A generating function in the theory of order statistics

By I. VINCZE (Budapest)

Dedicated to Professor Béla Szőkefalvi-Nagy at the occasion of his 50th birthday

Introduction

Let $\xi_1, \xi_2, \dots, \xi_n$ and $\eta_1, \eta_2, \dots, \eta_n$ be independent sample elements taken from populations with continuous distribution functions $F(x)$ and $G(x)$ resp. As usual put

$$D^+_{n,n} = \max_{(x)} \left(F_n(x) - G_n(x)\right) \quad \text{and} \quad D_{n,n} = \max_{(x)} |F_n(x) - G_n(x)|.$$

Let us denote in the following by $\varrho^+_{n,n}$ and $\varrho_{n,n}$ the „first" values of x for which the random functions $F_n(x) - G_n(x)$ resp. $|F_n(x) - G_n(x)|$ take their maximum, i. e.

$$\varrho^+_{n,n} = \inf\{x: F_n(x) - G_n(x) = D^+_{n,n}\}$$

$$\varrho_{n,n} = \inf\{x: |F_n(x) - G_n(x)| = D_{n,n}\}.$$

If $\zeta_1^* \leqq \zeta_2^* \leqq \dots \leqq \zeta_{2n}^*$ respresent the ordered union of the above two samples, further

$$R^+_{n,n} = \frac{1}{2}\left(F_n(\varrho^+_{n,n} + 0) + G_n(\varrho^+_{n,n} + 0)\right)$$

and

$$R_{n,n} = \frac{1}{2}\left(F_n(\varrho_{n,n} + 0) + G_n(\varrho_{n,n} + 0)\right),$$

then $2nR^+_{n,n}$ resp. $2nR_{n,n}$ are the first indices of the $\zeta_i^* - s$ for which the deviations take their maximal values $D^+_{n,n}$ and $D_{n,n}$ resp.

In paper [5] the joint distribution and joint limiting distribution of the pairs of random variables $(D^+_{n,n}, R^+_{n,n})$ and $(D_{n,n}, R_{n,n})$, in paper [6], the three variate generating function of the probabilities belonging to the pair of random variables $(D^+_{n,n}, R^+_{n,n})$ was determined under the assumption $F(x) \equiv G(x)$. The present paper deals with the determination of the generating function of the probabilities belonging to the pair of random variables $(D_{n,n}, R_{n,n})$.

Well known basic results concerning relations of generating functions for more general variables are due to E. SPARRE ANDERSEN [1, 2], F. SPITZER [4] and

W. FELLER [3], these are however valid for the one-sided case (for $D^+_{n,n}$ and $R^+_{n,n}$ in our case) and the two sided case does not immediately follows from them. Our problem suggests a direct derivation of the double generating function and this is contained in § 1. In § 2 the distribution of $D_{n,n}$ is treated.

§ 1. Derivation of the twovariate generating function

1. According to Theorem 2 of the paper [5] in case $F(x)\equiv G(x)$

$$P^{(n)}_{k,r}=P\left\{D_{n,n}=\frac{k}{n},\ R_{n,n}=\frac{r}{2n}\right\}=$$

$$=\frac{2}{\binom{2n}{n}}\frac{k(k+1)}{r(2n-r+1)}\sum_{\nu=1}^{\infty}\sum_{\mu=1}^{\infty}(-1)^{\nu+\mu}(2\nu+1)(2\mu+1)\binom{r}{\frac{r+k}{2}+\nu k}\cdot$$

$$\cdot\binom{2n-r+1}{n-\frac{r-k}{2}-\mu(k+1)}$$

for $k=1, 2, \ldots, n$, $r=k, k+2, \ldots, 2n-k$, while the probability $P^{(n)}_{k,r}$ is zero, if $k=0$ or $k+r$ odd.

