The pure mathematics behind infinite confidence intervals for percentiles

Roger B. Newson

January 4, 2019

1 Introduction

Applied statisticians frequently have good reasons for calculating infinite confidence intervals (with a lower bound of $-\infty$ and/or an upper bound of $+\infty$) for estimating scalar parameters. Some practical examples are mentioned in Newson (2006)[4], in which it is stated that confidence intervals for Theil–Sen percentile slopes (including Hodges–Lehmann percentile differences) can have lower confidence limits of $-\infty$, upper confidence limits of $+\infty$, or both. This is because these confidence intervals are derived by inverting finite confidence intervals for versions of Somers' D or Kendall's τ_a[3].

In the Stata statistical software[6], the quantities $-\infty$ and $+$ are represented by the storable double–precision numbers `c(minddouble)` and `c(maxdouble)`, respectively. These are currently listed as

```
-8.9884656743e+307
```

and as

```
8.9884656743e+307
```

respectively.

If the user has nstalled the optional add–on packages `somersd`, `expgen`, `scsomersd` and `rcentile`[5] (downloadable within Stata using the `ssc` command), and then types, in Stata,

```
sysuse auto, clear
rcentile weight, ce(0(25)100) transf(asin) tdist
```

then the user will see confidence intervals for percentiles 0, 25, 50, 75 and 100 of the variable `weight` in the `auto` dataset, which has 1 observation for each of 74 1970s car models. And Percentile 0 (also known as the minimum) has a lower confidence limit of $-\infty$, and Percentile 100 (also known as the maximum) has a lower confidence limit of $+\infty$. These values are sensible, as there is no evidence in the data that a car cannot have a higher weight than the sample maximum, or that a car cannot have a lower weight than the lower confidence limit (or even a negative weight).

And, if the user then types, in Stata,

```
rcentile weight in 1/4, transf(asin) tdist
```

then the user will see a confidence interval for the population median, estimated from the first 4 car weights, with a lower confidence interval of $-\infty$ and an upper confidence limit of $+\infty$. This is not unreasonable. After all, if we assume nothing about the population distribution except that we are sampling individuals independently from it, then we cannot make many inferences from a sample of 4. For instance, if the population is a mixture of 2 equal–sized and widely–separated subpopulations ('Subpopulation A" and "Subpopulation B"), then 1/16 of samples of 4 will be entirely from Subpopulation A and 1/16 of samples of 4 will be entirely from Subpopulation B, implying that 1/8 of samples will be totally unrepresentative of the population as a whole. Under these conditions, the user cannot really be "95% confident" about the location of the population median, except that it is somewhere on the real line.

Stata makes it easy for the user to talk about infinite confidence limits. However, it is less easy to explain these confidence limits rigorously, in a way that might be understood by a first–year student of real analysis, or of naive set theory. It might therefore be useful to explain the extensions of real–number theory necessary to talk about infinite confidence limits. After all, in a set–theoretic definition of the real numbers such as Stoll (1963), we rarely see any discussion of any such numbers as $-\infty$ and $+\infty$, although real analysis textbooks frequently contain notation such as $x \to \infty$ and $x \to -\infty$.

We will explain the necessary extensions under the following headings:

1. Extending the real numbers to include $-\infty$ and $+\infty$.
2. Extending the binary inequality operators of real analysis to all ordered pairs of extended real numbers.
3. Extending the infimum and supremum operators to all subsets of the extended real numbers.
4. Extending the binary pairwise mean to all pairs of extended real numbers.

We will then introduce the concept of a **ridit function** for an arbitrarily–distributed scalar random variable. Ater this, we will use the extended analytical notation previously introduced to define the left and right inverse ridit functions, and their extended binary pairwise mean, the central inverse ridit funtion. Once these extensions have been added to the notation, we will be able to define extended percentiles, and even extended percentile slopes, differences and ratios[4]. The aim of this exercise is partly to enable applied statisticians to ward off unjustified criticism about insufficient rigor from pettifogging pure mathematicians and career methodologists. Other benefits might include helping programmers to think clearly when writing programs to calculate confidence intervals for rank statistics.

