# Supplementary Material: Towards Practical Mean Bounds for Small Samples

- In Section A we describe the computation of Anderson's bound and present more experiments.
- In Section B, as noted in Section 1, we show a log-normal distribution where the sample mean distribution is visibly skewed when $n = 80$.
- In Section C we present the proofs of Section 2.1.2.
- In Section D we discuss the Monte Carlo convergence result of our approximation in Section 3.
- In Section E.1 we show that our bound reduces to the Clopper-Pearson bound for binomial distributions as mentioned in Section 4.1. In Section E.2 we present the proofs of Section 4.2.

## A. Other Experiments

In this section we perform experiments to find an upper bound of the mean of distributions given a finite upper bound of the support, or to a lower bound of the mean of distributions given a finite lower bound of the support. We find the lower bound of the mean of a random variable $X$ by finding the upper mean bound of $-X$ and negating it to obtain the lower mean bound of $X$.

First we describe the computation of Anderson's bound with $\mathbf{u}^{\text{And}}$ defined in Eq. 51. We compute $\beta(n)$ through Monte Carlo simulations. $\beta(n)$ is the value such that:

$$\mathbb{P}_{\mathbf{U}}(\, \forall i : 1 \leq i \leq n, U_{(i)} \geq i/n - \beta(n)) = 1 - \alpha. \tag{58}$$

Therefore

$$\mathbb{P}_{\mathbf{U}}(\beta(n) \geq \max_{i:1 \leq i \leq n} (i/n - U_{(i)})) = 1 - \alpha. \tag{59}$$

For each sample size $n$, we generate $L = 1{,}000{,}000$ samples $\mathbf{U}^j \in [0,1]^n, 1 \leq j \leq L$. For each sample $\mathbf{U}^j \in [0,1]^n$ we compute

$$\beta(n)_j = \max_{i:1 \leq i \leq n} i/n - U^j_{(i)}.$$

Let $\beta(n)_1 \leq \cdots \leq \beta(n)_L$ be the sorted values from $L$ samples. We output $\hat{\beta}(n) = \beta(n)_{(\lceil (1-\alpha)L \rceil)}$ as an approximation of $\beta(n)$.

For each experiment, we used $\alpha = 0.05$ unless specified otherwise. We plot the following:

- The expected value of the bounds versus the sample size. For each sample size, we draw 10,000 samples of $\mathbf{x}$, compute the bound for each $\mathbf{x}$ and compute the average.
- For the upper bound of the mean, we plot the $\alpha$-quantile of the bound distribution versus the sample size. For each sample size, we draw 10,000 samples of $\mathbf{x}$, compute the bound for each $\mathbf{x}$ and take the $\alpha$ quantile. If the $\alpha$-quantile is below the true mean, the bound does not have guaranteed coverage.

  For the lower bound of the mean, we plot the $1-\alpha$-quantile of the bound distribution versus the sample size. For each sample size, we draw 10,000 samples of $\mathbf{x}$, compute the bound for each $\mathbf{x}$ and take the $1-\alpha$ quantile. If the $1-\alpha$-quantile is above the true mean, the bound does not have guaranteed coverage.
- Coverage of the bounds. For each value of $\alpha$ from 0.02 to 1 with a step size of 0.02, we draw 10,000 samples of $\mathbf{x}$, compute the bound for each $\mathbf{x}$ and plot the percentage of the bounds that are greater than or equal to the true mean (denoted *coverage*). If this percentage is larger than $1-\alpha$, the bound has guaranteed coverage.

We perform the following experiments:

1. For the case in which we know a superset $D^+$ of the distribution's support with a finite lower bound and a finite upper bound (the 2-ended support setting), we compare the following bounds:
   - Anderson's bound.
   - New bound with the function $T$ being Anderson's bound.
   - Student's $t$.
   - Hoeffding's bound.
   - Maurer and Pontil's bound.

   We find an upper bound of the mean for the following distributions:
   - $\beta(1, 5)$, $\texttt{uniform}(0, 1)$ and $\beta(5, 1)$. The known superset of the support is $[0, 1]$. The result is in Figure 6.
   - $\beta(0.5, 0.5)$, $\beta(1, 1)$ and $\beta(2, 2)$. The known superset of the support is $[0, 1]$. The result is in Figure 7.
   - $\texttt{binomial}(10, 0.1)$, $\texttt{binomial}(10, 0.5)$ and $\texttt{binomial}(10, 0.9)$. The known superset of the support is the interval $[0, 10]$. The result is in Figure 8.
2. We also consider the case in which we want an upper bound of the mean without knowing the lower bound of the support (or to find a lower bound without knowing an upper bound of the support). In the main paper we referred to this as the 1-ended support setting. Since Hoeffding's and Maurer and Pontil's bounds require knowing both a finite lower bound and upper bound, they are not applicable in this setting. We compare the following bounds:
   - Anderson's bound.
   - New bound with $T$ being Anderson's bound.
   - Student's $t$

   We address the following distributions:
   - $\beta(1, 5)$, $\texttt{uniform}(0, 1)$ and $\beta(5, 1)$. The known superset of the support is $(-\infty, 1]$. We find the upper bound of the mean. The result is in Figure 9.
   - $\texttt{binomial}(10, 0.1)$, $\texttt{binomial}(10, 0.5)$ and $\texttt{binomial}(10, 0.9)$. The known superset of the support is $(-\infty, 10]$. We find the upper bound of the mean. The result is in Figure 10.
   - $\texttt{poisson}(2)$, $\texttt{poisson}(10)$ and $\texttt{poisson}(50)$. The known superset of the support is $[0, \infty)$. We find the lower bound of the mean. The result is in Figure 11.

All the experiments confirm that our bound has guaranteed coverage and is equal to or tighter than Anderson's and Hoeffding's.

From the experiments, our upper bound performs the best in distributions that are skewed right (respectively, our lower bound will perform the best in distributions that are skewed left), when we know a tight lower bound and upper bound of the support.

## B. Discussion on Section 1: Skewed Sample Mean Distribution with $n = 80$

In this section, as noted in Section 1, we show a log-normal distribution where the sample mean distribution is visibly skewed when $n = 80$ (Figure 12). Student's $t$ is not a good candidate in this case because the sample mean distribution is not approximately normal. This example is a variation on the one provided by Frost (2021).

While the log-normal distribution is an extreme example of skew, this example illustrates the danger of assuming the validity of arbitrary thresholds on the sample size, such as the traditional threshold of $n = 30$, for using the Student's $t$ method. Clearly there are cases where such a threshold, and even much larger thresholds, are not adequate.

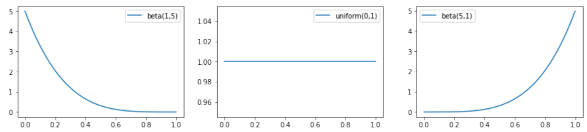
## C. Proof of Section 2.1.2

We restate the lemma and theorem statements for convenience.

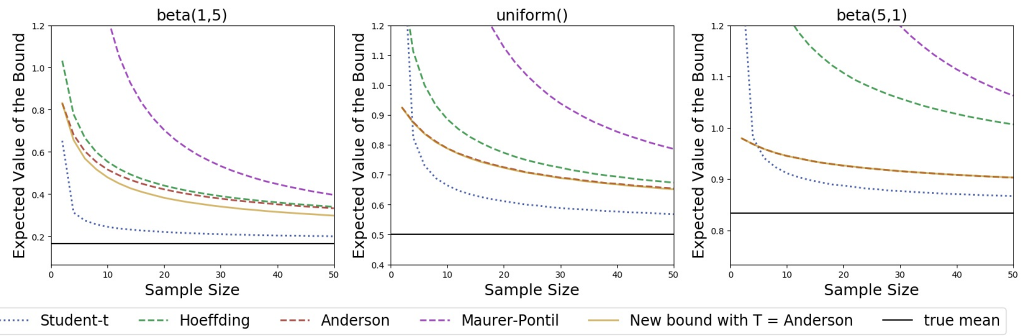
**Lemma C.1** (Lemma 2.2)**.** *Let $X$ be a random variable with CDF $F$ and $Y \stackrel{\text{def}}{=} F(X)$, known as the probability integral transform of $X$. Let $U$ be a uniform random variable on $[0, 1]$. Then for any $0 \leq y \leq 1$,*

$$\mathbb{P}(Y \leq y) \leq \mathbb{P}(U \leq y). \tag{60}$$

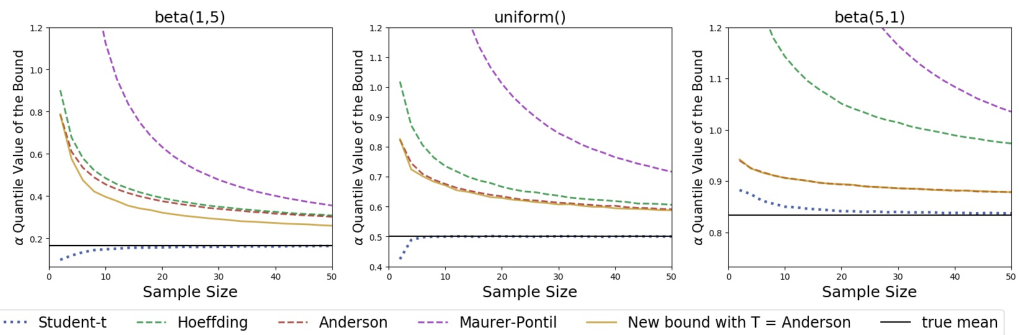
*If $F$ is continuous, then $Y$ is uniformly distributed on $(0, 1)$.*



