Journal of Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cjas20

Mode type quasi-range and its applications

Lin-An Chen [a], Hsien-Chueh Peter Yang [b] & Chau-Shyun Tang [c]

[a] Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan

[b] Department of Risk Management and Insurance, National Kaohsiung First University of Science and Technology, Kaohsiung, Taiwan

[c] Department of Management Science, National Chiao Tung University, Hsinchu, Taiwan

Published online: 15 Aug 2008.

Journal of Applied Statistics
Vol. 35, No. 10, October 2008, 1181–1192

Routledge
Taylor & Francis Group

Mode type quasi-range and its applications

Lin-An Chen[a]*, Hsien-Chueh Peter Yang[b] and Chau-Shyun Tang[c]

[a] *Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan;* [b] *Department of Risk Management and Insurance, National Kaohsiung First University of Science and Technology, Kaohsiung, Taiwan;* [c] *Department of Management Science, National Chiao Tung University, Hsinchu, Taiwan*

Building from the consideration of closeness, we propose the mode quasi-range as an alternative scale parameter. Application of this scale parameter to formulate the population standard deviation is investigated leading to an efficient sample estimator of standard deviation from the point of asymptotic variance. Monte Carlo studies, in terms of finite sample efficiency and robustness of breakdown point, have been performed for the sample mode quasi-range. This study reveals that this closeness consideration-based mode, quasi-range, is satisfactory because these statistical procedures based on it are efficient and are less misleading for drawing conclusion from the sample results.

Keywords: breakdown point; range; robustness; quasi-range; scale parameter

1. Introduction

Measuring the center and variability of a distribution for the random variable are two very important topics in statistical inferences. Basically variability indicates how closely and widely the values of variable X are spreading out. In statistics, it is often as important to study the variability as the center of a random variable. For example, when products that fit together (such as pipes) are manufactured, it is important to minimize the variations of the diameters of the products so they will fit together properly.

Although there are numerous techniques proposed in the literature for measuring variability, very few of them are designed with a clear explanation for their roles in individually measuring either the wideness or the closeness for values of X being spreading out. One exceptional case is that the variance can be explained as a measure of maximum dispersion from the mean. For most other variability measuring techniques, it is not possible to classify their roles for measuring either wideness or closeness. Let us consider an example for interpretation. Suppose a class of students has taken an examination for a course and we obtain two interval estimates (10,85) and (65,95) that predict, respectively, two population intervals which cover the students' scores with the same probability 0.9. Then two quasi-ranges, 75 and 28, are both estimates of variability, one measuring the interval spreading out most widely, i.e. the longest population interval, and the other measuring the interval spreading out most closely, i.e. the shortest population interval, both with

*Corresponding author. Email: lachen@stat.netu.edu.tw

ISSN 0266-4763 print/ISSN 1360-0532 online

DOI: 10.1080/02664760802271082
http://www.informaworld.com

the same coverage probability. Classifying a measure of variability for its role in interpretation of either wideness or closeness does make sense depending on the application. Our interest in this study is the interval concerning the side of closeness.

Traditionally there are many applications relying on the statistical inferences for a coverage interval (quantile interval), an interval with fixed proportion of underlying distribution such as $(F_\theta^{-1}(\alpha/2), F_\theta^{-1}(1-\alpha/2))$. Three examples are: its point estimator, which has been called a reference interval in laboratory chemistry; Shewhart's control chart in engineering quality control (see [5]); and naive prediction interval (see [4]) in reliability theory. The confidence interval is implicitly used as the tolerance interval (see [8]) in engineering statistics. Although the idea of treating a coverage interval as an interval-like parameter and making inferences for it has popularly been used in several statistics-related disciplines, their treatment with a general theory of estimation and hypothesis testing has received only little attention. One exception is the study by Huang *et al.* [2].

Considering the class of widths for the coverage intervals having a fixed coverage probability, we define the minimum width as the mode quasi-range. We then show that this width is appropriate to act as a scale parameter for measuring the closeness of the distribution. Several studies for this scale parameter are conducted. First, we formulate a representation of the population standard deviation based on this quasi-range and study its sample version in terms of asymptotic variances. Second, a non-parametric simulation analysis of this quasi-range in terms of mean squares error and breakdown point is also conducted. These studies of the new quasi-range indicate that the classical quasi-range-based techniques are not efficient and robust, so the statistical inferences based on them may be misleading.

2. Mode quasi-range

Traditionally we say that τ_0, a non-negative function of a random variable X and percentage γ, $0 < \gamma < 1$, is a measure of dispersion if it satisfies: (a) $\tau_0(X + b, \gamma) = \tau_0(X, \gamma)$ for $b \in R$, and (b) $\tau_0(aX, \gamma) = |a|\tau_0(X, \gamma)$ for $a \in R$. Intuitively the members in the following width family of γ coverage intervals $(F^{-1}(\alpha), F^{-1}(\gamma + \alpha))$,

$$\{F^{-1}(\gamma + \alpha) - F^{-1}(\alpha) \colon 0 < \alpha < 1 - \gamma\},$$

may serve as the quasi-range for the distribution function F. However not every member of the family satisfies the requirements for a measure of dispersion. It is well known that the symmetric quantile difference $\tau_{\text{med}}(1 - \alpha) = F^{-1}(1 - \alpha/2) - F^{-1}(\alpha/2)$, $0 < \alpha < 1$, is a measure of dispersion (see proof in [7]). We call τ_{med} the median quasi-range because the symmetric interval $(F^{-1}(\alpha/2), F^{-1}(1 - \alpha/2))$ shrinks to the median $F^{-1}(0.5)$ when α increases to 1. We are interesting in a measure of dispersion that is a quantile combination solving a minimization problem as

$$\tau_0(X, \gamma) = \inf_{0<\alpha<1-\gamma} \{cF^{-1}(\alpha) + dF^{-1}(\gamma + \alpha)\} \tag{1}$$

with $d > 0$, $c \in R$.

