Reprot20200319

Linyi Guo

2020/3/19

Hi Professor,

The following report is basically what I have done after our last meeting.

Part 1: Decomposition

This part is to compare the errors among different methods and X-11. Since we mainly care about *trend* and *seasonal adjusted* series, I defined the error as

$$Error = \sum (Trend_{X-11} - Trend_{Object})^2 + \sum (Seasonal_{X-11} - Seasonal_{Object})^2$$

Because SA = Data - Seasonal, the sum of difference of seasonal series is equal to that of SA series.

Also, according to *Andrew Gelman(2006)*, I took the half-Cauchy distribution as a weakly-informative prior at the same time. The following table is the `summary` result:

SORRY, I realized the half-Cauchy they suggested is for standard error not the variance when I was writing the part 3, but I applied it as the prior for variance σ^2. So all results related to MAP_halfCauchy are incorrect.

```
##       MLE             LOSS            MAP         MAP_halfcauchy
##  Min.   : 172.5  Min.   :132.4  Min.   : 176.7  Min.   :173.1
##  1st Qu.: 341.9  1st Qu.:300.9  1st Qu.: 316.6  1st Qu.:331.2
##  Median : 405.8  Median :357.1  Median : 373.4  Median :386.6
##  Mean   : 422.6  Mean   :363.0  Mean   : 390.5  Mean   :402.4
##  3rd Qu.: 487.0  3rd Qu.:414.5  3rd Qu.: 444.3  3rd Qu.:460.5
##  Max.   :1012.3  Max.   :726.7  Max.   :1029.6  Max.   :941.5
```

As we can see, the error based on MLE is the worst and the result after considering the informative prior(the third column, MAP) is much better. As a contrast, the wearkly-informative prior half-Cauchy distribution also works but not as much as our informative prior. The second column here is just the result from the ideal variance using grid search and our loss function, which have one more term w.r.t. the derivation of trend component compared with the error defination above.

The following figures are the distributions of these errors and their differences:

Decomposition error comparison(X11 is standard)
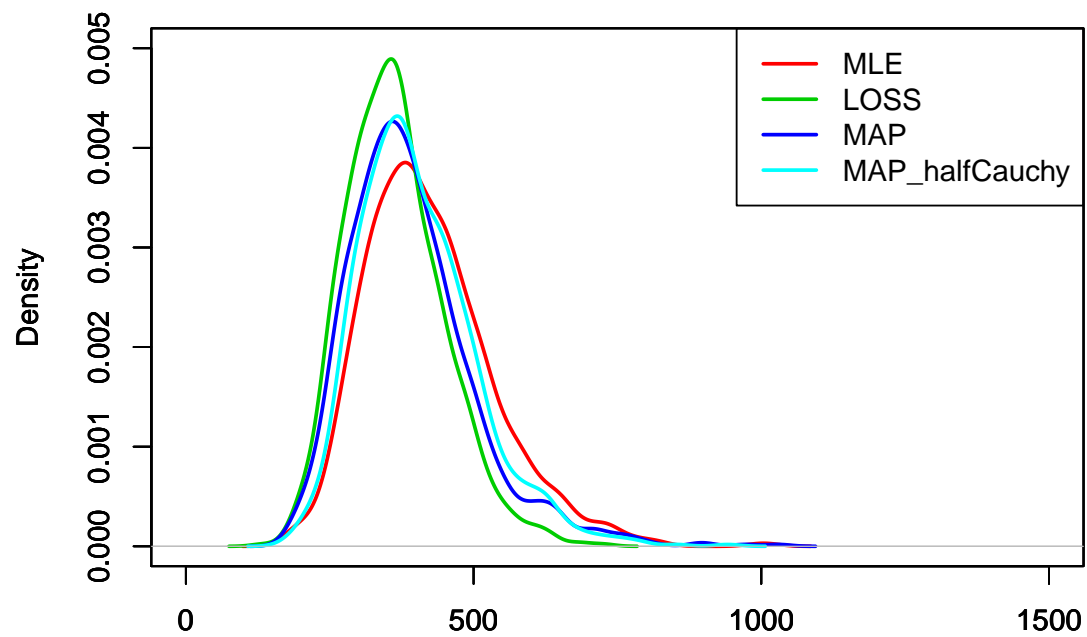
Distribution of differences among decomposition errors
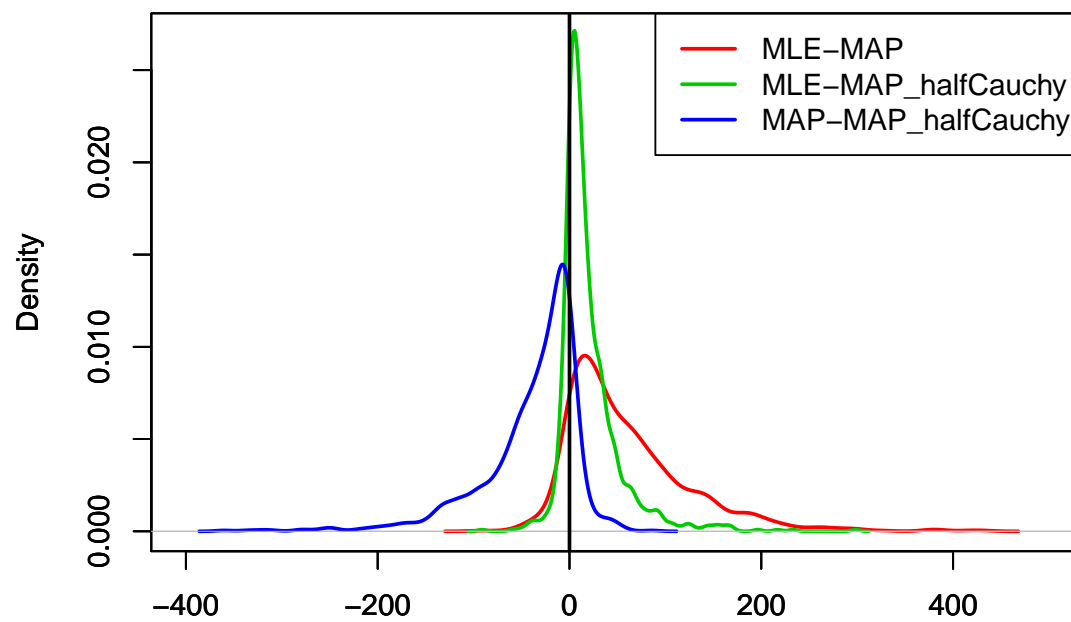
From the second figure, we can also derive that the decomposition from the informative prior is closer to that from X-11.

