Circular statistics for modelling seasonal patterns

Jolien Cremers

September 8th, 2021

Why?

Influenza

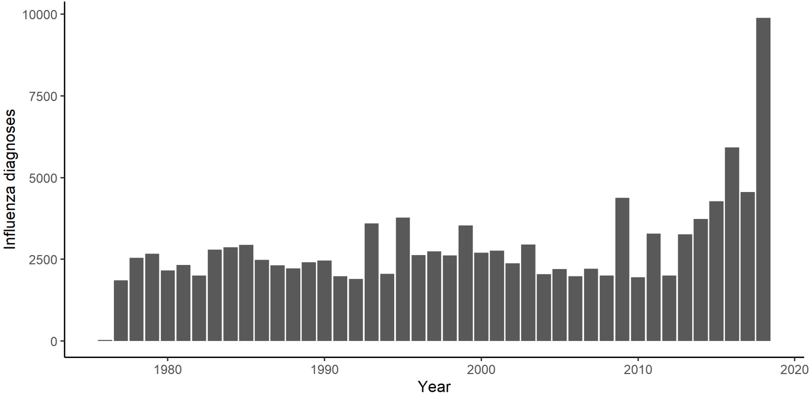


Figure 1: Influenza diagnoses per year in Denmark (altered dataset).

Standard approach

log-linear Poisson regression model:

$$\log(\lambda_t) = \alpha + \beta_1 \cos(\theta_t) + \beta_2 \sin(\theta_t),$$

- θ_t angles corresponding to each day of the year ($\frac{2\pi t}{365} \mod 2\pi$)
- λ_t = number of cases for each timepoint $t = 1, \ldots, 365$ (day of the year).

Peak-to-trough ratio (relative risk):

$$\exp(2\sqrt{\beta_1^2 + \beta_2^2})$$

Peak date:

$$\frac{\text{atan2}(\beta_2, \beta_1) * 365}{2\pi}$$

Poisson regression

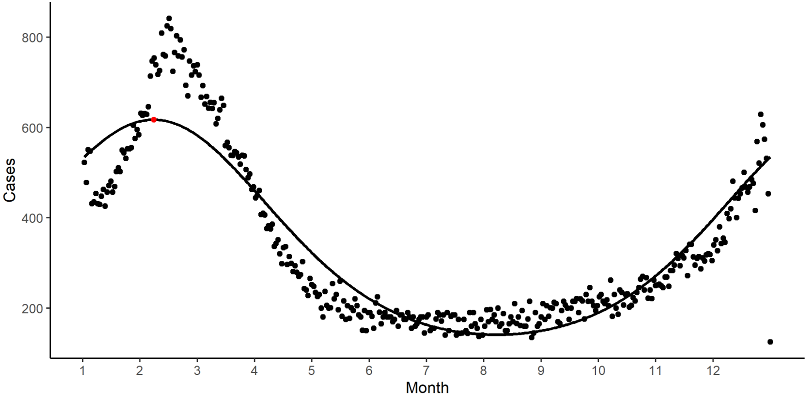


Figure 2: Aggregated daily counts of influenza cases (dots) in Denmark together with the estimated seasonal trend (black line) from a Bayesian Poisson model with seasonal trend. The red dot indicates the peak time as estimated by the model (7th of February).

Circular

Figure 3: Circular plot of a random sample influenza cases in Denmark (left) together with a von Mises kernel density function fit to these data (right).

