Statistics
A Journal of Theoretical and Applied Statistics

ISSN: 0233-1888 (Print) 1029-4910 (Online) Journal homepage: http://www.tandfonline.com/loi/gsta20

# Maximizing the expected range from dependent observations under mean–variance information

Nickos Papadatos

**To cite this article:** Nickos Papadatos (2016) Maximizing the expected range from dependent observations under mean–variance information, Statistics, 50:3, 596-629, DOI: 10.1080/02331888.2015.1074234

**To link to this article:** http://dx.doi.org/10.1080/02331888.2015.1074234

Published online: 03 Nov 2015.

*Statistics*, 2016
Vol. 50, No. 3, 596–629, http://dx.doi.org/10.1080/02331888.2015.1074234

Taylor & Francis
Taylor & Francis Group



# Maximizing the expected range from dependent observations under mean–variance information

Nickos Papadatos*

*Department of Mathematics, Section of Statistics and O.R., University of Athens, Panepistemiopolis, 157 84 Athens, Greece*

(*Received 24 June 2014; accepted 9 July 2015*)

In this article we derive the best possible upper bound for $\mathbb{E}[\max_i\{X_i\} - \min_i\{X_i\}]$ under given means and variances on $n$ random variables $X_i$. The random vector $(X_1, \ldots, X_n)$ is allowed to have any dependence structure, provided $\mathbb{E}X_i = \mu_i$ and $\operatorname{Var} X_i = \sigma_i^2$, $0 < \sigma_i < \infty$. We provide an explicit characterization of the $n$-variate distributions that attain the equality (extremal random vectors), and the tight bound is compared to other existing results.

**Keywords:** range; dependent observations; tight expectation bounds; extremal random vectors; probability matrices; characterizations

*Mathematics Subject Classification*: 62G30; 60E15; 62E10

## 1. Introduction

The problem of determining best possible expectation bounds on linear functions of order statistics in terms of means and variances of the observations has a long history. Especially for the sample range based on $n \geq 2$ independent identically distributed (i.i.d.) random variables, the problem goes back to Plackett,[1] Gumbel [2] and Hartley and David [3] who derived the inequality

$$\mathbb{E}\left[\max_{1\le i\le n}\{X_i\} - \min_{1\le i\le n}\{X_i\}\right] \le n\sigma\sqrt{\frac{2}{2n-1}\left(1-\frac{1}{\binom{2n-2}{n-1}}\right)}, \tag{1}$$

where $\sigma^2$ is the common variance of $X_i$. This bound is best possible in the sense that for any given values of $\mu \in \mathbb{R}$ and $\sigma \in (0,\infty)$ there exist $n$ i.i.d. random variables with mean $\mu$ and variance $\sigma^2$ that attain the equality in Inequality (1).

Since then, a lot of research has been developed in order to drop the assumptions of independence and/or identical distributions on the observations, and also to extend the results to any $L$-statistic of the form $L = \sum_{i=1}^n c_i X_{i:n}$, where $c_i$ are given constants and $X_{1:n} \le \cdots \le X_{n:n}$ are the order statistics corresponding to the random vector $(X_1, \ldots, X_n)$. When the components $X_i$ are merely assumed to be i.d. (identically distributed but not necessarily independent) with mean $\mu$ and variance $\sigma^2$, the best possible bounds for $\mathbb{E}L$ were established by Rychlik.[4] In particular,

*Email: npapadat@math.uoa.gr

setting $c_1 = -1$, $c_n = 1$ and $c_i = 0$ for any other $i$ in Rychlik's result, we get the optimal upper bound for the expected range:

$$\mathbb{E}[X_{n:n} - X_{1:n}] \le \sigma\sqrt{2n}. \tag{2}$$

For a comprehensive review of related results and extensions, the reader is referred to Rychlik's [5] monograph; see also [6–8]. Dropping both assumptions of independence and i.d. Arnold and Groeneveld [9] obtained the upper bound

$$\mathbb{E}\left(\sum_{i=1}^{n} c_i X_{i:n}\right) \le \bar{\mu}\sum_{i=1}^{n} c_i + \sqrt{\sum_{i=1}^{n}(c_i - \bar{c})^2}\sqrt{\sum_{i=1}^{n}\{(\mu_i - \bar{\mu})^2 + \sigma_i^2\}}, \tag{3}$$

which is valid for any random vector with $\mathbb{E}X_i = \mu_i$ and $\operatorname{Var} X_i = \sigma_i^2$, where $\bar{\mu} = (1/n)\sum_{i=1}^{n}\mu_i$, $\bar{c} = (1/n)\sum_{i=1}^{n} c_i$. For other inequalities related to Equation (3) the reader is referred to Nagaraja,[10] Aven,[11] Lefèvre,[12] Papadatos [13] and Kaluszka et al. [14]; see also the monograph by Arnold and Balakrishnan.[15] Applied to the range, Inequality (3) yields the inequality

$$\mathbb{E}[X_{n:n} - X_{1:n}] \le \mathrm{AG}_n := \sqrt{2\sum_{i=1}^{n}\{(\mu_i - \bar{\mu})^2 + \sigma_i^2\}}, \tag{4}$$

which, in the homogeneous case $\mu_i = \mu$, $\sigma_i^2 = \sigma^2$, reduces to Inequality (2). However, the upper bound in Equation (4) is not tight under general mean–variance information, and the purpose of the present work is to replace the RHS of Inequality (4) by its best possible value.

Recently, Bertsimas et al. [16,17] applied convex optimization techniques in order to replace the RHS of Inequality (3) by its tight counterpart in some particular cases of interest. They obtained, among other things, the best possible upper bound for the expected maximum under any mean–variance information and any dependence structure, namely

$$\mathbb{E}X_{n:n} \le \mathrm{BNT}_n := -\frac{n-2}{2}y_0 + \frac{1}{2}\sum_{i=1}^{n}\mu_i + \frac{1}{2}\sum_{i=1}^{n}\sqrt{(\mu_i - y_0)^2 + \sigma_i^2}, \tag{5}$$

where $y_0$ is the unique solution to the equation

$$\sum_{i=1}^{n}\frac{y_0 - \mu_i}{\sqrt{(\mu_i - y_0)^2 + \sigma_i^2}} = n - 2. \tag{6}$$

The equality in Inequality (5) is attained by the maximally dependent random vector with

$$\mathbb{P}[X_1 = y_0 - \alpha_1, \ldots, X_j = y_0 + \alpha_j, \ldots, X_n = y_0 - \alpha_n] = p_j, \quad j = 1, \ldots, n,$$

where

$$\alpha_j = \sqrt{(\mu_j - y_0)^2 + \sigma_j^2}, \quad p_j = \frac{1}{2}\left(1 - \frac{y_0 - \mu_j}{\sqrt{(\mu_j - y_0)^2 + \sigma_j^2}}\right), \quad j = 1, \ldots, n.$$

Note that $p_j > 0$ and, by Equation (6), $\sum_{j=1}^{n} p_j = 1$.

In the present work we extend the techniques of Lai and Robbins [18] and of Bertsimas et al.,[17] in order to obtain the best possible upper bound for the expected range. Also, we characterize the *extremal random vectors*, i.e. the vectors that attain the equality in the bound, and we provide simple conditions (on $\mu_i$ and $\sigma_i$) under which the $\mathrm{AG}_n$ bound of Inequality (4) is already sharp. The main result is given in Theorem 6.1. Particular cases of interest are presented as examples. All missing proofs are included in the appendix.

## 2. An upper bound for the expected range

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be an arbitrary random vector with $\mathbb{E}\boldsymbol{X} = \boldsymbol{\mu} := (\mu_1, \ldots, \mu_n)$ and $(\operatorname{Var} X_1, \ldots, \operatorname{Var} X_n) = (\sigma_1^2, \ldots, \sigma_n^2)$ where $0 < \sigma_i < \infty$ for all $i$. For notational simplicity we write $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_n^2)$ and $\operatorname{Var} \boldsymbol{X} = \boldsymbol{\sigma}^2$; that is, $\operatorname{Var} \boldsymbol{X} := \operatorname{diag}(\boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is the dispersion matrix of $\boldsymbol{X}$. The class of random vectors satisfying the above moment requirements will be denoted by

$$\mathcal{F}_n(\boldsymbol{\mu}, \boldsymbol{\sigma}) := \{\boldsymbol{X} : \mathbb{E}\boldsymbol{X} = \boldsymbol{\mu},\ \operatorname{Var} \boldsymbol{X} = \boldsymbol{\sigma}^2\}.$$

In particular, $X \in \mathcal{F}_1(\mu, \sigma)$ means that $\mathbb{E}X = \mu$ and $\operatorname{Var} X = \sigma^2$.

Let $X_{1:n} \le \cdots \le X_{n:n}$ be the order statistics corresponding to $\boldsymbol{X}$ and set $R_n = X_{n:n} - X_{1:n}$ for the range. Our main interest is in calculating

$$\inf_{X \in \mathcal{F}_n(\boldsymbol{\mu}, \boldsymbol{\sigma})} \mathbb{E}R_n, \qquad \sup_{X \in \mathcal{F}_n(\boldsymbol{\mu}, \boldsymbol{\sigma})} \mathbb{E}R_n, \tag{7}$$

for any given $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\boldsymbol{\sigma} \in \mathbb{R}_+^n$. However, the result is known for the infimum:

$$\inf_{X \in \mathcal{F}_n(\boldsymbol{\mu}, \boldsymbol{\sigma})} \mathbb{E}R_n = \max_i\{\mu_i\} - \min_i\{\mu_i\}.$$

Indeed, since $R_n = R_n(\boldsymbol{X})$ is a convex function of $\boldsymbol{X}$ we have $\mathbb{E}R_n(\boldsymbol{X}) \ge R_n(\boldsymbol{\mu}) = \max_i\{\mu_i\} - \min_i\{\mu_i\}$ from Jensen's inequality. Bertsimas et al. [19] showed that this lower bound is best possible even for the narrowed class of random vectors with given mean vector $\boldsymbol{\mu}$ and (any) given nonnegative defined dispersion matrix $\boldsymbol{\Sigma}$. For clarity of the presentation we provide here the construction of Bertsimas et al.[19] Define

$$\boldsymbol{X}_\epsilon = \boldsymbol{\mu} + \frac{I_\epsilon}{\sqrt{\epsilon}} \boldsymbol{V} \boldsymbol{\Sigma}^{1/2}, \quad 0 < \epsilon < 1,$$

where $\boldsymbol{V} = (V_1, \ldots, V_n)$ with $V_i$ being i.i.d. with zero mean and variance one and $I_\epsilon$ is a Bernoulli random variable, independent of $\boldsymbol{V}$, with probability of success equal to $\epsilon$. Then it is easy to verify that for all $\epsilon \in (0,1)$, $\boldsymbol{X}_\epsilon$ has mean $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Sigma}$. Let $A \subseteq \mathbb{R}^n$ be the finite collection of vectors of the form $\boldsymbol{e}(i) - \boldsymbol{e}(j)$, $i \ne j$, $i, j \in \{1, \ldots, n\}$, where $\boldsymbol{e}(i) = (0, \ldots, 1, \ldots, 0)$ is the unitary vector of the $i$th axis. With $\boldsymbol{x}^{\mathrm{t}}$ denoting the transpose of any $1 \times n$ vector $\boldsymbol{x}$, we have

$$R_n(\boldsymbol{X}_\epsilon) = \max_{\boldsymbol{\alpha} \in A}\{\boldsymbol{\alpha} \boldsymbol{X}_\epsilon^{\mathrm{t}}\} \le \max_{\boldsymbol{\alpha} \in A}\{\boldsymbol{\alpha}\boldsymbol{\mu}^{\mathrm{t}}\} + \frac{I_\epsilon}{\sqrt{\epsilon}} \max_{\boldsymbol{\alpha} \in A}\{\boldsymbol{\alpha}\boldsymbol{\Sigma}^{1/2}\boldsymbol{V}^{\mathrm{t}}\}.$$

Clearly, $\max_{\boldsymbol{\alpha} \in A}\{\boldsymbol{\alpha}\boldsymbol{\mu}^{\mathrm{t}}\} = \max_i\{\mu_i\} - \min_i\{\mu_i\}$, while

$$\mathbb{E}\left(\frac{I_\epsilon}{\sqrt{\epsilon}} \max_{\boldsymbol{\alpha} \in A}\{\boldsymbol{\alpha}\boldsymbol{\Sigma}^{1/2}\boldsymbol{V}^{\mathrm{t}}\}\right) = \sqrt{\epsilon}\,\mathbb{E}\left(\max_{\boldsymbol{\alpha} \in A}\{\boldsymbol{\alpha}\boldsymbol{\Sigma}^{1/2}\boldsymbol{V}^{\mathrm{t}}\}\right) \le \sqrt{\epsilon} \sum_{\boldsymbol{\alpha} \in A} \mathbb{E}|\boldsymbol{\alpha}\boldsymbol{\Sigma}^{1/2}\boldsymbol{V}^{\mathrm{t}}| = \gamma\sqrt{\epsilon},$$

where $\gamma \ge 0$ is a finite constant independent of $\epsilon$. It follows that

$$\mathbb{E}R_n(\boldsymbol{X}_\epsilon) \le \max_i\{\mu_i\} - \min_i\{\mu_i\} + \gamma\sqrt{\epsilon}$$

and thus,

$$\lim_{\epsilon \searrow 0} \mathbb{E}R_n(\boldsymbol{X}_\epsilon) = \max_i\{\mu_i\} - \min_i\{\mu_i\}.$$

Hence, the best possible lower bound for $\mathbb{E}R_n$ is $\max_i\{\mu_i\} - \min_i\{\mu_i\}$.

Regarding the supremum in Equation (7), we shall make use of the following definition.

DEFINITION 2.1 *A random vector $\boldsymbol{X} \in \mathcal{F}_n(\boldsymbol{\mu}, \boldsymbol{\sigma})$ of dimension $n \geq 2$ will be called extremal random vector (for the range) if $\mathbb{E}R_n(\boldsymbol{X}) = \sup \mathbb{E}R_n$, where the supremum is taken over $\mathcal{F}_n(\boldsymbol{\mu}, \boldsymbol{\sigma})$. The class of extremal random vectors is denoted by $\mathcal{E}_n(\boldsymbol{\mu}, \boldsymbol{\sigma})$.*

To the best of our knowledge, the value of the supremum and the nature of the set $\mathcal{E}_n(\boldsymbol{\mu}, \boldsymbol{\sigma})$ have not been analysed elsewhere; it is not even known whether $\mathcal{E}_n(\boldsymbol{\mu}, \boldsymbol{\sigma})$ in nonempty for general $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$. In the present article we shall address both issues.

We start with a deterministic inequality which is the range analogue of the inequality given by Lai and Robbins [18]:

LEMMA 2.1 *For any $\boldsymbol{X} \in \mathbb{R}^n$, $c \in \mathbb{R}$ and $\lambda > 0$,*

$$R_n \leq -(n-2)\lambda + \frac{\lambda}{2} \sum_{i=1}^{n} \left\{ \left| \frac{X_i - c}{\lambda} - 1 \right| + \left| \frac{X_i - c}{\lambda} + 1 \right| \right\}. \tag{8}$$

*The equality in Inequality* (8) *is attained if and only if*

$$X_{1:n} \leq c - \lambda \leq X_{2:n} \leq \cdots \leq X_{n-1:n} \leq c + \lambda \leq X_{n:n}. \tag{9}$$

The lemma entails that the use of two decision variables is sufficient for the proper handling of $R_n$. Also, it suggests the investigation of $\sup \mathbb{E}\{|X-1| + |X+1|\}$ when $X$ is a random variable with given mean and variance:

LEMMA 2.2 *For any $X \in \mathcal{F}_1(\mu, \sigma)$ $(0 < \sigma < \infty)$,*

$$\mathbb{E}\{|X-1| + |X+1|\} \leq U(\mu, \sigma), \tag{10}$$

*where*

$$U(\mu, \sigma) := \begin{cases} 2\sqrt{\mu^2 + \sigma^2} & \text{if } \mu^2 + \sigma^2 \geq 4, \\ 2 + \frac{1}{2}(\mu^2 + \sigma^2) & \text{if } 2|\mu| < \mu^2 + \sigma^2 < 4, \\ |\mu| + 1 + \sqrt{(|\mu| - 1)^2 + \sigma^2} & \text{if } \mu^2 + \sigma^2 \leq 2|\mu| < 4. \end{cases} \tag{11}$$

*The equality in Inequality* (10) *is attained by a unique random variable $X^* \in \mathcal{F}_1(\mu, \sigma)$. Depending on $(\mu, \sigma)$, $X^*$ assumes two or three supporting values. More precisely:*

(a) *For $\mu^2 + \sigma^2 \geq 4$,*

$$\mathbb{P}[X^* = \sqrt{\mu^2 + \sigma^2}] = \frac{1}{2}\left(1 + \frac{\mu}{\sqrt{\mu^2 + \sigma^2}}\right) = 1 - \mathbb{P}[X^* = -\sqrt{\mu^2 + \sigma^2}].$$

(b) *For $2|\mu| < \mu^2 + \sigma^2 < 4$,*

$$\mathbb{P}[X^* = 0] = 1 - \frac{\mu^2 + \sigma^2}{4}, \quad \mathbb{P}[X^* = -2] = \frac{\mu^2 + \sigma^2 - 2\mu}{8}, \quad \mathbb{P}[X^* = 2] = \frac{\mu^2 + \sigma^2 + 2\mu}{8}.$$

(c) *For $\mu^2 + \sigma^2 \leq 2\mu$ (and hence, $0 < \mu < 2$),*

$$\mathbb{P}[X^* = 1 + \sqrt{(\mu - 1)^2 + \sigma^2}] = \frac{1}{2}\left(1 + \frac{\mu - 1}{\sqrt{(\mu - 1)^2 + \sigma^2}}\right)$$
$$= 1 - \mathbb{P}[X^* = 1 - \sqrt{(\mu - 1)^2 + \sigma^2}].$$

(d) *For* $\mu^2+\sigma^2 \le -2\mu$ (*and hence,* $-2<\mu<0$),

$$\mathbb{P}[X^* = -1+\sqrt{(\mu+1)^2+\sigma^2}] = \frac{1}{2}\left(1+\frac{\mu+1}{\sqrt{(\mu+1)^2+\sigma^2}}\right)$$
$$= 1-\mathbb{P}[X^* = -1-\sqrt{(\mu+1)^2+\sigma^2}].$$

*Remark 2.1* Isii [20] presented general results that include inequalities of the form of Lemma 2.2; see also [21, Theorem 2.1, p. 472]. The univariate mean–variance inequality in Isii's paper can be stated as follows: If $h:\mathbb{R}\to\mathbb{R}$ is a Borel function, $\mu\in\mathbb{R}$ and $\sigma>0$ then

$$\sup_{X\in\mathcal{F}_1(\mu,\sigma)} \mathbb{E}h(X) = \inf_{\alpha_0,\alpha_1,\alpha_2}\{\alpha_0+\alpha_1\mu+\alpha_2(\mu^2+\sigma^2) : \alpha_0+\alpha_1 x+\alpha_2 x^2 \ge h(x) \text{ for all } x\}.$$

Isii showed that the above infimum is attained by some $\boldsymbol{\alpha}^* = (\alpha_0^*,\alpha_1^*,\alpha_2^*)\in A$, where

$$A = \{(\alpha_0,\alpha_1,\alpha_2) : \alpha_0+\alpha_1 x+\alpha_2 x^2 \ge h(x) \text{ for all } x\in\mathbb{R}\}\subseteq\mathbb{R}^3,$$

provided that the infimum is finite. However, usually it is not an easy task to specify the subset $A$ and the extremal point(s) $\boldsymbol{\alpha}^*$. Lemma 2.2 shows that this is possible for $h(x)=|x-1|+|x+1|$ and, more importantly, characterizes the case of equality.

The following corollary is a straightforward consequence of Lemma 2.2.

Corollary 2.1 *Let* $X\in\mathcal{F}_1(\mu,\sigma)$ $(0<\sigma<\infty)$. *Fix* $c\in\mathbb{R}$ *and* $\lambda>0$. *Then,*

$$\mathbb{E}\{|(X-c)-\lambda|+|(X-c)+\lambda|\} \le \lambda U\left(\frac{\mu-c}{\lambda},\frac{\sigma}{\lambda}\right), \tag{12}$$

*with* $U(\cdot,\cdot)$ *given by Equation* (11). *The equality in Inequality* (12) *is attained by a unique two- or three-valued random variable. Setting*

$$\xi=\mu-c,\quad \theta=\sqrt{(\mu-c)^2+\sigma^2},\quad \alpha=\sqrt{(\xi-\lambda)^2+\sigma^2},\quad \beta=\sqrt{(\xi+\lambda)^2+\sigma^2},$$

*the distribution that attains the equality is described by the following table:*

| No: Condition on $\mu,\sigma,c,\lambda$<br>Tight Upper Bound $\lambda U(\frac{\mu-c}{\lambda},\frac{\sigma}{\lambda})$ | value $x^-$<br>probability $p^-$ | value $x^o$<br>probability $p^o$ | value $x^+$<br>probability $p^+$ |
|---|---|---|---|
| 1: $(\mu-c)^2+\sigma^2\ge 4\lambda^2$<br>$2\sqrt{(\mu-c)^2+\sigma^2}$ | $c-\theta$<br>$\frac{1}{2}(1-\frac{\xi}{\theta})$ | | $c+\theta$<br>$\frac{1}{2}(1+\frac{\xi}{\theta})$ |
| 2: $2\lambda\lvert\mu-c\rvert<(\mu-c)^2+\sigma^2<4\lambda^2$<br>$2\lambda+\frac{1}{2\lambda}[(\mu-c)^2+\sigma^2]$ | $c-2\lambda$<br>$\frac{1}{8\lambda^2}[\theta^2-2\lambda\xi]$ | $c$<br>$1-\frac{1}{4\lambda^2}\theta^2$ | $c+2\lambda$<br>$\frac{1}{8\lambda^2}[\theta^2+2\lambda\xi]$ |
| 3: $(\mu-c)^2+\sigma^2\le 2\lambda(\mu-c)$<br>$\mu-c+\lambda+\sqrt{(\mu-c-\lambda)^2+\sigma^2}$ | | $c+\lambda-\alpha$<br>$\frac{1}{2}(1-\frac{\xi-\lambda}{\alpha})$ | $c+\lambda+\alpha$<br>$\frac{1}{2}(1+\frac{\xi-\lambda}{\alpha})$ |
| 4: $(\mu-c)^2+\sigma^2\le 2\lambda(c-\mu)$<br>$c-\mu+\lambda+\sqrt{(c-\mu-\lambda)^2+\sigma^2}$ | $c-\lambda-\beta$<br>$\frac{1}{2}(1-\frac{\xi+\lambda}{\beta})$ | $c-\lambda+\beta$<br>$\frac{1}{2}(1+\frac{\xi+\lambda}{\beta})$ | |

*Proof* Write $|(X-c)-\lambda|+|(X-c)+\lambda|=\lambda\{|(X-c)/\lambda-1|+|(X-c)/\lambda+1|\}$. Since $Y=(X-c)/\lambda\in\mathcal{F}_1((\mu-c)/\lambda,\sigma/\lambda)$, Lemma 2.2 yields Inequality (12) as follows:

$$\mathbb{E}\{|(X-c)-\lambda|+|(X-c)+\lambda|\}=\lambda\mathbb{E}\{|Y-1|+|Y+1|\}\le\lambda U\left(\frac{\mu-c}{\lambda},\frac{\sigma}{\lambda}\right).$$

Since $\lambda>0$, Lemma 2.2 asserts that the equality is attained by a unique random variable $Y^*\in\mathcal{F}_1((\mu-c)/\lambda,\sigma/\lambda)$. Thus, $X^*=c+\lambda Y^*$ is the unique random variable in $\mathcal{F}_1(\mu,\sigma)$ that attains the equality in Inequality (12). Substituting the probability function of $Y^*$ in the four distinct cases of Lemma 2.2 we obtain the probabilities and supporting points as in the table. ■

It is important to observe that, whatever the values of $\mu,\sigma,c,\lambda$ are, the supporting points satisfy the relation $x^-<c-\lambda<x^o<c+\lambda<x^+$.

