A Cantelli-type inequality for constructing non-parametric p-boxes based on exchangeability

Matthias C. M. Troffaes MATTHIAS.TROFFAES@DURHAM.AC.UK
Tathagata Basu TATHAGATA.BASU@DURHAM.AC.UK
Department of Mathematical Sciences, Durham University, UK

Abstract

In this paper we prove a new probability inequality that can be used to construct p-boxes in a non-parametric fashion, using the sample mean and sample standard deviation instead of the true mean and true standard deviation. The inequality relies only on exchangeability and boundedness.

Keywords: probability inequality, p-box, exchangeability, Cantelli, Chebyshev

1. Introduction

Probability boxes (or p-boxes) [3, 4] provide a well known method for modelling uncertainty where the probability distribution cannot be fully identified. They do so by bounding probabilities through a lower and upper cumulative distribution function. More precisely, consider a random variable X, and let $\mathscr{F}$ denote the set of all cumulative distribution functions on $\mathbb{R}$. Instead of using a single cumulative distribution function to model our information about X, a p-box specifies two cumulative distribution functions, $\underline{F}$ and $\overline{F}$, and then considers the set of all cumulative distribution functions that lie in between these two functions:

$$\{F \in \mathscr{F} : (\forall x \in \mathbb{R})(\underline{F}(x) \leq F(x) \leq \overline{F}(x)\}. \tag{1}$$

Combining p-boxes, under a variety of assumptions and binary operations, can be done extremely efficiently, as was demonstrated by Williamson and Downs [10]. The work of Destercke et al. [2], Troffaes and Destercke [9] generalized p-boxes to arbitrary pre-ordered spaces. Montes and Miranda [7] specifically introduced bivariate p-boxes.

In case we only know the mean μ and variance σ^2 of a random variable X, but we have no further information, a p-box can be constructed [3] using Cantelli's inequality [1, 5]:

$$0 \leq P\left(\tfrac{X-\mu}{\sigma} \leq \lambda\right) \leq \frac{1}{1+\lambda^2} \qquad \text{if } \lambda \leq 0, \tag{2a}$$

$$\frac{\lambda^2}{1+\lambda^2} \leq P\left(\tfrac{X-\mu}{\sigma} \leq \lambda\right) \leq 1 \qquad \text{if } \lambda \geq 0, \tag{2b}$$

or, equivalently,

$$0 \leq P(X \leq x) \leq \tfrac{\sigma^2}{\sigma^2+(x-\mu)^2} \qquad \text{if } x \leq \mu, \tag{3a}$$

$$\tfrac{(x-\mu)^2}{\sigma^2+(x-\mu)^2} \leq P(X \leq x) \leq 1 \qquad \text{if } x \geq \mu. \tag{3b}$$

Often, we may not know the mean μ and variance σ^2. In many situations however, we may have access to a sample mean $\overline{X}$ and a sample standard deviation S^2. Saw et al. [8] derived a version of Chebyshev's inequality (which is closely related to Cantelli's inequality) based on the sample mean and sample variance, assuming just exchangeability. On an abstract level, Saw et al. [8] identified a function f such that

$$P\left(\frac{|X_{n+1}-\overline{X}|}{S} \geq \lambda\right) \leq f(\lambda, n) \tag{4}$$

where $\overline{X}$ and S are the sample mean and sample standard deviation of $X_1, \dots, X_n$. The only assumption made is that $X_1, \dots, X_n, X_{n+1}$ are exchangeable.

The goal of this paper is to prove a Cantelli-like version of this inequality. In particular, we will find functions f_* and f^* such that

$$f_*(\lambda, n) \leq P\left(\frac{X_{n+1}-\overline{X}}{S+\varepsilon_n} \leq \lambda\right) \leq f^*(\lambda, n) \tag{5}$$

where $\varepsilon_n > 0$ goes to zero as n goes to infinity (obviously we wish we could avoid this ε_n but unfortunately we have not found a way to do so). To identify these functions, we closely follow the work of Saw et al. [8]. This allows us to construct a p-box purely based on boundedness and exchangeability, with no further assumptions on the form of the distribution.

Currently, unfortunately, we have not identified better uses for this p-box besides the unlikely case where one would be directly interested in the random variable

$$\frac{X_{n+1}-\overline{X}}{S+\varepsilon_n}, \tag{6}$$

or where one would use it to produce a random one-sided prediction interval based on the sample mean and sample standard deviation only. We will explain and discuss these issues at the end of the paper.

In Section 2, we introduce basic definitions and notation. In Section 3, we provide a proof of the inequality by Saw et al. [8]. In Section 4, we prove our inequality, which is a one-sided version of the inequality of Saw et al. [8]. Section 5 concludes the paper with a brief discussion.

2. Definitions and Notation

For any set A, $\#\{A\}$ denotes the number of elements of A.

