Judge Shopping Roadmap

March 17, 2021

Contents

1 What has been done **1**
- 1.1 Data . . . 1
  - 1.1.1 Sentencing Data . . . 1
  - 1.1.2 Schedule Data . . . 2
- 1.2 Parameter Estimation . . . 2
  - 1.2.1 Conviction Probability at Trial - θ . . . 2
  - 1.2.2 Expected Sentence Length if Convicted - τ . . . 3
  - 1.2.3 Defendant Cost of Trial - c_d . . . 3
  - 1.2.4 Judge maximum and minimum plea - $l_j(\theta\tau), u_j(\theta\tau)$ . . . 3
  - 1.2.5 County Arrival Rates - λ_c . . . 3
  - 1.2.6 Service Rates - μ_p, μ_t . . . 3
- 1.3 Simulation . . . 3
- 1.4 Analysis . . . 3

2 What remains to be done **4**
- 2.1 Data . . . 4
  - 2.1.1 Schedule Data . . . 4
- 2.2 Parameter Estimation . . . 4
  - 2.2.1 Conviction Probability at Trial - θ . . . 4
  - 2.2.2 Expected Sentence Length if Convicted - τ . . . 4
  - 2.2.3 Defendant Cost of Trial - c_d . . . 4
  - 2.2.4 Judge maximum and minimum plea - $l_j(\theta\tau), u_j(\theta\tau)$ . . . 4
  - 2.2.5 Service Rates - μ_p, μ_t . . . 4
- 2.3 Simulation . . . 5
- 2.4 Analysis . . . 5

1 What has been done

1.1 Data

1.1.1 Sentencing Data

Cleaning

- Dropped all sentencing events with Judge 1
- Changed dates to be in the yyyy-mm-dd format

- There were two statutes that were read in as dates, '12/21/90' and '12/21/92', all other statutes in the format 'XX-XX-XXXX'. There were two statutes in the STATA data that were '12-21-2790' and '12-21-2792'. We renamed the two 'date' statutes based on this.

Getting expmin variable from STATA dataset We noticed that the "expmin" variable in the STATA dataset was a function of the variables (Date,County,Circuit,Counts,Offense Seriousness,Offense Code, Offense Type, Sentence, Statute). In other words, sentencing events that had identical values for those variables had identical values of the "expmin" variable. So, given a sentencing event in the sentencing data, to determine its value of "expmin", I would find a sentencing event in the STATA data that had the same values for the variables mentioned above. There are 17,516 events in the STATA data with the "expmin" variable defined (the rest have missing values for it). Using this approach, we are able to find the "expmin" variable for 17,516 variables in our sentencing data.

Merging with schedule data We assign each Judge ID to the judge name in the schedule data whose schedule overlaps the most with that Judge ID. We consider the sentencing data and schedule data to overlap for a Judge ID and a Judge name in a specific week if the Judge ID in consideration has a sentencing event in a county that that judge name is assigned to in that week. So, if in Week 12 judge 'John Doe' is assigned to 'Aitken' county, and we see a sentencing event for Judge ID n in 'Aitken' county on Week 12, we would consider their schedules to be overlapping in that week. Note, this procedure yields a mapping in which Judge ID's are assigned to Judges based on alphabetical order.

1.1.2 Schedule Data

Cleaning

- We renamed Judge "COOPER" to "COOPER TW"
- We removed non-alphabetic characters from judge names
- We removed leading and trailing whitespace from judge names
- We parsed the judge assignments where multiple counties are listed and created a separate row in the data for each assignment. So, for example, if Judge Cooper was assigned to Williamsburg and Aitken county on Week 12, There would be one row showing only the Williamsburg assignment and another row showing only the Aitken assignment, both with the same week date.

1.2 Parameter Estimation

1.2.1 Conviction Probability at Trial - θ

We currently estimate this using logistic regression. We currently are using the following variables to predict this: Black, Offense Type, Offense Seriousness. **Note:** We are only using the defendants that went to trial to estimate this, there are only 270 defendants that went to trial in the data.

