

UTILISATION D'UNE ECHELLE FONCTIONNELLE EMPIRIQUE POUR L'ANALYSE
DE DISTRIBUTIONS POLYMODALES. APPLICATION A UN PROBLEME DE
RADIOBIOLOGIE.

par

A. SILVY

Service de Radioagronomie -
Commissariat à l'Energie Atomique (1)

et

J. DEJARDIN

Service de Biométrie -
Office de la Recherche Scientifique et Technique Outre-Mer (2)

RESUME : A propos d'un problème de radiobiologie on décrit une méthode graphique de décomposition des distributions polymodales dont la population de référence n'est pas gaussienne. Cette méthode utilise une échelle fonctionnelle, tracée de façon empirique, qui jouit des mêmes propriétés vis-à-vis de la population témoin que l'échelle gaussienne vis-à-vis de la distribution normale.

Le développement des tests statistiques basés sur la loi normale résulte des propriétés mathématiques de sa formulation. En particulier les méthodes qui permettent de décomposer les distributions polymodales se rapportent toujours à la loi normale, que ce soient les méthodes basées sur le calcul ou les méthodes graphiques qui utilisent le papier à échelle gaussienne (BHATTACHARYA, 1967; CASSIE, 1954; HARDING, 1949; LEWIS *et al.*, 1967; TANAKA, 1962). Sur ce papier, les fréquences relatives cumulées d'une distribution normale sont représentées par une droite tandis qu'une distribution bimodale, mélange de deux populations normales, donne une courbe sigmoïde. L'ordonnée du point d'inflexion correspond aux proportions des deux populations composantes dont les caractéristiques, effectif - moyenne - variance, peuvent être estimées après décomposition de la courbe sigmoïde en deux droites. Cependant ces méthodes sont inutilisables quand les populations de référence ne sont pas gaussiennes ou ne peuvent pas être rendues gaussiennes par une transformation des données. Le travail présenté ici concerne ce problème de décomposition. Il est applicable à de nombreux caractères quantitatifs étudiés en biologie qui, bien souvent, ne sont pas distribués normalement.

(1) B.P. 1, F-13115 St-Paul-lez-Durance.

(2) 70-74, Route d'Aulnay F - 93140 Bondy.

MATERIEL

Dans les recherches de radiobiologie effectuées sur grain d'Orge, on a l'habitude d'utiliser la longueur de la première feuille des plantules comme critère de sensibilité. Suivant les conditions de traitement, on observe une réduction générale de longueur sans modification de la forme de l'histogramme ou un partage en deux sous-populations qui évoluent indépendamment l'une de l'autre en fonction de la dose d'exposition (SILVY). La perte de signification de la moyenne, par suite de l'étalement ou même du partage de la population traitée, nous a conduit à étudier les différentes méthodes permettant l'analyse de telles distributions polymodales.

La figure 1 représente les distributions des longueurs de la première feuille en pourcentage du nombre de grains germés après exposition de caryopses d'Orge, variété Piroline, au rayonnement gamma du ^{60}Co en présence d'air et culture en serre. Les doses d'exposition sont 0, 10 et 20 kR. Les feuilles sont mesurées en fin de croissance et les mesures sont groupées en classes d'amplitude 1 cm. L'essai est conduit en 20 blocs complets randomisés. Chaque parcelle élémentaire étant constituée de 50 grains, les effectifs totaux théoriques sont donc de 1000 grains par traitement. L'exposition à des doses de 10 et 20 kR ne modifie pas de façon significative le pourcentage de germination, mais les histogrammes sont différents. La distribution correspondant à 20 kR est bimodale et on désire savoir si on peut déjà mettre en évidence les deux sous-populations après une exposition à 10 kR. Les plantules les plus petites ne survivent pas après un traitement à 20 kR, la décomposition de la distribution 10 kR permettrait d'étudier le phénomène responsable du partage.

En première approximation, les allures des diagrammes probit (fig. 2), comme les coefficients de Fisher du témoin et de la première dose (fig. 1), paraissent semblables. La concavité tournée vers le haut ainsi que la valeur négative du coefficient g_1 sont les caractéristiques de distribution dissymétriques dont le mode est déporté vers la droite. D'autre part la forme en S inversé, correspondant à 80% de l'effectif, et la valeur positive de g_2 indiquent les distributions plus aplaties que la normale (DAGNELIE, 1969). L'ajustement à différents modèles a été tenté sans résultat.