We can now obtain the proposed upper bound for the expected range.

THEOREM 2.1 *If* $\mathbb{E}\boldsymbol{X}=\boldsymbol{\mu}$ *and* $\operatorname{Var}\boldsymbol{X}=\boldsymbol{\sigma}^2$ *then*

$$\mathbb{E}R_n\le\inf_{c\in\mathbb{R},\lambda>0}\left\{-(n-2)\lambda+\frac{\lambda}{2}\sum_{i=1}^n U\left(\frac{\mu_i-c}{\lambda},\frac{\sigma_i}{\lambda}\right)\right\},\tag{13}$$

*where the function* $U(\cdot,\cdot):\mathbb{R}\times(0,\infty)\to(2,\infty)$ *is given by Equation* (11).

*Proof* Fix $c\in\mathbb{R}$, $\lambda>0$. We take expectations in Inequality (8) and then use Inequality (12) to get

$$\begin{aligned}\mathbb{E}R_n&\le-(n-2)\lambda+\frac{1}{2}\sum_{i=1}^n\mathbb{E}\{|(X_i-c)-\lambda|+|(X_i-c)+\lambda|\}\\&\le-(n-2)\lambda+\frac{\lambda}{2}\sum_{i=1}^n U\left(\frac{\mu_i-c}{\lambda},\frac{\sigma_i}{\lambda}\right).\end{aligned}$$

Since for all $c\in\mathbb{R}$ and $\lambda>0$ the last quantity is an upper bound for $\mathbb{E}R_n$, its infimum is an upper bound too. ■

*Remark 2.2* It is not clear at this stage whether the upper bound (13) is tight, and it is not an obvious task to find $c=c_0$ and $\lambda=\lambda_0$ (if exist) that realize the infimum in the RHS of Inequality (13). However, the substitution of any (convenient) arguments $c$ and $\lambda$ in the function

$$\phi_n(c,\lambda):=-(n-2)\lambda+\frac{\lambda}{2}\sum_{i=1}^n U\left(\frac{\mu_i-c}{\lambda},\frac{\sigma_i}{\lambda}\right)\tag{14}$$

will produce an upper bound for $\mathbb{E}R_n$. For example, one can choose $c=\bar{\mu}$ and $\lambda=\frac{1}{4}\mathrm{AG}_n$ (see Equation (4)). A simple way to produce a closed-form upper bound is the following: First observe that

$$\lambda U\left(\frac{\mu_i-c}{\lambda},\frac{\sigma_i}{\lambda}\right)\le 2\lambda+\frac{1}{2\lambda}[(\mu_i-c)^2+\sigma_i^2],$$

because the RHS is an upper bound for the expectation $\mathbb{E}\{|(X_i-c)-\lambda|+|(X_i-c)+\lambda|\}$ (since $|(X_i-c)-\lambda|+|(X_i-c)+\lambda|\le2\lambda+(1/2\lambda)(X_i-c)^2$ and $\mathbb{E}\{2\lambda+(1/2\lambda)(X_i-c)^2\}=$

$2\lambda + (1/2\lambda)[(\mu_i - c)^2 + \sigma_i^2]$), while the LHS is the least upper bound for the same expectation as $X_i$ varies in $\mathcal{F}_1(\mu_i, \sigma_i)$. It follows that

$$\phi_n(c,\lambda) \le \bar{\phi}_n(c,\lambda) := 2\lambda + \frac{1}{4\lambda}\sum_{i=1}^{n}\{(\mu_i - c)^2 + \sigma_i^2\}.$$

Minimizing $\bar{\phi}_n(c,\lambda)$ is a simple fact: it suffices to take $c = \bar{\mu}$ and $\lambda = \frac{1}{4}\mathrm{AG}_n$ as before. Observing that $\sum_{i=i}^{n}\{(\mu_i - \bar{\mu})^2 + \sigma_i^2\} = \frac{1}{2}\mathrm{AG}_n^2$, we get

$$\mathbb{E}R_n \le \inf_{c\in\mathbb{R},\lambda>0}\phi_n(c,\lambda) \le \inf_{c\in\mathbb{R},\lambda>0}\bar{\phi}_n(c,\lambda) = \bar{\phi}_n\left(\bar{\mu}, \frac{1}{4}\mathrm{AG}_n\right) = \frac{1}{2}\mathrm{AG}_n + \frac{1}{2}\mathrm{AG}_n = \mathrm{AG}_n.$$

Now it became clear that the bound in Inequality (13) is reasonable, since it outperforms the bound in Inequality (4) for any given values of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$. As a result, the $\mathrm{AG}_n$ bound need not be tight; e.g. the infimum of $\phi_n(c,\lambda)$ need not be attained at $(c,\lambda) = (\bar{\mu}, \frac{1}{4}\mathrm{AG}_n)$. We shall prove in the sequel that the new bound is always tight, and (for $n \ge 3$) the infimum in the RHS of Inequality (13) is attained by a unique value $(c_0, \lambda_0)$.

*Remark 2.3* Fixing $\mu$ in Inequality (11) and taking limits for $\sigma \searrow 0$ we see that

$$\lim_{\sigma\searrow 0} U(\mu,\sigma) = 2\max\{|\mu|, 1\} = |\mu - 1| + |\mu + 1|, \quad \mu \in \mathbb{R}.$$

Let us now set $\sigma_{n:n} = \max\{\sigma_1, \ldots, \sigma_n\}$ and fix $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$. Then,

$$\lim_{\sigma_{n:n}\searrow 0}\phi_n(c,\lambda) = -(n-2)\lambda + \frac{\lambda}{2}\sum_{i=1}^{n}\left\{\left|\frac{\mu_i - c}{\lambda} - 1\right| + \left|\frac{\mu_i - c}{\lambda} + 1\right|\right\}, \ \boldsymbol{\mu} \in \mathbb{R}^n, \ c \in \mathbb{R}, \ \lambda > 0.$$

Let $\mu_{1:n} \le \cdots \le \mu_{n:n}$ be the ordered values of $\mu_1, \ldots, \mu_n$, and assume that the $\mu$'s are not all equal, that is, $\mu_{1:n} < \mu_{n:n}$. Substituting in the above limit $c = c_0 = (\mu_{1:n} + \mu_{n:n})/2$, $\lambda = \lambda_0 = (\mu_{n:n} - \mu_{1:n})/2 > 0$, we obtain

$$\lim_{\sigma_{n:n}\searrow 0}\phi_n(c_0,\lambda_0) = -(n-2)\lambda_0 + \frac{\lambda_0}{2}\sum_{i=1}^{n}\left\{\left|\frac{\mu_i - c_0}{\lambda_0} - 1\right| + \left|\frac{\mu_i - c_0}{\lambda_0} + 1\right|\right\} = \mu_{n:n} - \mu_{1:n}.$$

Note that the last equality follows from Inequality (8) and (9), applied to $\boldsymbol{X} = \boldsymbol{\mu}$ (with $R_n(\boldsymbol{\mu}) = \mu_{n:n} - \mu_{1:n}$), observing that for the particular choice of $(c_0, \lambda_0)$,

$$\mu_{1:n} \le c_0 - \lambda_0 \le \mu_{2:n} \le \cdots \le \mu_{n-1:n} \le c_0 + \lambda_0 \le \mu_{n:n}.$$

For any $\boldsymbol{X} \in \mathcal{F}_n(\boldsymbol{\mu}, \boldsymbol{\sigma})$ it is true that $\mu_{n:n} - \mu_{1:n} \le \mathbb{E}R_n(\boldsymbol{X}) \le \inf_{c\in\mathbb{R},\lambda>0}\{\phi_n(c,\lambda)\}$. Therefore,

$$\mu_{n:n} - \mu_{1:n} \le \lim_{\sigma_{n:n}\searrow 0}\mathbb{E}R_n(\boldsymbol{X}) \le \lim_{\sigma_{n:n}\searrow 0}\left\{\inf_{c\in\mathbb{R},\lambda>0}\phi_n(c,\lambda)\right\} \le \lim_{\sigma_{n:n}\searrow 0}\phi_n(c_0,\lambda_0) = \mu_{n:n} - \mu_{1:n},$$

and we conclude that

$$\lim_{\sigma_{n:n}\searrow 0}\left\{\inf_{c\in\mathbb{R},\lambda>0}\phi_n(c,\lambda)\right\} = \mu_{n:n} - \mu_{1:n}. \tag{15}$$

The limit (15) continue to hold even if all $\mu_i$'s are equal. Then $\mu_{1:n} = \mu_{n:n}$ and the inequality $\inf_{c\in\mathbb{R},\lambda>0}\{\phi_n(c,\lambda)\} \le \mathrm{AG}_n$ (see Remark 2.2) shows that

$$0 \le \inf_{c\in\mathbb{R},\lambda>0}\{\phi_n(c,\lambda)\} \le \mathrm{AG}_n = \sqrt{2\sum_{i=1}^{n}\sigma_i^2} \le \sigma_{n:n}\sqrt{2n} \to 0, \quad \text{as } \sigma_{n:n} \searrow 0.$$

From these considerations, it is again clear that the $\mathrm{AG}_n$ bound is not tight in general; for example,

$$\lim_{\sigma_{n:n}\searrow 0} \mathrm{AG}_n = \sqrt{2\sum_{i=1}^{n}(\mu_i-\bar{\mu})^2} > \mu_{n:n}-\mu_{1:n},$$

whenever ($n \geq 3$ and) $\mu_{1:n}+\mu_{n:n} \neq 2\bar{\mu}$. The $\mathrm{AG}_n$ bound need not be tight even for equal $\mu_i$'s; see Theorem 3.1 and Example 3.2 .



## 3. When is the Arnold–Groeneveld bound tight?

Arnold and Groeneveld,[9] Rychlik [4] and Papadatos [13] showed that if $\mu_i=\mu$ and $\sigma_i=\sigma$ for all $i$, the $\mathrm{AG}_n$ bound of Equation (4), which reduces to Inequality (2), is attainable. In the present section we provide an exact characterization of the attainability of the $\mathrm{AG}_n$ bound under any mean–variance information.

The proof of Theorem 3.1 is based on the construction of particular bivariate probability distributions supported in a subset of $\{1,\dots,n\}^2$. A distribution of this kind corresponds to an $n\times n$ matrix with nonnegative elements having sum 1; a probability matrix. Matrices of this form with integer-valued entries have been extensively studied; for a recent review, see [22]. The actual question, related to our problem, is whether there exist probability matrices with given marginals and vanishing trace.

The following notation and terminology will be used in the sequel.

Definition 3.1 *An $n\times m$ matrix $Q=(q_{ij})$ $(n\geq 1,\ m\geq 1)$ is called a probability matrix if it has nonnegative elements summing to 1. In particular, a n-variate probability vector $\boldsymbol{p}=(p_1,\dots,p_n)$ is a probability matrix with dimension $1\times n$, and $X\sim\boldsymbol{p}$ is a convention for $\mathbb{P}[X=i]=p_i$ for all i. The marginals of Q, say $\boldsymbol{p}$, $\boldsymbol{q}$, are the probability vectors obtained by summing the rows and columns of Q, respectively; and $\mathcal{M}(\boldsymbol{p},\boldsymbol{q})$ denotes the class of probability matrices with given marginals $\boldsymbol{p},\boldsymbol{q}$. Moreover, $(X,Y)\sim Q$ is a convention for $\mathbb{P}[X=i,Y=j]=q_{ij}$ for all $i,j$.*

We now state a characterization for the $\mathrm{AG}_n$ bound.

Theorem 3.1 *Assume that $\mathbb{E}\boldsymbol{X}=\boldsymbol{\mu}$ and $\operatorname{Var}\boldsymbol{X}=\boldsymbol{\sigma}^2$. Then the equality in Inequality* (4) *is attainable if and only if both conditions* (i) *and* (ii) *below are satisfied.*

$$\left.\begin{aligned} &\text{(i)}\quad |\mu_i-\bar{\mu}| \leq \frac{\sqrt{2}[(\mu_i-\bar{\mu})^2+\sigma_i^2]}{\sqrt{\sum_{j=1}^{n}\{(\mu_j-\bar{\mu})^2+\sigma_j^2\}}},\\ &\text{(ii)}\quad \frac{(\mu_i-\bar{\mu})^2+\sigma_i^2}{\sum_{j=1}^{n}\{(\mu_j-\bar{\mu})^2+\sigma_j^2\}} \leq \frac{1}{2}, \end{aligned}\right\}\quad i=1,\dots,n. \tag{16}$$

*Provided that* (i) *and* (ii) *are fulfilled, any extremal random vector $\boldsymbol{X}\in\mathcal{E}_n(\boldsymbol{\mu},\boldsymbol{\sigma})$ has the representation*

$$\boldsymbol{X}=g(X,Y):=\bar{\mu}\,\mathbf{1}+\frac{\boldsymbol{e}(X)-\boldsymbol{e}(Y)}{\sqrt{2}}\sqrt{\sum_{j=1}^{n}\{(\mu_j-\bar{\mu})^2+\sigma_j^2\}}, \tag{17}$$

*where* $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^n$, $\boldsymbol{e}(i) = (0, \ldots, 1, \ldots, 0) \in \mathbb{R}^n$ *is the unitary vector of the ith axis, and* $(X, Y)$ *is a discrete random pair satisfying* $\mathbb{P}[X = Y] = 0$, *with marginal distributions*

$$
\begin{aligned}
p_i^+ &= \mathbb{P}[X = i] = \frac{(\mu_i - \bar{\mu})^2 + \sigma_i^2 + (1/2)(\mu_i - \bar{\mu})\mathrm{AG}_n}{\sum_{j=1}^n \{(\mu_j - \bar{\mu})^2 + \sigma_j^2\}}, \\
p_i^- &= \mathbb{P}[Y = i] = \frac{(\mu_i - \bar{\mu})^2 + \sigma_i^2 - (1/2)(\mu_i - \bar{\mu})\mathrm{AG}_n}{\sum_{j=1}^n \{(\mu_j - \bar{\mu})^2 + \sigma_j^2\}}, \quad i = 1, \ldots, n.
\end{aligned} \tag{18}
$$

*Moreover, if the inequalities in Conditions* (16) *are strict for all i, we can find infinitely many extremal random vectors; and if Conditions* (16) *is satisfied and for some i we have equality in* (ii), *then the extremal random vector is unique.*

*Remark 3.1* Let $(\mu_1, \mu_2, \mu_3) = (-1, 0, 1)$, $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (1, 3, 2)$, so that (16) holds. However, Condition (16)(ii) is satisfied with strict inequalities for all $i$, while this is not true for Condition (16)(i). We find $\mathrm{AG}_3 = 4$ and $\boldsymbol{p}^+ = (0, \frac{3}{8}, \frac{5}{8})$, $\boldsymbol{p}^- = (\frac{1}{2}, \frac{3}{8}, \frac{1}{8})$. It is easily seen that the distribution of $(X_1, X_2, X_3)$ (given in Equation (17)) is uniquely defined: it assigns probabilities $\frac{2}{8}, \frac{2}{8}, \frac{3}{8}, \frac{1}{8}$, to the points $(-2, 2, 0)$, $(-2, 0, 2)$, $(0, -2, 2)$, $(0, 2, -2)$, respectively. It follows that a random vector that attains the $\mathrm{AG}_n$ bound can be unique even if Conditions (16)(ii) is satisfied with strict inequalities for all $i$.

*Example 3.1* *The homogeneous case* $\mu_i = \mu$, $\sigma_i = \sigma > 0$. Conditions (16) are obviously satisfied with strict inequalities (for $n \geq 3$) and the $\mathrm{AG}_n$ bound is sharp (see also Inequality (2)):

$$\sup \mathbb{E}R_n = \mathrm{AG}_n = \sigma\sqrt{2n}.$$

Moreover, $p_i^+ = p_i^- = 1/n$ and from Theorem 3.1 we see that infinitely many random vectors attain the equality. The totality of them is characterized by Equation (17) via the probability matrices $Q$ of $(X, Y)$. Recall that $X$ and $Y$ are, respectively, the positions where $\mu + \sigma\sqrt{n/2}$ and $\mu - \sigma\sqrt{n/2}$ appears in the extremal vector $(X_1, \ldots, X_n)$; the rest entries are equal to $\mu$. Thus, $Q$ has uniform marginals and vanishing principal diagonal. A famous theorem of Birkhoff on *magic matrices* asserts that any matrix with nonnegative elements having row/column sums equal to 1 is a convex combination of *permutation matrices*, i.e. matrices with entries 0 or 1, having exactly one 1 in each row and in each column; see [23, Theorem 2.54]. From Birkhoff's result it is evident that the probability matrix $Q$ of $(X, Y)$, corresponding to any extremal random vector $\boldsymbol{X} = \mu\mathbf{1} + \sigma[\boldsymbol{e}(X) - \boldsymbol{e}(Y)]\sqrt{n/2}$, can be written as

$$Q = \sum_{i=1}^k \lambda_i D_i, \quad \lambda_i \geq 0, \quad \sum_{i=1}^k \lambda_i = \frac{1}{n},$$

where the $D_i$'s are *derangement matrices*, i.e. permutation matrices with vanishing diagonal entries. It is well known that there exist $n! \sum_{k=0}^n ((-1)^k/k!) \approx \mathrm{e}^{-1} n!$ different derangement matrices; they coincide with the extremal points of the convex polytope $\{D = (d_{ij}) : \sum_i d_{ij} = \sum_j d_{ij} = 1,\ d_{ij} \geq 0,\ d_{ii} = 0 \text{ for all } i, j\}$. In general, a convex polytope has a finite (often quite large) number of extremal points, but it is rather difficult to evaluate them exactly, since their total number depends on the marginals in an ambiguous way (cf. Example 3.2).

*Example 3.2* *The case* $\mu_i = \mu$. Assume $0 < \sigma_1 \leq \cdots \leq \sigma_n$ without loss of generality. From Theorem 3.1, we see that if the larger variance does not dominate the sum of the other variances then the $\mathrm{AG}_n$ bound is tight:

$$\sup \mathbb{E}R_n = \mathrm{AG}_n = \sqrt{2\sum_{i=1}^{n}\sigma_i^2}, \quad \text{whenever } \sigma_n^2 \le \sum_{i=1}^{n-1}\sigma_i^2.$$

Moreover, if $\sigma_n^2 = \sum_{i=1}^{n-1}\sigma_i^2$, the equality is uniquely attained by the random vector $\boldsymbol{X}$ taking values

$$\boldsymbol{x}_i = \left(\mu,\ldots,\mu,\mu+\frac{\mathrm{AG}_n}{2},\mu,\ldots,\mu;\mu-\frac{\mathrm{AG}_n}{2}\right), \quad \text{with probability } p_i,$$
$$\boldsymbol{y}_i = \left(\mu,\ldots,\mu,\mu-\frac{\mathrm{AG}_n}{2},\mu,\ldots,\mu;\mu+\frac{\mathrm{AG}_n}{2}\right), \quad \text{with probability } p_i,$$

where $p_i = \sigma_i^2/\sum_{j=1}^{n}\sigma_j^2$, $i = 1,\ldots,n-1$. Of course, if $\sigma_n^2 < \sum_{i=1}^{n-1}\sigma_i^2$ then there exist infinitely many extremal random vectors. They have the form $g(X,Y)$ (see Equation (17)), with $\mathbb{P}[X = Y] = 0$, $X \sim \boldsymbol{p}$, $Y \sim \boldsymbol{p}$, where $\boldsymbol{p} = (p_1,\ldots,p_{n-1},p_n)$.

