INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0107

POLYNOMIAL STRUCTURES IN ORDER STATISTICS DISTRIBUTIONS

M. DENUIT, C. LEFEVRE and Ph. PICARD

http://www.stat.ucl.ac.be

Polynomial structures in order statistics distributions

M. Denuit [a], **Cl. Lefèvre** [b, 1] , **Ph. Picard** [c]

[a] *Institut de Statistique, Université Catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgique*

[b] *Institut de Statistique et de Recherche Opérationnelle, Université Libre de Bruxelles, Boulevard du Triomphe C.P.210, B-1050 Bruxelles, Belgique*

[c] *Université de Lyon 1, Institut de Science Financière et d'Assurance, 43 Boulevard du 11 Novembre 1918, F-69622 Villeurbanne Cedex, France*

February 23, 2001

Abstract

This paper is concerned with the exact joint tail distributions of order statistics for i.i.d. random variables with arbitrary common distribution function. It is pointed out that the left tail distributions can be expressed in terms of Abel-Gontcharoff polynomials, while the right tail distributions can be written using Appell polynomials. This property is exploited to obtain, in a simple and unified way, closed forms and recursions for the evaluation of the corresponding probabilities. The Kolmogorov-Smirnov goodness-of-fit test for discrete models is discussed as a special case.

MSC: primary 62G30; secondary 33A65

Keywords: Order statistics; Exact distributions; Empirical function; First crossing; Kolmogorov-Smirnov statistics; Abel-Gontcharoff and Appell polynomials; Sheffer functions

1 Introduction

Considerable research has been done on the distributions, exact or approximate, of order statistics for i.i.d. samples. Comprehensive surveys are given, e.g., in the books by David (1981), Galambos (1987), Reiss (1989) and Arnold et al. (1992). The topic is also of direct relevance to the study of empirical processes. Much can be found in the masterful book by Shorack and Wellner (1986) and in the papers by Niederhausen (1981) and Csäki (1981).

The present work is concerned with the exact joint distributions of order statistics for i.i.d. random variables with arbitrary common law. Our interest will be focused essentially on the left and right tail distributions. We purpose to point out that these one-sided distributions rely on an underlying polynomial structure which is of Abel-Gontcharoff type for the left tail probabilities and of Appell type for the right tail probabilities. This property will allow us to obtain, in a simple and unified way, closed forms and recursive methods for evaluating these probabilities. For the two-sided distributions, we will briefly indicate that, as shown by

[1]Corresponding author.
E-mail address: clefevre@ulb.ac.be (Cl. Lefèvre).

Niederhausen (1981), the key algebraic tool is a sequence of functions of Sheffer type. Thus, our work provides, in a sense, a further look at the general approach developed by Niederhausen (1981). Moreover, we underline that the analysis is valid for all models, whether continuous or not.

The paper is organized as follows. Section 2 shows the polynomial structures that underly the joint one-sided distribution functions of the order statistics. Induced determinantal and recursive formulas are derived in Section 3. Section 4 deals similarly with some subsets among the order statistics. In Section 5, the problem of first crossing of an empirical distribution function through a boundary, lower or upper, is discussed and then applied to the one-sided Kolmogorov-Smirnov statistics. Section 6 examines the joint two-sided distributions. A short presentation of Abel-Gontcharoff and Appell polynomials is provided in Appendix.

2 One-sided distributions and remarkable polynomials

2.1. Starting point: the uniform continuous case. Let $U_1, \ldots, U_n$ be n independent uniform $(0,1)$ r.v.'s, and let $U_{(1)}, \ldots, U_{(n)}$ be the associated order statistics.

We are going to show that the joint left tail distributions of $(U_{(1)}, \ldots, U_{(n)})$ can be written in terms of Abel-Gontcharoff polynomials, while the joint right tail distributions can be expressed using Appell polynomials.

Lemma 2.1 *For $x \leq a_1 \leq \ldots \leq a_n \leq y \in [0,1]$,*

$$P[U_{(1)} \leq a_1, \ldots, U_{(n)} \leq a_n \text{ and } x \leq U_{(1)}] = n!(-1)^n G_n(x \mid a_1, \ldots, a_n), \tag{2.1}$$

$$P[U_{(1)} \geq a_1, \ldots, U_{(n)} \geq a_n \text{ and } y \geq U_{(n)}] = n! A_n(y \mid a_1, \ldots, a_n). \tag{2.2}$$

Proof. By definition, the former probability, denoted by p_1 say, is given by

$$\begin{aligned} p_1 &= n! P[x \leq U_1 \leq a_1, U_1 \leq U_2 \leq a_2, \ldots, U_{n-1} \leq U_n \leq a_n] \\ &= n! \int_{\nu_1=x}^{a_1} \int_{\nu_2=\nu_1}^{a_2} \ldots \int_{\nu_n=\nu_{n-1}}^{a_n} d\nu_n \ldots d\nu_2 d\nu_1. \end{aligned} \tag{2.3}$$

Comparing (2.3) with the definition (7.2) of the Abel-Gontcharoff polynomials then yields (2.1). For the latter probability, p_2 say, we have similarly

$$\begin{aligned} p_2 &= n! P[U_2 \geq U_1 \geq a_1, U_3 \geq U_2 \geq a_2, \ldots, y \geq U_n \geq a_n] \\ &= n! \int_{\nu_n=a_n}^{y} \int_{\nu_{n-1}=a_{n-1}}^{\nu_n} \ldots \int_{\nu_1=a_1}^{\nu_2} d\nu_1 \ldots d\nu_{n-1} d\nu_n. \end{aligned} \tag{2.4}$$

¿From (2.4) and the definition (7.3) of the Appell polynomials, we then obtain (2.2).□

The integral representations (2.3) and (2.4) were already given by Wald and Wolfowitz (1939). Whittle (1961) indicated that the probability in (2.4) corresponds to an Appell polynomial; see also Daniels (1945) and Niederhausen (1981) for a more general framework. Abel-Gontcharoff polynomials are much less classical and have been first used by Picard and Lefèvre

(1989) for expressing the probability in (2.3); see also Picard and Lefèvre (1996) and Ball and O'Neill (1999) in an epidemic context.

As underlined in Appendix, the two families of polynomials G and A rely on mathematical structures that are basically different. They are closely related, however, through the relation (7.6). Thus, the formulas (2.1) and (2.2) can be directly connected by using (7.6).

Remark. Let us deduce (2.2) from (2.1), for instance. By symmetry reasons, we have

$$\begin{aligned} &P[U_{(1)} \geq a_1, \ldots, U_{(n)} \geq a_n \text{ and } x \geq U_{(n)}] \\ &\quad = P[U_{(1)} \leq 1 - a_n, \ldots, U_{(n)} \leq 1 - a_1 \text{ and } 1 - x \leq U_{(1)}]. \end{aligned} \tag{2.5}$$

Expressing (2.5) with (2.1) and applying Property (7.5), we then get

$$= n!(-1)^n G_n(1 - x \mid 1 - a_n, \ldots, 1 - a_1), \tag{2.6}$$

$$= n! G_n(x \mid a_n, \ldots, a_1), \tag{2.7}$$

which becomes (2.2) by (7.6).

In the particular situation where the a_i's are defined by an affine transformation, (2.1) and (2.2) provide quite explicit expressions. Indeed, the polynomials G are then given explicitly by (7.4), and the polynomials A follow from (7.6).

Special case 2.2. Let $a_i = a + b(i-1)$, $1 \leq i \leq n$, with $a, b \geq 0$ and $a + b(n-1) \leq 1$. Then, for $x \in [0, a_1]$ and $y \in [a_n, 1]$,

$$P[U_{(1)} \leq a_1, \ldots, U_{(n)} \leq a_n \text{ and } x \leq U_{(1)}] = (a_1 - x)(a_{n+1} - x)^{n-1}, \tag{2.8}$$

$$P[U_{(1)} \geq a_1, \ldots, U_{(n)} \geq a_n \text{ and } y \geq U_{(n)}] = (y - a_n)(y - a_0)^{n-1}, \tag{2.9}$$

where $a_0 \equiv a - b$ and $a_{n+1} \equiv a + bn$.

For illustration, taking $a = b = 1/cn$, with $c \geq 1$, we find from (2.9) that

$$P[U_{(i)} \geq i/cn, 1 \leq i \leq n, \text{ and } y \geq U_{(n)}] = (y - 1/c)y^{n-1}, \tag{2.10}$$

a result obtained by Daniels (1945) when $y = 1$.

2.2. Extension to an arbitrary parent distribution. Let F be the distribution function of an arbitrary r.v. X, i.e. $F(x) = P(X \leq x)$ for $x \in \mathbf{R}$. Note that there is no continuity assumption about F. As usual, the left-continuous inverse of F is defined to be $F^{-1}(t) = \inf\{x \in \mathbf{R} \mid F(x) \geq t\}$, $t \in (0, 1)$. Obviously, $F^{-1}(t) \leq x$ if and only if $t \leq F(x)$, so that if U is a uniform r.v. on $(0, 1)$, then the r.v. $F^{-1}(U)$ has the same distribution as X.

We now show that Lemma 2.1 is easily generalized to any parent distribution. Let $X_1, \ldots, X_n$ be n i.i.d. r.v.'s with distribution function F, and let $X_{(1)}, \ldots, X_{(n)}$ be the associated order statistics. Given real numbers s_i, $1 \leq i \leq n$, we put $F_i \equiv F(s_i)$ and $F_i^- \equiv F(s_i-)$.

Proposition 2.2 *For* $s_1 \leq \ldots \leq s_n \in \mathbf{R}$,

$$P[X_{(1)} \leq s_1, \ldots, X_{(n)} \leq s_n] = n!(-1)^n G_n(0 \mid F_1, \ldots, F_n), \tag{2.11}$$

$$P[X_{(1)} > s_1, \ldots, X_{(n)} > s_n] = n! A_n(1 \mid F_1, \ldots, F_n). \tag{2.12}$$

Proof. The function F^{-1} being non-decreasing, $[X_{(1)}, \ldots, X_{(n)}]$ has the same joint distribution as $[F^{-1}(U_{(1)}), \ldots, F^{-1}(U_{(n)})]$, where $U_{(1)}, \ldots, U_{(n)}$ are the order statistics of n independent uniform $(0,1)$ r.v.'s. Therefore,

$$P[X_{(1)} \le s_1, \ldots, X_{(n)} \le s_n] = P[U_{(1)} \le F_1, \ldots, U_{(n)} \le F_n],$$

and using (2.1), we obtain (2.11). The formula (2.12) follows from (2.2) in the same way.□

The analytical derivation of (2.11), (2.12) is especially simple and emphasizes the central role of the uniform case. Other approaches are possible that consist in rewriting the probabilities of interest by an appropriate probabilistic argument (as, e.g., in Pitman (1972) and Conover (1972)). Such methods do not proceed from the uniform case and are much more intricate. Furthermore, they conceal the underlying polynomial structure of the solution.

