# Assessing the coverage probabilities of fixed-margin confidence intervals for the tail conditional allocation

N.V. Gribkova *,[1,2], J. Su †,[3,4], and R. Zitikis ‡,[4,5]

[1]Saint Petersburg State University, Saint Petersburg, 199034 Russia
[2]Saint Petersburg State Transport University, Saint Petersburg, 190031 Russia
[3]Purdue University, West Lafayette, Indiana 47907, United States
[4]Risk and Insurance Studies Centre, York University, Toronto, Ontario M3J 1P3, Canada
[5]Western University, London, Ontario N6A 5B7, Canada

**Abstract.** The tail conditional allocation plays an important role in a number of areas, including economics, finance, insurance, and management. Fixed-margin confidence intervals are of particular interest, and the assessment of their coverage probabilities is of much interest in practice. In this paper we offer a convenient way to achieve this goal via resampling. The theoretical part of the paper, which is technically demanding, is rigorously established under minimal conditions to facilitate the widest possible practical use. A numerical study illustrates the performance of the estimator on data.

*Key words and phrases*: tail conditional allocation, order statistics, concomitants, resampling, coverage probability.

---

This research has been supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada, and the national research organization Mathematics of Information Technology and Complex Systems (MITACS) of Canada.

*e-mail: n.gribkova@spbu.ru

†Corresponding author; e-mail: jianxi@purdue.edu

‡e-mail: rzitikis@uwo.ca

# 1 Introduction

The tail conditional allocation (TCA) of a pair of real-valued random variables $X$ and $Y$ is the conditional expectation

$$\mathrm{TCA}(p) = \mathbb{E}\big(X \mid Y > G^{-1}(p)\big),$$

where $p \in [0, 1)$ is a fixed parameter and $G^{-1}$ is the generalized (left-continuous) inverse of the cumulative distribution function (cdf) $G$ of $Y$, defined by

$$G^{-1}(p) = \inf\{y \in \mathbb{R} : G(y) \geq p\}.$$

In general, both $X \sim F$ and $Y \sim G$ are dependent. Denote their joint cdf by $H$. The special case $X = Y$ leads to what is known in the literature as the expected shortfall, also known by several other names such as the tail conditional expectation and the conditional value at risk. The TCA has played important roles in areas such as economics, finance, insurance, and management.

Since the aforementioned cdf's are not known in practice, they are estimated using the realized values of random pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$. Assume that the pairs are independent and identically distributed according to the joint cdf $H$ of $(X, Y)$. The empirical TCA is

$$\mathrm{TCA}_n(p) = \frac{1}{(1-p)n} \sum_{i=1}^{n} X_i \mathbb{1}_{(G_n^{-1}(p), \infty)}(Y_i), \tag{1.1}$$

where $G_n^{-1}(p)$ is the $p$-th quantile of the empirical cdf $G_n$ based on $Y_1, \ldots, Y_n$, that is,

$$G_n^{-1}(p) = \inf\{y \in \mathbb{R} : G_n(y) \geq p\}.$$

Using a specially designed example, Gribkova et al. (2021) showed that establishing consistency of the estimator $\mathrm{TCA}_n(p)$ is not possible without a continuity-type assumption on the cdf $G$. Subsequently, Gribkova et al. (2022) proved that it is enough to have the continuity of $G$ only in an open neighbourhood of the interval $[G^{-1}(p), G^{\leftarrow}(p)]$, where the right-continuous quantile function $G^{\leftarrow}$ is defined by

$$G^{\leftarrow}(p) = \inf\{y \in \mathbb{R} : G(y) > p\}.$$

Since $G^{-1}(p) \leq G^{\leftarrow}(p)$ for all $p \in (0, 1)$, the functions $G^{-1}$ and $G^{\leftarrow}$ are often called the lower and upper quantile functions, respectively.

As discussed by Gribkova et al. (2021, 2022), the fulfillment of practical needs starts with the construction of confidence intervals for the population $\mathrm{TCA}(p)$, followed by the assessment of the coverage probability. That is, given $\varepsilon > 0$, which could be very large in absolute terms but only a fraction of the (estimated value of the) population $\mathrm{TCA}(p)$, we wish to assess the magnitude of

$$\Pi_n(\varepsilon) := \mathbb{P}\Big(\big|\mathrm{TCA}_n(p) - \mathrm{TCA}(p)\big| \le \varepsilon\Big).$$

This probability relies on the population distribution and is therefore unknown. Often, the magnitude of such probabilities is assessed by deriving CLT-type results, which in our case would mean establishing the CLT for the estimator $\mathrm{TCA}_n(p)$. Gribkova et al. (2021) postulated such a result in the concluding section of their paper, deferring its rather complex proof to the follow-up paper by Gribkova et al. (2022). It is shown in the latter paper that, under certain conditions, there is a finite constant $\sigma > 0$ such that

$$\Pi_n(\varepsilon) \approx \Phi\big(\varepsilon\sqrt{n}/\sigma\big) - \Phi\big(-\varepsilon\sqrt{n}/\sigma\big) \tag{1.2}$$

for large $n$, where $\Phi$ denotes the standard normal cdf. Under further conditions, Gribkova et al. (2022) have derived a consistent estimator of $\sigma$ and thus enabled researchers to empirically assess the magnitude of $\Pi_n(\varepsilon)$ from available data.

While the above stratagem for the assessment of $\Pi_n(\varepsilon)$ is appealing to theoretical minds, it nevertheless requires more time and energy than practitioners could often afford for its understanding and implementation, because the stratagem relies on additional conditions and also on tuning of various parameters, and all these tasks are in addition to what is required to show that $\Pi_n(\varepsilon)$ tends to 1 when the sample size $n$ increases. This is of course natural because the right-hand side of statement (1.2) reveals the fastest rate of convergence to 1 that the probability $\Pi_n(\varepsilon)$ can achieve. Hence, a natural question arises: can we assess the magnitude of $\Pi_n(\varepsilon)$ in a more adaptable way? With the help of a resampling technique, we shall next show that the answer is in affirmative.

# 2 The main result and its numerical illustration

To avoid the unnecessary at this stage complication of technicalities, we restrict ourselves to the Efron's $m$-out-of-$n$ non-parametric bootstrap. For this, we generate a bootstrap sample $(X_1^*, Y_1^*), \ldots, (X_m^*, Y_m^*)$ of size $m := m_n$ from the empirical cdf $H_n$ that arises from the original sample $(X_1, Y_1), \ldots, (X_n, Y_n)$. That is, we have the following probability distribution

$$\mathbb{P}^*\Big((X_j^*, Y_j^*) = (X_i, Y_i)\Big) = \frac{1}{n}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m,$$

where $\mathbb{P}^*$ denotes the conditional on the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ probability $\mathbb{P}$. The bootstrapped version of the TCA is given by the formula

$$\mathrm{TCA}^*_{n,m}(p) = \frac{1}{(1-p)m} \sum_{i=1}^{m} X_i^* \mathbb{1}_{((G_m^*)^{-1}(p), \infty)}(Y_i^*).$$

A serious question now arises: when does $\mathrm{TCA}^*_{n,m}(p)$ estimate $\mathrm{TCA}(p)$? To see the non-triviality of this problem, we first recall the work of Gribkova et al. (2021) where it is shown that some kind of a continuity assumption is necessary but, unavoidably, the bootstrapped version of the TCA lives in the "bootstrap world," which is inherently discrete, as it is based on the discrete cdf $H_n$. The following theorem settles the matter.

**Theorem 2.1.** *Suppose that the following three conditions hold:*

($\mathrm{C}_1$) $\mathbb{E}(X \mid Y > G^{-1}(p - \delta))$ *is finite for a (small)* $\delta > 0$.

($\mathrm{C}_2$) $G$ *is continuous on an (open) neighbourhood of the interval* $[G^{-1}(p), G^{\leftarrow}(p)]$.

($\mathrm{C}_3$) $\limsup(m/n) < \infty$ *as* $n \wedge m := \min\{n, m\} \to \infty$.

*Then, for every* $\varepsilon > 0$ *and when* $n \wedge m \to \infty$,

$$\mathbb{P}^*\Big(\big|\mathrm{TCA}^*_{n,m}(p) - \mathrm{TCA}(p)\big| \le \varepsilon\Big) \xrightarrow{\mathbb{P}} 1. \tag{2.1}$$

Clearly, instead of relying on statement (1.2) to assess $\Pi_n(\varepsilon)$, we can now rely on statement (2.1), which immediately leads to the in-probability approximation

$$\Pi_n(\varepsilon) \approx \Pi^*_{n,m}(\varepsilon) := \mathbb{P}^*\Big(\big|\mathrm{TCA}^*_{n,m}(p) - \mathrm{TCA}_n(p)\big| \le \varepsilon\Big) \tag{2.2}$$

for all sufficiently large $n$ and $m$. This gives an easily implementable in practice way to assess $\Pi_n(\varepsilon)$. We shall next explore the performance of this approximation in a simulation study. The proof of Theorem 2.1, which is rather complex, is in Section 3, followed by Section 4 containing a host of auxiliary lemmas and their proofs.