Using the notation

$$A_{k,r}=\frac{k}{r}\sum_{\nu=0}^{\infty}(-1)^{\nu}(2\nu+1)\binom{r}{\frac{r+k}{2}+\nu k}$$

the mentioned probability has the form

$$P^{(n)}_{k,r}=2\cdot A_{k,r}A_{k+1,2n-r+1}.$$

With the notation

$$\omega(z)=\frac{1-\sqrt{1-4z}}{1+\sqrt{1-4z}}=\frac{4z}{(1+\sqrt{1-4z})^2},$$

we formulate the following

Theorem 1.1 *In the case* $F(x)\equiv G(x)$

$$(1.1)\quad G_k(z,w)=\sum_{n=k}^{\infty}\sum_{(r)}{}^{*}\binom{2n}{n}P^{(n)}_{k,r}z^{r-k}w^{n-k}=2\,\frac{[1+\omega(w)]^{k+1}}{1+[\omega(w)]^{k+1}}\,\frac{[1+\omega(z^2w)]^{k}}{1+[\omega(z^2w)]^{k}}$$

where $\sum\limits_{(r)}^{*}$ *denotes summation over* $r=k, k+2, \ldots, 2n-k$.

This relation holds for $|w|<\frac{1}{4}$ and $|z^2w|<\frac{1}{4}$. As throughout the convergence conditions are of such type, in the following we shall not mention this fact. In the following sections we carry out the proof of this theorem, with the aid of some lemmas.

2. A simple and often used fact is expressed in the following

Lemma 1. 1 *Let be given the quantities* $A_{k,r}$ *for* $k=1,2,\ldots,\ r=k,k+2,\ldots$ *and let be*

$$\sum_{(r)}{}^{*} A_{k,r}u^{r-k}=g_k(u^2)$$

where $\sum_{(r)}^{*}$ *denotes summation over* $r=k,k+2,\ldots$. *If now*

$$\binom{2n}{n}P_{k,r}^{(n)}=2A_{k,r}A_{k+1,2n-r+1}, \tag{1. 2}$$

then

$$\sum_{n=k}^{\infty}\sum_{(r)}{}^{*}\binom{2n}{n}P_{k,r}^{(n)}z^{n-k}w^{n-k}=2g_k(z^2w)g_{k+1}(w).$$

Proof. From (1. 2) follows the identity

$$\binom{2n}{n}P_{k,r}^{(n)}z^{r-k}w^{n-k}=2A_{k,r}(z^2w)^{\frac{r-k}{2}}A_{k+1,2n-r+1}w^{\frac{2n-r-k}{2}},$$

summation over n for $n=\frac{r+k}{2},\ \frac{r+k}{2}+1,\ldots$ and the use of the relation

$$\sum_{(n)}A_{k+1,2n-r+1}w^{\frac{2n-r-k}{2}}=\sum_{(\alpha)}{}^{*}A_{k+1,\alpha}w^{\frac{\alpha-k-1}{2}}$$

$\alpha=k+1,k+3,\ldots$ results in lemma 1. 1.

On the ground of this lemma it suffices to determine the generating function of

$$A_{k,r}=\frac{k}{r}\sum_{\nu=0}^{\infty}(-1)^{\nu}(2\nu+1)\binom{r}{\frac{r+k}{2}+\nu k}.$$

Introducing here instead of r the new variable $s=\frac{r-k}{2}$ and using the notation $A_{k,s}$ we obtain:

$$A_{k,s}=\sum_{\nu=0}^{\infty}(-1)^{\nu}\frac{(2\nu+1)k}{k+2s}\binom{k+2s}{(\nu+1)k+s}.$$

We shall make use of the second part of the following

Lemma 1. 2 *Let be* a *and* b *integers,* $b>0$, *then*

$$\sum_{s=-b}^{\infty}\binom{a+2s}{b+s}v^{s+b}=\frac{1}{\sqrt{1-4v}}\left(\frac{2}{1+\sqrt{1-4v}}\right)^{a-2b},\quad \textit{if}\quad a\geqq 2b,$$

$$\sum_{s=b-a}^{\infty}\binom{a+2s}{b+s}v^{s-b}=\frac{1}{\sqrt{1-4v}}\left(\frac{2}{1+\sqrt{1-4v}}\right)^{2b-a},\quad \textit{if}\quad a\leqq 2b.$$