Converting time-series to circular data

$$\theta = \frac{2\pi t}{P} \mod 2\pi$$

- t = timepoints
- P = period , e.g. 365 for yearly periodicity

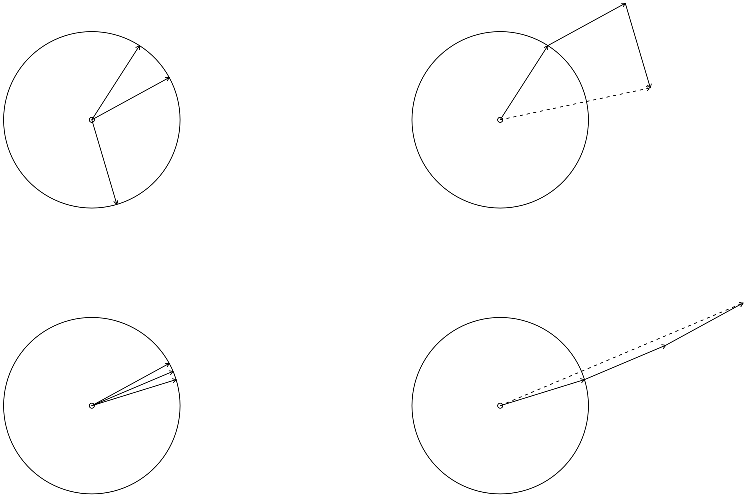
Circular summary statistics

$$\bar{\theta} = \begin{cases} \tan^{-1}\left(\frac{\sum_{i=1}^{n}\sin\theta_i}{\sum_{i=1}^{n}\cos\theta_i}\right) & \sum_{i=1}^{n}\sin\theta_i > 0, \sum_{i=1}^{n}\cos\theta_i > 0 \\ \tan^{-1}\left(\frac{\sum_{i=1}^{n}\sin\theta_i}{\sum_{i=1}^{n}\cos\theta_i}\right) + \pi & \sum_{i=1}^{n}\cos\theta_i < 0 \\ \tan^{-1}\left(\frac{\sum_{i=1}^{n}\sin\theta_i}{\sum_{i=1}^{n}\cos\theta_i}\right) + 2\pi & \sum_{i=1}^{n}\sin\theta_i < 0, \sum_{i=1}^{n}\cos\theta_i > 0 \end{cases}$$

$$\bar{R} = \frac{1}{N}\sqrt{\left(\sum_{i=1}^{N}\cos\theta_i\right)^2 + \left(\sum_{i=1}^{N}\sin\theta_i\right)^2}$$

see Fisher (1995) for an overview of other summary statistics.

Visualisation



The model

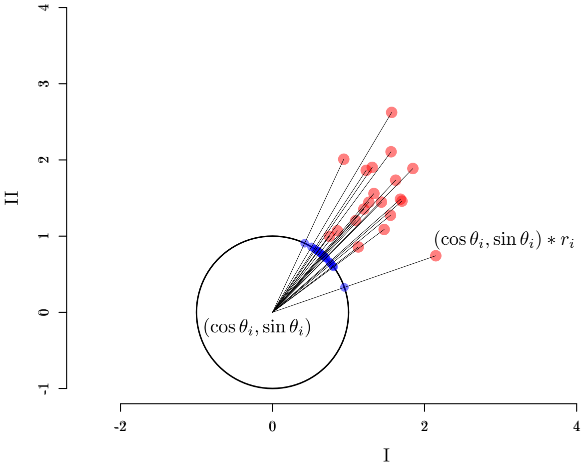
Projected normal distribution

$$PN(\theta \mid \boldsymbol{\mu}, \boldsymbol{I}) = \frac{1}{2\pi} e^{-0.5||\boldsymbol{\mu}||^2} \left[1 + \frac{\boldsymbol{u}'\boldsymbol{\mu}\Phi(\boldsymbol{u}'\boldsymbol{\mu})}{\phi(\boldsymbol{u}'\boldsymbol{\mu})}\right]$$

- $0 < \theta \leq 2\pi$
- $\boldsymbol{u} = (\cos\theta, \sin\theta)$
- $\boldsymbol{\mu} = \boldsymbol{B}'\boldsymbol{x} =$ mean vector (regression)
- $\boldsymbol{B} = [\beta^{\boldsymbol{I}}, \beta^{\boldsymbol{II}}]$
- $\boldsymbol{I} =$ variance-covariance matrix
- $\Phi(\cdot)$ and $\phi(\cdot)$ are standard normal cdf and pdf.