Hence, to illustrate the bootstrap methodology, and to do so efficiently and also in the context of the earlier developed methodologies of Gribkova et al. (2021, 2022), we employ the set-up of the latter two papers. Namely, we consider a multi-peril insurance product which consists of two insurance coverage. We assume that the associated losses random variables, denoted by $L_1$ and $L_2$, follow the bivariate Pareto distribution

$$\mathbb{P}(L_1 > l_1, L_2 > l_2) = \big(1 + l_1/\theta_1 + l_2/\theta_2\big)^{-\alpha}, \qquad l_1 > 0,\ l_2 > 0, \tag{2.3}$$

where $\theta_1 > 0$ and $\theta_2 > 0$ are scale parameters, and $\alpha > 0$ is a shape parameter: the smaller the value of $\alpha$, the heavier the distribution tails of $L_1$ and $L_2$. Due to the presence of policy deductibles $d_1 > 0$ and $d_2 > 0$, the payment random variables are

$$W_i = (L_i - d_i) \times \mathbb{1}_{(d_i,\infty)}(L_i), \quad i = 1, 2.$$

To analyze the risk contribution of, say, the first insurance coverage out of the total loss of the insurance product, we calculate the TCA with $X = W_1$ and $Y = W_1 + W_2$. We already know from Gribkova et al. (2022) that $\mathrm{TCA}_n(p)$ yields an asymptotically precise estimate of $\mathrm{TCA}(p)$. Concerning the performance of $\mathrm{TCA}_n(p)$, Gribkova et al. (2022) used the asymptotic normality of the estimator to approximate the coverage probability $\Pi_n(\varepsilon)$. We shall next contrast the currently developed bootstrap approach with the normality approach of Gribkova et al. (2022) in the context of evaluating $\Pi_n(\varepsilon)$.

If the population distribution (2.3) is known, then we can apply the matrix analytic method of Furman et al. (2021) to compute $\mathrm{TCA}(p)$ explicitly. Moreover, the coverage probability $\Pi_n(\varepsilon)$ can be assessed via Monte Carlo as follows:

$$\widehat{\Pi}_n^{\mathrm{MC}}(\varepsilon) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}_{[-\varepsilon,\varepsilon]}\Big(\mathrm{TCA}_n^t(p) - \mathrm{TCA}(p)\Big),$$

where $\mathrm{TCA}_n^t(p)$ is the empirical estimate of $\mathrm{TCA}(p)$ obtained from the $t^{\mathrm{th}}$ set of simulated data, $t = 1, \dots, T$. We choose large $T = 100\,000$ to ensure that $\Pi_n(\varepsilon)$ is calculated accurately.

In the more practical scenario when the population distribution is unknown and only a single data set is available, statement (2.2) advocates sampling from the empirical distribution obtained from the observed data in order to compute $\Pi_n(\varepsilon)$. Specifically, we obtain $T$ sets of re-samples with replacement from the original data, and then within each newly obtained dataset, we calculate $\mathrm{TCA}_{n,m}^{*,t}$, where $m = n$ and $t = 1, \dots, T$. The coverage probability $\Pi_n(\varepsilon)$

is estimated by

$$\widehat{\Pi^*_{n,n}}(\varepsilon) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}_{[-\varepsilon,\varepsilon]}\Big(\mathrm{TCA}^{*;t}_{n,n}(p) - \widehat{\mathrm{TCA}^*}(p)\Big),$$

where

$$\widehat{\mathrm{TCA}^*}(p) = \frac{1}{T}\sum_{t=1}^{T} \mathrm{TCA}^{*;t}_{n,n}(p).$$

In what follows, we fix $\theta_1 = 100$ and $\theta_2 = 50$, but vary the shape parameter $\alpha \in \{1.8,\ 3,\ 4\}$ in order to see how the tail behavior of distribution (2.3) is impacting the performance of $\Pi^*_{n,m}(\varepsilon)$ as an estimator of $\Pi_n(\varepsilon)$, with $n = i \times 10\,000$ and $i \in \{0.5,\ 1,\ 1.5,\ 2\}$. The margin of error $\varepsilon$ is set to $5\% \times \mathrm{TCA}(p)$, and the policy deductables are set to $d_1 = 25$ and $d_2 = 12$, which correspond to the medians of $L_1$ and $L_2$, respectively, under the shape parameter value $\alpha = 3$. We are interested in estimating $\mathrm{TCA}(p)$ at the confidence levels $p = 97.5\%$ and $p = 99\%$.

Figure 2.1 compares the coverage probabilities based on the herein proposed bootstrap method and the CLT-based method of Gribkova et al. (2022). Next are several observations:

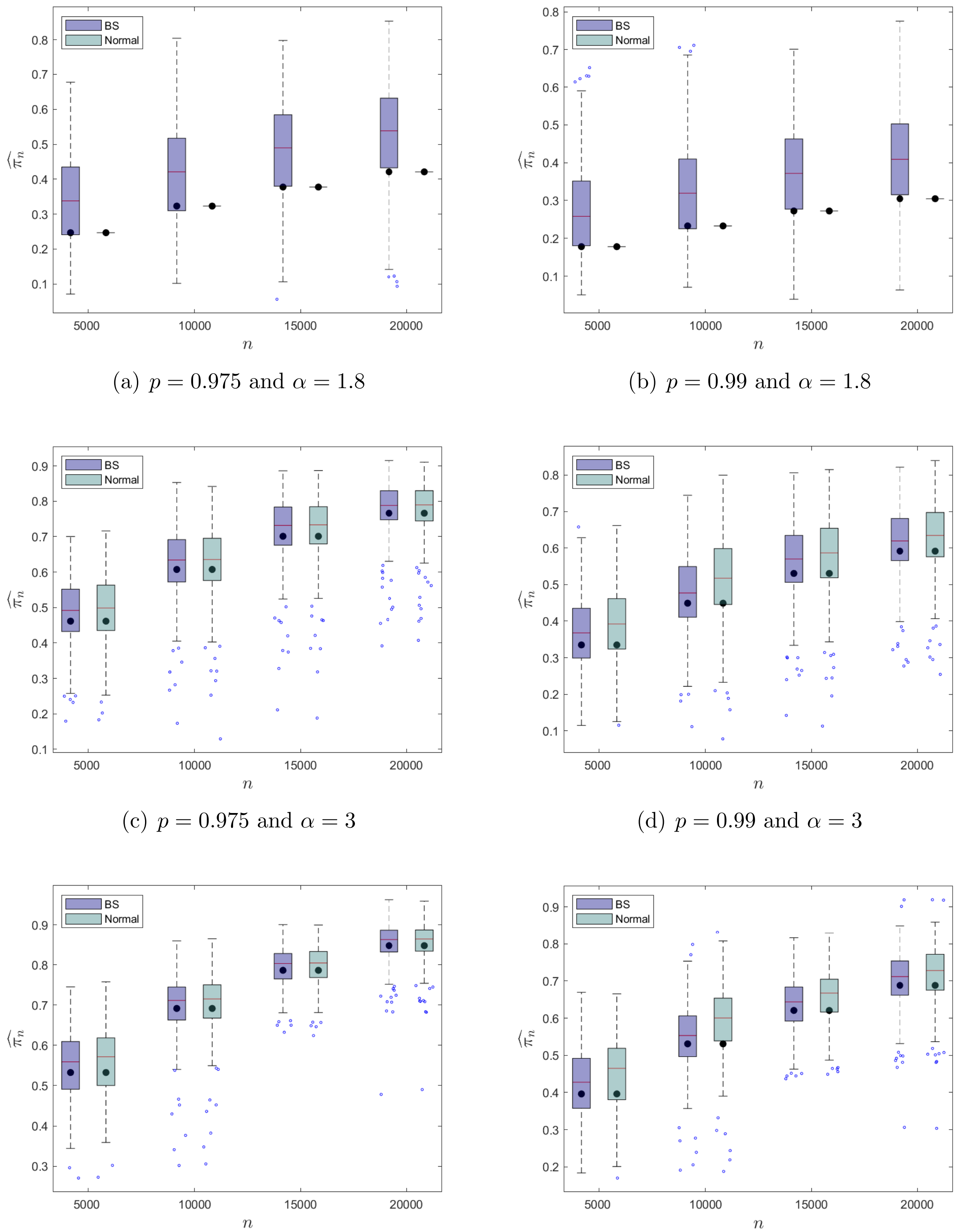
(i) The actual coverage probabilities are increasing with respect to $\alpha$ and $n$, but decreasing with respect to $p$. These patterns are well captured by both estimation methods.

(ii) For fixed $n$ and $p$, smaller values of the shape parameter $\alpha$, that is, more heavily tailed distributions (2.3) diminish the performance of the two coverage-probability estimators. When $\alpha \leq 2$, normal approximation is unavailable because of the lack of a finite second moment (see Gribkova et al. (2022) for technical details).