PROOF. There are known the following generating functions (see e. g. [6] p. 31) valid for $h=0, 1, 2, \ldots$

$$\sum_{s=0}^{\infty}\binom{h+2s}{s}v^s = \frac{1}{\sqrt{1-4v}}\left(\frac{2}{1+\sqrt{1-4v}}\right)^h,$$

$$\sum_{s=h}^{\infty}\binom{-h+2s}{s}v^s = \frac{1}{\sqrt{1-4v}}\left(\frac{2v}{1+\sqrt{1-4v}}\right)^h.$$

Substituting $a-2b$ and $2b-a$ instead of h in the first and second relation resp., and $s+b$ instead of s lemma 1. 2 is obtained.

3. Due to this lemma, we may put down the following relation valid for $v=0, 1, 2, \ldots$.

$$(1.\,3)\qquad h_k(v) = \sum_{s=vk}^{\infty}\binom{k+2s}{(v+1)k+s}v^s = \frac{v^{-(v+1)k}}{\sqrt{1-4v}}\left(\frac{2v}{1+\sqrt{1-4v}}\right)^{(2v+1)k}.$$

We shall now evaluate the following integral:

$$I_k(v) = \int_0^v t^{k-1}h_k(t^2)\,dt.$$

Using the transformation

$$t' = \frac{t}{1+\sqrt{1-4t^2}},$$

we get

$$dt' = \frac{dt}{\sqrt{1-4t^2}\,(1+\sqrt{1-4t^2})}$$

and the new integration limits $\left(0, \dfrac{v}{1+\sqrt{1-4v^2}}\right)$.

This results in

$$(1.\,4)\qquad I_k(v) = \frac{1}{(2v+1)k}\left(\frac{2v}{1+\sqrt{1-4v^2}}\right)^{(2v+1)k}.$$

Substituting now in 1. 2 v^2 instead of v and multiplying by v^{k-1} we obtain an identity, the integration of which leads to the following relation

$$\int_0^v t^{k-1}h_k(t^2)\,dt = I_k(v) = \sum_{s=vk}^{\infty}\frac{1}{k+2s}\binom{k+2s}{(v+1)k+s}v^{k+2s}.$$

A slight modification gives

$$\sum_{s=vk}^{\infty}\frac{(2v+1)k}{k+2s}\binom{k+2s}{(v+1)k+s}v^s = \left(\frac{2}{1+\sqrt{1-4v}}\right)^k\left(\frac{4v}{1+\sqrt{1-4v})^2}\right)^{vk}.$$

Multiplication by $(-1)^{\nu}$ and summation over ν results in the generating function of $A_{k,s}$

$$\sum_{s=0}^{\infty}\sum_{\nu=0}^{\infty}(-1)^{\nu}\frac{(2\nu+1)k}{k+2s}\binom{k+2s}{(\nu+1)k+s}v^{s}=\left(\frac{2}{1+\sqrt{1-4v}}\right)^{k}\frac{1}{1+[\omega(v)]^{k}}.$$

Substituting $k+2s$ by r we obtain the generating function of $A_{k,r}$

$$\sum_{(r)}{}^{*}A_{k,r}v^{r-k}=\frac{[1+\omega(w)]^{k}}{1+[\omega(w)]^{k}},$$

which gives on the ground of Lemma 1. 1 the proof of theorem 1. 1.

§ 2. Special cases

1. *Generating function of the Gnedenko—Koroljuk probabilities.* For given k the summation of the probabilities $P_{k,r}^{(n)}$ with respects to r gives the probabilities of Gnedenko—Koroljuk

$$P_{k,\cdot}^{(n)}=\sum_{(r)}{}^{*}P_{k,r}^{(n)}=P\left(D_{n,n}=\frac{k}{n}\right),\qquad k=1,2,\ldots,n$$

where $\sum\limits_{(r)}{}^{*}$ denotes summation for $r=k, k+2, \ldots, 2n-k$.

