

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

ROBERT PAOLI

À la recherche des caractères d'une distribution

Journal de la société statistique de Paris, tome 125, n° 2 (1984), p. 98-111

http://www.numdam.org/item?id=JSFS_1984__125_2_98_0

© Société de statistique de Paris, 1984, 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/>

A LA RECHERCHE DES CARACTÈRES D'UNE DISTRIBUTION

Robert PAOLI

*ancien élève de l'École Polytechnique,
administrateur de l'« Entreprise industrielle »*

Les offres présentées par les entreprises en réponse aux appels à la concurrence lancés par les Maîtres d'ouvrages constituent une population dont l'auteur s'est proposé d'analyser les caractères. Tout nouvel appel d'offres peut donner lieu à comparaison avec la distribution-type, notamment en vue de déterminer si l'offre la plus basse constitue ou non un prix de niveau acceptable.

Partant des résultats de près de 300 appels d'offres, l'auteur décrit les propriétés de la distribution des valeurs des offres autour de leur moyenne et calcule la probabilité pour qu'au cours d'une nouvelle consultation apparaisse une offre s'écartant de la moyenne dans une proportion déterminée.

Cet exposé complète un article publié en septembre 1983 dans la Revue « Marchés Publics » et dont tout développement mathématique avait été volontairement exclu.

The bids filled by contractors in response to tender invitations issued by owners form a population the characteristics of which the author proposes to analyse. Any new tender invitation may give rise to a comparison with the standard distribution, specially with a view to determining whether the lowest bid is a price on an acceptable level.

Starting from the results of nearly 300 tender notices, the author describes the properties of the distribution of the amounts of the bids around their mean and calculates the probability of a bid deviating from the mean in a given proportion if new tenders are called for.

This paper supplements an article published in september 1983 in the magazine « Marchés Publics » from which all mathematical arguments were deliberately omitted.

J'ai eu l'occasion d'exposer (voir la Revue « Marchés publics » n° 196 de septembre 1983) que j'avais étudié la répartition des fréquences de la valeur des offres présentées par les entrepreneurs consultés à l'occasion d'un certain nombre d'appels d'offres et que j'avais cherché à identifier la forme de la fonction de distribution correspondante.

L'article en question a été volontairement exempt de tout développement mathématique afin d'en alléger le texte et d'en souligner davantage les conclusions. Il ne m'a cependant pas paru sans intérêt de le compléter par l'exposé des raisonnements qui conduisent à ces conclusions, raisonnements qui reposent sur un emploi quelque peu inhabituel — m'a-t-il semblé — des théories classiques et, en particulier, de celle dite des « petits échantillons ».

Après avoir indiqué dans une première partie le détail de la méthode utilisée pour dépouiller les appels d'offres et après avoir examiné les caractères analytiques de la distribution des fréquences ainsi obtenue, je préciserai les essais effectués en vue de rapprocher cette distribution des lois de répartition connues et, en particulier, de celle de Laplace-Gauss. Je montrerai pour conclure comment un tel rapprochement devrait permettre de diminuer le risque d'erreur lorsque l'on cherche si une offre moins disante doit être considérée comme anormalement basse.

I — MÉTHODE UTILISÉE POUR DÉPOUILLER LES RÉSULTATS DES APPELS D'OFFRES

Les appels d'offres dont j'ai pu disposer sont au nombre de 296 et comprennent 2 364 offres. Comme chacun d'entre eux porte sur des travaux différents donc d'une valeur *a priori* différente, il était nécessaire pour comparer entre elles les offres issues de ces diverses consultations de faire usage d'une base de comparaison ayant une définition commune. La base qui a été choisie est la moyenne arithmétique des offres appartenant à un même appel d'offres, à laquelle on affecte la valeur 100. Chacune des offres est ainsi évaluée par son pourcentage de la moyenne correspondante et c'est sous cette forme qu'elle est introduite dans la statistique générale.

On aurait pu choisir un autre point de comparaison, par exemple prendre comme unité l'offre la plus basse de chaque appel d'offres; mais cela n'aurait pas dispensé de calculer la moyenne arithmétique et d'en introduire aussi la valeur dans le dépouillement en raison du rôle particulier joué par cette variable dans les raisonnements qui vont suivre.

On dispose donc de 2 364 valeurs d'offres exprimées en pourcentage de la moyenne correspondante. Comme il est également intéressant d'analyser le comportement des moins disants, on a noté au passage la valeur des 296 offres les plus basses (toujours en pourcentage de la moyenne correspondante).

Enfin, on s'est préoccupé du nombre d'offres de chaque consultation.

Les résultats de ces trois analyses font l'objet de trois tableaux.

— Le premier est obtenu en répartissant les pourcentages des offres par rapport à leur moyenne

