Supplementary Materials to Tight tail probability bounds for distribution-free decision making

Ernst Roos[a,*], Ruud Brekelmans[a], Wouter van Eekelen[a], Dick den Hertog[b], Johan S.H. van Leeuwaarden[a]

[a] *Tilburg University, Department of Econometrics and Operations Research, Warandelaan 2, 5037 AB Tilburg, The Netherlands*
[b] *Amsterdam Business School, University of Amsterdam, Plantage Muidergracht 12, 1018 TV Amsterdam, The Netherlands*

Appendix A. Proofs of tail bounds

Proof of Theorem 2. We solve problems (12) and (13) by considering the four scenarios depicted in Figure 3.

Scenario 1a implies $F_m(0) = 0$, $F_m(t) = F_m(b) = 1$, which gives

$$\lambda_0 = \frac{m}{t}, \quad \lambda_1^- = 0, \quad \lambda_1^+ = \frac{t-2m+b}{t}, \quad \lambda_2 = -\frac{1}{t},$$

and objective value

$$\lambda_0 + \lambda_1^+ \frac{1}{2} + \lambda_2 d_m = \frac{b-2d_m}{t} + \frac{1}{2}.$$

Solving the primal problem (12) with probability masses on the points $\{0, t, b\}$ gives

$$\int_x \mathbb{1}\{x \geq t\} \mathrm{d}\mathbb{P}(x) = p_t + p_b = \frac{b-2d_m}{2t} + \frac{1}{2}.$$

Scenario 1b implies that $F(0) = F(t) = F(m) = F(b) = 1$, and hence $\lambda_0 = 1$, $\lambda_1^+ = \lambda_1^- = \lambda_2 = 0$ with objective value 1.

Scenario 2a implies $F(m) = 0$, $F(t) = 1$ which gives

$$\lambda_0 = \lambda_1^- = \lambda_1^+ = 0, \quad \lambda_2 = \frac{1}{t-m},$$

*Corresponding author

Email addresses: e.j.roos@tilburguniversity.edu (Ernst Roos), r.c.m.brekelmans@tilburguniversity.edu (Ruud Brekelmans), w.j.e.c.vaneekelen@tilburguniversity.edu (Wouter van Eekelen), d.denhertog@uva.nl (Dick den Hertog), j.s.h.vanleeuwaarden@tilburguniversity.edu (Johan S.H. van Leeuwaarden)

Preprint submitted to European Journal of Operational Research *November 28, 2021*

with objective value

$$\lambda_2 d_m = \frac{d_m}{t-m}.$$

Solving the optimal probabilities for the primal problem (12) indeed shows that $p_t = \frac{d_m}{t-m}$.

Scenario 2b implies that $F(0) = 0$, $F(m) = F(t) = F(b) = 1$, and

$$\lambda_0 = 1, \quad \lambda_1^- = -1, \quad \lambda_1^+ = \lambda_2 = 0,$$

with objective value

$$\lambda_0 + (\lambda_1^- + \lambda_1^+)\frac{1}{2} = \frac{1}{2}.$$

Solving (12), with support $\{m, t, b\}$, gives $p_m = 1/2$.

□

Proof of Theorem 3. For a random variable X with distribution $\mathbb{P} \in \mathcal{P}_{(\mu,b,d,\beta)}$, we now solve

$$\begin{aligned} \sup_{\mathbb{P}\in\mathcal{M}^+} \quad & \int_x \mathbb{1}_{\{x\geq t\}}\,\mathrm{d}\mathbb{P}(x) \\ \text{s.t.} \quad & \int_x \mathrm{d}\mathbb{P}(x) = 1,\ \int_x x\,\mathrm{d}\mathbb{P}(x) = \mu, \\ & \int_x |x-\mu|\,\mathrm{d}\mathbb{P}(x) = d,\ \int_x \mathbb{1}_{\{x\geq\mu\}}\,\mathrm{d}\mathbb{P}(x) = \beta, \end{aligned} \tag{A.1}$$

which is a semi-infinite linear program with four equality constraints.