(iii) For fixed $\alpha$ and $n$, smaller values of $p$ improve the performance of both estimators, which is natural because more data in the tail portion are used.

(iv) Across all the choices of $\alpha$ and $p$, the estimation intervals of $\Pi_n(\varepsilon)$ shrink as the sample size $n$ increases, and the relationship is more noticeable when $\alpha$ is large.

In Tables 2.1 and 2.2 we compare the performance of the bootstrap and normal estimators based on their biases and mean absolute errors (MAE) with respect to the actual value of $\Pi_n(\varepsilon)$. We see that the bootstrap method outperforms the normal approximation in terms of smaller bias and MAE across all considered scenarios. As the sample size $n$ grows, the performance of the two estimators becomes very similar, especially when $p = 97.5\%$.




(a) $p = 0.975$ and $\alpha = 1.8$

(b) $p = 0.99$ and $\alpha = 1.8$

(c) $p = 0.975$ and $\alpha = 3$

(d) $p = 0.99$ and $\alpha = 3$

(e) $p = 0.975$ and $\alpha = 4$

(f) $p = 0.99$ and $\alpha = 4$

Figure 2.1: Box plots comparing the coverage-probability estimators based on the bootstrap (BS) and normal-approximation methods.

| $\alpha$ | $\frac{n}{10\,000}$ | $\widehat{\Pi_n^{\mathrm{MC}}}(\varepsilon)$ | Bootstrap | | Normal | |
|---|---|---|---|---|---|---|
| | | | Bias | MAE | Bias | MAE |
| 1.8 | 0.5 | 0.247 | 0.094 (38%) | 0.13 (52%) | - | - |
| | 1.0 | 0.323 | 0.091 (28%) | 0.139 (43%) | - | - |
| | 1.5 | 0.377 | 0.096 (25%) | 0.148 (39%) | - | - |
| | 2.0 | 0.421 | 0.096 (23%) | 0.150 (36%) | - | - |
| 3 | 0.5 | 0.461 | 0.022 (4.8%) | 0.082 (17%) | 0.029 (6.4%) | 0.086 (19%) |
| | 1.0 | 0.607 | 0.016 (2.6%) | 0.077 (12%) | 0.019 (3.1%) | 0.079 (13%) |
| | 1.5 | 0.701 | 0.017 (2.5%) | 0.073 (10%) | 0.019 (2.7%) | 0.073 (11%) |
| | 2.0 | 0.766 | 0.013 (1.6%) | 0.058 (7.6%) | 0.013 (1.7%) | 0.059 (7.7%) |
| 4 | 0.5 | 0.533 | 0.016 (3.0%) | 0.070 (13%) | 0.025 (4.8%) | 0.073 (13.8%) |
| | 1.0 | 0.692 | 0.010 (1.5%) | 0.057 (8.2%) | 0.013 (1.9%) | 0.058 (8.4%) |
| | 1.5 | 0.786 | 0.009 (1.1%) | 0.041 (5.2%) | 0.011 (1.4%) | 0.041 (5.2%) |
| | 2.0 | 0.847 | 0.006 (0.7%) | 0.038 (4.5%) | 0.007 (0.8%) | 0.038 (4.5%) |

Table 2.1: Performance of the bootstrap and normal estimators of $\Pi_n(\varepsilon)$ in terms of their biases and MAE's when $p = 97.5\%$. (The normal approximation is unavailable when $\alpha = 1.8$.)

| $\alpha$ | $\frac{n}{10\,000}$ | $\widehat{\Pi_n^{\mathrm{MC}}}(\varepsilon)$ | Bootstrap | | Normal | |
|---|---|---|---|---|---|---|
| | | | Bias | MAE | Bias | MAE |
| 1.8 | 0.5 | 0.178 | 0.094 (53%) | 0.118 (66%) | - | - |
| | 1.0 | 0.233 | 0.091 (39%) | 0.124 (53%) | - | - |
| | 1.5 | 0.272 | 0.096 (35%) | 0.133 (49%) | - | - |
| | 2.0 | 0.305 | 0.098 (32%) | 0.137 (45%) | - | - |
| 3 | 0.5 | 0.335 | 0.031 (9.2%) | 0.084 (25%) | 0.056 (17%) | 0.099 (29%) |
| | 1.0 | 0.449 | 0.025 (5.6%) | 0.085 (19%) | 0.064 (14%) | 0.105 (23%) |
| | 1.5 | 0.530 | 0.028 (5.2%) | 0.085 (16%) | 0.045 (8.4%) | 0.095 (18%) |
| | 2.0 | 0.591 | 0.022 (3.7%) | 0.074 (13%) | 0.034 (5.8%) | 0.081 (14%) |
| 4 | 0.5 | 0.396 | 0.026 (6.5%) | 0.080 (20%) | 0.054 (14%) | 0.093 (23%) |
| | 1.0 | 0.531 | 0.019 (3.6%) | 0.073 (14%) | 0.064 (12%) | 0.096 (18%) |
| | 1.5 | 0.621 | 0.018 (2.9%) | 0.059 (9.6%) | 0.039 (6.2%) | 0.068 (11%) |
| | 2.0 | 0.688 | 0.015 (2.2%) | 0.062 (8.9%) | 0.029 (4.1%) | 0.068 (9.7%) |

Table 2.2: Performance of the bootstrap and normal estimators of $\Pi_n(\varepsilon)$ in terms of their biases and MAE's when $p = 99\%$. (The normal approximation is unavailable when $\alpha = 1.8$.)

# 3 Proof of Theorem 2.1

Unless noted otherwise, all convergence statements are when $n \wedge m \to \infty$, where $\wedge$ denotes the minimum, and we shall also use the notation $\vee$ for the maximum.

Statement (2.1) is equivalent to

$$\mathbb{E}\Big(\mathbb{P}^*\Big(\big|\mathrm{TCA}^*_{n,m}(p) - \mathrm{TCA}(p)\big| > \varepsilon\Big)\Big) \to 0. \tag{3.1}$$

By Theorem 1.1 of Gribkova et al. (2022), the empirical estimator $\mathrm{TCA}_n(p)$ consistently estimates $\mathrm{TCA}(p)$, and thus proving statement (3.1) is equivalent to showing that, for every $\varepsilon > 0$,

$$\mathbb{E}\Big(\mathbb{P}^*\Big(\big|\mathrm{TCA}^*_{n,m}(p) - \mathrm{TCA}_n(p)\big| > \varepsilon\Big)\Big) \to 0. \tag{3.2}$$

We write

$$(1-p)\Big(\mathrm{TCA}^*_{n,m}(p) - \mathrm{TCA}_n(p)\Big) = I_{n,m} + J_{n,m},$$

where

$$I_{n,m} = \frac{1}{m}\sum_{j=1}^{m} X_i^* \mathbb{1}_{((G_m^*)^{-1}(p),\infty)}(Y_i^*) - \frac{1}{m}\sum_{j=1}^{m} X_i^* \mathbb{1}_{(G_n^{-1}(p),\infty)}(Y_i^*)$$

and

$$J_{n,m} = \frac{1}{m}\sum_{j=1}^{m} X_i^* \mathbb{1}_{(G_n^{-1}(p),\infty)}(Y_i^*) - \frac{1}{n}\sum_{i=1}^{n} X_i \mathbb{1}_{(G_n^{-1}(p),\infty)}(Y_i).$$

Hence, we need to prove that in the "bootstrap world," both $I_{n,m}$ and $J_{n,m}$ converge to 0 in the sense specified in the following two statements, where, and throughout the rest of the paper, we use the notation

$$\delta_m = \sqrt{\frac{\log m}{m}}.$$

**Statement 3.1.** *For every $\varepsilon > 0$ and when $n \wedge m \to \infty$, we have*

$$\mathbb{E}\Big(\mathbb{P}^*\big(|I_{n,m}| > \varepsilon\big)\Big) \to 0. \tag{3.3}$$

*Proof.* Denote the event

$$\mathcal{A}_{n,m} = \Big\{G^{-1}(p - 2\delta_m) < G_n^{-1}(p - \delta_m)\Big\} \cap \Big\{G_n^{-1}(p + \delta_m) \le G^{-1}(p + 2\delta_m)\Big\}.$$

By Lemma 4.1 (statements (4.2) and (4.3) with $c_1 = 1$ and $c_2 = 2$), we have

$$\mathbb{P}\big(\mathcal{A}_{n,m}^c\big) \to 0. \tag{3.4}$$

Hence,

$$\begin{aligned}\mathbb{E}\Big(\mathbb{P}^*\big(|I_{n,m}| > \varepsilon\big)\Big) &\le \mathbb{E}\Big(\mathbb{P}^*\big(|I_{n,m}| > \varepsilon\big)\mathbb{1}_{\mathcal{A}_{n,m}}\Big) + \mathbb{P}\big(\mathcal{A}_{n,m}^c\big) \\ &= \mathbb{E}\Big(\mathbb{P}^*\big(|I_{n,m}| > \varepsilon\big)\mathbb{1}_{\mathcal{A}_{n,m}}\Big) + o(1). \end{aligned} \tag{3.5}$$