The following theorem provides the condition that the minimization quantile combination in Equation (1) is a measure of dispersion.

THEOREM 2.1 *For given $c \in R$, $d > 0$, τ_0 of Equation* (1) *is a measure of dispersion if $c = -d$.*

Proof For convenience in this proof, we re-denote $F^{-1}(\alpha)$ for random variable X by $F^{-1}(X, \alpha)$. We know that the population quantile F^{-1} satisfies $F^{-1}(X + b, \alpha) = F^{-1}(X, \alpha) + b$ for $b \in R$

and $F^{-1}(aX, \alpha) = aF^{-1}(X, \alpha)$ if $a > 0$ and $aF^{-1}(X, 1-\alpha)$ if $a \leq 0$. Now,

$$\begin{aligned}\tau_0(X+b, \gamma) &= \inf_{0<\alpha<1-\gamma}\{cF^{-1}(X+b, \alpha) + dF^{-1}(X+b, \gamma+\alpha)\}\\ &= \inf_{0<\alpha<1-\gamma}\{cF^{-1}(X, \alpha) + dF^{-1}(X, \gamma+\alpha) + (c+d)b\}\\ &= \tau_0(X, \gamma) + (x+d)b, \quad \text{for } b \in R,\end{aligned}$$

which is equal, for satisfying condition (a), to $\tau_0(X, \gamma)$ only if $c + d = 0$. Then, $c = -d$. It is obvious that (b) holds for $a > 0$. To prove (b) for $a \leq 0$, we say $d = 1$.

$$\begin{aligned}\tau_0(aX, \gamma) &= \inf_{0<\alpha<1-\gamma}\{F^{-1}(aX, \gamma+\alpha) - F^{-1}(aX, \alpha)\}\\ &= \inf_{0<\alpha<1-\gamma}\{aF^{-1}(X, 1-(\gamma+\alpha)) - aF^{-1}(X, 1-\alpha)\}\\ &= \inf_{\gamma<\beta<1}\{aF^{-1}(X, \beta-\gamma) - aF^{-1}(X, \beta)\}\\ &= -a\inf_{\gamma<\beta<1}\{F^{-1}(X, \beta) - F^{-1}(X, \beta-\gamma)\}\\ &= |a|\inf_{0<\beta-\gamma<1-\gamma}\{F^{-1}(X, \gamma+(\beta-\gamma)) - F^{-1}(X, \beta-\gamma)\}\\ &= |a|\tau_0(aX, \gamma),\end{aligned}$$

which finishes condition (b). ■

Since $\tau_0(X, \gamma)$ is a measure of dispersion only if it is of the form $d\inf_{0<\alpha<1-\gamma}\{F^{-1}(\gamma+\alpha) - F^{-1}(\alpha)\}$, we may let $d = 1$. In this situation, the quantile difference in Equation (1) turns out to be the width of the interval $C(\gamma) = (F^{-1}(\alpha^*), F^{-1}(\alpha^*+\gamma))$ with $\alpha^* = \arg_\alpha \inf_{0<\alpha<1-\gamma}\{F^{-1}(\gamma+\alpha) - F^{-1}(\alpha))\}$. We call $C(\gamma)$ the mode interval because it may been shown to converge to the distribution mode as γ converges to zero. We then denote

$$\tau_{\text{mod}}(\gamma) = \inf_{0<\alpha<1-\gamma}\{F^{-1}(\gamma+\alpha) - F^{-1}(\alpha)\}. \tag{2}$$

By setting $\alpha^* = \arg_\alpha \inf_{0<\alpha<1-\gamma}\{F^{-1}(\gamma+\alpha) - F^{-1}(\alpha)\}$, we have $\tau_{\text{mod}}(\gamma) = F^{-1}(\gamma+\alpha^*) - F^{-1}(\alpha^*)$. We now are ready to define this new quasi-range.

Definition 2.2 *For* $0 < \gamma < 1$, *we call quantity* $\tau_{\text{mod}}(\gamma)$ *of Equation* (2) *the mode quasi-range.*

We next consider that the scale parameter τ_{mod} can interpret distribution closeness in some sense. This is stated in the following theorem, which is a direct result from the mode interval $C(\gamma)$ setting.

Theorem 2.3 *Suppose that F is the distribution function of a continuous distribution with density function f and a unique mode. Then, with* γ*-level mode interval* $C(\gamma)$, *we have*

$$f(x) \geq f(x') \quad \textit{for all } x \in C(\gamma),\ \ x' \notin C(\gamma).$$

When X has a continuous and unimodel distribution, the mode quasi-range represents the width of most closest coverage interval since its corresponding coverage interval collects sample points from the sample space with relatively higher densities.

