Polynomial structures in rank statistics distributions

- Claude LEFEVRE (Université libre de Bruxelles)
- Philippe PICARD (Université Lyon 1, Laboratoire SAF)

2010.5 (WP 2124)
Polynomial structures in rank statistics distributions

Claude Lefèvre $^a$, Philippe Picard $^b$

$^a$Université Libre de Bruxelles, Département de Mathématique, Campus de la Plaine C.P. 210, B-1050 Bruxelles, Belgium
$^b$Université de Lyon, Institut de Science Financière et d’Assurances, 50 Avenue Tony Garnier, F-69007 Lyon, France

Abstract

This paper deals with the classical problem of how to evaluate the joint rank statistics distributions for two independent i.i.d. samples from a common continuous distribution. It is pointed that these distributions rely on an underlying polynomial structure of negative binomial type. This property is exploited to obtain, in a systematic and unified way, various recursions, some well established, for computing the joint tail and rectangular probabilities of interest.

MSC: 62G30; secondary 33A65

Keywords: Rank statistics; Kolmogorov-Smirnov two-sample tests; Generalized Appell and Abel-Gontcharoff polynomials; Negative binomial type polynomials; Lattice path enumeration.

1 Introduction

A previous paper (Denuit et al. (2003)) was concerned with the joint distributions of the order statistics for i.i.d. samples. It was shown that the left and right tail distributions can be expressed in terms of Abel-Gontcharoff polynomials and Appell (Sheffer) polynomials, respectively. The joint rectangular distributions were also discussed. In that paper, the basic sequence of polynomials was the family of monomials \( \left\{ x^n/n!, \ n \geq 0 \right\} \).

In the present work, we deal with a related classical topic, namely the joint distributions of the rank statistics for two independent i.i.d. samples from a common continuous distribution. The paper is organized as follows. First, in Section 3, we prove that the left and right tail rank statistics distributions can be written by using a (different) special case of the wide family of generalized Abel-Gontcharoff and Appell polynomials, respectively. This time, the basic sequence of polynomials is the class of negative binomial type polynomials \( \left\{ \binom{x+n-1}{n}, \ n \geq 0 \right\} \). A short presentation of the theory of generalized Abel-Gontcharoff and Appell polynomials is provided in the Appendix. In Section 2, we study with more details these families of polynomials in the negative binomial case of interest. Then, we show in Section 4 that the joint rectangular rank statistics distributions can

---

$^1$Corresponding author.

E-mail address: clefevre@ulb.ac.be (Claude Lefèvre).
be computed by means of two remarkable recursions that rely on a similar, but more complex, algebraic structure. Finally, in Section 5, we illustrate the efficiency of such recursions on two data sets in biostatistics.

The distributions of rank statistics and related topics have been discussed by many authors (see e.g. Conover (1999)). The approach developed in this note is closely related to the pioneering paper by Steck (1969) and the nice works by Mohanty (1980) and Niederhausen (1981). It exploits a (known) direct connection with some problems of lattice path enumeration (for this topic, see e.g. Mohanty (1979), Niederhausen (1997) and Krattenthaler (1997)). Various simple recursions will be obtained for computation of the joint tail and rectangular distributions. Some are well established and all will be put in a unified way by pointing out the existence of a common polynomial structure.

Following Steck (1969), we introduce (and will illustrate) the problem through the Kolmogorov-Smirnov (K-S) two-sample test. Let \( X_1, \ldots, X_m \) and \( Y_1, \ldots, Y_n \) be the order statistics for two independent samples from continuous distributions \( F \) and \( G \). Let \( F_n(x) \) and \( G_m(x) \) be the corresponding empirical distribution functions. To test the null hypothesis \( H_0 : F = G \) against the one-sided alternatives \( H_1 : F < G \) or \( H_1 : F > G \), the K-S one-sided statistics are \( D^- = \max_x [G_m(x) - F_n(x)] \) or \( D^+ = \max_x [F_n(x) - G_m(x)] \). For the two-sided alternative \( H_1 : F \neq G \), the K-S two-sided statistics is \( D = \max(D^-, D^+) \).

We now suppose that \( H_0 \) is true. Let us combine the two samples into a single one of size \( m + n \), and denote by \( i + R_i \) the rank of \( X_{(i)} \), \( 1 \leq i \leq n \), in this new ordered sample. By construction, \( 0 \leq R_1 \leq \ldots \leq R_n \leq m \). Our purpose will be to determine the joint tail or rectangular distributions of the random vector \( (R_1, \ldots, R_n) \).

It is rather easy to see that the distribution functions of \( D^- \), \( D^+ \) and \( D \) can be rewritten as probabilities of this form. Specifically, \( P(D^- \leq d) = P(R_1 \leq t_1, \ldots, R_n \leq t_n) \), \( P(D^+ \leq d) = P(R_1 \geq s_1, \ldots, R_n \geq s_n) \) and \( P(D \leq d) = P(s_1 \leq R_1 \leq t_1, \ldots, s_n \leq R_n \leq t_n) \), where \( t_i = \min([m(i-1)/n + md], m) \) and \( s_i = \max(0, [mi/n - md]) \), \([z]\) representing the largest integer \( \leq z \) and \([z]\) the smallest integer \( \geq z \).

Note that \( D^- \) and \( D^+ \) then have the same distribution. This can be checked by observing that the two vectors \( (m - R_1, \ldots, m - R_n) \) and \( (R_n, \ldots, R_1) \) have the same distribution, and \( m - s_i = t_{n-i+1}, 1 \leq i \leq n \).

## 2 Negative binomial type polynomials

The Appendix provides a succinct presentation of the generalized Appell polynomials, \( \{A_n(x \mid U), n \geq 0\} \), and generalized Abel-Gontcharoff, \( \{G_n(x \mid U), n \geq 0\} \), of degree \( n \) associated to any real family \( U = \{u_i, i \geq 0\} \). The construction of both families uses a basic family of polynomials \( \{e_n, n \geq 0\} \) of degree \( n \) satisfying (6.1) and a linear operator \( \Delta \) defined by (6.2) on the space of polynomials.

From now, our attention will be focused on the special case where that family \( \{e_n, n \geq 0\} \) is given by

\[
e_n(x) = \binom{x + n - 1}{n} = \frac{x(x + 1)\ldots(x + n - 1)}{n!} = \left(\frac{-x}{n}\right)(-1)^n, \quad n \geq 0. \tag{2.1}\]
Observe that the condition (6.1) is well satisfied. Note also that $e_n(1) = 1$ for all $n \geq 0$ and $e_n(x) = 0$ if $x = 0, -1, \ldots, -n + 1$.

The previous general theory is then applicable and allows us to derive the associated generalized Appell and Abel-Gontcharoff polynomials. The polynomials $e_n$, $\bar{A}_n$ and $\bar{G}_n$ will be called of negative binomial type. The reason is that the combinatorial coefficients (2.1) remind us of, for fixed $x$, the negative binomial distribution and, as a function of $x$, the negative binomial Lévy process (e.g. Kozubowski and Podgórski (2007)). In the sequel, it will be be enough for us to define these polynomials only on $\mathbb{Z}$ (the set of integers), instead of $\mathbb{R}$.

**2.1.** To begin with, however, let us examine the polynomials $e_n(x)$ on $\mathbb{R}$. Their generating function $\phi(x, s)$ is given by

$$
\phi(x, s) = \sum_{n=0}^{\infty} \binom{x + n - 1}{n} s^n = (1 - s)^{-x}, \quad |s| < 1.
$$

It is thus of the form $\exp[xg(s)]$ with $g(s) = \ln[1/(1-s)]$, as indicated in (6.5). Obviously, it satisfies $\phi(x+y, s) = \phi(x, s)\phi(y, s)$ for all $x, y \in \mathbb{R}$, hence the convolution type property (6.3) holds true and the operators $S$ and $\Delta$ commute. By (6.6), $D = g(\Delta) = \ln[I/(I-\Delta)]$, giving $I - \Delta = \exp(-D)$, and noticing that $\exp(-D) = S^{-1}$, we get that

$$
\Delta = I - S^{-1},
$$

i.e. $\Delta$ corresponds to the standard backward difference operator.