The formulas (2.11) and (2.12) are, here too, closely related. To get (2.12) from (2.11) for example, it suffices to work with the auxiliary r.v. $Z_i = -X_i$, $1 \le i \le n$, and to use the equivalence (7.6).

Variants of (2.11), (2.12) can be directly deduced. Thus, if in the left-hand side of (2.11) the strict inequality $X_{(1)} < s_1$ is substituted for $X_{(1)} \le s_1$, passing to the limit in the right-hand side yields the formula (2.13) below. Similarly, substituting in (2.12) $X_{(1)} \ge s_1$ for $X_{(1)} > s_1$ leads to (2.14). The next formulas (2.15), (2.16) are also straighforward from Lemma 2.1.

Corollary 2.3 *For* $s_1 \le \ldots \le s_n \in \mathbf{R}$*,*

$$\begin{aligned} P[X_{(1)} < s_1, X_{(2)} \le s_2, \ldots, X_{(n)} \le s_n] &= n!(-1)^n G_n(0 \mid F_1^-, F_2, \ldots, F_n), && (2.13)\\ P[X_{(1)} \ge s_1, X_{(2)} > s_2, \ldots, X_{(n)} > s_n] &= n! A_n(1 \mid F_1^-, F_2, \ldots, F_n). && (2.14) \end{aligned}$$

Moreover, for $x < s_1$ *and* $y > s_n$*,*

$$\begin{aligned} P[X_{(1)} \le s_1, \ldots, X_{(n)} \le s_n \text{ and } x < X_{(1)}] &= n!(-1)^n G_n[F(x) \mid F_1, \ldots, F_n], && (2.15)\\ P[X_{(1)} > s_1, \ldots, X_{(n)} > s_n \text{ and } y > X_{(n)}] &= n! A_n[F(y-) \mid F_1, \ldots, F_n]. && (2.16) \end{aligned}$$

3 Recursive computation of the distributions

By (2.11) and (2.12), the evaluation of the joint tail distributions reduces to the computation of the polynomials G or A. We are going to show that central properties of the polynomials allow us to determine easily these probabilities.

To begin with, we notice that using the determinantal formula (7.13) for $G_n(0 \mid U)$, we get the following determinantal expressions for the joint tail distributions.

Proposition 3.1 *For* $s_1 \le \ldots \le s_n \in \mathbf{R}$*,*

$$\begin{aligned} P[X_{(1)} \le s_1, \ldots, X_{(n)} \le s_n] &= n! \det \left[(F_{n-i+1})^{i-j+1} / (i-j+1)! \right]_{i,j=1,\ldots,n}, && (3.1)\\ &= \det \left[\binom{i}{j-1} (F_j)^{i-j+1} \right]_{i,j=1,\ldots,n}, \end{aligned}$$

while $P[X_{(1)} > s_1, \ldots, X_{(n)} > s_n]$ *is given by (3.1) with* $1 - F_i$ *substituted for* F_{n-i+1}.

These determinantal formulas, of Hessenberg form, are standard (see, e.g., Wald and Wolfowitz (1939)) and have been extended by Steck (1969, 1971) to rectangle probabilities (see also Section 6). Their expansion gives a sum of n terms with alternating signs, which can generate problems of numerical accuracy when n is large.

Now, several recursive methods have been proposed to evaluate the probabilities. A review is provided in Chapter 9 of Shorack and Wellner (1986). These recursions have been obtained by various probabilistic arguments, sometimes astute or clumsy. We point out hereafter that they can be unified and extended in a direct way.

Specifically, given a sample of r.v.'s $X_1, \ldots, X_k$, $1 \leq k \leq n$, let $X_{(1)}, \ldots, X_{(k)}$ be the associated order statistics, and denote their tail distributions by

$$l_k = P[X_{(1)} \leq s_1, \ldots, X_{(k)} \leq s_k], \quad \text{with } l_0 \equiv 1, \tag{3.2}$$

$$r_k = P[X_{(1)} > s_1, \ldots, X_{(k)} > s_k], \quad \text{with } r_0 \equiv 1. \tag{3.3}$$

Proposition 3.2 *The l_k's satisfy the recursion, for any $x \in \mathbf{R}$,*

$$l_k = x^k - \sum_{j=0}^{k-1} \binom{k}{j} (x - F_{j+1})^{k-j} l_j, \quad k = 1, \ldots, n. \tag{3.4}$$

Proof. By (2.11) the l_k's correspond to values of G. A main interest of these polynomials is to allow an Abelian-type expansion for any polynomial. So, given any family of reals $U = \{u_i, i \geq 1\}$, the monomials x^k, with $x \in \mathbf{R}$, can be expanded by formula (7.9), yielding

$$k!G_k(x \mid U) = x^k - \sum_{j=0}^{k-1} \binom{k}{j} (u_{j+1})^{k-j} j! G_j(x \mid U), \quad k \in \mathbf{N}. \tag{3.5}$$

Let us take $u_{i+1} = x - F_{i+1}$, $0 \leq i \leq k-1$, in (3.5). From (2.11) and using Property (7.5), we then deduce (3.4). $\square$

Putting $x = F_k$ or $x = 1$ in (3.4) gives the recursions due to Steck (1971, Formula (2.3)) and Bol'shev [see Kotel'nikova and Chmaladze (1983), Formula (5)], respectively. Note that they involve sums of non-negative terms only. For $x = 0$, (3.4) yields a recursion that was extended to rectangle probabilities by Epanechnikov (1968, Formula (2)); see also Niederhausen (1981, Formula (2.8)). Remark that the terms in the sum have here alternating signs. Moreover, it can be checked that this recursion corresponds also to the expansion of the determinant (3.1) by its first column.

Proposition 3.3 *The r_k's satisfy the recursion, for any $x \in \mathbf{R}$,*

$$r_k = k!A_k(x \mid F_1, \ldots, F_k) - \sum_{j=0}^{k-1} \binom{k}{j} (x - 1)^{k-j} r_j, \quad k = 1, \ldots, n. \tag{3.6}$$

Proof. By (2.12) the r_k's are given by values of A. A key result for these polynomials is the binomial formula (7.14). Putting $u_i = F_i$, $1 \leq i \leq k$, it states that for $x, y \in \mathbf{R}$,

$$k!A_k(x \mid F_1, \ldots, F_k) = \sum_{j=0}^{k} \binom{k}{j} (x - y)^{k-j} j! A_j(y \mid F_1, \ldots, F_j), \quad k \in \mathbf{N}. \tag{3.7}$$

¿From (3.7) with $y = 1$, we then deduce (3.6).□

Notice that the recursion (3.6) for r_k relies on the preliminary evaluation of A_k. By comparison, the recursion (3.4) for l_k is thus more easily tractable.

Since $A_k(F_k \mid F_1, \ldots, F_k) = 0$, taking $x = F_k$ in (3.6) yields a simple recursion with terms of alternating signs. It corresponds to the other one-sided special case of the recursion by Epanechnikov (1968). This recursion follows also by expanding the modified determinant (3.1) by its last row.

Now, let us take $x = 1$ in (3.6), so that the problem is to find $A_k(1 \mid F_1, \ldots, F_k)$. For that, we consider again the binomial formula (3.7). Choosing $x = 1$ and $y = 0$, we obtain that r_k can be expressed as

$$r_k = \sum_{j=0}^{k} \binom{k}{j} j! A_j(0 \mid F_1, \ldots, F_j), \quad k = 1, \ldots, n. \tag{3.8}$$

To determine $A_j(0 \mid F_1, \ldots, F_j)$, $1 \le j \le k$, we write (3.7) with indices j and i substituted for k and j, and we put $x = F_j$ and $y = 0$, which yields the recursion

$$j! A_j(0 \mid F_1, \ldots, F_j) = -\sum_{i=0}^{j-1} \binom{j}{i} (F_j)^{j-i} i! A_i(0 \mid F_1, \ldots, F_i), \quad j = 1, \ldots, k. \tag{3.9}$$

That method was proposed by Wald and Wolfowitz (1939, Formulas (22), (24)). Note that the terms in the sums are not of constant sign.

Noé and Vandewiele (1968, Formulas (12), (13)), with a corrigenda by Noé (1972, formula (9)), developed the alternative recursive method below that involves non-negative terms only. First, (3.7) with $x = 1$ and $y = F_k$ gives the relation

$$r_k = \sum_{j=0}^{k-1} \binom{k}{j} (1 - F_k)^{k-j} j! A_j(F_k \mid F_1, \ldots, F_j), \quad k = 1, \ldots, n. \tag{3.10}$$

Writing (3.7) with indices j and i substituted for k and j, and taking $x = F_k$ and $y = F_{k-1}$, we get that $A_j(F_k \mid F_1, \ldots, F_j)$, $1 \le j \le k-1$, can be expressed in function of $A_i(F_{k-1} \mid F_1, \ldots, F_i)$, $1 \le i \le j$. Reiterating for $A_i(F_{k-1} \mid F_1, \ldots, F_i)$, $1 \le i \le k-2$, and continuing the procedure then leads to the recursion, for $m = 1, ..., k$,

$$j! A_j(F_m \mid F_1, \ldots, F_j) = \sum_{i=0}^{j} \binom{j}{i} (F_m - F_{m-1})^{j-i} i! A_i(F_{m-1} \mid F_1, \ldots, F_i), \quad j = 1, \ldots, m-1. \tag{3.11}$$

The above polynomial approach is purely analytic and has the advantage to be simple and systematic. Nevertheless, for certain recursions, a probabilistic argument, when adequately chosen, can work too rather quickly. This is illustrated with Bol'shev's recursion.

Remark. Denote by E_k the event $[X_{(1)} \le s_1, \ldots, X_{(k)} \le s_k]$, and let $\bar{E}_k$ be the complementary event. Clearly, $P(\bar{E}_k) = 1 - l_k$, and

$$P(\bar{E}_k) = \sum_{j=0}^{k-1} P[X_{(1)} \le s_1, \ldots, X_{(j)} \le s_j, X_{(j+1)} > s_{j+1}]. \tag{3.12}$$

The event $[X_{(1)} \leq s_1, \ldots, X_{(j)} \leq s_j, X_{(j+1)} > s_{j+1}]$, $0 \leq j \leq k-1$, means that any fixed group of j r.v.'s among the k possible variables (in number $\binom{k}{j}$) verifies the event E_j (with probability l_j), and the remaining $k-j$ variables take values larger than s_{j+1} (with probability $(1-F_{j+1})^{k-j}$). Thus, we get from (3.12) that

$$l_k = 1 - \sum_{j=0}^{k-1} \binom{k}{j} (1-F_{j+1})^{k-j} l_j, \quad k = 1, \ldots, n,$$

that is (3.4) with $x = 1$.

4 Extension to subsets of order statistics

In this section, we will focus our attention to some subsets among the n order statistics $(X_{(1)}, \ldots, X_{(n)})$. More precisely, let us consider a partial vector $(X_{(k+1)}, \ldots, X_{(m)})$, with $1 \leq k+1 \leq m \leq n$. By Proposition 2.2, we directly obtain that its tail distributions have the following polynomial expression.