For any $x \in \mathbb{R}$, $\lfloor x \rfloor$ denotes the largest integer that is less than or equal to x:

$$\lfloor x \rfloor := \max\{n \in \mathbb{Z} \colon n \leq x\} \tag{7}$$

and $\lceil x \rceil$ denotes the smallest integer that is greater than or equal to x:

$$\lceil x \rceil := \min\{n \in \mathbb{Z} \colon n \geq x\} \tag{8}$$

Similarly, $\lfloor x \rfloor_*$ denotes the largest integer that is strictly less than x:

$$\lfloor x \rfloor_* := \max\{n \in \mathbb{Z} \colon n < x\}. \tag{9}$$

For any $n \in \mathbb{N}$, $n \geq 1$, let Σ_n denote the set of all permutations of $\{1, 2, \ldots, n\}$.

For convenience, all random variables considered will be assumed to be discrete, and will be assumed to take only a finite number of values. For brevity, we will not repeat this latter finite assumption each time, and we will simply call them *discrete random variables*. The results can be generalised to more general random variables.

Definition 1 (Exchangeability) *A finite sequence X_1, $X_2, \ldots, X_n$ of discrete random variables is said to be* exchangeable *if, for all $\sigma \in \Sigma_n$ and all $x_1, \ldots, x_n \in \mathbb{R}$,*

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_{\sigma(1)}, \ldots, X_n = x_{\sigma(n)}). \tag{10}$$

For example, independent and identically distributed random variables are exchangeable. However, exchangeability is weaker.

For any $n \in \mathbb{N}$ and $V \subseteq \mathbb{R}^n$, let $V_{\leq}$ denote the set of all ordered vectors in V, i.e.

$$V_{\leq} := \{v \in V \colon v_1 \leq v_2 \leq \cdots \leq v_n\}. \tag{11}$$

3. Proof of the Inequality of Saw et al.

Before we prove our inequality, we provide our own proof of the inequality of Saw et al. [8], as the proof in [8] lacks many tricky details. This also allows the reader to fully compare both proofs. Note that, for ease of presentation and comparison to our inequality, here we only provide a slightly weaker bound (also provided in [8, Eq. (2.5)]. For the original (and rather complicated) bound we refer to [8].

Theorem 2 *Let $X_1, \ldots, X_n$, X_{n+1} be a sequence of discrete exchangeable random variables. Define $\overline{X} := \frac{1}{n}\sum_{j=1}^n X_j$, $S^2 := \frac{1}{n-1}\sum_{j=1}^n (X_j - \overline{X})^2$ and $Q^2 := \frac{n+1}{n} S^2$. Then for every $\lambda \geq 1$*

$$P\left(\left|X_{n+1} - \overline{X}\right| > \lambda Q\right) \leq \frac{1}{n+1}\left\lfloor \frac{n+1}{\lambda_n^2} \right\rfloor_* \tag{12}$$

where $\lambda_n := \sqrt{\frac{n\lambda^2}{(n-1+\lambda^2)}}$.

Note that $\lim_{n\to\infty} \lambda_n = \lambda$, and corresponds to what Saw et al. [8] define as k.

Proof We define,

$$\overline{X}_* := \frac{1}{n+1}\sum_{j=1}^{n+1} X_j \tag{13}$$

and

$$L^2 := \frac{1}{n+1}\sum_{j=1}^{n+1} (X_j - \overline{X}_*)^2. \tag{14}$$

Then we have,

$$P\left(\left|X_{n+1} - \overline{X}\right| > \lambda Q\right) \tag{15}$$

$$= P\left((X_{n+1} - \overline{X})^2 > \lambda^2 Q^2\right) \tag{16}$$

and because $\lambda^2 = \frac{(n-1)\lambda_n^2}{n-\lambda_n^2}$,

$$= P\left((X_{n+1} - \overline{X})^2 > \frac{\lambda_n^2(n-1)}{(n-\lambda_n^2)} Q^2\right) \tag{17}$$

$$= P\left((X_{n+1} - \overline{X})^2 > \frac{\lambda_n^2(n^2-1)}{(n^2-\lambda_n^2 n)} S^2\right) \tag{18}$$

$$= P\left((n^2 - \lambda_n^2 n)(X_{n+1} - \overline{X})^2 > \lambda_n^2(n^2-1)S^2\right) \tag{19}$$

$$= P\left(n^2(X_{n+1} - \overline{X})^2 > \lambda_n^2(n^2-1)S^2 + \lambda_n^2 n(X_{n+1} - \overline{X})^2\right) \tag{20}$$

so by Eq. (58) in Lemma 5,

$$= P\left(n^2(X_{n+1} - \overline{X})^2 > \lambda_n^2(n+1)^2 L^2\right) \tag{21}$$

and now by Eq. (57) in Lemma 5,

$$= P\left((X_{n+1} - \overline{X}_*)^2 > \lambda_n^2 L^2\right) \tag{22}$$

$$= P\left(\left|\frac{X_{n+1} - \overline{X}_*}{L}\right| > \lambda_n\right). \tag{23}$$

