

ÉCHANTILLONNAGE AU HASARD DES DÉBITS MOYENS MENSUELS*

MARCEL ROCHE

Ingénieur en Chef à Électricité de France

*Chef du Département Recherche Fondamentale au Service Hydrologique de l'ORSTOM
Professeur à l'ENGREF*

RÉSUMÉ

Après un bref aperçu sur la méthodologie de l'échantillonnage fictif et sur ses domaines d'application, on examine en détail le problème de l'échantillonnage au hasard des débits moyens mensuels. La difficulté posée par la série chronologique à variations saisonnières cycliques est tournée en considérant deux composantes : l'abondance annuelle et sa répartition dans l'année. Tandis que l'abondance est représentée par le module, la répartition peut se traduire par la suite des coefficients mensuels. Le module est traité par les méthodes habituelles de la statistique des variables aléatoires : loi de distribution, estimation des paramètres de la loi. Pour la répartition, on constitue un réservoir ou ensemble de fonctions aléatoires dans lequel la forme à appliquer au module est tirée au sort. On examine ensuite les corrélations qui peuvent exister entre certains indices caractéristiques de la forme et le module. L'existence de ces corrélations conduit à établir plusieurs réservoirs de formes, suivant une technique qui est indiquée, dans lesquels on doit puiser suivant la valeur du module préalablement tiré au sort.

SUMMARY

After a short look onto the methodology of the random sampling and its fields of application, the problem of the random sampling of the average monthly discharges is examined in detail. Difficulties arose from chronological sequence having periodic seasonal variations are turned over by considering two components: annual yield and its distribution during the year. While yield is given as the yearly mean discharge, the distribution may be considered through the annual sequence of the monthly coefficients. The yearly mean discharge is operated by the usual methods of the statistic of the variates: distribution law, estimation of the coefficients of the law. As for the distribution along the year, a "storage" or "set" of aleatory functions is constituted, from what the "shape" to be used for the yearly mean discharge, is drawn at random. Then, the correlations which may exist between certain shape indexes and the yearly mean discharge, are examined. If there exist such correlations, taking them into account leads to constitute several sets of shapes, according to a method which is explained in the paper, in what sets shapes are to be drawn, according to the value which has been previously drawn at random.

1. RAPPEL SUR LA CONSTITUTION AU HASARD D'UN ÉCHANTILLON DE VARIABLES INDÉPENDANTES

Soit une variable X dont on connaît la loi de répartition statistique $F(x)$, x étant une valeur numérique de X . Cette loi est définie par un certain nombre de paramètres estimés d'après un échantillon de valeurs observées de X . On peut aussi, soit analytiquement, soit empiriquement, déterminer la variabilité d'échantillonnage de ces paramètres et estimer les intervalles de confiance en résultant.

* Communication présentée au Colloque de l'AIHS à Fort Collins, septembre 1967.

Supposons qu'on dispose d'une collection de nombres « y » compris entre 0 et 1, tirés « au hasard » par un procédé répondant à l'idée qu'on se fait du terme hasard, de telle manière que les épreuves successives du tirage soient parfaitement indépendantes les unes des autres et que la densité de probabilité soit indépendante de la valeur du nombre (distribution rectangulaire ou équirépartition), ce qu'on vérifie par l'application de tests adéquats. Dans ces conditions, la fréquence de dépassement d'un nombre quelconque pris dans la table a même valeur que ce nombre : $G(y) = y$ (fig. 1).

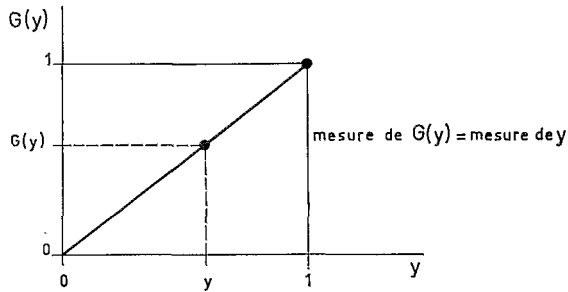


Fig. 1

Si l'on a lu sur la table de nombres au hasard, puis dressé la liste d'un certain nombre de valeurs y , auxquelles correspondent des valeurs $G(y)$, on a réalisé un échantillon de fréquences dont on peut attribuer les valeurs à un échantillon imaginaire de X . Autrement dit, pour chaque valeur de y lue dans la table, on fait :

$$F(x) = G(y) = y.$$

L'inversion $x = H(y)$ donne la valeur de x correspondante. Autrement dit, un tirage de y dans une table de nombres au hasard équirépartis, suivi de l'inversion $x = H(y)$, équivaut à un tirage au hasard de x dans la loi de probabilité $F(x)$.

Exemple

$F(x)$ est une loi normale définie par un écart-type $\sigma_x = 235$ et une moyenne $\bar{x} = 768$. On tire, dans une table de nombres au hasard $y = 3\,556$, ce qui peut se traduire par une fréquence de 0,3556. Si on convient qu'il s'agit d'une fréquence de dépassement, elle correspond dans la loi de Gauss à une variable réduite $u = 0,37$, d'où $x = 768 + 0,37 \cdot 235 = 855$. En continuant à lire des nombres à la suite sur la table, on peut ainsi constituer un échantillon de valeurs de x dans la distribution $F(x)$.

Lorsque les calculs sont effectués sur ordinateur, les nombres au hasard sont générés au fur et à mesure des besoins par une méthode arithmétique.

Un échantillon fictif, aussi grand soit-il, n'améliore absolument pas la connaissance que l'on a par ailleurs des paramètres de la loi de distribution. Ces paramètres sont estimés à partir de l'échantillon des valeurs observées et la précision avec laquelle ils sont connus dépend essentiellement de leur variabilité propre, de la variabilité de la variate d'origine et de la taille de l'échantillon.

À quoi peut donc servir l'échantillonnage fictif ? On peut lui assigner en gros trois champs d'application.

1.1 Détermination empirique de la variabilité d'un paramètre d'estimation ou d'un quantile de la variable d'origine; calcul des intervalles de confiance

On sait qu'un paramètre d'une loi statistique est déterminé à partir de l'échantillon des valeurs observées avec une certaine chance d'avoir une erreur d'une certaine valeur. C'est ainsi que la moyenne \bar{x} d'une loi de Gauss est estimée à partir d'un échantillon de taille N avec une variance d'estimation égale à σ_x/\sqrt{N} . Dans beaucoup de cas, cette variabilité peut être déterminée analytiquement au moins de façon approchée, comme c'est le cas de \bar{x} et de la plupart des moments de la plupart des lois, tout au moins si N est assez grand. Mais il peut se faire que l'estimation de la variabilité soit très laborieuse, ou trop imprécise s'il s'agit d'une évaluation approchée, pour que la méthode directe soit applicable. La technique est alors de tirer au hasard un grand nombre d'échantillons de taille N dans la loi d'origine, d'en déduire autant d'estimations du paramètre étudié et d'étudier empiriquement la loi de distribution de ces estimations.

