UP - US, distance between nearby points and segments

Tracks

1. Track 1: unmarked points
2. Track 2: unmarked segments

Question

Where in the genome are the points in track 1 closer to/further apart from the segments in track 2 than expected by chance?

Comment:

- The test is valid for all combinations of the alternative combinations of preservation and randomization of points in track 1 and segments in track 2. The test is not symmetric in the two tracks.
- Significance is determined by means of p-values. Small p-values identify regions where the points in track 1 are closer to or further apart from the closest segment in track 2 than expected. P-values are computed as explained below, where the null hypothesis is explained in detail.
- The p-values are found by simulation.

Bins

The genome (or the areas of the genome under study) are divided into small regions, called bins. The tests are performed in each bin.

Hypothesis tested

Hypothesis tested

For each bin i we have the null hypothesis:
H$_0$: *The points in track 1 are independent of the segments in track 2.*
and the following alternative hypotheses:
H$_1$: *Points in track 1 are closer to the segments in track 2 than expected* or
H$_2$: *Points in track 1 are further apart from the segments in track 2 than expected.*

Define the distance d_i as the smallest distance between point i in track 1 and a segment in track 2 for $i = 1, 2, \cdots, n$. If the point i is inside a segment, then $d_i = 0$. We use the test statistics $X = \frac{1}{n}\sum_{i=1}^{n} d_i$. The distribution for this test statistics is not know and it is necessary with MC simulation in order to decide whether to reject the hypothesis.