However, if $\sigma_n^2 > \sum_{i=1}^{n-1}\sigma_i^2$ then the $\mathrm{AG}_n$ is no longer tight: The infimum in Inequality (13) is attained at $c_0 = \mu$, $\lambda_0 = \frac{1}{2}\sqrt{\sum_{i=1}^{n-1}\sigma_i^2} < \frac{1}{4}\mathrm{AG}_n$, and we get the inequality

$$\mathbb{E}R_n \le \phi_n(c_0,\lambda_0) = \sigma_n + \sqrt{\sum_{i=1}^{n-1}\sigma_i^2} \quad \left(\sigma_n^2 > \sum_{i=1}^{n-1}\sigma_i^2\right).$$

From $\sqrt{x}+\sqrt{y} < \sqrt{2(x+y)}$ for $x \ne y$, we conclude that this bound is strictly better than $\mathrm{AG}_n$. Moreover, the new bound is tight; one can verify that the equality is (uniquely) attained by the random vector $\boldsymbol{X}$ taking values

$$\boldsymbol{x}_i = \left(\mu,\ldots,\mu,\mu+\sqrt{\sum_{i=1}^{n-1}\sigma_i^2},\mu,\ldots,\mu;\mu-\sigma_n\right), \quad \text{with probability } p_i,$$
$$\boldsymbol{y}_i = \left(\mu,\ldots,\mu,\mu-\sqrt{\sum_{i=1}^{n-1}\sigma_i^2},\mu,\ldots,\mu;\mu+\sigma_n\right), \quad \text{with probability } p_i,$$

where $p_i = \sigma_i^2/2\sum_{j=1}^{n-1}\sigma_j^2$, $i = 1,\ldots,n-1$. Thus, the tight upper bound on the expected range from dependent observations with equal means admits a simple closed form:

$$\sup \mathbb{E}R_n = \begin{cases} \sqrt{2\sum_{i=1}^{n}\sigma_i^2} & \text{if } 2\max_i\{\sigma_i^2\} \le \sum_{i=1}^{n}\sigma_i^2, \\ \max_i\{\sigma_i\} + \sqrt{\sum_{i=1}^{n}\sigma_i^2 - \max_i\{\sigma_i^2\}} & \text{if } 2\max_i\{\sigma_i^2\} \ge \sum_{i=1}^{n}\sigma_i^2. \end{cases} \tag{19}$$

Assuming that one variance tends to infinity (and keeping all other variances bounded), the limit $\lim_{\sigma_i\to\infty}(\sup \mathbb{E}R_n/\mathrm{AG}_n) = \frac{1}{\sqrt{2}} \approx .707$ says that we can gain of an up to 30% improvement over the $\mathrm{AG}_n$ bound.

The following lemma will play an important role in verifying existence of extremal random vectors.

LEMMA 3.1 *Let* $\boldsymbol{p} = (p_1, \dots, p_n)$ *and* $\boldsymbol{q} = (q_1, \dots, q_n)$ *be two probability vectors. A necessary and sufficient condition for the existence of a random pair* $(X, Y)$ *with*

$$\mathbb{P}[X = Y] = 0, \quad X \sim \boldsymbol{p}, \quad Y \sim \boldsymbol{q} \tag{20}$$

*is the following:*

$$\max_{1 \le i \le n} \{p_i + q_i\} \le 1. \tag{21}$$

*If the equality holds in Inequality* (21), *the random pair* $(X, Y)$ *is uniquely defined. If strict inequality holds in Inequality* (21) *and, furthermore,* $\min_i\{p_i\} > 0$, $\min_i\{q_i\} > 0$, *then there exist infinitely many random pairs satisfying Equation* (20).

## 4. Convexity

The purpose of the present section is to verify that for any given values of $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, the function $\phi_n(c, \lambda)$ of Equation (14) is convex. For convenience we set $T := \mathbb{R} \times (0, \infty)$ for the domain of both functions $U$ (of Equation (11)) and $\phi_n$.

We begin with a simple lemma.

LEMMA 4.1 *The function* $U(x, y) : T \to (2, \infty)$ *of Equation* (11) *has continuous partial derivatives, that is,* $U \in C^1(T)$.

We also need another simple lemma; see, e.g. [23].

LEMMA 4.2 *Let* $K$ *be a convex subset of* $\mathbb{R}^n$ *and* $f : K \to \mathbb{R}$. *For* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *in* $K$ *consider the function* $g : [0, 1] \to \mathbb{R}$ *given by*

$$g(t) := f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x})), \quad 0 \le t \le 1.$$

*Then,* $f$ *is convex if and only if* $g$ *is convex for any choice of* $\boldsymbol{x}$ *and* $\boldsymbol{y}$ *in* $K$.

Also, we shall make use of the following lemma.

LEMMA 4.3 *Consider a finite interval* $[\alpha, \beta]$, *a partition*

$$\alpha = t_0 < t_1 < \cdots < t_k < t_{k+1} = \beta$$

*and the convex functions* $g_i : [\alpha, \beta] \to \mathbb{R}$ (*or, merely,* $g_i : [t_{i-1}, t_i] \to \mathbb{R}$), $i = 1, \dots, k + 1$. *Assume that*

$$g_i(t_i) = g_{i+1}(t_i) \quad \text{and} \quad g_i'(t_i-) \le g_{i+1}'(t_i+), \quad i = 1, \dots, k, \tag{22}$$

*where* $g'(t-)$ *and* $g'(t+)$ *denote, respectively, the left- and right-hand side derivatives of* $g$ *at* $t$. *Then, the function*

$$g(t) := \begin{cases} g_1(t), & \alpha \le t \le t_1, \\ g_2(t), & t_1 \le t \le t_2, \\ \vdots & \\ g_k(t), & t_{k-1} \le t \le t_k, \\ g_{k+1}(t), & t_k \le t \le \beta, \end{cases} \tag{23}$$

*is convex.*

*Proof* Since all $g_i$ have non-decreasing left- and right-hand side derivatives, it is easily seen that the same is true for $g$. ■

Now we can verify the following result.

PROPOSITION 4.1 *The function $U : T \to (2, \infty)$ in (11) is convex.*

Finally, we shall make use of the following property, which seems to be of some independent interest.

LEMMA 4.4 *Let $f(x, y) : T \to \mathbb{R}$ and for fixed $x_0 \in \mathbb{R}$, $y_0 > 0$, consider the function $h(c, \lambda) : T \to \mathbb{R}$ with*

$$h(c, \lambda) := \lambda f\left(\frac{x_0 - c}{\lambda}, \frac{y_0}{\lambda}\right), \quad (c, \lambda) \in \mathbb{R} \times (0, \infty).$$

(i) *If $f$ is convex then $h$ is convex for all choices of $x_0 \in \mathbb{R}$, $y_0 > 0$.*
(ii) *If $h$ is convex for a particular choice of $x_0 \in \mathbb{R}$, $y_0 > 0$, then $f$ is convex.*

We can now state and prove the final conclusion of the present section:

THEOREM 4.1 *For any given $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, the function $\phi_n(c, \lambda)$ in Equation (14) is convex and belongs to $C^1(T)$, $T = \mathbb{R} \times (0, \infty)$.*

*Proof* The fact that $\phi_n \in C^1(T)$ follows by an obvious application of Lemma 4.1. Also, the function $U(x, y)$ in Equation (11) is convex by Proposition 4.1. Hence, by Lemma 4.4, the same is true for the function $h_i(c, \lambda) = \frac{1}{2}\lambda U((\mu_i - c)/\lambda, \sigma_i/\lambda)$ $(i = 1, \ldots, n)$. Since $h(c, \lambda) = -(n-2)\lambda$ is trivially convex, $\phi_n(c, \lambda)$ is a sum of convex functions. ■

## 5. Attainability of the infimum in Inequality (13) at a unique point

From now on we assume that $n \geq 3$. The simple (but interesting) case $n = 2$ is deferred to the last section, noting that the optimal upper bound for $\mathbb{E}R_2$ is closely related to the bound $\text{BNT}_2$ of Inequality (5).

In the present section we shall prove that the minimum value of $\phi_n(c, \lambda)$ is achieved at a unique point $(c_0, \lambda_0) \in T$. Of course, since $\phi_n$ is differentiable, a minimizing point (if exists) has to satisfy the system of equations

$$\frac{\partial}{\partial c}\phi_n(c, \lambda) = 0, \quad \frac{\partial}{\partial \lambda}\phi_n(c, \lambda) = 0. \tag{24}$$

However, due to the complicated form of the derivatives (see Equations (A1), (A2)), it is not a trivial fact to solve the System (24), or even to verify its consistency analytically. On the other hand, as we shall see in the sequel, it is important to know the existence (and uniqueness) of a minimizing point; it will be used in an essential way in the construction of extremal random vectors, concluding tightness of the bound (13).

The attainability of the infimum can be seen as follows:

Set $\epsilon_0 := \frac{1}{4}\min_i\{\sigma_i\} > 0$. For $c \in \mathbb{R}$ and $\lambda \in (0, \epsilon_0]$, $(\mu_i - c)^2 + \sigma_i^2 \ge 4\lambda^2$ $(i = 1, \ldots, n)$. Thus, $\lambda U((\mu_i - c)/\lambda, \sigma_i/\lambda) = 2\sqrt{(\mu_i - c)^2 + \sigma_i^2}$ for all $i$, and

$$\begin{aligned}\phi_n(c, \lambda) &= -(n-2)\lambda + \sum_{i=1}^{n} \sqrt{(\mu_i - c)^2 + \sigma_i^2} \\ &\ge -(n-2)\epsilon_0 + \sum_{i=1}^{n} \sqrt{(\mu_i - c)^2 + \sigma_i^2} = \phi_n(c, \epsilon_0).\end{aligned}$$

The function $c \mapsto \sum_{i=1}^{n} \sqrt{(\mu_i - c)^2 + \sigma_i^2}$ is strictly convex, tending to $\infty$ as $|c| \to \infty$; thus, its minimum is attained at a unique $c = c_1$. From $\phi_n(c, \epsilon_0) \ge \phi_n(c_1, \epsilon_0)$, we get

$$\phi_n(c, \lambda) \ge \phi_n(c_1, \epsilon_0) = -(n-2)\epsilon_0 + \sum_{i=1}^{n} \sqrt{(\mu_i - c_1)^2 + \sigma_i^2}, \quad c \in \mathbb{R}, \ 0 < \lambda \le \epsilon_0.$$

We now chose $\lambda_1 := \frac{1}{2}\min_i\{\sigma_i\}$, so that $\lambda_1 > \epsilon_0$ and $(\mu_i - c_1)^2 + \sigma_i^2 \ge 4\lambda_1^2$ for all $i$. Therefore, $\lambda_1 U((\mu_i - c_1)/\lambda_1, \sigma_i/\lambda_1) = 2\sqrt{(\mu_i - c_1)^2 + \sigma_i^2}$ $(i = 1, \ldots, n)$, and it follows that $\phi_n(c_1, \lambda_1) = -(n-2)\lambda_1 + \sum_{i=1}^{n} \sqrt{(\mu_i - c_1)^2 + \sigma_i^2}$. Since $\lambda_1 > \epsilon_0$ and $n \ge 3$, the inequality $-(n-2)\epsilon_0 > -(n-2)\lambda_1$ leads to $\phi_n(c_1, \epsilon_0) > \phi_n(c_1, \lambda_1)$. Moreover, $U(x, y) \ge U(0, y) = 2 + \int_0^y \min\{t, 2\}\mathrm{d}t > 2$ for all $x \in \mathbb{R}$ and $y > 0$. We thus obtain $\phi_n(c, \lambda) = -(n-2)\lambda + (\lambda/2)\sum_{i=1}^{n} U((\mu_i - c)/\lambda, \sigma_i/\lambda) > -(n-2)\lambda + n\lambda = 2\lambda$ for all $c$ and $\lambda > 0$. Setting $M_0 := \frac{1}{2}\phi_n(c_1, \epsilon_0) > \epsilon_0$ we see that

$$\phi_n(c, \lambda) \ge \phi_n(c_1, \epsilon_0) > \phi_n(c_1, \lambda_1) \quad \text{for all } c \in \mathbb{R}, \ \lambda \in (0, \epsilon_0] \cup [M_0, \infty).$$

Assume now that $\lambda \in (\epsilon_0, M_0)$ with $\epsilon_0$, $M_0$ as above. From the obvious inequality $U(x, y) \ge 2\max\{|x|, 1\} \ge 2|x|$, we get

$$\phi_n(c, \lambda) \ge -(n-2)\lambda + \sum_{i=1}^{n} |\mu_i - c| \ge -(n-2)M_0 + \sum_{i=1}^{n} |\mu_i - c|.$$

The last inequality shows that $\phi_n(c, \lambda) \to \infty$ as $|c| \to \infty$, uniformly in $\lambda \in (\epsilon_0, M_0)$; thus, we can find a constant $C_0$ such that

$$\phi_n(c, \lambda) \ge \phi_n(c_1, \epsilon_0) \quad \text{for all } |c| \ge C_0, \ \lambda \in (\epsilon_0, M_0).$$

Since $\phi_n(c_1, \epsilon_0) > \phi_n(c_1, \lambda_1)$, we arrived at the conclusion

$$\phi_n(c, \lambda) > \phi_n(c_1, \lambda_1) \quad \text{for all } (c, \lambda) \text{ with } |c| \ge C_0 \text{ or } \lambda \le \epsilon_0 \text{ or } \lambda \ge M_0.$$

This inequality shows that any minimizing point $(c_0, \lambda_0)$ of (the continuous function) $\phi_n(c, \lambda)$ over the compact rectangle $R := [-C_0, C_0] \times [\epsilon_0, M_0]$ must lie in the interior of $R$. The convexity of $\phi_n$ implies that its global minimum is attained at $(c_0, \lambda_0)$. On the other hand, the differentiability of $\phi_n$ shows that $(c, \lambda) = (c_0, \lambda_0)$ is a solution to the System (24); and the convexity of $\phi_n$ implies that any such solution is a minimizing point.

Let us now define

$$T_0 := \{(c, \lambda) \in T : (c, \lambda) \text{ is a solution to the System (24)}\}, \tag{25}$$

so that $T_0 \neq \emptyset$. The minimizing points of the convex function $\phi_n$ are exactly the points of $T_0$; thus, $T_0$ is a convex compact subset of $T$, and we have shown the following.

PROPOSITION 5.1 *If $n \geq 3$ then for any given values of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, the system* (24) *is consistent, and the set of solutions, $T_0$, is a convex compact subset of T. Moreover, for any* $(c_0, \lambda_0) \in T_0$,

$$\phi_n(c, \lambda) \geq \phi_n(c_0, \lambda_0) \quad \textit{for all } (c, \lambda) \in T = \mathbb{R} \times (0, \infty),$$

*with equality if and only if* $(c, \lambda) \in T_0$.

We now proceed to show that $T_0$ is a singleton. Let as fix $c = c_1 \in \mathbb{R}$. For this particular value $c_1$, we consider the function

$$\psi_n(\lambda) := \phi_n(c_1, \lambda) = -(n-2)\lambda + \sum_{i=1}^{n} u_i(\lambda), \quad \lambda > 0,$$

where

$$u_i(\lambda) := \frac{1}{2}\lambda U\left(\frac{\mu_i - c_1}{\lambda}, \frac{\sigma_i}{\lambda}\right), \quad \lambda > 0 \ \ (i = 1, \ldots, n).$$

The function $u_i$ can be written more precisely as follows:

$$u_i(\lambda) = \begin{cases} \sqrt{(\mu_i - c_1)^2 + \sigma_i^2}, & 0 < \lambda \leq t_i, \\ \lambda + \dfrac{1}{4\lambda}[(\mu_i - c_1)^2 + \sigma_i^2], & t_i \leq \lambda < \gamma_i, \\ \dfrac{1}{2}\{|\mu_i - c_1| + \lambda + \sqrt{(|\mu_i - c_1| - \lambda)^2 + \sigma_i^2}\}, & \lambda \geq \gamma_i, \end{cases}$$

where $t_i = t_i(c_1)$ and $\gamma_i = \gamma_i(c_1)$ are given by

$$t_i := \frac{1}{2}\sqrt{(\mu_i - c_1)^2 + \sigma_i^2}, \quad \gamma_i := \frac{(\mu_i - c_1)^2 + \sigma_i^2}{2|\mu_i - c_1|}, \quad 0 < t_i < \gamma_i \leq \infty \ \ (i = 1, \ldots, n). \tag{26}$$

Each function $u_i$ is continuously differentiable with derivative

$$u_i'(\lambda) = \begin{cases} 0, & 0 < \lambda \leq t_i, \\ 1 - \dfrac{(\mu_i - c_1)^2 + \sigma_i^2}{4\lambda^2}, & t_i \leq \lambda < \gamma_i, \\ \dfrac{1}{2}\left(1 + \dfrac{\lambda - |\mu_i - c_1|}{\sqrt{(\lambda - |\mu_i - c_1|)^2 + \sigma_i^2}}\right), & \lambda \geq \gamma_i, \end{cases} \qquad i = 1, \ldots, n. \tag{27}$$

Obviously, $u_i(\lambda)$ is constant (equal to $2t_i$) in the interval $(0, t_i]$ and then it is strictly increasing; its non-decreasing continuous derivative $u_i'(\lambda)$ satisfies $0 \leq u_i'(\lambda) < 1$ for all $\lambda$, and $\lim_{\lambda \to \infty} u_i'(\lambda) = 1$. It follows that

$$\psi_n'(\lambda) = -(n-2) + \sum_{i=1}^{n} u_i'(\lambda)$$

is non-decreasing and, thus, $\psi_n$ is convex. Let $t_{1:n}, \ldots, t_{n:n}$ be the ordered values of $t_1, \ldots, t_n$. Noting that $n \geq 3$ and $0 < t_{1:n} \leq \cdots \leq t_{n:n} < \infty$, we see that $\psi_n'(\lambda) = -(n-2) < 0$ for $\lambda \leq t_{1:n}$, and the function $\psi_n$ is strictly decreasing in the interval $(0, t_{1:n}]$. Also, $\psi_n(\lambda)$ is strictly convex in the interval $(t_{1:n}, \infty)$, because $\psi_n'(\lambda)$ is strictly increasing in that interval. Observe that $\psi_n$ is eventually strictly increasing: $\lim_{\lambda\to\infty} \psi_n'(\lambda) = -(n-2) + \sum_{i=1}^n \lim_{\lambda\to\infty} u_i'(\lambda) = 2$. It follows that $\psi_n(\lambda)$ attains its minimum value at a unique point $\lambda = \lambda_1 > t_{1:n}$; clearly, $\lambda_1 = \lambda_1(c_1)$ is the unique solution to the equation $\psi_n'(\lambda) = 0$, $0 < \lambda < \infty$.

We can further verify that the unique solution, $\lambda = \lambda_1$, of $\sum_{i=1}^n u_i'(\lambda) = n-2$ lies in the interval $(t_{n-1:n}, \sum_{i=1}^n t_i)$. Indeed, first observe that if $\lambda \le t_{n-1:n}$, then we can find two indices $s \neq r$ with $\lambda \le t_s$ and $\lambda \le t_r$. Since $u_r'(\lambda) = u_s'(\lambda) = 0$, the sum $\sum_{i=1}^n u_i'(\lambda)$ contains at most $n-2$ strictly positive terms $u_i'(\lambda)$; from $u_i'(\lambda) < 1$ it follows that $\sum_{i=1}^n u_i'(\lambda) < n-2$. This shows that $\lambda_1 > t_{n-1:n}$. Next, we observe that $\lim_{\lambda\searrow 0} \psi_n(\lambda) = 2\sum_{i=1}^n t_i$. Thus, $\psi_n(\lambda_1) \le 2\sum_{i=1}^n t_i$ (because $\lambda = \lambda_1$ minimizes $\psi_n(\lambda)$). However, we know that $\phi_n(c, \lambda) > 2\lambda$ for all $(c, \lambda) \in T$, so that $2\sum_{i=1}^n t_i \ge \psi_n(\lambda_1) = \phi_n(c_1, \lambda_1) > 2\lambda_1$.

Hence, we have shown the following.