Corollary 4.1 *For $s_{k+1} \leq \ldots \leq s_m \in \mathbf{R}$,*

$$P[X_{(k+1)} \leq s_{k+1}, \ldots, X_{(m)} \leq s_m] = n!(-1)^n G_n(0 \mid F_{k+1}, \ldots, F_{k+1}, F_{k+2}, \ldots, F_m, 1, \ldots, 1), \tag{4.1}$$

$$P[X_{(k+1)} > s_{k+1}, \ldots, X_{(m)} > s_m] = n! A_n(1 \mid 0, \ldots, 0, F_{k+1}, \ldots, F_m, F_m, \ldots, F_m). \tag{4.2}$$

These probabilities can be evaluated by the methods described before. We are going to derive a simplified procedure that accounts for their specific forms. Firstly, we show that for the first m order statistics, $1 \leq m \leq n$, (4.1) and (4.2) can be expanded in terms of the l_j's and r_j's introduced in (3.2) and (3.3), for $j = 1, ..., m-1$ only.

Proposition 4.2 *For $s_1 \leq \ldots \leq s_m \in \mathbf{R}$,*

$$P[X_{(1)} \leq s_1, \ldots, X_{(m)} \leq s_m] = 1 - \sum_{j=0}^{m-1} \binom{n}{j} (1-F_{j+1})^{n-j} l_j, \tag{4.3}$$

$$P[X_{(1)} > s_1, \ldots, X_{(m)} > s_m] = \sum_{j=0}^{m-1} \binom{n}{j} (1-F_m)^{n-j} r_j \left[\sum_{i=j}^{m-1} \binom{n-j}{i-j} (-1)^{i-j} \right]. \tag{4.4}$$

Proof. By (4.1) with $k = 0$, we have

$$P[X_{(1)} \leq s_1, \ldots, X_{(m)} \leq s_m] = n!(-1)^n G_n(0 \mid F_1, \ldots, F_m, 1, \ldots, 1). \tag{4.5}$$

Let us consider the Abelian expansion of the polynomial $(1-x)^n$, with $x \in \mathbf{R}$. By (7.9), we get, for any family $U = \{u_1, \ldots, u_n\}$,

$$n!(-1)^n G_n(x \mid U) = (1-x)^n - \sum_{j=0}^{n-1} \binom{n}{j} (1-u_{j+1})^{n-j} j!(-1)^j G_j(x \mid U). \tag{4.6}$$

Therefore, choosing in (4.6) $U = \{F_1, \dots, F_m, 1, \dots, 1\}$ and $x = 0$ yields

$$n!(-1)^n G_n(0 \mid F_1, \dots, F_m, 1, \dots, 1) = 1 - \sum_{j=0}^{m-1} \binom{n}{j} (1 - F_{j+1})^{n-j} j!(-1)^j G_j(0 \mid F_1, \dots, F_j). \tag{4.7}$$

¿From (4.5) and (4.7), and using (3.2) and (2.11) for l_j, we then deduce (4.3).

By (4.2) with $k = 0$, we have

$$P[X_{(1)} > s_1, \dots, X_{(m)} > s_m] = n! A_n(1 \mid F_1, \dots, F_m, F_m, \dots, F_m). \tag{4.8}$$

Let us consider the binomial formula (3.7). Putting $x = 1$ and $y = F_m$, we get, for any family $U = \{u_1, \dots, u_n\}$,

$$n! A_n(1 \mid U) = \sum_{j=0}^{n} \binom{n}{j} (1 - F_m)^{n-j} j! A_j(F_m \mid U). \tag{4.9}$$

Thus, taking $U = \{F_1, \dots, F_m, F_m, \dots, F_m\}$ in (4.9), we obtain, since $A_m(F_m \mid F_1, \dots, F_m) = 0$, that

$$n! A_n(1 \mid F_1, \dots, F_m, F_m, \dots, F_m) = \sum_{j=0}^{m-1} \binom{n}{j} (1 - F_m)^{n-j} j! A_j(F_m \mid F_1, \dots, F_j). \tag{4.10}$$

Now, to determine the $A_j(F_m \mid F_1, \dots, F_j)$'s, we apply (3.6) with $x = F_m$ and j substituted for k, which yields

$$j! A_j(F_m \mid F_1, \dots, F_j) = \sum_{i=0}^{j} \binom{j}{i} (F_m - 1)^{j-i} r_i. \tag{4.11}$$

Inserting (4.11) in (4.10) and permuting the tho sums involved, we then deduce that (4.8) reduces to (4.4).□

Note that in the special case where $m = n$, (4.3) becomes Bolshev's recursion [(3.4) with $k = n$ and $x = 1$], and (4.4) becomes Epanechnikov's recursion [(3.6) with $k = n$ and $x = F_n$].

We now show that for the last $n - k$ order statistics, $1 \leq n - k \leq n$, (4.1) and (4.2) can be calculated by a recurrence of order $n - k - 1$ only. Specifically, given a sample $X_1, \dots, X_{k+1}, \dots, X_i$ of size i, $1 \leq k + 1 \leq i \leq n$, let $X_{(1)}, \dots, X_{(k+1)}, \dots, X_{(i)}$ be the associated order statistics, and denote the tail distributions of the last $i - k$ order statistics by

$$f_i = P[X_{(k+1)} \leq s_{k+1}, \dots, X_{(i)} \leq s_i], \tag{4.12}$$
$$g_i = P[X_{(k+1)} > s_{k+1}, \dots, X_{(i)} > s_i]. \tag{4.13}$$

Proposition 4.3 *The f_i's satisfy the recursion*

$$f_i = (F_{k+1})^i - \sum_{j=k+1}^{i-1} \binom{i}{j} (F_{k+1} - F_{j+1})^{i-j} f_j, \quad i = k + 1, \dots, n, \tag{4.14}$$

and the g_i's satisfy the recursion

$$1 - g_i = (F_i)^i - \sum_{j=k+1}^{i-1} \binom{i}{j} (F_i - 1)^{i-j} (1 - g_j), \quad i = k+1, \ldots, n. \tag{4.15}$$

Proof. We will proceed as for (4.3) and (4.4). First, by (4.1),

$$f_i = i!(-1)^i G_i(0 \mid F_{k+1}, \ldots, F_{k+1}, F_{k+1}, \ldots, F_i). \tag{4.16}$$

Let us consider the Abelian expansion of the polynomial $(F_{k+1} - x)^i$, with $x \in \mathbf{R}$. Then, taking $U = \{F_{k+1}, \ldots, F_{k+1}, F_{k+1}, \ldots, F_i\}$ and putting $x = 0$, we get

$$(F_{k+1})^i = \sum_{j=k+1}^{i} \binom{i}{j} (F_{k+1} - F_{j+1})^{i-j} (-1)^j j! G_j(0 \mid F_{k+1}, \ldots, F_{k+1}, F_{k+1}, \ldots, F_j),$$

and using (4.16), we deduce the recursion (4.14).

Now, by (4.2) we have

$$g_i = i! A_i(1 \mid 0, \ldots, 0, F_{k+1}, \ldots, F_i). \tag{4.17}$$

Let us consider the binomial formula (3.7) with index i substituted for k. Choosing $U = \{0, \ldots, 0, F_{k+1}, \ldots, F_i\}$, $x = F_i$ and $y = 1$ then yields, using (4.17),

$$0 = \sum_{j=0}^{k} \binom{i}{j} (F_i - 1)^{i-j} + \sum_{j=k+1}^{i} \binom{i}{j} (F_i - 1)^{i-j} g_j,$$

or equivalently,

$$0 = (F_i)^i + \sum_{j=k+1}^{i} \binom{i}{j} (F_i - 1)^{i-j} (g_j - 1),$$

hence the recursion (4.15).□

In the case where $k = 0$, then $f_i = l_i$ and $g_i = r_i$ for all i, and we see that (4.14) corresponds to the recursion (3.4) for $x = F_1$ (with $k = i$), while (4.15) reduces to Epanechnikov's recursion [(3.6) with $k = i$ and $x = F_i$].

Finally, going back to the partial vector $(X_{(k+1)}, \ldots, X_{(m)})$, we show that (4.1) and (4.2) can be computed from the f_i's and g_i's for $i = k+1, \ldots, m-1$.

Proposition 4.4 *For $s_{k+1} \leq \ldots \leq s_m \in \mathbf{R}$,*

$$P[X_{(k+1)} \leq s_{k+1}, \ldots, X_{(m)} \leq s_m] = 1 - \sum_{j=0}^{k} \binom{n}{j} (1 - F_{k+1})^{n-j} (F_{k+1})^j - \sum_{j=k+1}^{m-1} \binom{n}{j} (1 - F_{j+1})^{n-j} f_j, \tag{4.18}$$

$$P[X_{(k+1)} > s_{k+1}, \ldots, X_{(m)} > s_m] = \sum_{j=0}^{k} \binom{n}{j} (1 - F_m)^{n-j} d_j + \sum_{j=k+1}^{m-1} \binom{n}{j} (1 - F_m)^{n-j} d_j g_j, \tag{4.19}$$

where

$$d_j = \sum_{i=j}^{m-1} \binom{n-j}{i-j} (-1)^{i-j}, \quad j = 0, \ldots, m-1.$$

Proof. To evaluate (4.1), we may apply (4.3) with $s_1 = \ldots = s_k = s_{k+1}$. It then remains to observe that in this case, $l_j = (F_{k+1})^j$ for $j = 0, \ldots, k$, and $l_j = f_j$ for $j = k+1, \ldots, m-1$. As for (4.2), we use (4.4) with $s_1 = \ldots = s_k = -\infty$, and we notice that here, $r_j = 1$ for $j = 0, \ldots, k$, and $r_j = g_j$ for $j = k+1, \ldots, m-1$.□

For illustration, we go on with the study of the particular situation, started in Special case 2.2, where the parent distribution is uniform continuous on $(0,1)$, and the boundary formed by the a_i's increases linearly.

