Overlapping Bins Gilles Gnacadja

a: lower bound of interval

b: upper bound of interval

n: number of bins, $n \ge 1$

c: bin width

r: bin overlap ratio

 \bar{r} : bin non-overlap ratio

 $s\,$: bin width to interval length ratio

 r_k : bin k-overlap ratio, $k \ge 1$

 a_i : bin lower bound, $i = 1, \ldots, n$

 b_i : bin upper bound, $i = 1, \ldots, n$

 $\gamma\,$: Ayasdi gain

$$0 < c \leqslant b - a$$

$$0 \leqslant r < 1$$

$$a = a_1 < \dots < a_n < b$$

$$a < b_1 < \cdots < b_n = b$$

$$b_i - a_i = c , \quad i = 1, \dots, n$$

$$b_i - a_{i+1} = rc$$
, $i = 1, \dots, n-1$

$$r_1 = r$$

$$b_i - a_{i+k} = r_k c$$
, $i = 1, ..., n-1, k \ge 1$

$$a_{i+1} - a_i = b_{i+1} - b_i = (1-r)c$$
, $i = 1, ..., n-1$

$$a_i = a + (i-1)(1-r)c$$
, $i = 1, ..., n$

$$b_i = b - (n-i)(1-r)c$$
, $i = 1, ..., n$

$$c = b_1 - a_1$$

$$=b_1-a$$

$$= b - (n-1)(1-r)c - a$$

$$b - a = (1 + (n-1)(1-r))c$$

= $(1 + (1-r)n - (1-r))c$
= $(r + (1-r)n)c$

$$b - a = (r + (1 - r)n)c$$

$$n = \frac{1}{1-r} \left(\frac{b-a}{c} - r \right) = \frac{1}{1-r} \left(\frac{1}{s} - r \right) = 1 + \left(\frac{1}{s} - 1 \right) \gamma$$

$$c = \frac{b-a}{r+(1-r)n}$$

$$r = \frac{1}{n-1} \left(n - \frac{b-a}{c} \right) = \frac{1}{n-1} \left(n - \frac{1}{s} \right)$$

$$s = \frac{1}{r + (1 - r)n} = \frac{1}{1 + (n - 1)/\gamma}$$

Overlapping Bins Gilles Gnacadja

$$r_k c = b_i - a_{i+k}$$

$$= (b - (n-i)(1-r)c) - (a + (i+k-1)(1-r)c)$$

$$= (b-a) - ((n-i) + (i+k-1))(1-r)c$$

$$= (r + (1-r)n)c - (n+k-1)(1-r)c$$

$$r_k = r + (1-r)n - (n+k-1)(1-r)$$

$$= 1 - k(1-r)$$

$$r_k = 1 - k(1 - r) = 1 - k/\gamma$$

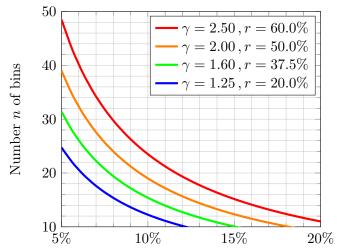
$$\boxed{r_k < 0 \,,\, \forall \, k \geqslant 2 \, \Leftrightarrow \, r_2 < 0 \, \Leftrightarrow \, r < 0.5 \, \Leftrightarrow \, \gamma < 2}$$

$$\bar{r} = 1 - 2r = -r_2$$

The bin non-overlap ratio \bar{r} is a sensible indicator only if r < 0.5, i.e. $\gamma < 2$. With $r \ge 0.5$, the inner bins (indices i = 2, ..., n - 1) have no proper elements.

A few parameter settings of interest

γ	r	\bar{r}	r_2
1.25	20.0%	60%	
1.60	37.5%	25%	•
2.00	50.0%	00%	00%
2.50	60.0%	•	20%



Bin width to interval length ratio s