S'il s'agit d'étudier la dispersion de l'estimation d'un quantile, le problème est sensiblement le même. La seule différence, c'est que l'on doit réestimer les valeurs de tous les paramètres de la loi pour chaque échantillon tiré au hasard, et calculer à chaque fois le quantile correspondant.

1.2 Application d'une chronique fictive à un modèle déterministe

C'est le cas par exemple de la simulation du fonctionnement d'un ouvrage hydraulique pour lequel on a adopté des consignes d'exploitation fixes ou dépendant uniquement des apports (et non pas de la consommation).

Le rôle de l'échantillon fictif, pris alors de très longue durée, est de fournir une image du plus grand nombre de cas possibles, par une transformation qu'il ne serait pas toujours aisé d'opérer directement. Il permet en outre, lorsque le processus engendré par la simulation est évolutif mais converge vers la stabilité (ergodicité), d'assurer une durée de simulation suffisante pour atteindre cette stabilité et travailler en processus stable pendant une durée suffisante : par exemple, les aménagements hydrauliques avec régularisation interannuelle.

1.3 Application d'une chronique fictive ou d'un échantillon fictif à un modèle stochastique; variables aléatoires résultant d'une composition de plusieurs autres variables aléatoires

Le problème est le suivant : on connaît la loi de probabilité des variates X_1, X_2, \dots ; on veut déterminer la loi de probabilité d'une variate résultante $Y = f(X_1, X_2, \dots)$. La détermination directe de la loi de probabilité de Y , c'est-à-dire sa forme et la valeur de ses paramètres, à partir des lois de probabilité de X_1, X_2, \dots définies par leur forme analytique et la valeur de leurs paramètres, est pratiquement impossible ou à tout le moins pénible dans les cas les plus simples.

Si les variates X_1, X_2, \dots sont indépendantes, la construction de l'échantillon fictif est simple : il suffit d'effectuer un tirage dans chacune des lois des X , puis de calculer la valeur de Y correspondante, de tirer à nouveau des valeurs des X etc.

Si les variates X sont liées, le problème est plus compliqué et ne peut être résolu que si l'on connaît les lois conditionnelles de X_2 , sachant que X_1 , de X_3 , sachant que X_1 et X_2 , etc.

2. CONSTRUCTION D'UN ÉCHANTILLON FICTIF DE DÉBITS MENSUELS

On a choisi, comme exemple d'échantillon naturel de débits moyens mensuels, les débits du Sénégal à Bakel qui figurent sur le tableau I. Un tel échantillon peut être considéré de plusieurs façons :

— série chronologique d'une variate unique X , comportant, outre une composante aléatoire, une composante cyclique qui tient compte de la variation saisonnière et telle que deux valeurs successives soient liées : $x(i)$ à $x(i-1)$, elle-même liée à $x(i-2)$ etc.;

TABLEAU I

*Le Sénégal à Bakel*Débits moyens mensuels (en m³/s)

Année	M	J	J	A	S	O	N	D	J	F	M	A	Module
1903-04	(10)	(120)	637	1762	2535	934	(380)	(142)	(84)	(53)	(28)	(10)	(559)
1904-05	(5)	(11)	578	2408	2933	986	481	208	121	80	(47)	(20)	(659)
1905-06	(10)	177	807	2503	2090	2180	951	292	145	92	(54)	(23)	(782)
1906-07	21	75	872	4700	3551	1315	603	(310)	(170)	(100)	(55)	(25)	(990)
1907-08	(10)	(120)	(403)	905	2194	1282	613	(340)	(185)	(110)	(62)	(28)	(521)
1908-09	(10)	81	799	2195	3691	1395	(500)	(235)	(130)	(75)	(42)	(18)	(767)
1909-10	(10)	(286)	949	2967	4144	1296	590	(255)	(140)	(83)	(46)	(20)	(902)
1910-11	(10)	(120)	(590)	2134	3004	1221	472	(215)	(120)	(70)	(38)	(16)	(670)
1911-12	(10)	(120)	(590)	1455	2439	930	431	(220)	(125)	(72)	(38)	(16)	(537)
1912-13	(10)	(120)	(590)	1425	2348	1305	(436)	(230)	(135)	(78)	(43)	(18)	(564)
1913-14	(10)	(120)	(497)	902	1131	876	(372)	(195)	(112)	(65)	(36)	(14)	(363)
1914-15	(10)	(120)	(590)	(1323)	1423	1035	(360)	(200)	(115)	(70)	(40)	(16)	(444)
1915-16	(10)	(90)	636	1896	2442	1261	(350)	(190)	(105)	(62)	(34)	(12)	(592)
1916-17	(5)	(4)	726	1782	3223	1664	(400)	(210)	(120)	(70)	(38)	(16)	(691)
1917-18	(10)	(20)	(250)	(2012)	3219	1132	(330)	(185)	(100)	(58)	(32)	(11)	(615)
1918-19	(10)	(160)	687	3072	4628	2307	608	(300)	(175)	(100)	(55)	(25)	(1014)
1919-20	(10)	(140)	(404)	(1704)	2261	1026	(356)	(210)	(115)	(70)	(38)	(15)	(530)
1920-21	(10)	(120)	(540)	2535	4252	1311	596	(290)	(160)	(95)	(52)	(23)	(834)
1921-22	(10)	(120)	(396)	(1201)	(2100)	736	(270)	(150)	(90)	(50)	(26)	(10)	(431)
1922-23	(10)	(40)	402	3213	6746	2778	778	316	158	(95)	(53)	(23)	(1219)
1923-24	(10)	(90)	628	(1808)	3764	1463	741	272	138	(80)	(44)	(19)	(754)
1924-25	(10)	(171)	1469	(4172)	(5602)	2597	806	384	(210)	(125)	(70)	(32)	(1309)
1925-26	(14)	101	397	2280	3275	2506	765	(325)	(185)	(110)	(65)	(30)	(841)
1926-27	(10)	(140)	(507)	(1607)	1741	(973)	715	270	(130)	(76)	(43)	(18)	(521)
1927-28	(10)	(120)	(777)	2800	(4745)	(2743)	878	(380)	(205)	(120)	(70)	(32)	(1075)
1928-29	(10)	(100)	458	3296	5046	1872	825	(332)	(190)	(110)	(65)	(27)	(1030)
1929-30	(10)	(350)	992	3244	4857	1510	543	(260)	(145)	(85)	(45)	(20)	(1008)
1930-31	(10)	(170)	(649)	(2621)	(3412)	(1929)	605	(290)	(167)	(97)	(58)	(25)	(839)
1931-32	(10)	(170)	940	1755	2715	2119	550	(270)	(155)	(90)	(50)	(22)	(739)
1932-33	(10)	(130)	(780)	2780	3181	1369	(445)	(227)	(130)	(75)	(43)	(19)	(770)
1933-34	(10)	(153)	1087	3302	3571	1066	(386)	(200)	(115)	(68)	(35)	(12)	(838)
1934-35	(10)	(20)	270	2339	3496	1315	(440)	(220)	(130)	(75)	(43)	(19)	(700)
1935-36	(10)	(120)	(896)	(4269)	(4971)	2487	(630)	(265)	(152)	(88)	(50)	(20)	(1166)
1936-37	(10)	(85)	599	4593	(5825)	2261	707	334	172	105	62	(25)	(1236)
1937-38	(10)	(120)	397	1748	3108	1339	504	(230)	(130)	(75)	(41)	(17)	(645)
1938-39	(10)	(120)	479	1826	3995	1870	800	(270)	(150)	(88)	(48)	(21)	(808)
1939-40	(2,3)	(28)	362	1935	2089	1377	(435)	(220)	(125)	(72)	(40)	(16)	(560)
1940-41	(10)	(50)	210	1316	1343	1254	(529)	(200)	(120)	(68)	(38)	(15)	(432)
1941-42	(10)	(120)	339	1158	2115	740	(247)	(130)	(75)	(44)	(19)	(6,5)	(418)
1942-43	(10)	(120)	385	1896	1715	539	(266)	(140)	(80)	(45)	(25)	(8)	(438)
1943-44	(10)	(120)	366	1867	2951	1801	443	(195)	(110)	(65)	(35)	(14)	(666)
1944-45	(10)	(120)	225	814	1444	663	(339)	(160)	(95)	(55)	(28)	(10)	(331)