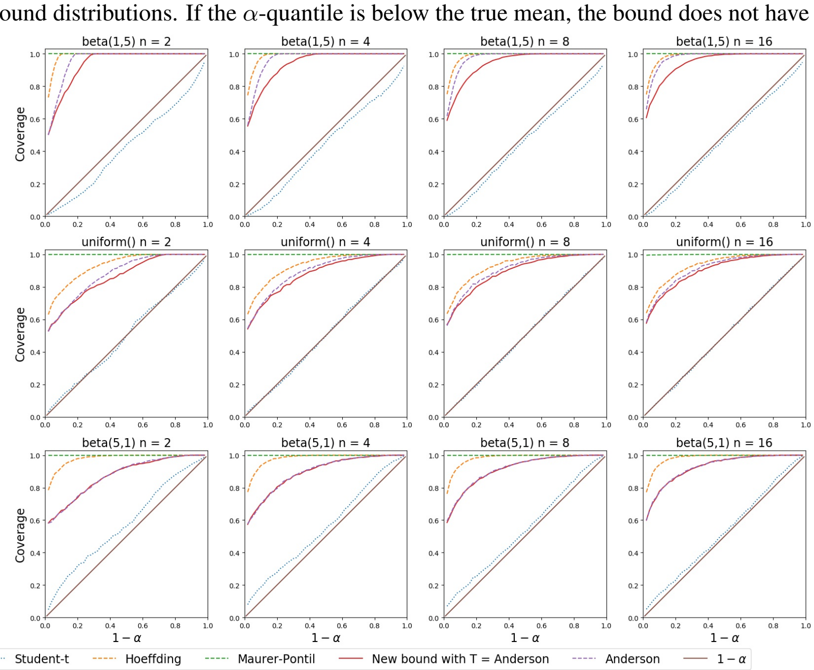

(a) The PDFs of the test distributions.




(b) Expected values of the bounds versus sample size.

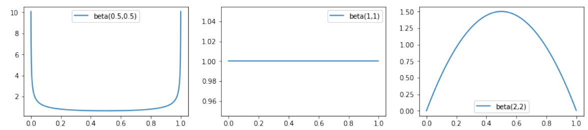



(c) The $\alpha$-quantiles of bound distributions. If the $\alpha$-quantile is below the true mean, the bound does not have guaranteed coverage.

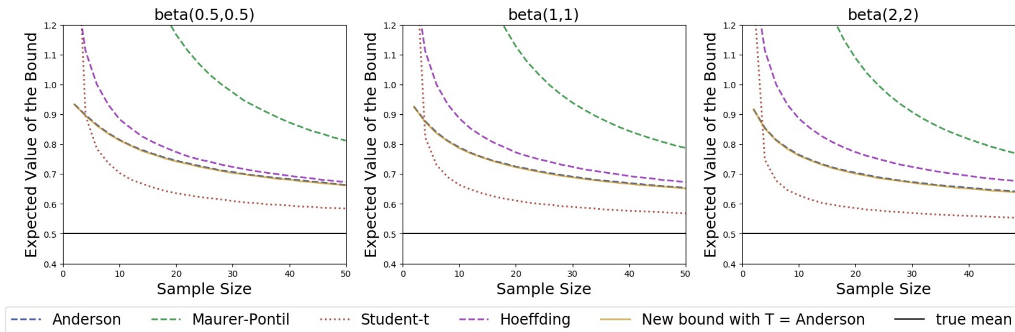



(d) The coverage of the bound. If the coverage is below the line $1 - \alpha$, the bound does not have guaranteed coverage.

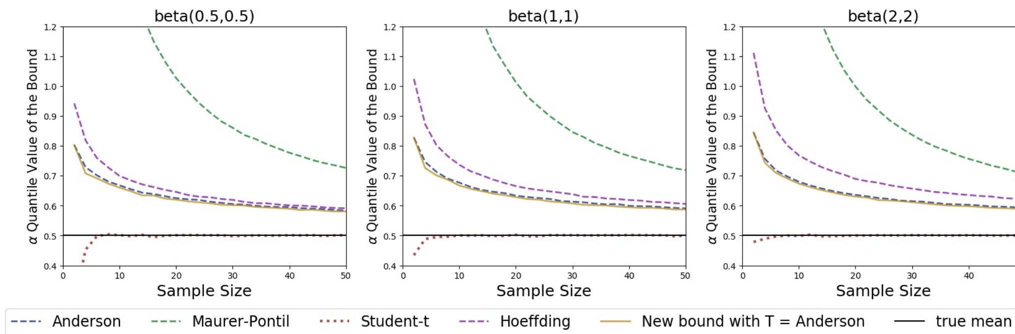
*Figure 6.* Finding the upper bound of the mean with $D^+ = [0, 1]$



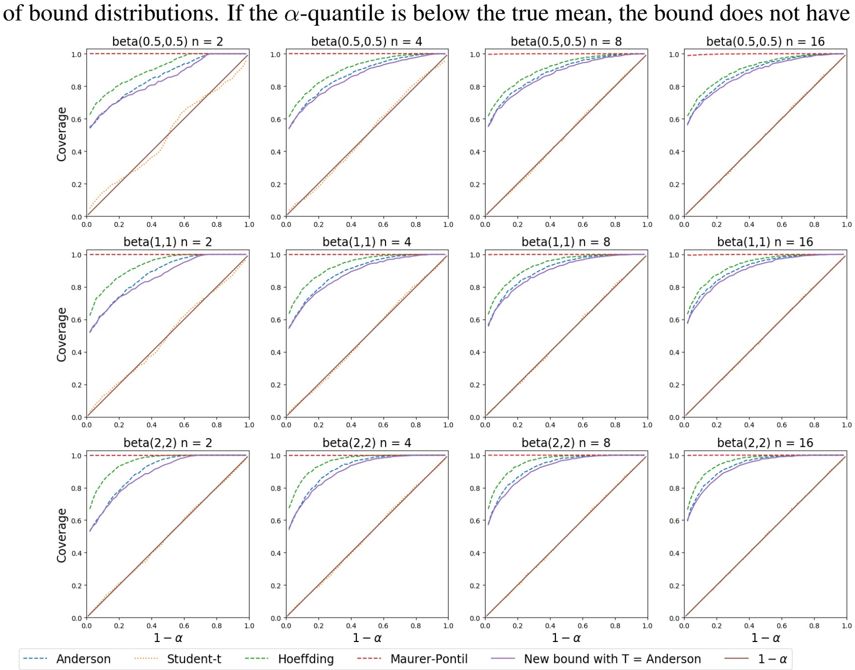

(a) The PDFs of the test distributions.




(b) Expected values of bounds versus sample size.

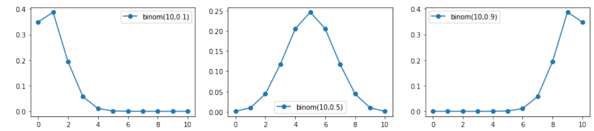



(c) The $\alpha$-quantiles of bound distributions. If the $\alpha$-quantile is below the true mean, the bound does not have guaranteed coverage.

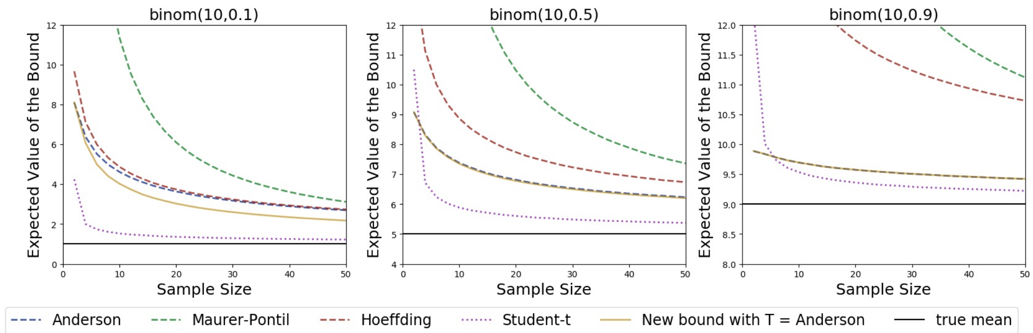



(d) The coverage of the bound. If the coverage is below the line $1 - \alpha$, the bound does not have guaranteed coverage.