Special case 4.5. Let $a_i = a + b(i-1)$, $k+1 \le i \le m$, with $a_{k+1} \ge 0$, $b \ge 0$ and $a_m \le 1$. Then, when $k = 0$,

$$\begin{aligned} P[U_{(1)} \le a_1, \ldots, U_{(m)} \le a_m] &= 1 - a_1 \sum_{j=0}^{m-1} \binom{n}{j} (1-a_{j+1})^{n-j} (a_{j+1})^{j-1} && (4.20) \\ &= a_1 \sum_{j=m}^{n} \binom{n}{j} (1-a_{j+1})^{n-j} (a_{j+1})^{j-1}, && (4.21) \end{aligned}$$

$$P[U_{(1)} \ge a_1, \ldots, U_{(m)} \ge a_m] = \sum_{j=0}^{m-1} \binom{n}{j} (1-a_m)^{n-j} (a_m - a_j)(a_m - a_0)^{j-1}. \tag{4.22}$$

When $m = n$, the left and right-tail distributions follow directly from (2.5) with (4.20), (4.21) and (4.22). In general,

$$P[U_{(k+1)} \le a_{k+1}, \ldots, U_{(m)} \le a_m] = 1 - \sum_{j=0}^{k} \binom{n}{j} (1-a_{k+1})^{n-j} (a_{k+1})^j - \sum_{j=k+1}^{m-1} \binom{n}{j} (1-a_{j+1})^{n-j} f_j^*, \tag{4.23}$$

where for $j = k+1, \ldots, m-1$,

$$f_j^* = \sum_{i=0}^{j-k-1} \binom{j}{i} (a_{k+1})^{j-i} (a_{j-i-1} - a_{k+1})(a_{j+1} - a_{k+1})^{i-1}; \tag{4.24}$$

$$P[U_{(k+1)} \ge a_{k+1}, \ldots, U_{(m)} \ge a_m] = \sum_{j=0}^{k} \binom{n}{j} (1-a_m)^{n-j} (a_m)^j + \sum_{j=k+1}^{m-1} \binom{n}{j} (1-a_m)^{n-j} g_j^*, \tag{4.25}$$

where for $j = k+1, \ldots, m-1$,

$$\begin{aligned} g_j^* &= (a_m)^j - (a_m - a_j) \sum_{i=0}^{j-k-1} \binom{j}{i} (a_{j-i})^{j-i} (a_m - a_{j-i+1})^{i-1} && (4.26) \\ &= (a_m - a_j) \sum_{i=j-k}^{j} \binom{j}{i} (a_{j-i})^{j-i} (a_m - a_{j-i+1})^{i-1}. && (4.27) \end{aligned}$$

Proof. The formula (4.20) is immediate from (4.3) since by (2.8) we have

$$l_j = a_1(a_{j+1})^{j-1}, \quad j = 0, \ldots, n.$$

Moreover, taking in (4.6) $U = \{a_1, \ldots, a_n\}$ and $x = 0$, we get

$$1 = \sum_{j=0}^{n} \binom{n}{j} (1 - a_{j+1})^{n-j} l_j,$$

so that (4.20) is equivalent to (4.21).

The tail probability in (4.22) can be expressed as in (4.4), and using (2.9) for r_j, $j = 0, \ldots, n$. However, the formula given in (4.22) is more compact and follows easily from (4.10) by noting that by (7.4) and (7.6),

$$A_j(a_m \mid a_1, \ldots, a_j) = (a_m - a_j)(a_m - a_0)^{j-1}/j!, \quad j = 0, \ldots, n.$$

For (4.23) and (4.24), it suffices to apply (4.18) and to evaluate f_j^*, $j = k+1, \ldots, m-1$, by combining (2.5) and (4.22).

The tail probability in (4.25) could be computed from (4.19). It is simpler to use again (4.10), which leads to the formula provided in (4.25) where g_j^*, $j = k+1, \ldots, m-1$, is given by

$$g_j^* = j! A_j(a_m \mid 0, \ldots, 0, a_{k+1}, \ldots, a_j). \tag{4.28}$$

Now, we rewrite (4.28) as

$$g_j^* = j!(-1)^j G_j(1 - a_m \mid 1 - a_j, \ldots, 1 - a_{k+1}, 1, \ldots, 1),$$

and from (4.6), we then obtain

$$g_j^* = (a_m)^j - \sum_{i=0}^{j-k-1} \binom{j}{i} (a_{j-i})^{j-i} i!(-1)^i G_i(1 - a_m \mid 1 - a_j, \ldots, 1 - a_{j-i+1}),$$

which reduces to (4.26) by linearity of the a_i's. From (4.6), we see also that (4.26) and (4.27) are equivalent.□

This special case has received much attention in the literature (see, e.g., Takács (1967), Eicker (1970), Niederhausen (1981) and Shorack and Wellner (1986)). We mention that (4.20) and/or (4.21) are given, e.g., in Birnbaum and Pyke (1958) and Dempster (1959) (and often named as Dempster's formula), and that (4.25) with (4.27) or (4.26) correspond, e.g., to (1.4) or (1.17a) in Eicker (1970). Another example is examined below.

Remark. Let us rederive a result by Chang (1955) which states that for $0 < c < 1$, and denoting by $[cn]$ the integer part of cn $(< n)$,

$$P[U_{(i)} \leq (i-1)/cn, 2 \leq i \leq [cn]+1] = 1 - \left(1 - \frac{1}{cn}\right)^n - \sum_{j=1}^{[cn]} \binom{n}{j} \left(1 - \frac{j}{cn}\right)^{n-j} \frac{(j-1)^{j-1}}{(cn)^j}. \tag{4.29}$$

For that, we use (4.23) and (4.24) with $k = 1$, $m = [cn] + 1$ and $a = 0$, $b = 1/cn$. We then get

$$\begin{aligned} P[U_{(i)} &\le (i-1)/cn, 2 \le i \le [cn]+1] \\ &= 1 - \left(1 - \frac{1}{cn}\right)^n - n\left(1 - \frac{1}{cn}\right)^{n-1}\frac{1}{cn} - \sum_{j=2}^{[cn]} \binom{n}{j}\left(1 - \frac{j}{cn}\right)^{n-j} f_j^*, \end{aligned} \tag{4.30}$$

where for $j = 2, \dots, [cn]$,

$$f_j^* = \frac{1}{(cn)^j} \sum_{i=0}^{j-2} \binom{j}{i} (j-1)^{i-1}(j-i-1).$$

Now, we observe that

$$\sum_{i=0}^{j} \binom{j}{i} (j-1)^{i}(j-i-1) = 0,$$

so that f_j^* can be rewritten as

$$f_j^* = (j-1)^{j-1}/(cn)^j. \tag{4.31}$$

Inserting (4.31) in (4.30) then yields (4.29).

5 First crossing for an empirical distribution function

Let $X_1, X_2, \dots, X_n$ be n i.i.d. r.v.'s with distribution function F (continuous or not). The associated distribution function is $F_n(x) = (1/n)$ (the number of X_i which are $\le x$), $x \in \mathbf{R}$. The problem of first crossing of F_n through a boundary has often been discussed, mainly when the boundary is linear. Important applications are related to Kolmogorov-Smirnov and similar statistics for goodness-of-fit tests (see, e.g., Shorack and Wellner (1986)).

Hereafter we examine a few representative questions in this context which will allow us to enlight the relevance of the previous theory. We mention that various other illustrations are possible but will not be developed for saving space.

5.1. Lower versus upper boundaries. Let $s(x)$ be a given real function that plays the role of lower boundary for $F_n(x)$. We assume that $s(x)$ is nondecreasing and right-continuous. This restriction is made for clarity and could be largely weakened. To avoid trivialities, we further assume that $s(x) \le 1$ for all x.

F_n is said to cross s as soon as $F_n(x) \le s(x)$ for some x. Note that this event can happen only at a step of F_n. Let L denote the first-crossing level of s by F_n. Since $s \le 1$, $L = 1$ means that crossing does not occur. Otherwise, L is a discrete r. v. with possible values m/n, $0 \le m \le n-1$. We are going to derive the distribution of L; for that, we define the points

$$s_i^{(l)} = s^{-1}\left(\frac{i-1}{n}\right), \quad \text{with} \quad F_i^{(l)} = F(s_i^{(l)}), \quad i = 1, \dots, n, \tag{5.1}$$

where, as usual, $s^{-1}(x) = \inf\{y \in \mathbf{R} : s(y) \ge x\}$.

Corollary 5.1 *In the above notation,*

$$P(L = m/n) = \binom{n}{m}(1 - F_{m+1}^{(l)})^{n-m} l_m, \quad m = 0, \ldots, n. \tag{5.2}$$

Proof. For $m = n$, we have

$$\begin{aligned} P(L = 1) &= P[F_n(x) > s(x) \text{ for all } x] \\ &= P[X_{(1)} \leq s_1^{(l)}, \ldots, X_{(n)} \leq s_n^{(l)}], \end{aligned} \tag{5.3}$$

which corresponds to l_n by (3.2). For $m \leq n-1$, we see that

$$P(L = m/n) = P[X_{(1)} \leq s_1^{(l)}, \ldots, X_{(m)} \leq s_m^{(l)}, X_{(m+1)} > s_{m+1}^{(l)}], \tag{5.4}$$

or equivalently,

$$P(L = m/n) = P[X_{(1)} \leq s_1^{(l)}, \ldots, X_{(m)} \leq s_m^{(l)}] - P[X_{(1)} \leq s_1^{(l)}, \ldots, X_{(m)} \leq s_m^{(l)}, X_{(m+1)} \leq s_{m+1}^{(l)}],$$

which becomes (5.2) by using (4.3).□

Consider the particular situation where F is uniform continuous on $(0, 1)$, and s is a nondecreasing straight line. We put $s(x) = (x - a)/bn$ with $a, b \geq 0$ and $a + b(n-1) \leq 1$, yielding $s_i^{(l)} = a + b(i-1) \equiv a_i$ say, for $1 \leq i \leq n$. From (5.2) and (2.8), we then deduce the following explicit formula for the law of L:

$$P(L = m/n) = a_1 \binom{n}{m} (1 - a_{m+1})^{n-m} (a_{m+1})^{m-1}, \quad m = 0, \ldots, n. \tag{5.5}$$

This distribution is called a quasi-binomial distribution by Consul (1974); see also, e.g., Charalambides (1990) and Berg and Mutafchiev (1990).

Alternatively, we can say that F_n crosses s strictly as soon as $F_n(x) < s(x)$ for some x. The first strict crossing level is denoted by Λ. Define now the points

$$\sigma_i^{(l)} = s^{-1}\left(\frac{i-1}{n}+\right), \text{ with } \Phi_i^{(l)} = F(\sigma_i^{(l)}), \quad i = 1, \ldots, n, \tag{5.6}$$

where we put $s^{-1}(x+) = \inf\{y \in \mathbf{R} : s(y) > x\}$. Then, we have

$$\begin{aligned} P[F_n(x) \geq s(x) \text{ for all } x] &= P[X_{(1)} \leq \sigma_1^{(l)}, \ldots, X_{(n)} \leq \sigma_n^{(l)}] \\ &= n!(-1)^n G_n(0 \mid \Phi_1^{(l)}, \ldots, \Phi_n^{(l)}). \end{aligned} \tag{5.7}$$

More generally, arguing as before, we obtain the following result.

Corollary 5.2 *The distribution of Λ is given by (5.2) with $\Phi_i^{(l)}$ substituted for $F_i^{(l)}$.*

Now, let $s(x)$ be a given real function that is viewed as an upper boundary for $F_n(x)$. $s(x)$ is taken again to be nondecreasing and right-continuous, and we further assume that $s(x) \geq 0$ for all x.

