Supplemental material

Sidelining the mean: The relative variability index as a generic mean-corrected variability measure for bounded variables

Merijn Mestdagh

September 20, 2017

In this supplemental material, we will explain how we calculate the maximum attainable variability for bounded measurements. In all situations, we assume N data points in a vector $\boldsymbol{x} = (x_1, ..., x_N)$ with sample average $\bar{x}$. Each measurement is bounded between a and b (i.e., $a \leq x_i \leq b$ for each x_i). Next, we seek to find the maximum variability (e.g., variance, range) given a certain sample average $\bar{x} = c$. First, we will discuss the variance (and the standard deviation), next the (root) mean squared successive difference and finally the inter percentile distance (including the range and the interquartile range).

A Maximum variance

Formally, the problem can be defined as a so-called constrained optimization problem:

$$\begin{aligned} &\underset{\boldsymbol{x}}{\text{maximize}} && \text{var}(\boldsymbol{x}) = \frac{\sum(x_i - \bar{x})^2}{N-1} \\ &\text{subject to the constraints} && a \leq x_i \leq b,\ i = 1, \ldots, N. \\ & && \frac{\sum x_i}{N} = \bar{x} = c, \end{aligned} \tag{1}$$

the maximization of the sample variance of $\boldsymbol{x} = (x_1, ..., x_N)$ for a given $\bar{x} = c$, where x_i is bounded between a and b. Obviously, also $\bar{x} = c$ lies in between a and b. So this means, we are free to change the values of $\boldsymbol{x} = (x_1, ..., x_N)$ in order to maximize the variance, as long as we satisfy the constraints. In the next steps, we will show the following:

1. **Structure of the maximum:** If the data set $\boldsymbol{x}$ has not the properties of a certain structure S, one can always increase the variance by converging to this structure.

2. **Uniqueness of this structure:** There is only one possible realization for this structure.

It logically follows that $\boldsymbol{x}$ will be the solution to the optimization problem (Equation 1) if and only if it possesses structure S. If the data points do not have this structure, one can always bring them in this unique structure to increase the variance. Note that a very similar proof can be constructed using Bauer's maximum principle [2].

A.1 Structure of the maximum

Assume that, for a certain feasible $\boldsymbol{x}$ (satisfying all equality and inequality constraints of Equation 1), there are two elements, say x_i and x_j that are not equal to neither a nor b: $a < x_i < b$ and $a < x_j < b$. We can now create a second data set $\boldsymbol{x'}$ set equal to $\boldsymbol{x}$ except for elements i and j: $x'_i = x_i + \Delta_i$ and $x'_j = x_j + \Delta_j$. To ensure feasibility of $\boldsymbol{x'}$ we must first ensure that $\bar{x}' = c$. Therefore it is necessary that $\Delta_i = -\Delta_j$. Second, we must have that $a \leq x'_i \leq b$ and $a \leq x'_j \leq b$.

The obtained change in variance is now:

$$\begin{aligned}
\text{var}(\boldsymbol{x'}) - \text{var}(\boldsymbol{x}) &= \frac{1}{N-1}\left((x'_i - c)^2 + (x'_j - c)^2 - (x_i - c)^2 - (x_j - c)^2\right) \\
&= \frac{1}{N-1}\left((x_i + \Delta_i - c)^2 + (x_j + \Delta_j - c)^2 - (x_i - c)^2 - (x_j - c)^2\right) \\
&= \frac{1}{N-1}\left((x_i + \Delta_i - c)^2 + (x_j - \Delta_i - c)^2 - (x_i - c)^2 - (x_j - c)^2\right) \\
&= \frac{1}{N-1}\left(2\Delta_i^2 + \Delta_i(x_i - c) - \Delta_i(x_j - c)\right)
\end{aligned} \tag{2}$$

As this is a convex function of Δ_i, the maximum of this increase in variance is only limited by the fact that Δ_i is limited by $a \leq x_i + \Delta_i \leq b$ and $a \leq x_j - \Delta_i \leq b$. This means that the maximum increase always results in $x'_i = a$, $x'_i = b$, $x'_j = a$ or $x'_j = b$ (meaning that one or more of the elements are moved to the bounds). As both x_i and x_j do not have values equal to neither a nor b the maximum increase is strictly positive.