2 Extending the real numbers to include $-\infty$ and $+\infty$

In a set–theoretic definition of the real numbers such as that of Stoll (1963)[7], real numbers are defined in terms of rational numbers, which have previously been defined in terms of integers, which have previously been defined in terms of natural numbers, which have previously been defined in terms of sets (using the Axiom of Infinity). Once natural numbers have been defined, real numbers are typically defined as equivalence classes of bounded sequences $\{a_n\}$ of rational numbers. These classes are defined using the binary equivalence relation

$$\lim_{n\to\infty} |a_n - b_n| = 0 \tag{1}$$

as a relation (that is to say a set of ordered pairs $(\{a_n\}, \{b_n\})$ of bounded infinite sequences of rational numbers. This binary relation can be shown to be an equivalence relation, as it is reflexive, symmetric and transitive. (Note that the notation $n \to \infty$ has usually been defined earlier, without acknowledging the existence of an infinity for n to tend to.)

In a similar way, we can define $-\infty$ and $+\infty$ as classes of sequences. $-\infty$ is the class of sequences $\{a_n\}$ tending to $-\infty$, and $+\infty$ is the class of sequences $\{b_n\}$ tending to $+\infty$. (Once again, we have usually defined "n tending to infinity" without having defined an infinity for n to tend to.)

Once we have defined the set of real numbers $\mathbb{R}$, we can then define an extended set of real numbers

$$\mathbb{R}^* = \mathbb{R} \cup \{-\infty, +\infty\}. \tag{2}$$

3 Extending inequality operators to ordered pairs of extended reals

Real numbers have the feature that a subset S of $\mathbb{R}$ that is bounded below has a maximum lower bound $\inf(S)$, and a subset S of $\mathbb{R}$ has a minimum upper bound $\sup(S)$.

We will extend the binary operators $<$ and $>$ to ordered pairs in the Cartesian product $\mathbb{R}^* \times \mathbb{R}^*$. Recall that a relation is a set of ordered pairs, definable as a subset of a Cartesian product. We will extend the $<$ relation from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}^* \times \mathbb{R}^*$ in such a way that $-\infty < +\infty$ is true, and $-\infty < x < +\infty$ for $x \in \mathbb{R}$. Similarly, we will extend the $>$ relation from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}^* \times \mathbb{R}^*$ in such a way that $+\infty > -\infty$ is true, and $+\infty > x > -\infty$ for $x \in \mathbb{R}$. We can then extend the $\leq$ and $\geq$ operators in the usual way to $\mathbb{R}^* \times \mathbb{R}^*$ by defining $x \leq y$ as $(x = y) \vee (x < y)$, and defining $x \geq y$ as $(x = y) \vee (x > y)$, where $\vee$ is the logical "or" operator. It is easy (but time–consuming and space–consuming) to verify that the $<$ and $>$ operators satisfy the strict ordering axioms (as they are irreflexive, anti–symmetric and transitive), and that the $\leq$ and $\geq$ operators satisfy the partial ordering axioms (as they are reflexive and transitive).

4 Extending the supremum and infimum to subsets of extended reals

The set of real numbers $\mathbb{R}$ has the feature that a non–empty subset S of $\mathbb{R}$ that has a lower bound has a maximum lower bound $\inf(S)$, and a non–empty subset S of $\mathbb{R}$ with an upper bound has a minimum upper bound $\sup(S)$. These definitions can be extended to subsets of $\mathbb{R}^*$, but will not be the same definitions, because *every* subset of $\mathbb{R}^*$ will then have an infimum and a supremum. Therefore, the extended infimum and supremum functions will have new names, distinct from those of the unextended supremum and infimum functions.