Visualisation

Underlying bivariate normal datapoint $\boldsymbol{Y}_i = (\cos\theta_i, \sin\theta_i) r_i$, $r_i = ||\boldsymbol{Y}_i||$ is latent



Transformation of coefficients (categorical x)

For a variable x with k categories, $\hat{\theta}_k = \text{atan2}\left(\beta_0^{II} + \beta_k^{II},\ \beta_0^{I} + \beta_k^{I}\right)$

$$\frac{\text{atan2}\left(\beta_0^{II} + \beta_k^{II},\ \beta_0^{I} + \beta_k^{I}\right) * 365}{2\pi}$$

$$\begin{aligned}
\text{atan2}(Y^{II}, Y^{I}) &= \arctan\left(\frac{Y^{II}}{Y^{I}}\right) && \text{if} \quad Y^{I} > 0 \\
&= \arctan\left(\frac{Y^{II}}{Y^{I}}\right) + \pi && \text{if} \quad Y^{I} < 0 \quad \& \ Y^{II} \geq 0 \\
&= \arctan\left(\frac{Y^{II}}{Y^{I}}\right) - \pi && \text{if} \quad Y^{I} < 0 \quad \& \ Y^{II} < 0 \\
&= +\frac{\pi}{2} && \text{if} \quad Y^{I} = 0 \quad \& \ Y^{II} > 0 \\
&= -\frac{\pi}{2} && \text{if} \quad Y^{I} = 0 \quad \& \ Y^{II} < 0 \\
&= \text{undefined} && \text{if} \quad Y^{I} = 0 \quad \& \ Y^{II} = 0 \qquad (1)
\end{aligned}$$

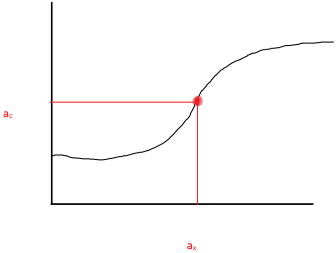
Peak-to-trough ratio (categorical x)

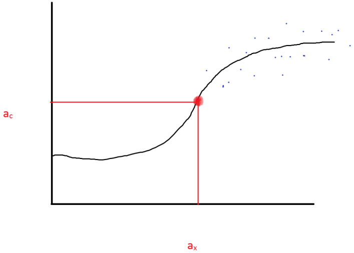
$$\mathrm{PTR}_k = PN(\hat{\theta}_k)/PN(\hat{\theta}_k - \pi)$$

$$PN(\theta \mid \boldsymbol{\mu}, \boldsymbol{I}) = \frac{1}{2\pi} e^{-0.5||\boldsymbol{\mu}||^2} \left[1 + \frac{\boldsymbol{u}'\boldsymbol{\mu}\Phi(\boldsymbol{u}'\boldsymbol{\mu})}{\phi(\boldsymbol{u}'\boldsymbol{\mu})}\right]$$

Transformation of coefficients (continuous x)

$$\hat{\theta} = \text{atan2}\left(Y^{II},\ Y^{I}\right) = \text{atan2}\left(\beta_0^{II} + \beta_1^{II}x,\ \beta_0^{I} + \beta_1^{I}x\right)$$
$$= a_c + \arctan\left[b_c\left(x - a_x\right)\right].$$





Interpretation of coefficients, see Cremers et al. (2018), R-package Cremers (2018).

