EXTREME EXPECTATIONS OF BERNOULLI CONVOLUTIONS
GIVEN THEIR FIRST FEW MOMENTS ARE ATTAINED
AT SHIFTED CONVOLUTIONS OF AS FEW BINOMIALS

LUTZ MAOTNER

Abstract. A result of Chebyshev (1846) and Hoeffding (1956), on bounding an expectation of a given function with respect to a Bernoulli convolution (also called Poisson binomial law, or law of the number of successes in independent trials) with any given first moment, is here generalised to the case of any given first few moments, as indicated in the title.

A nonprobabilistic, and perhaps more obvious, reformulation is: Every permutation invariant and separately affine-linear function of \( n \) real variables \( x_i \in [a, b] \) assumes its extremal values given the power sums \( \sum_{i=1}^n x_1^i, \ldots, \sum_{i=1}^n x_r^i \) at vectors \( x \) with at most \( r \) coordinate values different from \( a \) and \( b \).

1. Introduction and result

\textit{Bernoulli convolutions}, by which we mean the laws

\begin{equation}
\text{BC}_p := \bigstar_{i \in I} \text{B}_{p_i} \quad \text{for } p \in [0, 1]^I \text{ with } I \text{ finite}
\end{equation}

and with each \( \text{B}_{p_i} \) being Bernoulli (details about our terminology and notation are mostly given below between (4) and Theorem 1.1), occur naturally for two quite different reasons:

First, \( \text{BC}_p \) as in (1) is the law of the number of successes in \( \#I \) independent trials with possibly varying success probabilities \( p_i \). As such it has been studied by Laplace (1812, §38, pp. 421–423, the case of \( \nu(i) = 1; \ 1886, \text{ pp. 430–432} \)), who derived a normal approximation based on the mean and the variance

\begin{equation}
\mu(\text{BC}_p) = \sum_{i \in I} p_i, \quad \sigma^2(\text{BC}_p) = \sum_{i \in I} p_i(1 - p_i),
\end{equation}

and again by Poisson (1837) as described by Hald (1998, in particular pp. 570, 579). Because of Poisson’s work, \( \text{BC}_p \) is often called a \textit{Poisson binomial law}, to distinguish it from the special case of an ordinary binomial law

\begin{equation}
\text{B}_{n,p} := \text{B}^\ast_n := \bigstar_{i=1}^n \text{B}_{p_i} = \text{BC}(p, \ldots, p) \quad \text{for } n \in \mathbb{N}_0 \text{ and } p \in [0, 1].
\end{equation}

Second, many combinatorial probability distributions unexpectedly turn out to be Bernoulli convolutions, with in general the \( p_i \) in (1) neither explicitly known nor rational, but nevertheless the mean and the variance, and perhaps also higher moments or cumulants, easy to obtain. For example, each hypergeometric law \( \text{H}_{n,r,b} \) of the number of red balls obtained by drawing \( n \) balls successively without replacement from an urn containing \( r \) red and \( b \) blue balls, for which Mattner and Schulz (2018, pp. 729–733) recall some basic facts and references, is a Bernoulli convolution, as proved by Vatutin and Mikhailov (1983, Corollary 5 with \( n = 2 \)). Here it is easy to check that in particular

\[ \text{H}_{2,r,b} = \text{B}_{\frac{r+1}{r+\varepsilon}} \ast \text{B}_{\frac{b}{r+\varepsilon}} \quad \text{with} \quad \varepsilon := \frac{1}{r+b} \sqrt{\frac{rb}{r+b-1}} \quad \text{for } r, b \in \mathbb{N}, \]
with the Bernoulli success parameters irrational, and then without any combinatorial meaning, for example if \( r = b = 2 \). Pitman (1997, p. 279) gives references to further combinatorial Bernoulli convolutions, there named \( \text{Pólya frequency distributions} \). Warren (1999) treats recursion techniques for proving Poisson-binomiality, applicable in particular to give another proof for the hypergeometric case mentioned above. Tang and Tang (2019) provide a more recent review of Bernoulli convolutions.