For a subset S of $\mathbb{R}^*$, we will define

$$\inf^*(S) = \begin{cases} +\infty, & S = \emptyset \\ -\infty, & S \neq \emptyset \text{ and } -\infty \in S, \\ -\infty, & S \neq \emptyset \text{ and } -\infty \notin S \text{ and } S \cap \mathbb{R} \text{ unbounded below,} \\ \inf(S \cap \mathbb{R}), & S \neq \emptyset \text{ and } -\infty \notin S \text{ and } S \cap \mathbb{R} \text{ bounded below.} \end{cases} \tag{3}$$

And, similarly, we will define

$$\sup^*(S) = \begin{cases} -\infty, & S = \emptyset \\ +\infty, & S \neq \emptyset \text{ and } +\infty \in S, \\ +\infty, & S \neq \emptyset \text{ and } +\infty \notin S \text{ and } S \cap \mathbb{R} \text{ unbounded above,} \\ \sup(S \cap \mathbb{R}), & S \neq \emptyset \text{ and } +\infty \notin S \text{ and } S \cap \mathbb{R} \text{ bounded above.} \end{cases} \tag{4}$$

Note that, if we define $\inf^*(\cdot)$ and $\sup^*(\cdot)$ in this way, then *every* subset $S \subseteq \mathbb{R}^*$ (even the empty set) has an extended infimum and an extended supremum, as a subset S of $\mathbb{R}^*$ cannoot be "unbounded", even if it is empty (in which case $\sup^*(S) = -\infty < +\infty = \inf^*(S)$). This feature of the $\inf^*(\cdot)$ and $\sup^*(\cdot)$ functions will be useful later, when we define inverse ridits.

5 Extending the binary pairwise mean to all pairs of extended reals

Not all functions defined from the extended reals $\mathbb{R}$ can be extended in any simple way to define functions from $\mathbb{R}^*$. And not all binary operations defined from the Cartesian product $\mathbb{R} \times \mathbb{R}$ can be extended in any simple way to define binary operations from $\mathbb{R}^* \times \mathbb{R}^*$. However, we can do this with the binary mean function

$$\mu(x, y) = (x + y)/2, \tag{5}$$

for x and y in $\mathbb{R}$. Note that this binary function is commutative (meaning that $(x, y) = \mu(y, x)$ for all x and y in $\mathbb{R}$). This is as we expect a pairwise mean to be.

In this case, there are 3 classes of possibilities for x, namely $x = -\infty$, $x = +\infty$, and $x \in \mathbb{R}$. And, there are 3 classes of possibilities for y, namely $y = -\infty$, $y = +\infty$, and $y \in \mathbb{R}$. This implies $3 \times 3 = 9$ classes of possibilities for (x, y)–pairs. So the extended function $\mu^*(\cdot, \cdot)$ is best defined using a table, with 1 row for each combination of possibilities. These possibilities, with the corresponding values of $\mu^*(x, y)$, are listed in Table 1.

Table 1: Values of $\mu^*(x, y)$ for combinations of x and y in $\mathbb{R}^*$.

x	y	$\mu^*(x,y)$
$-\infty$	$-\infty$	$-\infty$
$-\infty$	$+\infty$	0
$-\infty$	$y \in \mathbb{R}$	y
$+\infty$	$-\infty$	0
$+\infty$	$+\infty$	$+\infty$
$+\infty$	$y \in \mathbb{R}$	y
$x \in \mathbb{R}$	$-\infty$	x
$x \in \mathbb{R}$	$+\infty$	x
$x \in \mathbb{R}$	$y \in \mathbb{R}$	$\mu(x,y)$

We see that the extended mean function $\mu^*(\cdot, \cdot)$ is commutative, the extended mean of 2 infinite values in the same direction is the same infinite value, the extended mean of 2 infinite values in opposite directions is zero, the extended mean of an infinite value and a finite value is the finite value, and the extended mean of 2 finite values is their unextended mean. This implies that the extended mean function has the feature that, for $x \leq y$, $x \leq \mu^*(x, y) = \mu^*(y, x) \leq y$. So, the extended mean function behaves as we would expect an extended mean function to behave.

