bssm: Bayesian exponential family state space models in R

Jouni Helske and Matti Vihola
University of Jyväskylä, Department of Mathematics and Statistics
jouni.helske@jyu.fi

UNIVERSITY OF JYVÄSKYLÄ

The bssm package supports general univariate state space models of form

$$p(y_t \mid x_t'\beta, Z_t\alpha_t, \phi), \quad p = \begin{cases} \text{Gaussian} \\ \text{Poisson} \\ \text{Binomial} \\ \text{Negative binomial} \end{cases}$$

$$\alpha_{t+1} = T_t\alpha_t + R_t\eta_t, \quad \eta_t \sim N(0,1), \quad \alpha_1 \sim N(a_1, P_1),$$

where $x_t'\beta$ is an optional regression component, and Z_t, T_t, R_t, β, and ϕ can depend on unknown parameter vector θ.

Main features of bssm

- User-friendly functions for Bayesian inference of exponential family time series
- Fully Bayesian inference of $p(\theta, \alpha|y)$ based on the robust adaptive Metropolis random walk algorithm
- Pseudo-marginal MCMC with delayed acceptance for non-Gaussian models using importance sampling
- Written with C++ using the Rcpp, RcppArmadillo and BH packages
- Efficient computation and plotting of prediction intervals with the ggplot2 package
- Results are easily analyzed using the coda package

References

J. Andrés Christen and Colin Fox. Markov chain Monte Carlo using an approximation. *Journal of Computational and Graphical Statistics*, 14(4):795–810, 2005.

Dirk Eddelbuettel and Romain François. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18, 2011.

Dirk Eddelbuettel and Conrad Sanderson. RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics and Data Analysis*, 71:1054–1063, 2014.

Dirk Eddelbuettel, John W. Emerson, and Michael J. Kane. *BH: Boost C++ Header Files*, 2016. R package version 1.60.0-2.

Jouni Helske and Matti Vihola. bssm: Bayesian inference of exponential family state space models in R. 2016. R package vignette.

Martyn Plummer, Nicky Best, Kate Cowles, and Karen Vines. CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6(1):7–11, 2006.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

Matti Vihola. Robust adaptive Metropolis algorithm with coerced acceptance rate. *Statistics and Computing*, 22(5):997–1008, 2012.

Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.

github.com/helske/bssm tinyurl.com/bssmVignette tinyurl.com/bssmPoster

Yearly deaths by drowning in Finland

$$p(y_t \mid \beta x_t, \mu_t, \phi) = \text{Poisson}(\phi_t \exp(\beta x_t + \mu_t))$$
$$\mu_{t+1} = \mu_t + \sigma_\eta \eta_t, \quad \eta_t \sim N(0,1)$$
$$x_t = \text{mean air temperature in June to August at year } t$$
$$\phi_t = \text{population at year } t$$

Parameter estimation and prediction

```
model <- ng_bsm(y = deaths, xreg = summer_temp, slope = FALSE,
  distribution = "poisson", sd_level = 0.1, phi = population)

out <- run_mcmc(model, n_iter = 1e5, nsim_states = 100)
summary(out$theta)
## Iterations = 50001:1e+05
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 50000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##                Mean      SD  Naive SE Time-series SE
## sd_level    0.1074 0.01952 8.730e-05      0.0002628
## summer_temp 0.1106 0.01725 7.714e-05      0.0002228
##
## 2. Quantiles for each variable:
##
##                2.5%     25%    50%    75%  97.5%
## sd_level    0.07410 0.09376 0.1058 0.1192 0.1508
## summer_temp 0.07708 0.09881 0.1106 0.1223 0.1447
pred <- predict(model, n_ahead = 10, n_iter = 1e5,
  probs = seq(0.05, 0.95, by = 0.05), nsim_states = 100)

trend <- rowMeans(exp(out$alpha[, "level", ]))

autoplot(pred, fit = trend, plot_mean = FALSE,
  obs_color = "#0038A8", interval_color = "#61C813",
  median_color = "#C85300", fit_color = "#C85300") +
  ylab("drownings per 100,000 persons in Finland") + theme_bw()
```