Results

Table 1: Estimated peak day (day of the year) in influenza occurrence in the circular and poisson regression models.

	mean	mode	sd	lb HPD	ub HPD
circular	38.22	38.07	0.31	37.72	38.93
poisson	37.89	37.96	0.33	37.26	38.57

Table 2: Estimated PTR for influenza in the circular and poisson regression models.

	mean	mode	sd	lb HPD	ub HPD
circular	4.42	4.43	0.04	4.36	4.49
poisson	4.38	4.37	0.04	4.31	4.46

Results peak date (age covariate)

Table 3: Circular regression model with age covariate.

	mean	mode	sd	lb HPD	ub HPD
0-16	35.65	35.76	0.54	34.55	36.71
17+	39.40	39.29	0.38	38.69	40.15

Table 4: Poisson regression model with age covariate.

	mean	mode	sd	lb HPD	ub HPD
0-16	35.34	35.30	0.57	34.17	36.40
17+	39.11	39.06	0.42	39.93	39.93

Results peak-to-trough ratio (age covariate)

Table 5: Circular regression model with age covariate.

	mean	mode	sd	lb HPD	ub HPD
0-16	4.76	4.71	0.07	4.63	4.90
17+	4.29	4.30	0.04	4.22	4.39

Table 6: Poisson regression model with age covariate.

	mean	mode	sd	lb HPD	ub HPD
0-16	4.78	4.79	0.08	4.62	4.93
17+	4.22	4.23	0.05	4.13	4.31

Other use of circular statistics

Circular approach to identify periods

$$\theta = \frac{2\pi t}{P} \mod 2\pi$$

$$\bar{R} = \frac{1}{N}\sqrt{\left(\sum_{i=1}^{N} \cos\theta_i\right)^2 + \left(\sum_{i=1}^{N} \sin\theta_i\right)^2}$$

Circular approach to identify periods

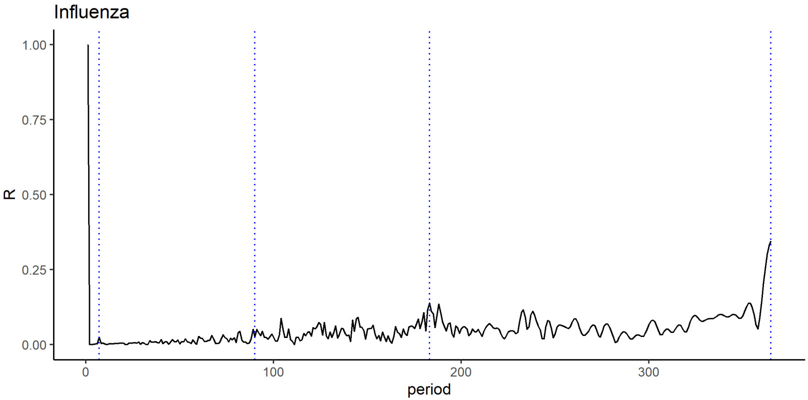


Figure 5: R-values for different periods for the Danish influenza data. The vertical blue lines are drawn at a period of 7, 90, 183 and 365 days respectively.

Poisson model with one categorical covariate

$$\log(\lambda_t) = \alpha + \beta_1 \cos(\theta_t) + \beta_2 \sin(\theta_t) + x(\beta_3 + \beta_4 \cos(\theta_t) + \beta_5 \sin(\theta_t)),$$

Peak-to-trough ratio (relative risk):

$$\exp\left(2\sqrt{\beta_1^2 + \beta_2^2 + x(\beta_4 + \beta_5)}\right)$$

Peak date:

$$\frac{\text{atan2}(\beta_2 + x\beta_5, \beta_1 + x\beta_4) * 365}{2\pi}$$

References

Cremers, J. (2018). *Bpnreg: Bayesian projected normal regression models for circular data*. Retrieved from https://CRAN.R-project.org/package=bpnreg

Cremers, J., Mulder, K., & Klugkist, I. (2018). Circular interpretation of regression coefficients. *British Journal of Mathematical and Statistical Psychology*, *71*(1), 75–95. doi:10.1111/bmsp.12108

Fisher, N. I. (1995). *Statistical analysis of circular data*. Cambridge: Cambridge University Press.