ANNALES DE L'I. H. P.

EUGENE LUKACS

Certains tests indépendants de la distribution initiale

Annales de l'I. H. P., tome 15, nº 4 (1957), p. 252-265
<http://www.numdam.org/item?id=AIHP_1957__15_4_252_0>

Certains tests indépendants de la distribution initiale

par

Eugene LUKACS (1),

The Catholic University of America.

1. **Introduction.** — Désignons par Ω une famille des fonctions de répartition et considérons une population statistique dont la fonction de répartition est un élément de Ω. Soit $X_1, X_2, \ldots, X_n$ un échantillon tiré d'une population dont la fonction de répartition $F(x)$ appartient à Ω.

Nous introduisons pour chaque n-uple $X_1, X_2, \ldots, X_n$ de nombres réels et pour chaque $F(x) \in \Omega$ l'expression

$$S = S(X_1, X_2, \ldots, X_n \mid F)$$

et nous supposons que S est pour chaque $F \in \Omega$ une fonction mesurable des $X_1, \ldots, X_n$ et une fonctionnelle en F. L'expression $S(X_1, \ldots, X_n \mid F)$ s'appelle alors une statistique en Ω.

Nous disons qu'une statistique $S = S(X_1, \ldots, X_n \mid F)$ en Ω est indépendante de la distribution initiale, si la probabilité que $S \leqq s$, pourvu que $F(x)$ soit la fonction de répartition de la population, est indépendante de $F(x)$ pour chaque $F \in \Omega$, c'est-à-dire si la probabilité

$$P\{S(X_1, X_2, \ldots, X_n \mid F) \leqq s; F\} = \varphi(s) \qquad (F \in \Omega)$$

est une fonction seulement de s.

Remarquons que cette propriété d'indépendance de la distribution initiale est toujours relative à une famille de fonctions de répartitions.

Dans la plus grande partie des applications on se sert de statistiques qui ne dépendent pas de la fonction $F(x)$. Nous appellerons de telles

---

(1) Recherche subventionnée par l'Office of Naval Research et aussi par la National Science Foundation (NSF. G 2781 et NSF. G 4220).

statistiques des statistiques de partition. On peut expliquer cette terminologie par la considération suivante. L'échantillon $(X_1, X_2, \ldots, X_n)$ peut être envisagé comme un point dans un espace euclidien E_n à n dimensions. Soit $S(X_1, \ldots, X_n)$ une statistique de partition. L'équation $S(X_1, \ldots, X_n) = c$ (où c est un paramètre réel) détermine alors une famille des surfaces qui partagent l'espace statistique E_n.

Z. W. Birnbaum [1] a introduit le concept de statistique indépendante de la distribution initiale que nous avons présenté plus haut et il a aussi proposé (²) le problème de déterminer les statistiques de partition qui sont indépendantes de la distribution initiale dans la famille des fonctions de répartition laplaciennes.

Dans la première partie de cet article nous examinerons la famille des répartitions laplaciennes et nous donnerons une caractérisation des statistiques de partition qui sont indépendantes de la distribution initiale. Dans la seconde partie nous nous occuperons de la famille des fonctions de répartition à rang fini. Cette famille est l'objet de recherches récentes de E. B. Dynkin [2], [3]. Notre résultat sera établi à l'aide d'un lemme qui fut obtenu dans un autre Mémoire [4].

2. Le lemme. — Soient X et Y deux variables aléatoires et désignons par $\mathbf{E}(Y \mid X)$ l'espérance conditionnelle de Y par rapport à X et par $\mathbf{E}(Y)$ l'espérance mathématique de Y. En général, les expectations $\mathbf{E}(Y \mid X)$ et $\mathbf{E}(Y)$ seront différentes. Nous introduisons la définition suivante :

Soit Y une variable aléatoire dont l'espérance mathématique est finie, nous disons que Y a une régression constante en X, si la relation

$$\mathbf{E}(Y \mid X) = \mathbf{E}(Y)$$

est satisfaite presque partout.

On peut maintenant formuler le lemme.

Lemme 1. — *Soient* X *et* Y *deux variables aléatoires et supposons que l'espérance mathématique de* Y *existe. Pour que* Y *ait la régression constante en* X, *il faut et il suffit que la relation*

$$\mathbf{E}(e^{itX}\, Y) = \mathbf{E}(e^{itX})\, \mathbf{E}(Y)$$

soit satisfaite pour toute valeur réelle de t.

(²) Dans une conversation.