TABLEAU I
Répartition des offres autour de leur moyenne

Classe	Fréquence	Fréquence cumulée	Classe	Fréquence	Fréquence cumulée	Classe	Fréquence	Fréquence cumulée
45-46	1	1	93-94	61	563	128-129	0	2 323
...	—	—	94-95	77	640	129-130	3	2 326
60-61	1	2	95-96	102	742	130-131	1	2 327
61-62	1	3	96-97	96	838	131-132	1	2 328
62-63	—	3	97-98	100	938	132-133	3	2 331
63-64	—	3	98-99	128	1 066	133-134	0	2 331
64-65	3	6	99-100	147	1 213	134-135	4	2 335
65-66	—	6	100-101	121	1 334	135-136	2	2 337
66-67	—	6	101-102	137	1 471	136-137	2	2 339
67-68	4	10	102-103	119	1 590	137-138	2	2 341
68-69	4	14	103-104	112	1 702	138-139	1	2 342
69-70	2	16	104-105	88	1 790	139-140	2	2 344
70-71	2	18	105-106	66	1 856	140-141	1	2 345
71-72	3	21	106-107	70	1 926	141-142	1	2 346
72-73	3	24	107-108	52	1 978	142-143	2	2 348
73-74	7	31	108-109	42	2 020	143-144	1	2 349
74-75	5	36	109-110	42	2 062	144-145	1	2 350
75-76	5	41	110-111	30	2 092	145-146	3	2 353
76-77	11	52	111-112	34	2 126	146-147	0	2 353
77-78	12	64	112-113	28	2 154	147-148	1	2 354
78-79	6	70	113-114	36	2 190	148-149	1	2 355
79-80	11	81	114-115	20	2 210	149-150	1	2 356
80-81	16	97	115-116	17	2 227	150-151	0	2 356
81-82	12	109	116-117	12	2 239	151-152	1	2 357
82-83	23	132	117-118	9	2 248	152-153	0	2 357
83-84	20	152	118-119	11	2 259	153-154	1	2 358
84-85	18	170	119-120	10	2 269	154-155	1	2 359
85-86	20	190	120-121	12	2 281	155-156	0	2 359
86-87	35	225	121-122	11	2 292	156-157	1	2 360
87-88	26	251	122-123	7	2 299	157-158	0	2 360
88-89	39	290	123-124	9	2 308	158-159	1	2 361
89-90	41	331	124-125	4	2 312	159-160	0	2 361
90-91	40	371	125-126	4	2 316	160-161	1	2 362
91-92	67	438	126-127	3	2 319	166-167	1	2 363
92-93	64	502	127-128	4	2 323	175-176	1	2 364

en classes de 1 %. La classe la plus basse comprend une offre située entre 45 et 46 %, la plus élevée une offre comprise entre 175 et 176 %.

— Le second utilise les mêmes données; la classe inférieure est toujours la classe 45-46 %, la plus élevée, 97-98 %.

— Le troisième tableau donne le nombre d'appels d'offres ayant fait l'objet de 2, 3, etc. réponses. Le nombre des réponses s'échelonne de 2 à 19.

II — EXAMEN DES RÉSULTATS OBTENUS

II.1. La distribution des offres autour de leur moyenne est caractérisée par un écart-type de 11,19 %; mais cet écart-type ne présente pas la propriété de celui de la distribution de Laplace-Gauss : en particulier, la proportion d'offres comprises entre la moyenne moins un écart-type et la moyenne plus un écart-type (c'est-à-dire entre 89 et 111) est de 76 % au lieu de 68. La répartition est donc plus concentrée autour de sa moyenne qu'une répartition normale.

On peut avoir une confirmation plus précise de ces résultats en calculant les moments des divers ordres :

$$\mu_2 = 125,16$$

$$\beta_1 = \frac{u_3^2}{u_2^3} = 0,607 \quad \gamma_1 = \sqrt{\beta_1} = 0,779$$

$$\mu_3 = 1.091,14$$

$$\beta_2 = \frac{u_4}{u_2^2} = 7,600 \quad \gamma_2 = \beta_2 - 3 = 4,600$$

$$\mu_4 = 119.060,76$$

La distribution est leptokurtique ($\gamma_2 > 0$). Quant à sa dissymétrie, elle est reflétée par le coefficient de Pearson.

$$Sk (Skewness) = \frac{\gamma_1 (\beta_2 + 3)}{2 (5 \beta_2 - 6 \beta_1 - 9)} = 0,163$$

Si la distribution était symétrique ce coefficient serait nul.

En tout état de cause, et pour illustrer graphiquement ce qui vient d'être dit, il est intéressant de comparer la courbe empirique des fréquences à celle qui représente la fonction de Laplace-Gauss de même moyenne et de même écart-type. C'est l'objet de la figure 1. Pour cette comparaison, la répartition empirique a été lissée en portant en ordonnée au droit de l'abscisse médiane de la classe n la moyenne des ordonnées des classes $n - 1, n, n + 1$. L'écart le plus net entre les 2 courbes se situe aux environs de la moyenne : ordonnée maximale de la courbe empirique 135, ordonnée maximale de la fonction normale 84,2. J'aurai l'occasion de revenir par la suite sur ce point.

II.2. Le second tableau donne la distribution des fréquences des offres moins-disantes. Sa représentation graphique après, comme précédemment, lissage sur 3 classes est donnée par la figure 2. Cette distribution est éminemment dissymétrique puisqu'elle est bornée par l'abscisse 100. On notera seulement que la moyenne des moins-disants est de 86,3 % (soit un écart de 13,7 avec la moyenne générale) et que la médiane 88,2 % s'en trouve éloignée de 2 points.

Le coefficient de Pearson Sk est égal à 0,668.