**Remark.** This operator $\Delta$ has various simple properties. For instance, it is easily shown that if $\Phi$ is an analytical function, then for any polynomial $B$,

$$
\Delta \Phi(B) = \Phi(B) - \Phi(B - \Delta B).
$$

Let us now pursue the discussion by restricting these polynomials $e_n$ on $\mathbb{Z}$. Note that the zeros of the $e_n$’s are kept, and the operator $\Delta$ remains internal (as $\Delta e_n$ is valued on $\mathbb{Z}$). Here is a useful property on the $e_n$’s on $\mathbb{Z}$ (where $\text{sign}(a)$ denotes the sign of $a$).

**Lemma 2.1** For $n \geq 1$ and $x, a \in \mathbb{Z}$,

$$
e_n(x + a) = e_n(x) + \text{sign}(a) \sum_{j=1+a\wedge 0}^{a\vee 0} e_{n-1}(x + j).
$$

In particular,

$$
e_n(x + 1) = \sum_{j=0}^{x\vee 0} e_{n-j}(x),
$$

$$
e_n(x) = \text{sign}(x) \sum_{j=1+x\wedge 0}^{x\vee 0} e_{n-1}(j).
$$
Proof. If \( a \geq 0 \), we write
\[
S^a - I = \sum_{j=1}^{a} (S^j - S^{j-1}) = \sum_{j=1}^{a} S^j (I - S^{-1}) = \sum_{j=1}^{a} S^j \Delta.
\]
Applying it to \( e_n(x) \) then yields
\[
e_n(x + a) - e_n(x) = \sum_{j=1}^{a} e_{n-1}(x + j).
\]
Similarly, if \( a \leq 0 \),
\[
S^a - I = \sum_{j=a+1}^{0} (S^{j-1} - S^j) = \sum_{j=a+1}^{0} S^j (S^{-1} - I) = - \sum_{j=a+1}^{0} S^j \Delta,
\]
which implies that
\[
e_n(x + a) - e_n(x) = - \sum_{j=1+a}^{0} e_{n-1}(x + j),
\]
so that (2.3) follows.

Now, choosing \( a = 1 \) in (2.3) gives
\[
e_n(x + 1) = e_n(x) + e_{n-1}(x + 1), \quad (2.6)
\]
and iterating (2.6) on \( n \), we then obtain (2.4). Finally, taking \( x = 0 \) and substituting \( x \) for \( a \) in (2.3) leads to (2.5).

Remark. By the general theory given in the Appendix, it is known that \( \Delta(ne_n(x)/x) = (n-1)e_{n-1}(x)/x, n \geq 1 \). One can prove that here, this result can be generalized as follows: for all \( r \geq 1 \),
\[
\Delta^r \left( \binom{n}{k} e_n \right) = \binom{n-r}{k} e_{n-r}, \quad 1 \leq k \leq n.
\]

2.2. Thanks to (2.3), we are going to derive a closed form expression for the associated \( \tilde{A}_n \) and \( \tilde{G}_n \) polynomials on \( \mathbb{Z} \). To begin with, let us build explicitly the inverse-kind operator \( I_u \) defined in (6.7).

Lemma 2.2 Given any \( u \in \mathbb{Z} \), then for any polynomial \( B \) on \( \mathbb{Z} \),
\[
I_u B(x) = \text{sign}(x-u) \sum_{j=1+x\land u}^{x\lor u} B(j). \quad (2.7)
\]

Proof. First, choose \( B(x) = e_n(x), n \geq 0 \). Putting \( u = x + a \), we get by (6.7) and (2.3) that
\[
I_u e_n(x) = -[e_{n+1}(x + a) - e_{n+1}(x)]
\]
\[
= -\text{sign}(a) \sum_{i=1+a \land 0}^{a \land 0} e_n(x + i)
\]
\[
= \text{sign}(x-u) \sum_{j=1+x+(u-x) \land 0}^{x+(u-x) \lor 0} e_n(j),
\]
which is equivalent to (2.7) for \( e_n(x) \).

Suppose now that \( B = \sum_{n \geq 0} \alpha_n e_n \). Using the previous result, we then find that

\[
I_u B(x) = \sum_{n \geq 0} \alpha_n I_u e_n(x)
\]

\[
= \text{sign}(x - u) \sum_{n \geq 0} \alpha_n \sum_{j=1+u}^{x \vee u} e_n(j)
\]

\[
= \text{sign}(x - u) \sum_{j=1+u}^{x \vee u} \sum_{n \geq 0} \alpha_n e_n(j),
\]

i.e. formula (2.7). \( \diamondsuit \)

An explicit formula for the \( \tilde{A}_n \) and \( \tilde{G}_n \)'s then easily follows. In view of (6.10), only \( G_n \) may be examined.

**Proposition 2.3** For \( n \geq 1 \) and \( x, u_0, u_1, \ldots \in \mathbb{Z} \),

\[
\tilde{G}_n(x \mid U) = \prod_{i=0}^{n-1} \left( \text{sign}(j_{i-1} - u_i) \sum_{j_{i-1} = 1+j_{i-1} \wedge u_i}^{j_{i-1} \vee u_i} e_0(x) \right),
\]

(2.8)

with the convention \( j_{-1} = x \). In particular, if \( x > u_0 \geq u_1 \geq u_2 \geq \ldots \),

\[
\tilde{G}_n(x \mid U) = \prod_{i=0}^{n-1} \sum_{j_i = 1+u_i}^{u_i} e_0(x),
\]

(2.9)

and if \( x < u_0 \leq u_1 \leq u_2 \leq \ldots \),

\[
\tilde{G}_n(x \mid U) = (-1)^n \prod_{i=0}^{n-1} \sum_{j_i = 1+u_i}^{u_i} e_0(x).
\]

(2.10)

**Proof.** Combining (6.17) and (2.7), we get

\[
G_n(x \mid U) = \text{sign}(x - u_0) \sum_{j=1+u_0}^{x \vee u_0} G_{n-1}(j \mid EU),
\]

which yields (2.8) by successive iterations. If \( x > u_0 \geq u_1 \geq u_2 \geq \ldots \), one has \( j_{i-1} \geq 1+u_{i-1} > u_i \) for \( 0 \leq i \leq n-1 \), so that (2.8) becomes (2.9). Similarly for (2.10). \( \diamondsuit \)

Formula (2.9) is remarkable by its simplicity. It can be extended to an arbitrary integer sequence \( U \) as follows.

**Corollary 2.4** For \( n \geq 1 \) and \( u_0, u_1, \ldots \in \mathbb{Z} \), put \( W_{n-1} = \{w_{i,n-1}, 0 \leq i \leq n-1\} \) where \( w_{i,n-1} = \max\{u_i, u_{i+1}, \ldots, u_{n-1}\} \). Then, for any integer \( x > u_{0,n-1} \),

\[
G_n(x \mid W_{n-1}) = \prod_{i=0}^{n-1} \sum_{j_i = 1+u_i}^{j_{i-1}} e_0(x).
\]

(2.11)

In particular, if \( u_0 \leq u_1 \leq u_2 \leq \ldots \), then for \( x \geq 1+u_{n-1} \) (resp. \(<\)),

\[
\tilde{G}_n(x \mid W_{n-1}) = e_n(x - u_{n-1}) \text{ (resp. } = 0)\).
\]

(2.12)
Proof. Let $J$ be the right-hand side of (2.11). Consider any given index $j_i$ in $J$. To have non-null terms, one needs $j_i \geq 1 + u_i, j_i \geq j_{i+1} \geq 1 + u_{i+1}, j_i \geq j_{i+1} \geq j_{i+2} \geq 1 + u_{i+2}, \ldots$ Thus, the only terms retained are $j_i \geq 1 + \max\{u_i, u_{i+1}, \ldots, u_{n-1}\} = 1 + w_{i,n-1}$. As a consequence, $J$ reduces to

$$J = \prod_{i=0}^{n-1} \sum_{j_i=1+w_{i,n-1}}^{j_{i+1}-1} e_0(x),$$

and this time, the $w_{i,n-1}$’s form a non-increasing sequence (with respect to $i$). From (2.9), we then deduce that if $x \leq w_{0,n-1}$, $J = \bar{C}_n(x \mid W_{n-1})$ as stated in (2.11).