One situation that the mode and median quasi-ranges are identical is stated in the following theorem.

THEOREM 2.4 *When the distribution function F has a symmetric continuous probability density function f that is unimodal, then* $\tau_{\mathrm{med}}(\gamma) = \tau_{\mathrm{mod}}(\gamma)$ *for* $0 < \gamma < 1$.

Proof From Equation (2), $0 = (\partial/\partial\alpha)(F^{-1}(\gamma+\alpha) - F^{-1}(\alpha)) = (1/f(F^{-1}(\gamma+\alpha))) - (1/f(F^{-1}(\alpha)))$. The symetric property and unimodal of f implies $\gamma + \alpha = 1 - \alpha$, which results in $\alpha = (1 - \gamma/2)$. Then $\tau_{\mathrm{mod}}(\gamma) = F^{-1}(\gamma + 1 - \gamma/2) - F^{-1}(1 - \gamma/2) = F^{-1}(1 + \gamma/2) - F^{-1}(1 - \gamma/2) = \tau_{\mathrm{med}}(\gamma)$. ■

We show that continuous type location scale distributions will have simpler form of mode quasi-range that helps to make statistical inferences.

THEOREM 2.5 *The family of continuous location scale distributions with probability density function of the form* $f(x, \theta_1, \theta_2) = 1/\theta_2 f_0((x - \theta_1)/\theta_2)$ *with parameter space* $\theta_1 \in R$ *and* $\theta_2 > 0$ *has*

$$\tau_{\mathrm{mod}}(\gamma) = \theta_2 \inf_{0<\alpha<1-\gamma} (F_0^{-1}(\alpha+\gamma) - F_0^{-1}(\alpha)), \tag{3}$$

where F_0 *is the distribution function of probability density function* f_0.

Proof The result follows from the fact that $F^{-1}(\alpha) = \theta_1 + \theta_2 F_0^{-1}(\alpha)$. ■

Suppose now that we have a random sample $X_1, \ldots, X_n$ drawn from a distribution with probability density function $f(x, \theta)$ that follows a location scale distribution family. With the fact that F_0 in Equation (3) is free of parameters θ_1 and θ_2, the mode quasi-range is simply a linear function of parameter θ_2. Then the statistical inferences for the mode quasi-range τ_{mod} are straightforward if the statistical inference techniques for scale parameter θ_2 are available. This means that when there is a statistical procedure for θ_2 with some desirable property such as unbiasedness, uniformly minimum variance unbiased estimation or most powerful, etc.; then, there is a procedure for τ_{mod} with the same desirable property. We then would not extend this any further and turn to study some properties, efficiencies, and robustness of mode quasi-range procedure that involves non-parametric estimation techniques.

3. Application of mode quasi-range to the representation of standard deviation

We often face situations where we need to formulate the population standard deviation of a random variable in terms of its quasi-range. Let σ_F be the standard deviation of a random variable X that has a distribution function F with probability density function $f(x, \theta)$ where the functional form of f is known but it involves the unknown parameter θ. The classical quasi-range-based formulation of σ_F is

$$\sigma_{\mathrm{med}} = d_{\mathrm{med}} \tau_{\mathrm{med}}(1 - \alpha), \tag{4}$$

where d_{med} is a constant chosen to meet the requirement that σ_{med} is identical with σ_F. In the case that X is a normal random variable, then constant d_{med} is equal to $(2z_{\alpha/2})^{-1}$. With the formulation of σ_F in Equation (4), the classical sample estimator of σ_{med} replaces τ_{med} by the difference of empirical quantiles, i.e.

$$\hat{\sigma}_{\mathrm{med}} = d_{\mathrm{med}} \left(F_n^{-1}\left(1 - \frac{\alpha}{2}\right) - F_n^{-1}\left(\frac{\alpha}{2}\right)\right), \tag{5}$$

where F_n is the empirical distribution function.

The general interest for the median quasi-range formulation of σ_F is searching for $\alpha = \alpha_{\text{med}}$ that maximizes the asymptotic variance $\text{Var}[\hat{\sigma}_{\text{med}}]$. We then call

$$\hat{\sigma}_{\text{med}} = d_{\text{med}}(F_n^{-1}(1-\alpha_{\text{med}}/2) - F_n^{-1}(\alpha_{\text{med}}/2)) \quad (6)$$

the median quasi-range estimator. When F is the normal distribution $N(\mu, \sigma^2)$, it has been noted that when $\alpha = \alpha_{\text{med}} = 0.14$ the asymptotic variance $\text{Var}[\hat{\sigma}_{\text{med}}]$ achieves a minimum (see [7]).

Next, consider a generalization of the quasi-range-based formulation of standard deviation σ_F and we fix the coverage probability as $\gamma \in (0, 1)$. For a given $\alpha \in (0, 1-\gamma)$, there is a constant d_F such that

$$d_F(F^{-1}(\gamma+\alpha) - F^{-1}(\alpha)) \quad (7)$$

is also identical to σ_F. Then Equation (7) provides a general quasi-range formulation of standard deviation σ_F. We then need to choose an appropriate quasi-range formulation of σ_F. For this, we propose a version based on mode quasi-range.