Consider the dual of (A.1),

$$\begin{aligned} \inf_{\lambda_0,\lambda_1,\lambda_2,\lambda_3} \quad & \lambda_0 + \lambda_1\mu + \lambda_2 d + \lambda_3\beta \\ \text{s.t.} \quad & \mathbb{1}_{\{x\geq t\}} \leq \lambda_0 + \lambda_1 x + \lambda_2|x-\mu| + \lambda_3\mathbb{1}_{\{x\geq\mu\}},\ \forall x \in [0,b]. \end{aligned} \tag{A.2}$$

Define $F(x) = \lambda_0 + \lambda_1 x + \lambda_2|x-\mu| + \lambda_3\mathbb{1}_{\{x\geq\mu\}}$. Then the inequality in (A.2) can be written as $\mathbb{1}_{\{x\geq t\}} \leq F(x)$, $\forall x$, i.e. $F(x)$ majorizes $\mathbb{1}_{\{x\geq t\}}$. Note that $F(x)$ has both a 'kink' and a jump discontinuity at μ. There are four candidate scenarios, which are described in Figure A.1. When $t \in [0,\mu)$, $F(x)$ touches $\mathbb{1}_{\{x\geq t\}}$ in $\{0,t\} \cup [\mu,b]$ (scenario 1a), or $F(x) = 1$ and touches in $[t,b]$ (scenario 1b). When $t \in [\mu,b]$, $F(x)$ touches in $[0,\mu] \cup \{t\}$ (scenario 2a), or in $[0,\mu) \cup [t,b]$ (scenario 2b).

Scenario 1a implies $F(0) = 0$, $F(t) = F(\mu) = F(b) = 1$, which gives

$$\lambda_0 = \frac{\mu}{2t}, \quad \lambda_1 = \frac{1}{2t}, \quad \lambda_2 = -\frac{1}{2t}, \quad \lambda_3 = \frac{\mu - t}{t},$$

and objective value

$$\lambda_0 + \lambda_1\mu + \lambda_2 d + \lambda_3\beta = \frac{(1-\beta)\mu + \beta t}{t} - \frac{d}{2t}.$$

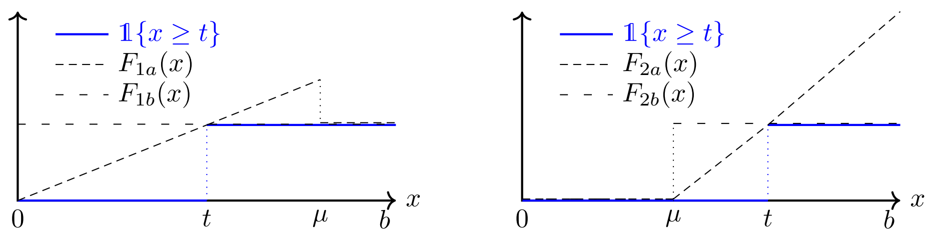

Figure A.1: Scenario 1 and the majorizing functions $F_{1a}(x)$ and $F_{1b}(x)$ under scenarios 1a and 1b, respectively. Scenario 2 and the majorizing functions $F_{2a}(x)$ and $F_{2b}(x)$ under scenarios 2a and 2b, respectively.

Solving the primal problem (A.1) with probability masses on the points $\{0, t, \mu, b\}$ gives

$$\int_x \mathbb{1}\{x \geq t\} \mathrm{d}\mathbb{P}(x) = p_t + p_\mu + p_b = \frac{(1-\beta)\mu + \beta t}{t} - \frac{d}{2t}.$$

Since primal and dual feasible solutions have the same objective value we have strong duality and hence found the optimal solutions.

Scenario 1b implies that $F(0) = F(t) = F(\mu) = F(b) = 1$, and hence $\lambda_0 = 1$, $\lambda_1 = \lambda_2 = \lambda_3 = 0$ with objective value 1. It is clear that the optimal primal objective value is also equal to 1.

