

4.5.5 - Uniformity Measurements

The Flux Simulator provides some statistical indicators to measure the uniformity of the read distribution along a transcript produced by *in silico* sequencing.

Fraction Covered

The fraction of a transcript that is covered by reads reflects its expression and the degree of coverage fluctuation provoked by biases.

$\frac{1}{L} \sum_{i=1}^L \mathbb{1}_{i \in R}$ where L is the length of the transcript, and the sign function

$\mathbb{1}_{i \in R}$

indicating whether a position is covered by at least one sequenced read.

Chi-square statistics (χ^2)

Pearson's chi-square can be used to test the goodness of fit of a given sample to a theoretical distribution. Given a transcript of length

L and coverage c_i at position

i , the test statistic is defined as follows:

$$\chi^2 = \sum_{i=1}^L \frac{(c_i - \bar{c})^2}{\bar{c}}$$

where \bar{c} is the average coverage along the molecule.

Coefficient of variation (CV)

A CV is defined as the ratio of the standard deviation and the standard deviation of a probability distribution:

$\frac{\sigma}{\mu}$, with standard deviation σ and mean

of transformed coverage values $\frac{\sqrt{c_i}}{c_i}$

Latter Anscombe transformation of coverage values has been proposed [Hansen et al. 2010] under the assumption that the distribution of reads along a transcript follows a Poisson distribution, which is to be transformed to a Gaussian distribution.