Taylor & Francis
Taylor & Francis Group

When Tukey Meets Chauvenet: A New Boxplot Criterion for Outlier Detection

Hongmei Lin[a], Riquan Zhang[a], and Tiejun Tong[b]

[a]School of Statistics and Data Science, Shanghai University of International Business and Economics, Shanghai, China; [b]Department of Mathematics, Hong Kong Baptist University, Hong Kong, China

ABSTRACT

The box-and-whisker plot, introduced by Tukey, is one of the most popular graphical methods in descriptive statistics. On the other hand, however, Tukey's boxplot is free of sample size, yielding the so-called "one-size-fits-all" fences for outlier detection. Although improvements on the sample size adjusted boxplots do exist in the literature, most of them are either not easy to implement or lack justification. As another common rule for outlier detection, Chauvenet's criterion uses the sample mean and standard derivation to perform the test, but it is often sensitive to the included outliers and hence is not robust. In this article, by combining Tukey's boxplot and Chauvenet's criterion, we introduce a new boxplot, namely the Chauvenet-type boxplot, with the fence coefficient determined by an exact control of the outside rate per observation. Our new outlier criterion not only maintains the simplicity of the boxplot from a practical perspective, but also serves as a robust Chauvenet's criterion. Simulation study and a real data analysis on the civil service pay adjustment in Hong Kong demonstrate that the Chauvenet-type boxplot performs extremely well regardless of the sample size, and can therefore be highly recommended for practical use to replace both Tukey's boxplot and Chauvenet's criterion. Lastly, to increase the visibility of the work, a user-friendly R package named "ChauBoxplot" has also been officially released on CRAN.

1. Introduction

The box-and-whisker plot, introduced by Tukey (1977), is also known as Tukey's boxplot or, simply, the boxplot. It is one of the most popular graphical methods in descriptive statistics, which provides a very efficient tool to visualize the distribution of the sample data. The boxplot not only displays a box ranging from the first quartile (Q_1) to the third quartile (Q_3) of the data, but also provides two whiskers extending from the edge of the box to the lower and upper fences for outlier detection. More specifically, by letting $IQR = Q_3 - Q_1$ be the interquartile range of the data, the lower and upper fences of the boxplot are given as $LF = Q_1 - k \times IQR$ and $UF = Q_3 + k \times IQR$, where $k > 0$ is the fence coefficient. Tukey (1977) further referred to the fences with $k = 1.5$ as the "inner fences", and with $k = 3$ as the "outer fences". The two visualized whiskers are then drawn from the smallest value within the inner fences to the first quartile, and from the third quartile to the largest value within the inner fences. Lastly, the outliers that differ from the majority of the data will be plotted as individual points beyond the whiskers on the boxplot. According to the degree of abnormality, Tukey (1977) also declared the observations between the inner and outer fences as the "outside" outliers, and those beyond the outer fences as the "far out" or "extreme" outliers. For more details on Tukey's boxplot and its variants for univariate data, one may refer to (Frigge, Hoaglin, and Iglewicz 1989; Schwertman, Owens,

and Adnan 2004; Sim, Gan, and Chang 2005 Wickham and Stryjewski 2011; Walker et al. 2018; Rodu and Kafadar 2022), and the references therein.

Apart from the boxplot, the identification of outliers itself is another important issue in robust statistics and related areas. Statistically, an outlier may be defined as an abnormal observation or contaminated data that does not follow the same distribution as the majority of the data. If the outliers are not identified and further excluded from the data, they may have an unexpected impact on the subsequent data analysis, yielding a misleading or erroneous conclusion. Because of this, there is a vast and growing literature on the identification and accommodation of suspected outliers since the 19th century. To name a few, the first rigorous test was proposed by Peirce (1852), which is also known as Peirce's criterion. Due to its complicated procedure, however, Peirce's criterion was soon superseded by an alternative called Chauvenet's criterion. Given a sample of data from a normal distribution with $\bar{X}$ the sample mean and S the sample standard deviation, Chauvenet (1863) declared an observation as an outlier if its value is outside the interval $[\bar{X} - c_n S, \bar{X} + c_n S]$, where $c_n = \Phi^{-1}(1 - 0.25/n)$ is the upper $0.25/n$ quantile of the standard normal distribution. Thanks to its simplicity and effectiveness, Chauvenet's criterion has been widely used in different fields including, but not limited to, astronomy, nuclear technology, geology, epidemiology, molecular biology,

---

CONTACT Riquan Zhang ✉ rqzhang@suibe.edu.cn ▣ School of Statistics and Data Science, Shanghai University of International Business and Economics, Shanghai, China; Tiejun Tong ✉ tongt@hkbu.edu.hk ▣ Department of Mathematics, Hong Kong Baptist University, Hong Kong, China.

radiology, and physical science (Ross 2003; Maples et al. 2018; Vermeesch 2018).

It is noteworthy that there exist many other outlier testing procedures in the literature (Andrews 1974; Andrews and Pregibon 1978; Beckman and Cook 1983; Rosner 1983; Davies and Gather 1993; Fung 1993; Atkinson 1994; Carey et al. 1997; Bates et al. 2023; Bhattacharya, Kamper, and Beirlant 2023), as well as those being summarized in the monographs (Hawkins 1980; Barnett and Lewis 1994; Atkinson and Riani 2000; Hawkins 2006). To the best of our knowledge, Tukey's boxplot and Chauvenet's criterion are still among the most popular methods for outlier detection, mainly thanks to their simplicity as well as the reasonable accuracy (Brant 1990; Milton 1999). Despite their popularity, each of these two methods has well-known disadvantages. Specifically for Tukey's boxplot, we note that the fence coefficient k is free of sample size, yielding the so-called "one-size-fits-all" fences to label the outliers (McGill, Tukey, and Larsen 1978; Sim, Gan, and Chang 2005). Although improvements on the sample size adjusted boxplots do exist in the literature, most of them are either not easy to implement (Sim, Gan, and Chang 2005; Banerjee and Iglewicz 2007) or lack justification (Barbato et al. 2011). For more details, see Section 2. On the other hand, Chauvenet's criterion uses the sample mean and standard derivation to perform the test, and consequently, it is often very sensitive to the included outliers and, hence, may not provide a robust method for outlier detection. For further illustration, see the toy example in Section 3.1.

In this article, by combining Tukey's boxplot and Chauvenet's criterion, we introduce a new boxplot, namely the *Chauvenet-type boxplot*, with the fence coefficient determined by an exact control of the outside rate per observation. Our new outlier criterion not only maintains the simplicity of the boxplot from a practical perspective, but also serves as a robust Chauvenet's criterion. Simulation study and a real data analysis on the civil service pay adjustment in Hong Kong also demonstrate that the Chauvenet-type boxplot performs extremely well regardless of the sample size, and can therefore be highly recommended for practical use to replace both Tukey's boxplot and Chauvenet's criterion. The remainder of the article is organized as follows. In Section 2, we review the existing boxplots for outlier detection that incorporate the sample size into the construction of the lower and upper fences. In Section 3, we first review Chauvenet's criterion and then propose our new boxplot that is a combination of Chauvenet's criterion and the classic boxplot for detecting outliers. In Section 4, we conduct simulation study and real data analysis to compare the performance of the new and existing boxplots as well as Chauvenet's criterion. In Section 5, we further extend the Chauvenet-type boxplot to handle nonnormal data for outlier detection and demonstrate by simulations that our new criterion performs also very well. Lastly, the article is concluded in Section 6 with some discussion and future directions.

2. Existing Boxplots for Outlier Detection

As is known, Tukey's box-and-whisker plot is very appealing in descriptive statistics, mainly thanks to its simplicity and low sensitivity to outlier distortion. On the other hand, however, the whiskers in Tukey's boxplot or some of its variants are free,

or nearly free, of the sample size of the data so that they yield a *one-size-fits-all* rule for outlier detection. In this section, we give a brief review of the major developments of boxplots that aim to incorporate the sample size into the fences. For ease of presentation, we will assume the data is normally distributed in the next three sections, whereas the extensions to nonnormal or skewed data will be presented in Section 5.

Let $X_1, \ldots, X_n$ be a sample of size n from the normal distribution $N(\mu, \sigma^2)$, where μ is the mean and $\sigma^2 > 0$ is the variance. Let also Q_1 and Q_3 be the first and third quartiles of the data, respectively. To construct the boxplot, we define the lower and upper fences as

$$\text{LF}_n = Q_1 - k_n \times \text{IQR} \quad \text{and} \quad \text{UF}_n = Q_3 + k_n \times \text{IQR}, \quad (1)$$

where $\text{IQR} = Q_3 - Q_1$ is the interquartile range and k_n is the fence coefficient depending on the sample size. Hoaglin, Iglewicz, and Tukey (1986) termed the *some-outside rate per sample*, denoted by α, as the probability that one or more observations in the normal sample will be wrongly classified as outliers. To control the some-outside rate, the outlier region is then set to satisfy

$$P \text{ (one or more of } X_1, \ldots, X_n \in (-\infty, \text{LF}_n) \cup (\text{UF}_n, \infty)) = \alpha.$$
$$(2)$$

In other words, if an observation X is either smaller than LF_n or larger than UF_n, then it will be labeled as an α-*outlier* in the boxplot. Hoaglin, Iglewicz, and Tukey (1986) further termed the *outlier labeling rule* to formally describe Tukey's outlier flagging procedure.

Observing that the exact solution of k_n satisfying (2) is rather complicated, Hoaglin, Iglewicz, and Tukey (1986) proposed a multistep approximation procedure to determine the choices of k_n in the boxplot for samples with normal data. One main disadvantage of their approximation procedure is that the statistical expertise and decision are needed in each step, which will unavoidably result in the propagation of approximation errors. In view of this, Hoaglin and Iglewicz (1987) further applied the simulation to obtain the numerical values of k_n for normal samples. Specifically, by converting the boxplot outlier labeling rule to a formal outlier identification testing procedure, they found the value of k_n to be close to 2.2 for $\alpha = 0.05$ and n up to 300.

2.1. Boxplots with Exact Some-Outside Rate

For the boxplots with some-outside rate per sample, Sim, Gan, and Chang (2005) made some extra efforts to derive the exact solution of k_n satisfying (2). Let $X_{(1)} = \min\{X_1, \ldots, X_n\}$ and $X_{(n)} = \max\{X_1, \ldots, X_n\}$ be the minimum and maximum values of the sample. Note that the event {one or more of $X_1, \ldots, X_n \in (-\infty, \text{LF}_n) \cup (\text{UF}_n, \infty)$} is the union of two disjoint compound events $A_1 = \{X_{(n)} > \text{UF}_n\}$ and $A_2 = \{X_{(1)} < \text{LF}_n, X_{(n)} \leq \text{UF}_n\}$. Under the exact control of the some-outside rate, the requirement on (2) implies that $P(A_1) + P(A_2) = \alpha$. In addition, the standardization procedure to $X_1, \ldots, X_n$ can be further applied so that the sample is free of μ and σ. In this section, for simplicity of notation, we directly regard $X_1, \ldots, X_n$ to be the standard normal data. Finally, by the property that the order

statistics of the sample form a Markov chain David and Nagaraja (2003), an exact expression for k_n satisfying (2) can be derived as

$$\int_{-\infty}^{\infty} \int_{x_1}^{\infty} \left\{ 1 - I_{G_u(y_u)}(n-u, 1) \left[1 - I_{G_l(y_l)}(1, l-1) \right] \right.$$
$$\left. \times f_{Q_1, Q_3}(x_1, x_3) \right\} dx_3 dx_1 = \alpha,$$

where $G_l(y_l) = \Phi(y_l)/\Phi(x_1)$ with $y_l = x_1 - k_n(x_3 - x_1)$, $G_u(y_u) = [\Phi(y_u) - \Phi(x_3)]/[1 - \Phi(x_3)]$ with $y_u = x_3 + k_n(x_3 - x_1)$, $I_p(a, b) = \int_0^p t^{a-1}(1-t)^{b-1}dt/B(a, b)$ is the incomplete beta function, and $f_{Q_1, Q_3}(x_1, x_3)$ is the joint probability density function of Q_1 and Q_3. For more details, see eqs. (7) and (8) in Sim, Gan, and Chang (2005) together with the associated text.