Definition 3.1 *Let X be a random vairable with distribution function F having a standard deviation σ_F. Let d_{mod} be a constant such that $\sigma_{\text{mod}} = d_{\text{mod}}\tau_{\text{mod}}$ is identical with σ_F. We that call σ_{mod} a mode quasi-range formulation of σ_F.*

Before making statistical analysis of the sample mode quasi-range formulation of σ_F, we introduce the median and mode quasi-ranges for the comparison of several distributions.

Mode quasi-range representation for some distributions. We say that X has a distribution that is either exponential distribution $\text{Exp}(\theta)$, of gamma distribution $\text{Gamma}(k/2, \theta)$ or uniform distribution $U(0, \theta)$, if it has probability density function either $f(x, \theta) = (1/\theta)e^{-x/\theta}, x > 0$, or $f(x, (k/2), \theta) = (1/\Gamma(k/2)\theta^{k/2})x^{k/2-1}e^{-x/\theta}, x > 0$ or $f(x, \theta) = (1/\theta), 0 < x < \theta$, respectively. The standard deviations for $\text{Exp}(\theta)$, $\text{Gamma}(k/2, \theta)$, and $U(0, \theta)$ are θ, $\sqrt{k/2}\theta$, and $\theta/\sqrt{12}$, respectively. Their corresonding mode and median quasi-range formulations of σ_F are listed in Tables 1 and 2, respectively.

With a mode range formulation of the standard deviation σ_F, we may consistently apply empirical distribution function F_n to estimate F that induces an estimator of σ_F as

$$\hat{\sigma}_{\text{mod}} = d_{\text{mod}}(F_n^{-1}(\gamma+\alpha^*) - F_n^{-1}(\alpha^*)). \quad (8)$$

Table 1. Mode quasi-range formulation of σ_F.

Distribution	$\tau_{\text{mod}} = F^{-1}(\alpha^*+\gamma) - F^{-1}(\alpha^*)$	σ_{mod}
$\text{Exp}(\theta)$	$-\theta \ln(1-\gamma)$	$\dfrac{F^{-1}(\gamma)}{-\ln(1-\gamma)}$
$\text{Gamma}\left(\dfrac{k}{2}, \theta\right)$	$\dfrac{\theta}{2}(G^{-1}(\gamma+\alpha^*) - G^{-1}(\alpha^*))$	$\dfrac{\sqrt{2k}(F^{-1}(\gamma+\alpha^*) - F^{-1}(\alpha^*))}{G^{-1}(\gamma+\alpha^*) - G^{-1}(\alpha^*)}$
$U(0, \theta)$	$\theta\gamma$	$\dfrac{F^{-1}(\gamma+\alpha^*) - F^{-1}(\alpha^*)}{\gamma\sqrt{12}}$

Table 2. Median quasi-range formulation of σ_F.

Distribution	$\tau_{\mathrm{med}} = F^{-1}(1-\alpha/2) - F^{-1}(\alpha/2)$	σ_{med}
$\mathrm{Exp}(\theta)$	$\theta \ln \dfrac{2-\alpha}{\alpha}$	$\dfrac{F^{-1}(1-\alpha/2) - F^{-1}(\alpha/2)}{\ln \dfrac{2-\alpha}{\alpha}}$
$\mathrm{Gamma}\left(\dfrac{k}{2}, \theta\right)$	$\dfrac{\theta}{2}(G^{-1}(1-\alpha/2) - G^{-1}(\alpha/2))$	$\dfrac{\sqrt{2k}(F^{-1}(1-\alpha/2) - F^{-1}(\alpha/2))}{G^{-1}(1-\alpha/2) - G^{-1}(\alpha/2)}$
$U(0,\theta)$	$\theta(1-\alpha)$	$\dfrac{F^{-1}(1-\alpha/2) - F^{-1}(\alpha/2)}{\sqrt{12}(1-\alpha)}$

We call $\hat{\sigma}_{\mathrm{mod}}$ the mode quasi-range estimator. These are two issues related to the quasi-range estimators to be addressed which are as follows.

(a) There is no need to study σ_{mod} when the distribution is symmetric since, in this situation, the median and mode quasi-ranges are identical. Then we need to search $\binom{\gamma_{\mathrm{mod}}}{\alpha_{\mathrm{mod}}}$ to maximize the asymptotic variance $\mathrm{Var}(\hat{\sigma}_{\mathrm{mod}})$ when X has asymmetric distribution.
(b) We also should compare asymptotic variances for median and mode quasi-range estimators.

With median and mode quasi-range estimators, $\hat{\sigma}_{\mathrm{med}}$ and $\hat{\sigma}_{\mathrm{mod}}$, both for estimation of the same parameter σ_F, to answer question (b) it is appropriate to compare the asymptotic variances of $n^{1/2}(\hat{\sigma}_{\mathrm{mod}} - \sigma_F)$ and $n^{1/2}(\hat{\sigma}_{\mathrm{med}} - \sigma_F)$. We will do this for X that follows the gamma distribution $\mathrm{Gamma}\big((k/2), \theta\big)$ as an example. Its population quantile has the representation $F^{-1}(\alpha) = (\theta/2)G^{-1}(\alpha)$ where G is the distribution function of $\chi^2(k)$. The mode and median type formulations of $\sigma_F = \sqrt{(k/2)}\theta$ are with $d_{\mathrm{mod}} = (\sqrt{2k}/G^{-1}(\gamma+\alpha^*) - G^{-1}(\alpha^*))$ and $d_{\mathrm{med}} = (\sqrt{2k}/G^{-1}(1-\alpha/2) - G^{-1}(\alpha/2))$, respectively. From the large sample theory of empirical quantile F_n^{-1} (see, e.g. [6]), we have representations