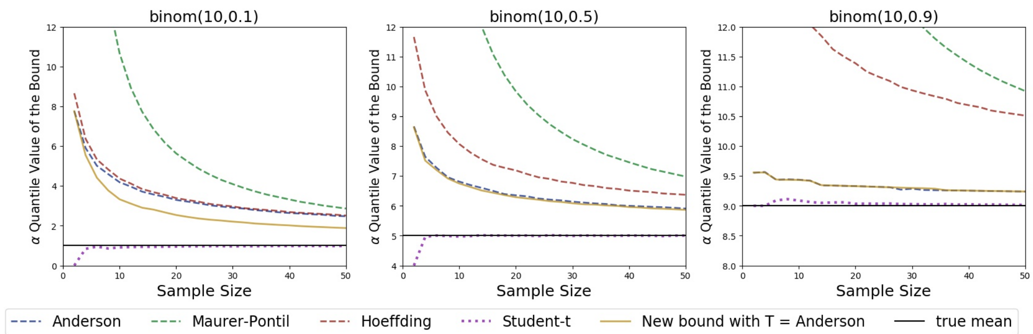
*Figure 7.* Finding the upper bound of the mean with $D^+ = [0, 1]$



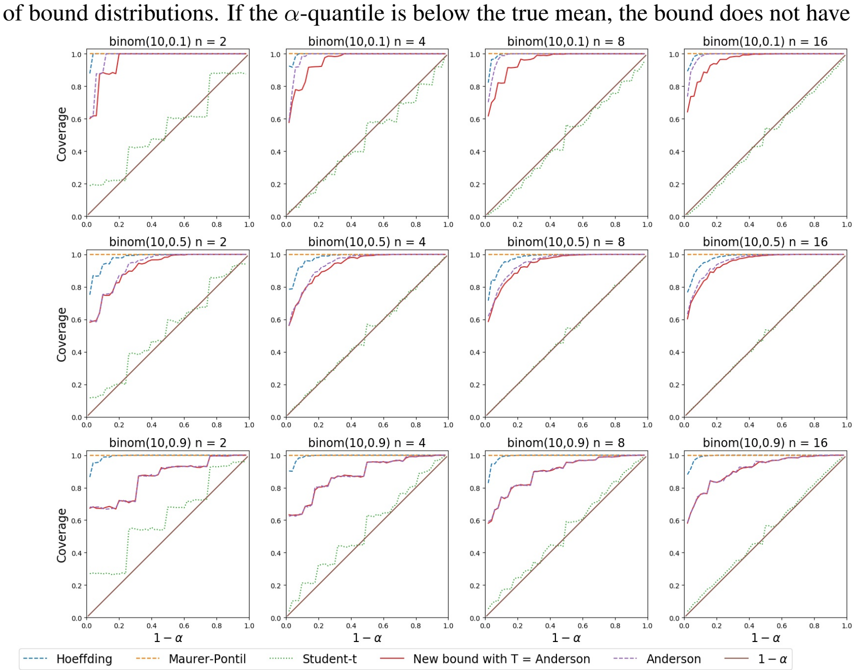

(a) The PMFs of the test distributions.




(b) Expected values of bounds versus sample size.

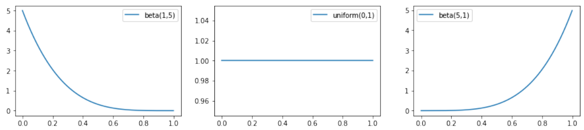



(c) The $\alpha$-quantiles of bound distributions. If the $\alpha$-quantile is below the true mean, the bound does not have guaranteed coverage.

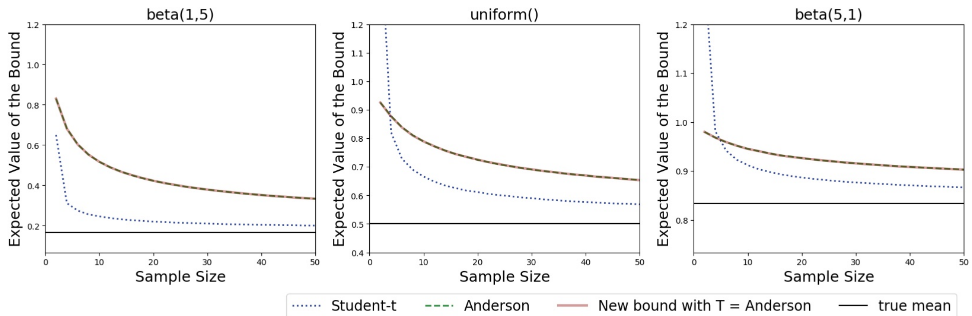



(d) The coverage of the bound. If the coverage is below the line $1 - \alpha$, the bound does not have guaranteed coverage.

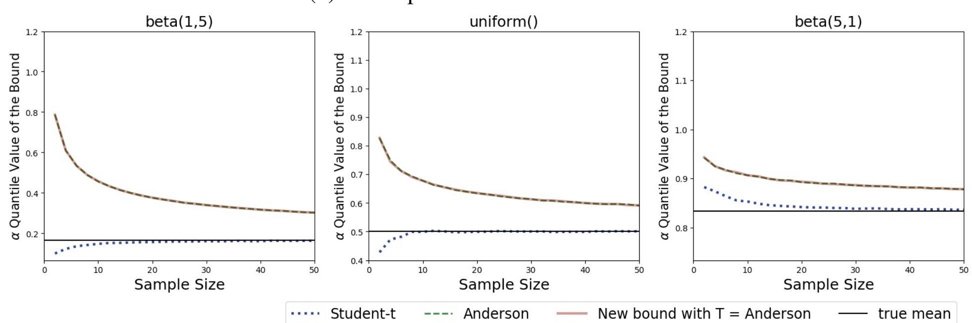
*Figure 8.* Finding the upper bound of the mean with $D^+ = [0, 10]$



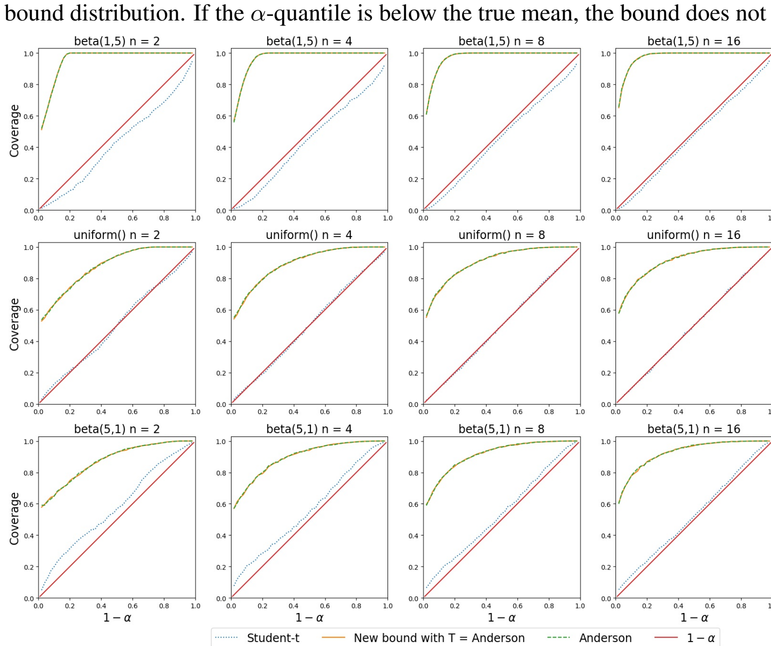

(a) The PDFs of the test distributions.




(b) The expected value of the bounds.

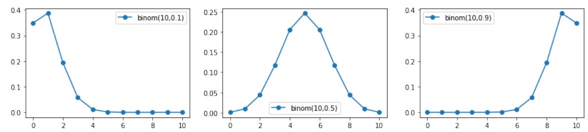



(c) The $\alpha$-quantile of the bound distribution. If the $\alpha$-quantile is below the true mean, the bound does not have guaranteed coverage.

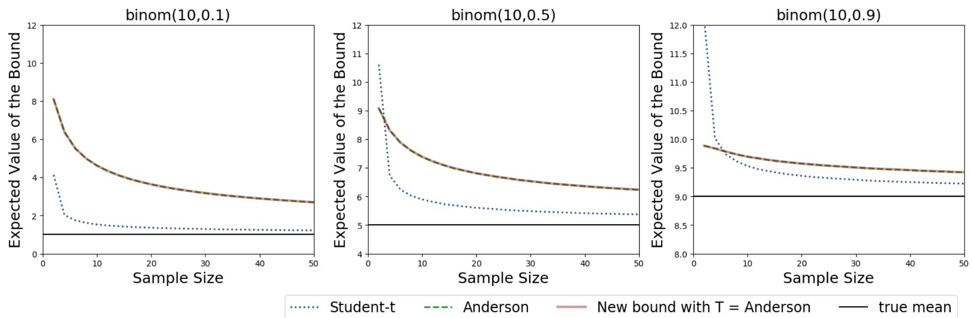



(d) The coverage of the bound. If the coverage is below the line $1 - \alpha$, the bound does not have guaranteed coverage.