La nécessité de la condition du lemme est immédiate, pour la démonstration qu'elle est aussi suffisante nous renvoyons le lecteur à [4].

3. **La famille des fonctions de répartitions laplaciennes.** — Nous supposons maintenant que la famille Ω est la famille des fonctions de répartitions laplaciennes, c'est-à-dire que Ω est la famille des fonctions

$$F(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x}\exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\}dy. \tag{3.1}$$

Chaque membre de cette famille est déterminé par les valeurs numériques des paramètres μ et σ et nous écrivons $\Omega=N(\mu,\sigma)=N$. On peut aussi considérer les sous-familles de N obtenues en fixant ou bien μ ou bien σ. Nous désignons par N^{σ_0} la famille des distributions laplaciennes avec moyenne arbitraire qui ont toutes la même variance σ_0^2. De même façon nous écrivons N_{μ_0} pour la sous-famille des répartitions laplaciennes dont la moyenne μ_0 est constante.

Soit $S=S(X_1, X_2, \ldots, X_n)$ une statistique de partition pour une population laplacienne. La fonction caractéristique de S est alors

$$\begin{aligned} f(t;\mu,\sigma)=\left(\frac{1}{\sqrt{2\pi}}\right)^{n}\sigma^{-n}\int_{-\infty}^{\infty}\cdots \\ \times\int_{-\infty}^{\infty}\exp\left[it\,S(x_1,\ldots,x_n)-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j-\mu)^2\right]dx_1\ldots dx_n. \end{aligned} \tag{3.2}$$

Supposons maintenant que le premier moment $\mathbf{E}(S)$ de la statistique S existe pour toute fonction de répartition de la famille N.

Pour que S soit indépendante de la distribution initiale, il faut et il suffit que

$$\frac{\partial f(t;\mu,\sigma)}{\partial\mu}=0, \tag{3.3.1}$$

$$\frac{f(t;\mu,\sigma)}{\partial\sigma}=0. \tag{3.3.2}$$

Chacune de ces équations a aussi un sens. La première signifie que S est indépendante de la distribution initiale en N^{σ} tandis que la seconde indique que la même chose est vraie pour la famille N_{μ}.

On voit aisément qu'il est permis de différentier l'expression $f(t;\mu,\sigma)$

sous l'intégrale et l'on obtient ainsi de (3.3.1)

$$\frac{\partial f(t;\mu,\sigma)}{\partial\mu}=\left(\frac{1}{\sqrt{2\pi}}\right)^n\sigma^{-n}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial}{\partial\mu}\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j-\mu)^2\right]$$
$$\times\exp\left[it\mathrm{S}(x_1,\ldots,x_n)-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j-\mu)^2\right]dx_1\ldots dx_n=0.$$

En observant que

$$\frac{\partial}{\partial\mu}\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j-\mu)^2\right]=\frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j-\mu),$$

nous transformons l'équation précédente en

$$\mathbf{E}\{(\overline{\mathrm{X}}-\mu)\,e^{it\mathrm{S}}\}=0.$$

Ici $\overline{\mathrm{X}}=\frac{1}{n}\sum_{j=1}^{n}\mathrm{X}_j$ et, par conséquent, $\mathbf{E}\overline{\mathrm{X}}=\mu$. Ainsi

$$\mathbf{E}(\overline{\mathrm{X}}\,e^{it\mathrm{S}})=\mathbf{E}(\overline{\mathrm{X}})\,\mathbf{E}(e^{it\mathrm{S}}). \tag{3.4.1}$$

De la même façon on obtient à partir de la condition (3.3.2) l'équation

$$\mathbf{E}\left\{e^{it\mathrm{S}}\sum_{j=1}^{n}(\mathrm{X}_j-\mu)^2\right\}=\mathbf{E}\left\{\sum_{j=1}^{n}(\mathrm{X}_j-\mu)^2\right\}\mathbf{E}(e^{it\mathrm{S}}). \tag{3.4.2}$$

Maintenant il est déjà possible de formuler nos premiers résultats.

Théorème 1 *a*. — *Supposons que* S *soit une statistique de partition telle que* $\mathbf{E}(\mathrm{S})$ *existe. Pour que* S *soit indépendante de la distribution initiale en* N^σ *il faut et il suffit que la moyenne* $\overline{\mathrm{X}}$ *de l'échantillon ait la régression constante en* S.

