ppdiag: Diagnostic Tools for Temporal Point Processes

Sally Sun[1]**, Owen G. Ward**[1]**, Jing Wu**[1]**, Lihao Xiao**[1]**, Xiaoxi Zhao**[1]**, and Tian Zheng**[1,2]

1 Department of Statistics, Columbia University **2** Data Science Institute, Columbia University

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 described in Wu et al. (2021).

Statement of Need

Temporal point processes can be used to model data arising in a wide variety of fields, including but not limited to neuroscience, seismology and finance. As such, tools for analysing the fit of point processes to data can be used by many practitioners. 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 and Poisson Processes.

- 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 computing raw and Pearson residuals, a Kolmogrov-Smirnov test (based on the time rescaling theorem) and corresponding diagnostic plots. These diagnostics can be used to compare the fit of multiple point processes to data and to identify reasons for lack of fit of a process to data.

Though there are some existing packages that contain functions for simulating point processes, (Zaatour, 2014),(Harte, 2017), and for computing some simple residuals (Harte, 2010), 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. These diagnostic tools are useful to identify problems in model fitting and understand the causes of this lack of fit.
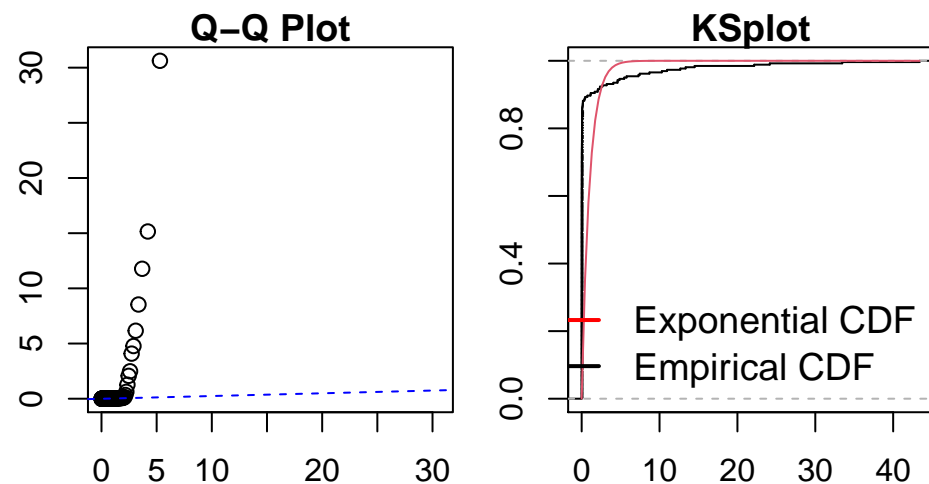
Example

We illustrate the use of this functionality with a simple example, using earthquake aftershock data (Harte, 2010). We consider all after shocks within a year of the original Phuket earthquake and wish to see if they can be described by a temporal point process.

```
set.seed(2021)
library(ppdiag)
library(PtProcess)
data("Phuket")
end <- 365
event_times <- Phuket$time[Phuket$time < end]
```

We can fit a homogeneous Poisson process and look at the goodness of fit of this model to the data. From the resulting diagnostics it is clear that a homogeneous Poisson process is not suitable for modelling the temporal patterns in this data.

```
shock_hpp <- fithpp(event_times, end = end)
pp_diag(shock_hpp, events = event_times)
```



```
#>
#> Raw residual: 0.07176529
#> Pearson residual: 0.08470523
#>
#>  One-sample Kolmogorov-Smirnov test
#>
#> data:  r
#> D = 0.78626, p-value < 2.2e-16
#> alternative hypothesis: two-sided
```

We can fit a self exciting Hawkes process to this data and examine the results of that fit. We see that from our diagnostic tools a Hawkes process is much more capable of describing the temporal structure of this data.

```
shock_hp <- fithp(events = event_times, end = end)
pp_diag(shock_hp, events = event_times)
```

Q–Q Plot

KSplot

Exponential CDF
Empirical CDF
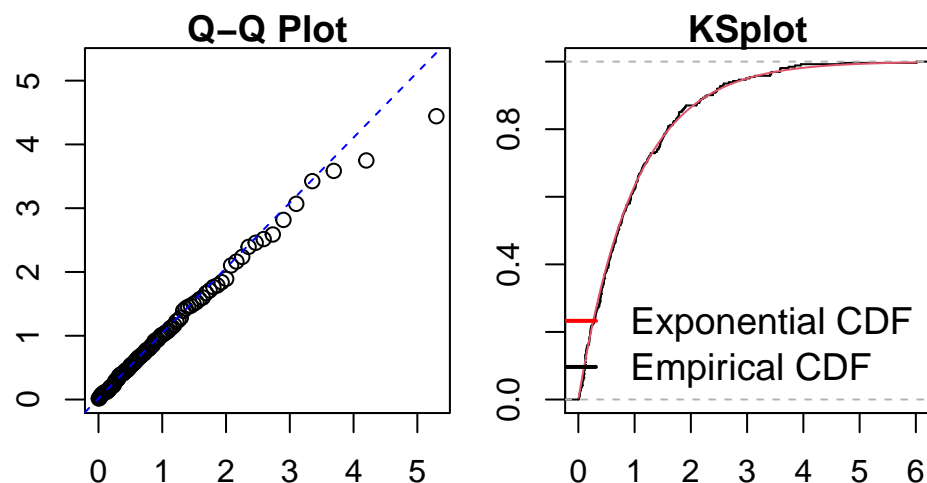
```
#> Raw residual: 1.026715
#> Pearson residual: -3.715193
#>
#>  One-sample Kolmogorov-Smirnov test
#>
#> data:  r
#> D = 0.0357, p-value = 0.8922
#> alternative hypothesis: two-sided
```

References

Harte, D. (2010). PtProcess: An R package for modelling marked point processes indexed by time. *Journal of Statistical Software*, *35*(8), 1–32. https://doi.org/10.18637/jss.v035.i08

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

Wu, J., Smith, A. L., & Zheng, T. (2021). Diagnostics and visualization of point process models for event times on a social network. In *Applied modeling techniques and data analysis 1* (pp. 129–145). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119821588.ch7

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