Now, for each $j \in \{1, 2, \ldots, n+1\}$, define

$$U_j := (X_j - \overline{X}_*)/L. \tag{24}$$

Note that,

$$\sum_{j=1}^{n+1} U_j = 0, \qquad \sum_{j=1}^{n+1} U_j^2 = n+1. \tag{25}$$

Also, note that the $U_1, \ldots, U_{n+1}$ are exchangeable (but not independent!), by Lemma 4(i). We have so far shown that

$$P\left(\left|X_{n+1} - \overline{X}\right| > \lambda Q\right) = P\left(|U_{n+1}| > \lambda_n\right). \tag{26}$$

Now, let V be the finite set of values in $\mathbb{R}^{n+1}$ that the random variables $(U_1, \ldots, U_{n+1})$ can jointly take. Note that $\sum_{j=1}^{n+1} v_j = 0$ and $\sum_{j=1}^{n+1} v_j^2 = 1$ for every $v \in V$, by Eq. (25).

For every $v \in V_{\leq}$, define

$$A(v) := \bigcup_{\sigma \in \Sigma_{n+1}} \left\{U_1 = v_{\sigma(1)}, \ldots, U_{n+1} = v_{\sigma(n+1)}\right\}. \tag{27}$$

Then, by Lemma 4(ii), we know that $\{A(v)\colon v \in V_{\le}\}$ is a partition. We now apply the partition theorem to the right-hand side of Eq. (26):

$$P\left(\left|X_{n+1}-\overline{X}\right| > \lambda Q\right) \tag{28}$$
$$= \sum_{v\in V_{\le}} P(|U_{n+1}| > \lambda_n \mid A(v))P(A(v)) \tag{29}$$

and so, by Lemma 4(i),

$$= \sum_{v\in V_{\le}} \sum_{j=1}^{n+1} \frac{\mathbf{I}_{|v_j|>\lambda_n}}{n+1} P(A(v)) \tag{30}$$
$$= \sum_{v\in V_{\le}} \frac{\#\{j : |v_j| > \lambda_n\}}{n+1} P(A(v)) \tag{31}$$

and so, finally, by Lemma 3,

$$\le \frac{1}{n+1} \left\lfloor \frac{n+1}{\lambda_n^2} \right\rfloor_* \sum_{v\in V_{\le}} P(A(v)) \tag{32}$$
$$= \frac{1}{n+1} \left\lfloor \frac{n+1}{\lambda_n^2} \right\rfloor_* \tag{33}$$

where once more we used the fact that the $A(v)$ form a partition. ■

Lemma 3 *For every $m \in \mathbb{N}$, every $u \in \mathbb{R}^m$ such that $\sum_{i=1}^m u_i = 0$ and $\sum_{i=1}^n u_i^2 = m$, and every $k \ge 1$, we have that*

$$\#\{j\colon |u_j| > k\} \le \left\lfloor \frac{m}{k^2} \right\rfloor_* . \tag{34}$$

Proof Immediate, from [8, Lemma 2]. Note that the bound in [8, Lemma 2] is tighter than Eq. (34). Here, we consider the simplified bound for ease of presentation; this simplified bound is also mentioned in [8, Eq. (2.5)]. ■

Lemma 4 *Let $m \in \mathbb{N}$, let $U_1, U_2, \dots, U_m$ be any finite sequence of discrete exchangeable random variables. Let V denote the finite set of values in $\mathbb{R}^m$ that the $(U_1,\dots,U_m)$ can jointly take, and for every $v \in V_{\le}$, we define*

$$A(v) := \bigcup_{\sigma\in\Sigma_m} \left\{U_1 = v_{\sigma(1)},\dots,U_m = v_{\sigma(m)}\right\}. \tag{35}$$

Then, the following statements hold.

(i) For every $v \in V_{\le}$, and every function f, we have that

$$E(f(U_m) \mid A(v)) = \frac{1}{m} \sum_{j=1}^m f(v_j). \tag{36}$$

(ii) $\{A(v)\colon v \in V_{\le}\}$ is a partition.

Note that a proof of the first statement can also be found in [6, Eq. (2.1)].