Next, to solve k_n from the above question, Sim, Gan, and Chang (2005) further applied the multivariate quadrature rule to numerically approximate the double integral, and then applied a direct search or Monte Carlo integration algorithm to search for the value of k_n that yields an exact control of the some-outside rate per sample. For practical use, they also provided some numerical values of k_n in their Table 1 together with an approximation formula. Taken the normal distribution with $\alpha = 0.05$ and $n = 4m + 1$ with $m \in \{2, 3, \ldots, 124\}$ as an example, the fence coefficient with exact some-outside rate (ER) is approximated as

$$k_n^{\mathrm{ER}} \approx \exp\{4.01761 - 2.35363 \ln(n) + 0.64618 \ln^2(n)$$
$$- 0.07893 \ln^3(n) + 0.00368 \ln^4(n)\}. \tag{3}$$

For more numerical approximations with other values of α and n, one may refer to Table A.1 from their Appendix, in which the maximum absolute deviation δ between the true and approximated values of k_n^{ER} were also given.

2.2. Boxplots with Tolerance Limits

As an alternative, Sim, Gan, and Chang (2005) further applied the tolerance interval, an important statistical tool in reliability and quality control, to construct the fences of the boxplot. A tolerance interval is an interval determined from a random sample in such a way that one may have a specified $100\gamma\%$ level of confidence that the interval will cover at least a specified β proportion of the sampled population. For more on tolerance intervals, one may refer to Guttman (1970), Patel (1986), and Liao and Iyer (2004). For a certain distribution, let $f(x)$ be the probability density function and $F(x)$ be the cumulative distribution function. Let also $L(\boldsymbol{X}) = L(X_1, \ldots, X_n)$ and $U(\boldsymbol{X}) = U(X_1, \ldots, X_n)$ be two statistics with $L(\boldsymbol{X}) < U(\boldsymbol{X})$. Then with the probability coverage of the random interval $W = \int_{L(\boldsymbol{X})}^{U(\boldsymbol{X})} f(x)dx = F(U(\boldsymbol{X})) - F(L(\boldsymbol{X}))$, the interval $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ defines a two-sided (γ, β) tolerance interval if $P(W \geq \beta) = \gamma$, where γ and β are two specified probability values.

As a good example, the tolerance interval with $L(\boldsymbol{X}) = \bar{X} - kS$ and $U(\boldsymbol{X}) = \bar{X} + kS$ has been widely used in quality control Guenther (1977), where $\bar{X}$ is the sample mean and S is the sample standard deviation. Now with the boxplot, Sim, Gan, and Chang (2005) constructed a (γ, β) tolerance interval with $L(\boldsymbol{X})$ and $U(\boldsymbol{X})$ being the lower and upper fences in (1), respectively. Observations that fall outside the tolerance limits are labeled as outliers. Then by letting $\beta = 1 - \alpha$, the value of k_n is defined to

satisfy $P(P(\text{all of } X_1, \ldots, X_n \in (\mathrm{LF}_n, \mathrm{UF}_n) \geq 1 - \alpha)) = \gamma$, or equivalently,

$$P\left(\int_{\mathrm{LF}_n}^{\mathrm{UF}_n} f(x)dx \geq 1 - \alpha_n \right) = \gamma,$$

where $\alpha_n = 1 - (1-\alpha)^{1/n}$. Moreover, by the symmetry of the normal distribution, the value of k_n can be determined from the equation $P(P(X < \mathrm{LF}_n) \leq \alpha_n/2) = \gamma$. Finally, numerical approximation using the quadrature rule needs to be implemented for computing the integral, followed by a direct search algorithm or Monte Carlo integration algorithm to search for the value of k_n. For practical use, the authors also provided the numerical values of k_n in their Table 2 for different combinations of (α, γ, n), together with the approximation formulas in Table A.2. Taken the normal distribution with $\alpha = 0.05$, $\gamma = 0.9$ and $n = 4m + 1$ with $m \in \{2, 3, \ldots, 124\}$ as an example, the fence coefficient with tolerance limits (TL) is approximated as

$$k_n^{\mathrm{TL}} \approx \exp\{4.45171 - 2.44501 \ln(n) + 0.64990 \ln^2(n)$$
$$- 0.07851 \ln^3(n) + 0.00365 \ln^4(n)\}. \tag{4}$$

2.3. Boxplots with Asymptotic Fences

In view of the complexity in deriving the exact expression of k_n satisfying (2), Iglewicz and Banerjee (2001) and Banerjee and Iglewicz (2007) proposed an alternative solution based on the asymptotics. Specifically, they first derived the limiting solution of k_n from (2) when n goes to infinity, and then regress the fitted values back to smaller sample sizes. By the symmetry of the normal distribution and the fact that {one or more of $X_1, \ldots, X_n \in (-\infty, \mathrm{LF}_n) \cup (\mathrm{UF}_n, \infty)$} $= \{X_{(1)} < \mathrm{LF}_n \text{ or } X_{(n)} > \mathrm{UF}_n\}$, Banerjee and Iglewicz (2007) thus, simplified the solution to (2), approximately, as the solution to the equation of

$$P\left(X_{(n)} \leq Q_3 + k_n \times \mathrm{IQR} \right) = 1 - \alpha/2.$$

Finally, since $P(X_{(n)} \leq x) = [P(X_1 \leq x)]^n$, it suffices to solve $P(X_1 \leq Q_3 + k_n \times \mathrm{IQR}) = (1 - \alpha/2)^{1/n}$ for the value of k_n.

For normal data, as mentioned in Section 2.1, the solution of k_n to the above question does not depend on μ and σ. Thus for simplicity, we assume that the data is sampled from $N(0, 1)$ with $\Phi(x)$ being the cumulative distribution function. When n is sufficiently large, by replacing Q_1 by $\Phi^{-1}(0.25)$ and Q_3 by $\Phi^{-1}(0.75)$, it leads to the limiting solution as $k_n = [\Phi^{-1}(1 - \alpha/2)^{1/n}) - \Phi^{-1}(0.75)]/(\Phi^{-1}(0.75) - \Phi^{-1}(0.25)) = [\Phi^{-1}(1 - \alpha/2)^{1/n}) - 0.6745]/1.349$. Finally, on the basis of a simulation study for smaller sample sizes, Banerjee and Iglewicz (2007) recommended their fence coefficient with asymptotic fences (AF) as

$$k_n^{\mathrm{AF}} \approx a_n \times \frac{\Phi^{-1}\left((1 - \alpha/2)^{1/n} \right) - 0.6745}{1.349}, \tag{5}$$

where a_n is the smoothed adjustment value with $a_n = 1 + 8.9764n^{-1} - 126.6262n^{-2} + 1531.7064n^{-3} - 10729.3439n^{-4}$ for $n < 2000$, and $a_n = 1$ for $n \geq 2000$.

2.4. Boxplots with Empirical Considerations

Recall that Tukey's outlier flagging procedure takes a constant k when constructing the lower and upper fences, yielding a one-size-fits-all boxplot for labeling the potential outliers. To take into account the sample size, unlike those complicated procedures in the previous sections that control either the some-outside rate or the tolerance limit, Barbato, Genta, and Levi (2009) suggested a modified IQR method based on some empirical considerations (EC). Specifically, by modifying IQR with IQR$[1 + 0.1\ln(n/10)]$, their fence coefficient associated with $k = 1.5$ is given as

$$k_n^{\text{EC}} = 1.5 \times [1 + 0.1\ln(n/10)]. \tag{6}$$

Note that k_n^{EC} is monotonically increasing with n, and so is the interval length consisting of the two fences, in such a way that the dependence on the sample size is introduced into the boxplot. In addition, as stated by Barbato et al. (2011), their boxplot outlier labeling rule retains its primary advantages, including simplicity and the lack of need for iteration, thanks to the robust estimate of spread achieved by excluding both tails.

We are not aware of any newer developments for constructing the sample size adjusted boxplots with univariate normal data, expect for those with nonnormal or skewed data for which we will discuss later in Section 6. It is also noteworthy that, during the last decade, researchers have turned their main attention to other types of boxplots including, for example, boxplots for functional data (Hyndman and Shang 2010; Sun and Genton 2011; Dai and Genton 2018; Qu and Genton 2022), boxplots for circular data (Abuzaid, Mohamed, and Hussin 2012; Buttarazzi, Pandolfo, and Porzio 2018), boxplots for contours, curves and paths (Whitaker, Mirzargar, and Kirby 2013; Mirzargar, Whitaker, and Kirby 2014; Raj et al. 2017), and boxplots for large data (Hofmann, Wickham, and Kafadar 2017).

3. Chauvenet-Type Boxplot

As reviewed in Section 2, the whiskers in Tukey's boxplot are free of the sample size of the data, yielding the so-called "one-size-fits-all" criterion for outlier detection. Although improvements on the sample size adjusted boxplots do exist in the literature, most of them are either not easy to implement or lack justification. As another example for illustration, on the basis of Hoaglin, Iglewicz, and Tukey (1986) and Hoaglin and Iglewicz's (1987) boxplot outlier labeling rules, Frigge, Hoaglin, and Iglewicz (1989) further conducted a survey on popular software packages to standardize the fence coefficient selection for boxplot construction. It is noted, however, that they finally recommended to use Tukey's (1977) fence constants of 1.5 or 3 as the default value in the boxplot. This, from another perspective, shows that the sample size adjusted whiskers in their boxplot papers in 1986 and 1987 may not be very practical for outlier detection.

3.1. Chauvenet's Criterion

Apart from the boxplot that uses the lower and upper quartiles, another common approach is to construct the outlier regions based on the mean ($\bar{X}$) and standard deviation (S) of the sample data. Among this category, the 3-*sigma limits* are the best known in the quality control chart (Pukelsheim 1994; Wheeler and Chambers 2010), by which a data point will be labeled as a suspected outlier if outside the interval $[\bar{X} - 3S, \bar{X} + 3S]$. As can be seen, this method is relatively simple to implement, and in addition, the constant 3 is not a must as the multiplier of the standard deviation. For instance, $[\bar{X} - 2S, \bar{X} + 2S]$, $[\bar{X} - 2.5S, \bar{X} + 2.5S]$ and $[\bar{X} - 4S, \bar{X} + 4S]$ are also occasionally used, depending on whether a more relaxed or strict control is desired. More interestingly, with *a constant time of sigma* clipping, we note that the 3-sigma limits work in a similar way as Tukey's whiskers with *a constant time* of IQR as the length. Taking the normal data as an example, Tukey's outlier labeling rule with $k_n = 1.5$ or $k_n = 1.72$ is asymptotically equivalent to the 2.7-sigma or 3-sigma clipping, respectively Morales, Giraldo, and Torres (2021).