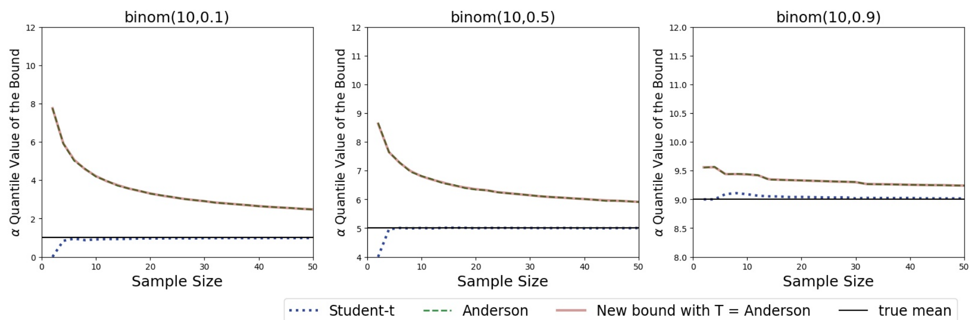
*Figure 9.* Finding the upper bound of the mean with $D^+ = (-\infty, 1]$



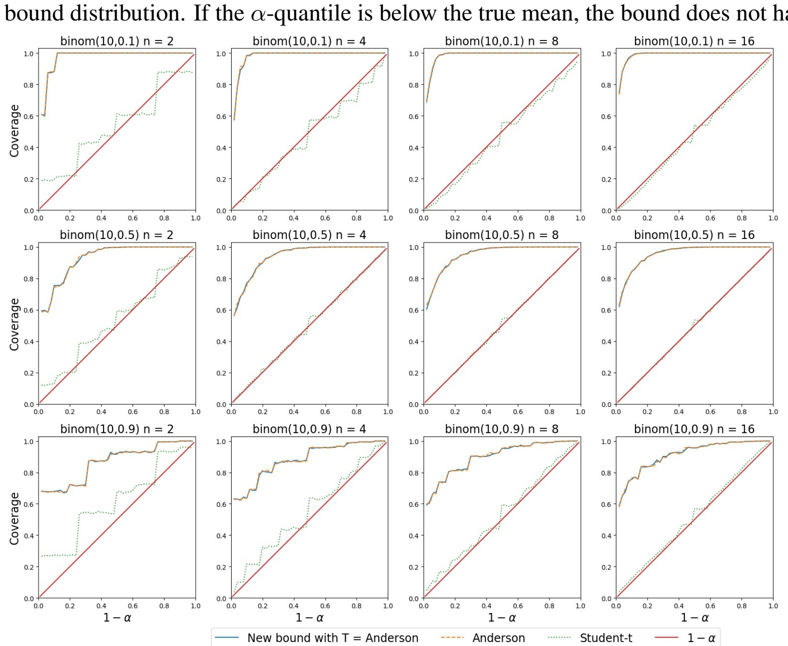

(a) The PMF of the test distributions.




(b) The expected value of the bounds.

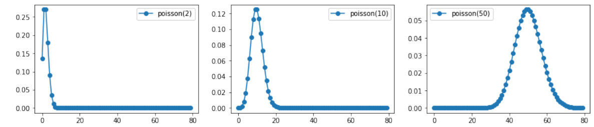



(c) The $\alpha$-quantile of the bound distribution. If the $\alpha$-quantile is below the true mean, the bound does not have guaranteed coverage.

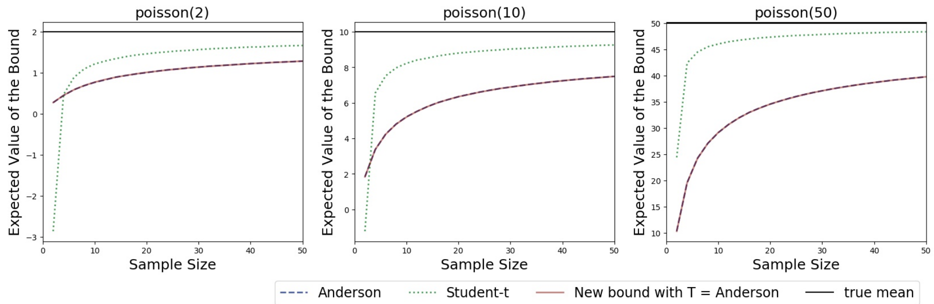



(d) The coverage of the bound. If the coverage is below the line $1-\alpha$, the bound does not have guaranteed coverage.

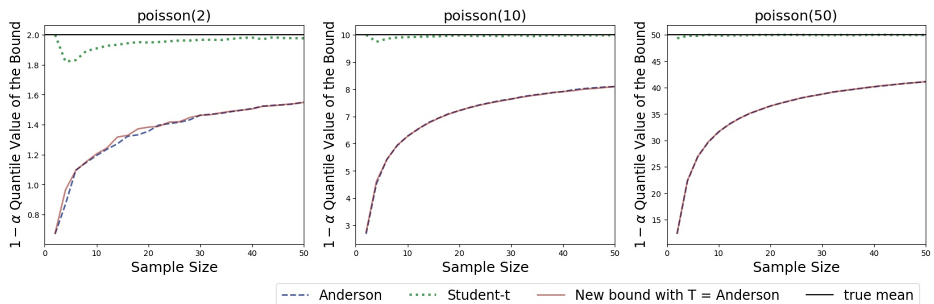
*Figure 10.* Finding the upper bound of the mean with $D^+ = (-\infty, 10]$



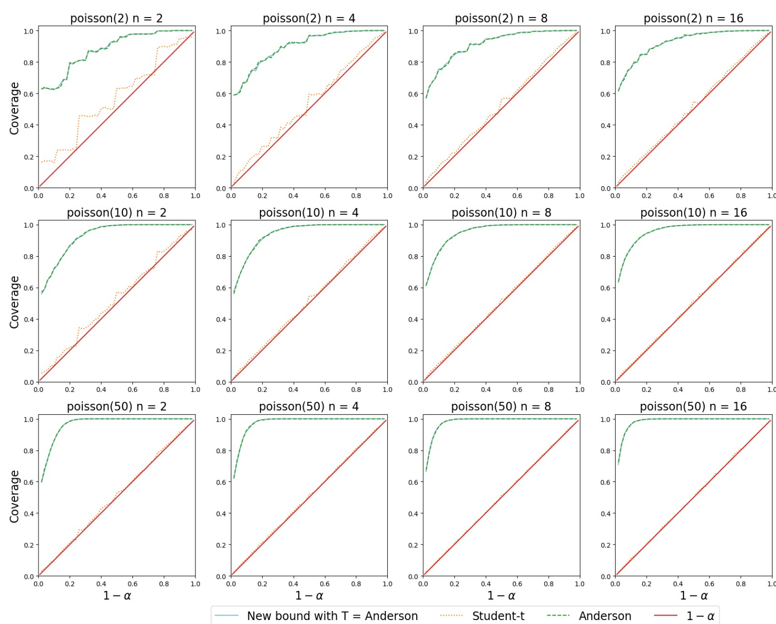

(a) The PMF of the test distributions.




(b) The expected value of the bounds.




(c) The $1-\alpha$-quantile of the bound distribution. If the $1-\alpha$-quantile is above the true mean, the bound does not have guaranteed coverage.




(d) The coverage of the bound. If the coverage is below the line $1-\alpha$, the bound does not have guaranteed coverage.

*Figure 11.* Finding the lower bound of the mean with $D^+ = [0, \infty)$

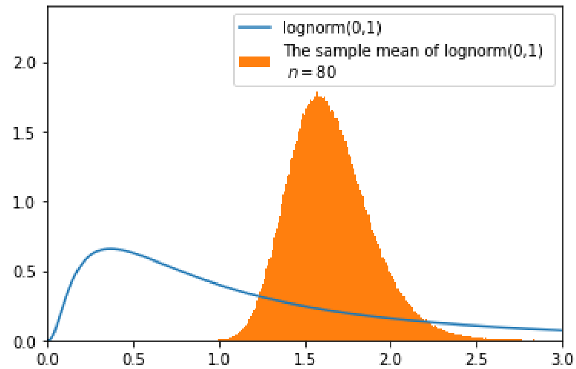



*Figure 12.* The PDFs of $lognorm(0, 1)$ and the sample mean distribution of $lognorm(0, 1)$. The sample mean distribution of $lognorm(0, 1)$ is visibly skewed when the sample size $n = 80$.

*Proof.* Since $\mathbb{P}(U \leq y) = y$, we will show that $\mathbb{P}(Y \leq y) \leq y$.

We will first show that if $F(x) \leq y$, then $x \leq \sup\{z : F(z) \leq y\}$. Suppose that $x > \sup\{z : F(z) \leq y\}$. Then, $F(x) > y$. Therefore,

$$F(x) \leq y \text{ implies } x \leq \sup\{z : F(z) \leq y\}. \tag{61}$$

Now we have

$$\begin{aligned}
\mathbb{P}(Y \leq y) &= \mathbb{P}(F(X) \leq y) &(62)\\
&\leq \mathbb{P}(X \leq \sup\{z : F(z) \leq y\}) &(63)\\
&= F(z^*) \text{ where } z^* = \sup\{z : F(z) \leq y\} &(64)\\
&\leq y. &(65)
\end{aligned}$$

If $F$ is continuous, Angus (1994) shows that $Y$ is uniformly distributed on $(0, 1)$. □

**Lemma C.2** (Lemma 2.3)**.** *For any* $\mathbf{x} \in D^n$,

$$m_D(\mathbf{x}, F(\mathbf{x})) \geq \mu. \tag{66}$$

*Proof.* We present the proof with the notation for when $F$ is continuous. The proof when $F$ has discontinuity is a trivial extension. We have:

$$m_D(\mathbf{x}, F(\mathbf{x})) = \sum_{i=1}^{n+1} x_{(i)}(F(x_{(i)}) - F(x_{(i-1)})) \tag{67}$$

where $x_{(0)} \overset{\text{def}}{=} -\infty$ and $x_{(n+1)} \overset{\text{def}}{=} s_D$. Then:

$$\begin{aligned}
\mu &= \int x \, dF(x) &(68)\\
&= \sum_{i=1}^{n+1} \int_{x_{(i-1)}}^{x_{(i)}} x \, dF(x) &(69)\\
&\leq \sum_{i=1}^{n+1} \int_{x_{(i-1)}}^{x_{(i)}} x_{(i)} \, dF(x) &(70)\\
&= \sum_{i=1}^{n+1} x_{(i)}(F(x_{(i)}) - F(x_{(i-1)})) &(71)\\
&= m_D(\mathbf{x}, F(\mathbf{x})). &(72)
\end{aligned}$$