Lemma 5.1 *Let $n \ge 3$ and fix an arbitrary $c_1 \in \mathbb{R}$. The function $\psi_n : (0, \infty) \to (0, \infty)$, with $\psi_n(\lambda) := \phi_n(c_1, \lambda)$, attains its minimum value at a unique point $\lambda_1 = \lambda_1(c_1)$. The minimizing point $\lambda_1$ is the unique solution of the equation*

$$\sum_{i=1}^n u_i'(\lambda) = n-2, \quad t_{n-1:n} < \lambda < \sum_{i=1}^n t_i, \tag{28}$$

*where $t_i = t_i(c_1)$ are as in Equation (26), $0 < t_{1:n} \le \cdots \le t_{n:n}$ are the ordered values of $t_i$ in Equation (26), and the functions $u_i'(\lambda)$ are given by Equation (27).*

*Remark 5.1* Fix a point $(c_1, \lambda_1) \in T_0$ and define the following (possibly empty) sets of indices:

$$\begin{aligned}
I_1 &:= \{i \in \{1, \ldots, n\} : (\mu_i - c_1)^2 + \sigma_i^2 \ge 4\lambda_1^2\} = \{i : \lambda_1 \le t_i\},\\
I_2 &:= \{i \in \{1, \ldots, n\} : 2\lambda_1|\mu_i - c_1| < (\mu_i - c_1)^2 + \sigma_i^2 < 4\lambda_1^2\} = \{i : t_i < \lambda_1 < \gamma_i\},\\
I_3 &:= \{i \in \{1, \ldots, n\} : (\mu_i - c_1)^2 + \sigma_i^2 \le 2\lambda_1(\mu_i - c_1)\} = \{i : \lambda_1 \ge \gamma_i \text{ and } \mu_i > c_1\},\\
I_4 &:= \{i \in \{1, \ldots, n\} : (\mu_i - c_1)^2 + \sigma_i^2 \le 2\lambda_1(c_1 - \mu_i)\} = \{i : \lambda_1 \ge \gamma_i \text{ and } \mu_i < c_1\}.
\end{aligned} \tag{29}$$

By definition, $I_i \cap I_j = \emptyset$ for $i \neq j$ and $I_1 \cup I_2 \cup I_3 \cup I_4 = \{1, \ldots, n\}$. Since $(c_1, \lambda_1) \in T_0$ it follows that $\lambda_1$ must solve Equation (28) (for this particular value of $c_1$), that is,

$$\sum_{i\in I_2} \left\{1 - \frac{(\mu_i - c_1)^2 + \sigma_i^2}{4\lambda_1^2}\right\} + \sum_{i\in I_3\cup I_4} \frac{1}{2}\left\{1 + \frac{\lambda_1 - |\mu_i - c_1|}{\sqrt{(\lambda_1 - |\mu_i - c_1|)^2 + \sigma_i^2}}\right\} = n-2,$$

where an empty sum should be treated as zero. Observe that all summands are (strictly positive and) strictly less than 1; thus, $N(I_2) + N(I_3) + N(I_4) \ge n-1$, and it follows that $N(I_1) \le 1$, where $N(I)$ denotes the cardinality of $I$. Furthermore, $(c, \lambda) = (c_1, \lambda_1)$ is a solution to $(\partial/\partial c)\phi_n(c, \lambda) = 0$. Using $(\partial/\partial c)\phi_n(c, \lambda) = -\frac{1}{2}\sum_{i=1}^n U_1((\mu_i - c)/\lambda, \sigma_i/\lambda)$ and the explicit form of $U_1$, given by Equation (A1), we obtain

$$\begin{aligned}
&\sum_{i\in I_1} \frac{\mu_i - c_1}{\sqrt{(\mu_i - c_1)^2 + \sigma_i^2}} + \sum_{i\in I_2} \frac{\mu_i - c_1}{2\lambda_1}\\
&\quad + \sum_{i\in I_3\cup I_4} \frac{\operatorname{sign}(\mu_i - c_1)}{2}\left\{1 + \frac{|\mu_i - c_1| - \lambda_1}{\sqrt{(|\mu_i - c_1| - \lambda_1)^2 + \sigma_i^2}}\right\} = 0.
\end{aligned} \tag{30}$$

This equality shows that $N(I_3) \le n-1$ and $N(I_4) \le n-1$; for if, e.g. $N(I_3) = n$ then we would have $I_1 = I_2 = I_4 = \emptyset$ and, since $\mu_i > c_1$ whenever $i \in I_3$, the above equation leads to the (obviously impossible) relation

$$\sum_{i=1}^{n} \frac{1}{2}\left\{1 + \frac{(\mu_i - c_1) - \lambda_1}{\sqrt{((\mu_i - c_1) - \lambda_1)^2 + \sigma_i^2}}\right\} = 0.$$

We have thus concluded the following key-property of a minimizing point:

$$\text{If } (c_1, \lambda_1) \in T_0 \text{ then } \max\{N(I_3), N(I_4)\} \le n-1 \text{ and } N(I_1) \le 1. \tag{31}$$

Most cases suggested by Relation (31) may appear for some values of $\boldsymbol{\mu}, \boldsymbol{\sigma}$ (one of the rare exceptions is $N(I_1) = N(I_2) = 0$, $\max\{N(I_3), N(I_4)\} = n-1$). Note that Theorem 3.1 is, in fact, concerned with the particular situation where $N(I_2) = n$ (thus, $N(I_1) = N(I_3) = N(I_4) = 0$). It is, essentially, the unique situation in which the $\mathrm{AG}_n$ bound is tight (plus boundary subcases). Due to Equation (31), it seems that this particular (but plausible) case is quite restricted.

Behind the tedious calculations, the rough meaning of the argument that led to Equation (31) is the following: For a particular $(c, \lambda)$ to be optimal (i.e. to minimize $\phi_n$) it is necessary that $c$ is not 'too far away' from the $\mu_i$'s and $\lambda$ is not 'too small' or 'too large' compared to $\frac{1}{2}\sum_{i=1}^{n} \sigma_i$. In particular, Equation (30) shows that an optimal $c$ can never lie outside the interval $[\min_i\{\mu_i\}, \max_i\{\mu_i\}]$, and it is located in an interior point when the $\mu_i$'s are not all equal; of course this fact is intuitively obvious.

Lemma 5.2 *If the set $T_0$ of Equation* (25) *contains two different elements, then it must be a compact line segment which is not parallel to the $\lambda$-axis. That is, $T_0$ has to be of the form $T_0 = [\boldsymbol{x}, \boldsymbol{y}] = \{\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}), 0 \le t \le 1\}$, for some $\boldsymbol{x} = (c_1, \lambda_1) \in T$ and $\boldsymbol{y} = (c_2, \lambda_2) \in T$ with $c_1 \ne c_2$.*

Lemma 5.3 *Let $A_0 = (c_0, \lambda_0) \ne A_1 = (c_1, \lambda_1)$ be two points in $T$. Fix $\mu \in \mathbb{R}$, $\sigma > 0$ and consider the points $B_0 = ((\mu - c_0)/\lambda_0, \sigma/\lambda_0) \in T$, $B_1 = ((\mu - c_1)/\lambda_1, \sigma/\lambda_1) \in T$. Let $A = (c, \lambda)$ and $B = ((\mu - c)/\lambda, \sigma/\lambda)$. As the point $A$ is moving linearly in the line segment* $[A_0, A_1]$ (*from $A_0$ to $A_1$*), *the point $B = B(A)$ is moving continuously in the line segment* $[B_0, B_1]$ (*from $B_0$ to $B_1$*).

We are now ready to state the conclusion of the present section.

Theorem 5.1 *If $n \ge 3$ then for any given values of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, there exists a unique solution $(c, \lambda) = (c_0, \lambda_0)$ of the System* (24), *and*

$$\phi_n(c, \lambda) \ge \phi_n(c_0, \lambda_0) \quad \text{for all } (c, \lambda) \in \mathbb{R} \times (0, \infty),$$

*with equality if and only if $(c, \lambda) = (c_0, \lambda_0)$.*

*Remark 5.2* The AG-bound is general because it is based on the universal inequality $|x-1| + |x+1| \le 2 + \frac{1}{2}x^2$ (in contrast to the other ones appearing in the proof of Lemma 2.2), it is sharp if $n_2 = N(I_2) = n$, and then always $(c_0, \lambda_0) = (\bar{\mu}, \frac{1}{4}\mathrm{AG}_n)$. Then conditions (16) immediately follow from these observations.

*Remark 5.3* For $n = 2$, Theorem 5.1 (as well as several conclusions of the present section) is no longer true. It is again true that the convex function $\phi_2(c, \lambda)$ attains its minimum value, $\rho_2 = \sqrt{(\mu_2 - \mu_1)^2 + (\sigma_1 + \sigma_2)^2}$, at the solutions of the system (24), but now $T_0$ is not a singleton: it contains points arbitrarily close to the boundary of the domain of $\phi_2$. More precisely, one can verify that for $n=2$, the exact set of minimizing points is the line segment $T_0=\{(c_0,\lambda_0); 0<\lambda_0\le\lambda^*\}$, where

$$c_0=\frac{\sigma_1}{\sigma_1+\sigma_2}\mu_2+\frac{\sigma_2}{\sigma_1+\sigma_2}\mu_1, \quad \lambda^*=\frac{\rho_2}{2(\sigma_1+\sigma_2)}\min\{\sigma_1,\sigma_2\}.$$

However, the set $\mathcal{E}_2(\mu_1,\mu_2,\sigma_1,\sigma_2)$ is a singleton, and this fact can be seen directly (see Section 7). Also, it is worth pointing out that, for $n=2$, $N(I_1)=2=n$ (compare with Equation (31)).

## 6. Tightness and characterization of extremal random vectors

Let $n\ge 3$, $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ be fixed (with $0<\sigma_i<\infty$ for all $i$). Let $(c_0,\lambda_0)$ be the unique solution of the System (24). With the help of $(c_0,\lambda_0)$, we shall give a complete description of the set $\mathcal{E}_n(\boldsymbol{\mu},\boldsymbol{\sigma})$ of extremal random vectors in $\mathcal{F}_n(\boldsymbol{\mu},\boldsymbol{\sigma})$. These are the random vectors $\boldsymbol{X}$ satisfying $\mathbb{E}\boldsymbol{X}=\boldsymbol{\mu}$, $\operatorname{Var}\boldsymbol{X}=\boldsymbol{\sigma}^2$ and $\mathbb{E}R_n(\boldsymbol{X})=\rho_n=\rho_n(\boldsymbol{\mu},\boldsymbol{\sigma})$, where

$$\rho_n:=\phi_n(c_0,\lambda_0)=-(n-2)\lambda_0+\frac{\lambda_0}{2}\sum_{i=1}^n U\left(\frac{\mu_i-c_0}{\lambda_0},\frac{\sigma_i}{\lambda_0}\right); \tag{32}$$

recall that $U(\cdot,\cdot)$ is given by Equation (11). The construction, though more complicated, follows parallel arguments as for the attainability of the AG$_n$ bound (Theorem 3.1).

We start by considering the partition $I_1,\ldots,I_4$ of $\{1,\ldots,n\}$ as in Equation (29), and the corresponding cardinalities $n_1,\ldots,n_4$. The main difference from Remark 5.1 is that, now, each $I_j$ has been stabilized, because $(c_0,\lambda_0)$ is unique; thus, one has to substitute $c_1=c_0$ and $\lambda_1=\lambda_0$ in Equation (29). Clearly some of the sets $I_j$ may be empty; then $n_j=0$. The situation with all $I_j$ being nonempty may also appear; this is the case, e.g. for $\boldsymbol{\mu}=(4,0,4,0)$, $\boldsymbol{\sigma}=(10,5,1,1)$. From Remark 5.1 (see Equation (31)), we know that $n_1$ $n_2$, $n_3$, $n_4$ (with $n_j\ge 0$, $\sum n_j=n$) cannot be completely arbitrary; they have to satisfy the restrictions:

$$n_3=N(I_3)\le n-1, \quad n_4=N(I_4)\le n-1, \quad n_1=N(I_1)\le 1. \tag{33}$$

Other impossible cases are given by $n_3=1, n_4=n-1$ and $n_3=n-1, n_4=1$; this is a by-product of Lemma 6.1 .

For notational simplicity, it is helpful to consider the following numbers $\xi_i$, $\theta_i$:

$$\xi_i:=\begin{cases}\mu_i-c_0, & i\in I_1\cup I_2;\\ \mu_i-c_0-\lambda_0, & i\in I_3;\\ c_0-\mu_i-\lambda_0, & i\in I_4;\end{cases} \qquad \theta_i:=\sqrt{\xi_i^2+\sigma_i^2}, \quad i=1,\ldots,n. \tag{34}$$

We note that $|\xi_i|<\theta_i$ for all $i$ and $2\lambda_0|\xi_i|<\theta_i^2<4\lambda_0^2$ for all $i\in I_2$ (if any). Following Corollary 2.1, we define the probabilities

$$p_i^-:=\frac{1}{2}\left(1-\frac{\xi_i}{\theta_i}\right), \quad p_i^o:=0, \quad p_i^+:=\frac{1}{2}\left(1+\frac{\xi_i}{\theta_i}\right), \quad i\in I_1;$$

$$p_i^-:=\frac{1}{8\lambda_0^2}\left[\theta_i^2-2\lambda_0\xi_i\right], \quad p_i^o:=1-\frac{\theta_i^2}{4\lambda_0^2}, \quad p_i^+:=\frac{1}{8\lambda_0^2}\left[\theta_i^2+2\lambda_0\xi_i\right], \quad i\in I_2; \tag{35}$$

$$p_i^- := 0, \quad p_i^o := \frac{1}{2}\left(1 - \frac{\xi_i}{\theta_i}\right), \quad p_i^+ := \frac{1}{2}\left(1 + \frac{\xi_i}{\theta_i}\right), \quad i \in I_3;$$
$$p_i^- := \frac{1}{2}\left(1 + \frac{\xi_i}{\theta_i}\right), \quad p_i^o := \frac{1}{2}\left(1 - \frac{\xi_i}{\theta_i}\right), \quad p_i^+ := 0, \quad i \in I_4,$$

and the corresponding (univariate) supporting points

$$\begin{aligned} x_i^- &:= c_0 - \theta_i, \quad x_i^+ := c_0 + \theta_i, \quad i \in I_1; \\ x_i^- &:= c_0 - 2\lambda_0, \quad x_i^o := c_0, \quad x_i^+ := c_0 + 2\lambda_0, \quad i \in I_2; \\ x_i^o &:= c_0 + \lambda_0 - \theta_i, \quad x_i^+ := c_0 + \lambda_0 + \theta_i, \quad i \in I_3; \\ x_i^- &:= c_0 - \lambda_0 - \theta_i, \quad x_i^o := c_0 - \lambda_0 + \theta_i, \quad i \in I_4. \end{aligned} \tag{36}$$

By definition, each $\boldsymbol{p}_i := (p_i^-, p_i^o, p_i^+)$ is a probability vector. Clearly, one could assign an arbitrary value to a missing point, since its corresponding probability is 0. The most convenient choice is to assign the respective values $c_0 - 2\lambda_0$, $c_0$, $c_0 + 2\lambda_0$, whenever $x_i^-$, $x_i^o$, $x_i^+$ is not specified from Equation (36). With this convention,

$$x_i^- < c_0 - \lambda_0 < x_i^o < c_0 + \lambda_0 < x_i^+, \quad i = 1, \dots, n. \tag{37}$$

Let $X_i$ be a random variable which assumes values $x_i^-, x_i^o, x_i^+$ with respective probabilities $p_i^-, p_i^o, p_i^+$. Corollary 2.1 asserts that (the distribution of) $X_i$ is characterized be the fact that maximizes the expectation of $|(X - c_0) - \lambda_0| + |(X - c_0) + \lambda_0|$ as $X$ varies in $\mathcal{F}_1(\mu_i, \sigma_i)$.

The following lemma provides the most fundamental tool for the main result.

Lemma 6.1 *The probabilities $p_i^+$, $p_i^-$ in Equation* (35) *satisfy the relation*

$$\sum_{i=1}^n p_i^+ = \sum_{i=1}^n p_i^- = 1. \tag{38}$$

Lemma 6.1 enables us to define the $n$-variate probability vectors

$$\boldsymbol{p}^+ = (p_1^+, \dots, p_n^+), \quad \boldsymbol{p}^- = (p_1^-, \dots, p_n^-). \tag{39}$$

By definition, $\boldsymbol{p}^+$ has its zero elements at exactly the positions $i$ where $i \in I_4$ (if $I_4 = \emptyset$, all $p_i^+$'s are positive), and $\boldsymbol{p}^-$ has its zero elements at exactly the positions $i$ where $i \in I_3$ (if any).

Proposition 6.1 *Assume we are given $n \geq 3$, $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$. Then,* (i) *and* (ii) *are equivalent:*

(i) *We can find a random vector $\boldsymbol{X} \in \mathcal{F}_n(\boldsymbol{\mu}, \boldsymbol{\sigma})$ such that $\mathbb{E}R_n = \rho_n$.*
(ii) *There exists an $n \times n$ probability matrix $Q \in \mathcal{M}(\boldsymbol{p}^+, \boldsymbol{p}^-)$ such that $q_{ii} = 0$ for all $i \in \{1, \dots, n\}$.*

*Moreover, with $\mathcal{L}(\boldsymbol{X})$ denoting the probability law of the random vector $\boldsymbol{X}$, the correspondence $\mathcal{L}(\boldsymbol{X}) \rightleftarrows Q$ is a bijection; the explicit formula for the transformation $Q = (q_{ij}) \mapsto \mathcal{L}(\boldsymbol{X})$ is given by*

$$\begin{aligned} &\mathbb{P}[\boldsymbol{X} = \boldsymbol{x}_{ij}] = q_{ij}, \quad \text{where } \boldsymbol{x}_{ij} := (x_1^o, \dots, x_{i-1}^o, x_i^+, x_{i+1}^o, \dots, x_{j-1}^o, x_j^-, x_{j+1}^o, \dots, x_n^o), \\ &\qquad i \neq j, \;\; i,j = 1, \dots, n. \end{aligned} \tag{40}$$

The main result of the present work reads as follows:

THEOREM 6.1 *Let $n \geq 3$, $\mu_i \in \mathbb{R}$, $\sigma_i > 0$ $(i = 1, \dots, n)$. Then,*
(a)

$$\sup \mathbb{E}R_n = \rho_n, \tag{41}$$

*where the supremum is taken over $\boldsymbol{X} \in \mathcal{F}_n(\boldsymbol{\mu}, \boldsymbol{\sigma})$ and $\rho_n = \rho_n(\boldsymbol{\mu}, \boldsymbol{\sigma})$ is given by Equation* (32), *with $(c_0, \lambda_0) = (c_0(\boldsymbol{\mu}, \boldsymbol{\sigma}), \lambda_0(\boldsymbol{\mu}, \boldsymbol{\sigma}))$ being the unique solution to the system of Equations* (24).

(b) *The set $\mathcal{E}_n(\boldsymbol{\mu}, \boldsymbol{\sigma})$ is nonempty. Any extremal $\boldsymbol{X} \in \mathcal{E}_n(\boldsymbol{\mu}, \boldsymbol{\sigma})$ is produced by Equation* (40), *with $x_i^-$, $x_i^o$, $x_i^+$ as in Equation* (36), *and corresponds uniquely to an $n \times n$ probability matrix $Q \in \mathcal{M}(\boldsymbol{p}^+, \boldsymbol{p}^-)$ with zero diagonal entries, where $\boldsymbol{p}^+, \boldsymbol{p}^-$ are given by Equation* (39).

*Proof* From Theorem 2.1 we know that $\mathbb{E}R_n \leq \rho_n$ and it suffices to prove (b). In view of Proposition 6.1, it remains to verify that the class of $n \times n$ probability matrices with zero diagonal entries and marginals $\boldsymbol{p}^+, \boldsymbol{p}^-$ is nonempty. However, this fact follows immediately from Lemma 3.1, because $\max_i\{p_i^+ + p_i^-\} \leq 1$ (see Equation (35)), and the proof is complete. ∎

*Remark 6.1* Since $\mathbb{E}R_n(\boldsymbol{X}) = \rho_n$ for any $\boldsymbol{X} \in \mathcal{E}_n(\boldsymbol{\mu}, \boldsymbol{\sigma})$,

$$\rho_n = \sum_{i=1}^{n} \{(x_i^+ - c_0)p_i^+ + (c_0 - x_i^-)p_i^-\}.$$

COROLLARY 6.1 *If $I_1 \neq \emptyset$ (see Equation* (29)) *then $I_1 = \{k\}$ for some $k \in \{1, \dots, n\}$, and the equality in Inequality* (41) *characterizes the random vector $\boldsymbol{X}$ with probability law*

$$\mathbb{P}[\boldsymbol{X} = \boldsymbol{x}_{ik}] = p_i^+, \quad \mathbb{P}[\boldsymbol{X} = \boldsymbol{x}_{ki}] = p_i^-, \quad i \neq k, \ \ i = 1, \dots, n. \tag{42}$$

*Proof* From Equation (33) we know that $N(I_1) \leq 1$, and thus, $I_1 = \{k\}$ for some $k$. Since $k \in I_1$, Equation (35) shows that $\max_i\{p_i^+ + p_i^-\} = p_k^+ + p_k^- = 1$. [Note that, by Lemma 6.1, $\sum_{i \neq k} p_i^- + \sum_{i \neq k} p_i^+ = (1 - p_k^-) + (1 - p_k^+) = 1$ and, hence, Equation (42) defines a probability law.] Lemma 3.1 implies uniqueness of $Q$, hence of $\mathcal{L}(\boldsymbol{X})$ (see Equation (40)). It is easily seen that the matrix $Q$, obtained by Equations (42) through (A5), is indeed the unique probability matrix with vanishing diagonal entries and marginals $\boldsymbol{p}^+, \boldsymbol{p}^-$. ∎

Corollary 6.1 implies uniqueness (denoted by (U)) for the second counterpart of the bound (19) in Example 3.2. It should be noted that the converse of Corollary 6.1 does not hold; that is, the condition $I_1 \neq \emptyset$ is not necessary for concluding uniqueness of the extremal random vector $\boldsymbol{X}$. A particular example was given by Remark 3.1.

Clearly, the most interesting situations in practice arise when $I_1 = \emptyset$. In such cases, it is fairly expected that there will be infinitely many extremal vectors, as in Theorem 3.1. This is, indeed, true in general, but not always. Lemma 3.1 guarantees infiniteness (denoted by (I)) only if all $p_i^+, p_i^-$ are nonzero, and this corresponds to the quite restricted case where $I_2 = \{1, \dots, n\}$. Of course, given the existence of two extremal vectors, one can deduce (I) by considering convex combinations of the corresponding matrices; cf. Example 3.1. If $I_1 = \emptyset$, the complete distinction between (U) and (I) depends upon the values of $n$, $n_3 = N(I_3)$ and $n_4 = N(I_4)$ (see Equations (29) and (33)); and if $n_3 = n_4 = 0$ we already know that (I) results.

We briefly discuss all remaining situations where $I_1 = \emptyset$: If $n_2 = N(I_2) = 0$ and $n_3 \geq 2, n_4 \geq 2$, it is obvious that (I) holds; note that $n_3 = 1, n_4 = n - 1$ and $n_3 = n - 1, n_4 = 1$ are impossible by Lemma 6.1. If $n_2 = n_3 = n_4 = 1$ or $n_2 = 2, n_3 = 1, n_4 = 0$ or $n_2 = 2, n_3 = 0, n_4 = 1$ then we are in (U), while (I) results if $n_2 = n_3 = 1, n_4 \geq 2$ or $n_2 = n_4 = 1, n_3 \geq 2$. If $n_2 = 1, n_3 \geq 2, n_4 \geq 2$ then we get (I), as well as in all remaining cases where $n_2 \geq 2, n_3 \geq 0, n_4 \geq 0$.

The final conclusion is as follows: If $I_1 = \emptyset$, the situations where the extremal distribution is uniquely defined are described by $n_2 = n_3 = n_4 = 1$ or $n_2 = 2, n_3 = 1, n_4 = 0$ or $n_2 = 2, n_3 = 0, n_4 = 1$ (and thus, $n = 3$); this provides an explanation to Remark 3.1. However, we note that knowledge of the values $n_j$ actually requires knowledge of the region where the optimal $(c_0, \lambda_0)$ appears, and this may be, or may not be, an easy task for particular $\boldsymbol{\mu}, \boldsymbol{\sigma}$.