Now, if the $w_i$’s are non-decreasing, $w_{i,n-1} = u_{n-1}$ for all $i$. So, if $x > u_{n-1}$, we obtain, using (6.24), that

$$J = \bar{C}_n(x \mid \{u_{n-1}, \ldots, u_n\} = e_n(x - u_{n-1}),$$

i.e. (2.12). Note that the identity $J = e_n(x - u_{n-1})$ could also be proved by a direct combinatorial argument. ◯

3 One-sided rank statistics distributions

Let us return to the rank statistics framework described in the Introduction. It is well known that the distribution of the random vector $(R_1, \ldots, R_n)$ can be obtained by counting lattice paths in a particular grid of $\mathbb{N}^2$. Specifically, the $m + n$ ordered random variables are followed one by one and, starting at point $(0,0)$, one associates a step $(0,1)$ to each occurrence of an $X_{(i)}$ and a step $(1,0)$ to each occurrence of an $Y_{(i)}$. Such a path, random by construction, is non-decreasing and always stops at point $(n,m)$. In fact, it can be simply represented by the sequence of points $\{(0,R_1),(1,R_2),\ldots,(n-1,R_n),(n,m)\}$. The number of paths from $(0,0)$ to $(n,m)$ being equal to $\binom{m+n}{n}$, one gets, for $0 \leq r_1 \leq r_2 \leq \ldots \leq r_n \leq m$,

$$P(R_1 = r_1, R_2 = r_2, \ldots, R_n = r_n) = 1/\binom{m+n}{n}.$$  \hspace{1cm} (3.1)

In this Section, our main purpose is to determine the joint left or right tail distributions of $(R_1, \ldots, R_n)$. This means that a boundary, upper or lower, will now be imposed to the lattice paths in the above grid. Denote such a lattice by $\{(0,y_0),(1,y_1)\ldots(n-1,y_{n-1})\}$ (the last step $(n,m)$ being omitted in the notation).

Let $\{s_0, \ldots, s_{n-1}\}$ be a non-decreasing family of non-negative integers with $s_{n-1} \leq m$, and consider an associated boundary $F = \{(0,s_0),(1,s_1),\ldots,(n-1,s_{n-1})\}$. Given some integer $z \in [0,s_0]$, we define by $N_{u,n}(z,s_0,s_1,\ldots,s_{n-1})$ the number of lattice paths satisfying

$$y_0 \leq s_0, y_1 \leq s_1, \ldots, y_{n-1} \leq s_{n-1} \quad \text{and} \quad z \leq y_0.$$  

When $z = 0$, this means that $F \equiv F_u$ is an upper boundary to the lattice paths in the grid. Then, one has

$$P(z \leq R_1 \leq s_0, R_2 \leq s_1, \ldots, R_n \leq s_{n-1}) = N_{u,n}(z,s_0,\ldots,s_{n-1})/\binom{m+n}{n}.$$  \hspace{1cm} (3.2)
Similarly, given some integer \( z \in [s_{n-1}, m] \), we write \( N_{l,n}(s_0, s_1, \ldots, s_{n-1}, z) \) for the number of lattice paths such that
\[
y_0 \geq s_0, y_1 \geq s_1, \ldots, y_{n-1} \geq s_{n-1} \quad \text{and} \quad y_{n-1} \leq z.
\]
When \( z = m, F \equiv F_i \) corresponds to a lower boundary in the grid. Then,
\[
P(R_1 \geq s_0, \ldots, R_{n-1} \geq s_{n-2}, z \geq R_n \geq s_{n-1}) = N_{l,n}(s_0, \ldots, s_{n-1}, z) f\left(\frac{m+n}{n}\right). \tag{3.3}
\]

3.1. In case of solely an upper horizontal boundary at level \( x - 1 \) (\( \leq m \)), then \( N_{u,n}(z = 0, x - 1, \ldots, x - 1) = N_{l,n}(0, \ldots, 0, z = x - 1) \) and this number is obviously equal to the basic negative binomial type polynomial \( e_n(x) = (x+n-1) \).

We are going to establish that more generally, the numbers \( N_{u,n} \) and \( N_{l,n} \) correspond to some negative binomial type \( \bar{G}_n \) and \( \bar{A}_n \) polynomials, respectively.

**Proposition 3.1** Given the above upper and lower boundaries (with \( n \geq 1 \)),
\[
N_{u,n}(z, s_0, \ldots, s_{n-1}) = \bar{G}_n(1 - z \mid \{-s_0, \ldots, -s_{n-1}\}), \tag{3.4}
\]
\[
N_{l,n}(s_0, \ldots, s_{n-1}, z) = \bar{A}_n(1 + z \mid \{s_0, \ldots, s_{n-1}\}), \tag{3.5}
\]
where the polynomials \( \bar{G}_n \) and \( \bar{A}_n \) are of negative binomial type.

**Proof.** Let us begin with \( N_{u,n} \). We directly see that this number can be represented under the form
\[
N_{u,n}(z, s_0, \ldots, s_{n-1}) = \sum_{y_0 = z}^{s_0} \sum_{y_1 = y_0}^{s_1} \ldots \sum_{y_{n-1} = y_{n-2}}^{s_{n-1}} 1. \tag{3.6}
\]
Choose an arbitrary integer \( a \) satisfying \( a > s_{n-1} \). Then, for \( i = 0, \ldots, n - 1 \), let us operate the change of indices \( j_i = a - y_i \) and put \( 1 + u_i = a - s_i \), where \( u_i \geq 0 \) as \( a > s_i \).

The summation (3.6) so becomes
\[
N_{u,n}(z, s_0, \ldots, s_{n-1}) = \sum_{j_0 = 1+u_0}^{a-z} \sum_{j_1 = 1+u_1}^{j_0} \ldots \sum_{j_{n-1} = 1+u_{n-1}}^{j_{n-2}} 1.
\]
Note that the \( u_i \)'s are non-increasing and \( a - z \geq a - s_0 = 1 + u_0 \). From (2.9), we then find that
\[
N_{u,n}(z, s_0, \ldots, s_{n-1}) = \bar{G}_n(a - z \mid \{u_0, \ldots, u_{n-1}\}).
\]
Inserting the expression of the \( u_i \)'s gives
\[
N_{u,n}(z, s_0, \ldots, s_{n-1}) = \bar{G}_n(a - z \mid \{a - 1 - s_0, \ldots, a - 1 - s_{n-1}\}),
\]
and applying property (6.20) finally yields (3.4).

Let us turn to the number \( N_{l,n} \). We can express it as
\[
N_{l,n}(s_0, \ldots, s_{n-1}, z) = \sum_{y_n = 0}^{y_1} \sum_{y_{n-1} = y_n}^{y_{n-1}} \ldots \sum_{y_0 = y_1}^{y_0} 1. \tag{3.7}
\]
Choose now any integer $b > x$. For $i = 0, \ldots, n-1$, let us make the transforms $j_i = b-x+y_i$ and put $1 + v_i = b - x + s_i$, where $v_i \geq 0$ by construction. Then, (3.7) becomes

$$N_{l,n}(s_0, \ldots, s_{n-1}, z) = \sum_{j_{n-1}=1+v_{n-1}}^{b-x+z} \sum_{j_{n-2}=1+v_{n-2}}^{j_{n-1}} \cdots \sum_{j_0=1+v_0}^{j_1} 1,$$

and using (2.9) and (6.10), we obtain

$$N_{l,n}(s_0, \ldots, s_{n-1}, z) = \bar{A}_n(b-x+z \mid \{v_0, \ldots, v_{n-1}\}).$$

Substituting the expression of the $v_i$’s and applying property (6.14) then leads to (3.5).

**Remarks.** (i) For an horizontal barrier at level $x-1$, the results (3.4) and (3.5) reduce to the polynomial $e_n(x)$, as expected. Indeed, from (3.4) and the explicit formula (6.24), one gets

$$N_{u,n}(0, x-1, \ldots, x-1) = \bar{G}_n(1 \mid \{-x+1, \ldots, -x+1\}) = e_n(x).$$

Similarly, using (6.10) and (6.24), $N_{l,n}(0, \ldots, 0, x-1)$ given by (3.5) becomes $e_n(x)$.