To incorporate the sample size into the sigma clipping rules, the first rigorous test using the probability theory was proposed by Peirce in 1852, which is also known as Peirce's criterion Peirce (1852), Gould (1855), Peirce (1878). Peirce's test is based on what are now called the Z-scores, by which any observations with $|Z| > c$ will be rejected, where c depends on both the sample size and the number of suspected outliers. Due to the complicated form of c, however, Peirce's test is not easy to implement which greatly hinders its popularity for practical use. And in fact, it was soon superseded by an alternative called *Chauvenet's criterion*, which identifies a data point as an outlier if the probability of obtaining a value at least as extreme as the observed one is less than $0.5/n$, where n is the size of the sample. Statistically, by letting $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ with $\bar{X}$ as the sample mean and S as the sample standard deviation, Chauvenet (1863) first computed the deviation of each data point from the mean as $D_i = |X_i - \bar{X}|/S$, where $i = 1, \ldots, n$. He then set $c_n = \Phi^{-1}(1 - 0.25/n)$ as the threshold, and declared each data point with deviation D_i greater than c_n as a suspected outlier. Or equivalently, if we present the criterion using the sigma clipping, then a data point will be labeled as an outlier if outside the interval

$$[\bar{X} - c_nS, \bar{X} + c_nS]. \tag{7}$$

To summarize, Chauvenet's criterion will reject, on average, half an observation of genuine data from the normal distribution regardless of the sample size.

As mentioned in Section 1, Chauvenet's criterion has been widely used in many different fields thanks to its simplicity and effectiveness Ross (2003), as well as its usage spread across government laboratories, industry, and universities. As a more recent example, Chauvenet's criterion and its modified version have been adopted for outlier detection by the very popular toolbox, *IsoplotR*, for geochronology Vermeesch (2018), which has been cited more than 2500 citations in Google Scholar as of December 2024. Despite its popularity, Chauvenet's criterion has drawbacks. An obvious limitation is that it can be very sensitive to the included outliers and, hence, may not provide a robust method for outlier detection.

To illustrate why Chauvenet's criterion is lack of robustness, we also consider a toy example with $n = 9$ observations, where $X_1, \ldots, X_7$ are a random sample of size 7 from the standard normal distribution and $X_8 = X_9 = 100$ are two contaminated data points. For ease of presentation, we further sort the

observations in ascending order so that the whole data set is $X = \{-1.938, -1.177, -0.854, -0.353, 0.890, 0.916, 1.741, 100, 100\}$. Then with the sample mean $\bar{X} = 22.136$ and standard deviation $S = 44.160$, the deviations of the data are given as

$$D = \frac{|X - \bar{X}|}{S}$$
$$= \{0.545, 0.528, 0.521, 0.509, 0.481, 0.481, 0.462, 1.763, 1.763\}.$$

Finally, since the maximum deviation $D_{\max} = 1.763$ is less than the threshold $c_9 = \Phi^{-1}(1 - 0.25/9) = 1.915$, Chauvenet's criterion will hence declare no data point as suspected outliers, although the last two are clearly contaminated.

3.2. A New Boxplot Criterion for Outlier Detection

As seen from the above toy example, Chauvenet's criterion, and more generally the sigma clipping rules, suffer from the included outliers via the contaminated sample mean and standard deviation. To improve the robustness, better estimates of the mean and standard deviation ought to be considered, for example using the quantile information as given in Tukey's boxplot. On the other hand, as reviewed in Section 2, Tukey's outlier labeling rule and its existing variants are themselves not perfect in the way of incorporating the sample size. To further improve boxplots for outlier detection, effective strategies involving the sample size adjusted sigma clipping can be incorporated.

Inspired by this, we propose to take advantage of both Tukey's boxplot and Chauvenet's criterion. More specifically, through the fusion of these two classic technologies, we are able to introduce a new boxplot criterion for more accurately and more robustly labeling the outliers of the data. By formula (1), recall that the lower and upper fences in Tukey's boxplot are $\mathrm{LF}_n = Q_1 - k_n \times \mathrm{IQR}$ and $\mathrm{UF}_n = Q_3 + k_n \times \mathrm{IQR}$. Now to apply Chauvenet's criterion to determine the k_n value, we first transform the boxplot fences from the two quartiles, Q_1 and Q_3, back to the sample mean and standard deviation. For normal data, by Tukey (1977) we can apply the midhinge $(Q_1 + Q_3)/2$ to estimate the sample mean, yielding $Q_1 + Q_3 \approx 2\bar{X}$; Further by Higgins et al. (2019), we apply $(Q_3 - Q_1)/1.35$ to estimate the sample standard deviation, yielding $Q_3 - Q_1 \approx 1.35S$. Solving the above two equations, we have $Q_1 \approx \bar{X} - 0.675S$ and $Q_3 \approx \bar{X} + 0.675S$. Finally, by plugging them into (1), the new lower and upper fences represented by the sample mean and standard deviation can be expressed as $\mathrm{LF}_n = \bar{X} - (1.35k_n + 0.675)S$ and $\mathrm{UF}_n = \bar{X} + (1.35k_n + 0.675)S$.

Now to fuse Tukey's boxplot with Chauvenet's criterion, we apply the upper $0.25/n$ quantile of the standard normal distribution, that is $c_n = \Phi^{-1}(1 - 0.25/n)$, as the threshold for labeling anomalous observations. This leads to $1.35k_n + 0.675 = \Phi^{-1}(1 - 0.25/n)$, or equivalently, $k_n = \Phi^{-1}(1 - 0.25/n)/1.35 - 0.5$. Taken together, the lower and upper fences of our new boxplot criterion for outlier detection are, respectively,

$$\mathrm{LF}_n^{\mathrm{Chau}} = Q_1 - k_n^{\mathrm{Chau}} \times \mathrm{IQR} \quad \text{and}$$
$$\mathrm{UF}_n^{\mathrm{Chau}} = Q_3 + k_n^{\mathrm{Chau}} \times \mathrm{IQR}, \tag{8}$$

where the fence coefficient associated with Chauvenet's criterion is

$$k_n^{\mathrm{Chau}} = \frac{\Phi^{-1}(1 - 0.25/n)}{1.35} - 0.5. \tag{9}$$

We refer to the above boxplot with new whiskers as the *Chauvenet-type boxplot*. With the new fence coefficient k_n^{Chau} in (9), the Chauvenet-type boxplot not only provides an exact control of the outside rate per observation, but also well maintains the simplicity of the boxplot from the perspective of practice.

Lastly, to demonstrate that our new boxplot criterion is more robust for outlier detection than Chauvenet's criterion, we revisit the toy example in Section 3.1. We have $Q_1 = -0.854$ and $Q_3 = 1.741$, which are indeed the 3rd and 7th observations respectively. Further by (8) and (9), we have $k_n^{\mathrm{Chau}} = 0.918$, and consequently, $[\mathrm{LF}_n^{\mathrm{Chau}}, \mathrm{UF}_n^{\mathrm{Chau}}] = [-3.237, 4.124]$. Finally, observing that the two contaminated data points with value 100 are far beyond the upper fence at 4.124, our new boxplot criterion for outlier detection will thus declare them as extreme outliers. On the other hand, the lower and upper limits using Chauvenet's criterion are given as $\bar{X} \pm 1.915S = [-62.430, 106.702]$, providing a very wide interval that contains all nine observations, and hence fail to declare any anomalous data points. To conclude, it is evident that our new boxplot criterion has provided a robust version of Chauvenet's criterion, in such a way that the lower and upper fences will no longer be sensitive to the extreme values when they present.

3.3. Further Exploration

When the data is normal, the following theorem establishes an asymptotic equivalence between our new boxplot criterion and Chauvenet's criterion for outlier detection, with the proof being relegated to Section S1 of the online supplementary material.

Theorem 1. For normal data, the outlier region of the Chauvenet-type boxplot is asymptotically equivalent to that of Chauvenet's criterion. More specifically by (7) and (8), we have

$$\bar{X} - c_n S \simeq Q_1 - k_n^{\mathrm{Chau}} \times \mathrm{IQR}, \tag{10}$$
$$\bar{X} + c_n S \simeq Q_3 + k_n^{\mathrm{Chau}} \times \mathrm{IQR}, \tag{11}$$

where $\simeq$ denotes the asymptotic equivalence such that if $a_n \simeq b_n$, then $a_n/b_n \to 1$ as $n \to \infty$.

By Theorem 1, the Chauvenet-type boxplot provides a comparable performance to Chauvenet's criterion when the normal data is not contaminated. But when the data contains outliers, both $\bar{X}$ and S may be dramatically affected so that the lower and upper thresholds of Chauvenet's criterion in (7) tends be wider, or much wider, than is expected. In contrast, our new boxplot criterion will be statistically resistant to the presence of outliers. Specifically by the definition in Huber (2004), the breakdown point of our new boxplot criterion will be as large as 25% of the whole data. In other words, when the number of outliers do not exceed 25%, the lower and upper fences in (8) for our new boxplot criterion will remain unchanged. Taking the toy example for illustration, with two contaminated data being 100, the percentage of the outliers is $2/9 = 22.2\% < 25\%$ and so they will not affect the outlier region. This explains why our new boxplot criterion can successfully detect the two anomalous data points, whereas Chauvenet's criterion fails to do so due to the very wide interval (7) caused by the contaminated sample mean and standard deviation.
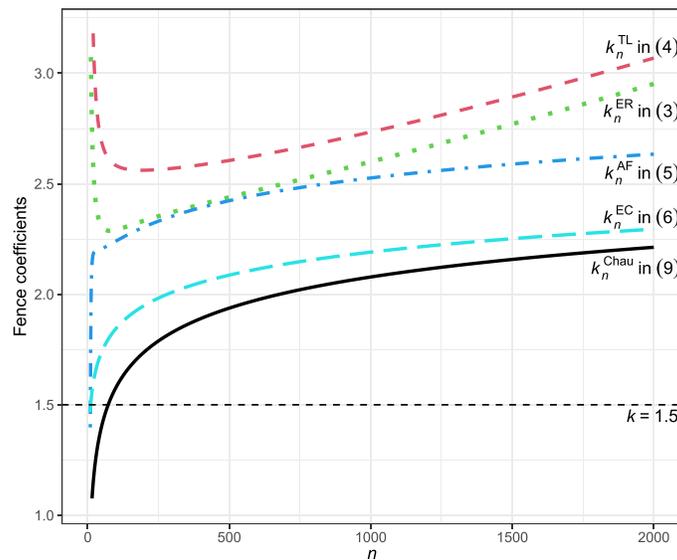
Figure 1. The sample size adjusted fence coefficients for the new and existing boxplots, where n ranges from 8 to 2000. The dotted line represents the fence coefficient k_n^{ER} in (3), the dashed line represents the fence coefficient k_n^{TL} in (4), the dash-dotted line represents the fence coefficient k_n^{AF} in (5), the long-dashed line represents the fence coefficient k_n^{EC} in (6), and the solid line represents our new fence coefficient k_n^{Chau} in (9). Lastly, the fence constant of $k = 1.5$ in Tukey's boxplot is also given for reference.