$$\sqrt{n}(\hat{\sigma}_{\mathrm{mod}} - \sigma_F) = d_{\mathrm{mod}}(\gamma)[f^{-1}(F^{-1}(\gamma+\alpha^*))n^{-1/2}\sum_{i=1}^{n}(\gamma+\alpha^* - I(X_i \le F^{-1}(\gamma+\alpha^*))) \\ - f^{-1}(F^{-1}(\alpha^*))n^{-1/2}\sum_{i=1}^{n}(\alpha^* - I(X_i \le F^{-1}(\alpha^*)))] + o_p(1)$$

and

$$\sqrt{n}(\hat{\sigma}_{\mathrm{med}} - \sigma_F) \\ = d_{\mathrm{med}}(1-\alpha)\left[f^{-1}\left(F^{-1}\left(\frac{1-\alpha}{2}\right)\right)n^{-1/2}\sum_{i=1}^{n}\left(\frac{1-\alpha}{2} - I\left(X_i \le F^{-1}\left(\frac{1-\alpha}{2}\right)\right)\right) \right. \\ \left. - f^{-1}\left(F^{-1}\left(\frac{\alpha}{2}\right)\right)n^{-1/2}\sum_{i=1}^{n}\left(\frac{\alpha}{2} - I\left(X_i \le F^{-1}\left(\frac{\alpha}{2}\right)\right)\right)\right] + o_p(1).$$

Then the asymptotic variances of $\sqrt{n}(\hat{\sigma}_{\text{mod}} - \sigma_F)$, denoted by $V_{\text{mod}}(\gamma)$, and $\sqrt{n}(\hat{\sigma}_{\text{med}} - \sigma_F)$, denoted by $V_{\text{med}}(1-\alpha)$, are

$$V_{\text{mod}}(\gamma) = \frac{k\theta^2}{2(G^{-1}(\gamma+\alpha^*) - G^{-1}(\alpha^*))^2}[(\gamma+\alpha^*)(1-(\gamma+\alpha^*))g^{-2}(G^{-1}(\gamma+\alpha^*)) + \alpha^*(1-\alpha^*)g^{-2}(G^{-1}(\alpha^*)) - 2\alpha^*(1-(\gamma+\alpha^*))g^{-1}(G^{-1}(\alpha^*))g^{-1}(G^{-1}(\gamma+\alpha^*))]$$

and

$$V_{\text{med}}(1-\alpha) = \frac{k\theta^2}{2(G^{-1}(1-\alpha/2) - G^{-1}(\alpha/2))^2}\left[\left(1-\frac{\alpha}{2}\right)\frac{\alpha}{2}\left(g^{-2}\left(G^{-1}\left(1-\frac{\alpha}{2}\right)\right) - g^{-2}\left(G^{-1}\left(\frac{\alpha}{2}\right)\right)\right) - \frac{\alpha^2}{2}g^{-1}\left(G^{-1}\left(\frac{1-\alpha}{2}\right)\right)g^{-1}\left(G^{-1}\left(\frac{\alpha}{2}\right)\right)\right].$$

Table 3. Representations of mode and median quasi-range formulations and their asymptotic variances.

k	$\alpha_{\text{mod}}\gamma_{\text{mod}}$	V_{mod}	$1-\alpha_{\text{med}}$	V_{med}
2	$\begin{pmatrix}0.0000\\0.7970\end{pmatrix}$	1.5441	0.7432	1.8105
3	$\begin{pmatrix}0.0029\\0.7870\end{pmatrix}$	1.6104	0.7638	2.1620
4	$\begin{pmatrix}0.0111\\0.8110\end{pmatrix}$	1.8794	0.7794	2.5365
5	$\begin{pmatrix}0.0164\\0.8310\end{pmatrix}$	2.2150	0.7908	2.9181
6	$\begin{pmatrix}0.0212\\0.8370\end{pmatrix}$	2.5722	0.7996	3.3023
7	$\begin{pmatrix}0.0257\\0.8370\end{pmatrix}$	2.9392	0.8062	3.6875
8	$\begin{pmatrix}0.0270\\0.8470\end{pmatrix}$	3.3116	0.8116	4.0731
9	$\begin{pmatrix}0.0299\\0.8470\end{pmatrix}$	3.6863	0.8160	4.4587
10	$\begin{pmatrix}0.0330\\0.8450\end{pmatrix}$	4.0639	0.8198	4.8442
11	$\begin{pmatrix}0.0345\\0.8470\end{pmatrix}$	4.4421	0.8230	5.2296
12	$\begin{pmatrix}0.0364\\0.8470\end{pmatrix}$	4.8218	0.8256	5.6148
13	$\begin{pmatrix}0.0374\\0.8490\end{pmatrix}$	5.2018	0.8280	6.0000
14	$\begin{pmatrix}0.0376\\0.8530\end{pmatrix}$	5.5820	0.8300	6.3850
15	$\begin{pmatrix}0.0382\\0.8550\end{pmatrix}$	5.9636	0.8318	6.7698

By letting $\theta = 1$ for simplication, we compute $V_{\text{mod}} = \min_{0<\gamma<1} V_{\text{mod}}(\gamma)$ and $V_{\text{med}} = \min_{0<\alpha<1} V_{\text{med}}(1-\alpha)$. We also let $\binom{\alpha_{\text{mod}}}{\gamma_{\text{mod}}}$ and α_{med} be the values of $\binom{\alpha^*}{\gamma}$ and α, respectively, corresponding with V_{mod} and V_{med}. We list these values for some gamma distributions in Table 3.