1.2.2 Expected Sentence Length if Convicted - τ

We currently estimate this using Negative Binomial Regression. We use the Cameron-Trivedi test for overdispersion to choose the overdispersion parameter. **Note:** We are only using the defendants that went to trial to estimate this, there are only 270 defendants that went to trial in the data.

1.2.3 Defendant Cost of Trial - c_d

We are currently estimating this using only the first method described in Nasser's document. In this method, we use the subset of cases where the sentence, s is less than $u_j(\theta\tau)$. In these cases, our model implies that $c_d = s - \theta\tau$.

1.2.4 Judge maximum and minimum plea - $l_j(\theta\tau), u_j(\theta\tau)$

We currently estimate this using a K-nearest neighbors approach. Given a defendant's $\theta\tau$, we find the K defendants in the judge's past pleas with the most similar values of $\theta\tau$. We then pick the minimum and maximum of these pleas to determine $l_j(\theta\tau), u_j(\theta\tau)$.

1.2.5 County Arrival Rates - λ_c

We set each county's arrival rate to be equal to the average number of sentencing events per week in that county, as observed in the data.

1.2.6 Service Rates - μ_p, μ_t

Service rates are currently being calculated as described in Nasser's document.

1.3 Simulation

- We first assign each county one judge for each time period. The time periods here are discrete and we think of them as weeks. There are 50 judges and 46 counties, in each week, we randomly draw 46 judges (without replacement) and assign them to a county in the order they are drawn. Counties are sorted alphabetically. So the first judge to be drawn is assigned to County A, the second to County B, and so on. Judges capacity is determined by our estimation of the plea service rate.
- In each time period, t, we iterate over the different counties. We simulate defendant arrivals for a given county, c, as follows: first, we determine the number of arrivals, n_{ct} by drawing from a Poisson distribution with mean $lambda_c$. We then draw n_{ct} defendants from county c's past defendants.
- Each defendant then chooses from the available judges that will be in county c in the next r weeks. The defendant chooses the judge, j that minimizes his expected cost, $\min(\theta\tau + c_d, u_j(\theta\tau)) + k(j)d$, where c_d is the defendant's cost of going to trial, $k(j)$ is the number of time periods until judge j will be in county c, and d is the cost of delay.
- Once a defendant chooses a judge, we reduce that judge's capacity for the week in which he will sentence that defendant.

1.4 Analysis

We currently only compute defendant backlog, and mean and variance.

2 What remains to be done

2.1 Data

2.1.1 Schedule Data

- Parse the days from the schedule data, so we know the exact dates judges were assigned to counties
- Parse the assignment types

2.2 Parameter Estimation

2.2.1 Conviction Probability at Trial - θ

- Look into Hurdle model used by Hester.
- Use more covariates to predict this.

2.2.2 Expected Sentence Length if Convicted - τ

- Look into Hurdle model used by Hester.
- Use more covariates to predict this.

2.2.3 Defendant Cost of Trial - c_d

- Figure out how to implement the MLE approach described in Nasser's document that uses information from sentencing events other than those in which $s < u_j$

2.2.4 Judge maximum and minimum plea - $l_j(\theta\tau), u_j(\theta\tau)$

- Look into convex hull approach.
- Remove each defendant from K nearest neighbors.

2.2.5 Service Rates - μ_p, μ_t

- Study documentation to refine exclusion criteria for "clean days".
- Try approach 1 for estimating plea service rate. Figure out how to enumerate the tuples of interest.
- Try approach 2 for estimating plea service rate.
- Try approach 3 for estimating plea service rate.
- Use updated plea service rate estimate to re-estimate trial service rate.

2.3 Simulation

- Confirm that current way of simulating arrivals works.
- Fix the processing of backlogged defendants.
- Consider only assigning judges to the counties they are assigned to in the data. Or taking into consideration geography when assigning judges.
- Incorporate defendants' decision to go to trial.
- Incorporate capacity reductions for judges from time spent on trials.
- Consider assigning more than one judge per week for busier counties.

2.4 Analysis

- Think about what other metrics to compute.
- Consider analyzing county level outcomes (eg intra and inter county variance).