Next, to explore the finite sample performance of the Chauvenet-type boxplot, we further plot the values of k_n^{Chau} in Figure 1 with the sample size ranging from 8 to 2000. And for comparison, the existing fence coefficients reviewed in Section 2 are also ploted, including the fence coefficient k_n^{ER} in (3), the fence coefficient k_n^{TL} in (4), the fence coefficient k_n^{AF} in (5), and the fence coefficient k_n^{EC} in (6). By Figure 1, it is evident that the existing sample size adjusted methods all provide a longer whisker, and hence a more stringent criterion, than our new boxplot criterion for labeling the suspected outliers. We also note that our new fence coefficient is much closer to Tukey's fence constant with $k = 1.5$, and in particular, the two fence coefficients will be the same when $n = 72$. In other words, with the sample size less than 72, we recommend a whisker even shorter than that in Tukey's boxplot for outlier detection, in a way to more effectively increase the detection power for outliers. In contrast, we observe that the two fence coefficients in Sim, Gan, and Chang (2005), including k_n^{ER} in (3) and k_n^{TL} in (4), both display a decreasing pattern when n is small which may not be reasonable. While for the fence coefficient k_n^{AF} in (5), it has a rapidly increasing trend when n is less than 20, thus, yielding a less-smooth function of the sample size.

3.4. R package

To increase the visibility of the work, a user-friendly R package named "ChauBoxplot" has also been officially released on CRAN. To install and load the package, one may enter the following commands in the R console:

```
> install.packages("ChauBoxplot")
> library(ChauBoxplot)
```

The new package consists of two main functions, *chau_boxplot()* and *geom_chau_boxplot()*, for graphically drawing the Chauvenet-type boxplot. For the first function, it can be operated the same way as *boxplot()* in "base R" except that the new fence coefficient k_n^{Chau} in (9) is now adopted rather than the default $k = 1.5$. For

the second function, it can also be operated the same way as *geom_boxplot()* in "ggplot2". Lastly, for reference, the source files and example codes are freely accessible on GitHub at *https://github.com/tiejuntong/ChauBoxplot*.

4. Comparison of the boxplots

We compare the performance of the new and existing boxplots for outlier detection using normal data in Section 4.1, nonnormal data in Section 4.2, and real data in Section 4.3. Note that the sample size adjusted whiskers, as reviewed in Section 2, are seldom used in practice due to the difficulty to implement or lack of justification. Thus to save space, we will only present the comparison results for the Chauvenet-type boxplot with the fence coefficient k_n^{Chau} in (9) and the classic Tukey's boxplot with the fence coefficient $k = 1.5$. We will demonstrate that our new boxplot criterion associated with the Chauvenet-type boxplot is not only simple and practical, but also more accurate and robust in labeling outliers.

4.1. Normal Data

Given a total sample size of n, we generate the first $n-2$ observations as a random sample from the standard normal distribution, and designate the last two observations as contaminated data with $X_{n-1} = 5$ and $X_n = 6$. Then with $n = 50, 500, 5000$, and 50,000, we plot the simulated data in Figure 2 using Tukey's boxplot (T.boxplot) and the Chauvenet-type boxplot (C.boxplot). From the figure, we note that Tukey's boxplot will label more and more genuine data mistakenly as outliers when the sample size becomes larger. As an example, with *set.seed(1863)* in R, Tukey's boxplot will label a total of 3, 48, and 357 suspected outliers when $n = 500, 5000$, and 50,000, respectively. By excluding the two contaminated data, the remaining 1, 46, and 355 outliers are indeed genuine data from the standard normal distribution and so are wrongly labeled, which coincides with
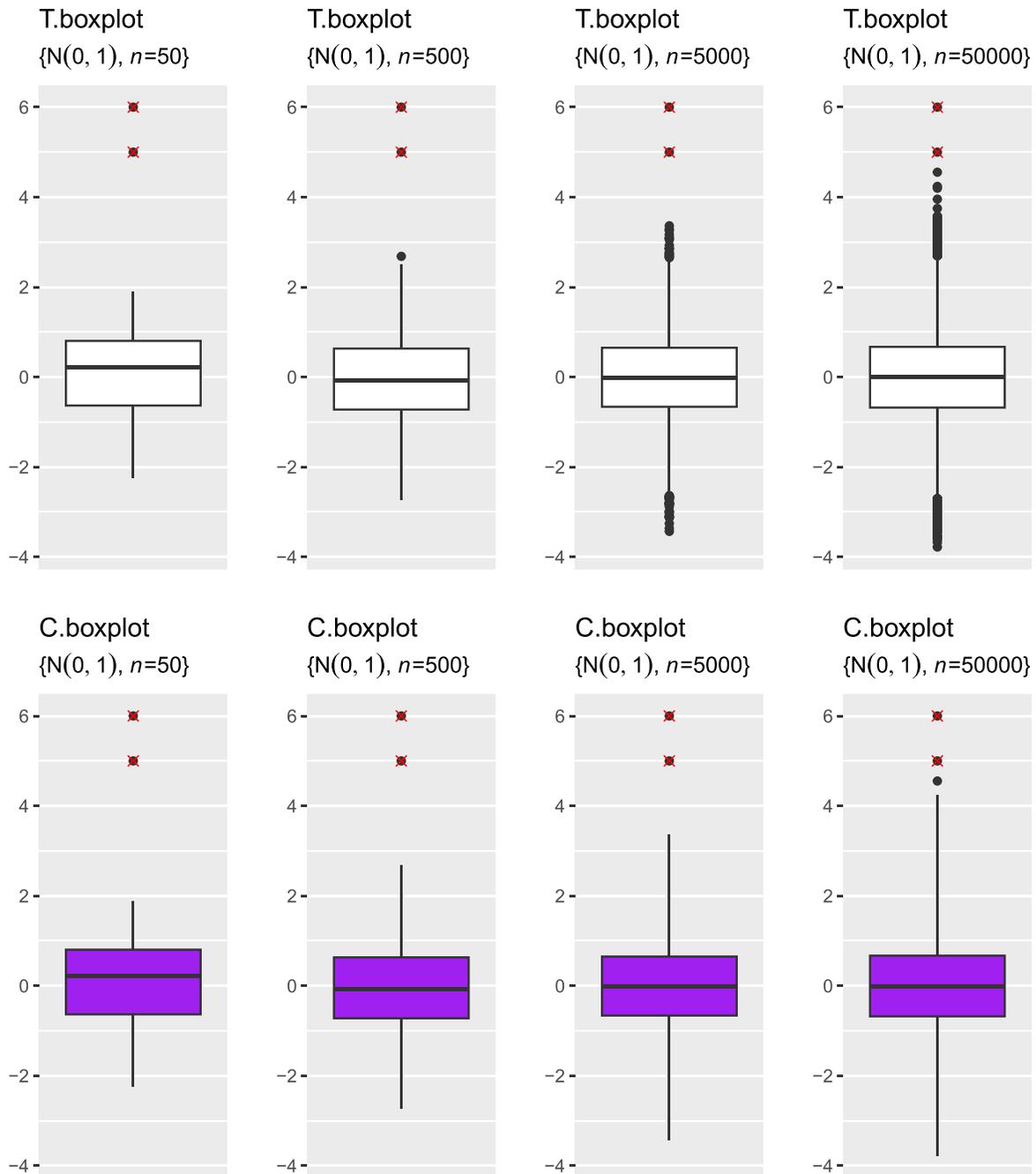
Figure 2. Outlier detection for normal data using Tukey's boxplot (T.boxplot) and the Chauvenet-type boxplot (C.boxplot), where the data is generated from the standard normal distribution with two contaminated data as 5 and 6. With *set.seed(1863)* in R, Tukey's boxplot labels a total of 2, 3, 48, and 357 suspected outliers, and the Chauvenet-type boxplot labels a total of 2, 2, 2, and 3 suspected outliers, for the simulated data with $n = 50, 500, 5000$ and 50,000, respectively. Lastly, the double glyph, consisting of a cross and a dot, indicates that the observation is both contaminated data and labeled as an outlier.

the well-known statement that "if the data is sampled from the normal distribution, then about 0.7% of them will lie outside the two fences of the boxplot" (Hubert and Vandervieren 2008). While for the Chauvenet-type boxplot, it labels a total of 2, 2, 2, and 3 suspected outliers for the simulated data with four different sample sizes, respectively. If we further exclude the two contaminated data, it then only labels 0, 0, 0, and 1 genuine data as outliers by mistake for each boxplot. This coincides with Chauvenet's criterion that it will reject, on average, half an observation of genuine data from the normal distribution, regardless of how large the sample size is.