Let $X_{1,m}^*, \dots, X_{m,m}^*$ denote the concomitants (a.k.a. induced order statistics) of $X_1^*, \dots, X_m^*$ corresponding to the order statistics $Y_{1:m}^*, \dots, Y_{m:m}^*$ (e.g., Bhattacharya, 1974; Yang, 1976; David and Nagaraja, 2003, Section 6.8). Then we write

$$\begin{aligned} I_{n,m} &= \frac{1}{m}\left(\sum_{j=1}^m X_{j,m}^* \mathbb{1}_{((G_m^*)^{-1}(p),\infty)}\big(Y_{j:m}^*\big) - \sum_{j=1}^m X_{j,m}^* \mathbb{1}_{(G_n^{-1}(p),\infty)}\big(Y_{j:m}^*\big)\right) \\ &= \frac{1}{m}\left(\sum_{j=K^*}^m X_{j,m}^* - \sum_{j=M^*}^m X_{j,m}^*\right) \\ &= \frac{\operatorname{sgn}\big(M^* - K^*\big)}{m} \sum_{j=K^*\wedge M^*}^{K^*\vee M^*-1} X_{j,m}^*, \end{aligned} \tag{3.6}$$

where

$$\begin{aligned} K^* &= \min\big\{j \,:\, Y_{j:m}^* > (G_m^*)^{-1}(p), 1 \le j \le m\big\}, \\ M^* &= \min\big\{j \,:\, Y_{j:m}^* > G_n^{-1}(p), 1 \le j \le m\big\}, \end{aligned}$$

with $\operatorname{sgn}(a)$ denoting the sign function, which is equal to 0 when $a = 0$ and $a/|a|$ when $a \neq 0$. In the definitions of $K^*$ and $M^*$, we take the minimum over those $j \in \{1, \dots, m\}$ that satisfy the noted inequalities, but if there is not any such $j$, then the corresponding minimum is set to $m+1$. Hence,

$$\begin{aligned} \mathbb{P}^*\big(|I_{n,m}| > \varepsilon\big) &\le \mathbb{P}^*\left(\frac{1}{m}\sum_{j=K^*\wedge M^*}^{K^*\vee M^*-1} |X_{j,m}^*| > \varepsilon\right) \\ &\le \mathbb{P}^*\left(\frac{1}{m}\sum_{j=K^*\wedge M^*}^{K^*\vee M^*-1} |X_{j,m}^*| > \varepsilon,\, \mathcal{B}_{K^*,n,m} \cap \mathcal{B}_{M^*,n,m}\right) \\ &\qquad + \mathbb{P}^*\big(\mathcal{B}_{K^*,n,m}^c\big) + \mathbb{P}^*\big(\mathcal{B}_{M^*,n,m}^c\big), \end{aligned} \tag{3.7}$$

where

$$\mathcal{B}_{k,n,m} = \left\{ Y^*_{k:m} \in \left( G_n^{-1}(p-\delta_m), G_n^{-1}(p+\delta_m) \right] \right\}$$

with $k = K^*$ and $k = M^*$. We estimate the first probability on the right-hand side of bound (3.7) as follows:

$$\begin{aligned}
&\mathbb{P}^* \left( \frac{1}{m} \sum_{j=K^* \wedge M^*}^{K^* \vee M^* - 1} |X^*_{j,m}| > \varepsilon,\ \mathcal{B}_{K^*,n,m} \cap \mathcal{B}_{M^*,n,m} \right) \\
&\quad = \mathbb{P}^* \left( \frac{1}{m} \sum_{j=K^* \wedge M^*}^{K^* \vee M^* - 1} |X^*_{j,m}| \mathbb{1}_{\left( G_n^{-1}(p-\delta_m), G_n^{-1}(p+\delta_m) \right]} \left( Y^*_{j:m} \right) > \varepsilon,\ \mathcal{B}_{K^*,n,m} \cap \mathcal{B}_{M^*,n,m} \right) \\
&\quad \le \mathbb{P}^* \left( \frac{1}{m} \sum_{j=1}^{m} |X^*_{j,m}| \mathbb{1}_{\left( G_n^{-1}(p-\delta_m), G_n^{-1}(p+\delta_m) \right]} \left( Y^*_{j:m} \right) > \varepsilon \right) \\
&\quad \le \frac{1}{\varepsilon m} \sum_{j=1}^{m} \mathbb{E}^* \left( |X^*_{j,m}| \mathbb{1}_{\left( G_n^{-1}(p-\delta_m), G_n^{-1}(p+\delta_m) \right]} \left( Y^*_{j:m} \right) \right) \\
&\quad = \frac{1}{\varepsilon} \mathbb{E}^* \left( |X^*_1| \mathbb{1}_{\left( G_n^{-1}(p-\delta_m), G_n^{-1}(p+\delta_m) \right]} \left( Y^*_1 \right) \right) \\
&\quad = \frac{1}{\varepsilon n} \sum_{i=1}^{n} |X_i| \mathbb{1}_{\left( G_n^{-1}(p-\delta_m), G_n^{-1}(p+\delta_m) \right]} \left( Y_i \right).
\end{aligned}$$

Consequently,

$$\begin{aligned}
&\mathbb{E} \left( \mathbb{P}^* \left( \frac{1}{m} \sum_{j=K^* \wedge M^*}^{K^* \vee M^* - 1} |X^*_{j,m}| \mathbb{1}_{\left( G_n^{-1}(p-\delta_m), G_n^{-1}(p+\delta_m) \right]} \left( Y^*_{j:m} \right) > \varepsilon \right) \mathbb{1}_{\mathcal{A}_{n,m}} \right) \\
&\quad \le \mathbb{E} \left( \frac{1}{\varepsilon n} \sum_{i=1}^{n} |X_i| \mathbb{1}_{\left( G^{-1}(p-2\delta_m), G^{-1}(p+2\delta_m) \right]} \left( Y_i \right) \right) \\
&\quad = \frac{1}{\varepsilon} \mathbb{E} \left( |X| \mathbb{1}_{\left( G^{-1}(p-2\delta_m), G^{-1}(p+2\delta_m) \right]} \left( Y \right) \right) \\
&\quad = \frac{1}{\varepsilon} \int_{p-2\delta_m}^{p+2\delta_m} h(G^{-1}(t)) \mathrm{d}t,
\end{aligned} \tag{3.8}$$

where $h(y) = \mathbb{E}(|X| \mid Y = y)$. The integral on the right-hand side of bound (3.8) converges to 0 because of condition ($\mathrm{C}_1$) and $\delta_m \to 0$. It remains to reach analogous conclusions about the expected values of the last two probabilities on the right-hand side of bound (3.7).

We start with the statement

$$\mathbb{E} \left( \mathbb{P}^* \left( \mathcal{B}^c_{M^*,n,m} \right) \right) \to 0. \tag{3.9}$$

Using the notation

$$\mathcal{C}_{n,m} = \Big\{ G_n(G_n^{-1}(p)) \le p + 0.25\delta_m \Big\} \cap \Big\{ p + 0.75\delta_m \le G_n(G_n^{-1}(p+\delta_m)) \Big\},$$

we have

$$\begin{aligned} \mathbb{E}\left(\mathbb{P}^*\big(\mathcal{B}^c_{M^*,n,m}\big)\right) &\le \mathbb{E}\left(\mathbb{P}^*\big(\mathcal{B}^c_{M^*,n,m}\big)\mathbb{1}_{\mathcal{C}_{n,m}}\right) + \mathbb{P}\big(\mathcal{C}^c_{n,m}\big) \\ &= \mathbb{E}\left(\mathbb{P}^*\big(\mathcal{B}^c_{M^*,n,m}\big)\mathbb{1}_{\mathcal{C}_{n,m}}\right) + o(1), \end{aligned} \tag{3.10}$$

where the right-most equation holds due to Lemmas 4.2 and 4.3. We have

$$\begin{aligned} \mathbb{P}^*\big(\mathcal{B}^c_{M^*,n,m}\big) &= \mathbb{P}^*\Big(Y^*_{M^*:m} \notin \big(G_n^{-1}(p-\delta_m), G_n^{-1}(p+\delta_m)\big]\Big) \\ &= \mathbb{P}^*\Big(Y^*_{M^*:m} \notin \big(G_n^{-1}(p), G_n^{-1}(p+\delta_m)\big]\Big) \\ &= \mathbb{P}^*\Big(Y^*_j \notin \big(G_n^{-1}(p), G_n^{-1}(p+\delta_m)\big], 1 \le j \le m\Big) \\ &= \bigg(1 - \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\big(G_n^{-1}(p), G_n^{-1}(p+\delta_m)\big]}(Y_i)\bigg)^m \\ &= \bigg(1 - \Big(G_n(G_n^{-1}(p+\delta_m)) - G_n(G_n^{-1}(p))\Big)\bigg)^m. \end{aligned}$$

Consequently,

$$\begin{aligned} \mathbb{P}^*\big(\mathcal{B}^c_{M^*,n,m}\big)\mathbb{1}_{\mathcal{C}_{n,m}} &\le \big(1 - 0.5\delta_m\big)^m \\ &\le e^{-0.5m\delta_m} \end{aligned} \tag{3.11}$$

with the right-hand side converging to 0 because $m\delta_m \to \infty$. Bounds (3.10) and (3.11) imply statement (3.9)

To complete the proof of Statement 3.1, we are left to show

$$\mathbb{E}\left(\mathbb{P}^*\big(\mathcal{B}^c_{K^*,n,m}\big)\right) \to 0. \tag{3.12}$$

For this, we start with the equations

$$\begin{aligned} \mathbb{P}^*\big(\mathcal{B}^c_{K^*,n,m}\big) &= \mathbb{P}^*\Big(Y^*_{K^*:m} \notin \big(G_n^{-1}(p-\delta_m), G_n^{-1}(p+\delta_m)\big]\Big) \\ &= \mathbb{P}^*\Big(Y^*_{K^*:m} \le G_n^{-1}(p-\delta_m)\Big) + \mathbb{P}^*\Big(Y^*_{K^*:m} > G_n^{-1}(p+\delta_m)\Big) \end{aligned} \tag{3.13}$$

and then work with the two probabilities on the right-hand side separately.