*Remark 6.2* The range $R_n(\boldsymbol{X})$ of an extremal vector $\boldsymbol{X}$ need not be a degenerate random variable. An example is provided by $\boldsymbol{\mu} = (-2, 0, 2)$, $\boldsymbol{\sigma} = (1, 3, 1)$. Then, $n_1 = 0$, $n_2 = n_3 = n_4 = 1$ and it can be shown that

$$\lambda_0 \approx 1.737, \quad \rho_3 = \frac{64\lambda_0^3 - 72\lambda_0 - 81}{4\lambda_0(4\lambda_0^2 - 9)} \approx 6.066$$

($\lambda_0$ is the unique solution of $4\lambda^2(2-\lambda) = (4\lambda^2 - 9)\sqrt{\lambda^2 - 4\lambda + 5}$, and this reduces to a four-degree polynomial equation). The range $R_3$ of the unique extremal vector assumes values $2\lambda_0 + \lambda_0(8\lambda_0 - 9)/(4\lambda_0^2 - 9) \approx 5.542$ and $2\lambda_0(8\lambda_0 - 9)/(4\lambda_0^2 - 9) \approx 6.245$ with respective probabilities $9/4\lambda_0^2 \approx .254$ and $1 - 9/4\lambda_0^2 \approx .746$. However, the improvement over the bound $\mathrm{AG}_3 = \sqrt{38} \approx 6.164$ is negligible. As a general observation, even for small $n$, the value of $\rho_n$ is difficult to evaluate when more than two index sets $I_j$ are nonempty.

*Example 6.1 Homoscedastic observations from two balanced groups.* Let $n = 2k$, $\sigma_i^2 = \sigma^2$ and $\mu_i = -\mu$ or $\mu$ according to $i \le k$ or $i > k$, respectively ($\mu \ge 0$). The Arnold–Groeneveld bound (4) takes here the form

$$\mathbb{E}R_{2k} \le \mathrm{AG}_{2k} = 2\sqrt{k(\mu^2 + \sigma^2)},$$

and it is tight if $\mu \le \sigma/\sqrt{k-1}$ (in particular, if $n = 2$ or $\mu = 0$). Also, we know from Theorem 3.1, the nature of the random vectors that attain the equality. However, for $\mu \ge \sigma/\sqrt{k-1}$ one finds $N(I_3) = N(I_4) = k$, and the tight bound of Theorem 6.1 becomes

$$\mathbb{E}R_{2k} \le \rho_{2k} = 2\mu + 2\sigma\sqrt{k-1} \quad \left(\mu \ge \frac{\sigma}{\sqrt{k-1}}\right);$$

note that $\rho_{2k}$ is equal to $\mathrm{AG}_{2k}$ only in the boundary case $\sigma = \mu\sqrt{k-1}$. For $\mu \ge \sigma/\sqrt{k-1}$ the nature of extremal random vectors is different: They assume values

$$\boldsymbol{y}_{ij} = (-x, \ldots, -x, -y, -x, \ldots, -x;\ x, \ldots, x, y, x, \ldots, x), \quad i, j = 1, \ldots, k,$$

where $-y$ is located at the $i$th place and $y$ is located at the $(k+j)$th place of the vector. Here, $0 \le x = \mu - \sigma/\sqrt{k-1} < y = \mu + \sigma\sqrt{k-1}$. The respective probabilities $p_{ij} = \mathbb{P}[\boldsymbol{X} = \boldsymbol{y}_{ij}]$, $i, j = 1, \ldots, k$, correspond to a probability matrix $P_{k\times k}$ with uniform marginals. Both limits

$$\lim_{\mu\to\infty} \frac{\rho_{2k}}{\mathrm{AG}_{2k}} = \frac{1}{\sqrt{k}} \quad (k, \sigma \text{ fixed}), \qquad \lim_{k\to\infty} \frac{\rho_{2k}}{\mathrm{AG}_{2k}} = \frac{\sigma}{\sqrt{\mu^2 + \sigma^2}} \quad (\mu, \sigma \text{ fixed})$$

show that, under some circumstances, the improvement that is achieved by using $\rho_n$ instead of $\mathrm{AG}_n$ can become arbitrarily large.

*Example 6.2 Homoscedastic data with a single outlier.* Let $\sigma_i^2 = \sigma^2$ for all $i$, $\mu_i = 0$ ($i = 1, \ldots, n-1$) and $\mu_n = \mu \geq 0$. Theorem 3.1 asserts that the bound

$$\mathbb{E}R_n \leq \mathrm{AG}_n = \sqrt{2\frac{n-1}{n}\mu^2 + 2n\sigma^2}$$

is not tight for $n \geq 3$ and $\mu > n/\sqrt{n-1}\sigma$. In this case we have $I_1 = \{n\}$, $I_2 = \{1, \ldots, n-1\}$, so that $n_1 = 1$, $n_2 = n-1$, and the tight bound has the form

$$\mathbb{E}R_n \leq \rho_n = \sqrt{(n-1)(c_0^2 + \sigma^2)} + \sqrt{(\mu - c_0)^2 + \sigma^2},$$

where $c_0$ is the unique root of the equation

$$\frac{c\sqrt{n-1}}{\sqrt{c^2+\sigma^2}} = \frac{\mu - c}{\sqrt{(\mu-c)^2+\sigma^2}}, \quad 0 < c < \min\left\{\frac{\mu}{n}, \frac{\sigma}{\sqrt{n-2}}\right\}.$$

[It can be checked that $\lambda_0 = (\sqrt{n-1}/2)\sqrt{c_0^2+\sigma^2}$.] Although $\rho_n < \mathrm{AG}_n$ (for $\mu\sqrt{n-1} > n\sigma$), it is not easy to make direct comparisons. However, the relations $c_0^2 < \sigma^2/(n-2)$ and $(\mu - c_0)^2 < \mu^2$ imply that $\rho_n < \rho_n' := ((n-1)/\sqrt{n-2})\sigma + \sqrt{\mu^2+\sigma^2}$. Hence, for the (non-tight) upper bound $\rho_n'$,

$$\lim_{\mu\to\infty} \frac{\rho_n'}{\mathrm{AG}_n} = \frac{\sqrt{n}}{\sqrt{2n-2}} \quad (n \geq 3,\ n, \sigma \text{ fixed}).$$

*Remark 6.3* Example 6.2 and Remark 6.2 entail that $\rho_n$ may have a rather complicated form when the $\mu_i$'s are not all equal. On the other hand, $\rho_n$ becomes quite plausible in the case of equal $\mu_i$'s; see Example 3.2. This particular case is useful in concluding some facts about the behaviour of $\rho_n$ in general. Indeed, taking into account the obvious relation $U(x,y) \geq U(0,y)$, we see that for any given $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$,

$$\rho_n = \phi_n(c_0, \lambda_0) \geq -(n-2)\lambda_0 + \frac{\lambda_0}{2}\sum_{i=1}^n U\left(0, \frac{\sigma_i}{\lambda_0}\right) = \hat{\phi}_n(0, \lambda_0) \geq \hat{\rho}_n := \inf_{x\in\mathbb{R}, y>0} \hat{\phi}_n(x,y),$$

where $\hat{\rho}_n$ is the upper bound of Theorem 2.1, calculated under $\mu_i = \mu$ for all $i$, and for the given $\boldsymbol{\sigma}$. Since $\hat{\rho}_n = \min_{y>0} \hat{\phi}_n(0,y)$ admits a simple closed form, see (19), we get the following lower bound:

$$\rho_n \geq \hat{\rho}_n = \begin{cases} \sqrt{2\sum_{i=1}^n \sigma_i^2} & \text{if } 2\max_i\{\sigma_i^2\} \leq \sum_{i=1}^n \sigma_i^2, \\ \max_i\{\sigma_i\} + \sqrt{\sum_{i=1}^n \sigma_i^2 - \max_i\{\sigma_i^2\}} & \text{if } 2\max_i\{\sigma_i^2\} \geq \sum_{i=1}^n \sigma_i^2, \end{cases} \quad \text{any } \boldsymbol{\mu}, \boldsymbol{\sigma}.$$

Since $U(x,y) > U(0,y)$ for $x \neq 0$, the equality holds only if all the $\mu_i$'s are equal. Despite its weakness, this lower bound provides an idea of what can be expected for the actual size of $\rho_n$. It is also helpful in giving some light to the observation that, provided the means are small compared to the variances, the $\mathrm{AG}_n$ bound tends to be tight. More precisely, assume that $\min_i\{\sigma_i^2\} \to \infty$ and $(\sum_{i=1}^n(\mu_i - \bar{\mu})^2)/(\sum_{i=1}^n \sigma_i^2) \to 0$ (in particular, $\max_i |\mu_i - \bar{\mu}| \leq C < \infty$ suffices for this). Then, the homogeneity assumption $\max_i\{\sigma_i^2\} \leq (n-1)\min_i\{\sigma_i^2\}$ is sufficient for the asymptotic tightness of the $\mathrm{AG}_n$ bound (for fixed $n \geq 3$). Indeed, from this assumption we get $\sum_{i=1}^n \sigma_i^2 \geq \max_i\{\sigma_i^2\} + (n-1)\min_i\{\sigma_i^2\} \geq 2\max\{\sigma_i^2\}$, and thus, $\hat{\rho}_n = \sqrt{2\sum_{i=1}^n \sigma_i^2}$. Hence,

$$1 \geq \left(\frac{\rho_n}{\mathrm{AG}_n}\right)^2 \geq \left(\frac{\hat{\rho}_n}{\mathrm{AG}_n}\right)^2 = \frac{\sum_{i=1}^n \sigma_i^2}{\sum_{i=1}^n\{(\mu_i-\bar{\mu})^2+\sigma_i^2\}} = \frac{1}{1+(\sum_{i=1}^n(\mu_i-\bar{\mu})^2)/(\sum_{i=1}^n\sigma_i^2)} \to 1.$$

Therefore, under the above circumstances, the improvement achieved by using $\rho_n$ instead of $\mathrm{AG}_n$ becomes negligible.

## 7. The case $n = 2$ and further remarks

For $n = 2$, $c \in \mathbb{R}$ and $0 < \lambda \leq \lambda^*(c) := \frac{1}{2}\min\{\sqrt{(\mu_1-c)^2+\sigma_1^2}, \sqrt{(\mu_2-c)^2+\sigma_2^2}\}$, it is easily seen that

$$\phi_2(c,\lambda) = \sqrt{(\mu_1-c)^2+\sigma_1^2} + \sqrt{(\mu_2-c)^2+\sigma_2^2}.$$

The last quantity attains its minimum value at the unique solution of the equation $(\partial/\partial c)\phi_2(c,\lambda) = 0$, i.e. at $c = c_0 = (\sigma_1/(\sigma_1+\sigma_2))\mu_2 + (\sigma_2/(\sigma_1+\sigma_2))\mu_1$. Thus, the bound $\rho_2$ admits a closed form. More precisely, Theorem 2.1 shows that

$$\mathbb{E}R_2 \leq \rho_2, \quad \text{where } \rho_2 := \inf_{c\in\mathbb{R},\lambda>0} \phi_2(c,\lambda) = \sqrt{(\mu_1-\mu_2)^2+(\sigma_1+\sigma_2)^2}. \tag{43}$$

Observe that $\lambda^*(c_0) = (\rho_2/2(\sigma_1+\sigma_2))\min\{\sigma_1,\sigma_2\}$ and that $\phi_2(c_0,\lambda)$ assumes the constant value $\rho_2$ for all $\lambda \in (0, \lambda^*(c_0)]$. It follows that the set of minimizing points form the line segment given in Remark 5.3. Clearly, $N(I_1) = 2 = n$, in contrast to (31) which guaranties that $N(I_1) \leq 1$ for every $n \geq 3$.

The inequality (43) is tight, since the equality is attained by (and characterizes) the random pair $(X_1, X_2)$ with distribution given by

$$\begin{aligned}
&\mathbb{P}\left[X_1 = \frac{\sigma_2\mu_1+\sigma_1\mu_2}{\sigma_1+\sigma_2} + \frac{\sigma_1}{\sigma_1+\sigma_2}\rho_2,\ X_2 = \frac{\sigma_2\mu_1+\sigma_1\mu_2}{\sigma_1+\sigma_2} - \frac{\sigma_2}{\sigma_1+\sigma_2}\rho_2\right] = \frac{1}{2}\left(1+\frac{\mu_1-\mu_2}{\rho_2}\right),\\
&\mathbb{P}\left[X_1 = \frac{\sigma_2\mu_1+\sigma_1\mu_2}{\sigma_1+\sigma_2} - \frac{\sigma_1}{\sigma_1+\sigma_2}\rho_2,\ X_2 = \frac{\sigma_2\mu_1+\sigma_1\mu_2}{\sigma_1+\sigma_2} + \frac{\sigma_2}{\sigma_1+\sigma_2}\rho_2\right] = \frac{1}{2}\left(1+\frac{\mu_2-\mu_1}{\rho_2}\right).
\end{aligned} \tag{44}$$

Therefore, $\mathcal{E}_2(\boldsymbol{\mu},\boldsymbol{\sigma})$ is a singleton. Also, $\mathrm{AG}_2 = \sqrt{(\mu_1-\mu_2)^2+2\sigma_1^2+2\sigma_2^2}$, and it is worth pointing out that the bound $\mathrm{AG}_2$ is tight if and only if $\sigma_1 = \sigma_2$. Another observation is that the extremal random vector for the expected range coincides with the (unique) extremal random vector for the expected maximum (see Inequality (5)). However, this is not a coincidence. In view of the obvious relationship

$$R_2 = |X_1 - X_2| = 2\max\{X_1, X_2\} - X_1 - X_2 = 2X_{2:2} - X_1 - X_2, \tag{45}$$

a bound for the maximum can be translated to a bound for the range, and vice-versa (provided that the expectations, $\mu_1, \mu_2$, of $X_1, X_2$, are known). In this sense, the bound $\rho_2$ turns to be a particular case of the results given by Bertsimas et al.,[16,17] namely

$$\rho_2 = \sup \mathbb{E}R_2 = \sup \mathbb{E}\{2X_{2:2} - X_1 - X_2\} = 2\sup \mathbb{E}X_{2:2} - \mu_1 - \mu_2 = 2\mathrm{BNT}_2 - \mu_1 - \mu_2,$$

and the equality characterizes the same extremal distribution as for the maximum. Consequently, it is of some interest to observe that the bound $\mathrm{BNT}_2$ admits a closed form, namely

$$\mathrm{BNT}_2 = \tfrac{1}{2}(\mu_1 + \mu_2) + \tfrac{1}{2}\sqrt{(\mu_1 - \mu_2)^2 + (\sigma_1 + \sigma_2)^2}.$$

Note also that the $\mathrm{BNT}_2$–bound improves the corresponding Arnold–Groeneveld bound (3) for the expected maximum only in the case where $\sigma_1 \neq \sigma_2$.

It is also worth pointing out that a particular application of the main result in [13] yields an even better (than $\mathrm{BNT}_2$, $\mathrm{AG}_2$ and $\rho_2$) bound. Indeed, setting $\rho := \mathrm{Corr}\,(X_1, X_2)$, it follows from Papadatos' results that for any $(X_1, X_2) \in \mathcal{F}_2(\boldsymbol{\mu}, \boldsymbol{\sigma})$,

$$\mathbb{E}R_2 \leq \gamma_2 := \sqrt{(\mu_1 - \mu_2)^2 + (\sigma_1 + \sigma_2)^2 - 2(1+\rho)\sigma_1\sigma_2}. \tag{46}$$

Obviously, $\gamma_2 \leq \rho_2$ with equality if and only if $\rho = -1$. This inequality explains the fact that the extremal random pair $(X_1, X_2)$ (that attains the bounds $\rho_2$ and $\mathrm{BNT}_2$) has correlation $\rho = -1$; see Equation (44).

The preceding inequalities have some interest because they provide a basis for the investigation of the dependence structure of an ordered pair. This kind of investigation is particularly useful for its application to reliability systems; see [24]. On the other hand, in view of the obvious facts $X_{1:2} + X_{2:2} = X_1 + X_2$ and $X_{1:2}X_{2:2} = X_1X_2$, we get the relation

$$\mathrm{Cov}\,[X_{1:2}, X_{2:2}] = \rho\sigma_1\sigma_2 - \tfrac{1}{4}(\mu_2 - \mu_1)^2 + \tfrac{1}{4}(\mathbb{E}R_2)^2, \quad (X_1, X_2) \in \mathcal{F}_2(\boldsymbol{\mu}, \boldsymbol{\sigma}), \tag{47}$$

where $\rho = \mathrm{Corr}\,(X_1, X_2)$. Thus, any bound (upper or lower) for $\mathbb{E}R_2$ can be translated to a bound for $\mathrm{Cov}\,(X_{1:2}, X_{2:2})$ as well as for $\mathbb{E}X_{2:2}$; see [25,26]. Therefore, it is of some interest to know whether the bound in (46) is tight for given $\rho$. This is indeed the case but, to the best of our knowledge, this elementary fact does not seem to be well known, and we shall provide a simple proof here. To this end, let $\boldsymbol{\mu} = (\mu_1, \mu_2)$, $\boldsymbol{\sigma} = (\sigma_1, \sigma_2)$ (with $\sigma_1 > 0$, $\sigma_2 > 0$), $-1 \leq \rho \leq 1$, and define the section

$$\mathcal{F}_2(\boldsymbol{\mu}, \boldsymbol{\sigma}; \rho) := \{(X_1, X_2) \in \mathcal{F}_2(\boldsymbol{\mu}, \boldsymbol{\sigma}) : \mathrm{Corr}\,(X_1, X_2) = \rho\}.$$

Then we have the following.

Theorem 7.1 *As $(X_1, X_2)$ varies in $\mathcal{F}_2(\boldsymbol{\mu}, \boldsymbol{\sigma}; \rho)$,*

$$\inf \mathbb{E}R_2 = |\mu_2 - \mu_1|, \quad \sup \mathbb{E}R_2 = \sqrt{(\mu_2 - \mu_1)^2 + (\sigma_1 + \sigma_2)^2 - 2(1+\rho)\sigma_1\sigma_2}.$$

*Remark 7.1* From the proof it follows that (the probability law of) the extremal vector $(X_1, X_2) \in \mathcal{F}_2(\boldsymbol{\mu}, \boldsymbol{\sigma}; \rho)$ that attains the equality in Inequality (46) is unique if and only if either (i) $\rho = -1$ or (ii) $\sigma_1 \neq \sigma_2$ and $\rho = 1$. With this in mind, let us keep $\mu_1, \mu_2, \sigma_1, \sigma_2$ constant, and write $\gamma_2 = \gamma_2(\rho)$ for the quantity defined by (46). Then, $\gamma_2(\rho)$ is strictly decreasing in $\rho$ (recall that $\sigma_1 > 0$, $\sigma_2 > 0$), attaining its maximum value at $\rho = -1$. By definition, $\gamma_2(-1) = \rho_2$ (see Equation (43)), and thus, for the equality $\mathbb{E}R_2 = \rho_2$ it is necessary that $\rho = -1$. This observation verifies that the unique distribution that attains the equality in Inequality (43) is the $\mathrm{BNT}_2$–distribution, given by Equation (44).

In view of (45), (47), the following result is straightforward from Theorem 7.1.

Corollary 7.1 *Let* $(X_1, X_2) \in \mathcal{F}_2(\mu_1, \mu_2, \sigma_1, \sigma_2; \rho)$ *with* $\sigma_1 > 0, \sigma_2 > 0$. *Then,*

$$\max\{\mu_1, \mu_2\} \le \mathbb{E}\{\max\{X_1, X_2\}\} \le \tfrac{1}{2}(\mu_1 + \mu_2) + \tfrac{1}{2}\sqrt{(\mu_2 - \mu_1)^2 + \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2},$$

$$\rho\sigma_1\sigma_2 \le \mathrm{Cov}\,[\min\{X_1, X_2\}, \max\{X_1, X_2\}] \le \tfrac{1}{4}(\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2).$$

*All bounds are best possible.*

It is worth pointing out that, as Corollary 7.1 shows, the covariance of an ordered pair can never be smaller than the covariance of the observations and, in particular, an ordered pair formed from nonnegatively correlated observations is nonnegatively correlated. While these facts, as well as the lower covariance bound of an ordered pair, are well-known (see [24, Equation's (2.9), (2.11)]), the upper bound seems to be of some interest.

There are some propositions and questions for further research. An obvious one is in extending the main result of Theorem 6.1 and of (5) to more general $L$-statistics. Recall that the tight bound for any $L$-statistic under the i.d. assumption is known from the work of Rychlik.[4] However, Rychlik's result is not applicable if arbitrary multivariate distributions are allowed for the data.

A second one concerns extension to other $L$-statistics of the bounds given in Corollary 7.1 and Theorem 7.1 for $n \ge 3$, noting that these bounds have a different nature, because they use covariance information from the data. It is particularly interest to know the tight bounds for the expected range and the expected maximum under mean–variance–covariance information on the observations. Non-tight bounds of this form are given, e.g. in [11,13]. It is worth pointing out that some sophisticated optimization techniques (semidefinite programming) have been fruitfully applied to this kind of problems, especially for the maximum and the range. The interested reader is referred to Natarajan and Teo,[27] where some financial applications of the range bounds are also included. However, note that one would hardly discover the simple formula (46) from the (reduced) semidefinite program in Natarajan and Teo's Section 4.

A lot of research has been devoted in deriving distribution and expectation bounds for $L$-statistics based on random vectors with given marginals; see [8,28–41]. The results by Lai and Robbins,[18] Nagaraja [10] and Arnold and Balakrishnan [15] show that some deterministic inequalities play an important role in the derivation of tight bounds for $L$-statistics; see [42]. On the other hand, the deterministic inequality (8) can be viewed as a range analogue of the inequality from Lai and Robbins.[18] Noting that the Lai-Robbins inequality yields the tight bound for the expected maximum under completely known marginal distributions (see [17,35]), it would not be surprising if (8) could produce the best possible bound for the expected range. Thus, a natural question is whether it is true that for all multivariate vectors with given marginal distributions $F_1, \ldots, F_n$ and finite first moment,

$$\sup \mathbb{E}R_n = \inf_{c\in\mathbb{R},\lambda>0} \left\{ -(n-2)\lambda + \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}[|(X_i - c) - \lambda| + |(X_i - c) + \lambda|] \right\}.$$

Note that the RHS is an upper bound for the LHS, and depends only on $F_1, \ldots, F_n$.

