PharmaSUG 2019 - Paper PO-324

Bayesian Methods for Treatment Design in Rare Diseases

Xuan(Kate) Sun and Ruohan Wang, Ultragenyx Pharmaceutical Inc. Novato, CA

Abstract

During the last thirty years, Bayesian methods have been developed rapidly with the explosively growth of high-speed computers. Bayesian Methods are quite useful and easy to interpret with the graphical displays of treatment effects by modeling. In rare diseases, we could not have the access to large samples for a parametric data analysis. Under this circumstance, Bayesian Methods might be a more flexible framework for rare diseases treatment modeling. In this article, the challenge of rare diseases is introduced in section I. The Bayesian Method and how to use Bayesian Method in rare diseases treatment prediction is introduced in section II. And Bayesian Method examples are shown in section IV.

I. Introduction

A. Challenge of Rare Diseases

The challenges of rare diseases are much greater than other diseases. The solutions for rare diseases are not distinguished as other diseases since there might be few experts in rare diseases. The limitation will cause diagnosis and treatment delay while it may not have corresponding medication for those rare diseases.

There are approximately 7,000 rare diseases and disorders. 30 million people, which is 10% of total population in the United Stated, are living with rare diseases. Such diseases usually have a genetic basis, 80% often affecting patients early in childhood, and are frequently progressive, disabling and life threatening in nature. These characteristics can have a devastating psychological impact on families of children suffering from these diseases.

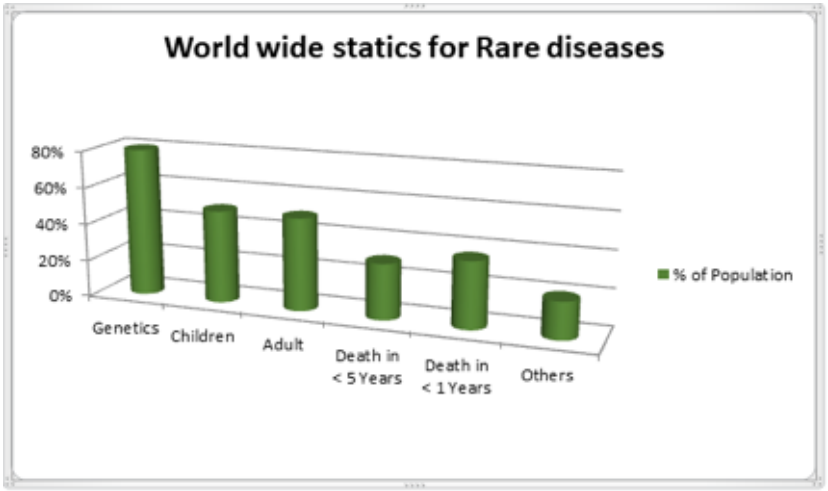

Figure 1: World wide statics for Rare diseases

II. Bayesian Methods

When applying extremely complicated Bayesian Models, various problems have been discovered for those data we could not apply a precise distribution on. We can use Bayesian Methods to solve those kinds of problems and find the accurate convergence.

Bayesian Methods can be very powerful when dealing with small sample sized data sets. In rare disease cases, Bayesian Methods could incorporate with historical data, which may come from previous studies (if it is a rollover study) used as a prior distribution, and build an appropriate model for explanation of the treatment results. Under suitable conditions, Bayes factor rates are of the same order as those that would be obtained under the correct model, but we point out a potential loss of sensitivity to detect truly active covariates.

Bayesian Method considers that parameters are random and the data is fixed. It can be extended in machine learning process. If we would like to train the machine some language or knowledge to learn, we always use known data set as the training data. In this case, the dataset is already fixed and it is not random anymore.

The prior distribution in Bayesian Method expresses the uncertainty about treatment effect without gathering the data. For rollover data sets, we can use previous data set to derive the prior distribution for current study. Otherwise, the prior distribution needs modeling assumption based on similar previous studies or concurrent data for those diseases.

The posterior distribution for the treatment effect (model parameters) given the data is found by combining the prior distribution with the likelihood for the parameters given the data.

This is done using Bayes Rule:

$$P(\text{parameters} \mid \text{data}) = \frac{P(\text{parameters})\, P(\text{data} \mid \text{parameters})}{P(\text{data})} \tag{1}$$

The denominator is just the required normalizing constant, and can often be filled in at the end, if necessary. So as a proportionality, we can write

$$P(\text{parameters} \mid \text{data}) \propto P(\text{parameters})\, P(\text{data} \mid \text{parameters}) \tag{2}$$

which can be written schematically as

$$\text{Posterior} \quad \propto \quad \text{Prior} \times \text{Likelihood} \tag{3}$$

We make predictions by integrating with respect to the posterior:

$$P(\text{new data} \mid \text{data}) = \int_{\text{parameters}} P(\text{new data} \mid \text{parameters})\, P(\text{parameters} \mid \text{data}) \tag{4}$$

III. Bayesian Analysis in SAS

SAS provides two kinds of Bayesian analysis[1]:

- built-in Bayesian analysis in several modeling procedures like GENMOD, FMM and PHREG.
- the MCMC procedure for general-purpose modeling.