F_n is said to cross s strictly as soon as $F_n(x) > s(x)$ for some x. Note that this event can happen only at a jump of F_n. Let T denote the first strict crossing time of s by F_n. T is a defective r.v. valued in $\mathbf{R}$, with $T = \infty$ meaning that reaching does not occur. To obtain the distribution of T, we define the points

$$s_i^{(u)} = s^{-1}\left(\frac{i}{n}\right), \quad \text{with} \quad F_i^{(u)-} = F(s_i^{(u)}-), \quad i = 1, \dots, n. \tag{5.8}$$

and we denote, for any $t \in \mathbf{R}$,

$$n_t = \max\{i \leq n : s_i^{(u)} \leq t\}. \tag{5.9}$$

Corollary 5.3 *Putting* $r_j^- = j!A_j(1 \mid F_1^{(u)-}, \dots, F_j^{(u)-})$, $1 \leq j \leq n$,

$$P(T > t) = \sum_{j=0}^{n_t} \binom{n}{j} [1 - F(t)]^{n-j} r_j^- \left[\sum_{i=j}^{n_t} \binom{n-j}{i-j} (-1)^{i-j}\right], \quad t \in \mathbf{R}. \tag{5.10}$$

Proof. For $t = \infty$, we have

$$\begin{aligned} P(T = \infty) &= P[F_n(x) \leq s(x) \text{ for all } x] \\ &= P[X_{(1)} \geq s_1^{(u)}, \dots, X_{(n)} \geq s_n^{(u)}], \end{aligned} \tag{5.11}$$

which is equal to r_n^- by (2.14). Moreover, we see that (5.10) when $t = \infty$ gives r_n^- too, after substituting $n_\infty = n$ and $F(\infty) = 1$. For $t < \infty$, we write

$$P(T > t) = \sum_{j=0}^{n_t} P[X_{(1)} \geq s_1^{(u)}, \dots, X_{(j)} \geq s_j^{(u)} \text{ and } X_{(j)} \leq t, X_{(j+1)} > t], \tag{5.12}$$

or equivalently,

$$\begin{aligned} P(T > t) \;=\; \sum_{j=0}^{n_t} \{ & P[X_{(1)} \geq s_1^{(u)}, \dots, X_{(j-1)} \geq s_{j-1}^{(u)}, X_{(j)} \geq s_j^{(u)}, X_{(j+1)} > t] \\ & - P[X_{(1)} \geq s_1^{(u)}, \dots, X_{(j-1)} \geq s_{j-1}^{(u)}, X_{(j)} > t, X_{(j+1)} > t]\}. \end{aligned}$$

Arguing as for (4.4), we then find that

$$P(T > t) = \sum_{j=0}^{n_t} \binom{n}{j} [1 - F(t)]^{n-j} \left\{ r_j^- - j!A_j[1 \mid F_1^{(u)-}, \dots, F_{j-1}^{(u)-}, F(t)] \right\}. \tag{5.13}$$

But from the integral representation of A_j, we get that

$$r_j^- = j!A_j[F(t) \mid F_1^{(u)-}, \dots, F_{j-1}^{(u)-}] + j!A_j[1 \mid F_1^{(u)-}, \dots, F(t)],$$

so that (5.13) reduces to

$$P(T > t) = \sum_{j=0}^{n_t} \binom{n}{j} [1 - F(t)]^{n-j} A_j[F(t) \mid F_1^{(u)-}, \ldots, F_j^{(u)-}]. \tag{5.14}$$

Expanding $A_j[F(t) \mid F_1^{(u)-}, \ldots, F_j^{(u)-}]$ in (5.14) by the binomial formula yields

$$P(T > t) = \sum_{j=0}^{n_t} \binom{n}{j} [1 - F(t)]^{n-j} \left\{ \sum_{i=0}^{j} \binom{j}{i} [F(t) - 1]^{i-j} r_i^- \right\},$$

which becomes (5.10) after permutation of the two sums.□

A result of this type has been derived by Stadje (1993) in the special case where F is continuous. In that work, however, the Appell polynomial structure is not recognized, which leads to quite involved formulas.

The particular situation with F continuous uniform on $(0, 1)$ and s linear nondecreasing, has been investigated extensively; see, e.g., Eicker (1970). Notice that a totally explicit expression can be directly deduced using (2.9).

Remark. Alternatively, we say that F_n crosses s as soon as $F_n(x) \geq s(x)$ for some x. The first crossing time is denoted by τ. Define the points

$$\sigma_i^{(u)} = s^{-1}\left(\frac{i}{n}+\right), \quad \text{with} \quad \Phi_i^{(u)} = F(\sigma_i^{(u)}), \quad i = 1, \ldots, n. \tag{5.15}$$

A priori, it could be natural to conjecture that

$$\begin{aligned} P[F_n(x) < s(x) \text{ for all } x] &= P[X_{(1)} > \sigma_1^{(u)}, \ldots, X_{(n)} > \sigma_n^{(u)}] \\ &= n! A_n(1 \mid \Phi_1^{(u)}, \ldots, \Phi_n^{(u)}), \end{aligned} \tag{5.16}$$

and more generally, $P(\tau > t)$, $t \in \mathbf{R}$, is given by (5.10) with $j!A_j(1 \mid \Phi_1^{(u)}, \ldots, \Phi_n^{(u)})$ substituted for r_j^-. These results, however, are not true in general. Indeed, by the assumption of right-continuity of F, we see, for example, that $X_{(i)} = \sigma_i^{(u)}$ for some i may be in agreement with $F_n(x) < s(x)$ for all x. There does not seem to exist any simple expression for the law of τ.

5.2. Kolmogorov-Smirnov and similar statistics. Let us consider a goodness-of-fit test with null hypothesis $H_0 : F(x) = H(x)$ for all $x \in \mathbf{R}$, where $F(x)$ is the unknown population distribution function and $H(x)$ is the hypothesized distribution function with all parameters specified. The two possible one-sided alternatives are either $H_1^- : F(x) < H(x)$ for some x, or $H_1^+ : F(x) > H(x)$ for some x, the two-sided alternative being $H_1 : F(x) \neq H(x)$ for some x. We underline that H might be continuous, discrete or mixed.

It is well-known that in all cases, the empirical distribution function F_n is a strongly consistent estimator of F, uniformly in x. Thus, appropriate measures of discrepancy betwen H and F (through F_n) are the one-sided Kolmogorov-Smirnov statistics

$$\begin{aligned} D_n^- &= \sup_x \{H(x) - F_n(x)\} \ \text{ for testing } \ H_1^-, \qquad (5.17) \\ D_n^+ &= \sup_x \{F_n(x) - H(x)\} \ \text{ for testing } \ H_1^+, \end{aligned}$$

the two-sided statistic for H_1 being $D_n = \max(D_n^-, D_n^+)$.

Historically, the Kolmogorov-Smirnov statistics have been proposed for tests of fit of continuous models. In fact, when H is continuous, these statistics are strictly distribution-free under H_0. Then $F = H$ can be taken as continuous uniform on $(0,1)$, and the tests rely essentially on the computation of some probabilities for uniform order statistics (see, e.g., Corollary 5.4 below).

Later, several authors have recommended the use of the Kolmogorov-Smirnov statistics for all models, even discrete (see, e.g., Conover (1972, 1980), Horn (1977), Pettitt and Stephens (1977), Wood and Altavela (1978) and Gleser (1985)). When H is not continuous, however, these statistics are no longer distribution-free. Nevertheless, the tests are then conservative in the sense that the associated significance levels are larger than the exact ones (see, e.g., Noether (1963) and Guilbaud (1986) for a general argument).

Now, as indicated, e.g., in Gleser (1985), rejection regions of H_0 in the one-tailed tests are of the form $D_n^- > d$ for H_1^- and $D_n^+ > d$ for H_1^+, d being a specified positive constant. We are going to point out that for any model, whether continuous or not, the probabilities of these regions correspond to some polynomials G_n and A_n, respectively.

Corollary 5.4 *Let $K(x)$ be any given distribution function for $F(x)$, $x \in \mathbf{R}$. Then, for $d \geq 0$,*

$$P(D_n^- \geq d \mid F = K) = 1 - n!(-1)^n G_n(0 \mid K_1^{(l)}, \ldots, K_n^{(l)}), \tag{5.18}$$

where

$$K_i^{(l)} = K(s_i^{(l)}), \quad \textit{with} \quad s_i^{(l)} = H^{-1}\left(\frac{i-1}{n} + d\right), \quad i = 1, \ldots, n, \tag{5.19}$$

and

$$P(D_n^- > d \mid F = K) = 1 - n!(-1)^n G_n(0 \mid \chi_1^{(l)}, \ldots, \chi_n^{(l)}), \tag{5.20}$$

where

$$\chi_i^{(l)} = K(\sigma_i^{(l)}), \quad \textit{with} \quad \sigma_i^{(l)} = H^{-1}\left(\frac{i-1}{n} + d+\right), \quad i = 1, \ldots, n, \tag{5.21}$$

whilst

$$P(D_n^+ > d \mid F = K) = 1 - n! A_n(1 \mid K_1^{(u)-}, \ldots, K_n^{(u)-}), \tag{5.22}$$

where

$$K_i^{(u)-} = K(s_i^{(u)}-), \quad \textit{with} \quad s_i^{(u)} = H^{-1}\left(\frac{i}{n} - d\right), \quad i = 1, \ldots, n. \tag{5.23}$$

Proof. By definition,

$$P(D_n^- \geq d \mid F = K) = 1 - P[F_n(x) > -d + H(x) \text{ for all } x \mid F = K].$$

¿From (5.1), (5.3), (2.11), and since $(-d + H)^{-1}(.) = H^{-1}(. + d)$, we then deduce (5.18) and (5.19). Analogously, we obtain (5.20) and (5.21) using (5.6) and (5.7). Furthermore,

$$P(D_n^+ > d \mid F = K) = 1 - P[F_n(x) \leq d + H(x) \text{ for all } x \mid F = K].$$

¿From (5.8), (5.11) and (2.14), we then get (5.22) and (5.23).□

By choosing for K the hypothesized H, (5.20)-(5.23) provide the exact levels of significance of the two one-sided Kolmogorov-Smirnov tests. For K different from H, they give the exact powers of the tests. We easily check that (5.20)-(5.23) correspond to the result obtained by Gleser (1985, Formulas (2.3), (2.4)). Notice that the polynomial expressions derived here are especially clear and convenient.

We mention that for the two-sided test, a usual procedure is to approximate $P(D_n > d \mid F = K)$ by $P(D_n^- > d \mid F = K) + P(D_n^+ > d \mid F = K)$. The error made by that way is on the conservative side and is generally very small (see, e.g., Conover (1980)).