6 Definitions of ridit functions

Given a scalar random variable X, the **Bross ridit function**[2] of X is defined for real (or extended real) x by the formula

$$R_X(x) = \Pr(X < x) + \frac{1}{2}\Pr(X = x), \tag{6}$$

which is on a scale from 0 to 1. Note that the Bross ridit function is an alternative to the cumulative distribution function

$$F_X(x) = \Pr(X \leq x), \tag{7}$$

which also uniquely specifies the distribution of the variable X. Both $F_X(x)$ and $R_X(x)$ are monotonically non–decreasing in x, and both tend to 0 in the limit as $x \to -\infty$, and both tend to 1 as $x \to \infty$. However, $F_X(\cdot)$ is right–continuous, but $R_X()$ is not always either right–continuou or left–continuous, although both functions are continuous (and equal) for a continuously distributed random variable X.

An alternative definition of ridits is the **Brockett–Levene ridit function**[1], defined for real (or extended real) x as

$$R^*_X(x) = \Pr(X < x) - \Pr(X > x) = 2R_X(x) - 1, \tag{8}$$

which is on a scale from -1 to 1, and which is mostly used as a means of computing the Bross ridit function by averaging the Brockett–Levene ridit function with 1, avoiding the computational precision issues associated with adding tiny half–probabilities to huge probabilities. We will be focussing on the Bross ridit function.

In mathematical statistics, ridits are mainly used in deriving the asymptotic distribution theory of rank statistics (which are really ridit statistics). We will focus on inverse ridits, also known as percentiles. For a quantity $q \in [0, 1]$, we might want to estimate the $100q$th percentile for a sampled population distribution, with confidence limits, and with a point estimate equal to the $100q$th percentile of the sample distribution. For continuous distributions, the $100q$th percentile is defined simply as the inverse ridit function $R_X^{-1}(q)$, also known as $F_X^{-1}(q)$, and can be computed by inverting the distribution function $F_X(\cdot)$. However, applied statisticians in real life do not encounter continuous distributions, because they always sample finite samples from a population with a finite number of possible X–values (limited by the huge but finite number of values representable in finite precision in a finite number of bytes). In real life, we can assume, without loss of generality, that there are N possible values of a variable X, known as **mass points**, which we can order in ascending order as $x_1, \ldots, x_N$. The unkown population distribution is defined by a probability mass function $f_X(\cdot)$, with the feature that, for real (or extended real) x,

$$Pr(X = x) = \begin{cases} f_X(x), & x \in \{x_i : 1 \leq i \leq N\}, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

(We can assume, without loss of generality, that $f_X(x_i)$ is strictly positive for all positive integers i from 1 to N.) A value x on the real line (or the extended real line) is either less than x_1, greater than x_N, in the set of mass points $x_1, \ldots, x_N$, or in an open interval (x_i, x_{i+1}) for some i such that $1 \leq i < N$. In this case, the Bross ridit function can be redefined as