## Acknowledgments

The author would like to cordially thank three Referees and the AE for their careful reading and for correcting some previous mistakes.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

Research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program 'Education and Lifelong Learning' of the National Strategic Reference Framework (NSRF) – Research Funding Program: ARISTEIA, Grant No: 4357. Also, this work is partially supported by the University of Athens Research Grant 70/4/5637.

## References

[1] Placket RL. Limits of the ratio of mean range to standard deviation. Biometrika. 1947;34:120–122.
[2] Gumbel EJ. The maxima of the mean largest value and of the range. Ann Math Statist. 1954;25:76–84.
[3] Hartley HO, David HA. Universal bounds for mean range and extreme observation. Ann Math Statist. 1954;25: 85–99.
[4] Rychlik T. Sharp bounds on *L*-estimates and their expectations for dependent samples. Commun Statist–Theory Methods. 1993;22:1053–1068. Erratum **23**, 305–306.
[5] Rychlik T, Projecting statistical functionals, Lecture Notes in Statistics, Vol. 160. New York: Springer; 2001.
[6] David HA. Order statistics. 2nd ed. New York: Wiley; 1981.
[7] David HA, Nagaraja HN. Order statistics. 3rd ed. New York: Wiley; 2003.
[8] Rychlik T, Bounds on expectations of *L*-estimates. In: Balakrishnan N and Rao CR, editors. Order statistics: theory and methods. Handbook of statistics, Vol. 16. Amsterdam: North-Holland; 1998. p. 105–145.
[9] Arnold BC, Groeneveld RA. Bounds on expectations of linear systematic statistics based on dependent samples. Ann Statist. 1979;7:220–223. Correction: **8**, 1401.
[10] Nagaraja HN. Some finite sample results for the selection differential. Ann Inst Statist Math. 1981;33:437–448.
[11] Aven T. Upper (lower) bounds on the mean of the maximum (minimum) of a number of random variables. J Appl Probab. 1985;22:723–728.
[12] Lefèvre C. Bounds for the expectation of linear combinations of order statistics with application to pert networks. Stochast Anal Appl. 1986;4:351–356.
[13] Papadatos N. Expectation bounds on linear estimators from dependent samples. J Statist Plan Inference. 2001;93:17–27.
[14] Kaluszka M, Okolewski A, Szymanska K. Sharp bounds for *L*-statistics from dependent samples of random length. J Statist Plan Inference. 2005;127:71–89.
[15] Arnold BC, Balakrishnan N, Relations, bounds and approximations for order statistics, Lecture Notes in Statistics, Vol. 53. New York: Springer; 1989.
[16] Bertsimas D, Natarajan K, Teo C-P. Probabilistic combinatorial optimization: moments, semidefinite programming and asymptotic bounds. SIAM J Optim. 2004;15(1):185–209.
[17] Bertsimas D, Natarajan K, Teo C-P. Tight bounds on expected order statistics. Prob Engineer Inform Sci. 2006;20:667–686.
[18] Lai TL, Robbins H. Maximally dependent random variables. Proc Nat Acad Sci USA. 1976;73(2):286–288.
[19] Bertsimas D, Doan XV, Natarajan K, Teo C-P. Models for minimax stochastic linear optimization problems with risk aversion. Math O R. 2010;35(3):580–602.
[20] Isii K. On the sharpness of Chebyshev-type inequalities. Ann Inst Statist Math. 1963;14:185–197.
[21] Karlin S, Studden WJ. Tchebycheff systems: with applications in analysis and statistics. New York: Wiley-Interscience; 1966.
[22] Barvinok A. Matrices with prescribed row and column sums. Linear Algebra Appl. 2012;436:820–844.
[23] Giaquinta M, Modica G. Mathematical analysis. Foundations and advanced techniques for functions of several variables. Birkhäuser: Springer; 2012.
[24] Navarro J, Balakrishnan N. Study of some measures of dependence between order statistics and systems. J Multivariate Anal. 2010;101:52–67.
[25] Balakrishnan N, Balasubramanian K. Equivalence of Hartley–David–Gumbel and Papathanasiou bounds and some further remarks. Statist Probab Lett. 1993;16:39–41.
[26] Papathanasiou V. Some characterizations of distributions based on order statistics. Statist Probab Lett. 1990;9: 145–147.
[27] Natarajan K, Teo C-P, Semidefinite programming reformulation of completely positive programs: range estimation and best-worst choice modeling, (2014). Available from: http://people.sutd.edu.sg.
[28] Arnold BC. Distribution-free bounds on the mean of the maximum of a dependent sample. SIAM J Appl Math. 1980;38:163–167.
[29] Arnold BC. *p*-Norm bounds on the expectation of the maximum of a possibly dependent sample. J Multivariate Anal. 1985;17:316–332.
[30] Arnold BC. Bounds on the expected maximum. Commun Statist–Theory Methods. 1988;17:2135–2150.
[31] Caraux G, Gascuel O. Bounds on distribution functions of order statistics for dependent variates. Statist Probab Lett. 1992;14:103–105.
[32] Gajek L, Rychlik T. Projection method for moment bounds on order statistics from restricted families. I. Dependent case. J Multivariate Anal. 1996;57:156–174.

[33] Gajek L, Rychlik T. Projection method for moment bounds on order statistics from restricted families. II. Independent case. J Multivariate Anal. 1998;64:156–182.
[34] Gascuel O, Caraux G. Bounds on expectations of order statistics via extremal dependences. Statist Probab Lett. 1992;15:143–148.
[35] Meilijson I, Nadas A. Convex majorization with an application to the length of critical paths. J Appl Probab. 1979;16:671–677.
[36] Papadatos N. Distribution and expectation bounds on order statistics from possibly dependent variates. Statist Probab Lett. 2001;54:21–31.
[37] Rychlik T. Stochastically extremal distributions of order statistics for dependent samples. Statist Probab Lett. 1992;13:337–341.
[38] Rychlik T. Bounds for expectations of *L*-estimates for dependent samples. Statistics. 1993;24:9–15.
[39] Rychlik T. Distributions and expectations of order statistics for possibly dependent random variables. J Multivariate Anal. 1994;48:31–42.
[40] Rychlik T. Bounds for order statistics based on dependent variables with given nonidentical distributions. Statist Probab Lett. 1995;23:351–358.
[41] Rychlik T, Optimal deterministic bounds on *L*-statistics. In: Ahsanullah M and Raqab MZ, editors. Recent developments in ordered random variables. New York: Nova Science Publishers; 2007. p. 1–18.
[42] Rychlik T. Sharp inequalities for linear combinations of elements of monotone sequences. Bull Polish Acad Sci Math. 1992;40:247–254.



## Appendix. Proofs

*Proof of Lemma* 2.1. Fix $c \in \mathbb{R}$ and $\lambda > 0$ and set $y_1 = c - \lambda$, $y_2 = c + \lambda$, so that $y_1 < y_2$. Observe that $R_n = X_{n:n} - X_{1:n}$ and

$$\sum_{i=1}^{n}\{|X_i - y_1| + |X_i - y_2|\} = \sum_{i=1}^{n}\{|X_{i:n} - y_1| + |X_{i:n} - y_2|\}.$$

Hence,

$$\begin{aligned}\sum_{i=1}^{n}\{|X_i - y_1| + |X_i - y_2|\} - (n-2)(y_2 - y_1) - 2R_n &= \sum_{i=2}^{n-1}\{|X_{i:n} - y_1| + |X_{i:n} - y_2| - (y_2 - y_1)\}\\ &\quad + \{|X_{1:n} - y_1| + |X_{n:n} - y_1| - (X_{n:n} - X_{1:n})\}\\ &\quad + \{|X_{1:n} - y_2| + |X_{n:n} - y_2| - (X_{n:n} - X_{1:n})\}.\end{aligned}$$

For each $i \in \{2, \ldots, n-1\}$, we have

$$y_2 - y_1 = |y_2 - y_1| = |(X_{i:n} - y_1) - (X_{i:n} - y_2)| \le |X_{i:n} - y_1| + |X_{i:n} - y_2|,$$

with equality if and only if $y_1 \le X_{i:n} \le y_2$. Since the sum $\sum_{i=2}^{n-1}\{|X_{i:n} - y_1| + |X_{i:n} - y_2| - (y_2 - y_1)\}$ contains only nonnegative terms, it follows that

$$\sum_{i=2}^{n-1}\{|X_{i:n} - y_1| + |X_{i:n} - y_2| - (y_2 - y_1)\} \ge 0,$$

with equality if and only if $y_1 \le X_{2:n} \le \cdots \le X_{n-1:n} \le y_2$. Also, for $y = y_1$ or $y_2$,

$$X_{n:n} - X_{1:n} = |(X_{n:n} - y) - (X_{1:n} - y)| \le |X_{1:n} - y| + |X_{n:n} - y|$$

with equality if and only if $X_{1:n} \le y \le X_{n:n}$. Therefore,

$$-2R_n - (n-2)(y_2 - y_1) + \sum_{i=1}^{n}\{|X_i - y_1| + |X_i - y_2|\} \ge 0$$

with equality if and only if $X_{1:n} \le y_1 \le X_{2:n} \le \cdots \le X_{n-1:n} \le y_2 \le X_{n:n}$. ■

*Proof of Lemma* 2.2. In case $\mu^2 + \sigma^2 \geq 4$ it suffices to use the inequality

$$|X - 1| + |X + 1| \leq \sqrt{\mu^2 + \sigma^2} + \frac{X^2}{\sqrt{\mu^2 + \sigma^2}},$$

where the equality holds if and only if $X \in \{-\sqrt{\mu^2 + \sigma^2}, \sqrt{\mu^2 + \sigma^2}\}$. Taking expectations, we get

$$\mathbb{E}\{|X - 1| + |X + 1|\} \leq \sqrt{\mu^2 + \sigma^2} + \frac{\mathbb{E}X^2}{\sqrt{\mu^2 + \sigma^2}} = 2\sqrt{\mu^2 + \sigma^2}.$$

For equality $X$ has to assume the values $\pm\sqrt{\mu^2 + \sigma^2}$. Set $p = \mathbb{P}[X = \sqrt{\mu^2 + \sigma^2}]$ so that $1 - p = \mathbb{P}[X = -\sqrt{\mu^2 + \sigma^2}]$. The relation $\mathbb{E}X^2 = \mu^2 + \sigma^2$ is satisfied for any value of $p \in [0, 1]$, while the condition $\mathbb{E}X = \mu$ specifies $p$ to be as in (a).

Next, we assume that $2|\mu| < \mu^2 + \sigma^2 < 4$ and use the inequality

$$|X - 1| + |X + 1| \leq 2 + \tfrac{1}{2}X^2,$$

in which the equality holds if and only if $X \in \{-2, 0, 2\}$. Taking expectations we again conclude Inequality (10) with $U(\mu, \sigma)$ given by the second line of Equation (11). It is easy to see that the unique random variable in $\mathcal{F}_1(\mu, \sigma)$ that assumes values in the set $\{-2, 0, 2\}$ is the one given by (b).

Next, suppose that $\mu^2 + \sigma^2 \leq 2\mu$, and hence, $0 < \mu < 2$. Working as before, it suffices to take expectations in the inequality

$$|X - 1| + |X + 1| \leq 2 + \frac{(X - 1 + \sqrt{(\mu - 1)^2 + \sigma^2})^2}{2\sqrt{(\mu - 1)^2 + \sigma^2}},$$

in which the equality holds if and only if $X \in \{x_1, x_2\}$, where $x_1 = 1 - \sqrt{(\mu - 1)^2 + \sigma^2}$, $x_2 = 1 + \sqrt{(\mu - 1)^2 + \sigma^2}$. Note that $0 < (\mu - 1)^2 + \sigma^2 = 1 - [2\mu - (\mu^2 + \sigma^2)] \leq 1$; thus, $0 \leq x_1 < 1 < x_2 \leq 2$. Now it is easily seen that the unique random variable in $\mathcal{F}_1(\mu, \sigma)$ that assumes values in the set $\{x_1, x_2\}$ is the one given by (c). Observing that $|X - 1| + |X + 1|$ is even, the case $\mu^2 + \sigma^2 \leq -2\mu$ is reduced to the previous one by considering $-X \in \mathcal{F}_1(-\mu, \sigma)$. ■

*Proof of Lemma* 3.1. For $n = 1$ both Equations (20) and (21) are invalid, so we have nothing to prove. For $n = 2$, the result is trivial (we have uniqueness if Equation (20) is satisfied; we have equality in Inequality (21) whenever it is fulfilled). Assume $n \geq 3$ and consider the set of all probability matrices with the given marginals,

$$\mathcal{M}(\boldsymbol{p}, \boldsymbol{q}) = \left\{ Q = (q_{ij}) \in \mathbb{R}^{n\times n} : q_{ij} \geq 0, \quad \sum_{i=1}^{n} q_{ij} = q_j, \quad \sum_{j=1}^{n} q_{ij} = p_i \text{ for all } i, j \right\}.$$

The set $\mathcal{M}(\boldsymbol{p}, \boldsymbol{q})$ is nonempty since, e.g. it contains the matrix $Q = (p_i q_j)$. Also, the function $f(Q) := \text{trace}(Q) = \sum_{i=1}^{n} q_{ii}$ is continuous with respect to the total variation distance, $d(Q, \tilde{Q}) = \sum_{i,j} |q_{ij} - \tilde{q}_{ij}|$ (or any other equivalent metric on $\mathbb{R}^{n\times n}$). Moreover, $\mathcal{M}(\boldsymbol{p}, \boldsymbol{q})$ is a compact subset of $\mathbb{R}^{n\times n}$, since it is obviously closed, and it is contained in a ball with centre the null matrix $O_{n\times n}$ and (total variation) radius 1. It follows that $f(Q)$ attains its minimum value for some $Q^* \in \mathcal{M}(\boldsymbol{p}, \boldsymbol{q})$.

Let $(X, Y) \sim Q^* = (q^*_{ij})$ where $Q^* \in \mathcal{M}(\boldsymbol{p}, \boldsymbol{q})$ is a minimizing matrix. Then, $X \sim \boldsymbol{p}$, $Y \sim \boldsymbol{q}$ and $f(Q^*) = \mathbb{P}[X = Y]$. A simple argument shows that the principal diagonal of any minimizing matrix $Q^*$ can contain at most one nonzero entry. Indeed, if $q^*_{ii} > 0$ and $q^*_{jj} > 0$ with $i \neq j$, set $\gamma = \min\{q^*_{ii}, q^*_{jj}\} > 0$, and consider the matrix $\tilde{Q} = (\tilde{q}_{ij})$ which differs from $Q^*$ only in the following four entries: $\tilde{q}_{ii} = q^*_{ii} - \gamma$, $\tilde{q}_{jj} = q^*_{jj} - \gamma$, $\tilde{q}_{ij} = q^*_{ij} + \gamma$, $\tilde{q}_{ji} = q^*_{ji} + \gamma$. Since the row/column sums are unaffected and the elements of $\tilde{Q}$ are nonnegative, it is clear that $\tilde{Q} \in \mathcal{M}(\boldsymbol{p}, \boldsymbol{q})$ and we arrived at the contradiction $f(\tilde{Q}) = f(Q^*) - 2\gamma < f(Q^*)$. Therefore, all diagonal entries of a minimizing matrix $Q^*$ have to be zero, with the possible exception of at most one of them.

*Sufficiency*: Assume that Condition (21) is satisfied, and suppose that $\min_Q f(Q) = f(Q^*) = \theta > 0$. Let $q^*_{kk} = \theta$ and thus, $q^*_{ii} = 0$ for all $i \neq k$. Then,

$$\mathbb{P}[\{X = k\} \cup \{Y = k\}] = \mathbb{P}[X = k] + \mathbb{P}[Y = k] - \mathbb{P}[X = k, Y = k] = p_k + q_k - \theta.$$

Since $1 - p_k - q_k \geq 0$ (from Condition (21)) we thus obtain

$$\mathbb{P}[X \neq k, Y \neq k] = 1 - \mathbb{P}[\{X = k\} \cup \{Y = k\}] = \theta + (1 - p_k - q_k) \geq \theta > 0.$$

On the other hand, since $q^*_{ii} = 0$ for all $i \neq k$, we have

$$\mathbb{P}[X \neq k, Y \neq k] = \sum_{(i,j): i\neq k, j\neq k, i\neq j} q^*_{ij}.$$

The above probability is at least $\theta$, and thus, strictly positive. It follows that the sum contains at least one positive term. Hence, we can find two indices $r, s$ with $r \neq k$, $s \neq k$, $r \neq s$, such that $q^*_{rs} > 0$. Set $\delta = \min\{\theta, q^*_{rs}\} > 0$ and consider the matrix $\tilde{Q} = (\tilde{q}_{ij})$ which differs from $Q^*$ only in the elements $\tilde{q}_{kk} = q^*_{kk} - \delta = \theta - \delta$, $\tilde{q}_{rs} = q^*_{rs} - \delta$, $\tilde{q}_{rk} = q^*_{rk} + \delta$, $\tilde{q}_{ks} = q^*_{ks} + \delta$. Since the row/column sums are unaffected and the elements of $\tilde{Q}$ are nonnegative, it is clear that $\tilde{Q} \in \mathcal{M}(\boldsymbol{p}, \boldsymbol{q})$, and this results to the contradiction $f(\tilde{Q}) = \theta - \delta < \theta$. Thus, $f(Q^*) = \mathbb{P}[X = Y] = 0$; this proves the existence of random vectors satisfying (20).

*Necessity*: This is entirely obvious. For, if a random vector $(X, Y)$ satisfies Equation (20) then $(X, Y) \sim Q$ for some $Q \in \mathcal{M}(\boldsymbol{p}, \boldsymbol{q})$ with $q_{ii} = 0$ for all $i$. Thus, for any $i$,

$$p_i + q_i = p_i + q_i - q_{ii} = \mathbb{P}[X = i] + \mathbb{P}[Y = i] - q_{ii} = \mathbb{P}[\{X = i\} \cup \{Y = i\}] \leq 1.$$

*Uniqueness*: Assume that $\max_i\{p_i + q_i\} = 1$ and choose $k$ with $p_k + q_k = 1$. If $(X, Y) \sim Q$ satisfies Equation (20), we have $\mathbb{P}[\{X = k\} \cup \{Y = k\}] = p_k + q_k - q_{kk} \geq p_k + q_k - \mathbb{P}[X = Y] = p_k + q_k = 1$. It follows that $Q$ can have nonzero entries only in its $k$th row and in its $k$th column. Thus, $q_{ik} = p_i$ for all $i \neq k$, $q_{kj} = q_j$ for all $j \neq k$ and $q_{ij} = 0$ otherwise; hence, $Q$ is uniquely determined from $\boldsymbol{p}, \boldsymbol{q}$. Note that $k$ need not be unique, but $Q$ is always unique. For example, if $\boldsymbol{p} = (1 - p, p, 0, \ldots, 0)$ and $\boldsymbol{q} = (p, 1 - p, 0, \ldots, 0)$ with $0 \leq p \leq 1$, we obtain the unique solution to Equation (20) as $\mathbb{P}[X = 2, Y = 1] = p = 1 - \mathbb{P}[X = 1, Y = 2]$. In fact, one can easily verify that this example describes the most general case (modulo the positions of $p, 1 - p$) where the relation $p_k + q_k = 1$ can hold for more than one index $k$.