We first prove

$$\mathbb{E}\Big(\mathbb{P}^*\big(Y^*_{K^*:m} \le G_n^{-1}(p-\delta_m)\big)\Big) \to 0. \tag{3.14}$$

Since $Y^*_{K^*:m} > (G^*_m)^{-1}(p)$, we have

$$\begin{aligned}\mathbb{P}^*\Big(Y^*_{K^*:m} \le G_n^{-1}(p-\delta_m)\Big) &\le \mathbb{P}^*\Big((G^*_m)^{-1}(p) \le G_n^{-1}(p-\delta_m)\Big) \\ &= \mathbb{P}^*\Big(G^*_m(G_n^{-1}(p-\delta_m)) \ge p\Big) \\ &= \mathbb{P}^*\Big(\sqrt{m}\,\big(G^*_m(G_n^{-1}(p-\delta_m)) - p_{n,m}\big) \ge \sqrt{m}\,(p-p_{n,m})\Big), \end{aligned} \tag{3.15}$$

where

$$\begin{aligned} p_{n,m} &= \mathbb{E}^*\big(G^*_m(G_n^{-1}(p-\delta_m))\big) \\ &= G_n(G_n^{-1}(p-\delta_m)). \end{aligned}$$

By Lemma 4.4, $\sqrt{m}\,(p-p_{n,m})$ tends in probability to $+\infty$, which means that for every $L \in (0,\infty)$,

$$\mathbb{P}\Big(\sqrt{m}\,(p-p_{n,m}) \le L\Big) \to 0.$$

Denote the event

$$\mathcal{F}_{n,m} = \Big\{\sqrt{m}\,(p-p_{n,m}) > L\Big\}.$$

Continuing with bound (3.15), we therefore have

$$\begin{aligned}\mathbb{E}\Big(\mathbb{P}^*\big(Y^*_{K^*:m} \le G_n^{-1}(p-\delta_m)\big)\Big) &= \mathbb{E}\Big(\mathbb{P}^*\big(Y^*_{K^*:m} \le G_n^{-1}(p-\delta_m)\big)\mathbb{1}_{\mathcal{F}_{n,m}}\Big) + o(1) \\ &\le \mathbb{E}\Big(\mathbb{P}^*\big(\kappa^*_{n,m} > L\big)\Big) + o(1), \end{aligned} \tag{3.16}$$

where

$$\kappa^*_{n,m} = \sqrt{m}\sup_{x\in\mathbb{R}} \big|G^*_m(x) - G_n(x)\big|.$$

By the Dvoretzky-Kiefer-Wolfowitz (DKW) inequality (Dvoretzky et al., 1956; Massart, 1990), we have

$$\mathbb{P}^*\big(\kappa^*_{n,m} > L\big) \le 2e^{-2L^2}. \tag{3.17}$$

Hence, statement (3.14) follows from inequality (3.16).

Finally, we show that the expectation of the last probability on the right-hand side of equation (3.13) vanishes asymptotically, that is,

$$\mathbb{E}\Big(\mathbb{P}^*\Big(Y^*_{K^*:m} > G_n^{-1}(p+\delta_m)\Big)\Big) \to 0. \tag{3.18}$$

We start with the equation

$$\mathbb{P}^*\Big(Y^*_{K^*:m} > G_n^{-1}(p+\delta_m)\Big) = \mathbb{P}^*\Big(S^*_{n,m} < K^*\Big), \tag{3.19}$$

where

$$\begin{aligned} S^*_{n,m} =& \sum_{i=1}^{m} \mathbb{1}_{\left(-\infty, G_n^{-1}(p+\delta_m)\right]}(Y_i^*) \\ =& m G_m^*(G_n^{-1}(p+\delta_m)). \end{aligned}$$

Since

$$\begin{aligned} K^* &= \min\left\{j \,:\, Y^*_{j:m} > (G_m^*)^{-1}(p), 1 \le j \le m\right\} \\ &= 1 + \max\left\{j \,:\, Y^*_{j:m} \le (G_m^*)^{-1}(p), 1 \le j \le m\right\} \\ &= 1 + \sum_{i=1}^{m} \mathbb{1}_{\left(-\infty, (G_m^*)^{-1}(p)\right]}(Y_i^*) \\ &= 1 + m G_m^*\big((G_m^*)^{-1}(p)\big), \end{aligned}$$

equation (3.19) implies

$$\begin{aligned} &\mathbb{P}^*\Big(Y^*_{K^*:m} > G_n^{-1}(p+\delta_m)\Big) \\ &\quad = \mathbb{P}^*\Big(G_m^*(G_n^{-1}(p+\delta_m)) < m^{-1} + G_m^*((G_m^*)^{-1}(p))\Big) \\ &\quad = \mathbb{P}^*\Big(\sqrt{m}\,\big(G_m^*(G_n^{-1}(p+\delta_m)) - q_{n,m}\big) < m^{-1/2} + \sqrt{m}\,\big(G_m^*((G_m^*)^{-1}(p)) - q_{n,m}\big)\Big), \end{aligned} \tag{3.20}$$

where

$$\begin{aligned} q_{n,m} &= \mathbb{E}^*\big(G_m^*(G_n^{-1}(p+\delta_m))\big) \\ &= G_n(G_n^{-1}(p+\delta_m)). \end{aligned}$$

By Lemma 4.6, we have, for every $L \in (0, \infty)$,

$$\mathbb{E}\bigg(\mathbb{P}^*\Big(\sqrt{m}\,\big(G_m^*((G_m^*)^{-1}(p)) - q_{n,m}\big) \geq -L\Big)\bigg) \to 0.$$

Denote the event

$$\mathcal{G}_{n,m} = \Big\{\sqrt{m}\,\big(G_m^*((G_m^*)^{-1}(p)) - q_{n,m}\big) < -L\Big\}.$$

Continuing with bound (3.20), we therefore obtain

$$\begin{aligned}
\mathbb{E}\Big(\mathbb{P}^*\Big(Y^*_{K^*:m} > G_n^{-1}(p+\delta_m)\Big)\Big) &= \mathbb{E}\Big(\mathbb{P}^*\Big(Y^*_{K^*:m} > G_n^{-1}(p+\delta_m)\Big)\mathbb{1}_{\mathcal{G}_{n,m}}\Big) + o(1) \\
&\leq \mathbb{E}\Big(\mathbb{P}^*\big(\kappa^*_{n,m} > L\big)\Big) + o(1).
\end{aligned}$$

Consequently, in view of inequality (3.17), statement (3.18) follows. This establishes Statement 3.1. $\square$

**Statement 3.2.** *For every $\varepsilon > 0$ and when $n \wedge m \to \infty$, we have*

$$\mathbb{E}\Big(\mathbb{P}^*\big(|J_{n,m}| > \varepsilon\big)\Big) \to 0. \tag{3.21}$$

*Proof.* Since (e.g., Shorack, 2000, Inequality 3.1, p. 350)

$$\begin{aligned}
\mathbb{P}^*\big(|J_{n,m}| > \varepsilon\big) &\leq 7\varepsilon \int_0^{1/\varepsilon} \big(1 - \mathrm{Re}\,\big(\mathbb{E}^*\big(e^{itJ_{n,m}}\big)\big)\big)\mathrm{d}t \\
&\leq 7\varepsilon \int_0^{1/\varepsilon} \big|\mathbb{E}^*\big(e^{itJ_{n,m}}\big) - 1\big|\mathrm{d}t,
\end{aligned}$$

statement (3.21) follows if, for every real $t \in \mathbb{R}$,

$$\mathbb{E}\Big(\big|\mathbb{E}^*\big(e^{itJ_{n,m}}\big) - 1\big|\Big) \to 0. \tag{3.22}$$