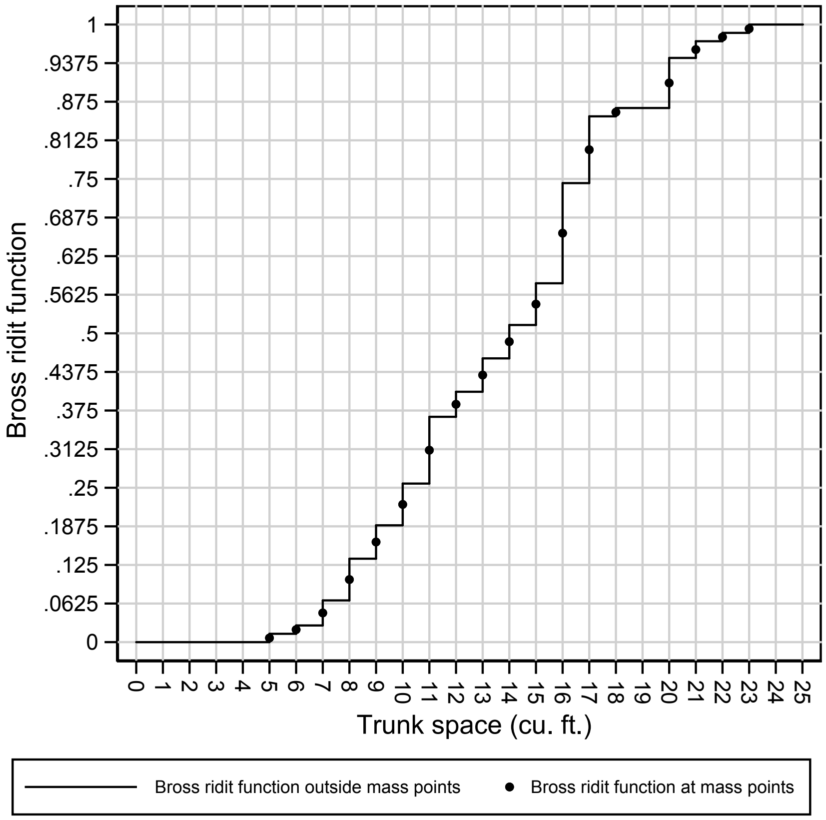
$$R_X(x) = \begin{cases} \frac{1}{2}f_X(x) + \sum_{i:x_i<X} f_X(x_i), & x \in \{x_i : 1 \leq i \leq N\}, \\ \sum_{i:x_i<x} f_X(x_i), & \text{otherwise.} \end{cases} \tag{10}$$

An example of a ridit function defined in this way is illustrated in Figure 1, which shows the Bross ridit function as defined for the variable `trunk` (trunk space in cubic feet) for the `auto` data distributed with the Stata statistical software[6]. The ridit function at the mass points is given as small black circles on the vertical parts of the line, and the ridit function outside the mass points is given as the horizontal parts of the line. We see that there are mass points at all $N = 17$ integers from $x_1 = 5$ to $x_n = 23$ inclusive, and no other mass points, implying that the ridit function is constant within each of the empty intervals $(-\infty, x_i)$, $(x_N, +\infty)$, and (x_i, x_{i+1}) for $1 \leq i \leq N$. As implied by (10), each mass point has a Bross ridit halfway between the constant value for the previous empty interval and the constant value for the successive empty interval. Note that the ridit function defined as in (10) has a value 0 at $x = -\infty$ and a value 1 at $x = +\infty$, although these values are not included in Figure 1.

7 Inverses for ridit functions

We might want to define an inverse function for the ridit function, with a view to defining inverse ridits (also known as percentiles) for X. That is to say, we might want to find a solution in θ to the equation $R_X(\theta) = q$ for q in the open or closed interval from 0 to 1.

Figure 1: Bross ridit function for the variable `trunk` in the `auto` data.



This would be easy for a continuous variable X with a continuous distribution function $F_X(\cdot)$, synonymous with $R_X(\cdot)$, because $F_X(x)$ is then continuous and monotonically increasing in x, implying that there is a unique inverse function $F_X^{-1}(\cdot)$, which is also strictly monotonically increasing. In that case, the $100q$th percentile, for $q \in (0,1)$, is equal to $F_X^{-1}(q)$. *However*, with the ridit function of Figure 1, we see that multiple non–mass points x share the same Bross ridit function value. Also, there are values of $q \in (0,1)$ (on the vertical connection lines between the flat regions of the non–mass intervals) that are not the Bross ridit of any number x. So, the Bross ridit function only appears to be invertible for a finite number of values of q that are ridits of mass points.