Théorème 1 *b*. — *Supposons que* S *soit une statistique de partition telle que* $\mathbf{E}(\mathrm{S})$ *existe. Pour que* S *soit indépendante de la distribution initiale en* N_μ *il faut et il suffit que* $\sum_{j=1}^{n}(\mathrm{X}_j-\mu)^2$ *ait la régression constante en* S.

La nécessité de ces conditions est une conséquence des équations (3.4.1) et (3.4.2) et du lemme. Réciproquement, supposons que

$\bar{X}\left[\text{ou } \sum_{j=1}^{n}(X_j-\mu)^2\right]$ ait la régression constante en S. On déduit alors du lemme que la relation (3.4.1) [ou (3.4.2)] est valide. Puis il faut retracer le raisonnement pour obtenir (3.3.1) et (3.3.2).

On peut modifier les conditions du théorème si la statistique S est indépendante de la distribution initiale en N. Alors (3.4.1) et (3.4.2) sont valides simultanément. Développant le carré dans (3.4.2) et utilisant (3.4.1), nous obtenons aisément

$$(3.4.3) \qquad \mathbf{E}\left[e^{itS}\sum_{j=1}^{n}X_j^2\right] = n(\sigma^2+\mu^2)\,\mathbf{E}(e^{itS}) = \mathbf{E}\left(\sum_{j=1}^{n}X_j^2\right)\mathbf{E}(e^{itS}).$$

Mais cela veut dire que nous avons établi le résultat suivant :

Théorème 2. — *Soit* S *une statistique de partition telle que* $\mathbf{E}(S)$ *existe. Pour que* S *soit indépendante de la distribution initiale en* N, *il faut et il suffit que* $\sum_{j=1}^{n}X_j$ *aussi bien que* $\sum_{j=1}^{n}X_j^2$ *aient la régression constante en* S.

Il nous reste encore à démontrer qu'il existe des statistiques de partition qui sont indépendantes de la distribution initiale en N. Un résultat un peu plus général nous permettra de construire de telles statistiques.

Considérons une famille de fonctions de répartition absolument continues qui dépend d'un paramètre de position μ et d'un paramètre d'échelle $\sigma \neq 0$. Désignons cette famille par $L = L(\mu, \sigma)$. Chaque répartition $G(x; \mu, \sigma)$ de L a alors la forme

$$G(x; \mu, \sigma) = \frac{1}{\sigma}\int_{-\infty}^{x} F'\left(\frac{t-\mu}{\sigma}\right) dt,$$

où $F(y)$ est une fonction de répartition absolument continue. Ainsi $L(\mu, \sigma)$ est la famille des fonctions de répartition du type de $F(x)$.

La famille $N = N(\mu, \sigma)$ des répartitions laplaciennes est une telle famille, avec

$$F'(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}.$$

Théorème 3. — *Soit* $L = L(\mu, \sigma)$ *une famille de fonctions de répar-*

tition absolument continues qui dépend d'un paramètre de position μ *et d'un paramètre d'échelle* $\sigma \neq 0$. *La condition nécessaire et suffisante pour que la statistique de partition* $S(X_1, \ldots, X_n)$ *soit indépendante de la distribution initiale en* L *est que les deux statistiques* $S(X_1, \ldots, X_n)$ *et* $S(aX_1+b, aX_2+b, \ldots, aX_n+b)$ *aient la même fonction de répartition quel que soit* $a \neq 0$ *et* b *(*a, b *réels).*

Soit $\mu^\star = a\mu + b$ et $\sigma^\star = a\sigma$. Nous déterminons les fonctions caractéristiques $f(t; \mu, \sigma)$ et $f(t; \mu^\star, \sigma^\star)$ de S pour les valeurs (μ, σ) et $(\mu^\star, \sigma^\star)$ des paramètres. On obtient

$$f(t; \mu, \sigma) = \frac{1}{\sigma^n}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp[it\,S(x_1, \ldots, x_n)] \times F'\left(\frac{x_1-\mu}{\sigma}\right)\ldots F'\left(\frac{x_n-\mu}{\sigma}\right)dx_1\ldots dx_n \tag{3.5}$$

et

$$f(t; \mu^\star, \sigma^\star) = \frac{1}{(\sigma^\star)^n}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp[it\,S(y_1, \ldots, y_n)] \times F'\left(\frac{y_1-\mu^\star}{\sigma^\star}\right)\ldots F'\left(\frac{y_n-\mu^\star}{\sigma^\star}\right)dy_1\ldots dy_n.$$