The built-in Bayesian procedures are ideal for data analysts who just start to use Bayesian methods, and they suffice for many analysis objectives. Simply adding the BAYES statement generates Bayesian analyses without the need to program priors and likelihoods for the GENMOD, PHREG, LIFEREG, and FMM procedures. Thus, you can obtain Bayesian results for the many statistical models like the following

- linear regression
- logistic regression
- Poisson regression
- Cox proportional models

- finite mixture models

The built-in Bayesian procedures apply the appropriate Markov chain Monte Carlo sampling technique. They also provide default prior distributions depending on what models are specified. You can choose from other available priors by using the CPRIOR= option (for coefficient parameters) and SCALEPRIOR= option (for scale parameters). Other options allow you to choose the numbers of burn-ins, the number of iterations, and so on[1].

Although the built-in Bayesian procedures provide Bayesian analyses for many standard techniques, they only go so far. You might want to include priors that are not available, or you might want to perform a Bayesian analysis for a model that isnt offered. For example, the GENMOD procedure doesnt presently offer Bayesian analysis for the proportional odds model. However, you can fit nearly any model that you want, for any prior and likelihood you can program, with the MCMC procedure.

The following sections describe how to use the built-in Bayesian procedures to perform Bayesian analyses.

IV. Bayesian Method Example: Rare Diseases (DATASET size is about 30)

We applied Bayesian methods on an example data set about a rare disease. The data set includes the values of 25 instances and 10 attributes. Because in rare diseases, there may not be so many patients or enrolled subjects. We use small sample size to explain the variables and we can expand the theory to other rare diseases. The response is whether the patient is recovered from the disease. Explanatory variables include the age, gender and the duration of the disease before the treatment began.

Logistic regression is considered for this data set.

$$y = logit(\beta_0 + \beta_1 \cdot age + \beta_2 \cdot bp + ...) \tag{5}$$

A normal prior distribution with large variance is used here as a noninformative prior distribution on all the regression coefficients

$$\pi \sim normal(0, var = 1e6) \tag{6}$$

We fit this logistic regression model into the built-in PROC GENMOD. The SAS code is shown below.

```
proc stdize data=newbay1 reponly method=mean out=baymean;
        var mcv_lc platlc cesd base_cd4 partsc age_sc c342le;
run;
data baye1;
set baymean;
array _nums {*} _numeric_;
do i = 1 to dim(_nums);
  _nums{i} = round(_nums{i},.01);
end;
drop i;
run;

ods graphics on;
proc genmod data=baye1;
model y=mcv_lc platlc cesd base_cd4 partsc age_sc c342le smokbh status
racesc/dist=normal;
bayes seed=1 outpost=post;
run;
```

Figure 2 displays the posterior summaries and Figure 3 displays the posterior intervals.

The intervals suggest that treatment is very influential. Some parameters do not appear to be that important. However, all terms are kept in the model.

Bayesian Analysis

Posterior Summaries						
				Percentiles		
Parameter	N	Mean	Standard Deviation	25%	50%	75%
Intercept	10000	1.3196	0.8253	0.7939	1.3304	1.8487
MCV_LC	10000	0.0160	0.0121	0.00834	0.0160	0.0234
PLATLC	10000	-0.00089	0.00160	-0.00192	-0.00088	0.000138
CESD	10000	-0.00669	0.00853	-0.0123	-0.00690	-0.00119
Base_CD4	10000	0.000216	0.0157	-0.00997	0.000291	0.0104
PARTSC	10000	-0.0723	0.2339	-0.2219	-0.0718	0.0784
AGE_SC	10000	-0.0226	0.0175	-0.0339	-0.0227	-0.0114
C342LE	10000	0.000981	0.0154	-0.00910	0.00104	0.0108
SMOKBH	10000	-0.0156	0.2527	-0.1786	-0.0178	0.1459
Dispersion	10000	0.3008	0.1210	0.2181	0.2759	0.3530

Figure 2: Posterior Summaries

Posterior Intervals					
Parameter	Alpha	Equal-Tail Interval		HPD Interval	
Intercept	0.050	-0.2921	2.9772	-0.3389	2.9204
MCV_LC	0.050	-0.00787	0.0402	-0.00788	0.0402
PLATLC	0.050	-0.00407	0.00233	-0.00399	0.00239
CESD	0.050	-0.0233	0.0104	-0.0230	0.0106
Base_CD4	0.050	-0.0312	0.0306	-0.0309	0.0307
PARTSC	0.050	-0.5387	0.3968	-0.5396	0.3937
AGE_SC	0.050	-0.0578	0.0116	-0.0572	0.0120
C342LE	0.050	-0.0290	0.0311	-0.0283	0.0316
SMOKBH	0.050	-0.5187	0.4844	-0.5003	0.4964
Dispersion	0.050	0.1471	0.6053	0.1270	0.5411

Figure 3: Posterior Intervals

V. Conclusions

From the above results, we can conclude that Bayesian method can be a powerful framework for rare diseases treatment modeling and analysis, especially when the data size is very small, in which case most other statistical methods like machine learning are not that useful.

References

[1] Maura Stokes, Fang Chen, and Funda Gunes. An Introduction to Bayesian Analysis with SAS/STAT Software, *Paper SAS400-2014*.

Disclaimer: This paper is using simulation data, it represents the options of the authors and not meant to represent the opinions from the company.

Contact Information

Your comments and questions are valued and encouraged. Contact the author:

Name: Xuan(Kate) Sun
Enterprise: Untragenyx Pharmaceutical Inc.
Address: 60 Leveroni Ct, Novato, CA 94949
Work Phone: 415-483-8974
Email: xsun@ultragenyx.com

Name: Ruohan Wang
Enterprise: Untragenyx Pharmaceutical Inc.
Address: 60 Leveroni Ct, Novato, CA 94949
Work Phone: 415-483-8810