Fortunately, using the definitions of the previous Sections, we can define "generalized inverse ridit functions", which have some of the features that we would like an inverse ridit function to have. These features might include mapping values of q equal to the ridits of the mass points back to the mass points, and mapping values of q in between the ridits of 2 consecutive mass points either to the most sensible mass point or to a sensible intermediate value. They might also include mapping $q = 0$ to x_1 as the "0th percentile", and mapping $q = 1$ to x_N as the "100th percentile", enabling us to define inverse ridits from the closed interval $q \in [0,1]$, instead of from the open interval $q \in (0,1)$. And, ideally, they might even include methods enabling us to define confidence limits, or at least conservative confidence limits, around the sample percentiles, for the corresponding population percentiles, under a wide range of sampling schemes, possibly involving clustering and/or sampling–probability weighting.

We will use the extended real line $\mathbb{R}^*$ of (2), the extended definitions $\inf^*(\cdot)$ and $\sup^*(\cdot)$ of (3), and the extended pairwise mean function $\mu^*(\cdot,\cdot)$ of Table 1 to define, for the ridit function $R_X(\cdot)$, a left–inverse ridit function $\mathrm{linvr}_X(\cdot)$, a right–inverse ridit function $\mathrm{rinvr}_X(\cdot)$, and a central–inverse ridit function $\mathrm{cinvr}_X(\cdot)$, from the closed interval $[0,1]$ to $\mathbb{R}^*$. These are defined as

$$\begin{aligned} \mathrm{linvr}_X(q) &= \sup{}^*\left(\{x : x \in \mathbb{R}^* \wedge R_X(x) < q\}\right), \\ \mathrm{rinvr}_X(q) &= \inf{}^*\left(\{x : x \in \mathbb{R}^* \wedge R_X(x) > q\}\right), \\ \mathrm{cinvr}_X(q) &= \mu^*\left[\mathrm{linvr}_X(q), \mathrm{rinvr}_X(q)\right], \end{aligned} \tag{11}$$

where $\wedge$ is the logical "and" operator.

Note that, for our discrete distribution with N finite mass points, the finite extended suprema and infima of the ridit function can only be at mass points, as the ridit function is locally constant outside the mass points. So, for $q \in [0,1]$, $\text{linvr}_X(q)$ is the highest mass point x_i such that $R_X(x_i - \delta) < q$ for positive δ sufficiently small (if such a mass point exists), and otherwise is $-\infty$. And $\text{rinvr}_X(q)$ is the lowest mass point x_i such that $R_X(x_i + \delta) > q$ for positive δ sufficiently small (if such a mass point exists), and otherwise is $+\infty$. And $\text{cinvr}_X(q)$ is the mean of the left and right inverses (if both are finite), or the finite left or right inverse (if only one of them is finite), or $-\infty$ (if both the left and the right inverse are $-\infty$), or $+\infty$ (if both the left and the right inverse are $+\infty$), or zero (if the left inverse is $-\infty$ and the right inverse is $+\infty$).

We define the 100qth quantile as $\text{cinvr}_X(q)$. Note that, if all the mass points are finite, then so are all the quantiles, which can be either mass points or midpoints (also known as means) between 2 consecutive mass points. Note, also, that, if we have a continuous distribution instead of a discrete distribution, then all the 100qth quantiles reduce to the simple formula $F_X^{-1}(q)$.

Figure 2: Inverse Bross ridit function for the variable `trunk` in the `auto` data.