To verify statement (3.22), we first write

$$J_{n,m} = \sum_{j=1}^{m} \zeta_{j,n,m},$$

where

$$\zeta_{j,n,m} = \frac{1}{m}\bigg(X_j^*\mathbb{1}_{(G_n^{-1}(p),\infty)}(Y_j^*) - \frac{1}{n}\sum_{i=1}^{n} X_i\mathbb{1}_{(G_n^{-1}(p),\infty)}(Y_i)\bigg).$$

Note that for fixed $n$ and $m$, conditionally on $(X_1, Y_1), \dots, (X_n, Y_n)$, the random variables

$\zeta_{j,n,m}$, $j = 1, \dots, m$, are independent and identically distributed. They are also centered, that is, $\mathbb{E}^*(\zeta_{j,n,m}) = 0$. Let $\mathcal{D}_{n,m}$ denote an event (to be specified later) whose complement satisfies

$$\mathbb{P}\big(\mathcal{D}_{n,m}^c\big) \to 0. \tag{3.23}$$

Then

$$\begin{aligned}
\mathbb{E}\left(\left|\mathbb{E}^*\left(e^{itJ_{n,m}}\right) - 1\right|\right) &\le \mathbb{E}\left(\left|\mathbb{E}^*\left(e^{itJ_{n,m}}\right) - 1\right| \mathbb{1}_{\mathcal{D}_{n,m}}\right) + o(1) \\
&= \mathbb{E}\left(\left|\prod_{j=1}^{m} \mathbb{E}^*\left(e^{it\zeta_{j,n,m}}\right) - \prod_{j=1}^{m} 1\right| \mathbb{1}_{\mathcal{D}_{n,m}}\right) + o(1) \\
&\le 2\max\{|t|, t^2\}\Delta_{n,m} + o(1), \qquad (3.24)
\end{aligned}$$

where we obtained the right-most inequality by following the proof of Theorem 1 of Borovkov (1988, p. 518–519) and using the notation

$$\Delta_{n,m} = \mathbb{E}\left(\sum_{j=1}^{m} \mathbb{E}^*\big(g_1(\zeta_{j,n,m})\big)\mathbb{1}_{\mathcal{D}_{n,m}}\right)$$

with $g_1(x) = \min\{|x|, x^2\}$. Furthermore, following Borovkov (1988, inequality (3), p. 517), we have

$$\Delta_{n,m} \le \tau \sum_{j=1}^{m} \mathbb{E}\bigg(\mathbb{E}^*\Big(|\zeta_{j,n,m}|\Big)\mathbb{1}_{\mathcal{D}_{n,m}}\bigg) + \sum_{j=1}^{m} \mathbb{E}\bigg(\mathbb{E}^*\Big(|\zeta_{j,n,m}|\mathbb{1}_{[\tau,\infty)}(|\zeta_{j,n,m}|)\Big)\mathbb{1}_{\mathcal{D}_{n,m}}\bigg)$$

for every $\tau \in (0, 1)$. Hence, due to bound (3.24), statement (3.22) holds if

$$\sum_{j=1}^{m} \mathbb{E}\bigg(\mathbb{E}^*\Big(|\zeta_{j,n,m}|\Big)\mathbb{1}_{\mathcal{D}_{n,m}}\bigg) = O(1) \tag{3.25}$$

and, for every $\tau \in (0, 1)$,

$$\sum_{j=1}^{m} \mathbb{E}\bigg(\mathbb{E}^*\Big(|\zeta_{j,n,m}|\mathbb{1}_{[\tau,\infty)}(|\zeta_{j,n,m}|)\Big)\mathbb{1}_{\mathcal{D}_{n,m}}\bigg) = o(1). \tag{3.26}$$

Note that the left-hand side of equation (3.25) is equal to the left-hand side of equation (3.26) if we set $\tau = 0$. This explains why we next temporarily shift from $\tau \in (0, \infty)$ to $\tau \in [0, \infty)$.

To prove statements (3.25) and (3.26), we define the aforementioned event $\mathcal{D}_{n,m}$ as follows:

$$\mathcal{D}_{n,m} := \Big\{G_n^{-1}(p) > G^{-1}(p-\delta_m)\Big\} \cap \Big\{\frac{1}{n}\sum_{i=1}^{n} |X_i| \mathbb{1}_{(G_n^{-1}(p),\infty)}(Y_i) \le \Lambda\Big\},$$

where $\Lambda \in (0,\infty)$ is a sufficiently large constant that ensures condition (3.23); see Lemma 4.1 (statement (4.2) with $c_1 = 0$ and $c_2 = 1$) and Lemma 4.7. Then, for every $\tau \in [0,1)$,

$$\begin{aligned}
&\mathbb{E}^*\Big(|\zeta_{j,n,m}| \mathbb{1}_{[\tau,\infty)}(|\zeta_{j,n,m}|)\Big) \mathbb{1}_{\mathcal{D}_{n,m}} \\
&\qquad = \frac{1}{mn}\sum_{k=1}^{n} \Bigg(\Big|X_k \mathbb{1}_{(G_n^{-1}(p),\infty)}(Y_k) - \frac{1}{n}\sum_{i=1}^{n} X_i \mathbb{1}_{(G_n^{-1}(p),\infty)}(Y_i)\Big| \\
&\qquad\qquad \times \mathbb{1}_{[m\tau,\infty)}\Big(\Big|X_k \mathbb{1}_{(G_n^{-1}(p),\infty)}(Y_k) - \frac{1}{n}\sum_{i=1}^{n} X_i \mathbb{1}_{(G_n^{-1}(p),\infty)}(Y_i)\Big|\Big)\Bigg) \mathbb{1}_{\mathcal{D}_{n,m}} \\
&\qquad \le \frac{1}{mn}\sum_{k=1}^{n} |X_k| \mathbb{1}_{(G_n^{-1}(p),\infty)}(Y_k) \mathbb{1}_{[m\tau-\Lambda,\infty)}\Big(|X_k| \mathbb{1}_{(G_n^{-1}(p),\infty)}(Y_k)\Big) \mathbb{1}_{\mathcal{D}_{n,m}} \\
&\qquad\qquad + \frac{\Lambda}{mn}\sum_{k=1}^{n} \mathbb{1}_{[m\tau-\Lambda,\infty)}\Big(|X_k| \mathbb{1}_{(G_n^{-1}(p),\infty)}(Y_k)\Big) \mathbb{1}_{\mathcal{D}_{n,m}} \\
&\qquad \le \frac{1}{mn}\sum_{k=1}^{n} |X_k| \mathbb{1}_{(G^{-1}(p-\delta_m),\infty)}(Y_k) \mathbb{1}_{[m\tau-\Lambda,\infty)}\Big(|X_k| \mathbb{1}_{(G^{-1}(p-\delta_m),\infty)}(Y_k)\Big) \\
&\qquad\qquad + \frac{\Lambda}{mn}\sum_{k=1}^{n} \mathbb{1}_{[m\tau-\Lambda,\infty)}\Big(|X_k| \mathbb{1}_{(G^{-1}(p-\delta_m),\infty)}(Y_k)\Big). \qquad (3.27)
\end{aligned}$$

We next show that the just derived inequality implies statements (3.25) and (3.26), for which we again resume considering the cases $\tau = 0$ and $\tau \in (0,\infty)$ separately.

To prove statement (3.25), we set $\tau = 0$ in inequality (3.27) and have

$$\begin{aligned}
\sum_{j=1}^{m} \mathbb{E}\Big(\mathbb{E}^*\Big(|\zeta_{j,n,m}| \mathbb{1}_{\mathcal{D}_{n,m}}\Big)\Big) &\le \sum_{j=1}^{m} \mathbb{E}\Big(\frac{1}{mn}\sum_{k=1}^{n} |X_k| \mathbb{1}_{(G^{-1}(p-\delta_m),\infty)}(Y_k)\Big) + \Lambda \\
&= \mathbb{E}\Big(\frac{1}{n}\sum_{k=1}^{n} |X_k| \mathbb{1}_{(G^{-1}(p-\delta_m),\infty)}(Y_k)\Big) + \Lambda \\
&= \mathbb{E}\big(|X| \mathbb{1}_{(G^{-1}(p-\delta_m),\infty)}(Y)\big) + \Lambda. \qquad (3.28)
\end{aligned}$$

The right-hand side is finite for all sufficiently large $m$, due to condition $(C_1)$. This establishes statement (3.25).