TABLEAU I (suite)

Le Sénégal à Bakel

Débits moyens mensuels (en m³/s)

Année	M	J	J	A	S	O	N	D	J	F	M	A	Module
1945-46	(10)	(120)	396	3260	4738	1909	(464)	(195)	(110)	(65)	(35)	(14)	(946)
1946-47	(10)	(120)	362	2505	3024	1819	580	(238)	(130)	(75)	(41)	(17)	(747)
1947-48	(10)	(120)	343	1860	3363	1509	(397)	(180)	(105)	(60)	(32)	(12)	(666)
1948-49	(5)	44	712	2041	2919	1104	495	(191)	(105)	(60)	(32)	(12)	(646)
1949-50	(5)	(9)	377	2272	2123	947	(282)	(155)	(90)	(50)	(27)	(10)	(532)
1950-51	(5)	(3)	545	2914	5891	3071	778	304	153	86	43	13	1154
1951-52	4	57	387	1418	2331	3581	1455	423	214	125	(64)	(27)	843
1952-53	(5)	22	524	1395	2421	3126	(597)	(246)	(134)	(71)	37	17	721
1953-54	3	101	788	1547	2926	1236	464	219	140	81	41	13	632
1954-55	12	253	963	3987	4419	1655	681	396	197	116	68	42	1070
1955-56	32	207	612	3563	4004	2615	770	347	203	119	69	34	1051
1956-57	13	40	495	2210	5237	2159	634	285	163	99	60	24	953
1957-58	8	215	608	2668	4227	2904	935	351	197	118	67	32	1031
1958-59	18	175	568	3985	4028	1916	785	444	237	139	84	40	1040
1959-60	19	164	583	2434	4047	1242	489	223	126	76	42	17	788
1960-61	5	82	789	1790	2508	1301	504	213	120	75	41	16	623
1961-62	3,5	102	781	2956	5201	1360	458	207	121	74	40	12	945
1962-63	2,7	85	511	2220	3632	1620	594	262	138	86	43	18	770
1963-64	8	7	473	1620	2772	1988	636	230	129	72	36	18	667
1964-65	(3,2)	171	602	1973	5680	1989	580	285	166	105	58	26	970

- ensemble de 12 variates représentant le débit moyen mensuel de chacun des 12 mois de l'année, ayant chacune sa propre loi de probabilité et non tout à fait indépendantes entre elles;
- série chronologique d'une fonction aléatoire correspondant à l'ensemble des 12 débits mensuels de l'année, caractérisée par une valeur moyenne qui est le module ou débit moyen annuel et une répartition mois par mois de ce module au sein des 12 mois.

C'est ce dernier point de vue, qui se prête très aisément à la constitution d'un échantillon fictif, que nous avons adopté. On peut en effet traiter séparément, par les méthodes habituelles de la statistique, les deux composantes de la fonction aléatoire :

- sa grandeur ou module : chronique d'une variate très souvent normale dont les valeurs successives sont pratiquement indépendantes, sauf dans le cas assez rare d'un important stockage interannuel dans le bassin;
- sa répartition dont on peut se constituer un échantillon à partir de l'échantillon des débits observés. À vrai dire, l'utilisation de ce « réservoir de formes » pose quelques petits problèmes qui seront évoqués ultérieurement.

La constitution d'un échantillon fictif de modules ne présente aucune difficulté, que la loi de probabilité envisagée soit normale ou pas. Dans l'exemple qui est proposé, on supposera que la distribution est normale; les valeurs des paramètres, estimées d'après l'échantillon naturel, sont alors :

- moyenne $\bar{Q} = 768 \text{ m}^3/\text{s}$;
- écart-type $\sigma_Q = 235 \text{ m}^3/\text{s}$.

Le tableau II montre les résultats obtenus par tirage au sort; ils constituent le début d'un échantillon fictif. La première colonne contient les nombres lus dans un ordre prédéterminé sur la table de nombres au hasard; les colonnes suivantes correspondent aux trois phases du calcul de Q .

Le tableau III contient les coefficients mensuels du Sénégal à Bakel pour la même période d'observations; ce sont, pour chaque mois, les pourcentages des débits moyens mensuels au module de l'année correspondante : la somme des 12 coefficients d'une année est égale à 100. Il s'agit donc d'une collection de formes possibles de répartition du débit annuel sur les 12 mois de l'année. On peut affecter à chacune de ces formes un numéro, n'importe lequel pourvu qu'il soit inférieur à un nombre entier fixé arbitrairement mais au moins égal au nombre maximal d'années d'observation, et à la condition qu'aucun numéro ne soit répété.