We can draw several conclusions from the results in Table 3 as follows.

(a) The asymptotic variances of the median and mode quasi-range estimators are both increasing in k. This reveals that both quasi-range estimators perform similarly as the variance of a gamma distribution.
(b) The asymptotic variances of the mode quasi-range estimator $\hat{\sigma}_{\text{mod}}$ are smaller than those of the median quasi-range estimator $\hat{\sigma}_{\text{med}}$. This shows that the mode quasi-range τ_{mod} is satisfactory for constructing qausi-range formulation of standard deviation σ_F.
(c) When a gamma distribution serves an underlying distribution where constant k is known, this table provides the optimal choice of percentage pair $\binom{\gamma}{\alpha^*}$ and α for the mode and median quasi-range formulations of σ_F. This is only an example and extension of this result to many other distributions of interest is straightforward.

4. Non-parametric Estimation of Mode Quasi-Range

We now consider a non-parametric estimation technique for an unknown mode quasi-range. Parametric methods of data analysis rely on distributional assumptions that presumably gave rise to the observed data. Non-parametric methods however, are fully data-driven and hence are particularly suited for less understood random experiments of high complexity.

Let $X_1, \ldots, X_n$ be the order statistics of a random sample of size n drawn from a distribution F. By letting $h = [n\gamma] + 1$, we define the mode quasi-range estimator as the minimum width of the h consecutive samples as

$$\hat{\tau}_{\text{mod}} = \arg_{h,h+1,\ldots,n} \min\{X_{(h)} - X_{(1)}, X_{(h+1)} - X_{(2)}, \ldots, X_{(n)} - X_{(n-h+1)}.$$

For comparison, we also define the median quasi-range estimator as

$$\hat{\tau}_{\text{med}} = X_{(n(1+\gamma)/2)} - X_{(n(1-\gamma)/2)}.$$

Having introduced the mode quasi-range τ_{mod} and as an alternative for the traditional median quasi-range τ_{med}, we now examine their non-parametric estimators $\hat{\tau}_{\text{med}}$ and $\hat{\tau}_{\text{mod}}$ in two aspects. First, the aim for using a quasi-range is essentially for robustness consideration. It is useful to see if the mode quasi-range estimator $\hat{\tau}_{\text{mod}}$ is more efficient than the median quasi-range estimator $\hat{\tau}_{\text{med}}$. Second, most traditional statistical methods are efficient when the underlying distribution is symmetric. It is then useful to see the results of these two quasi-range estimators for the sample drawn from asymmetric distributions.

We answer these two questions through Monte Carlo studies, using replications $m = 10{,}000$. In the first study, we choose sample sizes $n = 50$ and 100, and for number ith replication, we compute two quasi-range estimates $\hat{\tau}^i_{\text{med}}$ and $\hat{\tau}^i_{\text{mod}}$.

Then we define the mean square errors (MSE),

$$\text{MSE}_{\text{med}} = \frac{1}{m}\sum_{i=1}^{m}(\hat{\tau}^i_{\text{med}} - \tau_{\text{med}})^2 \quad \text{and} \quad \text{MSE}_{\text{mod}} = \frac{1}{m}\sum_{i=1}^{m}(\hat{\tau}^i_{\text{mod}} - \tau_{\text{mod}})^2,$$

where τ_{med} and τ_{mod} are true quasi-ranges computed from the underlying distribution.

Table 4. Mean square errors (MSEs) of the median and mode quasi-range estimators under the exponential distribution.

	$n = 50$		$n = 100$	
γ	MSE_{mod}	MSE_{med}	MSE_{mod}	MSE_{med}
0.5	0.0208	0.0548	0.0096	0.0272
0.6	0.0321	0.0839	0.0149	0.0373
0.7	0.0469	0.1031	0.0234	0.0508
0.8	0.1010	0.1579	0.0433	0.0838
0.9	0.1503	0.2522	0.0819	0.1547
0.95	0.3507	0.5979	0.1503	0.3085

Table 5. Efficiencies (Eff) of two quasi-range estimators for chi-square distribution.

	$k = 3$		$k = 10$	
γ	Eff_{mod}	Eff_{med}	Eff_{mod}	Eff_{med}
0.6	1	0.2702	1	0.4445
0.7	1	0.3387	1	0.5881
0.8	1	0.3821	1	0.6318
0.9	1	0.4910	1	0.7181
0.95	1	0.5449	1	0.6661

We then compare the MSEs of these two quasi-range estimators when the underlying distribution is the exponential distribution with probability density function

$$f(x) = e^{-x}, \quad x > 0.$$

The results of the MSEs are displayed in Table 4.

It is interesting that the mode quasi-range estimator has MSEs all less than the corresponding values of the median quasi-range estimator.