Part 2: Prediction

Now let's look at the behaviors of these methods for prediction. Still we shall use the simulated data. *Review*: the simulated data contains 1000 datasets, and each one's length is 15 years. The *Part 1: Decomposition* is build on the first 14 yrs, and the data of the last year is to serve the prediction part.

First, the summary of their prediction error:

```
##      x11                 MLE                 MAP                 map2
## Min.   :  85.71   Min.   :  73.0   Min.   :  68.69   Min.   :  69.84
## 1st Qu.: 316.29   1st Qu.: 298.3   1st Qu.: 293.66   1st Qu.: 288.82
## Median : 480.10   Median : 478.1   Median : 475.07   Median : 475.22
## Mean   : 735.43   Mean   : 619.9   Mean   : 621.93   Mean   : 618.39
## 3rd Qu.: 862.97   3rd Qu.: 706.2   3rd Qu.: 699.53   3rd Qu.: 699.47
## Max.   :6949.56   Max.   :5305.3   Max.   :5133.59   Max.   :5186.45
```

I DID NOT post the result of LOSS here.

When we compare these prediction errors, we may find that the differences among three state space models are not great, but X11 here really did a bad job. Although the result after considering our informative prior seems to be equivalent to that from weakly-informative prior, on my end it makes sense, because the information we used to build our prior is only related to the decomposition, we didn't consider the prediction error in our *LOSS* function. From this perspective, if we add this consideration into our loss function, the prediction could be better perhaps. But in my opinion, this is a little worthless to check since the differences between these SSMs are really small.

These are the figures of distributions of differences among these prediction errors:

