COMBINING INDIVIDUALLY VALID AND CONDITIONALLY I.I.D. P-VARIABLES

BY LUTZ MATTNER

Universität Trier
August 22, 2011

Combining_P-variables_SWEAVE_arXiv_20110819.tex

For a given testing problem, let \( U_1, \ldots, U_n \) be individually valid and conditionally on the data i.i.d. P-variables (often called P-values). For example, the data could come in groups, and each \( U_i \) could be based on subsampling just one datum from each group, in order to justify an independence assumption under the hypothesis. The problem is then to deterministically combine the \( U_i \) into a valid summary P-variable. Restricting here our attention to functions of a given order statistic \( U_k \) of the \( U_i \), we compute the function \( f_{n,k} \) which is smallest among all increasing functions \( f \) such that \( f(U_k; n) \) is always a valid P-variable under the stated assumptions. Since

\[
f_{n,k}(u) \leq 1 \land (\frac{k}{n} u),
\]

with the right hand side being a good approximation for the left when \( k \) is large, one may in particular always take the minimum of 1 and twice the left sample median of the given P-variables.

We sketch the original application of the above in a recent study of associations between various primate species by Astaras et al.

1. Introduction and main result.

1.1. The problem and a summary of its tentative solution. This paper is motivated by a practical example of the following kind: Suppose that, for testing a hypothesis of interest, we have many observations collected over time and a test, or more precisely a family of tests indexed by the observed sample size, which is judged to be reasonable if independence of the observations could be assumed. If this latter assumption is in doubt only for observations close in time, say those taken on the same day, one might consider choosing one per day at random in order to apply the test only to those chosen. While this would be theoretically correct, or at least less faulty under the null hypothesis than ignoring the dependence issue, its reported result could be challenged by asking whether perhaps, after several trials, specifically a

---

AMS 2000 subject classifications: Primary 62F03, 62G10; secondary 60E15.

Keywords and phrases: Derandomization, multiple testing, P-value, sub-sampling, binomial distribution inequalities.
subset of the observations leading to the desired rejection had been chosen. Hence, given the data, one might consider repeating the random choice independently \(n\) times. If the test statistics are written as P-variables, see Subsection 1.3 below for formal definitions, this yields an example of the following more general problem: Given \(n\) individually valid and conditionally on the data i.i.d. P-variables \(U_1, \ldots, U_n\), find an appropriate summary P-variable, that is, find a permutation invariant measurable function \(F : [0,1]^n \to [0,1]\) such that \(V := F(U_1, \ldots, U_n)\) is a reasonable P-variable for our testing problem.

The main result of this paper, Theorem 1.1 below, gives simple solutions to this general problem, with each optimal within the set of all summary P-variables being isotone functions of a fixed order statistic \(U_{kn}\).

For practical purposes the results of this paper may be summarized as follows: If \(k \in \{1, \ldots, n\}\) is fixed in advance, then \(V := 1 \wedge (\frac{2}{n} U_{kn})\) is a P-variable under the stated assumptions. For \(U_{kn}\) small, and this is the only case of interest, and for \(k\) rather large, this \(V\) is approximately the smallest isotone function of \(U_{kn}\) being a P-variable. For moderate \(k\), the constant \(\frac{2}{n}\) may be replaced by a somewhat smaller one.

Given \(n\), it thus remains to select \(k\). We tentatively suggest on intuitive grounds \(k = \lceil \frac{n+1}{2} \rceil\), so that practically \(V\) is twice the left sample median of the \(U_i\), but we have to admit the mathematical arbitrariness of this choice. We applied this choice in [1] as explained in Example 2.2 below. This application triggered off the present paper.

Before stating our results formally, let us finish this introduction by stating the obvious unsolved problems: Is there any convincing rationale for choosing \(k\) in a particular way? Are there competitive summary P-variables which are not functions of single order statistics? Are there significantly better summary P-variables for interesting particular cases of our general problem, say in the examples of Section 2 below?

1.2. Analysis and probability notations. We assume \(n \in \mathbb{N} := \{1, 2, 3, \ldots\}\) and put \(\frac{0}{0} := 0\), unless noted otherwise. Terms like “positive” or “increasing” are used in the wide sense. We write Argmax \(f\) for the set of all global maximizers of the real-valued function \(f\), and argmax \(f\) for its unique element if existing. We use indicator notation for sets, like \(1_A\), and for statements, like \((x \in A) = 1_A(x)\). Measurable spaces \((\mathcal{X}, \mathcal{A})\) are simply denoted by \(\mathcal{X}\) if the \(\sigma\)-algebra \(\mathcal{A}\) is clear from the context or irrelevant. \(\mathcal{B}([a, b])\) denotes the Borel \(\sigma\)-algebra of the interval \([a, b]\).