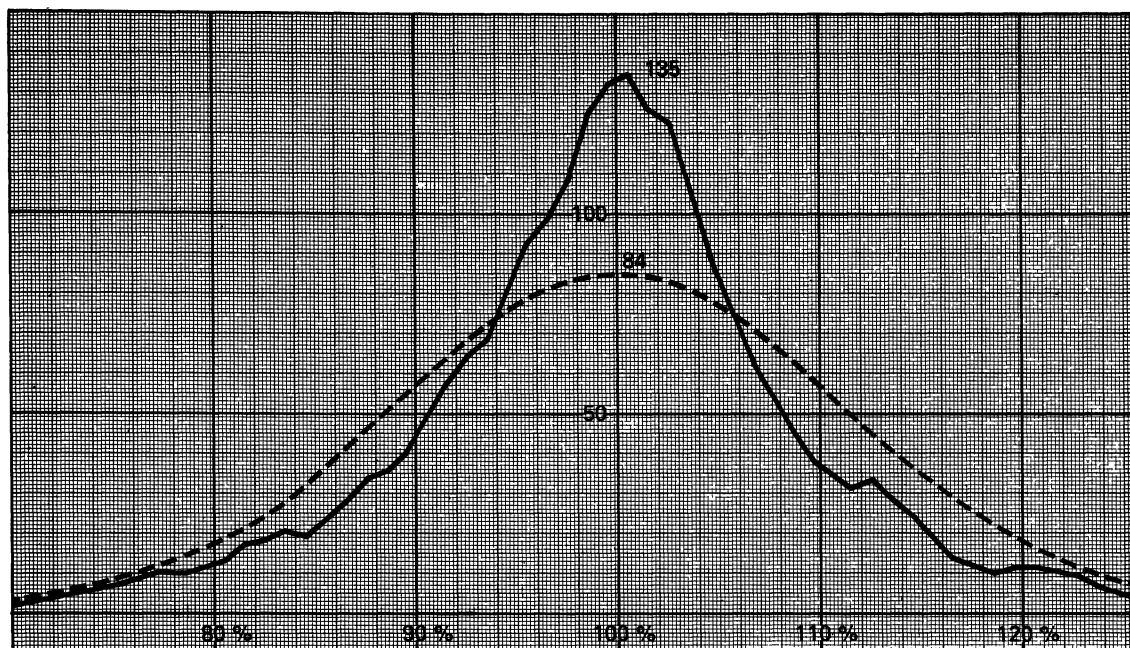


FIG. 1. — En trait plein, nombre d'offres par classe de 1 % en fonction de la position de la classe exprimée en pourcentage de la moyenne. En tireté, courbe des fréquences de Laplace-Gauss (loi normale) de même moyenne et de même écart-type.

TABLEAU II

Distribution des rapports moins-disant/moyenne en %

Classe	Fréquence	Fréquence cumulée	Classe	Fréquence	Fréquence cumulée	Classe	Fréquence	Fréquence cumulée
45-46	1	1	75-76	3	34	87-88	13	147
60-61	1	2	76-77	8	42	88-89	15	162
61-62	1	3	77-78	6	48	89-90	14	176
64-65	2	5	78-79	1	49	90-91	11	187
67-68	4	9	79-80	9	58	91-92	21	208
68-69	4	13	80-81	13	71	92-93	27	235
69-70	2	15	81-82	4	75	93-94	15	250
70-71	0	15	82-83	13	88	94-95	15	265
71-72	3	18	83-84	11	99	95-96	22	287
72-73	2	20	84-85	11	110	96-97	8	295
73-74	7	27	85-86	10	120	97-98	1	296
74-75	4	31	86-87	14	134			

II.3. Le nombre de réponses reçues à chaque appel d'offres s'étend de 2 à 19. Leur fréquence est consignée dans le troisième tableau. La moyenne des offres reçues est égale à 8 et le nombre de réponses le plus probable est 7. On remarquera que cette distribution est très régulière.

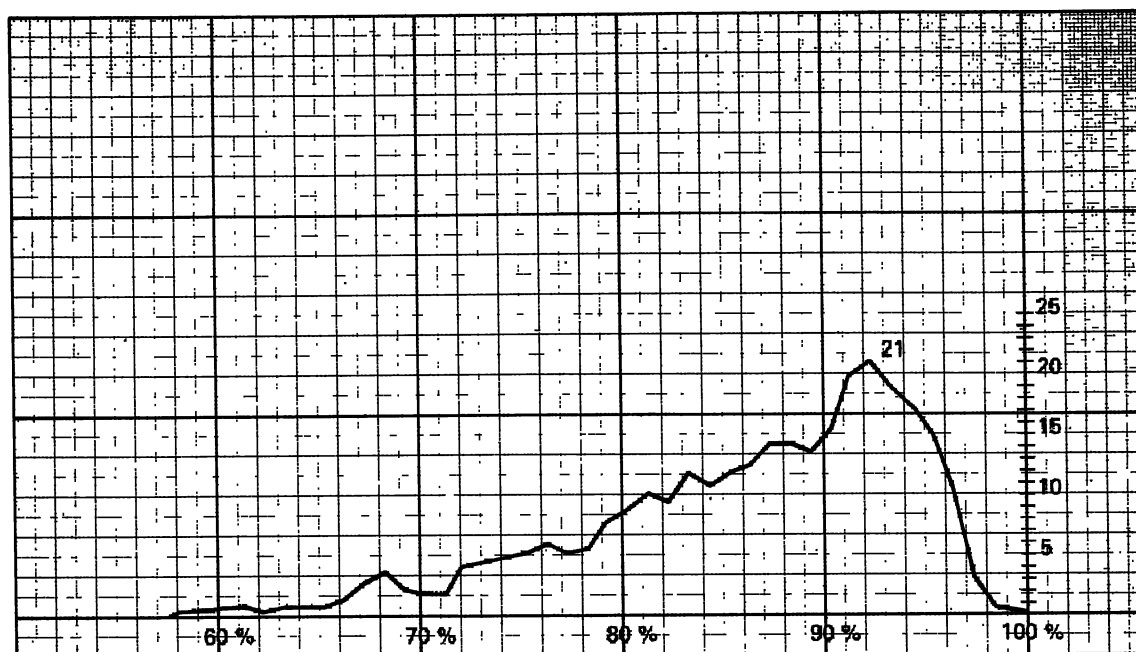


FIG. 2. — Nombre de moins disants par classe de 1 %
en fonction de la position de la classe exprimée en pourcentage de la moyenne

TABLEAU III
Nombre de réponses par appel d'offres

Nombre de réponses	Fréquence	Fréquence cumulée	Nombre de réponses	Fréquence	Fréquence cumulée	Nombre de réponses	Fréquence	Fréquence cumulée	Nombre de réponses	Fréquence	Fréquence cumulée
2	5	5	6	34	102	10	25	234	14	8	287
3	15	20	7	38	140	11	21	255	15	6	293
4	18	38	8	35	175	12	15	270	16	2	295
5	30	68	9	34	209	13	9	279	19	1	296

III.1. Les résultats précédents concernent, on l'a vu, 296 appels d'offres groupant 2 364 offres. La courbe de répartition de ces dernières traduit d'une manière très générale l'allure de la loi qui gouverne cette répartition. Mais il faut bien voir qu'en dépit du nombre relativement important des offres classées (2 364) on n'a dépouillé que 296 appels d'offres, qu'au demeurant le nombre d'offres par appel d'offres est assez dispersé (il s'étend de 2 à 19) et que même si on voulait se limiter à comparer entre eux les appels d'offres ayant donné lieu au même nombre de réponses — ce qui bornerait à 38 appels d'offres les comparaisons possibles, dans le cas le plus favorable — on aurait affaire à des consultations portant sur des travaux très différents, tant par leur nature que par leur montant.