4.2. Nonnormal Data

For nonnormal data, we first consider two common distributions. One is the Chi-square distribution, which is right skewed, and the other is Student's t distribution, which is symmetric but heavy tailed. We also consider 8 and 30 to represent the small and large degrees of freedom associated with each distribution, but to save space, we only present the simulation results for 8 degrees of freedom in the main text. Then with $n = 50$, 500, 5000, and 50,000, we generate sample data from the two distributions and then plot them in Figures 3 and 4, respectively, using Tukey's boxplot (T.boxplot) and the Chauvenet-
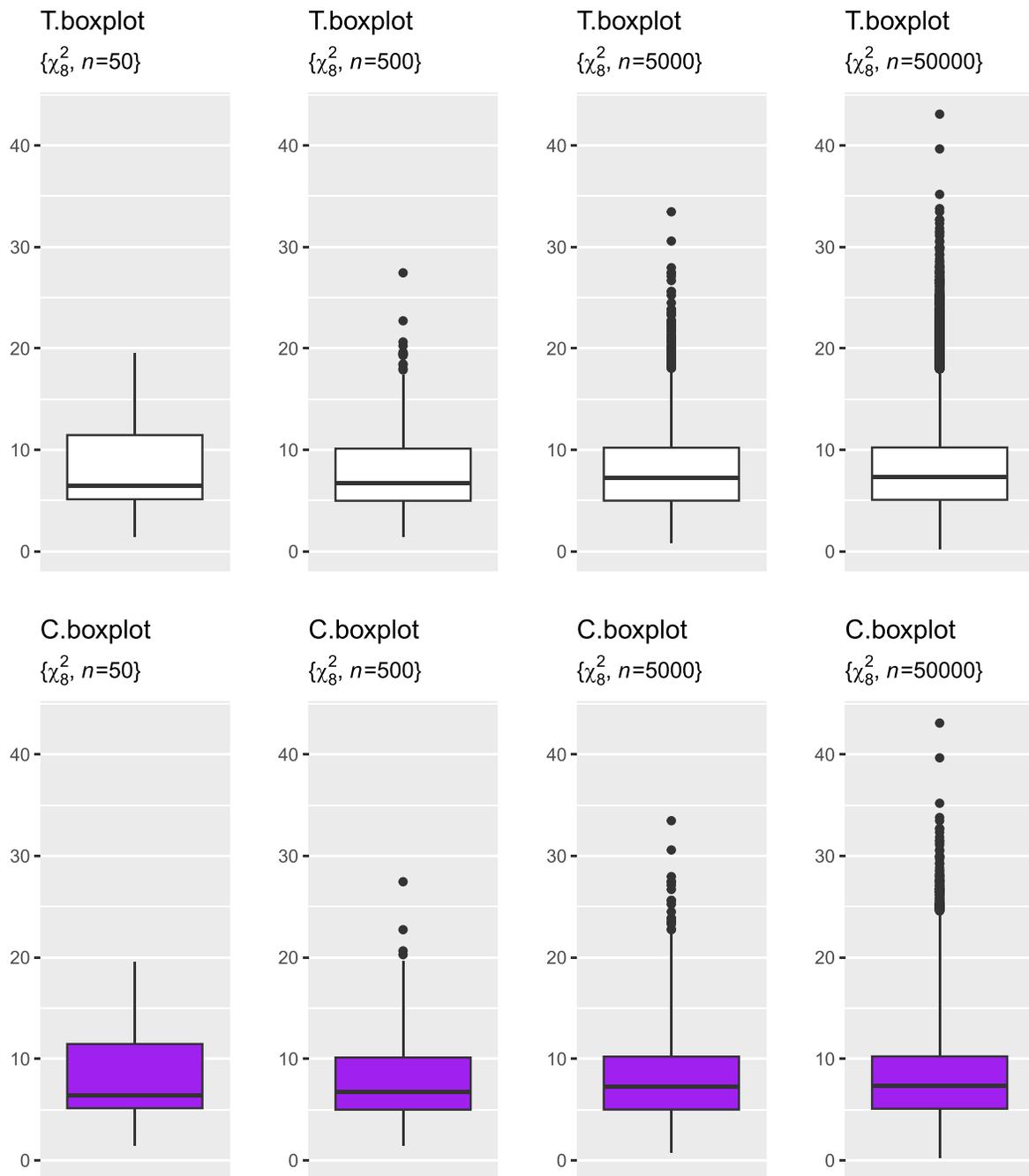
Figure 3. Outlier detection for skewed data using Tukey's boxplot (T.boxplot) and the Chauvenet-type boxplot (C.boxplot), where the data is generated from the Chi-square distribution with 8 degrees of freedom. With *set.seed(1863)* in R, Tukey's boxplot labels a total of 0, 13, 102, and 1102 suspected outliers, and the Chauvenet-type boxplot labels a total of 0, 4, 18 and 106 suspected outliers, for the simulated data with $n = 50$, 500, 5000 and 50,000, respectively.

type boxplot (C.boxplot). As is seen, there is no contaminated data involved in this simulation; in other words, this study does not suffer from any real outliers. Nevertheless, the two figures show that both Tukey's boxplot and the Chauvenet-type boxplot have wrongly labeled many outliers, especially when the sample size is large. Notably, a similar pattern has also been observed from other skewed or heavy tailed distributions, including, but not limited to, the beta, exponential, gamma, and log-normal distributions. For more details, please refer to the boxplots in the online supplementary material.

For a closer comparison, we note that the two boxplots perform similarly when $n = 50$, mainly because the fence coef-

ficient by Chauvenet's criterion, $k_{50}^{\text{Chau}} = 1.41$, is very close to Tukey's coefficient at $k = 1.5$. When the sample size becomes larger, both of the boxplots will increasingly label more genuine data as outliers, but our new boxplot criterion is relatively much better. Taking *set.seed(1863)* again in R, for the χ_8^2 distribution with $n = 500$, 5000, and 50,000, respectively, Tukey's boxplot labels a total of 13, 102 and 1102 suspected outliers, whereas our new boxplot only labels 4, 18, and 106 suspected outliers. And for the t_8 distribution with $n = 500$, 5000, and 50,000, respectively, Tukey's boxplot labels a total of 8, 117, and 1104 suspected outliers, whereas our new boxplot only labels 3, 18, and 90 suspected outliers. To conclude, although not perfect,
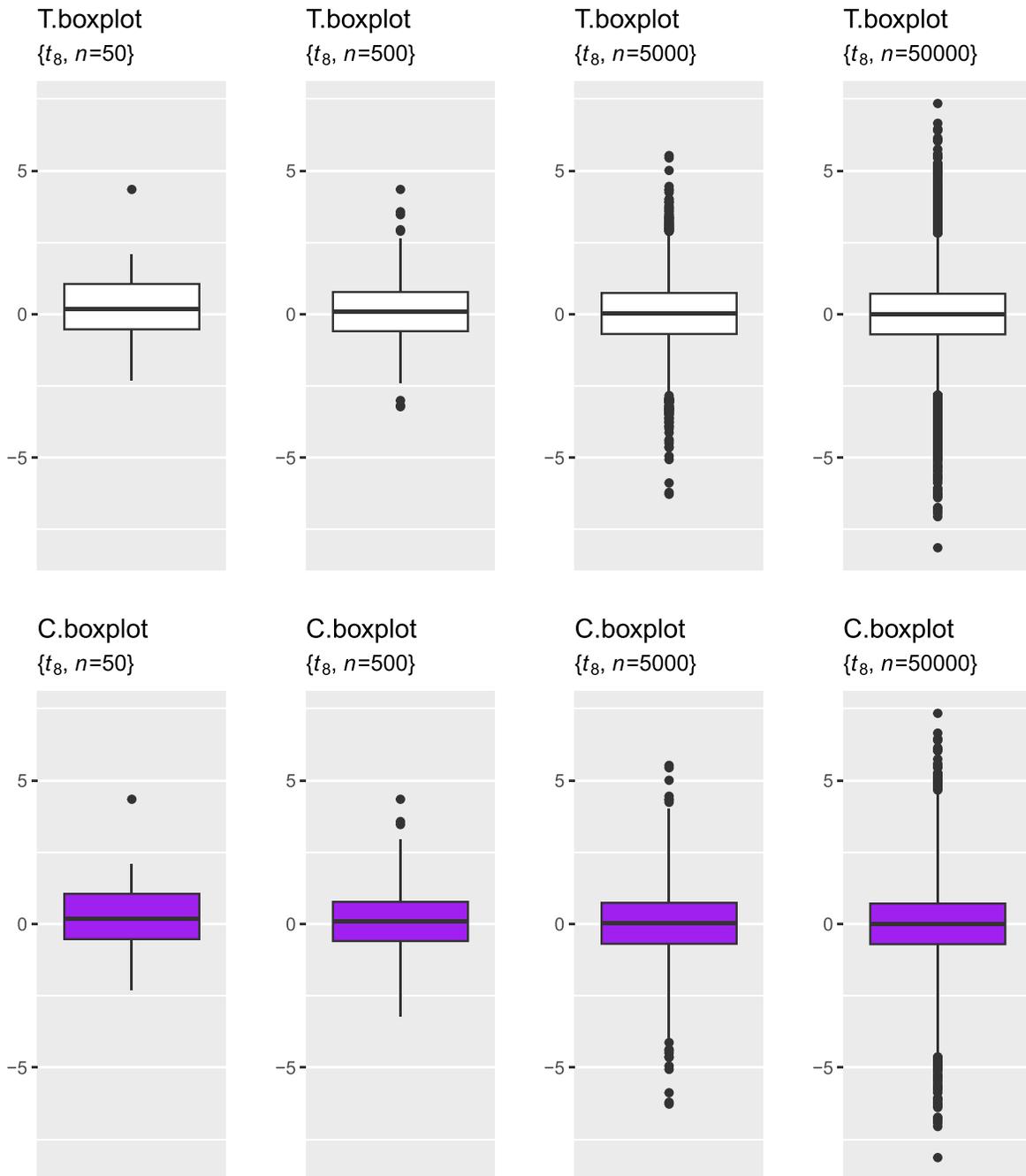
Figure 4. Outlier detection for heavy tailed data using Tukey's boxplot (T.boxplot) and the Chauvenet-type boxplot (C.boxplot), where the data is generated from Student's t distribution with 8 degrees of freedom. With *set.seed(1863)* in R, Tukey's boxplot labels a total of 1, 8, 117, and 1104 suspected outliers, and the Chauvenet-type boxplot labels a total of 1, 3, 18, and 90 suspected outliers, for the simulated data with $n = 50$, 500, 5000 and 50,000, respectively.

our newly developed Chauvenet-type boxplot has the capacity of labeling less "false positive" outliers thanks to the sample size adjusted fences, especially for large sample sizes.

4.3. Real Data

For real data, we consider the civil service pay adjustment in Hong Kong, where the data can be downloaded from *https:// www.csb.gov.hk/english/admin/pay/55.html* for the annual tax year which runs from April 1 to March 31 of the following year. In the 24 tax years since 2001, civil servants in Hong Kong had their pay frozen or cut a total of eight times. In particular, as

a consequence of the SARS outbreak, they suffered from two immediate pay cuts in 2003–2004 and 2004–2005, followed by two pay freezes in 2005–2006 and 2006–2007. Since 2007, the economic situation of Hong Kong has improved significantly, but still encountered two adverse events thereafter. The first was the global financial crisis between mid 2007 and early 2009, which resulted in a pay cut of 5.38% for the upper-level staff and a pay freeze for the lower and mid-level staff in the tax year of 2009–2010. And the second was the economic turmoil following the COVID-19 pandemic, which also prompted a pay freeze for two consecutive years in 2020–2021 and 2021–2022.

As can be seen, the annual rate of pay adjustment serves as an important index for the economic development of Hong Kong. In arriving at this pay adjustment rate, a thorough consideration was often taken into account including six key factors under the established annual civil service pay adjustment mechanism. In this study, using the annual pay adjustment, we are interested in examining whether the existing or new outlier criteria can identify the pay freeze or pay cut, caused by the global financial crisis or the COVID-19 pandemic, as contaminated data. Noting that the pay adjustment for the lower and mid-level staff are always the same, we thus combine these two bands and refer to them as the "junior civil servants" for simplicity. While for the staff in the upper band, we refer to them as the "senior civil servants". Lastly, for outlier detection using the three criteria, we report the pay adjustment rates since 2007 in Table 1 for both the junior and senior groups, each with $n = 18$ annual observations.

For junior civil servants, the rates of pay adjustment range from 0% to 6.16%, together with $Q_1 = 2.61\%$, $Q_3 = 4.70\%$, and $n = 18$. The lower and upper fences of Tukey's boxplot with $k = 1.5$ are LF $= Q_1 - k \times$ IQR $= -0.53\%$ and UF $= Q_3 + k \times$ IQR $= 7.84\%$. Since no rate is outside the interval [LF, UF], Tukey's boxplot fails to detect the 3 pay freezes with rate 0% as contaminated data. Next, to apply the Chauvenet-type boxplot, with $k_n^{\text{Chau}} = \Phi^{-1}(1 - 0.25/n)/1.35 - 0.5 = 1.13$, the lower and upper fences are given as

$$\text{LF}_n^{\text{Chau}} = Q_1 - k_n^{\text{Chau}} \times \text{IQR} = 0.25\%,$$
$$\text{UF}_n^{\text{Chau}} = Q_3 + k_n^{\text{Chau}} \times \text{IQR} = 7.07\%.$$

Noting that the rate 0% is below the lower fence, we therefore claim the 3 pay freezes in 2009–2010, 2020–2021, and 2021–2022 as contaminated data, or suspected outliers, that were caused by either the global financial crisis or the COVID-19 pandemic. One main reason for this success is that, due to the small sample size of $n = 18$, we are able to apply a smaller fence coefficient $k_n^{\text{Chau}} = 1.13$ to construct the outlier region.

