\sigma_\xi^2 = \frac{(0,063 n_i + 0,045 \sum_1 x) \sum_1 z}{(\sum_1 x + \sum^2 z)}$$

de même

$$\sigma_\eta^2 = \frac{(0,063 n_i + 0,045 \sum_1 y) \sum_1 z}{(\sum_1 x + \sum^2 z)^2}$$

Le tableau V résume les valeurs obtenues pour ξ , η , σ_ξ et σ_η pour les folios analysés du registre. On a distingué dans le folio en litige, d'une part, la mention suspecte 441 Q, d'autre part, le contexte 441 C et enfin l'ensemble des deux textes 441 T supposé provenir d'une origine unique.

TABLEAU V

N° DU FOLIO	ξ	η	σ_ξ^2	σ_η^2
441 Q	1,1006	0,7428	0,001865	0,007579
441 C	1,1514	0,6238	0,001630	0,017688
441 T	1,1286	0,7165	0,0007598	0,005880
1	1,1927	1,1175	0,002196	0,004041
52	1,2186	1,0032	0,002012	0,009508
124	1,1709	0,8664	0,001998	0,01102
168	1,1807	0,8412	0,003200	0,02329
240	0,4574	0,8060	0,003520	0,01467
305	1,1386	0,8844	0,001184	0,005730
367	1,1842	1,0845	0,002565	0,007291
417	1,2147	0,8636	0,002180	0,02041
442	1,1281	0,5723	0,0007105	0,008199
443	0,6227	0,9050	0,018998	0,03314
451	0,2641	0,0506	0,02467	0,02550
485	0,8140	0,6074	0,002860	0,007333
Moyennes	0,9777	0,7860		

RÉPARTITION STATISTIQUE DES VALEURS ξ ET η .

Nous ne connaissons pas les valeurs vraies de ξ et de η pour chaque folio mais seulement des valeurs approchées.

Les écarts ne modifient pas systématiquement les moyennes $\bar{\xi}$ et $\bar{\eta}$. Par contre, elles introduisent des erreurs sur la dispersion des valeurs par rapport aux moyennes.

Appelons Θ_{ξ} et Θ_{η} la racine carré de la valeur moyenne vraie de $(\xi - \bar{\xi})^2$ et $(\eta - \bar{\eta})^2$. On voit facilement que les valeurs les plus probables de Θ_{ξ} et de Θ_{η} sont :

$$\Theta_{\xi}^2 = \frac{\Sigma (\xi - \bar{\xi})^2 - \Sigma \sigma_{\xi}^2}{n - 1}$$

$$\Theta_{\eta}^2 = \frac{\Sigma (\eta - \bar{\eta})^2 - \Sigma \sigma_{\eta}^2}{n - 1}$$

D'autre part, la valeur la plus probable de $(\xi - \bar{\xi})(\eta - \bar{\eta})$ est :

$$\frac{\Sigma (\xi - \bar{\xi})(\eta - \bar{\eta}) - \gamma \Sigma \sigma_{\xi} \sigma_{\eta}}{n - 2}$$

d'où l'on tire facilement la valeur du coefficient de corrélation ρ entre ξ et η . Finalement on trouve :

$$\begin{aligned} \Theta_{\xi} &= 0,3185 \\ \Theta_{\eta} &= 0,2414 \\ \rho &= 0,7373 \end{aligned} \quad (8)$$

LES ÉCARTS DES VALEURS ENTRE ξ ET η POUR LE TRAIT SUSPECT ET LE CONTEXTE DU FOLIO 441 SONT-ILS CARACTÉRISTIQUES?

On sait que la loi de répartition des écarts se rapproche d'autant plus d'une loi de Laplace-Gauss que l'on opère sur des moyennes provenant d'un nombre plus élevé de mesures. Nous pouvons donc admettre sans erreur importante qu'il en est ainsi dans le cas présent quelle que soit la loi de répartition réelle des écarts pour chaque mesure isolée. Dans ces conditions, la loi de répartition des différences entre deux séries de mesures est encore une loi de Laplace-Gauss. Le carré de l'écart quadratique sur ${}_1x$ ou ${}_1y$ est la somme des carrés des écarts quadratiques pour chaque série et le coefficient de corrélation reste le même. Les valeurs calculées dans ces conditions sont portées à la troisième ligne du tableau V. Ajoutons que l'on a :

$$\lambda = \xi_Q - \xi_C = 0,0508 \quad \text{et} \quad \mu = \eta_Q - \eta_C = -0,1190$$