□

**Lemma C.3** (Lemma 2.4). *Let $\mathbf{Z}$ be a random sample of size $n$ from $F$. Let $\mathbf{U} = U_1, ..., U_n$ be a sample of size $n$ from the continuous uniform distribution on $[0,1]$. For any function $T : D^n \to R$ and any $\mathbf{x} \in D^n$,*

$$\mathbb{P}_{\mathbf{Z}}(T(\mathbf{Z}) \leq T(\mathbf{x})) \leq \mathbb{P}_{\mathbf{U}}(b(\mathbf{x}, \mathbf{U}) \geq \mu). \tag{73}$$

*Proof.* Let $\cup$ denote the union of events and $\{\}$ denote an event. Let $\mathbf{Z}$ be a sample from $F$. Then for any sample $\mathbf{x}$:

$$\mathbb{P}_{\mathbf{Z}}(T(\mathbf{Z}) \leq T(\mathbf{x})) \tag{74}$$
$$= \mathbb{P}_{\mathbf{Z}}(\mathbf{Z} \in \mathbb{S}(\mathbf{x})) \tag{75}$$
$$= \mathbb{P}_{\mathbf{Z}}(\cup_{\mathbf{y} \in \mathbb{S}(\mathbf{x})} \{\mathbf{Z} = \mathbf{y}\}) \tag{76}$$
$$\leq \mathbb{P}_{\mathbf{Z}}(\cup_{\mathbf{y} \in \mathbb{S}(\mathbf{x})} \mathbf{Z} \preceq \mathbf{y}) \text{ because } \mathbf{Z} = \mathbf{y} \text{ implies } \mathbf{Z} \preceq \mathbf{y} \tag{77}$$
$$\leq \mathbb{P}_{\mathbf{Z}}(\cup_{\mathbf{y} \in \mathbb{S}(\mathbf{x})} \{F(\mathbf{Z}) \preceq F(\mathbf{y})\}) \tag{78}$$

because $F$ is non-decreasing, so $Z_{(i)} \leq y_{(i)}$ implies $F(Z_{(i)}) \leq F(y_{(i)})$. Let $U_1, \cdots, U_n$ be $n$ samples from the uniform distribution on $(0,1)$. From Lemma 2.2, for any $u \in (0,1)$, $\mathbb{P}(F(Z_i) \leq u) \leq \mathbb{P}(U_i \leq u)$. Therefore

$$\mathbb{P}_{\mathbf{Z}}(\cup_{\mathbf{y} \in \mathbb{S}(\mathbf{x})} \{F(\mathbf{Z}) \preceq F(\mathbf{y})\}) \tag{79}$$
$$\leq \mathbb{P}_{\mathbf{U}}(\cup_{\mathbf{y} \in \mathbb{S}(\mathbf{x})} \{\mathbf{U} \preceq F(\mathbf{y})\}). \tag{80}$$

Recall that $m_D(\mathbf{y}, \mathbf{U}) = s_D - \sum_{i=1}^n U_{(i)}(y_{(i+1)} - y_{(i)})$ where $\forall i, y_{(i+1)} - y_{(i)} \geq 0$. Therefore if $\forall i, U_{(i)} \leq F(y_{(i)})$ then $m_D(\mathbf{y}, \mathbf{U}) \geq m_D(\mathbf{y}, F(\mathbf{y}))$:

$$\mathbb{P}_{\mathbf{U}}(\cup_{\mathbf{y} \in \mathbb{S}(\mathbf{x})} \{\mathbf{U} \preceq F(\mathbf{y})\}) \tag{81}$$
$$\leq \mathbb{P}_{\mathbf{U}}(\cup_{\mathbf{y} \in \mathbb{S}(\mathbf{x})} \{m_D(\mathbf{y}, \mathbf{U}) \geq m_D(\mathbf{y}, F(\mathbf{y}))\}) \tag{82}$$
$$\leq \mathbb{P}_{\mathbf{U}}(\cup_{\mathbf{y} \in \mathbb{S}(\mathbf{x})} \{m_D(\mathbf{y}, \mathbf{U}) \geq \mu\}), \text{by Lemma 2.3} \tag{83}$$
$$\leq \mathbb{P}_{\mathbf{U}}(\sup_{\mathbf{y} \in \mathbb{S}(\mathbf{x})} m_D(\mathbf{y}, \mathbf{U}) \geq \mu) \tag{84}$$
$$= \mathbb{P}_{\mathbf{U}}(b(\mathbf{x}, \mathbf{U}) \geq \mu) \tag{85}$$

The inequality in Eq. 84 is because if there exists $\mathbf{y} \in \mathbb{S}(\mathbf{x})$ such that $m_D(\mathbf{y}, \mathbf{U}) \geq \mu$, then $\sup_{\mathbf{y} \in \mathbb{S}(\mathbf{x})} m_D(\mathbf{y}, \mathbf{U}) \geq \mu$. Therefore the event $\cup_{\mathbf{y} \in \mathbb{S}(\mathbf{x})} \{m_D(\mathbf{y}, \mathbf{U}) \geq \mu\}$ is a subset of the event $\sup_{\mathbf{y} \in \mathbb{S}(\mathbf{x})} m_D(\mathbf{y}, \mathbf{U}) \geq \mu$, and Eq. 84 follows.

From Eqs. 78, 80 and Eq. 85:

$$\mathbb{P}_{\mathbf{Z}}(T(\mathbf{Z}) \leq T(\mathbf{x})) \quad \leq \quad \mathbb{P}_{\mathbf{U}}(b(\mathbf{x}, \mathbf{U}) \geq \mu). \tag{86}$$

□

## D. Discussion on Section 3: Monte Carlo Convergence

In Section 3, we discussed the use of Monte Carlo sampling of the induced mean function $b(\mathbf{x}, \mathbf{U})$ via sampling of the uniform random variable $U$, to approximate the $1-\alpha$ quantile of $b(\mathbf{x}, \mathbf{U})$. Let $\hat{q}_\ell$ denote the output of the Monte Carlo algorithm (Algorithm 1) using $\ell$ Monte Carlo samples. In this section we show that our estimator converges to the true quantile as the number of Monte Carlo samples grows, and, given a desired threshold $\epsilon$, we can compute an upper bound at most $Q(1-\alpha, b(\mathbf{x}, \mathbf{U})) + \epsilon$ with guaranteed coverage.

**Theorem D.1.** *Let $\epsilon > 0$. Let $\gamma = \min\left(\alpha, \left(\frac{\epsilon}{3(s_D - \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} z_{(1)})}\right)^n\right)$. Use $\ell = \left\lceil \frac{-\ln(\gamma/2)}{2} \left(\frac{3(s_D - \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} z_{(1)})}{\epsilon}\right)^n \right\rceil$ Monte Carlo samples to compute $Q(1-\alpha+\gamma, b(\mathbf{x}, \mathbf{U}))$ using Algorithm 1. Let $\hat{q}_\ell$ be the output of the algorithm. We output $\hat{q}_\ell + \epsilon/3$ as the final estimator.*

*Then with probability at least $1-\alpha$:*

$$\mu \leq \hat{q}_\ell + \epsilon/3 \leq Q(1-\alpha, b(\mathbf{x}, \mathbf{U})) + \epsilon. \tag{87}$$

To prove Theorem D.1, we first show some lemmas.

The Monte Carlo approximation error is quantified in the following lemma due to Serfling (1980). Let $F(m-) \stackrel{\text{def}}{=} \lim_{x \to m^-} F(x)$.

**Lemma D.2** (Theorem 2.3.2 in Serfling (1980))**.** *Let $0 < p < 1$. If $Q(p, M)$ is the unique solution $m$ of $F(m-) \leq p \leq F(m)$, then for every $\epsilon > 0$,*

$$\mathbb{P}(|M_{\lceil pl \rceil} - Q(p, M)| > \epsilon) \leq 2e^{-2l\delta}, \tag{88}$$

*where*

$$\delta = \min\left(p - F(Q(p, M) - \epsilon), F(Q(p, M) + \epsilon) - p\right).$$

Note that when the condition that $Q(p, M)$ is the unique solution $m$ of $F(m-) \leq p \leq F(m)$ is satisfied, $\delta > 0$. Let $M \stackrel{\text{def}}{=} b_{D,T}(\mathbf{x}, \mathbf{U}) \in [0, 1]$. In Lemma D.3 we will show that the CDF of $M$ satisfies the condition in Lemma D.2. Therefore the error incurred by computing the bound via Monte Carlo sampling can be decreased to an arbitrarily small value by choosing a large enough number of Monte Carlo samples $l$. The Monte Carlo estimation of $b^{\alpha}_{D^+,T}(\mathbf{x})$ where $D^+ = [0, 1]$ is presented in Algorithm 1.