C'est cette remarque qui m'a amené à suggérer que l'Administration qui dispose d'un champ d'expérimentation très étendu procède à une analyse beaucoup plus détaillée en prenant en compte la nature des travaux, leur montant, leur localisation, et naturellement le nombre des réponses reçues.

Ces réserves ne modifient heureusement en rien les méthodes qui seront exposées maintenant.

III.2. Au paragraphe II-1 on a signalé que la courbe empirique des fréquences et celle qui représente la fonction de Laplace-Gauss de même moyenne et de même écart-type n'étaient guère voisines. On sait pourtant que cette dernière fonction est la plus vraisemblable.

On peut alors se demander s'il ne conviendrait pas de rechercher une fonction qui soit, dans l'ensemble, la plus voisine de la distribution des fréquences observées. Soit f_i la fréquence observée dans l'intervalle $\left(x_i - \frac{\Delta x}{2}, x_i + \frac{\Delta x}{2}\right)$ et soit $p(x_i)$ la fonction de probabilité cherchée. Au droit de l'abscisse x_i l'écart est $f_i - np(x_i)$ ou pour simplifier $f_i - np_i$. Notons que l'on a déjà par hypothèse $\sum (f_i - np_i) = 0$, puisque $\sum f_i = n$ et $\sum p_i = 1$. C'est donc par la somme des carrés des écarts que nous définirons notre fonction :

$$\sum (f_i - np_i)^2 \text{ minimale}$$

Dans cette expression p_i est une fonction absolument quelconque, mais pour aller plus loin, nous devons en préciser la forme, faisant appel à un certain nombre de paramètre, α, β, \dots et la somme en question sera minimale si les dérivées partielles par rapport à $\alpha, \beta \dots$ sont nulles :

$$\sum (f_i - np_i) \frac{\partial p_i}{\partial \alpha} = 0 \quad \sum (f_i - np_i) \frac{\partial p_i}{\partial \beta} = 0 \dots$$

Si nous choisissons pour p_i une loi de Laplace-Gauss de moyenne m et d'écart-type σ , les équations qui précèdent s'écrivent :

$$\begin{cases} \sum f_i \frac{\partial p_i}{\partial m} = n \sum p_i \frac{\partial p_i}{\partial m} \\ \sum f_i \frac{\partial p_i}{\partial \sigma} = n \sum p_i \frac{\partial p_i}{\partial \sigma} \end{cases}$$

Les seconds membres ne contiennent pas f_i et peuvent être remplacés, compte tenu des approximations habituelles, par des intégrales.

$$n \int_{-\infty}^{+\infty} p \frac{\partial p}{\partial m} dx \text{ et } n \int_{-\infty}^{+\infty} p \frac{\partial p}{\partial \sigma} dx, \text{ avec } p = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

La première de ces deux intégrales est nulle; quant à la seconde, elle se calcule sans difficulté et conduit à :

$$-\frac{1}{4\sigma^2\sqrt{\pi}}$$

La fonction cherchée sera donc définie par la valeur de ses deux paramètres m et σ tirés du système d'équations

$$\begin{cases} \sum f_i \frac{\partial p_i}{\partial m} = 0 \\ \sum f_i \frac{\partial p_i}{\partial \sigma} = \frac{-n}{4\sigma^2\sqrt{\pi}} \end{cases}$$

En appliquant cette méthode à la distribution des fréquences observées, et en résolvant par rapport à m et σ les équations qui précèdent, ce qui ne peut se faire que par approximations successives, on trouve

$$\begin{aligned} m &= 99,73 \% \\ \sigma &= 7,83 \% \end{aligned}$$

La somme des carrés des écarts est de 7 308.

III.3. Ce qui précède appelle diverses observations.

III.3.1. Les sommes Σ s'étendent en principe à la totalité du domaine de variation de la variable x_i c'est-à-dire dans le cas d'espèce de $-\infty$ à $+\infty$. Mais les fréquences observées ne sont différentes de zéro que dans un intervalle limité (dans le cas qui nous occupe de la classe dont le milieu est 45,5 jusqu'à celle dont le milieu est 175,5). Il n'y a aucune contradiction entre l'intégration du second membre de $-\infty$ à $+\infty$ et la sommation du premier dans l'intervalle limité au domaine des fréquences non nulles.

III.3.2. L'écart-type ainsi déterminé est très différent de celui de la population-mère et, partant, de celui de la loi de Laplace-Gauss qui conserve ce dernier. Pour matérialiser ces 3 répartitions, on les a représentées sur un même graphique (fig. 3). Il saute aux yeux que la distribution dont l'écart-type est 7,83 suit la distribution empirique de bien plus près que celle qui a le même écart-type. La somme des carrés des écarts passe de 9 215 à 7 308. Il paraît nécessaire de rappeler les raisons intrinsèques de cette différence. Je m'abstiendrai de reprendre en détail les raisonnements correspondants et me contenterai d'énoncer les résultats.