Dans l'exemple cité, on peut entre autres affecter aux formes du tableau III le n° 01 pour l'année 1903-04, le n° 02 pour 1904-05, etc. . . jusqu'au n° 62 pour 1964-65. Chaque fois que l'on a tiré un module suivant la méthode indiquée plus haut, on lit un nombre de deux chiffres dans une table de nombres au hasard; si ce nombre est égal à 00 ou supérieur à 62, on le rejette et on va en lire un autre, jusqu'à ce qu'on tombe sur un nombre compris entre 01 et 62 ou égal à l'un de ces nombres. On affecte alors au module que l'on vient de tirer la forme dont le numéro est égal au premier nombre de deux chiffres répondant au critère imposé. Autrement dit, on multiplie par le module les 12 coefficients de la forme de répartition mensuelle, au préalable divisés par 100.

TABLEAU II

*Tirage au hasard d'un échantillon de modules
dans une loi normale de moyenne 768 m³/s et d'écart-type 235 m³/s*

Nombre extrait de la table	Fréquence	Variable réduite de Gauss (u)	Module (m ³ /s)
46 675	$F(u) = 0,46675$	- 0,085	748
00 526	$F(u) = 0,00526$	- 2,56	168
04 111	$F(u) = 0,04111$	- 1,74	358
49 788	$F(u) = 0,49788$	- 0,001	766
14 254	$F(u) = 0,14254$	- 1,07	517
31 844	$F(u) = 0,31844$	- 0,47	658
12 905	$F(u) = 0,12905$	- 1,13	502
09 799	$F(u) = 0,09799$	- 1,29	465
49 899	$F(u) = 0,49899$	0,00	768
17 194	$F(u) = 0,17194$	- 0,95	545
66 104	$F_1(u) = 0,33896$	+ 0,41	862
94 249	$F_1(u) = 0,05751$	+ 1,58	1 139
99 341	$F_1(u) = 0,00651$	+ 2,48	1 351
etc.			

TABLEAU III

Le Sénégal à Bakel

Coefficients mensuels

Année	M	J	J	A	S	O	N	D	J	F	M	A
1903-04	(0,14)	(1,8)	9,5	26,3	37,9	13,9	(5,7)	(2,1)	(1,3)	(0,8)	(0,41)	(0,15)
1904-05	(0,06)	(0,12)	7,3	30,5	37,2	12,7	6,1	2,6	1,6	1,1	(0,50)	(0,23)
1905-06	(0,10)	1,9	8,7	26,8	22,4	23,3	10,2	3,1	1,7	0,97	(0,58)	(0,25)
1906-07	0,17	0,63	7,5	39,8	30,0	11,2	5,2	(2,6)	(1,4)	(0,84)	(0,46)	(0,20)
1907-08	(0,16)	(1,9)	(6,4)	14,5	35,1	20,5	9,8	(5,4)	(3,0)	(1,8)	(1,0)	(0,44)
1908-09	(0,11)	0,89	8,7	23,9	40,2	15,3	(5,4)	(2,6)	(1,4)	(0,82)	(0,47)	(0,21)
1909-10	(0,10)	(2,6)	8,8	27,5	38,4	12,0	5,5	(2,4)	(1,3)	(0,77)	(0,43)	(0,20)
1910-11	(0,13)	(1,5)	(7,4)	26,6	37,5	15,2	5,9	(2,7)	(1,5)	(0,88)	(0,48)	(0,21)
1911-12	(0,15)	(1,9)	(9,2)	22,6	37,8	14,4	6,7	(3,4)	(1,9)	(1,1)	(0,60)	(0,25)
1912-13	(0,16)	(1,8)	(8,7)	21,2	34,8	19,3	(6,5)	(3,4)	(2,0)	(1,2)	(0,66)	(0,28)
1913-14	(0,23)	(2,8)	(11,5)	20,8	26,1	20,2	(8,6)	(4,5)	(2,6)	(1,5)	(0,84)	(0,33)
1914-15	(0,20)	(2,3)	(11,1)	24,9	26,8	19,5	(6,8)	(3,8)	(2,2)	(1,3)	(0,77)	(0,33)
1915-16	(0,15)	(1,3)	9,0	26,7	34,4	17,8	(4,9)	(2,7)	(1,5)	(0,88)	(0,49)	(0,18)
1916-17	(0,06)	(0,05)	8,9	21,6	39,0	20,2	(4,8)	(2,5)	(1,4)	(0,84)	(0,45)	(0,20)
1917-18	(0,14)	(0,27)	(3,4)	(27,4)	43,7	15,3	(4,5)	(2,5)	(1,4)	(0,80)	(0,44)	(0,15)
1918-19	(0,09)	(1,3)	5,7	25,3	38,1	19,1	5,0	(2,5)	(1,4)	(0,83)	(0,46)	(0,22)
1919-20	(0,16)	(2,2)	(6,4)	(26,8)	35,7	16,1	(5,6)	(3,3)	(1,8)	(1,1)	(0,60)	(0,24)
1920-21	(0,10)	(1,2)	(5,4)	25,4	42,6	13,1	6,0	(2,9)	(1,6)	(0,95)	(0,52)	(0,23)
1921-22	(0,20)	(2,3)	(7,7)	(23,2)	(40,7)	14,3	(5,2)	(2,9)	(1,8)	(0,98)	(0,52)	(0,20)
1922-23	(0,07)	(0,27)	2,8	22,0	46,1	19,0	5,3	2,2	1,1	(0,65)	(0,36)	(0,15)
1923-24	(0,11)	(0,99)	6,9	(20,0)	41,6	16,1	8,2	3,0	1,5	(0,89)	(0,49)	(0,22)
1924-25	(0,06)	(1,1)	9,5	(26,7)	(35,8)	16,6	5,1	2,4	(1,3)	(0,80)	(0,44)	(0,20)
1925-26	(0,14)	(1,0)	3,9	22,7	32,6	24,9	7,7	(3,2)	(1,8)	(1,1)	(0,65)	(0,31)
1926-27	(0,15)	(2,2)	(8,2)	(25,8)	27,9	(15,7)	11,5	4,3	(2,1)	(1,2)	(0,68)	(0,27)
1927-28	(0,07)	(0,93)	(6,0)	21,8	36,9	(21,3)	6,8	(2,9)	(1,6)	(0,92)	(0,54)	(0,24)
1928-29	(0,08)	(0,80)	3,7	26,8	40,9	15,2	6,7	(2,7)	(1,5)	(0,89)	(0,52)	(0,21)
1929-30	(0,08)	(2,9)	8,2	26,9	40,3	12,6	4,5	(2,1)	(1,2)	(0,70)	(0,37)	(0,15)
1930-31	(0,10)	(1,7)	(6,5)	(26,1)	(34,0)	(19,2)	6,0	(2,9)	(1,7)	(0,97)	(0,58)	(0,25)
1931-32	(0,10)	(1,9)	10,6	19,9	30,7	24,0	6,2	(3,0)	(1,8)	(1,0)	(0,56)	(0,24)
1932-33	(0,11)	(1,4)	(8,6)	30,2	34,6	14,9	(4,8)	(2,5)	(1,4)	(0,81)	(0,47)	(0,21)
1933-34	(0,09)	1,5	10,9	33,0	35,8	10,7	(3,8)	(2,0)	(1,1)	(0,67)	(0,34)	(0,10)
1934-35	(0,12)	(0,24)	3,2	27,9	41,7	15,7	(5,3)	(2,6)	(1,6)	(0,90)	(0,51)	(0,23)
1935-36	(0,06)	(0,85)	(6,4)	(30,6)	(35,6)	17,9	(4,5)	(1,9)	(1,1)	(0,61)	(0,34)	(0,14)
1936-37	(0,06)	(0,57)	4,1	31,1	(39,4)	15,2	4,8	2,3	1,2	0,70	0,41	(0,16)
1937-38	(0,12)	(1,6)	5,1	22,7	40,3	17,3	6,5	(3,0)	(1,7)	(0,96)	(0,52)	(0,20)
1938-39	(0,10)	(1,2)	4,9	18,9	41,3	19,3	8,3	(2,8)	(1,6)	(0,90)	(0,50)	(0,20)
1939-40	(0,03)	(0,42)	5,4	28,8	31,2	20,5	(6,5)	(3,3)	(1,9)	(1,1)	(0,60)	(0,25)
1940-41	(0,20)	(0,97)	4,1	25,6	26,0	24,3	(10,3)	(3,9)	(2,3)	(1,3)	(0,74)	(0,29)
1941-42	0,20	2,4	6,8	23,1	42,3	14,8	4,9	2,6	1,5	0,88	0,38	0,14
1942-43	0,20	2,3	7,4	36,2	32,7	10,4	5,1	2,7	1,5	0,86	0,48	0,16
1943-44	0,13	1,5	4,6	23,5	37,0	22,5	5,5	2,4	1,4	0,83	0,46	0,18
1944-45	0,25	3,0	5,7	20,6	36,5	16,7	8,5	4,0	2,4	1,4	0,70	0,25