*Non-uniqueness*: Suppose that all $p_i$ and $q_i$ are positive and that Condition (21) holds as a strict inequality, that is, $p_i + q_i < 1$ for all $i$. [The last assumption is possible only if $n \geq 3$.] Set $\beta = (1/n^2)[1 - \max_i\{p_i + q_i\}] > 0$, $\delta = \min_{i,j}\{p_i q_j\} > 0$ and $\epsilon = \min\{\beta, \delta\} > 0$. Define

$$\mathcal{M}_\epsilon(\boldsymbol{p}, \boldsymbol{q}) := \{Q \in \mathcal{M}(\boldsymbol{p}, \boldsymbol{q}) : q_{ij} \geq \epsilon \text{ for all } i, j \text{ with } i \neq j\}.$$

Observe that $\mathcal{M}_\epsilon(\boldsymbol{p}, \boldsymbol{q})$ is a nonempty (since it contains $Q = (p_i q_j)$) compact subject of $\mathbb{R}^{n\times n}$. Applying the same arguments as in the beginning of the proof we see that the continuous function $f(Q) = \text{trace}(Q)$ attains its minimum value at a matrix $Q^*_\epsilon = (q^*_{ij}) \in \mathcal{M}_\epsilon(\boldsymbol{p}, \boldsymbol{q})$; $Q^*_\epsilon$ has at most one nonzero diagonal entry while, by the definition of $\mathcal{M}_\epsilon(\boldsymbol{p}, \boldsymbol{q})$, all off-diagonal entries are at least $\epsilon$. Let $(X, Y) \sim Q^*_\epsilon$. Assuming $\mathbb{P}[X = Y] = \theta > 0$ we can find a unique index $k$ such that $q^*_{kk} = \theta$; then, $\mathbb{P}[\{X = k\} \cup \{Y = k\}] = p_k + q_k - \theta$. Since $q^*_{ii} = 0$ for $i \neq k$, we have

$$\begin{aligned}\sum_{(i,j): i\neq k, j\neq k, i\neq j} q^*_{ij} &= \mathbb{P}[X \neq k, Y \neq k] = \theta + (1 - p_k - q_k) \geq \theta + [1 - \max_i\{p_i + q_i\}] \\ &= \theta + n^2\beta \geq \theta + n^2\epsilon > n^2\epsilon.\end{aligned}$$

This sum contains $(n - 1)(n - 2) < n^2$ terms and the inequality shows that at least one of them is greater than $\epsilon$. Thus, we can find two indices $r, s$ with $r \neq k$, $s \neq k$, $r \neq s$, such that $q^*_{rs} > \epsilon$; say $q^*_{rs} = \epsilon + \gamma$ with $\gamma > 0$. Set $\lambda = \min\{\theta, \gamma\} > 0$ and consider the matrix $\tilde{Q}_\epsilon = (\tilde{q}_{ij})$, which differs from $Q^*_\epsilon$ at exactly the four elements $\tilde{q}_{kk} = q^*_{kk} - \lambda = \theta - \lambda \geq 0$, $\tilde{q}_{rs} = q^*_{rs} - \lambda = \epsilon + (\gamma - \lambda) \geq \epsilon$, $\tilde{q}_{rk} = q^*_{rk} + \lambda$, $\tilde{q}_{ks} = q^*_{ks} + \lambda$. It is clear that $\tilde{Q}_\epsilon \in \mathcal{M}_\epsilon(\boldsymbol{p}, \boldsymbol{q})$ and, once again, it contradicts the definition of $Q^*_\epsilon$: $f(\tilde{Q}_\epsilon) = \theta - \lambda < \theta = f(Q^*_\epsilon)$. Thus, $f(Q^*_\epsilon) = \mathbb{P}[X = Y] = 0$. This shows the existence of random vectors $(X, Y)$ satisfying (20) with the additional property $\mathbb{P}[X = i, Y = j] \geq \epsilon > 0$ for all $i \neq j$, provided that $\epsilon > 0$ is sufficiently small. Given a probability matrix $Q^*_\epsilon = (q^*_{ij})$ of this form, it is easy to construct a second solution, $Q = (q_{ij})$, to (20); e.g. set $q_{12} = q^*_{12} - \epsilon/2$, $q_{13} = q^*_{13} + \epsilon/2$, $q_{21} = q^*_{21} + \epsilon/2$, $q_{23} = q^*_{23} - \epsilon/2$, $q_{31} = q^*_{31} - \epsilon/2$, $q_{32} = q^*_{32} + \epsilon/2$, and leave the rest entries unchanged. Finally, it is easy to see that if $Q_0$, $Q_1$ both solve (20), the same is true for $Q_t = tQ_1 + (1 - t)Q_0$, $0 \leq t \leq 1$, and the proof is complete. ■

*Proof of Theorem* 3.1. Assume that $\mathbb{E}R_n = \text{AG}_n$ for some random vector $\boldsymbol{X}$ with $\mathbb{E}\boldsymbol{X} = \boldsymbol{\mu}$ and $\text{Var}\,\boldsymbol{X} = \boldsymbol{\sigma}^2$. Set $c = \bar{\mu}$, $\lambda = \frac{1}{4}\text{AG}_n > 0$ and take expectations in (8) to get (cf. Remark 2.2)

$$\text{AG}_n = \mathbb{E}R_n \leq \frac{-(n-2)\text{AG}_n}{4} + \frac{\text{AG}_n}{8}\sum_{i=1}^n \mathbb{E}\left\{\left|\frac{X_i - \bar{\mu}}{\text{AG}_n/4} - 1\right| + \left|\frac{X_i - \bar{\mu}}{\text{AG}_n/4} + 1\right|\right\}.$$

Next, from $|y - 1| + |y + 1| \leq 2 + \frac{1}{2}y^2$ with equality if and only if $y \in \{-2, 0, 2\}$ we get

$$\sum_{i=1}^n \mathbb{E}\left\{\left|\frac{X_i - \bar{\mu}}{\text{AG}_n/4} - 1\right| + \left|\frac{X_i - \bar{\mu}}{\text{AG}_n/4} + 1\right|\right\} \leq 2n + \frac{1}{2}\sum_{i=1}^n \mathbb{E}\left\{\left(\frac{X_i - \bar{\mu}}{\text{AG}_n/4}\right)^2\right\} = 2n + 4.$$

Since $-(n - 2)\text{AG}_n/4 + (\text{AG}_n/8)(2n + 4) = \text{AG}_n$, it follows that the preceding inequalities are, in fact, equalities. Therefore, $\mathbb{E}R_n = \text{AG}_n$ is equivalent to (9) (with $c = \bar{\mu}$, $\lambda = \frac{1}{4}\text{AG}_n$) and $(X_i - \bar{\mu})/\text{AG}_n/4 \in \{-2, 0, 2\}$, $i = 1, \ldots, n$ (of course, it suffices to hold with probability 1). Hence, $\mathbb{E}R_n = \text{AG}_n$ if and only if

(a) $X_{1:n} \leq \bar{\mu} - \frac{1}{4}\text{AG}_n \leq X_{2:n} \leq \cdots \leq X_{n-1:n} \leq \bar{\mu} + \frac{1}{4}\text{AG}_n \leq X_{n:n}$ and

$$\text{(b)}\quad X_i \in \left\{\bar{\mu} - \frac{\mathrm{AG}_n}{2}, \bar{\mu}, \bar{\mu} + \frac{\mathrm{AG}_n}{2}\right\}, \quad i = 1, \ldots, n,$$

with probability 1. Therefore, the (essential) support of any extremal random vector is a subset of

$$S := \left\{\left(\bar{\mu}, \ldots, \bar{\mu}, \bar{\mu} + \frac{\mathrm{AG}_n}{2}, \bar{\mu}, \ldots, \bar{\mu}, \bar{\mu} - \frac{\mathrm{AG}_n}{2}, \bar{\mu}, \ldots, \bar{\mu}\right)\right\},$$

where the plus and minus signs can appear at any two (different) places. Clearly, $S$ has $n(n-1)$ elements and can be written as

$$S = \left\{\bar{\mu}\,\mathbf{1} + \frac{\boldsymbol{e}(i) - \boldsymbol{e}(j)}{2}\mathrm{AG}_n : (i,j) \in \{1, \ldots, n\}^2,\ i \neq j\right\}.$$

Let $S' := \{(i,j) \in \{1, \ldots, n\}^2 : i \neq j\}$. The function $g : S' \to S$, that sends $(i,j)$ to $g(i,j) = \bar{\mu}\,\mathbf{1} + ((\boldsymbol{e}(i) - \boldsymbol{e}(j))/2)\mathrm{AG}_n$, is a bijection. It follows that $(X, Y) := g^{-1}(\boldsymbol{X})$ is a random pair with values in a subset of $S'$, and $\boldsymbol{X} = g(X, Y)$; this verifies the representation (17). For $i \in \{1, \ldots, n\}$ we set

$$p_i^+ := \mathbb{P}\left[X_i = \bar{\mu} + \frac{\mathrm{AG}_n}{2}\right] = \mathbb{P}[X = i], \quad p_i^- = \mathbb{P}\left[X_i = \bar{\mu} - \frac{\mathrm{AG}_n}{2}\right] = \mathbb{P}[Y = i],$$

so that $\mathbb{P}[X_i = \bar{\mu}] = 1 - p_i^+ - p_i^-$. From $\mathbb{E}X_i = \mu_i$ we get $p_i^+ - p_i^- = 2(\mu_i - \bar{\mu})/\mathrm{AG}_n$ and from $\mathbb{E}\{(X_i - \bar{\mu})^2\} = (\mu_i - \bar{\mu})^2 + \sigma_i^2$ we obtain $p_i^+ + p_i^- = 4[(\mu_i - \bar{\mu})^2 + \sigma_i^2]/\mathrm{AG}_n^2$. Hence,

$$p_i^+ = \frac{2[(\mu_i - \bar{\mu})^2 + \sigma_i^2] + (\mu_i - \bar{\mu})\mathrm{AG}_n}{\mathrm{AG}_n^2}, \quad p_i^- = \frac{2[(\mu_i - \bar{\mu})^2 + \sigma_i^2] - (\mu_i - \bar{\mu})\mathrm{AG}_n}{\mathrm{AG}_n^2},$$

and (18) follows. Therefore, we can find a random vector $\boldsymbol{X}$ with $\mathbb{E}\boldsymbol{X} = \boldsymbol{\mu}$, $\operatorname{Var}\boldsymbol{X} = \boldsymbol{\sigma}^2$ and $\mathbb{E}R_n = \mathrm{AG}_n$ if and only if the above construction of a random pair $(X, Y)$, with $\mathbb{P}[X = Y] = 0$, is possible. According to Lemma 3.1, this is equivalent to $\max_i\{p_i^+ + p_i^-\} \le 1$, which gives (16)(ii) (it also guarantees that $\mathbb{P}[X_i = \bar{\mu}] = 1 - p_i^+ - p_i^- \ge 0$), while (16)(i) follows from $p_i^+ \ge 0$ and $p_i^- \ge 0$.

Finally, the inequalities (16) are strict for all $i$ if and only if $p_i^+ + p_i^- < 1$, $p_i^+ > 0$ and $p_i^- > 0$ for all $i$. Lemma 3.1 shows that there exist infinitely many vectors $(X, Y)$ in this case. Also, if (16) is satisfied and we have equality in (16)(ii) for some $i$, uniqueness follows again from Lemma 3.1. ■

*Proof of Lemma* 4.1. The functions $f_i : T \to (0, \infty)$ $(i = 1, 2, 3, 4)$ given by $f_1(x, y) := 2\sqrt{x^2 + y^2}$, $f_2(x, y) := 2 + \frac{1}{2}(x^2 + y^2)$, $f_3(x, y) := x + 1 + \sqrt{(x-1)^2 + y^2}$ and $f_4(x, y) := 1 - x + \sqrt{(x+1)^2 + y^2}$ are obviously $C^\infty(T)$. The function $U$ can be defined as the restriction of $f_1$ in $A_1 := \{(x, y) \in T : x^2 + y^2 \ge 4\}$, of $f_2$ in $A_2 := \{(x, y) \in T : 2|x| \le x^2 + y^2 \le 4\}$, of $f_3$ in $A_3 := \{(x, y) \in T : x^2 + y^2 \le 2x\}$ and of $f_4$ in $A_4 := \{(x, y) \in T : x^2 + y^2 \le -2x\}$. Observe that $A_3$ and $A_4$ are the closed (with respect to $T$) semidisks $T \cap D((1, 0), 1)$, $T \cap D((-1, 0), 1)$; also, $A_2 = T \cap [D((0, 0), 2) \smallsetminus A_3^o \cup A_4^o]$, and $A_1 = T \smallsetminus A_2^o \cup A_3 \cup A_4$. Therefore, $A_1 \cap A_3 = \emptyset$, $A_1 \cap A_4 = \emptyset$, $A_3 \cap A_4 = \emptyset$, $\partial A_1 = A_1 \cap A_2 = \{(x, y) \in T : x^2 + y^2 = 4\}$, $\partial A_3 = A_2 \cap A_3 = \{(x, y) \in T : (x-1)^2 + y^2 = 1\}$, $\partial A_4 = A_2 \cap A_4 = \{(x, y) \in T : (x+1)^2 + y^2 = 1\}$ and $\partial A_2 = \partial A_1 \cup \partial A_3 \cup \partial A_4$. It is easy to check that both partial derivatives of $f_1$ and $f_2$ coincide at $\partial A_1$, that both partial derivatives of $f_2$ and $f_3$ coincide at $\partial A_3$ and that both partial derivatives of $f_2$ and $f_4$ coincide at $\partial A_4$. We conclude that for $(x, y) \in T$,

$$U_1(x, y) := \frac{\partial}{\partial x}U(x, y) = \begin{cases} \dfrac{2x}{\sqrt{x^2 + y^2}} & \text{if } x^2 + y^2 \ge 4, \\ x & \text{if } 2|x| \le x^2 + y^2 \le 4, \\ \dfrac{x-1}{\sqrt{(x-1)^2 + y^2}} + 1 & \text{if } (x-1)^2 + y^2 \le 1, \\ \dfrac{x+1}{\sqrt{(x+1)^2 + y^2}} - 1 & \text{if } (x+1)^2 + y^2 \le 1, \end{cases} \tag{A1}$$

and

$$U_2(x, y) := \frac{\partial}{\partial y}U(x, y) = \begin{cases} \dfrac{2y}{\sqrt{x^2 + y^2}} & \text{if } x^2 + y^2 \ge 4, \\ y & \text{if } 2|x| \le x^2 + y^2 \le 4, \\ \dfrac{y}{\sqrt{(x-1)^2 + y^2}} & \text{if } (x-1)^2 + y^2 \le 1, \\ \dfrac{y}{\sqrt{(x+1)^2 + y^2}} & \text{if } (x+1)^2 + y^2 \le 1, \end{cases} \tag{A2}$$

and the above functions are obviously continuous. ■

*Proof of Proposition* 4.1. Fix $\boldsymbol{x}$ and $\boldsymbol{y}$ in $T$. The set $\partial A_2$ (where $U$ changes type) is a union of three disjoint semicircles, and the line segment $[\boldsymbol{x}, \boldsymbol{y}] = \{\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}), 0 \le t \le 1\}$ can have at most six common points with $\partial A_2 = \{(x, y) \in T : x^2 + y^2 = 4 \text{ or } (x-1)^2 + y^2 = 1 \text{ or } (x+1)^2 + y^2 = 1\}$; for the definition of $A_2$ see the proof of Lemma 4.1. Consider now the function $g : [0,1] \to \mathbb{R}$ with $g(t) := U(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}))$, $0 \le t \le 1$, which is continuously differentiable from Lemma 4.1. Also, $g$ is of the form of Equation (23) with $k \in \{0, \dots, 6\}$, where $g_i(t) = f_j(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}))$, $0 \le t \le 1$, for some $j = j(i) \in \{1, 2, 3, 4\}$ (the functions $f_j : T \to (0, \infty)$ are defined in the proof of Lemma 4.1). It is easy to verify that each $f_j$ has nonnegative definite Hessian matrix and, thus, is convex. Lemma 4.2 asserts that $g_i(t) : [0,1] \to (0, \infty)$ $(i = 1, \dots, k+1)$ is convex. Since $g$ is continuously differentiable, Condtition (22) is automatically satisfied, and we conclude from Lemma 4.3 that $g$ is convex. Therefore, $g$ is convex for any choice of $\boldsymbol{x}$ and $\boldsymbol{y}$ in $T$, and a final application of Lemma 4.2 completes the proof. ∎

*Remark A.1* It is worth pointing out that the function $U$ is $C^1(T)$, convex and not $C^2(T)$. Specifically,

$$U(x, y) = \begin{cases} 4 & \text{on } A_1 \cap A_2 = \partial A_1, \\ 2 + x & \text{on } A_2 \cap A_3 = \partial A_3, \\ 2 - x & \text{on } A_2 \cap A_4 = \partial A_4. \end{cases}$$

Moreover,

$$\frac{\partial}{\partial x} U(x, y) = x \quad \text{and} \quad \frac{\partial}{\partial y} U(x, y) = y \quad \text{on } \partial A_2,$$

and

$$H_1(x, y) = \frac{2}{(x^2 + y^2)^{3/2}} \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix},$$

$H_2(x, y) = I_2$, $H_3(x, y) = H_1(x - 1, y)$, $H_4(x, y) = H_1(x + 1, y)$, where $H_i$ is the Hessian of $U$ restricted to $A_i$, $i = 1, \dots, 4$.

*Proof of Lemma* 4.4. (i) Fix $x_0 \in \mathbb{R}$, $y_0 > 0$ and let $\alpha \in (0, 1)$, $c_1, c_2 \in \mathbb{R}$, $\lambda_1, \lambda_2 > 0$. Write $\beta_1 = \alpha\lambda_1/(\alpha\lambda_1 + (1-\alpha)\lambda_2) > 0$, $\beta_2 = (1-\alpha)\lambda_2/(\alpha\lambda_1 + (1-\alpha)\lambda_2) > 0$, so that $\beta_1 + \beta_2 = 1$. We have

$$\begin{aligned} \frac{h(\alpha c_1 + (1-\alpha)c_2, \alpha\lambda_1 + (1-\alpha)\lambda_2)}{\alpha\lambda_1 + (1-\alpha)\lambda_2} &= f\left(\frac{x_0 - [\alpha c_1 + (1-\alpha)c_2]}{\alpha\lambda_1 + (1-\alpha)\lambda_2}, \frac{y_0}{\alpha\lambda_1 + (1-\alpha)\lambda_2}\right) \\ &= f\left(\beta_1\left(\frac{x_0 - c_1}{\lambda_1}\right) + \beta_2\left(\frac{x_0 - c_2}{\lambda_2}\right), \quad \beta_1\left(\frac{y_0}{\lambda_1}\right) + \beta_2\left(\frac{y_0}{\lambda_2}\right)\right) \\ &\le \beta_1 f\left(\frac{x_0 - c_1}{\lambda_1}, \frac{y_0}{\lambda_1}\right) + \beta_2 f\left(\frac{x_0 - c_2}{\lambda_2}, \frac{y_0}{\lambda_2}\right) = \frac{\alpha h(c_1, \lambda_1) + (1-\alpha)h(c_2, \lambda_2)}{\alpha\lambda_1 + (1-\alpha)\lambda_2}, \end{aligned}$$

showing that $h$ is convex.

(ii) Suppose that for a particular $(x_0, y_0) \in T$, the function $h_0(c, \lambda) = \lambda f((x_0 - c)/\lambda, y_0/\lambda)$ is convex. Set $x = (x_0 - c)/\lambda$, $y = y_0/\lambda > 0$, so that

$$c = x_0 - y_0\frac{x}{y}, \quad \lambda = \frac{y_0}{y}, \quad y_0 f(x, y) = y h_0\left(x - y_0\frac{x}{y}, \frac{y_0}{y}\right), \quad (x, y) \in T.$$

Let $\alpha \in (0, 1)$, $x_1, x_2 \in \mathbb{R}$ and $y_1, y_2 > 0$. Let us now write $\beta_1 = \alpha y_1/(\alpha y_1 + (1-\alpha)y_2) > 0$, $\beta_2 = (1-\alpha)y_2/(\alpha y_1 + (1-\alpha)y_2) > 0$, so that $\beta_1 + \beta_2 = 1$. It follows that

$$\begin{aligned} & y_0 f(\alpha x_1 + (1-\alpha)x_2, \alpha y_1 + (1-\alpha)y_2) \\ &= [\alpha y_1 + (1-\alpha)y_2] h_0\left(\beta_1\left(x_1 - y_0\frac{x_1}{y_1}\right) + \beta_2\left(x_2 - y_0\frac{x_2}{y_2}\right), \beta_1\left(\frac{y_0}{y_1}\right) + \beta_2\left(\frac{y_0}{y_2}\right)\right) \\ &\le [\alpha y_1 + (1-\alpha)y_2]\left\{\beta_1 h_0\left(x_1 - y_0\frac{x_1}{y_1}, \frac{y_0}{y_1}\right) + \beta_2 h_0\left(x_2 - y_0\frac{x_2}{y_2}, \frac{y_0}{y_2}\right)\right\} \\ &= \alpha y_1 h_0\left(x_1 - y_0\frac{x_1}{y_1}, \frac{y_0}{y_1}\right) + (1-\alpha)y_2 h_0\left(x_2 - y_0\frac{x_2}{y_2}, \frac{y_0}{y_2}\right) \\ &= y_0[\alpha f(x_1, y_1) + (1-\alpha)f(x_2, y_2)], \end{aligned}$$

and the proof is complete. ∎

*Proof of Lemma* 5.2. If $(c_0,\lambda_0)\in T_0$ then, by Proposition 5.1, $\phi_n(c,\lambda)\ge\phi_n(c_0,\lambda_0)$ for all $(c,\lambda)\in T$. On the other hand, for this $c_0$ we can define the function $\psi_n(\lambda)=\phi_n(c_0,\lambda)$; by Lemma 5.1, the function $\psi_n(\lambda)$ is minimized at a unique $\lambda=\lambda_1=\lambda_1(c_0)$. Thus,

$$\psi_n(\lambda_0)=\phi_n(c_0,\lambda_0)\le\phi_n(c_0,\lambda_1)=\psi_n(\lambda_1)\le\psi_n(\lambda_0);$$

the first inequality follows from $(c_0,\lambda_0)\in T_0$ and the second from the definition of $\lambda_1$. Therefore, $\psi_n(\lambda_0)=\psi_n(\lambda_1)$, so that $\lambda=\lambda_0$ is a minimizing point for $\psi_n(\lambda)$. By uniqueness, $\lambda_1=\lambda_0$. Thus, $\lambda_0=\lambda_1(c_0)$, where $\lambda_1(\cdot):\mathbb{R}\to(0,\infty)$ is a well-defined function; it is described (implicitly) in Lemma 5.1. Hence, if $(c_0,\lambda_0)\ne(c_2,\lambda_2)$ are any two points in $T_0$ then $c_0\ne c_2$; indeed, $c_0=c_2$ implies $\lambda_0=\lambda_1(c_0)=\lambda_1(c_2)=\lambda_2$, contradicting the assumption $(c_0,\lambda_0)\ne(c_2,\lambda_2)$.

Let $L$ be the straight line that passes through the points $(c_0,\lambda_0)$ and $(c_2,\lambda_2)$. We now verify that if $(c_3,\lambda_3)\in T_0$ then $(c_3,\lambda_3)\in L$. Indeed, assume $(c_3,\lambda_3)\in T_0\setminus L$ and let $B$ be the convex hull of the points $\{(c_0,\lambda_0),(c_2,\lambda_2),(c_3,\lambda_3)\}$ (i.e. a triangle). Then $B$ must be a subset of $T_0$, because $T_0$ is convex. Since, however, $(c_3,\lambda_3)\notin L$, the triangle $B$ contains a line segment of positive length, parallel to the $\lambda$-axis and, by the previous argument, this is impossible. It follows that $T_0\subseteq L\cap T$, and since $T_0$ is compact and convex, it must be a compact line segment. ■

*Proof of Lemma* 5.3. By assumption, $A$ is moving linearly in the line segment $[A_0,A_1]$ from $A_0$ to $A_1$, thus we may write $A=A(t):=(c(t),\lambda(t))$ where $c(t)=c_0+t(c_1-c_0)$, $\lambda(t)=\lambda_0+t(\lambda_1-\lambda_0)$, $0\le t\le 1$. Then $B=B(t)=((\mu-c(t))/\lambda(t),\sigma/\lambda(t))$, so that $B(0)=B_0$, $B(1)=B_1$ and $B(t)$ is continuous in $t$. Under notation $\mathbf{B}(t)=(B(t),1)$, we have

$$\det[\mathbf{B}(0),\mathbf{B}(t),\mathbf{B}(1)]:=\begin{vmatrix}\dfrac{\mu-c_0}{\lambda_0} & \dfrac{\sigma}{\lambda_0} & 1\\ \dfrac{\mu-c(t)}{\lambda(t)} & \dfrac{\sigma}{\lambda(t)} & 1\\ \dfrac{\mu-c_1}{\lambda_1} & \dfrac{\sigma}{\lambda_1} & 1\end{vmatrix}=\frac{\sigma}{\lambda_0\lambda(t)\lambda_1}\begin{vmatrix}c_0 & \lambda_0 & 1\\ c(t) & \lambda(t) & 1\\ c_1 & \lambda_1 & 1\end{vmatrix}=0,$$

which means that $B(t)$ is a convex combination of $B(0)=B_0$ and $B(1)=B_1$. ■

*Proof of Theorem* 5.1. According to Proposition 5.1, it remains to verify that $T_0$ in Equation (25) is a singleton. Assume, in contrary, that $T_0$ contains two points $(c_0,\lambda_0)\ne(c_1,\lambda_1)$. From Lemma 5.2, we know that $c_0\ne c_1$, and that all points $(c,\lambda)\in T_0$ can be written as $(c,\lambda)=(c,\alpha c+\beta)$, $c_2\le c\le c_3$, for some $\alpha,\beta,c_2,c_3\in\mathbb{R}$ with $c_2<c_3$. Therefore, we can write $\lambda(c)=\alpha c+\beta$, $c_2\le c\le c_3$, and

$$T_0=\{(c,\alpha c+\beta),\ c_2\le c\le c_3\},\quad \alpha,\beta,c_2,c_3\in\mathbb{R},c_2<c_3.$$

Note that the parameters $\alpha$, $\beta$, $c_2$, $c_3$ have to fulfill additional restrictions so that $\lambda(c)>0$ for all $c\in[c_2,c_3]$; namely, $\alpha c_2+\beta>0$ and $\alpha c_3+\beta>0$.