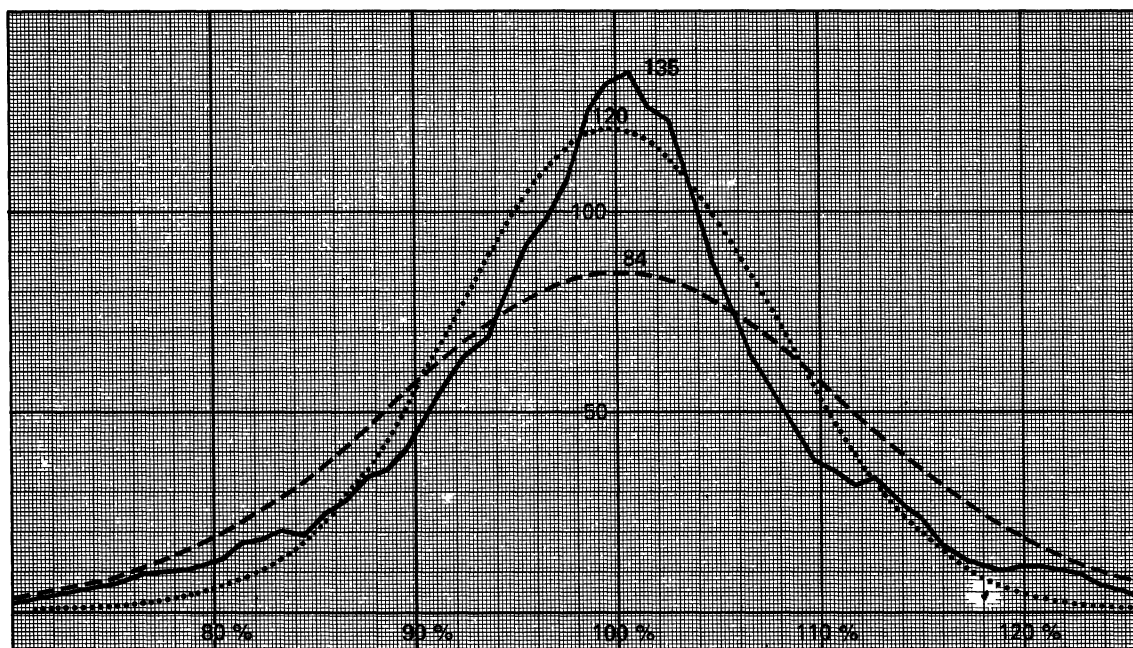


FIG. 3 —

- Nombre d'offres par classe de 1 % en fonction de la position de la classe exprimée en pourcentage de la moyenne.
- courbe des fréquences de Laplace-Gauss (loi normale) de même moyenne et de même écart-type.
- — — courbe des fréquences de Laplace-Gauss (loi normale) la plus voisine.

La fonction de Laplace-Gauss qui a même moyenne et même écart-type que la population-mère est celle qui satisfait au système d'équations

$$\begin{cases} \Sigma \frac{f_i}{p_i} \frac{\partial p_i}{\partial m} = 0 \\ \Sigma \frac{f_i}{p_i} \frac{\partial p_i}{\partial \sigma} = 0 \end{cases}$$

Celle qui minimise la somme des carrés des écarts répond de son côté au système

$$\begin{cases} \sum f_i \frac{\partial p_i}{\partial m} = n \sum p_i \frac{\partial p_i}{\partial m} \\ \sum f_i \frac{\partial p_i}{\partial \sigma} = n \sum p_i \frac{\partial p_i}{\partial \sigma} \end{cases}$$

Considérons les premiers membres des équations de ces deux systèmes et plus particulièrement les termes de rang i . Dans le 1^{er} système le poids du terme de rang i est $\frac{f_i}{p_i}$ dans le second f_i . Le rapport du second de ces poids au premier est donc égal à p_i et par conséquent très petit lorsqu'on s'éloigne de la moyenne. Autrement dit, la fonction de Laplace-Gauss qui a même moyenne et même écart-type que la population-mère tient compte des termes extrêmes que la fonction qui rend minimale la somme des carrés des écarts néglige pratiquement.

Cette remarque revêt toute son importance dans le cas qui nous occupe, car elle montre que cette dernière fonction élimine pratiquement les « queues » donc les valeurs de la variable qu'on peut considérer comme aberrantes pour s'attacher au contraire à celles qui sont voisines de la moyenne.

III.3.3 On peut se demander s'il est légitime d'envisager une fonction $p(x_i)$ dont la moyenne m n'est pas *a priori* identique à celle de la distribution empirique d'où l'on est parti. Dans la méthode qui vient d'être décrite, nous avons considéré m comme un des deux paramètres qui définissent la fonction cherchée. Si les deux équations indiquées sont satisfaites, elles assigneront à m et σ des valeurs telles que la fonction de Laplace-Gauss correspondante sera la plus voisine de la répartition empirique étudiée. Le résultat obtenu sera mathématiquement indiscutable. L'est-il également du point de vue du domaine étudié, celui des offres répondant aux appels à la concurrence? En raison de leur formation les différentes valeurs de la variable aléatoire x_i ont obligatoirement une moyenne égale à 100. En cherchant une loi de Laplace-Gauss aussi voisine que possible de la répartition observée, on tente de définir la loi idéale à laquelle obéirait — peut-être — la population-mère. Une propriété fondamentale de cette loi idéale, est bien, par définition, d'avoir une moyenne égale à 100. Il semblerait donc qu'une loi qui ne respecterait pas cette condition ne peut constituer cette loi idéale. A la question ainsi posée, il est possible de répondre de deux façons :

- on peut toujours ne faire dépendre la loi cherchée que du seul paramètre σ en s'imposant $m = 100$. Il ne subsiste alors qu'une équation et l'on trouve $\sigma = 7,81$ au lieu de 7,83; la somme des carrés des écarts remonte à 7 429 au lieu de 7 308. En tout état de cause, dans le cas particulier, les différences ne sont pas considérables.
- on peut aussi faire remarquer que le principe même de la méthode utilisée qui diminue d'autant plus l'influence des valeurs de la variable qu'elles s'éloignent davantage de la moyenne conduit obligatoirement à une révision de celle-ci. C'est cette moyenne révisée qui constituerait le juste prix ou vraie valeur des travaux.