We will show that for any $\mathbf{x}$, for any $T$, for any $p \in (0, 1)$, $F_M(m-) \leq p \leq F_M(m)$ has a unique solution by showing that for any $\mathbf{x}$ and $T$, $F_M$ is strictly increasing on its support. To do so, for any $c_1, c_2$ in the support such that $c_1 < c_2$ we will show that

$$F_M(c_2) - F_M(c_1) > 0. \tag{89}$$

**Lemma D.3.** *Let $M \stackrel{\text{def}}{=} b(\mathbf{x}, \mathbf{U})$. Let $F_M$ be the CDF of $M$.*

*For any $\mathbf{x}$, for any scalar function $T$, either:*

1. *$M$ is a constant, or*
2. *For any $c_1, c_2$ such that $0 \leq c_1 < c_2 \leq 1$,*

$$F_M(c_2) - F_M(c_1) \geq \left(\frac{c_2 - c_1}{s_D - \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} z_{(1)}}\right)^n > 0. \tag{90}$$

*Proof.* Recall the definition of the induced mean as

$$b(\mathbf{x}, \boldsymbol{\ell}) = \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} \sum_{i=1}^{n+1} z_{(i)}(\ell_{(i)} - \ell_{(i-1)}), \tag{91}$$

$$= \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} s_D - \sum_{i=1}^{n} \ell_{(i)}(z_{(i+1)} - z_{(i)}), \tag{92}$$

where $\ell_{(0)} \stackrel{\text{def}}{=} 0$, $\ell_{(n+1)} \stackrel{\text{def}}{=} 1$ and $z_{(n+1)} \stackrel{\text{def}}{=} s_D$.

We now find the support of $M$. Let $\phi \stackrel{\text{def}}{=} \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} z_{(1)}$. We will show that for any $\mathbf{u}$ where $0 \leq u_i \leq 1$, we have $\phi \leq b(\mathbf{x}, \mathbf{u}) \leq s_D$, and therefore the support of $M$ is a subset of $[\phi, s_D]$. We have

$$b(\mathbf{x}, \mathbf{u}) = \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} s_D - \sum_{i=1}^{n} u_{(i)}(z_{(i+1)} - z_{(i)}) \tag{93}$$

$$\leq \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} s_D - \sum_{i=1}^{n} 0(z_{(i+1)} - z_{(i)}) \tag{94}$$

$$= s_D. \tag{95}$$

Similarly we have

$$b(\mathbf{x}, \mathbf{u}) = \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} s_D - \sum_{i=1}^{n} u_{(i)}(z_{(i+1)} - z_{(i)}) \tag{96}$$
$$\geq \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} s_D - \sum_{i=1}^{n} 1(z_{(i+1)} - z_{(i)}) \tag{97}$$
$$= \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} s_D - (z_{(n+1)} - z_{(1)}) \tag{98}$$
$$= \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} z_{(1)} \tag{99}$$
$$= \phi. \tag{100}$$

Therefore $M = b(\mathbf{x}, \mathbf{U}) \in [\phi, s_D]$. We consider two cases: where $\phi = s_D$ and where $\phi < s_D$.

**Case 1: $\phi = s_D$.**

Then for all $i$, $\sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} z_{(i)} \geq \phi = s_D$. Since $z_{(i)} \leq s_D$, we have for all $i$, $\sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} z_{(i)} = s_D$. Therefore

$$b(\mathbf{x}, \boldsymbol{\ell}) = \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} \sum_{i=1}^{n+1} z_{(i)}(\ell_{(i)} - \ell_{(i-1)}) \tag{101}$$
$$= \sum_{i=1}^{n+1} s_D(\ell_{(i)} - \ell_{(i-1)}) \tag{102}$$
$$= s_D. \tag{103}$$

Therefore $M = b(\mathbf{x}, \mathbf{U})$ is a constant $s_D$, and the $1 - \alpha$ quantile of $M$ is $s_D$.

**Case 2: $\phi < s_D$.**

Let $c_1, c_2 \in \mathcal{R}$ be such that $\phi \leq c_1 < c_2 \leq s_D$. We will now show that

$$F_M(c_2) - F_M(c_1) > 0. \tag{104}$$

Let $v \stackrel{\text{def}}{=} \frac{s_D - c_2}{s_D - \phi}$ and $w \stackrel{\text{def}}{=} \frac{s_D - c_1}{s_D - \phi}$. If $\phi \leq c_1 < c_2 \leq s_D$ then $v < w$ and $v, w \in [0, 1]$.

Let $\mathbf{v} \stackrel{\text{def}}{=} (v_1, \cdots, v_n)$ and $\mathbf{w} \stackrel{\text{def}}{=} (w_1, \cdots, w_n)$ where $\forall i, v_i = v$ and $w_i = w$. Then

$$b(\mathbf{x}, \mathbf{v}) = \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} \sum_{i=1}^{n+1} z_{(i)}(v_{(i)} - v_{(i-1)}) \tag{105}$$
$$= \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} z_{(n+1)}(v_{(n+1)} - v_{(n)}) + z_{(1)}(v_{(1)} - v_{(0)}) \tag{106}$$
$$= \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} s_D(1 - v) + z_{(1)}(v - 0) \tag{107}$$
$$= \sup_{\mathbf{z} \in \mathbb{S}(\mathbf{x})} s_D - (s_D - z_{(1)})\frac{s_D - c_2}{s_D - \phi} \tag{108}$$
$$= s_D - (s_D - \phi)\frac{s_D - c_2}{s_D - \phi} \text{ because } \frac{s_D - c_2}{s_D - \phi} \geq 0 \tag{109}$$
$$= c_2. \tag{110}$$

Similarly,

$$
\begin{aligned}
b(\mathbf{x},\mathbf{w}) &= \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} \sum_{i=1}^{n+1} z_{(i)}(w_{(i)} - w_{(i-1)}) && (111)\\
&= \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} z_{(n+1)}(w_{(n+1)} - w_{(n)}) + z_{(1)}(w_{(1)} - w_{(0)}) && (112)\\
&= \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} s_D(1-w) + z_{(1)}(w-0) && (113)\\
&= \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} s_D - (s_D - z_{(1)})\frac{s_D - c_1}{s_D - \phi} && (114)\\
&= s_D - (s_D - \phi)\frac{s_D - c_1}{s_D - \phi} \text{ because } \frac{s_D - c_1}{s_D - \phi} \geq 0 && (115)\\
&= c_1. && (116)
\end{aligned}
$$

Since $b(\mathbf{x},\mathbf{u})$ is constructed from a linear function of $\mathbf{u}$ with non-positive coefficients, for any $\mathbf{u}$ such that $v \leq u_{(1)} \leq \cdots \leq u_{(n)} < w$ we have:

$$
b(\mathbf{x},\mathbf{w}) < b(\mathbf{x},\mathbf{u}) \leq b(\mathbf{x},\mathbf{v}), \tag{117}
$$

which is equivalent to:

$$
c_1 < b(\mathbf{x},\mathbf{u}) \leq c_2. \tag{118}
$$

So we have $v \leq u_{(1)} \leq \cdots \leq u_{(n)} < w$ implies $c_1 < b(\mathbf{x},\mathbf{u}) \leq c_2$. Therefore for any $c_1, c_2$ such that $\phi \leq c_1 < c_2 \leq s_D$:

$$
\begin{aligned}
& F_M(c_2) - F_M(c_1) && (119)\\
&= \mathbb{P}(c_1 < M \leq c_2) && (120)\\
&= \mathbb{P}_{\mathbf{U}}(c_1 < b(\mathbf{x},\mathbf{U}) \leq c_2) && (121)\\
&\geq \mathbb{P}_{\mathbf{U}}(v \leq U_{(1)} \leq \cdots \leq U_{(n)} < w) && (122)\\
&= \mathbb{P}_{\mathbf{U}}(\,\forall i, 1 \leq i \leq n : v \leq U_i < w) && (123)\\
&= (w - v)^n && (124)\\
&= \left(\frac{c_2 - c_1}{s_D - \phi}\right)^n && (125)\\
&> 0 \text{ because } c_1 < c_2. && (126)
\end{aligned}
$$

Since the support of $M$ is in $[\phi, s_D]$ we have that $F_M$ is strictly increasing on the support. ☐

In summary, the Monte Carlo estimate of our bound will converge to the correct value as the number of samples grows.

Now we prove Theorem D.1.

