The Journal of Open Source Software

ppdiag: Diagnostic Tools for Temporal Point Processes

Sally Sun, Owen G. Ward, Xiaoxi Zhao, Jing Wu, and Tian Zheng

Columbia University

DOI:

Software

- Review
- Repository
- Archive

Submitted:
Published:

Summary

Temporal Point Processes are widely used to model phenomena in many fields, such as finance and neuroscience. However, tools to evaluate the fit of these point processes to data, and to identify reasons for lack of fit, are not readily implemented in common software. Here we provide `ppdiag`, an `R` package containing a selection of statistically motivated tools to analyse the goodness of fit of point processes to data, as have been utilised in Wu, Smith, & Zheng (2020).

Statement of Need

This package provides functions to evaluate the fit of univariate temporal point processes. These functions allow:

- Simulation of data from a range of common univariate point processes including Homogeneous Poisson Process, Hawkes Process, and Markov-modulated Hawkes Process.
- Fitting common univariate point processes to data, and plotting the intensity function over data.
- After fitting a point process model to data, evaluating the ability of that model to capture the temporal structure present in data. Methods for diagnostics include raw and Pearson residuals, a Kolmogrov-Smirnov test and corresponding diagnostic plots.

Demonstration

We illustrate the use of this functionality with a simple example of readily available email data (Çetinkaya-Rundel et al., 2020).

```
library(ppdiag)
library(openintro)
library(lubridate)
first <- ymd_hms(email$time[1],tz="EST")
second <- ymd_hms(email$time[2],tz="EST")
time <- (as.duration(interval(first,
                    ymd_hms(email$time[-1],tz="EST")))/3600)[1:300]
time_data <- sort(unique(time))
```

We can fit a homogeneous Poisson process and look at the goodness of fit of this model to the data.

```
email_hpp <- fithpp(time_data)
pp_diag(email_hpp, events = time_data)
#> Please input the right model. Select from hp, hpp, mmpp and mmhp.
```

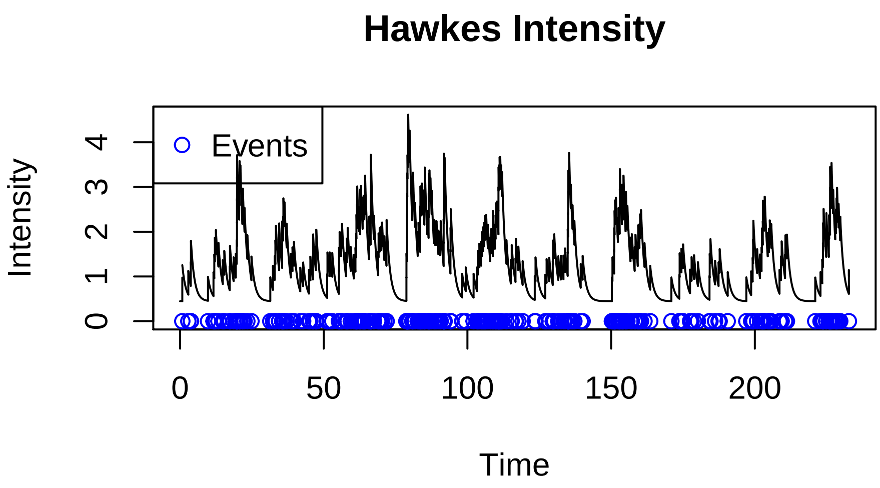
Similarly, we can fit a self exciting Hawkes process to this data and examine the results of that fit.

```
email_hp <- fithp(events = time_data)
pp_diag(email_hp, time_data)
#> Please input the right model. Select from hp, hpp, mmpp and mmhp.
```

Examining the results of the Kolmogrov-Smirnov test, based on the time rescaling theorem (Brown, Barbieri, Ventura, Kass, & Frank, 2002), indicates that a Hawkes process better describes these events.

We can then examine the estimated intensity of this Hawkes process to this data.

```
drawHPIntensity(email_hp, events = time_data,
                plot_events = TRUE)
```



intensity-1.bb

Comparison with Other Packages

Though there are some packages that contain functions for simulating point processes, (Zaatour, 2014),(Harte, 2017), to the best of our knowledge, there are no existing packages that provide methods for diagnostics of different temporal point processes and for comparing the fit of these point processes.

References

Brown, E. N., Barbieri, R., Ventura, V., Kass, R. E., & Frank, L. M. (2002). The time-rescaling theorem and its application to neural spike train data analysis. *Neural computation*, *14*(2), 325–346. doi:10.1162/08997660252741149

Çetinkaya-Rundel, M., Diez, D., Bray, A., Kim, A., Baumer, B., Ismay, C., & Barr, C. (2020). *Openintro: Data sets and supplemental functions from 'OpenIntro' textbooks and labs*. Retrieved from https://CRAN.R-project.org/package=openintro

Harte, D. (2017). *HiddenMarkov: Hidden markov models*. Retrieved from https://CRAN.R-project.org/package=HiddenMarkov

Wu, J., Smith, A. L., & Zheng, T. (2020). Diagnostics and visualization of point process models for event times on a social network. *arXiv preprint arXiv:2001.09359*.

Zaatour, R. (2014). *Hawkes: Hawkes process simulation and calibration toolkit*. Retrieved from https://CRAN.R-project.org/package=hawkes