Au demeurant, dans le cas particulier que nous évoquons, la valeur trouvée pour m (99,73) est si peu différente de 100, que la question du présent paragraphe procède plus d'un souci de rigueur de raisonnement que de la recherche d'un intérêt pratique.

III.3.4. On peut enfin être encore plus ambitieux dans la recherche d'identification de la distribution des fréquences observées, en considérant que celles-ci n'appartiennent pas à une population pure, mais composée, c'est-à-dire résultant de plusieurs groupes de valeurs de la variable ayant chacun sa moyenne et son écart-type propres, issus d'une loi de Laplace-Gauss particulière.

Bien qu'à mon avis les conclusions à tirer d'une telle recherche ne présentent pas un intérêt considérable, notamment pour ce qui va suivre, et que leur interprétation soit des plus délicates, il ne m'a pas paru inutile d'exposer comment on peut procéder.

J'ai effectué cette recherche dans l'hypothèse où la population-mère serait composée de 3 groupes, relevant chacun d'une loi normale. Soient g_1, g_2 et g_3 ces fonctions; la probabilité p_i serait la somme :

$$p_i = \alpha g_1(x_i) + \beta g_2(x_i) + \gamma g_3(x_i)$$

Il est nécessaire que $\alpha + \beta + \gamma = 1$.

La probabilité p_i dépend donc de 8 paramètres qui sont les moyennes et les écarts-types des fonctions g_1, g_2 et g_3 : $m_1, m_2, m_3, \sigma_1, \sigma_2$ et σ_3 et les proportions α et β, γ étant égal à $1 - \alpha - \beta$.

Comme précédemment, nous voulons que $\Sigma (f_i - np_i)^2$ soit minimale, ce qui revient à annuler les dérivées par rapport aux paramètres, conduisant à 8 équations à 8 inconnues. De ces 8 équations, celles qui résultent de la dérivation par rapport à α et β présentent un intérêt particulier. Elles s'écrivent en effet

$$\begin{cases} \Sigma f_i \frac{\partial p_i}{\partial \alpha} = n \Sigma p_i \frac{\partial p_i}{\partial \alpha} \\ \Sigma f_i \frac{\partial p_i}{\partial \beta} = n \Sigma p_i \frac{\partial p_i}{\partial \beta} \end{cases}$$

Les seconds membres ne contiennent pas f_i et peuvent être remplacés par des intégrales de la forme

$$n \int_{-\infty}^{+\infty} p(x) \frac{\partial p}{\partial \alpha} dx$$

Observant que $p(x) = \alpha g_1 + \beta g_2 + (1 - \alpha - \beta)g_3$

et que

$$\frac{\partial p}{\partial \alpha} = g_1 - g_3, \quad \frac{\partial p}{\partial \beta} = g_2 - g_3,$$

on constate que le second membre est une somme d'intégrales de la forme $\int g_j g_k dx$ dont la valeur est

$$\frac{1}{\sqrt{2\pi(\sigma_j^2 + \sigma_k^2)}} e^{-\frac{(m_j - m_k)^2}{2(\sigma_j^2 + \sigma_k^2)}}$$

Dans le cas où $j = k$ cette expression se réduit à $\frac{1}{2\sigma_j\sqrt{\pi}}$.

On a donc tous les éléments pour écrire les équations correspondant aux dérivées partielles par rapport à α et β , qui sont de la forme :

$$\Sigma f_i (g_1 - g_3) = A\alpha + B\beta + C$$

$$\Sigma f_i (g_2 - g_3) = B\alpha + D\beta + E$$

Ainsi à tout système de paramètres $m_1 \dots \sigma_3$ on peut immédiatement adjoindre les valeurs α et β correspondantes.

En ce qui concerne les autres équations, on notera que $\frac{\partial g_j}{\partial m_j} = \frac{x - m_j}{\sigma_j^2} g_j$ et que les fonctions de la forme $(x - m_j) g_j^2$ sont impaires par rapport à l'ordonnée $x = m_j$ et ont par conséquent des intégrales nulles de $-\infty$ à $+\infty$.

Finalement, on peut arriver à déterminer les 8 paramètres de manière à rendre minimale la somme des carrés des écarts. Les valeurs de ces paramètres sont

$$\begin{array}{lll} \alpha = 29,4 \% & \beta = 15,8 \% & \gamma = 1 - \alpha - \beta = 54,8 \% \\ m_1 = 90,9 & m_2 = 100,25 & m_3 = 112,5 \\ \sigma_1 = 9,1 & \sigma_2 = 4,8 & \sigma_3 = 7,5 \end{array}$$

La somme des carrés des écarts est alors ramenée à 2 400.

IV. — APPLICATION DE LA THÉORIE A LA DÉTECTION D'OFFRES ANORMALES

Dans l'article qu'a publié la revue « Marchés publics », j'ai exposé que l'intérêt général commandait que les marchés de travaux soient traités au juste prix, c'est-à-dire à leur vraie valeur. J'y indiquais que si lors de chaque consultation, celle-ci pouvait faire appel à un très grand nombre d'entreprises, la vraie valeur serait égale à la moyenne des offres remises. Une consultation aussi large n'est ni possible, ni souhaitable : le maître de l'ouvrage à l'ouverture des plis ne dispose donc que d'un échantillon. Quelle conclusion peut-il en induire sur la vraie valeur des travaux à entreprendre?

IV.1. Deux cas sont à considérer :

- ou bien l'on n'admet pas que les éléments d'information recueillis par voie statistique sur la distribution des fréquences de la population-mère soient suffisants pour que l'on puisse en faire état dans la recherche de la vraie valeur;
- ou bien au contraire les études préalables ont été suffisamment complètes pour que l'on puisse avoir une bonne idée de la forme de la population-mère et par conséquent s'y référer.