In the second study, we compute the efficiencies (Eff) of quasi-range estimators defined by

$$\text{Eff}_{\text{med}} = \frac{\min\{\text{MSE}_{\text{med}}, \text{MSE}_{\text{mod}}\}}{\text{MSE}_{\text{med}}} \quad \text{and} \quad \text{Eff}_{\text{mod}} \frac{\min\{\text{MSE}_{\text{med}}, \text{MSE}_{\text{mod}}\}}{\text{MSE}_{\text{mod}}},$$

where the underlying distribution is $\chi^2(k)$ with sample size $n = 50$. The simulation results are displayed in Table 5.

Table 5 provides the relative efficiencies of these two quasi-range estimators. The median quasi-range estimator could have efficiencies as small as 0.27 and all less than 0.72. This further supports the use of the mode quasi-range estimator when there is a chi-square distribution.

5. Breakdown point analysis for mode quasi-range

The classical statistical techniques are designed to be the best possible when stringent assumptions apply. However, experience and further research have shown that classical techniques can behave badly when the practical situation departs from the ideal described by such assumptions. The more recently developed robust and exploratory methods are broadening the effectiveness of statistical analyses. One aspect to evaluate the effectiveness is to compare the breakdown points for the estimators.

The breakdown points for an estimator, loosely speaking, is the largest proportion of gross errors that can never carry the estimator over all bounds. We design a simulation study to evaluate the performance of the median and mode quasi-range estimators in terms of their breakdown points. We will see that one property is very useful to be developed.

We consider that the sample is drawn from the following distribution model,

$$X_i = \begin{cases} Z_i & \text{if outlier does not occur} \\ Z_i + v_i & \text{if outlier does occur} \end{cases},$$

where z_i is independent and identically distributed drawn from an ideal distribution and v_i is the contaminated error with $v_i = 1000 + 10 * i$. If $X_i = Z_i + v_i$, then this x represents an extreme point. In the simulation process, we will determine when X_i will be an extreme point.

In this simulation, the sample size is 1000 with replications $m = 10{,}000$. For jth replication $1 \leq j \leq m$, we first generate a sample $z_1, \ldots, z_n$. By denoting $\hat{\tau}$ as a quasi-range estimator, we specify the formulation, based on $z_1, \ldots, z_n$, of sample $X_1, \ldots, X_n$ and the method of computing the breakdown point of $\hat{\tau}$ at this replication.

Given random variable v_1 and a random number i_1 from $\{1, \ldots, n\}$, let $x_i = \begin{cases} z_i & \text{if } i \neq i_1 \\ z_i + v_1 & \text{if } i = i_1 \end{cases}$ and we compute $|\hat{\tau}^{i_1}(x_1, \ldots, x_n) - \tau|$, denoting $\hat{\tau}^{i_1}$ for $\hat{\tau}$ in this observation.

Given random variables v_1, v_2 and random numbers i_1, i_2 from $\{1, \ldots, n\}$, let $x_i = \begin{cases} z_i & \text{if } i \neq i_1, i_2 \\ z_i + v_1 & \text{if } i = i_1 \\ z_i + v_2 & \text{if } i = i_2 \end{cases}$ and we compute $|\hat{\tau}^{i_1,i_2}(x_1, \ldots, x_n) - \tau|$, denoting $\hat{\tau}^{i_1,i_2}$ for $\hat{\tau}$ in this observation, and so on.

We say that $\hat{\tau}^{i_1,\ldots,i_j}(x_1, \ldots, x_n)$ is the breakdown point if $\hat{\tau}^{i_1,\ldots,i_j}(x_1, \ldots, x_n) - \tau| \geq a$ for some specied constant a.

Then we define the breakdown numbers bd^j_{med} and bd^j_{mod}, respectively, for median quasi-range $\hat{\tau}_{\text{med}}$ And mode quasi-range $\hat{\tau}_{\text{mod}}$ as

$$\begin{aligned} bd^j_{\text{med}}(z_1, \ldots, z_n) &= \frac{k}{n} I \quad (\text{there is } k, k \text{ is the smallest } h,\ 1 \leq h \leq n,\ \text{such that} \\ &\qquad |\hat{\tau}^{i_1,\ldots,i_h}_{\text{med}}(x_1, \ldots, x_n) - \tau_{\text{med}}| \geq a), \\ bd^j_{\text{mod}}(z_1, \ldots, z_n) &= \frac{k}{n} I \quad (\text{there is } k, k \text{ is the smallest } h,\ 1 \leq h \leq n,\ \text{such that} \\ &\qquad |\hat{\tau}^{i_1,\ldots,i_h}_{\text{mod}}(x_1, \ldots, x_n) - \tau_{\text{med}}| \geq a). \end{aligned}$$

The average breakdown points, respectively for $\hat{\tau}_{\text{med}}$ and $\hat{\tau}_{\text{mod}}$ are then defined as

$$\text{BD}_{\text{med}} = \frac{1}{m} \sum_{j=1}^{m} bd^j_{\text{med}}(z_1, \ldots, z_n) \quad \text{and}$$

$$\text{BD}_{\text{mod}} = \frac{1}{m} \sum_{j=1}^{m} bd^j_{\text{mod}}(z_1, \ldots, z_n).$$

In Table 6, we present the average breakdown points of median and mode quasi-range estimators in the case where the ideal distribution of sample $z_1, \ldots, z_n$ is standard normal distribution.