We now have shown that we can increase $\text{var}(\boldsymbol{x})$ if more than one data point lies between a and b. We can repeat this procedure, step by step, until we have a data set with structure S. Structure S has the following property: At most one of the data points lies strictly in between a and b ($a < x_i < b$) and all the other data points are exactly equal to one of the constraints a or b. If we violate this structure one will always be able to change data points to increase the variance until structure S is achieved again.

A.2 Uniqueness of structure S

A feasible data set with structure S and N elements has n_a elements with values equal to a, n_b elements with values equal to b and one element with value m for which we have that

$a \leq m \leq b$. We know that

$$n_a a + n_b b + m = N\bar{x} = Nc, \tag{3}$$

with $n_a, n_b \in \mathbb{N}$ and

$$n_a + n_b = N - 1. \tag{4}$$

If $c = a$ we must have that $n_a = N - 1$, $n_b = 0$ and $m = a$ and if $c = b$ we must have that $n_b = N - 1$, $n_a = 0$ and $m = b$. In other cases, where $a < c < b$ we can rewrite Equation 3 and 4 as

$$\begin{aligned}
(N - n_b - 1)a + n_b b &= Nc - m \\
\Leftrightarrow (b - a)n_b &= Nc - m - (N - 1)a \\
\Leftrightarrow n_b &= \frac{Nc - m - (N - 1)a}{b - a} \\
\Leftrightarrow \frac{Nc - Na}{b - a} - 1 \leq n_b &\leq \frac{Nc - Na}{b - a}
\end{aligned}$$

In the last step we use that $a \leq m \leq b$.

Now there are two possibilities. First, if $\frac{Nc-Na}{b-a}$ is not a natural number, then n_b is uniquely defined (because $n_b \in \mathbb{N}$) as

$$n_b = \left\lfloor \frac{Nc - Na}{b - a} \right\rfloor, \tag{5}$$

where $\lfloor\rfloor$ is the flooring sign. One can then find n_a using Equation 4 as

$$n_a = N - 1 - n_b \tag{6}$$

and m using Equation 3 as

$$m = Nc - n_a a - n_b b. \tag{7}$$

Second, if $\frac{Nc-Na}{b-a}$ is a natural number and $a < c < b$ two solutions are possible. One where $m = a$ and one when $m = b$. Choosing $m = a$ instead of $m = b$ increases n_b by one and logically also decreases n_a by one. In the end, the same number of points have a value equal to a or b. This solution is also found by using Equation 5.

As we can uniquely define the value for each data point, given structure S, structure S is unique.

A.3 The maximum

The maximum variance is given by

$$\max(\mathrm{var}(\boldsymbol{x})) = \frac{n_a(a - c)^2 + n_b(b - c)^2 + (m - c)^2}{N - 1} \tag{8}$$

with m, n_m, n_a and n_b found as previously described.

A.4 A special case: Only bounded at one side

We only handle the case where measurements are bounded by a lower bound (the case where they are bound by only an upper bound can be handled similarly). When measurements are not bounded from above then $b = +\infty$. Again, if $c = a$ we have that $n_a = N - 1$, $n_b = 0$ and $m = a$. When this is not the case we have that

$$n_b = \left\lfloor \frac{Nc - Na}{b - a} \right\rfloor = \left\lfloor \frac{Nc - Na}{\infty} \right\rfloor = 0,$$

so that $n_a = N - 1$ and $m = Nc - (N - 1)a = N(c - a) + a$. This gives the following result:

$$\max(\mathrm{var}(\boldsymbol{x})) = \frac{(N-1)(a-c)^2 + (N(c-a) + a - c)^2}{N-1} = N(a-c)^2.$$

If we further assume that $a = 0$, then the maximum variance is given by $\max(\mathrm{var}(\boldsymbol{x})) = Nc^2$. In such a case, the relative standard deviation, SD^*, is

$$SD^* = \frac{SD}{\sqrt{N}c} = \frac{CV}{\sqrt{N}}$$

where CV is the coefficient of variation.