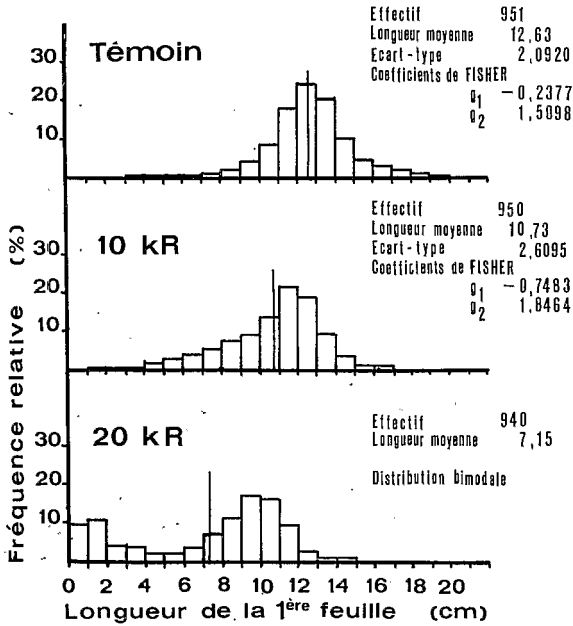


Figure 1. - Distribution de la longueur de la première feuille après exposition de grains d'Orge au rayonnement γ du ^{60}Co .

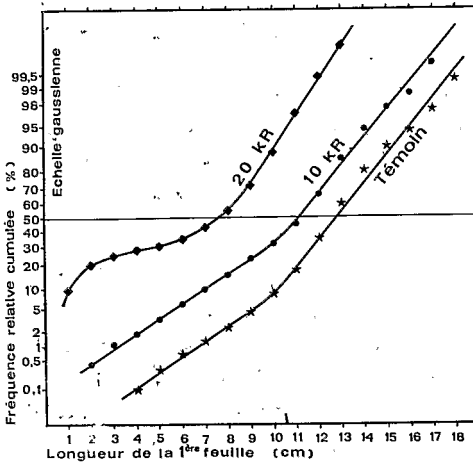


Figure 2. - Diagramme probit.

ETUDE DE LA DISTRIBUTION DE REFERENCE

Les grains d'Orge proviennent de sélections particulièrement surveillées pour leur pureté génétique. Ils sont triés par calibrage puis stockés dans des dessiccateurs en présence d'un mélange de glycérine et d'eau pour être amenés à une teneur en eau déterminée, 9,2% par rapport au poids frais, avant la radioexposition. Le lot initial de Pirolina a été ainsi divisé en sept parties dont les poids de 100 grains s'échelonnent de 3,12 à 5,36 g (tableau II). Pour l'essai de la figure 1, on a utilisé le calibre 5. Le diagramme probit du témoin (fig. 2) n'est pas linéaire mais les points d'inflexion n'y sont pas suffisamment marqués pour qu'il soit possible d'utiliser une méthode graphique d'analyse. La méthode de BHATTACHARYA (1967) permet tout de même de le décomposer de proche en proche en quatre sous-populations normales (fig. 3 et tableau I). Cependant les proportions $9 + 82 + 7 + 2 = 100$ ne correspondent pas au partage $70 + 30 = 100$ observé après une exposition à 20 kR. De plus, une telle décomposition, d'ordre plus physiologique que génétique quand on a affaire à une lignée pure, paraît difficilement interprétable et utilisable.

Les calibres 1 - 4 - 7 étudiés en essai comparatif ainsi que des échantillons d'autres variétés donnent, comme le calibre 5, des coefficients de Fisher de signes contraires, $g_1 < 0$ et $g_2 > 0$ (tableau II). Devant la stabilité de ces résultats, la décomposition de la population témoin en quatre sous-populations paraît assez artificielle et il semble plus logique de la considérer globalement comme une distribution unique en admettant que le caractère "longueur de la première feuille" n'est pas distribué suivant une loi normale, ni un mélange de lois normales.

Nous avons d'ailleurs observé le même type de distribution chez le maïs et le riz.