(ii) The numbers $N_{u,n}$ and $N_{l,n}$ are closely related. For instance, consider all the paths counted by $N_{l,n}$. Making a rotation of $180^\circ$ and following the paths in reversed time, we observe that

$$N_{l,n}(s_0, \ldots, s_{n-1}, z) = N_{u,n}(0, z-s_{n-1}, \ldots, z-s_0).$$

Formulas (3.4) and (3.5) can then be connected through properties of the polynomials $\bar{G}_n$ and $\bar{A}_n$. For instance, starting from (3.4) and thanks to the identity (6.10), we get

$$N_{l,n}(s_0, \ldots, s_{n-1}, z) = \bar{G}_n(1 \mid \{-z+s_{n-1}, \ldots, -z+s_0\}) = \bar{A}_n(1 \mid \{-z+s_0, \ldots, -z+s_{n-1}\}),$$

which becomes (3.5) using (6.14).

If the boundary is linear, (3.4) and (3.5) provide quite explicit formulas. Indeed, $\bar{G}_n$ and $\bar{A}_n$ can then be found explicitly from (6.24) with (6.10).

**Corollary 3.2** Let $s_i = a + bi$, $0 \leq i \leq n-1$, with $a, b \geq 0$. Then,

$$N_{u,n}(z, s_0, \ldots, s_{n-1}) = (1 - z + a) \frac{e_n(1 - z + a + b n)}{1 - z + a + b n}, \quad (3.8)$$

$$N_{l,n}(s_0, \ldots, s_{n-1}, z) = [1 + z - a - b(n - 1)] \frac{e_n(1 + z - a + b)}{1 + z - a + b} \quad (3.9)$$

In case of an arbitrary (non-decreasing) boundary, a closed form for (3.4) and (3.5) follows from (6.25) with (6.10). We note that (3.10) below with $z = 0$ corresponds to the formula (21) derived in Mohanty (1980).
Corollary 3.3

\[ N_{u,n}(z, s_0, \ldots, s_{n-1}) = \sum_{r=1}^{n} \sum_{c_r} e_{j_1} \left( 1 - z + s_0 \right) \prod_{i=2}^{r} e_{j_i} (s_{\sigma_i-1} - s_{\sigma_i-2}), \quad (3.10) \]

\[ N_{l,n}(s_0, \ldots, s_{n-1}, z) = \sum_{r=1}^{n} \sum_{c_r} e_{j_1} \left( 1 + z - s_{n-1} \right) \prod_{i=2}^{r} e_{j_i} (s_{n-1-\sigma_{i-2}} - s_{n-1-\sigma_{i-1}}), \quad (3.11) \]

where \( \sigma_0 = 0, \sigma_i = j_1 + \ldots + j_i, i \geq 1, \) and \( C_r = \{ (j_1 \geq 1, \ldots, j_r \geq 1) : \sigma_r = n \}. \)

3.2. Thanks to the properties of the \( \bar{G}_n \) and \( \bar{A}_n \)'s, the numbers of paths \( N_u \) and \( N_l \) can be easily determined by recursion. Specifically, let \( U = \{ s_i, i \geq 0 \} \) be a non-decreasing sequence of non-negative integers, and consider the successive boundaries \( F_k = \{ (0, s_0), (1, s_1) \ldots (k - 1, s_{k-1}) \} \) for \( k = 1, 2, \ldots \). Write \( F_{u,k} \) in case of an upper boundary and let \( N_{u,k}(z, U) \equiv N_{u,k}(z, s_0, \ldots, s_{k-1}) \) be the number of paths below this boundary with, as before, \( z \leq y_0 \) for some integer \( z \in [0, s_0] \); put \( N_{u,0}(z, U) = 1 \). Similarly, let \( N_{l,k}(z, U) \equiv N_{l,k}(s_0, \ldots, s_{k-1}, z) \) be the number of paths above the lower boundary \( F_{l,k} \) such that \( y_{k-1} \leq z \) for some integer \( z \geq s_{k-1} \); put \( N_{l,0}(z, U) = 1 \).

Proposition 3.4 The \( N_{u,k}(z, U) \)'s satisfy the following recursion: for any \( y \in \mathbb{Z} \),

\[ e_k(y + 1 - z) = \sum_{j=0}^{k} e_{k-j} (y - s_j) N_{u,j}(z, U), \quad k \geq 1. \quad (3.12) \]

Proof. Given any real family \( \tilde{U} \), we have by (6.22) that

\[ e_k(y + 1 - z) = \left[ S^y e_k(x) \right]_{x=1-z} = \sum_{j=0}^{k} e_{k-j}(\tilde{u}_j) \bar{G}_j(y + 1 - z | \tilde{U}). \]

Let us take \( \tilde{u}_i = y - s_i, 0 \leq i \leq k - 1 \). Thanks to (6.20), we find that

\[ e_k(y + 1 - z) = \sum_{j=0}^{k} e_{k-j}(y - s_j) \bar{G}_j(1 - z | \tilde{U}), \]

hence (3.12) by virtue of (3.4). \( \diamond \)

In particular, choosing \( y = z - 1 \) in (3.12) yields, by (6.1),

\[ 0 = \sum_{j=0}^{k} e_{k-j}(z - 1 - s_j) N_{u,j}(z, U), \quad k \geq 1. \quad (3.13) \]

On another hand, taking \( y = s_{k-1} \) gives

\[ e_k(s_{k-1} + 1 - z) = \sum_{j=0}^{k} e_{k-j}(s_{k-1} - s_j) N_{u,j}(z, U), \quad k \geq 1. \quad (3.14) \]
When \( z = 0 \), both recursions were derived by Mohanty (1980) (formulas (16) and (18)) by using a different combinatoric approach. Note that in the recursion (3.13), the terms in the sum have alternating signs, while in (3.14), all the terms are non-negative which is numerically more stable.

**Proposition 3.5** The \( N_{l,k}(z,U) \)'s satisfy the following recursion: for any \( y \in \mathbb{Z} \),

\[
\bar{A}_k(y + 1 + z | U) = \sum_{j=0}^{k} e_{k-j}(y) N_{l,j}(z,U), \quad k \geq 1.
\] (3.15)

**Proof.** By (6.15), for any real family \( \tilde{U} \),

\[
\bar{A}_k(y + 1 + z | \tilde{U}) = \sum_{j=0}^{k} e_{k-j}(y) \bar{A}_j(1 + z | \tilde{U}).
\]

So, choosing \( \tilde{u}_i = s_i, \) \( 0 \leq i \leq k - 1 \), yields (3.15), using (3.5). \( \diamond \)

In particular, since \( \bar{A}_k(s_{k-1} \mid \{s_0, \ldots, s_{k-1}\}) = 0 \) (see (6.13) with \( r = 0 \)), taking \( y = s_{k-1} - 1 - z \) in (3.15) gives the recursion

\[
0 = \sum_{j=0}^{k} e_{k-j}(s_{k-1} - 1 - z) N_{l,j}(z,U), \quad k \geq 1.
\] (3.16)

Observe that the terms in the sum are of alternating signs. Another possible method is to put \( z = -1 \) and then \( y = 1 + z \) in (3.15); using (3.5), this yields the formula

\[
N_{l,k}(z,U) = \sum_{j=0}^{k} e_{k-j}(1 + z) \bar{A}_j(0 \mid U), \quad k \geq 1,
\] (3.17)

where the \( \bar{A}_j(0 \mid U) \)'s can now be determined recursively through (6.16). Note that the terms in these sums are not of constant sign.

### 4 Two-sided rank statistics distributions

We pursue the analysis by examining the joint rectangular probabilities for the random vector \((R_1, \ldots, R_n)\). Let \( U = \{s_i, i \geq 0\} \) and \( V = \{t_i, i \geq 0\} \) be two non-decreasing sequences of non-negative integers satisfying \( s_i \leq t_i \) for all \( i \). Then, with respect to the previous grid, one has

\[
P(s_0 \leq R_1 \leq t_0, \ldots, s_{n-1} \leq R_n \leq t_{n-1}) = N_{l_u,n}(U,V)/\binom{m+n}{n};
\] (4.1)

where \( N_{l_u,n}(U,V) \equiv N_{l_u,n}(s_0, \ldots, s_{n-1}, t_0, \ldots, t_{n-1}) \) is the number of paths of the form \((0, y_0), \ldots, (n-1, y_{n-1})\) such that \( s_0 \leq y_0 \leq t_0, \ldots, s_{n-1} \leq y_{n-1} \leq t_{n-1} \).
Furthermore, that joint probability is allowed to contain, as before, the supplementary event that either \((z \leq R_1)\) for some integer \(z \in [0, t_0]\), or \((R_n \leq z)\) for some integer \(z \in [s_{n-1}, m]\). This addition implies an obvious adaptation of formula (4.1).

