Bayesian model in JAGS

Tutorial 3 for transition

Zhe Zheng (Gigi)[1]

2023-01-16

[1]zhe.zheng@yale.edu

Outline

- ▶ Section 1: Background knowledge of estimating respiratory virus attributable disease burden (hospitalizations/mortality etc) with statistical models.

- ▶ Section 2: Background knowledge of using Hierarchical Bayesian regression to estimate RSV attributable hospitalizations.

- ▶ Section 3: R code for running the Hierarchical Bayesian regression to estimate RSV attributable hospitalizations in older adults by age and risk groups.

Note: In this tutorial, we will focus on using time-series data. Therefore, we will not introduce other methods to estimate disease burden such as the disease pyramid. For other methods, please check out Ginny's class *quantitative method in infectious disease epidemiology* and Dan's class *Public Health Surveillance*.

Section 1: Background knowledge of estimating respiratory virus attributable disease burden with statistical models.

There are 5 main statistical methods to estimate the hospitalization/mortality attributable to respiratory virus infection. In Dan's class, Public Health Surveillance, he teaches all four methods in details. Here, we will only give a brief introduction to each of the method and provide the link to the initial publications.

1. **Serfling regression**
2. **Periseason differences**
3. **Poisson regression with log link**
4. **Negative binomial regression with identity link**
5. **Box-Jenkins transfer function (ARIMA model)**

Brief introduction to serfling regression and periseason differences

1.Serfling regression

This type of model identifies the epidemic season and establish a epidemic threshold (seasonal baseline) using historical data. After predicting the expected diseases baseline, observed diseases above the baseline (epidemic threshold) during the epidemic season will be attributable to the viral infection. The basic Serfling regression is given by:

$Y_t = \mu + bt + \sum \alpha_i cos\theta + \sum \beta_i sin\theta$ where Y_t is the expected hospitalizations at time t. μ is the baseline. bt captures time trend. $\sum \alpha_i \cos \theta + \sum \beta_i \sin \theta$ are for the seasonal variations. Serfling regression is a linear model initially using ordinary least square to fit. Click to see an R code example

2.Periseason differences

This method calculate the hospitalizations attributable to influenza based on the differences in the hospitalization rates when respiratory virus was circulating and the hospitalization rates when there was no respiratory virus circulation in the community.

Brief introduction to poisson regression with log link

3.**Poisson regression with log link**

Poisson regression is used to model count variables. Therefore, it can be used to model the number of hospitalizations over time. The logarithm of the expected value of the count variable is modeled as a linear combination of the predictor variables. For example, in Thompson's paper[2], the poisson regression with log link is given by:

$$Y = \alpha \exp^{(\beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 \sin(2\pi/52) + \beta_4 \cos(2\pi/52) + \beta_5 Flu_A(H1N1) + \beta_6 Flu_A(H3N2) + \beta_7 Flu_B + \beta_8 RSV)}$$

In this example, Y is the observed number of deaths/hospitalization in a particular week, α is the population size as an offset term, β_0 is the baseline hospitalization incidence, $\beta_1 t + \beta_2 t^2$ represents the long-term trend of hospitalizations, $\sin(2\pi/52) + \beta_4 \cos(2\pi/52)$ captures the annual seasonal variations of hospitalizations, the rest are the pathogens that can contribute to the increase of the observed hospitalizations.

---

[2]Thompson WW, Shay DK, Weintraub E, et al. Mortality Associated With Influenza and Respiratory Syncytial Virus in the United States. JAMA. 2003;289(2):179–186. doi:10.1001/jama.289.2.179

Brief introduction to negative binomial regression with identity link

4.**Negative binomial regression with identity link**

Negative binomial regression is also used to model count variables. It is a generalization of the poisson regression, in which the variance is assumed to be equal to the mean. The new approach use an identity link instead of an log link because log link corresponds to a multiplicative relationship among covariates (including baseline hospitalizations, seasonality, and several respiratory viruses) while identity link assumes additive effects. Apparently, additive effects are more realistic than multiplicative effects. The negative binomial regression with identity link is given by:

$$Y \sim NB(p, r), p = \frac{r}{r + \lambda}$$

$$\lambda = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 \sin(2\pi/52) + \beta_4 \cos(2\pi/52) + ... + \beta_8 RSV$$

where Y is the number of observed failures (diseases), p is the probability of success (survival/not infected), r is the number of successes (survival/not infected) and λ is the expected value of failures (diseases). All other parameters remain the same as what we introduced in Poisson regression.