A.5 The relative standard deviation and its relation to other distributions.

If N goes to infinity there are some interesting relations with existing indices. Because most bounded distributions are originally defined between 0 and 1 (and every measurement instrument can be rescaled to this interval), also we will adopt these bounds in the following paragraphs, leading to $a = 0$ and $b = 1$. If N is going to infinity and $a = 0$ and $b = 1$, we find, using Equations 5 and 6, that

$$n_b = Nc$$

and that

$$n_a = N - 1 - Nc = N(1 - c).$$

Using this results in Equation 8 gives us that

$$\begin{aligned}
\lim_{N\to\infty} \max(\mathrm{var}(\boldsymbol{x})) &= \frac{N(1-c)c^2 + Nc(1-c)^2 + (m-c)^2}{N-1} \\
&= \frac{N((1-c)c^2 + c(1-c)^2) + (m-c)^2}{N-1} \\
&= (1-c)c^2 + c(1-c)^2 \\
&= (1-c)c \\
&= (1-\bar{x})\bar{x}
\end{aligned} \tag{9}$$

We can now use this result in combination with existing indices.

A.5.1 The Bernouilli distribution

The probability distribution defined by a Bernoulli distribution is given by

$$\begin{aligned} Pr(x_i = 1) &= p \\ Pr(x_i = 0) &= 1 - p \end{aligned}$$

For the mean of the Bernouilli distribution one has that

$$\mathbf{E}(\bar{x}) = p$$

while, as $N\bar{x}$ is the number ones in $\boldsymbol{x}$, the variance $\mathrm{var}(\boldsymbol{x})$ is given by

$$\begin{aligned} \mathrm{var}(\boldsymbol{x}) &= \frac{\sum(x_i - \bar{x})^2}{N-1} \\ &= \frac{N\bar{x}(1-\bar{x})^2 + (N - N\bar{x})(0-\bar{x})^2}{N-1} \\ &= \frac{N\bar{x}(1-\bar{x})^2 + N(1-\bar{x})(\bar{x})^2}{N-1} \\ &= \frac{N\bar{x}(1-\bar{x})(1-\bar{x}+\bar{x})}{N-1} \\ &= \bar{x}(1-\bar{x})\frac{N}{N-1} \end{aligned}$$

There is an exact mathematical relation between the sample average and the sample variance. In the Bernoulli distribution all data points are equal to one of the bounds, 0 or 1. As we have shown this means that the variance is always equal to the maximum possible variance given the mean. The relative standard deviation (and the relative variance) is therefore always equal to one:

$$SD^* = 1$$

A.5.2 The Binomial distribution

The Binomial distribution is a generalization of the Bernoulli distribution and is given by

$$\mathrm{Pr}(x_i = m) = \binom{n}{m} p^m (1-p)^{(n-m)}$$

where n is the number of trials (which is different from N the number data points in a vector $\boldsymbol{x} = (x_1, ..., x_N)$) and m is the number of successes. To scale these distributions between the bounds of $a = 0$ and $b = 1$ we define the Binomial proportion distribution as

$$\mathrm{Pr}(x_i = \frac{m}{n}) = \binom{n}{m} p^m (1-p)^{(n-m)}.$$

For N going to infinity we have that

$$\bar{x} = p$$

and

$$\mathrm{var}(\boldsymbol{x}) = \frac{p(1-p)}{n}$$

As the maximum variance is given by Equation 9. The standard deviation and the relative standard deviation are given by.

$$SD = \sqrt{\frac{p(1-p)}{n}}$$
$$SD^* = \frac{1}{\sqrt{n}}$$

The relative standard deviation is no function of p (or indirectly of $\bar{x}$) as opposed to the normal standard deviation. This is illustrated in Figure 1.