Let us now focus on the associated lattice path counting problem. We start by introducing a family of functions, \(\{f_n, n \geq 0\}\), defined on \(\mathbb{Z}\) by

\[
f_0 = 1 \quad \text{and} \quad f_n(x) = \binom{-x}{n}(n-1)^n, \quad n \geq 1, \tag{4.2}
\]

where we put \(\binom{-x}{n} = \binom{-x}{n}\) (resp. = 0) if \(-x \geq n\) (resp. \(-x < n\)). These \(f_n\)'s remind us, of course, of the negative binomial type polynomials \(e_n\). In fact, it is directly checked that when \(n \geq 1\), an equivalent expression is

\[
f_n(x) = e_n(x) 1_{(x<0)}, \quad n \geq 1, \tag{4.3}
\]

i.e. \(f_n\) corresponds to a truncated \(e_n\) with value 0 on \(\mathbb{Z}^+\).

Here too, we are going to consider successively the (two-sided) boundaries \(F_{lu,k} = \{(s_0, t_0), \ldots, (s_{k-1}, t_{k-1})\}\), for \(k = 1, 2, \ldots\) We will mainly derive two simple recursions (given by Propositions 4.1 and 4.2) which, once combined, allow us to determine rather easily the numbers of paths desired. Both recursions are expressed in terms of the negative binomial type polynomials \(e_n\). Their algebraic structures are roughly similar to those pointed out in the one-side case.

**Proposition 4.1**

\[
f_k(s_{k-1} - z) = \sum_{j=0}^{k} f_{k-j}(s_{k-1} - t_j - 1) N_{lu,j}(z, U, V), \quad k \geq 1. \tag{4.4}
\]

**Proof.** By definition,

\[
N_{lu,k}(z, U, V) = \sum_{y_0 = s_0 \lor z}^{t_0} \sum_{y_1 = s_1 \lor y_0}^{t_1} \ldots \sum_{y_{k-1} = s_{k-1} \lor y_{k-2}}^{t_{k-1}} 1
\]

\[
\equiv Q_0 Q_1 \ldots Q_{k-1} 1, \tag{4.5}
\]

where

\[
Q_j = \sum_{y_j = s_j \lor y_{j-1}}^{t_j} ,
\]

after putting \(y_{-1} = z\). Each \(Q_j\) can be decomposed under the form

\[
Q_j = 1_{(y_{j-1} \leq s_j)} \sum_{y_j = s_j}^{t_j} + 1_{(y_{j-1} > s_j)} \sum_{y_j = y_{j-1}}^{t_j}
\]

\[
= \sum_{y_j = s_j}^{t_j} - 1_{(y_{j-1} \geq 1 + s_j)} \sum_{y_j = s_j}^{y_{j-1} - 1} \equiv \Sigma_j - \Sigma'_j. \tag{4.6}
\]
Thus, one sees that $Q_0 = \Sigma_0 - \Sigma'_0, \ Q_0Q_1 = Q_0\Sigma_1 - Q_0\Sigma'_1 = Q_0\Sigma_1 - \Sigma_0\Sigma'_1 + \Sigma'_0\Sigma'_1, \ldots$ In

|general, making successive decompositions (4.6) in (4.5), from the right to the left, until the first appearance of a $\Sigma_j$ leads to the expansion

$$Q_0Q_1 \cdots Q_{k-1} = \sum_{j=0}^{k-1} (-1)^{k-j} Q_0 \cdots Q_{j-1} \Sigma_j \Sigma'_{j+1} \cdots \Sigma'_{k-1} + (-1)^k \Sigma'_0 \cdots \Sigma'_{k-1}.$$  \hspace{1cm} \hspace{1cm} (4.7)

By (4.5), one knows that $Q_0 \cdots Q_{j-1} = N_{lu,j}(z, U, V).$ Now, putting $z_j = 1 + j + y_j$ allows us to express $\Sigma'_j$ as

$$\Sigma'_j = 1_{(1+j+s_j \leq z_{j-1})} \sum_{z_j=1+j+s_j}^{z_{j-1}}.$$  \hspace{1cm} \hspace{1cm} (4.8)

From (2.11), (2.12), we then find that

$$\Sigma'_0 \cdots \Sigma'_{j-1} = [e_j(y_{j-1} - j + 1 - s_{j-1})]_+ = \begin{pmatrix} y_{j-1} - s_{j-1} \\ j \end{pmatrix}.$$  \hspace{1cm} (4.9)

Consider the operator $\mathcal{E}$ that maps $s_j$ (resp. $y_j$) to $s_{j+1}$ (resp. $y_{j+1}$), for all $j$ (as e.g. in (6.17)). This leads us to write that

$$\Sigma'_j \cdots \Sigma'_{k-1} 1 = \mathcal{E}^{k-j} (\Sigma'_0 \cdots \Sigma'_{k-j-2} 1) = [e_{k-j}(y_j - k + j + 2 - s_{k-1})]_+.$$  \hspace{1cm} (4.10)

Observe that $\Sigma_j$ is equivalent to $\Sigma'_j$ when $t_j+1$ is substituted for $y_{j-1}$. Writing $\Sigma'_j \cdots \Sigma'_{k-1} 1$ by means of (4.8), we so get that

$$\Sigma_j \Sigma'_j \cdots \Sigma'_{k-1} 1 = [e_{k-j}(t_j - k + j + 2 - s_{k-1})]_+.$$  \hspace{1cm} (4.11)

Now, inserting (4.9),(4.10) in (4.5), (4.7) yields

$$N_{lu,k}(z, U, V) = (-1)^k [e_k(z - k + 1 - s_{k-1})]_+ + \sum_{j=0}^{k-1} (-1)^{k-j} [e_{k-j}(t_j - k + j + 2 - s_{k-1})]_+ N_{lu,j}(z, U, V).$$  \hspace{1cm} (4.12)

From the definitions (2.1) of $e_n$ and (4.2) of $f_n$, one can then write (4.11) as in (4.4). $\diamond$

Expressing $f_n$ by means of (4.3), the recursion (4.4) becomes

$$e_k(s_{k-1} - z) 1_{(s_{k-1} \leq z)} = \sum_{j=n_k(U, V)}^k e_{k-j}(s_{k-1} - t_j - 1) N_{lu,j}(z, U, V), \quad k \geq 1.$$  \hspace{1cm} (4.13)
where $r_k(U, V) = \min\{j \geq 0 : t_j + 1 > s_{k-1}\} \leq k-1$ by definition. Now, let us consider the special case $z = 0$. Remember that $N_{lu,k}(0, U, V) \equiv N_{lu,k}(U, V)$, the number of paths inside $F_{lu,k}$. Then, (4.12) provides the simple recursion

$$0 = \sum_{j=r_k(U, V)}^{k} e_{k-j}(s_{k-1} - t_j - 1) N_{lu,j}(U, V), \quad k \geq 1. \tag{4.13}$$

The same result was obtained by Mohanty (1980) (formula (23)) using a different method.

Let us try to compare the recursions derived here for $N_{lu,k}$ and in Section 3 for $N_{u,k}$ (where the upper boundary is now defined through the sequence $\{t_i, i \geq 0\}$). We first notice that the formula (4.4) has an algebraic shape that looks like an Abelian expansion of type (6.21). In the same vein, we also see that the recursion (4.12) presents a structure roughly similar to (3.12) with $y = s_{k-1} - 1$ (and $t_i$ is substituted for $s_i$). Clearly, the algebraic differences observed come from that the basic elements in the two-sided case are the functions $f_n$ which are truncated polynomials $e_n$.