Let \(\mathcal{X}\) and \(\mathcal{Y}\) be measurable spaces. Then \(\text{Prob}(\mathcal{X})\) denotes the set of all probability measures on \(\mathcal{X}\) and \(\text{Mark}(\mathcal{X}, \mathcal{Y})\) the set of all Markov kernels from \(\mathcal{X}\) to \(\mathcal{Y}\). Let \(P \in \text{Prob}(\mathcal{X})\) and \(K \in \text{Mark}(\mathcal{X}, \mathcal{Y})\). Then their product is
denoted by $P \otimes K$, that is, $(P \otimes K)(C) := \int_{\mathcal{X}} \int_{\mathcal{Y}} 1_{C}(x, y) K(x, dy) P(dx)$ for $C$ belonging to the appropriate product $\sigma$-algebra. The same symbol $\otimes$ is also used for products of probability measures and, accordingly, $P^{\otimes n} := \bigotimes_{j=1}^{n} P$ denotes the product of the $n$ identical factors $P \in \Prob(\mathcal{X})$. Thus, e.g. in (5) in below, an expression like $K(x, \cdot)^{\otimes n}$ with $K \in \Mark(\mathcal{X}, [0, 1])$ and $x \in \mathcal{X}$, is a product of $n$ probability measures on $[0, 1]$, not a product really involving kernels.

For $P \in \Prob(\mathcal{X})$ and $S : \mathcal{X} \to \mathcal{Y}$ measurable, we use the nonstandard notation $S \Box P$ for the distribution, or image measure, of $S$ with respect to $P$. We write $B_{n,p}$ for the binomial distribution with counting density $b_{n,p}$, that is, $b_{n,p}(k) = \binom{n}{k} p^{k} (1 - p)^{n-k}$ for $n \in \mathbb{N} \cup \{0\}$, $p \in [0, 1]$, and $k \in \{0, \ldots, n\}$, and further $U_{[a,b]}$ and $U_{\{1,\ldots,n\}}$ for continuous or discrete uniform distributions on the indicated sets. Our use of nonitalic letters here allows us to use e.g. $U$ for a random variable.

1.3. Hypotheses, P-variables, P-kernels, P-values. In this paper, we call hypothesis any set $\mathcal{P}_0 \subseteq \Prob(\mathcal{X})$ of probability measures on the same sample space $\mathcal{X}$. We might think of $\mathcal{P}_0$ as a subset of some strictly larger set $\mathcal{P} \subseteq \Prob(\mathcal{X})$, with then $\mathcal{P} \setminus \mathcal{P}_0$ called the alternative, but only $\mathcal{P}_0$ matters in the present investigation.

Let $\mathcal{P}_0 \subseteq \Prob(\mathcal{X})$ be a hypothesis. A P-variable for $\mathcal{P}_0$ is a statistic $U : \mathcal{X} \to [0, 1]$ which is under $\mathcal{P}_0$ stochastically larger than the uniform distribution on $[0, 1]$, that is,

$$P(U \leq \alpha) \leq \alpha \quad \text{for } P \in \mathcal{P}_0 \text{ and } \alpha \in [0, 1]$$

A P-kernel for $\mathcal{P}_0$ is a kernel $K \in \Mark(\mathcal{X}, [0, 1])$ with

$$(P \otimes K)(\mathcal{X} \times [0, \alpha]) \leq \alpha \quad \text{for } P \in \mathcal{P}_0 \text{ and } \alpha \in [0, 1]$$

Finally, a P-value is any number $\in [0, 1]$ resulting from any process we happen to model by a P-variable or a P-kernel.

Clearly, a kernel $K \in \Mark(\mathcal{X}, [0, 1])$ is a P-kernel for $\mathcal{P}_0$ iff the coordinate projection $\mathcal{X} \times [0, 1] \ni (x, u) \mapsto u$ is a P-variable for the hypothesis $\{P \otimes K : P \in \mathcal{P}_0\}$ on the sample space $\mathcal{X} \times [0, 1]$; we refer to the latter hypothesis and sample space as extended versions of the original ones. Conversely, a statistic $U : \mathcal{X} \to [0, 1]$ is a P-variable for $\mathcal{P}_0$ iff the kernel $\mathcal{X} \times \mathcal{B}([0, 1]) \ni (x, B) \mapsto \delta_{U(x)}(B)$ is a P-kernel for $\mathcal{P}_0$, and we refer to such a $K$ as deterministic. In statistical practice, non-deterministic P-kernels often arise through the application of Monte Carlo tests, as in Example 2.2 below.
1.4. The main result. Let \( n \in \mathbb{N} \) and \( k \in \{1, \ldots, n\} \) be fixed in this subsection, except for the penultimate sentence. Let \( U_{k:n} \) denote the \( k \)-th order statistic on \([0,1]^n\), so \( U_{k:n}(u) := \min\{v \in [0,1] : \sum_{i=1}^n (u_i \leq v) \geq k\} \) for \( u \in [0,1]^n \). We put

\[
\psi_{n,k}(p) := p^{-1}B_{n,p}(\{k, \ldots, n\}) \quad \text{for } p \in [0,1]
\]

with \( \psi_{n,k}(0) \) defined by continuity. Let \( p_{1,1} := 0 \). If not \( n = k = 1 \), let \( p_{n,k} := \arg\max \psi_{n,k} \) which exists according to Lemma 3.2 below. Finally, let

\[
c_{n,k} := \max_{p \in [0,1]} \psi_{n,k}(p)
\]

and

\[
f_{n,k}(u) := u\psi_{n,k}(p_{n,k} \lor u) = \begin{cases} c_{n,k}u & \text{for } u \in [0,p_{n,k}] \\ B_{n,u}(\{k, \ldots, n\}) & \text{for } u \in [p_{n,k}, 1] \end{cases}
\]

The following theorem and its accompanying proposition, both proved in Section 3 below, make precise the claims from Subsection 1.1 and constitute the main technical result of this paper.

