MEMO: How to correct the final concentration by the LOB?

When the limit of blank is non-zero, the estimated quantity of target nucleic acids in a test sample should be corrected by "subtracting the false positives" weighted by their probability distribution.

a) An exact calculation method:

1) Determine the limit of blank $LOB(95\%)$ of your experiment by applying the method of the memo "How to calculate the limit of blank" on R replicates.

2) From the numbers $x(i)$ of "false positive" partitions observed in the well of each replicate i , calculate the probability $P(FP = k)$ of observing k "false positive" partitions in a replicate for each integer k ranging from 0 to $K = \max_i \{ x(i), i = 1..R \}$. To do this, simply count the number of replicates in which k "false positive" partitions have been observed.

3) Count the number q of positive partitions observed in the test sample.

   These observed positive partitions include true positive partitions and possibly false positive partitions.

   According to the Bayes Rule, the probability of having q true positive partitions in a test sample knowing the probability $\Lambda = e^{-\lambda}$ that a partition is negative (where λ is the average number of copies per partition), is the following:

$$P(q/\Lambda) = \sum_{k=0}^{K} P(FP = k)\, P(q/\Lambda\,, FP = k)$$

$$f(\Lambda) = P(q/\Lambda) = Cste \sum_{k=0}^{K} P(FP = k)\, (1-\Lambda)^{q-k}\, \Lambda^{N-(q-k)}$$

4) The true concentration of target nucleic acids in the well is given by:

$$C_{true} = -\frac{1}{v}\ln(\Lambda_{true})$$

   where Λ_{true} is the value of Λ which maximizes the function $f(\Lambda) = P(q/\Lambda)$.

   If we denote C the estimated concentration without correction by the possible false positives, we have the following inequality: $C_{true} \leq C$.

   There is no explicit formula for Λ_{true} but it is possible to determine it by discretizing a set of possible values for Λ between 0 and 1 (for example using Excel), by plotting the curve of the function $f(\Lambda)$ and by taking the abscissa from the highest point on the curve.

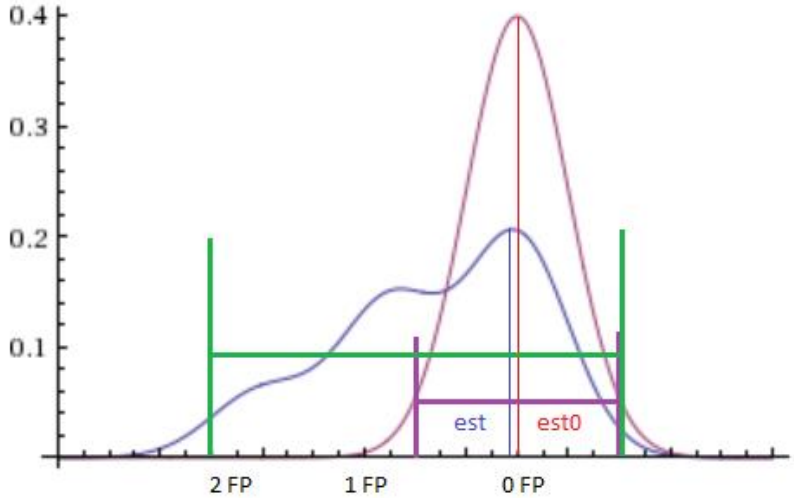
- Special case: if there are never false positives, then $K = 0$ and $P(FP = 0) = 1$, which implies $C_{true} = C = -\frac{1}{v}\ln(1 - q/N)$

5) The true 95% confidence interval, denoted $[C_{true}^{min}, C_{true}^{max}]$, is larger than the interval calculated without correction, denoted $[C^{min}, C^{max}]$.

There is no explicit formula for the true confidence interval, but it is possible to determine it:

 - Take as the true upper bound the upper bound of the uncorrected confidence interval: $C_{true}^{max} = -\frac{1}{v}\ln(\Lambda_{true}^{max}) = C^{max}$
 - The value $C_{true}^{min} = -\frac{1}{v}\ln(\Lambda_{true}^{min})$ is determined such that the area located under the curve $f(\Lambda)$ and included between the vertical lines $x = \Lambda_{true}^{min}$ and $x = \Lambda_{true}^{max}$ is equal to 95% of the total area located under the curve $f(\Lambda)$. The area calculations can be approximated by sums of rectangle areas of rectangles following the discretization of Λ.

- Example of curve $f(\Lambda)$ (blue curve) obtained for $K = 3$ with $P(FP = 0) = 0.75$, $P(FP = 1) = 0.20$ and $P(FP = 2) = 0.05$.



 - The purple curve represents the distribution of the case where there are no false positives, with the uncorrected concentration given by the red vertical line and the uncorrected confidence interval given by the purple horizontal segment.

 - The blue curve represents the true case which is a weighted sum of distributions, with the corrected concentration given by the blue vertical line and the corrected confidence interval given by the green horizontal segment.

- Special case: if there are never false positives then $[C_{true}^{min}, C_{true}^{max}] = [C^{min}, C^{max}]$.

b) An approximate calculation method:

Count the number q of positive partitions observed in the test sample. These observed positive partitions include true positive partitions and possibly false positive partitions.

An approximate calculation method consists in:

- Taking this lower-bound for the true concentration in the well:

$$C_{approx,\ lower-bound} = -\frac{1}{v}\ln\left(1 - \frac{q - LOB(95\%)}{N}\right)$$

- Taking this upper-bound for the true concentration in the well:

$$C_{approx,\ upper-bound} = -\frac{1}{v}\ln\left(1 - \frac{q}{N}\right)$$

- Taking this lower-bound for the true 95% confidence interval:

$$C_{approx}^{min}(95\%) = -\frac{1}{v}\ln\left(1 - \frac{q - LOB(95\%)}{N} - 1.96\sqrt{\frac{\frac{q}{N}\left(1 - \frac{q}{N}\right)}{N}}\right)$$

- Taking this upper-bound for the true 95% confidence interval:

$$C_{approx,\ upperbound} = -\frac{1}{v}\ln\left(1 - \frac{q}{N}\right)$$