Consider now the points $A(c):=(c,\lambda(c))$ and $B_i(c):=((\mu_i-c)/\lambda(c),\sigma_i/\lambda(c))$, $i=1,\dots,n$, $c_2\le c\le c_3$. As $c$ varies in $[c_2,c_3]$, the point $A=A(c)$ is moved from $A(c_2)$ to $A(c_3)$, generating the line segment $[A(c_2),A(c_3)]=T_0\subset T$. It follows from Lemma 5.3 that each point $B_i=B_i(c)$, $i=1,\dots,n$, produces a line segment too; that is, $B_i$ generates its corresponding segment $L_i:=[B_i(c_2),B_i(c_3)]\subset T$. Consider now the region $A_2=\{(x,y)\in T:2|x|\le x^2+y^2\le 4\}\subset T$. The function $U(x,y)$ (see Equation (11)) changes types (and it is not even $C^2$) only at the boundary points of $A_2$, i.e. at those $(x,y)\in T$ that belong to the set

$$C:=\{x^2+y^2=4\}\cup\{(x-1)^2+y^2=1\}\cup\{(x+1)^2+y^2=1\}\subset\mathbb{R}^2.$$

The set $\partial A_2=C\cap T$ is a union of three (disjoint) semicircles, and thus, any line segment can have at most six common points with it. It follows that only of finite number of points of the set $\cup_{i=1}^n L_i=\cup_{i=1}^n\cup_{c_2\le c\le c_3}B_i(c)$ can intersect $\partial A_2$. Let $\Gamma_1,\dots,\Gamma_k$ be all these points. Each $\Gamma_j$ belongs to some $L_i$; that is, for any $j\in\{1,\dots,k\}$ we can find an index $i=i(j)\in\{1,\dots,n\}$, and then a unique number $t=t_{ij}\in[c_2,c_3]$ such that $B_i(t)=\Gamma_j$. Clearly, for a particular index $j$, the maximal number of different $t$'s that can be found (satisfying $B_i(t)=\Gamma_j$ for some $i$) is $n$, because $B_i(t_1)\ne B_i(t_2)$ if $t_1\ne t_2$. Therefore, the set

$$N:=\{t\in[c_2,c_3]:B_i(t)=\Gamma_j \text{ for some } i \text{ and } j\}$$

is finite, say $N=\{t_1,\dots,t_m\}$ with $c_2\le t_1<\cdots<t_m\le c_3$. Fix now an interval $[t,s]\subseteq(c_2,c_3)$, of positive length, such that $[t,s]\cap N=\emptyset$. Since $[t,s]$ has no common points with $N$, it is clear that the line segment $J_i:=[B_i(t),B_i(s)]\subseteq L_i$

does not intersect $\partial A_2$, and this is true for all $i \in \{1,\ldots,n\}$. In this way we obtain a subset $T_1$ of $T_0$, namely

$$T_1 := \{(c, \alpha c + \beta),\ t \le c \le s\}, \quad \text{with } c_2 < t < s < c_3.$$

The boundary of $A_2$ divides $T$ into four disjoint open regions, namely

$$G_1 := \{(x,y) \in T : x^2 + y^2 > 4\}, \quad G_2 := \{(x,y) \in T : 2|x| < x^2 + y^2 < 4\},$$

$$G_3 := \{(x,y) \in T : (x-1)^2 + y^2 < 1\}, \quad G_4 := \{(x,y) \in T : (x+1)^2 + y^2 < 1\}.$$

Compared to $T_0$, the set $T_1$ has the additional property that, as $c$ varies, every line segment $\{B_i(c), t \le c \le s\}$ stays in the same open region. This means that the sets of indices $I_1, I_2, I_3, I_4$, defined in Remark 5.1, do not depend on $c$. Recall that

$$B_i(c) \in G_1 \Leftrightarrow (\mu_i - c)^2 + \sigma_i^2 > 4\lambda^2 \Rightarrow i \in I_1,$$

$$B_i(c) \in G_2 \Leftrightarrow 2\lambda|\mu_i - c| < (\mu_i - c)^2 + \sigma_i^2 < 4\lambda^2 \Rightarrow i \in I_2,$$

$$B_i(c) \in G_3 \Leftrightarrow (\mu_i - c)^2 + \sigma_i^2 < 2\lambda(\mu_i - c) \Rightarrow i \in I_3,$$

$$B_i(c) \in G_4 \Leftrightarrow (\mu_i - c)^2 + \sigma_i^2 < -2\lambda(\mu_i - c) \Rightarrow i \in I_4,$$

where $\lambda = \lambda(c) = \alpha c + \beta$.

Consider now the function $g_n : (t,s) \to \mathbb{R}$ with

$$g_n(c) := \phi_n(c, \lambda(c)) = \phi_n(c, \alpha c + \beta), \quad t < c < s.$$

The explicit form of $g_n$ is quite complicated:

$$\begin{aligned} g_n(c) = &-(n-2)(\alpha c + \beta) + \sum_{i \in I_1} \sqrt{(\mu_i - c)^2 + \sigma_i^2} \\ &+ \sum_{i \in I_2} \left\{ (\alpha c + \beta) + \frac{1}{4(\alpha c + \beta)} [(\mu_i - c)^2 + \sigma_i^2] \right\} \\ &+ \sum_{i \in I_3} \frac{1}{2} \{\mu_i - c + (\alpha c + \beta) + \sqrt{[\mu_i - c - (\alpha c + \beta)]^2 + \sigma_i^2}\} \\ &+ \sum_{i \in I_4} \frac{1}{2} \{c - \mu_i + (\alpha c + \beta) + \sqrt{[c - \mu_i - (\alpha c + \beta)]^2 + \sigma_i^2}\}. \end{aligned}$$

Since, however, the sets $I_j$ do not depend on $c$, it is obvious that $g_n \in C^\infty(t,s)$. By assumption, $(c, \lambda(c))$ minimizes $\phi_n(c, \lambda)$ for all $c \in (t,s)$, and this means that $g_n(c)$ is constant, implying that $g_n''(c) = 0$, $t < c < s$. A straightforward computation shows that for all $c \in (t,s)$,

$$\begin{aligned} g_n''(c) = &\sum_{i \in I_1} \frac{\sigma_i^2}{[(\mu_i - c)^2 + \sigma_i^2]^{3/2}} + \frac{1}{2[\lambda(c)]^3} \sum_{i \in I_2} \{\alpha^2 \sigma_i^2 + (\alpha \mu_i + \beta)^2\} \\ &+ \frac{(\alpha+1)^2}{2} \sum_{i \in I_3} \frac{\sigma_i^2}{[(\beta + (\alpha+1)c - \mu_i)^2 + \sigma_i^2]^{3/2}} + \frac{(\alpha-1)^2}{2} \sum_{i \in I_4} \frac{\sigma_i^2}{[(\beta + (\alpha-1)c + \mu_i)^2 + \sigma_i^2]^{3/2}}. \end{aligned}$$

Obviously, all summands are nonnegative. If $\alpha \ne 0$, the only two possibilities which are compatible with $g_n''(c) = 0$ are the following: (i) either $I_1 = I_2 = I_4 = \emptyset$ (and thus, $N(I_3) = n$) and $\alpha = -1$ or (ii) $I_1 = I_2 = I_3 = \emptyset$ (and $N(I_4) = n$) and $\alpha = 1$. However, because of Condition (31), neither (i) nor (ii) is allowed for a minimizing point $(c, \lambda)$, and in particular for $(c, \lambda(c))$. Finally, if $\alpha = 0$ then we must have $I_1 = I_3 = I_4 = \emptyset$ and, therefore, $N(I_2) = n$. The condition $\lambda(c) > 0$ now yields $\beta > 0$; thus, $g_n''(c) = n/2\beta > 0$ and $g_n(c)$ could not be a constant function in the interval $t < c < s$.

The resulting contradiction implies that the set $T_0$ cannot contain two distinct elements, and the proof is complete. ■

*Proof of Lemma* 6.1. From (Table of) Corollary 2.1 and Equation (14), and in view of Equations (34), (35),

$$\begin{aligned} \left.\frac{\partial}{\partial \lambda} \phi_n(c, \lambda)\right|_{c=c_0, \lambda=\lambda_0} &= -(n-2) + \frac{1}{2} \sum_{i \in I_2} \left\{ 2 - \frac{1}{2\lambda_0^2} [(\mu_i - c_0)^2 + \sigma_i^2] \right\} \\ &\quad + \frac{1}{2} \sum_{i \in I_3} \left\{ 1 - \frac{\mu_i - c_0 - \lambda_0}{\sqrt{(\mu_i - c_0 - \lambda_0)^2 + \sigma_i^2}} \right\} + \frac{1}{2} \sum_{i \in I_4} \left\{ 1 - \frac{c_0 - \mu_i - \lambda_0}{\sqrt{(c_0 - \mu_i - \lambda_0)^2 + \sigma_i^2}} \right\} \\ &= -(n-2) + \sum_{i \in I_2 \cup I_3 \cup I_4} p_i^o. \end{aligned}$$

Since $(\partial/\partial\lambda)\phi_n(c, \lambda)|_{c=c_0, \lambda=\lambda_0} = 0$ and $p_i^o = 0$ for $i \in I_1$, it follows that $\sum_{i=1}^n p_i^o = n - 2$. Taking into account the fact that $p_i^o = 1 - p_i^+ - p_i^-$, we obtain $\sum_{i=1}^n p_i^- + \sum_{i=1}^n p_i^+ = 2$.

Similarly, we have

$$\begin{aligned}\left.\frac{\partial}{\partial c}\phi_n(c,\lambda)\right|_{c=c_0,\lambda=\lambda_0} &= -\sum_{i\in I_1}\frac{\mu_i-c_0}{\sqrt{(\mu_i-c_0)^2+\sigma_i^2}}-\sum_{i\in I_2}\frac{\mu_i-c_0}{2\lambda_0}\\ &\quad -\sum_{i\in I_3}\frac{1}{2}\left\{1+\frac{\mu_i-c_0-\lambda_0}{\sqrt{(\mu_i-c_0-\lambda_0)^2+\sigma_i^2}}\right\}+\sum_{i\in I_4}\frac{1}{2}\left\{1+\frac{c_0-\mu_i-\lambda_0}{\sqrt{(c_0-\mu_i-\lambda_0)^2+\sigma_i^2}}\right\}\\ &= -\sum_{i\in I_1}(p_i^+-p_i^-)-\sum_{i\in I_2}(p_i^+-p_i^-)-\sum_{i\in I_3}p_i^+ +\sum_{i\in I_4}p_i^-,\end{aligned}$$

that is, $(\partial/\partial c)\phi_n(c,\lambda)|_{c=c_0,\lambda=\lambda_0}=-\sum_{i\in I_1\cup I_2\cup I_3}p_i^+ +\sum_{i\in I_1\cup I_2\cup I_4}p_i^-$. From the fact that $p_i^+=0$ for $i\in I_4$ and $p_i^-=0$ for $i\in I_3$ (see Equation (35)), the relation $(\partial/\partial c)\phi_n(c,\lambda)|_{c=c_0,\lambda=\lambda_0}=0$ implies the equality $\sum_{i=1}^n p_i^+=\sum_{i=1}^n p_i^-$, and Equation (38) follows. ■

*Proof of Proposition* 6.1. [(ii) ⇒ (i)]. Suppose we are given a probability matrix $Q$ satisfying (ii). By assumption, $Q$ has vanishing principal diagonal. Define $\boldsymbol{X}=(X_1,\ldots,X_n)$ as in Equation (40). Since $\sum_{(i,j):\,i\neq j}q_{ij}=\sum_{i,j}q_{ij}=1$, this procedure maps $Q$ to a well-defined probability law $\mathcal{L}(\boldsymbol{X})$ on $\mathbb{R}^n$, and the map $Q\mapsto\mathcal{L}(\boldsymbol{X})$ is, obviously, one to one. Due to Inequalities (37), the order statistics of $\boldsymbol{X}$ satisfy

$$X_{1:n}<c_0-\lambda_0<X_{2:n}\le\cdots\le X_{n-1:n}<c_0+\lambda_0<X_{n:n}\quad\text{with probability 1.}$$

Thus, from Lemma 2.1 it follows that, with probability 1,

$$R_n=-(n-2)\lambda_0+\frac{1}{2}\sum_{i=1}^n\{|(X_i-c_0)-\lambda_0|+|(X_i-c_0)+\lambda_0|\}. \tag{A3}$$

The assumptions $Q\in\mathcal{M}(\boldsymbol{p}^+,\boldsymbol{p}^-)$ and $q_{ii}=0$ for all $i$ now show that for any fixed $j$, $\mathbb{P}[X_j=x_j^-]=\sum_{i\neq j}q_{ij}=\sum_{i=1}^n q_{ij}=p_j^-$. Similarly we conclude that for any fixed $i$, $\mathbb{P}[X_i=x_i^+]=\sum_{j\neq i}q_{ij}=\sum_{j=1}^n q_{ij}=p_i^+$. Thus, $\mathbb{P}[X_i=x_i^o]=1-p_i^--p_i^+=p_i^o$, and the marginal $X_i$ of $\boldsymbol{X}$ is the extremal random variable in $\mathcal{F}_1(\mu_i,\sigma_i)$. That is, it has mean $\mu_i$, variance $\sigma_i^2$, and maximizes $\mathbb{E}\{|(X-c_0)-\lambda_0|+|(X-c_0)+\lambda_0|\}$ as $X$ varies in $\mathcal{F}_1(\mu_i,\sigma_i)$. Since this holds for all $i$, taking expectations in Equation (A3) we see that

$$\mathbb{E}R_n=-(n-2)\lambda_0+\frac{\lambda_0}{2}\sum_{i=1}^n U\left(\frac{\mu_i-c_0}{\lambda_0},\frac{\sigma_i}{\lambda_0}\right)=\rho_n,$$

completing the proof.

[(i) ⇒ (ii)]. Assumptions $\boldsymbol{X}\in\mathcal{F}_n(\boldsymbol{\mu},\boldsymbol{\sigma})$ and $\mathbb{E}R_n=\rho_n$ imply that (repeat the proof of Theorem 2.1)

$$\begin{aligned}\rho_n=\mathbb{E}R_n&\le\mathbb{E}\left\{-(n-2)\lambda_0+\frac{1}{2}\sum_{i=1}^n[|(X_i-c_0)-\lambda_0|+|(X_i-c_0)+\lambda_0|]\right\}\\ &=-(n-2)\lambda_0+\frac{1}{2}\sum_{i=1}^n\mathbb{E}\{|(X_i-c_0)-\lambda_0|+|(X_i-c_0)+\lambda_0|\}\\ &\le-(n-2)\lambda_0+\frac{\lambda_0}{2}\sum_{i=1}^n U\left(\frac{\mu_i-c_0}{\lambda_0},\frac{\sigma_i}{\lambda_0}\right)=\phi_n(c_0,\lambda_0)=\rho_n.\end{aligned}$$

Thus, all displayed inequalities are attained as equalities. In view of Lemma 2.1 and Corollary 2.1, this can happen only if the law $\mathcal{L}(\boldsymbol{X})$ of the given random vector $\boldsymbol{X}=(X_1,\ldots,X_n)$ satisfies

$$\begin{aligned}&\text{(a)}\quad \mathbb{P}[X_{1:n}\le c_0-\lambda_0\le X_{2:n}\le\cdots\le X_{n-1:n}\le c_0+\lambda_0\le X_{n:n}]=1\quad\text{and}\\ &\text{(b)}\quad X_i\text{ is extremal in }\mathcal{F}_1(\mu_i,\sigma_i)\text{ for all }i,\ \text{or, equivalently,}\\ &\mathbb{P}[X_i=x_i^-]=p_i^-,\ \mathbb{P}[X_i=x_i^o]=p_i^o,\ \mathbb{P}[X_i=x_i^+]=p_i^+,\ i=1,\ldots,n.\end{aligned} \tag{A4}$$

Taking into account (37) we conclude that Condition (A4) can happen only if the (essential) support of $\boldsymbol{X}$ is contained in the set

$$S:=\{\boldsymbol{x}_{ij},\ i\neq j,\ i,j=1,\ldots,n\},$$

with $\boldsymbol{x}_{ij}$ as in Equation (40). We can thus define the $n\times n$ matrix $Q$ as follows:

$$Q:=(q_{ij}),\quad\text{with } q_{ii}:=0,\quad q_{ij}:=\mathbb{P}[\boldsymbol{X}=\boldsymbol{x}_{ij}],\quad i\neq j,\ i,j=1,\ldots,n. \tag{A5}$$

By definition, $Q$ has vanishing principal diagonal and nonnegative entries, and the relation $\mathbb{P}[\boldsymbol{X}\in S]=1$ implies that $Q$ is a probability matrix. By the assumption $\boldsymbol{X}\in\mathcal{F}_n(\boldsymbol{\mu},\boldsymbol{\sigma})$ and $\mathbb{E}R_n=\rho_n$, the marginal $X_i$ of $\boldsymbol{X}$ has to fulfil

Condition (A4)(b), that is, $\sum_{j=1}^{n} q_{ij} = \sum_{j\neq i} q_{ij} = \sum_{j} \mathbb{P}[\boldsymbol{X} = \boldsymbol{x}_{ij}] = \mathbb{P}[X_i = x_i^+] = p_i^+$; similarly, $\sum_{i=1}^{n} q_{ij} = p_j^-$. Therefore, we have constructed a matrix $Q \in \mathcal{M}(\boldsymbol{p}^+, \boldsymbol{p}^-)$ with $q_{ii} = 0$ for all $i$. Clearly, if two random vectors $\boldsymbol{X}$, $\boldsymbol{Y}$, with $\mathcal{L}(\boldsymbol{Y}) \neq \mathcal{L}(\boldsymbol{X})$, satisfy the assumptions in (i), the corresponding matrices (obtained through Equation (A5)) will be distinct. Consequently, the above procedure determines a one to one mapping $\mathcal{L}(\boldsymbol{X}) \mapsto Q$, completing the proof. ∎

*Proof of Theorem* 7.1. For the infimum, a proof (for any $n \geq 2$) is given in the beginning of Section 2, following the arguments of Bertsimas et al.[19] Regarding the supremum: The key-observation is that Inequality (46) is a special application of the Cauchy–Schwarz inequality,

$$\mathbb{E}R_2 = \mathbb{E}|X_1 - X_2| \leq \sqrt{\mathbb{E}[(X_1 - X_2)^2]} = \sqrt{(\mu_1 - \mu_2)^2 + \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} = \gamma_2.$$

This means that, in order to justify the equality, we have to construct a vector $(X_1, X_2) \in \mathcal{F}_2(\boldsymbol{\mu}, \boldsymbol{\sigma}; \rho)$ such that the random variable $|X_1 - X_2|$ is degenerate. Let $\delta := \operatorname{Var}[X_1 - X_2] = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2 \geq 0$. We distinguish cases $\delta > 0$, $\delta = 0$.

Assume $\delta > 0$, so that $\gamma_2 > 0$. First, we consider a 0–1 Bernoulli random variable $I_p$ with probability of success $p := \frac{1}{2}(1 + (\mu_1 - \mu_2)/\gamma_2)$. Next, we consider another random variable $T$ with mean $\mu_T := \mu_1\sigma_2^2 + \mu_2\sigma_1^2 - \rho\sigma_1\sigma_2(\mu_1 + \mu_2)$ and variance $\sigma_T^2 := \delta\sigma_1^2\sigma_2^2(1 - \rho^2) \geq 0$, stochastically independent of $I_p$. Finally, we define

$$(X_1, X_2) := \frac{1}{\delta}\,[\gamma_2(\sigma_1^2 - \rho\sigma_1\sigma_2)(2I_p - 1) + T,\ \gamma_2(\rho\sigma_1\sigma_2 - \sigma_2^2)(2I_p - 1) + T].$$

It is easily seen that $(X_1, X_2) \in \mathcal{F}_2(\boldsymbol{\mu}, \boldsymbol{\sigma}; \rho)$ and $|X_1 - X_2| = \gamma_2$ with probability 1.

Let us now assume $\delta = 0$. This implies that $X_1 - X_2 = \mu_1 - \mu_2$ with probability 1, and hence, $\sigma_1 = \sigma_2$ and $\rho = 1$. Let $\sigma^2 > 0$ be the common variance and consider the pair $(X_1, X_2) := (\mu_1 + T, \mu_2 + T)$, where $T$ is any random variable with mean zero and variance $\sigma^2$. It follows that $(X_1, X_2)$ satisfies the moment requirements and $|X_1 - X_2| = |\mu_1 - \mu_2| = \gamma_2$ with probability 1. This completes the proof. ∎