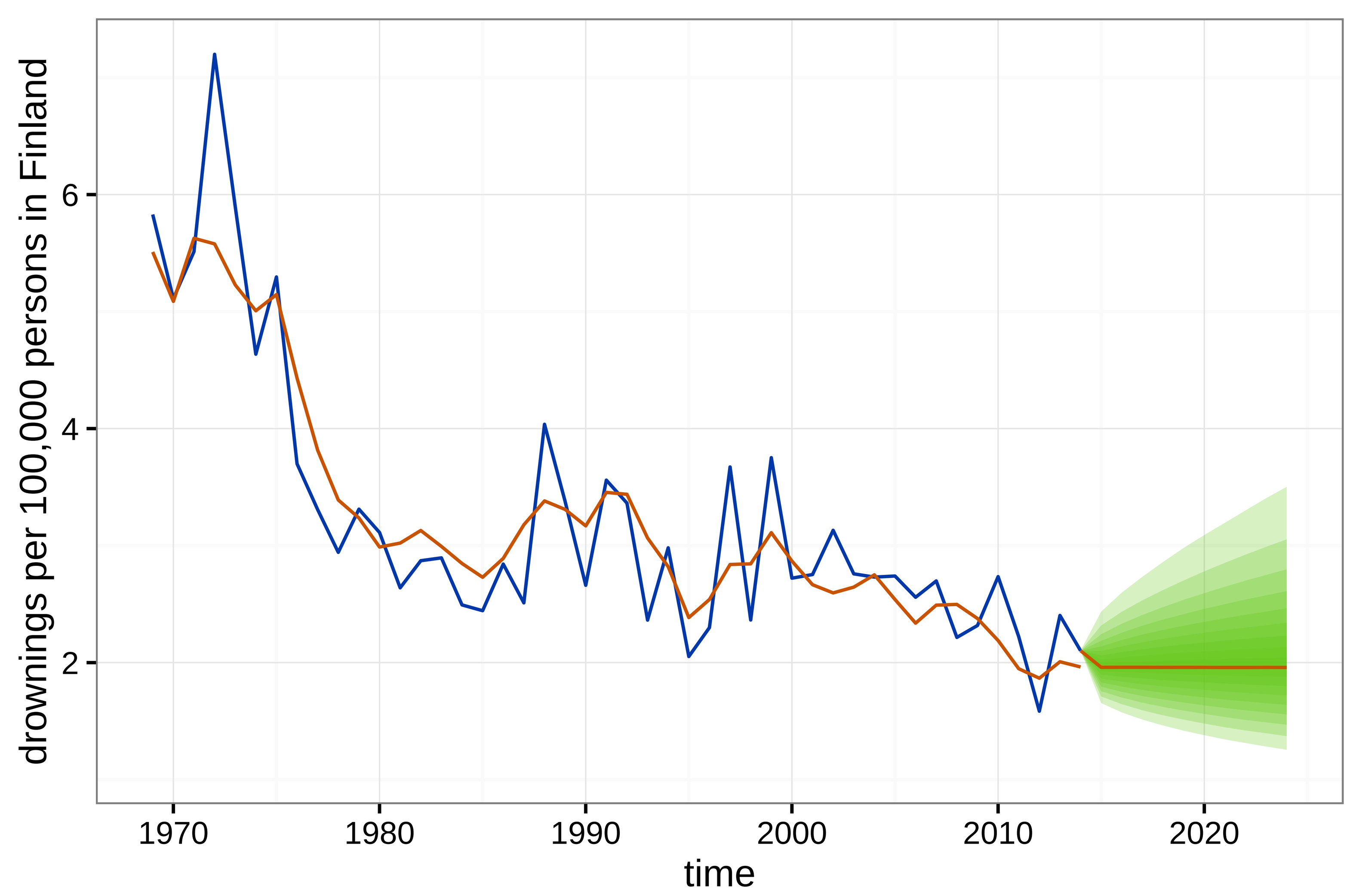

Figure 1: Observed number of deaths by drowning in Finland (blue), smoothed temperature-adjusted estimates and median count predictions for years 2015–2024 (burnt orange), together with prediction intervals (green).

Acknowledgements

This research is part of the project "Exact approximate Monte Carlo methods for complex Bayesian inference" funded by the Academy of Finland.