Hence, in either case, bounds for probabilities \( BC_p(A) \) or more general expectations \( BC_p g \), with a given set \( A \) or function \( g \) and given any first few moments or cumulants of \( BC_p \), are of interest. Theorem 1.1, stated below after explaining some terminology and notation and recalling some standard facts, reduces the search for such bounds to certain special Bernoulli convolutions, namely shifted convolutions of few binomial laws as in (8).

We put \( N := \{1, 2, \ldots \} \), \( N_0 := \mathbb{N} \cup \{0\} \), and

\[
\mathbb{N} := \{1, \ldots, n\} \quad \text{and} \quad \mathbb{N}_0 := \mathbb{N} \cup \{0\} = \{0, \ldots, n\} \quad \text{for} \ n \in \mathbb{N}_0,
\]

so in particular \( \emptyset = \emptyset \). We write \# \( A \) for the number of elements of a set \( A \), and \( |a, b[ \) for an open interval.

We write \( \text{Prob}(\mathbb{R}) \) for the set of all probability measures (or \textit{distributions}, or \textit{laws}) on the Borel sets of \( \mathbb{R} \), and \( * \) for convolution on \( \text{Prob}(\mathbb{R}) \). The special laws occuring in this paper are Dirac laws \( \delta_x \) for \( x \in \mathbb{R} \), Bernoulli laws \( B_p := (1 - p)\delta_0 + p\delta_1 \) for \( p \in [0, 1] \), and the convolutions \( BC_p \) and \( B_{n,p} \) defined by (1) and (3), where of course \( \bigstar_{i \in \mathbb{N}} P_i := \delta_0 \).

For \( r \in \mathbb{N}_0 \) we put \( \text{Prob}_r(\mathbb{R}) := \{ P \in \text{Prob}(\mathbb{R}) : \int |x|^r \, dP(x) < \infty \} \), and \( \mu_r(P) := \int |x|^r \, dP(x) \) for \( P \in \text{Prob}_r(\mathbb{R}) \); then \( \mu := \mu_1 \) on \( \text{Prob}_1(\mathbb{R}) \), \( \sigma^2 := \mu_2 - \mu_1^2 \) on \( \text{Prob}_2(\mathbb{R}) \). For \( r \in \mathbb{N} \), we let \( \kappa_r \) denote the \( r \)-th cumulant functional on \( \text{Prob}_r(\mathbb{R}) \), that is,

\[
\kappa_r(P) := \left. i^{-r} \left( \frac{d}{dt} \right)^r \log \left( \int e^{itx} \, dP(x) \right) \right|_{t=0} \quad \text{for} \ P \in \text{Prob}_r(\mathbb{R}).
\]

We recall Thiele’s (1889, 1899) recursion

\[
\mu_{r+1} = \sum_{\ell=0}^{r} \binom{r}{\ell} \mu_{r-\ell} \kappa_{\ell+1} \quad \text{on} \ \text{Prob}_{r+1}(\mathbb{R}), \ \text{for each} \ \ell \in \mathbb{N}_0
\]

given in Hald (2000, pp. 144, 150) with essentially its modern proof as, for example, in Mattner (1999); this shows that the first \( r \) moments \( \mu_1, \ldots, \mu_r \) are (polynomial) functions of the first \( r \) cumulants \( \kappa_1, \ldots, \kappa_r \) and conversely; in particular \( \kappa_1 = \mu, \kappa_2 = \sigma^2 \). Hence it is purely a matter of taste whether one fixes some first few moments or the same number of first cumulants of a law on \( \mathbb{R} \), and we have chosen the latter in Theorem 1.1 below, and the perhaps better known former in the title of this paper. The main argument in favour of cumulants is of course their additivity with respect to convolution,

\[
\kappa_r(P * Q) = \kappa_r(P) + \kappa_r(Q) \quad \text{for} \ r \in \mathbb{N} \ \text{and} \ P, Q \in \text{Prob}_r(\mathbb{R}),
\]

first observed by Thiele (1889) according to Hald (1998, pp. 345, 347), and later shown to be characteristic in a quite strong sense by Mattner (1999, 2004).

