Something related to Bayesian data analysis

Abstract

1. Methods

The cut-off of biomarker AI and its predictive values were determined using a Bayesian model described by Vradi et al. (2018). The model can be expressed as:

$$Y|X \sim Bernoulli(p)$$

$$p(x) = P(Y = 1|X = x) = \begin{cases} p_1 = P(Y = 1|X \leq cp) = 1 - NPV, & \text{if } x \leq cp \\ p_2 = P(Y = 1|X > cp) = PPV, & \text{if } x > cp \end{cases}$$

The binary response variable Y took the value 1 when a patient had NPSLE and 0 otherwise. X was the continuous measurement of the biomarker AI. In this study, the positive predictive value (PPV) of the cutoff cp was expected to $\geq 70\%$, while the negative predictive value (NPV) was $\geq 50\%$. Thus, prior distributions of 1-NPV and PPV were chosen as follows:

$$p_1 \sim Uniform(0, 0.5)$$

$$p_2 \sim Uniform(0.7, 1)$$

The Bayesian analysis was carried out in SAS Studio (SAS Institute Inc., 2015) through PROC MCMC, which uses Markov chain Monte Carlo (MCMC) algorithm. The model was run for a total of 2×10^6 iterations. The first 10^6 iterations were discarded as burn-in. The remaining were kept 1 in 10 samples (thinning) to reduce autocorrelation. The MCMC representativeness and accuracy were assessed by trace plot, sample autocorrelation, effective sample size (ESS), and Monte Carlo standard error (MCSE).

2. Results

- Diagnostic results
  - Trace plot of all parameters mixed well => showed no potential problem with convergence.
  - Sample autocorrelation of all parameters dropped quickly to zero with increasing lag.

- ESS of all parameters were > 17500. (recommended: > 10000)

- MCSE of all parameters were < 0.0059.
  => Conclusion: MCMC chains were stable and accurate.

- Posterior distribution summary

- report: the mean and 95% highest density interval (HDI) of cp, p1, p2. (in Table **Posterior Summaries and Intervals** of the file **Results_cutoff_VEleni_final.pdf**)

Table 1: this table is used to compare mean and median value when necessary

Parameter	Median	Mean
cp	5.1522	5.259981
p1	0.4471	0.439842
p2	0.8800	0.870422

- 95% HDI of cutoff value was in range 3.6796-7.1657 but 89% HDI of cutoff was in this small range 4.2134-5.9498.

3. Caption of Figure:

Fig1. Summary of posterior distribution of the cutoff and its predictive value (1-NPV and PPV). The vertical red dotted lines denote the median of the distribution.

4. Supplementary

4.1. Diagnostic results

4.2. SAS MCMC code

```
PROC import datafile="/path/to/data.csv"
DBMS=csv out=Data replace;
RUN;


PROC MCMC
    data=Data outpost=Dataoutput
        nbi=1000000
        nmc=1000000
        thin=10
        seed=1
        diag=all
```

```
        monitor=(p1 p2 cp I w);
PARMS cp1 cp2 p1 p2 w I;
prior cp1 ~ uniform(1,10);
prior cp2 ~ normal(5,sd=1);
hyperprior I ~ beta(1,1);
prior w ~ binary(I);
cp = w*cp1 + (1-w)*cp2;
prior p2 ~ uniform(0.7, 1);
prior p1 ~ uniform(0, 0.5);
p = (AI<=cp)*p1 + (AI>cp)*p2;
model Y ~ binary(p);
RUN;
```

References

SAS Institute Inc., 2015. SAS OnDemand for Academics: User's Guide 524.

Vradi, E., Jaki, T., Vonk, R., Brannath, W., 2018. A Bayesian model to estimate the cutoff and the clinical utility of a biomarker assay. Statistical Methods in Medical Research 1–19. doi:10.1177/0962280218784778