Table 1. The annual rates of pay adjustment in Hong Kong for the lower and mid-level staff (Junior civil servants) and the upper-level staff (Senior civil servants) since 2007.

Tax year	Junior civil servants	Senior civil servants
2024–2025	3.00%	3.00%
2023–2024	4.65%	2.87%
2022–2023	2.50%	2.50%
2021–2022	0.00%	0.00%
2020–2021	0.00%	0.00%
2019–2020	5.26%	4.75%
2018–2019	4.51%	4.06%
2017–2018	2.94%	1.88%
2016–2017	4.68%	4.19%
2015–2016	4.62%	3.96%
2014–2015	4.71%	5.96%
2013–2014	3.92%	2.55%
2012–2013	5.80%	5.26%
2011–2012	6.16%	7.24%
2010–2011	0.56%	1.60%
2009–2010	0.00%	−5.38%
2008–2009	5.29%	6.30%
2007–2008	4.62%	4.96%

NOTE: Among the 18 tax years, junior civil servants had suffered a total of 3 pay freezes in 2009–2010, 2020–2021, and 2021–2022. And senior civil servants had suffered a pay cut of 5.38% in 2009–2010, followed by 2 pay freezes in 2020–2021 and 2021–2022.

Finally, it is also interesting to point out that, if Chauvenet's criterion is considered, then $\bar{X} = 3.51\%$, $S = 2.08\%$, and $c_n = \Phi^{-1}(1 - 0.25/n) = 2.20$. Further by (7), the sigma clipping interval is given as $[\bar{X} - c_n S, \bar{X} + c_n S] = [-1.07\%, 8.09\%]$, which also fails to detect 0% as a suspected outlier due to its sensitivity to the included outliers.

For senior civil servants, the rates of pay adjustment range from -5.38% to 7.24%. In addition, we have $Q_1 = 2.04\%$, $Q_3 = 4.91\%$, $\bar{X} = 3.09\%$, $S = 2.92\%$, and $n = 18$. For Tukey's boxplot, the lower and upper fences are LF $= -2.27\%$ and UF $= 9.22\%$. For the Chauvenet-type boxplot, the lower and upper fences are LF$_n^{\text{Chau}} = -1.20\%$ and UF$_n^{\text{Chau}} = 8.15\%$. And for Chauvenet's criterion, the sigma clipping interval is given as $[\bar{X} - c_n S, \bar{X} + c_n S] = [-3.33\%, 9.52\%]$. It is evident that all three criteria can detect the pay cut with rate -5.38% as an outlier of the pay adjustment data, but none of them identifies the 3 pay freezes as suspected outliers. To clarify, with the range from -5.38% to 7.24%, a rate of 0% is indeed close to the mid-range and hence does not look abnormal. In other words, as also seen in Figure 5, the question of whether or not the rate 0% is contaminated has largely been overshadowed by the abnormal rate -5.38%. Lastly, it is worth pointing out that our lower fence LF$_n^{\text{Chau}} = -1.20\%$ is much closer to 0% compared to the lower fence LF $= -2.27\%$ and the lower limit $\bar{X} - c_n S = -3.33\%$. This shows that, among the three criteria, our new boxplot criterion does have the greatest potential to label the pay freeze as unusual pay adjustment for the civil servants in Hong Kong.

5. Chauvenet-Type Boxplot for Nonnormal Data

In this section, we extend the Chauvenet-type boxplot to handle nonnormal data for outlier detection. We first show how to construct the lower and upper fences of the boxplot using the gamma distribution as an example, chosen to illustrate the method on a two-parameter distribution. Let $\{X_1, \ldots, X_n\}$ be a random sample of size n from Gamma(α, β), where α is the shape parameter and β is the scale parameter (Casella and Berger 2002). This leads to the method of moments estimates of the two parameters as

$$\hat{\alpha} = \frac{n\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad \text{and} \quad \hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n\bar{X}}.$$

We then treat Gamma $(\hat{\alpha}, \hat{\beta})$ as the distribution of the data, and let $\hat{F}^{-1}(\cdot)$ represent its quantile function. Note that $\hat{F}^{-1}(\cdot)$ is a generic notation that also applies to other nonnormal distributions, as long as the distribution parameters can be properly estimated from the data.

5.1. A New Boxplot Criterion for Nonnormal Data

Now to construct the new boxplot criterion for nonnormal data, we let the lower and upper fences of the boxplot be LF$_n = Q_1 - k'_n \times$ IQR and UF$_n = Q_3 + k''_n \times$ IQR, where k'_n and k''_n are the fence coefficients. By the same spirit of Chauvenet's criterion, we further set $\hat{F}^{-1}(0.25/n)$ and $\hat{F}^{-1}(1 - 0.25/n)$ as the lower and upper thresholds for outlier detection, respectively. In other words, Chauvenet's criterion suggests to set the lower fence as LF$_n = \hat{F}^{-1}(0.25/n)$ and the upper fence as UF$_n = \hat{F}^{-1}(1 -$

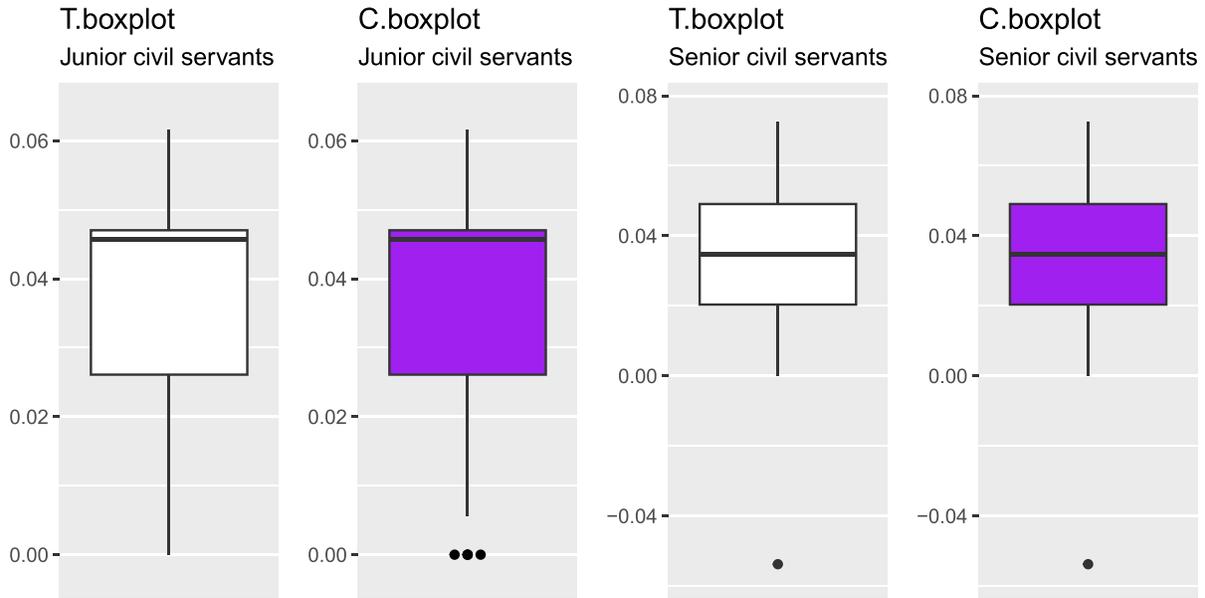| T.boxplot | C.boxplot | T.boxplot | C.boxplot |
| Junior civil servants | Junior civil servants | Senior civil servants | Senior civil servants |

Figure 5. Tukey's boxplot and the Chauvenet-type boxplot of the civil service pay adjustment in Hong Kong since 2007. The left two panels are for junior civil servants, where the three outliers with jitter in the second boxplot represent the 3 pay freezes in 2009–2010, 2020–2021 and 2021–2022. The right two panels are for senior civil servants, with the outlier in both the boxplots represents the pay cut in 2009–2010.

$0.25/n$). By doing so, our new criterion will wrongly reject, on average, a quarter observation from the upper tail and another quarter observation from the lower tail. In addition, noting that $Q_1 \approx \hat{F}^{-1}(0.25)$ and $Q_3 \approx \hat{F}^{-1}(0.75)$, we can derive that

$$k'_n \approx \frac{\hat{F}^{-1}(0.25) - \hat{F}^{-1}(0.25/n)}{\hat{F}^{-1}(0.75) - \hat{F}^{-1}(0.25)} \quad \text{and}$$

$$k''_n \approx \frac{\hat{F}^{-1}(1 - 0.25/n) - \hat{F}^{-1}(0.75)}{\hat{F}^{-1}(0.75) - \hat{F}^{-1}(0.25)}. \quad (12)$$

In the special case when the data is normal, it is easy to verify that $k'_n = k''_n = k_n^{\text{Chau}} = \Phi^{-1}(1 - 0.25/n)/1.35 - 0.5$. Lastly, with the fence coefficients in (12), the lower and upper fences of our new boxplot criterion can be constructed as

$$\text{LF}_n^{\text{Chau}} = Q_1 - \frac{\hat{F}^{-1}(0.25) - \hat{F}^{-1}(0.25/n)}{\hat{F}^{-1}(0.75) - \hat{F}^{-1}(0.25)} \times \text{IQR}, \quad (13)$$

and

$$\text{UF}_n^{\text{Chau}} = Q_3 + \frac{\hat{F}^{-1}(1 - 0.25/n) - \hat{F}^{-1}(0.75)}{\hat{F}^{-1}(0.75) - \hat{F}^{-1}(0.25)} \times \text{IQR}. \quad (14)$$

Lastly, to distinguish it with the Chauvenet-type boxplot in Section 3.2, we refer to the new boxplot with fences (13) and (14) as the *Chauvenet-type boxplot for nonnormal data*.

5.2. Simulation Study

To evaluate the performance of the new boxplot for nonnormal data, we revisit the two distributions in Section 4.2. We first consider the Chi-square distribution with v degrees of freedom. Let $\{X_1, \ldots, X_n\}$ be the sample data of size n, with $\bar{X}$ being the sample mean and Q_1 and Q_3 being the first and third quartiles. Then by the method of moments, we have

$$\hat{v} = \bar{X}.$$

Letting also $\hat{F}^{-1}(\cdot)$ be the quantile function of the $\chi_{\hat{v}}^2$ distribution, we can obtain the fence coefficients k'_n and k''_n in (12), yielding further the outlier region of the Chauvenet-type boxplot as $\left(-\infty, Q_1 - k'_n \times \text{IQR}\right) \cup \left(Q_3 + k''_n \times \text{IQR}, \infty\right)$. Finally, with $v = 8$ and $n = 50,000$, we generate the sample data from the χ_8^2 distribution without contaminated data, and then plot the data in Figure 6 using Tukey's boxplot (T.boxplot), the Chauvenet-type boxplot (C.boxplot), Tukey's boxplot for nonnormal data (T.boxplot.NN), and the Chauvenet-type boxplot for nonnormal data (C.boxplot.NN). More specifically for T.boxplot.NN, it is implemented using the "litteR" package in CRAN as the adjusted boxplot for skewed distributions by Hubert and Vandervieren (2008). Now if we take *set.seed(1863)* in R, the simulated data will range from 0.22 to 43.09, with $Q_1 = 5.08$, $Q_3 = 10.24$ and $\hat{v} = \bar{X} = 8.02$. Further by (12)–(14), it leads to $k'_n = 0.94$, $k''_n = 5.58$, and $[\text{LF}_n^{\text{Chau}}, \text{UF}_n^{\text{Chau}}] = [0.20, 39.02]$. Lastly, observing that there are only two observations (39.66 and 43.09) above the upper fence and no observation below the lower fence, C.boxplot.NN thus only labels two suspected outliers among a total of 50,000 observations. Among the other three boxplots, however, the Chauvenet-type boxplot performs the best since it only labels 106 suspected outliers as already seen in Section 4.2.