Various goodness-of-fit tests of this type can be discussed in the same way. So, a weighted statistic proposed by Anderson and Darling (1952) is, for testing H_1^- say,

$$AD_n^- = \sup_x \left\{[H(x) - F_n(x)]/[H(x)(1 - H(x))]^{1/2}\right\}. \tag{5.24}$$

We then see that $P(AD_n^- \geq d \mid F = K)$, for instance, is still given by (5.18) where $K_i^{(l)} = K(s_i^{(l)})$ with

$$s_i^{(l)} = \left\{-d[H(1-H)]^{1/2} + H\right\}^{-1}\left(\frac{i-1}{n}\right), \quad i = 1, \ldots, n. \tag{5.25}$$

Another simple illustration happens with a type II censored sample, that is when observations are made until the r-th order statistic $X_{(r)}$, r fixed. A version of the Kolmogorov-Smirnov statistic could be, for H_1^-,

$$C_n^- = \sup_{x \leq X_{(r)}} \{H(x) - F_n(x)\}. \tag{5.26}$$

Then, $P(C_n^- \geq d \mid F = K)$, for instance, is given by (5.18) where $K_i^{(l)}, 1 \leq i \leq r$, is defined by (5.19) and

$$K_{r+1}^{(l)} = \ldots = K_n^{(l)} = 1. \tag{5.27}$$

Guilbaud (1986) pointed out that when H is continuous, such a test is again distribution-free but is no longer necessarily conservative. This observation reinforces the interest of the above exact calculations (using (4.3)).

5.3. A few numerical examples. We will only illustrate the formulas (5.18) and (5.19) for the statistic D_n^-. Firstly, let us consider the insulin dependent diabetes data given in Table 1 and studied by Horn (1977):

Table 1.

The alternative of interest in this problem is H_1^-. Obviously, $D_{30}^- = 25/30 - 19/30 = 0.2$. Let us determine the p-value $P(D_{30}^- \geq 0.2 \mid F = H)$. By (5.18), (5.19), we have

$$P(D_{30}^- \geq 0.2 \mid F = H) = 1 - 30!(-1)^{30} G_{30}(0 \mid H_1^{(l)}, \ldots, H_{30}^{(l)}), \tag{5.28}$$

where $H_1^{(l)} = \ldots = H_{13}^{(l)} = H(2) = 18/30$, $H_{14}^{(l)} = \ldots = H_{20}^{(l)} = H(3) = 25/30$, $H_{21}^{(l)} = \ldots = H_{23}^{(l)} = H(4) = 28/30$, $H_{24}^{(l)} = H(5) = 29/30$ and $H_{25}^{(l)} = \ldots = H_{30}^{(l)} = H(6) = 1$. We can evaluate (5.28) using the recursions of Section 4. Indeed, by (4.18) we express (5.28) as

$$P(D_{30}^- \geq 0.2 \mid F = H) = \sum_{j=0}^{12} \binom{30}{j} \left(1 - \frac{18}{30}\right)^{30-j} \left(\frac{18}{30}\right)^{j} + \sum_{j=13}^{23} \binom{30}{j} \left(1 - H_{j+1}^{(l)}\right)^{30-j} f_j^{(l)},$$

where the $f_j^{(l)}$'s satisfy the recursion (4.14) with $H_i^{(l)}$ substituted for F_i. We then find that the p-value is equal to 0.02612364.

Now, let us examine two specific discrete models where the hypothesized distribution is a Poisson law with mean 1 (denoted by $P(1)$) or a binomial distribution with mean 1 and parameter 0.2 (denoted by $Bin(5, 0.2)$). For comparison purposes, consider a continuous model with arbitrary distribution function F^*. The 0.95 quantiles $w_{0.95}^*$ for D_n^- in the continuous case are provided, e.g., by Table A.14 in Conover (1980), for $n = 1, \ldots, 40$. Using Corollary 5.4, we easily compute the probability that D_n^- is as large as $w_{0.95}^*$ for the two discrete models. The results are given in Table 2. As indicated earlier, the continuous model yields a conservative approximation. Observe that the probabilities in the discrete cases are considerably smaller than the exact level of 5% in the continuous case.

Table 2.

The two Figures below give the graph of the p-values for the discrete models, as well as for the continuous model, when $n = 10$. Here too, we notice that the differences can be important.

Figure 1.

Figure 2.

6 Turning to two-sided distributions

To close, we move to the joint rectangle distribution of the order statistics. Niederhausen (1981) has shown, in a general framework, that the key algebraic tool here is a sequence of functions of Sheffer type. His paper provides also many applications and a review. In this Section, we only purpose to derive several basic results by following a quite simple approach.

Let $a_1 \leq \ldots \leq a_n, b_1 \leq \ldots \leq b_n$, and $a_1 \leq b_1, \ldots, a_n \leq b_n \in [0, 1]$. For a sample of size k, $1 \leq k \leq n$, let $X_{(1)}, \ldots, X_{(k)}$ be the order statistics and denote the two-sided probabilities by

$$d_k = P[a_1 < X_{(1)} \leq b_1, \ldots, a_k < X_{(k)} \leq b_k], \quad \text{with} \quad d_0 \equiv 1. \tag{6.1}$$

As before, we can write that

$$d_k = P[F_1^a \leq U_{(1)} \leq F_1^b, \ldots, F_k^a \leq U_{(k)} \leq F_k^b],$$

where $F_i^a \equiv F(a_i)$ and $F_i^b \equiv F(b_i)$, $1 \leq i \leq k$, and $U_{(1)}, \ldots, U_{(k)}$ are the order statistics for a sample of k uniform $(0, 1)$ r.v.'s.

Proposition 6.1 *For $x \in (F_n^a, 1]$,*

$$P[a_1 < X_{(1)} \leq b_1, \ldots, a_n < X_{(n)} \leq b_n \wedge F^{-1}(x)] = \sum_{j=0}^{n} \binom{n}{j} (F_{j+1}^b - x)^{n-j} (-1)^{n-j} d_j. \tag{6.2}$$

Proof. By definition, this probability, p_3 say, is given by

$$\begin{aligned} p_3 &= P[F_1^a \le U_{(1)} \le F_1^b, \ldots, F_n^a \le U_{(n)} \le F_n^b \wedge x] \\ &= n! \int\limits_{\nu_n=F_n^a}^{F_n^b\wedge x} \int\limits_{\nu_{n-1}=F_{n-1}^a}^{F_{n-1}^b\wedge\nu_n} \cdots \int\limits_{\nu_1=F_1^a}^{F_1^b\wedge\nu_2} d\nu_1 \ldots d\nu_{n-1} d\nu_n. \end{aligned} \tag{6.3}$$

The right-hand side member of (6.3) is denoted by $n!S_n(x \mid F_1^a, F_1^b, \ldots, F_n^a, F_n^b)$, or more simply by $n!S_n(x)$. Obviously,

$$d_k = k!S_k(F_k^b), \quad k = 1, \ldots, n. \tag{6.4}$$

Let us vary x from 1 ($\equiv F_{n+1}^b$) to F_n^a ($\equiv F_0^b$). For that, consider $x \in (F_{n-i}^b, F_{n-i+1}^b]$ for $i = 0, \ldots, n$ successively, provided that $F_n^a \le F_{n-i}^b$. We see that on this interval, the first i derivatives of $S_n(x)$ satisfy

$$S_n^{(k)}(x) = \begin{cases} S_{n-k}(x) & \text{for } k = 1, \ldots, i-1, \\ S_{n-i}(F_{n-i}^b) & \text{for } k = i. \end{cases} \tag{6.5}$$

On the other hand, we can check that the polynomials $t_{n,i}(x), x \in \mathbf{R}$, defined by

$$t_{n,i}(x) = \sum_{j=n-i}^{n} \binom{n}{j} \frac{1}{n!} (x - F_{j+1}^b)^{n-j} d_j \tag{6.6}$$

verify (6.5) on $\mathbf{R}$ for any fixed $i = 0, \ldots, n$. Therefore, we get that

$$n!S_n(x) = n!t_{n,i}(x) \ \text{ for } \ x \in (F_{n-i}^b, F_{n-i+1}^b], \tag{6.7}$$

which is equivalent to (6.2).□

We recall that a sequence of functions $\{S_n(x), n \ge 0\}$ defined by (6.6), (6.7) corresponds to a F^b-Sheffer sequence in the sense indicated in the appendix of Niederhausen (1981).

The probability d_n can now be directly evaluated by considering the result (6.2) for sample sizes $k = 1, \ldots, n$. As mentioned in Section 3, the formulas (6.8) and (6.9) below were obtained by Epanechnikov (1968) and Steck (1971).

Corollary 6.2 *The d_k's satisfy the recursion*

$$d_k = -\sum_{j=0}^{k-1} \binom{k}{j} (F_{j+1}^b - F_k^a)_+^{k-j} (-1)^{k-j} d_j, \quad k = 1, \ldots, n. \tag{6.8}$$

Under determinantal form,

$$d_n = \det \left[\binom{i}{j-1} (F_j^b - F_i^a)_+^{i-j+1} \right]_{i,j=1,\ldots,n}. \tag{6.9}$$

Proof. Putting $x = F_n^a$ in (6.2) gives

$$0 = \sum_{j=0}^{n} \binom{n}{j} (F_{j+1}^b - F_n^a)^{n-j} (-1)^{n-j} d_j,$$

which leads to (6.8). Then, by Cramer's rule, and after inverting the order of the lines and rows of the determinant, we deduce (6.9).□

It is worth noticing that for $x \in (F_{n-i}^b, F_{n-i+1}^b], i = 0, \ldots, n$, with $F_n^a \leq F_{n-i}^b$ (as in the proof), the probability of interest p_3 reduces to

$$\begin{aligned} &P[a_1 < X_{(1)} \leq b_1, \ldots, a_{n-i} < X_{(n-i)} \leq b_{n-i}, \\ &\quad a_{n-i+1} < X_{(n-i+1)}, \ldots, a_{n-1} < X_{(n-1)}, a_n < X_{(n)} \leq F^{-1}(x)]. \end{aligned} \tag{6.10}$$

We examine below the special situation where $F_n^a \leq F_1^b$. This case is rather frequent when computing significance points (see Niederhausen (1981, p. 926)).

Corollary 6.3 *If $F_n^a \leq F_1^b$, the d_k's satisfy the recursion, for any $x \in \mathbf{R}$,*

$$d_k = k! A_k(x \mid F_1^a, \ldots, F_k^a) - \sum_{j=0}^{k-1} \binom{k}{j} (x - F_{j+1}^b)^{k-j} d_j, \quad k = 1, \ldots, n. \tag{6.11}$$

Proof. For $x \in (F_n^a \leq F_1^b]$, we get from (6.2) and (6.10) that

$$P[a_1 < X_{(1)}, \ldots, a_{n-1} < X_{(n-1)}, a_n < X_{(n)} \leq F^{-1}(x)] = \sum_{j=0}^{n} \binom{n}{j} (x - F_{j+1}^b)^{n-j} d_j. \tag{6.12}$$

But we know by Corollary 2.4 that this probability is also equal to $n! A_n(x \mid F_1^a, \ldots, F_n^a)$. Combining these two representations then leads to (6.11) for $x \in (F_n^a, F_1^b]$, and therefore for any $x \in \mathbf{R}$.□

In particular, if $a_1 = \ldots = a_n = -\infty$, (6.11) becomes the recursion (3.4) for the l_k's, while if $b_1 = \ldots = b_n = \infty$, (6.11) gives the recursion (3.6) for the r_k's.

