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).

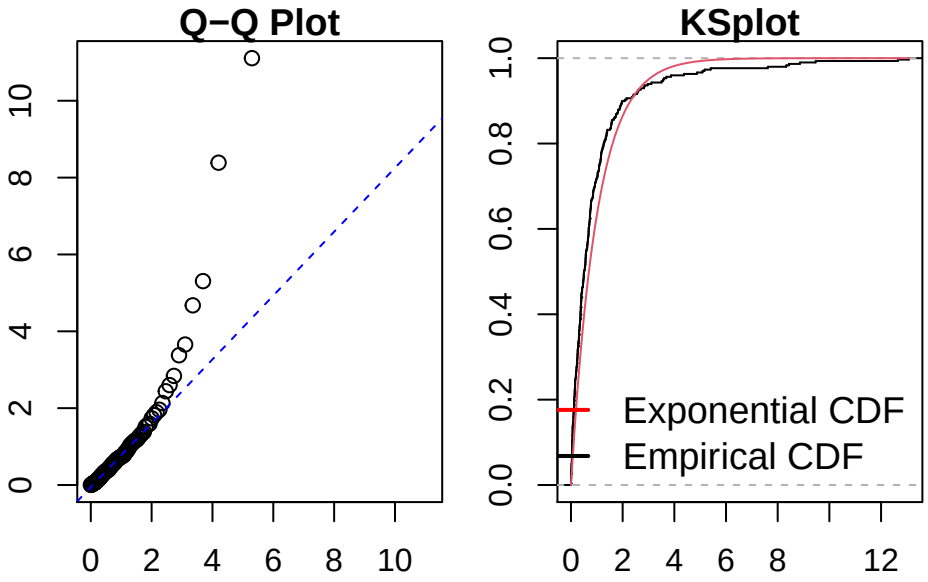
Example

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)
#> Loading required package: airports
#> Loading required package: cherryblossom
#> Loading required package: usdata
library(lubridate)
#>
#> Attaching package: 'lubridate'
#> The following objects are masked from 'package:base':
#>
#>     date, intersect, setdiff, union
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)
diagpp(email_hpp, events = time_data)
#> Warning in ks.test(r, "pexp"): ties should not be present for the Kolmogorov-
#> Smirnov test
```

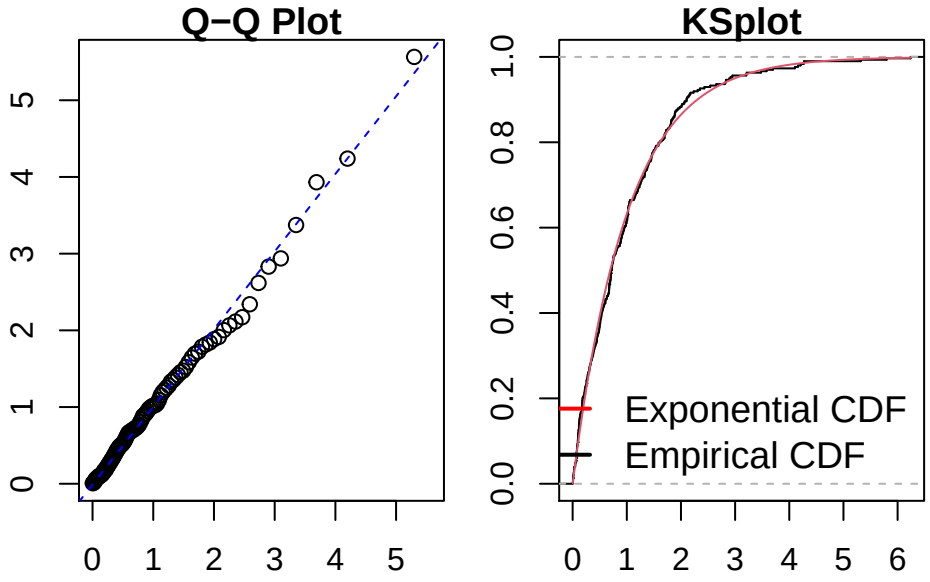


hpp-1.bb

```
#>
#> Raw residual: -1.012849
#> Pearson residual: -0.8935674
#>
#>  One-sample Kolmogorov-Smirnov test
#>
#> data:  r
#> D = 0.12391, p-value = 0.0002123
#> alternative hypothesis: two-sided
```

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)
diagpp(email_hp, time_data)
```



hawkes-1.bb

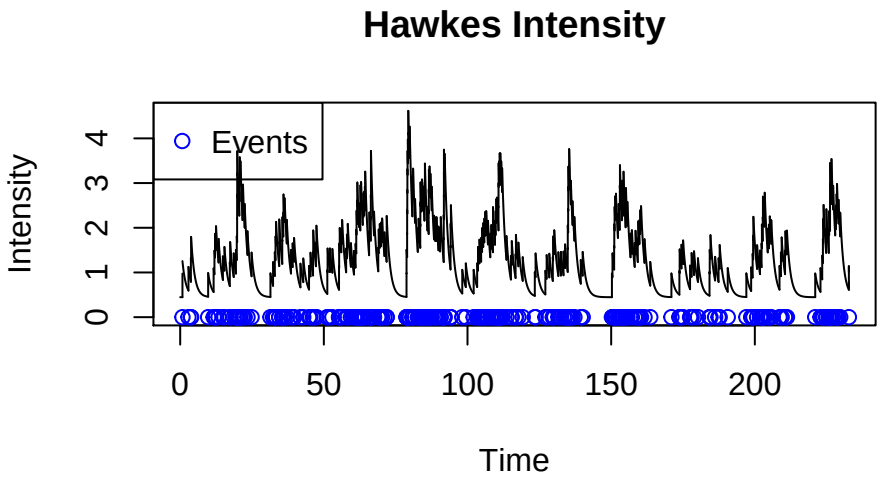
```
#> Raw residual: 0.0001270847
#> Pearson residual: 1.560215
#>
```

```
#>  One-sample Kolmogorov-Smirnov test
#>
#> data:  r
#> D = 0.041649, p-value = 0.6795
#> alternative hypothesis: two-sided
```

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)
#> Using the hp object. Set fit=TRUE to fit events provided.
```



intensity-1.bb

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.

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

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*.