ETABLISSEMENT D'UNE ECHELLE FONCTIONNELLE EMPIRIQUE

Alors que les méthodes basées sur le calcul nécessitent de connaître la formulation mathématique de la loi de distribution, la méthode graphique par les points d'inflexion est utilisable pour n'importe quelle sorte de distribution à condition de disposer d'une échelle fonctionnelle ayant pour cette distribution

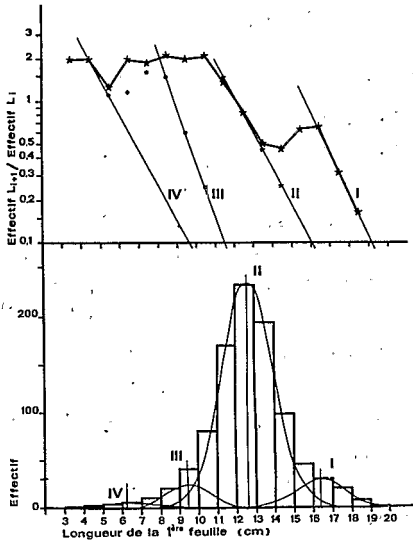


Figure 3. - Décomposition de la distribution témoin en composants gaussiens par la méthode Battacharya.

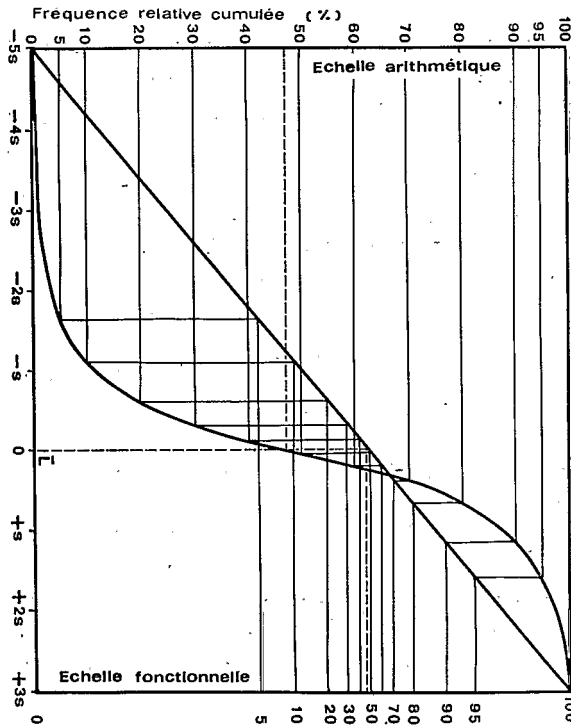


Figure 4. - Tracé d'une échelle fonctionnelle.

Tableau II

Décomposition de la distribution témoin en composantes Gaussiennes

	Effectif		Longueur moyenne (cm)	Ecart type	Coefficients de Fisher	
	Nombre	%			g_{11}	g_{22}
TOTAL	9511	1000	122,633	2,09200	-0,23777	1,50988
II	888	99	169,377	1,20599	0,00355	3,00655
III	7799	822	122,622	1,32111	0,00083	2,97066
IIII	688	77	99,455	1,05200	0,00888	2,91711
IVV	166	22	66,288	1,28277	-0,12155	2,43066

Test de KOLMOGOROV et SMIRNOV: Non significatif.

Maximum des valeurs absolues des différences :: 0,377 %

Valeur critique approximative à 55 %

:: 2,877 %

Tableau III

Caractéristiques des plusieurs échantillons d'Orge

VARIÉTÉ				Effectif	Longueur moyenne (cm)	Ecart type	Coefficients de Fisher	
calibre	Poids total (g)	Poids 100 grains (g)	g_{11}				g_{22}	
Pivroline	1	477	33,122	4699	112,29	2,5718	-0,2469	1,6521
	2	1455	33,444					
	3	2233	33,783					
	4	3333	44,288	4866	112,99	2,4552	-0,3921	2,5790
	5	3400	44,799					
	6	1555	55,122					
	7	611	55,366					
Himalaya	n°1 (Cadache)			974	112,511	1,7066	-2,0232	8,8846
	n°2 (Euratom)			464	111,35	2,2674	-1,1098	2,2288
Rika	(Sté Desprez)			4911	110,29	1,2042	-1,3378	6,7897
Ingrid	(Sté Desprez)			4777	112,55	1,3378	-2,0290	6,3823
Bérénice	(I N R A)			496	15,68	1,2703	-1,2322	4,4834

la même propriété que le papier probit pour la distribution normale. Cette échelle peut être tracée assez simplement de façon empirique en projetant parallèlement à l'axe des y sur la droite reliant 0% à 100% les différents points de la courbe des fréquences relatives cumulées, sigmoïde en coordonnées arithmétiques (fig. 4).