**Remark.** The decompositions (4.6) might be inserted in (4.5) from the left to the right until the first appearance of a $\Sigma_j$. That leads to the expansion

$$Q_0 Q_1 \cdots Q_{k-1} = \sum_{j=0}^{k-1} (-1)^j \Sigma_0' \cdots \Sigma_{j-1}' \Sigma_j Q_{j+1} \cdots Q_{k-1} + (-1)^k \Sigma_0' \cdots \Sigma_{k-1}'.$$

By adopting analogous arguments, one then obtains the following relations:

$$N_{lu,k}(z, U, V) = \sum_{j=0}^{k} f_j (s_j - z) N_{lu,k-j}(E^j U, E^j V), \quad k \geq 1, \tag{4.14}$$

where $s_{-1} \equiv z$ and $E^j U$ (resp. $V$) $= \{s_{j+i} \text{ (resp. } t_{j+i}, \quad i \geq 0\}, \quad j \geq 0$. The use of (4.14) requires to first evaluate the $N_{lu,k-j}(E^j U, E^j V)$’s, which can be done from the recursion (4.25) below.

**4.2.** Alternatively, define by $N_{lu,k}(U, V, z) \equiv N_{lu,k}(s_0, \ldots, s_{k-1}, t_0, \ldots, t_{k-1}, z)$ the number of paths inside the boundary $F_{lu,k}$ such that $y_{k-1} \leq z$ for some integer $z \geq s_{k-1}$. Put $N_{lu,0}(U, V, z) = 1$.

**Proposition 4.2**

$$N_{lu,k}(U, V, z) = \sum_{j=0}^{k} f_{k-j}(z - t_j) N_{lu,j}(U, V), \quad k \geq 1. \tag{4.15}$$

**Proof.** We follow a similar approach but on basis of the formula

$$N_{lu,k}(U, V, z) = \sum_{y_{k-1}=s_{k-1}}^{z} \sum_{y_{k-2}=s_{k-2}}^{y_{k-1}} \cdots \sum_{y_0=s_0}^{y_{k-2}} 1$$

$$= \chi_k \cdots \chi_0 1, \tag{4.16}$$

13
where
\[ \chi_j = \sum_{y_j=s_j}^{t_j} \chi_{j+1}, \]
with \( y_k \equiv z \). Each \( \chi_j \) is expressed as
\[ \chi_j = \sum_{y_j=s_j}^{t_j} \chi_{j+1} = \sum_{y_j=s_j}^{t_j} \chi_{j+1} \]
\( \equiv \sigma_j - \sigma'_j. \) \hspace{1cm} (4.17)

Making successive decompositions (4.17) in (4.16) from the left to the right until the first appearance of a \( \sigma_j \) then gives
\[ \chi_{k-1} \chi_{k-2} \ldots \chi_0 = \sum_{j=0}^{k-1} (-1)^{k-1-j} \sigma'_{k-1} \ldots \sigma'_{j+1} \sigma_j \chi_{j-1} \ldots \chi_0 + (-1)^k \sigma'_{k-1} \ldots \sigma'_0. \] \hspace{1cm} (4.18)

Evidently, one knows that \( \chi_{j-1} \ldots \chi_0 \equiv N_{lu,j}(U, V, z) \), and therefore,
\[ \sigma_j \chi_{j-1} \ldots \chi_0 \equiv N_{lu,j+1}(U, V, t_j) \]
\[ = N_{lu,j+1}(U, V), \] \hspace{1cm} (4.19)
the latter equality coming from that, by definition, \( N_{lu,j+1}(U, V, t_j) \) gives the number of paths inside \( F_{lu,j+1} \). Moreover, putting \( z_l = l + t_j - y_j - z_j + 2 \) for \( 0 \leq l \leq j - 1 \), one has
\[ \sigma'_{j-1} \ldots \sigma'_0 = \sum_{z_0=1}^{t_j-1-y_j} \sum_{z_1=1+t_j-1-y_j}^{z_0} \ldots \sum_{z_{j-1}=1+t_j-1-z_j}^{z_{j-2}} \]
so that by (2.11), (2.12),
\[ \sigma'_{j-1} \ldots \sigma'_0 \equiv [e_j(t_0 - y_j - j + 1)]^+ = \left( \frac{t_0 - y_j}{j} \right)^+. \] \hspace{1cm} (4.20)

Furthermore, for \( j \leq k - 2 \),
\[ \sigma'_{k-1} \ldots \sigma'_{j+1} \equiv \mathcal{E}^{j+1}(\sigma'_{k-1} \ldots \sigma'_0) \equiv [e_{k-1-j}(t_{j+1} - z - k + j + 2)]^+. \] \hspace{1cm} (4.21)

By combining (4.18), (4.19), (4.20), (4.21), we then obtain that
\[ N_{lu,k}(U, V, z) = \sum_{j=0}^{k-2} (-1)^{k-1-j} [e_{k-1-j}(t_{j+1} - z - k + j + 2)]^+ N_{lu,j+1}(U, V) \]
\[ + N_{lu,k}(U, V) + (-1)^k \left( \begin{array}{c} t_0 - z \\ k \end{array} \right)^+. \] \hspace{1cm} (4.22)

Using (2.1) and (4.2), this formula (4.22) can be rewritten under the form (4.15).
By (4.3), the relations (4.15) can also be expressed as

$$N_{lu,k}(U, V, z) = \sum_{j=k \wedge \rho_z(V)}^{k} e_{k-j}(z - t_j) N_{lu,j}(U, V), \quad k \geq 1,$$  \hspace{1cm} (4.23)

where $\rho_z(V) = \min\{j \geq 0 : t_j > z\}$. Thus, to determine $N_{lu,k}(U, V, z)$, it suffices to calculate the $N_{lu,j}(U, V)$'s, which will be provided by the recursion (4.13).

Formula (4.15), or (4.23), is especially simple. It has exactly the same structure as the formula (A.16) given by Niederhausen (1981) in the framework of an extended Sheffer (generalized Appell) theory. In that context, the family $\{N_{lu,k}(U, V, z), k \geq 0\}$ corresponds to a so-called Sheffer-sequence for the operator $\Delta$ and with basic polynomials $\{e_n, n \geq 0\}$. The reason for this appellation is that, looking at the right hand side of (4.15) and defining the polynomials

$$t_{k,i}(z) = \sum_{j=i}^{k} e_{k-j}(z - t_j) N_{lu,j}(U, V), \quad 0 \leq i \leq k,$$

one directly sees that each family $\{t_{i+n,i}, n \geq 0\}$, $i \geq 0$, is a generalized Appell sequence.

Remark. Making successive decompositions (4.17) in (4.16) from the right to the left until a first $\sigma_j$ would give the expansion

$$\chi_{k-1} \chi_{k-2} \cdots \chi_0 = \sum_{j=0}^{k-1} (-1)^j \chi_{k-1} \cdots \chi_{j+1} \sigma_j \sigma_{j-1} \cdots \sigma_0' + (-1)^k \sigma_k' \cdots \sigma_0'.$$

One can then deduce the following relations

$$f_k(z - t_0) = \sum_{j=0}^{k} f_j(s_{j-1} - t_0 - 1) N_{lu,k-j}(\mathcal{E}^j U, \mathcal{E}^j V, z), \quad k \geq 1.$$  \hspace{1cm} (4.24)

In particular, choosing $z = t_{k-1}$ in (4.24) yields

$$0 = \sum_{j=0}^{k \wedge \rho_{t_0}(U)} e_j(s_{j-1} - t_0 - 1) N_{lu,k-j}(\mathcal{E}^j U, \mathcal{E}^j V), \quad k \geq 1,$$  \hspace{1cm} (4.25)

where $\rho_{t_0}(U) = \max\{j \geq 1 : s_{j-1} < t_0 + 1\}$.

5 Some numerical illustrations

The recursions obtained in the paper are simple to implement and, in general, do not raise any serious stability problems. For instance, one can easily construct a table of quantiles for the K-S two-sample test such as given in the book by Conover (1999).

In this Section, we will apply this test to two particular data sets in biostatistics. The variable under study is the CRP level (mg/l) measured in two independent samples, one
of size $n = 8$ for patients that receive a drug (prednisone) and the other of size $m = 10$ for patients that get a placebo (CRP is a c-reactive protein that reflects acute inflammation). The measures are presented for two different days.

(i) Here are the observations obtained at day 1:

with prednisone: 2.2, 3.8, 4.6, 8.3, 8.4, 18.5, 32.0, 58.3,

with placebo: 0.9, 2.5, 3.5, 4.5, 9.4, 18.5, 26.3, 37.2, 39.5, 79.0.