Prediction error comparison
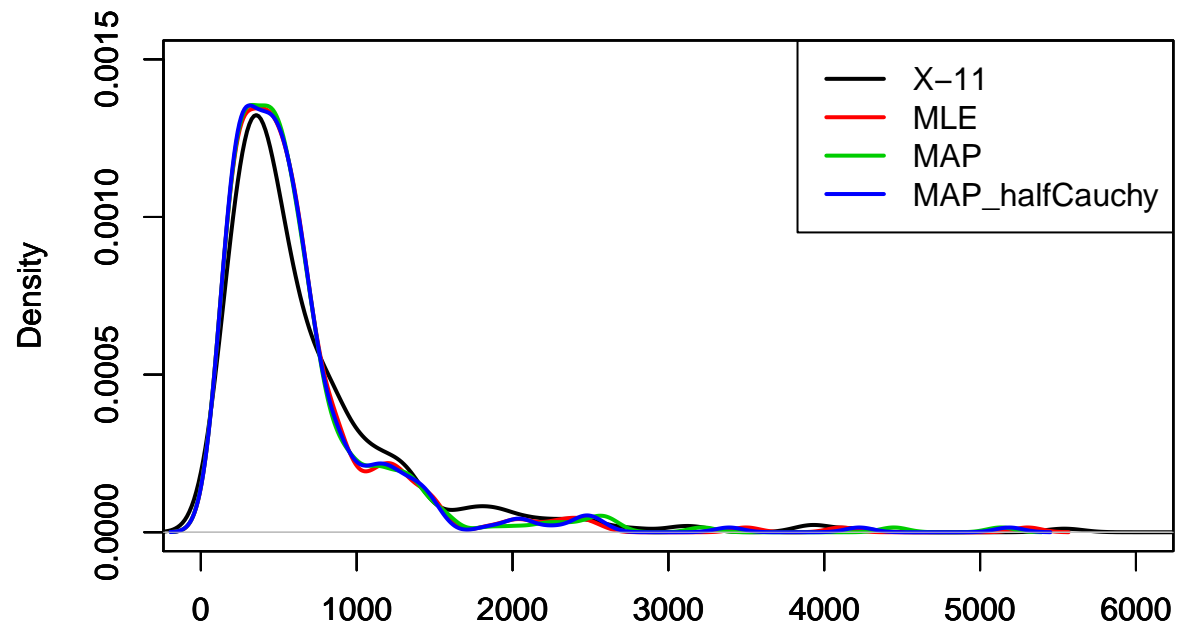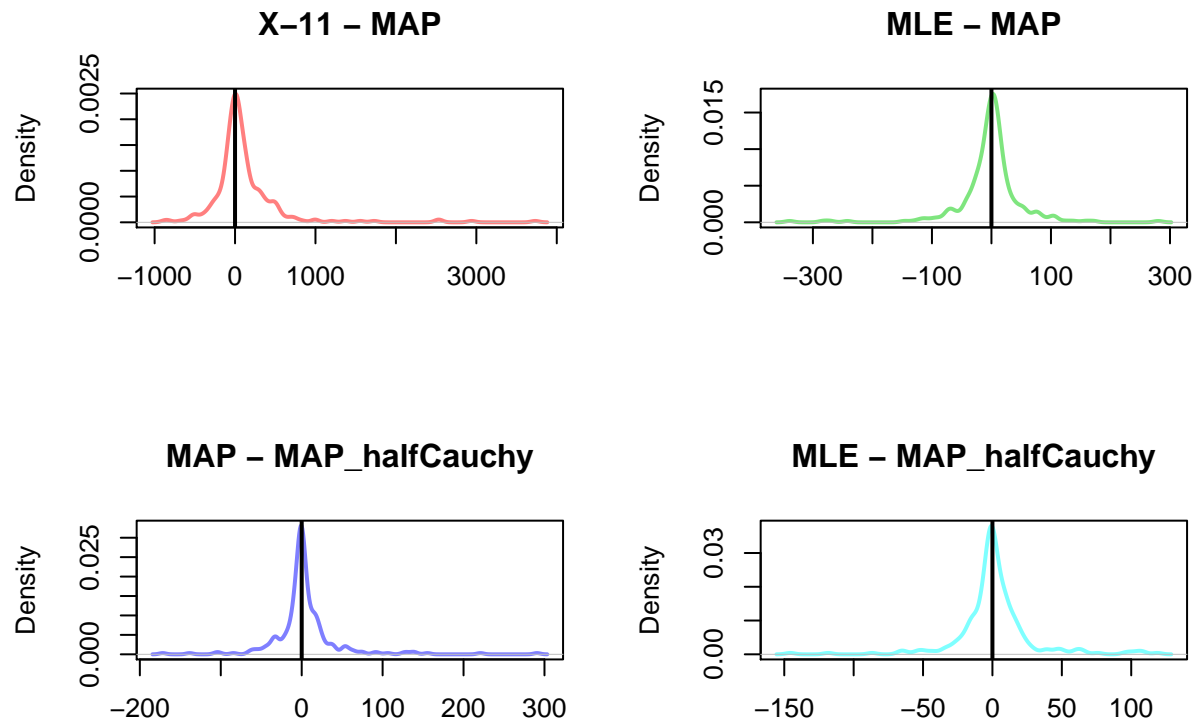
Part 3: Summary

We talked about the difference between SSM and hierarchical models last time. It seems that SSM is one special case where the number of clusters/groups equals to 1, and this is the reason why I chose half-Cauchy, cause Andrew Gelman and others suggest using half-t family under this case.

SORRY, I just realized the half-Cauchy they suggested is for standard error not the variance when I was writing this part, but I applied it as the prior for variance σ^2. So all results related to MAP_halfCauchy are incorrect. I will fix this ASAP.

I haven't moved to check the real data, but I will try them later this weeK once I finished the problem above :)