Proof (i). First note that

$$\sum_{u\in A(v)} f(u_m)P(U = u) \tag{37}$$
$$= \sum_{u\in A(v)} f(u_m)P(U_1 = u_1,\dots,U_j = u_j,\dots,U_m = u_m) \tag{38}$$

and now, by exchangeability (see Theorem 1),

$$= \sum_{u\in A(v)} f(u_m)P(U_1 = u_1,\dots,U_j = u_m,\dots,U_m = u_j) \tag{39}$$

and now, swapping u_m and u_j,

$$= \sum_{u\in A(v)} f(u_j)P(U_1 = u_1,\dots,U_j = u_j,\dots,U_m = u_m) \tag{40}$$
$$= \sum_{u\in A(v)} f(u_j)P(U = u) \tag{41}$$

Therefore,

$$E(f(U_m) \mid A(v)) \tag{42}$$
$$= \sum_{u\in V} f(u_m)P(U = u \mid A(v)) \tag{43}$$
$$= \frac{1}{P(A(v))} \sum_{u\in V} f(u_m)P(\{U = u\} \cap A(v)) \tag{44}$$

and because $\{U = u\} \cap A(v) = \{U = u\}$ whenever $u \in A(v)$, and $\emptyset$ otherwise,

$$= \frac{1}{P(A(v))} \sum_{u\in A(v)} f(u_m)P(U = u) \tag{45}$$

and now because of what we proved earlier,

$$= \frac{1}{P(A(v))} \sum_{u\in A(v)} f(u_j)P(U = u) \tag{46}$$
$$= E(f(U_j) \mid A(v)) \tag{47}$$

Taking the sum over all j of the above equality, and using linearity of expectation, we find that

$$mE(f(U_m) \mid A(v)) = E\left(\sum_{j=1}^m f(U_j) \mid A(v)\right) \tag{48}$$

and because if we know that $A(v)$ has obtained, we know the values of the U_j but not their order, so consequently we do know the value of the sum, and thereby,

$$= \sum_{j=1}^m f(v_j). \tag{49}$$

This establishes the desired equality.

(ii). We have to show that

$$\bigcup_{v\in V_\leq} A(v) = \Omega \tag{50}$$

and that for all v and $u \in V_\leq$ such that $v \neq u$, we have that

$$A(v) \cap A(u) = \emptyset. \tag{51}$$

Consider any $\omega \in \Omega$, Then there is a $u \in V$ such that $\omega \in \{U = u\}$. But obviously, $\{U = u\} \subseteq A(v)$ where v is u but with components rearranged in order. Consequently, for every ω, there is a $v \in V_\leq$ such that $\omega \in A(v)$. Therefore,

$$\bigcup_{v\in V_\leq} A(v) \supseteq \Omega. \tag{52}$$

Clearly, also $A(v) \subseteq \Omega$ for all $v \in V_\leq$, so both sides must be equal.

We will prove the final part by contraposition. Consider any v and $u \in V_\leq$ such that

$$A(v) \cap A(u) \neq \emptyset. \tag{53}$$

Then, there is at least one $w \in V$ such that

$$\{U = w\} \subseteq A(v) \cap A(u). \tag{54}$$

But this implies that both v and u are equal to w with components rearranged in order. Consequently, it must hold that $v = u$. ■

Lemma 5 *Let X_1, X_2, ..., X_n, X_{n+1} be any sequence of random variables. Define*

$$\overline{X} := \frac{1}{n}\sum_{j=1}^{n} X_j \qquad S^2 := \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2 \tag{55}$$

and

$$\overline{X}_* := \frac{1}{n+1}\sum_{j=1}^{n+1} X_j \quad L^2 := \frac{1}{n+1}\sum_{j=1}^{n+1}(X_j - \overline{X}_*)^2. \tag{56}$$

Then,

$$X_{n+1} - \overline{X}_* = \frac{n}{n+1}(X_{n+1} - \overline{X}) \tag{57}$$

$$(n+1)^2 L^2 = (n^2-1)S^2 + n(X_{n+1} - \overline{X})^2 \tag{58}$$

and

$$L - \sqrt{\frac{n-1}{n+1}}S \leq \frac{|X_{n+1} - \overline{X}_*|}{\sqrt{n}}. \tag{59}$$

Equations (57) and (58) are also stated in [8], and are used to prove Theorem 2. Equation (59) is a new result which we will use to prove Theorem 6 further.

Proof By definition of $\overline{X}_*$,

$$(n+1)(X_{n+1} - \overline{X}_*) = (n+1)X_{n+1} - \sum_{j=1}^{n+1} X_j \tag{60}$$

$$= nX_{n+1} - \sum_{j=1}^{n} X_j \tag{61}$$

and so, by definition of $\overline{X}$,

$$= n(X_{n+1} - \overline{X}). \tag{62}$$

This proves Eq. (57).

To prove Eq. (58), we will repeatedly use that, for any sequence $a_1, \ldots, a_m$,

$$\sum_{j=1}^{m}(a_j - a)^2 = \left(\sum_{j=1}^{m} a_j^2\right) - ma^2 \tag{63}$$

where $a := \frac{1}{m}\sum_{j=1}^{m} a_j$.