Nous substituons dans la seconde intégrale $y_j = ax_j + b$, alors

$$\frac{y_j-\mu^\star}{\sigma^\star} = \frac{x_j-\mu}{\sigma}$$

et

$$f(t; \mu^\star, \sigma^\star) = \frac{1}{\sigma^n}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp[it\,S(ax_1+b, \ldots, ax_n+b)] \times F'\left(\frac{x_1-\mu}{\sigma}\right)\ldots F'\left(\frac{x_n-\mu}{\sigma}\right)dx_1\ldots dx_n.$$

Soit maintenant S une statistique de partition qui est indépendante de la distribution initiale en L. Alors pour chaque a et b, et par conséquent pour chaque $\mu^\star$ et $\sigma^\star$, on a

$$f(t; \mu, \sigma) = f(t; \mu^\star, \sigma^\star).$$

Mais cela veut dire que les statistiques $S(X_1, \ldots, X_n)$ et $S(aX_1+b, \ldots, aX_n+b)$ ont la même fonction de répartition, pourvu que la fonction de répartition de la population soit un membre de L. La réciproque est ainsi démontrée aisément : Si les statistiques $S(X_1, X_2, \ldots, X_n)$ et $S(aX_1+b, \ldots, aX_n+b)$ ont la même fonction de répartition, alors S est indépendante de la distribution initiale.

Le théorème 3 permet la construction des statistiques de partition qui sont indépendantes de la distribution initiale en L. Soit, par exemple, $S(x_1, \ldots, x_n)$ une fonction qui est invariante relativement à des transformations linéaires des variables, c'est-à-dire

$$S(ax_1 + b, ax_2 + b, \ldots, ax_n + b) = S(x_1, x_2, \ldots, x_n) \tag{3.6}$$

pour chaque $a \neq 0$ et b. Alors S satisfait à la condition du théorème 3 et est indépendante de la distribution initiale en L et l'on obtient le corollaire suivant :

COROLLAIRE 1 DU THÉORÈME 3. — *Soit* L *une famille de répartitions absolument continues qui dépend d'un paramètre de position et d'un paramètre d'échelle. Supposons que la statistique de partition* S *est invariante relativement à des transformations linéaires. Alors* S *est indépendante de la distribution initiale en* L.

Il est immédiat qu'on peut établir des résultats analogues pour les familles L^{σ}, dépendant seulement d'un paramètre de position et aussi pour les familles L_{μ} dépendant d'un paramètre d'échelle.

COROLLAIRE 2 DU THÉORÈME 3. — *Une statistique* S *qui est invariante relativement à des translations [transformations d'échelle] est indépendante de la distribution initiale pour chaque famille qui dépend seulement d'un paramètre de position [d'échelle].*

4. Les fonctions de répartition à rang fini. — La théorie de l'estimation forme une partie essentielle de la statistique mathématique. R. A. Fisher a introduit dans cette théorie une idée importante et bien utile : celle des statistiques exhaustives. Mais de telles statistiques n'existent pas toujours. G. Darmois [5] et D. Dugué [6] ont démontré que des statistiques exhaustives existent seulement pour des familles spéciales de fonctions de répartition. Récemment E. B. Dynkin [2], [3] a généralisé la définition des statistiques exhaustives et a déterminé les familles de fonctions de répartition qui admettent une statistique exhaustive non triviale.

Suivant Dynkin, nous introduisons le concept de famille régulière de fonctions de répartition à rang fini. Puis nous donnons une condition nécessaire et suffisante pour qu'une statistique de partition soit indépendante de la distribution initiale dans une famille de fonctions de répartition à rang fini.

Dans ce qui suit nous considérons seulement des fonctions de répartition absolument continues et nous donnons ces fonctions à l'aide de leurs densités de probabilité. Une famille F de fonctions de répartitions absolument continues est déterminée par une densité $p(x, \theta)$ qui dépend d'un paramètre θ. L'ensemble Π des valeurs admissibles du paramètre est donné. Le paramètre n'est pas nécessairement un nombre, en admettant que θ soit un vecteur nous pouvons considérer des familles qui dépendent de plusieurs paramètres. Chaque $\theta \in \Pi$ détermine alors une densité de probabilité $p(x, \theta)$ et ainsi un élément de la famille F.

Soit Δ un intervalle (fini ou infini) sur l'axe réel et soit $p(x)$ une fonction définie pour $x \in \Delta$. Nous disons que la fonction $p(x)$ est régulière en Δ si elle satisfait aux conditions suivantes :

(i) $p(x)$ est continue en Δ;

(ii) Il y a un sous-ensemble A de Δ dont la fermeture est égale à la fermeture de Δ, tel que la dérivée $p'(x)$ existe et est continue en A.