We use the notation \( Pg := \int g \, dP \) for \( P \)-integrable functions \( g \).

\[ \textbf{Theorem 1.1.} \text{ Let } I \text{ be a set with } n := \#I \in \mathbb{N}_0, \ g : \mathbb{N}_0 \to \mathbb{R} \text{ a function}, \ r \in \mathbb{N}_0, \ \text{and} \ c \in \mathbb{R}^r \text{ with} \]

\[
D := \left\{ p \in [0, 1]^r : \kappa_\ell(BC_p) = c_\ell \text{ for } \ell \in \mathbb{L} \right\} \neq \emptyset.
\]
Then \( \max_{p \in D} BC_p g \) is attained at some \( p \in D \) with \( r' := \#(\{p_i : i \in I\} \setminus \{0, 1\}) \leq r \). For such a \( p \), we have

\[
BC_p = \delta_{n_0} \ast \bigoplus_{j=1}^{r'} B_{n_j, q_j}
\]

with \( \{q_j : j \in \mathbb{N}_0^r\} := \{p_i : i \in I\} \setminus \{0, 1\}, \quad q_0 := 1, \text{ and } n_j := \#(\{i \in I : p_i = q_j\}) \) for \( j \in \mathbb{N}_0^r \).

The corresponding minimisation result is of course obtained by applying the above to \(-g\).

For \( r = 0 \), Theorem 1.1 is a rather obvious and not very interesting consequence of the separate affine-linearity of \( BC_p g \) as a function of \( p \in D = [0, 1]^I \). For \( r = 1 \), Theorem 1.1 is Hoeffding (1956, p. 717, Corollary 2.1), proved explicitly before by Tchebichef (1846, p. 260, second Théorème) for \( g \) an indicator of a set of the form \( \{x \in \mathbb{N}_0 : x \geq m\} \), but with a method applicable to general \( g \) as well; examples of its application, with the determination of an appropriate pair \((n_0, n_1)\) not quite trivial, include Mattner and Roos (2007, p. 202) and Mattner and Tasto (2015, p. 307).

We hope that the present Theorem 1.1 will find comparable applications for \( r \geq 2 \), perhaps in the following nonprobabilistic reformulation. For \( r \in \mathbb{N}_0, I \) a finite set, and \( x \in \mathbb{R}^I \), we denote the \( r \)th power sum of \( x \) by

\[
S_r(x) := \sum_{i \in I} x_i^r.
\]

**Theorem 1.2.** Let \( I \) be a finite set, \( a, b \in \mathbb{R}, f : [a, b]^I \to \mathbb{R} \) a function permutation invariant in its \#\( I \) variables and affine-linear in each of these, \( r \in \mathbb{N}_0 \), and \( c \in \mathbb{R}^r \) with

\[
D := \{x \in [a, b]^I : S_\ell(x) = c_\ell \text{ for } \ell \in \mathbb{R}\} \neq \emptyset.
\]

Then \( \max_{x \in D} f(x) \) is attained at some \( x \in D \) with \( \#(\{x_i : i \in I\} \setminus \{a, b\}) \leq r \).

Theorem 1.2 is closely related to Kovačec, Kuhlmann and Riener (2012, p. 221, Theorem 3, there given without details of a proof). In fact, the present Lemma 2.1 is, after the first two sentences of its proof, a special case of the claim just cited.

The proof of Theorem 1.1 given near the end of the next section rests on Lemmas 2.1–2.5. These are “well–known”, as just indicated in case of Lemma 2.1, but we attempt to provide appropriate references and, if deemed advisable, details of proofs. The idea for Theorem 1.1 is then to apply Lemma 2.1 to any \( r + 1 \) coordinates of a suitable maximiser \( p \).