The corresponding empirical distribution functions (d.f.), $F_n$ and $G_m$, are drawn in Figure 1. One sees that $d^- = \max_x \{ G_m(x) - F_n(x) \} = 7/40$, $d^+ = \max_x \{ F_n(x) - G_m(x) \} = 9/40$ and $d = \max\{d^-, d^+\} = 9/40$. As recalled in the Introduction, the d.f. of $D^-$ and $D^+$ can be expressed in terms of the left and right tail distributions of the ranks ($R_1, \ldots, R_n$).

With our data, that gives

\[
P(D^- \leq d^-) = P(R_1 \leq 1, R_2 \leq 3, R_3 \leq 4, R_4 \leq 5, R_5 \leq 6, R_6 \leq 8, R_7 \leq 9, R_8 \leq 10),
\]

\[
P(D^+ \leq d^+) = P(R_1 \geq 0, R_2 \geq 1, R_3 \geq 2, R_4 \geq 3, R_5 \geq 4, R_6 \geq 6, R_7 \geq 7, R_8 \geq 8).
\]

The d.f. $P(D \leq d)$ is obtained by requiring the ranks to be between the two boundaries above. Under $H_0$, these probabilities can be computed by means of the recurrences (3.13), (3.16) and (4.13). So, we find that the $p$-values for the three tests are $P(D^- \geq d^-) = 0.6225$, $P(D^+ \geq d^+) = 0.5135$ and $P(D \geq d) = 0.89823$. That implies that no significant difference between prednisone and placebo was detected in CRP levels at level 0.05 ($RH_0$ with both one and two-sided tests). This is not surprising as at day 1, the drug has not yet produced any effect.

(ii) Let us examine the observations obtained at day 8:

with prednisone: 0.9, 1.0, 1.1, 1.3, 2.0, 4.3, 6.4, 12.9,

with placebo: 0.4, 2.1, 3.3, 5.0, 7.7, 9.2, 26.7, 29.3, 39.4, 44.5.

Figure 2 shows the associated empirical d.f. Here, $d^- = 1/10$ and $d^+ = 21/40$. Using the
vector \((R_1, \ldots, R_8)\), one gets
\[
P(D^- \leq d^-) = P(R_1 \leq 1, R_2 \leq 2, R_3 \leq 3, R_4 \leq 4, R_5 \leq 6, R_6 \leq 7, R_7 \leq 8, R_8 \leq 9),
\]
\[
P(D^+ \leq d^+) = P(R_1 \geq 0, R_2 \geq 0, R_3 \geq 0, R_4 \geq 0, R_5 \geq 1, R_6 \geq 3, R_7 \geq 4, R_8 \geq 5).
\]
This yields the \(p\)-values \(P(D^- \geq 1/10) = 0.80808\), \(P(D^+ \geq 21/40) = 0.04756\) and \(P(D \geq 21/40) = 0.09511\). At the 0.05 level of significance, prednisone significantly reduces the CRP level \((RH_0\) with a one-sided test). With a two-sided test, no effect of prednisone is detected at the same level \((RH'_0\) at level 0.05, but \(RH_0\) at level 0.10). Thus, the effect of the drug is not yet sufficiently strong at day 8. For data at day 15 (not given here), the decision is clearly \(RH_0\) at level 0.05.

![Figure 2](image_url)

**Figure 2.** Empirical d.f. \(F_n\) (prednisone) and \(G_m\) (placebo), day 8.

(iii) To check the efficiency of the approach, we have also evaluated, under \(H_0\), the d.f. \(P(D^+ \leq d)\) \((= P(D^- \leq d))\) and \(P(D \leq d)\), \(d \in [0, 1]\). The corresponding graphs are given in Figure 3. The simplicity of the recursions obtained allows us to perform such computations in a fast and precise manner.

6 **Appendix: \(\bar{A}_n\) and \(\bar{G}_n\) polynomials**

The Appell polynomials are classical in the literature and cover the subfamilies of Hermite, Laguerre, Bernoulli and Euler polynomials (see e.g. Boas and Buck (1958) and Kaz’min (1988)). Generalized Appell polynomials, often called Sheffer polynomials, are also quite standard (see e.g. Mullin and Rota (1970) and Niederhausen (1981)). Recently, these polynomials were shown to provide a key mathematical tool in the theory of ruin for insurance (see e.g. Picard and Lefèvre (1997), De Vylder (1999), Ignatov and Kaishev (2000), (2004), Lefèvre and Picard (2006) and Picard et al. (2003)).

Although of similar construction, the Abel-Gontcharoff polynomials, introduced by Gontcharoff (1937), seem to be less popular. In applied probability, they were found to
be useful for the analysis of various stochastic epidemic processes (see e.g. Lefèvre and Picard (1990) and Ball and O’Neill (1999)). Generalized Abel-Gontcharoff polynomials were introduced by Picard and Lefèvre (1996) for a wider approach to certain first-passage time problems (see also. Lefèvre and Picard (1999) and Picard and Lefèvre (2003)).

As pointed out in Picard and Lefèvre (1996), both generalized families of polynomials can be presented in a unified framework. Moreover, they can be extended, in a similar way, to functions (named pseudopolynomials) of Appell and Abel-Gontcharoff types.

Let us mention that both families are closely related for their applications to first-crossing problems. Roughly speaking, Appell type (pseudo)polynomials are relevant, if they are, in case of an upper boundary, while Abel-Gontcharoff type (pseudo)polynomials are appropriate in case of a lower boundary. For instance, Denuit et al. (2003) showed that the simple Appell and Abel-Gontcharoff polynomials provide the joint right-tail and left-tail distributions of order statistics, respectively.

Hereafter, we summarize the main points of the theory such as introduced in Picard and Lefèvre (1996). Some complements are also given. Only the polynomial case is of interest for our purpose.

6.1. Consider the vector space $\mathcal{P}$ of polynomials $B$ on $\mathbb{R}$. Let $\{e_n, n \geq 0\}$ be a family of polynomials of degree (exactly) equal to $n$. Such a family forms a basis of $\mathcal{P}$. It is assumed that

$$e_0 = 1 \quad \text{and} \quad e_n(0) = 0, \quad n \geq 1, \quad (6.1)$$

and a linear operator $\Delta$ on $\mathcal{P}$ is defined by writing that

$$\Delta e_0 = 0 \quad \text{and} \quad \Delta e_n = e_{n-1}, \quad n \geq 1. \quad (6.2)$$

In many practical cases, the $e_n$‘s satisfy a convolution type property, i.e. for $x, y \in \mathbb{R}$,

$$e_n(x + y) = \sum_{j=0}^{n} e_j(x) e_{n-j}(y), \quad n \geq 0. \quad (6.3)$$

Figure 3. Theoretical d.f. of $D^+$ and $D$ under $H_0$, when $n = 8$ and $m = 10$. 

As indicated in Section 2.1, this property holds for the problem studied here. Let \( S^a, a \in \mathbb{R}, \) be the shift operator defined by \( (S^aB)(x) = B(x + a). \) Under (6.3), it can be proved that the operator \( \Delta \) is shift-invariant, i.e. \( S^a\Delta = \Delta S^a. \) As a consequence, any polynomial \( B \) of degree \( n \) can be expanded by the following Taylor type formula:

\[
B(x) = \sum_{j=0}^{n} \Delta^j B(a) e_j(x - a),
\]

(6.4)

where \( \Delta^j \) is the \( j \)-th iterate of the operator \( \Delta. \) Moreover, the formal generating function of the \( e_n \)'s is given by

\[
\sum_{n=0}^{\infty} e_n(x)s^n = e^{[xg(s)]},
\]

(6.5)

where \( g(s) \) is a formal series such that \( g(0) = 0. \) The operator \( \Delta \) commutes with the differentiation operator \( D, \) and it can be expressed as a power series with respect to \( D \) (and reciprocally):

\[
\Delta = g^{-1}(D) \quad \text{[reciproc.} \ D = g(\Delta)].
\]

(6.6)

6.2. Let us now introduce some kind of inverse operator to \( \Delta. \) Given any real \( u, \) such an operator, \( I_u \) say, is a linear operator on \( \mathcal{P} \) defined by writing that

\[
I_u e_n = e_{n+1} - e_{n+1}(u)e_0, \quad n \geq 0.
\]

(6.7)