A.5.3 The Beta distribution

Another well-known, somewhat related bounded distribution, is the beta distribution. Here the probability density function is mostly parametrized using α and β and is given by

$$f(x_i) = \frac{1}{\mathbf{B}(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

Another way to parametrize this is by using the mean μ and the concentration or precision v (in Bayesian statistics, this parameter can also be seen as the sample size, the number of trials) [3, 1]. This precision v is a measure of how concentrated the distribution is around a certain mean μ. The relation between this parametrization and the original parametrization is the following:

$$\mu = \frac{\alpha}{\alpha+\beta}$$
$$v = \alpha + \beta$$

For N going to infinity we have that the sample mean and sample standard deviation are given by

$$\bar{x} = \mu$$
$$SD = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}$$
$$= \sqrt{\frac{\mu(1-\mu)}{1+v}}$$

Using Equation 9 we find that the relative standard deviation is given by

$$SD^* = \frac{1}{\sqrt{1+v}}.$$

While the standard deviation is a function of the mean μ and the concentration v, the relative standard deviation is only a function of the concentration v. This can also be seen in Figure 1.

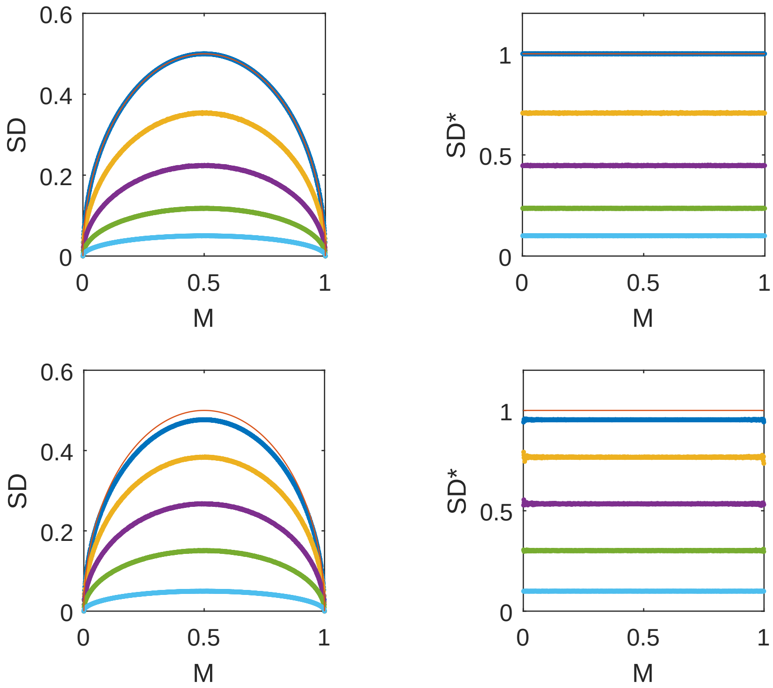

Figure 1: Simulations of variability measures of different series $\boldsymbol{x} = (x_1, ..., x_N)$ where N is going to infinity. Top panels: Samples from binomial distributions with random p (points with the same color come from a distribution with the same number of trials n). Note that we display the proportions (number of successes m divided by number of trials n). The top distributions in blue are the Bernoulli distributions (i.e., a binomial with a single trial, $n = 1$). Middle panels: beta distributions with random μ and different concentrations or precisions v (same color refers to panels with same concentration).

B Maximum mean squared successive difference

If we have one long uninterrupted time series, the mean squared successive difference ($MSSD$) of time series data $\boldsymbol{x} = (x_1, ..., x_N)$ is given by

$$MSSD(\boldsymbol{x}) = \frac{\sum_{i=1}^{N-1}(x_{i+1} - x_i)^2}{N-1}.$$

However, in psychological studies, these time series are often divided in P parts, with N_p elements for part p, because of missing data or day-night interruptions (as is commonly the case in experience sampling). In this case, the $MSSD$ is given by

$$MSSD(\boldsymbol{x}) = \frac{\sum_{p=1}^{P}\sum_{i=1}^{N_p-1}(x_{p,i+1} - x_{p,i})^2}{\sum_{p=1}^{P}(N_p - 1)}, \tag{10}$$

where $x_{p,i}$ is time point i of part p of the time series. Here we will find the solution to the following optimization problem:

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{maximize}} \quad & MSSD(\boldsymbol{x}) = \frac{\sum_{p=1}^{p=P} \sum_{i=1}^{N_p-1} (x_{p,i+1} - x_{p,i})^2}{\sum_{p=1}^{P} (N_p - 1)} \\ \text{subject to} \quad & a \le x_{p,i} \le b, \ p = i \dots P \text{ and } i = 1, \dots, N. \\ & \frac{\sum_{p=1}^{p=P} \sum_{i=1}^{N_p} x_{p,i}}{\sum_{p=1}^{P} (N_p)} = \bar{x} = c. \end{aligned} \tag{11}$$

For this optimization problem, not only the values of elements $x_{p,i}$ are important, but also the order. First we will discuss how the values of elements $x_{p,i}$ can be found independently of the order and then we will discuss the order of these elements.