2. Proofs

**Lemma 2.1.** Let \( I \) be a finite set, \( a, b \in \mathbb{R}, r \in \mathbb{N}_0, c \in \mathbb{R}^r \), and let \( x \) locally minimise or maximise \( S_{r+1} \) in \([a, b]^I\) subject to the constraints \( S_\ell(x) = c_\ell \) for \( \ell \in \mathbb{R} \). Then

\[
\#(\{x_i : i \in I\} \setminus \{a, b\}) \leq r.
\]

**Proof.** Let \( I_0 := \{i \in I : x_i \in ]a, b[\} \). Then \( (x_i : i \in I_0) \) locally minimises or maximises

\[
\sum_{i \in I_0} x_i^{r+1}
\]

in the open set \([a, b]^{I_0}\) subject to the constraints \( \sum_{i \in I_0} x_i^\ell = d_\ell := c_\ell - \sum_{i \in I \setminus I_0} x_i^\ell \) for \( \ell \in \mathbb{R} \). Hence there exists a Lagrange multiplier \( \lambda \in \mathbb{R}^{2r} \), not identically zero, with

\[
\lambda_0(r + 1)x_i^r + \sum_{\ell=1}^{2r} \lambda_\ell x_i^{\ell-1} = 0 \quad \text{for } i \in I_0,
\]

for example by Tichomirov (1982, p. 63). Hence the \( x_i \) with \( i \in I_0 \) are zeros of a common nonzero polynomial of degree at most \( r \), hence \( \#(\{x_i : i \in I_0\}) \leq r \), and hence we have (10).
Lemma 2.2. For \( r \in \mathbb{N} \), there are unique \( a_{r,\ell} \in \mathbb{R} \) with
\[
(11) \quad \kappa_r(B_p) = \sum_{\ell=1}^r a_{r,\ell} p^\ell \quad \text{for} \ p \in [0, 1],
\]
and then \( a_{r,r} = (-1)^{r-1}(r-1)! \). In particular we have \( \kappa_1(B_p) = p, \ \kappa_2(B_p) = p(1-p), \ \kappa_3(B_p) = p(1-p)(1-2p), \ \kappa_4(B_p) = p(1-p)(1-6p(1-p)) \).

First proof. Since obviously \( \mu_r(B_p) = p^{1+r} \) for \( r \in \mathbb{N}_0 \) and \( p \in [0, 1] \), applying \( (5) \) to \( B_p \) yields
\[
\kappa_{r+1}(B_p) = p \left( 1 - \sum_{\ell=0}^{r-1} \binom{r}{\ell} \kappa_{\ell+1}(B_p) \right) \quad \text{for} \ r \in \mathbb{N}_0 \text{ and } p \in [0, 1],
\]
and hence the claim by induction. \( \square \)

Second proof. One easily verifies
\[
\kappa_1(B_p) = p, \quad \kappa_{r+1}(B_p) = p(1-p) \frac{d}{dp} \kappa_r(B_p) \quad \text{for} \ r \in \mathbb{N} \text{ and } p \in [0, 1]
\]
as indicated in Stuart and Ord (1987, p. 203, Exercise 5.1), with reference to Frisch (1926) and Haldane (1940, pp. 392–393, there \( p \) and \( 1-p \) inconsistently interchanged). \( \square \)

We fix the notation \( a_{r,\ell} \), defined by \( (11) \), up to the proof of Lemma 2.4. In generalisation of \( (2) \) we hence get:

Lemma 2.3. For \( r \in \mathbb{N} \) we have
\[
(12) \quad \kappa_r(BC_p) = \sum_{i \in I} \kappa_r(B_{p_i}) = \sum_{\ell=1}^r a_{r,\ell} S_\ell(p) \quad \text{for} \ p \in [0, 1]^I \text{ with } \#I \text{ finite.}
\]

Proof. First \( (6) \), then \( (11) \), then \( (9) \). \( \square \)

For \( r \in \mathbb{N}_0 \), \( I \) a finite set, and \( x \in \mathbb{R}^I \), we put
\[
(13) \quad K_r(x) := \sum_{\ell=1}^r a_{r,\ell} S_\ell(x),
\]
\[
(14) \quad E_r(x) := \sum_{\alpha \in \{0,1\}^I : |\alpha| = r} x^\alpha = \sum_{J \subseteq I : \#J = r} \prod_{i \in J} x_i,
\]
using here multi-index notation, as explained for example by John (1982, pp. 54–56). So the \( K_r \) are, in case of \( r \geq 1 \), the polynomial extensions of the left hand sides of \( (12) \), and the \( E_r \) are the elementary symmetric functions, \( E_0(x) = 1, E_1(x) = \sum_{i \in I} x_i = S_1(x), \ldots, E_{\#I}(x) = \prod_{i \in I} x_i, E_r(x) = 0 \) for \( r > \#I \).