**Theorem 1.1.** The following two conditions on increasing functions \( f : [0,1] \rightarrow [0,1] \) are equivalent:

(i) If \( K \) is a \( P \)-kernel for a hypothesis \( \mathcal{P}_0 \) on a sample space \( \mathcal{X} \), then \( K_{n,k,f} \in \text{Mark}(\mathcal{X}, [0,1]) \) defined by

\[
K_{n,k,f}(x, \cdot) := (f \circ U_{k:n}) \circ K(x, \cdot)^\otimes n \quad \text{for } x \in \mathcal{X}
\]

is a \( P \)-kernel for \( \mathcal{P}_0 \).

(ii) \( f \geq f_{n,k} \).

Since \( f_{n,k} : [0,1] \rightarrow [0,1] \) is increasing, Theorem 1.1 states that \( f_{n,k} \) is the smallest, and thus optimal, eligible function satisfying condition (i).

**Proposition 1.1.** \( f_{n,k} \) is a continuous and increasing bijection of \([0, 1]\) onto \([0, 1]\), and is linear with slope \( c_{n,k} \) on \([0, p_{n,k}] \supseteq [0, \frac{k-1}{n-1}]\). We have

\[
c_{n,k} \leq \frac{n}{k} \leq \frac{n}{k+1} 
\]

and, for \( u \in [0,1] \),

\[
\frac{1 \land \frac{nu}{k}}{1 + 5k^{-1/3}} \leq f_{n,k}(u) \leq 1 \land (c_{n,k}u) \leq 1 \land \frac{nu}{k}
\]
In (6) and (7), the main interest is in the upper bounds for $c_{n,k}$ and $f_{n,k}$.

The probably improvable lower bounds show that $f_{n,k}(u)$ is asymptotically equivalent to its upper bound $1 \land \frac{nu}{k}$, for $k \to \infty$, uniformly in $u$ and in $n \geq k$. A numerical computation of $p_{n,k}$, and hence of $c_{n,k}$ and $f_{n,k}(u)$, is straightforward due to the monotonicity properties of $\psi_{n,k}$ given in Lemma 3.2 below.

2. Examples.

2.1. This is a formalization of the fictitious example from the beginning of Subsection 1.1. With a notation slightly different from that of Subsections 1.3 and 1.4 (here $X^N$ and $Q_0$, there $\mathcal{X}$ and $\mathcal{P}_0$), let $\mathcal{X}$ be a sample space and let $\mathcal{P}_0 \subseteq \text{Prob}(\mathcal{X})$. Informally speaking, we want to test $\mathcal{P}_0$ based on several but possibly dependent observables. We suppose we know how to handle the i.i.d. case: For $m \in \mathbb{N}$, let $U^{(m)}$ be a P-variable for the hypothesis $\{P^{\otimes m} : P \in \mathcal{P}_0\}$ on the sample space $\mathcal{X}^m$. We further assume that we have $N = \sum_{j=1}^m N_j$ observations $X_{ji}$ coming in $m$ independent groups as

$$(X_{1i}, \ldots, X_{N_1}), \ldots, (X_{mi}, \ldots, X_{mN_m})$$

with possibly arbitrary dependencies within each group, so that our hypothesis can be formalized as

$$Q_0 := \left\{ \bigotimes_{j=1}^m Q_j : Q_j \in \text{Prob}(\mathcal{X}^{N_j}), \text{ all } \mathcal{X}\text{-marginals equal and } \in \mathcal{P}_0 \right\}$$

don the sample space $\mathcal{X}^N$. Then randomly picking just one observable from each of the $m$ blocks and applying $U^{(m)}$ to these yields a valid P-variable. Formally put,

$$K(x, \cdot) := \left( \bigotimes_{j=1}^m \{1, \ldots, N_j\} \ni i \mapsto (x_{1i}, \ldots, x_{mi}) \right) \otimes \bigotimes_{j=1}^m U_{\{1, \ldots, N_j\}}$$

for $x = ((x_{11}, \ldots, x_{1N_1}), \ldots, (x_{m1}, \ldots, x_{mN_m})) \in \mathcal{X}^N$ defines a P-kernel for $Q_0$. Repeating the random picking $n$ times amounts to considering the kernel $K_n$ from $\mathcal{X}^N$ to $[0, 1]^n$ defined by $K_n(x, \cdot) := K(x, \cdot)^\otimes n$, and Theorem 1.1 shows how to transform it into a P-kernel.