Scenario 2a implies $F(0) = F(\mu) = 0$, $F(t) = 1$ which gives

$$\lambda_0 = -\frac{\mu}{2(t-\mu)}, \quad \lambda_1 = \lambda_2 = \frac{1}{2(t-\mu)}, \quad \lambda_3 = 0,$$

with objective value

$$\lambda_0 + \lambda_1\mu + \lambda_2 d + \lambda_3\beta = \frac{d}{2(t-\mu)}.$$

Solving (A.1) with probabilities masses on $\{0, \nu, \mu, b\}$, with $\nu \in (0, \mu)$, indeed shows that $p_t = \frac{d}{2(t-\mu)}$.

Scenario 2b implies that $F(0) = 0$, $F(\mu) = F(t) = F(b) = 1$, which gives as the dual feasible solution

$$\lambda_0 = \lambda_1 = \lambda_2 = 0, \quad \lambda_3 = 1,$$

and objective value

$$\lambda_0 + \lambda_1\mu + \lambda_2 d + \lambda_3\beta = \beta.$$

Finding the optimal probabilities of (A.1) confirms that $p_0 = (1-\beta)$. $\square$

Proof Corollary 4. We solve

$$\begin{aligned} \inf_{\mathbb{P}\in\mathcal{M}^+} \quad & \int_x \mathbb{1}_{\{x>t\}}\,\mathrm{d}\mathbb{P}(x) \\ \text{s.t.} \quad & \int_x \mathrm{d}\mathbb{P}(x) = 1,\ \int_x x\,\mathrm{d}\mathbb{P}(x) = \mu, \\ & \int_x |x-\mu|\,\mathrm{d}\mathbb{P}(x) = d,\ \int_x \mathbb{1}_{\{x\geq\mu\}}\,\mathrm{d}\mathbb{P}(x) = \beta, \end{aligned} \tag{A.3}$$

which is a semi-infinite linear program with four equality constraints. The dual problem is given by

$$\begin{aligned} \sup_{\lambda_0,\lambda_1,\lambda_2,\lambda_3} \quad & \lambda_0 + \lambda_1\mu + \lambda_2 d + \lambda_3\beta \\ \text{s.t.} \quad & \mathbb{1}_{\{x>t\}} \geq \lambda_0 + \lambda_1 x + \lambda_2|x-\mu| + \lambda_3\mathbb{1}_{\{x\geq\mu\}} =: F(x),\ \forall x\in[0,b]. \end{aligned} \tag{A.4}$$

The proof is similar to that of Theorem 3 but, since we are minimizing, $F(x)$ is a minorizing function. Note that $F(x)$ has both a 'kink' and a jump discontinuity at μ. There are four candidate solutions, which are depicted in Figure A.2. When $t\in[0,\mu)$, $F(x)$ touches $\mathbb{1}_{\{x>t\}}$ in $\{t\}\cup[\mu,b]$ (scenario 1a) or in $[0,t]\cup[\mu,b]$ (scenario 1b). When $t\in[\mu,b]$, $F(x)$ touches in $[0,\mu)\cup\{t,b\}$ (scenario 2a), or $F(x)=0$ and touches in $[0,t]$ (scenario 2b).

Scenario 1a implies $F(t)=0$, $F(\mu)=F(b)=1$, which gives the dual solution

$$\lambda_0 = \frac{2t-\mu}{2(t-\mu)}, \quad \lambda_1 = -\frac{1}{2(t-\mu)}, \quad \lambda_2 = \frac{1}{2(t-\mu)}, \quad \lambda_3 = 0,$$

and objective value

$$\lambda_0 + \lambda_1\mu + \lambda_2 d + \lambda_3\beta = 1 - \frac{d}{2(\mu-t)}.$$

Solving the primal problem (A.3) with probability masses on the points $\{t,\mu,\nu,b\}$, with $\nu\in(\mu,b)$, gives

$$\int_x \mathbb{1}\{x>t\}\,\mathrm{d}\mathbb{P}(x) = 1-p_t = 1-\frac{d}{2(\mu-t)}.$$