Nous envisageons successivement les deux cas. Dans le premier, seules les caractéristiques des résultats de l'appel d'offres peuvent intervenir. On considérera alors que les offres remises constituent un petit échantillon d'une population-mère dont on cherche la moyenne. Soit m la moyenne cherchée, μ la moyenne de l'échantillon σ son écart-type. La moyenne m est encadrée par

$$\mu - \frac{ts}{\sqrt{n}} \leq m \leq \mu + \frac{ts}{\sqrt{n}}$$

s est la meilleure estimation de l'écart-type de la population : $\sigma \sqrt{\frac{n}{n-1}}$ et t est extrait de la table de Student en fonction du degré de confiance que l'on s'impose, avec $n - 1$ comme degrés de liberté.

Rappelons l'exemple donné dans l'article susvisé :

Entreprise	montant de l'offre %
—	—
A	60,18
B	81,47
C	81,60
D	87,61
E	91,16
F	92,03
moyenne	100,00
G	114,82
H	117,45
I	129,78
J	143,90

Ici on a $n = 10$ $\mu = 100$ $\sigma = 24,29 \%$ $s = 25,61 \%$.

La limite inférieure qui seule nous intéresse est donnée par $100 - 8,097 t$. La table de Student donne :

risque d'erreur unilatéral	t	$100 - 8,097 t$
—	—	—
1 %	2,821	77,16 %
0,5 %	3,250	73,68 %
0,1 %	4,297	65,21 %

La conclusion est donc qu'il y a 99 chances sur 100 pour que le juste prix soit supérieur aux 77 centièmes de la moyenne des offres reçues, 99,5 chances sur 100 pour qu'il soit supérieur à 74 centièmes de cette moyenne ou 999 chances sur mille pour qu'il soit supérieur à 65 centièmes. Autrement dit, l'offre moins-disante a vraiment très peu de chances de représenter la vraie valeur des travaux.

On peut d'ailleurs, sans plus de difficulté, raisonner en sens inverse et se demander quelle est la probabilité pour que l'offre A , ou l'offre B , soit à la limite inférieure de la vraie valeur des travaux. On détermine alors t par

$$t = \frac{100 - A}{8,097} \text{ ou } \frac{100 - B}{8,097}$$

De ces valeurs de t et par interpolation de la table on déduit la probabilité qui pour A est de 0,4 chances sur mille, pour B est de 2,4 chances sur 100.

IV.2. Il est évident que l'entrée en possession des éléments caractéristiques de la population donne aux estimations un degré de probabilité bien supérieur. Cela provient d'abord du fait que la connaissance de l'écart-type de la population réduit l'amplitude de la fourchette dans le rapport $\sqrt{\frac{n-1}{n}}$, ce qui est loin d'être négligeable lorsque n reste peu élevé. Mais cela n'est pas l'essentiel.

Je raisonnerai comme si l'étude statistique décrite dans la première partie possédait les qualités suffisantes pour être reconnue comme conduisant aux caractéristiques véritables de la population dont l'échantillon défini au paragraphe IV.1 est extrait. En partant de cet échantillon, nous avons évalué l'écart-type à introduire dans la formule de Student à $24,29 \sqrt{\frac{10}{9}} = 25,61$ %. Cette fois celui qui interviendra ne sera au plus que de 11,19 % ou même si l'on se réfère à la distribution des fréquences la plus voisine, à 7,83 %. Pour une même valeur de t la fourchette limitant la vraie valeur des travaux va se trouver très sensiblement réduite. Il faut remarquer cependant, avant d'entreprendre le calcul, que l'écart-type de la population-mère est connu comme une proportion de la moyenne m . Posons $s = \alpha m$. Les inégalités qui définissent la fourchette s'écrivent donc

$$\mu - t \frac{\alpha m}{\sqrt{n}} \leq m \leq \mu + t \frac{\alpha m}{\sqrt{n}}$$

de sorte que la limite inférieure de m s'écrit cette fois

$$\frac{\mu}{1 + \frac{t\alpha}{\sqrt{n}}} \leq m$$

Par exemple pour $n = 10$, $\alpha = 11,19$ % et pour $t = 4,297$ qui correspond à un risque de une chance sur mille, on trouve $m \geq 86,8$ %.

Si on décide de prendre pour écart-type de la population-mère la valeur 7,83 % qui minimise les valeurs extrêmes, on trouve $m \geq 90,38$.

On remarquera que la dispersion de l'échantillon des offres n'intervient pas dans le calcul. Les seuls éléments de l'appel d'offres qui interviennent sont sa moyenne (100 dans le cas particulier) et le nombre des offres n .

On ne trouve pas toujours dans les tables de Student qui sont publiées les valeurs de t correspondant à de petites valeurs du risque. Aussi ai-je cru devoir accompagner cet exposé d'une table dont l'argument est le risque unilatéral et où les valeurs du degré de liberté ont été limitées à 20 (ce qui correspond à $n = 21$).

IV.3. Toujours dans l'article publié par la revue « Marchés publics », j'ai indiqué que faute de disposer de renseignements statistiques suffisants, on pourrait dans une phase préliminaire, utiliser une règle de détection des offres anormales consistant à fixer le seuil à la moyenne des offres reçues diminuée de l'écart-type multiplié par

$$\sqrt{3 \frac{n-1}{n+1}}$$

Il y a lieu d'expliquer, sinon de justifier, cette suggestion.