TABLEAU III (suite)

Le Sénégal à Bakel

Coefficients mensuels

Année	M	J	J	A	S	O	N	D	J	F	M	A
1945-46	0,10	1,1	3,5	28,8	41,8	16,9	4,1	1,7	0,97	0,59	0,31	0,13
1946-47	0,11	1,3	4,1	28,1	33,8	20,4	6,5	2,7	1,5	0,84	0,45	0,20
1947-48	0,15	1,5	4,3	23,3	42,1	18,9	5,0	2,2	1,3	0,73	0,40	0,12
1948-49	0,06	0,58	9,2	26,4	37,8	14,3	6,4	2,5	1,4	0,78	0,42	0,16
1949-50	0,08	0,14	5,9	35,8	33,5	15,0	4,4	2,4	1,4	0,79	0,43	0,16
1950-51	0,04	0,02	3,9	21,1	42,6	22,3	5,7	2,2	1,1	0,63	0,32	0,09
1951-52	0,04	0,56	3,8	14,1	23,1	35,4	14,4	4,2	2,1	1,2	0,63	0,27
1952-53	0,06	0,26	6,2	16,2	28,1	36,3	7,0	2,9	1,5	0,83	0,44	0,21
1953-54	0,06	1,3	10,4	20,5	38,7	16,4	6,1	2,9	1,8	1,1	0,56	0,18
1954-55	0,11	2,0	7,5	31,2	34,5	13,0	5,3	3,1	1,5	0,91	0,55	0,33
1955-56	0,25	1,6	4,9	28,4	31,9	20,7	6,1	2,8	1,6	0,94	0,54	0,27
1956-57	0,12	0,36	4,3	19,3	45,9	18,9	5,6	2,5	1,4	0,88	0,53	0,21
1957-58	0,06	1,7	4,9	21,6	34,3	23,6	7,6	2,9	1,6	0,95	0,54	0,25
1958-59	0,13	1,4	4,6	32,1	32,4	15,5	6,3	3,6	1,9	1,1	0,66	0,31
1959-60	0,20	1,7	6,2	25,7	42,8	13,1	5,2	2,4	1,3	0,80	0,43	0,17
1960-61	0,06	1,1	10,6	24,1	33,7	17,5	6,7	2,9	1,6	1,0	0,54	0,20
1961-62	0,05	0,92	6,9	26,1	45,9	12,0	4,1	1,8	1,1	0,66	0,36	0,11
1962-63	0,03	0,90	5,5	24,2	39,5	17,6	6,4	2,8	1,5	0,92	0,46	0,19
1963-64	0,10	0,09	5,9	20,3	34,7	24,9	8,0	2,9	1,6	0,90	0,44	0,17
1964-65	0,03	1,5	5,1	17,0	48,8	17,1	5,0	2,4	1,4	0,92	0,51	0,24
1965-66												

S'il y a indépendance entre le module et la répartition saisonnière, on a ainsi réalisé un tirage au sort parfaitement correct d'un élément d'une fonction aléatoire représentant les 12 débits moyens mensuels d'une année. En poursuivant l'opération pour l'ensemble de l'échantillon fictif des modules, on réalise ainsi la constitution d'un échantillon fictif des débits moyens mensuels correspondant à la structure statistique définie par l'échantillon des débits observés.

Il reste à savoir si la condition d'indépendance est effectivement réalisée et, dans le cas contraire, comment tenir compte de la dépendance, ou tout au moins quel palliatif utiliser pour atténuer la distorsion systématique qui risque d'en résulter pour l'échantillon fictif.

Une fonction aléatoire se prêtant malaisément à la recherche de corrélations, on doit d'abord rechercher un ou des critères capables de caractériser la forme de cette fonction. Le choix de ces critères pourra dépendre assez largement du but final de l'opération. En gros, ils seront de deux types : ceux qui se rapportent à la forme plus ou moins aiguë de la courbe de répartition, ceux qui traduisent l'arrivée plus ou moins tardive de la crue. Nous appellerons les premiers « indices de pointe », les seconds « indices d'occurrence ».