Nous disons que la famille F des fonctions de répartition est régulière en Δ si les densités $p(x, \theta)$ sont régulières en $\Delta (\theta \in \Pi)$.

Soit F une famille des fonctions de répartition régulière en Δ et soit θ_0 un élément fixe de Π. Formons pour chaque $\theta \in \Pi$ la fonction

$$g_x(\theta) = \ln p(x, \theta) - \ln p(x, \theta_0).$$

Soit Λ le plus petit espace linéaire qui contient les constantes numériques et toutes les fonctions $g_x(\theta) (\theta \in \Pi)$. Désignons par $r + 1$ la dimension de Λ. Le nombre r, qui peut aussi être infini, s'appelle le rang de la famille F en Δ.

Dynkin a démontré le théorème suivant :

Théorème de Dynkin. — *Soit* F *une famille des fonctions de répartition qui est régulière dans l'intervalle* Δ *et qui a un rang* r *fini. La densité* $p(x, \theta)$ *admet alors la représentation*

$$(4.1) \quad p(x, \theta) = \exp\left\{ \sum_{j=1}^{r} \varphi_j(x) c_j(\theta) + c_0(\theta) + \varphi_0(x) \right\} \qquad (x \in \Delta, \ \theta \in \Pi).$$

Ici les fonctions $\varphi_j(x) (j = 1, \ldots, r)$ *sont régulières en* Δ *et les fonc-*

tions $[1, \varphi_1(x), \ldots, \varphi_r(x)]$ *aussi bien que les fonctions* $[1, c_1(\theta), \ldots, c_r(\theta)]$ *sont linéairement indépendantes.*

La réciproque est aussi vraie : Faisons l'hypothèse que la densité $p(x, \theta)$ *a la forme* (4.1), *où les* $\varphi_j(x)$ *sont régulières et supposons que les* $[1, \varphi_1(x), \ldots, \varphi_r(x)]$ *ainsi que les* $[1, c_1(\theta), \ldots, c_r(\theta)]$ *sont linéairement indépendantes. Alors le rang de la famille déterminée par les* $p(x, \theta)$ *est fini et égal à* r.

Dynkin a aussi démontré qu'une famille qui admet une statistique exhaustive non triviale ([3]) est toujours de rang fini. Soit F une famille de répartitions régulières de rang fini; et soit $(X_1, X_2, \ldots, X_n)$ un échantillon. Supposons $n \geqq r$, alors le système des fonctions

$$\mathrm{T}_j(x_1, \ldots, x_n) = \sum_{j=1}^{n} \varphi_j(x_j) \qquad (j = 1, \ldots, r) \tag{4.2}$$

est une statistique exhaustive pour F.

Les T_j s'appellent les composantes de la statistique exhaustive pour F.

Dans ce qui suit nous étudions une famille F de fonctions de répartition qui est régulière et de rang fini r et qui dépend de $p \geqq 1$ paramètres. Alors nous dirons que F est une famille du type D. Il sera quelquefois plus commode de mettre les composants de θ en évidence, nous écrivons donc, au lieu de (4.1),

$$p(x; \theta_1, \ldots, \theta_p) = \exp\left\{\sum_{j=1}^{r} \varphi_j(x)\, c_j(\theta_1, \ldots, \theta_p) + c_0(\theta_1, \ldots, \theta_p) + \varphi_0(x)\right\}. \tag{4.1.1}$$

Soit $n \geqq r$ et prenons un échantillon $X_1, \ldots, X_n$. La fonction de vraisemblance a la forme

$$\mathrm{L}(X_1, \ldots, X_n \mid \theta_1, \ldots, \theta_p) = \exp\left\{\sum_{j=1}^{r} \mathrm{T}_j(X_1, \ldots, X_n) c_j(\theta_1, \ldots, \theta_p) + n c_0(\theta_1, \ldots, \theta_p) + \sum_{\nu=1}^{n} \varphi_0(X_\nu)\right\}, \tag{4.3}$$

où le système des fonctions $\mathrm{T}_j = \mathrm{T}_j(X_1, \ldots, X_n)$ $(j = 1, 2, \ldots, r)$ est donné par (4.2) et forme une statistique exhaustive pour la famille F.

([3]) Une statistique triviale est équivalente à l'ensemble des observations.