Referring to the use of the symbols \( S, K, E \) in \((9,13,14)\), we have:

Lemma 2.4. Let \( r \in \mathbb{N} \). For \( (A, B) \in \{S, K, E\} \) there exist a constant \( c_{A,B,r} \in \mathbb{R} \setminus \{0\} \) and a polynomial function \( T_{A,B,r} : \mathbb{R}^{r-1} \to \mathbb{R} \) with
\[
A_r(x) = c_{A,B,r} B_r(x) + T_{A,B,r}(B_1(x), \ldots, B_{r-1}(x)) \quad \text{for} \ x \in \mathbb{R}^I \text{ with } I \text{ finite.}
\]
In particular, \( c_{E,S,r} = \frac{(-1)^{r-1}}{r} \).

Proof. For \( (A, B) = (K, S) \), the claim is true with \( c_{K,S,r} = a_{r,r} = (-1)^{r-1}(r-1)! \) and \( T_{K,S,r} \) even linear, by \( (13) \) and Lemma 2.2.

For \( (A, B) = (E, S) \), the claim follows from the Newton identities
\[
E_r = \frac{1}{r} \sum_{\ell=1}^r (-1)^{\ell-1} S_\ell E_{r-\ell} \quad \text{for} \ r \in \mathbb{N}
\]
proved, for example, by Uspensky (1948, p. 261) and by Macdonald (1995, p. 23, (2.11')).

The other cases follow from these two, with \( c_{A,C,r} = c_{A,B,r}c_{B,C,r} \) for \( A, B, C \in \{ S, K, E \} \). □

As noted by Hoeffding (1956, pp. 713-714), we have:

**Lemma 2.5.** Let \( I \) be a finite set with \( n := \# I \in \mathbb{N}_0 \), and let \( f : [0,1]^I \to \mathbb{R} \) be a function. Then the following three statements are equivalent:

(i) There is a function \( g : \mathbb{R}^n \to \mathbb{R} \) with \( f(p) = BC_p g \) for \( p \in [0,1]^I \).

(ii) \( f \) is permutation invariant in its \( n \) variables, and affine-linear in each of these.

(iii) There is a vector \( b \in \mathbb{R}^\mathbb{R} \) with \( f(p) = \sum_{k=0}^n b_k E_k(p) \) for \( p \in [0,1]^I \).

**Proof.** (i) ⇒ (ii): Obvious by permutation invariance and multilinearity of convolution.

(ii) ⇒ (iii): If \( f \) is separately affine-linear as stated, then \( f \) is in particular polynomial (jointly in its variables, not merely separately), as can be seen by induction: Assuming here \( I = \emptyset \) and \( n \geq 1 \), we have \( f(p) = f(p_1, \ldots, p_{n-1}, 0) + (f(p) - f(p_1, \ldots, p_{n-1}, 0)) = g(p_1, \ldots, p_{n-1}) + h(p_1, \ldots, p_{n-1})p_n \) for \( p \in [0,1]^n \) with some functions \( g \) and \( h \), which by considering for example first \( p_n = 0 \) and then \( p_n = 1 \), are seen to be affine-linear in each of their respectively \( n - 1 \) variables, hence polynomial by assumption, yielding polynomiality of \( f \).

Hence then \( f \) is representable by its Taylor expansion about \( 0 \in [0,1]^I \), which under permutation invariance and separate affine-linearity reduces to the form claimed in (iii).