Starting at the right-hand side of Eq. (58), we get

$$(n^2-1)S^2 + n(X_{n+1} - \overline{X})^2 \tag{64}$$

$$= (n+1)\left(\sum_{j=1}^{n} X_j^2 - n\overline{X}^2\right)$$

$$+ nX_{n+1}^2 - 2nX_{n+1}\overline{X} + n\overline{X}^2 \tag{65}$$

where we applied the definition of S as well as Eq. (63), and we also expanded the square. So, expanding all terms, we see that the term $n\overline{X}^2$ cancels out, to obtain

$$= (n+1)\sum_{j=1}^{n} X_j^2 - n^2\overline{X}^2 + nX_{n+1}^2 - 2nX_{n+1}\overline{X} \tag{66}$$

so, after rearranging the terms,

$$= (n+1)\sum_{j=1}^{n} X_j^2 + nX_{n+1}^2 - 2nX_{n+1}\overline{X} - n^2\overline{X}^2 \tag{67}$$

and now, absorbing nX_{n+1}^2 into the sum,

$$= (n+1)\sum_{j=1}^{n+1} X_j^2 - X_{n+1}^2 - 2nX_{n+1}\overline{X} - n^2\overline{X}^2 \tag{68}$$

$$= (n+1)\sum_{j=1}^{n+1} X_j^2 - (X_{n+1} + n\overline{X})^2 \tag{69}$$

and now, after noting that $X_{n+1} + n\overline{X} = (n+1)\overline{X}_*$,

$$= (n+1)\sum_{j=1}^{n+1} X_j^2 - (n+1)^2\overline{X}_*^2 \tag{70}$$

$$= (n+1)\left(\sum_{j=1}^{n+1} X_j^2 - (n+1)\overline{X}_*^2\right) \tag{71}$$

and now, applying the definition of L as well as Eq. (63),

$$= (n+1)^2 L^2. \tag{72}$$

This proves Eq. (58).

For any non-negative A and B, the following inequality holds:

$$\sqrt{A} + \sqrt{B} \geq \sqrt{A+B}. \tag{73}$$

Taking $A = (n^2-1)S^2$ and $B = n(X_{n+1} - \overline{X})^2$, we get, after applying Eq. (58),

$$\sqrt{n^2-1}S + \sqrt{n}|X_{n+1} - \overline{X}| \geq (n+1)L. \tag{74}$$

Consequently, after dividing both sides by $n+1$ and rearranging terms,

$$L - \sqrt{\frac{n-1}{n+1}}S \leq \frac{\sqrt{n}}{n+1}|X_{n+1} - \overline{X}| \tag{75}$$

$$= \frac{1}{\sqrt{n}} \frac{n}{n+1}|X_{n+1} - \overline{X}| \tag{76}$$

and so, by Eq. (57),

$$= \frac{1}{\sqrt{n}}|X_{n+1} - \overline{X}_*|. \tag{77}$$

This proves Eq. (59). ■

4. Main Result

Theorem 6 *Let, $X_1, \ldots, X_n$, X_{n+1} be a finite sequence of discrete exchangeable random variables. Let $\Delta \in \mathbb{R}$ denote the range of the X_j i.e. $\Delta := \max X_j - \min X_j$ where $\max X_j$ is the maximum value that can be attained by X_j, and $\min X_j$ is the minimum value. Let $\overline{X} := \sum_{j=1}^n X_j/n$, and $S^2 := \sum_{j=1}^n (X_j - \overline{X})^2/(n-1)$. Then for every $\lambda \geq 0$,*

$$\frac{1}{n+1}\left\lceil \frac{(n+1)\lambda_n^2}{\lambda_n^2+1} \right\rceil \leq P\left(\frac{X_{n+1} - \overline{X}}{S + \frac{\Delta_n}{\sqrt{n}}} < \lambda\right) \leq 1 \tag{78}$$

where $\lambda_n := \frac{n}{\sqrt{n^2-1}}\lambda$ and $\Delta_n := \sqrt{\frac{n+1}{n-1}}\Delta$. Similarly, for $\lambda \leq 0$,

$$0 \leq P\left(\frac{X_{n+1} - \overline{X}}{S + \frac{\Delta_n}{\sqrt{n}}} \leq \lambda\right) \leq \frac{1}{n+1}\left\lfloor \frac{n+1}{\lambda_n^2+1} \right\rfloor. \tag{79}$$

Note that $\lim_{n\to\infty} \lambda_n = \lambda$, and $\lim_{n\to\infty} \Delta_n = \Delta$.

Proof We define $\overline{X}_*$ and L^2 as before (see Eqs. (13) and (14)).