*Proof of Theorem D.1.* To simplify the notation, we use $Q(\alpha)$ to denote $Q(\alpha, M)$. From Lemma D.3, since $F_M$ is strictly increasing on the support, for $\gamma$ such that $0 < \gamma \leq \alpha$, $Q(1-\alpha) < Q(1-\alpha+\gamma)$ and:

$$
\gamma = F(Q(1-\alpha), Q(1-\alpha+\gamma)) \geq \left(\frac{Q(1-\alpha+\gamma) - Q(1-\alpha)}{s_D - \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} z_{(1)}}\right)^n. \tag{127}
$$

Therefore, letting $\gamma = \min\left(\alpha, \left(\frac{\epsilon}{3(s_D - \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} z_{(1)})}\right)^n\right)$ we have that

$$
\begin{aligned}
Q(1-\alpha+\gamma) &\leq Q(1-\alpha) + \gamma^{1/n}(s_D - \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} z_{(1)}) && (128)\\
&\leq Q(1-\alpha) + \epsilon/3. && (129)
\end{aligned}
$$

Let $p \stackrel{\text{def}}{=} 1-\alpha+\gamma$. From Lemma D.2 and Lemma D.3,

$$\mathbb{P}(|\hat{q}_\ell - Q(p)| > \epsilon/3) \leq 2e^{-2l\delta}, \tag{130}$$

where

$$\begin{aligned}
\delta &= \min\left(p - F(Q(p)-\epsilon/3), F(Q(p)+\epsilon/3) - p\right) && (131)\\
&= \min\left(F(Q(p)-\epsilon/3, Q(p)), F(Q(p), Q(p)+\epsilon/3)\right) && (132)\\
&\geq \left(\frac{\epsilon}{3(s_D - \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} z_{(1)})}\right)^n. && (133)
\end{aligned}$$

Therefore letting $\ell = \left\lceil \frac{-\ln(\gamma/2)}{2}\left(\frac{3(s_D - \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} z_{(1)})}{\epsilon}\right)^n \right\rceil$,

$$\begin{aligned}
\mathbb{P}(|\hat{q}_\ell - Q(1-\alpha+\gamma)| > \epsilon/3) &\leq 2e^{-2l\left(\frac{\epsilon}{3(s_D - \sup_{\mathbf{z}\in\mathbb{S}(\mathbf{x})} z_{(1)})}\right)^n} && (134)\\
&\leq \gamma. && (135)
\end{aligned}$$

Since $\mathbb{P}(Q(1-\alpha+\gamma) < \mu) \leq \alpha - \gamma$, using the union bound we have

$$\begin{aligned}
\mathbb{P}(|\hat{q}_\ell - Q(1-\alpha+\gamma)| > \epsilon/3 \text{ OR } Q(1-\alpha+\gamma) < \mu) &\leq \gamma + \alpha - \gamma && (136)\\
&= \alpha. && (137)
\end{aligned}$$

And therefore,

$$\begin{aligned}
1-\alpha &\leq \mathbb{P}(|\hat{q}_\ell - Q(1-\alpha+\gamma)| \leq \epsilon/3 \text{ AND } Q(1-\alpha+\gamma) \geq \mu) && (138)\\
&\leq \mathbb{P}(Q(1-\alpha+\gamma) \leq \hat{q}_\ell + \epsilon/3 \leq Q(1-\alpha+\gamma) + 2\epsilon/3 \text{ AND } Q(1-\alpha+\gamma) \geq \mu) && (139)\\
&\leq \mathbb{P}(\mu \leq \hat{q}_\ell + \epsilon/3 \leq Q(1-\alpha+\gamma) + 2\epsilon/3) && (140)\\
&\leq \mathbb{P}(\mu \leq \hat{q}_\ell + \epsilon/3 \leq Q(1-\alpha) + \epsilon) \text{ from Eq. 129}. && (141)
\end{aligned}$$

□

## E. Discussion on Section 4

We discuss the case when the distribution is Bernoulli in Section E.1, and present the proofs of Section 4.2 in Section E.2.

### E.1. Special Case: Bernoulli Distribution

When we know that $D = \{0, 1\}$, the distribution is Bernoulli. If we choose $T$ to be the sample mean, we will show that our bound becomes the same as the Clopper-Pearson confidence bound for binomial distributions (Clopper & Pearson, 1934).

If $\mathbf{x}, \mathbf{z} \in \{0,1\}^n$ and $T(\mathbf{z}) \leq T(\mathbf{x})$ then $m(\mathbf{z}, \mathbf{u}) \leq m(\mathbf{x}, \mathbf{u})$. Therefore for any $\mathbf{u} \in [0,1]^n$,

$$b_{D,T}(\mathbf{x}, \mathbf{u}) = \sup_{\mathbf{z}\in\{0,1\}^n : T(\mathbf{z}) \leq T(\mathbf{x})} m_D(\mathbf{z}, \mathbf{u}) = m_D(\mathbf{x}, \mathbf{u}). \tag{142}$$

Let $p_{\mathbf{x}}$ be the number of 0's in $\mathbf{x}$. Therefore the bound becomes the $1-\alpha$ quantile of $m_D(\mathbf{x}, \mathbf{U})$ where

$$m_D(\mathbf{x}, \mathbf{U}) = 1 - \sum_{i=1}^{n} U_{(i)}(x_{(i+1)} - x_{(i)}) = 1 - U_{(p_{\mathbf{x}})}. \tag{143}$$

Therefore the bound is the $1-\alpha$ quantile of $1 - U_{(p_{\mathbf{x}})}$. Then

$$\mathbb{P}(U_{(p_{\mathbf{x}})} \leq 1 - b^\alpha(\mathbf{x})) = \mathbb{P}(1 - U_{(p_{\mathbf{x}})} \geq b^\alpha(\mathbf{x})) = \alpha. \tag{144}$$

Let $\beta(i,j)$ denote a beta distribution with parameters $i$ and $j$. We use the fact that the order statistics of a uniform distribution are beta-distributed. Since $U_{(p_{\mathbf{x}})} \sim \beta(p_{\mathbf{x}}, n+1-p_{\mathbf{x}})$, we have $1 - U_{(p_{\mathbf{x}})} \sim \beta(n - p_{\mathbf{x}} + 1, p_{\mathbf{x}})$

$$b^\alpha(\mathbf{x}) = Q(1-\alpha, \beta(n - p_{\mathbf{x}} + 1, p_{\mathbf{x}})). \tag{145}$$

This is the same as the Clopper-Pearson upper confidence bound for binomial distributions.

### E.2. Proof of Section 4.2

**Lemma E.1** (Lemma 4.1). *Let $\mathbf{X} = (X_1, \cdots, X_n)$ be a sample of size $n$ from a distribution with mean $\mu$. Let $\boldsymbol{\ell} \in [0,1]^n$. If $G_{\mathbf{X},\boldsymbol{\ell}}$ is a $(1-\alpha)$ lower confidence bound for the CDF then*

$$\mathbb{P}_{\mathbf{X}}(m(\mathbf{X}, \boldsymbol{\ell}) \geq \mu) \geq 1 - \alpha. \tag{146}$$

*Proof.* [9] If $\forall y \in \mathcal{R}, F(y) \geq G_{\mathbf{X},\boldsymbol{\ell}}(y)$ then

$$\forall i : 1 \leq i \leq n, F(X_{(i)}) \geq \ell_{(i)}. \tag{147}$$

Recall that $m_D(\mathbf{X}, \boldsymbol{\ell}) = s_D - \sum_{i=1}^{n} \ell_{(i)}(z_{(i+1)} - z_{(i)})$. Therefore if $\forall i : 1 \leq i \leq n,\ F(X_{(i)}) \geq \ell_{(i)}$ then $m(\mathbf{X}, \boldsymbol{\ell}) \geq m(\mathbf{X}, F(\mathbf{X}))$.

From Lemma 2.3, $m(\mathbf{X}, F(\mathbf{X})) \geq \mu$. Therefore $m(\mathbf{X}, \boldsymbol{\ell}) \geq \mu$. And hence, finally,

$$\begin{aligned} \mathbb{P}(m(\mathbf{X}, \boldsymbol{\ell}) \geq \mu) &\geq \mathbb{P}_{\mathbf{X}}\left(\forall y \in \mathcal{R}, F(y) \geq G_{\mathbf{X},\boldsymbol{\ell}}(y)\right) && (148)\\ &= 1 - \alpha. && (149) \end{aligned}$$

□

We now show that if $G_{\mathbf{X},\boldsymbol{\ell}}$ (Figure 1) is a lower confidence bound, then the order statistics of $\boldsymbol{\ell}$ are element-wise smaller than the order statistics of a sample of size $n$ from the uniform distribution with high probability:

**Lemma E.2.** *Let $\mathbf{U} = U_1, ..., U_n$ be a sample of size $n$ from the continuous uniform distribution on $[0, 1]$. Let $\boldsymbol{\ell} \in [0,1]^n$ and $\alpha \in (0, 1)$. If $\mathcal{D}^+$ is continous and $G_{\mathbf{X},\boldsymbol{\ell}}$ is a $(1-\alpha)$ lower confidence bound for the CDF then:*

$$\mathbb{P}_{\mathbf{U}}(\ \forall i : 1 \leq i \leq n, U_{(i)} \geq \ell_{(i)}) \geq 1 - \alpha. \tag{150}$$

*Proof.* Let $K$ be the CDF of a distribution such that $K$ is continuous and strictly increasing on $\mathcal{D}^+$ (since $\mathcal{D}^+$ is continuous, $K$ exists). Let $\mathbf{X} = (X_1, \cdots, X_n)$ be a sample of size $n$ from the distribution with CDF $K$. By Lemma 2.2, $K(X)$ is uniformly distributed on $[0, 1]$.

