Minimum Required Data Length to Reconstruct GC Network

November 8, 2013

1 Task

Compute the minimum required data length when calculating GC. Use IF neural model as an example.

2 Analysis

Recall:

Distribution of GC obey (say there are two random variables x and y)

$$L \cdot \hat{F}_{x\to y} \overset{a}{\sim} \chi'^2(m,\ L \cdot F_{x\to y})$$

in the large L limit ($\overset{a}{\sim}$). Where L is data length, $\hat{F}_{x\to y}$ is calculation value of true GC $F_{x\to y}$, m is fitting order. χ'^2 is Noncentral chi-squared distribution (see ref. [1] for definition and properties).

For convenience, we denote the probability density function of $\hat{F}$ as $\rho_{m,L,F}(x)$.

Now we want to know: for a given L, m and true $F_{x\to y}$, $F_{y\to x}$, what's the expected correct rate? Obviously, 25% will be the lowest bound (random guess).

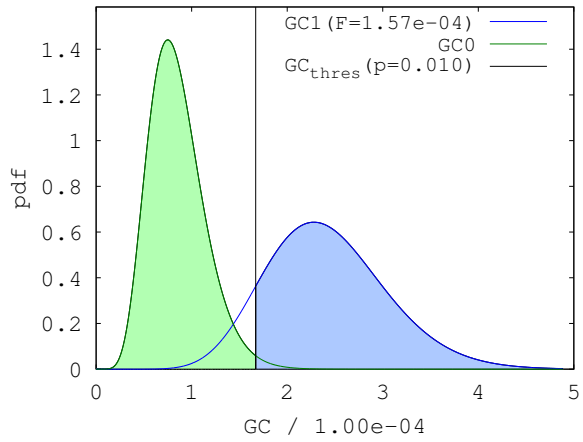

Figure 1: GC pdf under one set of typical parameter. The black line represent the GC thresholding value that we used to judge whether there is connection or not. $\mu = 1\,\text{kHz}$, $f = 0.012$, $S = 0.01$, using $L = 2\times10^5 (T = 1\times10^5\,\text{ms})$, $m = 17$, the true GC is $F_{x\to y} \approx 1.57\times10^{-4}$ and $F_{y\to x} \approx 0.01\times10^{-4}$ (obtained by $L = 1\times10^8$).

References

[1] http://en.wikipedia.org/wiki/Non-central_chi-square_distribution