For the second simulation, we consider Student's t distribution with v degrees of freedom. Let $\{X_1, \ldots, X_n\}$ be the sample data with S^2 being the sample variance and Q_1 and Q_3 being the first and third quartiles. By equating the sample variance to the population variance, that is $v/(v - 2) = S^2$, we have the estimated degrees of freedom as

$$\hat{v} = 2S^2/(S^2 - 1).$$

Letting also $\hat{F}^{-1}(\cdot)$ be the quantile function of the $t_{\hat{v}}$ distribution, we can thus have the outlier region of the Chauvenet-type boxplot as $\left(-\infty, Q_1 - k'_n \times \text{IQR}\right) \cup \left(Q_3 + k''_n \times \text{IQR}, \infty\right)$. Finally, with $v = 8$ and $n = 50,000$, we generate data from the t_8
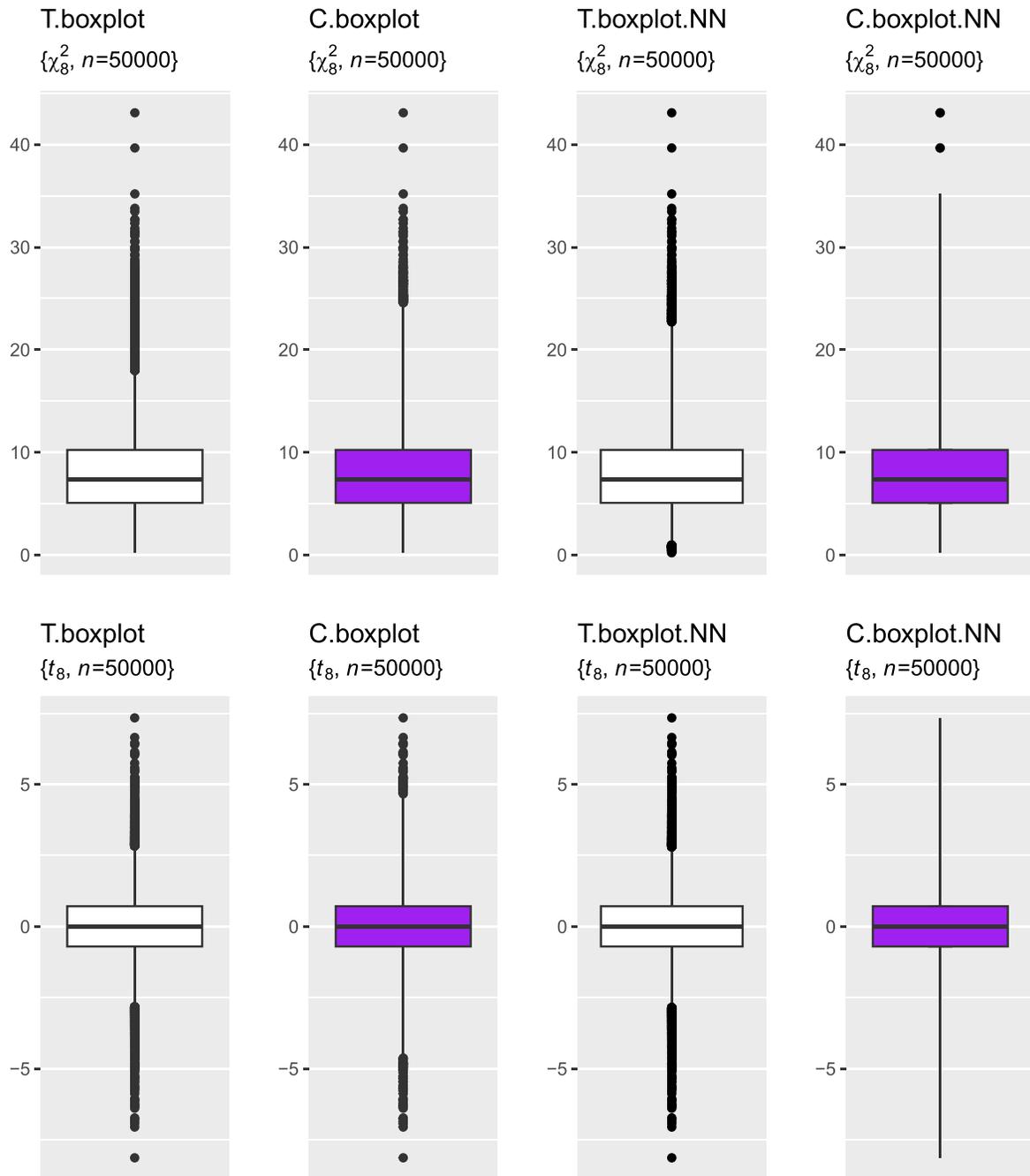
Figure 6. Outlier detection for nonnormal data using Tukey's boxplot (T.boxplot), the Chauvenet-type boxplot (C.boxplot), Tukey's boxplot for nonnormal data (T.boxplot.NN), and the Chauvenet-type boxplot for nonnormal data (C.boxplot.NN), where the data is generated from the χ_8^2 and t_8 distributions, respectively. With $n = 50{,}000$ and *set.seed(1863)* in R, the Chauvenet-type boxplot for nonnormal data labels only 2 and 0 suspected outliers for each distribution. Among the other three boxplots, the Chauvenet-type boxplot labels the least number of observations as outliers.

distribution and also plot the four boxplots in Figure 6 as in the previous simulation. Taking *set.seed(1863)* in R again, the simulated data ranges from -8.13 to 7.34, with $Q_1 = -0.70$, $Q_3 = 0.71$, $\hat{v} = 2S^2/(S^2 - 1) = 8.02$ and $k'_n = k''_n = 6.41$. This leads to $[\text{LF}_n^{\text{Chau}}, \text{UF}_n^{\text{Chau}}] = [-9.77, 9.78]$, thus, explaining why C.boxplot.NN labels no outlier. Lastly, we note that the comparison results among the other three boxplots remain the same. For additional results on the boxplots with $n = 50, 500$ and 5000 from the χ_8^2 and t_8 distributions, please refer to the online supplementary material.

6. Conclusion

Tukey's box-and-whisker plot is one of the most popular methods used for descriptive statistics, but it is also known to be free of sample size, yielding the so-called "one-size-fits-all" fences for outlier detection. The main purpose of this article was to introduce a new boxplot criterion that can more accurately, or more robustly, label the suspected outliers or contaminated data. In contrast to Tukey's boxplot with the whiskers being 1.5 or 3 times of the interquartile range (IQR), our new boxplot tailored

the whiskers' length as $k_n^{\text{Chau}} \times \text{IQR}$, where

$$k_n^{\text{Chau}} = \frac{\Phi^{-1}(1 - 0.25/n)}{1.35} - 0.5$$

is the sample size adjusted fence coefficient determined by Chauvenet's criterion, with n being the sample size and Φ^{-1} being the quantile function of the standard normal distribution. Observing that our new boxplot takes advantage of both Tukey's boxplot and Chauvenet's criterion, we referred to it as the *Chauvenet-type boxplot*, with the lower and upper fences being specified as $\text{LF}_n^{\text{Chau}} = Q_1 - k_n^{\text{Chau}} \times \text{IQR}$ and $\text{UF}_n^{\text{Chau}} = Q_3 + k_n^{\text{Chau}} \times \text{IQR}$ as in (8). In addition, we noted that k_n^{Chau} is an increasing function of n, with $k_n^{\text{Chau}} = 1.5$ when $n = 72$, and $k_n^{\text{Chau}} = 3$ when $n = 217{,}282$. And in particular, when the sample size is less than 72, we even recommended a whisker shorter than $1.5 \times \text{IQR}$ for outlier detection, in a way to more effectively increase the detection power of outliers. With the fence coefficient k_n^{Chau}, the Chauvenet-type boxplot not only provides an exact control of the outside rate per observation, but also maintains the simplicity of the boxplot from a practical perspective. Moreover, from the perspective of Chauvenet's criterion, our new boxplot criterion can also serve as a robust Chauvenet's criterion. Simulation study and a real data analysis on the civil service pay adjustment in Hong Kong also demonstrated that the Chauvenet-type boxplot performs extremely well regardless of the sample size, and can therefore be highly recommended for practical use to replace both Tukey's boxplot and Chauvenet's criterion.

Lastly, we also presented in Section 5 that the Chauvenet-type boxplot can be further extended to handle nonnormal data for outlier detection. By numerical studies with the data generated from the Chi-square and t distributions, the *Chauvenet-type boxplot for nonnormal data* can perfectly do its job, and in particular by providing two different whiskers, our new criterion also effectively prevents to label too many outliers from one side of the skewed data. On the other hand, we noted that the fence coefficients k_n' and k_n'' derived in (12) require the nonnormal distribution to be known, which however may not be realistic in practice. If the underlying distribution of the sample data is unknown, one may need to apply other statistical techniques, for example the nonparametric methods, to estimate the distribution function. As another alternative, one may also apply some existing techniques in the literature to label outliers for nonnormal and in particular for skewed data (Schwertman, Owens, and Adnan 2004; Dümbgen and Riedwyl 2007; Schwertman and de Silva 2007; Hubert and Vandervieren 2008; Bruffaerts, Verardi, and Vermandele 2014; Dovoedo and Chakrabort 2015; Shein and Fitrianto 2017; Walker et al. 2018; Zhao and Yang 2019; Rodu and Kafadar 2022). In particular, we noted that Kimber (1990) applied the semi-interquartile ranges to construct the lower and upper fences, yielding $\text{LF} = Q_1 - k(M - Q_1)$ and $\text{UF} = Q_3 + k(Q_3 - M)$, and Carling (2000) further constructed the lower and upper fences as a form of $\text{LF} = M - k \times \text{IQR}$ and $\text{UF} = M + k \times \text{IQR}$, where M is the sample median and k is the fence coefficient. Further research is warranted in this direction to build more robust Chauvenet-type boxplots.

Supplementary Materials

The online supplementary material gives the proof of Theorem 1 together with a supporting lemma from the literature. It also provides a total of nine additional figures for nonnormal boxplots with other skewed or heavy tailed distributions to supplement the simulation studies in the main text.