By the definition of $G_{\mathbf{X},\boldsymbol{\ell}}$, if $\ \forall x \in C, K(y) \geq G_{\mathbf{X},\boldsymbol{\ell}}(y)$ then:

$$\begin{aligned} &K(y) \geq 0, && \text{if } y < X_{(1)} && (151)\\ &K(y) \geq \ell_{(i)}, && \text{if } X_{(i)} \leq y < X_{(i+1)} && (152)\\ &K(y) \geq 1, && \text{if } y \geq s_C. && (153) \end{aligned}$$

which is equivalent to:

$$\forall i : 1 \leq i \leq n, K(y) \geq \ell_{(i)}, \text{if } X_{(i)} \leq y < X_{(i+1)}. \tag{154}$$

Since $K(y)$ is non-decreasing, this is equivalent to:

$$\forall i : 1 \leq i \leq n, K(X_{(i)}) \geq \ell_{(i)} \tag{155}$$

Since $G_{\mathbf{X},\boldsymbol{\ell}}$ is a lower confidence bound,

$$\begin{aligned} 1 - \alpha &\leq \mathbb{P}_{\mathbf{X}}(\ \forall y \in \mathcal{R}, K(y) \geq G_{\mathbf{X},\boldsymbol{\ell}}(y)) && (156)\\ &= \mathbb{P}_{\mathbf{X}}(\forall i : 1 \leq i \leq n, K(X_{(i)}) \geq \ell_{(i)}) && (157)\\ &= \mathbb{P}_{\mathbf{U}}(\forall i : 1 \leq i \leq n, U_{(i)} \geq \ell_{(i)}) \text{ by Lemma 2.2.} && (158) \end{aligned}$$

□

To prove Theorem 4.3, we prove the more general version where $G_{\mathbf{X},\boldsymbol{\ell}}$ is a (possibly not exact) lower confidence bound for the CDF.

[9] The proof is implied in (Anderson, 1969b) but we provide it here for completeness

**Theorem E.3.** *Let $\boldsymbol{\ell} \in [0,1]^n$. Let $D^+ = [-\infty, b]$. If $G_{\mathbf{X},\boldsymbol{\ell}}$ is a $1-\alpha$ lower confidence bound for the CDF, then for any sample size $n$, for all sample values $\mathbf{x} \in D^n$ and all $\alpha \in (0,1)$, using $T(\mathbf{x}) = m_{D^+}(\mathbf{x}, \boldsymbol{\ell})$ to compute $b^{\alpha}_{D^+,T}(\mathbf{x})$ yields:*

$$b^{\alpha}_{D^+,T}(\mathbf{x}) \leq m_{D^+}(\mathbf{x}, \boldsymbol{\ell}). \tag{159}$$

*Proof.* Since $G_{\mathbf{X},\boldsymbol{\ell}}$ is a lower confidence bound for the CDF $F$, from Lemma E.2,

$$\mathbb{P}(\forall i, U_{(i)} \geq \ell_{(i)}) \geq 1-\alpha. \tag{160}$$

First we note that

$$b_{D^+,T}(\mathbf{x}, \boldsymbol{\ell}) = \sup_{\mathbf{y}:\mathbf{y} \in \mathbb{S}_{D^+,T}(\mathbf{x})} m_{D^+}(\mathbf{y}, \boldsymbol{\ell}) \tag{161}$$

$$= \sup_{m_{D^+}(\mathbf{y},\boldsymbol{\ell}) \leq m_{D^+}(\mathbf{x},\boldsymbol{\ell})} m_{D^+}(\mathbf{y}, \boldsymbol{\ell}) \tag{162}$$

$$= m_{D^+}(\mathbf{x}, \boldsymbol{\ell}). \tag{163}$$

Recall that $b^{\alpha}_{D^+,T}(\mathbf{x})$ is the $1-\alpha$ quantile of $b_{D^+,T}(\mathbf{x}, \mathbf{U})$. In order to show that $b^{\alpha}_{D^+,T}(\mathbf{x}) \leq b_{D^+,T}(\mathbf{x}, \boldsymbol{\ell})$, we will show that

$$\mathbb{P}(b_{D^+,T}(\mathbf{x}, \mathbf{U}) \leq b_{D^+,T}(\mathbf{x}, \boldsymbol{\ell})) \geq 1-\alpha. \tag{164}$$

Recall that $b_{D^+,T}(\mathbf{x}, \mathbf{U}) = \sup_{\mathbf{y} \in \mathbb{S}_T(\mathbf{x})} t_{D^+} - \sum_{i=1}^{n} U_{(i)}(x_{(i+1)} - x_{(i)})$. Then if $\forall i, U_{(i)} \geq \ell_{(i)}$ then $b_{D^+,T}(\mathbf{x}, \mathbf{U}) \leq b_{D^+,T}(\mathbf{x}, \boldsymbol{\ell})$. Therefore,

$$\mathbb{P}(b_{D^+,T}(\mathbf{x}, \mathbf{U}) \leq b_{D^+,T}(\mathbf{x}, \boldsymbol{\ell})) \tag{165}$$

$$\geq \mathbb{P}(\forall i, U_{(i)} \geq \ell_{(i)}) \tag{166}$$

$$\geq 1-\alpha, \text{ by Lemma E.2.} \tag{167}$$

□

We can now show the comparison with Anderson's bound and Hoeffding's bound.

**Theorem E.4** (Theorem 4.3)**.** *Let $\boldsymbol{\ell} \in [0,1]^n$ be a vector such that $G_{\mathbf{X},\boldsymbol{\ell}}$ is an exact $(1-\alpha)$ lower confidence bound for the CDF.*

*Let $D^+ = (-\infty, b]$. For any sample size $n$, for any sample value $\mathbf{x} \in D^n$, for all $\alpha \in (0,1)$, using $T(\mathbf{x}) = b^{\alpha,\textit{Anderson}}_{\boldsymbol{\ell}}(\mathbf{x})$ yields*

$$b^{\alpha}_{D^+,T}(\mathbf{x}) \leq b^{\alpha,\textit{Anderson}}_{\boldsymbol{\ell}}(\mathbf{x}). \tag{168}$$

*Proof.* We have $b^{\alpha,\text{Anderson}}_{\boldsymbol{\ell}}(\mathbf{x}) = m_{D^+}(\mathbf{x}, \boldsymbol{\ell})$ where $\boldsymbol{\ell}$ satisfies $G_{\mathbf{X},\boldsymbol{\ell}}$ is a $1-\alpha$ lower confidence bound for the CDF. Therefore applying Theorem E.3 yields the result. □

**Theorem E.5** (Theorem 4.4)**.** *Let $D^+ = (-\infty, b]$. For any sample size $n$, for any sample value $\mathbf{x} \in D^n$, for all $\alpha \in (0, 0.5]$, using $T(\mathbf{x}) = b^{\alpha,\textit{Anderson}}_{\boldsymbol{\ell}}(\mathbf{x})$ where $\boldsymbol{\ell} = \mathbf{u}^{\textit{And}}$ yields*

$$b^{\alpha}_{D^+,T}(\mathbf{x}) \leq b^{\alpha,\textit{Hoeffding}}(\mathbf{x}), \tag{169}$$

*where the inequality is strict when $n \geq 3$.*

*Proof.* The proof follows directly from Lemma 4.2 and Theorem 4.3.

Recall that $G_{\mathbf{X},\mathbf{u}^{\text{And}}}$ is an exact $(1-\alpha)$ lower confidence bound for the CDF and therefore:

$$\mathbb{P}_{\mathbf{U}}(\forall i : 1 \leq i \leq n, U_{(i)} \geq u^{\text{And}}_{(i)}) = 1-\alpha. \tag{170}$$

From Theorem 4.3, using $T(\mathbf{x}) = b^{\alpha,\text{Anderson}}_{\mathbf{u}^{\text{And}}}(\mathbf{x})$ yields

$$b^{\alpha}_{D^+,T}(\mathbf{x}) \leq b^{\alpha,\text{Anderson}}_{\mathbf{u}^{\text{And}}}(\mathbf{x}). \tag{171}$$

Let $\boldsymbol{\ell} \in [0,1]^n$ be defined such that

$$\ell_i \stackrel{\text{def}}{=} \max\left\{0, i/n - \sqrt{\ln(1/\alpha)/(2n)}\right\}. \tag{172}$$

Since $G_{\mathbf{X},\boldsymbol{\ell}}(x)$ is an $1-\alpha$ lower confidence bound, from Lemma E.2:

$$\mathbb{P}_{\mathbf{U}}(\forall i: 1 \leq i \leq n, U_{(i)} \geq \ell_{(i)}) \geq 1-\alpha \tag{173}$$

$$= \mathbb{P}_{\mathbf{U}}(\forall i: 1 \leq i \leq n, U_{(i)} \geq u^{\text{And}}_{(i)}). \tag{174}$$

Since $u_i^{\text{And}} = \max\{0, i/n - \beta(n)\}$ and $\ell_i = \max\left\{0, i/n - \sqrt{\ln(1/\alpha)/(2n)}\right\}$, we have $\beta(n) \leq \sqrt{\ln(1/\alpha)/(2n)}$, and therefore $u_i^{\text{And}} \geq \ell_i$ for all $i$. Therefore $m(\mathbf{x}, \boldsymbol{\ell}) \geq m(\mathbf{x}, \mathbf{u}^{\text{And}})$, i.e.

$$b^{\alpha,\text{Anderson}}_{\mathbf{u}^{\text{And}}}(\mathbf{x}) \leq b^{\alpha,\text{Anderson}}_{\boldsymbol{\ell}}(\mathbf{x}). \tag{175}$$

From Eq. 171, Eq. 175 and Lemma 4.2 we have the result. $\square$