First, recall Eq. (59):

$$L - \sqrt{\frac{n-1}{n+1}}S \leq \frac{|X_{n+1} - \overline{X}_*|}{\sqrt{n}}. \tag{80}$$

By definition of Δ, it follows also that $|X_{n+1} - \overline{X}_*| \leq \Delta$. So,

$$L \leq \sqrt{\frac{n-1}{n+1}}S + \frac{\Delta}{\sqrt{n}} = \sqrt{\frac{n-1}{n+1}}\left(S + \frac{\Delta_n}{\sqrt{n}}\right) \tag{81}$$

and consequently,

$$\sqrt{\frac{n+1}{n-1}}L \leq S + \frac{\Delta_n}{\sqrt{n}}. \tag{82}$$

We can now start with proving the desired inequality. First, note that, by Eq. (82),

$$P\left(X_{n+1} - \overline{X} < \lambda\left(S + \tfrac{\Delta_n}{\sqrt{n}}\right)\right) \tag{83}$$

$$\geq P\left(X_{n+1} - \overline{X} < \lambda L\sqrt{\frac{n+1}{n-1}}\right) \tag{84}$$

and now applying Eq. (57) and moving L to the other side of the inequality,

$$= P\left(\frac{n+1}{n}\frac{X_{n+1} - \overline{X}_*}{L} < \lambda\sqrt{\frac{n+1}{n-1}}\right) \tag{85}$$

$$= P\left(\frac{X_{n+1} - \overline{X}_*}{L} < \frac{n\lambda}{\sqrt{n^2-1}}\right) \tag{86}$$

and, with $U_j := (X_j - \overline{X}_*)/L$ and $\lambda_n := \frac{n\lambda}{\sqrt{n^2-1}}$,

$$= P(U_{n+1} < \lambda_n) \tag{87}$$

and, as before, applying the partition theorem,

$$= \sum_{v\in V_\leq} P(U_{n+1} < \lambda_n \mid A(v))P(A(v)) \tag{88}$$

and so, by Lemma 4(i),

$$= \sum_{v\in V_\leq}\sum_{j=1}^{n+1}\frac{\mathbf{I}_{v_j<\lambda_n}}{n+1}P(A(v)) \tag{89}$$

$$= \sum_{v\in V_\leq}\frac{\#\{j\colon v_j < \lambda_n\}}{n+1}P(A(v)) \tag{90}$$

and so, finally, by Lemma 7,

$$\geq \frac{1}{n+1}\left\lceil \frac{(n+1)\lambda_n^2}{\lambda_n^2+1} \right\rceil \sum_{v\in V_\leq} P(A(v)) \tag{91}$$

$$= \frac{1}{n+1}\left\lceil \frac{(n+1)\lambda_n^2}{\lambda_n^2+1} \right\rceil. \tag{92}$$

For the other part when $\lambda \leq 0$, note that

$$P\left(X_{n+1} - \overline{X} \leq \lambda\left(S + \tfrac{\Delta_n}{\sqrt{n}}\right)\right) \tag{93}$$

$$= 1 - P\left(X_{n+1} - \overline{X} > \lambda\left(S + \tfrac{\Delta_n}{\sqrt{n}}\right)\right) \tag{94}$$

$$= 1 - P\left(-X_{n+1} + \overline{X} < -\lambda\left(S + \tfrac{\Delta_n}{\sqrt{n}}\right)\right) \tag{95}$$

and therefore, now applying the previous result on the random variables $-X_1, \ldots, -X_{n+1}$,

$$\leq 1 - \frac{1}{n+1}\left\lceil \frac{(n+1)\lambda_n^2}{\lambda_n^2+1} \right\rceil \tag{96}$$

$$= \frac{1}{n+1}\left(n+1-\left\lceil \frac{(n+1)\lambda_n^2}{\lambda_n^2+1}\right\rceil\right) \tag{97}$$

$$= \frac{1}{n+1}\left(n+1+\left\lfloor \frac{-(n+1)\lambda_n^2}{\lambda_n^2+1}\right\rfloor\right) \tag{98}$$

$$= \frac{1}{n+1}\left(\left\lfloor n+1+\frac{-(n+1)\lambda_n^2}{\lambda_n^2+1}\right\rfloor\right) \tag{99}$$

$$= \frac{1}{n+1}\left\lfloor \frac{(n+1)}{\lambda_n^2+1}\right\rfloor. \tag{100}$$

■

Lemma 7 *For every $m \in \mathbb{N}$, every $u \in \mathbb{R}^m$ such that $\sum_{i=1}^m u_i = 0$ and $\sum_{i=1}^n u_i^2 = m$, and every $k \geq 0$, we have that*

$$\#\{j \colon u_j < k\} \geq \left\lceil \frac{mk^2}{k^2+1} \right\rceil. \tag{101}$$

Proof The condition holds trivially for $k = 0$, so without loss of generality, we can assume that $k > 0$.

Consider any $u \in \mathbb{R}^m$ satisfying the conditions of the lemma. Let $J := \{j \colon u_j < k\}$. Note that J must be non-empty, because at least one u_j must be strictly negative, because the u_j sum to zero, and they cannot all be zero because their squares sum to m. Let $p := \#J$. Because J is non-empty, we know that $p \geq 1$. We also use the notation $J^c = \{1, \ldots, m\} \setminus J$.