Scenario 1b implies that $F(0)=F(t)=0$, $F(\mu)=F(b)=1$, and hence $\lambda_0=\lambda_1=\lambda_2=0$, $\lambda_3=1$ with objective value β. Now solving the primal problem with probability masses on $\{0,t,\mu,b\}$ gives us

$$\int_x \mathbb{1}\{x>t\}\,\mathrm{d}\mathbb{P}(x) = p_\mu + p_b = \beta.$$

Scenario 2a implies that $F(0)=F(t)=0$ and $F(b)=1$, which results in

$$\lambda_0 = \frac{\mu}{2(t-b)}, \quad \lambda_1 = \lambda_2 = \frac{1}{2(b-t)}, \quad \lambda_3 = -\frac{(\mu-t)}{(b-t)},$$

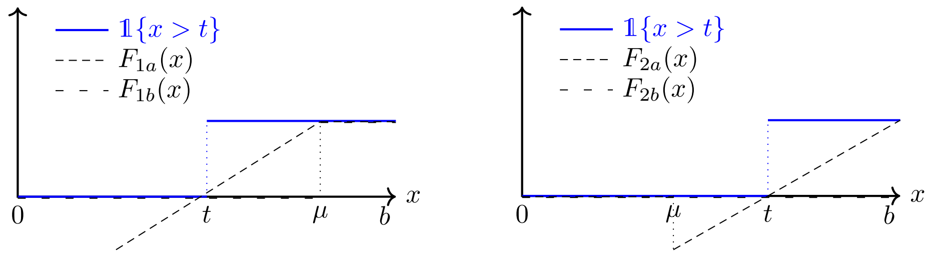

Figure A.2: Scenario 1 and the minorizing functions $F_{1a}(x)$ and $F_{1b}(x)$ under scenarios 1a and 1b, respectively. Scenario 2 and the minorizing functions $F_{2a}(x)$ and $F_{2b}(x)$ under scenarios 2a and 2b, respectively.

with objective value

$$\lambda_0 + \lambda_1\mu + \lambda_2 d + \lambda_3\beta = \frac{\beta(\mu - t)}{(b-t)} + \frac{d}{2(b-t)}.$$

Indeed, solving the primal problem with probability masses on $\{0, \nu, t, b\}$, with $\nu \in (0, \mu)$, gives $p_b = \frac{\beta(\mu-t)}{(b-t)} + \frac{d}{2(b-t)}$.

Scenario 2b implies that $F(0) = F(\mu) = F(t) = F(b) = 0$, which gives the dual feasible solution

$$\lambda_0 = \lambda_1 = \lambda_2 = \lambda_3 = 0,$$

with objective value 0. All probability mass is placed on points that are less than or equal to t. Hence, the optimal primal objective value is also equal to 0. □

Appendix B. Comparison with tight bounds for mean-variance information

Next to the comparison with Cantelli's inequality in Section 2.3, we also look at the tight bounds for the (μ, b, σ)-ambiguity set. De Schepper & Heijnen (1995) provide expressions for these bounds. The upper bound is given by

$$\sup_{\mathbb{P}\in\mathcal{P}_{(\mu,b,\sigma)}} \mathbb{P}(X \geq t) = \begin{cases} 1, & t \in \left[0, \mu - \frac{\sigma^2}{b-\mu}\right], \\ 1 - \frac{\sigma^2 + (b-\mu)(t-\mu)}{bt}, & t \in \left[\mu - \frac{\sigma^2}{b-\mu}, \mu + \frac{\sigma^2}{\mu}\right], \\ \frac{\sigma^2}{\sigma^2 + (t-\mu)^2}, & t \in \left[\mu + \frac{\sigma^2}{\mu}, b\right]. \end{cases} \tag{B.1}$$

The lower bound equals

$$\inf_{\mathbb{P}\in\mathcal{P}_{(\mu,b,\sigma)}} \mathbb{P}(X \geq t) = \begin{cases} \frac{(\mu-t)^2}{(\mu-t)^2+\sigma^2}, & t \in \left[0, \mu - \frac{\sigma^2}{b-\mu}\right], \\ \frac{\sigma^2 + \mu(\mu-t)}{b(b-t)}, & t \in \left[\mu - \frac{\sigma^2}{b-\mu}, \mu + \frac{\sigma^2}{\mu}\right], \\ 0, & t \in \left[\mu + \frac{\sigma^2}{\mu}, b\right]. \end{cases} \tag{B.2}$$

Comparing these bounds with their MAD equivalents is again not straightforward, and we will be using a similar numerical example to compare our results with those of De Schepper & Heijnen (1995). We use the following parameter setting: $a = 0$, $\mu = 1$, $b = 2$, $d = 1/4$. Furthermore, we consider three values for σ: $\sigma = d = 1/4$, $\sigma = 1/3$, and $\sigma = 1/2$.