Brief introduction to ARIMA model

5.**Autoregressive integrated moving average (ARIMA) model**
ARIMA model can account for the autorcorrelation of the time-series and also the time-delayed associations. For infectious diseases, the number of the infected individuals in the next time point is correlated with the number of the infected individuals currently. This phenomenon is described as autocorrelated. The influenza infection reported this week may contribute to hospitalizations next week, which means the time-delayed associations. For example, in Gilca's paper[3], ARIMA model is given by:

$$Y_t = \alpha_0 + \alpha_1 Temp_t + \alpha_2 Holiday_t + \beta_1 Flu_t + \sum_{i=0}^{2} \omega_i RSV_{t-i}$$

Y_t is the observed number of hospitalizations in a particular week t, α_0 is the baseline hospitalization incidence, $\alpha_1 Temp_t + \alpha_2 Holiday_t$ are the hospitalizations attributable to seasonal changes and holiday gathering, $\beta_1 Flu_t$ is the flu attributable hospitalizations, and $\sum_{i=0}^{2} \omega_i RSV_{t-i}$ is the RSV attributable hospitalizations. **This example suggested that not only RSV infections this week but also RSV infections reported 1 week and 2 weeks ago associates with hospitalizations reported this week.**

[3] Rodica Gilca, Gaston De Serres, Danuta Skowronski, Guy Boivin, David L. Buckeridge, The Need for Validation of Statistical Methods for Estimating Respiratory Virus–Attributable Hospitalization, American Journal of Epidemiology, Volume 170, Issue 7, 1 October 2009, Pages 925–936, https://doi.org/10.1093/aje/kwp195

Initial publications of the method

1. **Serfling regression**

- ▶ Methods for Current Statistical Analysis of Excess Pneumonia-influenza Deaths
- ▶ The impact of influenza epidemics on mortality: introducing a severity index
- ▶ Impact of Influenza Vaccination on Seasonal Mortality in the US Elderly Population

2. **Periseason differences**

- ▶ The effect of influenza on hospitalizations, outpatient visits, and courses of antibiotics in children
- ▶ Respiratory illness associated with influenza and respiratory syncytial virus infection
- ▶ Influenza and the Rates of Hospitalization for Respiratory Disease among Infants and Young Children
- ▶ Impact of influenza and respiratory syncytial virus on mortality in England and Wales from January 1975 to December 1990

Initial publications of the method (continued)

3. **Poisson regression with log link**

▶ Mortality Associated With Influenza and Respiratory Syncytial Virus in the United States

▶ The Association of Respiratory Syncytial Virus Infection and Influenza with Emergency Admissions for Respiratory Disease in London: An Analysis of Routine Surveillance Data

4. **Negative binomial regression with identity link**

▶ Modelling the unidentified mortality burden from thirteen infectious pathogenic microorganisms in infants

▶ Hospitalization Attributable to Influenza and Other Viral Respiratory Illnesses in Canadian Children

5. **Autoregressive integrated moving average (ARIMA) model**

▶ Community influenza outbreaks and emergency department ambulance diversion

▶ Time-Series Analysis of the Relation between Influenza Virus and Hospital Admissions of the Elderly in Ontario, Canada, for Pneumonia, Chronic Lung Disease, and Congestive Heart Failure

Other useful references

- The Need for Validation of Statistical Methods for Estimating Respiratory Virus–Attributable Hospitalization

Section 2: Background knowledge of using Hierarchical Bayesian regression to estimate RSV attributable hospitalizations

Background knowledge of Hierarchical Bayesian regression

Hierarchical Bayesian regression can make population inference based on group and individual observations. The benefits of using a Hierarchical Bayesian structure is that it can shrink the uncertainty of inference while allow between-group variability.

Background knowledge of Hierarchical Bayesian regression: example

For example, if we would like to estimate the baseline hospitalizations β_{0jk} in age j and socioeconomic group k, we can treat this parameter as (1) independent, meaning that the baseline hospitalizations are different in each group and we will need to estimate 27 parameters $\beta_{0j=1,k=1}$, $\beta_{0j=1,k=2} \ldots \beta_{0j=9,k=3}$, (2) the same, meaning that the baseline hospitalizations are the same in all age and socioeconomic group and we will only need to estimate one parameter β_0, and (3) random draws from the a normal distribution, meaning that baseline hospitalizations are normally distributed in the whole population and baseline hospitalization of each subpopulation jk is an realization of the distribution (a point in the distribution). In this case, we will estimate the mean and variations of the normal distribution and then get the subpopulation estimates.

Further readings on Hierarchical Bayesian regression

- ▶ Bayesian hierarchical modeling by wikipedia
- ▶ Chapter 10 Bayesian Hierarchical Modeling by Drs. Jim Albert and Jingchen Hu
- ▶ Hierarchical Bayesian Modeling by Dr. Angie Wolfgang