Alors

$$\frac{\partial \mathrm{L}}{\partial \theta_k} = \left[\sum_{j=1}^{r} \mathrm{T}_j \frac{\partial c_j}{\partial \theta_k} + n \frac{\partial c_0}{\partial \theta_k} \right] \mathrm{L}. \tag{4.4}$$

Soit $\mathrm{S} = \mathrm{S}(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ une statistique de partition en F et désignons la fonction caractéristique de S par

$$f(t) = f(t \mid \theta_1, \ldots, \theta_p) = \mathbf{E}(e^{it\mathrm{S}}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{it\mathrm{S}} \mathrm{L}\, dx. \tag{4.5}$$

Dans les formules (4.4) et (4.5) nous avons écrit L au lieu de $\mathrm{L}(\mathrm{X}_1, \ldots, \mathrm{X}_n \mid \theta_1, \ldots, \theta_p)$, $\frac{\partial c_j}{\partial \theta_k}$ au lieu de $\frac{\partial c_j(\theta_1, \ldots, \theta_k)}{\partial \theta_k}$ et dx au lieu de $dx_1, \ldots, dx_n$.

Dans ce qui suit, nous ferons toujours les deux hypothèses suivantes :

(I) Les moments $\mathbf{E}(\mathrm{T}_j)(j = 1, 2, \ldots, r)$ existent;

(II) On peut différentier $f(t)$ relativement à $\theta_k(k = 1, \ldots, p)$ sous l'intégrale (4.5).

Alors

$$\frac{\partial f}{\partial \theta_k} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{it\mathrm{S}} \frac{\partial \mathrm{L}}{\partial \theta_k} dx = \mathbf{E}\left\{ e^{it\mathrm{S}} \left[\sum_{j=1}^{r} \mathrm{T}_j \frac{\partial c_j}{\partial \theta_k} + n \frac{\partial c_0}{\partial \theta_k} \right] \right\}. \tag{4.6}$$

Pour que la statistique S soit indépendante de la distribution initiale en F, il faut et il suffit que $\frac{\partial f}{\partial \theta_k} = 0$ pour $k = 1, 2, \ldots, p$ ou

$$\sum_{j=1}^{r} \frac{\partial c_j}{\partial \theta_k} \mathbf{E}(e^{it\mathrm{S}}\mathrm{T}_j) = -n \frac{\partial c_0}{\partial \theta_k} \mathbf{E}(e^{it\mathrm{S}}) \qquad (k = 1, 2, \ldots, p) \tag{4.7}$$

pour chaque valeur réelle t et chaque $(\theta_1, \ldots, \theta_p) \in \Pi$. Par conséquent, on a pour $t = 0$ la relation

$$\sum_{j=1}^{r} \frac{\partial c_j}{\partial \theta_k} \mathbf{E}(\mathrm{T}_j) = -n \frac{\partial c_0}{\partial \theta_k}. \tag{4.7.1}$$

Il s'ensuit de (4.7) et de (4.7.1) que

$$\sum_{j=1}^{r} \frac{\partial c_j}{\partial \theta_k} \left\{ \mathbf{E}(e^{it\mathrm{S}}\mathrm{T}_j) - \mathbf{E}(e^{it\mathrm{S}})\mathbf{E}(\mathrm{T}_j) \right\} = 0 \tag{4.8}$$

pour $k = 1, 2, \ldots, p$ et $(\theta_1, \theta_2, \ldots, \theta_p) \in \Pi$.

Nous avons besoin d'un lemme; pour le formuler, il est avantageux d'introduire une notation plus condensée.

Soit $\theta^\alpha = (\theta_1^\alpha, \theta_2^\alpha, \ldots, \theta_p^\alpha)$ un vecteur de **Π**. Nous substituons dans la dérivée partielle $\frac{\partial c_j}{\partial \theta_k}$ de la fonction $c_j(\theta_1, \ldots, \theta_p)$ pour les variables les valeurs des composants du vecteur θ^α et nous obtenons

$$\left.\frac{\partial c_j}{\partial \theta_k}\right|_{\theta_\alpha} = \frac{\partial c_j}{\partial \theta_k}(\theta_1^\alpha, \ldots, \theta_p^\alpha)$$

et nous désignons cette quantité par C_{jk}^α.