Figure B.3 shows the upper bounds for mean-MAD ambiguity and the (μ, b, σ)-ambiguity sets. The difference with Cantelli's inequality essentially lies in the behavior of the bound near the mean μ. Expression (B.1) sharpens the bounds of the tail probability for $t \in [\mu - \sigma^2/(b - \mu), \mu + \sigma^2/\mu]$ by using information about the upper bound of the support. Note that Cantelli's inequality and the tight upper bound are equivalent in the tail. Another interesting observation is the equivalence of the mean-variance and mean-MAD bound for $t = \mu$ and $\sigma = \sqrt{db/2}$. Figure B.4 paints a similar picture for the lower bounds.

In Figure B.5, all three obtained upper bounds as well as Cantelli's bound are depicted for four different sets of parameters from different continuous distributions. The first two distributions, Beta(2,2) and Uniform(0,1), are symmetrical, i.e., have equal mean and median, which causes some overlap between the (μ, b, d, β) and (m, b, d_m) bounds. For these four distributions, the (μ, b, d, β) bound generally dominates Cantelli's bound for most t. The only exception to this is found in Figure B.5c, where the mean-variance bound is tighter in the tail, i.e., for large t.

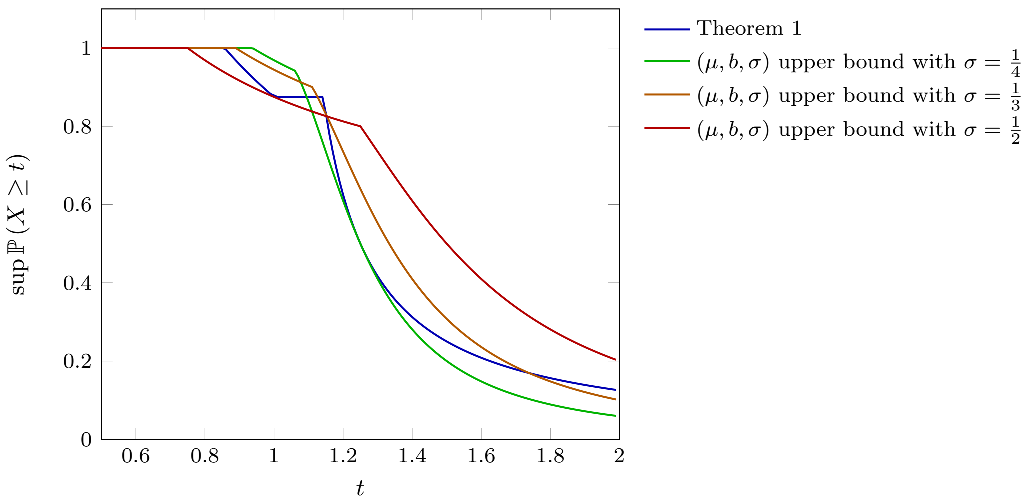

Figure B.3: A comparison of the mean-MAD upper bound with the upper bound of De Schepper & Heijnen (1995) for three different values of σ with the parameter chosen as follows: $a = 0$, $\mu = 1$, $b = 2$ and $d = \frac{1}{4}$.