(iii) ⇒ (i): For \( k \in \mathbb{R}_0 \) and \( p \in [0,1]^I \), we get

\[
E_k(p) = \int_{\{0,1\}^I} E_k \, d\otimes_{i \in I} B_{p_i} = \int_{\{0,1\}^I} \left( \begin{array}{c} S_1 \\ k \end{array} \right) \, d\otimes_{i \in I} B_{p_i} = \int_{\mathbb{R}^I} \left( \begin{array}{c} x \\ k \end{array} \right) \, dB_C(x)
\]

by using in the third step that the convolution \( BC_p \) is by definition the image under \( S_1 \) of the product measure \( \otimes_{i \in I} B_{p_i} \), in the second \( E_k(y) = \left( \begin{array}{c} S_1(y) \\ k \end{array} \right) \) for \( y \in \{0,1\}^I \), and in the first \( f \sum = \sum f \) and Fubini. Hence, given (iii), we take \( g(x) := \sum_{k=0}^n b_k \left( \begin{array}{c} x \\ k \end{array} \right) \) for \( x \in \mathbb{R}_0 \). □

Instead of the proof of polynomiality in (ii) ⇒ (iii) above, it would appear to be more elegant to use a general fact like the generalisation by Palais (1978, Theorem in section 5, Theorem 8.3) of a result of Carroll (1961), but we are not aware of a short exposition yielding easily what is needed here.

**Proof of Theorem 1.1.** Under the assumptions of the theorem up to (7), \( \max_{p \in D} BC_p g \) exists by continuity, compactness, and nonemptiness. So we may choose a maximiser of \( S_{r+1} \) among the perhaps several maximisers of \( D \ni p \mapsto BC_p g \) and, changing notation, call this maximiser \( \hat{p} \), rather than \( p \) as in the theorem.

We recall that an \( x \in \mathbb{R}^I \) is a function on \( I \), hence a set \( \{ (i, x_i) : i \in I \} \) of ordered pairs; hence the notation \( \hat{p}|_J := \{ (j, \hat{p}_j) : j \in J \} \) and \( q \cup \hat{s} \) occurring below.

Let us assume that \( J \subseteq I \) with \( \# J = r + 1 \), and let \( \hat{q} := \hat{p}|_J \) and \( \hat{s} := \hat{p}|_{I \setminus J} \). With

\[
f(q) := BC_{q,\hat{s}} g \quad \text{for } q \in [0,1]^I
\]

then \( \hat{q} \) maximises \( f \) on

\[
D \hat{s} := \{ q \in [0,1]^I : q \cup \hat{s} \in D \} = \{ q \in [0,1]^I : \kappa_\ell(BC_q) = c_\ell - \kappa_\ell(BC_{\hat{s}}) \text{ for } \ell \in \mathbb{L} \}
\]

with a certain \( \hat{d} = d(c, \hat{s}) \in \mathbb{R}^r \). Here the second representation of \( D \hat{s} \) follows from \( BC_{q,\hat{s}} = BC_q \ast BC_{\hat{s}} \) and (6), and the third from (12,13) and Lemma 2.4 applied to \( (S, K) \) and to \( (K, S) \), with \( q \) in the role of \( x \), and each \( \ell \in \mathbb{L} \) here in the role of \( r \) there.
Now \( f \) is a permutation invariant function of its \( r + 1 \) arguments, and affine-linear in each of these, and hence the implication (ii) \( \Rightarrow \) (iii) in Lemma 2.5, followed by Lemma 2.4 applied to \( (E,S) \) and to each \( \ell \in \mathbb{R}^{r+1} \) here in the role of \( r \) there, yields

\[
(16) \quad f(q) = \sum_{\ell=0}^{r+1} b_\ell E_\ell(q) = \frac{(-1)^r b_{r+1}}{r+1} S_{r+1}(q) + h(S_1(q), \ldots, S_r(q)) \quad \text{for } q \in [0,1]^J
\]

for some \( b = b(\tilde{\ell}) \in \mathbb{R}^{r+1} \) and some function \( h = h_0 \).

If \( \tilde{\ell} \) is a maximiser or minimiser of \( S_{r+1} \) on \( D_\delta \), then \#\{\( i \in J : \tilde{\ell}_i \in ]0,1[ \} \leq r \) by Lemma 2.1. If \( \tilde{\ell} \) is neither, then (16) and \( f(\tilde{\ell}) = \max_{q \in D_\delta} f(q) \) imply \( b_{r+1} = 0 \), hence \( f(q) = f(\tilde{\ell}) \) for every \( q \in D_\delta \), so that by (15) each \( q \cup \tilde{s} \) with \( q \in D_\delta \) maximises \( D \ni p \mapsto BC_p g \), and hence the second maximality property of \( \tilde{\ell} \) yields \( S_{r+1}(\tilde{\ell}) = S_{r+1}(\hat{\tilde{\ell}}) - S_{r+1}(\tilde{s}) \geq S_{r+1}(q \cup \tilde{s}) - S_{r+1}(\tilde{s}) = S_{r+1}(q) \) for every \( q \in D_\delta \), a contradiction.

Hence \( r' \leq r \), and the remaining claims are then obvious. \( \square \)

**Proof of Theorem 1.2.** We may assume \( a < b \), put \( n := \#I \), and

\[
\tilde{f}(p) := f(a + (b - a)p_i : i \in I) \quad \text{for } p \in [0,1]^I.
\]

Then \( \tilde{f} \) is a permutation invariant and separately affine-linear function on \([0,1]^I\), so that Lemma 2.5 yields a function \( g : n_0 \to \mathbb{R} \) with \( \tilde{f}(p) = BC_p g \) for \( p \in [0,1]^I \). Hence, using also \( S_\ell(a + (b - a)p_i : i \in I) = \sum_{k=0}^\ell (\begin{array}{c} \ell \\ k \end{array} ) (b - a)^k S_k(p) \) for \( \ell \in \mathbb{R} \) and \( p \in [0,1]^I \), and Lemma 2.4 with \( \{A, B\} = \{S, K\} \), the claim follows from Theorem 1.1. \( \square \)

**Acknowledgements**

We thank Norbert Henze and Patrick van Nerven for helpful remarks.

**References**

Links to full texts, if provided here, are green if they are free, and red if paywalled, as experienced by us at various times starting in March 2022. The asterisk * marks references I have taken from secondary sources, without looking at the original. Links in blue are for navigating within this paper.

**Carroll, F.W.** (1961). A polynomial in each variable separately is a polynomial. *American Mathematical Monthly* 68(1), 42. [http://www.jstor.org/stable/2311361](http://www.jstor.org/stable/2311361).

**Chebyshev, P.L.** (1846). See Chebichef (1846).

**Frisch, R.** (1926). Sur les semi-invariants et moments employé dans l'étude des distributions statistiques. Oslo, *Skripter af det Norske Videnskaps Akademie, II, Hist.-Filos. Klasse*, No. 3.

**Hald, A.** (1998). A *History of Mathematical Statistics* from 1750 to 1930. Wiley.

**Hald, A.** (2000). The early history of cumulants and the Gram–Charlier series. *International Statistical Review* 68, 137–153 (2000). [https://www.jstor.org/stable/1403665](https://www.jstor.org/stable/1403665).

**Haldane (1940).** The cumulants and moments of the binomial distribution, and the cumulants of \( \chi^2 \) for a \((n \times 2)\)-fold table. *Biometrika* 31, 392–396. [https://www.jstor.org/stable/2332619](https://www.jstor.org/stable/2332619).

**Hoeffding, W.** (1956). On the distribution of the number of successes in independent trials. *Annals of Mathematical Statistics* 27, 713–721. [https://doi.org/10.1214/aoms/1177728178](https://doi.org/10.1214/aoms/1177728178).

**John, F.** (1982). *Partial Differential Equations*, Fourth Edition, Springer.

**Kovačec, A., Kuhlmann, S. and Rienner (2012).** A note on extrema of linear combinations of elementary symmetric functions. *Linear and Multilinear Algebra* 60(2), 219–224. [https://doi.org/10.1080/03081087.2011.588438](https://doi.org/10.1080/03081087.2011.588438).

**Laplace, P.-S.** (1812, 1814, 1820). *Théorie analytique des probabilités*. Three editions, Courcier. First edition [https://gallica.bnf.fr/ark:/12148/btv1b8625611h](https://gallica.bnf.fr/ark:/12148/btv1b8625611h), third edition reprinted in *Laplace (1886)*.

