STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

AFDELING MATHEMATISCHE STATISTIEK

S 224

Les fonctions génératrices liées à quelques tests non-paramétriques

par

Prof.Dr. D. van Dantzig

herdruk 1964

Fl. 0,30

STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

AFDELING MATHEMATISCHE STATISTIEK

S 224

Les fonctions génératrices liées à quelques tests non-paramétriques

par

Prof.Dr. D. van Dantzig

herdruk 1964

Printed at the Mathematical Centre at Amsterdam, 49, 2nd Boerhaavestraat. The Netherlands.

The Mathematical Centre, founded the 11th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications, and is sponsored by the Netherlands Government through the Netherlands Organization for Pure Scientific Research (Z.W.O.) and the Central National Council for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

§1. Soient $\underline{x}_1, \ldots, \underline{x}_n$ [1] n variables aléatoires indépendantes, ayant la même fonction de répartition continue. Soit l'ensemble $I = \{1, \ldots, n\}$ divisé en m sous-ensembles disjoints $S_1, \ldots, S_m$. Soit n_λ [2] le nombre des éléments de S_λ, de sorte que $\sum n_\lambda = n$. Soit enfin $\underline{T}$ le nombre des pairs (i,j) où $i \in S_\lambda$, $j \in S_\mu$, et où simultanément

$$(1) \qquad \lambda < \mu \qquad , \quad \underline{x}_i > \underline{x}_j \ .$$

Soit P_T la probabilité pour que la variable aléatoire $\underline{T}$ prenne la valeur numérique T, de sorte que $P_T > 0$ entraîne que T est entier, ≥ 0 et $\leq N \overset{\text{def}}{=} \sum_{i<j} n_i n_j$.

A cause de la continuité de la fonction de répartition des $\underline{x}_i$, une égalité $\underline{x}_i = \underline{x}_j$ a la probabilité zéro, sauf dans le cas $i = j$ (donc $\lambda = \mu$).

Le but de cette note est d'exposer une manière bien simple pour trouver la fonction génératrice (comme d'ailleurs, voir §4) de l'aléatoire $\underline{T}$.

En 1946, 1947 nous avons exposé les principes d'une méthode qui admet parfois d'obtenir une fonction génératrice (ou aussi une fonction caractéristique à argument réel) sans recourir d'abord à une équation recurrente.

L'aspect peut-être le plus caractéristique de notre méthode consiste en une interprétation probabiliste des variables auxiliaires intervenants dans la fonction génératrice, et de cette fonction elle-même.

---

1) Les aléatoires seront designées par des lettres soulignées, les valeurs numériques qu'elles prennent par des lettres, par préférence correspondantes, non soulignées.

2) Les indices i,j,k parcouront toujours l'ensemble I; les indices λ, μ prendront les valeurs $1, \ldots, m$.

§2. Supposons qu'on lance successivement n particules matérielles à dimensions négligeables, partant du "point" $+\infty$ de l'axe des x et tellement que pour chaque particule la probabilité qu'elle s'arrête dans un intervalle de cette axe soit égale à la probabilité qu'un $\underline{x}_i$ y soit contenu.

Supposons d'ailleurs que toutes les fois qu'une de ces particules passe au dessus d'une autre, lancée antérieurement, il y a une probabilité $1-z$ qu'un événement bien défini se réalise. Nous appellerons cet événement une "catastrophe". Il peut, par example, consister en ce qu'un signal rouge s'allume. Soit enfin $G_n(z)$ la probabilité que le feu rouge ne se présente <u>aucune fois</u> à cause du lancement des n particules.

Supposons que $k-1$ particules ont déjà été lancées et que jusque là aucune "catastrophe" n' a eu lieu. Alors les $k-1$ particules se trouvent sauf une probabilité 0 dans $k-1$ points différents de l'axe des x, en divisant cette axe en k intervalles disjoints. Lorsque la $k^{\text{ième}}$ particule sera lancée, toutes les permutations des k particules auront des probabilités égales. Donc la probabilité que la $k^{\text{ième}}$ particule s'arrête à l'intérieur du $i^{\text{ième}}$ des k intervalles est pour chaque i égal à $\frac{1}{k}$. La probabilité pour que la $k^{\text{ième}}$ particule ne cause aucune catastrophe sera donc égale à $\frac{1}{k}(1+z+z^2+\ldots+z^{k-1})$. Il s'ensuit immédiatement que

$$G_n(z) = \prod_1^n k^{-1}(1+\ldots+z^{k-1}) = \prod_1^n \frac{1-z^k}{k(1-z)} . \tag{2}$$

§3. Soit maintenant

$$C_{n_1,\ldots,n_m}(z) \stackrel{\text{def}}{=} \sum_0^N P_T z^T \tag{3}$$

la fonction génératrice cherchée.

