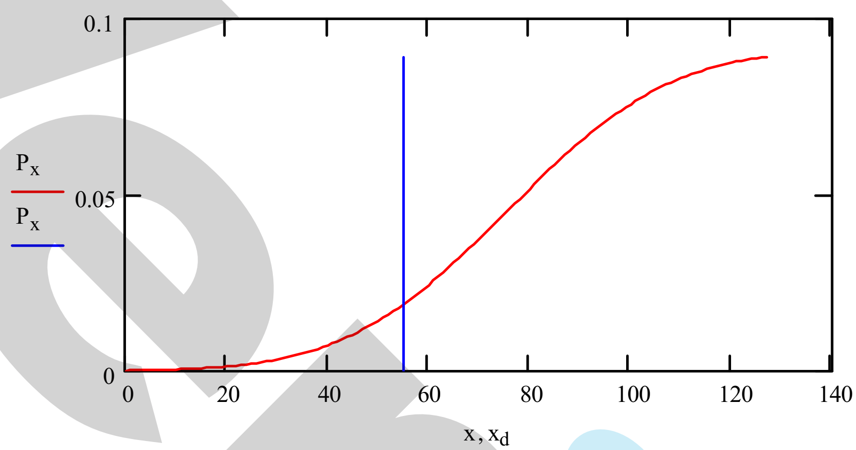
Super Enhanced D-Value Method (pos_{sed}):

The Enhanced D-value routine can be improved even more by including a robust estimate for P(x), the fraction of positives at x. As discussed above, an exact solution to the positive fraction for all x with Cx>0 is described the the equation,

$$POS(x) := \frac{D_x + P_x}{C_x}$$

The enhanced D-value routine assumes Px is zero and evaluates D_x/C_x at $x=x_d$.

One way of possibly improving the enhanced D-value routine would be to try to estimate the $P(x=x_d)$. Let's look at this question graphically.

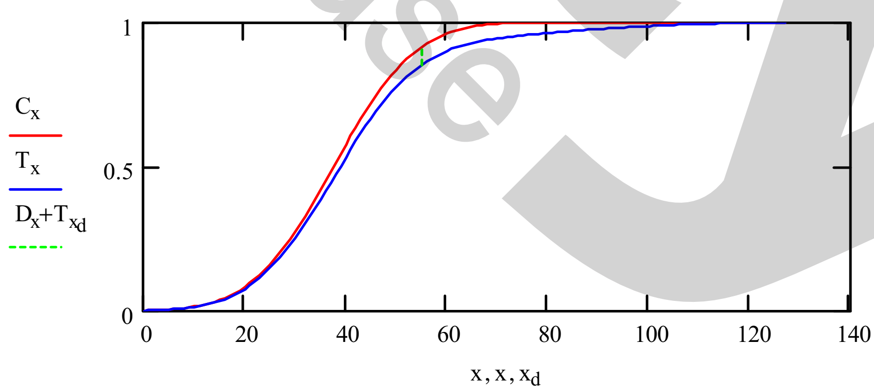


If one could estimate P(x=xd), then the function,

$$\frac{\left(D_{x_d} + P_{x_d}\right)}{C_{x_d}} = 0.089 \qquad pos = 0.089$$

would more precisely estimate the positive fraction. Note, the exact positive fraction value, pos, equals the above expression.

Recall that the D-value is given by the maximum difference between the normalized cumulative control (C) and test (T) histograms.



As shown above, we can improve upon this estimate by dividing by $C(x=x_d)$ (enhanced D-value). Thus, the algorithm for estimating $P(x=x_r)$, where r is the last channel in the histogram or channel domain, is 1) normalize the cumulative distributions, C and T, so that they are unity at the end of the x-axis (r), 2) find the maximum difference value between these two cumulative distributions and 3) divide by the cumulative control evaluated at this location of maximum difference.

We can potentially estimate P(x=xd) by applying the same idea to a subportion of the two cumulative distributions. The example below will make this point clearer.
Suppose we renormalize C and T such that they are unity at $x=x_d$. Our new x domain will range from 0 to xd. Mathematically we are just substituting xd for r in the above described algorithm.

$$x2 := 0 .. x_d$$

$$C2_{x2} := \frac{C_{x2}}{C_{x_d}} \qquad T2_{x2} := \frac{T_{x2}}{T_{x_d}} \qquad C2_{x_d} = 1 \qquad T2_{x_d} = 1$$

Let's find the difference between these two new cumulative distributions over x2.

$$D2_{x2} := C2_{x2} - T2_{x2}$$

The maximum difference difference is given by

$$d2_{max} := \max(D_2)$$

and its location is

$$x_{d2} := LOC(D_2, d2_{max})$$

By the same reasoning as shown for the enhanced d-value, if we divide this maximum difference, $D2_{max}$, by $C2(x=x_{d2})$, we should have a reasonable estimate of $P(x=x_d)$, Pxd.

$$Pxd := \frac{d2_{max}}{C2_{x_{d2}}} \qquad Pxd = 8.064 \times 10^{-3} \qquad {}_{x_d} = 0.019$$

Note that our estimate Pxd is an estimate of P(x=xd). We can now plug this new estimate into the general cumulative distribution formula to better estimate the positive fraction.

$$\frac{D_{x_d} + Pxd}{C_{x_d}} = 0.077 \qquad pos = 0.089 \qquad pos_{ed} = 0.068$$

As we hoped, the estimate is approaching the actual pos value. Let's plug in the above logic into the equation to simplify the equation.

Expanding Dxd, we obtain,

$$\frac{C_{x_d} - T_{x_d} + Pxd}{C_{x_d}} = 0.077$$

Substituting,

$$\frac{C_{x_d} - T_{x_d} + \frac{\left(C2_{x_{d2}} - T2_{x_{d2}}\right)}{C2_{x_{d2}}}}{C_{x_d}} = 0.077$$

Further substituting,

$$\frac{C_{x_d} - T_{x_d} + \frac{\left(\frac{C_{x_{d2}}}{C_{x_d}} - \frac{T_{x_{d2}}}{T_{x_d}}\right)}{\frac{C_{x_{d2}}}{C_{x_d}}}}{C_{x_d}} = 0.077$$

Simplifying,

$$\frac{C_{x_d} - T_{x_d} + \frac{\left(C_{x_{d2}} - \frac{T_{x_{d2}} \cdot C_{x_d}}{T_{x_d}}\right)}{C_{x_{d2}}}}{C_{x_d}} = 0.077$$

Further simplifying,

$$1 - \frac{T_{x_d}}{C_{x_d}} + \frac{1}{C_{x_d}} - \frac{T_{x_{d2}}}{T_{x_d} \cdot C_{x_{d2}}} = 0.077$$

One could conceptionally add additional recursive elements to the above equation to improve the accuracy, but the errors will also likely increase.

$$pos_{sep} := 1 - \frac{T_{x_d}}{C_{x_d}} + \frac{1}{C_{x_d}} - \frac{T_{x_{d2}}}{T_{x_d} \cdot C_{x_{d2}}} \qquad pos_{sep} = 0.077$$

$$\varepsilon_{sep} := \frac{\left(pos_{sep} - pos\right) \cdot 100}{pos} \qquad \varepsilon_{sep} = -13.156$$

The other errors were

$\varepsilon_i = 66.284$

$\varepsilon_d = -29.629$

$\varepsilon_{ed} = -23.06$

$\varepsilon_{ns} = -17.645$