2.2. Primate associations. In [1] we wanted to test a hypothesis of “no association” between seven primate species in a certain area (38 km$^2$ within a national park in Cameroon). The data, obtained by patrolling 3284 km
LUTZ MATTNER

over 217 days and recording the species composition of 612 observed primate clusters, is a matrix

$$x = (x_{ij}) \in \{0, 1\}^{612 \times 7} =: \mathcal{X}$$

with

$$x_{ij} := \begin{cases} 1 & \text{if species } j \text{ is present} \\ 0 & \text{absent} \end{cases} \text{ in cluster } i$$

For example, the clusters observed on the first two and last two observation days yielded

| date       | putty | mona | redear | crowned | drill | mangabey | redcol |
|------------|-------|------|--------|---------|-------|----------|--------|
| 1 2006-02-03 | 1 0   | 1    | 0      | 0       | 0     | 0        | 0      |
| 2 2006-02-03 | 1 0   | 0    | 0      | 0       | 0     | 0        | 0      |
| 3 2006-02-03 | 0 0   | 1    | 1      | 0       | 0     | 1        |        |
| 4 2006-02-04 | 0 0   | 0    | 1      | 0       | 0     | 0        | 0      |
| 5 2006-02-04 | 1 0   | 1    | 0      | 0       | 0     | 0        | 0      |
| 6 2006-02-04 | 1 0   | 0    | 0      | 0       | 0     | 0        | 0      |
| 7 2006-02-04 | 0 0   | 0    | 0      | 0       | 1     | 0        |        |
| 8 2006-02-04 | 1 0   | 0    | 0      | 0       | 0     | 0        | 0      |
| 9 2006-02-04 | 1 0   | 0    | 1      | 0       | 0     | 1        |        |
| 607 2008-01-17 | 1 0 | 1   | 0      | 0       | 1     | 0        |        |
| 608 2008-01-17 | 1 0   | 0    | 0      | 0       | 0     | 0        | 0      |
| 609 2008-01-17 | 1 0  | 1    | 1      | 0       | 0     | 1        |        |
| 610 2008-01-17 | 1 1   | 0    | 0      | 0       | 0     | 0        | 0      |
| 611 2008-01-18 | 0 1   | 0    | 1      | 0       | 0     | 0        | 0      |
| 612 2008-01-18 | 1 0   | 0    | 1      | 0       | 0     | 0        | 0      |

Here the columnnames starting from “putty” are abbreviations for English species names, with the corresponding scientific names obtainable from [1, p. 130, Fig. 2]. For example, “putty” is an abbreviation for “putty-nosed guenon” or “Cercopithecus nictitans”.

We write $x_+$ and $x_+$ for the row and column sums vectors of $x \in \mathcal{X}$. The “no association” hypothesis is formalized as

$$\mathcal{P}_0 := \{ P \in \text{Prob}(\mathcal{X}) : P(\{x\}) = P(\{y\}) \text{ if } x_+ = y_+ \text{ and } x_+ = y_+ \}$$

This formalization is common in ecology and in social network analysis, see [5] for some references, it is certainly debatable, but it is for instance larger, and hence more interesting to reject, than assuming that rows are independent and each represents a random sample from the species’ which is simple (in the usual sense of all samples being equally likely) if conditioned on its possibly random size. A submodel of $\mathcal{P}_0$, consisting of all laws of $\mathcal{X}$-valued random variables $X = (X_{ij})$ with independent indicators $X_{ij}$ with success probabilities $\alpha_i \beta_j / (1 + \alpha_i \beta_j)$ for some $\alpha_i, \beta_j \in [0, \infty[$, was originally proposed by Rasch for a completely different situation, see [7] page 75].
Let $T : \mathcal{X} \to \mathbb{R}$ be a statistic to be used for testing $P_0$, with sufficiently large values to be regarded as significant. The corresponding P-variable $U$ appears impossible to compute, but Besag and Clifford constructed in [2] a P-kernel for $P_0$ using Markov chains as follows:

Fix a length $N \in \mathbb{N}$. Given the observation $x \in \mathcal{X}$, let $\tau$ be uniformly distributed on $\{1, \ldots, N\}$, then let $X_{\tau} := x$, then construct $X_{\tau+1}, \ldots, X_N$ as a Markov chain starting from $X_{\tau}$ and using stationary transition probabilities with the uniform distribution on $\mathcal{X}(x) := \{y \in \mathcal{X} : y_+ = x_+ \text{ and } y_{++} = x_{++}\}$ as their stationary distribution, then analogously construct $X_{\tau-1}, \ldots, X_1$ using the corresponding reversed transition probabilities. Then take the observed value of

$$\frac{1}{N} \sharp \{t : T(X_{\tau}) \leq T(X_t)\}$$

as our P-value. This is modelled by a valid P-variable, on a suitably extended sample space, since $\tau$ and the chain $(X_1, \ldots, X_N)$ are independent under the hypothesis $P_0$. See [2] and [5] for more details.

With a certain $T$ and with $N = 10^8$, the above yielded a P-value of 0.006 for our data. (This may be regarded as a P-value obtained with a practically, though not theoretically, deterministic P-kernel, as running a few repetitions of the described Besag-Clifford procedure showed.) But perhaps the null model we then wanted to reject is inappropriate not due to an interesting scientific reason but, for example, due to a possible reobservation of the same primate cluster on the same day. Hence we reanalyzed the data by randomly picking just one cluster per day. Doing this $n = 1000$ times yielded P-values with sample quartiles 0.01, 0.03, 0.09, maximum 0.7, and the following histogram:
Now what to report as an appropriate summary P-value? With the mathematical formalization analogous to Example 2.1 we decided to use Theorem 1.1 with \( k = \frac{n}{2} \), yielding \( c_{n,k} = 1.846 \), and we thus reported \( 0.03 \times 1.846 = 0.06 \) in [1, Table II, last column].

2.3. Testing Hardy-Weinberg equilibrium using complex survey data. In [6], Li and Graubard propose certain tests for Hardy-Weinberg equilibrium and related hypotheses for humans, based on data obtained with certain complex survey designs, often leading to samples with individuals belonging to the same household. To address the dependence problem due to people living in the same household possibly being blood related, they “selected one person randomly from each household to form a subsample”. Their “analyses using 10 subsamples produce(d) consistent results . . . and in general agreed with results using the full sample analyses”. See [6, pages 1103, 1101]. So here the P-variable combination problem is not properly addressed, at least not explicitly. Perhaps one could interpret “produce(d) consistent results” as looking for the maximum of the P-variables and multiplying it by the then trivial correction constant \( c_{10,10} = 1 \), but such a procedure should have been fixed before looking at the 10 actual P-values.

3. Auxiliary results and proofs. As in Subsection 1.4 let \( n \in \mathbb{N} \) and \( k \in \{1, \ldots, n\} \) be fixed. In Lemma 3.1 and in the proofs of Lemma 3.2 and Proposition 1.1, we write

\[
\varphi(p) := \varphi_{n,k}(p) := B_{n,p}(\{k, \ldots, n\})
\]

for \( p \in [0,1] \), so that \( \psi_{n,k}(p) = p^{-1} \varphi(p) \), with the right hand side defined by continuity for \( p = 0 \). We recall that \( \frac{k-1}{n-1} := 0 \) in case of \( k = n = 1 \).

**Lemma 3.1.** The function \( \varphi \) is continuous, strictly increasing, strictly convex on \( [0, \frac{k-1}{n-1}] \), concave on \( [\frac{k-1}{n-1}, 1] \), and satisfies \( \varphi(0) = 0 \).

**Proof.** For \( p \in ]0,1[ \), we have \( \varphi'(p) = n b_{n-1,p}(k-1) \) and

\[
\varphi''(p) = \frac{n b_{n-1,p}(k-1)}{p(1-p)} (k - 1 - (n - 1)p)
\]

and everything is obvious.

**Lemma 3.2.** We have \( \psi_{1,1}(p) = 1 \) for \( p \in [0,1] \), while for \( n \geq 2 \), \( p_{n,k} = \arg\max \psi_{n,k} \) exists and belongs to \( [\frac{k-1}{n-1}, 1] \), with \( \psi_{n,k} \) strictly increasing on \( [0, p_{n,k}] \) and strictly decreasing on \( [p_{n,k}, 1] \). We further have

\[
f_{n,k}(u) = \max_{p \in [u,1]} \psi_{n,k}(p) \quad \text{for } u \in [0,1]
\]
Proof. The case of \( n = 1 \) is trivial, so let us assume \( n \geq 2 \). Let us abbreviate \( \psi := \psi_{n,k} \). On \( [0, \frac{k-1}{n-1}] \), the function \( \varphi \) is strictly convex by Lemma 3.1, hence \( \psi(p) = \frac{\varphi(p) - \varphi(0)}{p-0} \) is strictly increasing there. For \( p \in [0, 1] \), we have}

\[
\psi'(p) = p^{-2} \left( p \varphi'(p) - \varphi(p) \right) = p^{-2} \omega(p)
\]

\[
\omega'(p) = p \varphi''(p) = \frac{n \ln p (k-1)}{1-p} (k-1-(n-1)p)
\]

by (8). So \( \omega'(p) < 0 \) for \( p \in [\frac{k-1}{n-1}, 1] \). Hence there is at most one \( p_0 \in [\frac{k-1}{n-1}, 1] \) with \( \psi'(p_0) = 0 \), and if such a \( p_0 \) exists, then \( \psi' \) changes sign from plus to minus at \( p_0 \). Thus \( p_{n,k} = p_0 \) if \( p_0 \) as above exists, and otherwise \( \psi \) is strictly monotone on \( [\frac{k-1}{n-1}, 1] \) and \( p_{n,k} \) is the appropriate element of \( \{ \frac{k-1}{n-1}, 1 \} \). This proves all claims about \( \psi \), and (9) follows, see the by now justified definition in (4). \( \square \)

**Lemma 3.3.** We have

\[
(10) \quad \sup_{p \in [0,1]} \frac{k}{np} B_{n,p}(\{k, \ldots, n\}) \leq 1
\]

\[
(11) \quad \max_{p \in [\frac{k}{n},1]} \frac{k}{np} B_{n,p}(\{k, \ldots, n\}) \geq (1 + 5k^{-1/3})^{-1}
\]

Proof. Relation (10), following from \( k B_{n,p}(\{k, \ldots, n\}) \leq \sum_{j=0}^{n} j b_{n,p}(j) \), is a simple Markov inequality. To prove (11), let \( g(p) := \frac{k}{np} B_{n,p}(\{k, \ldots, n\}) \) for \( p \in [\frac{k}{n}, 1] \). If also \( p \neq \frac{k}{n} \), then \( k < \mu(B_{n,p}) \) and the one-sided Chebyshev inequality [1, page 476, (3.2)] yields

\[
B_{n,p}(\{k, \ldots, n\}) \geq 1 - B_{n,p}(\{0, \ldots, k\}) \geq \frac{(np-k)^2}{np(1-p) + (np-k)^2}
\]

and hence

\[
g(p) \geq \left( \frac{np}{k} + \frac{p^2(1-p)}{k(p-k)^2} \right)^{-1} \geq \left( \frac{np}{k} + \frac{p^2}{k(p-k)^2} \right)^{-1} =: h(p)
\]

If now in particular \( p := (k + k^{2/3})/n \) satisfies \( p \leq 1 \), then we get

\[
h(p) = \left( 1 + k^{-1/3} + k^{-1/3} (1 + k^{-1/3})^2 \right)^{-1} \geq \left( 1 + 5k^{-1/3} \right)^{-1}
\]

and otherwise we have \( \frac{k}{n} > (1+k^{-1/3})^{-1} \) and hence \( g(1) = \frac{k}{n} > (1+5k^{-1/3})^{-1} \). Thus (11) holds in every case. \( \square \)
Proof of Proposition 1.1. Let us abbreviate \( f := f_{n,k} \). By the definition in (4), \( f \) is positive and continuous, since \( \psi_{n,k} \) is so, and also \( f(0) = 0 \). By the second representation in (4), \( f \) is strictly increasing, using \( c_{n,k} > 0 \) and \( \varphi \) strictly increasing, and we have \( f(1) = 1 \). With \( p_{n,k} \geq \frac{k-1}{n-1} \) from Lemma 3.2, this proves the claimed mapping properties of \( f \).

Inequalities (6) follow from (10) and (11). For \( u \leq \frac{k}{n} \), inequality (11) also yields the first inequality in (7), since then (9) shows that \( f(u) \) is at least \( \frac{nu}{k} \) times the left hand side of (11), while for \( u \geq \frac{k}{n} \) the inequality to be proved persists, since the left hand side is constant and the right hand side increasing. Next, \( f(u) \leq c_{n,k} u \) for \( u \in [0,1] \) follows from \( \psi_{n,k}(p_{n,k} \vee u) \leq c_{n,k} \), and \( f(u) \leq 1 \) was proved above. The final inequality in (7) follows from (6). \( \square \)

Now let us recall the definition and a well-known lemma concerning the convex order: For \( P,Q \in \text{Prob}(\mathbb{R}) \) with finite means \( \mu(P) \) and \( \mu(Q) \), one writes \( P \leq_{cx} Q \) if \( \int \varphi \, dP \leq \int \varphi \, dQ \) for every convex function \( \varphi : \mathbb{R} \to ]-\infty, \infty] \). Then necessarily \( \mu(P) = \mu(Q) \).

Lemma 3.4. Let \( a,b \in \mathbb{R} \) with \( a < b \) and let \( P \in \text{Prob}([a,b]) \) with mean \( \mu \). Then \( \delta_\mu \leq_{cx} P \leq_{cx} \frac{b-a}{b-a} \delta_a + \frac{\mu-a}{b-a} \delta_b \).

Proof. This is trivial if \( \mu \in \{a,b\} \). Otherwise, for \( \varphi \) convex on \([a,b] \), apply \( \int \ldots dP(t) \) to \( \varphi(\mu) + \varphi'(\mu+)(t-\mu) \leq \varphi(t) \leq \frac{b-t}{b-a} \varphi(a) + \frac{t-a}{b-a} \varphi(b) \). \( \square \)

Lemma 3.5. Let \( \alpha,c \in [0,1] \) and let \( \varphi : [0,1] \to \mathbb{R} \) be a continuous and increasing function, convex on \([0,c]\) and concave on \([c,1]\). Then the set

\[
\text{Argmax} \left( \int_{[0,1]} \varphi \, dP : P \in \text{Prob}([0,1]) \text{ with } \mu(P) \leq \alpha \right)
\]

has nonempty intersection with

\[
\left\{ \left( 1 - \frac{\alpha}{t} \right) \delta_0 + \frac{\alpha}{t} \delta_t : t \in [\alpha \vee c, 1] \right\}
\]

Proof. Let \( \mathcal{P} \) denote the set in (12). As \( \{P \in \text{Prob}([0,1]) : \mu(P) \leq \alpha \} \) is nonempty and closed with respect to convergence in distribution, the compactness of \([0,1]\) and the continuity of \( \varphi \) imply by Prohorov’s theorem (in the elementary Helly case, see e.g. [3] Section 25) that \( \mathcal{P} \) is nonempty.

So let \( P \in \mathcal{P} \). If \( \mu(P) < \alpha \), then \( P_\varepsilon := (1-s)P + s \delta_1 \in \text{Prob}([0,1]) \) with \( s := (\alpha - \mu(P))/(1 - \mu(P)) \) satisfies \( \mu(P_\varepsilon) = \alpha \) and, since \( \varphi \) is increasing, \( \int \varphi \, dP \leq \int \varphi \, dP_\varepsilon \). Hence we may assume \( \mu(P) = \alpha \) in what follows. If \( P([0,c]) > 0 \) and \( c > 0 \), then we apply the second inequality in Lemma 3.4 with \( a := 0 \) and \( b := c \) to the law \( \mathcal{B}([0,c]) \ni B \mapsto P(B)/P([0,c]) \) with mean \( \lambda \), say, to see that
\[
\int \varphi \, dP \leq \int \varphi \, dQ \quad \text{where} \quad Q := P([0, c]) (\frac{\lambda}{c} \delta_0 + \frac{\lambda}{c} \delta_c) + 1_{[c, 1]} P \in \mathcal{P} \quad \text{also satisfies} \quad \mu(Q) = \alpha. \]
Hence we may also assume \(P([0, c]) = 0\) in what follows. If, finally, \(P([c, 1]) > 0\) and \(c < 1\), then we apply the first inequality in Lemma 3.4 with \(a := c\) and \(b := 1\) to the law \(\mathcal{B}([c, 0]) \ni B \mapsto P(B)/P([c, 1])\) with mean \(\varrho\), say, to see that \(\int \varphi \, dP \leq \int \varphi \, dR\) where \(R := P(\{0\}) \delta_0 + P([c, 1]) \delta_c\) if \(c > 0\), and \(R := \delta_{\varrho}\) if \(c = 0\), so that \(R\) belongs to \(\mathcal{P}\) and to the set in (13).

From Lemmas 3.3 and 3.1, recalling (9) and observing that each member of the set in (13) has mean \(\alpha\), we get

**Lemma 3.6.** For every \(\alpha \in [0, 1]\), we have

\[
\max \left\{ \int B_{n,t}(\{k, \ldots, n\}) \, dP(t) : P \in \text{Prob}([0, 1]), \mu(P) \leq \alpha \right\} = f_{n,k}(\alpha)
\]

**Proof of Theorem 1.1.** Let \(f : [0, 1] \to [0, 1]\) be an increasing function, fixed in the entire proof. If \(\alpha \in [0, 1]\) and

\[
I := f^{-1}([0, \alpha]), \quad \beta := \sup I, \quad \sup \emptyset := 0
\]

so that \(I = [0, \beta] \text{ or } I = [0, \beta]\), and if further \(K\) is a P-kernel for a hypothesis \(\mathcal{P}_0\) on a sample space \(\mathcal{X}\), \(K_{n,k,f}\) is defined as in (5), \(P \in \mathcal{P}_0\) and \(\tilde{P} := K(\cdot, I) \circ P \in \text{Prob}([0, 1])\)

then we have

\[
(P \otimes K_{n,k,f})(\mathcal{X} \times [0, \alpha]) = \int_{\mathcal{X}} (U_{k,n} \circ K(x, \cdot) \otimes n)(I) \, dP(dx)
\]

(16)

\[
= \int_{\mathcal{X}} B_{n,K(x,I)}(\{k, \ldots, n\}) \, dP(dx)
\]

(17)

\[
\leq f_{n,k}(\beta)
\]

where at (16) we have used the standard formula for the distribution functions of order statistics [3, Exercise 14.7, unfortunately missing in the third edition from 1995], and where inequality (18) follows from Lemma 3.6 applied to \(\beta\) and \(\tilde{P}\), as \(\mu(\tilde{P}) = \int_{\mathcal{X}} K(\cdot, I) \, dP = \left( P \otimes K \right)(\mathcal{X} \times I) \leq \left( P \otimes K \right)(\mathcal{X} \times [0, \beta]) \leq \beta\) by the P-kernel assumption (1).

Now let us assume condition (ii). Then, for every \(\alpha \in [0, 1]\), the corresponding \(\beta\) from (14) satisfies \(\beta \leq \sup f_{n,k}^{-1}([0, \alpha]) = f_{n,k}^{-1}(\alpha)\), using Proposition 1.1 for the last equality, and hence

\[
f_{n,k}(\beta) \leq f_{n,k}(f_{n,k}^{-1}(\alpha)) = \alpha
\]
so that \((P \otimes K_{n,k,f})(X \times [0,\alpha]) \leq \alpha\). Hence (i) follows.

To prepare for the proof of the converse implication, let us consider \(X := [0,1], \mathcal{P}_0 := \{P\}\) with \(P := U_{[0,1]}\) and \(t \in [0,1]\). Then

\[
K(x,\cdot) := t\delta_{xt} + (1-t)U_{[t,1]} \quad \text{for} \quad x \in X
\]
defines a \(P\)-kernel \(K \in \text{Mark}(X,[0,1])\), since for \(\alpha \in [0,1]\) we have

\[
\begin{align*}
K(x,[0,\alpha]) &= t \cdot (xt \leq \alpha) + (\alpha - t)_+ \\
K(x,[0,\alpha]) &= t \cdot (xt < \alpha) + (\alpha - t)_+
\end{align*}
\]
for \(x \in X\) and hence, using just (19),

\[
(P \otimes K)(X \times [0,\alpha]) = \alpha \wedge t + (\alpha - t)_+ = \alpha
\]
The identities (19) and (20) further yield

\[
K(\cdot,[0,\alpha]) \circ P = K(\cdot,[0,\alpha]) \circ P = (1 - \frac{\alpha}{t} \wedge 1)\delta_{t+(\alpha - t)_+}
\]
for \(\alpha \in [0,1]\). Let us now fix \(\alpha \in [0,1]\) for the rest of this paragraph, define \(I\) and \(\beta\) as in (14) above, recall \(\psi_{n,k}\) and \(p_{n,k}\) from Section 1.4, and specialize to \(t := p_{n,k} \lor \beta\). Then (15) and (21), the latter with \(\beta\) in place of \(\alpha\), yield

\[
\tilde{P} = (1 - \frac{\beta}{t})\delta_0 + \frac{\beta}{t}\delta_t
\]
and the integral in (17) reduces to \(\beta \psi_{n,k}(t) = f_{n,k}(\beta)\) by the definition in (4), and hence we get

\[
(P \otimes K_{n,k,f})(X \times [0,\alpha]) = f_{n,k}(\beta)
\]
Now let us assume condition (i). Then the left hand side of (22) is at most \(\alpha\), and hence we get

\[
f_{n,k}(\sup f^{-1}([0,\alpha])) \leq \alpha \quad \text{for} \quad \alpha \in [0,1]
\]
We recall Proposition 1.1 for the properties of \(f_{n,k}\) used below. If now \(f(\alpha_0) < f_{n,k}(\alpha_0)\) for some \(\alpha_0 \in [0,1]\), then \(\alpha_0 > 0\) as \(f_{n,k}(0) = 0\), and hence the continuity of \(f_{n,k}\) yields an \(\alpha_1 \in [0,\alpha_0]\) with \(f(\alpha_0) \leq f_{n,k}(\alpha_1)\), and thus

\[
\sup f^{-1}([0,f_{n,k}(\alpha_1)]) \geq \sup f^{-1}([0,f(\alpha_0)]) \geq \alpha_0 > \alpha_1
\]
and hence

\[
f_{n,k}(\sup f^{-1}([0,f_{n,k}(\alpha_1)])) > f_{n,k}(\alpha_1)
\]
in contradiction to (23) with \(\alpha := f_{n,k}(\alpha_1)\). Thus (ii) holds. \(\square\)
Acknowledgements. Thanks are due to my coauthors from [1] for the cooperation leading to the problem treated in the present paper, in particular to Christos Astaras for collecting the data and initiating our joint work, and to Stefan Krause for numerous discussions about the data analysis, and also for actually doing the computations mentioned in Example 2.2. I further thank Jannis Dimitriadis for proofreading and Todor Dinev for help with Sweave.

References.

[1] Astaras, C., Krause, S., Mattner, L., Rehse, C. and Waltert, M. (2011). Associations between the drill (Mandrillus leucophaeus) and sympatric primates in Korup National Park, Cameroon. American Journal of Primatology 73, 127-134.
[2] Besag, J. and Clifford, P. (1989). Generalized Monte Carlo significance tests. Biometrika 76, 633-642.
[3] Billingsley, P. (1986). Probability and Measure. Second edition, Wiley.
[4] Karlin, S. and Studden, W.J. (1966). Tschebycheff Systems: With Applications in Analysis and Statistics. Wiley.
[5] Krause, S., Mattner, L., James, R., Guttridge, T., Corcoran, M.J., Gruber, S.H. and Krause, J. (2009). Social network analysis and valid Markov Chain Monte Carlo tests of null models. Behavioral Ecology and Sociobiology 63, 1089-1096.
[6] Li, Y. and Graubard, B.I. (2009). Testing Hardy-Weinberg equilibrium and homogeneity of Hardy-Weinberg disequilibrium using complex survey data. Biometrics 65, 1096-1104.
[7] Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests, Danish Institute for Educational Research. Reprinted 1980 “with a Foreword and Afterword by Benjamin Wright”, The University of Chicago Press. Reprinted 2010, The Institute for Objective Measurement, Chicago.