Lemme 2. — *Soit* F *une famille de fonctions de répartition du type* D *et soit* r *le rang de* F. *Alors on peut déterminer* r *nombres entiers* k_1, $k_2, \ldots, k_r$ *et* r *vecteurs* $\theta^\alpha = (\theta_1^\alpha, \ldots, \theta_p^\alpha)(\alpha = 1, \ldots, r)$ *tels que :*

(i) $$1 \leqq k_j \leqq r;$$

(ii) $$\theta^\alpha \in \Pi;$$

(iii)

$$\begin{vmatrix} C_{1k_1}^1 & \ldots & C_{rk_1}^1 \\ C_{1k_2}^2 & \ldots & C_{rk_2}^2 \\ \ldots & \ldots & \ldots \\ C_{1k_r}^r & \ldots & C_{rk_r}^r \end{vmatrix} \neq 0.$$

Ici les nombres $k_1, \ldots, k_r$ *et les vecteurs* $\theta^1, \ldots, \theta^r$ *ne sont pas nécessairement différents.*

Nous donnons la démonstration de ce lemme dans la section suivante. Maintenant nous l'acceptons pour étudier ses conséquences.

Supposons que les entiers $k_1, \ldots, k_r$ et les vecteurs $\theta^1, \ldots, \theta^r$ satisfont aux conditions du lemme 2. On déduit alors de la formule (4.8) les équations

$$(4.9) \qquad \sum_{k=1}^{r} C_{jk_\nu}^\nu \{\mathbf{E}(e^{itS}T_j) - \mathbf{E}(e^{itS})\mathbf{E}(T_j)\} = 0 \qquad (\nu = 1, 2, \ldots, r).$$

Le système (4.9) est un système de r équations linéaires et homogènes. En vertu du lemme 2, le déterminant de ce système est différent de zéro. Ainsi (4.9) n'a aucune solution, sauf la solution triviale et l'on a nécessairement

$$(4.10) \qquad \mathbf{E}(e^{itS}T_j) = \mathbf{E}(e^{itS})\mathbf{E}(T_j) \qquad (j = 1, \ldots, r).$$

Mais cela veut dire que chaque fonction T_j a la régression constante en S.

Réciproquement on déduit aisément de (4.7.1) et de l'hypothèse que chaque $T_j (j = 1, \ldots, k)$ a la régression constante en S la relation $\frac{\partial f}{\partial \theta_k} = 0$ pour $k = 1, \ldots, p$ et l'on obtient le théorème suivant :

Théorème 4. — *Soit* F *une famille du type* D *de rang* r *et soit* S *une statistique de partition. Supposons que :*

(I) *Les moments* $\mathbf{E}(T_j)$ *existent;*

(II) *On peut différentier* $f(t)$ *relativement à* $\theta_k (k = 1, \ldots, p)$ *sous l'intégrale* (4.5);

(III) $\sum_{j=1}^{r} \frac{\partial c_j}{\partial \theta_k} \mathbf{E}(T_j) = -n \frac{\partial c_0}{\partial \theta_k}$.

Pour que S *soit indépendante de la distribution initiale en* S, *il faut et il suffit que les composantes* T_j *de la statistique exhaustive pour* F *aient la régression constante en* S.

Il nous reste encore à démontrer le lemme 2.

5. Démonstration du lemme 2. — Pour cette démonstration nous avons besoin d'un autre lemme.

Lemme 3. — *Soient* $c_j = c_j(\theta_1, \theta_2, \ldots, \theta_p) (j = 1, 2, \ldots, r)$ *les fonctions introduites par le théorème de Dynkin et soient* $\alpha_1, \ldots, \alpha_r$ *des constantes. L'ensemble des p équations*

$$\sum_{j=1}^{r} \alpha_j \frac{\partial c_j}{\partial \theta_k} = 0 \qquad (k = 1, 2, \ldots, p) \tag{5.1}$$

entraîne les relations

$$\alpha_j = 0 \quad \text{pour} \quad j = 1, \ldots, r. \tag{5.1.1}$$

Ce lemme est une conséquence immédiate de l'indépendance des fonctions $1, c_1(\theta), \ldots, c_r(\theta)$.