Plutôt que de refaire ce travail pour chaque essai, nous avons construit une courbe moyenne à partir des témoins de différents essais effectués avec le même stock de grains. Comme la longueur moyenne et l'étalement de la distribution varient légèrement d'un essai à un autre en fonction des conditions de traitement ou de culture et des caractéristiques du grain, la distribution moyenne a été calculée par l'intermédiaire des fréquences relatives cumulées lues sur les graphiques de chaque essai en prenant la longueur moyenne comme origine et le 1/4 de l'écart-type comme unité sur l'axe des x. C'est cette courbe sigmoïde qui a été ensuite transformée en droite donnant l'échelle fonctionnelle figurant sur le côté droit de la figure 4. Elle a été tracée à partir des mesures individuelles de 13 738 plantules provenant des témoins de 13 essais. Il faut noter que la distribution étant dissymétrique, la longueur moyenne doit être lue au niveau 47,5% de la fréquence relative cumulée qui correspond sensiblement au point d'inflexion de la sigmoïde.

DECOMPOSITION DES DISTRIBUTIONS POLYMODALES

Sur la figure 5, qui utilise l'échelle fonctionnelle, la courbe correspondant à la dose 20 kR a la même allure que sur le papier probit (fig. 2). En revanche, les courbes du témoin et du traitement 10 kR sont différentes non seulement des précédentes mais aussi entre elles : alors que le témoin paraît assimilable à une droite, la première dose donne déjà une courbe sigmoïde caractéristique d'une distribution bimodale.

La décomposition des distributions polymodales s'effectue par la même méthode que pour le papier probit (HARDING, 1949; LEWIS *et al.*, 1967). Dans le cas de la distribution bimodale correspondant à la dose 10 kR, le point d'inflexion de la courbe sigmoïde peut être situé par approximations successives dans la classe 9,5 cm au point partageant l'effectif total dans les proportions qui permettent le meilleur alignement sur des droites représentatives des deux sous-populations (fig. 6). La validité

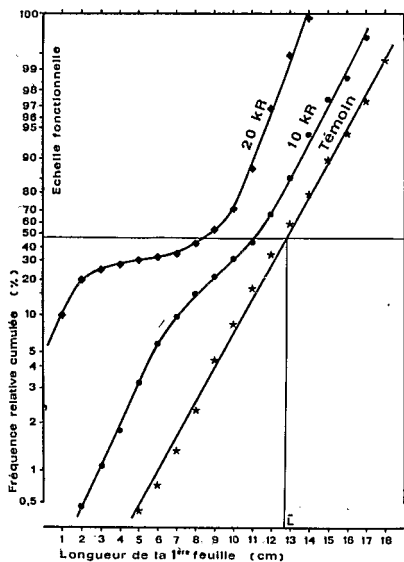


Figure 5. - Diagramme fonctionnel.

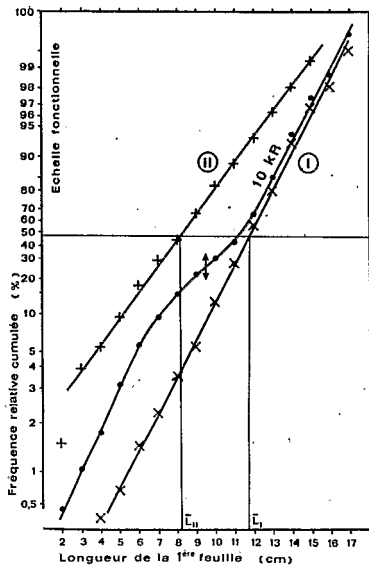


Figure 6. - Décomposition d'une distribution bimodale.

Tableau III

Décomposition de la distribution correspondant au traitement 10 kR

Centre des classes	Fréquence observée	Fréquence calculée			Fréquence relative cumulée (%)		
		I	II	Total	observée	calculée	$ F_o - F_c $
1,5	4		5,8	5,8	0,42	0,73	0,31
2,5	6		2,4	2,4	1,05	1,02	0,03
3,5	7		4,4	4,4	1,79	1,57	0,22
4,5	13	4,8	8,1	12,9	3,16	3,08	0,08
5,5	26	4,8	12,0	16,8	5,89	5,08	0,81
6,5	35	6,2	26,2	32,4	9,58	9,00	0,58
7,5	50	8,3	39,2	47,5	14,84	14,76	0,08
8,5	72	19,4	46,9	66,3	22,42	22,64	0,22
9,5	86	42,9	31,6	74,5	31,47	31,09	0,38
10,5	128	107,3	15,3	122,6	44,95	44,29	0,66
11,5	205	193,8	12,0	205,8	66,53	66,19	0,34
12,5	177	173,0	6,5	179,3	85,16	85,21	0,05
13,5	86	70,6	3,7	74,3	94,21	93,11	1,10
14,5	31	34,6	2,4	37,0	97,47	97,04	0,43
15,5	11	16,6	0,9	17,5	98,63	98,89	0,26
16,5	10	7,6	0,9	8,2	99,68	99,78	0,10
17,5	3	2,1		2,1	100	100,00	0,00
Total	950	692	258	950			
Longueur	10,73	11,73	8,21	10,44			

Test de KOLMOGOROV et SMIRNOV : Non significatif.