B.1 Value of the elements of the maximum

The values of the elements $x_{p,i}$ which solve the optimization problem can be found in the same way as for the maximum variance (or by using Bauer's maximum principle [2]). Assume again that for a certain feasible $\boldsymbol{x}$ (satisfying all equality and inequality constraints of Equation 11), there are two elements, say $x_{p,i}$ and $x_{q,j}$ that are not equal to neither a nor b: $a < x_{p,i} < b$ and $a < x_{q,j} < b$. We can now create a second data set $\boldsymbol{x'}$ set equal to $\boldsymbol{x}$ except for elements (p, i) and (q, j): $x'_{p,i} = x_{p,i} + \Delta_i$ and $x'_{q,j} = x_{q,j} + \Delta_j$. To ensure feasibility of $\boldsymbol{x'}$ we must first ensure that $\bar{x'} = c$. Therefore it is necessary that $\Delta_i = -\Delta_j$.

As the formula to compute the $MSSD$ (Equation 10) is a convex function of the elements of $\boldsymbol{x}$, it follows that also the change in $MSSD$: $MSSD(\boldsymbol{x'}) - MSSD(\boldsymbol{x})$ is a convex function of Δ_i (as in Equation 2). This means that the maximum of this increase in $MSSD$ is only limited by the fact that Δ_i is limited by $a \le x_{p,i} + \Delta_i \le b$ and $a \le x_{q,j} - \Delta_i \le b$. This means that the maximum increase always results in $x'_{p,i} = a$, $x'_{p,i} = b$, $x'_{q,j} = a$ or $x'_{q,j} = b$ (meaning that one or more of the elements are moved to the bounds). As b oth $x_{p,i}$ and $x_{q,j}$ do not have values equal to neither a nor b the maximum increase is strictly positive.

Therefore, we achieve the same conclusion as with the maximum variance. One can always increase the $MSSD(\boldsymbol{x})$, independent of the order of the elements, until we have a data set with structure S. Structure S has the following property: At most one of the data points lies strictly in between a and b ($a < x_{p,i} < b$) and all the other data points are exactly equal to one of the constraints a or b. If we violate this structure one will always be able to change data points to increase the $MSSD$ until structure S is achieved again. As this structure S is unique (see Section 'Uniqueness of structure S'), it follows that $\boldsymbol{x}$ can only maximize the $MSSD$ if it possesses structure S.

B.2 Order of the maximum

This does however not mean that any $\boldsymbol{x}$ with structure S is a solution to Equation 11. Structure S defines only the values of the elements the maximum, but not the order. The order of the elements is more difficult. We do not provide any analytical expression to solve this problem. In Matlab and R we solve this using a brute force algorithm.

B.3 Bounded at one side

We only handle the case where measurements are bounded by a lower bound (the case where they are bound by only an upper bound can be handled similarly). When measurements are not bounded from above then $b = +\infty$. Again, if $c = a$ we have that $n_a = N - 1$, $n_b = 0$ and $m = a$. When this is not the case we have that

$$n_b = \left\lfloor \frac{Nc - Na}{b - a} \right\rfloor = \left\lfloor \frac{Nc - Na}{\infty} \right\rfloor = 0,$$

so that $n_a = N - 1$ and $m = Nc - (N-1)a = N(c-a) + a$. If there exists at least one part with $N_p > 2$, the optimal order of the elements will be one where m is neighboring a at two sides so that the maximum $MSSD$ is given by

$$\max(MSSD(\boldsymbol{x})) = \frac{2(N(c-a)+a-a)^2}{\sum_{p=1}^{p=P}(N_p - 1)} = \frac{2N^2(c-a)^2}{\sum_{p=1}^{p=P}(N_p - 1)}$$

If $a = 0$, the $RMSSD = \sqrt{MSSD}$ becomes

$$\max(RMSSD(\boldsymbol{x})) = \frac{\sqrt{2}N}{\sqrt{\sum_{p=1}^{p=P}(N_p - 1)}} c.$$

Also for the $RMSSD^*$, where one has to divide the $RMSSD$ by this function of $c = \bar{x}$, the link with the CV is obvious.