Considérons la fonction $c_1 = c_1(\theta_1, \ldots, \theta_p)$. Comme le système $(1, c_1, \ldots, c_p)$ est linéairement indépendant, c_1 ne peut pas être constant, il y a alors nécessairement un indice j_1 et un vecteur θ_1 tel que

$$C^1_{1.j_1} = \left. \frac{\partial c_1}{\partial \theta_{j_1}} \right|_{\theta = \theta_1} \neq 0.$$

Nous introduisons le déterminant

$$D_{j_1\nu}(\theta \mid \theta^1) = \begin{vmatrix} C^1_{1j_1} & C^1_{2j_1} \\ \dfrac{\partial c_1}{\partial \theta_\nu} & \dfrac{\partial c_2}{\partial \theta_\nu} \end{vmatrix}.$$

Ce déterminant est une fonction des $\theta_1, \ldots, \theta_p$. Désignons par $\alpha_1 = - C^1_{2j_1}$ et $\alpha_2 = C^1_{1j_1}$, alors

$$D_{j_1\nu}(\theta \mid \theta^1) = \alpha_1 \frac{\partial c_1}{\partial \theta_\nu} + \alpha_2 \frac{\partial c_2}{\partial \theta_\nu}.$$

Comme $\alpha_2 \neq 0$, nous déduisons du lemme 3 que la fonction $D_{j\nu}(\theta \mid \theta^1)$ ne peut pas être o identiquement. Il est alors possible de déterminer un indice j_2 et un vecteur θ^2 tel que

$$\alpha_3 = D_{j_1 j_2}(\theta^{(2)} \mid \theta^{(1)}) \neq 0.$$

Nous continuons le raisonnement de cette façon et démontrons inductivement le lemme 2.

6. **Exemples**. — La fonction de répartition a la densité

$$(6.1) \qquad p(x, \theta_1, \theta_2) = \frac{(\theta_2)^{\theta_1}}{\Gamma(\theta_1)} x^{\theta_1 - 1} e^{-\theta_2 x}.$$

On voit aisément que $p(x, \theta_1, \theta_2)$ a la forme (4.1.1), avec

$$(6.2) \qquad \begin{cases} \varphi_0(x) = -\ln x, & c_0(\theta_1, \theta_2) = \theta_1 \ln \theta_2 - \ln \Gamma(\theta_1), \\ \varphi_1(x) = \ln x, & c_1(\theta_1, \theta_2) = \theta_1, \\ \varphi_2(x) = x, & c_2(\theta_1, \theta_2) = -\theta_2. \end{cases}$$

Les conditions (I), (II), (III) du théorème 4 sont satisfaites et l'on obtient le résultat :

Corollaire 1 du théorème 4. — *Soit* G *la famille des fonctions de répartition* (6.1) *et soit* S *une statistique de partition. Pour que* S *soit indépendante de la distribution initiale en* G, *il faut et il suffit que les composantes*

$$T_1 = \sum_{j=1}^{n} X_j \qquad \text{et} \qquad T_2 = \sum_{j=1}^{n} \ln X_j$$

de la statistique exhaustive de G *aient la régression constante en* S.

Désignons par G_{θ_2} (resp. par G^{θ_1}) la sous-famille de G obtenue en donnant θ_2 (resp. θ_1). On peut alors montrer que la condition pour la sous-famille G_{0_2} (resp. G^{0_1}) est que T_2 (resp. T_1) ait la régression constante en S. La sous-famille G^{0_1} dépend d'un paramètre d'échelle, ainsi on peut se servir du théorème 3 pour construire des statistiques indépendantes de la distribution initiale en G^{0_1}.

On peut aussi obtenir les théorèmes 1 *a*, 1 *b* et 2 comme corollaires du théorème 4.

BIBLIOGRAPHIE.

[1] Z. W. BIRNBAUM, *Distribution free tests of fit for continuous distribution functions* (*Ann. Math. Stat.*, t. 24, 1955, p. 1-8).

[2] E. B. DYNKIN, *Les statistiques nécessaires et suffisantes pour les familles de répartition* (*Usp. Mat. Nauk*, t. 6, n° 41, 1951, p. 68-90; en russe).

[3] E. B. DYNKIN, *Les statistiques suffisantes et nécessaires pour les familles de répartition* [*Dokl. Akad. Nauk S. S. S. R.* (nouv. série), t. 75, 1950, p. 161-164; en russe].

[4] E. LUKACS, *Characterization of populations by properties of suitable statistics* (*Proc. Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. II, p. 195-214, University of California Press, Berkeley, California, 1956).

[5] G. DARMOIS, *Sur les lois de probabilité à estimation exhaustive* (*C. R. Acad. Sc.*, t. 200, 1935, p. 1265-1266).

[6] D. DUGUÉ, *Applications des propriétés de la limite au sens du calcul des probabilités à l'étude de diverses questions d'estimation* (*J. Éc. Polytechnique*, t. 3, n° 4, 1937, p. 305-373).