**Laplace, P.-S.** (1886). *Oeuvres complètes*, Tome 7. Gauthier-Villars, [https://gallica.bnf.fr/ark:/12148/bpt6k775950](https://gallica.bnf.fr/ark:/12148/bpt6k775950).

**Macdonald, I.G.** (1995). *Symmetric Functions and Hall Polynomials*. Second Ed., Oxford University Press.
Mattner, L. (1999). What are cumulants? Documenta Mathematicae 4, 601–622, https://www.emis.de/journals/DMJDMV/vol-04/18.html.

Mattner, L. (2004). Cumulants are universal homomorphisms into Hausdorff groups. Probability Theory and Related Fields 130, 151–166, https://doi.org/10.1007/s00440-004-0354-y.

Mattner, L. and Roos, B. (2007). A shorter proof of Kanter's Bessel function concentration bound. Probability Theory and Related Fields 139, 191–205, https://doi.org/10.1007/s00440-006-0043-0.

Mattner, L. and Schulz, J. (2018). On normal approximations to symmetric hypergeometric laws. Transactions of the American Mathematical Society 370, 727–748, https://doi.org/10.1090/tran/6986.

Mattner, L. and Tasto, C. (2015). Confidence intervals for average success probabilities. Probability and Mathematical Statistics 35(2), 301–312, https://www.math.uni.wroc.pl/~pms/publications.php?nr=35.2.

Palais, R.S. (1978). Some analogues of Hartogs' theorem in an algebraic setting. American Journal of Mathematics 100(2), 387–405, http://www.jstor.org/stable/2373854.

Pitman, J. (1997). Probabilistic bounds on the coefficients of polynomials with only real zeros. Journal of Combinatorial Theory, Series A 77, 279–303, https://doi.org/10.1006/jcta.1997.2747.

Poisson, S.D. (1837). Recherche sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités. Bachelier, https://gallica.bnf.fr/ark:/12148/bpt6k110193z.

Stuart, A. and Ord, J.K. (1987). Kendall’s Advanced Theory of Statistics. Fifth Edition of Volume I. Distribution Theory. Griffin.

Tang, W. and Tang, F. (2019). The Poisson binomial distribution - old & new. https://arxiv.org/abs/1908.10024

Tchebichef, P. (1846). Démonstration élémentaires d’une proposition générale de la théorie des probabilités. Journal für die reine und angewandte Mathematik 33, 259–267, http://www.digizeitschriften.de/dms/resolverppn/?PID=PPN243919689_0033|log17.

*Thiele, T.N. (1889). Almindelig Iagttagelseslære: Sandsynlighedsregnig og mindste Kvadraters Methode. Reitzel. English translation (1903) reprinted as Thiele (1931).

*Thiele, T.N. (1899). Om Iagttagelseslærens Halvinvarianter. English translation “On the halfinvariants in the theory of observations” in Hald (2000, pp. 149–153).

Thiele, T.N. (1931). The theory of observations. Annals of Mathematical Statistics 2, 165–307, https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-2/issue-2.

Tichomirow, V.M. (1982). Grundprinzipien der Theorie der Extremalaufgaben. Teubner, Leipzig.

Uspensky, J.V. (1948). Theory of Equations. McGraw-Hill.

Vatutin, V.A. and Mikhailov, V.G. (1983). Limit theorems for the number of empty cells in an equiprobable scheme for group allocation of particles. Theory of Probability and Its Applications 27, 734–743, https://doi.org/10.1137/1127084. Russian original in Teoriya Veroyatnostei i ee Primeneneniya 27, 684–692 (1982), http://mi.mathnet.ru/eng/tvp2405.

Warren, D. (1999). The Frobenius-Harper technique in a general recurrence model. Journal of Applied Probability 36(1), 30–47, http://www.jstor.org/stable/3215400.

Universtität Trier, Fachbereich IV – Mathematik, 54286 Trier, Germany
Email address: mattner@uni-trier.de