C Interpercentile distance

In this section we will find the solution of the following problem:

$$\begin{aligned}
\underset{\boldsymbol{x}}{\text{maximize}} \quad & p_1, p_2 \text{ interpercentile distance}(\boldsymbol{x}) = x_{n_2} - x_{n_1} \\
\text{subject to} \quad & a \leq x_i \leq x_{i+1} \leq b,\ i = 1, \ldots, N. \\
n_1 &= \lceil p_1 N \rceil \\
n_2 &= \lceil p_2 N \rceil \\
\frac{\sum x_i}{N} &= \bar{x} = c,
\end{aligned} \tag{12}$$

The maximum distance between percentile p_2 and p_1 for an ordered set $\boldsymbol{x} = x_1, ..., x_N$ where $a \leq x_i \leq x_{i+1} \leq b$, for a given $\bar{x} = c$. This is obviously only possible if also c lies in between a and b. In the following proof, it is not important how n_1 and n_2 are computed.

Therefore, this proof will also hold for other methods that define a percentile differently (e.g. $n_1 = \lfloor p_1 N \rfloor$).

The maximum interpercentile distance is always bounded by $b-a$. If this distance $(b-a)$ is achieved, at least n_1 elements have values equal to a and at least $N-n_2+1$ elements have values equal to b. In such a case, for a feasible $\bar{x}$ this means that we can write that

$$Nc = N\bar{x} = n_1 a + (N - n_2 + 1)b + \sum_{i=n_1+1}^{n2-1} x_i. \tag{13}$$

Because $a \leq x_i \leq b$ we can look at the range of possible c for which this maximum distance is possible:

$$\begin{aligned} n_1 a + (N - n_2 + 1)b + (n_2 - n_1 - 1)a \leq Nc \leq n_1 a + (N - n_2 + 1)b + (n_2 - n_1 - 1)b \\ \Leftrightarrow (n_2 - 1)a + (N - n_2 + 1)b \leq Nc \leq n_1 a + (N - n_1)b \end{aligned} \tag{14}$$

This means that when Nc satisfies Equation 14, the maximum interpercentile distance will be $b-a$. For example, when $Nc = (n_2-1)a+(N-n_2+1)b$ the maximum distance $(b-a)$ will be reached when the first (n_2-1) elements have values equal to a and the last $(N-n_2+1)$ elements have values equal to b. All data points are equal to one of the constraints. Now lets look at the case where $Nc < (n_2-1)a+(N-n_2+1)b$ (the case where $Nc > n_1 a + (N-n_1)b$ is exactly symmetrical and will not be discussed here). We will show the following:

1. **Structure of the maximum:** If the data set $\boldsymbol{x}$ has not the properties of a certain structure T, one can always increase the interpercentile range by converging to this structure.

2. **Uniqueness of this structure:** This structure exists if $Nc < (n_2-1)a+(N-n_2+1)b$ and is unique.

It logically follows that $\boldsymbol{x}$ will be the solution to the optimization problem (Equation 12) if and only if it possesses structure T and $Nc < (n_2-1)a + (N-n_2+1)b$. If the data points do not have this structure, one can always bring them in this unique structure to increase the interpercentile range.

C.1 Structure of the maximum

We can now define two rules which will always increase the maximum interpercentile range. We start with ordered data set $\boldsymbol{x}$ and change it into ordered data set $\boldsymbol{x'}$.

1. If there exists any element $x_i > a$ with $i < n_2$ and $x_{n_2} < b$ [1] the interpercentile range can be increased by making sure that $x'_i < x_i$ and $x'_{n_2} > x_{n_2}$ while $\bar{x'} = c$. This way $\boldsymbol{x'}$ will have a higher interpercentile range as $x'_{n_1} \leq x_{n_1}$ while $x'_{n_2} > x_{n_2}$.