Substituting now in our generating function the value $z=1$, we obtain the generating function of the probabilities $P_{k,\cdot}^{(n)}$

$$\sum_{n=k}^{\infty}\binom{2n}{n}P_{k,\cdot}^{(n)}w^{n-k}=2\frac{[1+\omega(w)]^{2k+1}}{(1+[\omega(w)]^{k})(1+[\omega(w)]^{k+1})}.$$

2. *Generating function of the Gnedenko—Koroljuk distribution.* Omitting an inmediate derivation we shall verify the following

Theorem 2. 1 *Using the notation*

$$\overline{P}_{k}^{(n)}=\sum_{s=1}^{k-1}P_{s,\cdot}^{(n)}=P\left(D_{n,n}<\frac{k}{n}\right)$$

for $n=k, k+1, \ldots$ *and* $P_{k}^{(n)}=1$ *if* $n<k$ *the relation*

$$G_{k}(w)=\sum_{n=0}^{\infty}\binom{2n}{n}\overline{P}_{k}^{(n)}w^{n}=\frac{1+\omega(w)}{1-\omega(w)}\,\frac{1-[\omega(w)]^{k}}{1+[\omega(w)]^{k}}\tag{2. 1}$$

is valid.

Proof. We shall show that from (2. 1) for $\overline{P}_{k}^{(n)}$ the known formula of the Gnedenko—Koroljuk distribution follows. The coefficient of w^{n} in the series expansion of the function on the left hand side is

$$\frac{1}{2\pi i}\oint G_{k}(w)\frac{dw}{w^{n+1}},\tag{2. 2}$$

where the integration is to be extended for a simple closed curve around the origin inside the circle $|w| = \frac{1}{4}$. Using the transformation

$$\omega = \frac{1-\sqrt{1-4w}}{1+\sqrt{1-4w}}$$

from this

$$w = \frac{\omega}{(1+\omega)^2} \quad \text{and} \quad dw = \frac{1-\omega}{(1+\omega)^3}\, d\omega$$

follows. Now we have for our integral (2. 2)

$$\frac{1}{2\pi i} \oint \frac{1-\omega^k}{1+\omega^k} \frac{(1+\omega)^{2k}}{\omega^{n+1}}\, d\omega$$

where the integration is to be extended for a circle small enough around the origin $\omega = 0$. This means that we have to expand the function

$$\frac{1-\omega^k}{1+\omega^k}(1+\omega)^{2n}$$

into a power series according to ω and to determine the coefficient of ω^n

$$\sum_{i=0}^{\infty} \sum_{j=0}^{2n} (-1)^i \binom{2n}{j} (\omega^{ik+j} - \omega^{(i+1)k+j}) =$$

$$= \sum_{s=0}^{2n} \binom{2n}{s} \omega^s - 2 \sum_{i=1}^{\infty} \sum_{j=0}^{2n} (-1)^j \binom{2n}{j} \omega^{ik+j}.$$

Taking $s = n$ and $j + ik = n$ in the first and second term resp., we obtain the well known formula for $\binom{2n}{n} \bar{P}_k^{(n)}$

$$\sum_{i=-\infty}^{\infty} (-1)^i \binom{2n}{n-ik}$$

q. e. d.

Bibliography

[1] E. S. Andersen, On sums of symmetrically dependent random variables, *Skand. Aktuarietidskr.* (1953), 123—138.

[2] E. S. Andersen, On the fluctuations of sums of random variables I and II., *Math. Scand.* **1** (1953), 263—285 and **2** (1954), 195—223.

[3] W. Feller, On combinatorial methods in fluctuation theory, Probability and statistics: The Herald Cramér Volume *Stockholm, New York,* 1959, 75—91.

[4] F. Spitzer, A combinatorial lemma and its applications to probability theory, *Trans. Amer. Math. Soc.* **82** (1956), 323—339.

[5] I. Vincze, Einige zweidimensionale Verteilungs- und Grenzverteilungssätze in der Theorie der geordneten Stichproben, *Publ. Math. Inst. Hung. Acad. Sci.* **2** (1957), 183—209.

[6] I. Vincze, On some joint distributions and joint limiting distributions in the theory of order statistics II., *Publ. Math. Inst. Hung. Acad. Sci.* **4** (1959), 29—47.

(Received August 22, 1962.)