Thus, \( \Delta I_u = \Delta \) \((\text{but } I_u \Delta \neq \Delta \text{ in general}).\) Note also that \( I_u e_n(u) = 0, \) giving \( I_u B(u) = 0 \) for any polynomial \( B. \)

We are then able to give the two central definitions. Let \( U = \{u_i, i \geq 0\} \) be a family of real numbers. Attached to \( U, \) the family of generalized Appell polynomials, \( \{\bar{A}_n(\ | U), n \geq 0\} \) of degree \( n, \) is defined by

\[
\bar{A}_0(\ | U) = 1 \quad \text{and} \quad \bar{A}_n(\ | U) = I_{u_{n-1}}I_{u_{n-2}} \ldots I_{u_0} e_0, \quad n \geq 1,
\]

(6.8)

while the family of generalized Abel-Gontcharoff polynomials, \( \{G_n(\ | U), n \geq 0\} \) of degree \( n, \) is defined by

\[
\bar{G}_0(\ | U) = 1 \quad \text{and} \quad \bar{G}_n(\ | U) = I_{u_{n-1}} \ldots I_{u_1} e_0, \quad n \geq 1.
\]

(6.9)

Notice that for \( n \) fixed, \( \bar{A}_n(\ | U) \) and \( G_n(\ | U) \) depend on the family \( U \) only through its first \( n \) members \( u_0, \ldots, u_{n-1}. \) Furthermore, these polynomials are linked by the identity

\[
\bar{A}_n(\ | u_0, \ldots, u_{n-1}) = \bar{G}_n(\ | u_{n-1}, \ldots, u_0), \quad n \geq 0.
\]

(6.10)

Both families of polynomials have simple operational properties. For \( \bar{A}_n(\ | U), n \geq 1, \) (6.8) implies that

\[
\bar{A}_n(\ | U) = I_{u_{n-1}} \bar{A}_{n-1}(\ | U),
\]

(6.11)

\[
\Delta^r \bar{A}_n(\ | U) = \bar{A}_{n-r}(\ | U), \quad r = 0, \ldots, n,
\]

(6.12)

\[
\Delta^r \bar{A}_n(u_{n-1-r} \ | U) = \delta_{n,r}, \quad r = 0, \ldots, n,
\]

(6.13)

\[
\bar{A}_n(x \ | x + U) = \bar{A}_n(0 \ | U), \quad x \in \mathbb{R},
\]

(6.14)
where \( x + U = \{ x + u_i, i \geq 0 \} \).

Under the condition (6.3), the Taylor type formula (6.4) yields the following expansion:
for \( x, y \in \mathbb{R} \),
\[
\bar{A}_n(x + y \mid U) = \sum_{j=0}^{n} \bar{A}_{n-j}(y \mid U) e_j(x), \quad n \geq 0. \tag{6.15}
\]
This yields a simple algorithm to evaluate the polynomials. Indeed, (6.15) with \( y = 0 \) gives \( \bar{A}_n(x \mid U) \) in function of the terms \( \bar{A}_j(0 \mid U) \), \( 0 \leq j \leq n \). These terms can then be determined recursively from the relations
\[
\delta_{n,0} = \sum_{j=0}^{n} \bar{A}_{n-j}(0 \mid U) e_j(u_{n-1}), \quad n \geq 0, \tag{6.16}
\]
which follow by choosing \( x = u_{n-1} \) and \( y = 0 \) in (6.15) and using (6.13) with \( r = 0 \).

For \( \bar{G}_n \mid U \), \( n \geq 1 \), putting \( \mathcal{E}^r U = \{ u_{r+i}, i \geq 0 \} \) where \( r \geq 0 \), one gets
\[
\bar{G}_n \mid U = I_0 \bar{G}_{n-1} \mid \mathcal{E}U, \tag{6.17}
\]
\[
\Delta^r \bar{G}_n \mid U = \bar{G}_{n-r} \mid \mathcal{E}^r U, \quad r = 0, \ldots, n, \tag{6.18}
\]
\[
\Delta^r \bar{G}_n(u_r \mid U) = \delta_{n,r}, \quad r = 0, \ldots, n, \tag{6.19}
\]
\[
\bar{G}_n(x \mid x + U) = \bar{G}_n(0 \mid U), \quad x \in \mathbb{R}. \tag{6.20}
\]

Using (6.19), one can obtain the following generalized Abelian type expansion for any polynomial \( B \) of degree \( n \):
\[
B(x) = \sum_{j=0}^{n} \Delta^j B(u_j) \bar{G}_j(x \mid U). \tag{6.21}
\]
Let us underline that the condition (6.3) is not required for (6.21). This gives us an algorithm to evaluate the polynomials. Indeed, taking \( B = e_n \) in (6.21) yields the recursion:
\[
e_n(x) = \sum_{j=0}^{n} e_{n-j}(u_j) \bar{G}_j(x \mid U), \quad n \geq 1. \tag{6.22}
\]
If (6.3) is satisfied, an alternative method for calculating \( \bar{G}_n(x \mid U) \) is to use the Taylor type formula (6.4) around \( a = 0 \). It then suffices to find the coefficients \( \bar{G}_{n-j}(0 \mid \mathcal{E}^j U) \), \( 0 \leq j \leq n \). For that, taking \( x = u_r \) in (6.4) (with \( a = 0 \)) and applying (6.19) yields the recursion
\[
\delta_{n,r} = \sum_{j=r}^{n} \bar{G}_{n-j}(0 \mid \mathcal{E}^j U) e_{j-r}(u_r), \quad r = 0, \ldots, n. \tag{6.23}
\]

6.3. A simple particular case for \( \bar{G}_n(x \mid U) \) is when \( U \) is an affine sequence, i.e. \( u_i = a + bi, i \geq 0 \). If \( x \neq u_n \), then a quite explicit formula holds true:
\[
\bar{G}_n(x \mid U) = (x - u_0) \frac{e_n(x - u_n)}{x - u_n}, \quad n \geq 0. \tag{6.24}
\]
The identity (6.24) can be proved by showing that the right hand side, $H_n(x \mid U)$ say, is a polynomial that satisfies the conditions (6.19), as $\bar{G}_n(x \mid U)$.

A closed form formula still exists in the general case (we did not realize it so far):

$$
\bar{G}_n(x \mid U) = \sum_{r=1}^{n} \sum_{C_r} e_{j_1}(x - u_0) \prod_{i=2}^{r} e_{j_i}(u_{\sigma_{i-2}} - u_{\sigma_{i-1}}), \quad n \geq 1,
$$

(6.25)

where $\sigma_i = j_1 + \ldots + j_i$, $i \geq 1$, with $\sigma_0 = 0$, and $C_r = \{(j_1 \geq 1, \ldots, j_r \geq 1) : \sigma_r = n\}$. To prove this, it suffices to apply the Taylor type formula (6.4) by chain, around the first term of the successive shifted families $U$. So, firstly $\bar{G}_n(x \mid U)$ is expanded around $u_0$ (a zero of the polynomial), which provides coefficients $\bar{G}_{n-j_1}(u_0 \mid E^{j_1}U)$ for $1 \leq j_1 \leq n$. Then, each coefficient is expanded, with $u_0$ as a variable, around $u_{j_1}$ (a zero of the new polynomial), hence the coefficients $\bar{G}_{n-j_1-j_2}(u_{j_1} \mid E^{j_1+j_2}U)$ for $1 \leq j_2 \leq n - j_1$, which are now expanded, with $u_{j_1}$ as variable, around $u_{j_1+j_2}$ and so on.

To close, let us recall that the simple Appell and Abel-Gontcharoff polynomials, denoted by $A$ and $G$, are obtained in the special situation where

$$
e_n(x) = x^n/n!, \quad n \geq 0.
$$

(6.26)

Then, $\Delta$ corresponds to the differentiation operator $D$, and $I_u$ becomes the usual integration operator (i.e. $I_u e_n(x) = \int_u^x e_n(t)dt$, $n \geq 0$).

Acknowledgment

This research was made while C. Lefèvre was visiting the Institut de Science Financière et d’Assurances, Université de Lyon. We would like to thank M. Gathy for her valuable help with the numerical illustrations. We are also grateful to Pr. S.G. Mohanty for motivating our interest in this problem through his nice paper of 1980.

References