To prove statement (3.26), we set any $\tau \in (0,\infty)$ in inequality (3.27) and have

$$\begin{aligned}
&\sum_{j=1}^{m} \mathbb{E}\bigg( \mathbb{E}^*\Big( |\zeta_{j,n,m}| \mathbb{1}_{[\tau,\infty)}(|\zeta_{j,n,m}|) \mathbb{1}_{\mathcal{D}_{n,m}} \Big) \bigg) \\
&\qquad \le \mathbb{E}\bigg( \frac{1}{n} \sum_{k=1}^{n} |X_k| \mathbb{1}_{(G^{-1}(p-\delta_m),\infty)}(Y_k) \mathbb{1}_{[m\tau-\Lambda,\infty)}\Big( |X_k| \mathbb{1}_{(G^{-1}(p-\delta_m),\infty)}(Y_k) \Big) \bigg) \\
&\qquad\quad + \mathbb{E}\bigg( \frac{\Lambda}{n} \sum_{k=1}^{n} \mathbb{1}_{[m\tau-\Lambda,\infty)}\Big( |X_k| \mathbb{1}_{(G^{-1}(p-\delta_m),\infty)}(Y_k) \Big) \bigg) \\
&\qquad = \mathbb{E}\bigg( |X| \mathbb{1}_{(G^{-1}(p-\delta_m),\infty)}(Y) \mathbb{1}_{[m\tau-\Lambda,\infty)}\Big( |X| \mathbb{1}_{(G^{-1}(p-\delta_m),\infty)}(Y) \Big) \bigg) \\
&\qquad\quad + \Lambda \mathbb{E}\bigg( \mathbb{1}_{[m\tau-\Lambda,\infty)}\Big( |X| \mathbb{1}_{(G^{-1}(p-\delta_m),\infty)}(Y) \Big) \bigg).
\end{aligned} \tag{3.29}$$

No matter what the (fixed) value of $\Lambda \in (0,\infty)$ is, the right-most side of bound (3.29) converges to 0 when $m \to \infty$ because $\mathbb{E}\big(|X| \mathbb{1}_{(G^{-1}(p-\delta),\infty)}(Y)\big) < \infty$ for some $\delta > 0$. Hence, statement (3.26) holds. This, in turn, completes the proof of statement (3.22) and establishes Statement 3.2. Theorem 2.1 is proved. $\square$

# 4 Auxiliary lemmas

Throghout this section, we frequently use the notation

$$\kappa_n = \sqrt{n} \sup_{x \in \mathbb{R}} \big| G_n(x) - G(x) \big|$$

as well as the DKW inequality (Dvoretzky et al., 1956; Massart, 1990), according to which the bound

$$\mathbb{P}\big(\kappa_n > L\big) \le 2e^{-2L^2} \tag{4.1}$$

holds for every $L \in (0,\infty)$.

**Lemma 4.1.** *If conditions* ($C_2$) *and* ($C_3$) *are satisfied, then, for all real constants* $c_1 < c_2$ *and when* $n \wedge m \to \infty$*, we have*

$$\mathbb{P}\Big( G_n^{-1}(p - c_1\delta_m) \le G^{-1}(p - c_2\delta_m) \Big) \to 0 \tag{4.2}$$

*and*

$$\mathbb{P}\Big( G_n^{-1}(p + c_1\delta_m) > G^{-1}(p + c_2\delta_m) \Big) \to 0. \tag{4.3}$$

*Proof.* We have

$$\begin{aligned}\mathbb{P}\Big(G_n^{-1}(p-c_1\delta_m) \le G^{-1}(p-c_2\delta_m)\Big) &= \mathbb{P}\Big(p-c_1\delta_m \le G_n(G^{-1}(p-c_2\delta_m))\Big)\\ &\le \mathbb{P}\Big(p-c_1\delta_m \le G(G^{-1}(p-c_2\delta_m)) + \kappa_n/\sqrt{n}\Big)\\ &\le \mathbb{P}\Big(\kappa_n \ge (c_2-c_1)\sqrt{n}\delta_m\Big), \qquad (4.4)\end{aligned}$$

where we used condition (C$_2$) to have $G(G^{-1}(p-c_2\delta_m)) = p - c_2\delta_m$ for all sufficiently large $m$. The right-hand side of bound (4.4) converges to 0 due to DKW inequality (4.1) and $(c_2-c_1)\sqrt{n}\delta_m \to \infty$. This establishes statement (4.2).

Furthermore, we have

$$\begin{aligned}\mathbb{P}\Big(G_n^{-1}(p+c_1\delta_m) > G^{-1}(p+c_2\delta_m)\Big) &= \mathbb{P}\Big(p+c_1\delta_m > G_n(G^{-1}(p+c_2\delta_m))\Big)\\ &\le \mathbb{P}\Big(p+c_1\delta_m > G(G^{-1}(p+c_2\delta_m)) - \kappa_n/\sqrt{n}\Big)\\ &\le \mathbb{P}\Big(\kappa_n > (c_2-c_1)\sqrt{n}\delta_m\Big), \qquad (4.5)\end{aligned}$$

where we used condition (C$_2$) to have $G(G^{-1}(p+c_2\delta_m)) = p + c_2\delta_m$ for all sufficiently large $m$. The right-hand side of bound (4.5) converges to 0 due to DKW inequality (4.1) and $(c_2 - c_1)\sqrt{n}\delta_m \to \infty$. This establishes statement (4.3) and completes the proof of Lemma 4.1. □

**Lemma 4.2.** *If conditions* (C$_2$) *and* (C$_3$) *are satisfied, then, when* $n \wedge m \to \infty$*,*

$$\mathbb{P}\Big(G_n(G_n^{-1}(p)) \ge p + 0.25\delta_m\Big) \to 0.$$

*Proof.* By Lemma 4.1 (statement (4.3) with $c_1 = 0$ and $c_2 = 0.1$) we have

$$\mathbb{P}\Big(G_n^{-1}(p) > G^{-1}(p+0.1\delta_m)\Big) \to 0.$$

Therefore,

$$\begin{aligned}\mathbb{P}\Big(G_n(G_n^{-1}(p)) \ge p+0.25\delta_m\Big) &\le \mathbb{P}\Big(G_n(G_n^{-1}(p)) \ge p+0.25\delta_m, G_n^{-1}(p) \le G^{-1}(p+0.1\delta_m)\Big) + o(1)\\ &\le \mathbb{P}\Big(G_n(G^{-1}(p+0.1\delta_m)) \ge p+0.25\delta_m\Big) + o(1)\\ &\le \mathbb{P}\Big(G(G^{-1}(p+0.1\delta_m)) \ge p+0.25\delta_m - \kappa_n/\sqrt{n}\Big) + o(1)\\ &= \mathbb{P}\Big(\kappa_n \ge 0.15\sqrt{n}\delta_m\Big) + o(1).\end{aligned}$$

By DKW inequality (4.1), the right-hand side converges to 0 because $\sqrt{n}\delta_m \to \infty$. This establishes Lemma 4.2. $\square$

**Lemma 4.3.** *If conditions* (C$_2$) *and* (C$_3$) *are satisfied, then, when* $n \wedge m \to \infty$*,*

$$\mathbb{P}\Big(G_n(G_n^{-1}(p+\delta_m)) \le p + 0.75\delta_m\Big) \to 0.$$

*Proof.* By Lemma 4.1 (statement (4.2) with $c_1 = -1$ and $c_2 = -0.9$), we have

$$\mathbb{P}\Big(G_n^{-1}(p+\delta_m) \le G^{-1}(p+0.9\delta_m)\Big) \to 0.$$

Therefore,

$$\begin{aligned}
&\mathbb{P}\Big(G_n(G_n^{-1}(p+\delta_m)) \le p + 0.75\delta_m\Big) \\
&\qquad \le \mathbb{P}\Big(G_n(G_n^{-1}(p+\delta_m)) \le p + 0.75\delta_m, G_n^{-1}(p+\delta_m) > G^{-1}(p+0.9\delta_m)\Big) + o(1) \\
&\qquad \le \mathbb{P}\Big(G_n(G^{-1}(p+0.9\delta_m)) \le p + 0.75\delta_m\Big) + o(1) \\
&\qquad \le \mathbb{P}\Big(G(G^{-1}(p+0.9\delta_m)) \le p + 0.75\delta_m + \kappa_n/\sqrt{n}\Big) + o(1) \\
&\qquad = \mathbb{P}\Big(\kappa_n \ge 0.15\sqrt{n}\delta_m\Big) + o(1).
\end{aligned}$$

The right-hand side converges to 0 because of DKW inequality (4.1) and $\sqrt{n}\delta_m \to \infty$. This establishes Lemma 4.3. $\square$

**Lemma 4.4.** *If conditions* (C$_2$) *and* (C$_3$) *are satisfied, then, for every* $L \in \mathbb{R}$ *and when* $n \wedge m \to \infty$*,*

$$\mathbb{P}\Big(\sqrt{m}\left(p - G_n(G_n^{-1}(p-\delta_m)\right) \le L\Big) \to 0.$$

*Proof.* By Lemma 4.1 (statement (4.3) with $c_1 = -1$ and $c_2 = -0.5$), we have

$$\mathbb{P}\Big(G_n^{-1}(p-\delta_m) > G^{-1}(p-0.5\delta_m)\Big) \to 0.$$

Hence,

$$
\begin{aligned}
\mathbb{P}&\Big(G_n(G_n^{-1}(p-\delta_m)) \geq p - L/\sqrt{m}\Big) \\
&\leq \mathbb{P}\Big(G_n(G_n^{-1}(p-\delta_m)) \geq p - L/\sqrt{m}, G^{-1}(p-0.5\delta_m) \geq G_n^{-1}(p-\delta_m)\Big) + o(1) \\
&\leq \mathbb{P}\Big(G_n(G^{-1}(p-0.5\delta_m)) \geq p - L/\sqrt{m}\Big) + o(1) \\
&\leq \mathbb{P}\Big(G(G^{-1}(p-0.5\delta_m)) \geq p - L/\sqrt{m} - \kappa_n/\sqrt{n}\Big) + o(1) \\
&\leq \mathbb{P}\Big(\kappa_n \geq 0.5\sqrt{n}\delta_m - L\sqrt{n/m}\Big) + o(1) \\
&\leq \mathbb{P}\Big(\kappa_n \geq 0.5\sqrt{n}\big(\delta_m - L\sqrt{1/m}\big)\Big) + o(1),
\end{aligned}
$$

where we used condition (C$_2$) to have $G(G^{-1}(p-0.5\delta_m)) = p - 0.5\delta_m$ for all sufficiently large $m$. The right-hand side of the above bound converges to 0 because of DKW inequality (4.1) and $\sqrt{n}\delta_m \to \infty$. This establishes Lemma 4.4. $\square$

The following lemma is a bootstrap-version of statement (4.3), which we shall later need for establishing Lemma 4.6.