Pour le Sénégal à Bakel, on a choisi, dans un but comparatif, deux indices de pointe différents : IP1 qui est égal à la somme des coefficients mensuels des mois d'août, septembre et octobre, IP2 qui est égal au coefficient mensuel du mois ayant le débit le plus fort de l'année. Ces deux coefficients, pour chaque année du tableau III, sont représentés sur le graphique

de la figure 2 en regard du module correspondant. Les droites de régression ont été tracées en s'efforçant de minimiser la somme des écarts absolus, sans toutefois procéder à un réglage systématique des paramètres. Ces droites montrent que la valeur des indices de pointe tend à croître avec la grandeur du module. Il n'y a donc pas totalement indépendance dans l'exemple cité et les années fortes ont tendance à avoir des hydrogrammes de forme plus aiguë que ceux des années faibles.

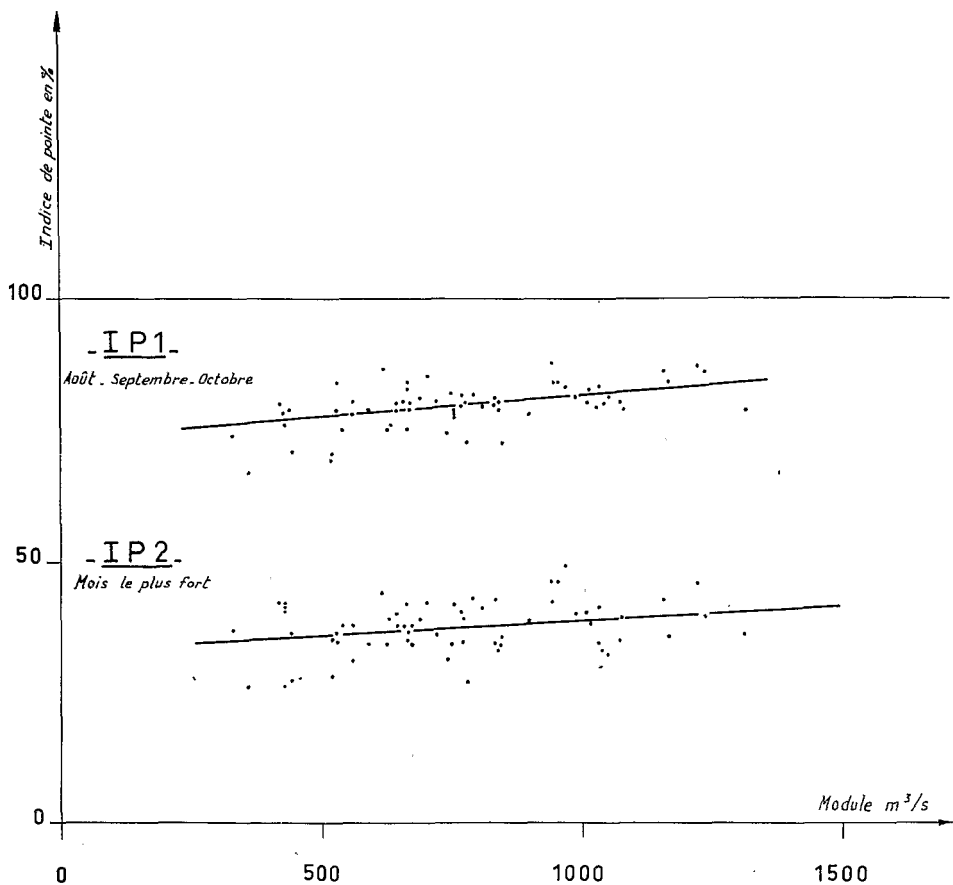


Fig. 2 — Sénégal à Bakel (indices de pointe en regard du module).

D'une façon générale, la relation entre l'indice adopté et le module peut aller de l'indépendance stochastique totale à la relation fonctionnelle, bien que cette dernière hypothèse soit peu vraisemblable. Dans le cas d'une relation fonctionnelle ou quasi fonctionnelle, le choix d'un module par tirage au sort entraîne obligatoirement le choix d'un indice parfaitement défini ou tout au moins dont la valeur est comprise dans une bande étroite, ce qui revient pratiquement à fixer de façon impérative la forme à adopter pour la répartition dans l'année des débits moyens mensuels. Si, au contraire, il y a indépendance, le choix du module n'a aucune influence sur la forme de la répartition et on doit procéder par tirage au sort dans le réservoir des formes comme il a été dit précédemment.

Dans le cas présent, et c'est souvent qu'il en sera ainsi, on se trouve entre ces deux extrêmes. Reprenons l'indice IP2 calculé pour le Sénégal et reportons les valeurs trouvées sur la figure 3 en regard des modules, les ordonnées étant dilatées deux fois par rapport à celles du graphique de la figure 2. Chaque point représentatif est affecté d'un nombre correspondant à la numérotation des formes du tableau III.