Maximum de la valeur absolue de la différence : 1,10%

Valeur critique approximative à 5% : 2,87%.

Tableau IV

Analyse de l'essai

Traite- ment	Total		Sous-population I			Sous-population II		
	Effec- tif	Longueur moyenne (cm)	Effectif Nombre	%	Longueur moyenne (cm)	Effectif Nombre	%	Longueur moyenne (cm)
0	951	12,63						
10 kR	950	10,73	692	72,8	11,7	258	27,2	8,2
20 kR	940	7,15	663	70,5	9,6	277	29,5	1,6

de l'ajustement est confirmée par le test de Kolmogorov et Smirnov (DAGNELIE, 1969) (tableau III). Les longueurs moyennes peuvent être calculées à partir de chaque distribution ou lues sur les droites I et II.

Après avoir estimé également les caractéristiques des deux sous-populations correspondant à la dose 20 kR, on constate que la radioexposition provoque dès 10 kR un partage dans des proportions qui restent constantes quand la dose augmente, tandis que la longueur moyenne diminue (tableau IV).

CONCLUSIONS

Plusieurs conclusions peuvent être présentées à partir de ces résultats. Du point de vue biologique, on constate que la forme non gaussienne de la distribution de la longueur de la première feuille n'est pas modifiée quand on effectue un calibrage du grain, des modifications de teneur en eau, et même une radioexposition aux neutrons ou au rayonnement gamma en présence d'azote (SILVY). De plus, après traitement au rayonnement gamma en présence d'air, les composantes de la distribution bimodale sont du même type que la distribution de référence: les points de I en II peuvent être considérés comme bien alignés. Il semble donc que cette distribution, considérée comme unique au départ, a une réalité biologique qui permet de caractériser un lot de graines. Toutefois, des variations ont été observées entre des variétés ou entre des lots de la même variété provenant d'origines différentes.

Du point de vue statistique, alors que l'approximation à la loi normale ne permet pas de mettre en évidence de différence entre le témoin et la dose 10 kR, la construction d'une échelle fonctionnelle empirique à partir de la population témoin nous a permis de montrer l'hétérogénéité de la distribution "10 kR" et l'influence de l'augmentation de la dose d'exposition. Cette méthode d'analyse, de réalisation très simple, peut être utilisée chaque fois que la distribution de référence n'est pas conforme à un modèle théorique pour lequel un papier fonctionnel a été dessiné, cas du modèle gaussien dans l'exemple que nous avons traité. Il faut toutefois signaler que pour établir une échelle de référence dans nos conditions expérimentales, nous devons disposer d'échantillons comprenant au moins 500 plantules.

SUMMARY : With regard to a radiobiological problem a graphic method of decomposition of polymodal distributions with not gaussian reference population is described. The functional scale which is used in this method is empirically plotted and presents the same properties with respect to the control population as the gaussian scale does with respect to the normal distribution.

BIBLIOGRAPHIE

- BHATTACHARYA C.G. - A simple method of resolution of a distribution into gaussian components. *Biometrics*, 23, 1967, p. 115-135.
- CASSIE R.M. - Some uses of probability paper in the analysis of size frequency distributions. *Aust. J. Mar. Freshw. Res.* 5, 1954, p. 513-522.
- DAGNELIE P. - *Théorie et Méthodes statistiques*. Ed. J. Duculot, Gembloux, 1969, Vol. 1 et 2.
- HARDING J.P. - The use of probability paper for the graphical analysis of polymodal frequency distributions. *J. Mar. Biol. Ass. U.K.* 28, 1949, p. 141-153.
- LEWIS et TAYLOR - *Introduction to experimental ecology*. Academic Press, London 1967, p. 63-66.
- SILVY A. - Mise en évidence de deux types de réaction après exposition de grains d'orge au rayonnement gamma du ^{60}Co en présence d'air. *C.R. Acad. Sci. Fr., Série D.* 264, 1975, p. 799-802.
- TANAKA S. - A method of analysis a polymodal frequency distribution and its application to the length distribution of the Porgy, Tauius tumifrons (T and S). *J. Fish. Res. Bd. Can.* 19, 1962, p. 1143-1159.