Considérons n offres rangées dans l'ordre croissant $A_1, A_2, A_3, \dots, A_n$. Leur écart-type est égal à

$$\sigma = \sqrt{\frac{\sum A_i^2}{n} - \left(\frac{\sum A_i}{n}\right)^2}$$

Examinons le cas particulièrement simple où les offres seraient en progression arithmétique de raison r . On aura

$$\sum A_i = \sum [A_1 + (i-1)r] = nA_1 + r \sum (i-1)$$

$$\sum A_i^2 = \sum [A_1 + (i-1)r]^2 = nA_1^2 + 2A_1r \sum (i-1) + r^2 \sum (i-1)^2$$

or

$$\sum (i-1) = \frac{n(n-1)}{2} \quad \sum (i-1)^2 = \frac{n(n-1)(2n-1)}{6}$$

L'écart-type est donc égal à :

$$\sigma = r \sqrt{\frac{(n-1)(n+1)}{12}}$$

Or r est égal à $\frac{A_n - A_1}{n-1}$ et la moyenne est égale à $\frac{A_1 + A_n}{2}$. L'écart entre la moyenne et la plus petite offre est donc égal à $\frac{A_n - A_1}{2} = \frac{n-1}{2} r = \sigma \sqrt{3 \frac{n-1}{n+1}}$

Autrement dit lorsque les offres sont en progression arithmétique la plus basse coïncide avec la moyenne dont on a retranché $\sqrt{3 \frac{n-1}{n+1}}$ fois l'écart-type. L'institution d'une telle règle correspond à la volonté de considérer comme acceptable la plus basse d'une série en progression arithmétique.

Si l'on compare ce choix à celui qui s'inspire de la formule de Student on est amené à comparer

$$\frac{t}{\sqrt{n-1}} \text{ à } \sqrt{3 \frac{n-1}{n+1}}$$

autrement dit à comparer t à $(n-1) \sqrt{\frac{3}{n+1}}$

Cela permet de calculer le risque que l'on court en utilisant la formule sus-indiquée. Ce risque est fonction du nombre n d'offres de l'échantillon analysé. Il décroît rapidement avec n .

Ainsi pour $n = 3$, il y a 11 chances sur 100 pour que la vraie valeur soit inférieure au plancher

calculé; pour $n = 4$, 5 chances sur 100; pour $n = 5$: 2; pour $n = 6$: 1; pour $n = 7$: 1/2; pour $n = 8$: 1/4; pour $n = 9$: 1 chance sur mille; pour $n = 10$: 0,6 chance sur mille. A partir de $n = 11$, le risque d'erreur est tout à fait négligeable.

IV.4. On trouve, ci-après, 2 tables *A* et *B*.

L'une, table *A*, et celle de Student-Fisher donnant la valeur de t en fonction du nombre n d'offres reçues (et non du degré de liberté qui est égal à $n - 1$, dans le cas particulier) et du pourcentage de risque accepté. Le risque est le risque unilatéral. Elle a été limitée à 21 pour le nombre d'offres et aux petits risques (compris entre 0,05 % et 2,5 %).

La table *B* est la table inverse et permet de résoudre par interpolation plus commodément que par une recherche dans la table *A* la question posée au IV.1 (dernier alinéa).

Table *A*
Valeurs de t pour un nombre d'offres n et diverses valeurs
du risque unilatéral P exprimé en %

$n \backslash P$	1 %	0,5 %	0,1 %	0,05 %
2	31,8	63,7	318,3	636,6
3	6,97	9,93	22,3	31,6
4	4,54	5,84	10,2	12,9
5	3,75	4,60	7,17	8,61
6	3,37	4,03	5,89	6,87
7	3,14	3,71	5,21	5,96
8	3,00	3,50	4,79	5,41
9	2,90	3,36	4,50	5,04
10	2,82	3,25	4,30	4,78
11	2,76	3,17	4,14	4,59
12	2,72	3,11	4,03	4,44
13	2,68	3,06	3,93	4,32
14	2,65	3,01	3,85	4,22
15	2,62	2,98	3,79	4,14
16	2,60	2,95	3,73	4,07
17	2,58	2,92	3,69	4,02
18	2,57	2,90	3,65	3,97
19	2,55	2,88	3,61	3,92
20	2,54	2,86	3,58	3,88
21	2,53	2,85	3,55	3,85

Table B

Pourcentage de risque unilatéral correspondant à diverses valeurs de t et pour un nombre n d'offres

$t \backslash n$	1,5	2	2,5	3	3,5	4	4,5	5	6	7	8	9	10	12	15	20	30	50
3	13,62	9,18	6,48	4,77	3,64	2,86	2,30	1,89	1,33	0,990	0,763	0,606	0,492	0,344	0,221	0,125	0,056	0,020
4	11,53	6,97	4,39	2,88	1,97	1,40	1,03	0,770	0,464	0,299	0,204	0,145	0,106	0,062	0,032	0,014		
5	10,40	5,81	3,34	2,00	1,25	0,807	0,541	0,375	0,194	0,110	0,066	0,042	0,028	0,014				
6	9,70	5,10	2,73	1,51	0,864	0,516	0,320	0,205	0,092	0,046	0,025	0,014						
7	9,21	4,62	2,33	1,20	0,641	0,356	0,205	0,123	0,048	0,021	0,010							
8	8,86	4,28	2,05	1,09	0,500	0,259	0,140	0,078	0,027	0,011								
9	8,60	4,03	1,85	0,854	0,404	0,197	0,100	0,053	0,016									
10	8,39	3,83	1,69	0,748	0,336	0,156	0,074	0,037	0,010									
11	8,23	3,67	1,57	0,667	0,286	0,126	0,057	0,027										
12	8,09	3,54	1,48	0,604	0,249	0,104	0,045	0,021										
13	7,97	3,43	1,40	0,553	0,219	0,088	0,036	0,016										
14	7,88	3,34	1,33	0,512	0,196	0,076	0,030	0,012										
15	7,79	3,26	1,27	0,478	0,177	0,066	0,025	0,010										
16	7,72	3,20	1,23	0,449	0,161	0,058	0,021											
17	7,65	3,14	1,18	0,424	0,148	0,052	0,018											
18	7,60	3,09	1,15	0,403	0,137	0,046	0,016											
19	7,55	3,04	1,12	0,384	0,128	0,042	0,014											
20	7,50	3,00	1,09	0,368	0,120	0,038	0,012											
21	7,46	2,96	1,06	0,354	0,113	0,035	0,011											

Nota : Cette table donne en principe trois chiffres significatifs, mais pas plus de trois décimales. Les valeurs inférieures à 0,01 % n'ont pas été notées.