Since $p \in \mathbb{N}$, it suffices to show that

$$p \geq \frac{mk^2}{k^2+1}. \tag{102}$$

First, by Jensen's inequality, we have that

$$\frac{1}{p}\sum_{j\in J} u_j^2 \geq \left(\frac{1}{p}\sum_{j\in J} u_j\right)^2 = \frac{1}{p^2}\left(\sum_{j\in J} u_j\right)^2 \tag{103}$$

and because $\sum_{j=1}^m u_j = 0$,

$$= \frac{1}{p^2}\left(\sum_{j\in J^c} u_j\right)^2 \geq \frac{1}{p^2}\left(\sum_{j\in J^c} k\right)^2 \tag{104}$$

$$= \frac{1}{p^2}(m-p)^2 k \tag{105}$$

Also,

$$\sum_{i\in J^c} u_i^2 \geq \sum_{i\in J^c} k^2 = (m-p)k^2. \tag{106}$$

Therefore, combining Eqs. (105) and (106), and using the fact that $\sum_{j=1}^m u_j^2 = m$,

$$m = p\left(\frac{1}{p}\sum_{j\in J} u_j^2\right) + \sum_{i\in J^c} u_i^2 \tag{107}$$

$$\geq p\frac{1}{p^2}(m-p)^2k^2 + (m-p)k^2 \tag{108}$$

$$= \left(\frac{m-p}{p}+1\right)(m-p)k^2 \tag{109}$$

$$= \frac{m(m-p)k^2}{p}. \tag{110}$$

Therefore, after rearranging the terms, we find that

$$p \geq (m-p)k^2 \tag{111}$$

which implies

$$p(1+k^2) \geq mk^2 \tag{112}$$

which then evidently proves Eq. (102). ■

5. Discussion and Conclusion

As noted in the introduction, Cantelli's inequality can be used to construct a p-box [3, 4] on a random variable, provided we know its expectation and variance. Our main result, Theorem 6, allows us to construct a p-box directly on the quantity

$$Z_{n+1} := \frac{X_{n+1} - \overline{X}}{S + \frac{\Delta_n}{\sqrt{n}}}. \tag{113}$$

Figure 1 shows how the p-box bounds given by Theorem 6 compare to the bounds from Cantelli's inequality given by Eq. (2). We can see that with higher number of samples, our bounds converge to Cantelli's bounds, as expected.

A critical difference between our bounds and the bounds by Saw et al. [8] is that we need to add a constant offset of $\frac{\Delta_n}{\sqrt{n}}$ to the sample standard deviation. Fortunately, this offset converges to zero as n goes to infinity. So far, we have not found a way to avoid this offset.

Note that the bounds are achieved from the inside, i.e. we have tighter p-boxes for smaller sample sizes. This seems counter-intuititive at first. The reason why there is no contradiction here is the correction term in the denominator $\frac{\Delta_n}{\sqrt{n}}$ is larger for smaller sample sizes. For example, for $n = 4$ this correction is approximately $0.645 \times \Delta$, whereas for $n = 256$ it is only $0.063 \times \Delta$ (where Δ denotes the range of X_n).

We do note that this p-box *cannot* be turned into a p-box directly on X_{n+1}, that is, one cannot simply substitute observed values for $\overline{X}$ and S into this equation, as our p-box

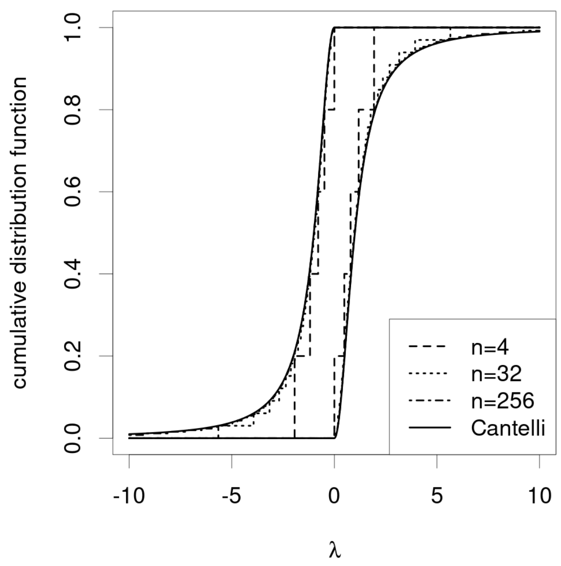

Figure 1: Comparison of our bounds with Cantelli's bounds.

is fully unconditional, and does not condition on $\overline{X}$ and S. Finding bounds on the cumulative distribution function of X_{n+1}, conditional on $\overline{X}$ and S, using exchangeability only, remains an open problem.

