The Q are still the same, since (1) Q is the prior process noise & (2) in real fitting, we always use no window smoother to select Q (so the Q is always the same for different methods).

See the code for grid calculation in <https://github.com/weigcdsb/COM_POISSON/blob/main/demo/smoother_fail/testLlhdGrid.m>.

The code for special cases (examples): <https://github.com/weigcdsb/COM_POISSON/blob/main/demo/smoother_fail/special_Cases.m>

The window size is selected by **forward chaining**.

Chart

Description automatically generated

All the results are average of 100 replications. In each replica, models are fitted by 1 training dataset and tested by 1 test dataset.

# Observations & (Possible) Explanation

1. The exact-Hessian smoother (denoted as “exact”) is not robust. When it gives positive-definite variance matrix, it will improve the Fisher-Hessian smoother (denoted as “Fisher”). To ensure the stableness, it’s necessary to use Fisher.
2. When parameters are constant, all models are good.
3. When there are large jumps in parameters, usually the “window” & “MAP”/ “NR” will improve the estimation. The improvement by “MAP” is more robust than “window” (Sometimes, the “window” will be worse than “Fisher”)
4. The improvement is related to mean firing rate (this is why the values of parameters will also influence the results, when there’s a jump).
   1. Since , then change in results in larger change in mean. Therefore, the change is more sensitive to change in .
   2. **In the case with very small mean firing rate, the “window” will beat “MAP”. Since with window, we are essentially enlarge the sample size at each step, by “cheating”/ “borrowing” from the future.**

# Questions

I don’t know if my current criteria, i.e. test log-likelihood per spike, is fair enough for different mean firing rate.

Maybe the values of llhd/spk for different dataset just make no sense (we can only compare performance of different models within one test dataset, but not for performance/ improvement of one model across different values)?

# Keep both and constant

|  |  |
| --- | --- |
|  |  |
|  |  |

It seems that if window size >1, the performance is always worse than the raw Fisher smoother. But the differences are small (~ 1e-4).

A special case: true beta = 2 & true gamma = 2.5

|  |  |
| --- | --- |
|  |  |

# Keep constant, change

When

|  |  |
| --- | --- |
|  |  |
|  |  |

A special case

|  |  |
| --- | --- |
|  |  |

When

|  |  |
| --- | --- |
|  |  |
|  |  |

A special case

|  |  |
| --- | --- |
|  |  |

When

|  |  |
| --- | --- |
|  |  |
|  |  |

A special case

|  |  |
| --- | --- |
|  |  |

# Keep constant, change

When

|  |  |
| --- | --- |
|  |  |
|  |  |

A special case

|  |  |
| --- | --- |
|  |  |

When

|  |  |
| --- | --- |
|  |  |
|  |  |

A special case

|  |  |
| --- | --- |
|  |  |

When

|  |  |
| --- | --- |
|  |  |
|  |  |

A special case

|  |  |
| --- | --- |
|  |  |

# Change both and

|  |  |
| --- | --- |
|  |  |
|  |  |

3 special cases (the last row)

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |