stochastics is a little Java library I've put together to do estimation, simulation and prediction of a class of 'self-exciting' stochastic processes called Hawkes processes.

For example, you can download SPY.mat which is the data corresponding to a marked point process of trades of the SPDR S&P 500 ETF on a particular 6.5 hour trading day sometime in 2016 and run the program ProcessEstimator and you will see something like this

```
Estimating parameters for SPY.mat
E[dt]=180.0731913776857     <Average time between trades across the entire 6.5 hour session , in units of milliseconds )
E_0[dt]=101.08133459977739    <Average time vetween trades of the first 30 minutes of regularly scheduled trading, in units of milliseconds )
ff[31;107mspawning ff[95m24ff[32m ExtendedApproximatePowerlawSelfExcitingProcessff[31;107mes to estimate the model parameters ff[94m[,,,bff[31;107m] most likely to have generated the observed sequence of ff[95m17804ff[31;107m timestampsff[m
ff[31m          thread-13       #1/24  Storing 1.0000000000 0.0000000000 3.0255799155 1.7241985792  with LL score 1718889.4428270236ff[39m
ff[31m          thread-3        #3/24  Storing 1.0000000000 0.0000000000 3.0251966081 1.7246375499  with LL score 1718889.442804567ff[39m
ff[31m          thread-15       #2/24  Storing 1.0000000000 0.0000000000 3.0255466416 1.7242359036  with LL score 1718889.4428268662ff[39m
ff[31m          thread-8        #5/24  Storing 1.0000000000 0.0000000000 3.0295636875 1.7206915112  with LL score 1718889.4412852346ff[39m
ff[31m          thread-6        #4/24  Storing 1.0000000000 0.0000000000 3.0255781444 1.7241991514  with LL score 1718889.4428270238ff[39m
..........
ff[31m          thread-8        #22/24  Storing 1.0000000000 0.0000000000 3.0255731777 1.7242194415  with LL score 1718889.442826963ff[39m
ff[31m          thread-2        #21/24  Storing 1.0000000000 0.0000000000 3.0255960057 1.7241934397  with LL score 1718889.4428270042ff[39m
estimation completed in 0.542400 minutes at 74.084317 evals/sec
parameter estimates for ExtendedApproximatePowerlawSelfExcitingProcess{=1.0,=0.0,=3.0295636875196883,b=1.7206915111784111,Z=20.212944519689493,Edt=135.725254656012}
 ---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
 |  |  |                  | b                  | Log-Lik            | KS(Lambda)              | mean(Lambda)            | var(Lambda)             | MM(Lambda)              | LB(Lambda)              | MMLB(Lambda)            | E[dt]            |
 |=======================================================================================================================================================================================
1. | 1.0| 0.03.0295636875196883| | 1.7206915111784111| 1718889.4412852346| 0.866404942169227 | 0.9994319031632158| 0.9601678843210508| 0.02105209014603504 | 1300.3947948630428| 0.15096857806824288| 135.725254656012  |
2. | 1.0| 0.0| 3.0255558095751067| 1.7241933312893514| 1718889.4428269528| 0.8663128815718412| 0.9994320070964131| 0.9597795977901094| 0.02124602560415001 | 1299.8196125605334| 0.15234993419140458| 135.67557364607737|
3. | 1.0| 0.0| 3.0255794239386953| 1.7241903564322218| 1718889.4428270115| 0.8663131085461994| 0.9994320079874562| 0.9597784615931496| 0.021246591921049607| 1299.8388017200318| 0.15235430852694057| 135.67514849842505|
4. | 1.0| 0.0| 3.0255810148945113| 1.724191059476018 | 1718889.442827015 | 0.8663131076913555| 0.9994320081242861| 0.9597782101066586| 0.021246717390713044| 1299.8409012782388| 0.15235524253260319| 135.67508318612306|
5. | 1.0| 0.0| 3.025566171530826 | 1.7242003859853978| 1718889.4428270126| 0.8663128317578   | 0.9994320081980411| 0.959777480464975 | 0.021247082064086764| 1299.8354973653923| 0.15235776925594738| 135.67504778074579|
,,,,,,,,,
19. | 1.0| 0.0| 3.0255808158401374| 1.724205552880895 | 1718889.4428270133| 0.8663128472027217| 0.9994320093466433| 0.9597754181804553| 0.021248110909794615| 1299.8536587223991| 0.15236544351290007| 135.67449954231628|
20. | 1.0| 0.0| 3.025574541481823 | 1.7242116484463734| 1718889.4428270014| 0.8663126920826797| 0.9994320095608489| 0.9597746928372614| 0.0212484731531023  | 1299.853297077291 | 0.15236803517130917| 135.6743971852654 |
21. | 1.0| 0.0| 3.025613200916328 | 1.7241987316644454| 1718889.4428269141| 0.8663132074745514| 0.9994320103355935| 0.9597743908429659| 0.02124862260120053 | 1299.8775285846236| 0.15236950263288776| 135.67402776061152|
22. | 1.0| 0.0| 3.0255731776959305| 1.724219441518698 | 1718889.442826963 | 0.8663125427688295| 0.9994320101572236| 0.959773282760179 | 0.021249176999232344| 1299.858994166482 | 0.1523731753546449 | 135.67411241608536|
23. | 1.0| 0.0| 3.02554664157473  | 1.7242359035798402| 1718889.4428268662| 0.8663120532459874| 0.9994320102710825| 0.9597720193294557| 0.021249808486941024| 1299.8491439720772| 0.15237754270850248| 135.6740577131032 |
24. | 1.0| 0.0| 3.025196608086467 | 1.7246375499354134| 1718889.442804567 | 0.8663022985818155| 0.9994320274511848| 0.9597196394329696| 0.02127596408473742 | 1299.8839023458731| 0.15256566703137792| 135.66585037341952|

writing timestamp data, compensator, intensity, and innovation to stochastics/test0.mat and parameters to stochastics/test0.mat.eapl.model
```

The output of this process is two files called 'test0.mat' and 'test0.mat.eapl.model', the first one is a matlab compatible file that has the input data, along with the estimated intensity and compensator of the process which can be tested for goodness-of-fit and verifying certain hypothesis about the data such as the compensator being a unit-rate Poisson process with no auto-correlation, that is, if the model is a good fit to the data then the mean and variance of the variable '**compensator**' (otherwise denoted Λ since matlab doesn't support UTF characters in variable names) in test0.mat are both equal to 1 and there will be no detectable autocorrelation for any lags other than 0. These tests are determined by comparing the statistics of each of the candidate solutions

#	τ	ε	η	b	$\text{Log}-\text{Lik}$	$\text{KS}(\Lambda)$	$\text{mean}(\Lambda)$	$\text{var}(\Lambda)$	$MM(\Lambda)$
1	1.0	0.0	3.0255794239386953	1.7241903564322218	1718889.4428270115	0.8663131085461994	0.9994320079874562	0.9597784615931496	0.021246591921049607
2	1.0	0.0	3.0255810148945113	1.724191059476018	1718889.442827015	0.8663131076913555	0.9994320081242861	0.9597782101066586	0.021246717390713044
3	1.0	0.0	3.025566171530826	1.7242003859853978	1718889.4428270126	0.8663128317578	0.9994320081980411	0.959777480464975	0.021247082064086764
4	1.0	0.0	3.025560746696718	1.7242055316875902	1718889.4428270059	0.8663126998659925	|0.9994320083726566	0.9597768774468013	0.02124738322404185
...	....	...	...	...	...	...	...	...	...
24	1.0	0.10164...	0.1	0.0	1718423.379808068	0.8685636719501787	0.9995379499758154	1.0588055428354197	0.029402878162822277

The column-labels indicate the parameters and summary statistics of a set of candidate solutions which are local minimima of the likelihood surface of the data given the parameters $\tau, \varepsilon, \eta, b$ and they are ranked in order of best fit to worst fit according to the column MMLB(Λ). The summary statistics are

- Log-Lik=$\ln\mathcal{L}(N(t)_{t\in[0,T]})$ is the logarithm of the likelihood score which is defined in closed-form by

$$\begin{aligned}\ln\mathcal{L}(N(t)_{t\in[0,T]}) &= \int_0^T (1-\lambda(s))\mathrm{d}s + \int_0^T \ln\lambda(s)\mathrm{d}N_s \\ &= T - \int_0^T \lambda(s)\mathrm{d}s + \int_0^T \ln\lambda(s)\mathrm{d}N_s\end{aligned}$$

- KS(Λ) is actually 1 minus the Kolomogorov-Smirnov statistic

$$D_n = \sup_x |F_n(x) - F(x)|$$

- mean(Λ)=$\frac{1}{n}\sum_{i=1}^n \Lambda_i$

- var(Λ) = $\frac{1}{n}\sum_{i=1}^n (\Lambda_i - \text{mean}(\Lambda))^2$

- MM(Λ) is a moment-matching measure defined by

$$MM(\Lambda) = |\text{mean}(\Lambda) - 1| + |\text{var}(\Lambda) - 1|$$

- LB(Λ) is the Ljung-Box statistic, which tests for serial autocorrelation, defined by

$$Q = n(n+2)\sum_{k=1}^{h} \frac{\hat{\rho}_\kappa^2}{n-k}$$

where $\hat{\rho}_k$ is the sample autocorrelation at lag k given by

$$\hat{\rho}_k = \frac{1}{(n-k)}\sum_{i=1}^{n-k} \Lambda_i \Lambda_{i+k}$$

- MMLB(Λ) is a combination moment-matching autocorrelation minimizing metric

$$\text{MMLB}(\Lambda) = MM(\Lambda)\log(1 + \text{LB}(\Lambda))$$

- and finally $E[\text{dt}]$ is the unconditional expected mean time between points of the process

$$\begin{aligned} E[\text{dt}] &= \int_0^\infty t f(t)\text{d}t \\ &= \int_0^\infty \frac{t}{Z}\sum_{j=1}^{P} \alpha_j e^{-\beta_j t}\text{d}t \\ &= \frac{\sum_{j=1}^{P} \gamma(j,2)}{(\prod_{j=1}^{P} \beta_j)(\sum_{j=1}^{P} \gamma(j,1))Z} \end{aligned}$$

where

$$\gamma^n(k) = \prod_{j=1}^{P} \begin{cases} \alpha_j & j = k \\ \beta_j^n & j \neq k \end{cases} \tag{1}$$

and

$$\gamma(k) = \gamma^1(k) \tag{2}$$

when n is not denoted and the exponential powerlaw weights α_j and β_j are a function of the parameters $\tau, \varepsilon, \eta, b$ defined by

$$\alpha_j = \begin{cases} (\tau m^i)^{-(1+\varepsilon)} & j < M \\ b & j = P \end{cases} \tag{3}$$

$$\beta_j = \begin{cases} \tau m^i & j < P \\ \eta^{-1} & j = P \end{cases}$$

and

$$Z = \sum_{j=1}^{P} \frac{\alpha_j}{\beta_j}$$

with $P = M + 1$ and $M = 15$ and $m = e^{\frac{\ln(60000)}{M}}$ so that the exponential powerlaw-approximation covers about 15 minutes with a resolution of 1 minute. The additional weight b to allow for a smooth-drop to zero as described in Critical reflexivity in financial markets: a Hawkes process analysys

All times are in units of milliseconds unless otherwise specified. As we can see, the theoretical mean of the estimated model is 135.67..ms which is not far from the emperical sample mean of $E_0[\mathrm{dt}]=101.08133459977739$ which is the sample mean of the mean time between trades in the first 30 minutes. The mean and variance of the compensator is also very close to 1, and the autocorrelation is relatively small compared to the input data.