D'autre part, les probabilités élémentaires de deux valeurs particulières de λ et μ sont données par la formule de Laplace-Gauss à deux variables :

$$\frac{1}{2\pi \sigma_{\xi} \sigma_{\mu} \sqrt{1 - \gamma^2}} e^{-\frac{1}{2(1 - \gamma^2)} \left(\frac{\lambda^2}{\sigma_{\xi}^2} - \frac{2\lambda\mu\gamma}{\sigma_{\xi}\sigma_{\eta}} + \frac{\mu^2}{\sigma_{\mu}^2} \right)} d\lambda d\mu$$

qui devient dans le cas particulier :

$$21,38 e^{-1,1662} d\lambda d\mu$$

On a $e^{-1.1662} = 0,3115$; si λ et μ étaient nuls, l'exponentielle serait égale à 1. La probabilité trouvée est donc 0,3115 fois la valeur maximum de la probabilité également élémentaire. On peut encore dire en remarquant que la valeur moyenne de cette probabilité est $d\lambda d\mu/2 \pi \sigma_\xi \sigma_\eta \sqrt{1 - \gamma^2}$, que la valeur trouvée est 0,62 fois la valeur moyenne des écarts. La différence observée n'est donc pas caractéristique d'une différence d'origine. Rien ne s'oppose donc à ce que le tracé suspect ait été écrit en même temps que le contexte.

Peut-on aller plus loin et affirmer que les deux tracés ont bien été écrits en même temps? Pour répondre à cette question, nous allons comparer les probabilités calculées dans les deux hypothèses suivantes :

1° Il s'agit d'un même tracé;

2° Il s'agit de deux tracés n'ayant aucun rapport.

Pour cela il était nécessaire de connaître la probabilité d'un texte ayant des coordonnées ξ et η données ce qui supposait connue la loi de répartition des différentes valeurs de ξ et de η . Ce n'est évidemment pas les treize feuillets étudiés qui permettent de trouver cette loi et nous en serons réduits à cet égard à une hypothèse. Supposons donc qu'il s'agisse encore d'une loi de Laplace-Gauss à deux variables. Les valeurs trouvées en (8) permettent alors de trouver la probabilité élémentaire pour chaque couple.

$$dP_{\xi\eta} = 3,071 e^{-10,80 (\xi - \bar{\xi})^2 - 21,06 (\xi - \bar{\xi})(\eta - \bar{\eta}) + 18,88 (\eta - \bar{\eta})^2} d\xi d\eta$$

Si l'on admet la première hypothèse, la probabilité totale sera le produit de la probabilité de l'écart par la probabilité du folio 441. Pour calculer la probabilité du feuillet 441 nous avons pris les valeurs correspondant à 441 T.

Dans la deuxième hypothèse la probabilité totale est le produit des probabilités pour chaque texte. Le rapport des probabilités totales pour les deux hypothèses est finalement :

$$\frac{dP_1}{dP_2} = 1,61.$$

Autrement dit l'hypothèse de l'identité des encres n'a que 1,61 fois plus de chance de conduire à l'écart observé que l'hypothèse de deux textes écrits au hasard. On ne peut donc pas dire que les calculs justifient l'identité d'origine des deux textes.

Les analyses et les calculs effectués peuvent se résumer ainsi : rien ne s'oppose à l'identité d'origine des deux textes incriminés, mais rien non plus ne l'a justifiée. Une telle conclusion peut évidemment paraître décevante. Il arrive malheureusement quelquefois qu'une expertise judiciaire n'aboutisse pas à une précision plus grande. Heureusement qu'il n'en est pas toujours ainsi. On conçoit que les mêmes calculs appliqués à d'autres exemples numériques aient pu conduire au contraire à une précision très grande et que l'une des deux éventualités, origine identique et origine différente, ait pu être pratiquement éliminée avec une grande probabilité. Les calculs précédents, facilement transposables dans chaque cas particulier, permettent de connaître alors avec une bonne précision la valeur exacte d'une telle probabilité.

Laboratoire municipal (Paris).

L. AMY.