**Lemma 4.5.** *If conditions (C$_2$) and (C$_3$) are satisfied, then, for all real constants $c_1 < c_2$ and when $n \wedge m \to \infty$, we have*

$$
\mathbb{E}\bigg(\mathbb{P}^*\Big((G_m^*)^{-1}(p + c_1\delta_m) > G^{-1}(p + c_2\delta_m)\Big)\bigg) \to 0.
$$

*Proof.* We have

$$
\begin{aligned}
\mathbb{P}^*&\Big((G_m^*)^{-1}(p + c_1\delta_m) > G^{-1}(p + c_2\delta_m)\Big) \\
&= \mathbb{P}^*\Big(p + c_1\delta_m > G_m^*(G^{-1}(p + c_2\delta_m))\Big) \\
&\leq \mathbb{P}^*\left(p + c_1\delta_m > G(G^{-1}(p + c_2\delta_m)) - \frac{1}{\sqrt{m}}\kappa_{n,m}^* - \frac{1}{\sqrt{n}}\kappa_n\right) \\
&\leq \mathbb{P}^*\left(\kappa_{n,m}^* + \sqrt{\frac{m}{n}}\kappa_n > (c_2 - c_1)\sqrt{m}\delta_m\right),
\end{aligned}
$$

where we used condition (C$_2$) to have $G(G^{-1}(p + c_2\delta_m)) = p + c_2\delta_m$ for all sufficiently large $m$. Hence,

$$
\mathbb{E}\bigg(\mathbb{P}^*\Big((G_m^*)^{-1}(p) > G^{-1}(p + \delta_m)\Big)\bigg) \leq \mathbb{E}\bigg(\mathbb{P}^*\left(\kappa_{n,m}^* + \sqrt{\frac{m}{n}}\kappa_n > (c_2 - c_1)\sqrt{m}\delta_m\right)\bigg).
$$

To prove that the right-hand side converges to 0, we use DKW inequalities (3.17) and (4.1), and also the fact that $(c_2 - c_1)\sqrt{m}\delta_m \to \infty$. This finishes the proof of Lemma 4.5. $\square$

**Lemma 4.6.** *If conditions* (C$_2$) *and* (C$_3$) *are satisfied, then, for every* $L \in \mathbb{R}$ *and when* $n \wedge m \to \infty$*,*

$$\mathbb{E}\Big(\mathbb{P}^*\Big(\sqrt{m}\left(G_m^*((G_m^*)^{-1}(p)) - G_n(G_n^{-1}(p+\delta_m))\right) \geq L\Big)\Big) \to 0.$$

*Proof.* We have

$$\begin{aligned}
G_m^*((G_m^*)^{-1}(p)) - G_n(G_n^{-1}(p+\delta_m)) &= G_m^*((G_m^*)^{-1}(p)) - G_n((G_m^*)^{-1}(p)) \\
&\quad + G_n((G_m^*)^{-1}(p)) - G((G_m^*)^{-1}(p)) \\
&\quad + G((G_m^*)^{-1}(p)) - G(G_n^{-1}(p+\delta_m)) \\
&\quad + G(G_n^{-1}(p+\delta_m)) - G_n(G_n^{-1}(p+\delta_m)) \\
&\leq G((G_m^*)^{-1}(p)) - G(G_n^{-1}(p+\delta_m)) \\
&\quad + \frac{1}{\sqrt{m}}\kappa_{n,m}^* + \frac{2}{\sqrt{n}}\kappa_n.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathbb{E}\Big(\mathbb{P}^*\Big(\sqrt{m}\left(G_m^*((G_m^*)^{-1}(p)) - G_n(G_n^{-1}(p+\delta_m))\right) \geq L\Big)\Big) \\
&\qquad \leq \mathbb{E}\Big(\mathbb{P}^*\Big(\sqrt{m}\left(G((G_m^*)^{-1}(p)) - G(G_n^{-1}(p+\delta_m))\right) + \kappa_{n,m}^* + 2\sqrt{\frac{m}{n}}\,\kappa_n \geq L\Big)\Big).
\end{aligned}$$

To prove that the right-hand side converges to 0, we use DKW inequalities (3.17) and (4.1), and in this way reduce the problem to showing

$$\mathbb{E}\Big(\mathbb{P}^*\Big(\sqrt{m}\left(G((G_m^*)^{-1}(p)) - G(G_n^{-1}(p+\delta_m))\right) \geq L\Big)\Big) \to 0. \tag{4.6}$$

By Lemma 4.1 (statement (4.2) with $c_1 = -1$ and $c_2 = -0.75$), we have

$$\mathbb{P}\Big(G_n^{-1}(p+\delta_m) \leq G^{-1}(p+0.75\delta_m)\Big) \to 0, \tag{4.7}$$

and by Lemma 4.5 (with $c_1 = 0$ and $c_2 = 0.5$), we have

$$\mathbb{E}\Big(\mathbb{P}^*\Big((G_m^*)^{-1}(p) > G^{-1}(p+0.5\delta_m)\Big)\Big) \to 0. \tag{4.8}$$

Therefore, statement (4.6) follows because, due to condition ($C_2$) and $\sqrt{m}\delta_m \to \infty$, the inequality

$$\sqrt{m}\Big(G(G^{-1}(p+0.5\delta_m)) - G(G^{-1}(p+0.75\delta_m))\Big) \geq L \tag{4.9}$$

is never satisfied for all sufficiently large $m$. This completes the proof of Lemma 4.6. $\square$

**Lemma 4.7.** *If conditions ($C_1$) and ($C_2$) are satisfied, then, for every constant $\Lambda > 0$ such that*

$$\Lambda > \mathbb{E}\Big(|X|\mathbb{1}_{(G^{-1}(p),\infty)}(Y)\Big),$$

*we have, when $n \wedge m \to \infty$,*

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}|X_i|\mathbb{1}_{(G_n^{-1}(p),\infty)}(Y_i) > \Lambda\right) \to 0.$$

*Proof.* This is an immediate consequence of Gribkova et al. (2022) who proved that the empirical TCA is a consistent estimator of the population TCA under conditions ($C_1$) and ($C_2$). $\square$

# References


Ahsanullah, M., Nevzorov, V.B. and Shakil, M. (2013). *An Introduction to Order Statistics.* Atlantis Press, Paris.

Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (2008). *A First Course in Order Statistics.* Society for Industrial and Applied Mathematics, Philadelphia.

Bhattacharya, P.K. (1974). Convergence of sample paths of normalized sum of induced order statistics. *Annals of Statistics*, 2, 1034–1039.

Borovkov, A.A. (1998). *Mathematical Statistics.* Routledge, London.

David, H.A. and Nagaraja, H.N. (2003). *Order Statistics.* (Third Edition.) Wiley, Hoboken.

Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Annals of Mathematical Statistics*, 27, 642–669.

Furman, E., Kye, Y. and Su, J. (2021). Multiplicative background risk models: Setting a course for the idiosyncratic risk factors distributed phase-type. *Insurance: Mathematics and Economics*, 96, 153–167.

Gribkova, N., Su, J. and Zitikis, R. (2021). Empirical tail conditional allocation and its consistency under minimal assumptions. *Annals of the Institute of Statistical Mathematics*, 74, 713–735.

Gribkova, N., Su, J. and Zitikis, R. (2022). Inference for the tail conditional allocation: large sample properties, insurance risk assessment, and compound sums of concomitants. *Insurance: Mathematics and Economics*, 107, 199–222.

Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Annals of Probability*, 18, 1269–1283.

Shorack, G.R. (2000). *Probability for Statisticians.* Springer, New York.

Yang, S.S. (1976). *Concomitants of Order Statistics.* PhD Dissertation, Iowa State University, Ames.