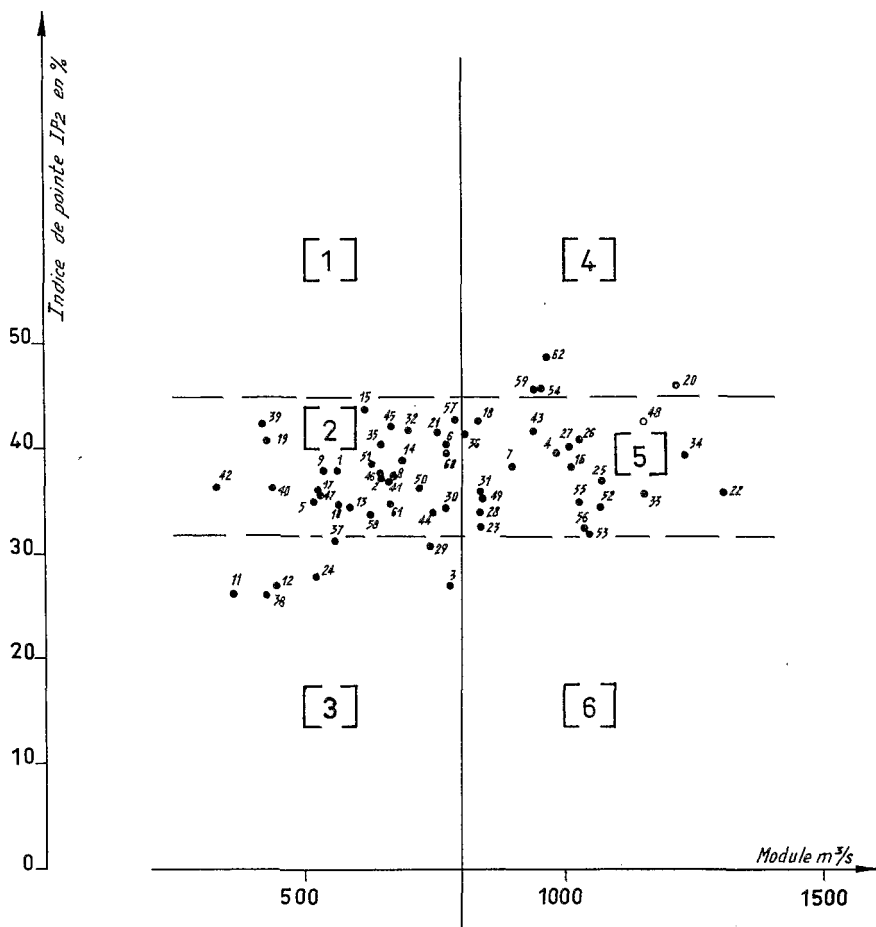


Fig. 3

Coupons le plan par un trait vertical tracé à peu près à égale distance des points correspondant aux valeurs extrêmes des modules. Traçons ensuite deux traits horizontaux qui, avec le trait vertical, diviseront le plan en 6 zones, numérotées de 1 à 6 sur le graphique (chiffres entre crochets), de telle manière que la zone [1] et la zone [6] ne contiennent aucun point, mais que la zone [3] et la zone [4] contiennent le plus de points possible. Toutefois, on peut admettre un ou deux points nettement détachés dans les zones [1] et [6]. Faisons le décompte des points qui se trouvent alors dans chacune des six zones :

- [1] : zéro
- [2] : 30
- [3] : 7
- [4] : 4
- [5] : 21
- [6] : zéro

Les zones [2] et [5] correspondent à une même gamme de valeurs de l'indice de pointe IP2. Si elles seules contenaient des points, il y aurait indépendance stochastique et on pourrait effectuer un tirage au hasard dans le réservoir des formes sans se préoccuper de la valeur du module.

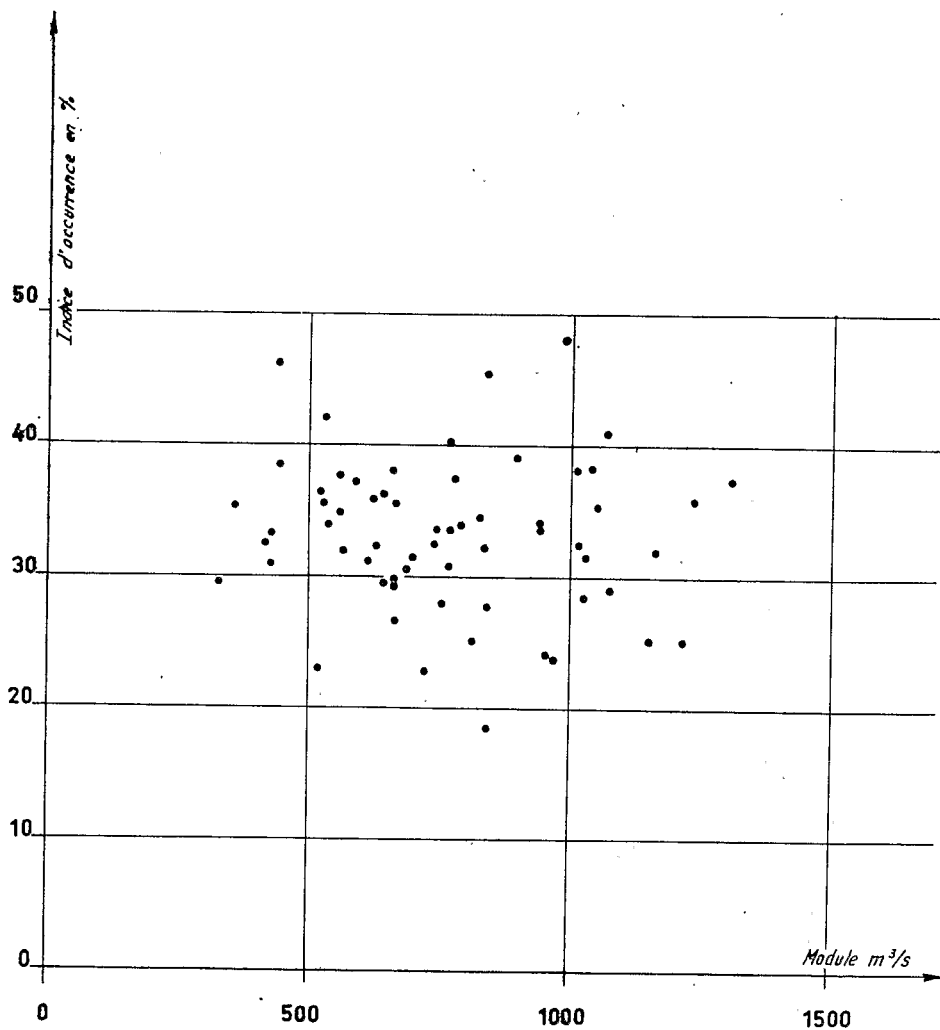


Fig. 4 — Sénégal à Bakel (indices d'occurrence en regard du module).

Si seules les zones [3] et [4] contenaient des points, on devrait tirer les formes dans la zone [3] pour les modules inférieurs à 800 m³/s et dans la zone [4] pour les modules supérieurs à 800 m³/s. Il en résulte que si l'on veut reconstituer un échantillon de débits mensuels tout à fait conforme au modèle fourni par les valeurs observées, on devra constituer deux réservoirs, l'un contenant les formes figurant dans les zones [2], [3] et [5], dans lequel on puisera, toujours par tirage au hasard, quand on aura amené un module inférieur à 800 m³/s, l'autre contenant les formes des zones [2], [4] et [5], à utiliser pour les modules supérieurs à 800 m³/s.

Dans l'exemple donné, l'effet de corrélation entre forme et module est faible; on aurait pu se contenter du tirage dans un réservoir unique contenant l'ensemble des formes. Par contre, il pourrait se faire aussi que le partage en deux classes de modules soit insuffisant; il n'est guère plus compliqué de considérer trois classes ou plus auxquelles on puisse faire correspondre autant de réservoirs de formes.

Le contrôle précédent ne s'adresse qu'à une des caractéristiques de la répartition. Pour la seconde, on doit choisir ce que nous avons appelé un indice d'occurrence et que nous désignerons par 10C; pour le Sénégal à Bakel, on a pris, pour 10C, la somme des coefficients mensuels de mai à août. Les valeurs trouvées ont été portées sur le graphique de la fig. 4, en regard des modules. Dans le cas particulier, il paraît évident, à simple vue du graphique et sans calculer de corrélation, que l'indice d'occurrence choisi n'est d'aucune manière influencé par l'abondance des apports annuels. S'il en était autrement, il conviendrait, pour obtenir un échantillonnage fictif vraiment conforme au modèle fourni par les observations, de compliquer un peu le schéma des réservoirs dans lesquels doivent être effectués les tirages de formes.