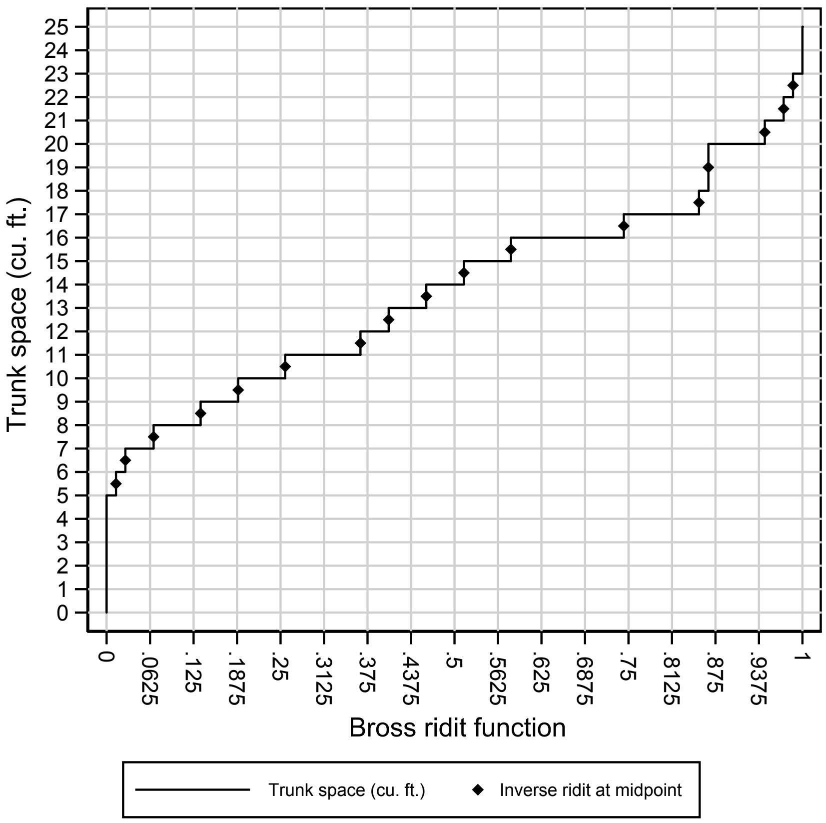


Figure 2 shows a plot of inverse Bross ridits of the `trunk` variable. The line is an inversion of the line in Figure 1, with the horizontal and vertical axes interchanged. The horizontal parts of the line represent inverse ridits which are mass points. We can see that percentiles 25, 50 and 75 are 10, 14 and 17 cubic feet, respectively. The diamonds on the vertical parts of the line represent midpoint inverse ridits, each of which is the midpoint between 2 mass points.

The left and right inverse ridits have other uses, apart from defining the central inverse ridit. We frequently define confidence intervals for percentiles (and for percentile slopes, differences and ratios) by inverting a confidence interval for a mean ridit[4]. The lower confidence interval for the percentile is the left inverse ridit of the lower confidence limit for the mean ridit, and the right confidence limit for the percentile is the right inverse of the upper confidence limit for the mean ridit. These left and right inverse ridits can be infinite.

And inverse ridits can also be defined for Brockett–Levene ridits, defined on a scale from -1 to 1 instead of on a scale from 0 to 1. These are used to define confidence intervals for percentile differences by inverting confidence intervals for differences between mean ridits.

References

[1] Brockett PL, Levene A. 1977. On a characterization of ridits. *The Annals of Statistics* **5(6)**: 1245–1248.

[2] Bross IDJ. 1958. How to use ridit analysis. *Biometrics* **14(1)**: 18–38.

[3] Newson R. Confidence intervals for rank statistics: Somers' D and extensions. *The Stata Journal* 2006; **6(3)**: 309–334. Download from *http://www.stata-journal.com/article.html?article=snp15_6*

[4] Newson R. Confidence intervals for rank statistics: Percentile slopes, differences, and ratios. *The Stata Journal* 2006; **6(4)**: 497–520. Download from *http://www.stata-journal.com/article.html?article=snp15_7*

[5] Newson RB. Easy–to–use packages for estimating rank and spline parameters. Presented at the 20th UK Stata User Meeting, 11–12 September, 2014. Download from *https://ideas.repec.org/p/boc/usug14/01.html*

[6] StataCorp. *Stata: Release 15. Statistical Software.* College Station, TX: StataCorp LLC; 2017.

[7] Stoll RR. *Set Theory and Logic.* New York, NY: Dover Publications Inc.; 1963.