7 Appendix: Abel-Gontcharoff and Appell polynomials

Both families of polynomials correspond to natural generalizations of the central family of monomials $\{x^n, n \in \mathbf{N}\}$. Specifically, let us consider the integral representation

$$\frac{x^n}{n!} = \int_{\nu_n=0}^{x} \int_{\nu_{n-1}=0}^{\nu_n} \ldots \int_{\nu_1=0}^{\nu_2} d\nu_1 \ldots d\nu_{n-1} d\nu_n, \quad n \geq 1, \tag{7.1}$$

and let $U = \{u_1, u_2, \ldots\}$ be any given family of real numbers (often given, in practice, in increasing order).

Integral definitions. To U is attached a family of Abel-Gontcharoff polynomials $G_n(x \mid U)$ of degrees $n \in \mathbf{N}$, defined by $G_0(x \mid U) = 1$ and

$$G_n(x \mid U) = \int\limits_{\nu_1=u_1}^{x} \int\limits_{\nu_2=u_2}^{\nu_1} \dots \int\limits_{\nu_n=u_n}^{\nu_{n-1}} d\nu_n \dots d\nu_2 d\nu_1, \quad n \geq 1, \tag{7.2}$$

and a family of Appell polynomials $A_n(x \mid U)$ of degrees $n \in \mathbf{N}$, defined by $A_0(x \mid U) = 1$ and

$$A_n(x \mid U) = \int\limits_{\nu_n=u_n}^{x} \int\limits_{\nu_{n-1}=u_{n-1}}^{\nu_n} \dots \int\limits_{\nu_1=u_1}^{\nu_2} d\nu_1 \dots d\nu_{n-1} d\nu_n, \quad n \geq 1. \tag{7.3}$$

Appell polynomials are classical in the mathematical literature and cover many remarquable subfamilies, e.g. Hermite, Laguerre, Bernoulli and Euler polynomials. The reader is referred, e.g., to Appell (1880), Boas and Buck (1958) and Kaz'min (1988).

Although of similar construction, the polynomials G are much less popular. They were introduced by Gontcharoff (1930, 1937) for the interpolation of entire functions, and extend the classical Abel polynomials obtained when the u_i's are defined by an affine transformation (hence our appellation of Abel-Gontcharoff polynomials).

Abel polynomials. If $u_i = a + b(i-1)$, $i \geq 1$, with $a, b \in \mathbf{R}$, then

$$G_n(x \mid U) = (x - u_1)(x - u_{n+1})^{n-1}/n!, \quad n \geq 0. \tag{7.4}$$

In particular, when $b = 0$, $G_n(x \mid U) = (x - u_1)^n/n!$, $n \geq 0$.

To the best of our knowledge, the first use of the polynomials G for stochastic problems is due to Daniels (1967) in a context of epidemic theory. Since then, they have been widely exploited for the study of various epidemic models (see, e.g., Picard (1980), Lefèvre and Picard (1990), Ball and O'Neill (1999)), as well as in risk theory (Picard and Lefèvre (1994)).

Let us fix any degree n. By construction, G_n and A_n depend on U only through the first n terms $u_1, \dots, u_n$. Thus, they may be denoted as $G_n(x \mid u_1, \dots, u_n)$ and $A_n(x \mid u_1, \dots, u_n)$. Notice also that for $a, b \in \mathbf{R}$,

$$G_n(ax + b \mid aU + b) = a^n G_n(x \mid U), \tag{7.5}$$

where $aU + b = \{au_i + b, i \geq 1\}$. Furthermore, we observe that G_n and A_n are closely related by the identity

$$A_n(x \mid u_1, \dots, u_n) = G_n(x \mid u_n, \dots, u_1). \tag{7.6}$$

The equivalence (7.6) between G_n and A_n is valid for fixed n only. The two families of polynomials, however, are really distinct. This is clearly seen by examining their derivatives. Indeed, differentiating (7.2) r times, $1 \leq r \leq n$, gives

$$G_n^{(r)}(x \mid U) = G_{n-r}(x \mid E^r U), \tag{7.7}$$

where $E^rU = \{u_{r+i}, i \geq 1\}$ is the family U deprived of its first r terms, while for (7.3),

$$A_n^{(r)}(x \mid U) = A_{n-r}(x \mid U). \tag{7.8}$$

In other words, the differentiation shifts U to the right for G, but has no influence on U for A (which justifies that U is generally omitted in the notation of A). This difference is crucial for the properties of the polynomials.

A fundamental result for the family $\{G_n, n \in \mathbf{N}\}$ is that it provides a basis for Abelian-type expansions of arbitrary polynomials.

Abelian-type expansion. Any polynomial $R(x)$ of degree n can be expanded as

$$R(x) = \sum_{j=0}^{n} R^{(j)}(u_{1+j})G_j(x \mid U). \tag{7.9}$$

For the proof, it suffices to write $R(x) = \sum_{j=0}^{n} \alpha_j G_j(x \mid U)$ for some coefficients α_j, and to identify the α_j's by using the property, immediate from (7.2) and (7.7), that

$$G_n^{(r)}(u_{1+r} \mid U) = \delta_{n,r}, \quad r = 0, \ldots, n. \tag{7.10}$$

Formula (7.9) allows us to generate the G_n's recursively. So, taking $R(x) = x^n/n!$, we get

$$G_n(x \mid U) = \frac{x^n}{n!} - \sum_{j=0}^{n-1} \frac{(u_{1+j})^{n-j}}{(n-j)!} G_j(x \mid U), \quad n \geq 1. \tag{7.11}$$

A disadvantage of (7.11) is that if x is changed, then the n polynomials $G_j(x \mid U), 1 \leq j \leq n$, have to be recomputed. To avoid this, an alternative method relies on the Taylor expansion of $G_n(x \mid U)$ at point 0. For the evaluation of the coefficients $G_n^{(j)}(0 \mid U), 0 \leq j \leq n$, we have recourse to (7.10), which yields the following triangular system of linear equations:

$$\sum_{j=r}^{n} G_n^{(j)}(0 \mid U) \frac{(u_{1+r})^{j-r}}{(j-r)!} = \delta_{n,r}, \quad r = n, \ldots, 0. \tag{7.12}$$

For instance, by Cramer's rule and after some algebraic manipulations, we find from (7.12) that $G_n(0 \mid U)$ is given under determinantal form by

$$\begin{aligned} G_n(0 \mid U) &= (-1)^n \det \left[(u_{n-i+1})^{i-j+1}/(i-j+1)! \right]_{i,j=1,\ldots,n}, \\ &= \frac{(-1)^n}{n!} \det \left[\binom{i}{j-1} (u_j)^{i-j+1} \right]_{i,j=1,\ldots,n}. \end{aligned} \tag{7.13}$$

Let us turn to the family $\{A_n, n \in \mathbf{N}\}$. The Taylor expansion of $A_n(x+y \mid U)$ at point y with (7.8) leads directly to the following important formula.

Binomial theorem. For $x, y \in \mathbf{R}$,

$$A_n(x+y \mid U) = \sum_{j=0}^{n} A_{n-j}(y \mid U) x^j / j! \tag{7.14}$$

For $y = 0$, (7.14) allows us to evaluate $A_n(x \mid U)$ from the coefficients $A_{n-j}(0 \mid U), 0 \leq j \leq n$. These are computed using the property, immediate from (7.3) and (7.8), that

$$A_n^{(r)}(u_{n-r} \mid U) = \delta_{n,r}, \quad r = 0, \ldots, n. \tag{7.15}$$

Note that the generating function of the $A_n(x \mid U)$'s has the simple factorization

$$\sum_{n=0}^{\infty} A_n(x \mid U) s^n = g(s, U) e^{xs}, \tag{7.16}$$

$g(s, U)$ being the generating function of the $A_n(0 \mid U)$'s.

Remark. It may be worth mentioning that recently, Picard and Lefèvre (1996) have generalized the polynomials G and A into functions $\bar{G}$ and $\bar{A}$, called pseudopolynomials, with similar structures. Firstly, the differential operator D is replaced by a more general operator Δ. This step is not really original and has often been made since Sheffer (1939) to extend Appell polynomials (see, e.g., Mullin and Rota (1970) and Niederhausen (1981)). The second step, more ambitious, consists in working with the vector space generated by an arbitrary family of linearly independent functions $\{e_n, n \in \mathbf{N}\}$, instead of $\{x^n/n!, n \in \mathbf{N}\}$. These new functions $\bar{G}$ and $\bar{A}$ constitute essential tools of study in epidemic and risk theories (see, e.g., Picard and Lefèvre (1997) and Lefèvre and Picard (1999)).

Acknowledgements

The first author author would like to thank the support of the Belgian Government under "Projet d'Actions de Recherche Concertées" (No. 98/03-217)

References

Anderson, T.W., Darling, D.A., 1952. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. Ann. Math. Statist. 23, 193-212.

Appell, P.E., 1880. Sur une classe de polynômes. Ann. Sci. Ecole Normale Sup. 9, 119-144.

Arnold, B.C., Balakrishnan, N., Nagaraja, H.N., 1992. A First Course in Order Statistics. Wiley, New York.

Ball, F.G., O'Neill, P.D., 1999. The distribution of general final state random variables for stochastic epidemic models. J. Appl. Probab. 36, 473-491.

Berg, S., Mutafchiev, L., 1990. Random mapping with an attracting center: Lagrangian distributions and regression function. J. Appl. Probab. 27, 622-636.

Birnbaum, Z.W., Pyke, R., 1958. On some distributions related to the statistic D_n^+. Ann. Math. Statist. 29, 179-187.

Boas, R.P.Jr., Buck, R.C., 1958. Polynomial Expansions of Analytic Functions. Springer, Heidelberg.

Chang, L., 1955. On the ratio of the empirical distribution to the theoretical distribution function. Acta Math. Sinica 5, 347-368 (English translation in IMS-AMS Sel. Transl. Math. Statist. Prob. 4 (1963), pp. 17-38).

Charalambides, C.A., 1990. Abel series distributions with applications to fluctuations of sample functions of stochastic processes. Comm. Statist.-Theory Meth. 19, 317-335.

Conover, W.J., 1972. A Kolmogorov goodness-of-fit test for discontinuous distributions. J. Amer. Statist. Ass. 67, 591-596.

Conover, W.J., 1980. Practical Nonparametric Statistics (2nd edition). Wiley, New York.

Consul, P.C., 1974. A simple urn model dependent on pre-determined strategy. Sankhya 36, 391-399.

Csäki, E., 1981. Investigations concerning the empirical distribution function. IMS-AMS Sel. Transl. Math. Statist. Prob. 15, 229-317.

Daniels, H.E., 1945. The statistical theory of the strength of bundles of threads. Proc. Roy. Soc. London Ser. A 183, 405-435.

Daniels, H.E., 1967. The distribution of the total size of an epidemic. Proc. Fifth Berkeley Symp. Math. Statist. Prob. 4, 281-293.