[1] In case of $Nc < (n_2-1)a + (N-n_2+1)b$, one always has that $x_{n_2} < b$

2. If there exists any element $x_j > x_{n_2}$ we can increase the interpercentile range by making sure that $x'_j = x'_{n_2}$ and $x'_{n_2} > x_{n_2}$ while $\bar{x'} = c$. This way we have that $x'_{n_1} = x_{n_1}$ while $x'_{n_2} > x_{n_2}$.

We can not apply this rules if the data set has structure T:

$$\forall i < n_2 : x_i = a$$
$$\forall i, j \geq n_2 : x_i = x_j$$

In any other case we can apply one of the two rules to converge to structure T and to increase the interpercentile range.

C.2 Uniqueness of the maximum

For a data set with structure T we have that $n_2 - 1$ elements are equal to a and $N - n_2 + 1$ elements are equal to m. For a feasible data set we have that

$$(n_2 - 1)a + (N - n_2 + 1)m = N\bar{x} = Nc$$

which uniquely defines m as

$$m = \frac{Nc - (n_2 - 1)a}{N - n_2 + 1}$$

as long as $Na \leq Nc < (n_2 - 1)a + (N - n_2 + 1)b$ this results indeed leads to a m that is within the constraints: $a \leq m \leq b$. The maximum interpercentile distance is

$$m - a = \frac{Nc - (n_2 - 1)a}{N - n_2 + 1} - a = \frac{N(c - a)}{N - n_2 + 1}$$

C.3 The maximum

The maximum interpercentile range is given by

$$\text{range}(\boldsymbol{x}) = \begin{cases} \dfrac{N(c-a)}{N - n_2 + 1} & \text{if } a \leq Nc < (n_2 - 1)a + (N - n_2 + 1)b \\ b - a & \text{if } (n_2 - 1)a + (N - n_2 + 1)b \leq Nc \leq n_1 a + (N - n_1)b \\ \dfrac{N(b-c)}{n_1} & \text{if } n_1 a + (N - n_1)b < Nc \leq b \end{cases} \tag{15}$$

Note that the case where $a + (N - 1)b < Nc$ was not explicitly discussed but is exactly symmetrical to the case where $Nc < (N - 1)a + b$.

C.4 Maximum range

In case of the maximum range, we have that $n_1 = 1$ and $n_2 = N$. The maximum is now given by

$$\text{range}(\boldsymbol{x}) = \begin{cases} N(c-a) & \text{if } a \leq Nc < (N-1)a+b \\ b-a & \text{if } (N-1)a+b \leq Nc \leq a+(N-1)b \\ N(b-c) & \text{if } a+(N-1)b < Nc \leq b \end{cases}$$

C.5 Maximum interquartile range

In case of the interquartile range we have that $p_1 = 0.25$ and $p_2 = 0.75$. This means that $n_1 = \lceil 0.25N \rceil$ and $n_2 = \lceil 0.75N \rceil$. Now we can use Equation 15 to find the maximum interquartile range.

C.6 Bounded at one side

We only handle the case where measurements are bounded by a lower bound (the case where they are bound by only an upper bound can be handled similarly). When measurements are not bounded from above then $b = +\infty$. Using Equation 15 we find the maximum using the part where $Nc \leq (n_2 - 1)a + (N - n_2 + 1)b$. If $a = 0$ the maximum becomes $\frac{Nc}{N-n_2+1}$. A division by this maximum is again very similar to the procedure of the calculation of the coefficient of variation.

References

[1] J. K. Kruschke. *Doing Bayesian data analysis: A tutorial with R and BUGS.* Academic Press / Elsevier, 2015.

[2] M. Kruzik. Bauer's maximum principle and hulls of sets. *Calculus of Variations and Partial Differential Equations*, 11(3):321–332, Nov. 2000.

[3] A. V. Rocha and F. Cribari-Neto. Beta autoregressive moving average models. *TEST*, 18(3):529–545, June 2008.