Acknowledgments

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

References

Abuzaid, A. H., Mohamed, I. B., and Hussin, A. G. (2012), "Boxplot for Circular Variables," *Stat*, 27, 381–392. [201]

Andrews, D. F. (1974), "A Robust Method for Multiple Linear Regression," *Technometrics*, 16, 523–531. [199]

Andrews, D. F., and Pregibon, D. (1978), "Finding the Outliers that Matter," *Journal of the Royal Statistical Society*, Series B, 44, 1–36. [199]

Atkinson, A. C. (1994), "Fast Very Robust Methods for the Detection of Multiple Outliers," *Journal of the American Statistical Association*, 89, 1329–1339. [199]

Atkinson, A. C., and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, New York: Springer. [199]

Banerjee, S., and Iglewicz, B. (2007), "A Simple Univariate Outlier Identification Procedure Designed for Large Samples," *Communications in Statistics–Simulation and Computation*, 36, 249–263. [199,200]

Barbato, G., Barini, E. M., Genta, G., and Levi, R. (2011), "Features and Performance of Some Outlier Detection Methods," *Journal of Applied Statistics*, 38, 2133–2149. [199,201]

Barbato, G., Genta, G., and Levi, R. (2009), *Outlier Detection*, Paris: CIRP STC P–Precision Engineering and Metrology. [201]

Barnett, V., and Lewis, T. (1994), *Outliers in Statistical Data* (3rd ed.), Chichester: Wiley. [199]

Bates, S., Candès, E., Lei, L., Romano, Y., and Sesia, M. (2023), "Testing for Outliers with Conformal *p*-values," *Annals of Statistics*, 51, 149–178. [199]

Beckman, R. J., and Cook, R. D. (1983), "Outliers..........s," *Technometrics*, 25, 119–149. [199]

Bhattacharya, S., Kamper, F., and Beirlant, J. (2023), "Outlier Detection based on Extreme Value Theory and Applications," *Scandinavian Journal of Statistics*, 50, 1466–1502. [199]

Brant, R. (1990), "Comparing Classical and Resistant Outlier Rules," *Journal of the American Statistical Association*, 85, 1083–1090. [199]

Bruffaerts, C., Verardi, V., and Vermandele, C. (2014), "A Generalized Boxplot for Skewed and Heavy-Tailed Distributions," *Statistics and Probability Letters*, 95, 110–117. [210]

Buttarazzi, D., Pandolfo, G., and Porzio, G. C. (2018), "A Boxplot for Circular Data," *Stat*, 74, 1492–1501. [201]

Carey, V. J., Walters, E. E., Wager, C. G., and Rosner, B. A. (1997), "Resistant and Test-based Outlier Rejection: Effect on Gaussian One- and Two-Sample Inference," *Technometrics*, 39, 320–330. [199]

Carling, K. (2000), "Resistant Outlier Rules and the Non-Gaussian Case," *Computational Statistics and Data Analysis*, 33, 249–258. [210]

Casella, G., and Berger, R. L. (2002), *Statistical Inference*, Pacific Grove: Duxbury. [207]

Chauvenet, W. (1863), *A Manual of Spherical and Practical Astronomy*, Philadelphia: J. B. Lippincott Company. [198,201]

Dai, W., and Genton, M. G. (2018), "Functional Boxplots for Multivariate Curves," *Stat*, 7, e190. [201]

David, H. A., and Nagaraja, H. N. (2003), *Order Statistics* (3rd ed.), New York: Wiley. [200]

Davies, L., and Gather, U. (1993), "The Identification of Multiple Outliers," *Journal of the American Statistical Association*, 88, 782–792. [199]

Dovoedo, Y. H., and Chakrabort, S. (2015), "Boxplot-based Outlier Detection for the Location-Scale Family," *Communications in Statistics-Simulation and Computation*, 44, 1492–1513. [210]

Dümbgen, L., and Riedwyl, H. (2007), "On Fences and Asymmetry in Box-and-Whiskers Plots," *The American Statistician*, 61, 356–359. [210]

Frigge, M., Hoaglin, D. C., and Iglewicz, B. (1989), "Some Implementation of the Boxplot," *The American Statistician*, 43, 50–54. [198,201]

Fung, W. (1993), "Unmasking Outliers and Leverage Points: A Confirmation," *Journal of the American Statistical Association*, 88, 515–519. [199]

Gould, B. A. (1855), "On Peirce's Criterion for the Rejection of Doubtful Observations, with Tables for Facilitating its Application," *Astronomical Journal*, 4, 81–87. [201]

Guenther, W. C. (1977), *Sampling Inspection in Statistical Quality Control*, London: Charles Griffin. [200]

Guttman, I. (1970), *Statistical Tolerance Regions: Classical and Bayesian*, London: Charles Griffin. [200]

Hawkins, D. M. (1980), *Identification of Outliers*, London: Chapman and Hall. [199]

——— (2006), *Outliers. Encyclopedia of Statistical Sciences*, New York: Wiley. [199]

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., and Welch, V. A. (2019), *Cochrane Handbook for Systematic Reviews of Interventions* (2nd ed.), Chichester: Wiley. [202]

Hoaglin, D. C., and Iglewicz, B. (1987), "Fine-Tuning Some Resistant Rules for Outlier Labelling," *Journal of the American Statistical Association*, 82, 1147–1149. [199,201]

Hoaglin, D. C., Iglewicz, B., and Tukey, J. W. (1986), "Performance of Some Resistant Rules for Outlier Labelling," *Journal of the American Statistical Association*, 81, 991–999. [199,201]

Hofmann, H., Wickham, H., and Kafadar, K. (2017), "Letter-Value Plots: Boxplots for Large Data," *Journal of Computational and Graphical Statistics*, 26, 469–477. [201]

Huber, P. J. (2004), *Robust Statistics*, New York: Wiley-Interscience. [202]

Hubert, M., and Vandervieren, E. (2008), "An Adjusted Boxplot for Skewed Distributions," *Computational Statistics and Data Analysis*, 52, 186–5201. [204,208,210]

Hyndman, R. J., and Shang, H. L. (2010), "Rainbow Plots, Bagplots, and Boxplots for Functional Data," *Journal of Computational and Graphical Statistics*, 19, 29–45. [201]

Iglewicz, B., and Banerjee, S. (2001), "A Simple Univariate Outlier Identification Procedure," *Proceedings of the American Statistical Association, Section on Statistical Computing [CD-ROM]* [200]

Kimber, A. C. (1990), "Exploratory Data Analysis for Possibly Censored Data from Skew Distributions," *Applied Statistics*, 39, 21–30. [210]

Liao, C. T., and Iyer, H. K. (2004), "A Tolerance Interval for the Normal Distribution with Several Variance Components," *Communications in Statistics - Theory and Methods*, 14, 217–229. [200]

Maples, M. P., Reichart, D. E., Konz, N. C., Berger, T. A., Trotter, A. S., Martin, J. R., Dutton, D. A., Paggen, M. L., Joyner, R. E., and Salemi, C. P. (2018), "Robust Chauvenet Outlier Rejection," *The Astrophysical Journal Supplement Series*, 238, 2. [199]

McGill, R., Tukey, J. W., and Larsen, W. A. (1978), "Variations of Box Plots," *The American Statistician*, 32, 12–16. [199]

Milton, J. S. (1999), *Statistical Methods in the Biological and Health Sciences*, New York: WCB/McGraw-Hill. [199]

Mirzargar, M., Whitaker, R. T., and Kirby, R. M. (2014), "Curve Boxplot: Generalization of Boxplot for Ensembles of Curves," *IEEE Transactions on Visualization and Computer Graphics*, 20, 2654–2663. [201]

Morales, C., Giraldo, R., and Torres, M. (2021), "Boxplot Fences in Proficiency Testing," *Accreditation and Quality Assurance*, 26, 193–200. [201]

Patel, J. K. (1986), "Tolerance Limits - A Review," *Communications in Statistics - Theory and Methods*, 15, 2719–2762. [200]

Peirce, B. (1852), "Criterion for the Rejection of Doubtful Observations," *Astronomical Journal*, 2, 161–163. [198,201]

——— (1878), "On Peirce's Criterion," *Proceedings of the American Academy of Arts and Sciences*, 13, 348–351. [201]

Pukelsheim, F. (1994), "The Three Sigma Rule," *The American Statistician*, 48, 88–91. [201]

Qu, Z., and Genton, M. G. (2022), "Sparse Functional Boxplots for Multivariate Curves," *Journal of Computational and Graphical Statistics*, 31, 976–989. [201]

Raj, M., Mirzargar, M., Ricci, R., Kirby, R. M., and Whitaker, R. T. (2017), "Path Boxplots: A Method for Characterizing Uncertainty in Path Ensembles on a Graph," *Journal of Computational and Graphical Statistics*, 26, 243–252. [201]

Rodu, J., and Kafadar, K. (2022), "The q-q Boxplot," *Journal of Computational and Graphical Statistics*, 31, 26–39. [198,210]

Rosner, B. (1983), "Percentage Points for a Generalized ESD Many-Outlier Procedure," *Technometrics*, 25, 165–172. [199]

Ross, S. M. (2003), "Peirce's Criterion for the Elimination of Suspect Experimental Data," *Journal of Engineering Technology*, 20, 38–41. [199,201]

Schwertman, N. C., and de Silva, R. (2007), "Identifying Outliers with Sequential Fences," *Computational Statistics and Data Analysis*, 51, 3800–3810. [210]

Schwertman, N. C., Owens, M. A., and Adnan, R. (2004), "A Simple More General Boxplot Method for Identifying Outliers," *Computational Statistics and Data Analysis*, 47, 165–174. [198,210]

Shein, W. H., and Fitrianto, A. (2017), "A Comparative Study of Outliers Identification Methods in Univariate Data Set," *Advanced Science Letters*, 23, 1422–1427. [210]

Sim, C. H., Gan, F. F., and Chang, T. C. (2005), "Outlier Labeling with Boxplot Procedures," *Journal of the American Statistical Association*, 100, 642–652. [198,199,200,203]

Sun, Y., and Genton, M. G. (2011), "Functional Boxplots," *Journal of Computational and Graphical Statistics*, 20, 316–334. [201]

Tukey, J. W. (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley. [198,201,202]

Vermeesch, P. (2018), "IsoplotR: A Free and Open Toolbox for Geochronology," *Geoscience Frontiers*, 9, 1479–1493. [199,201]

Walker, M. L., Dovoedo, Y. H., Chakraborti, S., and Hilton, C. W. (2018), "An Improved Boxplot for Univariate Data," *The American Statistician*, 72, 348–353. [198,210]

Wheeler, D. J., and Chambers, D. S. (2010), *Understanding Statistical Process Control* (3rd ed.), Knoxville: SPC Press. [201]

Whitaker, R. T., Mirzargar, M., and Kirby, R. M. (2013), "Contour Boxplots: A Method for Characterizing Uncertainty in Feature Sets from Simulation Ensembles," *IEEE Transactions on Visualization and Computer Graphics*, 19, 2713–2722. [201]

Wickham, H., and Stryjewski, L. (2011), "40 Years of Boxplots," Technical Report. *https://vita.had.co.nz/papers/boxplots.pdf*. [198]

Zhao, C., and Yang, J. (2019), "A Robust Skewed Boxplot for Detecting Outliers in Rainfall Observations in Real-Time Flood Forecasting," *Advances in Meteorology*, 2019, 1795673. [210]