Supposons que d'abord les particules de l'ensemble S_1, alors celles de l'ensemble S_2, etc. sont lancées.

Disons qu'une catastrophe est d'espèce μ, si elle est due au passage d'une particule de S_μ sur une autre appartenante au même sous-ensemble, et d'espèce 0, si elle est due au passage d'une particule sur une autre, appartenant à un sous-ensemble différant (donc antérieur). La probabilité qu'aucune catastrophe d'espèce 0 n'arrive est donc $C_{n_1,\ldots,n_m}(z)$, tandisque la probabilité qu'aucune catastrophe d'espèce μ arrive est égale à $G_{n_\mu}(z)$. Donc on a, à cause de l'indépendance des probabilités de manque de catastrophe au cas de passage:

$$(4) \qquad G_n(z) = C_{n_1,\ldots,n_m}(z) \cdot \prod_1^m G_{n_\mu}(z)$$

ou

$$C_{n_1,\ldots,n_m}(z) = \frac{G_n(z)}{\prod_1^m G_{n_\mu}(z)}$$

ou aussi, à cause de (4),

$$(5) \qquad C_{n_1,\ldots,n_m}(z) = \frac{[n]!}{n!} \cdot \prod_1^m \frac{n_\mu!}{[n_\mu]!}$$

si on pose pour abréviation

$$(6) \qquad [k]! \overset{\text{def}}{=} \prod_1^k (1-z^j) .$$

§4. La variable aléatoire

$$(7) \qquad \underline{S} \overset{\text{def}}{=} 2\underline{T} - N$$

est le <u>coefficient de corrélation de rang</u> d'après Kendall, dans le cas où tous les rangs sont différents dans l'un des arrangements, et s'arrangent en m groupes de $n_1,\ldots,n_m$ rangs égaux ("ties") dans l'autre.

La fonction caractéristique de $\underline{S}$ s'obtient donc par une transformation simple de la fonction génératrice de $\underline{T}$, à savoir

$$(8) \qquad E \exp(it\underline{S}) = e^{-itN} C_{n_1,\ldots,n_m}(e^{2it}) \quad .$$

Dans le cas $m = n$, c.à.d. $n_1 = \ldots = n_m = 1$ ("no ties") $C_{1,\ldots,1}(z)$ se réduit à $G_n(z)$, et la fonction caractéristique de $\underline{S}$ devient

$$\prod_1^n \frac{\sin kt}{k \sin t} \quad .$$

Elle a été obtenue par Kendall (1947), p.403.

Dans le cas $m = 2$ $\underline{T}$ se réduit à la variable aléatoire $\underline{U}$ déterminant d'après Mann et Whitney le test de Wilcoxon. En utilisant la relation de récurrence de Mann et Whitney j'ai obtenu en 1949 la fonction génératrice dans ce cas.

Dans le cas général un test basé sur $\underline{T}$ a été construit par M. Terpstra (1952). Sur ma demande M. Terpstra a publié dans son article la fonction caractéristique de T, déduite de celle pour $m = 2$ par induction complète d'après m. Quoique donc les résultats sont connus, la méthode de démonstration simple publiée ici [3], moins formelle et plus intuitive que l'ancienne, pourrait peut-être avoir quelqu'intérêt en elle-même.

3) Elle a été donnée dans mes cours depuis quelques années.

Litérature

D. van Dantzig (1947) Kadercursus Mathematische Statistiek, Hoofdstuk 2.

D. van Dantzig (1948) Kadercursus Mathematische Statistiek, Hoofdstuk 6.

D. van Dantzig (1949) Sur la méthode des fonctions génératrices, Colloques Internationaux du Centre National de la Recherche Scientifique. XIII. Le calcul des probabilités et ses applications.

M.G. Kendall (1947) The advanced theory of statistics, Volume I, London.

H.B. Mann and D.R. Whitney (1947) On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat. 18, 50-60.

T.J. Terpstra (1952) The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking, Proc. Kon. Ned. Akad. van Wetensch. 55, 327-333.