David, H.A., 1981. Order Statistics (2nd edition). Wiley, New York.

Dempster, A.P., 1959. Generalized D_n^+ statistics. Ann. Math. Statist. 30, 593-597.

Eicker, F., 1970. On the probability that a sample distribution function lies below a line segment. Ann. Math. Statist. 41, 2075-2092.

Epanechnikov, V.A., 1968. The significance level and power of the two-sided Kolmogorov test in the case of small sample sizes. Theory Probab. Appl. 13, 686-690.

Galambos, J., 1987. The Asymptotic Theory of Extreme Order Statistics (2nd edition). Krieger, Malabar, Florida.

Gleser, L.J., 1985. Exact power of goodness-of-fit tests of Kolmogorov type for discontinuous distributions. J. Amer. Statist. Ass. 80, 954-958.

Gontcharoff, W., 1930. Recherches sur les dérivées successives des fonctions analytiques. Généralisation de la série d'Abel. Ann. Sci. Ecole Normale Sup. 47, 1-78.

Gontcharoff, W., 1937. Détermination des Fonctions Entières par Interpolation. Hermann, Paris.

Guilbaud, O., 1986. Stochastic inequalities for Kolmogorov and similar statistics, with confidence region applications. J. Scand. Statist. 13, 301-305.

Horn, S.D., 1977. Goodness-of-fit tests for discrete data: a review and an application to a health impairment scale. Biometrics 33, 237-248.

Kaz'min, Y.A., 1988. Appel polynomials. In: Hazewinkel, M. (Ed.), Encyclopedia of Mathematics. Kluwer, Dordrecht, vol. 1, pp. 209-210.

Kotel'nikova, V.F., Chmaladze, E.V., 1983. On computing the probability of an empirical process not crossing a curvilinear boundary. Theory Probab. Appl. 27, 640-648.

Lefèvre, Cl., Picard, Ph., 1990. A non-standard family of polynomials and the final size distribution of Reed-Frost epidemic processes. Adv. Appl. Probab. 22, 25-48.

Lefèvre, Cl., Picard, Ph., 1999. Abel-Gontcharoff pseudopolynomials and the exact final outcome of SIR epidemic models (III). Adv. Appl. Probab. 31, 532-549.

Mullin, R., Rota, G.-C., 1970. On the foundation of combinatorial theory - III: Theory of binomial enumeration. In: Harris (Ed.), Graph Theory and its Applications. Academic Press, New York, pp. 167-213.

Niederhausen, H., 1981. Sheffer polynomials for computing exact Kolmogorov-Smirnov and Rényi type distributions. Ann. Statist. 9, 923-944.

Noé, M., Vandewiele, G., 1968. The calculation of distributions of Kolmogorov-Smirnov type statistics including a table of significance points for a particular case. Ann. Math. Statist. 39, 233-241.

Noé, M., 1972. The calculation of distributions of two-sided Kolmogorov-Smirnov type statistics. Ann. Math. Statist. 43, 58-64.

Noether, G.E., 1963. A note on the Kolmogorov-Smirnov statistics in the discrete case. Metrika 7, 115-116.

Pettitt, A.N., Stephens, M.A., 1977. The Kolmogorov-Smirnov goodness-of-fit statistic with discrete and group data. Technometrics 19, 205-210.

Picard, Ph., 1980. Applications of martingale theory to some epidemic models. J. Appl. Probab. 17, 583-599.

Picard, Ph., Lefèvre, Cl., 1994. On the first crossing of the surplus process with a given upper barrier. Insurance: Math. Econ. 14, 163-179.

Picard, Ph., Lefèvre, Cl., 1989. Sur des familles de polynômes non-standards intervenant dans des problèmes de premier passage. Contributed paper at the 47th ISI (International Statistical Institute) Session, Paris.

Picard, Ph., Lefèvre, Cl., 1996. Abel expansions and generalized Abel polynomials in stochastic models. In: Heyde, C.C., Prohorov, Yu V., Pyke, R., Rachev, S.T. (Eds.), Proceedings of the Athens Conference on Applied Probability and Time Series Analysis in Honor of J.M. Gani and in Memory of E.J. Hannan. Springer, New York, Lecture Notes in Statistics 114, pp. 55-69.

Picard, Ph., Lefèvre, Cl., 1996. First-crossing of basic counting processes with lower non-linear boundaries: A unified approach through pseudopolynomials (I). Adv. Appl. Probab. 28, 853-876.

Picard, Ph., Lefèvre, Cl., 1997. The probability of ruin in finite time with discrete claim size distribution. Scand. Actuarial J. 1, 58-69.

Pitman, E.J.G., 1972. Simple proofs of Steck's determinantal expressions for probabilities in the Kolmogorov-Smirnov tests. Bull. Austral. Math. Soc. 7, 227-232.

Reiss, R.-D., 1989. Approximate Distributions of Order Statistics with Applications to Nonparametric Statistics. Springer, New York.

Sheffer, I.M., 1939. Some properties of polynomials of type zero. Duke Math. J. 5, 590-622.

Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics. Wiley, New York.

Stadje, W., 1993. Distribution of first-exit times for empirical counting and Poisson processes with mouving boundaries. Comm. Statist.-Theory Meth. 9, 91-103.

Steck, G.P., 1969. The Smirnov two-sample tests as rank tests. Ann. Math. Statist. 40, 1449-1466.

Steck, G.P., 1971. Rectangle probabilities for uniform order statistics and the probability that the empirical distribution function lies between two distribution functions. Ann. Math. Statist. 42, 1-11.

Takács, L., 1967. Combinatorial Methods in the Theory of Stochastic Processes. Wiley, New York.

Wald, A., Wolfowitz, J., 1939. Confidence limits for continuous distribution functions. Ann. Math. Statist. 10, 105-118.

Wood, C.L., Altavela, M.M., 1978. Large-sample results for Kolmogorov-Smirnov statistics for discrete distributions. Biometrika 65, 235-239.

Whittle, P., 1961. Some exact results for one-sided distribution tests of the Kolmogorov-Smirnov type. Ann. Math. Statist. 32, 499-505.

Health impairment level x	Hypothesized d.f. $H(x)$	Empirical d.f. $F_{30}(x)$
1	1/30	0
2	18/30	15/30
3	25/30	19/30
4	28/30	26/30
5	29/30	28/30
6	1	1

Table 1: Insulin diabetes data (Horn (1977)).

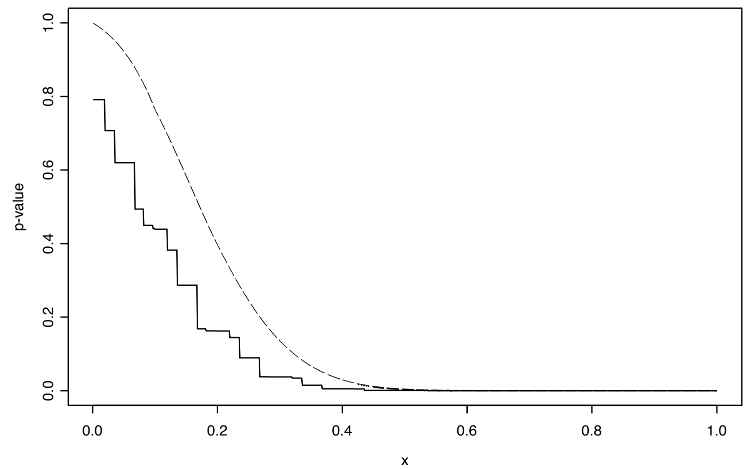

Figure 1: Graph of the functions $P[D_{10}^{-} \geq d \mid F = P(1)]$ (———) and $P[D_{10}^{-} \geq d \mid F = F^{*}]$ (- - - - - -).

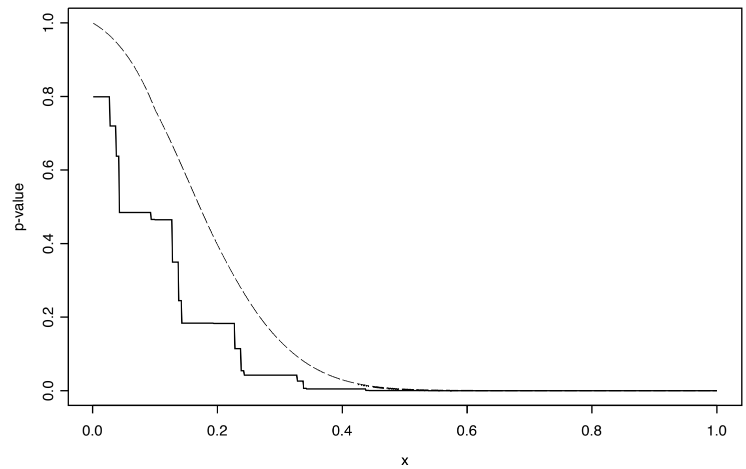

Figure 2: Graph of the functions $P[D_{10}^{-} \geq d \mid F = Bin(5, 0.2)]$ (———) and $P[D_{10}^{-} \geq d \mid F = F^{*}]$ (- - - - - -).

Sample size n	Quantiles $w^*_{0.95}$	$P[D_n^- \geq w^*_{0.95} \mid F = P(1)]$	$P[D_n^- \geq w^*_{0.95} \mid F = Bin(5, 0.2)]$
1	0.950	1.3%	0.7%
2	0.776	0.6%	0.3%
3	0.636	1.9%	1.8%
4	0.565	0.6%	0.5%
5	0.509	2.2%	2.0%
6	0.468	0.7%	0.6%
7	0.436	1.7%	1.6%
8	0.410	0.8%	0.6%
9	0.387	1.3%	1.3%
10	0.369	0.5%	0.5%
11	0.352	1.7%	1.0%
12	0.338	0.8%	0.4%
13	0.325	1.1%	1.3%
14	0.314	0.5%	0.7%
15	0.304	0.8%	0.8%
16	0.295	1.7%	1.2%
17	0.286	0.9%	0.6%
18	0.279	1.1%	0.9%
19	0.271	0.6%	0.9%
20	0.265	1.4%	1.0%
21	0.259	1.5%	1.2%
22	0.253	0.8%	0.6%
23	0.247	1.0%	0.8%
24	0.242	1.2%	0.9%
25	0.238	1.2%	0.9%
26	0.233	1.2%	1.1%
27	0.229	0.7%	0.6%
28	0.225	1.4%	0.7%
29	0.221	0.9%	0.8%
30	0.218	0.9%	0.8%
31	0.214	1.0%	1.0%
32	0.211	1.1%	0.6%
33	0.208	1.1%	0.7%
34	0.205	1.2%	1.2%
35	0.202	0.7%	0.7%
36	0.199	1.3%	0.8%
37	0.196	0.8%	1.0%
38	0.194	0.8%	0.9%
39	0.191	0.9%	1.0%
40	0.189	1.0%	0.6%

Table 2: 0.95 quantiles $w^*_{0.95}$ for an arbitrary continuous model, and probabilities of being as large as $w^*_{0.95}$ for two specific discrete models.