In the next simulation, we consider the gamma distribution with $\alpha = 0.25$ and $\beta = 2$ as the ideal distribution (Table 7).

Table 6. Breakdown points under normal distribution.

γ	BD_{mod}	BD_{med}
0.95	0.051	0.027
0.9	0.101	0.052
0.8	0.2	0.101
0.7	0.301	0.152
0.6	0.4	0.201
0.5	0.5	0.251
0.4	0.6	0.301
0.3	0.7	0.351
0.2	0.8	0.401
0.1	0.9	0.451

Table 7. Breakdown points under gamma distribution (Gamma(2.5, 2)).

γ	BD_{mod}	BD_{med}
0.95	0.0502	0.0267
0.9	0.0992	0.0514
0.8	0.1957	0.0996
0.7	0.2938	0.1502
0.6	0.3891	0.1986
0.5	0.4968	0.2484
0.4	0.597	0.2984
0.3	0.6973	0.3485
0.2	0.7976	0.3987
0.1	0.8986	0.4497

From Tables 6 and 7, we may draw several conclusions.

(a) The breakdown points for the mode quasi-range esitmator are about twice the values of the median quasi-range estimator. Those for mode quasi-range estimator are all about $1 - \gamma$ and those for median quasi-range estimator are about a half of $1 - \gamma$.
(b) The breakdown point for each type of quasi-range is increasing when γ decreases.
(c) Hampel *et al.* [1] claimed that the breakdown point for an estimator may not be larger than 0.5. However, the breakdown point of the mode quasi-range estimator available in this design is not only greater than 0.5, but is close to 1. This significant result has not been observed in the literature.

There are many situations where we use quasi-range estimators to construct statistical inference techniques. Low breakdown point estimates may cause the statistical inferences based on them to be misleading. In addition to using the quasi-range to formulate the standard deviation, the interquartile range and the process capability index are other two examples.

The interquartile range classically is defined as the median quasi-range as

$$\tau_{med}(0.5) = F^{-1}(0.75) - F^{-1}(0.25),$$

and is estimated through the empirical quantile function as $\hat{\tau}_{med}(0.5) = \hat{F}_n^{-1}(0.75) - \hat{F}_n^{-1}(0.25)$. Obviously, we may define an alternative interquartile range using the mode quasi-range as

$$\tau_{mod}(0.5) = \inf_{0<\alpha<0.5}\{F^{-1}(0.5 + \alpha) - F^{-1}(\alpha)\}.$$

When the distribution function F is unknown, the mode quasi-range-based interquartile range may be estimated through the non-parametric technique introduced in Section 4. From our studies, the estimator of the mode quasi-range-based interquartile range does provide a satisfactory version in consideration of MSE and breakdown point.

To understand what the process is actually doing and to see if the process meets the quality requirements or the consumer's expectations, a manufacturer often needs to provide an index for process improvement as a certificate for customers. The general formulation of a capability index is the following form:

$$\frac{\text{USL} - \text{LSL}}{\text{Process spreading limits}},$$

where *LSL* and *USL* represent, respectively, the lower and upper specification limits (see [3] for detailed description). In practice, median quasi-range is generally used to represent the spreading limit as

$$C_p^{\text{med}} = \frac{\text{USL} - \text{LSL}}{\tau_{\text{med}}(\gamma)}. \tag{9}$$

We may then modify the classical C_p to the one based on mode quasi-range as

$$C_p^{\text{mod}} = \frac{\text{USL} - \text{LSL}}{\tau_{\text{mod}}(\gamma)}, \tag{10}$$

which we may call as the mode type process capability index.

In a perfect world, all process data would be normally distributed, although that is not always the case. Although the usual process capability analysis provides some very powerful tools to describe the capability of processes, process data do not always follow a normal distribution. In this situation, with our breakdown point analysis shown, the classically used C_p^{med} may lead to misleading conclusions, whereas the mode type process capability index C_p^{mod} may avoid this deficit.

Acknowledgements

The authors would like to express their gratitute to the editor and a referee for their valuable comments. The constructive and insightful comments from the referee greatly improved the quality of the paper. For the computations involved in this paper the authors would like to thank Zhi-Xiang Jiang. This research work was partially supported by the National Science Council of Taiwan Grant No. NSC95-2118-M-009-007.

References

[1] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel, *Robust Statistics: The Approach Based on Influence Function*, Wiley, New York, 1986.
[2] J.-Y. Huang, L.-A. Chen, and A.H. Welsh, *The mode interval and its application to control charts and tolerance intervals*, submitted to JSPI for publication (in revision) 2007.
[3] S. Kotz and C.R. Lovelace, *Process Capability Indices in and Practice*, Arnold, London, 1998.
[4] W.Q. Meeker and L.A. Escobar, *Statistical Methods for Reliability Data*, Wiley, New York, 1998.
[5] D.C. Montegomery, *Statistical Quality Control*, Wiley, New York, 1991.
[6] R.J. Sering, *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.
[7] R.G. Staudle and S.J. Sheather, *Robust Estimation and Testing*, Wiley, New York, 1990.
[8] S.S. Wilks, Determination of sample sizes for setting tolerance limits, Ann. Math. Statist. 12 (1941), pp. 91–96.