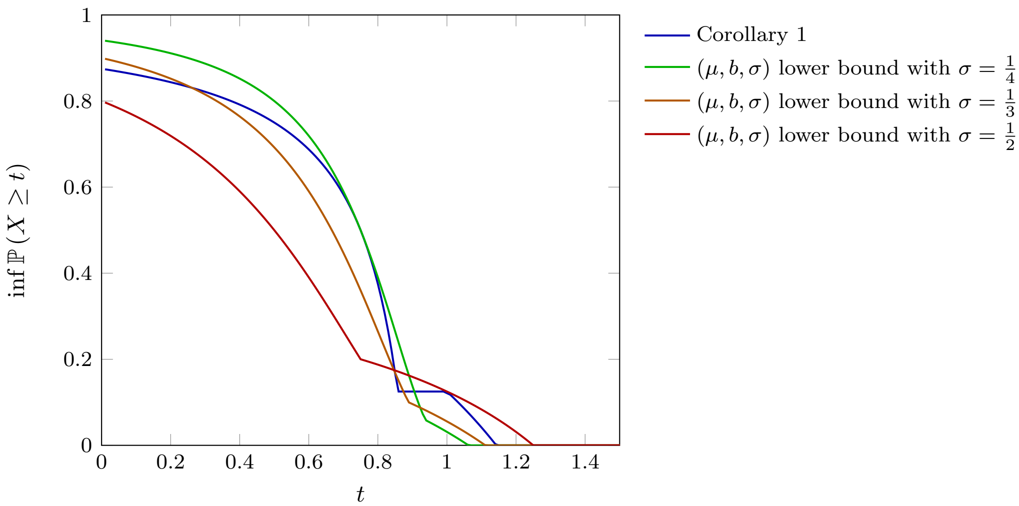

Figure B.4: A comparison of the mean-MAD lower bound with the lower bound of De Schepper & Heijnen (1995) for three different values of σ with the parameters chosen as follows: $a = 0$, $\mu = 1$, $b = 2$ and $d = \frac{1}{4}$.

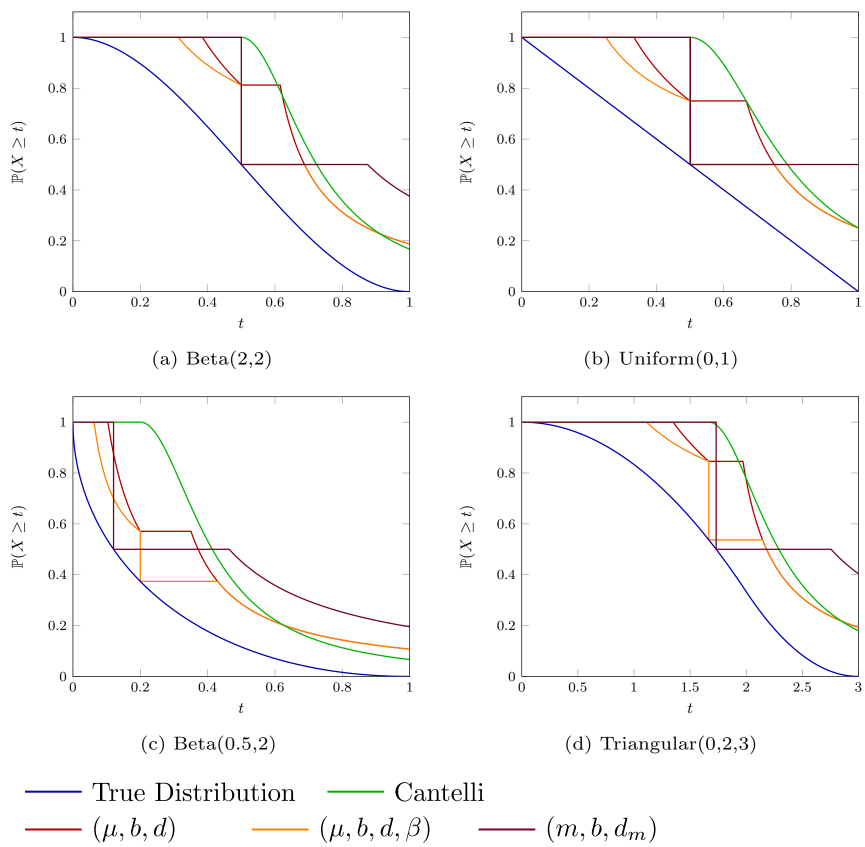

Figure B.5: Comparison of our bounds and variance based bounds for different distributions.