Prenez un exemple hypothétique pour lequel on aurait décidé de scinder l'échelle des modules en trois classes dans l'opération de régression indice de pointe — module, et en trois classes également dans la régression indice d'occurrence — module. Le modèle ainsi constitué est déjà d'une très grande souplesse et permet de tenir compte de corrélations très serrées pour ce genre de problème.

TABLEAU IV

Classes de modules	Ensembles de formes possibles		
	Suivant le critère de pointe	Suivant le critère d'occurrence	Répondant l'ensemble des exigences
0, q' ₁	f ₁	G ₁	f ₁ ∩ G ₁
q' ₁ , q ₁	f' ₁	g ₂	f' ₁ ∩ g ₂
q ₁ , q ₂	F ₂	g' ₂	F ₂ ∩ g' ₂
q ₂ , q' ₂	f ₃	g'' ₂	f ₃ ∩ g'' ₂
> q' ₂	f' ₃	G ₃	f' ₃ ∩ G ₃

On supposera que les classes de modules sont définies de la manière suivante :

Indice de pointe :

$$\begin{array}{ll}
 0 \leq \text{module} \leq q_1 & F_1 \\
 q_1 < \text{module} \leq q_2 & F_2 \\
 \text{module} > q_2 & F_3
 \end{array}$$

Indice d'occurrence :

$0 \leq \text{module} \leq q'_1$	G_1
$q'_1 < \text{module} \leq q'_2$	G_2
$\text{module} > q_2$	G_3

On désigne par F_1 l'ensemble des « formes » dans lequel il convient de puiser lorsque le module est compris entre 0 et q_1 pour satisfaire au critère de pointe; il s'établit de la manière qui a été indiquée dans l'exemple du Sénégal. G_1 est l'ensemble des « formes » disponibles lorsque le module est compris entre 0 et q'_1 , etc. . . pour satisfaire au second critère.

On supposera, pour fixer les idées, qu'on a :

$$q'_1 < q_1 < q_2 < q'_2.$$

Les positions respectives des classes de modules pour l'application des deux critères sont représentées graphiquement dans la fig. 5.

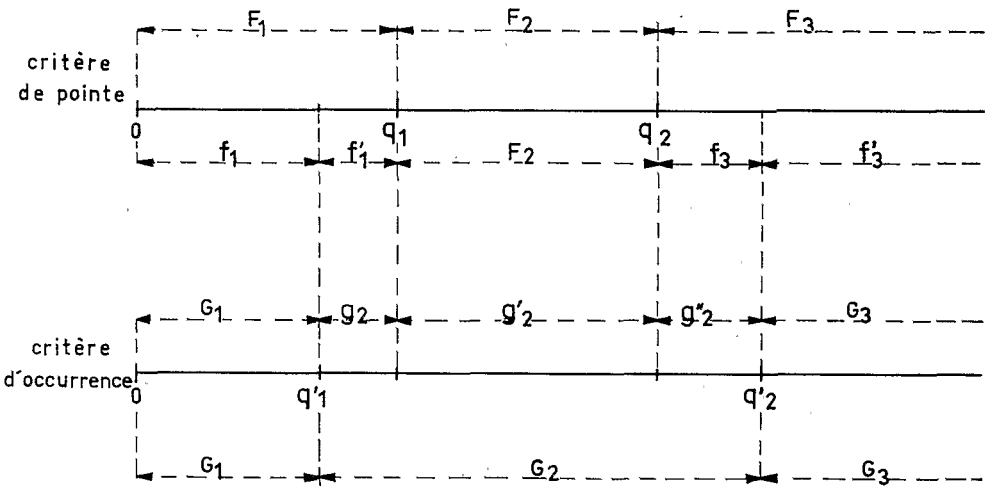


Fig. 5 — Répartition des classes de modules dans l'application des critères de forme.

Décomposons l'ensemble F_1 en deux sous-ensembles tels que le premier, f_1 , contienne les formes qui seraient à prendre en compte si on considérait pour le critère de pointe l'intervalle des modules (0, q'_1) seul, et que le second, f'_1 , contienne les formes qui seraient à prendre en compte si, pour la même opération, on considérait l'intervalle (q'_1 , q_1) seul. Il est à noter que les ensembles f_1 et f'_1 ne sont certainement pas disjoints, mais comportent des éléments communs en nombre d'autant plus grand que la corrélation entre l'indice utilisé et le module est plus lâche. La fig. 5 montre les autres décompositions qui doivent être effectuées pour obtenir le même nombre de classes de modules pour les deux critères de forme. Dans le cas cité, on obtiendrait donc en définitive la répartition en classes figurant sur le tableau IV avec, en regard, les ensembles de formes possibles conformément aux exigences de chacun des deux critères.

Considérons maintenant un de ces nouveaux intervalles, par exemple 0, q'_1 . Si un module tiré au hasard tombe dans cet intervalle, la considération du critère de pointe exige que la forme à appliquer pour obtenir la séquence des débits moyens mensuels soit tirée au sort dans le réservoir de formes constitué par l'ensemble f_1 . Mais la considération du critère d'occur-

rence exige que le tirage soit effectué dans l'ensemble G_1 . Comme on doit nécessairement imputer une seule forme de répartition mensuelle au module obtenu, il faudra que cette forme soit tirée au sort dans un réservoir ne contenant que des éléments communs à f_1 et G_1 , et tous les éléments communs pour éviter une perte d'information. Autrement dit, le tirage sera effectué dans un nouvel ensemble égal à l'intersection $f_1 \cap G_1$ des deux ensembles f_1 et G_1 . Le tableau IV montre ce qu'il en est pour l'ensemble des classes de modules.

En résumé, la suite des opérations à effectuer pour construire un échantillon fictif de débits moyens mensuels est la suivante :

- départ — liste de débits moyens mensuels et de modules observés;
- calcul des coefficients mensuels à partir du tableau précédent. Numérotage de chaque « forme » annuelle ainsi obtenue;
- étude statistique des modules, recherche de la loi de distribution la mieux adaptée, calcul des paramètres;
- choix d'indices de forme convenant le mieux au but final cherché. Dans la plupart des cas, ces indices se rapporteront à un critère de pointe et à un critère d'occurrence;
- recherche de régressions éventuelles entre les indices de forme et les modules;
- compte tenu de ces corrélations, constitution d'un ou plusieurs réservoirs de formes.

Les opérations de tirage se déroulent de la façon suivante :

- on tire un module dans la loi de répartition établie à partir de l'échantillon de débits observés;
- on recherche à quelle classe de modules appartient le module tiré; cette opération oriente vers le réservoir de formes à utiliser;
- on tire au hasard une forme dans le réservoir sélectionné;
- on applique la forme obtenue sur le module tiré;
- on tire un nouveau module, etc.