However, we still can use Theorem 6 to construct asymmetric prediction intervals. From Eqs. (78) and (79) for any $\ell_1 < \ell_2$, we can find values $\underline{p}_1$, $\overline{p}_1$, $\underline{p}_2$, and $\overline{p}_2$ such that

$$\underline{p}_1 \leq P(Z_{n+1} \leq \ell_1) \leq \overline{p}_1, \tag{114}$$

$$\underline{p}_2 \leq P(Z_{n+1} \leq \ell_2) \leq \overline{p}_2. \tag{115}$$

So, because

$$P(\ell_1 < Z_{n+1} \leq \ell_2) = P(Z_{n+1} \leq \ell_2) - P(Z_{n+1} \leq \ell_1) \tag{116}$$

we find the following bounds:

$$\underline{p}_2 - \overline{p}_1 \leq P(\overline{X} + \ell_1 S_n < X_{n+1} \leq \overline{X} + \ell_2 S_n) \leq \overline{p}_2 - \underline{p}_1 \tag{117}$$

where $S_n := S + \frac{\Delta_n}{\sqrt{n}}$. This gives us a prediction interval on X_{n+1} with an imprecise coverage probability.

Typically, we would be interested in $\ell_1 < 0$ and $\ell_2 > 0$. In that case $\underline{p}_1 = 0$ and $\overline{p}_2 = 1$, so we obtain:

$$\underline{p}_2 - \overline{p}_1 \leq P(\overline{X} + \ell_1 S_n < X_{n+1} \leq \overline{X} + \ell_2 S_n) \leq 1 \tag{118}$$

so we obtain a prediction interval with guaranteed minimal coverage probability of $\underline{p}_2 - \overline{p}_1$.

If $\ell_2 = -\ell_1$, because of symmetry,

$$\underline{p}_2 = 1 - \overline{p}_1 \tag{119}$$

and we obtain

$$2\underline{p}_2 - 1 \leq P\left(\overline{X} - \ell_2 S_n < X_{n+1} \leq \overline{X} + \ell_2 S_n\right). \tag{120}$$

In this case, however, Saw et al. [8] (i.e. Eq. (12)) provides a tighter bound. This is similar to how the classical Cantelli inequality compares to the classical Chebyshev inequality.

Acknowledgments

We thank Scott Ferson for pointing us to the paper by Saw et al. [8] and for suggesting that it might be used for constructing p-boxes based on exchangeability only. We also thank Jochen Einbeck for his valuable inputs during discussions of this work. Finally, we thank all reviewers for carefully considering and commenting on our paper.

This work is funded by the European Commission's H2020 programme, through the UTOPIAE Marie Curie Innovative Training Network, H2020-MSCA-ITN-2016, Grant Agreement number 722734.

References

[1] F. P. Cantelli. Sui confini della probabilità. *Atti del Congresso Internazional del Matematici*, 6:47–59, 1928. Bologna.

[2] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations: I. Generalized p-boxes. *International Journal of Approximate Reasoning*, 49(3):649–663, 2008. doi:10.1016/j.ijar.2008.07.003.

[3] Scott Ferson, Vladik Kreinovich, Lev Ginzburg, Davis S. Myers, and Kari Sentz. Constructing probability boxes and Dempster-Shafer structures. Technical Report SAND2002–4015, Sandia National Laboratories, January 2003. Unabridged version.

[4] Scott Ferson, Roger B. Nelsen, Janos Hajagos, Daniel Berleant, Jianzhong Zhang, W. Troy Tucker, Lev R. Ginzburg, and William Oberkampf. Dependence in probabilistic modeling, Dempster-Shafer theory, and probability bounds analysis. Technical Report SAND2004–3072, Sandia National Laboratories, October 2004.

[5] B. K. Ghosh. Probability inequalities related to Markov's theorem. *The American Statistician*, 56(3): 186–190, 2002. URL http://www.jstor.org/stable/3087296.

[6] J. F. C. Kingman. Uses of exchangeability. *The Annals of Probability*, 6(2):183–197, April 1978. doi:10.1214/aop/1176995566.

[7] Ignacio Montes and Enrique Miranda. Bivariate p-boxes and maxitive functions. *International Journal of General Systems*, 46(4):354–385, 2017. doi:10.1080/03081079.2017.1305960.

[8] John G. Saw, Mark C. K. Yang, and Tse Chin Mo. Chebyshev inequality with estimated mean and variance. *The American Statistician*, 38(2): 130–132, 1984. URL http://www.jstor.org/stable/2683249.

[9] Matthias C. M. Troffaes and Sébastien Destercke. Probability boxes on totally preordered spaces for multivariate modelling. *International Journal of Approximate Reasoning*, 52(6):767–791, 2011. doi:10.1016/j.ijar.2011.02.001.

[10] Robert C. Williamson and Tom Downs. Probabilistic arithmetic. I. Numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning*, 4(2):89–158, 1990. doi:10.1016/0888-613X(90)90022-T.